GENERALIZED GRADIENT METHODS
FOR SOLVING LOCALLY LIPSCHITZ
FEASIBILITY PROBLEMS

Dan Butnariu

TR #277
December 1990
GENERALIZED GRADIENT METHODS FOR SOLVING
LOCALLY LIPSCHITZ FEASIBILITY PROBLEMS

DAN BUTNARIU

Department of Mathematics and Computer Science
Haifa University, 31999 Haifa, Israel

ABSTRACT
In this paper we study the behavior of a class of iterative
algorithms for solving feasibility problems, that is finite
systems of inequalities $f_i(x) \leq 0, \ (i \in I)$, where each $f_i$ is a
locally Lipschitz functional on a Hilbert space $X$. We show that,
under quite mild conditions, the algorithms studied in this note,
if converge, then they approximate a solution of the feasibility
given problem, provided that the feasibility problem is
consistent. We prove several convergence criteria showing that,
when the envelope of the functionals $f_i, \ (i \in I)$, is sufficiently
"regular", then the algorithms converge. The class of algorithms
studied in this note contains, as special cases, many of the
subgradient and projection methods of solving convex feasibility
problems discussed in the literature.

December 1990

1 Work of Dan Butnariu was done while visiting the Department of
Mathematics of the University of Texas at Arlington, Arlington,
Texas 76012, USA.
1. Introduction

A feasibility problem is a problem of computing solutions of a system of inequalities

\[ f_i(x) \leq 0, \quad (i \in I), \]

where \( I \) is a finite set and, for each \( i \in I \), \( f_i \) is a continuous real functional on a Hilbert space \( X \). The feasibility problem (1) is called locally Lipschitz (respectively convex) if the functionals \( f_i \) \( (i \in I) \), involved in it are locally Lipschitz (respectively convex). Feasibility problems appear in applied mathematics in fields like Optimization (see, for instance, [5], [25], [40]), Image Reconstruction From Projections (cf. [41], [17], [11]), Game Theory (cf. [21], [22], [6]), etc. In practical cases feasibility problems may appear as intersection problems, that is problems of computing points in the intersection \( Q \) of a finite family of closed subsets \( \{Q_i \mid i \in I\} \) of \( X \). It is obvious that an intersection problem can be rewritten as a feasibility problem (1) with \( f_i(x) = d_{Q_i}(x), \quad (x \in X) \) where \( d_{Q_i}(x) \) denotes the distance of \( x \) to \( Q_i \), i.e.

\[ d_{Q_i}(x) := \inf \{ \| x - z \| \mid z \in Q_i \}, \quad (i \in I). \] (2)

Also, it is clear that a feasibility problem (1) can be reduced to an intersection problem with \( Q_i := C_i \), where

\[ C_i := \{ x \in X \mid f_i(x) \leq 0 \}, \quad (i \in I). \] (3)

In that follows, the feasibility problem

\[ d_{C_i}(x) \leq 0, \quad (i \in I), \] (4)

will be called the dual problem of (1) and the feasibility problem

\[ d_{Q_i}(x) \leq 0, \quad (i \in I), \] (5)

associated to a given intersection problem \( Q := \bigcap_{i \in I} Q_i \) will be called the normal form of the intersection problem.
In this note we study a class of algorithms of solving feasibility problems either by approaching the problems in their primal form (1) or via their duals (4). The algorithms we present are generated by what we call the generalized gradient method (GGM) and are essentially based on computing generalized gradients in the sense of Clarke [13]. The GGM is defined in Section 2. The main requirement for the existence of GGM generated algorithms for a feasibility problem (1) is that the functionals \( f_i \), \( i \in I \), are locally Lipschitz. This condition is always satisfied by the functionals involved in the dual (4) of the feasibility problem (1) provided that the problem is consistent. If the functionals \( f_i \) are continuous, then (1) and (4) have the same set of solutions. Hence, algorithms for solving locally Lipschitz feasibility problems are, implicitly, methods of solving feasibility problems involving continuous functionals via their duals. In Section 3 we study the behavior of GGM generated algorithms applied to feasibility problems (1) in primal form and we show (see Theorem 1) that, under quite mild conditions, such algorithms, if converge, then they approximate solutions of the given feasibility problem, provided that such solutions exist. In Sections 4 and 5 we prove convergence criteria for different classes of GGM generated algorithms. The convergence criterion proved in Section 4 applies to locally Lipschitz feasibility problems (1) when the functionals \( f_i \), \( i \in I \), have bounded (Clarke) generalized gradient multifunctions. The boundedness condition for the generalized gradients is restrictive but it is automatically satisfied by the functionals involved in the dual (4) of the problem (1) provided that \( X \) is finite dimensional. In this way we obtain convergence criteria for a class of iterative algorithms
for solving feasibility problems which we call normal methods because they are essentially based on computing normals of various subsets of $X$. The convergence criterion proved in Section 5 applies to intersection problems in their normal form (5) under the assumption that the sets $Q_i$ are convex and closed in $\mathbb{R}^n$. Thus, it implicitly provides tools of solving convex feasibility problems via their duals.

From an analytic point of view the GGM generated algorithms studied in this note are "descent" procedures simultaneously applied to all functionals involved in the given feasibility problem. Methods of this nature can be traced back to Fourier [23] who applied them to solve intersection problems (of polyhedral sets) in their normal forms. Cauchy [8] employed a method of the same kind in a frontal approach of a feasibility problem in its primal form.

