

D. WEXLER

**Prox-mappings associated with a pair of  
Legendre conjugate functions**

*Revue française d'automatique informatique recherche opérationnelle. Mathématique*, tome 7, n° R2 (1973), p. 39-65

[http://www.numdam.org/item?id=M2AN\\_1973\\_\\_7\\_2\\_39\\_0](http://www.numdam.org/item?id=M2AN_1973__7_2_39_0)

© AFCET, 1973, tous droits réservés.

L'accès aux archives de la revue « Revue française d'automatique informatique recherche opérationnelle. Mathématique » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques  
<http://www.numdam.org/>

## PROX-MAPPINGS ASSOCIATED WITH A PAIR OF LEGENDRE CONJUGATE FUNCTIONS <sup>(1)</sup>

par D. WEXLER <sup>(2)</sup>

Résumé. — *Etant donné un couple  $\phi, \psi$  de fonctions duales de Legendre, à tout couple  $f, g$  de fonctions duales de Fenchel on associe les applications  $\text{prox}_f^\phi, \text{prox}_g^\psi$ , généralisant aux espaces de Banach réflexifs les applications prox introduites par Moreau pour les espaces de Hilbert [17]. Les propriétés de continuité de l'application  $\text{prox}_f^\phi$  sont en rapport avec les propriétés de différentiabilité de la fonction  $\psi$ .*

### 1. INTRODUCTION

Let  $F$  and  $G$  be real vector spaces, in duality with respect to the bilinear form  $\langle \cdot, \cdot \rangle$ . Denote by  $\Gamma_0(F)$ , respectively  $\Gamma_0(G)$ , the collection of extended real-valued functions on  $F$ , respectively  $G$ , which are proper, convex and weakly lower semicontinuous. For any  $\Psi \in \Gamma_0(F)$ , any  $f \in \Gamma_0(F)$  and any  $X \in F$ , denote by  $\text{Prox}_f^\phi X$  the (possibly empty) set of minimum points of the function  $u \rightarrow \Phi(X - u) + f(u)$ ; if the set  $\text{Prox}_f^\phi X$  consists of exactly one point, denote this point by  $\text{prox}_f^\phi X$ . Define similarly  $\text{Prox}_g^\psi Y$  and  $\text{prox}_g^\psi Y$  for  $\Psi \in \Gamma_0(G)$ ,  $g \in \Gamma_0(G)$  and  $Y \in G$ . This concept was primarily introduced by Moreau [17] for  $F = G = \text{Hilbert space}$ ,  $\langle \cdot, \cdot \rangle$  as the scalar product and  $\Phi = \Psi = 2^{-1} \|\cdot\|^2$ . He showed that : (i) for any  $f \in \Gamma_0(F)$  and any  $X \in F$ ,  $\text{Prox}_f^\phi X$  consists of just one point and the mapping  $\text{prox}_f^\phi : F \rightarrow F$  has certain regularity properties (it is a contraction, a Fréchet gradient mapping etc.); (ii) the pairs of conjugate points associated with any pair of Fenchel conjugate functions  $f, g$  may be characterized in a simple way in terms of the pair of prox-mappings  $\text{prox}_f^\phi, \text{prox}_g^\psi$ .

Extensions of results of type (i) were stated by Lescarret [15] in certain reflexive normed spaces, taking  $\Phi$  as a composite function of the form  $\varphi \circ \|\cdot\|$ ,

(1) This research was completed while the author was working at the Institute of Mathematics of the Romanian Academy of Sciences, Bucharest.

(2) UER, 47, Université de Paris, VI.

where  $\varphi$  is a suitable function  $\mathbf{R} \rightarrow \mathbf{R}$ . Upper semicontinuity property of multivalued prox-mappings was discussed by Lescarret [16] and other problems concerning prox-mappings were considered by Castaing [9] and Aris [1].

We are here interested in the following problem : find general conditions on  $\Phi$  and  $\Psi$ , such that the pairs of conjugate points with respect to any pair of Fenchel conjugate functions  $f, g$  may be characterized in terms of the pair of (single-valued) proxmappings  $\text{prox}_f^\Phi, \text{prox}_g^\Psi$ , which are at least continuous. We found the following answer to the above stated problem : for reflexive spaces  $F$ , with  $G$  as the strong dual of  $F$ , a natural framework is that of normed reflexive spaces with  $\Phi \in \Gamma_0(F), \Psi \in \Gamma_0(G)$  as any pair of Legendre conjugate functions. We show in §5 that such pairs  $\Phi, \Psi$  exist for any reflexive normed space (consequence of the renorming theorems of Asplund [2] (1)); such pairs  $\Phi, \Psi$ , may arise in a natural way (as in the case of Orlicz spaces) in other forms than those considered earlier by Lescarret [15]. Thus, our results are more general than, but do not quite contain, the corresponding results stated in [15].

The proofs of our results in § 3 (characterisation of conjugate points) consist in an adaptation to our framework of the corresponding results of Moreau [17]; the proofs in § 4 (continuity properties of prox-mappings) make an essential use of the general duality between differentiability and rotundity, introduced by Asplund and Rockafellar [4], as well as of uniform versions (our lemmas 2, 3, 5 and 6) of certain results of [4]; in § 5 we discuss a few examples and relate our results to the projection on convex closed sets and to the theorem of Beurling-Livingstone [5].

## 2. BASIC ASSUMPTIONS

In what follows  $F$  is a reflexive normed space over  $\mathbf{R}$ ;  $F'$  is the dual space normed by the dual norm;  $G$  is any space isomorphic (algebraically and metrically) to  $F'$ ;  $\langle \cdot, \cdot \rangle$  is the canonical bilinear form on  $F \times G$ . Unless otherwise stated, all questions related to the topology of  $F$ , respectively  $G$ , refer to the norm topology.

For the theory of topological vector spaces we refer the reader to [8]. As known, the collection of convex closed sets (as well as that of bounded sets) in  $F$ , respectively  $G$ , is the same in all the topologies compatible with the duality between  $F$  and  $G$  (in particular in the norm topology of  $F$ , respectively, — by the reflexivity of  $F$ , — in the norm topology of  $G$ ). Hence  $\Gamma_0(F)$  and  $\Gamma_0(G)$  do not change if we replace the requirement of lower semicontinuity in

---

(1) We are thankful to Professor R. T. Rockafellar, who pointed out to us a very recent result of Troyanski [21], that enables one to construct pairs of Legendre conjugate functions, which are even Fréchet differentiable.

the weak topology by lower semicontinuity in any topology compatible with the given duality (in particular in the norm topology).

For the general theory of convex functions, Fenchel conjugacy (for short, conjugacy) and subgradients we refer the reader to [18] and [13]. Recall that : if  $f \in \Gamma_0(F)$  is Gâteaux differentiable at  $x$ , then the set  $\partial f(x)$  of subgradients of  $f$  at  $x$  consists of exactly one element, denoted by  $\nabla f(x)$ , which is the Gâteaux gradient of  $f$  at  $x$ ; the pair of points  $x \in F, y \in G$  is said to be conjugate with respect to the pair of conjugate functions  $f \in \Gamma_0(F), g \in \Gamma_0(G)$ , if

$$f(x) + g(y) = \langle x, y \rangle;$$

we have

$$(2.1) \quad f(x) + g(y) = \langle x, y \rangle \Leftrightarrow y \in \partial f(x) \Leftrightarrow x \in \partial g(y).$$

Let  $\mathcal{A}$  and  $\mathcal{B}$  be two collections of nonempty bounded sets in  $F$ , respectively  $G$ , such that  $\bigcup_{A \in \mathcal{A}} A$  generates  $F$  algebraically, respectively  $\bigcup_{B \in \mathcal{B}} B$  generates  $G$  algebraically, and  $A \in \mathcal{A}$  implies  $-A \in \mathcal{A}$ , respectively  $B \in \mathcal{B}$  implies  $-B \in \mathcal{B}$ . Denote by  $\mathcal{S}$ , respectively  $\mathcal{C}$ , the topology induced on  $F$ , respectively  $G$ , by uniform convergence of the linear functionals on sets in  $\mathcal{B}$ , respectively  $\mathcal{A}$ . The topologies  $\mathcal{S}$  on  $F$  and  $\mathcal{C}$  on  $G$  are locally convex, Hausdorff and compatible with the duality between  $F$  and  $G$ . We shall consider functions on  $F$ , respectively  $G$ , which are  $\mathcal{A}$  —, respectively  $\mathcal{B}$  —, differentiable in the sense of [4]. Recall that if  $h \in \Gamma_0(F)$  is  $\mathcal{A}$  — differentiable at  $x$  and  $\nabla h(x)$  is its  $\mathcal{A}$  — gradient, then  $h$  is Gâteaux differentiable at  $x$  and  $\nabla h(x)$  is its Gâteaux gradient. Indeed, the following two special cases present for us the main interest :

1)  $\mathcal{A}$  is the collection of all singleton sets in  $F$ , and  $\mathcal{B}$  is the collection of all singleton sets in  $G$ ; then  $\mathcal{A}$  —, respectively  $\mathcal{B}$  —, differentiability is Gâteaux (or weak) differentiability and  $\mathcal{S}$  respectively  $\mathcal{C}$ , is the weak topology on  $F$ , respectively  $G$ ;

2)  $\mathcal{A}$  is the collection of all bounded sets in  $F$ , and  $\mathcal{B}$  is the collection of all bounded sets in  $G$ ; then  $\mathcal{A}$  —, respectively  $\mathcal{B}$  —, differentiability is equivalent to Fréchet (or strong) differentiability and  $\mathcal{S}$ , respectively  $\mathcal{C}$ , is the norm topology on  $F$ , respectively  $G$ .

We assume that there is given a pair of conjugate functions  $\Phi \in \Gamma_0(F), \Psi \in \Gamma_0(G)$ , which satisfy the following two conditions :

- (A)  $\Phi$  is  $\mathcal{A}$  — differentiable on  $F$  and  $\Psi$  is  $\mathcal{B}$  — differentiable on  $G$ ;
- (B)  $\Phi(O_F) = 0$  and  $\nabla \Phi(O_F) = O_G$ .

Note that condition (B) is not essential : it is easy to verify that if (A) holds, the pair of functions

$$\Phi_1 : F \rightarrow R, \Phi_1(x) = \Phi(x) - \langle x, \nabla \Phi(O_F) \rangle - \Phi(O_F)$$

and its conjugate  $\Psi_1$  satisfy both (A) and (B).

If each function of a pair  $\Phi \in \Gamma_0(F)$ ,  $\Psi \in \Gamma_0(G)$  is the Legendre conjugate (transform) of the other (i.e.  $\Phi$  and  $\Psi$  are Gâteaux differentiable on  $F$ , respectively  $G$ , the mappings  $\nabla\Phi : F \rightarrow G$ ,  $\nabla\Psi : G \rightarrow F$  are bijective and

$$(2.2) \quad \begin{aligned} \Phi(x) &= \langle x, (\nabla\Psi)^{-1}(x) \rangle - \Psi((\nabla\Psi)^{-1}(x)), \quad \forall x \in F, \\ \Psi(y) &= \langle (\nabla\Phi)^{-1}(y), y \rangle - \Phi((\nabla\Phi)^{-1}(y)), \quad \forall y \in G, \end{aligned}$$

it follows by (2.1) that  $\Phi, \Psi$  is a pair of (Fenchel) conjugate functions and our assumptions hold with  $\mathcal{A}$  — and  $\mathcal{B}$  — differentiability as Gâteaux differentiability. On the other hand, if our assumptions hold, then  $\Phi$  and  $\Psi$  are (at least) Gâteaux differentiable, and, by (2.1), the mappings  $\nabla\Phi, \nabla\Psi$  are reciprocal to each other, (2.2) holds and hence each function of the pair  $\Phi, \Psi$  is the Legendre conjugate of the other. Thus, our assumptions are just more specific than the simple requirement that each function of the pair  $\Phi \in \Gamma_0(F)$ ,  $\Psi \in \Gamma_0(G)$  is the Legendre conjugate of the other.

