August 1986

LIDS-P-1597

CONVOLUTIONS OF MAXIMAL MONOTONE MAPPINGS

JAVIER LUQUE

Department of Mathematical Sciences
University of North Florida
4567 St. Johns Bluff Rd. So.
Jacksonville, FL 32216

July 27, 1986

Submitted to SIAM J. Control and Optimization

CONVOLUTIONS OF MAXIMAL MONOTONE MAPPINGS*

Abstract. Let X, Y be two real vector spaces, and let $S, T: X \to 2^{Y}$. A new internal law of composition, the convolution of S and T, $S \square T$, is introduced. It is shown that $(S \square T)^{-1} = S^{-1} + T^{-1}$, thus the commutative monoids $((2^{Y})^{X}, \square)$ and $((2^{X})^{Y}, +)$ are isomorphic. The proximal map of S with respect to T, $P(S,T): X \to 2^{X}$, is also introduced. A purely algebraic generalizationof Moreau's Proximal Point Theorem is proved. The convolution of subdifferential maps of extended real-valued functions, and of monotone maps is studied. If $f, g \in \overline{\mathbb{R}}^{X}$, $f \square g$ is their infimal convolution, and ∂f denotes the subdifferential map of f, then $\partial f \square \partial g \subseteqq \partial (f \square g)$, and $P(\partial f, \partial g) \subseteqq p(f,g): x \to \text{Arg min } (f + g(x - \cdot))$. When f, g are proper convex, sufficient conditions for equality in the preceding inclusions are given. The strict and strong monotonicity of $S \square T$, and the Lipschitz continuity of S T and P(S,T) are studied. Several generalizations of Moreau's Proximal Point Theorem are proved. These include the known relation between the Yosida approximant and the resolvent of a monotone mapping.

Key words. Monotone maps, convolution, proximal maps, Yosida approximant, resolvent.

Classifications. AMS (MOS) 46,47,49,90.

*This research is part of the author's dissertationwritten under the supervision of Prof. Dimitri P. Bertsekas of the Massachusetts Institute of Technology. It was supported by the ITP Foundation, Madrid, Spain, and the National Science Foundation under grant ECS 8217668.

1. Introduction. The object of this paper is a new internal law of composition for point-to-set mappings between real vector spaces. This law was introduced in Luque (1984), which contains essentially all of the results reported here, and announced in Luque (1986). Let X, Y be real vector spaces, and let $S, T: X \to 2^Y$. The convolution of S and T, $S \Box T$, is given by

$$(S \Box T)x = \cup\{Su \cap Tv \,|\, u + v = x\}.$$

Associated with it we have the proximal map of S with respect to T, $P(S,T)$, defined by

$$P(S,T)x = \{u \in X \,|\, v \in X,\ u + v = x,\ Su \cap Tv \neq \emptyset\}.$$

Note that these definitions are valid when $(X,+)$ is just a semigroup.

The contents of this paper are as follows. Section 2 introduces the definitions above and proves some general facts about them. It is shown that $\Box$ is a commutative associative operation with unit. Thus $(2^Y)^X$, the set of multivalued maps from X into Y, equipped with $\Box$, is an abelian monoid. For $S: X \to 2^Y$, let S^{-1} be defined in the obvious manner. It is shown (theorem 2.3) that $(S \Box T)^{-1} = S^{-1} + T^{-1}$, thus the commutative monoids $((2^Y)^X, \Box)$, $((2^X)^Y, +)$ are isomorphic. The section ends with a generalization of Moreau's (1962) Proximal Points Theorem. This generalization is purely algebraic in the sense that the only concepts used are those of addition and convolution of multivalued maps, and the notion of proximal mappings.

Section 3 quickly reviews some concepts of convex analysis. Its main purpose is to introduce the notation to be used.

In section 4 we turn our attention to the convolution of monotone and subdifferential maps. Let $f,g \in \overline{\mathbb{R}}^X$, $f \square g$ denote their inf-convolution, and ∂f, ∂g their respective subdifferential maps. It is shown that $\partial f \square \partial g \subseteq \partial(f \square g)$ in the sense of inclusion of graphs. The proximal map of f with respect to g at x, $p(f,g)x$, is the set of optimal solutions of the minimization defining $(f \square g)(x)$ when it is $< +\infty$, and empty otherwise. It is also shown that $P(\partial f, \partial g) \subseteq p(f,g)$, again in the sense of inclusion of graphs. When f,g are proper convex, sufficient conditions for equality in the preceding inclusions are given. The strict and strong monotonicity of $S \square T$, and the Lipschitz continuity of $S \square T$ and $P(S,T)$ are also studied.

Section 5 is devoted to proving several generalizations of Moreau's Proximal Points Theorem. These include the known relation between the Yosida approximant and the resolvent of a monotone map (see Pascali and Sburlan 1978, p. 128).

Several authors have studied particular cases of the concepts introduced in this chapter. Moreau (1962) introduced the proximal map of f with respect to g, $p(f,g)$, in the particular case in which f is a proper closed convex function on a real Hilbert space and $g = \frac{1}{2}|\cdot|^2$, $|\cdot|$ being the Hilbert space norm. There he proved his Proximal Points Theorem which corresponds to our corollary 5.9 for $a = b = 1$. The proximal map $p(f,g)$ when f,g are defined on a real Banach space has also been studied by Lescarret (1967) and Wexler (1972).

Rockafellar defined the convolution (inverse addition in his terminology) of sets (1970a,p. 21, 3), complete increasing curves in $\mathbb{R} \times \mathbb{R}$ (1967a,p. 553,(2.21)), and monotone processes from $\mathbb{R}^n$ to $\mathbb{R}^m$ (1967b,p. 43). He also proved for these particular cases our theorem 2.3 (Rockafellar 1967a,p. 553,(2.22), and 1967b,p. 49, th. 5).

Gol'shtein (1975,p. 1146, §3) used a regularization of monotone maps from $\mathbb{R}^n$ into $\mathbb{R}^n$, which is equivalent to convolving them with the gradient of a convex function. This gradient and its inverse map were assumed to be strongly monotone.

Finally let us mention (see section 5, after theorem 5.4) that the Yosida approximant and the resolvent of a monotone map T from a real Banach space to its dual can be expressed as $T \square \lambda^{-1} J$ and $P(T, \lambda^{-1} J)$ where $\lambda > 0$ and J is the normalized duality map of the norm of X.

2. Convolution of maps. Let X, Y be real vector spaces. Let us consider a map $S:X \to 2^Y$. Its effective domain is $\mathrm{dom}\, S = \{x \in X \mid Sx \neq \emptyset\}$, its range, $\mathrm{ran}\, S = \bigcup\{Sx \mid x \in X\}$, and its graph, $\mathrm{gph}\, S = \{(x,y) \in X \times Y \mid y \in Sx\}$. The set of all such maps will be denoted by $(2^Y)^X$. A map $S \in (2^Y)^X$ is single valued iff for all $x \in X$, Sx has at most one element. If $S \in (2^Y)^X$, its inverse point-to-set map $S^{-1} \in (2^X)^Y$ is such that for all $y \in Y$, $S^{-1}y = \{x \in X \mid y \in Sx\}$. It is elementary that $\mathrm{dom}\, S^{-1} = \mathrm{ran}\, S$, $\mathrm{ran}\, S^{-1} = \mathrm{dom}\, S$, and $\mathrm{gph}\, S^{-1} = \{(y,x) \in Y \times X \mid y \in Sx\}$. The inversion operation $S \to S^{-1}:(2^Y)^X \to (2^X)^Y$, is a bijection. If $S, T \in (2^Y)^X$, we write $S \subseteq T$ iff $Sx \subseteq Tx$ for all $x \in X$. Clearly $S \subseteq T$ iff $\mathrm{gph}\, S \subseteq \mathrm{gph}\, T$ iff $S^{-1} \subseteq T^{-1}$.

X and Y will play completely symmetrical roles in what follows. Therefore statements will only be made for one of the two possible cases.

In $(2^Y)^X$ it is possible to define a law of composition by means of the addition of images. Let $S,T \in (2^Y)^X$, then for all $x \in X$, $(S + T)x = \{y + y' \mid y \in Sx,\ y' \in Tx\}$. From the group properties of addition in Y, it follows that this operation is commutative, associative, and has a unit E, such that for all $x \in X$, $Ex = \{0_Y\}$, where 0_Y denotes the unit of addition in Y. When there is no danger of confusion, we will simply write 0. A set in which an associative internal law of composition with unit is defined, is called a monoid. Therefore $((2^Y)^X,+)$ is an Abelian (commutative) monoid.

Definition 2.1. For all $S,T \in (2^Y)^X$, the convolution of S and T, $S \square T \in (2^Y)^X$, is given by $x \to \bigcup\{Su \cap Tv \mid u + v = x\}$.

Proposition 2.2. $((2^Y)^X, \square)$ is an Abelian monoid.

Proof. Clearly $\square$ is a commutative internal law of composition.

The unit is the map $F \in (2^Y)^X$ such that $F0 = Y$, and for all $x \neq 0$, $Fx = \emptyset$. Let $R,S,T \in (2^Y)^X$, $x \in X$

$$(R\Box(S\Box T))x = \bigcup\{Ru\cap(S\Box T)z \mid u + z = x\} =$$
$$\bigcup\{Ru\cap(\bigcup\{Sv\cap Tw \mid v + w = z\}) \mid u + z = x\} =$$
$$\bigcup\{Ru\cap Sv\cap Tw \mid u + z = x,\ v + w = z\} =$$
$$\bigcup\{Ru\cap Sv\cap Tw \mid u + v + w = x\}.$$

QED

Theorem 2.3. The inversion map is an isomorphism between $((2^Y),\Box)$ and $((2^X)^Y,+)$. In particular

$$\forall\ S,T \in (2^Y)^X,\ (S\Box T)^{-1} = S^{-1} + T^{-1},$$
$$\forall\ U,V \in (2^X)^Y,\ (U + V)^{-1} = U^{-1}\Box V^{-1}.$$

Proof. As seen before the inversion is a bijection between $(2^Y)^X$ and $(2^X)^Y$. Let $F \in (2^Y)^X$ be, as above, the unit of the convolution in $(2^Y)^X$, then $F^{-1} \in (2^X)^Y$ is such that for all $y \in Y$, $F^{-1}y = \{0_X\}$, thus F^{-1} is the unit of $((2^X)^Y,+)$. Let $S,T \in (2^Y)^X$, and $x \in X$, then

$$(S\Box T)x = \{y \in Y \mid y \in Su\cap Tv,\ u + v = x\} =$$
$$\{y \in Y \mid u \in S^{-1}y,\ v \in T^{-1}y,\ u + v = x\} = (S^{-1} + T^{-1})^{-1}x.$$

QED

If $S \in (2^Y)^X$ and $a \in \mathbb{R}$, as usual, $aS \in (2^Y)^X$ is such that $(aS)x = a(Sx)$ for all $x \in X$. This is the left scalar multiplication. It is possible to define a dual (with respect to inversion) operation, the right scalar multiplication. The notation will follow the convention that gives priority in the evaluation of expressions, to the external operations over the internal law of composition.

Definition 2.4. For all $S \in (2^Y)^X$, and for all $a \in \mathbb{R}\setminus\{0\}$, $(Sa) \in (2^Y)^X$ is the map $x \to S(a^{-1}x)$.

Proposition 2.5. Let $S,T \in (2^Y)^X$, and $a,b \in \mathbb{R}\setminus\{0\}$, then

(1) $a(S+T) = aS + aT$; $((S+T)a) = (Sa) + (Ta)$; $(a+b)S \subseteq aS + bS$,

(2) $a(S\Box T) = aS \Box aT$; $((S\Box T)a) = (Sa)\Box(Ta)$; $(S(a+b)) \subseteq (Sa)\Box(Sb)$,

with equality in (1) (resp. (2)) if $a,b > 0$ and S (resp. S^{-1}) is convex-valued (i.e., for all x in X, Sx is convex).

