Analele ştiinţifice ale Universităţii „Al. I. Cuza" din Iaşi
Tomul XXV, s. I a, 1979, f. 2

CONSTRAINED MINIMIZATION OF NONSMOOTH FUNCTIONALS BY AN EXTENSION OF THE FRANK-WOLFE METHOD

BY

C. ILIOI

The F r a n k - W o l f e method is a simple and efficient procedure of minimization, especially when linear constraints are involved. This method was extended by J. C é a [1] in a Hilbert space setting, for twice differentiable strictly convex functionals. A slightly modified version of it, was also applied to some special non-differentiable functions [2]. Our aim is to prove the convergence of an extension of the Frank-Wolfe algorithm for a large class of nonsmooth convex functionals. An application is also considered to nonlinear programming problems, by using an extremal-value function studied by W. H o g a n [2].

1. Hypotheses and description of the method. Let $J : E \to \bar{R}$ be a proper convex, lower semicontinuous functional, defined on the reflexive space E, $K \subset E$ a convex closed set and suppose that

(1) $$\exists u_0 \in K,\ W(u_0) = \{v \in D(J) / J(v) \leqslant J(u_0)\} \text{ is bounded.}$$

Then the problem

(2) $$\min_{v \in K} J(v) = J(u) = \min_{v \in U} J(v),\ U = W(u_0) \cap K$$

has at least a solution $u \in U$. The iterative method is

(3) $$u_{n+1} = u_n + \lambda_n h_n,\ n = 0, 1, 2, \ldots,$$

where $h_n \in E$, $\| h_n \| = 1$ and $\lambda_n \in R$, $\lambda_n > 0$.

Under our assumptions, the directional differential $J'_+(v, h)$ exists for any $h \in E$ and it is convex in h. For the construction of h_n at each step, let us consider a closed bounded convex set $V \subset E$ containing O and y_n defined by :

(4) $$J'_+(u_n,\ y_n - u_n) = \min_{y \in U_n} J'_+(u_n,\ y - u_n),\ U_n = U \cap (u_n + V).$$

If y_n could be exactly calculated, we can take $h_n = (y_n - u_n) / \| y_n - u_n \|$. But, in the most cases, we get only an approximation z_n of y_n, obtained by a numerical method, such as the relative error at each step is kept under a given bound ε:

(5) $$| J'_+(u_n, z_n - u_n) - J'_+(u_n, y_n - u_n) |/| J'_+(u_n, y_n - u_n) | \leqslant \varepsilon.$$

Then we take

(6) $$h_n = (z_n - u_n)/\| z_n - u_n \|.$$

Lemma 1. *If u_n is not a solution of the problem* (2), *then h_n given by* (6), *is a descent direction of J at u_n (see Definition* 3.1 *from* [3]).

Proof. From (4) and (5) we get:

(7) $$J'_+(u_n, y - u_n) \geqslant J'_+(u_n, y_n - u_n) \geqslant J'_+(u_n, z_n - u_n)/(1 - \varepsilon)$$

for any $y \in U_n$. If $y = u_n$, we have $J'_+(u_n, z_n - u_n) \leqslant 0$. Suppose that $J'_+(u_n, z_n - u_n) = 0$, then $J'_+(u_n, y - u_n) \geqslant 0$, for any $y \in U_n$, that is u_n yield a local minimum for J on U. As J is convex, u_n will be a global minimum point on U, impossible. Therefore $J'_+(u_n, z_n - u_n) < 0$ and h_n is a descent direction at u_n.

Let us state now the choice of the steplength λ_n on the direction h_n. For that we assume in addition, the following hypothesis:

(8) $$\forall \varepsilon > 0,\ \exists \delta(\varepsilon) > 0,\ 0 < \lambda \leqslant \delta(\varepsilon) \Rightarrow 0 \leqslant J'_-(v + \lambda h, h) - J'_+(v, h) \leqslant \varepsilon,$$

for any $v \in W(u_0)$, $h \in E$, with $\| h \| = 1$. We have considered in a previous paper [4] some classes of convex functionals for which this assumption is satisfied. As we have proved there, under this hypothesis a sequence of real positive numbers μ_n can always be found, such that

