Second Order Cones for Maximal Monotone Operators via Representative Functions

A. C. Eberhard and J. M. Borwein*

November 23, 2007

Abstract

It is shown that various first and second order derivatives of the Fitzpatrick and Penot representative functions for a maximal monotone operator T, in a reflexive Banach space, can be used to represent differential information associated with the tangent and normal cones to the Graph T. In particular we obtain formula for the Proto-derivative, as well as its polar, the normal cone to the graph of T. First order derivatives are shown to be useful in recognising points of single-valuedness of T. We show that a strong form of Proto-differentiability to the graph of T, is often associated with single valuedness of T.

Dedicated to Boris Mordukhovich on the occasion of his 60th Birthday.

1 Introduction

A cornerstone of nonsmooth analysis is the construction of the limiting normal cone from more classical normal cone constructions. These robust, nonconvex limiting quantities are essential for practical applications and their construction was first introduced by Mordukhovich [10]. For a detailed history of these developments see [11]. By now an extensive calculus has been developed [11]. Still the construction of some basic quantities which are a starting point for the application of this powerful mathematical machinery is a challenging problem. This is particularly true when considering second order constructions where the normal cone to the graph of a multifunction is required. This program is necessary for calculation the second order subdifferential as introduce by Mordukhovich [12] and further developed in [13] and subsequent papers. In this paper we restrict attention to an important class of multifunctions, namely maximal monotone operators. Of course this class includes the subdifferential of a convex function but much more. Maximal monotone operators have wide application in optimization, differential equations and other areas of mathematics. As shown in many recent papers [3], [5], [17], [18] and many more, the representative function provides a powerful tool for the study of such operators. Also the construction of the Fitzpatrick representative functions appears to be tractable in many cases [7]. When $T : X \rightrightarrows X^*$ is maximal monotone, we denote the *Fitzpatrick representative function,* introduced in [6], by

$$\mathcal{F}_T(z, z^*) := \sup_{(x,x^*)\in \operatorname{Graph} T} \{\langle z, x^*\rangle + \langle x, z^*\rangle - \langle x, x^*\rangle\}$$

and the Penot representative function is given by $\mathcal{P}_T(z, z^*) = \mathcal{F}_T^*(y, y^*)^\dagger$ where $(y, y^*)^\dagger = (y^*, y)$ is the transpose operator.

*The second author's research was funded by NSERC and the Canada Research Chair programme, and the first author's by ARC grant number DP0664423. This study was commenced between August and December 2005 while the first author was visiting Dalhousie University.
2000 Mathematics Subject Classification: 47H05, 46N10, 47H04, 46A20, 49J53.
Keywords and Phrases: Second order cones, maximal monotone operators, Proto–differentiability

The Fitzpatrick function was developed precisely to provide a more transparent convex alternative to the earlier saddle function construction due to Krauss [9]. At the time, Fitzpatrick's interests were more centrally in the differentiation theory for convex functions and monotone operators. The search for results relating when a maximal monotone T is single-valued to differentiability of $\mathcal{F}_T$ did not yield fruit, and he put the function aside. This is still the one area where to the best of our knowledge $\mathcal{F}_T$ has so far proved of little help—in part because generic properties of $\operatorname{dom} \mathcal{F}_T$ and of $\operatorname{dom}(T)$ seem poorly related. In this paper we make some progress towards a theory through which the differentiability properties of the Fitzpatrick and its brother, the Penot representative function, yield insight into the single valuedness of T and also the differentiability properties of T.

As a beginning we opt to study the construction of the so called contingent tangent cone to the graph of T and its polar, the contingent normal cone. This quantity is of interest as a primitive for the construction of limit normal cones. As the theory of representative functions is now sufficiently mature in reflexive spaces, we assume for most part that X is a reflexive space. Suppose that $(z, z^*) \in M := \operatorname{Graph} T$ is a point where the Proto–derivative exists in the sense that the following limit exists:

$$T_M (z, z^*) := b\text{-}\lim_{t \downarrow 0} \frac{1}{t} (M - (z, z^*)),$$

where $b\text{-}\lim_{t\downarrow 0}$ refers to convergence in the Attouch–Wetts sense [1]. This is somewhat a restrictive assumption but the basic approach developed here is expected to be able to weakened as we develop a better understanding of the transmission of certain set limits through the Fenchel conjugate when applied to a specific class of *nonconvex* functions.

In this paper we show that the polar cone

$$T_M (z, z^*)^{\circ} := \{(y^*, y) \mid \langle (y, y^*), (h, h^*)\rangle \le 0, \forall (h, h^*) \in T_M (z, z^*)\}$$

is characterised by the identities

$$T_M (z, z^*)^{\circ} = \partial \mathcal{P}_T'' (z, z^*) (0, 0) = \{(v, v^*) \mid \mathcal{F}_T'' (z, z^*) (v, v^*) \le 0\}$$

where we show that the second order directional derivative given by

$$\begin{aligned} \mathcal{F}_T'' (z, z^*) (v, v^*) &:= b\text{-}e\text{-}\lim_{t\downarrow 0} \frac{1}{t^2} \{\mathcal{F}_T((z, z^*) + t (v, v^*)) - \mathcal{F}_T(z, z^*) - t\langle (v, v^*), (z, z^*)\rangle\} \\ &= \mathcal{F}_{T_M(z,z^*)} (v, v^*) \qquad \text{for all } (v, v^*) \in X \times X^*. \end{aligned}$$

Similarly $\mathcal{P}_T'' (z, z^*) (v, v^*) = \mathcal{P}_{T_M(z,z^*)} (v, v^*)$ for all $(v, v^*) \in X \times X^*$ under a strong Proto–differentiability assumption along with

$$\mathcal{P}_T'' (z, z^*) (v, v^*) = (\mathcal{F}_T'' (z, z^*))^* (v, v^*)^{\dagger}.$$

Moreover when M is strongly Proto-differentiable at (z, z^*) and $\mathcal{F}_{T_M(z,z^*)+(z,z^*)}$ is Fréchet differentiable at (z, z^*) then $T(z) = \{z^*\}$ with

$$d^* (z^*, T (y)) = O (\|y - z\|),$$

a very strong form of single valuedness of T. In addition $T_M (z, z^*)$ must then be maximal monotone. We also show that $\mathcal{F}_{T_M(z,z^*)+(z,z^*)}$ is Gâteaux differentiable at (z, z^*) for all $(z, z^*) \in M$ when $(\overline{\text{co}} T_M (z, z^*))^{\circ} \cap \overline{\text{co}} T_M (z, z^*)^{\dagger} = \{0\}$. Moreover the Fréchet differentiability of $\mathcal{F}_{T_M(z,z^*)+(z,z^*)}$ at $(z, z^*) \in M$ is equivalent to the condition

$$\inf_{(v,v^*) \in T_M(z,z^*)/(0,0)} \frac{1}{\|(v, v^*)\|^2} \langle v, v^*\rangle > 0.$$

In general we have

$$\operatorname{diam} \{z^* \mid (z^*, y) \in \partial \mathcal{F}_T(y, y^*)\} \le \varepsilon \quad \Longrightarrow \quad \operatorname{diam} T (y) \le \varepsilon.$$

2 The Subdifferential and Gradient of a Maximal Monotone Operator

For most of this paper we assume X is a reflexive Banach space although the results of this first section hold true in an arbitrary Banach space. We may view $X\times X^*$ paired with $X^*\times X$ using the coupling $\langle (y,y^*),(x^*,x)\rangle = \langle y,x^*\rangle + \langle x,y^*\rangle$. For convenience we will use $\langle (y,y^*),(x^*,x)\rangle \equiv \langle (y,y^*),(x,x^*)\rangle$. The indicator function of a set $C\subseteq X\times X^*$ is denoted by δ_C and the Fenchel conjugate of a convex function $F: X\times X^* \to X^*\times X$ is denote by $F^*(x^*,x)$. The epi-graph of F corresponds to $\operatorname{epi} F := \{(x,x^*,\alpha)\in X\times X^*\times \mathbf{R} \mid \alpha \geq F(x,x^*)\}$. For a multi–function $\mathcal{G}: X \rightrightarrows Y$ we denote its graph by $\operatorname{Graph}\mathcal{G} := \{(x,y)\in X\times Y \mid y\in \mathcal{G}(x)\}$. Denote the class of closed, proper, convex functions on $X\times X^*$ by $\Gamma(X\times X^*)$.

We say T is a monotone operator when $M := \operatorname{Graph} T$ is a monotone set i.e.

$$(\forall\,(x,x^*)\in M)\,(\forall\,(y,y^*)\in M)\quad \langle x-y, x^*-y^*\rangle \geq 0. \tag{1}$$

If M does not posses a proper extension that is still monotone then M is said to maximal monotone. We say (y,y^*) is monotonically related to M when $(\forall\,(x,x^*)\in M)$ we have $\langle x-y,x^*-y^*\rangle\geq 0$. When M is maximal then $(y,y^*)\notin M$ implies the existence of $(x,x^*)\in M$ such that $\langle x-y,x^*-y^*\rangle<0$.

