Optim Lett
DOI 10.1007/s11590-010-0178-x

REVIEW ARTICLE

Fifty years of maximal monotonicity

Jonathan M. Borwein

Received: 15 November 2009 / Accepted: 27 January 2010

Abstract Maximal monotone operator theory is about to turn (or just has turned) 50. I intend to briefly survey the history of the subject. I shall try to explain why maximal monotone operators are both interesting and important—culminating with a description of the remarkable progress made during the past decade.

Keywords Maximal monotonicity · Convex analysis · Fitzpatrick function · Minty surjectivity theorem · Optimization

1 Introduction and preliminaries

In Sect. 2 I shall retrace the high points—as I see them—of the study of maximal monotonicity in subsections for each of five decades: from 1959 to 2009. In Sect. 3 I shall actually prove a theorem before concluding in Sect. 4 with a discussion of what remain the main open questions for the theory.

1.1 Preliminaries

A *monotone operator* from a Hausdorff locally convex space X—for us almost always a normed space—to its topological dual X^* is identified with a subset T of $X^* \times X$ (a relation) such that

$$\langle x^* - y^*, x - y\rangle \geq 0 \tag{1}$$

J. M. Borwein's research supported by the Australian Research Council.

J. M. Borwein (✉)
Centre for Computer Assisted Research Mathematics and its Applications (CARMA),
University of Newcastle, Newcastle, Australia
e-mail: jonathan.borwein@newcastle.edu.au

Published online: 14 February 2010

for all (x^*, x) and (y^*, y) from T, where $\langle \cdot, \cdot \rangle$ denotes the duality pairing between X and X^*. Here we use $(x^*, x) \in T$ to identify a relation with the graph of a multifunction between X and X^* and interchangeably write $x^* \in T(x)$ or $x \in T^{-1}(x^*)$ in terms of the *inverse* relation. The *domain*, $D(T) := \{x \in X : \exists x^* \in X^* \text{ s.t. } (x^*, x) \in T\}$, and the *range*, $R(T) := \{x^* \in X^* : \exists x \in X \text{ s.t. } (x^*, x) \in T\}$, are both projections of T. We emphasize that when we write $T + S$ we intend the pointwise sum not the Minkowski sum.

A monotone operator T is *maximal* if T is maximal with respect to set inclusion amongst all monotone relations. This is to say that if $(y^*, y) \in X^* \times X$ is such that

$$\langle x^* - y^*, x - y \rangle \geq 0 \tag{2}$$

for all (x^*, x) in T, then $(y^*, y) \in T$ or alternatively $y^* \in T(y)$. Thus, maximality is a very useful surrogate for topological closure.

From now on we shall assume X is a Banach space since this is the case of most interest. We shall use H for a Hilbert space. When we have not specified matters that we follow the notation of [12, Ch. 9] in which full details of all assertions can be followed up (see also [8]) in which a larger reference list for more recent work can also be consulted. There are three core motivating examples.

1.1.1 Convex subgradients

Suppose that f is a closed convex function on a Banach space then the *subgradient* of f at x in the *domain* of f ($\{x : f(x) < +\infty\}$) is defined by

$$\partial f(x) := \{x^* \in X^* : \langle x^*, y - x \rangle \leq f(y) - f(x), \forall y \in X\}. \tag{3}$$

If we select $y^* \in \partial f(y), x^* \in \partial f(x)$ and use (3) twice we discover that ∂f is a monotone operator (for a potentially nonconvex proper function f) as studied in the smooth case by Minty in [36]. The modern subgradient was developed independently by Rockafellar [3] and by Moreau. A most important case is that of the *duality mapping*

$$J_X(x) := \frac{1}{2}\partial \|x\|^2 = \{x^* \in X^* : \|x\|^2 = \|x^*\|^2 = \langle x, x^* \rangle\}.$$

In particular $J_H = I$. The duality mapping's properties are central to the geometry of the corresponding Banach space.

1.1.2 Skew linear operators

Recall that a possibly nonsymmetric linear operator $T : X \to X^*$ is *positive semidefinite* (psd) if $\langle Tx, x \rangle \geq 0$ for all $x \in X$ and is *skew* if $\langle Tx, x \rangle = 0$ for all $x \in X$. Clearly any psd operator is monotone and, hence, so are all skew mappings. Let us say a single-valued mapping T is *hemicontinuous* on lines if $\langle \lim_{t \to 0^+} T(x + th) - T(x), x \rangle = 0$

for all $h \in X$. The next simple proposition is revealing and important in that it is highly representative of how monotonicity gets invoked.

Proposition 1 *A hemicontinuous monotone operator with full domain is maximal.*

Proof Using (2), we suppose $\langle y^* - T(x), y - x\rangle \geq 0$ for all $x \in X$. We specialize to deduce $\langle y^* - T(y + th), y - (y + th)\rangle \geq 0$ for all $h \in X$ and $t > 0$. Hence $\langle T(y + th) - y^*, h\rangle) \geq 0$ for all $h \in X$ and $t > 0$ and an appeal to hemicontinuity on lines shows that $\langle T(y) - y^*, h\rangle \geq 0$. A separation argument shows $(y^*, y) \in T$ as required. □

It is true but harder to prove that ∂f is maximal for any proper closed convex function f on a Banach space [42] (this fails in incomplete settings [12]). The continuous case is significantly easier. The proof of the maximality of the subgradient has a long history of simplification over the decades. A recent very simple proof using only basic convex analysis due to Alves and Svaiter is given in [12, Ch. 9].

One might think that skew operators are somewhat exotic for the theory but that is far from so. If one considers a linear program $\min\langle c, x\rangle$ s.t. $Ax \leq b, x \geq 0$ with respect to coordinate orderings then the associated *variational equality* [9, Sect. 7.3] relies on the skew matrix

$$T := \begin{bmatrix} 0 & A \\ -A^* & 0 \end{bmatrix}.$$

Thus, the entire duality theory of linear programming is concerned with the skew case. There is a whole taxonomy of n-monotone operators. The definition above is the 2-monotone case while subgradients are n-monotone for all $n \in \mathbf{N}$ and are called *cyclically monotone* [5,12]. For instance the noncyclic rotation

$$S_\theta := \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}$$

is precisely n-monotone for $\theta := \pi/n$. This result, originally proven by Asplund [2] in a paper with a misleading title, now has a reasonably direct though still quite technical Fitzpatrick-function proof [12].

