Journal of Global Optimization (2006) 35: 197–213
DOI 10.1007/s10898-005-3835-3

Variational Methods in Convex Analysis*

JONATHAN M. BORWEIN[1] and QIJI J. ZHU[2]
[1]*Faculty of Computer Science, Dalhousie University, Halifax, N. S. Canada, B3H 1W5 (e-mail: jborwein@cs.dal.ca)*
[2]*Department of Mathematics, Western Michigan University, Kalamazoo, MI 49008, USA (e-mail: zhu@wmich.edu)*

(Accepted 6 October 2005)

Abstract. We use variational methods to provide a concise development of a number of basic results in convex and functional analysis. This illuminates the parallels between convex analysis and smooth subdifferential theory.

Key words: convex analysis, fenchel duality, functional analysis, sandwich theorem, variational methods

1.

The purpose of this note is to give a concise and explicit account of the following folklore: several fundamental theorems in convex analysis such as the sandwich theorem and the Fenchel duality theorem may usefully be proven by variational arguments. Many important results in linear functional analysis can then be easily deduced as special cases. These are entirely parallel to the basic calculus of smooth subdifferential theory. Some of these relationships have already been discussed in [1, 2, 5, 6, 12, 18].

2.

By a ‘variational argument’ we connote a proof with two main components: (a) an argument that an appropriate auxiliary function attains its minimum and (b) a ‘decoupling’ mechanism in a sense we make precise below.

It is well known that this methodology lies behind many basic results of smooth subdifferential theory [6, 20]. It is known, but not always made explicit, that this is equally so in convex analysis. Here we record in an organized fashion that this method also lies behind most of the important theorems in convex analysis.

*Research was supported by NSERC and by the Canada Research Chair Program and National Science Foundation under grant DMS 0102496.

In convex analysis the role of (a) is usually played by the following theorem attributed to Fenchel and Rockafellar (among others) for which some preliminaries are needed.

Let X be a real locally convex topological vector space. Recall that the *domain* of an extended valued convex function f on X (denoted dom f) is the set of points with value less than $+\infty$. A subset T of X is *absorbing* if $X=\bigcup_{\lambda>0}\lambda T$ and a point s is in the *core* of a set $S\subset X$ (denoted by $s\in$ core S) provided that $S-s$ is absorbing and $s\in S$. A symmetric, convex, closed and absorbing subset of X is called a *barrel*. We say X is *barrelled* if every barrel of X is a neighborhood of zero. All Baire – and hence all complete metrizable – locally convex spaces are barrelled, but not conversely.

Recall that $x^*\in X^*$, the topological dual, is a *subgradient* of $f\colon X\to(-\infty,+\infty]$ at $x\in$ dom f provided that $f(y)-f(x)\geqslant\langle x^*, y-x\rangle$. The set of all subgradients of f at x is called the *subdifferential* of f at x and is denoted $\partial f(x)$, We use the standard convention that $\partial f(x)=\emptyset$ for $x\notin$ dom f. We use cont f to denote the set of all continuity points of f.

THEOREM 1 (Fenchel–Rockafellar). *Let X be a locally convex topological vector space and let $f\colon X\to(-\infty,+\infty]$ be a convex function. Then for every x in* cont f, $\partial f(x)\neq\emptyset$.

Combining this result with a decoupling argument we obtain the following lemma that can serve as a launching pad to develop many basic results in convex and in linear functional analysis.

LEMMA 2 (Decoupling). *Let X be a Banach space and let Y be a barrelled locally convex topological vector space. Let $f\colon X\to(-\infty,+\infty]$ and $g\colon Y\to(-\infty,+\infty]$ be lower semicontinuous convex functions and let $A\colon X\to Y$ be a closed linear map. Let*

$$p=\inf\{f(x)+g(Ax)\}.$$

Suppose that f, g and A satisfy the interiority condition

$$0\in\mathrm{core}(\mathrm{dom}\ g-A\,\mathrm{dom}\,f). \tag{1}$$

Then, there is a $\phi\in Y^*$ such that, for any $x\in X$ and $y\in Y$,

$$p\leqslant[f(x)-\langle\phi, Ax\rangle]+[g(y)+\langle\phi, y\rangle]. \tag{2}$$

Proof. Define an *optimal value function* $h\colon Y\to[-\infty,+\infty]$ by

$$h(u):=\inf_{x\in X}\{f(x)+g(Ax+u)\}.$$

It is easy to check that h is convex and that $\operatorname{dom} h = \operatorname{dom} g - A \operatorname{dom} f$. We may assume $f(0) = g(0) = 0$, and define

$$\mathcal{B} := \bigcup_{x \in B_X} \{u \in Y : f(x) + g(Ax + u) \leqslant 1\}.$$