A well-known class of algorithms for solving convex intersection problems in normal form and, implicitly, convex feasibility problems via their duals, consists of the so-called "projection methods". The algorithms in this class are based on computing, at each step, orthogonal projections on the sets $Q_i$ and $C_i$ respectively. The projection methods are particular GGM generated algorithms (see Section 2). The first interested in them was, probably, Fourier [24]. In the early 1930s Johannes von Neumann discovered the "alternating projection method" for solving the intersection problem for two closed linear subspaces of a Hilbert space (cf. [34]). Since von Neumann's result wasn't published till 1949, Nakano [33] rediscovered it independently. This method was further generalized by Halperin [29] to finite families of closed linear subspaces of a Hilbert space. Projection methods were used...
in the late 1930s by Kaczmarz [30] and Cimmino [12] in order to solve intersection problems of half-spaces and irrespective affine subsets of \( \mathbb{R}^n \) via their normal forms. Their results were further expanded by Agmon [1] and Motzkin and Scheönberg [32] and, under the impact of their works, a large class of "projection methods", namely the so called "relaxation methods" for solving intersection problems of half-spaces of \( \mathbb{R}^n \) and, equivalently, feasibility problems with affine left-hand sides, evolved (see Censor [9] for a survey of classical results in this area and see Goffin [24], Deutsch [18] and Kayalar and Weinert [31] for studies of the computational efficiency of relaxation methods). From the studies concerning the relaxation procedures for intersection problems of affine subsets in \( \mathbb{R}^n \) evolved yet a more general class of projection methods, namely the block-iterative projection methods for solving intersection problems for convex closed subsets in finite dimensional Hilbert spaces (see Iusem and De Pierro [28], Aharoni and Censor [3], and Butnariu and Censor [6] and the references therein). In Section 5 we use the general results on the GGM generated algorithms in order to prove a convergence criterion of a new class of projection algorithms for convex intersection and feasibility problems. The algorithms in this class tend to reduce to a minimum the number of projections needed at each computational step without affecting the rate of convergence of the global procedure. This is of interest since it may happen that the task of determining projections on the sets \( Q_i \) (or \( C_i \)) to be computationally complicated when these sets are not affine.

The projection algorithms approach convex intersection problems in normal form and convex feasibility problems in dual
form. However, feasibility problems can be approached in their primal form (1) whenever the functionals $f_i$, ($i \in I$), are satisfying sufficient "smoothness" conditions (see Section 3). A way of solving feasibility problems in primal form when the functionals $f_i$ are differentiable is suggested by Cauchy's [8] results on systems of equations involving differentiable functions. Techniques of essentially the same nature as those used in the 1847 paper of Cauchy were further extended for solving "convex feasibility problems", that is feasibility problems which involve convex functionals $f_i$ only, by subgradient methods (see Shor [39], Eremin [19], Polyak [36], Goffin [25], Censor and Lent [10] and the references therein). In Optimization Theory techniques of the same nature were employed for the treatment of convex and nonconvex optimization problems (see, for instance, Polyak [37], Goldstein [26]). From a formal point of view, the GGM algorithms presented in this note can be seen as a further extension of the subgradient method to a class of feasibility problems satisfying weakened convexity requirements.
2. The Generalized Gradient Method

In this section we present the generalized gradient method for the primal problem (1) and we show how it relates to the intersection problem in normal form and to the dual form of (1). To this end we first recall some definitions and basic properties of the functions involved in (1), (4) and (5).

Let $X$ be a Hilbert space and consider the feasibility problem (1). This problem is called consistent if it has at least one solution, that is if the solution set $C := \bigcap_{i \in I} C_i$ of (1) is nonempty.

A function $h: X \rightarrow \mathbb{R}$ is called locally Lipschitz if for each $x \in X$ there exists a neighborhood $U$ of $x$ and a positive real number $K = K(x)$ such that

$$|h(z) - h(y)| \leq K\|z - y\|$$

for any $z, y \in U$. If there exists a positive real number $K$ such that (5) holds for all $z, y \in X$, then $h$ is said to be globally Lipschitz. If $h$ is convex and locally bounded, then $h$ is locally Lipschitz (cf. Roberts and Varberg [37]). If $X$ is the Euclidean space $\mathbb{R}^n$ and $h$ is convex, then $h$ is locally Lipschitz (cf. Rockafellar [38, Section 10]).

Let $h: X \rightarrow \mathbb{R}$ be a locally Lipschitz functional and let $x$ and $v$ be vectors in $X$. We denote, as usual, $h^o(x, v)$ Clarke's generalized derivative of $h$ at $x$ in the direction $v$ (see Clarke [13, Chap. 2]), that is

$$h^o(x, v) := \lim_{Y \rightarrow x, Y \neq x, \lambda \rightarrow 0} \frac{h(Y + \lambda v) - h(Y)}{\lambda}.$$ 

We recall (cf. Clarke [13]) that, $h^o(x, v)$ is finite and it is locally Lipschitz, subadditive and positively homogeneous (hence convex) with respect to $v$. $h^o(x, \cdot)$ is the support function of the
set
\[ \partial h(x) := \{ \xi \in X^* \mid \langle \xi, x \rangle \leq h^0(x, u), (u \in X) \} \] (7)
called generalized gradient of \( h \) at \( x \). The generalized gradient \( \partial h(x) \) is nonempty convex and weak* compact. The functional \( h \) is said to be regular at \( x \) if, for each \( u \in X \), the generalized derivative \( h^0(x, u) \) coincides to the directional derivative of \( h \) in the direction \( u \), that is if
\[ h^0(x, u) = \lim_{t \to 0} \frac{h(x + tu) - h(x)}{t}, \quad (u \in X). \] (8)
Convex functionals as well as continuously differentiable functionals on \( X \) are regular. If \( h \) is convex, then its generalized gradient multifunction is exactly its subgradient multifunction (cf. Clarke [13, Prop. 2.2.7]), i.e. if \( h \) is convex we have
\[ \partial h(x) = \{ \xi \in X^* \mid h(y) - h(x) \geq \langle \xi, y - x \rangle, (y \in X) \}, \quad (x \in X). \] (9)
For a continuously differentiable functional \( h \) the set \( \partial h(x) \) is an one-element set consisting of the gradient of \( h \) at \( x \), i.e. \( \partial h(x) = \{ \nabla h(x) \} \).

Let \( S \) be a nonempty subset of \( X \). Then the distance function \( d_S(x) \) is globally Lipschitz of Lipschitz constant 1 (cf. [13, p. 50]). Therefore, its generalized derivative \( d_S^0(x, u) \) is always finite and satisfies the properties mentioned above. Note that, if \( S \) is convex, then \( d_S \) is convex too. \( d_S \) is not differentiable on \( X \) (even if \( S \) is convex). However, if \( S \) is convex and closed, then the function \( d_S^2 : X \longrightarrow \mathbb{R} \) defined by
\[ d_S^2(x) := \inf \{ \| x - z \|^2 \mid z \in S \} \] (10)
is continuously differentiable on \( X \) (cf. [4, p. 24]) and its gradient is given by
\[ \nabla d_S^2(x) = 2(x - P_S(x)), \quad (x \in X), \] (11)
where \( P_S(x) \) denotes the (orthogonal) projection of \( x \) on \( S \), that is
the closest point to $x$ in $S$.