Definition 1 *The Fitzpatrick function associated with an operator $T: X\rightrightarrows X^*$ is defined by*

$$\begin{aligned}\mathcal{F}_T(y,y^*) &:= \sup_{(x,x^*)\in \operatorname{Graph} T}\{\langle y,x^*\rangle + \langle x,y^*\rangle - \langle x,x^*\rangle\}\\ &= \left\{\langle y,y^*\rangle - \inf_{(x,x^*)\in\operatorname{Graph} T}\langle y-x,y^*-x^*\rangle\right\}\end{aligned}$$

From this definition it is easily see that when T is monotone for $y^*\in T(y)$ we have

$$\mathcal{F}_T(y,y^*) = \langle y,y^*\rangle. \tag{2}$$

When T is maximal we have $\mathcal{F}_T(y,y^*)\geq \langle y,y^*\rangle$ holding for all (y,y^*) with equality if and only if $y^*\in T(y)$. Define the 'transpose' operator $\dagger$: $(x^*,x)\to(x,x^*)$. One can easily show that $(\delta_{\operatorname{Graph} T}(\cdot)+\langle\cdot,\cdot\rangle)^*(y,y^*)^\dagger = \mathcal{F}_T(y,y^*)$. The second conjugate of $\delta_{\operatorname{Graph} T}(\cdot)+\langle\cdot,\cdot\rangle$ is of interest in that

$$\begin{aligned}&\mathcal{P}_T(y,y^*) := \mathcal{F}_T^*(y,y^*)^\dagger = (\delta_{\operatorname{Graph} T}(\cdot)+\langle\cdot,\cdot\rangle)^{**}(y,y^*) = \overline{p_T(y,y^*)} \quad \text{where}\\ &p_T(y,y^*) = \inf\{\sum_i \lambda_i\langle x_i,x_i^*\rangle \mid \sum_i \lambda_i\,(x_i,x_i^*,1) = (y,y^*,1),\,(x_i,x_i^*)\in\operatorname{Graph} T,\ \lambda_i\geq 0\}.\end{aligned}$$

Recall that a representative of a monotone mapping T on X is a convex function $\mathcal{H}_T$ on $X\times X^*$ such that $\mathcal{H}_T(y,y^*)\geq\langle y,y^*\rangle$ for all $(y,y^*)\in X\times X^*$ with $\mathcal{H}_T(y,y^*)=\langle y,y^*\rangle$ when $y^*\in T(y)$. The following is now well established.

Lemma 2 *[3] For any monotone mapping T the function $\mathcal{P}_T: X\times X^*\to\overline{\mathbf{R}}$ is a representative convex function for T.*

When T is maximal then $\mathcal{F}_T$ is the smallest representative functions and $\mathcal{P}_T$ is the largest.

Next we define an important multi-function in our study. Let

$$\mathcal{M}_T(z,z^*) := \{(a^*,a)\in X^*\times X \mid \mathcal{F}_T(z,z^*)\leq\langle z,a^*\rangle+\langle a,z^*\rangle-\langle a,a^*\rangle\}.$$

Note that for all $(z,z^*)\in\operatorname{Graph} T$ we have $(z^*,z)\in\mathcal{M}_T(z,z^*)$ (since $\mathcal{F}_T(z,z^*)=\langle z,z^*\rangle$). Recall that the subdifferential is defined as

$$\partial\mathcal{F}_T(z,z^*) := \{(y^*,y) \mid \mathcal{F}_T(z,z^*)+\mathcal{P}_T(y,y^*) = \langle (z,z^*),(y,y^*)\rangle\}.$$

We intend to characterise the subdifferential of the Fitzpatrick function in terms of $\mathcal{M}_T$ and T only but we first include some elementary properties of the multifunction $\mathcal{M}_T$.

Lemma 3 *Suppose $T : X \rightrightarrows X^*$ is monotone and $(a, a^*) \in \operatorname{Graph} T$. Then*

$$\mathcal{M}_T(a, a^*) = \{(z^*, z) \in X^* \times X \mid \langle z - a, z^* - a^* \rangle \le 0\}$$

and so

$$\begin{aligned}\mathcal{M}_T(a, a^*) \cap \operatorname{Graph} T^\dagger &= \left\{(z^*, z) \in \operatorname{Graph} T^\dagger \mid \langle z, a^* \rangle + \langle a, z^* \rangle - \langle z, z^* \rangle = \mathcal{F}_T(a, a^*)\right\} \\ &= \left\{(z^*, z) \in \operatorname{Graph} T^\dagger \mid \langle z - a, z^* - a^* \rangle = 0\right\} \\ &\subseteq \{(z^*, z) \in X^* \times X \mid \langle z - a, z^* - a^* \rangle = 0\}.\end{aligned}$$

Proof. If $(z^*, z) \in \mathcal{M}_T(a, a^*)$ then by definition

$$\begin{aligned}\langle z, a^* \rangle + \langle a, z^* \rangle - \langle z, z^* \rangle \ge \mathcal{F}_T(a, a^*) = \langle a, a^* \rangle \quad &(\text{since } (a, a^*) \in \operatorname{Graph} T) \\ \text{or } \langle z, a^* \rangle + \langle a, z^* \rangle - \langle z, z^* \rangle - \langle a, a^* \rangle &\ge 0 \\ \text{equivalently} \quad \langle z, a^* - z^* \rangle + \langle a, z^* - a^* \rangle \ge 0 \quad \text{or} \quad &\langle z - a, z^* - a^* \rangle \le 0.\end{aligned}$$

If $(z, z^*) \in \operatorname{Graph} T$ then $\langle z - a, z^* - a^* \rangle \ge 0$ by monotonicity giving equality.
When $T : X \rightrightarrows X^*$ is maximal monotone and $(a, a^*) \in \operatorname{Graph} T$ then the points (z, z^*) on the boundary of $\mathcal{M}_T(a, a^*)$ satisfy $\langle z - a, z^* - a^* \rangle = 0$. ■

Lemma 4 *Suppose $T : X \rightrightarrows X^*$ is maximal monotone. Then $\langle z - a, z^* - a^* \rangle = 0$ and $(z^*, z) \in \mathcal{M}_T(a, a^*)$ imply $(a, a^*) \in \operatorname{Graph} T$.*

Proof. It follows that

$$\begin{aligned}\langle a, a^* \rangle \le \mathcal{F}_T(a, a^*) &\le \langle (z, z^*), (a, a^*) \rangle - \langle z, z^* \rangle \\ &= \langle z, a^* \rangle + \langle a, z^* \rangle - \langle z, z^* \rangle = \langle a, a^* \rangle\end{aligned}$$

implying $\langle a, a^* \rangle = \mathcal{F}_T(a, a^*)$ or $(a, a^*) \in \operatorname{Graph} T$ (by maximality). ■
To our knowledge the subdifferential of the Fitzpatrick functions has not yet been full characterised. The next result is a step in that direction.

Proposition 5 *Suppose $T : X \rightrightarrows X^*$ is monotone and $(z, z^*) \in X \times X^*$. Then*

$$\begin{aligned}\mathcal{M}_T(z, z^*) \cap \overline{\text{co}}\ \operatorname{Graph} T^\dagger \supseteq \partial \mathcal{F}_T(z, z^*) \ &\supseteq \overline{\text{co}}\left(\mathcal{M}_T(z, z^*) \cap \operatorname{Graph} T^\dagger\right) \\ &\supseteq \partial \mathcal{F}_T(z, z^*) \cap \operatorname{Graph} T^\dagger = \mathcal{M}_T(z, z^*) \cap \operatorname{Graph} T^\dagger.\end{aligned}$$

Proof. Now $(a^*, a) \in \partial \mathcal{F}_T(z, z^*)$ iff we have

$$\mathcal{F}_T(z, z^*) + \mathcal{P}_T(a, a^*) = \langle (z, z^*), (a, a^*) \rangle. \tag{3}$$

Next note that as $\mathcal{F}_T(z, z^*) = (\langle \cdot, \cdot \rangle + \delta_T(\cdot, \cdot))^*(z, z^*)$ we have

$$\mathcal{P}_T(a, a^*) = \{\langle \cdot, \cdot \rangle + \delta_T(\cdot, \cdot)\}^{**}(a, a^*) = \overline{\text{co}}\{\langle \cdot, \cdot \rangle + \delta_T(\cdot, \cdot)\}(a, a^*).$$

Note that $(a, a^*) \notin \overline{\text{co}}\ \operatorname{Graph} T$ then then $\mathcal{P}_T(a, a^*) = +\infty$ which invalidates (3). If $(a^*, a) \in \partial \mathcal{F}_T(z, z^*)$ then

$$\begin{aligned}\mathcal{F}_T(z, z^*) &\le \langle (z, z^*), (a, a^*) \rangle - \mathcal{P}_T(a, a^*) \\ &\le \langle (z, z^*), (a, a^*) \rangle - \langle a, a^* \rangle \quad (\text{always as } \langle a, a^* \rangle \le \mathcal{P}_T(a, a^*)) \\ &\le \mathcal{F}_T(z, z^*) \quad (\text{when } (a, a^*) \in \operatorname{Graph} T).\end{aligned} \tag{4}$$

The second inequality shows that $\mathcal{M}_T(z, z^*) \supseteq \partial \mathcal{F}_T(z, z^*)$ and so

$$\mathcal{M}_T(z, z^*) \cap \overline{\text{co}}\ \operatorname{Graph} T^\dagger \supseteq \partial \mathcal{F}_T(z, z^*).$$