Let $T := \mathcal{B} \cap (-\mathcal{B})$. We check that cl T is a barrel and hence a neighborhood of 0. Clearly cl T is closed, convex and symmetric. We need only show that it is absorbing. In fact we will establish the stronger result that T is absorbing. Let $y \in Y$ be an arbitrary element. Since $0 \in$ core (dom $g -$ A dom f) there exists $t > 0$ such that $\pm ty \in \operatorname{dom} g - A \operatorname{dom} f$. Choose elements $x_+, x_- \in \operatorname{dom} f$ such that $Ax_\pm \pm ty \in \operatorname{dom} g$. Then, there exists $k > 0$ such that

$$f(x_\pm) + g(Ax_\pm \pm ty) \leqslant k < \infty. \tag{3}$$

Choose $m \geqslant \max\{\|x_+\|, \|x_-\|, k, 1\}$. Dividing (3) by m and observing f and g are convex and $f(0) = g(0) = 0$ we have

$$f\left(\frac{x_\pm}{m}\right) + g\left(A\frac{x_\pm}{m} \pm \frac{ty}{m}\right) \leqslant 1.$$

Thus, $\pm ty/m \in \mathcal{B}$ or $ty/m \in T$ which implies that T is absorbing.

Next we show that T is cs-closed (see [11]). Let $y = \sum_{i=1}^{\infty} \lambda_i y_i$ where $\lambda_i \geqslant 0$, $\sum_{i=1}^{\infty} \lambda_i = 1$ and $y_i \in T$. Since $T \subset \mathcal{B}$, for each i there exists $x_i \in B_X$ such that

$$f(x_i) + g(Ax_i + y_i) \leqslant 1. \tag{4}$$

Clearly $\sum_{i=1}^{\infty} \lambda_i x_i$ converges, say to $x \in B_X$. Multiplying (4) by λ_i and sum over all $i = 1, 2, \ldots$ we have

$$\sum_{i=1}^{\infty} \lambda_i f(x_i) + \sum_{i=1}^{\infty} \lambda_i g(Ax_i + y_i) \leqslant 1.$$

Since f and g are convex and lower semicontinuous and A is continuous we have

$$f(x) + g(Ax + y) \leqslant 1$$

or $y \in \mathcal{B}$. A similar argument shows that $y \in -\mathcal{B}$. Thus, $y \in T$ and T is cs-closed.

It follows that $0 \in$ core $T = \operatorname{int} T = \operatorname{int} \operatorname{cl} T$ (see [11]).

Note that h is bounded above by 1 on T and, therefore, continuous in a neighborhood of 0. By the Fenchel–Rockafellar theorem there is some $-\phi \in \partial h(0)$. Then, for all u in Y and x in X,

$$\begin{aligned} h(0) = p &\leqslant h(u) + \langle \phi, u \rangle \\ &\leqslant f(x) + g(Ax+u) + \langle \phi, u \rangle. \end{aligned} \tag{5}$$

For arbitrary $y \in Y$, setting $u = y - Ax$ in (5), we arrive at (2). □

Remark 3. (a) In the above proof we only used the fact that B_X is a cs-compact absorbing set [11]. Thus, the decoupling theorem and many of the results in the sequel are also valid when assuming X is locally convex with a cs-compact absorbing set. In particular, the previous result obtains when X is a completely metrizable locally convex topological vector space. Actually, with more work assuming X is a complete metrizable space is enough. We will not pursue this technical generalization. However, we do want to emphasize the fact that Y is not necessarily complete.

(b) The constraint qualification condition (1) and the lower semicontinuity of f and g are assumed to ensure that h is continuous at 0. An alternative and often convenient constraint qualification condition to achieve the same goal is

$$\operatorname{cont} g \cap A \operatorname{dom} f \neq \emptyset. \tag{6}$$

In this case one can directly deduce that h is bounded above and, therefore, continuous (actually locally Lipschitz) in a neighborhood of 0 without the lower semicontinuity assumptions on f and g.

3.

We now use Lemma 2 to recapture several basic theorems in convex analysis.

THEOREM 4 (Sandwich). *Let X be a Banach space and let Y be a barrelled locally convex topological vector space. Let $f: X \to (-\infty, +\infty]$ and g: $Y \to (-\infty, +\infty]$ be lower semicontinuous convex functions and let $A: X \to Y$ be a closed densely defined linear map* (*meaning that the adjoint A^* is well defined*)*. Suppose that*

$$f \geqslant -g \circ A$$

and f and g satisfy condition (1). *Then, there is an affine function $\alpha : X \to R$ of the form*

$$\alpha(x) = \langle A^*\phi, x\rangle + r$$

for some ϕ in Y^*, satisfying

$$f \geqslant \alpha \geqslant -g \circ A.$$

Moreover, for any $\bar{x}$ satisfying $f(\bar{x}) = -g \circ A(\bar{x})$, one has $-\phi \in \partial g(A\bar{x})$.

Proof. By Lemma 2 there exists $\phi \in X^*$ such that, for any $x \in X$ and $y \in Y$,

$$0 \leqslant p \leqslant [f(x) - \langle \phi, Ax\rangle] + [g(y) + \langle \phi, y\rangle]. \tag{7}$$

For any $z \in X$ setting $y = Az$ in (7) we have

$$f(x) - \langle A^*\phi, x\rangle \geqslant -g(Az) - \langle A^*\phi, z\rangle.$$

Thus,

$$a := \inf_{x\in X}[f(x) - \langle A^*\phi, x\rangle] \geqslant b := \sup_{y\in Y}[-g(Ay) - \langle A^*\phi, y\rangle].$$

Picking any $r \in [a, b]$, and defining $\alpha(x) := \langle A^*\phi, x\rangle + r$ yields an affine function that separates f and $-g \circ A$.