Proof.

$$a(S\Box T)x = a\bigcup\{Su \cap T(x-u) \mid u \in X\} =$$
$$\bigcup\{(aS)u \cap (aT)(x-u) \mid u \in X\} = (aS\Box aT)x.$$

Using theorem 2.3, we have

$$((S\Box T)a)^{-1} = a(S\Box T)^{-1} = a(S^{-1} + T^{-1}) = aS^{-1} + aT^{-1} =$$
$$(Sa)^{-1} + (Ta)^{-1} = ((Sa)\Box(Ta))^{-1}.$$

Also,

$$(S(a+b))^{-1} = (a+b)S^{-1} \subseteq aS^{-1} + bS^{-1} = (Sa)^{-1} + (Sb)^{-1} = ((Sa)\Box(Sb))^{-1}.$$

If $a,b > 0$ and S^{-1} is convex-valued, then $(a+b)S^{-1} = aS^{-1} + bS^{-1}$. QED

Let $S,T \in (2^Y)^X$, it is clear that $\text{dom }(S+T) = \text{dom } S \cap \text{dom } T$, and that $\text{ran }(S+T) \subseteq \text{ran } S + \text{ran } T$. Theorem 2.3 yields the following

Proposition 2.6. For all $S,T \in (2^Y)^X$, $\text{dom } S\Box T \subseteq \text{dom } S + \text{dom } T$, $\text{ran } S\Box T = \text{ran } S \cap \text{ran } T$.

When computing $(S \square T)x$, there may be some points $u \in X$ such that $Su \cap T(x - u) \neq \emptyset$. For each $x \in X$, the set of these points, for reasons that will become clear below, will be called the set of proximal points of S to x, along T. The point-to-set map that assigns to each $x \in X$ the set of its proximal points, will be denoted by $P(S,T)$, the proximal map of S with respect to T.

Definition 2.7. For all $S,T \in (2^Y)^X$, the proximal map of S with respect to T, $P(S,T) \in (2^X)^X$, is given by $x \to \{u \in X | Su \cap T(x-u) \neq \emptyset\}$.

We now turn to the relation between the convolution of two maps S,T, and their corresponding proximal maps $P(S,T)$, $P(T,S)$. First, let us note that if $S,T \in (2^Y)^X$, then $S^{-1}T$, $(S^{-1}T)^{-1} \in (2^X)^X$ satisfy

$$\forall x \in X, \quad S^{-1}Tx = \{u \in X | Su \cap Tx \neq \emptyset\},$$

$$\forall x \in X, \quad (S^{-1}T)^{-1}x = \{z \in X | x \in S^{-1}Tz\} = \{z \in X | Sx \cap Tz \neq \emptyset\} = T^{-1}Sx.$$

Proposition 2.8. Let I denote the identity map of X. Let $S,T \in (2^Y)^X$, then

(1) $P(S,T) = I - P(T,S)$,

(2) $P(S,T) = P(I,S^{-1}T) = P(T^{-1}S,I)$,

(3) $P(S,T) = I \square S^{-1}T = (T^{-1}S + I)^{-1}$,

(4) $S \square T \subseteq SP(S,T)$, $P(S,T) \subseteq S^{-1}(S \square T)$, with equality if S or S^{-1} is single-valued, respectively.

Proof. Let $x \in X$, then

$$u \in P(S,T)x \iff Su \cap T(x-u) \neq \emptyset \iff$$

$$[\exists v \in X, (u = x - v,\ S(x-v) \cap Tv \neq \emptyset)] \iff x - u \in P(T,S)x.$$

This proves (1). Let $u \in X$

$$Su \cap T(x - u) \neq \emptyset \iff u \in S^{-1}T(x - u) \iff Iu \cap S^{-1}T(x - u) \neq \emptyset,$$

and $P(S,T) = P(I,S^{-1}T)$. Using (1)

$$P(S,T) = I - P(T,S) = I - P(I,T^{-1}S) = P(T^{-1}S,I),$$

which shows (2). With everything as above

$$Su \cap T(x - u) \neq \emptyset \iff x - u \in T^{-1}Su \iff u \in (I + T^{-1}S)^{-1}x.$$

Therefore $P(S,T) = (I + T^{-1}S)^{-1}$, using theorem 2.3, $P(S,T) = (I \square S^{-1}T)$. From the definitions of $S \square T$ and $P(S,T)$, it is clear that $S \square T \subseteqq SP(S,T)$ with equality if S is single-valued. On the other hand, using theorem 2.3

$$\begin{aligned}
&u \in P(S,T)x \iff Su \cap T(x - u) \neq \emptyset \iff \\
&[\exists v \in Y, (u \in S^{-1}v, x \in u + T^{-1}v)] \Rightarrow \\
&[\exists v \in Y, (u \in S^{-1}v, x \in (S^{-1} + T^{-1})v)] \iff \\
&[\exists v \in (S \square T)x, u \in S^{-1}v] \iff u \in S^{-1}(S \square T)x.
\end{aligned}$$

If S^{-1} is single-valued, the reverse argument is also valid. QED

Proposition 2.9. Let $S,T \in (2^Y)^X$, then

(1) $\operatorname{dom} P(S,T) = \operatorname{dom} P(T,S) = \operatorname{dom} S \square T$,

(2) $\operatorname{ran} P(S,T) = \operatorname{ran} S^{-1}(S \square T)$.

Proof. (1) follows readily from definitions 2.1, 2.7, and proposition 2.8(1). From proposition 2.8(3) and elementary computations

$$\begin{aligned}
&\operatorname{ran} P(S,T) = \operatorname{dom} (I + T^{-1}S) = \operatorname{dom} T^{-1}S = \operatorname{ran} S^{-1}T = S^{-1}(\operatorname{ran} T) = \\
&S^{-1}(\operatorname{ran} S \cap \operatorname{ran} T) = S^{-1}(\operatorname{ran} S \square T) = \operatorname{ran} S^{-1}(S \square T).
\end{aligned}$$

QED

The expression of $P(S,T)$ as a convolution of two suitable maps in proposition 2.8(3) has the disadvantage, that even if S and T are monotone maps (see the following section for the definitions) $S^{-1}T$ does not have to be so and the theory available for these maps cannot be used to study $P(S,T)$ by means of the convolution operation. Therefore it is of interest to develop an alternative relation between the convolution operation and the induced proximal maps.

For any $u \in X$ let the map $D_u: X \to X$, be such that $x \to u - x$. Clearly $D_u D_u$ is the identity map of X and thus $D_u = D_u^{-1}$. Analogous maps defined in Y for $v \in Y$ will be denoted by D_v. The space on which these maps are acting will be clear from the context.

<u>Proposition 2.10.</u> Let $S,T \in (2^Y)^X$, $x \in X$, $y \in Y$. Then

(1) $(S\,\square\,T)x = P(S^{-1}, D_x T^{-1} D_y)y = P(T^{-1}, D_x S^{-1} D_y)y,$

(2) $P(S,T)x = (S^{-1}\,\square\,(D_x T^{-1} D_y))y = D_x(T^{-1}\,\square\,(D_x S^{-1} D_y))y.$

<u>Proof.</u>

$$(S\,\square\,T)x = \bigcup\{Su \cap T(x-u) \mid u \in X\} = \{v \in Y \mid v \in Su \cap T(x-u),\ u \in X\} =$$
$$\{v \in Y \mid u \in S^{-1}v,\ u \in x - T^{-1}v\} = \{v \in Y \mid S^{-1}v \cap (x - T^{-1}v) \neq \emptyset\}.$$

But $x - T^{-1}v = D_x T^{-1} D_y D_y v = D_x T^{-1} D_y (y - v)$, thus

$$(S\,\square\,T)x = \{v \in Y \mid S^{-1}v \cap D_x T^{-1} D_y (y - v) \neq \emptyset\} = P(S^{-1}, D_x T^{-1} D_y)y.$$

Using the commutativity of $\square$, (1) follows. To obtain (2), use (1) in conjunction with proposition 2.8(1), i.e., $P(S,T)x = D_x P(T,S)x$ for all $x \in X$.

QED

Let us suppose that we are interested in finding some $x \in X$ such that $Sx \cap B \neq \emptyset$ for some $S \in (2^Y)^X$ and $B \subseteq Y$. The following proposition shows how this problem is reduced to finding the fixed points, in the obvious generalized sense, of the multivalued map $P(S,T)$, for suitable choices of $T \in (2^Y)^X$.

<u>Proposition 2.11.</u> Let $T \in (2^Y)^X$, $A \subseteq X$, $B \subseteq Y$. Then

(1) $TA \subseteq B \iff \forall S \in (2^Y)^X,\ \{u \in X | u \in P(S,T)(u + A)\} \subseteq S^{-1}B$,

(2) $TA \supseteq B \iff \forall S \in (2^Y)^X,\ \{u \in X | u \in P(T,S)(u + A)\} \supseteq S^{-1}B$.

In particular, setting $A = \{0_X\}$, $B = \{0_Y\}$,

(3) $T0_X = \{0_Y\} \iff \forall S \in (2^Y)^X,\ \{u \in X | u \in P(S,T)u\} = S^{-1}0_Y$.

<u>Proof.</u> First let us note that

$$u \in P(S,T)(u + A) \iff (\exists v \in A,\ u \in P(S,T)(u + v)) \iff$$
$$(\exists v \in A,\ Su \cap Tv \neq \emptyset) \iff u \in S^{-1}TA.$$

This proves the forward direction of both (1) and (2). Let $TA \setminus B \neq \emptyset$, pick some $x_1 \in X$, let $S_1 \in (2^Y)^X$ be such that $S_1x_1 = TA \setminus B$, and for all $x \neq x_1$, $S_1x = \emptyset$. As ran $S_1 \cap B = \emptyset$,

$$\{x_1\} = S_1^{-1}(TA \setminus B) = S_1^{-1}TA = S_1^{-1}TA \setminus S_1^{-1}B \neq \emptyset.$$

This ends the proof of (1). Let $B \setminus TA \neq \emptyset$, pick some $x_2 \in X$, let $S_2 \in (2^Y)^X$ be such that $S_2x_2 = B \setminus TA$, and for all $x = x_2$, $S_2x = \emptyset$. As ran $S_2 \cap TA = \emptyset$,

$$\{x_2\} = S_2^{-1}(B \setminus TA) = S_2^{-1}B = S_2^{-1}B \setminus S_2^{-1}TA \neq \emptyset,$$

and the proof is concluded. QED

Theorem 2.12. For all $S,T \in (2^Y)^X$, $x, u \in X$, $y, v \in Y$, (1) implies (2)-(5). Furthermore, if all sets appearing in (2)-(5) are singletons, then (1)-(5) are equivalent.

(1) $v \in T(x-u)$, $u \in T^{-1}(y-v)$, $v \in Su$,

(2) $v \in (S \square T)x$, $u \in (S^{-1} \square T^{-1})y$, $x \in T^{-1}v + T^{-1}(y-v)$,

(3) $v \in (S \square T)x$, $u \in (S^{-1} \square T^{-1})y$, $y \in Tu + T(x-u)$,

(4) $u \in P(S,T)x$, $v \in P(S^{-1},T^{-1})y$, $x \in T^{-1}v + T^{-1}(y-v)$,

(5) $u \in P(S,T)x$, $v \in P(S^{-1},T^{-1})y$, $y \in Tu + T(x-u)$.

Proof. $u \in S^{-1}v$ and $x \in u + T^{-1}v$, imply $x \in (S^{-1} + T^{-1})v$. Using theorem 2.3, $v \in (S \square T)x$. By definition 2.7, $v \in Su$ and $v \in T(x-u)$ $u \in P(S,T)x$. Finally, $y \in v + Tu$ and $v \in T(x-u)$ imply $y \in Tu + T(x-u)$. This proves that (1) implies (2)-(5). Let us assume that all sets appearing in (3) are singletons (we will then use the same symbol to denote both the set and its unique element). As $u = P(S,T)x$, there is some v' $\in Y$ such that $v' \in Su$, $v' = y - Tu = T(x-u)$. Thus $v' \in (S \square T)x = \{v\}$, and (1) follows at once. Similarly, let us assume that all sets in (5) are singletons. Then so are Tu and $T(x-u)$, and $u = P(S,T)x$ implies that $T(x-u) \in Su$ or $y - Tu \in Su$, from which $y \in (S+T)u$, and via theorem 2.3, that $u \in (S^{-1} \square T^{-1})y$. Since $v = P(S^{-1},T^{-1})y$, $u \in S^{-1}v \cap T^{-1}(y-v)$. The remainder of (1) follows from $v = Tu - y = T(x-u)$. QED

3. Convex analysis. This section is an outline of some concepts of convex analysis. The presentation is informal, its main purpose being to introduce notation and terminology. Some references are Asplund (1969), Ekeland and Temam (1976), Holmes (1975), Moreau (1966), and Rockafellar (1970a, 1974).