From the results of [4], § 4, it follows that the reciprocal to each other mappings  $\nabla\Phi : F \rightarrow G$  and  $\nabla\Psi : G \rightarrow F$  are continuous from the norm topology to the  $\mathcal{C}$  —, respectively  $\mathcal{S}$  —, topology. In the above case 2),  $\nabla\Phi$  is a homeomorphism,  $\nabla\Psi$  being its reciprocal homeomorphism. Note that according to [4], the Fréchet differentiability of both  $\Phi$  and  $\Psi$  may be expressed in term of requirements on only one of them.

By  $\nabla\Psi(O_G) = O_F$  and (2.2),  $\Psi(O_G) = 0$ . Thus  $\Psi$  also satisfies (B) and the functions  $\Phi$  and  $\Psi$  play a symmetrical part. Therefore the results that follows still hold if we replace  $F, \mathcal{A}, \mathcal{S}$  and  $\Phi$  by  $G, \mathcal{B}, \mathcal{C}$ , respectively  $\Psi$ .

We mention some properties of  $\Phi$ . By Fenchel inequality,

$$\forall x \in F, \Phi(x) + \Psi(O_G) \geq 0,$$

and hence  $\Phi(x) \geq 0, \forall x \in F$ . By (2.1),

$$\Phi(x) = 0 \Leftrightarrow \Phi(x) + \Psi(O_G) = 0 \Leftrightarrow x = \nabla\Psi(O_G),$$

hence  $\Phi(x) = 0 \Leftrightarrow x = O_F$ .

According to [20],  $h \in \Gamma_0(F)$ ,  $\text{int dom } h \neq \emptyset$  and  $\partial h(x_1) \cap \partial h(x_2) = \emptyset, \forall x_1, x_2 \in \text{int dom } h$  with  $x_1 \neq x_2$  implies that  $h$  is strictly convex on  $\text{int dom } h$  (1). Thus  $\Phi$  is strictly convex on  $F$ , since the mapping  $\nabla\Phi$  is one-to-one.

According to [19],  $\text{dom } \Psi = G$  implies that the sets,

$$L(y, \lambda) = \{ u \in F : \Phi(u) - \langle u, y \rangle \leq \lambda \}, \quad y \in G, \lambda \in \mathbf{R},$$

are bounded. Since  $\Phi \in \Gamma_0(F)$ , these sets are weakly closed. In a reflexive space, a weakly closed and bounded set is weakly compact, hence the sets  $L(y, \lambda)$ ,

---

(1) This assertion was proved in [20] for finite-dimensional  $F$ , but the arguments still hold for any locally convex  $F$ .

$y \in G, \lambda \in R$ , are weakly compact. Since  $F$  is a Banach space,  $\text{dom } \Phi = F$  implies that  $\Phi$  is norm continuous on  $F$  (see [19]).

It is worth mentioning (although we do not make explicit use of this fact) that the norm and  $\mathcal{S}$ -topologies on  $F$  can be related to  $\Phi$  : it is easy to see that the function  $|||\cdot|||_F$  defined on  $F$  by

$$|||x|||_F = \sup \{ |\langle x, y \rangle| : y \in G, \Psi(y) \leq 1 \}$$

is a norm on  $F$ , which is equivalent to the initial one  $\|\cdot\|_F$ ; it is also easy to verify that if  $\mathcal{F}$  is a filter in  $F$  such that  $\lim_{\mathcal{F}} \Phi = 0$ , then  $\mathcal{F} \rightarrow O_F$  in the  $\mathcal{S}$ -topology (invoke the  $\mathcal{S}$ -rotundity of  $\Phi$  at  $O_F$  relative to  $O_G$ , which, according to [4], follows from the  $\mathcal{B}$ -differentiability of  $\Psi$  at  $O_G$ , with  $\nabla\Psi(O_G) = O_F$ ).

Note that the simple requirements that  $\Phi$  and  $\Psi$  are conjugate to each other,  $\text{dom } \Phi = F$  and  $\text{dom } \Psi = G$  (all the more condition (A)) and  $F$  reflexive does not allow to go beyond the framework of *normed* spaces. As a matter of fact, if we do not require explicitly the initial reflexive locally convex Hausdorff topology of  $F$  to be a norm topology, the condition  $\text{dom } \Psi = G$  ( $G$  being the strong dual space of  $F$ ) yields as above that the set

$$\{ u \in F : \Phi(u) - \Phi(O_F) \leq 1 \}$$

is bounded, while  $\text{dom } \Phi = F$  yields (since reflexive spaces are barrelled) that  $\Phi$  is continuous on  $F$ . Hence the above set is a bounded neighborhood of  $O_F$ , and, by the normability theorem of Kolmogorov, it follows that the initial topology of  $F$  is a norm topology.

### 3. PROX-MAPPINGS AND CONJUGATE POINTS

For any points  $X \in F, x \in F, y \in G$  and  $\lambda \in R_+$ , consider the sets

$$A(x, y, \lambda) = \{ u \in F : \Phi(x + u) - \Phi(x) - \langle u, y \rangle \leq \lambda \},$$

$$B(X, x, y, \lambda) = x - A(X - x, y, \lambda),$$

$$\partial_\lambda \Psi(y) = \{ u \in F : \Phi(u) - \langle u, y \rangle \leq \lambda - \Psi(y) \}.$$

By simple calculations we obtain

$$(3.1) \quad A(x, y, \lambda) = L(y, \lambda + \Phi(x) - \langle x, y \rangle) - x,$$

$$(3.2) \quad B(X, x, y, \lambda) = X - L(y, \lambda + \Phi(X - x) - \langle X - x, y \rangle),$$

$$(3.3) \quad \partial_\lambda \Psi(y) = L(y, \lambda - \Psi(y)),$$

$$(3.4) \quad A(x, \nabla\Phi(x), \lambda) = \partial_\lambda \Psi(\nabla\Phi(x)) - x.$$

Since  $O_F \in A(x, y, \lambda)$ ,  $x \in B(X, x, y, \lambda)$  and  $\nabla \Psi(y) \in \partial_\lambda \Psi(y)$ , the sets  $A(x, y, \lambda)$ ,  $B(X, x, y, \lambda)$  and  $\partial_\lambda \Psi(y)$  are nonempty; by the properties of  $L$  mentioned in § 2, they are also convex and weakly compact. Note that for any fixed  $X, x$  and  $y$  the above sets are increasing with respect to  $\lambda \in \mathbf{R}_+$ .

In what follows,  $f \in \Gamma_0(F)$ ,  $g \in \Gamma_0(G)$  is an arbitrary pair of conjugate functions. For any  $X \in F$ , respectively  $Y \in G$ , we define the function  $\Phi_X$  respectively  $\Psi_Y$ , from  $F$ , respectively  $G$ , into  $\mathbf{R} \cup \{+\infty\}$  by

$$\Phi_X(u) = \Phi(X - u) + f(u), \quad \text{respectively,} \quad \Psi_Y(v) = \Psi(Y - v) + g(v).$$

**Lemma 1.** *For any points  $X \in F$ ,  $x_0 \in \text{dom } f$ ,  $y_0 \in G$  and  $\lambda \geq 0$ , we have  $\Phi_X(u) > \Phi_X(x_0) + \lambda - [f(x_0) + g(y_0) - \langle x_0, y_0 \rangle]$ ,  $\forall u \notin B(X, x_0, y_0, \lambda)$ .*

*Proof.* If either  $f(u) = +\infty$  or  $g(y_0) = +\infty$ , the above assertion holds obviously.

Assume  $u \in \text{dom } f$ ,  $y_0 \in \text{dom } g$  and put  $u_1 = x_0 - u$ . Since

$$u_1 \notin A(X - x_0, y_0, \lambda)$$

we have

$$\Phi(X - x_0 + u_1) > \lambda + \Phi(X - x_0) + \langle u_1, y_0 \rangle$$

and thus

$$(3.5) \quad \Phi_X(u) > \lambda + \Phi(X - x_0) + \langle x_0, y_0 \rangle - \langle u, y_0 \rangle + f(u).$$

By Fenchel inequality,

$$(3.6) \quad f(u) - \langle u, y_0 \rangle \geq -g(y_0).$$

Substituting (3.6) in (3.5) we complete the proof of the lemma.

**Theorem 1.** *For any  $X \in F$ , the function  $\Phi_X$  has a unique minimum point on  $F$  (denote it by  $\text{prox}_f^{\Phi}$  or  $\text{prox}_f X$ ). Furthermore, if  $x_0 \in \text{dom } f$ ,  $y_0 \in \text{dom } g$  and*

$$\lambda \geq f(x_0) + g(y_0) - \langle x_0, y_0 \rangle,$$

then

$$(3.7) \quad \text{prox}_f X \in B(X, x_0, y_0, \lambda).$$

*Dual assertions hold for  $\Psi_Y$ .*

*Proof.* Since the conjugate functions  $f$  and  $g$  are proper, there exist  $x_0 \in \text{dom } f$  and  $y_0 \in \text{dom } g$ . For

$$\lambda_0 = f(x_0) + g(y_0) - \langle x_0, y_0 \rangle$$

we have  $0 \leq \lambda_0 < +\infty$  and, by lemma 1,

$$(3.8) \quad \Phi_X(u) > \Phi(x_0), \quad \forall u \notin B(X, x_0, y_0, \lambda_0).$$

Since  $\Phi_X$  is weakly lower semicontinuous on the nonempty and weakly compact set  $B(X, x_0, y_0, \lambda_0)$ , it has a minimum point  $x$  on this set. Then  $\Phi_X(x_0) \geq \Phi_X(x)$ , since  $x_0 \in B(X, x_0, y_0, \lambda_0)$ . Thus, by (3.8),  $\Phi_X(u) > \Phi_X(x)$ ,  $\forall u \notin B(X, x_0, y_0, \lambda_0)$ . Hence  $x$  is a minimum point of  $\Phi_X$  on all of  $F$ . The uniqueness of the minimum point follows immediately from the strict convexity of  $\Phi$ .

As was shown,  $x \in B(X, x_0, y_0, \lambda_0)$ , and, all the more  $x \in B(X, x_0, y_0, \lambda)$ , for any  $\lambda > \lambda_0$ . Thus (3.7) holds.

REMARK 1. In order to prove the existence of a minimum point, we could refer the reader to Moreau [18], section 7, since  $\inf_{u \in F} \Phi_X(u)$  is the value of the inf-convolution  $\Phi \square f$  at  $X$ . We gave a direct proof because in § 4 we shall make use of lemma 1 and (3.7).

EXAMPLES 1. If  $f$  is Gâteaux differentiable on  $F$ , then the mapping  $Id_F + \nabla \Psi \circ \nabla f$  ( $Id_F$  is the identity on  $F$ ) on  $F$  is bijective and

$$(3.8 a) \quad \text{prox}_f X = [Id_F + \nabla \Psi \circ \nabla f]^{-1}(X), \quad \forall X \in F.$$

As a matter of fact, since  $\text{prox}_f X$  is the unique minimum point of the Gâteaux differentiable function  $\Phi_X$  and  $\Phi_X \in \Gamma_0(F)$ , the equation  $\nabla \Phi_X(u) = 0_G$  has for any  $X \in F$  a unique solution, which is just  $\text{prox}_f X$ . This equation is equivalent to

$$[Id_F + \nabla \Psi \circ \nabla f]u = X.$$

2. Taking  $f = \Phi$  in example 1 we obtain  $\text{prox}_\Phi X = 2^{-1}X$ ,  $\forall X \in F$ .

3. Assume that  $f = \langle \cdot, y \rangle$ , with  $y \in G$ . Then by example 1, the mapping  $\text{prox}_f$  is the translation by  $-\nabla \Psi(y)$  :

$$\text{prox}_f X = X - \nabla \Psi(y), \quad \forall X \in F.$$

**Proposition 1.** For any  $X \in F$ , the points  $\text{prox}_f X$  and  $\nabla \Phi(X - \text{prox}_f X)$  are conjugate with respect to the pair of conjugate functions  $f, g$ .