Finally, when $f(\bar{x}) = -g \circ A(\bar{x})$, it follows from (7) that $-\phi \in \partial g(A\bar{x})$. □

THEOREM 5 (Fenchel duality). *Let X be a Banach space and let Y be a barrelled locally convex topological vector space. Let $f : X \to (-\infty, +\infty]$ and $g : Y \to (-\infty, +\infty]$ be lower semicontinuous convex functions and let $A : X \to Y$ be a closed densely defined linear map. Suppose that f and g satisfy the condition*

$$0 \in \operatorname{core}(\operatorname{dom} g - A \operatorname{dom} f)$$

Define

$$p = \inf_{x\in X}\{f(x) + g(Ax)\}, \tag{8}$$
$$d = \sup_{\phi\in Y^*}\{-f^*(A^*\phi) - g^*(-\phi)\}. \tag{9}$$

Then $p = d$, and the supremum in the dual problem (9) *is attained whenever finite.*

proof. It follows from Fenchel's inequality that

$$h(z)+h^*(\phi)\geqslant\langle\phi,z\rangle$$

for any function h, that $p\geqslant d$ always holds. This fact is usually referred to as *weak duality.*

If p is $-\infty$ there is nothing to prove, while if condition (1) holds and p is finite then by Lemma 2 there is a $\phi\in Y^*$ such that (2) holds. For any $u\in Y$, setting $y=Ax+u$ in (2) we have

$$\begin{aligned}p&\leqslant f(x)+g(Ax+u)+\langle\phi,u\rangle\\ &=\{f(x)-\langle A^*\phi,x\rangle\}+\{g(y)-\langle-\phi,y\rangle\}.\end{aligned}$$

Taking the infimum over all points y, and then over all points x, gives the inequalities

$$p\leqslant -f^*(A^*\phi)-g^*(-\phi)\leqslant d\leqslant p.$$

Thus ϕ attains the supremum in problem (9), and $p=d$. □

THEOREM 6 (Convex subdifferential sum and composition rule). *Let X be a Banach space and let Y be a barrelled locally convex topological vector space, let both $f:X\to(-\infty,+\infty]$ and $g:Y\to(-\infty,+\infty]$ be lower semicontinuous convex functions and let $A:X\to Y$ be a closed densely defined linear map. Then at any point x in X, the sum rule*

$$\partial(f+g\circ A)(x)\supset\partial f(x)+A^*\partial g(Ax) \tag{10}$$

holds, with equality if either condition (1) *or* (6) *holds.*

Proof. Inclusion (10) is easy. We prove the reverse inclusion under condition (1). Suppose $x^*\in\partial(f+g\circ A)(\bar{x})$. Since shifting by a constant does not change the subdifferential of a convex function, we may assume without loss of generality that

$$x\mapsto f(x)+g(Ax)-\langle x^*,x\rangle,$$

attains its minimum 0 at $x=\bar{x}$. By the sandwich theorem of Theorem 3 there exists an affine function $\alpha(x):=\langle A^*\phi,x\rangle+r$ with $-\phi\in\partial g(A\bar{x})$ such that

$$f(x)-\langle x^*,x\rangle\geqslant\alpha(x)\geqslant -g(Ax).$$

Since equality is attained at $x=\bar{x}$, we have $x^*+A^*\phi\in\partial f(\bar{x})$. Therefore,

$$x^*=(x^*+A^*\phi)+A^*(-\phi)\in\partial f(\bar{x})+A^*\partial g(A\bar{x}).$$ □

Recall that the *convex normal cone* to C at x is defined to be

$$N_C(\bar{x}) := \{\phi \in X^* : \langle \phi, c - \bar{x} \rangle \leqslant 0, \ \forall c \in C\}.$$

With this notation, suppose $g := i_C$ where C is a closed convex subset of X and i_C denotes the *convex indicator function of C,* which is zero on C and $+\infty$ otherwise, and A is the identity mapping on X. Then we derive:

THEOREM 7 (Pshenichnii–Rockafellar conditions [13]). *If the convex set C in a Banach space X satisfies the condition that* (i) *cont* $f \cap C \neq \emptyset$, *or the condition that* (ii) dom $f \cap \operatorname{int} C \neq \emptyset$, *and if f is bounded below on C, then there is an affine function $\alpha \leqslant f$ with*

$$\inf_C f = \inf_C \alpha.$$

In addition, the point $\bar{x}$ minimizes f on C if and only if it satisfies

$$0 \in \partial f(\bar{x}) + N_C(\bar{x}).$$

Combining Theorems 6 and 7 and Ekeland's variational principle [9] – applicable in the complete metrizable setting – we may next derive a convex version of the *multidirectional mean value theorem* [8, 12].