Let X,Y be two real vector spaces in duality by means of a bilinear form $(\cdot,\cdot):X\times Y \to \mathbb{R}$, satisfying

(1) For all $x \in X$, $x \neq 0$, there is some $y \in Y$, such that $(x,y) \neq 0$.

(2) For all $y \in Y$, $y \neq 0$, there is some $x \in X$, such that $(x,y) \neq 0$.

Two real vector spaces X,Y paired as above will be denoted (X,Y). Usually Y will be a subspace of the algebraic dual X' of X, and X, Y will be canonically paired. By this we mean that (x,y) will be y(x), the value of the linear functional y at x. This is the case if X is a topological vector space (e.g., a Banach space) and Y is its (topological) dual X*. If X is a Hilbert space or $\mathbb{R}^n$, then X* can be identified with X via the inner product which will then act as the bilinear form.

A locally convex topology on X (resp. Y) is compatible with the pairing (X,Y) iff the continuous linear functionals on X (resp. Y) are precisely $\{x \to (x,y) \mid y \in Y\}$ (resp. $\{y \to (x,y) \mid x \in X\}$). By (1) and (2), such topologies are Hausdorff, and each continuous linear functional has a unique representation.

Various topologies compatible with a given duality always exist and can be generated systematically. The weak topology on X, w(X,Y), is the coarsest topology on X compatible with the pairing (X,Y). The Mackey topology on X, m(X,Y), is the topology of uniform convergence on the w(Y,X)-compact convex subsets of Y. It is the finest topology compatible with the duality. The strong topology on X, $s(X,Y)$, is the topology of uniform convergence on the $w(Y,X)$-bounded subsets of Y. $s(X,Y)$ is finer than $m(X,Y)$, it is compatible with the duality iff $s(X,Y) = m(X,Y)$.

In the reflexive case

$$s(X,Y) = m(X,Y), \quad s(Y,X) = m(Y,X).$$

If X is a Banach space and $Y = X^*$, the norm and the weak, $w(X,X^*)$, topologies on X, as well as the weak* topology on X^*, $w(X^*,X)$, are compatible with the canonical pairing (X,X^*). The norm topology on X^* is compatible iff X is reflexive. Further details can be found in Kelley, Namioka et al. (1963).

Let $\overline{\mathbb{R}}^X$ denote the set of all functions defined on X with values in $\overline{\mathbb{R}} = \mathbb{R} \bigcup \{-\infty,\infty\}$. The epigraph and strict epigraph of any $f \in \overline{\mathbb{R}}^X$ are respectively

$$\text{epi } f = \{(x,r) \in X \times \mathbb{R} \,|\, f(x) \leqq r\},$$
$$\text{sep } f = \{(x,r) \in X \times \mathbb{R} \,|\, f(x) < r\}.$$

Its effective domain is $\text{dom } f = \{x \in X \,|\, f(x) < \infty\}$. Such a function f is proper iff it is not identically ∞, and never takes the value $-\infty$. Thus dom f is nonempty and f is finite there.

A function $f \in \overline{\mathbb{R}}^X$ is convex iff its (strict) epigraph is a convex subset of $X \times \mathbb{R}$. Assuming the computation rules

$$\infty + (-\infty) = \infty, \quad 0 \cdot \infty = 0 \cdot (-\infty) = 0,$$

it follows that f is convex iff for all $x,x' \in X$; $t \in [0,1]$

$$f((1-t)x + tx') \leqq (1-t)f(x) + tf(x').$$

Let X,Y be given topologies compatible with the pairing (X,Y). The continuous affine functions on X are precisely those of the form $x \to (x,y) - r$, where $y \in Y$, $r \in \mathbb{R}$. The pointwise supremum of any collection of such affine functions is convex. The set of all such convex functions is denoted $\Gamma(X,Y)$ (simply $\Gamma(X)$ if it is clear what the pairing is), the set of convex functions defined on X which are regular with respect to the pairing (X,Y). Furthermore, $f \in \Gamma(X,Y)$ iff f is convex and lower semicontinuous (lsc) in any topology compatible with the duality (X,Y) (Moreau 1966, p 28, prop 5.d). Let $\Gamma_o(X,Y)$ denote the set of functions $\Gamma(X,Y)$ less the two constant functions $\omega_X \equiv \infty$ and $-\omega_X \equiv -\infty$. If $f \in \Gamma(X,Y)$ takes the value $-\infty$, then it has no continuous affine minorants and $f \equiv -\infty$. Thus $\Gamma_o(X,Y)$ is the set of all proper lsc (in any topology compatible with (X,Y)) convex functions defined on X.

Given $y \in Y$, $r \in \mathbb{R}$, the continuous affine function $x \to (x,y) - r$, minorizes $f \in \overline{\mathbb{R}}^X$ iff

$$r \geqq \sup \{(x,y) - f(x) | x \in X\} \stackrel{d}{=} f^*(y).$$

The maximal elements (with respect to the usual partial ordering of extended real-valued functions) of the collection of continuous affine minorants of f, are those of the form $x \to (x,y) - f^*(y)$, such that $f^*(y)$ is finite. The Fenchel transformation $f \to f^*$ defined above, maps $\overline{\mathbb{R}}^X$ onto $\Gamma(Y,X)$. Furthermore, it is a bijection between $\Gamma(X,Y)$ and $\Gamma(Y,X)$. Its inverse $g \to g^*: \Gamma(Y,X) \to \Gamma(X,Y)$, is defined by

$$g^*(x) = \sup \{(x,y) - g(y) | y \in Y\}.$$

In fact $f^{**} = (f^*)^* = f$ iff $f \in \Gamma(X,Y)$ (see Asplund 1969, p 5, th 2.10). Taking into account that $\omega_X^* = -\omega_Y$, and $(-\omega_X)^* = \omega_Y$, the Fenchel transformation is also a bijection between $\Gamma_o(X,Y)$ and $\Gamma_o(Y,X)$. A pair of functions $f \in \Gamma(X,Y)$, $g \in \Gamma(Y,X)$ such that $f = g^*$, or equivalently, $g = f^*$, are called (Fenchel) dual or conjugate.

A function $f \in \overline{\mathbb{R}}^X$ is subdifferentiable at $x_o \in X$ iff there is a continuous affine function $x \to (x,y_o) - r_o$, which minorizes f and is exact (i.e., takes the same value as f) at x_o. If a function is subdifferentiable at some point, it is proper as it can never take the value $-\infty$, and it is finite wherever it is subdifferentiable. The slope y_o is a subgradient of f at x_o. The set of such subgradients is the subdifferential of f at x_o, $\partial f(x_o)$, which can be expressed as

$$\partial f(x_o) = \{y \in Y | \forall x \in X,\ f(x) \geq f(x_o) + (x - x_o, y)\}.$$

Being the solution set of a system of continuous linear inequalities, $\partial f(x_o)$ is a closed in any topology compatible with the pairing (X,Y).

Taking into account the definition of f^*, $\partial f(x_o)$ can be written

$$\partial f(x_o) = \{y \in Y | f(x_o) + f^*(y) = (x_o,y)\}.$$

If $\partial f(x_o) \neq \emptyset$, then $f(x_o) = f^{**}(x_o)$, and the above expression of $\partial f(x_o)$ implies

$$y_o \in \partial f(x_o) \Rightarrow x_o \in \partial f^*(y_o).$$

When $f \in \Gamma_o(X,Y)$, then $f^{**} = f$ and

$$y_o \in \partial f(x_o) \Leftrightarrow x_o \in \partial f^*(y_o) \Leftrightarrow f(x_o) + f^*(y_o) = (x_o,y_o).$$

f achieves a finite global minimum at x_o iff $0 \in \partial f(x_o)$.

Let $f, g \in \overline{\mathbb{R}}^X$, their inf-convolution $f \square g$ is given by (see Moreau 1966, p 15, ch 3)

$$(f \square g)(x) = \inf \{f(u) + g(v) \,|\, u + v = x\},$$

where the convention $\infty + (-\infty) = \infty$ remains in effect. The inf-convolution is closely related to the convolution of multivalued maps $X \to 2^Y$ introduced in § 2, as it will be seen in §4. $f \square g$ is exact at x iff the infimum in its definition is actually reached. This operation is commutative, associative and has as unit the function ψ such that $\psi(0) = 0$, and $\psi(x) = \infty$ whenever $x \neq 0$.

One can show that (ibid.)

$$\mathrm{dom}\, f \square g = \mathrm{dom}\, f + \mathrm{dom}\, g, \quad \mathrm{sep}\, f \square g = \mathrm{sep}\, f + \mathrm{sep}\, g.$$

The inf-convolution of two convex functions is convex. However the inf-convolution of two proper functions need not be proper, consider in $X = \mathbb{R}$ two linear functions with different slopes. If f,g are proper and $f \square g$ is exact, then it is proper. If f,g are weakly lsc and there is an $a \in Y$ where both f^*, g^* are finite and one of them is continuous, then $f \square g$ is weakly lsc and exact. When X is barrelled, i.e., each closed convex balanced absorbing subset is a neighborhood of 0 (which happens if X is reflexive or a Banach space), $f \in \Gamma_o(X,Y)$ is continuous on int(dom f) (Rockafellar 1966a, p 61, cor 7.C). Thus if $f, g \in \Gamma_o(X,Y)$, Y is barrelled in the Mackey topology $m(Y,X)$ and $\mathrm{dom}\, f^* \cap \mathrm{int}(\mathrm{dom}\, g^*) \neq \emptyset$, then $f \square g \in \Gamma_o(X,Y)$ and is exact.

For all $f, g \in \overline{\mathbb{R}}^X$, $(f \square g)^* = f^* + g^*$, if f,g never take the value

$-\infty$. Otherwise $f \square g \equiv -\infty$. If f,g are proper convex functions and both are finite at a point where one of them is continuous then $(f + g)^* = f^* \square g^*$, and the inf-convolution $f^* \square g^*$ is exact. These properties clearly resemble those proved in theorem 2.3 for the convolution in $(2^Y)^X$.

At each $x \in X$, $f \square g$ is defined via a minimization problem. The set of solutions to such minimization is termed the set of proximal points of f to x along g. The multivalued map that assigns to each $x \in X$ the set of its proximal points will be denoted $p(f,g) \in (2^X)^X$, the proximal map of f along g,

$$p(f,g)x = \{u \in X \mid f(u) + g(x - u) \leqq (f \square g)(x) < \infty\}.$$

When f,g are convex, so is $p(f,g)x$. If in addition f,g are lsc, then $p(f,g)$ is closed. If $(f \square g)(x)$ is finite, then

$$p(f,g)x = \{u \in X \mid 0 \in \partial(f + g(x - \cdot))(u)\}.$$

Proximal maps were introduced by Moreau (1962, p 2897) in the case $X = Y = H$ a real Hilbert space, $f \in \Gamma_o(H)$, and $g: x \to \frac{1}{2}|x|^2$. If f is the indicator function of a nonempty closed convex subset A, of X, that is f is equal to zero on the set A, and equal to ∞ elsewhere, and g is as above, $(f \square g)(x) = \frac{1}{2}d(x,A)^2$, where $d(x,A)$ is the distance from x to A.

Let f be subdifferentiable. For any positive integer n, and any $n + 1$ points in the graph of ∂f

$$(x_i, y_i) \in \text{gph } \partial f \quad (i = 0,1,\ldots,n),$$

one has

$$\forall i,j \in \{0,1,\ldots,n\}, \qquad f(x_j) \geqq f(x_i) + (x_j - x_i, y_i),$$

hence

$$0 \geqq (x_0 - x_n, y_n) + (x_n - x_{n-1}, y_{n-1}) + \ldots + (x_1 - x_0, y_0).$$

A map $S \in (2^Y)^X$ satisfying this inequality for every set of $n + 1$ points in its graph will be called monotone of degree n, n-monotone for short. 1-monotone maps are called monotone. A map is cyclically monotone iff it is n-monotone for every positive integer n. The sets of n-monotone, monotone, and cyclically monotone maps, will be respectively denoted by $M^n(X,Y)$, $M(X,Y)$, $M^\infty(X,Y)$. We shall simply write $M^n(X)$ et cetera, if the particular pairing in mind is clear. If $f \in \overline{\mathbb{R}}^X$, then $\partial f \in M^\infty(X,Y)$, although it is possible that $\partial f \equiv \emptyset$.