When $(a^*, a) \in \partial \mathcal{F}_T(z, z^*) \cap \operatorname{Graph} T^\dagger$ we have $(a^*, a) \in \mathcal{M}_T(z, z^*) \cap \operatorname{Graph} T^\dagger$. Conversely let

$$(a^*, a) \in \mathcal{M}_T(z, z^*) \cap \operatorname{Graph} T^\dagger = \{(a^*, a) \in \operatorname{Graph} T \mid \mathcal{F}_T(z, z^*) \leq \langle z, a^* \rangle + \langle a, z^* \rangle - \langle a, a^* \rangle\}$$

then as $\mathcal{P}_T(a, a^*) = \langle a, a^* \rangle$ we have

$$\mathcal{F}_T(z, z^*) \leq \langle (y, y^*), (a, a^*) \rangle - \langle a, a^* \rangle = \langle (y, y^*), (a, a^*) \rangle - \mathcal{P}_T(a, a^*)$$

and so

$$\mathcal{F}_T(z, z^*) + \mathcal{P}_T(a, a^*) \leq \langle (y, y^*), (a, a^*) \rangle$$

or $(a^*, a) \in \partial \mathcal{F}_T(z, z^*) \cap \operatorname{Graph} T^\dagger$. From convexity of the subdifferential it follows that

$$\begin{aligned} \partial \mathcal{F}_T(z, z^*) &\supseteq \overline{\operatorname{co}}\left(\mathcal{M}_T(z, z^*) \cap \operatorname{Graph} T^\dagger\right) \\ &\supseteq \partial \mathcal{F}_T(z, z^*) \cap \operatorname{Graph} T^\dagger = \mathcal{M}_T(z, z^*) \cap \operatorname{Graph} T^\dagger. \end{aligned}$$

■

One is lead to conjecture that $\partial \mathcal{F}_T(z, z^*) = \overline{\operatorname{co}}\left(\mathcal{M}_T(z, z^*) \cap \operatorname{Graph} T^\dagger\right)$ but a proof is still allusive.

Remark 6 *Note that when T is only monotone we can have the inequality $\mathcal{F}_T(z, z^*) \geq \langle z, z^* \rangle$ failing for some $(z, z^*) \notin \operatorname{Graph} T$. As $\partial \mathcal{F}_T(z, z^*) \supseteq \mathcal{M}_T(z, z^*) \cap \operatorname{Graph} T^\dagger$ and $(z^*, z) \in \mathcal{M}_T(z, z^*) \cap \operatorname{Graph} T^\dagger$ whenever $(z, z^*) \in \operatorname{Graph} T$ then $\partial \mathcal{F}_T(z, z^*) \supseteq \{(z^*, z)\}$ always when $(z, z^*) \in \operatorname{Graph} T$.*

Recall the fundamental inequality (see [17] or [3]) that

$$\mathcal{P}_T(y, z^*) + \mathcal{P}_T(z, y^*) \geq \langle y, y^* \rangle + \langle z, z^* \rangle. \tag{5}$$

We may now draw an interesting implication for $T(y)$.

Theorem 7 *Suppose $T : X \rightrightarrows X^*$ is monotone. If there exists $(y, y^*), (y, z^*) \in \operatorname{Graph} T$ with $y^* \neq z^*$ (i.e. $T(y) \supseteq \{y^*, z^*\}$ is not unique) then (y^*, y), $(y^*, z) \in \partial \mathcal{F}_T(y, y^*)$ and so $\partial \mathcal{F}_T(y, y^*) \cap \operatorname{Graph} T^\dagger$ is also not a singleton. Consequently when $(y, y^*) \in \operatorname{Graph} T$ and $\nabla \mathcal{F}_T(y, y^*)$ exists then $T(y)$ is a singleton. More generally we have*

$$\begin{aligned} &\operatorname{diam}\{z^* \mid (z^*, y) \in \partial \mathcal{F}_T(y, y^*)\} \leq \varepsilon \\ &\quad \Longrightarrow \operatorname{diam}\left\{z^* \mid (z^*, y) \in \mathcal{M}_T(y, y^*) \cap \operatorname{Graph} T^\dagger\right\} \\ &\quad \Longleftrightarrow \operatorname{diam}\left\{z^* \mid (z^*, y) \in \partial \mathcal{F}_T(y, y^*) \cap \operatorname{Graph} T^\dagger\right\} \leq \varepsilon \\ &\quad \Longrightarrow \quad \operatorname{diam}\{z^* \mid (z^*, y) \in \partial \mathcal{F}_T(y, y^*) \cap (T(y), y)\} \leq \varepsilon \quad \Longrightarrow \quad \operatorname{diam} T(y) \leq \varepsilon. \end{aligned} \tag{6}$$

Proof. Suppose $(y, z^*) \in \operatorname{Graph} T$ then on using (5) and the fact that $y^* \in T(y)$ we have

$$\begin{aligned} &\langle y, y^* \rangle + \langle y, z^* \rangle - \langle y, z^* \rangle \\ &= \{\mathcal{P}_T(y, y^*) + \mathcal{P}_T(y, z^*)\} - \langle y, z^* \rangle \\ &\geq \{\langle y, y^* \rangle + \langle y, z^* \rangle\} - \langle y, z^* \rangle = \langle y, y^* \rangle \\ &= \mathcal{P}_T(y, y^*) \geq \mathcal{F}_T(y, y^*) \end{aligned}$$

implying $(y^*, z) \in \mathcal{M}_T(y, y^*) \cap \operatorname{Graph} T^\dagger \subseteq \partial \mathcal{F}_T(y, y^*)$. We always have $(y^*, y) \in \partial \mathcal{F}_T(y, y^*)$ when $(y, y^*) \in \operatorname{Graph} T$. Thus when $\nabla \mathcal{F}_T(y, y^*) = (y^*, y)$ we cannot have $z^*, y^* \in T(y)$ with $z^* \neq y^*$ for otherwise we have $(y^*, y), (y^*, z) \in \partial \mathcal{F}_T(y, y^*)$, a contradiction. Finally note that due to Proposition 5 the first implication of (6) are obvious and then if $\operatorname{diam}\{z^* \mid (z^*, y) \in \partial \mathcal{F}_T(y, y^*) \cap (T(y), y)\} \leq \varepsilon$ there can't exist $T(y) \supseteq \{y^*, z^*\}$ with $\|y^* - z^*\| > \varepsilon$. ■

Remark 8 *When $\nabla \mathcal{F}_T(x, x^*)$ exists as a Fréchet derivative, $\varepsilon > 0$ and (y, y^*) is sufficiently close to (x, x^*) then one can show that $\operatorname{diam}\{z^* \mid (z^*, y) \in \partial \mathcal{F}_T(y, y^*)\} \leq \varepsilon$.*

Corollary 9 *Suppose $T : X \rightrightarrows X^*$ is monotone and $(y, y^*) \in \operatorname{Graph} T$. Suppose in addition $\nabla_x \mathcal{F}_T(y, y^*)$ exists then $T(y) = \{y^*\}$ is a singleton.*

Proof. Suppose $\nabla_x \mathcal{F}_T(y, y^*)$ exists. First note that we always have $(y, y^*) \in \mathcal{M}_T(z, z^*) \cap \operatorname{Graph} T$ and so by the subgradient inequality

$$\langle y^*, v - y\rangle + \langle y, v^* - y^*\rangle \leq \mathcal{F}_T(v, v^*) - \mathcal{F}_T(y, y^*).$$

Placing $v^* = y^*$ we obtain

$$\langle y^*, v - y\rangle \leq \mathcal{F}_T(v, y^*) - \mathcal{F}_T(y, y^*)$$

and so $y^* \in \partial_x \mathcal{F}_T(y, y^*)$ and so $\nabla_x \mathcal{F}(y, y^*) = \{y^*\}$.
Now suppose $\operatorname{Pr}_{X^*} \partial \mathcal{F}_T(y, y^*) \supseteq \{z^*, y^*\}$ then we have the existence of z such that $(z^*, z) \in \partial \mathcal{F}_T(y, y^*)$ and by the subgradient inequality for all (v, v^*) that

$$\langle z^*, v - y\rangle + \langle z, v^* - y^*\rangle \leq \mathcal{F}_T(v, v^*) - \mathcal{F}_T(y, y^*).$$

Place $v^* = y^*$ to obtain

$$\langle z^*, v - y\rangle \leq \mathcal{F}_T(v, y^*) - \mathcal{F}_T(y, y^*)$$

and so $z^* \in \partial_x F_T(y, y^*)$ implying $z^* = y^*$. Thus $\partial F_T(y, y^*) = \{y^*\} \times \partial_{x^*} F_T(y, y^*)$.
If $T(y)$ is not a singleton then there exists $(y, y^*), (y, z^*) \in \operatorname{Graph} T$ with $y^* \neq z^*$ which implies by Theorem 7 that

$$(y^*, y), (z^*, y) \in \partial F_T(y, y^*) = \{y^*\} \times \partial_{x^*} \mathcal{F}_T(y, y^*)$$

in which case $y^* = z^*$, a contradiction. Thus $T(y)$ is a singleton. ■

Recall that in an Asplund space a finite convex function defined on a open convex subset is Fréchet differentiable on a G_δ dense subset of its domain. Recall also that Asplund spaces include those that admit a Fréchet differentiable equivalent norm and these include reflexive spaces. The difficulty in using Lemma 9 to show generic single-valuednessis is that $\operatorname{co} \operatorname{Graph} T$ does not necessarily have an interior.