In that follows we adopt the usual convention of denoting by the same symbol an element $\xi$ of the (continuous) dual $X^*$ of $X$ and the representation of $\xi$ in $X$ by Riesz's Theorem (see, for instance, [14, Theorem 4.2]). Therefore, $i^* \xi \in X^*$ and if $v \in X$, we may write $\langle \xi, v \rangle$ instead of $\xi(v)$. Also, we may say that $v \in \partial h(x)$ meaning that the functional $y \mapsto \langle v, y \rangle : X \to \mathbb{R}$ belongs to $\partial h(x)$.

With these facts in mind we define the generalized gradient methods (GGM). Suppose that, for each $i \in I$, the functional $f_i$ involved in (1) is locally Lipschitz. Then, as noted above, $\partial f_i(x) \neq \emptyset$, $(x \in X, i \in I)$. The generalized gradient method consists of the following: Choose an initial point $x^0 \in X$ and do:

$$x^{k+1} := x^k - \lambda_k \cdot \sum_{i \in I} w_k(i) \xi_i^k, \quad (k \in \mathbb{N}),$$

(12)

where, for each $k \in \mathbb{N}$, $\lambda_k \in \mathbb{R}$ is called relaxation parameter, $w_k : I \to \mathbb{R}^+$ is a weight function (i.e. it has $\sum_{i \in I} w_k(i) = 1$) and $\xi_i^k \in \partial f_i(x^k)$, $(i \in I)$.

The GGM is a general scheme of generating "algorithms". An algorithms generated by the GGM is a sequence $(x^k \mid k \in \mathbb{N})$ obtained according to (12) for specific choices of the relaxation parameters $\lambda_k$ and of the weight functions $w_k$. We say that an algorithm $(x^k \mid k \in \mathbb{N})$ converges if the sequence $(x^k \mid k \in \mathbb{N})$ converges no matter how the initial point $x^0$ and the gradients $\xi_i^k$ are chosen in $X$ and in $\partial f_i(x^k)$ respectively. We will show latter that a large class of GGM generated algorithms are convergent to solutions of the given feasibility problem.

The GGM generated algorithms are of special practical interest when $X = \mathbb{R}^n$. In this case the computation of the generalized gradients involved in GGM generated algorithms can
sometimes be reduced to a standard computation of partial derivatives since, if the functionals $f_i$, ($i \in I$), are continuously differentiable at $x^k$, then the problem of computing the generalized gradients $\xi_i^k$ is a standard problem of computing partial derivatives. If $X = \mathbb{R}^n$, then, according to Rademacher's Theorem, locally Lipschitz functionals on $X$ are (Lebesgue) almost everywhere continuously differentiable and, therefore, at each stage $k$ of a GGM generated algorithm, there is plenty of points $x^{k+1}$ (which can be eventually achieved by adequately choosing the relaxation parameter $\lambda_k$ and the weight function $u_k$) at which all functionals $f_i$, ($i \in I$), are continuously differentiable. Hence, when $X = \mathbb{R}^n$, the computational procedure of "moving" from a step to another according to a GGM generated algorithm can be substantially simplified by appropriate choices of the relaxation parameters and of the weight functions. This is one of the reasons that convergence criteria for GGM generated algorithms have to leave as much freedom as possible for the choice of the $\lambda_k$'s and $u_k$'s.

It was noted above that the generalized gradient of a convex function is exactly its subgradient (see (9)). Thus, if all functionals $f_i$, ($i \in I$), in the feasibility problem (1) are convex, then the GGM algorithms are "subgradient methods". Subgradient methods for convex feasibility problems were studied, among others, by Eremin [19], Goffin [24] and Censor and Lent [10].

Now, consider the intersection problem $Q := \bigcap_{i \in I} Q_i$ in its normal form (5). The functionals $d_i := d_{a_i}$ involved in (5) are locally Lipschitz. Therefore, GGM generated algorithms for the
feasibility problem (5) always exist. In that follows, a GGM generated algorithm for the normal form of an intersection problem will be called a normal algorithm. In particular, normal algorithms for feasibility problems in dual form (4) always exist. A normal algorithm for the dual (4) of the feasibility problem (1) will be called a normal algorithm for the problem (1). It will be shown in Section 4 that normal algorithms in finitely dimensional spaces converge under quite unrestrictive conditions. Normal algorithms for solving convex feasibility problems via their duals were studied by Aharoni, Berman and Censor [2] and Flam and Zowe [22].

Another class of GGM generated algorithms consists of the so called "projection methods for solving convex intersection problems". A projection method is a GGM generated algorithm for the feasibility problem
\[ d_{Q_i}^2(x) \leq 0, \quad (i \in I), \tag{13} \]
associated to the intersection problem \( Q := \bigcap_{i \in I} Q_i \) involving convex closed sets \( Q_i, \quad (i \in I). \) It was noted above that, if \( Q_i, \quad (i \in I), \) are convex and closed, then the functionals involved in (13) are continuously differentiable and their gradients are given by (11). Therefore, a projection method is an algorithm which can be presented as follows: Choose the initial point \( x^0 \in X \) and do:
\[ x^{k+1} = x^k + \lambda_k \sum_{i \in I} w_k(i)(P_{Q_i}(x^k) - x^k), \quad (k \in \mathbb{N}), \tag{14} \]
where \( \lambda_k \) and \( w_k \) are as above. Projection methods for solving convex feasibility problems (1) via their duals (4) are well studied. General convergence criteria for projection methods can be found in [3], [6] and [28]. In Section 5 yet another convergence theorem for projection methods is proved.
3. The Behavior Of GGM Generated Algorithms In A Hilbert Space

In this section we show that convergent GGM generated algorithms necessarily approximate solutions of the primal feasibility problem (1) provided that such solutions exist and that the solution set of the given feasibility problem is sufficiently "regular".