(9a) $$J(u_n + \mu_n h_n) < J(u_n)$$

and

(9b) $$\lim_{n\to\infty} (J(u_n) - J(u_n + \mu_n h_n)) = 0 \Rightarrow \lim_{n\to\infty} J'_+(u_n, h_n) = 0.$$

Practical algorithms to compute such sequences have been described also in [4]: the maximum decreasing method and any approximate version of it, the generalized Goldstein's and Armijo's methods. It is easy to show that, for all these algorithms, we have

(10) $$(J(u_n) - J(u_n + \mu_n h_n))/\mu_n \geqslant \gamma(- J'_+(u_n, h_n)).$$

where $\gamma : R^+ \to R^+$ is a „forcing function", i.e. a real function convergent to zero only together with its argument. Let us denote:

(11) $$\nu_n = \max \{\lambda > 0, u_n + \lambda h_n \in U\},$$

and finally, to state:

(12) $$\lambda_n = \min \{\mu_n, \nu_n\},\ n = 0, 1, 2, \ldots,$$

whatever the choice for μ_n satisfying (9) and (10) will be.

2. Convergence of the method. In order to prove the convergence, the first step is the following lemma:

Lemma 2. *If no one of the iterations u_n given by* (3), (6) *and* (12), *is an optimal solution of the problem* (2), *then*:

(13) $$\lim_{n\to\infty} J'_+(u_n, z_n - u_n) = 0.$$

Proof. Since h_n is a descent direction for J at u_n and because of (9a) and (12), the sequence $\{J(u_n)\}$ is nonincreasing. Hence it is convergent and

(14) $$\lim_{n\to\infty} (J(u_n) - J(u_{n+1})) = 0.$$

If $\lambda_n = \mu_n \leqslant \nu_n$, the implication (9b) shows that $\lim\limits_{n\to\infty} J'_+(u_n, h_n) = 0$, which leads immediately to (13), because $z_n - u_n$ belongs to the bounded set V. Let us consider now that $\lambda_n = \nu_n < \mu_n$. The convexity of J implies the inequality:

$$J(u_n) - J(u_n + \nu_n h_n) \geqslant (\nu_n/\mu_n)(J(u_n) - J(u_n + \mu_n h_n)),$$

so that, according to (10), we have:

$$\lim_{n\to\infty} \nu_n \gamma(J'_+(u_n, h_n)) = 0.$$

If ν_n is lower bounded by a positive constant, then we get $\lim\limits_{n\to\infty} J'_+(u_n, h_n) = 0$, which implies, as before, the limit (13). Suppose now that $\lim\limits_{n\to\infty} \nu_n = 0$. From the convexity of U and by the definition (11) of ν_n, it follows easily the inequality: $\nu_n \geqslant \| z_n - u_n \|$, then $\lim\limits_{n\to\infty} \| z_n - u_n \| = 0$. But, because U is a bounded closed set, according to the properties of the convex functionals (see for example [6], Theorem 24.7), we have:

$$\alpha = \sup \{ \| v^* \|_* / v^* \in \partial J(U)\} < \infty.$$

Hence $| J'_+(u_n, z_n - u_n) \| \leqslant \alpha \| z_n - u_n \|$ and (13) holds again.

The main theorem of convergence is stated as follows:

Theorem 1. *If no iteration is an optimal solution, then every weak accumulation point of the sequence $\{u_n\}$ is an optimal solution of the problem* (2) *and*

(15) $$\lim_{n\to\infty} J(u_n) = \min_{v\in K} J(v).$$

Proof. Let $u \in U$ be a solution of the problem (2). For each n, if $u_n \notin U_n$ let $0 < \theta_n \leqslant 1$ be such that $w_n = u_n + \theta_n(u - u_n)$ is on the frontier of U_n. When $u \in U_n$ we can take $w_n = u$, $\theta_n = 1$. Anyway if we denote by r the radius of the largest sphere contained in V, we have $\| u - u_n \| \geqslant r/\theta_n$. Then there exists $M > 0$ such that $\theta_n \geqslant M$ for any natural n, because otherwise the sequence $\{u - u_n\} \subset V$ will be unbounded, which contradicts the hypotheses. Hence:

(16) $$u - u_n = (w_n - u_n)/\theta_n,\ w_n \in U_n,\ M \leqslant \theta_n \leqslant 1.$$

As a consequence of the convexity of J and according to (16) and (7) we can write:

$$J(u) \geqslant J(u_n) + J'_+(u_n, z_n - u_n)/M(1-\varepsilon). \tag{17}$$

Let u^* be a weak accumulation point of the weakly compact set $\{u_n\}$ and $\{u_{n'}\} \subseteq \{u_n\}$ a subsequence, weakly convergent to u^*. Then as a consequence of the Lemma 2 and of the lower semicontinuity of J, from the inequality (17) we obtain:

$$J(u) \geqslant \limsup_{n'\to\infty} J(u_{n'}) \geqslant \liminf_{n'\to\infty} J(u_{n'}) \geqslant J(u^*),$$

wherefrom we conclude that u^* is an optimal solution and $\lim_{n'\to\infty} J(u_{n'}) = J(u)$. Finally, since this argument is valid for any other sequence, which is weakly convergent, we get to (15).

When the solution of the problem is unique, we have got the weak convergence of $\{u_n\}$ to the unique optimal solution. Further, if a condition of strong compactity of U, or alternately a hypothesis of uniform convexity for J is assumed, then the strong convergence is achieved.

3. Applications. At each step of the method, we have to solve, with certain accuracy the problem (4) which is equivalent to

$$\max_{u_n^*\in\partial J(u_n)} < y_n - u_n, u_n^* > \min_{v\in U_n} \max_{u_n^*\in\partial J(u_n)} < y - u_n, u_n^* >, \tag{18}$$

where $\partial J(u_n)$ is the subdifferential of J at u_n. Thus the efficiency of our method depends on the simplicity of solution of this problem. Usually, even for differentiable functionals, the Frank-Wolfe method is applied to linearly constrained problems. When the set V is chosen such that to imply only linear constraints, the problem (4) is a linear programming problem. For a quite large class of nonsmooth functionals, as was shown by E. S. L e v i t i n [5] the problem (17) is a linear or quadratique programming problem, or a nonlinear programming problem of a very special kind, easy to solve.

We shall consider an extremal-value function, which was studied previously by W. H o g a n [2] and we shall prove the validity of the hypotheses assumed in the section 1. Let $f: R^p \times R^n \to \overline{R}$, $g: R^p \times R^n \to \overline{R}^m$, convex functions on $R^p \times R^n$ and X a nonempty convex compact set of R^p. Suppose that, for any $x \in X$ there exists a vector $v \in R^n$ such that $g(x, v) < 0$. Consider the function J given by

$$J(v) = \min_{x\in X} \{f(x, v)/g(x, v) \leqslant 0\}. \tag{19}$$

Let us denote $M(v) = \{x \in X / f(x, v) = J(v),\ g(x, v) \leqslant 0\}$, for each $v \in D(J)$.

Lemma 3. *The function J defined by (19) is convex.*

Proof. We show that the epigraph of J (epi J) is a convex set of $R^n \times R$. Let $[v_1, a_1], [v_2, a_2] \in$ epi J and $x_1 \in M(v_1)$, $x_2 \in M(v_2)$. Then $f(x_1, v_1) \leqslant a_1$, $f(x_2, v_2) \leqslant a_2$ and $g(x_1, v_1) \leqslant 0$, $g(x_2, v_2) \leqslant 0$. Since f, g are convex functions we have, for $\lambda \in [0, 1]$, $f(\lambda x_1 + (1-\lambda)x_2, \lambda v_1 + (1-\lambda)v_2) \leqslant a_1\lambda + (1-\lambda)a_2$ and $g(\lambda x_1 + (1-\lambda)x_2, \lambda v_1 + (1-\lambda)v_2) \leqslant 0$. Therefore $J(\lambda v_1 + (1-\lambda)v_2) \leqslant \lambda a_1 + (1-\lambda)a_2$, i.e. epi J is convex.