THEOREM 8 (Convex multidirectional mean value inequality). *Let X be an arbitrary Banach space, let C be a nonempty, closed and convex subset of X . Fix x in X and let $f : X \to \mathbb{R}$ be a continuous convex function, Suppose that f is bounded below on $[x, C]$ and*

$$\inf_{y \in C} f(y) - f(x) > r.$$

Then, for any $\varepsilon > 0$, there exist $z \in [x, C]$ and $z^* \in \partial f(z)z$, such that

$$f(z) < \inf_{[x,C]} f + |r| + \varepsilon$$

and

$$r < \langle z^*, y - x \rangle + \varepsilon \| y - x \|$$

for all y in C.

Proof. Using the auxiliary function $F(x, t) := f(x) - rt$ we can convert the general case to the special case when $r = 0$. So we will only prove this special case. Let $\tilde{f} := f + i_{[x,c]}$. Then $\tilde{f}$ is bounded below on X. By taking a smaller $\varepsilon > 0$ if necessary, We may assume that

$$\varepsilon < \inf_{y \in C} f(y) - f(x).$$

Applying Ekeland's variational principle [9] we conclude that there exists z such that

$$\tilde{f}(z) < \inf \tilde{f} + \varepsilon \tag{11}$$

and

$$\tilde{f}(z) \leqslant \tilde{f}(u) + \varepsilon \|u - z\|, \quad \forall u \in X. \tag{12}$$

That is to say

$$u \to f(u) + i_{[x,C]}(u) + \varepsilon \|u - z\|$$

attains a minimum at z. By (11) $\tilde{f}(z) < +\infty$ hence $z \in [x, C]$.

The sum rule for convex subdifferentials given in Theorem 6 (with A being the identity mapping) implies that there exists $z^* \in \partial f(z)$ such that $0 \leqslant \langle z^*, w - z\rangle + \varepsilon \|w - z\|$, $\forall w \in [x, C]$. Using a smaller ε to begin with if necessary we have, for $w \neq z$,

$$0 < \langle z^*, w - z\rangle + \varepsilon \|w - z\|, \quad \forall w \in [x, C] \setminus \{z\}. \tag{13}$$

Moreover by inequality (11) we have $f(z) = \tilde{f}(z) \leqslant f(x) + \varepsilon < \inf_c f$, so $z \notin C$. Thus we can write $z = x + \bar{t}(\bar{y} - x)$ where $\bar{t} \in [0, 1)$. For any $y \in C$ set $w = y + \bar{t}(\bar{y} - y) \neq z$ in (13) yields

$$0 < \langle z^*, y - x\rangle + \varepsilon \|y - x\|, \quad \forall y \in C. \tag{14}$$

□

Note that in the proof of this result besides using the subdifferential sum rule (which we have seen is a consequence of the decoupling lemma) we centrally used Ekeland's variational principle to locate the mean value point z.

The multidirectional mean value inequality can be used to prove a quite general *open mapping* theorem [12]. Recall that a multifunction $F{:}X \to 2^Y$ is a *closed convex multifunction* if the *graph of* $F(\{(x, y) : y \in F(x)\})$ is a closed convex set.

THEOREM 9 (Open mapping). *Let X and Y be Banach spaces. Let $F{:}X \to 2^Y$ be a closed convex multifunction. Suppose that*

$$y_0 \in \operatorname{core} F(X).$$

Then F is open at y_0; that is, for any $x_0 \in F^{-1}(y_0)$ and any $\eta > 0$,

$$y_0 \in \operatorname{int} F(x_0 + \eta B_X).$$

Proof. Let $T{:}X \times Y \to Y$ be a linear operator defined by $T(x, y) := y$ and let $G :=$Graph F. It is plain that we need only to show that $T|_A$ is open at (x_0, y_0). Since

$$0 \in \operatorname{core} T(G - (x_0, y_0)) = F(X) - y_0$$

and G is convex, a standard Baire category argument implies that there exists $\varepsilon > 0$ such that

$$\varepsilon B_Y \subset \operatorname{cl} T((G - (x_0, y_0)) \cap B_{X\times Y}). \tag{15}$$

We need to remove the closure above and so to show that

$$T(x_0, y_0) + (\varepsilon\eta/2)B_Y \subset T(((x_0, y_0) + \eta B_{X\times Y}) \cap G).$$

Let $z \in T(x_0, y_0) + (\varepsilon\eta/2)B_y$ and set $h(x, y) := \|T(x, y) - z\|$.

Applying the convex multidirectional mean value inequality of Theorem 8 to function h, set $Y := ((x_0, y_0) + \eta B_{X\times Y}) \cap G$ and point (x_0, y_0) yields that there exist $u \in ((x_0, y_0) + \eta B_{X\times Y}) \cap A$ and $u^* \in \partial h(u)$ such that

$$\inf_Y h - h(x_0, y_0) - \varepsilon\eta/4 \leqslant \langle u^*, (x, y) - (x_0, y_0)\rangle, \quad \forall x \in Y. \tag{16}$$

If $h(u) = 0$ then $T(u) = z$ and we are done. Otherwise $u^* = T^* y^*$ with $y^* \in \partial\|\cdot\|(T(u) - z)$ being a unit vector. Then we can rewrite (16) as

$$\begin{aligned} 0 &\leqslant \inf_Y h \leqslant h(x_0, y_0) + \varepsilon\eta/4 + \langle y^*, T((x, y) - (x_0, y_0))\rangle \\ &\leqslant \varepsilon\eta/2 + \varepsilon\eta/4 + \langle y^*, T((x, y) - (x_0, y_0))\rangle, \\ &\qquad \forall (x, y) \in ((x_0, y_0) + \eta B_{X\times Y}) \cap G. \end{aligned}$$

Observe that $\eta\varepsilon B_Y \subset \operatorname{cl} T((G - (x_0, y_0)) \cap \eta B_{X\times Y})$ the infimum of the right hand side of the above inequality is $-\varepsilon\eta/4$, a contradiction. □

As an easy corollary we have the following boundedness result for convex functions, which holds somewhat more generally in Baire or barrelled normed spaces.