*Proof.* Put  $x = \text{prox}_f X$  and  $y = \nabla \Phi(X - x)$ . Since  $x$  is a minimum point of  $\Phi_X$ , we have

$$\Phi_X(\lambda u + (1 - \lambda)x) \geq \Phi_X(x), \quad \forall u \in F, \quad \forall \lambda \in [0, 1].$$

Taking into account the convexity of  $f$ , we obtain

$$\Phi(X - x + \lambda(x - u)) - \Phi(X - x) \geq \lambda(f(x) - f(u)), \quad \forall u \in F, \forall \lambda \in [0, 1]$$

and thus

$$(3.9) \quad \lambda^{-1}[\Phi(X - x + \lambda(x - u)) - \Phi(X - x)] \\ \geq f(x) - f(u), \quad \forall u \in F, \forall \lambda \in ]0, 1].$$



Since  $\Phi$  is Gâteaux differentiable at  $X - x$ , with  $\nabla \Phi(X - x) = y$ , (3.9) yields

$$(3.9 a) \quad f(u) \geq f(x) + \langle u - x, y \rangle, \quad \forall u \in F$$

and that means  $y \in \partial f(x)$ .

(3.9 a) shows that proposition 1 expresses in fact the variational inequality associated with the problem of minimizing the function  $\Phi_x$ .

The theorem below gives a full characterisation by means of prox-mappings of pairs of points which are conjugate with respect to the pair of conjugate functions  $f, g$ .

**Theorem 2.** *For any two pairs of points  $(x, y), (X, Y)$  from  $F \times G$  the following three assertions are equivalent :*

- (i)  $f(x) + g(y) = \langle x, y \rangle$ , and  $x + \nabla \Psi(y) = X, y + \nabla \Phi(x) = Y$  ;
- (ii)  $x = \text{prox}_f X, y = \text{prox}_g Y$ , and  $Y = \nabla \Phi(x) + \nabla \Phi(X - x)$  ;
- (iii)  $x = \text{prox}_f X, y = \text{prox}_g Y$ , and  $X = \nabla \Psi(y) + \nabla \Psi(Y - y)$ .

*Proof.* (i)  $\Rightarrow$  (ii). We have

$$\Phi_x(u) - \Phi_x(x) = \Phi(X - u) - \Phi(X - x) + f(u) - f(x), \quad \forall u \in F.$$

By the Fenchel inequality,  $f(u) \geq \langle u, y \rangle - g(y)$ ,  $\forall u \in F$ , and thus, by (i)

$$f(u) - f(x) \geq \langle u - x, y \rangle, \quad \forall u \in F.$$

Hence

$$(3.10) \quad \Phi_x(u) - \Phi_x(x) \geq \Phi(X - u) - \Phi(X - x) + \langle u - x, y \rangle ; \quad \forall u \in F.$$

By simple calculations

$$\inf_{u \in F} (\Phi(X - u) + \langle u - x, y \rangle) = \langle X - x, y \rangle$$

$$- \sup_{u \in F} (\langle X - u, y \rangle - \Phi(X - u)) = \langle X - x, y \rangle - \Psi(y) = \Phi(X - x).$$

We used the fact that, according to (i), the pair  $X - x, y$  is conjugate with respect to the pair of conjugate functions  $\Phi, \Psi$ . Hence, by (3.10),  $\Phi_x(u) - \Phi_x(x) \geq 0$ ,  $\forall u \in F$ , and thus  $x = \text{prox}_f X$ .

By means of dual calculations (for  $\Psi_y$ ) we obtain  $y = \text{prox}_g Y$ .

By (i),  $X = x + \nabla \Psi(Y - \nabla \Phi(x))$ , hence  $Y = \nabla \Phi(x) + \nabla \Phi(X - x)$ .

For the implication (i)  $\Rightarrow$  (iii) we have only to note that dual calculation to the above one yields  $X = \nabla \Psi(y) + \nabla \Phi(Y - y)$ .

(ii)  $\Rightarrow$  (i). Put  $y_1 = \nabla \Phi(X - x)$ . By proposition 1,

$$(3.11) \quad f(x) + g(y_1) = \langle x, y_1 \rangle .$$

Hence, by the implication (i)  $\Rightarrow$  (ii) proved above,

$$(3.12) \quad x = \text{prox}_f X_1 \quad , \quad y_1 = \text{prox}_g Y_1,$$

where

$$(3.13) \quad X_1 = x + \nabla \Psi(y_1) \quad , \quad Y_1 = y_1 + \nabla \Phi(x)$$

and

$$(3.14) \quad Y_1 = \nabla \Phi(x) + \nabla \Phi(X_1 - x).$$

By (3.13) and  $\nabla \Psi = (\nabla \Phi)^{-1}$ ,  $X_1 = X$ . Then by (ii) and (3.14),  $Y_1 = Y$ . Hence, by (3.12)  $y_1 = y$ , and (3.11), (3.13) show that (i) holds.

Dual calculations yield the implication (iii)  $\Rightarrow$  (i).

As in [17], the above theorem yields some corollaries concerning the range of prox — mappings.

**Corollary 1.** *For any  $a \in F$ ,*

$$\text{prox}^{-1}(a) = a + \nabla \Psi(\partial f(a)).$$

*(In particular this set is empty if and only if  $\partial f(a) = \emptyset$ .)*

*Proof.* If  $X \in a + \nabla \Psi(\partial f(a))$ , there is  $y \in \partial f(a)$  such that  $X = a + \nabla \Psi(y)$ . Hence, by theorem 2,  $a = \text{prox}_f X$ .

Assume now that  $X \in F$  is such that  $\text{prox}_f X = a$ . Put  $y = \nabla \Phi(X - a)$ . By proposition 1,  $y \in \partial f(a)$  and, by  $(\nabla \Phi)^{-1} = \nabla \Psi$ ,  $X = a + \nabla \Psi(y)$ .

From corollary 1 follows immediately

**Corollary 2.**  *$\text{Prox}_f F$  is the set of points at which  $f$  is subdifferentiable.*

**Corollary 3.** *The set of minimum points of  $f$  coincides with the set of fixed-points of the mapping  $\text{prox}_f$ .*

*Proof.* Let  $x$  be a minimum point of  $f$ . Hence  $x \in \partial g(O_G)$ . By theorem 2,  $x = \text{prox}_f X$ , where  $X = x + \nabla \Psi(O_G) = x$ , and thus  $x = \text{prox}_f x$ .

Assume now  $x = \text{prox}_f x$ . By proposition 1, it follows  $x \in \partial g(O_G)$  and thus  $x$  is a minimum point of  $f$ .

REMARK 2. Corollaries 2 and 3 show that the set  $\text{prox}_f F$  and the set of fixed-points of  $\text{prox}_f$  do not depend on  $\Phi$ .

#### 4. CONTINUITY PROPERTIES

We shall now investigate some continuity properties of prox-mappings under various assumptions on  $\Phi$  and  $\Psi$ . As in § 3,  $f \in \Gamma_0(F)$ ,  $g \in \Gamma_0(G)$  is an arbitrary pair of conjugate functions.

**Theorem 3.** *The mapping  $\text{prox}_f : F \rightarrow F$  is continuous from the norm topology to the  $\mathcal{S}$  — topology.*

*Proof.* Let  $X$  be any point in  $F$  and  $V$  any  $\mathcal{S}$ -neighborhood of  $O_F$ . Since  $\mathcal{S}$  is a locally convex topology, there exists a convex  $\mathcal{S}$ -neighborhood  $W$  of  $O_F$ , with  $W \subset V$ .

Put  $x = \text{prox}_f X$  and  $y = \nabla \Phi(X - x)$ . Since  $\Psi$  is  $\mathcal{B}$ -differentiable at  $y$  with  $\nabla \Psi(y) = X - x$ , it follows by [4] that  $\Phi$  is  $\mathcal{S}$ -rotund at  $X - x$ , with respect to  $y$ . Hence, there is  $\delta > 0$  such that

$$(4.1) \quad A(X - x, y, \delta) \subset -2^{-1}W.$$

Since  $\Phi$  is norm continuous at  $X - x$ , there is  $\eta > 0$  such that

$$\Phi(X - x + u) \leq \delta/4 + \Phi(X - x) \quad \text{and} \quad \langle u, y \rangle \leq \delta/4,$$

for any  $u$  in the ball  $\Sigma(O_F, \eta)$  with centre  $O_F$  and radius  $\eta$ . Indeed, we can choose  $\eta$  so that

$$(4.2) \quad \Sigma(O_F, \eta) \subset 2^{-1}W$$

holds too. Then, taking into account that the pair  $X - x, y$  is conjugate with respect to the pair of conjugate functions  $\Phi, \Psi$ ,

$$(4.3) \quad \Phi(X - x + u) - \langle X - x + u, y \rangle \leq \delta/2 - \Psi(y), \quad \forall u \in \Sigma(O_F, \eta).$$

Let  $u$  be any point in  $\Sigma(O_F, \eta)$ . By proposition 1,  $y \in \partial f(x)$ . Hence, by theorem 1,

$$(4.4) \quad \text{prox}_f(X + u) \in B(X + u, x, y, \delta/2).$$

By (3.2), (4.3) and (3.3),

$$B(X + u, x, y, \delta/2) \subset X + \Sigma(O_F, \eta) - \partial_\delta \Psi(y).$$

Hence, by (3.4) and (4.4),

$$(4.5) \quad \text{prox}_f(X + u) \in \text{prox}_f X + \Sigma(O_F, \eta) - A(X - x, y, \delta), \quad \forall u \in \Sigma(O_F, \eta).$$

It follows by (4.1), (4.2), the convexity of  $W$  and  $W \subset V$  that

$$\text{prox}_f(X + u) \in \text{prox}_f X + V, \quad \forall u \in \Sigma(O_F, \eta).$$

**REMARK 3.** For any  $X \in F$ , there is a norm neighborhood of  $X$ , on which the mapping  $\text{prox}_f$  is bounded (invoke (4.5) and the fact that, as noted in 3 §, the set  $A(X - x, y, \delta)$  is bounded).

**Proposition 2.** *If  $\Phi$  is majorized on any bounded set in  $F$ , then  $\text{prox}_f$  is bounded on any bounded set in  $F$ .*

*Proof.* Let  $K$  be any bounded set in  $F$ . Choose  $x_0 \in \text{dom } f$ ,  $y_0 \in \text{dom } g$  and put

$$\lambda_0 = f(x_0) + g(y_0) - \langle x_0, y_0 \rangle .$$

By theorem 1,  $\text{prox}_f X \in B(X, x_0, y_0, \lambda_0)$ ,  $\forall X \in K$ . Since  $\Phi$  is majorized on  $K - x_0$ ,

$$k = \sup_{X \in K} \{ \lambda_0 + \Phi(X - x_0) - \langle X - x_0, y_0 \rangle \} < + \infty .$$

Thus, by (3.2),  $B(X, x_0, y_0, \lambda_0) \subset X - L(y_0, k)$ ,  $\forall X \in K$ , and it follows

$$\text{prox}_f X \in K - L(y_0, k), \quad \forall X \in K .$$

Since  $L(y_0, k)$  is bounded,  $\text{prox}_f$  is bounded on  $K$ .