1.1.3 The Laplacian

Much early impetus for the study of maximal monotone operators came out of partial differential equations [14] and takes place within the confines of Sobolev space—and so we give only an illustration of what is possible.

As an application of their study of existence of eigenvectors of second-order nonlinear elliptic equations in $L_2(\Omega)$, the authors of [32] assume that $\Omega \subset \mathbf{R}^n$, $(n > 1)$ is a bounded open set with boundary belonging to $C^{2,\alpha}$ for some $\alpha > 0$. They assume that one has functions $|a_i(x, u)| \leq \nu\,(1 \leq i \leq n)$ and $|a_0(x, u)| \leq \nu|u| + a(x)$ for some $a \in L_2(\Omega)$ and $\nu > 0$; where all a_i are measurable in x and continuous in u (a.e. x). They then consider the normalized eigenvalue problem

$$\begin{aligned} \triangle u + \lambda \left\{ \sum_{i=1}^{n} a_i(x,u) \frac{\partial u}{\partial x_i} + a_0(x,u) \right\} = 0, \quad x \in \Omega, \\ u(x) = 0 \qquad x \in \text{ bdry}\Omega \\ \|u\|_2 = 1 \end{aligned} \tag{4}$$

where $\triangle u = -\nabla^2 u = -\sum_{i=1}^{n} \frac{\partial^2 u}{\partial x_i^2}$ is the classical *Laplacian*.

To make this accessible to Sobolev theory, a weak solution is requested to (4) for $0 < \lambda \leq 1$ when $u \in W^{2,2}(\Omega) \cap W_0^{1,2}(\Omega)$. In this setting, a solution of

$$\triangle u + \tau u = f(x)$$

for all $\tau > 0$ and all $f \in L_2$ (and with $\|u\|_2 = 1$) is assured. Minty's surjectivity theorem 3 implies $T := \triangle$ is linear and maximal monotone on $L_2(\Omega)$ with domain $W^{2,2}(\Omega) \cap W_0^{1,2}(\Omega)$. Of course, one must first check *monotonicity* of $\triangle$ using integration by parts in the form

$$\int_\Omega \langle v, \triangle u \rangle = \int_\Omega \langle \nabla v, \nabla u \rangle,$$

for all $v \in W^{-1,2}(\Omega)$, $u \in C_0^\infty(\Omega) \subset W_0^{1,2}(\Omega)$. One is now able to provide a Fredholm alternative type result for (4) [32, Theorem 10]. In like-fashion one can make sense of the assertion that for $2 \leq p < \infty$ the *p-Laplacian* $\triangle_p$ is maximal monotone: $\triangle_p u$ is given by

$$\triangle_p u := -\text{div}(|\nabla u|^{p-2} \nabla u) \in W^{-1,q}(\Omega)$$

for $u \in W^{1,p}(\Omega)$ with $1/p + 1/q = 1$.

2 Maximal monotonicity: a five decade history

Monotone operators (or multifunctions) were first introduced by George Minty to aid in the abstract study of electrical networks [34], then in the setting of partial differential equations by Felix Browder and his school [16]. I personally learned a great deal from Zarantonello's early work, see [52]. Maximal monotone operators rapidly found uses for subgradients, optimization, variational inequalities, algorithms, mathematical economics, and much more. A useful survey of the state of play in 1968 was given by Minty [37] himself. By 1975, the main ideas were clear—if not easy—in Hilbert space and more generally in reflexive Banach space. Brézis's [14] monograph studied contraction semigroups of operators, related them to monotone operators and reprised what was by then a most satisfactory theory in Hilbert space—much of it original to Brézis. Outside of Hilbert space a similar theory was developed for accretive operators.

2.1 1959–1969

2.1.1 Minty

The inattentive reader might be forgiven for not seeing the modern subject clearly in Minty's seminal paper [34][1] but over the next few years researchers such as Minty, Browder (along notably with Brézis, Hess [17] and others), Asplund [2], Rockafellar, Zarantonello [51,53], laid down the foundations of the modern theory. The first really significant theorem was given and exploited to solve equations by Minty [35]:

Theorem 1 (Minty surjectivity theorem, 1962) *If T is continuous and monotone on a Hilbert space H then*

$$R(T + I) = H.$$

The application of monotonicity to elliptic PDE's as illustrated in Sect. 1.1.3 was largely by way of finite dimensional (*Galerkin*) *approximates*. In finite dimensions, the underlying surjectivity result could be proved by degree theory or other topological means and under milder hypotheses. Monotonicity became necessary when passing to the limit.

It is an easy consequence of maximality that any maximal monotone operator is *demiclosed* in that sequentially

$$x_n^* \in T(x_n), x_n^* \to_s x^*, x_n \to_w x \Rightarrow x^* \in T(x), \tag{5}$$

and this property is just what is needed for many approximation methods. A nice description of this approach can be found in [20]. A much more comprehensive accounting as of 20 years ago can be found in [54]. Banach spaces which have the property that every nonexpansive not necessarily self-map on every closed bounded convex set is demiclosed are said to have the *Browder-Göhde property*. We have shown that Hilbert space has the Browder-Goöhde property; interestingly, the Hilbert renorm on $\ell^2(N) \times R$ given by $\|(x, r)\| := \max(\|x\|_2, |r|)$ does not [29].

[1] In *Math. Reviews* it merited a three paragraph review from Frank Harary the graph theorist who gave two paragraphs of Minty's abstract and concluded.

> "The single new theorem of graph theory is as follows. Let N be a network (directed graph) whose branches (directed lines) are partitioned into three sets A, B, and C, and let one branch b of the set B be distinguished. Then N contains either a cycle containing b but no line of C in which all lines of B are similarly directed, or a cocycle containing b but no line of A in which all lines of B are similarly directed. This theorem, in addition to its intrinsic interest in graph theory, leads to important applications in both linear and nonlinear programming and the steady-state solution of nonlinear electric networks (the multitude of definitions required for presenting these consequences does not permit their inclusion here)."