THEOREM 10 (Boundedness of convex functions). *Let X be a Banach space and let $f: X \to \bar{R}$ be a lower semicontinuous convex function. Then f is continuous at every point in the core (equivalently interior) of its domain.*

In particular, f is everywhere continuous if and only if f is everywhere finite.

Proof. We need only prove the first assertion. Consider

$$F(x) := f(x) + [0, +\infty).$$

Then F and F^{-1} are closed convex multifunctions because graph $F :=$ epi f is a closed convex set. Let $x \in$ core (dom f) = core $F^{-1}(R)$. By the Open mapping Theorem 9, F^{-1} is open at x. Now, consider any open interval (a, b) that contains $f(x)$. The lower semicontinuity of f implies that $\{x: f(x) \leqslant a\}$ is closed. Thus, x is in the open set

$$f^{-1}((a,b)) = F^{-1}((a,b)) \backslash \{x : f(x) \leqslant a\}.$$

Therefore, f is continuous at x. □

4.

Much of linear functional analysis can be viewed as a special case of convex analysis. Below we recall how to derive the basic results of linear functional analysis from the results of the previous section.

THEOREM 11 (Hahn–Banach extension). *Let X be a Banach space. Suppose the function $f: X \to \mathbb{R}$, is lower semicontinuous, everywhere finite and sublinear, and suppose for some linear subspace L of X the function $h: L \to \mathbb{R}$ is linear and* dominated *by f, that is, $f \geqslant h$ on L. Then there is a linear function $\bar{h}: X \to \mathbb{R}$, dominated by f, which agrees with h on L.*

Proof. Let $X = Y$, let A be the identity mapping of X, let $g = -h + i_L$ and apply the sandwich result of Theorem 4. □

THEOREM 12 (Hahn–Banach separation). *Let X be a Banach space and let C_1 and C_2 be two convex subsets of X. Suppose that int $C_1 \neq \emptyset$ but that $C_2 \cap \operatorname{int} C_1 = \emptyset$. Then there exists an affine function α on X such that*

$$\sup_{c_1 \in C_1} \alpha(c_1) \geqslant \inf_{c_2 \in C_2} \alpha(c2).$$

Proof. Without loss of generality we may assume that $0 \in \operatorname{int} C_1$. Consider the *gauge* function of C_1 defined by

$$\gamma(x) := \inf\{r : x \in rC_1\}.$$

Then γ is convex and dom $\gamma = X$. Moreover, int $C_1 = \{x \in X : \gamma(x) < 1\}$ and, consequently $C_1 \subset \{x \in X : \gamma(x) \leqslant 1\}$. It follows that $0 \in$ int γ. Applying the Sandwich Theorem 4 with $f = i_{\mathrm{cl}C_2}$, A is the identity mapping of X and $g = \gamma - 1$ we have there exists an affine function α on X such that $f \geqslant \alpha \geqslant -g$. Now for any $c_1 \in C_1, \alpha(c_1) \geqslant 1 - \gamma(c_1) \geqslant 0$ and for any $c_2 \in C_2, \alpha(c_2) \leqslant i_{C_2}(c_2) = 0$. □

The following classical open mapping theorem for linear mappings is a direct corollary of Theorem 9 in which $F(x) = \{Ax\}$ and, as usual, a linear mapping A from X to Y is said to be *open* when it maps open sets in X to open sets in Y.

THEOREM 13 (Open mapping theorem for linear mappings). *Let X and Y be Banach spaces and let A be a closed linear mapping from X to Y such that $A(X) = Y$. Then A is an open mapping.*

Next, we recall how directly to deduce the *principle of uniform boundedness* of linear functional analysis from Theorem 10.

THEOREM 14 (Principle of uniform boundedness). *Let X and Y be Banach spaces. Let Γ be a set of bounded linear operators from X to Y such that for each $x \in X$,*

$$\sup\{\|Ax\| : A \in \Gamma\} < +\infty.$$

Then

$$\sup\{\|A\| : A \in \Gamma\} < +\infty.$$

Proof. Define

$$f(x) := \sup\{\|Ax\| : A \in \Gamma\}.$$

Then it is easy to verify that f is a lower semicontinuous convex function, as a supremum of convex continuous functions. Since, by assumption, $f(x) < +\infty$ for all $x \in X$, it follows by Theorem 10 that f is continuous. In particular, there exists a constant $\eta > 0$ such that $\sup\{f(x) : x \in \eta B_X\} < \infty$. Then

$$\begin{aligned}\sup\{\|A\| : A \in \Gamma\} &= \sup\{\|Ax\| : A \in \Gamma, x \in B_X\} \\ &= \frac{1}{\eta}\sup\{\|Ax\| : A \in \Gamma, x \in \eta B_X\} \\ &= \frac{1}{\eta}\sup\{f(x) : x \in \eta B_X\} < +\infty.\end{aligned}$$

□

5.