**Lemma 2.** *For any function  $h \in \Gamma_0(F)$ , with  $\text{dom } h = F$ , the following five assertions are equivalent :*

- (i)  *$h$  is majorized on any bounded set in  $F$  ;*
- (ii) *for any  $\mu > 0$  and any bounded set  $K$  in  $F$ , the set  $\bigcup_{x \in K, \lambda \leq \mu} \partial_\lambda h(x)$  is bounded in  $G$  ;*
- (iii) *for any bounded set  $K$  in  $F$  the set  $\bigcup_{x \in K} \partial h(x)$  is bounded in  $G$  ;*
- (iv)  *$h$  is Lipschitz on any bounded set in  $F$  ;*
- (v)  *$h$  is uniformly norm continuous on any bounded set in  $F$ .*

*Proof.* The implications (ii)  $\Rightarrow$  (iii), (iv)  $\Rightarrow$  (i) and (iv)  $\Rightarrow$  (v) hold obviously,

(i)  $\Rightarrow$  (ii). Since  $h$  is weakly lower semicontinuous on  $F$  and  $h(x) > -\infty \forall x \in F$ ,  $h$  has a lower bound on any weakly compact set in  $F$ . In a reflexive space, any bounded set is weakly relatively compact. Hence  $h$  has a lower bound on any bounded set in  $F$  and thus  $h$  is bounded on any bounded set in  $F$ .

Let  $K$  be any bounded set in  $F$ . Denote by  $l$  the function conjugate to  $h$ . Since  $l \in \Gamma_0(G)$ , there is  $y_0 \in \text{dom } l$ . Put  $W = \Sigma(O_F, 1)$  and

$$\mu_1 = \sup_{(x,u) \in K \times W} \{ h(x + u) - h(x) - \langle u, y_0 \rangle \} .$$

Since  $h$  is bounded on any bounded set in  $F$ ,  $0 \leq \mu_1 < +\infty$ . Put  $r = \max(\mu, \mu_1)$  and  $V = r^{-1}W$ . It follows

$$V \subset r^{-1} \{ u \in F : h(x + u) - h(x) - \langle u, y_0 \rangle \leq r \}, \quad \forall x \in K .$$

Let  $x$  be any point in  $K$  and put

$$\alpha = \langle x, y_0 \rangle - h(x) - l(y_0).$$

According to [4], proposition 3,

$$\begin{aligned} r^{-1} \{ u \in F : h(x+u) - h(x) - \langle u, y_0 \rangle \leq r \} \\ \subset 2 \{ v \in G : l(y_0+v) - l(y_0) - \langle x, v \rangle \leq r + \alpha \}^0 \\ = 2[\partial_r h(x) - y_0]^0. \end{aligned}$$

It follows  $V \subset 2[\partial_r h(x) - y_0]^0$ ,  $\forall x \in K$ . Taking polars,

$$\partial_r h(x) \subset y_0 + 2V^0 = y_0 + \Sigma(O_G, 2r), \quad \forall x \in K,$$

and thus (ii) holds.

(iii)  $\Rightarrow$  (iv) Let  $K$  be any bounded set in  $F$ . By (iii),

$$m = \sup \{ \|y\|_G : y \in \partial h(x), x \in K \} < +\infty.$$

Let  $x_1, x_2$  be any points in  $K$ . By  $\text{dom } h = F$ ,  $\partial h(x_1) \neq \emptyset$  and  $\partial h(x_2) \neq \emptyset$ . Choose  $y_1 \in \partial h(x_1)$  and  $y_2 \in \partial h(x_2)$ . Hence

$$h(x_1+u) \geq h(x_1) + \langle u, y_1 \rangle, \quad h(x_2+u) \geq h(x_2) + \langle u, y_2 \rangle, \quad \forall u \in F.$$

For  $u = x_2 - x_1$  and  $u = x_1 - x_2$  it follows

$$h(x_2) - h(x_1) \geq \langle x_2 - x_1, y_1 \rangle, \quad h(x_1) - h(x_2) \geq \langle x_1 - x_2, y_2 \rangle.$$

Hence

$$h(x_2) - h(x_1) \geq -m \|x_2 - x_1\|_F \quad \text{and} \quad h(x_2) - h(x_1) \leq m \|x_2 - x_1\|_F$$

Thus  $|h(x_2) - h(x_1)| \leq m \|x_2 - x_1\|_F$ ,  $\forall x \in K$  and (iv) holds.

(v)  $\Rightarrow$  (iii). Let  $K$  be any bounded set in  $F$ . Choose  $r > 0$  and  $y_0 \in \text{dom } l$ . Since  $h$  is uniformly norm continuous on any bounded set in  $F$ , there is  $\rho > 0$  such that

$$h(x+u) - h(x) \leq r/2, \quad \forall x \in K, \quad u \in \Sigma(O_F, \rho).$$

Put  $W = \Sigma(O_F, \rho)$ . Indeed, we can choose  $\rho$  such that  $\langle u, y_0 \rangle \leq r/2$ ,  $\forall u \in W$ , holds too. Put  $V = r^{-1}W$ . It follows

$$V \subset r^{-1} \{ u \in F : h(x+u) - h(x) - \langle u, y_0 \rangle \leq r \}, \quad \forall x \in K.$$

As in the final part of the proof of the implication (i)  $\Rightarrow$  (ii), we obtain

$$\partial_r h(x) \subset y_0 + 2V^0 = y_0 + \Sigma(O_G, 2r\rho^{-1}), \quad \forall x \in K$$

and thus (iii) holds. This completes the proof of lemma 2.

We shall say that the function  $h \in \Gamma_0(F)$  is  $\mathcal{A}$  — differentiable uniformly on the nonempty set  $K$  in  $F$ , if  $h$  is Gâteaux differentiable on  $K$  and, for any  $A \in \mathcal{A}$ ,

$$\lim_{\lambda \downarrow 0} \sup_{(u,x) \in A \times K} \left| \frac{h(x + \lambda u) - h(x)}{\lambda} - \langle x, \nabla h(x) \rangle \right| = 0.$$

**Lemma 3.** *If  $h$  is  $\mathcal{A}$  — differentiable uniformly on  $K$ , then for any  $\mathcal{C}$  — neighborhood  $B$  of  $O_G$  there exists  $\delta > 0$  such that*

$$(4.5 \text{ a}) \quad \{v \in G : l(\nabla h(x) + v) - l(\nabla h(x)) - \langle x, v \rangle \leq \delta\} \subset B, \quad \forall x \in K,$$

where  $l$  is the function conjugate to  $h$ .

The duality between uniform  $\mathcal{A}$ -differentiability and « uniform  $\mathcal{C}$ -rotundity » stated in lemma 3 can be proved in the same way as the duality between  $\mathcal{A}$ -differentiability and  $\mathcal{C}$ -rotundity, see [4].

We can now prove a uniform version of theorem 3.

**Theorem 4.** *Assume that  $\Phi$  is majorized on any bounded set in  $F$  and  $\Psi$  is  $\mathcal{B}$ -differentiable uniformly on any bounded set in  $G$ . Then, on any bounded set in  $F$ , the mapping  $\text{prox}_f : F \rightarrow F$  is uniformly continuous from the norm topology to the  $\mathcal{S}$ -topology.*

*Proof.* Let  $K$  be any nonempty bounded set in  $F$  and  $V$  any  $\mathcal{S}$ -neighborhood of  $O_F$ . There exists a convex  $\mathcal{S}$ -neighborhood  $W$  such that  $W \subset V$ .

By proposition 2, the set  $K_0 = \{X - \text{prox}_f X : X \in K\}$  in  $F$  is bounded and hence, by lemma 2, the set  $\nabla \Phi(K_0)$  in  $G$  is also bounded. Thus  $\Psi$  is  $\mathcal{B}$ -differentiable uniformly on  $\nabla \Phi(K_0)$ . It follows by lemma 3 that there exists  $\delta > 0$  such that, for any  $y \in \nabla \Phi(K_0)$

$$\{u \in F : \Phi(\nabla \Psi(y) + u) - \Phi(\nabla \Psi(y)) - \langle u, y \rangle \leq \delta\} \subset -2^{-1}W.$$

Hence

$$(4.6) \quad A(X - \text{prox}_f X, \nabla \Phi(X - \text{prox}_f X), \delta) \subset -2^{-1}W, \quad \forall X \in K.$$

Since, by lemma 2,  $\Phi$  is uniformly continuous on any bounded set in  $F$ , there exists  $\eta > 0$  such that

$$\Phi(z + u) - \Phi(z) \leq \delta/4, \langle u, y \rangle \leq \delta/4, \forall z \in K_0, \forall y \in \nabla \Phi(K_0), \forall u \in \Sigma(O_F, \eta).$$

Indeed, we can choose  $\eta$  such that (4.2) holds too. Then, using the fact that the points  $X - \text{prox}_f X, \nabla \Phi(X - \text{prox}_f X)$  are conjugate with respect to the pair of conjugate functions  $\Phi, \Psi$ ,

$$(4.7) \quad \begin{aligned} & \Phi(X - \text{prox}_f X + u) - \langle X - \text{prox}_f X + u, \nabla \Phi(X - \text{prox}_f X) \rangle \\ & \leq \delta/2 - \Psi(\nabla \Phi(X - \text{prox}_f X)), \quad \forall X \in K, \quad \forall u \in \Sigma(O_F, \eta). \end{aligned}$$

Let now  $X$  be any point in  $K$  and  $u$  any point in  $\Sigma(O_F, \eta)$ . Put  $x = \text{prox}_f X$  and  $y = \nabla \Phi(X - x)$ . As in the final part of the proof of theorem 3 we obtain

$$\text{prox}_f(X + u) \in \text{prox}_f X + V, \forall X \in K, \forall u \in \Sigma(O_F, \eta);$$

we have only to use (4.7) and (4.6) instead of (4.3), respectively (4.1).

**Lemma 4.** *Assume that on any bounded set in  $F$  the function  $\Phi$  is majorised and, on any bounded set in  $G$ , the mapping  $\nabla \Psi$  is Lipschitz of order  $\beta$ ,  $0 < \beta \leq 1$ . Let  $K_0$  be any nonempty bounded set in  $F$ ,  $\lambda_0$  any real number,  $\lambda_0 > 0$ , and put  $M_0 = \nabla \Phi(K_0) + \Sigma(O_G, \lambda_0)$ ,*

$$m_0 = \sup \left\{ \frac{\|\nabla \Psi(v_2) - \nabla \Psi(v_1)\|_F}{\|v_2 - v_1\|_G} : (v_1, v_2) \in M_0 \times M_0, v_1 \neq v_2 \right\} + 1.$$

Then  $1 < m_0 < +\infty$  and

$$(4.7 a) \quad A(x, \nabla \Phi(x), m_0 \lambda^{\beta+1}) \subset \Sigma(O_F, 2m_0 \lambda^\beta), \forall x \in K_0, \forall \lambda \in ]0, \lambda_0].$$

*Proof.* By lemma 2, the set  $M_0$  in  $G$  is bounded and hence  $m_0 < +\infty$ . Since  $\nabla \Psi$  is one-to-one,  $m_0 > 1$ .