As a convex function on R^n, J is continuous in the interior of the effective domain $D(J)$ and directionally differentiable. W. Hogan has proved [2] that, for any $v \in D(J)$ and $h \in R^n$ we have

$$J'_+(v, h) = \inf \{f'_+(x, v; y, h)/y \in H(x, v, h)\}, \tag{20}$$

where $f'_+(x, v; y, h)$ is the directional differential of f at $[x, v] \in R^p \times R^n$ in the direction $[y, h] \in R^p \times R^n$, $x \in M(v)$ and

$$\begin{aligned} H(x, v, h) = \{y \in R^p, \|y\| = 1/g(x+\lambda y,\ v+\lambda h) \leqslant 0, \\ x + \lambda y \in X, \textit{ for an interval } 0 < \lambda < \lambda(x, v, y, h)\}. \end{aligned} \tag{21}$$

We have to investigate now, when the essential hypothesis (8) is satisfied. For that, let us recall [4] that a convex functional $J: E \to \overline{R}$ is uniformly directionally differentiable on a set $D \subset D(J)$ if, for any $x \in D$ and $h \in E$, with $\|h\| = 1$, we have $0 \leqslant (J(x+\lambda h) - J(x))/\lambda - J'_+(x, h) \leqslant \varepsilon$, whenever $0 < \lambda < \delta(\varepsilon)$.

Theorem 2. *If, in addition to the above assumptions, f is uniformly directionally differentiable on any compact set $D_1 \times D \subset R^p \times R^n$, then, for the function J given at (18), the assertion (8) is satisfied, for any $v \in D$ and $h \in R^n$ with $\|h\| = 1$.*

Proof. Let $v \in D$, $h \in R^n$ with $\|h\| = 1$, arbitrary, and $x \in M(v)$. According to (20), for any $\varepsilon > 0$, there exists $y \in H(x, v, h)$ such that

$$J'_+(v, h) \geqslant f'_+(x, v; y, h) - \varepsilon/2. \tag{22}$$

On the other side, by the hypothesis of the theorem and because $D_1 = M(D)$ is compact in R^p, there exists $\delta(\varepsilon)$ such that, for any $0 < \lambda < \delta(\varepsilon)$,

$$(f(x+\lambda y,\ v+\lambda h) - f(x, v))/\lambda - f'_+(x, v; y, h) < \varepsilon/2. \tag{23}$$

By combining these relations we obtain:

$$\begin{aligned} 0 \leqslant (J(v+\lambda h) - J(v))/\lambda - J'_+(x, v) \leqslant (f(x+\lambda h,\ v+\lambda h) - f(x, v))/\lambda \\ - f'_+(x, v; y, h) + \varepsilon/2 < \varepsilon/2 + \varepsilon/2 = \varepsilon, \end{aligned}$$

for any $v \in D$, $h \in R^n$ with $\|h\| = 1$ whenever $0 < \lambda < \delta(\varepsilon)$. Now, as we proved in [4] (Theorem 6.1), this is equivalent to the hypothesis (8) on D, q.e.d.

Therefore the theory of the first two sections can be applied to the function J. The most interesting case is that, when the functions f and g are continuously differentiable on $X \times W(u_0)$. Then the derivatives are uniformly continuous on the compact set $X \times W(u_0)$ and, according to the equivalence which we proved in [4], the directional differential of J satisfies (8) and it is of the form (see [2]):

$$\begin{aligned} J'_+(v, h) = (\mathrm{grad}_v f(x, v), h) + \min \{(\mathrm{grad}_x f(x, v), y)/y \in R^p, \\ (\mathrm{grad}_v g_j(x, v), y) \leqslant -(\mathrm{grad}_x g_j(x, y), h), \textit{ for } x \in M(v) \textit{ and } j \in \\ \in \{1, 2, \ldots, m\} \textit{ for which } g_j(x, v) = 0\}. \end{aligned} \tag{24}$$