We consider the feasibility problem (1), the sets $C_i$, $(i \in I)$, as defined at (3) and the set $C = \bigcap_{i \in I} C_i$ of all solutions of (1). The function $f: X \longrightarrow \mathbb{R}$ defined by

$$f(x) := \max \{ f_i(x) \mid i \in I \}, \quad (x \in X), \quad (15)$$

will be called envelope of the family of functions $f_i$, $(i \in I)$.

For any $x \in X$ we denote

$$I(x) := \{ i \in I \mid f_i(x) = f(x) \}. \quad (16)$$

A weight function $w: I \longrightarrow \mathbb{R}_+$ is said to be proper at the point $x \in X$ if for any $j \in I \setminus I(x)$ we have $w(j) = 0$. Since for any $x \in X$ $I(x) \neq \emptyset$, it follows that for any $x \in X$ there exists at least one weight function $w_x: I \longrightarrow \mathbb{R}_+$ which is proper at $x$, namely, the weight function defined by

$$w_x(i) = \begin{cases} \frac{1}{|I(x)|} & \text{if } i \in I(x), \\ 0 & \text{otherwise}. \end{cases} \quad (17)$$

A GGM generated sequence $(x^k \mid k \in \mathbb{N})$ is called properly generated if, for each $k \in \mathbb{N}$, the weight function $w^k$ involved in (12) is proper at $x^k$. For instance, taking in (12) $w^k = w^k_x$ as defined at (17), we obtain proper GGM generated sequences.

The following result concerning the proper weight functions will be of use.

**Lemma 1.** Suppose that the functionals $f_i$, $(i \in I)$, involved in (1) are continuous. Let $(x^k \mid k \in \mathbb{N})$ be a convergent sequence
in \( X \) and let \( \bar{x} \) be its limit. If \( \{ w_k \mid k \in \mathbb{N} \} \) is a sequence of weight functions such that, for each \( k \in \mathbb{N} \), \( w_k \) is proper at \( x^k \), then each cluster point of \( \{ w_k \mid k \in \mathbb{N} \} \) in \( \mathbb{R}^I \) is a weight function which is proper at \( \bar{x} \).

Proof: Since sequences of weight functions on \( I \) are bounded sequences in \( \mathbb{R}^I \) it follows that they always have cluster points and, obviously, cluster points of a sequence of weight functions are weight functions. It remains to show that, if \( \{ w_k \mid k \in \mathbb{N} \} \) is a sequence of weight functions such that, for each \( k \in \mathbb{N} \), \( w_k \) is proper at \( x^k \) and if \( w_* \) is a cluster point of \( \{ w_k \mid k \in \mathbb{N} \} \), then \( w_* \) is proper at \( \bar{x} \). To this end, let \( i \) be in \( I \setminus I(\bar{x}) \) and consider a subsequence \( \{ w_{k_p} \mid p \in \mathbb{N} \} \) which converges to \( w_* \). Then \( f_i(\bar{x}) < f(\bar{x}) \). Since the functionals \( f_i \) are continuous, there exists a neighborhood \( U \) of \( \bar{x} \) such that \( f_i(y) < f(y) \) for all \( y \in U \). Therefore, there exists a positive integer \( k^* \) such that \( f_i(x^k) < f(x^k) \) for all integers \( k \geq k^* \). Hence, \( i \in I(x^k) \) for all integers \( k \geq k^* \). Thus, \( w_{k_p}(i) = 0 \) whenever \( k_p \geq k^* \) (because \( w_{k_p} \) is proper at \( x^{k_p} \)). By consequence, \( w_*(i) = \lim_{p \to \infty} w_{k_p}(i) = 0 \) and the proof is complete. \( \blacksquare \)

In order to prove the main theorem of this section we observe that the proof of Proposition 2.3.12 of Clarke [13] can be reproduced without essential modifications in order to deduce that, if the envelope of the family of functionals \( f_i \) (\( i \in I \)), is regular, then we have

\[
\partial f(x) = \text{conv} \left\{ \bigcup_{i \in I(x)} \partial f_i(x) \right\}, \quad (18)
\]

for any \( x \in X \).

Now we are in position to prove the following limit characterization for proper GGM generated sequences:

THEOREM 1: Suppose that (1) is a consistent locally Lipschitz feasibility problem and the envelope $f$ of the family of functionals $f_i$, $(i \in I)$, is convex. Let $(x^k \mid k \in \mathbb{N})$ be a GGM properly generated sequence whose relaxation parameters satisfy the condition

$$\inf \{\lambda_k \mid k \in \mathbb{N}\} = \alpha > 0,$$  \hspace{1cm} (19)

If the sequence $(x^k \mid k \in \mathbb{N})$ converges, then its limit is a solution of the feasibility problem (1).

Proof: According to (12) and (19) we have

$$\| \sum_{i \in I} w_k(i) \xi_i^k \| \leq \frac{1}{\alpha} \| x^{k+1} - x^k \|,$$

for any $k \in \mathbb{N}$. Since the sequence $(x^k \mid k \in \mathbb{N})$ converges, it follows that

$$\lim_{k \to \infty} \sum_{i \in I} w_k(i) \xi_i^k = 0.$$  \hspace{1cm} (20)

According to Lemma 1, the sequence $(w_k \mid k \in \mathbb{N})$ contains a convergent subsequence $(w_{k_p} \mid p \in \mathbb{N})$ which converges to a weight function $w_*$ which is proper at the point $x = \lim_{k \to \infty} x^k$. Since the functionals $f_i$, $(i \in I)$, are locally Lipschitz, then their envelope $f$ is locally Lipschitz. Locally Lipschitz and convex functionals are regular (cf. [13, Prop. 2.3.6(b)]). Hence, the functional $f$ satisfies (18). By consequence, for each $p \in \mathbb{N}$,

$$\sum_{i \in I} w_{k_p}(i) \xi_i^{k_p} \in \partial f(x^{k_p})$$

because $w_{k_p}$ is proper at $x^{k_p}$. According to (20), the sequence on the left hand side of (21) converges to zero. The generalized gradient multifunction $x \longrightarrow \partial f(x)$ is weak* closed (cf. [13, Prop. 2.1.5(b)]) and the subsequence $(x^{k_p} \mid p \in \mathbb{N})$ converges to $x$. Hence, $0 \in \partial f(\bar{x})$. This means that $\bar{x}$ is a global minimum of $f$ because $f$ is convex and, then, (9) holds. Since, by hypothesis, the feasibility problem (1) is consistent, we have $C \neq \emptyset$. For any
z \in C \text{ we have } f_i(\overline{x}) \leq f(x) \leq f(z) \leq 0, \quad (i \in I), \text{ showing that } \overline{x} \text{ is a solution of the feasibility problem (1) and this completes the proof.}

This result generalizes in some respects Theorem 3.2 in [6] and, implicitly, a limit characterization theorem due to Iusem and De Pierro [28].