In the same way, writing

$$0 \in T(x) + \partial\delta_C(x) \Leftrightarrow \exists x^* \in T(x), x \in C, \sup_{c \in C}\langle x^*, c - x\rangle \leq 0. \tag{6}$$

reduced a general *monotone variational inequality* on the right of (6) to a monotone set *inclusion* on the left. Here δ_C is the convex *indicator* function which is 0 on C and $+\infty$ otherwise.

To have any hope of using general theory it was necessary to know that $T_C(x) := T(x) + \partial\delta_C(x)$ was still maximal under reasonable conditions on T and C. Thus, it was natural to hope that subgradients were maximal and also to look for a 'sum'-theorem along the lines of Theorem 2 of the next subsection. Finally, we emphasize that even when T is single-valued and smooth T_C is neither.

2.2 1969–1979

2.2.1 Rockafellar

While maximal monotonicity now plays a large role in optimization and convex analysis there is only passing reference of monotonicity in Terry Rockafellar's now classic 1970 book *Convex Analysis* [43]. That said at the same time Rockafellar was making huge strides with the theory of monotone operators in reflexive space. In [41] he introduced the notion of *cyclic monotonicity*. Consider the convex potential

$$f_T(x) := \sup\left\{\langle x_n^*, x - x_n\rangle + \sum_{k=1}^{n-1}\langle x_{k-1}^*, x_k - x_{k-1}\rangle : x_k^* \in T(x_k), n \in \mathbf{N}\right\}, \tag{7}$$

where the sup is over all such chains. We now call f_T the *Rockafellar function*. Using it Rockafellar was able to show that maximal cyclically monotone operators were maximal monotone and possessed a subgradient.[2]

In [42] he proved the following theorem:

Theorem 2 (Rockafellar sum theorem) *Suppose that S and T are maximal monotone operators on a reflexive Banach space. Suppose that*

$$D(S) \cap \operatorname{int} D(T) \neq \emptyset. \tag{8}$$

Then $S + T$ is maximal monotone.

[2] The *Math Review* is again an author abstract:

> "The subdifferential of a lower semi-continuous proper convex function on a Banach space is a maximal monotone operator, as well as a maximal cyclically monotone operator. This result was announced by the author in a previous paper [same J. 17 (1966), 497–510; MR0193549 (33 #1769)], but the argument given there was incomplete; the result is proved here by a different method, which is simpler in the case of reflexive Banach spaces. At the same time, a new fact is established about the relation between the subdifferential of a convex function and the subdifferential of its conjugate in the nonreflexive case."

Rockafellar's very delicate proof was based on his improvement of Theorem 1:

Theorem 3 (Minty surjectivity theorem) *Suppose T is maximal monotone on a reflexive Banach space X (and both J_X and J_X^{-1} are single valued). Then*

$$R(T + J_X) = X^*.$$

The parenthetic condition was needed for technical reasons[3] (and was guaranteed by renorming X). In the next section we shall give a proof of Theorem 3 in full generality. The proof of Theorem 2 also required proving that *a monotone operator is locally bounded throughout the interior of its domain*. This also now has a simple convexity proof but was quite difficult to establish at the time.

A final fruit from Rockafellar's extraordinary year of 1970 was the proof that in a reflexive space the domain of a maximal monotone operator was *virtually convex* also called *nearly convex*: $\overline{D(T)}$ *is convex* [44]. This result applied to T^{-1} shows that in a reflexive setting $\overline{R(T)}$ is also convex. Outside of reflexive space this duality breaks down completely. In general $\operatorname{int} D(T)$ is always convex but in every nonreflexive space there is a maximal monotone operator with $\operatorname{int} R(T)$ nonconvex [8], [12, Ch 9.]. This is a consequence of R.C. James characterization of reflexive space.

Another signal achievement of the decade was Rockafellar's [45] remarkable paper on the *proximal point algorithm* first introduced by Martinet. It would take a complete paper to describe it and its impact. We settle for the following

Example 1 (Surjectivity) What reflexivity provides you is that *every coercive maximal monotone operator is surjective.* (We require T to be coercive in the sense that $\inf_{x* \in T(x)} \|x^*\| \to +\infty$ as $x \to \infty$. Indeed reflexivity characterizes Minty surjectivity theorem 3 (again via James' theorem). Moreover Theorem 3 ensures that in reflexive space the *resolvent* $R_\lambda := (T + \lambda J_X)^{-1}$, and *Yosida approximant*

$$T_\lambda := \left(T^{-1} + \lambda J_X^{-1}\right)^{-1}. \tag{9}$$

are everywhere defined [13,20] and well-behaved. It was these two objects that let Rockafellar prove both the maximality of the sum and the virtual convexity of the domain; neither is accessible in a nonreflexive setting. □

Example 2 (Nonexpansivity) Recall that P is *nonexpansive* if $\|x^* - y^*\| \le \|x - y\|$ for all $(x^*, x), (y^*, y) \in P$. Also P is necessarily single valued on its domain. What Hilbert space buys you is *nonexpansivity* since $J_H = I$. [We note that J_X and $(J_X)^{-1}$ are simultaneously smooth only on renormings of H.] Moreover, R_λ is nonexpansive. This is a consequence of the following remarkable observation which is easily established once discovered.

[3] It is needed for the equivalence of *hypermaximality*, that is $R(T + J_X) = X^*$, with maximality [46, Remark 10.8] but this can be avoided, see [12, Prop. 9.3.1].

Proposition 2 *Suppose H is a Hilbert space and T and P are related by*

$$(x, y) \in P \Leftrightarrow \left(\frac{x-y}{2}, \frac{x+y}{2}\right) \in T$$

then T is monotone iff P is nonexpansive (and necessarily singleton).

This is key to the analysis of the proximal point algorithm which is thus equivalent to *Krasnoselski's algorithm* [11] in H. We illustrate the power of Proposition 2 by showing how easily it implies the following fundamental result.