The sets of maps $M^n(X,Y)$, $1 \leqq n \leqq \infty$, can be partially ordered by means of the inclusion of graphs. Thus given $S,T \in M^n(X,Y)$, $S \subseteqq T$ iff $\text{gph } S \subseteqq \text{gph } T$, or equivalently, iff $S^{-1} \subseteqq T^{-1}$. The maximal elements of this partial order relation are called maximal n-monotone maps, maximal monotone maps if $n = 1$, or maximal cyclically monotone maps if "$n = \infty$." The sets of such maps will be denoted $\bar{M}^n(X,Y)$, $1 \leqq n \leqq \infty$. Clearly $S \in \bar{M}^n(X,Y)$ iff $S^{-1} \in \bar{M}^n(Y,X)$. From Zorn's lemma it follows that every $S \in M^n(X,Y)$ can be extended to a $T \in \bar{M}^n(X,Y)$ which contains it.

A map $S \in (2^Y)^X$ is contained in the subdifferential map ∂f, of some $f \in \overline{\mathbb{R}}^X$, i.e., $Sx \subseteqq \partial f(x)$ for all $x \in X$, iff $S \in M^\infty(X,Y)$. If $S \in \bar{M}^\infty(X,Y)$, then $S = \partial f$, and $f \in \Gamma_o(X,Y)$. When X is a (not necessarily reflexive) real Banach space, the converse statement is true, i.e., if $f \in \Gamma_o(X,X^*)$, then $\partial f \subseteqq \bar{M}^\infty(X,X^*)$. This fails in more general spaces.

4. Convolution of monotone maps. The inf-convolution $f \square g$ of functions $f,g \in \overline{\mathbb{R}}^X$ has some of the properties of of the convolution of maps $X \to 2^Y$, especially when f, g are proper convex (see §3). There are two cyclically monotone maps associated with f,g, their respective subdifferential maps $\partial f, \partial g \in (2^Y)^X$. Therefore one can compute the inf-convolution $f \square g$ then find its subdifferential $\partial(f \square g)$, and also directly compute the convolution of the subdifferential maps, $\partial f \square \partial g$.

Theorem 4.1. (1) Let $f,g \in \overline{\mathbb{R}}^X$, then

$$P(\partial f, \partial g) \subseteq p(f,g), \quad \partial f \square \partial g \subseteq \partial(f \square g).$$

(2) Sufficient conditions for $P(\partial f, \partial g)x = p(f,g)x$ are as follows. (a) $\partial(f + g(x - \cdot)) = \partial f + \partial g(x - \cdot)$. (b) f, g are proper convex and there is a vector in X where both f, $g(x - \cdot)$ are finite, and one of them is continuous. Equivalently, there is a vector in $\operatorname{dom} f \cap (x - \operatorname{dom} g)$ (resp. $(x - \operatorname{dom} f) \cap \operatorname{dom} g$) where f (resp. g) is continuous. (c) $f,g \in \Gamma_o(X,Y)$, X is barrelled in the Mackey topology $m(X,Y)$, and $x \in \operatorname{dom} f + \operatorname{int}(\operatorname{dom} g)$, or $x \in \operatorname{int}(\operatorname{dom} f) + \operatorname{dom} g$.

(3) If $f,g \in \Gamma_o(X,Y)$, then $\partial f \square \partial g = \partial(f \square g)$ iff $\partial f^* + \partial g^* = \partial(f^* + g^*)$. Sufficient conditions for these equalities are as follows. (a) There is a vector in Y where both f^*, g^* are finite, and one of them is continuous. (b) Y is barrelled in the Mackey topology $m(Y,X)$ and $\operatorname{dom} f^* \cap \operatorname{int}(\operatorname{dom} g^*) \neq \emptyset$.

Proof. (1) Let $x \in X$,

$$u \in P(\partial f, \partial g)x \iff \partial f(u) \cap \partial g(x - u) \neq \emptyset \iff 0 \in \partial f(u) - \partial g(x - u).$$

The last inclusion implies $0 \in \partial(f + g(x - \cdot))(u)$, and $u \in p(f,g)x$. If $y \in (\partial f \,\Box\, \partial g)(x)$, there is an $u \in X$ such that $y \in \partial f(u) \cap \partial g(x - u)$. Then, for all $x', u' \in X$

$$f(u') \geqq f(u) + (u' - u, y), \quad g(x' - u') \geqq g(x - u) + (x' - u' - (x - u), y).$$

Adding both inequalities and taking the infimum over $u' \in X$ in the left side, for all $x' \in X$

$$(f \,\Box\, g)(x') \geqq f(u) + g(x - u) + (x' - x, y) \geqq (f \,\Box\, g)(x) + (x' - x, y)$$

(actually $f(u) + g(x - u) = (f \,\Box\, g)(x)$, because $u \in P(\partial f, \partial g)x \subseteqq p(f,g)x$), and $y \in \partial(f \,\Box\, g)(x)$.

(2) The sufficiency of the condition is clear from the proof of the first half of (1). When f,g are proper convex one can apply a result of Rockafellar (1966b, p 85, th 3(b)) to show that the sufficient condition holds. Finally, Rockafellar (1966a, p 61, cor 7C) allows a further refinement.

(3) If $f,g \in \Gamma_0(X,Y)$, then

$$\begin{aligned} \partial f \,\Box\, \partial g &= (\partial f^{-1} + \partial g^{-1})^{-1} = (\partial f^* + \partial g^*)^{-1} \subseteqq (\partial(f^* + g^*))^{-1} \\ &= (\partial(f \,\Box\, g)^*)^{-1} = \partial(f \,\Box\, g). \end{aligned}$$

The first equality is from theorem 2.3, the inclusion is a well known fact (see the proof of (1)), for the rest see Moreau (1966, p 38, 6.15 and p 60, 10.b). This proves the equivalence between $\partial f \,\Box\, \partial g = \partial(f \,\Box\, g)$ and $\partial f^* + \partial g^* = \partial(f^* + g^*)$. The rest is as in (2). QED

This theorem has several consequences. First, the convolution of two cyclically monotone maps is cyclically monotone. If $S,T \in M^{\infty}(X,Y)$ then there are $f,g \in \Gamma_o(X,Y)$ with $S \subseteq \partial f$, $T \subseteq \partial g$ (Rockafellar 1966c, p 500, th 1), and $S \square T \subseteq \partial f \square \partial g \subseteq \partial(f \square g)$.

Second, let X be a Banach space (thus barrelled), and $f,g \in \Gamma_o(X,X^*)$ be such that $x \in \text{dom } f + \text{int}(\text{dom } g)$, then $P(\partial f,\partial g)x = p(f,g)x$, and $P(\partial f,\partial g)x$ is a closed convex subset of X.

Third, let X be a reflexive Banach space. If $f,g \in \Gamma_o(X,X^*)$, then $f^*,g^* \in \Gamma_o(X^*,X)$ and their respective subdifferentials are all maximal monotone. The condition $\text{dom } f^* \cap \text{int}(\text{dom } g^*) \neq \emptyset$, implies that $\partial f \square \partial g = \partial(f \square g)$, and also that $f \square g \in \Gamma_o(X,X^*)$ (Rockafellar 1966b, p 85, th 3(a)), hence $\partial(f \square g) \in \bar{M}^{\infty}(X,X^*)$. Clearly $\text{dom } \partial f \subseteq \text{dom } f$, also (see Pascali and Sburlan 1978, p 27, prop 2.6) $\text{int}(\text{dom } g) \subseteq \text{dom } \partial g$, thus $\text{int}(\text{dom } g) = \text{int}(\text{dom } \partial g)$. Hence $\text{dom } \partial f^* \cap \text{int}(\text{dom } \partial g^*) \neq \emptyset$ implies that $\partial f \square \partial g \in \bar{M}^{\infty}(X,X^*)$.

The following theorem extends results to monotone maps.

<u>Theorem 4.2.</u> Let $S \in M^m(X,Y)$, $T \in M^n(X,Y)$, with $m,n \in \mathbb{Z}_+ \cup \{\infty\}$. Then $S \square T \in M^p(X,Y)$ for some $p \geq \min\{m,n\}$.

Let X be a reflexive Banach space, and let $S,T \in \bar{M}(X,X^*)$. If $\text{ran } S \cap \text{int}(\text{ran } T) \neq \emptyset$, then $S \square T \in \bar{M}(X,Y)$. If $x \in \text{dom } S + \text{int}(\text{dom } T)$, then $P(S,T)x$ is a closed convex subset of X.

<u>Proof.</u> By theorem 2.3, $(S \square T)^{-1} = S^{-1} + T^{-1}$. It is easy to show that $S^{-1} \in M^m(Y,X)$ and $T^{-1} \in M^n(Y,X)$, and that $S^{-1} + T^{-1}$ is min $\{m,n\}$-monotone, hence so is $S \square T$.

The condition on the ranges of S,T is equivalent to $\text{dom } S^{-1} \cap \text{int}(\text{dom } T^{-1}) \neq \emptyset$, which implies (Rockafellar 1970b, p 76, th 1) that $S^{-1} +$

$T^{-1} = (S\square T)^{-1}$ is maximal

The condition on the domains of S,T is equivalent to dom $S \cap (x - \operatorname{int}(\operatorname{dom} T)) \neq \emptyset$. For all $y \in Y$, $x - \operatorname{int}(\operatorname{dom} T) = \operatorname{int}(\operatorname{dom} D_y T D_x)$ (the maps D_x, D_y were introduced in proposition 2.10), and $D_y T D_x$ is clearly maximal monotone. Thus the condition is equivalent to ran $S^{-1} \cap$ int-(ran $DxT^{-1}Dy) \neq \emptyset$, which by above guarantees the maximality of $S^{-1}\square (D_x T^{-1} D_y)$. Proposition 2.10(2) implies that for all $y \in Y$, $P(S,T)x = (S^{-1}\square(D_x T^{-1} D_y))y$, as the image at any point of a maximal monotone map is closed convex (Pascali and Sburlan 1978, p 105, 2.3), so is $P(S,T)x$.

QED

A monotone map S is strictly monotone iff for all $y_i \in Sx_i$ $(i=1,2)$, if $x_1 \neq x_2$ then $(x_1 - x_2, y_1 - y_2) > 0$. Contrapositively, if $(x_1 - x_2, y_1 - y_2) = 0$, then $x_1 = x_2$. The inverse S^{-1}, of a strictly monotone map S, is single-valued. Otherwise there would exist $x_1, x_2 \in X$, $y \in Y$, satisfying $x_1, x_2 \in S^{-1}y$, $x_1 \neq x_2$, but then $(x_1 - x_2, y - y) = 0$ contradicting the strict monotonicity of S. The converse statement is true in $X = \mathbb{R}$ but fails elsewhere. Consider in $X = \mathbb{R}^2$ the linear map S which rotates vectors counterclockwise through an angle of $\pi/2$. Clearly both S, $S^{-1} = -S$, are monotone and single -valued, but neither is strictly monotone.

<u>Proposition 4.3.</u> For all $S,T \in M(X,Y)$,

(1) If S^{-1} or T^{-1} is strictly monotone, then $S\square T$ is single-valued.

(2) If S or T is strictly monotone, then $P(S,T)$ and $P(T,S)$ are single-valued.

(3) Statements (1) and (2) are dual, and their converse statements are false.

(4) If S and T are strictly monotone, so is $S\square T$.

(5) If S^{-1} and T^{-1} are strictly monotone, so is $(S+T)^{-1}$.

(6) The strict monotonicity of either S or T (resp. S^{-1} or T^{-1}) is not sufficient for the strict monotonicity of $S\square T$ (resp. $(S+T)^{-1}$). Furthermore, the statements converse to (4), and to (5), are false.

<u>Proof.</u> (1) Let $y_1,y_2 \in (S\square T)x$ with $y_1 \neq y_2$. Then there are u_1, $u_2 \in X$ such that $y_i \in Su_i \cap T(x-u_i)$ $(i=1,2)$. As $S,T \in M(X,Y)$,

$$(u_1-u_2,y_1-y_2) \geqq 0, \text{ and } (x-u_1-(x-u_2),y_1-y_2) \geqq 0,$$

thus $(u_1-u_2,y_1-y_2) = 0$. As $y_1 \neq y_2$, neither S^{-1} nor T^{-1} can be strictly monotone.

(2) Let $u_1,u_2 \in P(S,T)x$ with $u_1 \neq u_2$. Let $y_i \in Su_i \cap T(x-u_i)$ $(i = 1,2)$. One has

$$0 = (u_1- u_2,y_1-y_2) + (x-u_1-(x-u_2),y_1-y_2).$$

$S,T \in M(X,Y)$, hence both terms in the sum vanish. As $u_1 \neq u_2$ neither S nor T can be strictly monotone.