Note that the condition of Theorem 1 that the envelope \( f \) of the family of functions \( f_i, \ (i \in I) \), is convex implies that the solution set \( C \) of (1) has to be convex but it doesn't imply that the functionals \( f_i, \ (i \in I) \), are convex. For instance, if \( X = \mathbb{R}, \ I = (1,2), \ f_1(x) := x(x-1) \) and \( f_2(x) := x^3 \), then the envelope \( f \) is convex on \( X \) in spite of the fact that \( f_2 \) is not convex on \( X \). However, if each \( f_i \) is convex we have the following

**COROLLARY 1:** Suppose that (1) is a convex locally Lipschitz consistent feasibility problem. If \( \{x^k \mid k \in \mathbb{N} \} \) is a GGM properly generated convergent sequence with relaxation parameters satisfying (19), then the limit of \( \{x^k \mid k \in \mathbb{N} \} \) is a solution of the feasibility problem (1).²

If \( X = \mathbb{R}^n \), then the condition that each \( f_i \) is locally Lipschitz can be removed from Corollary 1 since, in this case, finite convex functions are implicitly locally Lipschitz. The condition that each functional \( f_i \) is locally Lipschitz can be further avoided in Theorem 1 if one replace the primal problem (1) by its dual (4) since the distance functions \( d_i(x) = d_{\mathbb{R}^n}(x) \) are necessarily globally Lipschitz (see Section 2). Precisely, we have the following

**COROLLARY 2:** Suppose that the feasibility problem (1) is consistent and the functionals \( f_i, \ (i \in I), \) involved in it are continuous. If the envelope of the family of functionals \( d_i \)
(\ell \in I), is convex, then any convergent sequence \( (x^k \mid k \in \mathbb{N}) \) which is properly generated by the normal method with relaxation parameters \( \lambda_k \), \( (k \in \mathbb{N}) \), satisfying (19) converges to a solution of the feasibility problem (1).\]

Theorem 1 and its corollaries characterizes the limit of GGM properly generated sequences provided that they are convergent. In general, GGM properly generated sequences may be divergent even if (19) hold. An example in this sense can be found in [6]. Convergence of GGM generated sequences can be ensured under weak conditions provided that the generalized gradient multifunctions \( x \mapsto \partial f_i(x), (i \in I) \), are bounded. In this case, if \( \lim k=0 \lambda_k < +\infty \), then the GGM generated sequences with the relaxation parameters \( \lambda_k \) are Cauchy sequences, hence convergent (see, for instance, Goffin [23, Lemma 4.1]). Convergence by itself does not necessarily imply that the limit of a GGM generated sequence is a solution of the corresponding feasibility problem. However, we have the following convergence criterion:

**THEOREM 2:** Suppose that (1) is a consistent locally Lipschitz feasibility problem and that the envelope \( f \) of the functionals \( f_i, (i \in I) \), is convex. If \( (x^k \mid k \in \mathbb{N}) \) is a GGM properly generated sequence with relaxation parameters \( (\lambda_k \mid k \in \mathbb{N}) \) satisfying (19) and such that \( \sum_{k=0}^{\infty} \lambda_k \| \xi^k \| < +\infty, (i \in I) \), then the sequence \( (x^k \mid k \in \mathbb{N}) \) converges and its limit is a solution of the feasibility problem (1).

**Proof:** We first show that \( (x^k \mid k \in \mathbb{N}) \) is a Cauchy sequence, hence convergent. To this end, note that, according to (12), for each \( k \in \mathbb{N} \) we have

\[
\| x^{k+1} - x^k \| \leq \lambda_k \sum_{i \in A} v_k(i) \| \xi^k_i \|.
\]

Therefore, for any \( \rho, k \in \mathbb{N} \), we have
\[ \| x^{k+P} - x^k \| = \sum_{j=0}^{p-1} \| x^{k+j+1} - x^{k+j} \| \leq \sum_{j=0}^{p-1} \| x^{k+j+1} - x^{k+1} \| \]
\[ \leq \sum_{j=0}^{p-1} \lambda_{k+j} \sum_{i \in I} w_{k+j}(i) \| \xi_i^{k+j} \| = \sum_{i \in I} \sum_{j=0}^{p-1} w_{k+j}(i) \lambda_{k+j} \| \xi_i^{k+j} \| \]
\[ = \sum_{i \in I} \sum_{j=k}^{p+k-1} w_j(i) \lambda_j \| \xi_i^j \| \leq \sum_{i \in I} \sum_{j=k}^{\infty} w_j(i) \lambda_j \| \xi_i^j \| \]
\[ \leq \sum_{i \in I} \sum_{j=k}^{\infty} \lambda_j \| \xi_i^j \|. \]

Since the last sum is the reminder of a convergent series, it follows that \( \| x^{k+P} - x^k \| \) becomes arbitrarily small when \( k \) is arbitrarily large, that is the sequence \( (x^k \mid k \in \mathbb{N}) \) is a Cauchy sequence. Now, since \( (x^k \mid k \in \mathbb{N}) \) is convergent and Theorem 1 applies, it follows that the limit of \( (x^k \mid k \in \mathbb{N}) \) is a solution of (1).\textsuperscript{1}

It will be shown in a further paper that Theorem 2 can be used in combination with results in [6] and [31] in order to obtain an alternative proof for the von Neumann-Halperin convergence theorem concerning the alternating projection method (see [34] and [29]). However, from a purely applicative point of view, convergence criteria for GGM generated algorithms imposing conditions on the choices of the vectors \( \xi_i^k \) in the corresponding generalized gradients may be difficult to implement. That is why, in that follows, we concentrate on proving convergence criteria for GGM generated algorithms under assumptions which are easy to implement from a computational point of view.
4. A Convergence Theorem For GGM Generated Algorithms In $\mathbb{R}^n$

One of the difficulties which one encounters while trying to prove convergence criteria for GGM generated algorithms in a Hilbert space is determined by the fact that, in general, the unit ball in a Hilbert space is not compact and the convexity of the space may be not as "uniform" as we would like to. Sometimes, one may be able to overcome these difficulties by using special techniques which apply when the feasibility problem is sufficiently restricted as happens, for instance, in the case when the feasibility problem is a normal form intersection problem of closed linear subspaces of $X$ (see [29]). In this and in the next section we restrict $X$ to be the Euclidean space $\mathbb{R}^n$. In this setting, the space is sufficiently uniform for ensuring convergence for large classes of GGM generated algorithms. If $X$ is not finitely dimensional, then the following convergence criteria still hold if we replace "convergence" by "weak convergence".