Theorem 4 (Valentine-Kirszbraun theorem (1945–1932)) *Every nonexpansive mapping P on $A \subset H$ extends to a nonexpansive mapping $\widehat{P}$ with $D(\widehat{P}) = H$.*

Proof Associate a monotone T to P, as in Proposition 2, and let $\widehat{T}$ be any maximal extension of T as is assured by Zorn's lemma (actually recent work [7, Sect. 5] lets this be done more constructively). Then let $\widehat{P}$ denote the associated nonexpansive mapping, using Proposition 2 again. We observe that $D(\widehat{P}) = R(\widehat{T}) = H$, on applying Minty's surjectivity theorem 1. Finally we check that $\widehat{P}|_A = P$. □

This allows us to show in Hilbert space that R_λ is nonexpansive and in consequence T_λ is λ-Lipschitz [13,20]. We note also that in Hilbert space, it is easy to see that if P is nonexpansive with full domain then $I - P$ is maximal monotone and so, as observed above, demiclosed. This lets us deduce very quickly the Hilbert space case of a famous theorem: *Suppose $P: C \hookrightarrow C$ is a nonexpansive self-map of a closed bounded convex set $C \subset H$ then P has fixed point in C.* To see this we argue as follows. Fix $c_0 \in C$ and let $P_n(x) := (1 - 1/n)P(x) + 1/nc_0$ for each $0 < n \in \mathbf{N}$. As P_n is a contraction on C we can solve $P_n(x_n) = x_n \in C$ and since C is weakly compact there is a subsequence $x_{n_k} \rightharpoonup_w \overline{c} \in C$. Then $x_{n_k} - P(x_{n_k}) \to 0$ in norm (since C is bounded) and so by demiclosure $P(\overline{c}) = \overline{c}$.

It is unknown if this result holds in an arbitrary reflexive space—it is true in a uniformly convex space and much more is known. □

2.2.2 Brézis and Haraux

In [15] the authors considered multifunctions $T: H \to H$. satisfying the condition $(*)$ that $\sup_{(z,h)\in(T}(h - f, y - z) < \infty$ for any $f \in R(T)$ and $y \in D(T)$. Angle-bounded or coercive monotone operators satisfy $(*)$. If T is maximal monotone, then $T + \lambda I$, $(I + \lambda T)^{-1}$ and $(T^{-1} + \lambda I)^{-1}$ satisfy $(*)$ for each $\lambda > 0$. The remarkable consequences are that

$$\overline{R(S+T)} = \overline{R(S) + R(T)}$$
$$\text{int}[R(S+T)] = \text{int}[R(S) + R(T)],$$

whenever S and T satisfy one of the following conditions: (a) S and T both satisfy $(*)$; (b) T satisfies $(*)$ and $D(S) \subseteq D(T)$; (b) $S = \partial\varphi$ and $\varphi((I + \lambda T)^{-1}\cdot) \leq \varphi(\cdot)$.

2.2.3 Mignot

Motivated by the second-order analysis of convex functions, it makes sense to say a monotone operator T on a Euclidean space E is (Fréchet) *differentiable* at $x \in D(T)$ if $T(x)$ is singleton (this can be weakened but one will conclude $T(x)$ is singleton) and there is a matrix $A = \nabla T(x)$ such that

$$T(x') \subset T(x) + \nabla T(x)(x' - x) + o(\|x' - x\|)B_E. \tag{10}$$

Another great accomplishment was made by Mignot [33] who showed in finite dimensions that *a monotone operator is a.e. differentiable throughout the interior of its domain*. This now provides the canonical proof of Alexandrov's theorem on the twice-differentiability of convex functions in Euclidean space, and is central to the study of viscosity solutions for PDEs. It is vaguely possible that Mignot's result extends to separable Hilbert space. It works no more generally as the next example shows.

Example 3 The following example from [10] provides a continuous convex function d on any nonseparable Hilbert space which is nowhere second-order differentiable: *Let A be uncountable and let C the positive cone of $\ell_2(A)$. Denote by d the distance function to C and let $P := \nabla d$.*

Proof (a) Clearly, $P(a) = a^+$ for all $a \in \ell_2(A)$, where $a^+ = (a_\alpha^+)_{\alpha\in A}$ and $a_\alpha^+ = \max\{0, a_\alpha\}$. Pick $x \in \ell_2(A)$ and $\alpha \in A$, then P is differentiable in the direction e_α if and only if $x_\alpha \neq 0$. Here e_α stands for an element of the canonical basis. Since each $x \in \ell_2(A)$ has only countably many nonzero coordinates, d is nowhere second-order differentiable. Likewise the maximal monotone operator P is nowhere differentiable in the sense of Mignot.

(b) On the other hand the distance function to the positive cone in $\ell_2(N)$ is necessarily second-order differentiable on a dense set (again at points with all nonzero coordinates). Now suppose X is a separable subspace of $\ell_2(A)$. Let $I \subset A$ be the countable set of nonzero coordinates of elements from X. (Note that X is a subspace of $\ell_2(I)$ but $X \neq \ell_2(I)$ in general.) Then $d \restriction_{\ell_2(I)}$ coincides with the distance function to the positive cone of $\ell_2(I)$, hence is second-order differentiable on a dense set (at points with all nonzero coordinates). Since the elements of X with nonzero coordinates are dense in X, we have that $d \restriction_X$ is second-order differentiable on a dense subset of X. In summary, we see that d restricted to any separable subspace is second-order differentiable but d itself is not. So separable reduction techniques cannot apply. □

Similar counterexamples are given in [10,12] to show that Mignot's result fails in all separable $\ell^p(N)$ for $p \neq 2$. □

2.2.4 Zarantonello and Kenderov

Zarantonello [53] seems to have been the first to realize that differentiability results for convex functions were usefully viewed as single-valued results for monotone operators. In 1975, Kenderov [30] married this to his knowledge of set-theoretic topology to prove the following lovely result.