Let  $x$  be any point in  $K_0$ ,  $\lambda$  any point in  $]0, \lambda_0]$  and put  $y = \nabla \Phi(x)$ . Let  $v$  be any point in  $\Sigma(O_G, 1)$ . By the Lagrange formula, there exists  $\theta, 0 < \theta < 1$ , such that

$$\Psi(y + \lambda v) - \Psi(y) = \langle \nabla \Psi(y + \theta \lambda v), \lambda v \rangle.$$

Hence, by  $x = \nabla \Psi(y)$ ,

$$\begin{aligned} \Psi(y + \lambda v) - \Psi(y) - \langle x, \lambda v \rangle &= \langle \nabla \Psi(y + \theta \lambda v) - x, \lambda v \rangle \\ &\leq \|\nabla \Psi(y + \theta \lambda v) - \nabla \Psi(y)\|_F \|\lambda v\|_G \leq m_0 \lambda^{\beta+1}. \end{aligned}$$

It follows

$$\begin{aligned} \Sigma(O_G, (m_0 \lambda^{\beta+1})^{-1}) &\subset (m_0 \lambda^{\beta+1})^{-1} \{ w \in G : \Psi(y + w) \\ &\quad - \Psi(y) - \langle x, w \rangle \leq m_0 \lambda^{\beta+1} \}. \end{aligned}$$

By [4], Proposition 3 (with  $\alpha = 0$ , since  $y = \nabla \Phi(x)$ ),

$$\begin{aligned} (m_0 \lambda^{\beta+1})^{-1} \{ w \in G : \Psi(y + w) - \Psi(y) - \langle x, w \rangle \leq m_0 \lambda^{\beta+1} \} \\ \subset 2 \{ u \in F : \Phi(x + u) - \Phi(x) - \langle u, y \rangle \leq m_0 \lambda^{\beta+1} \}^0 \end{aligned}$$

and hence  $\Sigma(O_G, (m_0 \lambda^{\beta+1})^{-1}) \subset 2 A(x, \nabla \Phi(x), m_0 \lambda^{\beta+1})^0$ . Taking polars, we obtain (4.7 a).

We can now prove a Lipschitz version of theorem 3.

**Proposition 3.** *Assume that on any bounded set in  $F$  the function  $\Phi$  is majorised and, on any bounded set in  $G$  the mapping  $\Delta \Psi$  is Lipschitz of order*

$\beta$ ,  $0 < \beta \leq 1$ . Then on any bounded set in  $F$  the mapping  $\text{prox}_f$  is Lipschitz of order  $\beta/(\beta + 1)$ .

*Proof.* Let  $K$  be any bounded set in  $F$  containing at least two points. By proposition 2, there exists a bounded set  $U \subset F$  such that

$$(4.8) \quad \text{prox}_f^u X \in U, \quad \forall X \in K.$$

Since the set  $K - U$  is bounded, by lemma 2, there exists  $m > 0$  such that

$$(4.9) \quad |\Phi(u_2) - \Phi(u_1)| \leq m \|u_2 - u_1\|_F, \quad \forall u_1 \in K - U, \quad \forall u_2 \in K - U.$$

Put

$$\lambda_0 = \sup \{ (2m \|X_2 - X_1\|_F)^{1/(\beta+1)} : (X_1, X_2) \in K \times K \},$$

hence  $0 < \lambda_0 < +\infty$ . By lemma 4 with  $K_0 = K - \text{prox}_f K$ ,  $1 < m_0 < +\infty$  and

$$(4.10) \quad A(X - \text{prox}_f X, \nabla \Phi(X - \text{prox}_f X), m_0 \lambda^{\beta+1}) \subset \Sigma(O_F, 2m_0 \lambda^\beta), \quad \forall X \in K, \quad \forall \lambda \in ]0, \lambda_0].$$

Let now  $X_1, X_2$  be any two points in  $K$ . Put  $x_1 = \text{prox}_f X_1$ ,  $x_2 = \text{prox}_f X_2$ ,  $y_1 = \nabla \Phi(X_1 - x_1)$  and  $y_2 = \nabla \Phi(X_2 - x_2)$ . Choose

$$\lambda = (2m m_0^{-1} \|X_2 - X_1\|_F)^{1/(\beta+1)}.$$

Since  $m_0 > 1$ ,  $\lambda \in ]0, \lambda_0[$ . Put  $\delta = m_0 \lambda^{\beta+1}$  and  $\varepsilon = 2m_0 \lambda^\beta$ , hence

$$\delta = 2m \|X_2 - X_1\|_F \text{ and } \varepsilon = 2m_0 (2m m_0^{-1})^{\beta/(\beta+1)} \|X_2 - X_1\|_F^{\beta/(\beta+1)}$$

By (4.9),

$$(4.11) \quad |\Phi(X_2 - u) - \Phi(X_1 - u)| \leq \delta/2, \quad \forall u \in U$$

and by (4.10),

$$(4.12) \quad A(X_1 - x_1, y_1, \delta) \subset \Sigma(O_F, \varepsilon).$$

By proposition 1,  $y_1 \in \partial f(x_1)$ . Hence, by lemma 1,

$$(4.13) \quad \Phi_{X_1}(u) > \Phi_{X_1}(x_1) + \delta, \quad \forall u \notin B(X_1, x_1, y_1, \delta).$$

By (4.8),  $x_1 \in U$  and thus, by (4.11),

$$(4.14) \quad \Phi_{X_2}(x_1) \leq \Phi_{X_1}(x_1) + \delta/2.$$

By (4.11),  $\Phi_{X_2}(u) \geq \Phi_{X_1}(u) - \delta/2$ ,  $\forall u \in U$  and hence, by (4.13),

$$(4.15) \quad \Phi_{X_2}(u) > \Phi_{X_1}(x_1) + \delta/2, \quad \forall u \in U \setminus B(X_1, x_1, y_1, \delta).$$



Since  $x_2$  is the unique minimum point of  $\Phi_{x_2}$  on  $U$ , we deduce by  $x_1 \in U$ , (4.14) and (4.15) that  $x_2 \notin U \setminus B(X_1, x_1, y_1, \delta)$ . Hence  $x_2 \in B(X_1, x_1, y_1, \delta)$  and thus

$$x_2 \in x_1 - A(X_1 - x_1, y_1, \delta).$$

It follows by (4.12) and the expression of  $\varepsilon$  and  $\delta$  that for any  $X_1$  and any  $X_2$  in  $K$ ,

$$\|\text{prox}_f X_2 - \text{prox}_f X_1\|_F \leq 2m_0(2mm_0^{-1})^{\beta/(\beta+1)} \|X_2 - X_1\|_F^{\beta/(\beta+1)}.$$

We shall now investigate the function  $\tilde{f} : F \rightarrow \mathbf{R}$ ,  $\tilde{f}(X) = \Phi_X(\text{prox}_f X)$ , associated with  $f$ . Since

$$(4.15 a) \quad \tilde{f}(X) = \inf \{ \Phi(X - u) + f(u) : u \in F \},$$

$\tilde{f}$  is the inf-convolution  $\Phi \square f$  of  $\Phi$  and  $f$ .

**Theorem 5.** *The function  $\tilde{f}$  is finite, convex, norm continuous on  $F$  and its conjugate  $\tilde{f}^*$  is the function  $\Psi + g$ . Furthermore, if  $\Psi$  is Fréchet differentiable on  $G$ , the function  $\tilde{f}$  is  $\mathcal{A}$ -differentiable on  $F$  and*

$$\nabla \tilde{f} = \nabla \Phi \circ (\text{Id}_F - \text{prox}_f).$$

*Proof.* Since  $\Phi$  is norm continuous, the inf-convolution  $\tilde{f}$  is the conjugate of  $\Psi + g$ , see [18], section 9. Hence  $\tilde{f}$  is convex. By the same argument it follows, according to [18], section 4, that  $\tilde{f}$  is norm upper semicontinuous on  $F$ . Hence, by [8], chapter II,  $\tilde{f}$  is norm continuous on  $F$ . Since  $\tilde{f}$  is the conjugate of  $\Psi + g$  and  $\Psi + g \in \Gamma_0(G)$ ,  $\tilde{f}^* = \Psi + g$ .

To prove the final assertion of theorem 5 consider two arbitrary points  $X, X_1$  in  $F$  and put  $x = \text{prox}_f X$ ,  $x_1 = \text{prox}_f X_1$ ,  $y = \nabla \Phi(X - x)$ , and  $y_1 = \nabla \Phi(X_1 - x_1)$ .

By proposition 1.

$$(4.16) \quad f(x) + g(y) = \langle x, y \rangle \quad \text{and} \quad f(x_1) + g(y_1) = \langle x_1, y_1 \rangle.$$

By Fenchel inequality,  $\tilde{f}(X) + \tilde{f}^*(y_1) \geq \langle X, y_1 \rangle$ .

It follows by  $\tilde{f}^* = \Psi + g$  and (4.16)

$$(4.17) \quad \tilde{f}(X) \geq \langle X, y_1 \rangle + f(x_1) - \Psi(y_1) - \langle x_1, y_1 \rangle.$$

Since  $y_1 = \nabla \Phi(X_1 - x_1)$ ,

$$\Phi(X_1 - x_1) - \langle X_1, y_1 \rangle = -\Psi(y_1) - \langle x_1, y_1 \rangle$$

and substituting in (4.17),

$$(4.18) \quad \tilde{f}(X) - \tilde{f}(X_1) \geq \langle X - X_1, y_1 \rangle.$$

Symmetrically,  $\tilde{f}(X_1) - \tilde{f}(X) \geq \langle X_1 - X, y \rangle$

and hence

$$(4.19) \quad \tilde{f}(X_1) - \tilde{f}(X) - \langle X_1 - X, y \rangle \geq 0.$$

By (4.18),

$$(4.20) \quad \tilde{f}(X_1) - \tilde{f}(X) - \langle X_1 - X, y \rangle \leq \langle X_1 - X, y_1 - y \rangle.$$

Thus by (4.19) and (4.20)

$$(4.21) \quad 0 \leq \tilde{f}(X_1) - \tilde{f}(X) - \langle X_1 - X, y \rangle \leq \langle X_1 - X, y_1 - y \rangle.$$

Let now  $X$  be any point in  $F$ ,  $A$  any set in  $\mathcal{A}$  and put  $x = \text{prox}_f X$ ,  $y = \nabla \Phi(X - x)$ . For any  $u \in A$  and any  $\lambda > 0$ , put

$$x_{u,\lambda} = \text{prox}_f(X + \lambda u), y_{u,\lambda} = \nabla \Phi(X + \lambda u - x_{u,\lambda}).$$

By (4.21),

$$\sup_{u \in A} \left| \frac{\tilde{f}(X + \lambda u) - \tilde{f}(X)}{\lambda} - \langle u, y \rangle \right| \leq \sup_{u \in A} \langle u, y_{u,\lambda} - y \rangle.$$

Assume  $\lambda_n \downarrow 0$ . Since  $\Psi$  is Fréchet differentiable,  $\mathcal{S}$  is the norm topology of  $F$ . Hence, by theorem 3,  $\|x_{u,\lambda_n} - x\|_F \rightarrow 0$  uniformly with respect to  $u$  in the bounded set  $A$ . According to [4], the mapping  $\nabla \Phi$  is continuous from the norm topology to the  $\mathcal{C}$ -topology. Hence  $y_{u,\lambda_n} \rightarrow y$  in the  $\mathcal{C}$ -topology, uniformly with respect to  $u \in A$ . It follows  $\langle u, y_{u,\lambda_n} - y \rangle \rightarrow 0$  uniformly with respect to  $u \in A$ . Hence

$$\limsup_{\lambda \downarrow 0} \sup_{u \in A} \left| \frac{\tilde{f}(X + \lambda u) - \tilde{f}(X)}{\lambda} - \langle u, y \rangle \right| = 0$$

and this completes the proof of theorem 5.

**REMARK 4.** By theorem 5,  $\text{prox}_f = \text{Id}_F - \nabla \Psi \circ \nabla \tilde{f}$ . This is in our framework the substitute of the fact that in the framework of Moreau [17],  $\text{prox}_f$  is a gradient mapping.

As in [17], the above theorem yields some corollaries.

**Corollary 1.** Assume that  $\Psi$  is Fréchet differentiable on  $G$  and let  $f$  and  $f_1$  be two arbitrary functions in  $\Gamma_0(F)$ . Then  $\text{prox}_{f_1} = \text{prox}_f$  if and only if  $f_1 = f + \text{const}$ .