**Theorem 3:** Suppose that (1) is a locally Lipschitz feasibility problem in $X = \mathbb{R}^n$ with the solution set having nonempty interior and such that, for each $i \in I$, the generalized gradient multifunction $\partial f_i : x \rightarrow \partial f_i(x)$ is bounded. If the envelope $f$ of the family of functionals $f_i$, $(i \in I)$, is convex and if $(x^k \mid k \in \mathbb{N})$ is a GGM properly generated sequence with relaxation parameters $\lambda_k$ satisfying the condition

$$\max(0, f(x^k)) \leq \lambda_k M^2 \leq 2 \cdot \max(0, f(x^k)), \quad (k \in \mathbb{N}), \quad (23)$$

for some upperbound $M$ of all multifunctions $\partial f_i$, $(i \in I)$, then the sequence $(x^k \mid k \in \mathbb{N})$ converges to a solution of the feasibility problem (1).

*Proof:* Since $(x^k \mid k \in \mathbb{N})$ is GGM generated it has the form (12). We denote...
\[ u^k := \sum_{i \in I} w_k(i) \xi_i^k, \quad (k \in \mathbb{N}). \]

First we prove the following

**Claim 1.** For any \( z \in C \) and for any \( k \in \mathbb{N} \) we have

\[ \|x^{k+1} - z\| \leq \|x^k - z\|. \quad (24) \]

In order to prove that, note that

\[ \|x^{k+1} - z\|^2 = \|x^k - z\|^2 + \lambda_k \left[ \lambda_k \|\nu_k\|^2 - 2 \cdot \langle \nu_k, x^k - z \rangle \right]. \quad (25) \]

If \( \lambda_k = 0 \) or \( \nu_k = 0 \), then (24) clearly holds with equality. Assume that \( \lambda_k \neq 0 \) and \( \nu_k \neq 0 \). Then, according to (23), we have \( \lambda_k > 0 \) and \( f(z) \leq 0 < f(x^k) \) for any \( z \in C \). Since \( f \) is convex, then \( f \) is locally Lipschitz and regular and, therefore, we have

\[ f^{z}(x^k, z - x^k) = \lim_{t \to 0} \frac{f(x^k + t(z - x^k)) - f(x^k)}{t}. \quad (26) \]

Also, for any \( t \in [0,1] \) we have

\[ f(x^k + t(z - x^k)) - f(x^k) \leq t(f(z) - f(x^k)) \quad (27) \]

because \( f \) is convex. From (26) and (27) it follows that

\[ f^{z}(x^k, z - x^k) \leq f(z) - f(x^k) < 0, \quad (z \in C). \quad (28) \]

Again because \( f \) is regular (18) holds. Hence, for each \( k \in \mathbb{N} \), \( u^k \in \partial f(x^k) \) since the sequence \( \{x^k \mid k \in \mathbb{N}\} \) is properly generated (i.e., \( w_k \) is proper at \( x^k \)). By consequence, we have

\[ \langle u^k, z - x^k \rangle \leq f^{z}(x^k, z - x^k), \quad (z \in C). \quad (29) \]

By combining (28) and (29) we obtain

\[ \langle u^k, x^k - z \rangle \leq -f^{z}(x^k, z - x^k) \geq f(x^k) - f(z) \geq f(x^k) > 0, \]

for any \( z \in C \). From that and from (23) we obtain

\[ 2 \cdot \langle u^k, x^k - z \rangle \geq 2 \cdot f(x^k) \geq \lambda_k \|x^k\|^2 \geq \lambda_k \|\nu_k\|^2, \quad (z \in C), \quad (30) \]

since \( M \) is an upperbound of each \( \partial f_i \) and, therefore, an upperbound of \( \partial f \) (cf. (18)). Formula (30) shows that the expression between square brackets in (25) is nonpositive. Hence, (24) holds and the claim is completely proved.

Since \( C \neq \emptyset \) and (24) holds, it follows that the sequence \( \{\|x^k - z\| \mid k \in \mathbb{N}\} \) is convergent for any \( z \in C \) and that the
sequence \( \{x^k \mid k \in \mathbb{N}\} \) is bounded. Hence, there exists a convergent subsequence \( \{x^{k^p} \mid p \in \mathbb{N}\} \) of the sequence \( \{x^k \mid k \in \mathbb{N}\} \). Let \( x^* \) be the limit of this subsequence. The function \( f \) is continuous as being convex (cf. Rockafellar [38, Theorem 10.1]). Hence, the sequence \( \{f(x^{k^p}) \mid p \in \mathbb{N}\} \) converges to \( f(x^*) \). According to (23), the sequence \( \{\lambda^{k^p} \mid p \in \mathbb{N}\} \) is bounded because \( \{f(x^{k^p}) \mid p \in \mathbb{N}\} \) is bounded. The sequence \( \{\nu^k \mid k \in \mathbb{N}\} \) is bounded because of the boundedness of the multifunctions \( \partial f_i \) and because of the proper definition of the sequence \( \{x^k \mid k \in \mathbb{N}\} \). By consequence, using Lemma 1 we deduce that there exists a sequence \( \{s_t \mid t \in \mathbb{N}\} \) of nonnegative integers such that the following limits exist

\[
\begin{align*}
x^* &= \lim_{t \to \infty} x^{s_t}, \\
\lambda^* &= \lim_{t \to \infty} \lambda_{s_t}, \\
\nu^* &= \lim_{t \to \infty} \nu_{s_t}, \\
w^* &= \lim_{t \to \infty} w_{s_t},
\end{align*}
\]

the weight function \( w^* \) is proper at \( x^* \) and

\[
\nu^* = \sum_{i \in I} w^* \xi_i^*,
\]

where,

\[
\xi_i^* = \lim_{t \to \infty} \xi_i^{s_t}, \quad (i \in I).
\]