3 The Fitzpatrick Function of the Tangent cone as a Derived Fitzpatrick Function

In this section we discuss how one may use the Fitzpatrick function of a maximal monotone operator T to obtain the Fitzpatrick function for the multifunction whose graph is the tangent cone to the graph of T. The following characterisation of Mosco convergence follows from Proposition 5.4.8 of [2]. By $z_\beta \to z$ we denote the strong convergence of a net and by $z^*_\beta \to^w z^*$ weak convergence.

Definition 10 *Let $\{T_\beta\}_{\beta\in\Lambda}$ be a net of sets in a reflexive Banach space Z. Then $T = M\text{-}\lim_\beta T_\beta$ iff both*

1. $\liminf_\beta T_\beta := \left\{z \in Z \mid \forall z \in T \text{ and subnet } \beta_\gamma, \exists z_{\beta_\gamma} \in T_{\beta_\gamma} \text{ with } z_{\beta_\gamma} \to z\right\} \supseteq T$ *and*
2. $b\text{-}w\text{-}\limsup_\beta T_\beta$

$$:= \{z \in Z \mid \exists M > 0 \text{ and a subnet } \{\beta_\gamma\} \text{ with } z_{\beta_\gamma} \in T_{\beta_\gamma} \cap \overline{B}_M(0) \text{ s.t. } z_{\beta_\gamma} \to^w z\} \subseteq T.$$

Remark 11 *When we replace 2. with the stronger limit supremum*

$$\limsup_\beta T_\beta := \{z \in Z \mid \exists \text{ a subnet } \{\beta_\gamma\} \text{ with } z_{\beta_\gamma} \in T_{\beta_\gamma} \text{ s.t. } z_{\beta_\gamma} \to^w z\} \subseteq T$$

then we say that the Kuratowski–Painlevé limit exists.

Definition 12 *Let $\{T_\beta\}_{\beta\in\Lambda}$ be a net of sets in the Banach space $X \times X^*$. Then denote*

$$b\text{-}s \times w\text{-}\limsup_\beta T_\beta := \{(x, x^*) \in X \times X^* \mid \exists M > 0 \text{ and a subnet } \{\beta_\gamma\} \text{ with } (x_{\beta_\gamma}, x^*_{\beta_\gamma}) \in T_{\beta_\gamma} \cap \overline{B}_M(0) \text{ s.t. } (x_{\beta_\gamma}, x^*_{\beta_\gamma}) \to^{s\times w} (x, x^*)\}.$$

Alternatively one may define these limits via hit-miss topologies. Denote

$$V^{-} := \{W \neq \emptyset \mid W \text{ closed and } V \cap W \neq \emptyset\}$$

for V norm open and $(K^c)^+ := \{W \neq \emptyset \mid W \subseteq K^c\}$ for K weakly compact. The inclusion in 1. may be expressed as: if $T \in V^-$ for some norm open set then $T_\beta \in V^-$ eventually. The inclusion in 2. can be expressed as: if $T \in (K^c)^+$ then eventually $T_\beta \in (K)^c$ i.e. $T \cap K = \emptyset$ implies $T_\beta \cap K = \emptyset$ eventually.

In this paper we exclusively deal with families of sets parametrised by the positive number $\mathbf{R}_+$ i.e. $\{T_t\}_{t>0}$. As $\mathbf{R}_+$ is totally ordered the convergence of such families may be dealt with considering the convergence of all subfamilies $\{T_{t_n}\}_n$ where $t_n \downarrow 0$.

Proposition 13 *Let $\{T_\beta\}_{\beta\in\Lambda}$ be a net of sets of $X \times X^*$ for a reflexive Banach space X.*

1. *Then $T \supseteq b\text{-}s \times w\text{-}\limsup_\beta T_\beta$ implies the property:*

   (P): For all K weakly compact in X^* and C strongly compact in X, such that $(C \times K) \cap T = \emptyset$, implies that eventually $(C \times K) \cap T_\beta = \emptyset$.

2. *It $\{T_n\}_{n\in\mathbf{N}}$ is a sequence of sets (i.e. $\Lambda = \mathbf{N}$) then (P) implies $T \supseteq b\text{-}s \times w\text{-}\limsup_n T_n$.*

Proof. Assume $T \supseteq b\text{-}s \times w\text{-}\limsup_\beta T_\beta$. The contrapositive of the second proposition corresponds to: If, for some K weakly compact in X^* and C strongly compact in X, we have $(C \times K) \cap T_\beta \neq \emptyset$ infinitely often then $(C \times K) \cap T \neq \emptyset$. By the supposed compactness there exists a subnet $\left(x_{\beta_\gamma}, x^*_{\beta_\gamma}\right) \in (C \times K) \cap T_{\beta_\gamma}$ such that $\left(x_{\beta_\gamma}, x^*_{\beta_\gamma}\right) \to^{s\times w} (x, x^*)$. As $\left(x_{\beta_\gamma}, x^*_{\beta_\gamma}\right) \in \overline{B}_M(0) \cap T_\beta$ for

$$M = \sup\{\|(v, v^*)\| \mid (v, v^*) \in (C \times K)\} < \infty$$

we have $(x, x^*) \in T$ and $(C \times K) \cap T_\beta \neq \emptyset$.

Now suppose $\Lambda = \mathbf{N}$ and take

$$(x, x^*) \in b\text{-}s \times w - \limsup_n T_n.$$

Then, by definition, there exists a subsequence $(x_{n_m}, x^*_{n_m}) \in T_{n_m} \cap \overline{B}_M(0)$ with $(x_{n_m}, x^*_{n_m}) \to^{s \times w} (x, x^*)$. Note that for all N

$$\begin{aligned} &\{(x, x^*)\} \cup \left\{(x_{\beta_{\gamma_n}}, x^*_{\beta_{\gamma_n}}) \mid n \geq N\right\} \\ &\subseteq \left(\{x\} \cup \{x_{\beta_{\gamma_n}} \mid n \geq N\}\right) \times \left(\{x^*\} \cup \left\{x^*_{\beta_{\gamma_n}} \mid n \geq N\right\}\right) \end{aligned}$$

which is of the form $C_N \times K_N$ with C_N strongly compact and K_N weakly compact due to the fact that K_N is bounded weakly and closed. Thus $(C_N \times K_N) \cap T_\beta \neq \emptyset$ infinitely and so we may conclude that $(C_N \times K_N) \cap T \neq \emptyset$, irrespective of the choice of N. Thus we conclude that $(x, x^*) \in T$. ∎

Consider the following characterisation of the upper Mosco-limit. From [2] Proposition 5.4.8 that $M\text{-}\limsup_\beta T_\beta \subseteq T$ iff for every weakly compact set $K \subseteq X \times X^*$ we have

$$\begin{aligned} T &\supseteq \limsup_\beta (K \cap T_\beta) \\ &:= \left\{(x, x^*) \mid \exists \{\beta_\gamma\} \text{ and } \left(x_{\beta_\gamma}, x^*_{\beta_\gamma}\right) \in K \cap T_{\beta_\gamma} \text{ with } \left(x_{\beta_\gamma}, x^*_{\beta_\gamma}\right) \to^w (x, x^*)\right\}. \end{aligned}$$

In particular this implies $\limsup_\beta \left(\overline{B}_M(0) \cap T_\beta\right) \subseteq T$ for all $M > 0$. We may obtain a similar characterisation of our convergence notion.

Proposition 14 *Let $\{T_n\}_{n\in\mathbf{N}}$ be a sequence of sets in $X \times X^*$ for a reflexive Banach space X. Then $T \supseteq b\text{-}s \times w\text{-}\limsup_n T_n$ iff for all K weakly compact in X^* and C strongly compact in X we have*

$$\limsup_n ((K \times C) \cap T_n) \subseteq T. \tag{7}$$

Proof. The forward implication is obvious thus we only address the later and so assume (7) holds for all choices of K and C. Suppose K weakly compact in X^* and C strongly compact in X and $(K \times C) \cap T_n \neq \emptyset$ frequently. By the compactness assumption b-$s \times w$-$\limsup_n (K \times C) \cap T_n$ is a nonempty subset of $K \times C$. Now suppose (7) fails for some $K \times C$ then take

$$(x, x^*) \in \left[b\text{-}s \times w\text{-} \limsup_n (K \times C) \cap T_n \right] \backslash T$$

and a $s \times w$-closed neighbourhood $U_1 \times U_2$ disjoint from T. Let $K_1 := K \cap U_1$ and $C_1 := C \cap U_2$. Then K_1 is weakly compact and C_1 strongly compact with $[K_1 \times C_1] \cap T = \emptyset$. Yet by construction $((K_1 \times C_1) \cap T_n) \neq \emptyset$ frequently contrary to property (P) of Proposition 13. ■

Consequently we obtain the following:

Corollary 15 *Let $\{T_\beta\}_{\beta \in \Lambda}$ be a sequence of sets in $X \times X^*$ for a reflexive Banach space X. Then*

$$b\text{-}s \times w\text{-} \limsup_n T_n \subseteq M\text{-} \limsup_n T_n. \tag{8}$$

As is usual in variational analysis when we have a net of functions $\{f_\alpha\}$ we say

$$f \leq b\text{-}s \times w\text{-}e\text{-} \liminf_\alpha f_\alpha \quad \text{iff} \quad b\text{-}s \times w\text{-} \limsup_\alpha \operatorname{epi} f_\alpha \subseteq \operatorname{epi} f. \tag{9}$$

For a detailed account of such convergences one can consult [2], chapter 3 and also [16]. Consequently (8) implies b-$s \times w$-e-$\liminf_\alpha f_\alpha \geq e$-$\liminf_\alpha f_\alpha$. We will say that

$$f = b\text{-}s \times w\text{-}e\text{-} \lim_\alpha f_\alpha$$

if in addition to (9) we have $\liminf_\alpha \operatorname{epi} f_\alpha \supseteq \operatorname{epi} f$. We say the epi-limit e-$\lim_\alpha f_\alpha = f$ exists iff

$$m\text{-} \lim_\alpha \operatorname{epi} f_\alpha = \operatorname{epi} f.$$

It is clear that e-$\lim_\alpha f_\alpha = f$ implies $f = b$-$s \times w$-e-$\lim_\alpha f_\alpha$.