(3) Let $x,u_1,u_2,z_1,z_2 \in X$, $y,v_1,v_2,w_1,w_2 \in Y$ satisfy

$$x = u_i + z_i, \quad y = v_i + w_i \qquad (i = 1,2).$$

The strict monotonocity of S is equivalent to

$$(w_i \in (D_ySD_x)z_i \ (i = 1,2),\ z_1 \neq z_2) \Rightarrow (z_1-z_2,w_1-w_2) > 0,$$

which holds iff D_ySD_x is strictly monotone. Mutatis mutandi one proves that $S^{-1}$is strictly monotone iff $D_xS^{-1}D_y$ is so. From proposition 2.10

$$(S\square T)x = P(S^{-1},D_x T^{-1}D_y)y, \qquad P(S,T)x = (S^{-1}\square(D_x T^{-1}D_y))y.$$

If S^{-1} or T^{-1} is strictly monotone, then, respectively, S^{-1} or $D_x T^{-1}D_y$ is so. By (2) $P(S^{-1},D_x T^{-1}D_y)y$ is either a singleton or empty, and so is $(S\square T)x$ as (1) asserts. Similarly, if S or T is strictly monotone, then, respectively S or $D_y SD_x$ is so. By (1) $S^{-1}\square(D_x T^{-1}D_y)$ is either a singleton or empty, and so is $P(S,T)x$ as (2) asserts. Let $X = \mathbb{R}^2$, and let S rotate vectors counterclockwise through an angle of $\pi/2$. Then S and S^{-1} are single-valued, and so is $S\square S = (S^{-1}+S^{-1})^{-1} = (S2)$. But neither S nor S^{-1} is strictly monotone. Also for all $x \in \mathbb{R}^2$,

$$P(S,S)x = \{u \in \mathbb{R}^2 \mid Su = S(x-u)\} = \{2^{-1}x\},$$

thus $P(S,S)$ is single-valued whilst S is not strictly monotone.

(4) Let us suppose that $S\square T$ is not strictly monotone, then there are $x_1,x_2 \in X$, $x_1 \neq x_2$, and $y_i \in (S\square T)x_i$ $(i = 1,2)$, such that $(x_1 - x_2, y_1 - y_2) = 0$. Let $u_1,u_2 \in X$ be such that $y_i \in Su_i \quad T(x-u_i)$ $(i=1,2)$, then

$$0 = (x_1-x_2,y_1-y_2) = (u_1-u_2,y_1-y_2) + (x_1-u_1-(x_2-u_2),y_1-y_2).$$

Being S,T monotone, it follows that both terms in the right side vanish. If S is strictly monotone, then $u_1 = u_2$ and $x_1 - u_1 \neq x_2 - u_2$, which implies that T is not strictly monotone. If T is strictly monotone, then $u_1 - u_2 = x_1 - x_2 \neq 0$, and S is not strictly monotone.

(5) Apply (4) to $S^{-1}\square T^{-1} = (S + T)^{-1}$.

(6) In $X = \mathbb{R}$, consider I the identity map, and T given by

$$\text{gph } T = \{(x,x) \mid x \leqq 0\} \bigcup \{(x,0) \mid 0 \leqq x \leqq 1\} \bigcup \{(x,x-1) \mid 1 \leqq x\}.$$

Then for all $x \in [0,1]$, $(I \square T)x = \{0\}$. Again in $X = \mathbb{R}^2$, consider the map S introduced in (3) and the identity map I. An easy calculation shows that $I \square S = (I + S^{-1})^{-1} = (I - S)^{-1} = 2^{-1}(I + S)$ is strictly monotone without S being so. QED

Let X, Y be real normed spaces whose norms are both denoted $|\cdot|$. The bilinear form pairing X and Y, and the norms on both X and Y, are related by the Cauchy-Buniakowsky inequality $|(x,y)| \leqq |x| \cdot |y|$, for all $x \in X$, $y \in Y$. Typically X will be a real Banach space and Y will be a linear subspace of its dual X^*.

Definition 4.4. Let $S \in (2^Y)^X$. Let

$$\Lambda S = \{\lambda \in [0,\infty) \mid \forall (x,y),(x',y') \in \text{gph } S,\ |y - y'| \leqq \lambda |x - x'|\}.$$

The modulus of Lipschitz continuity of S is $\lambda S = \inf \Lambda S$, where one sets $\lambda S = \infty$ if $\Lambda S = \emptyset$.

Definition 4.5. Let $S \in M(X,Y)$. Let

$$MS = \{\mu \in [0,\infty) \mid \forall (x,y),(x',y') \in \text{gph } S,\ \mu |x - x'|^2 \leqq (x - x', y - y')\}.$$

The modulus of strong monotonicity of S is $\mu S = \sup MS$.

For any set A, let $|A|$ denote its cardinal number. If $|\text{gph } S| \leqq 1$, then $\Lambda S = [0,\infty)$, $MS = [0,\infty)$, and $\lambda S = 0$, $\mu S = \infty$. If $|\text{gph } S| > 1$ but $|\text{dom } S| = 1$, then $\Lambda S = \emptyset$, $MS = [0,\infty)$ and $\lambda S = \mu S = \infty$. Most results proved below depend on the assumption that gph S contains more than one point. Maps $S \in (2^Y)^X$ such that $|\text{gph } S| \leqq 1$ will be called trivial.

These definitions of the Lipschitz continuity and strong monotonicity moduli have several advantages over the usual ones (see Dolezal 1979, p 345, (36),(37))

$$\lambda' S = \sup\left\{\left|\frac{y - y'}{x - x'}\right| \,\middle|\, (x,y),(x',y') \in \text{gph } S,\ x \neq x'\right\},$$

$$\mu' S = \inf\left\{\frac{(x - x', y - y')}{|x - x'|^2} \,\middle|\, (x,y),(x',y') \in \text{gph } S,\ x \neq x'\right\}.$$

First, they allow the tratment of the case in which dom S is a singleton in a more satisfactory manner. As the set in the definition of $\mu' S$ is empty, $\mu' S = \infty$ which coincides with the value of μS as seen above. However, the set in the definition of $\lambda' S$ is also empty and there is no way to assign to $\lambda' S$ the value ∞, which is the most natural, when ran S is not a singleton and thus S is not single-valued. Second, the sets ΛS, MS are defined as the solution sets to systems of linear inequalities, thus they are closed and if not empty they contain λS, μS respectively. Third, they make the proof of the following propositions easier.

Proposition 4.6. (1) Let $S \in (2^Y)^X$ be nontrivial, if $\lambda S < \infty$, then S is single-valued, thus $|\text{dom } S| = |\text{gph } S| > 1$, and $\Lambda S = [\lambda S, \infty)$. If $\lambda S = 0$, then $|\text{ran } S| \leqq 1$.

(2) Let $S \in M(X,Y)$. If $\mu S > 0$, then S^{-1} is single-valued. Also,

$$\mu S < \infty \iff |\text{dom } S| > 1 \iff MS = [0, \mu S].$$

(3) Let $S \in M(X,Y)$ be nontrivial, then $\mu S \leqq \lambda S$, and the bound is reached.

Proof. (1) If $y, y' \in Sx$ and $y \neq y'$, ΛS is empty and $\lambda S = \infty$. Clearly

$$\infty > \lambda' \geqq \lambda \in \Lambda S \quad \Rightarrow \quad \lambda' \in \Lambda S,$$

and S is closed as seen above. Thus if $\Lambda S \neq \emptyset$, which happens if $\lambda S < \infty$, it contains its infimum, and $\Lambda S = [\lambda S, \infty)$. Let $\lambda S = 0 < \infty$, then $0 \in \Lambda S$ and for any pair $(x,y),(x',y') \in \text{gph } S$ one has $|y-y'| \leqq 0|x-x'| = 0$, thus $|\text{ran } S| \leqq 1$.

(2) If $\mu S > 0$, S is strictly monotone and S^{-1} is single-valued. Also if $y \in Sx \cap Sx'$, $x \neq x'$, the only solution in $[0,\infty)$ to

$$\mu|x - x'|^2 \leqq (x - x', y - y) = 0$$

is $\mu = 0$, thus $MS = \{0\}$ and $\mu S = 0$. Clearly,

$$0 \leqq \mu' \leqq \mu \in MS \quad \Rightarrow \quad \mu' \in MS,$$

MS is closed as seen above, and if bounded from above it contains its supremum, thus

$$\mu S = \sup MS < \infty \quad \Leftrightarrow \quad MS = [0, \mu S] \subseteqq [0,\infty).$$

If $|\text{dom } S| \leqq 1$, then $\mu S = \infty$. If $|\text{dom } S| > 1$, picking $x \neq x'$ in dom S and $y \in Sx$, $y' \in Sx'$, any μ in MS has to satisfy

$$\mu \leqq \frac{(x - x', y - y')}{|x - x'|^2} < \infty,$$

thus $\mu S < \infty$.

(3) Assume $\lambda S < \infty$, pick $\lambda \in \Lambda S = [\lambda S, \infty) \neq \emptyset$ by (1), $\mu \in MS \neq \emptyset$ as $0 \in MS$ always, and $x \neq x'$ in dom S which is possible as S is single-valued by (1) and $|\text{dom } S| = |\text{gph } S| > 1$ by the hypothesis. If $y \in Sx$, $y' \in Sx'$, $|x-x'|^2 \leqq (x-x', y-y') \leqq \mu|x-x'|^2$, and $\mu \leqq \lambda$. Therefore

$\mu S = \sup MS \leqq \inf \Lambda S = \lambda S$. To prove that the bound is reached, let $X = Y = H$ a real inner product space with identity map I, select any $S = aI$ with $a > 0$. QED

The objective is to find bounds for $\lambda(S \square T)$, $\mu(S \square T)$, $\lambda P(S,T)$. Thus we need bounds for λS^{-1}, μS^{-1}, $\lambda(S + T)$, $\mu(S + T)$, et cetera.

<u>Proposition 4.7.</u> Let $S \in M(X,Y)$ be nontrivial, then

(1) $(\lambda S)^{-1} \leqq \lambda S^{-1} \leqq (\mu S)^{-1}$,

(2) $\mu S(\lambda S)^{-2} \leqq \mu S^{-1} \leqq (\mu S)^{-1}$,

where $\mu S(\lambda S)^{-2} = \infty$ if $\mu S = \lambda S = 0$, and $\mu S(\lambda S)^{-2} = 0$ if $\mu S = \lambda S = \infty$. Furthermore, the bounds are reached.

<u>Proof.</u> (1) If λS or $\lambda S^{-1} = \infty$, then $(\lambda S)^{-1} \leqq \lambda S^{-1}$. Let $\lambda S < \infty$, then by proposition 4.6(1), $\lambda S \in \Lambda S$ and $|\text{dom } S| = |\text{gph } S| > 1$. Choose $(x,y),(x',y') \in \text{gph } S$, then $|y - y'| \leqq \lambda S|x - x'|$. If $\lambda S = 0$, $|\text{ran } S| = |\text{dom } S^{-1}| = 1$, as $|\text{dom } S| = |\text{ran } S^{-1}| > 1$, S^{-1} is not single-valued and $\lambda S^{-1} = \infty$ (proposition 4.6(1)). If $0 < \lambda S < \infty$ and $\lambda S^{-1} < \infty$, $\lambda S^{-1} \in \Lambda S^{-1}$. As $|\text{dom } S| > 1$, picking $x \neq x'$ above, yields $|y - y'| \leqq \lambda S|x - x'|$ and $|x - x'| \leqq \lambda S^{-1}|y - y'|$. From $x \neq x'$ and $\lambda S > 0$, it follows that $(\lambda S)^{-1} \leqq \lambda S^{-1}$. If $\mu S = 0$ there is nothing to prove, so let $\mu S > 0$. If $\mu S = \infty$, $|\text{dom } S| = |\text{ran } S^{-1}| = 1$ (proposition 4.6(2)), thus $\Lambda S^{-1} = [0,\infty)$ and $\lambda S^{-1} = 0$. If $\mu S < \infty$, pick $(x,y),(x',y') \in \text{gph } S$. As $\mu S \in MS$ (proposition 4.6(2)), $\mu S|x - x'|^2 \leqq (x - x',y - y') \leqq |x - x'|\cdot|y - y'|$. In any case $\mu S|x - x'| \leqq |y - y'|$, and as $\mu S > 0$, $|x - x'| \leqq (\mu S)^{-1}|y - y'|$, which implies $(\mu S)^{-1} \in \Lambda S^{-1}$ and $\lambda S^{-1} = \inf \Lambda S^{-1} \leqq (\mu S)^{-1}$.