Therefore in this case, the problem (4) is a linear programming problem (if we also take $V = \{v \in R^n, |v_i| \leqslant 1,\ i = 1, 2, \ldots, n\}$). The algorithm is especially efficient for nonlinear programming problems of the form

$$\min_{x,y} \{f(x, y)/g(x, y) \leqslant 0,\ h(x, y) = 0\}, \tag{25}$$

which are relatively easy to solve when y is fixed. For example if f, g and h are linear with respect to x, then it is natural to consider the equivalent problem of minimization of

$$J(y) = \min_x \{f(x, y)/g(x, y) \leqslant 0,\ h(x, y) = 0\}.$$

REFERENCES

1. Céa J. — *Optimisation. Théorie et algorithmes*, Dunod, Paris, 1971.
2. Hogan W. — *Directional derivatives for extremal-value functions with applications to the completely convex case*, Operations Research, Vol. 21, Nr. 1, 1973, 188—209.
3. Ilioi C. — *A unified approach of the descent method for minimization of convex functionals. Part I: General theory*, An. şt. Univ., Iaşi, Tom XXIV, 1978, 307—314.
4. Ilioi C. — *A unified approach of the descent methods for minimization of convex functionals. Part II: Steplength algorithms*, to appear in Bull. Math. Soc. Sci. Math. de la R. S. R., Tome 23, Nr. 1, 1979.
5. Levitin E. S. — *A general method of minimization for nonsmooth extremal problems* (in Russian), J. Vîscisl. Mat. i Mat. Fiz., 9, 2, 1969, 783—806.
6. Rockafellar R. T. — *Convex analysis*, Princeton Univ. Press, 1970.

Received 7.V.1979

Faculty of Mathematics
University of Iaşi
Iaşi, R. S. România

Analele ştiinţifice ale Universităţii „Al. I. Cuza" din Iaşi
Tomul XXV, s. I a, 1979, f. 2

CONNEXIONS SUR LES FIBRES SPINORIELS

PAR

MIHAI ANASTASIEI

Dans [1] on a défini et étudié les structures spinorielles sur les variétés hilbertiennes. Dans le présent article on considère les connexions adaptées aux structures spinorielles (les connexions spinorielles). Après quelques résultats relatifs aux connexions sur les fibrés principaux banachique (§1), on définit les connexions spinorielles (§2). En utilisant la dérivée covariante des spineurs, on met en évidence une famille d'opérateurs différentiels du premier ordre (au sens de [7] p. 91) sur les fibrés spinoriels. Si la variété est à dimension finie l'opérateur de Dirac peut être déduit de cette famille des opérateurs.

1. Connexions sur les fibrés principaux banachiques. Soient M une variété différentiable de classe C^∞, modelée sur un espace de Banach $\mathbf{M}$, G un groupe de Lie-Banach réel et $P(M, \pi_P, G)$ un fibré principal (abrégé f.p.) de base M, de groupe structural G et de projection π_P. Nous notons par $R_g : u \to ug$, $u \in P$, $g \in G$ l'action de G sur P et par $\sigma_u : G \to P$, $u \in P$, l'application $g \to ug$, $g \in G$. σ_u est un difféomorphisme de classe C^∞ entre le groupe G et la fibre au-dessus du point $p = \pi_P(u)$. Le foncteur tangent sera noté par T. Considérons la suite exacte de fibrés vectoriels au-dessus de P:

$$0 \to P \times \mathbf{G} \xrightarrow{I} TP \xrightarrow{T\pi_P!} \pi_P^* TM \to 0 \quad \text{où :} \tag{1.1}$$

- $\mathbf{G}$ est l'algèbre de Lie-Banach de G,
- $\pi_P^* TM$ est le fibré image réciproque de TM par π_P,
- $T\pi_P! = (\tau_P, T\pi_P)$, $\tau_P : TP \to P$,
- $I(u, A) = T\sigma_u(A)$, $u \in P$, $A \in \mathbf{G}$.

Un G-fibré vectoriel est un fibré vectoriel sur lequel G opère par des automorphismes de fibré vectoriel. Si nous considérons les actions naturelles de G sur TP et $\pi_P^* TM$ ainsi que l'action de G sur $P \times \mathbf{G}$, $(u, A)g = (ug, \mathrm{ad}(g^{-1})A)$, $u \in P$, $g \in G$, $A \in \mathbf{G}$, alors la suite (1.1) devient une suite exacte de G-fibrés vectoriels.

Définition 1.1. [8] *Une connexion (infinitésimale) sur le f.p. $P(M, \pi_P, G)$ est une scission de la suite exacte* (1.1) *de G-fibrés vectoriels.*