Recall that a monotone operator T is maximal monotone locally or of type VFP iff Graph $T \cap (U \times X^*)$ is maximal monotone in $U \times X^*$ for every open subset U of X. It is well known that all maximal monotone operators are of type VFP in a reflexive space [3] and it has been recently been shown that this also is true in a general Banach space when dom (T) is either closed or has non-empty interior [5].

Proposition 16 *Suppose $T : X \rightrightarrows X^*$ is maximal monotone and $(z, z^*) \in M :=$ Graph T. Let*

$$T_M(z, z^*) := b\text{-}s \times w\text{-} \limsup_{t \downarrow 0} \frac{1}{t} (M - (z, z^*)). \tag{10}$$

Then $T_M(z, z^*)$ is also a monotone set. If in addition T is norm to norm continuous at (z, z^*) then $T_M(z, z^*)$ is also maximal monotone subset of $X \times X^*$.

Proof. When $(h, h^*) \in T_M(z, z^*)$ then for $t_\beta \downarrow 0$ there exists a net $\left(h^i_\beta, h^{i*}_\beta\right) \to^{s \times w} \left(h^i, h^{i*}\right)$ (for $i = 1, 2$) such that $\left\| \left(h^i_\beta, h^{i*}_\beta\right) \right\| \leq K$ and

$$(z, z^*) + t_\beta \left(h^i_\beta, h^{i*}_\beta\right) = \left(z + t_\beta h^i_\beta, z^* + t_\beta h^{i*}_\beta\right) \in M$$

As M is monotone

$$\langle z + t_\beta h^1_\beta - \left(z + t_\beta h^2_\beta\right), z^* + t_\beta h^{1*}_\beta - \left(z^* + t_\beta h^{2*}_\beta\right) \rangle = t^2_\beta \langle h^1_\beta - h^2_\beta, h^{1*}_\beta - h^{2*}_\beta \rangle \geq 0 \quad \text{for all } \beta$$

Let $x_\beta := h^1_\beta - h^2_\beta$, $x := h^1 - h^2$ and $x^*_\beta := h^{1*}_\beta - h^{2*}_\beta$, $x^* := h^{1*} - h^{2*}$ then taking the strong limit in X and the bounded weak limit in X^* we have

$$\begin{aligned} &\left| \langle x_\beta, x^*_\beta \rangle - \langle x, x^* \rangle \right| \leq \left| \langle x_\beta - x, x^*_\beta \rangle \right| + \left| \langle x, x^*_\beta - x^* \rangle \right| \\ &\leq \left\| x^*_\beta \right\| \left\| x_\beta - x \right\| + \|x\| \left\| x^*_\beta - x^* \right\| \leq K \left(\|x_\beta - x\| + \left\| x^*_\beta - x^* \right\| \right) \to_\beta 0. \end{aligned}$$

Thus we obtain

$$\langle h^1 - h^2, h^{1*} - h^{2*}\rangle \geq 0 \quad \text{for all } \left(h^i, h^{i*}\right) \in T_M(z, z^*).$$

Now suppose $T_M(z, z^*)$ is not maximal monotone. Then there exists $(w, w^*) \notin T_M(z, z^*)$ such that (w, w^*) is monotonically related to $T_M(z, z^*)$. As $T_M(z, z^*)$ is closed there exists a $s \times w$ compact set $\mathcal{N}(w, w^*) = C \times K$ containing (w, w^*) such that $\mathcal{N}(w, w^*) \cap T_M(z, z^*) = \emptyset$ and a half-space $H(y, y^*) := \{(v, v^*) \mid \langle (y, y^*), (v, v^*)\rangle > 0\}$ such that $T_M(z, z^*) \cap H(y, y^*) = \emptyset$ and $\mathcal{N}(w, w^*) \subseteq H(y, y^*)$. Thus for all $(h, h^*) \in T_M(z, z^*)$

$$\begin{aligned}
\langle (w + \delta y) - h, (w^* + \delta y^*) - h^*\rangle &= \langle (w - h) + \delta y, (w^* - h^*) + \delta y^*\rangle \\
&= \langle w - h, w^* - h^*\rangle + \delta\left(\langle w, y^*\rangle + \langle y, w^*\rangle\right) \\
&\qquad - \delta\left(\langle h, y^*\rangle + \langle y, h^*\rangle\right) + \delta^2\langle y, y^*\rangle \\
&\geq \delta\{\langle (w, w^*), (y, y^*)\rangle + \\
&\qquad - \langle (h, h^*), (y, y^*)\rangle + \delta\langle y, y^*\rangle\} > 0 \qquad (11)
\end{aligned}$$

if $\delta > 0$ fixed but sufficiently small. Also $(w + \delta y, w^* + \delta y^*) \in H(y, y^*)$ because

$$\langle (y, y^*), (w + \delta y, w^* + \delta y)\rangle = \langle (y, y^*), (w, w^*)\rangle + \delta\langle (y, y^*), (y, y^*)\rangle > 0.$$

It follows that $(z + t(w + \delta y), z^* + t(w^* + \delta y^*)) \notin \operatorname{Graph} T = M$ for $t > 0$ small otherwise $(w + \delta y, w^* + \delta y) \in T_M(z, z^*)$ contradicting $T_M(z, z^*) \cap H(y, y^*) = \emptyset$.

Now use the norm to norm continuity of T to take $\eta > 0$ sufficiently small such that for $t = \eta$ and all $\|h\| \leq \eta$ with $z + th \in \operatorname{dom} T$ we have $\|h^*\| \leq 1$ for all $z^* + th^* \in T(z + th)$. Thus we have $z \in \operatorname{dom} T$ and for all $\|h\| \leq \eta$ with $z + th \in \operatorname{dom} T$ that $(h, h^*) \in \overline{B}_\eta(0) \times \overline{B}_1(0) \cap [\operatorname{Graph} T - (z, z^*)]/t$. Using (10) the norm to norm continuity of T at (z, z^*), (11) and the weak compactness of the unit ball, we have, for a possibly smaller value of η that

$$\begin{aligned}
&\langle (w + \delta y) - h, (w^* + \delta y^*) - h^*\rangle > 0 \\
&\text{for } t = \eta \text{ and all } \|h\| \leq \eta \text{ with } z + th \in \operatorname{dom} T \text{ and } z^* + th^* \in T(z + th) \\
&\text{or} \quad \langle z + t(w + \delta y) - x, z + t(w^* + \delta y^*) - x^*\rangle > 0 \qquad (12) \\
&\text{for all } x^* \in T(x) \text{ and } x \in B_{\eta^2}(z) \cap \operatorname{dom} T.
\end{aligned}$$

As all maximal monotone operators in reflexive spaces are maximal monotone locally and (12) implies $(z + t(w + \delta y), z + t(w^* + \delta y^*))$ is locally monotonically related to $\operatorname{Graph} T \cap \left(B_{\eta^2}(z) \times X^*\right)$ we must have $z + t(w^* + \delta y^*) \in T(z + t(w + \delta y))$, a contradiction. ■

Question: Under what other assumptions is $T_M(z, z^*)$ maximally monotone?

If $T_M(z, z^*)$ is monotone then we immediately have from $(0, 0) \in T_M(z, z^*)$ that

$$(\forall (h, h^*) \in T_M(z, z^*)) \quad \langle h, h^*\rangle \geq 0.$$

We will proceed with only the presumption that $T_M(z, z^*)$ is a monotone set.

Definition 17 *We say that $M \subseteq X \times X^*$ is Proto-differentiable at (z, z^*) if the following limit exists with respect to strong-topology on $X \times X^*$:*

$$T_M(z, z^*) := \lim_{t \downarrow 0} \frac{1}{t}\left(M - (z, z^*)\right).$$

One of the problems with dealing with convergence of Fitzpatrick functions is that of passing limits through conjugation operations applied to nonconvex functions. Namely to the function $\langle \cdot, \cdot\rangle + \delta_{\frac{1}{t}(\operatorname{Graph} T - (z, z^*))}(\cdot, \cdot)$. The only study applicable to this situation may be found in [15] and is applied in [16] to representative functions. These results only apply in reflexive space and also use stronger convergence notions than those discussed here. It is noted that a uniform coercivity condition is critical in obtaining these results.