(2) $|\text{gph } S^{-1}| = |\text{gph } S| > 1$, applying proposition 4.6(3) and (1),

$\mu S^{-1} \leqq \lambda S^{-1} \leqq (\mu S)^{-1}$. If $\lambda S = 0$, $|\text{ran } S| = |\text{dom } S^{-1}| = 1$ and by proposition 4.6(2), $\mu S^{-1} = \infty$. If $0 < \lambda S < \infty$, S is single-valued, using $|\text{gph } S| > 1$, one shows $|\text{dom } S| > 1$ and $\mu S < \infty$ (proposition 4.6(1,2)). Select $(x,y),(x',y') \in \text{gph } S$, as $\lambda S \in \Lambda S$, $\mu S \in MS$, by proposition 4.6(1,2),

$$\mu S|x - x'|^2 \leqq (x - x', y - y'), \qquad (\lambda S)^{-1}|y - y'| \leqq |x - x'|,$$

thus $\mu S(\lambda S)^{-2}|y - y'|^2 \leqq (x - x', y - y')$ and $\mu S(\lambda S)^{-2} \in MS^{-1}$ which implies $\mu S(\lambda S)^{-2} \leqq \sup MS^{-1} = \mu S^{-1}$. If $\mu S = \infty$, by proposition 4.6(2), $|\text{dom } S| = |\text{ran } S^{-1}| = 1$, as $|\text{gph } S| > 1$, $\lambda S^{-1} = 0$. Also (proposition 4.6(3)) $\lambda S \geqq \mu S = \infty$, $0 = \lambda S^{-1} \geq \mu S^{-1}$. As $\mu S \leq \lambda S$, $\mu S(\lambda S)^{-2} \leq (\lambda S)^{-1} = 0$.

To prove that the bounds are reached, it suffices to take in $X = Y = H$, a real inner product space, $S = aI$, $T = bI$, where I is the identity map of H and $a,b > 0$. QED

Proposition 4.8. For all $S,T \in (2^Y)^X$, $\lambda(S + T) \leq \lambda S + \lambda T$. For all $S,T \in M(X,Y)$, $\mu(S + T) \geq \mu S + \mu T$. Furthermore, the bounds are reached.

Proof. If $|\text{gph } (S + T)| \leq 1$, then $\lambda(S + T) = 0$ as seen above. If $|\text{gph } (S + T)| > 1$ and $|\text{dom } (S + T)| = 1$, either S or T is multivalued on dom $(S + T)$, and $\lambda S + \lambda T = \infty$ by proposition 4.6(1). If $|\text{dom } (S + T)| > 1$, assume $\lambda S + \lambda T < \infty$, pick $(x,y),(x',y') \in \text{gph } S$, $(x,z),(x',z') \in \text{gph } T$,

$$|y + z - (y' + z')| \leq |y - y'| + |z - z'| \leq (\lambda S + \lambda T)|x - x'|.$$

Thus $\lambda S + \lambda T \in \Lambda(S + T)$ and the bound follows.

If $|\text{dom } (S + T)| \leq 1$ then $\mu(S + T) = \infty$ as seen above. Otherwise $|\text{dom } S|$, $|\text{dom } T| > 1$ which by proposition 4.6(2) implies $\mu S \in MS$, $\mu T \in$

MT, $M(S+T) = [0,\mu(S+T)]$. Select vectors x,x',y,y',z,z' as above, then

$$(\mu S + \mu T)|x - x'|^2 \leq (x - x', y + z - (y' + z')),$$

and $\mu S + \mu T \in M(S + T)$ from which the bound follows.

To prove that the bounds are reached, proceed as in proposition 4.7.

QED

Theorem 4.9. Let $S,T \in M(X,Y)$ be such that S, T, $S\square T$ are nontrivial. Then

$$\frac{1}{(\mu S)^{-1} + (\mu T)^{-1}} \leq \lambda(S\square T) \leq \frac{1}{\mu S(\lambda S)^{-2} + \mu T(\lambda T)^{-2}},$$

$$\frac{\mu S(\lambda S)^{-2} + \mu T(\lambda T)^{-2}}{((\mu S)^{-1} + (\mu T)^{-1})^2} \leq \mu(S\square T) \leq \frac{1}{\mu S(\lambda S)^{-2} + \mu T(\lambda T)^{-2}}.$$

To evaluate the bounds when one or more of the moduli is either zero or infinite, replace every modulus equal to zero by r, every modulus equal to infinite by r^{-1}, and take the limit as $r \to 0$. The bounds are tight.

Proof. By theorem 2.3, $S\square T = (S^{-1} + T^{-1})^{-1}$, and $\lambda(S \square T) = \lambda(S^{-1} + T^{-1})^{-1}$. Being $S^{-1} + T^{-1}$ nontrivial, proposition 4.7(1) yields

$$(\lambda(S^{-1} + T^{-1}))^{-1} \leq \lambda(S\square T) \leq (\mu(S^{-1} + T^{-1}))^{-1}.$$

From propositions 4.7 and 4.8

$$\lambda(S^{-1} + T^{-1}) \leq \lambda S^{-1} + \lambda T^{-1} \leq (\mu S)^{-1} + (\mu T)^{-1},$$

$$\mu(S^{-1} + T^{-1}) \geq \mu S^{-1} + \mu T^{-1} \geq \mu S(\lambda S)^{-2} + \mu T(\lambda T)^{-2},$$

and the bounds on $\lambda(S\square T)$ follow.

Using proposition 4.6(3) and the above upper bound on $\lambda(S\square T)$

$$\mu(S\square T) \le (S\square T) \le (\mu S(\lambda S)^{-2} + \mu T(\lambda T)^{-2})^{-1}.$$

From proposition 4.7(2)

$$\mu(S\square T) = \mu(S^{-1} + T^{-1})^{-1} \ge \mu(S^{-1} + T^{-1})(\lambda(S^{-1} + T^{-1}))^{-2},$$

the two factors in the last term can be transformed, as done above, to obtain the final expression of the lower bound on $\mu(S\square T)$.

The procedure for resolution of indeterminate expressions works for both upper bounds by proposition 4.7(2). The lower bound of $\mu(S\square T)$, see above, comes from an expression of the type $\mu(\cdot)(\lambda(\cdot))^{-2}$. The above procedure is equivalent to evaluate $\mu(\cdot) \ge \mu S(\lambda S)^{-2} + \mu T(\lambda T)^{-2}$, and then $\mu(\cdot)(\lambda(\cdot))^{-2}$, by the rules of proposition 4.7(2).

Taking $X = Y = H$ a real inner product space, $S = aI$, $T = bI$, where $a,b > 0$ and I is the identity map of H, it is easy to show that all bounds are reached. QED

<u>Proposition 4.10.</u> Let X,Y be normed spaces, and $S,T \in (2^Y)^X$. Then $S \subseteq T$ implies $\lambda S \le \lambda T$.

Let X,Y,Z be normed spaces, $S \in (2^Y)^X$ and $T \in (2^Z)^Y$. If $T\circ S \in (2^Z)^X$ is nontrivial, then $\lambda(T\circ S) \le \lambda T\cdot\lambda S$, where by convention, $\lambda T\cdot\lambda S = 0$ whenever $\lambda T = 0$, and $\lambda T\cdot\lambda S = \infty$ if $\lambda S = 0$ and $\lambda T = \infty$. The bound is reached.

<u>Proof.</u> By definition 4.4, $S \subseteq T$ implies $\Lambda T \subseteq \Lambda S$, and $\lambda S = \inf \Lambda S \le \inf \Lambda T = \lambda T$.

The proof of the second part will be broken up into several cases. If $\lambda T = 0$ then ran T is a singleton by proposition 4.6(1), and ran $T\circ S \subseteq$

ran T is also a singleton. Being $T \circ S$ nontrivial, $|\text{gph } T \circ S| > 1$, thus $|\text{dom } T \circ S| > 1$ and picking two different points in dom $T \circ S$, one gets $\lambda T \circ S = 0$. If $\lambda T \in (0,\infty)$ and $\lambda S \in [0,\infty)$, let $y_i \in Sx_i$, $z_i \in Ty_i$ $(i = 1,2)$. By Proposition 4.6(1) $\lambda S \in \Lambda S$ and $\lambda T \in \Lambda T$, thus

$$|z_1 - z_2| \leq \lambda T|y_1 - y_2| \leq \lambda T \cdot \lambda S|x_1 - x_2|,$$

hence $\lambda T \cdot \lambda S \in \Lambda T \circ S$ and $\lambda T \circ S = \inf \Lambda T \circ S \leq \lambda T \cdot \lambda S$. If $\lambda S, \lambda T > 0$ and $\max\{\lambda S, \lambda T\} = \infty$, there is nothing to prove. The only case left is $\lambda S = 0$, $\lambda T = \infty$, in which it is easy to construct examples with $\lambda T \circ S = \infty$. Take $S \equiv \{y\}$ on X with $y \in \text{dom } T$, and T multivalued at y, then for any $x \in X$, $T \circ Sx = Ty$, and $\lambda T \circ S = \infty$. To show that the bound is reached, let $X = Y = Z$, let I be the identity map in X, $S = aI$, $T = bI$ with $a,b > 0$. QED

<u>Theorem 4.11.</u> Let $S,T \in M(X,Y)$ be such that $S, T, S\square T, S^{-1}(S\square T)$, $S \circ P(S,T)$ are nontrivial. Then

$$\frac{(\lambda S)^{-1}}{(\mu S)^{-1} + (\mu T)^{-1}} \leq \lambda P(S,T) \leq \frac{(\mu S)^{-1}}{\mu S(\lambda S)^{-2} + \mu T(\lambda T)^{-2}},$$

whenever the expressions giving the bounds are defined. If in addition T is linear and T + S is nontrivial

$$\lambda P(S,T) \leq \frac{\lambda T}{\lambda S + \lambda T},$$

whenever the right side is defined. All bounds are reached.

<u>Proof.</u> By proposition 2.8(1), $P(S,T) \subseteqq S^{-1} \circ (S\square T)$, $S\square T \subseteqq S \circ P(S,T)$. If the right sides are nontrivial, the above proposition yields

$$\lambda P(S,T) \leq \lambda S^{-1} \cdot \lambda(S \square T), \quad \lambda(S \square T) \leq \lambda S \cdot \lambda P(S,T).$$

If S is nontrivial and $\lambda S > 0$, proposition 4.7(1) gives

$$(\lambda S)^{-1} \cdot \lambda(S \square T) \leq \lambda P(S,T) \leq (\mu S)^{-1} \cdot \lambda(S \square T).$$

If in addition T, $S \square T$ are nontrivial, theorem 4.9 yields the bounds on $\lambda P(S,T)$. By proposition 2.8(3) and the linearity of T

$$P(S,T) = (I + T^{-1}S)^{-1} \underline{\underline{\subset}} (T^{-1}(T + S))^{-1} = (T + S)^{-1}T.$$

By the above proposition, the nontriviality of T + S, and proposition 4.8

$$\lambda P(S,T) \leq \lambda T \cdot (\mu(S + T))^{-1} \leq \frac{\lambda T}{\mu S + \mu T}.$$

To prove that the bounds are reached, take X = Y = H a real inner product space with identity map I, and S = aI, T = bI with a,b > 0. QED

5. Duality theorems. Given two real vector spaces X, Y in duality, and $f \in \Gamma_o(X,Y)$, the vectors $u \in X$, $v \in Y$ are said to be conjugate with respect to f iff $f(u) + f^*(v) = (u,v)$. Moreau (1962, p 2897) characterized such pairs of points, when $X = Y = H$ a real Hilbert space, by proving that the following two statements are equivalent,

(I) $\quad f(u) + f^*(v) = (u,v), \quad x = u + v,$

(II) $\quad u = p(f,q)x, \quad v = p(f^*,q)x,$

where $q = \frac{1}{2}|\cdot|^2$. The object of this section is to prove generalizations in several directions of the above and related results (Wexler 1972, p 1328, th 2). By using to the fullest extent theorem 2.12, we are able to prove the results in this section in an elementary and unified fashion.