Theorem 5 (Kenderov) *Suppose X is a Banach space which admits a dual norm that is strictly convex (as holds in many spaces including all reflexive and all separable spaces).*

Then every (maximal) monotone operator on the space is single-valued on a generic subset of the interior of its domain. In particular any continuous convex function on the space is generically Gâteaux differentiable.

Proof We consider the optimization problem

$$k_T(x) := \inf\{\|x^*\|_*\colon x^* \in T(x)\}, \tag{11}$$

and check (a) that k_T is lower-semicontinuous and (b) that k_T can only be continuous at x if all members of $T(x)$ have the same norm. With our hypotheses this means that T is singleton at points where k_T is continuous. Finally, an easy consequence of *Fort's theorem*[4] [24] assures us that k_T is generically continuous. □

In [31] Kenderov equally spectacularly connected Fréchet differentiability of convex functions to norm-norm continuity of monotone operators.

2.2.5 Gossez

In a series of papers in the early to mid-seventies Gossez [28] introduced the notion of *dense type*: $T \in X^{**} \times X^*$ is of dense type (D): if

$$\inf_{(x,x^*)\in T} \langle x^* - z^*, x - z^{**}\rangle \geq 0$$

(we say (z^*, z^{**}) is *monotonically related* to T) implies there is some bounded net $(x_a, x_a^*) \in T \to_{w^*\times s} (z^{**}, z^*)$. He showed dense-type included all convex ∂f and trivially all reflexive maximal monotone operators. Gossez was thence able to lift part of the theory from reflexive space despite the complete failure of $R(T + J_X) = X^*$. Perhaps more consequentially, he was able to show that a variety of pathologies could occur. In 1974 he produced the following nondense type bounded linear maximal monotone operator.

Example 4 (Gossez operator) The continuous linear map $S : \ell^1(\mathbf{N}) \to \ell^\infty(\mathbf{N})$ is given by

$$(Sx)_n := -\sum_{k<n} x_k + \sum_{k>n} x_k, \quad \forall x = (x_k) \in \ell_1, n \in \mathbf{N}. \tag{12}$$

Then $\mp S : \ell_1 \mapsto \ell_\infty$ is a skew bounded linear operator for which S^* is not monotone but $-S^*$ is. □

[4] **Theorem** (Fort) *An upper semicontinuous mapping of a topological space X into the set of all nonempty compact subsets of a metric space Y is continuous at all points of a residual set in X.*

In 1976 Gossez showed the nonuniqueness of monotone extensions (that is he produced a maximal monotone operator with infinitely many points monotonically related to it but not to each other). In 1977 he showed the first example of a maximal monotone operator whose range was not virtually closed. All of these results are described and extended in [12, Ch. 9].

2.2.6 Asplund

A maximal monotone operator T is *acylic* if whenever $T = \partial f + M$ where f is convex and M is maximal then f is affine. In [2] Asplund proves an interesting decomposition theorem for maximal monotone operators. A modern accounting of it is given in [12, Sect. 9.2]. In particular this implies that a maximal monotone operator (with $\mathrm{int} D(T)$ nonempty) can be written as a sum of a cyclic part (a subgradient) and an acylic part (with no nontrivial subgradient part). Clearly skew mappings are acyclic. A first example of an explicit nonlinear acyclic mapping was given only recently [12, Sect. 9.2].

Example 5 (Borwein-Wiersma (2007)) Define $S : \mathbf{R}^2 \to \mathbf{R}^2$ by

$$S(x, y) := (-y, x) \text{ for } x^2 + y^2 \leq 1. \tag{13}$$

Then the unique maximal monotone extension $\widehat{S}$ of S with range restricted to lie in the unit disc is acyclic and has:

$$\widehat{S}(x) = \sqrt{1 - \frac{1}{\|x\|^2}} \frac{x}{\|x\|} + \frac{1}{\|x\|} S\left(\frac{x}{\|x\|}\right) \tag{14}$$

for $\|x\| \geq 1$.

There are interesting recent extensions by Musev-Ribarska [38], but in general nonlinear acyclic mappings are little understood. Since cyclic and acyclic mappings are in some sense extremal within maximal monotone mappings, they merit more study.

2.2.7 Project independence

The use of monotone operators in intensive modeling and optimization of economic equilibria can be seen as early as the US Department of Energy's *Project Independence* of the Nixon and Ford administrations [4,26]. Originally secret, it was precipitated by the oil price shocks of the period but has clearly not yet come to fruition. The Project Independence Evaluation System (PIES) algorithm [1] in particular modeled demand via monotone operators.[5]

[5] I still have a copy of [1] on my bookshelf. I quote [26] as a testament to 35 years of wasted opportunities.

> "The Federal Energy Administration's Project Independence Blueprint held hopes that the US could reach energy independence by 1985. This study on US energy demand and supply and dependence on oil imports from 1975 to 1985 concludes that these Project Independence goals are unattainable."

My own introduction to monotone operators was the result of a happy mistake. In 1972 my supervisor in Oxford suggested I write my MSc thesis on a paper by Mosco. Fortunately, he directed me to the wrong journal and I studied the then-new-field of maximality rather than set convergence as he had intended.

2.3 1979–1989

2.3.1 Spingarn

From a theoretical point of view, one of the most fruitful algorithmic ideas can be seen in a 1983 paper [49] by Jon Spingarn (who had recently completed his PhD with Rockafellar). In this paper, Spingarn provides an appropriate generalization of the infimal convolution to allow provably-convergent decomposition methods for monotone operator inclusions in Hilbert space. At base is the problem:

For closed orthogonal vector subspaces A and B with $A \oplus B = H$ solve $b \in T(a)$ with $b \in B, a \in A$.

Motivated in part by the proximal point algorithm, Spingarn's iterative algorithm may be applied to efficiently solving monotone inclusions, complementarity problems, convex feasibility problems and more. The *partial inverse*as named by Spingarn [49] is an early example of a very rich and continuing stream of papers on *splitting algorithms* with applications to PDEs, inverse problems, image processing, etc. Some notable contributions are [19,21,25,27,50].