In this section, we use the Fitzpatrick function to give variational proofs of Rockafellar's results on the range of maximal monotone multifunctions and on maximality of the sum of two maximal monotone operators. Throughout this section, $(X, \|\cdot\|)$ is a reflexive Banach space with dual X^* and $T: X \to 2^{X^*}$ is maximal monotone. The *Fitzpatrick function* F_T [7, 10], associated with T, is the proper closed convex function defined on $X \times X^*$ by

$$\begin{aligned} F_T(x, x^*) &:= \sup_{y^* \in T_Y} [\langle y^*, x\rangle + \langle x^*, y\rangle - \langle y^*, y\rangle] \\ &= \langle x^*, x\rangle + \sup_{y^* \in T_Y} \langle x^* - y^*, y - x\rangle. \end{aligned}$$

Since T is maximal monotone

$$\sup_{y^* \in T_y} \langle x^* - y^*, y - x\rangle \geqslant 0$$

and the equality holds if and only if $x^* \in Tx$. It follows that

$$F_T(x, x^*) \geqslant \langle x^*, x\rangle \tag{17}$$

with equality holds if and only if $x^* \in Tx$. Thus, we capture much of the character of a maximal monotone operator from a convex associate function.

Using only the Fitzpatrick function and the decoupling Lemma we can prove the following fundamental result remarkably easily [19].

THEOREM 15 (Rockafellar). *Let X be a reflexive Banach space and let $T: X \to 2^{X^*}$ be a maximal monotone operator. Then $R(T + J) = X^*$. Here J is the duality map defined by $J(x) := \partial\|x\|^2/2$.*

Proof. The Cauchy inequality and (17) implies that, for all x, x^*,

$$F_T(x, x^*) + \frac{\|x\|^2 + \|x^*\|^2}{2} \geqslant 0. \tag{18}$$

Applying the Decoupling Lemma 2 to (18) we conclude that there exist points $w^* \in X^*$ and $w \in X$ such that

$$\begin{aligned} 0 \leqslant\, & F_T(x, x^*) - \langle w^*, x\rangle - \langle x^*, w\rangle \\ & + \frac{\|y\|^2 + \|y^*\|^2}{2} + \langle w^*, y\rangle + \langle y^*, w\rangle. \end{aligned} \tag{19}$$

Choose $y \in -Jw^*$ and $y^* \in -Jw$ in inequality (19) we have

$$F_T(x, x^*) - \langle w^*, x\rangle - \langle x^*, w\rangle \geqslant \frac{\|w\|^2 + \|w^*\|^2}{2}. \tag{20}$$

For any $x^* \in Tx$, adding $\langle w^*, w\rangle$ to both sides of (20) and noticing $F_T(x, x^*) = \langle x^*, x\rangle$ we have

$$\langle x^* - w^*, x - w\rangle \geqslant \frac{\|w\|^2 + \|w^*\|^2}{2} + \langle w^*, w\rangle \geqslant 0. \tag{21}$$

Since (21) holds for all $x^* \in Tx$ and T is maximal we must have $w^* \in Tw$. Now setting $x^* = w^*$ and $x = w$ in (21) yields

$$\frac{\|w\|^2 + \|w^*\|^2}{2} + \langle w^*, w\rangle = 0,$$

which implies $-w^* \in Jw$. Thus, $0 \in (T + J)w$. Since the argument applies equally well to all translations of T, we have $R(T + J) = X^*$ as required. □

Replacing the Cauchy inequality in the proof of Theorem 15 by the Fenchel–Young inequality one can derive an even stronger version of the surjectivity result that can deduce Rockafellar's theorem on the maximal monotonicity of the sum of two maximal monotone operators as an easy corollary. This was discovered very recently [3, 4].

PROPOSITION 16. *A monotone mapping T is maximal if and only if the mapping $T(\cdot + x) + J$ is surjective for all x in X. Moreover, when J and J^{-1} are both single valued, a monotone mapping T is maximal if and only if $T + J$ is surjective.*

Proof. The 'only if' is established in Theorem 15. We prove the 'if'. Assume (w, w^*) is monotonically related to the graph of T. By hypothesis, we may solve $w^* \in T(x + w) + J(x)$. Thus $w^* = t^* + j^*$ where $t^* \in T(x + w)$ and $j^* \in J(x)$ and

$$0 \leqslant \langle w^* - t^*, w - (w + x)\rangle = -\langle w^* - t^*, x\rangle = -\langle j^*, x\rangle = -\|x\|^2 \leqslant 0.$$

Hence $j^* = 0, x = 0$ and we are done. □

We now prove our central result:

THEOREM 17. *Let X be any reflexive space and let T be maximal and f closed and convex. Suppose that*

$$0 \in \operatorname{core}\{\operatorname{conv}\ \operatorname{dom}(T) - \operatorname{conv}\ \operatorname{dom} \partial(f)\}.$$