Theorem 5.1. Let $S,T \in M(X,Y)$, and let T,T^{-1} be strictly monotone. Then (1)-(5) are equivalent.

(1) $\quad v \in T(x-u), \quad u \in T^{-1}(y-v), \quad v \in Su,$

(2) $\quad v \in (S \square T)x, \quad u \in (S^{-1} \square T^{-1})y, \quad x \in T^{-1}v + T^{-1}(y-v),$

(3) $\quad v \in (S \square T)x, \quad u \in (S^{-1} \square T^{-1})y, \quad y \in Tu + T(x-u),$

(4) $\quad u \in P(S,T)x, \quad v \in P(S^{-1},T^{-1})y, \quad x \in T^{-1}v + T^{-1}(y-v),$

(5) $\quad u \in P(S,T)x, \quad v \in P(S^{-1},T^{-1})y, \quad y \in Tu + T(x-u).$

Proof. As T, T^{-1} are strictily monotone both are univoque, and by proposition 4.3 so are $S \square T$, $S^{-1} \square T^{-1}$, $P(S,T)$, $P(S^{-1},T^{-1})$. Apply theorem 2.12. QED

To assume that T, T^{-1} are univoque is not enough. Let $X = \mathbb{R}^2$, S

rotate vectors counterclockwise by an angle of $\pi/2$, and $T = S^{-1} = -S$. THen (1) and (2) of the theorem read

$$(1) \quad v = S(x - u), \qquad u = S(y - v), \qquad v = Su,$$

$$(2) \quad v \in (S \square (-S))x, \quad u \in ((-S) \square S)y, \quad x = Sv + S(y - v).$$

For any $x \in \mathbb{R}^2$

$$(S \square (-S))x = \bigcup \{Su \cap -S(x-u) \mid u \in \mathbb{R}^2\} =$$

$$\{Su \mid Sx = 0,\ u \in \mathbb{R}^2\} = \begin{cases} \mathbb{R}^2 & \text{if } x = 0, \\ \emptyset & \text{if } x \neq 0, \end{cases}$$

and analogously for $((-S) \square S)y$. The vectors $x,y,u,v \in \mathbb{R}^2$ satisfying (2) are $x = 0 = y$; $u,v \in \mathbb{R}^2$, which do not necessarily satisfy (1) unless $v = Su$.

A convex function is strictly convex iff for all $x,x' \in \operatorname{dom} f$, for all $t \in (0,1)$, $f((1 - t)x + tx') < (1 -t)f(x) + tf(x')$. The relation between the strict convexity of f and the strict monotonicity of ∂f is as follows.

Lemma 5.2. Let $f \in \overline{\mathbb{R}}^X$ be convex. If f is strictly convex, then ∂f is strictly monotone. If ∂f is strictly monotone, then f is strictly convex on any convex subset of $\operatorname{dom} \partial f$.

Proof. Let $y_i \in \partial f(x_i)$ $(i = 1,2)$ with $x_1 \neq x_2$. As f is convex and finite at x_1,

$$\forall t \in [0,1], \quad f(t(x_2 - x_1) + x_1) - f(x_1) \leq t(f(x_2) - f(x_1)).$$

The definition of subgradient implies

$$f(t(x_2 - x_1) + x_1) - f(x_1) \geq t(x_2 - x_1, y_1),$$
$$t(f(x_2) - f(x_1)) \leq t(x_2 - x_1, y_2).$$

If $(x_1 - x_2, y_1 - y_2) = 0$, f is not strictly convex on $[x_1, x_2]$.

Let f be not strictly convex on some convex subset of dom ∂f. Then there are $x_1, x_2 \in$ dom ∂f with $x_1 \neq x_2$ such that $[x_1, x_2] \subseteqq$ dom ∂f and

$$\forall t \in [0,1], \quad f((1 - t)x_1 + tx_2) = (1 - t)f(x_1) + tf(x_2).$$

Let $y_t \in \partial f(x_t)$, where $x_t = (1 - t)x_1 + tx_2 \in$ dom ∂f, for some $t \in [0,1]$, then

$$f(x_2) - f(x_t) \geq (x_2 - x_t, y_t), \quad f(x_1) - f(x_t) \geq (x_1 - x_t, y_t).$$

Replacing x_t by its expression as a convex combination of x_1 and x_2, and using the fact that f is not strictly convex on $[x_1, x_2]$, one can easily obtain $f(x_1) - f(x_2) = (x_1 - x_2, y_t)$. With this equality, and substituting for x_t, $f(x_t)$ their expressions as convex combinations of x_1, x_2, and $f(x_1), f(x_2)$, respectively, in

$$\forall z \in X, \quad f(z) \geq f(x_t) + (z - x_t, y_t),$$

one gets

$$\forall z \in X, \quad f(z) \geq f(x_i) + (z - x_i, y_t) \quad (i = 1,2).$$

Thus $y_t \in \partial f(x_1) \cap \partial f(x_2)$, and it follows that ∂f is not strictly monotone. QED

<u>Theorem 5.3.</u> Let X be a reflexive topological vector space. Let $g \in \Gamma_o(X, X^*)$ be strictly convex with strictly convex dual $g^* \in \Gamma_o(X^*, X)$.

Then for all $f \in \Gamma_o(X,X^*)$, $x \in X$, $y \in X^*$ such that

$$\operatorname{dom} f \cap \operatorname{int}(\operatorname{dom} g) \neq \emptyset, \qquad x \in \operatorname{dom} f + \operatorname{int}(\operatorname{dom} g),$$

$$\operatorname{dom} f^* \cap \operatorname{int}(\operatorname{dom} g^*) \neq \emptyset, \qquad y \in \operatorname{dom} f^* + \operatorname{int}(\operatorname{dom} g^*),$$

where the interiors are in the strong topologies, and any $u \in X$, $v \in X^*$, statements (1)-(5) are equivalent.

(1) $v \in \partial g(x - u)$, $\quad u \in \partial g^*(y - v)$, $\quad v \in \partial f(u)$,

(2) $v \in \partial(f \square g)(x)$, $\quad u \in \partial(f^* \square g^*)(y)$, $\quad x \in \partial g^*(v) + \partial g^*(y - v)$,

(3) $v \in \partial(f \square g)(x)$, $\quad u \in \partial(f^* \square g^*)(y)$, $\quad y \in \partial g(u) + \partial g(x - u)$,

(4) $u \in p(f,g)x$, $\quad v \in p(f^*,g^*)y$, $\quad x \in \partial g^*(v) + \partial g^*(y - v)$,

(5) $u \in p(f,g)x$, $\quad v \in p(f^*,g^*)y$, $\quad y \in \partial g(u) + \partial g(x - u)$.

<u>Proof.</u> That (1) implies (2)-(5) follows from definitions 2.1, 2.7 and theorem 4.1. To prove that any of (2)-(5) implies (1), one notices the following facts. First, the conditions imposed on the domains of f, g, f^*, g^* imply, by theorem 4.1, that

$$\partial(f \square g) = \partial f \square \partial g, \qquad \partial(f^* \square g^*) = \partial f^* \square \partial g^*,$$

$$p(f,g)x = P(\partial f, \partial g)x, \qquad p(f^*,g^*)y = P(\partial f^*, \partial g^*)y.$$

Second, by lemma 5.2, both $\partial g, \partial g^*$ are strictly monotone. The result follows from theorem 5.1. QED

Let X be a real Banach space with dual X^*. The norms on X and X^* will be denoted $|\cdot|$ and $|\cdot|_*$ respectively. If confusion does not arise, we will write $|\cdot|$ for $|\cdot|_*$. The normalized duality map of X, $J: X \to 2^{X^*}$ is given, for all $x \in X$, by

$$Jx = \{x^* \in X^* \mid (x,x^*) = |x|^2 = |x^*|^2\}.$$

It is straightforward that J can be also defined as the subdifferential of $q = \frac{1}{2}|\cdot|^2 \in \Gamma_o(X,X^*)$ (see Pascali and Sburlan 1978, p 109, §2.6),

$$Jx = \{x^* \in X^* \mid \forall y \in X,\ q(y) \geqq q(x) + (y - x,x^*)\}.$$

A Banach space is locally uniformly convex if for any $\varepsilon \in (0,2]$, $x \in X$ with $|x| = 1$, there exists a $\delta > 0$ such that whenever $y \in X$ with $|y| = 1$, $|x - y| \geqq \varepsilon$ implies $|x + y| \leqq 2(1 - \delta)$. Troyanski (1971, p 177, th 1) has proved that a reflexive Banach space can be renormed so that X and X* are locally uniformly convex. In what follows we will assume that every reflexive Banach space has been so renormed.

If X is reflexive, the normalized duality map of X*, J* has as expression for all $x^* \in X^*$

$$J^*x^* = \{x \in X \mid (x,x^*) = |x^*|^2 = |x|^2\} = J^{-1}(x^*).$$

This also follows from the fact that $J^* = \partial(\frac{1}{2}|\cdot|_*^2)$, and (Asplund 1969, p 15) $q^* = (\frac{1}{2}|\cdot|^2)^* = \frac{1}{2}|\cdot|_*^2$. Hence J, J^{-1} will have the same properties whenever $|\cdot|$ and $|\cdot|_*$ do so. The main facts about J (and J^{-1}) are summarized in the following

Theorem 5.4. Let X be a reflexive real Banach space which is normed so that X and X* are both locally uniformly convex. Let J be the normalized duality map of X. Then J (and J^{-1}) has the following properties.

(1) J is a bounded homeomorphism between $(X,|\cdot|)$ and $(X^*,|\cdot|_*)$.

(2) J is homogeneous of degree 1, i.e., $\forall x \in X$, $\forall r \in \mathbb{R}$, $J(rx) = rJ(x)$.

(3) J is strictly monotone.

(4) J is coercive, i.e., there is a function $\rho: \mathbb{R}_+ \to \mathbb{R}$, with $\lim_{r\to\infty} \rho(r) = \infty$, such that for all $x \in X$, $(x,Jx) \geqq |x|\rho(|x|)$. Take as ρ the injection of $\mathbb{R}_+$ into $\mathbb{R}$.

<u>Proof.</u> See Pascali and Sburlan 1978, p 109, §2.6. QED

Let $S \in M(X,X^*)$ and $a > 0$. The Yosida approximant and the resolvent of S are, respectively,

$$S_a = (S \,\square\, a^{-1}J), \quad J_a^S = P(S,a^{-1}J).$$

Clearly S_a is monotone. Using theorem 2.3 and proposition 2.8(3) one obtains the usual expressions

$$S_a = (S^{-1} + aJ^{-1})^{-1}, J_a^S = (aJ^{-1} \circ S + I)^{-1},$$

and the splitting, $aJ^{-1} \circ S_a + J_a^S \in I$, where I is the identity map of X.

<u>Theorem 5.5.</u> Under the assumptions of theorem 5.4, let $S \in \bar{M}(X,X^*)$ and $a > 0$. Then,

(1) $S_a \in \bar{M}(X,X^*)$ is single-valued, continuous $(X,|\cdot|) \to (X^*,|\cdot|_*)$, with $\mathrm{dom}\, S_a = X$ and $\mathrm{ran}\, S_a = \mathrm{ran}\, S$.

(2) J_a^S = is single-valued, continuous $(X,|\cdot|) \to (X,|\cdot|)$, with $\mathrm{dom}\, J_a^S = X$ and $\mathrm{ran}\, J_a^S = \mathrm{dom}\, S$.

<u>Proof.</u> (1) J is maximal monotone with $\mathrm{ran}\, J = X^*$, thus theorem 4.2 implies that S_a is maximal monotone. As J^{-1} is strictly monotone, by proposition 4.3, S_a is single-valued. By proposition 2.6, $\mathrm{ran}\, S_a = \mathrm{ran}\, S$. Since $S^{-1} \in \bar{M}(X^*,X)$, $\mathrm{dom}\, S_a = X$ and $S_a = (S^{-1} + aJ^{-1})^{-1}$, the continuity of S_a follows (see Pascali and Sburlan 1978, p 122, prop 2.11, note that

the conclusion of lemma 2.11 (ibid.) is actually $x_j \to x$ in norm, see for example Browder 1983, p 20, prop 8).