We must also introduce some coercivity in our approximates. We will assume from here on that by a translation of the graph of T we have $(0,0) \in T$and $\langle x, x^* \rangle \geq 0$ for all $(x,x^*) \in T$. Denote

$$s_\lambda(x,x^*) := \langle x, x^* \rangle + \frac{\lambda}{2}\left(\|x\|^2 + \|x^*\|^2\right) \qquad \text{and}$$
$$c_T(x,x^*) := \langle x, x^* \rangle + \delta_M(x,x^*).$$

Under our standing assumption we have

$$s_\lambda(x,x^*) \geq \frac{\lambda}{2}\|(x,x^*)\|^2 \quad \text{for all } (x,x^*) \in \bar{T}_N$$
$$\text{and so} \quad q_T^\lambda(x,x^*) := (s_\lambda + \delta_T)^{**}(x,x^*) \geq \frac{\lambda}{2}\|(x,x^*)\|^2 \quad \text{for all } (x,x^*). \tag{13}$$

The following is a small variation of Lemma 2.2 of [16].

Proposition 18 *Let T be maximal monotone with $(0,0) \in T$ and let $q_T^\lambda := (s_\lambda + \delta_T)^{**}$. Then*

$$\delta_T + s_\lambda \geq q_T^\lambda \geq \mathcal{P}_T + \frac{\lambda}{2}\|\cdot\|^2 \geq \mathcal{F}_T + \frac{\lambda}{2}\|\cdot\|^2 \geq s_\lambda \tag{14}$$

and $\mathcal{F}_T = \left(q_T^\lambda - \frac{\lambda}{2}\|\cdot\|^2\right)^*$ with $\mathcal{P}_T = \left(q_T^\lambda - \frac{\lambda}{2}\|\cdot\|^2\right)^{**}$. In particular q_T^λ is coercive for all $\lambda > 0$ and $q_T^\lambda(0,0) = 0$.
Finally

$$T = \left\{(x,x^*) \mid q_T^\lambda(x,x^*) = s_\lambda(x,x^*)\right\}.$$

Proof. The inequalities (14) follows as in [16] which we reproduce for completeness. The first inequality of (10) is clear while the second is due to $\delta_T + s_\lambda \geq \mathcal{P}_T + \frac{\lambda}{2}\|\cdot\|^2$ with $\mathcal{P}_T + \frac{\lambda}{2}\|\cdot\|^2 \in \Gamma_{x\times \mathrm{bw}^*}(X \times X^*)$. The other two use $\mathcal{P}_T \geq \mathcal{F}_T \geq c_T$. The first inequality in (14) implies $\left(q_T^\lambda - \frac{\lambda}{2}\|\cdot\|^2\right)^* \leq c_T^* = \mathcal{F}_T^\dagger$ with equality ensuing. As $\mathcal{P}_T = (\mathcal{F}_T^*)^\dagger$ we have $\mathcal{P}_T = \left(q_T^\lambda - \frac{\lambda}{2}\|\cdot\|^2\right)^{**}$. The coercivity follows from (13) along with $q_T^\lambda(0,0) \geq 0$. As $(s_\lambda + \delta_T)(0,0) = 0$ and q_T^λ is the largest convex function dominated by $s_\lambda + \delta_T$ we have $q_T^\lambda(0,0) \leq 0$.
Suppose (x,x^*) is such that $q_T^\lambda(x,x^*) = s_\lambda(x,x^*)$ then $q_T^\lambda(x,x^*) - \frac{\lambda}{2}\|(x,x^*)\|^2 = \langle x, x^* \rangle \geq \mathcal{P}_T(x,x^*)$ but as $\mathcal{P}_T$ is always representative $\langle x, x^* \rangle = \mathcal{P}_T(x,x^*)$ implying $(x,x^*) \in \operatorname{Graph} T$. Conversely when $(x,x^*) \in M$ one has $s_\lambda(x,x^*) + \delta_{\operatorname{Graph} T}(x,x^*) = s_\lambda(x,x^*)$ and so the inequalities (14) are equalities.
■

The function q_T^λ is more easily studied from a variational view point so we must relate this back to the Fitzpatrick function.

Corollary 19 *Let T be maximal monotone with $(0,0) \in T$ and let $q_T^\lambda := (s_\lambda + \delta_T)^{**} : X \times X^* \to \overline{\mathbf{R}}$.*

1. *For all $\lambda > 0$*

$$\left(\mathcal{F}_T \hat{\square} \frac{1}{2\lambda}\|\cdot\|^2\right)(x,x^*) := \inf\left\{\mathcal{F}_T(v,v^*) + \frac{1}{2\lambda}\|(x,x^*) - (v,v^*)\|^2\right\}$$
$$= \left(q_T^\lambda\right)^*(x^*,x) \geq \left((\delta_M + \langle \cdot,\cdot \rangle) + \frac{\lambda}{2}\|\cdot\|^2\right)^*(x^*,x).$$

2. *For all $\lambda > 0$*

$$\mathcal{F}_T(x,x^*) = \left(q_T^\lambda - \frac{\lambda}{2}\|\cdot\|^2\right)^*(x,x^*)$$
$$= \left(\left(q_T^\lambda\right)^* \check{\square} \frac{1}{2\lambda}\|\cdot\|^2\right)(x,x^*) := \sup_{(v^*,v^{**})}\left\{\left(q_T^\lambda\right)^*((x,x^*) + (v,v^*)) - \frac{1}{2\lambda}\|(v,v^*)\|^2\right\}. \tag{15}$$

Proof. By (10) we have

$$\left(\mathcal{P}_T + \frac{\lambda}{2}\left\|\cdot\right\|^2\right)^* \geq \left(q_T^\lambda\right)^* \geq \left(\delta_T + s_\lambda\right)^*$$
$$= \left(\left(\delta_T + \langle\cdot,\cdot\rangle\right) + \frac{\lambda}{2}\left\|\cdot\right\|^2\right)^*.$$

Since $\frac{\lambda}{2}\left\|\cdot\right\|^2$ is a finite continuous convex function with domain $X \times X^*$ we have (see page 253 of [2])

$$\left(q_T^\lambda\right)^* \leq \left(\mathcal{P}_T + \frac{\lambda}{2}\left\|\cdot\right\|^2\right)^* = \mathcal{P}_T^* \hat{\square} \left(\frac{\lambda}{2}\left\|\cdot\right\|^2\right)^* = \left(\mathcal{F}_T \hat{\square} \frac{1}{2\lambda}\left\|\cdot\right\|^2\right)$$

as $\left(\frac{\lambda}{2}\left\|\cdot\right\|^2\right)^* = \frac{1}{2\lambda}\left\|\cdot\right\|^2$.

The first and second inequalities in (10) imply

$$\left(\delta_M + \langle\cdot,\cdot\rangle\right)^* = \mathcal{F}_T \leq \left(q_T^\lambda - \frac{\lambda}{2}\left\|\cdot\right\|^2\right)^* \leq \mathcal{P}_T^* = \mathcal{F}_T.$$

Then (15) is implied by the Toland–Singer formula [22] for the dual of a difference of functions i.e.

$$\left(q_T^\lambda - \frac{\lambda}{2}\left\|\cdot\right\|^2\right)^* = \left(q_T^\lambda\right)^* \check{\square} \frac{1}{2\lambda}\left\|\cdot\right\|^2.$$

Finally consider

$$\left(\mathcal{F}_T \hat{\square} \frac{1}{2\lambda}\left\|\cdot\right\|^2\right)(x,x^*) = \left\{\left(\left(q_T^\lambda\right)^* \check{\square} \frac{1}{2\lambda}\left\|\cdot\right\|^2\right) \hat{\square} \frac{1}{2\lambda}\left\|\cdot\right\|^2\right\}(x,x^*)$$
$$= \inf_{(v,v^*)}\left\{\sup_{(w,w^*)}\left\{\left(q_T^\lambda\right)^*\left((x,x^*) - (v,v^*) + (w,w^*)\right) - \frac{1}{2\lambda}\left\|(w,w^*)\right\|^2\right\} + \frac{1}{2\lambda}\left\|(v,v^*)\right\|^2\right\}$$
$$\leq \sup_{(w,w^*)}\left(\left\{\left(q_T^\lambda\right)^*\left((x,x^*) - (w,w^*) + (w,w^*)\right) - \frac{1}{2\lambda}\left\|(w,w^*)\right\|^2\right\} + \frac{1}{2\lambda}\left\|(w,w^*)\right\|^2\right)$$
$$= \left(q_T^\lambda\right)^*(x,x^*).$$

■

The study in [16] uses a stronger form of convergence, namely the epi–distance or Attouch–Wetts topology. We would rather stay with Mosco–convergence but currently the literature lacks the machinery to achieve this end. Thus, following [16] we say a net of sets $\{M_\alpha\}$ boundedly converges to M iff both

$$M \subseteq b\text{-}\liminf_\alpha M_\alpha \qquad \text{or equivalently}$$
$$\forall \{(x_\alpha, x_\alpha^*)\} \subseteq M \text{ uniformly bounded we have } d\left((x_\alpha, x_\alpha^*), M_\alpha\right) \to_\alpha 0$$

and

$$M \supseteq b\text{-}\limsup_\alpha M_\alpha \qquad \text{or equivalently}$$
$$\forall (x_\alpha, x_\alpha^*) \in M_\alpha \text{ such that } \{(x_\alpha, x_\alpha^*)\} \text{ uniformly bounded we have } d\left((x_\alpha, x_\alpha^*), M\right) \to_\alpha 0.$$

We write $T_\alpha \to^b T$ to mean that $\operatorname{Graph} T_\alpha \to^b \operatorname{Graph} T$. It is clear that $T_\alpha \to^b T$ implies $T = m\text{-}\lim_\alpha T_\alpha$.