*Proof.* Assume  $\text{prox}_{f_1} = \text{prox}_f$ . By remark 4,  $\nabla \tilde{f}_1 = \nabla \tilde{f}$  and hence  $\tilde{f}_1 = \tilde{f} + \alpha$ , where  $\alpha = \text{const}$ . Denote by  $g, g_1$  the functions conjugate to  $f$ , respectively  $f_1$ . By theorem 5,  $\Psi + g_1 = \Psi + g - \alpha$ . Hence  $g_1 = g - \alpha$  and this yields  $f_1 = f + \alpha$ .

The reciprocal implication follows immediately from the definition of prox-mappings.

**Corollary 2.** *Assume that  $\Phi$  is Fréchet differentiable on  $F$ . For a function  $\varphi : F \rightarrow \bar{\mathbf{R}}$  the following two assertions are equivalent :*

(i)  $\varphi \in \Gamma_0(F)$  and  $\varphi$  is more convex than  $\Phi$  (i.e. there exists a convex function  $f : F \rightarrow \bar{\mathbf{R}}$  such that  $\varphi = \Phi + f$ ) ;

(ii) the function  $\psi$ , conjugate to  $\varphi$  is  $\mathcal{B}$ -differentiable on  $G$  and there exists  $g \in \Gamma_0(G)$  such that

$$(4.22) \quad \nabla\psi = \nabla\Psi \circ (\text{Id}_G - \text{prox}_g^\Psi).$$

*Proof.* (i)  $\Rightarrow$  (ii). Since  $\varphi \in \Gamma_0(F)$  and  $\Phi$  is norm continuous, by  $f = \varphi - \Phi$  we conclude that  $f \in \Gamma_0(F)$ . Denote by  $g$  the conjugate of  $f$ , hence  $g \in \Gamma_0(G)$ . According to [18], section 9, the function conjugate to  $\varphi$  is the inf-convolution  $\Psi \square g$ . Hence  $\psi = \Psi \square g$  and in order to prove that (ii) holds we have only to apply theorem 5 (with  $G$  and  $\Psi$  replaced by  $F$ , respectively  $\Phi$ ).

(ii)  $\Rightarrow$  (i). By theorem 5 the inf-convolution  $\tilde{g} = \Psi \square g$  is  $\mathcal{B}$ -differentiable on  $G$  and

$$\nabla\tilde{g} = \nabla\Psi \circ (\text{Id}_G - \text{prox}_g^\Psi).$$

Hence, by (4.22),  $\nabla\psi = \nabla\tilde{g}$  and it follows  $\psi = \tilde{g} - c$  where  $c$  is a constant. Taking conjugates,  $\varphi = \Phi + f + c$ , where  $f$  is the function conjugates, to  $g$ , and thus (i) holds.

**Lemma 5.** *Assume that  $h \in \Gamma_0(F)$  is majorized on any bounded set in  $F$ . Then, for any convex norm-neighborhood  $B$  of  $O_G$ , any bounded set  $K \subset F$  and any  $\delta > 0$ , there exist a norm neighborhood  $W$  of  $O_F$  and  $\varepsilon > 0$  such that*

$$\partial_\lambda h(x + u) \subset \partial_\delta h(x) + B, \quad \forall x \in K, \quad \forall u \in W, \quad \forall \lambda \in [\delta - \varepsilon, \delta + \varepsilon].$$

*Proof.* Let  $U$  be a bounded norm neighborhood of  $O_F$  and put

$$M = \bigcup \{ \partial_\lambda h(x) : x \in K + U, \lambda \leq 2\delta \}.$$

Since the set  $K + U$  in  $F$  is bounded, by lemma 2, the set  $M$  in  $G$  is also bounded. It follows that there exists  $\rho > 2\delta$  such that  $M - M \subset \rho B$ . Choose  $\varepsilon > 0$  such that  $2\varepsilon < \delta/(\rho + 1)$ . By lemma 2, there exists a norm neighborhood  $W$  of  $O_F$  such that  $W \subset U$  and

$$|h(x + u_1) - h(x + u_2)| \leq \varepsilon/2, \quad \forall x \in K, \quad \forall u_1 \in W, \quad \forall u_2 \in W.$$

Since  $M$  is bounded, we can choose  $W$  such that

$$\langle u, y \rangle \leq \varepsilon/4, \quad \forall u \in W, \quad \forall y \in M$$

holds too. Further on the proof follows as that of proposition 5 of [4].

**Lemma 6.** *If, on any bounded set in  $F$ , the function  $h \in \Gamma_0(F)$  is majorized and uniformly  $\mathcal{A}$ -differentiable, then, on any bounded set in  $F$ , the mapping  $\nabla h : F \rightarrow G$  is uniformly continuous from the norm topology to the  $\mathcal{G}$ -topology.*

*Proof.* Let  $K$  be any bounded set in  $F$  and  $V$  be any  $\mathcal{G}$ -neighborhood of  $O_G$ . There exists a convex  $\mathcal{G}$ -neighborhood  $B$  of  $O_G$  such that  $B + B \subset V$ . By lemma 3, there exists  $\delta > 0$  such that (4.5 a) holds,  $l$  being the function conjugate to  $h$ . By lemma 5 and  $\nabla h(z) \in \partial_\lambda h(z)$ ,  $\forall z \in F$ ,  $\forall \lambda > 0$ , there exists a norm neighborhood  $W$  of  $O_F$  such that

$$(4.23) \quad \nabla h(x + u) \in \partial_\delta h(x) + B, \quad \forall x \in K, \forall u \in W.$$

Since  $h(x) + l(\nabla h(x)) = \langle x, \nabla h(x) \rangle$ , we have for any  $x \in K$ ,

$$\begin{aligned} \partial_\delta h(x) &= \{ y \in G : l(y) - \langle x, y \rangle \leq \delta - h(x) \} \\ &= \{ y \in G : l(y) - l(\nabla h(x)) - \langle x, y - \nabla h(x) \rangle \leq \delta \} \\ &= \nabla h(x) + \{ v \in G : l(\nabla h(x) + v) - l(\nabla h(x)) - \langle x, v \rangle \leq \delta \}. \end{aligned}$$

Hence, by (4.5 a) and (4.23),

$$\nabla h(x + u) \in \nabla h(x) + B + B \subset h(x) + V, \quad \forall x \in K, \forall u \in W.$$

We can now make theorem 5 more precise under certain additional assumptions.

**Theorem 6.** *Assume that on any bounded set in  $F$ , the function  $\Phi$  is majorized and uniformly  $\mathcal{A}$ -differentiable and, on any bounded set in  $G$ , the function  $\Psi$  is uniformly Fréchet differentiable. Then on any bounded set in  $F$ , the function  $\tilde{f}$  is majorized and uniformly  $\mathcal{A}$ -differentiable.*

*Proof.* Since on any bounded set in  $F$  the function  $\Phi$  is majorized, it follows immediately by (4.15 a) that on any bounded set in  $F$  the function  $\tilde{f}$  is also majorized.

In order to show that on any bounded set in  $F$  the function  $\tilde{f}$  is uniformly  $\mathcal{A}$ -differentiable we have only to return to the final part of the proof of theorem 5 : invoke theorem 4 instead of theorem 3 and lemma 6 instead of the continuity of the mapping  $\nabla \Phi : F \rightarrow G$  from the norm topology to the  $\mathcal{G}$ -topology.

## 5. EXAMPLES AND FINAL REMARKS

Note first that if  $F$  is a Hilbert space over  $R$ ,  $G = F$ ,  $\langle \cdot, \cdot \rangle$  as the scalar product, and  $\Phi = \Psi = (1/2) \|\cdot\|^2$ , we are in the framework considered primarily by Moreau [17]. The assumptions of all our theorems hold. Theorems 1 (the first part), 2 and 5 reduce to the corresponding results of [17]. Concerning proposition 3, we point out that it furnishes a weaker result than

the corresponding one of [17] (the mapping  $\text{prox}_f$  is a contraction), obtained by means which seem specific to Hilbert spaces. That is why we think that our proposition 3 may be improved.

We shall now discuss a few examples that seem to motivate our framework and the investigation of the various continuity properties of prox-mappings in § 4.

EXAMPLES. 4. We show now that if the norms  $\|\cdot\|_F$  and  $\|\cdot\|_G$  have some smoothness properties (for the general theory of smoothness of the norm we refer the reader to [10] and [11]) and  $\varphi, \psi$  is a pair of Young conjugate functions on  $\mathbf{R}$ , then the pair of composite functions  $\varphi \circ \|\cdot\|_F, \psi \circ \|\cdot\|_G$  satisfies our conditions (A) and (B).

Let  $a$  be any function  $\mathbf{R}_+ \rightarrow \mathbf{R}_+$ , which is continuous and strictly increasing on  $\mathbf{R}_+$ , with  $a(0) = 0$  and  $a(\lambda) \rightarrow +\infty$  as  $\lambda \rightarrow +\infty$ . The reciprocal function  $b : \mathbf{R}_+ \rightarrow \mathbf{R}_+$  possesses the same properties as  $a$ . The functions  $\varphi$  and  $\psi$  from  $\mathbf{R}$  onto  $\mathbf{R}_+$ ,

$$\varphi(\lambda) = \int_0^{|\lambda|} a(t)dt, \psi(\lambda) = \int_0^{|\lambda|} b(t)dt$$

are differentiable, strictly convex and even on  $\mathbf{R}$ ; in addition  $\varphi'(0) = \psi'(0) = 0$ . They form a pair of (classical Young) conjugate functions on  $\mathbf{R}$ .

If the norms  $\|\cdot\|_F$  and  $\|\cdot\|_G$  are  $\mathcal{A}$  —, respectively  $\mathcal{B}$  —, differentiable on  $F \setminus \{O_F\}$ , respectively  $G \setminus \{O_G\}$ , then the pair of functions  $\Phi = \varphi \circ \|\cdot\|_F, \Psi = \psi \circ \|\cdot\|_G$  satisfies our conditions (A) and (B).

*Proof.* It is easy to verify that  $\Phi \in \Gamma_0(F), \Psi \in \Gamma_0(G)$  and that they are conjugate to each other (see also [18], section 14).  $\Phi(O_F) = \Psi(O_G) = 0$  holds obviously and an immediate calculation shows that  $\Phi$  and  $\Psi$  are Fréchet differentiable at  $O_F$ , respectively  $O_G$ , with  $\nabla \Phi(O_F) = O_G, \nabla \Psi(O_G) = O_F$ ; then, for any  $y \in G, y \neq O_G$ , any  $v \in G$  and any  $\lambda > 0$  we have (by the Lagrange formula for  $\psi$ )

$$(5.1) \quad \frac{\Psi(y + \lambda v) - \Psi(y)}{\lambda} = \frac{\|y + \lambda v\|_G - \|y\|_G}{\lambda} \psi'(c(y, v, \lambda)),$$

where  $c(y, v, \lambda)$  is a real number placed between  $\|y\|_G$  and  $\|y + \lambda v\|_G$ . Since  $\|\cdot\|_G$  is  $\mathcal{B}$ -differentiable at  $y$  and  $\psi'$  is continuous at  $\|y\|_G$ , it follows that  $\Psi$  is  $\mathcal{B}$ -differentiable at  $y$ , with

$$(5.2) \quad \nabla \Psi(y) = \psi'(\|y\|_G)[\nabla \|\cdot\|_G(y)] \quad , \quad \forall y \in G, y \neq O_G.$$

Similarly,  $\Phi$  is  $\mathcal{A}$ -differentiable on  $F$ .

Note that these functions  $\Phi$  and  $\Psi$  are bounded on any bounded set in  $F$ , respectively  $G$  and thus proposition 2 applies.

If the norm  $\|\cdot\|_G$  is uniformly smooth (i.e. uniformly Fréchet differentiable on any bounded closed set in  $G \setminus \{O_G\}$ ), then  $\Psi$  is uniformly Fréchet differentiable on any bounded set in  $G$ .