(2) By proposition 2.9, dom $J_a^S = X$, ran J_a^S = dom S. By proposition 4.3, J_a^S is single-valued. As dom S_a = dom $J_a^S = X$, one actually has the splitting $I = aJ^{-1} \circ S_a + J_a^S$. Given this plus the continuities of S_a and J, the continuity of J_a^S follows. QED

Let $f \in \Gamma_o(X,X^*)$, then $\partial f \in \bar{M}(X,X^*)$ and by theorem 4.1(3c)

$$(\partial f)_a = \partial f \,\square\, a^{-1}\partial q = \partial(f \,\square\, a^{-1}q) = \partial f_a,$$

where $f_a = f \,\square\, a^{-1}q$. As ∂f_a is single-valued, f_a is Gâteaux differentiable (Moreau 1966, p 66, prop 10.g), as f_a is continuous, f_a is actually Fréchet differentiable (Pascali and Sburlan 1978, p 11) with differential df_a. By theorem 4.1(2b) one can conclude that $J_a^{\partial f} = p(f,a^{-1}q)$.

If X is a Hilbert space, then $J = I$, S_a is Lipschitz continuous with constant a^{-1}, and J_s^S is nonexpansive. Just set $T = a^{-1}I$ in theorems 4.9, 4.11 to find $(S \,\square\, a^{-1}I) \leqq a^{-1}$, $P(S,a^{-1}I) \leqq 1$ (see also Pascali and Sburlan 1978, p 131). Using theorem 2.3, proposition 2.8 one can easily get

$$S_a = S \,\square\, a^{-1}I = (S^{-1} + aI)^{-1} = (a^{-1}S^{-1} + I)^{-1} \circ (a^{-1}I) = J_{a^{-1}}^{S^{-1}} \circ (a^{-1}I),$$

$$J_a^S = P(S,a^{-1}I) = (aS + I)^{-1} = (S + a^{-1}I)^{-1} \circ (a^{-1}I) = (S^{-1})_{a^{-1}} \circ (a^{-1}I),$$

$$aS_a + J_a^S \in I,$$

whenever $S \in M(H)$ and $a > 0$. If $S \in \bar{M}(H)$, then there is = instead of $\in$ in the last equality above, and

$$I = aS_a + (S^{-1})_{a^{-1}} \circ (a^{-1}I) = S_a \circ (aI) + a^{-1}(S^{-1})_{a^{-1}}$$

$$= a^{-1}J_a^S \circ (aI) + J_{a^{-1}}^{S^{-1}} = J_a^S + aJ_{a^{-1}}^{S^{-1}} \circ (a^{-1}I).$$

When $S = \partial f$ with $f \in \Gamma_o(H)$, the above expressions become

$$df_a = p(f^*, aq) \circ (a^{-1}I), \quad df^*_{a^{-1}} \circ (a^{-1}I) = p(f, a^{-1}q),$$

$$I = a \cdot df_a + df^*_{a^{-1}} \circ (a^{-1}I) = df_a \circ (aI) + a^{-1} \cdot df^*_{a^{-1}}$$

$$= a^{-1}p(f, a^{-1}q) \circ (aI) + p(f^*, aq)$$

$$= p(f, a^{-1}q) + ap(f^*, aq) \circ (a^{-1}I).$$

Theorem 5.6. Let X be a reflexive real Banach space with normalized duality map J. For all $S \in M(X,X^*)$; $x,u \in X$; $y,v \in X^*$; $a,b \in (0,\infty)$, (1)-(5) are equivalent.

(1) $v = a^{-1}J(x - u)$, $\quad u = b^{-1}J^{-1}(y - v)$, $\quad v \in Su$,

(2) $v = S_a(x)$, $\quad u = (S^{-1})_b(y)$, $\quad x = b^{-1}J^{-1}(y - v) + aJ^{-1}v$,

(3) $v = S_a(x)$, $\quad u = (S^{-1})_b(y)$, $\quad y = a^{-1}J(x - u) + bJu$,

(4) $u = J_a^S(x)$, $\quad v = J_b^{S^{-1}}(y)$, $\quad x = b^{-1}J^{-1}(y - v) + aJ^{-1}v$,

(5) $u = J_a^S(x)$, $\quad v = J_b^{S^{-1}}(y)$, $\quad y = a^{-1}J(x - u) + bJu$.

Proof. The proof is essentially the same as that of theorem 2.12. However there is a small subtlety associated with the fact that if $ab \neq 1$, then $(a^{-1}J)^{-1} \neq bJ^{-1}$. Luckily this does not present any difficulty due to the homogeneity of degree 1 of J (resp. J^{-1}). There is no difficulty, using definitions 2.1, 2.7 and the homogeneity of J, in proving that (1) implies (2)-(5).

Assuming that (2) holds, one has for some $u' \in X$, $v' \in X^*$,

$$v = a^{-1}J(x - u') \in Su', \quad u = b^{-1}J^{-1}(y - v') \in S^{-1}v'.$$

From the first equality it follows that $x = u' + aJ^{-1}v$ which together with the third equality of (2) yields $u' = b^{-1}J^{-1}(y - v) \in S^{-1}v$. Hence

$$u' \in (S^{-1} \square b^{-1}J^{-1})y, \quad v \in P(S^{-1}, b^{-1}J^{-1})y.$$

Being J^{-1} strictly monotone, these two sets have precisely one element u and v' respectively, from which (1) follows easily.

Assuming that (4) holds, there are $u' \in X$, $v' \in X^*$, such that

$$v' = a^{-1}J(x - u) \in Su, \quad u' = b^{-1}J^{-1}(y - v) \in S^{-1}v.$$

From the second equality, it follows that $u' = b^{-1}J^{-1}(y - v)$, which together with the third equality of (4) yields $v = a^{-1}J(x - u') \in Su'$. Hence

$$v = (S \square a^{-1}J)x, \quad u' \in P(S, a^{-1}J)x.$$

Being J strictly monotone, these two sets have precisely one element, v' and u respectively. QED

<u>Corollary 5.7.</u> Let X, X^*, J be as in the above theorem. For all $f \in \Gamma_o(X, X^*)$; $x, u \in X$; $y, v \in X^*$; $a, b \in (0, \infty)$, (1)-(5) are equivalent.

(1) $v = a^{-1}J(x - u)$, $\quad u = b^{-1}J^{-1}(y - v)$, $\quad f(u) + f^*(v) = (u, v)$,

(2) $v = df_a(x)$, $\quad u = df^*_b(y)$, $\quad x = b^{-1}J^{-1}(y - v) + aJ^{-1}v$,

(3) $v = df_a(x)$, $u = df_b^*(y)$, $y = a^{-1}J(X - u) + bJu$,

(4) $u = p(f, a^{-1}q)x$, $v = p(f^*, b^{-1}q^*)y$, $x = b^{-1}J^{-1}(y - v) + aJ^{-1}v$,

(5) $u = p(f, a^{-1}q)x$, $v = p(f^*, b^{-1}q^*)y$, $y = a^{-1}J(X - u) + bJu$.

Corollary 5.8. Let H be a real Hilbert space with identity map I. For all $S \in M(H)$; $x,y,u,v \in H$; $a,b \in (0,\infty)$, (1)-(5) are equivalent.

(1) $x - u = av$, $y - v = bx$, $v \in Su$,

(2) $v = S_a(x)$, $u = (S^{-1})_b(y)$, $x = b^{-1}(y - v) + av$,

(3) $v = S_a(x)$, $u = (S^{-1})_b(y)$, $y = a^{-1}(x - u) + bu$,

(4) $u = J_a^S(x)$, $v = J_b^{S^{-1}}(y)$, $x = b^{-1}(y - v) + av$,

(5) $u = J_a^S(x)$, $v = J_b^{S^{-1}}(y)$, $y = a^{-1}(x - u) + bu$.

If $ab = 1$, then (1') and (2') are equivalent

(1') $x = u + av$, $v \in Su$,

(2') $u = J_a^S(x) = (S^{-1})_{a^{-1}}(x/a)$, $v = J_{a^{-1}}^{S^{-1}}(x/a) = S_a(x)$.

Corollary 5.9. Let H be a real Hilbert space with identity map I. For all $f \in \Gamma_o(H)$; $x,y,u,v \in H$; $a,b \in (0,\infty)$, (1)-(5) are equivalent.

(1) $x - u = av$, $y - v = bu$, $f(u) + f^*(v) = (u,v)$,

(2) $v = df_a(x)$, $u = df_b^*(y)$, $x = b^{-1}(y - v) + av$,

(3) $v = df_a(x)$, $u = df_b^*(y)$, $y = a^{-1}(x - u) + bu$

$$(4) \quad u = p(f, a^{-1}q)x, \quad v = p(f^*, b^{-1}q^*)y, \quad x = b^{-1}(y - v) + av,$$

$$(5) \quad u = p(f, a^{-1}q)x, \quad v = p(f^*, b^{-1}q^*)y, \quad y = a^{-1}(x - u) + bu.$$

If $ab = 1$, then (1') and (2') are equivalent

$$(1') \quad x = u + av, \quad f(u) + f^*(v) = (u,v),$$

$$(2') \quad u = p(f, a^{-1}q)x = df^*_{a^{-1}}(x/a), \quad v = p(f^*, aq)(x/a) = df_a(x).$$

REFERENCES

[1] E. Asplund (1969), Topics in the theory of convex functions. In: Theory and applications of monotone operators, A. Ghizzetti, ed., Proceedings of a NATO Advanced Study Institute, Venice, Italy, June 17-30, 1968, Edizioni Oderisi, Gubio, Italy.

[2] F.E. Browder (1983), Fixed point theory and nonlinear problems. Bull. (N.S.) Amer. Math. Soc., 9: 1, 1-39.

[3] V. Dolezal (1979), Feedback systems described by monotone operators. SIAM J. Control and Optimization, 17: 3, 339-364.

[4] I. Ekeland, and R. Temam (1976), Convex analysis and variational problems. North-Holland, Amsterdam.

[5] E.G. Gol'shtein (1975), Method of modification of monotone mappings. Economics and Math. Methods, 11: 5, 1144-1159 (in Russian).

[6] R.B. Holmes (1975), Geometric functional analysis and its applications. GTM #24, Springer-Verlag, New York.

[7] J.L. Kelley, I. Namioka, et al. (1963), Linear topological spaces. GTM #36, Springer-Verlag, New York.

[8] C. Lescarret (1967), Applications "prox" dans un espace de Banach. C.R. Acad. Sci. Paris, 265A, 676-678.

[9] J. Luque (1984), Nonlinear proximal point algorithms. Dissertation, Operations Research Center, Massachusetts Institute of Technology, May 1984.

[10] J. Luque (1986), Convolutions of maximal monotone mappings, abstract. Abstracts of papers presented to the Amer. Math. Soc., 7: 1, 108, no. *825-49-541.

[11] J.-J. Moreau (1962), Fonctions convexes duales et points proximax dans un espace hilbertien. C.R. Acad. Sci. Paris, 255, 2897-2899.

[12] J.-J. Moreau (1966), Fonctionelles convexes. Lecture notes, séminaire sur les équations aux derivées partielles, Collège de France, Paris.

[13] D. Pascali, and S. Sburlan (1978), Nonlinear mappings of monotone type. Sijthoff and Noordhoff, Aalphen aan den Rijn, The Netherlands.

[14] R.T. Rockafellar (1966a), Level sets and continuity of conjugate convex functions. Trans. Amer. Math. Soc., 123, 46-63.

[15] R.T. Rockafellar (1966b), Extension of Fenchel's duality theorem for convex functions. Duke Math. J., 33, 81-89.

[16] R.T. Rockafellar (1966c), Characterization of the subdifferentials of convex functions. Pacific J. Math., 17: 3, 497-510.

[17] R.T. Rockafellar (1967a), Convex programming and systems of elementary monotonic relations. J. Math. Analysis Applications, 19: 3, 543-564.

[18] R.T. Rockafellar (1967b), Monotone processes of convex and concave type. Memoirs Amer. Math. Soc., #77.

[19] R.T. Rockafellar (1970a), Convex analysis. Princeton University Press.

[20] R.T. Rockafellar (1970b), On the maximality of sums of nonlinear monotone operators. Trans. Amer. Math. Soc., 149, 75-88.

[21] R.T. Rockafellar (1974), Conjugate duality and optimization. Regional Conference Series in Applied Mathematics #16, SIAM.

[22] S.L. Troyanski (1971), On locally uniformly convex and differentiable norms in certain non-separable Banach spaces. Studia Mathema-

tica, 37, 173-180.

[23] D. Wexler (1972), Applications prox associées à un couple de fonctions duales de Legendre. C.R. Acad. sci. Paris, 275A, 1327-1330.