Since, for each \( t \in \mathbb{N} \), we have

\[
x^{s_{t+1}} = x^t - \lambda_{s_t} \nu^t,
\]

it follows that \( \{x^{s_{t+1}} \mid t \in \mathbb{N}\} \) converges to \( x^* - \lambda^* \nu^* \). On the other hand, for each \( z \in C \), \( \{\|x^{s_{t+1}} - z\| \mid t \in \mathbb{N}\} \) and \( \{\|x^t - z\| \mid t \in \mathbb{N}\} \) are subsequences of the same convergent sequence \( \{\|x^k - z\| \mid k \in \mathbb{N}\} \) (cf. Claim 1) and, therefore, they have the same limit, i.e.,

\[
\|x^* - z\| = \lim_{t \to \infty} \|x^{s_t} - z\| = \lim_{k \to \infty} \|x^k - z\|
\]

\[
= \lim_{t \to \infty} \|x^{s_{t+1}} - z\| = \|x^* - \lambda^* \nu^* - z\|,
\]

no matter how \( z \) is chosen in \( C \). This implies
\[ \|x^* - z\|^2 = \|x^* - z\|^2 + \lambda_\ast \left[ \lambda_\ast \nu^* \|z\|^2 - 2 \cdot \langle \nu^*, x^* - z \rangle \right], \]
for any \( z \in C \). Hence, for any \( z \in C \), we have
\[ \lambda_\ast \left[ \lambda_\ast \nu^* \|z\|^2 - 2 \cdot \langle \nu^*, x^* - z \rangle \right] = 0. \]  
(37)

Using this fact, we can prove the following

**Claim 2.** The point \( x^* \) defined at (31) belongs to \( C \).

To this end, note that, \( \nu^* \in \partial f(x^*) \) since, according to (18), we have \( \nu^t \in \partial f(x^t) \), \( t \in \mathbb{N} \), (31) and (34) hold and the subgradient multifunction \( \partial f \) is weak* closed. Hence, if \( \nu^* = 0 \), then \( 0 \in \partial f(x^*) \), that is \( x^* \) is a global minimum of \( f \) (cf. (9)). Therefore, if \( \nu^* = 0 \), we have \( f(x^*) \leq f(z) \leq 0 \), \( z \in C \), and this implies \( x^* \in C \). Assume that \( \nu^* \neq 0 \). Then, according to (37), we have either \( \lambda_\ast = 0 \) or
\[ \lambda_\ast \nu^* \|z\|^2 - 2 \cdot \langle \nu^*, x^* - z \rangle = 0, \quad (z \in C). \]  
(38)

If (38) holds, then \( C \) is a subset of the hyperplane of equation
\[ \langle \nu^*, x \rangle = (1/2) \left[ 2 \cdot \langle \nu^*, x^* \rangle - \lambda_\ast \nu^* \|z\|^2 \right]. \]
This contradicts the assumption that \( \text{Int}(C) \neq \emptyset \). Hence, necessarily \( \lambda_\ast = 0 \). According to (23), (31) and (32) this implies \( f(x^*) \leq 0 \), i.e. \( x^* \in C \). The claim is proved.

Now, according to Claim 2 and Claim 1, the sequence \( \|x^k - x^*\| \mid k \in \mathbb{N} \) converges and it must have the same limit as its subsequence \( \|x^t - x^*\| \mid t \in \mathbb{N} \), i.e. \( \lim_{k \to \infty} \|x^k - x^*\| = 0 \). By consequence, the sequence \( x^k \mid k \in \mathbb{N} \) converges to the point \( x^* \) which is a solution of the feasibility problem (1). The proof of the theorem is complete.

It can be easily seen that Theorem 3 applies to intersection problems in normal form provided that their solution set has nonempty interior since the functionals involved in (5) are globally Lipschitz and have bounded (by 1) generalized gradients.

It was noted in Section 2 that convex functionals on \( \mathbb{R}^n \) are

D. Butnariu/Generalized Gradient Methods.../Dec. 1990/page 21
locally Lipschitz and that they satisfy (9). Therefore, if the feasibility problem (1) is "convex" (i.e., all functionals \( f_i \), \( i \in I \), are convex), then Theorem 3 can be restated as follows:

**COROLLARY 3:** Suppose that (1) is a convex feasibility problem and \( \text{Int}(C) \neq \emptyset \). If the subgradient multifunctions of the functionals \( f_i \), \( i \in I \), are bounded, then any GGM properly generated sequence with relaxation parameters satisfying (23) converges to a solution of the feasibility problem (1).■

Corollary 3 applies to convex intersection problems in normal form, that is we have the following:

**COROLLARY 4:** Suppose that \( Q_i \), \( i \in I \), are closed convex subsets of \( \mathbb{R}^n \) and that the set \( Q = \bigcap_{i \in I} Q_i \) has nonempty interior. If \( (x^k \mid k \in \mathbb{N}) \) is a sequence properly generated by the normal method such that the relaxation parameters satisfy

\[
\max_{i \in I} d_i(x^k) \leq \lambda_k \leq 2 \cdot \max_{i \in I} d_i(x^k), \quad (k \in \mathbb{N}),
\]  

(39)

then \( (x^k \mid k \in \mathbb{N}) \) converges to a point in \( Q \).■

In particular, Corollary 5 applies to feasibility problems (1) in dual form. Precisely, we have

**COROLLARY 5:** Suppose that (1) is a feasibility problem with continuous functionals \( f_i \), \( i \in I \). If the sets \( C_i \), \( i \in I \), are convex and \( \text{Int}(C) \neq \emptyset \), then any sequence \( (x^k \mid k \in \mathbb{N}) \) which is properly generated by the normal method for (1) and with relaxation parameters \( \lambda_k \), \( (k \in \mathbb{N}) \), satisfying (39) converges to a solution of the feasibility problem (1).■