The following requires very minor change to Proposition 3.1 of [16] and follows by the same proof method.

Proposition 20 ([16]) *Suppose* $\{T_n\}$ *is a sequence of maximal monotone operators with graphs* $M_n := \operatorname{Graph} T_n$ *such that* $(0,0) \in M_n$ *for all* n. *Then the following hold:*

1. $M \supseteq b\text{-}\limsup_n M_n \iff q_M^\lambda \leq b\text{-}\liminf_n q_{M_n}^\lambda$ *for any* $\lambda > 0$.
2. $M \subseteq b\text{-}\liminf_n M_n \iff q_M^\lambda \geq b\text{-}\limsup_m q_{M_n}^\lambda$ *for any* $\lambda > 0$.
3. $M_n \to^b M \iff q_M^\lambda = b\text{-}\lim_n q_{M_n}^\lambda$ *for any* $\lambda > 0$.

We will denote

$$\begin{aligned} b\text{-}\limsup_n \operatorname{epi} f_n = \operatorname{epi} f \quad &\text{by} \quad e\text{-}b\text{-}\liminf_n f_n = f \\ \text{and } b\text{-}\liminf_n \operatorname{epi} f_n = \operatorname{epi} f \quad &\text{by} \quad e\text{-}b\text{-}\limsup_n f_n = f. \end{aligned}$$

Also when $f \leq e\text{-}b\text{-}\liminf_n f_n \leq e\text{-}b\text{-}\limsup_n f_n \leq f$ we write

$$b\text{-}e\text{-}\lim_n f_n = f$$

Definition 21 *We say that* $M \subseteq X \times X^*$ *is strongly Proto-differentiable at* (z, z^*) *if the following limit exists:*

$$T_M(z, z^*) := b\text{-}\lim_{t \downarrow 0} \frac{1}{t}(M - (z, z^*)).$$

We now relate $\mathcal{F}_{\mathcal{B}}(z, z^*)$ back to the more fundamental Fitzpatrick function $\mathcal{F}_M(z, z^*)$ and its conjugate $\mathcal{F}_A^*(z^*, z)^\dagger = \mathcal{P}_A(y, y^*)$.

Theorem 22 *Suppose* $T : X \rightrightarrows X^*$ *is maximal monotone,* $M := \operatorname{Graph} T$ *and* $(z, z^*) \in M$. *Suppose in addition that* M *is strongly Proto-differentiable at* (z, z^*) *then for all* $(y, y^*) \in X \times X^*$ *we have*

$$\begin{aligned} \mathcal{F}_{T_M(z,z^*)}(y, y^*) &= b\text{-}e\text{-}\lim_{t \downarrow 0} \frac{1}{t^2}\left\{\mathcal{F}_T((z, z^*) + t(y, y^*)) - \mathcal{F}_T(z, z^*) - t\langle (y, y^*), (z, z^*)\rangle\right\} \qquad (16) \\ &:= \mathcal{F}_T''(z, z^*)(y, y^*). \end{aligned}$$

That is, $\mathcal{F}_{T_M(z,z^*)}(y, y^*)$ *corresponds to the second order epi–derivative of* $\mathcal{F}_M$ *at* (z, z^*) *in the direction* (y, y^*). *It also follows that*

$$\begin{aligned} \mathcal{P}_{T_M(z,z^*)}(y, y^*) &= b\text{-}e\text{-}\lim_{t \downarrow 0} \frac{1}{t^2}\left\{\mathcal{P}_T((z, z^*) + t(y, y^*)) - \mathcal{P}_T(z, z^*) - t\langle (y, y^*), (z, z^*)\rangle\right\} \qquad (17) \\ &:= \mathcal{P}_T''(z, z^*)(y, y^*). \end{aligned}$$

Proof. We use the bi–continuity of the Fenchel conjugate with respect to epi–convergence and the Toland-Singer duality formula for the difference of convex functions. Note that $(0,0) \in \frac{1}{t}(M - (z, z^*))$ for all $t > 0$ when $(z, z^*) \in M$. Also $\left\{\frac{1}{t}(M - (z, z^*))\right\}_{t>0}$ is a parametrised set of maximal monotone operators (inheriting its maximality from that of M). We assume that

$$T_M(z, z^*) := b\text{-}\lim_{t \downarrow 0} \frac{1}{t}(M - (z, z^*))$$

and so

$$\delta_{T_M(z,z^*)}(y, y^*) = b\text{-}e\text{-}\lim_{t \downarrow 0} \delta_{\frac{1}{t}(M-(z,z^*))}(y, y^*).$$

By [2] Theorem 7.1.5 we have for all $\lambda > 0$

$$s_\lambda(\cdot,\cdot) + b\text{-}e\text{-}\lim_{t \downarrow 0} \delta_{\frac{1}{t}(M-(z,z^*))}(\cdot,\cdot) = b\text{-}e\text{-}\lim_{t \downarrow 0}\left(s_\lambda(\cdot,\cdot) + \delta_{\frac{1}{t}(M-(z,z^*))}(\cdot,\cdot)\right).$$

On applying Proposition 20 and Corollary 19 we have for any $\lambda > 0$

$$\begin{aligned}\left(q^{\lambda}_{T_M(z,z^*)}\right)^*(y^*,y) &= \left(s_\lambda(\cdot,\cdot) + b\text{-}e\text{-}\lim_{t\downarrow 0}\delta_{\frac{1}{t}(M-(z,z^*))}(\cdot,\cdot)\right)^*(y^*,y) \\ &= b\text{-}e\text{-}\lim_{t\downarrow 0}\left(s_\lambda(\cdot,\cdot) + \delta_{\frac{1}{t}(M-(z,z^*))}(\cdot,\cdot)\right)^*(y^*,y). \\ &= b\text{-}e\text{-}\lim_{t\downarrow 0}\left(q^{\lambda}_{\frac{1}{t}(M-(z,z^*))}\right)^*(y^*,y).\end{aligned}$$

Applying the second part of Corollary 19 we have

$$\begin{aligned}\mathcal{F}_{T_M(z,z^*)}(x,x^*) &= \left(\left(q^{\lambda}_{T_M(z,z^*)}\right)^*\check{\square}\frac{1}{2\lambda}\|\cdot\|^2\right)(x,x^*) \\ &= \left(b\text{-}e\text{-}\lim_{t\downarrow 0}\left(q^{\lambda}_{\frac{1}{t}(M-(z,z^*))}\right)^*\check{\square}\frac{1}{2\lambda}\|\cdot\|^2\right)(x,x^*)\end{aligned}$$

and on applying [2] Theorem 7.3.8 and Corollary 19 again it follows that

$$\begin{aligned}\mathcal{F}_{T_M(z,z^*)}(x,x^*) &= b\text{-}e\text{-}\lim_{t\downarrow 0}\left(\left(q^{\lambda}_{\frac{1}{t}(M-(z,z^*))}\right)^*\check{\square}\frac{1}{2\lambda}\|\cdot\|^2\right)(x,x^*) \\ &= b\text{-}e\text{-}\lim_{t\downarrow 0}\mathcal{F}_{\frac{1}{t}(M-(z,z^*))}(x,x^*). \end{aligned} \tag{18}$$

Now consider $\delta_{\frac{1}{t}(M-(z,z^*))}(w,w^*) = \delta_M((z,z^*)+t(w,w^*))$. Then placing $(v,v^*) = (z+tw, z^*+tw^*)$

$$\begin{aligned}\langle w,w^*\rangle + \delta_{\frac{1}{t}(M-(z,z^*))}(w,w^*) &= \langle\frac{1}{t}(v-z), \frac{1}{t}(v^*-z^*)\rangle + \delta_M(v,v^*) \\ &= \frac{1}{t^2}\left\{\delta_M(v,v^*) + \langle v,v^*\rangle - (\langle z,v^*\rangle + \langle v,z^*\rangle - \langle z,z^*\rangle)\right\}.\end{aligned}$$

Consequently

$$\begin{aligned}&\left(\langle\cdot,\cdot\rangle + \delta_{\frac{1}{t}(M-(z,z^*))}(\cdot,\cdot)\right)^*(y^*,y) \\ =\ &\sup_{(w,w^*)}\left\{\langle(w,w^*),(y,y^*)\rangle - \left[\langle w,w^*\rangle + \delta_{\frac{1}{t}(M-(z,z^*))}(w,w^*)\right]\right\} \\ =\ &\sup_{(v,v^*)}\{\langle\left(\frac{1}{t}(v-z), \frac{1}{t}(v^*-z^*)\right),(y,y^*)\rangle \\ &\qquad - \left[\frac{1}{t^2}\{\delta_M(v,v^*) + \langle v,v^*\rangle - (\langle z,v^*\rangle + \langle v,z^*\rangle - \langle z,z^*\rangle)\}\right]\} \\ =\ &\frac{1}{t^2}\sup_{(v,v^*)}\{t\langle v-z,y^*\rangle + t\langle y,v^*-z^*\rangle \\ &\qquad - [\delta_M(v,v^*) + \langle v,v^*\rangle - (\langle z,v^*\rangle + \langle v,z^*\rangle - \langle z,z^*\rangle)]\} \\ =\ &\frac{1}{t^2}(\sup_{(v,v^*)}\{\langle(v,v^*),t(y,y^*)\rangle - [\delta_M(v,v^*) + \langle v,v^*\rangle - (\langle z,v^*\rangle + \langle v,z^*\rangle - \langle z,z^*\rangle)]\} \\ &\qquad\qquad - \langle t(y,y^*),(z,z^*)\rangle) \\ =\ &\frac{1}{t^2}\left\{(\delta_M(\cdot,\cdot) + \langle\cdot,\cdot\rangle - (\langle z,\cdot\rangle + \langle\cdot,z^*\rangle - \langle z,z^*\rangle))^*(t(y,y^*)) - t\langle(y,y^*),(z,z^*)\rangle\right\}.\end{aligned}$$