2.3.2 Fitzpatrick

Mathematical Reviews writes:

"In an earlier work E. Krauss (1985) found a representation of monotone operators with the help of subdifferentials of saddle functions on $E \times E$. In the paper under review the author studies a monotone operator $T \subset (E \times E^*)$ by using the **convex function** L_T $E \times E^* \to \mathbf{R} \cup \{+\infty\}$ defined by

$$L_T(x, x^*) = \sup\{\langle x^*, y\rangle + \langle y^*, x - y\rangle : (y, y^*) \in T\}."$$

Simon Fitzpatrick dreamed up this construction when he asked me to suggest a topic to consider for a forthcoming workshop in Canberra at which he was invited to speak. I pointed him to Krauss's paper and the rest as they say is history. Brézis had earlier used $\sup\{\langle x^*, y\rangle : (y, y^*) \in T\}$ without the final term which destroys the convexity but gives L_T its power. Likewise, Fitzpatrick and I had proved local boundedness on int $D(T)$ via continuity of the *convex* function

$$f_T(x) := \sup\{\langle y^*, y - x\rangle : (y^*, y) \in T\}. \quad (15)$$

What Fitzpatrick proved was the following:

Theorem 6 (Fitzpatrick function) *Let T be a maximal monotone operator. Then L_T is convex. Moreover $L_T(x, x^*) \geq \langle x^*, x\rangle$ with equality if and only if $x^*, x \in T$.*

This result lay largely fallow until the next century when it helped reshape the subject entirely.

2.4 1989–1999

2.4.1 Phelps

In 1989 Bob Phelps brought out the first edition of his monograph *Convex Functions, Monotone Operators and Differentiability* [39] which really brought together in one place for the first time what was known on the intersection of the three topics in the title. Ironically, differentiability is the one place where the Fitzpatrick function has so far proved of little use. It is hard to look at a function on $X \times X^*$ to obtain differentiability results in X.

In 1990, Preiss-Phelps-Namioka [40] drew together all the techniques developed since [30,31,53] to show among other things that a maximal monotone operator on a space with a smooth norm is generically single-valued in the interior of its domain, and so that all continuous convex functions on such a space are generically Gâteaux differentiable.

In 1992, Fitzpatrick and Phelps [23] introduced some significant ideas on how to build a replacement for the resolvent and Yosida approximant in a nonreflexive setting. While these ideas are developed a little further and exploited in [8,12] they cry out for further study.

2.4.2 Simons

Perhaps the most significant event of the decade was the publication of [46] in which Simons provided the first accounting of maximal monotonicity within the general confines of convex analysis. This involved the use of—in hindsight—quite complicated sequence spaces and the exploitation of minimax theory in which he is most expert. The second edition [47] a decade later provided a dramatically simpler accounting and was published with an appropriate name change. The convex analysis was now genuinely simple, thanks to Fitzpatrick's function [22].

Simons also introduced another fundamental class of maximal monotone operators: T is of type *negative infimum* (NI) if for all $(z^*, z^{**}) \in (X^*, X^{**})$ one has

$$\inf_{(x^*,x)\in T} \langle x^* - z^*, x - z^{**}\rangle = 0. \tag{16}$$

2.4.3 Bauschke

Another significant step forward was achieved by Bauschke in his 1996 thesis where he showed that *a bounded linear (maximal) monotone operator T has T^* is monotone* *if and only if T is dense type if and only if T is of type (NI)* [6]. It is now known that dense type and (N) coincide in full generality.

Example 6 (The Fitzpatrick-Phelps operator (1996)) The bounded linear operator of Volterra type: $T : L^1[0, 1] \to L^\infty[0, 1]$ given by

$$T(x)(t) := \int_0^t x(s)\,\mathrm{d}s - \int_t^1 x(s)\,\mathrm{d}s \tag{17}$$

is skew but neither $\pm T$ is of dense type since $\pm T^*$ is not monotone.

All known pathologies, such as the Gossez operator given in Example 4 and the Fitzpatrick-Phelps operator of Example 6 are of the form $T + \lambda J_X$ where T is linear monotone but T^* is not monotone. In particular they cannot exist on a Banach lattice X unless $X = \ell^1(\mathbf{N}) \bigoplus Y$ [6]. Thus, at the time of writing is possible that unless a Banach space contains a complemented copy of $\ell^1(\mathbf{N})$ every maximal monotone operator on the space is dense-type. □

2.5 1999–2009

Thanks to the work of a large number of talented researchers this decade has seen a remarkable advance in our understanding of maximal monotone operators. I mention Alves-Marques, Bot, Burachik, Eberhard, Iusem, Martinez-Legas, Penot, Simons, Svaiter, Thera, Vosei, and Zalinescu, who have contributed mightily and I am mindful that this list is incomplete. Full details and accurate citations of this recent work, up to mid-2009, are to be found in [12, Ch. 9].

Figure 1 captures the current situation schematically. Thanks especially to recent work of Alves-Marques with Svaiter and of Voisei with Zalinescu, the chart is pretty complete. It does not mention Simon's interesting class (ED) which is also now known to coincide with (NI) and (D) [48]. I refer yet again to [12] for the definitions of the acronyms used: (ANA) almost negatively aligned; (BR) Brønsted-Rockafellar; (LMM) locally maximal monotone; (FP) Fitzpatrick-Phelps; (MM) maximal monotone.

3 Maximal monotonicity after Fitzpatrick

To finish our idiosyncratic survey let me prove the full form of Minty's theorem. The proof is now marvelously simple and Rockafellar's sum theorem 2 is now almost equally easy (see [13, Sect. 5.1] and [8,12]).