Then

(a) $\partial f + T + J$ *is surjective.*
(b) $\partial f + T$ *is maximal monotone.*

Proof (Sketch). (a) We introduce $f_J(x) := f(x) + \|x\|^2/2$. Using the the *Fenchel–Young inequality* $-f_J(x) + f_J^*(-x^*) + \langle x^*, x\rangle \geqslant 0$, $\forall x, x^*$ to replace the Cauchy inequality in the proof of Theorem 15 we obtain

$$F_T(x, x^*) + f_J(x) + f_J^*(-x^*) \geqslant 0.$$

Now, the (CQ)

$$0 \in \operatorname{core}\{\operatorname{conv}\ \operatorname{dom}(T) - \operatorname{conv}\ \operatorname{dom}(\partial f)\}$$

assures that the *decoupling lemma* applies since f_J^* is everywhere finite. The rest of the proof is similar to that of Theorem 15.

(b) follows from (a) and Proposition 16. □

COROLLARY 18. *Let X be a reflexive Banach space and suppose that*

$$0 \in \operatorname{core}\{\operatorname{conv}\operatorname{dom}\partial(f)\}.$$

Then ∂f is maximal monotone.

Proof. Let $T = 0$ in (b) of Theorem 17. □

COROLLARY 19 [17]. *The sum of two maximal monotone operators T and S on a reflexive Banach space is maximal monotone if* $0 \in \operatorname{core}\,[\operatorname{conv}\ \operatorname{dom}(T) - \operatorname{conv}\ \operatorname{dom}(S)]$.

Proof. Theorem 17 applies to the maximal monotone mapping $T(z) := (T_l(x), T_2(y))$ and the indicator function $f(z) = \iota_{\{x=y\}}$. Finally, check that the given transversality condition implies the needed (CQ). □

Note that we have the *Fitzpatrick inequality*

$$F_{T_1}(x, x^*) + F_{T_2}(x, -x^*) \geqslant 0, \quad \forall x \in X, x^* \in X^*, \tag{22}$$

valid for *any* maximal monotone T_1, T_2. Since

$$F_{\partial f}(x, x^*) \leqslant f(x) + f^*(x^*)$$

the Fitzpatrick inequality is sharper than the Fenchel–Young inequality

$$F_T(x, x^*) + f(x) + f^*(-x^*) \geqslant 0, \quad \forall x \in X, x^* \in X^*, \tag{23}$$

valid for *any* maximal monotone T and *any* convex function f.

Going one step further and using the Fitzpatrick inequality in the place of the Fenchel–Young inequality in the proof of Theorem 17 we may establish:

THEOREM 20. *Let T_1 and T_2 be maximal monotone operators on a reflexive space. Suppose that* $0 \in \operatorname{core}\{\operatorname{dom}(F_{T_1}) \operatorname{conv} \operatorname{dom}(F_{T_2(-\cdot)})\}$ *as happens if* $0 \in \operatorname{core}\{\operatorname{conv}\operatorname{graph}(T_1) - \operatorname{conv}\operatorname{graph}(T_2(-\cdot))\}$. *Then* $0 \in R(T_1 + T_2)$.

Proof. Follow through the steps of Theorem 17. □

The original proofs in [16] were very extended and quite sophisticated – they used tools such as Brouwer's fixed point theorem and Banach space renorming theory, As with our proof of local boundedness, ultimately the result is reduced to much more accessible geometric convex analysis. The short proofs given here are a reworking and further simplification of those of [19]. They well illustrates the techniques of variational analysis: a properly constructed auxiliary function – the Fitzpatrick function – the variational principle with decoupling in the form of a decoupling lemma, followed by an appropriate decoding of the information.

6.

In Hilbert space, there is a tight relationship between nonexpansive mappings and monotone operators as described in the next lemma.

LEMMA 21. *Let H be a Hilbert space. Suppose that P and T are two multifunctions from subsets of H to 2^H whose graphs are related by $(x, y) \in$ graph P if and only if $(v, w) \in$ graph T where $x = w + v$ and $y = w - v$. Then*

(i) *P is nonexpansive (and single-valued) if and only if T is monotone.*
(ii) $D(P) = R(T + I)$.

Proof. Consider $v_n \in T w_n, n = 1, 2$. Then $y_n \in P x_n$ where $x_n = w_n + v_n$ and $y_n = w_n - v_n$. Direct computation yields

$$\begin{aligned}\langle v_1 - v_2, w_1 - w_2 \rangle &= \left\langle \frac{x_1 - x_2 - (y_1 - y_2)}{2}, \frac{x_1 - x_2 + (y_1 - y_2)}{2} \right\rangle \\ &= \frac{\|x_1 - x_2\|^2 - \|y_1 - y_2\|^2}{4}.\end{aligned}$$

It is easy to see that P is nonexpansive if and only if T is monotone.