*Proof.* First, the function  $\Psi$  is uniformly Fréchet differentiable on any ring  $K_r^\rho = \{y \in G : r \leq \|y\|_G \leq \rho\}$ , with  $0 < r \leq \rho < \infty$  (take account of

$$\begin{aligned} & \left| \frac{\Psi(y + \lambda v) - \Psi(y)}{\lambda} - \langle v, \nabla \Psi(y) \rangle \right| \\ \leq & \left| \frac{\|y + \lambda v\|_G - \|y\|_G}{\lambda} - \langle v, \nabla \|\cdot\|_G(y) \rangle \right| \psi'(\|y\|_G) \\ & + |\psi'(c(y, v, \lambda)) - \psi'(\|y\|_G)| \|v\|_G, \end{aligned}$$

which follows by (5.1) and (5.2); then invoke the uniform Fréchet differentiability of the function  $\|\cdot\|_G$  on  $K_r^\rho$ ,  $\psi'(\|y\|_G) \leq \psi'(\rho)$ ,  $\forall y \in K_r^\rho$ , and the uniform continuity of  $\psi'$  on any bounded interval in  $\mathbf{R}$ , which follows from the continuity of  $\psi'$  on  $\mathbf{R}$ ). Next, choose  $\rho > 0$  and  $\varepsilon > 0$  arbitrarily. By (5.1) and the continuity of  $\psi'$  at 0, there exist  $r, 0 < r \leq \rho$  and  $\lambda_1 > 0$ , such that

$$\lambda^{-1} |\Psi(y + \lambda v) - \Psi(y)| \leq \varepsilon/2, \quad \forall y \in \Sigma(O_G, r), \forall v \in \Sigma(O_G, \rho), \forall \lambda \in ]0, \lambda_1].$$

By the continuity of  $\nabla \Psi$  in the norm topologies (guaranteed, according to [4], by the Fréchet differentiability of  $\Psi$  on  $G$ ) and  $\nabla \Psi(O_G) = O_F$ , it follows that we can choose  $r$  such that

$$|\langle v, \nabla \Psi(y) \rangle| < \varepsilon/2, \quad \forall y \in \Sigma(O_G, r), \forall v \in \Sigma(O_G, \rho)$$

holds too. Thus

$$(5.3) \quad \sup_{(y,v) \in \Sigma(O_G,r) \times \Sigma(O_G,\rho)} \left| \frac{\Psi(y + \lambda v) - \Psi(y)}{\lambda} - \langle v, \nabla \Psi(y) \rangle \right| \leq \varepsilon, \quad \forall \lambda \in ]0, \lambda_1].$$

Since  $\Phi$  is uniformly Fréchet differentiable on the ring  $K_r^\rho$ , there exists  $\lambda_0, 0 < \lambda_0 < \lambda_1$  such that

$$(5.4) \quad \sup_{(y,v) \in K_r^\rho \times \Sigma(O_G,\rho)} \left| \frac{\Psi(y + \lambda v) - \Psi(y)}{\lambda} - \langle v, \nabla \Psi(y) \rangle \right| \leq \varepsilon, \quad \forall \lambda \in ]0, \lambda_0].$$

By (5.3) and (5.4) it follows that  $\Psi$  is uniformly Fréchet differentiable on any bounded set in  $G$ .

The above statements enable us to furnish various examples when our propositions and theorems hold (with  $\Phi, \Psi$  defined as above). For instance : if the norms  $\|\cdot\|_F$  and  $\|\cdot\|_G$  are smooth (i.e. Gateaux differentiable except at the origin), then (A) and (B) hold with  $\mathcal{A}$ - and  $\mathcal{B}$ -differentiability as Gateaux differentiability (hence  $\mathcal{S}$  and  $\mathcal{C}$  as the weak topologies) ; if the norms  $\|\cdot\|_F$  and  $\|\cdot\|_G$

are uniformly smooth, then (A) and (B) hold with  $\mathcal{A}$ - and  $\mathcal{B}$ -differentiability as Fréchet differentiability (hence  $\mathcal{S}$  and  $\mathcal{G}$  is the norm topologies). In the latter case all of our theorems and propositions apply, except proposition 3.

Recall that if  $M$  is a non empty convex closed set in  $F$  and  $f$  is its indicator function (i.e.  $f(x) = 0$ , if  $x \in F$  and  $f(x) = +\infty$ , if  $x \notin F$ ), then (for  $\Phi, \Psi$  defined as above)  $\text{prox}_f X$  is the projection of  $X$  on  $M$  (i.e. the unique minimum point on  $M$  of the function  $M \ni u \mapsto \|X - u\|_F$ ). Thus, from our results one can as well derive various continuity properties of the projection on  $M$  mapping, corresponding to certain smoothness properties of the norms  $\|\cdot\|_F, \|\cdot\|_G$ . For instance, if  $\|\cdot\|_F$  and  $\|\cdot\|_G$  are smooth, then by theorem 3, the projection on  $M$  mapping is continuous from the norm topology to the weak topology.

Finally, note that according to a result of [21] and the renorming theorems of [2], every reflexive normed space has an equivalent norm  $\|\cdot\|'_F$ , which is Fréchet differentiable and whose dual norm  $\|\cdot\|'_G$  is also Fréchet differentiable (except at the origin). Putting  $\Phi = \varphi \circ \|\cdot\|'_F, \Psi = \psi \circ \|\cdot\|'_G$ , we see that in reflexive normed spaces it is always possible to construct pairs of Legendre conjugate functions which satisfy our conditions (A) and (B), with  $\mathcal{A}$ - and  $\mathcal{B}$ -differentiability as Fréchet differentiability (hence  $\mathcal{S}$  and  $\mathcal{G}$  as the norm topologies). Note that for these  $\Phi$  and  $\Psi$ , theorem 5 and its corollaries apply.

Invoking theorems 2 and 3 (for the above  $\Phi$  and  $\Psi$ ), we may see that for any  $f \in \Gamma_0(F)$ , the graph of  $\partial f$  is homeomorphic to  $F$  (and to  $G$ ): as a matter of fact, the mapping  $(x, y) \mapsto x + \nabla \Psi(y)$  from the graph of  $\partial f$  to  $F$  is bijective and norm-to-norm continuous in both directions.

5. We shall now indicate an example when proposition 3 applies. Let  $T$  be any locally compact space,  $\mu$  any positive Radon measure on  $T$  and  $p, q$  any reals with  $p > 1$  and  $p^{-1} + q^{-1} = 1$ . As usual,  $L^p$  is the set of (classes of) real-valued measurable functions  $x$  such that the quantity

$$\|x\|_p = \left\{ \int_T |x|^p d\mu \right\}^{1/p}$$

is finite. Put  $F = L^p$  with the norm  $\|\cdot\|_p, G = L^q$  with the norm  $\|\cdot\|_q$ ; the canonical bilinear form is as usual

$$\langle x, y \rangle = \int_T xy d\mu.$$

Put  $\Phi = p^{-1} \|\cdot\|_p^p, \Psi = q^{-1} \|\cdot\|_q^q$ . Since the norms  $\|\cdot\|_p$  and  $\|\cdot\|_q$  are uniformly smooth, this example may be considered as a special case of example 4 with  $a$  and  $b$  defined by  $a(\lambda) = |\lambda|^{p-1} \text{sgn } \lambda, b(\lambda) = |\lambda|^{q-1} \text{sgn } \lambda$ . From

$$p^{-1} |\lambda_1|^p + q^{-1} |\lambda_2|^q \geq \lambda_1 \lambda_2, \forall \lambda_1 \in \mathbf{R}, \forall \lambda_2 \in \mathbf{R},$$

with equality if and only if  $\lambda_2 = |\lambda_1|^{p-1} \operatorname{sgn} \lambda_1$  it follows

$$\Phi(x) + \Psi(y) \geq \langle x, y \rangle, \forall x \in L^p, \forall y \in L^q,$$

with equality if and only if  $y = |x|^{p-1} \operatorname{sgn} x$ . Hence, by (2.1),

$$\nabla\Phi(x) = |x|^{p-1} \operatorname{sgn} x.$$

Similarly,  $\nabla\Psi(y) = |y|^{q-1} \operatorname{sgn} y$ .

Assume  $1 < p \leq 2$ . From the elementary inequality

$$\left| |v|^{p-1} \operatorname{sgn} v - |u|^{p-1} \operatorname{sgn} u \right| \leq 2^{2-p} |v - u|^{p-1}, \forall u \in \mathbf{R}, \forall v \in \mathbf{R} \quad (1),$$

it follows that

$$\|\nabla\Phi(x_1) - \nabla\Phi(x_2)\|_q \leq 2^{2-p} \|x_2 - x_1\|_p^{p-1}, \forall x_1 \in L^p, \forall x_2 \in L^p.$$

Thus the mapping  $\nabla\Phi : L^p \rightarrow L^q$  is Lipschitz of order  $p - 1$  on all of  $L^p$ . Assume now  $p > 2$ . In this case there exists  $k > 0$  such that

$$\left| |v|^{p-1} \operatorname{sgn} v - |u|^{p-1} \operatorname{sgn} u \right| \leq k(|u| + |v|)^{p-2} |v - u|, \forall u \in \mathbf{R}, \forall v \in \mathbf{R} \quad (2).$$

Let  $x_1, x_2$  be any two functions in  $L^p$ . From the above inequality,

$$(5.5) \quad \left| \nabla\Phi(x_2) - \nabla\Phi(x_1) \right|^q \leq k^q (|x_1| + |x_2|)^{p-q} |x_2 - x_1|^q.$$

Since  $|x_2 - x_1|^q \in L^{p/q}$ ,  $(|x_1| + |x_2|)^{p-q} \in L^{p/(p-q)}$  and the reals  $p/q, p/(p-q)$  are conjugate to each other, by the Hölder inequality it follows

$$(5.6) \quad \int_T (|x_1| + |x_2|)^{p-q} |x_2 - x_1|^q d\mu \leq \| |x_1| + |x_2| \|_p^{p-q} \|x_2 - x_1\|_p^q.$$

By (5.5) and (5.6)

$$\|\nabla\Phi(x_2) - \nabla\Phi(x_1)\|_q \leq k \| |x_1| + |x_2| \|_p^{p-2} \|x_2 - x_1\|_p, \forall x_1 \text{ et } x_2 \in L^p$$

and  $\nabla\Phi$  is Lipschitz on any bounded set in  $L^p$ .

(1) In order to prove this inequality consider the functions  $\chi_1$  and  $\chi_2$  from  $[0,1[$  into  $\mathbf{R}$ ,  $\chi_1(\lambda) = (1 + \lambda^{p-1})/(1 + \lambda)^{p-1}$ ,  $\chi_2(\lambda) = (1 - \lambda^{p-1})/(1 - \lambda)^{p-1}$  and take into account that for any  $\lambda \in [0,1[$  we have  $\chi_1(\lambda) \leq 2^{2-p}$ ,  $\chi_2(\lambda) \leq 2^{2-p}$ .

(2) In order to prove this inequality consider the function  $\chi_3 : [0,1[ \rightarrow \mathbf{R}$ ,  $\chi_3(\lambda) = (1 - \lambda^{p-1})/(1 - \lambda)(1 + \lambda)^{p-2}$  and put  $k = \sup \{ \chi_3(\lambda) : \lambda \in [0,1[ \}$ , hence  $2^{2-p} < k < +\infty$ ; then, take into account the estimates for  $\chi_1$  and  $\chi_2$ .



Thus the conditions of proposition 3 hold in this example.

6. In all of the above examples  $\Phi$  (and  $\Psi$ ) is a composite function of the form  $\varphi \circ \|\cdot\|$ , where  $\varphi$  is a function  $\mathbf{R} \rightarrow \mathbf{R}$ . It is easy to imagine pairs of functions which satisfy conditions (A) and (B) but are not of the above type (for instance put  $F = G = \mathbf{R}^n$ ,  $n \geq 2$ , and define  $\Phi, \Psi$  by

$$\Phi(z) = \sum_1^n p_j^{-1} |\zeta_j|^{p_j}, \Psi(z) = \sum_1^n q_j^{-1} |\zeta_j|^{q_j} \quad \text{for } z = (\zeta_1, \dots, \zeta_n)$$

where  $p_j, q_j$  are given reals with  $p_j > 1, p_j^{-1} + q_j^{-1} = 1$  and at least two of the  $p_j$  are different).