Theorem 7 (Minty surjectivity theorem) *Suppose T is maximal monotone on a reflexive Banach space X. Then*

$$R(T + J_X) = X^*.$$

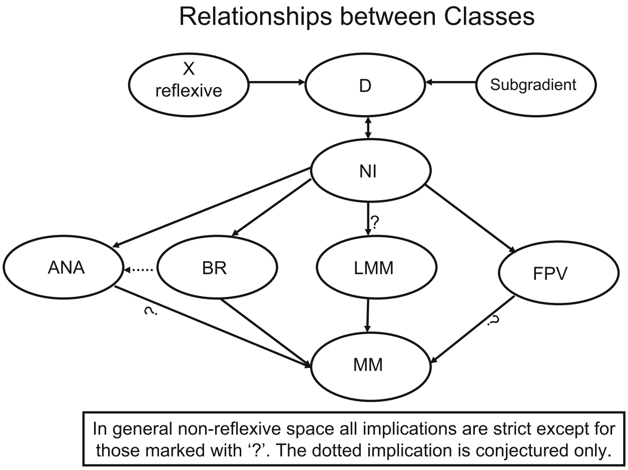

Fig. 1 What we now know

Proof Cauchy's inequality and Fitzpatrick's Theorem 6 implies that for all x, x^*,

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

The Hahn-Banach sandwich theorem [12, Ch. 4] lets us conclude that there exists a point $(w^*, w) \in X^* \times X$ such that

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

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

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

For any $x^* \in Tx$, adding $\langle w^*, w\rangle$ to both sides of the above inequality and noticing $L_T(x, x^*) = \langle x^*, x\rangle$ we obtain

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

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. □

I note that we also now have access to a family of Fitzpatrick functions that captures n-monotonicity and unifies the study of the Fitzpatrick function and of the Rockafellar function (7) as described in [5] and [12, Sect. 9.2]. There is also a rapidly growing corpus of knowledge regarding *enlargements* of maximal monotone operators [18] that may be viewed as analogues of the ε-subdifferential. As highlighted in [8,12,47] there are many other areas where the use of the Fitzpatrick function and other *representative functions* makes previously unanticipated progress possible.

4 What remains to be done

To my mind the most substantial open questions—in addition to the unresolved implications in Fig. 1 and the various questions mentioned in the text—are as follows:

- Does Rockafellar's sum theorem 2 hold in an arbitrary Banach space X? That is, if S and T are maximal monotone and

$$D(S) \cap \operatorname{int} D(T) \neq \emptyset,$$

  is $S + T$ maximal? If (a) int $D(S) \cap$ int $D(T) \neq \emptyset$ or if S, T are of dense type (Voisei-Zalinescu) then $S + T$ is indeed maximal; and this is known under various other hypotheses.
- Does every maximal monotone have $\overline{D(T)}$ convex? This is true for all operators of dense type.
- Are any nonreflexive spaces X of type (D)? That is, are there nonreflexive spaces on which all maximal monotones on X are type (D). I conjecture 'weakly' that if X contains no copy of $\ell^1(\mathbf{N})$ then X is type (D) as would hold in $X = c_0$
- Is there a way to attach a convex function f_T to a maximal monotone operator T so that if $f_T(x)$ is differentiable at x then $T(x)$ is singleton? (That is, is there a convex analogue of (11) which does not rely on the geometry of the norm?)
- Can one formalize the informal observation that the acyclic part of a maximal monotone operator contains all the possible pathologies. That is I conjecture: (a) if T has some bad property so does its acyclic part; (b) if T has some good property so does its cyclic part; (c) if T and ∂f share some good property so does $T + \partial f$.

5 Conclusion

Maximal monotone operator theory has had a recent reflowering. There are still many attractive open questions: theorems to be proved and counter-examples to be found. We are far from knowing the final story.

Acknowledgments I wish to thank two referees for their careful attention and two groups of researchers; those who have participated in this half-century of research and those who are about to contribute.

References

1. Ahn, B.-H.: Computation of Market Equilibria for Policy Analysis : The Project Independence Evaluation System (PIES) Approach. Models demand especially via (anti) monotone operators. Garland Publishing, New York (1979)
2. Asplund, E.: A monotone convergence theorem for sequences of nonlinear mappings. Proc. Symp. Pure Math. **18**, 1–9 (1970)
3. Asplund, E., Rockafellar, R.T.: Gradients of convex functions. Trans. Am. Math. Soc. **139**, 443–467 (1969)
4. Bahn, O., Haurie, A., Zachary, D.S.: Mathematical modelling and simulation models in energy systems. Les Cahier de GERAD, G-2004-41, (2004)
5. Bartz, S., Bauschke, H.H., Borwein, J.M., Reich, S., Wang, X.: Fitzpatrick functions, cyclic monotonicity and Rockafellar's antiderivative. Nonlinear Anal. **66**, 1198–1223 (2007)
6. Bauschke, H.H., Borwein, J.M.: Maximal monotonicity of dense type, local maximal monotonicity, and monotonicity of the conjugate are all the same for continuous linear operators. Pacific J. Math. **189**, 1–20 (1999)
7. Bauschke, H.H., Wang, X.: The kernel average for two convex functions and its application to extensions and representation of monotone operators. Trans. Am. Math. Soc **361**, 5947–5965 (2009)
8. Borwein, J.M.: Maximal monotonicity via convex analysis. J. Conv. Anal. **13**, 561–586 (2006)
9. Borwein, J.M., Lewis, A.S.: Convex Analysis and Nonlinear Optimization. Theory and Examples, CMS Books, vol. 3. Springer, Berlin (2000). Second extended edition (2005)
10. Borwein, J.M., Noll, D.: Second order differentiability of convex functions in Banach spaces. Trans. Am. Math. Soc. **342**, 43–82 (1994)
11. Borwein, J.M., Reich, S., Shafrir, I.: Krasnoselski-Mann iterations in normed spaces. Canad. Math. Bull. **35**, 21–28 (1992)
12. Borwein, J.M., Vanderwerff, J.D.: Convex Functions: Constructions, Characterizations and Counterexamples. Encyclopedia of Mathematics and Applications, vol. 109. Cambridge University Press, UK (2010)
13. Borwein, J.M., Zhu, Q.: Techniques of Variational Analysis, CMS Books, vol. 20. Springer, Berlin (2005)
14. Brézis, H.: Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert (French). North-Holland Mathematics Studies, No. 5. Notas de Matemática (50). North-Holland Publishing Co., Amsterdam, London; Elsevier, New York (1973)
15. Brézis, H., Haraux, A.: Image d'une somme d'opŕateurs monotones et applications. (English summary). Israel J. Math. **23**, 165–186 (1976)
16. Browder, F.E.: Fixed point theory and nonlinear problems. Bull. Am. Math. Soc. **9**, 1–39 (1983)
17. Browder, F.E., Hess, P.: Nonlinear mappings of monotone type in Banach spaces. J. Funct. Anal. **11**, 251–294 (1972)
18. Burachik, R.S., Iusem, A.N.: Set-Valued Mappings and Enlargements of Monotone Operators. Springer Optimization and Its Applications, vol. 8. Springer, Berlin (2008)
19. Combettes, P.L., Pesquet, J.-C.: A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery. IEEE J. Select. Topics Signal Process. **1**, 564–574 (2007)
20. Deimling, K.: Nonlinear Functional Analysis. Springer, Berlin (1985)
21. Eckstein, J., Bertsekas, D.P.: On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program. **55**, 293–318 (1992)
22. Fitzpatrick, S.: Representing monotone operators by convex functions. Proc. Centre Math. Anal. Austral. Nat. Univ. **20**, 59–65 (1988)
23. Fitzpatrick, S., Phelps, R.R.: Bounded approximants to monotone operators on Banach spaces. Ann. Inst. Henri Poincaré, Analyse Non linéaire **9**, 573–595 (1992)
24. Fort, M.K. Jr.: Points of continuity of semi-continuous functions. Publ. Math. Debrecen **2**, 100–102 (1951)
25. Fortin, M., Glowinski, R. (eds.): Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems. North-Holland Publishing Co., Amsterdam (1983)
26. Franssen, H.T.: Towards project independence: energy in the coming decade. Prepared for the Joint Committee on Atomic Energy, United States Congress, OSTI ID: 7283709. Library of Congress, Congressional Research Service, Ocean and Coastal Resources Project (1975)