To prove (ii), note that if $x \in D(P)$ and $y = Px$ then $\frac{x+y}{2} \in (\frac{x-y}{2})$ by definition so that $x = \frac{x+y}{2} \in \frac{x-y}{2} \in R(T + 1)$. Conversely if $w \in R(T + I)$ then there exists v such that $w \in (T + I)v$ or $w - v \in Tv$ which implies that $w = (w - v) + v \in D(P)$. □

This very easily leads to the Kirszbraun–Valentine theorem on the existence of nonexpansive extensions to all of Hilbert space of nonexpansive mappings on subsets of Hilbert space.

THEOREM 22 (Kirszbraun–Valentine). *Let H be a Hilbert space and let D be a nonempty subset of H. Suppose that $P:D \to H$ is a nonexpansive mapping. Then there exists a nonexpansive mapping $\hat{P}:H \to H$ defined on all of H such that $\hat{P}|_D = P$.*

Proof. Associate P to a monotone multifunction T as in Lemma 21. Extend T to a maximal monotone multifunction $\hat{T}$. Define $\hat{P}$ from $\hat{T}$ using Lemma 21 again, Applying Rockafellar's Theorem 15 we have $D(\hat{P}) = R(\hat{T}+I) = H$. It is easy to check that $\hat{P}$ is indeed an extension of P. □

Alternatively [14], one may directly associate a convex Fitzpatrick function F_p with a nonexpansive mapping P, and thereby derive the Kirszbraun–Valentine theorem.

7.

We have seen in the proofs of all the results discussed herein that the decoupling Lemma is either explicitly or implicitly involved. Thus, a variational argument is indeed a common thread behind many of the fundamental results of convex and functional analysis. Such matters are also discussed in [7] where additional examples are to be found. In particular, similar proofs are given the local boundedness of maximal monotone operators throughout the core of their domains, and of the surjectivity of coercive maximal monotone operators in reflexive space.

This is by no means a claim of the intrinsic superiority of the treatment herein, which in part follows [5]. For example, recently S. Simons showed an elegant and different way of explaining a similar basket of results starting from a generalized Hahn-Banach extension theorem [18]. Indeed [18] was in part what stimulated us to record the present perspective.

Acknowledgments

We thank J.-P. Penot and C. Zalinescu for pointing out inaccuracies in an earlier version of the paper.

References

1. Borwein, J.M. (1981), Convex relations in analysis and optimization. In: Schaible, S. and Ziemba, W.T. (eds.), *Generalized Concavity in Optimization and Economics*, Academic Press, New York.
2. Borwein, J.M. (1986), Stability and regular points of inequality systems, *Jornal of Optimization Theory and Applications* 48, 9–52.
3. Borwein, J.M. (2006), Maximal monotonicity via convex analysis, to appear in PAMS.
4. Borwein, J.M. (2006), Maximal monotonicity via convex analysis, *J. Convex Analysis*, in press.

5. Borwein, J.M. and Lewis, A.S. (2000), *Convex Analysis and Nonlinear Optimization: Theory and Examples*, CMS Springer-Verlag Books, Springer-Verlag, New York.
6. Borwein, J.M. and Zhu, Q.J. (1999), A survey of subdifferential calculus with applications, *Nonlinear Analysis, TMA* 38, 687–773.
7. Borwein, J.M. and Zhu, Q.J. (2005), *Techniques of Variational Analysis*, CMS Springer-Verlag Books, Springer-Verlag, New York.
8. Clarke, F.H and Ledyaev, Yu. S. (1994), Mean value inequalities, *Proc. Amer, Math. Soc.*, 122, 1075–1083.
9. Ekeland, I. (1974), On the variational principle, *J. Math. Anal. Appl.* 47, 324–353.
10. Fitzpatrick, S. (1988), Representing monotone operators by convex functions. *Workshop/Miniconference on Functional Analysis and Optimization* (Canberra, 1988), Proc. Centre Math. Anal. Austral. Nat. Univ., 20, Austral. Nat. Univ., Canberra, pp. 59–65.
11. Jameson, G.J.O. (1974), *Topology and Normed Spaces*, Chapman and Hall, New York.
12. Ledyaev, Yu. S. and Zhu, Q.J. (1999), Implicit multifunction theorems, *Set-Valued Analysis* 7, 209–238.
13. Pshenichnii, B. (1971), *Necessary Conditions for an Extremum*, Marcel Dekker, New York.
14. Reich, S. and Simons, S. (2004), Fenchel duality, Fitzpatrick functions and the Kirszbraun-Valentine extension theorem, preprint.
15. Rockafellar, R.T. (1970), *Convex Analysis*, Princeton University Press, Princeton, NJ.
16. Rockafellar, R.T. (1970), On the maximality of sums of nonlinear monotone operators, *Trans. Amer. Math, Soc.* 149, 75–88.
17. Simons, S. (1998), *Minimax and Monotonicity*, Lecture Notes in Mathematics, 1693, Springer-Verlag, Berlin.
18. Simons, S., (2003), "A new version of the Hahn-Banach theorem," www.math.ucsb.edu/simons/preprints/.
19. Simons, S. and Zalinescu, C.A. (2004), New Proof for Rockafellar's Characterization of Maximal Monotonicity. *Proc Amer. Math. Soc.*, 132, 2969–2972.
20. Zhu, Q.J. (1998), The equivalence of several basic theorems for subdifferentials, *Set- Valued Analysis*, 6, 171–185.