Solving intersection problems via their normal forms by normal method generated algorithms involves computing generalized gradients of the functionals \( d_i \), that is normal vectors of the corresponding sets. This can be done using the "chain rules" in
[13, Chapter 2]. If the intersection problem is convex, then one may be able to avoid computing normal vectors while solving the intersection problem by a projection algorithm. A convergence theorem for projection algorithms is given in the next section.
5. A Convergence Criterion For Projection Algorithms in $\mathbb{R}^n$

In this section we consider the intersection problem of a family of closed convex subsets $Q_i$, $(i \in I)$, of $\mathbb{R}^n$ in the form (13). It was noted in Section 2 that the functionals involved in (13) are continuously differentiable on $\mathbb{R}^n$ and that their gradients are given by (11). Nevertheless, Theorem 3 and its corollaries do not apply to the feasibility problem (13) since, in general, the gradients of the functionals $g_i$ are not bounded. In this section we show that, in spite of this fact, a large class of projection algorithms converge to solutions of the intersection problem provided that such solutions exist. However, it has to be pointed out that projection algorithms converge slower than their normal counterparts.

**THEOREM 4:** Let $Q_i$, $(i \in I)$, be a family of closed convex subsets of $\mathbb{R}^n$ and suppose that $Q = \bigcap_{i \in I} Q_i$ has nonempty interior. If $(x^k | k \in \mathbb{N})$ is a sequence properly generated by the projection method (14) with relaxation parameters $\lambda_k \in [\alpha, 2]$, $(k \in \mathbb{N})$, for some $\alpha > 0$, then the sequence $(x^k | k \in \mathbb{N})$ converges to a point in $Q$, no matter how the initial point $x^0$ is chosen.

**Proof:** First note that the sequence $(x^k | k \in \mathbb{N})$ satisfies (24) for any $z \in Q$ because of Corollary 4.3 in [6]. Since $Q$ is not empty, this implies that the sequence $(x^k | k \in \mathbb{N})$ is bounded. Hence, it has a convergent subsequence $(x^{k_p} | p \in \mathbb{N})$. Let $\bar{x}$ be the limit of this subsequence. Since $(\lambda_{k_p} | p \in \mathbb{N})$ is bounded by hypothesis and the sequence $(x^k | k \in \mathbb{N})$ is properly generated, it follows (cf. Lemma 1) that there exist a sequence $(s_t | t \in \mathbb{N})$ of nonnegative integers such that (31), (32) and (34) still hold and such that $w^*_t$ is proper at $\bar{x}$. Therefore, the sequence $(x^{s_t+1} | ...
\( \lambda \in \mathbb{N} \) converges to \( x^* + \lambda \sum_{i \in E} w_i(i)(P_{Q_i}(x^*) - x^*) \). Since, for each \( z \in Q \), the sequences \( \{\|x^{*k} - z\| \mid k \in \mathbb{N}\} \) and \( \{\|x^{*k+1} - z\| \mid k \in \mathbb{N}\} \) must have the same limit because of (24), it follows that, for any \( z \in Q \), we have

\[
\|x^* + \lambda \sum_{i \in E} w_i(i)(P_{Q_i}(x^*) - x^*) - z\| = \|x^* - z\|. \tag{40}
\]

This implies

\[
\lambda \left[ \lambda \|\nu^*\|^2 + 2 \cdot \langle \nu^*, x^* - z \rangle \right] = 0, \quad (z \in Q), \tag{41}
\]

where

\[
\nu^* = \sum_{i \in E} w_i(i)(P_{Q_i}(x^*) - x^*). \]

By hypothesis, \( \lambda \neq 0 \). Therefore, it follows from (41) that either \( \nu^* = 0 \) or \( Q \) is a subset of a hyperplane in \( \mathbb{R}^n \) of normal \( \nu^* \). Since \( Q \) has nonempty interior, the latter is not possible. Hence, \( \nu^* = 0 \).

Now, denote \( J := \{i \in I \mid w_i(i) > 0\} \) and let \( w \) be the restriction of the weight function \( w^* \) to \( J \). Obviously, \( w \) is still a weight function. Since \( \nu^* = 0 \), it follows that \( x^* \) is a fixed point of the operator \( P_{\nu} : \mathbb{R}^n \rightarrow \mathbb{R}^n \) defined by

\[
P_{\nu}(x) = \sum_{i \in J} w_i(i)P_{Q_i}(x). \]

Note that Proposition 2.2(iii) in [6] applies to \( P_{\nu} \) and it shows that \( x^* \in \bigcap_{i \in J} Q_i \). Hence, \( \partial_i^2(x^*) = 0 \) for any \( i \in J \) (cf. (10)).

Observe that \( J \subseteq I(x^*) \) because \( w^* \) is proper at \( x^* \). Hence, if \( j \in J \), then \( \partial_j^2(x^*) = \max \{\partial_i^2(x^*) \mid i \in I\} \). Therefore, we have \( \partial_i^2(x^*) = 0 \) for any \( i \in I \) and this completes the proof. \( \blacksquare \)
REFERENCES


30. S. Kaczmarz, Angenaherte aufl"ossung von systemen linearer
31. S. Kayalar and H. Wienert, Error bounds for the method of
alternating projections, Mathematics of Control, Signals and
32. T. S. Motzkin and I. J. Sche"oenberg, The relaxation method for
linear inequalities, Canadian Journal of Mathematics, 6, (1954),
393-404.
33. H. Nakano, Spectral Theory in Hilbert Space, Japanese Society
For Promotion Of Science, Tokyo, 1953.
34. J. von Neumann, On rings of operators. Reduction theory,
35. B. T. Polyak, Minimization of nonsmooth functionals, USSR
Computational Mathematics and Mathematical Physics, 9, (1969),
509-521.
36. B. T. Polyak, A general method for solving extremum problems,
37. A. W. Roberts and D. E. Varberg, Another proof that convex
functions are locally Lipschitz, American Mathematical Monthly,
38. R. T. Rockafellar, Convex Analysis, Princeton University
39. N. Z. Shor, An application of the method of gradient descent
to the solution of the network transportation problem, in:
Materialy Naucnovo Seminara po Theoret i Priklad. Voprosam