27. Glowinski, R., Le Tallec, P.: Augmented Lagrangian and Operator-splitting Methods in Nonlinear Mechanics. SIAM, USA (1989)
28. Gossez, J.-P.: Opérateurs monotones non linaires dans les espaces de Banach non réflexifs. J. Math. Anal. Appl. **34**, 371–376 (1971)
29. Gulevich, N.M.: Fixed points of nonexpansive mappings. J. Math. Sci. **79**, 765–815 (1996)
30. Kenderov, P.S.: Semi-continuity of set-valued monotone mappings. Fund. Math. **88**, 61–69 (1975)
31. Kenderov, P.S.: Monotone operators in Asplund spaces. C.R. Acad. Bulgare Sci. **30**, 963–964 (1977)
32. Kartsatos, A.G., Skrypnik, I.V.: On the eigenvalue problem for perturbed nonlinear maximal monotone operators in reflexive Banach spaces. Trans. Am. Math. Soc. **358**, 3851–3881 (2006)
33. Mignot, F.: Contrôle dans les inéquations variationelles elliptiques. J. Funct. Anal. **22**, 130–185 (1976)
34. Minty, G.J.: Monotone Networks. Proc. Roy Soc. Lond. **257**, 194–212 (1960)
35. Minty, G.J.: Monotone (nonlinear) operators in Hilbert space. Duke Math. J. **29**, 341–346 (1962)
36. Minty, G.J.: On the monotonicity of the gradient of a convex function. Pacific J. Math. **14**, 243–247 (1964)
37. Minty, G.J.: On some aspects of the theory of monotone operators. In: Theory and Application of Monotone Operators (Proceedings of NATO Advanced Study Institute, Venice, 1968), pp. 67–82 (1969)
38. Musev, B., Ribarska, N.: On a question of J. Borwein and H. Wiersma. Preprint (2009)
39. Phelps, R.R.: Convex Functions, Monotone Operators and Differentiability. Lecture Notes in Mathematics, vol. 1364. Springer, Berlin (1989). Second edition 1993
40. Preiss, D., Phelps, R.R., Namioka, I.: Smooth Banach spaces, weak Asplund spaces and monotone or usco mappings. Israel J. Math. **72**, 257–279 (1990)
41. Rockafellar, R.T.: On the maximal monotonicity of subdifferential mappings. Pacific J. Math. **33**, 209–216 (1970)
42. Rockafellar, R.T.: On the maximality of sums of nonlinear monotone operators. Trans. Am. Math. Soc. **149**, 75–88 (1970)
43. Rockafellar, R.T.: Convex Analysis. Princeton University Press, USA (1970)
44. Rockafellar, R.T.: On the virtual convexity of the domain and range of a nonlinear maximal monotone operator. Math. Ann. **185**, 81–90 (1970)
45. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**, 877–898 (1976)
46. Simons, S.: Minimax and Monotonicity. Lecture Notes in Mathematics, vol. 1693. Springer, Berlin (1998)
47. Simons, S.: Minimax and Monotonicity. Lecture Notes in Mathematics, vol. 1693, 2nd edition: Renamed From Hahn-Banach to Monotonicity. Springer, Berlin (2008)
48. Simons, S.: Banach SSD spaces and classes of monotone sets. http://arxiv/org/abs/0908.0383v2 posted 26 August (2009)
49. Spingarn, J.E.: Partial inverse of a monotone operator. Appl. Math. Optim. **10**, 247–265 (1983)
50. Tseng, P.: Applications of a splitting algorithm to decomposition in convex programming and variational inequalities. SIAM J. Control Optim. **29**, 119–138 (1991)
51. Zarantonello, E.H.: Projections on Convex Sets in Hilbert Space and Spectral Theory I Projections on Convex Sets. In: Contributions to Nonlinear Functional Analysis. Academic Press, Dublin (1971)
52. Zarantonello, E.H. (ed.): Contributions to Nonlinear Functional Analysis. pp. 237–341. Academic Press, Dublin (1971)
53. Zarantonello, E.H.: Dense single-valuedness of monotone operators. Israel J. Math. **15**, 158–166 (1973)
54. Zeidler, E.: Nonlinear Functional Analysis and its Applications: Part 2 A: Linear Monotone Operators and Part 2 B: Nonlinear Monotone Operators. Springer, Berlin (1990)