Other types of pairs  $\Phi, \Psi$  arise in a very natural way in Orlicz spaces (for the theory of Orlicz spaces we refer the reader to [12] and [14]). Let  $T$  and  $\mu$  be as in example 5 and  $\varphi, \psi$  as in example 4. Assume in addition that both functions  $\varphi$  and  $\psi$  satisfy the condition  $(\Delta_2)$  of [14], (denoted by  $(\Delta)$  in [12]). For any measurable real-valued function  $z$ , put

$$\Phi(z) = \int_T \varphi \circ z \, d\mu, \Psi(z) = \int_T \psi \circ z \, d\mu$$

(the above integrals are in the sense of [12], i.e. essential integrals in the sense of [7], chapter 5). The Orlicz space  $L^\varphi$ , respectively  $L^\psi$ , is the space of (classes of) functions  $z$ , such that  $\Phi(z) < +\infty$ , respectively  $\Psi(z) < +\infty$ . Under the Orlicz norms  $\|\cdot\|_\varphi$ , respectively  $\|\cdot\|_\psi$ ,  $L^\varphi$  and  $L^\psi$  are reflexive Banach spaces, each of them being isomorphic (algebraically and topologically) to the dual space of the other. The canonical bilinear form on  $L^\varphi \times L^\psi$  is given by  $\langle x, y \rangle = \int_T xy \, d\mu$ . Thus  $F = L^\phi$  with the Orlicz norm  $\|\cdot\|_\varphi$  and  $G = L^\psi$  with the norm  $\|\cdot\|_\psi$ , dual to  $\|\cdot\|_\varphi$ , satisfy the conditions of § 2.

Since  $\varphi$  satisfies condition  $(\Delta_2)$ , there exists  $k > 0$  such that  $\lambda\varphi'(\lambda) \leq k\varphi(\lambda), \forall \lambda \in \mathbf{R}$  (see [14]). On the other hand,  $0 \leq \lambda\varphi'(\lambda), \forall \lambda \in \mathbf{R}$ . Hence

$$0 \leq x(\varphi' \circ x) \leq k(\varphi \circ x), \forall x \in L^\phi$$

and thus  $x(\varphi' \circ x)$  is an integrable function,  $\forall x \in L^\phi$ . Since the functions  $\varphi, \psi$  are differentiable and conjugate to each other on  $\mathbf{R}$ ,

$$(5.7) \quad \varphi(\lambda_1) + \psi(\lambda_2) \geq \lambda_1\lambda_2 \quad , \quad \forall \lambda_1 \in \mathbf{R}, \forall \lambda_2 \in \mathbf{R},$$

with equality if and only if  $\lambda_2 = \varphi'(\lambda_1)$  (which is equivalent to  $\lambda_1 = \psi'(\lambda_2)$ ). Hence

$$\psi \circ (\varphi' \circ x) = x(\varphi' \circ x) - \varphi \circ x, \forall x \in L^\phi,$$

and thus  $\varphi' \circ x \in L^\psi, \forall x \in L^\phi$ . Similarly  $\psi' \circ y \in L^\phi, \forall y \in L^\psi$ .

By (5.7),

$$\Phi(x) + \Psi(y) \geq \langle x, y \rangle, \forall x \in L^\phi, \forall y \in L^\psi,$$

with equality if and only if  $y = \varphi' \circ x$  (which is equivalent to  $x = \psi' \circ y$ ). Hence the functions  $\Phi$  and  $\Psi$  are conjugate to each other and  $\partial\Phi(x) = \{ \varphi' \circ x \}$ ,  $\partial\Psi(y) = \{ \psi' \circ y \}$ . Since  $\text{dom } \Phi = L^\phi$ , the function  $\Phi$  is continuous on  $L^\phi$ , see [19]. According to [18], section 10, the continuity of  $\Phi$  and  $\partial\Phi(x) = \{ \varphi' \circ x \}$ , implies that  $\Phi$  is Gâteaux differentiable at any  $x \in L^\phi$  with  $\nabla\Phi(x) = \varphi' \circ x$ . Similarly,  $\Psi$  is Gâteaux differentiable at any  $y \in L^\psi$  with  $\nabla\Psi(y) = \psi' \circ y$ . Then, by the continuity of  $\varphi'$  and  $\psi'$  on  $\mathbb{R}$ , arguments similar to those used in [14], theorem 17.3, yield that the mappings  $\nabla\Phi$  and  $\nabla\Psi$  are continuous from the norm topology to the norm topology. Hence, by [4],  $\Phi$  and  $\Psi$  are Fréchet differentiable on  $F$ , respectively  $G$ . Thus our conditions (A) and (B) hold with  $\mathcal{A}$ - and  $\mathcal{B}$ -differentiability as Fréchet differentiability (hence  $\mathcal{S}$  and  $\mathcal{C}$  as the norm topologies).

7. We shall now express a Beurling-Livingstone type theorem in terms of prox-mappings. We replace the conditions on the mapping  $T : F \rightarrow G$  in [5], [6], [3], by the requirement that  $T$  is the gradient mapping of a function  $\Phi$  such that  $\Phi$  and its conjugate  $\Psi$  satisfy our conditions in § 2 (i.e.  $\Phi, \Psi$  is a pair of Legendre conjugate functions) : *let  $P$  be a closed subspace in  $F$  and  $Q$  its annihilator in  $G$ . For any couple  $(\xi, \eta)$  in  $F \times G$  there exists a unique couple  $(p, q)$  in  $P \times Q$  such that*

$$(5.8) \quad \nabla\Phi(\xi + p) = \eta + q.$$

As a matter of fact,

$$(5.9) \quad p = -\text{prox}_f \xi, \quad q = \text{prox}_g [\nabla\Phi(-p) + \nabla\Phi(\xi + p)] - \eta,$$

where the conjugate to each other functions  $f, g$  are defined as  $f(u) = \langle u, \eta \rangle$  if  $u \in P, f(u) = +\infty$  if  $u \notin P$  and  $g$  as the indicator function of  $Q + \eta$ .

*Proof.* Let  $p$  and  $q$  be defined by (5.9). Obviously  $p \in P$  and  $q \in Q$ . By theorem 2, the pair  $-p, \eta + q$  is conjugate with respect to the pair  $f, g$  and  $-p + \nabla\Psi(\eta + q) = \xi$ . This equality yields immediatly (5.8).

To prove uniqueness, assume that  $p' \in P, q' \in Q$  satisfy

$$(5.10) \quad \nabla\Phi(\xi + p') = \eta + q'$$

By  $p' \in P, q' \in Q$  it follows that the pair  $-p', \eta + q'$  is conjugate with respect to the pair  $f, g$ . Then by theorem 2,

$$-p' = \text{prox}_f X, \quad \text{where } X = -p' + \nabla\Psi(\eta + q').$$

By (5.10) it follows  $X = \xi$ , hence  $p' = p$ . Then by (5.8) and (5.10),  $q' = q$ .

If the functions  $\Phi$  and  $\Psi$  are Fréchet differentiable on  $F$ , respectively  $G$ , theorem 5 may be used to express the points  $p, q$  in a form different from (5.9). The continuity properties of prox-mappings furnish some continuity properties of the mapping  $(\xi, \eta) \mapsto (p, q)$ , where  $(p, q) \in P \times Q$  is the unique solution of (5.8). For instance : *if  $\Phi$  and  $\Psi$  are Fréchet differentiable, then the mapping  $(\xi, \eta) \mapsto (p, q)$  is norm-to-norm continuous with respect to each argument* (to prove continuity with respect to the first argument invoke theorem 3 and the norm-to-norm continuity of  $\nabla \Phi$ ; to prove continuity with respect to the second argument, replace  $F$  and  $P$  by  $G$ , respectively  $Q$  and take into account the uniqueness of  $p, q$  satisfying (5.8)).

As above, one may prove the following more general statement : let  $P$  be a convex closed cone in  $F$  and  $Q = P^0$ . For any couple  $(\xi, \eta) \in F \times G$  there exists a unique couple  $(p, q) \in (-P) \times Q$ , such that

$$\nabla \Phi(\xi + p) = \eta + q \quad \text{and} \quad \langle p, q \rangle = 0;$$

$p$  and  $q$  are defined as above by (5.9).

#### REFERENCES

- [1] H. ARIS, *Condition suffisante de dérivabilité de l'application prox*, Faculté des sciences de Montpellier, Travaux du séminaire d'analyse unilatérale, 1969, exposé n° 4.
- [2] E. ASPLUND, *Averaged norms*, Israel J. Math., 1967, 5, 227-233.
- [3] E. ASPLUND, *Positivity of duality mappings*, Bull. Amer. Math. Soc., 1967, 73, 200-203.
- [4] E. ASPLUND and R. T. ROCKAFELLAR, *Gradients of convex functions*, Trans. Amer. Math. Soc., 1969, 139, 443-467.
- [5] A. BEURLING and A. E. LIVINGSTONE, *A theorem on duality mappings in Banach spaces*, Ark. Mat., 1962, 4, 405-411.
- [6] F. E. BROWDER, *On a theorem of Beurling and Livingstone*, Canad. J. Math., 1965, 17, 367-372.
- [7] N. BOURBAKI, *Intégration*, Hermann, Paris, 1952 and 1956.
- [8] N. BOURBAKI, *Espaces vectoriels topologiques*, Hermann, Paris, 1953 and 1955.
- [9] C. CASTAING, *Proximité et mesurabilité*, Faculté des sciences de Montpellier. Travaux du séminaire d'analyse unilatérale, 1968, exposé n° 1.
- [10] D. F. CUDIA, *The geometry of Banach spaces. Smoothness*, Trans. Amer. Math. Soc., 1964, 110, 284-314.
- [11] M. M. DAY, *Normed linear spaces*, Springer, Berlin, 1958.
- [12] N. DINCULEANU, *Integrarea pe spatii local compacte*, Editura Acad. R.S.R., Bucharest, 1965.
- [13] A. D. IOFFE and V. M. TIHOMIROV, *Conjugacy of convex functionals and extremum problems* (Russian), Uspehi mat. nauk, 23, 1968, 6, 51-116.
- [14] M. A. KRASNOSEL'SKII and Ja. B. PUTICKII, *Convex functions and Orlicz spaces* (Russian), GIF-ML, Moscow, 1958.
- [15] C. LESCARET, *Applications « prox » dans un espace de Banach*, C. R. Acad. Sci., Paris, 1967, 265, 676-678.

- [16] C. LESCARRET, *Semicontinuité supérieure de la multi-application « prox »*, Faculté des sciences de Montpellier, Travaux du séminaire d'analyse unilatérale, 1969, exposé n° 5.
- [17] J. J. MOREAU, *Proximité et dualité dans un espace hilbertien*, Bul. Soc. Math. France, 1965, 93, 273-299.
- [18] J. J. MOREAU, *Fonctionnelles convexes, mimeographed lecture notes*, Séminaire « Équations aux dérivées partielles », Collège de France, 1966.
- [19] R. T. ROCKAFELLAR, *Level sets and continuity of conjugate functions*, Trans. Amer. Math. Soc., 1966, 123, 46-63.
- [20] R. T. ROCKAFELLAR, *Conjugates and Legendre transforms of convex functions*, Canad. J. Math., 1967, 19, 200-205.
- [21] S. L. TROYANSKI, *On locally uniformly convex and differentiable norms in certain non separable Banach spaces*, Studia Math. 37, 1971, 2, 173-180.