Next apply the Toland-Singer duality formula [22] to the conjugate of the difference of the two functions g and h where $g(x,x^*) := \delta_M(x,x^*) + \langle x,x^*\rangle$ and $h(x,x^*) := \langle z,x^*\rangle + \langle x,z^*\rangle - \langle z,z^*\rangle$. Thence

$$(g-h)^*(t(y^*,y)) = \sup_{(v,v^*)}\{g^*(t(y^*,y) + (v^*,v)) - h^*(v^*,v)\}.$$

Note that this formula does not require g to be convex, only h to be proper and convex. By direct calculation $h^*(v,v^*) = \langle v^*, v\rangle + \delta_{(z^*,z)}(v^*,v)$ and clearly $g^*(t(y^*,y)+(v^*,v)) = \mathcal{F}_T((v^*,v)+t(y^*,y))$ thus

$$
\begin{aligned}
&\mathcal{F}_{\frac{1}{t}(M-(z,z^*))}(y,y^*) = \left(\langle\cdot,\cdot\rangle + \delta_{\frac{1}{t}(M-(z,z^*))}(\cdot,\cdot)\right)^*(y^*,y)\\
&= \frac{1}{t^2}\left(\sup_{(v,v^*)}\left\{\mathcal{F}_T((v,v^*)+t(y,y^*)) - \left\{\langle v,v^*\rangle + \delta_{(z,z^*)}(v,v^*)\right\}\right\} - t\langle (y,y^*),(z,z^*)\rangle\right)\\
&= \frac{1}{t^2}\left(\mathcal{F}_T((z,z^*)+t(y^*,y)) - \langle z,z^*\rangle - t\langle (y,y^*),(z,z^*)\rangle\right)\\
&= \frac{1}{t^2}\left(\mathcal{F}_T((z,z^*)+t(y,y^*)) - \mathcal{F}_T(z,z^*) - t\langle (y,y^*),(z,z^*)\rangle\right),
\end{aligned}
$$

using the fact that $(z,z^*)\in M$ and so $\mathcal{F}_T(z,z^*) = \langle z,z^*\rangle$. Applying (18) gives the result (16).
The second observation follows from the well known fact (see [14] and reference contained therein) that for

$$\left(\Delta_t^2\mathcal{F}_T\right)_{(z,z^*)}(y,y^*) := \frac{1}{t^2}\left(\mathcal{F}_T((z,z^*)+t(y,y^*)) - \mathcal{F}_T(z,z^*) - t\langle (y,y^*),(z,z^*)\rangle\right)$$

we have

$$
\begin{aligned}
\left[\Delta_t^2\mathcal{F}_T(z,z^*)\right]^*(y^*,y) &= \left(\Delta_t^2\mathcal{F}_T^*(z,z^*)\right)(y^*,y)\\
&= \left(\Delta_t^2\mathcal{P}_T(z,z^*)\right)(y,y^*).
\end{aligned}
$$

Applying the bi-continuity of the Fenchel conjugate we obtain

$$\mathcal{F}^*_{T_M(z,z^*)}(y,y^*)^\dagger = b\text{-}e\text{-}\lim_{t\downarrow 0}\left(\Delta_t^2\mathcal{P}_T\right)_{(z,z^*)}(y,y^*).$$

■

We are now able to use the formula relating Proto-derivative of a subgradient and the subdifferential of an second order epi-derivative, see [14].

Proposition 23 *Suppose $T : X \rightrightarrows X^*$ is maximal monotone, $M := \operatorname{Graph} T$ and $(z,z^*)\in M$. Suppose in addition that M is strongly Proto-differentiable at (z,z^*). Then*

$$\partial\mathcal{F}_{T_M(z,z^*)}(y,y^*) = \lim_{t\downarrow 0}\frac{1}{t}\left\{\partial\mathcal{F}_T((z,z^*)+t(y,y^*)) - (z,z^*)\right\}. \tag{19}$$

Proof. We note that $\partial\mathcal{F}_{T_M(z,z^*)}(y,y^*) = \partial\left(\frac{1}{2}\mathcal{F}''_T(z,z^*)\right)(y,y^*)$ corresponds to the subdifferential of a second order epi–derivative and the right hand side of (19) corresponds to a strong Proto-derivative of the subdifferential. Now recall that the bounded strong epi-limit in (16) implies the existence of and coincidence with the corresponding Mosco epi-limit. Using Theorem 3.9 of [14] we deduce that

$$\partial\left(\frac{1}{2}\mathcal{F}''_T(z,z^*)\right)(y,y^*) = \lim_{t\downarrow 0}\frac{1}{t}\left\{\partial\mathcal{F}_T((z,z^*)+t(y,y^*)) - (z,z^*)\right\}$$

where the limit on the right corresponds to a Kuratowski–Painlevé limit with respect to the strong topology. ■

4 Single-valuedness of T and Proto–Differentiability

Can we succeed in reducing the question of the differentiability of a monotone mapping to that of a convex function? For the remainder of this section let $M := \operatorname{Graph} T$ where $T : X \rightrightarrows X^*$ is maximal monotone. Define another monotone set as

$$\mathcal{B}(z,z^*) := T_M(z,z^*) + (z,z^*).$$

Note that we now have $(z, z^*) \in \mathcal{B}(z, z^*)$. When no confusion occurs we will suppress reference to (z, z^*) i.e. $\mathcal{B}(z, z^*) \equiv \mathcal{B}$. We note that as

$$\mathcal{F}_{T+(z,z^*)}(y, y^*) = \mathcal{F}_T(y - z, y^* - z^*) - (\langle y - z, y^* - z^*\rangle - \langle y, y^*\rangle)$$

results stated in terms of $\mathcal{B}(z, z^*)$ may be restated in terms of $T_M(z, z^*)$. For example,

$$\partial\mathcal{F}_{\mathcal{B}(z,z^*)}(y, y^*) = \partial\mathcal{F}_{T_M(z,z^*)}(y - z, y^* - z^*) + (z^*, z). \tag{20}$$

Example 24 *A simple instructive example is $f(x) := |x|$. Here it is easily shown that $\mathcal{F}_T(z, z^*) = |z|$ for $T = \partial f$ and $\mathcal{P}_T(z, z^*) = |z| + \delta_{\mathbf{R}\times[-1,1]}(z, z^*)$. Then for $(0,0) \in \partial f(0,0)$ we have $\mathcal{F}_{T_M(0,0)}(y, y^*) = \delta_{\mathbf{R}\times\{0\}}(y, y^*)$ and $\mathcal{P}_{T_M(0,0)}(y, y^*) = \delta_{\{0\}\times\mathbf{R}}(y, y^*)$. Thus differentiability is not assured.*

Recall

$$\mathcal{M}_{\mathcal{B}}(z, z^*) := \{(a, a^*) \in X \times X^* \mid \mathcal{F}_{\mathcal{B}}(z, z^*) = \langle z, a^*\rangle + \langle a, z^*\rangle - \langle a, a^*\rangle\}.$$

Note that $(z, z^*) \in \mathcal{M}_{\mathcal{B}}(z, z^*)$ since $\mathcal{F}_{\mathcal{B}}(z, z^*) = \langle z, z^*\rangle$ as $(z, z^*) \in \mathcal{B}$ which is a monotone set.

Lemma 25 *Suppose $T : X \rightrightarrows X^*$ is maximal monotone, $M := \operatorname{Graph} T$ and $(z, z^*) \in M$. Then*

$$\begin{aligned}\partial\mathcal{F}_{\mathcal{B}}(z, z^*) \cap \mathcal{B} &= \left\{(T_M(z, z^*))^\circ + (z, z^*)\right\} \cap \mathcal{B}^\dagger \\ &= \left\{(T_M(z, z^*))^\circ + (z^*, z)\right\} \cap \left\{T_M(z, z^*)^\dagger + (z^*, z)\right\} \\ &= \mathcal{M}_{\mathcal{B}}(z, z^*) \cap \mathcal{B}.\end{aligned} \tag{21}$$

If in addition $\mathcal{B}$ is maximal monotone (which is the case when $\nabla\mathcal{F}_T$ exists as a Fréchet derivative at (z, z^*)) then

$$(T_M(z, z^*))^\circ \subseteq \mathcal{M}_{\mathcal{B}}(z, z^*) - (z^*, z). \tag{22}$$

Proof. Suppose $(a, a^*) - (z, z^*) \in (T_M(z, z^*))^\circ$ with $(z, z^*) \in M$ and so $(z, z^*) \in \mathcal{B}(z, z^*) := \mathcal{B}$. Then for all $(y, y^*) \in \mathcal{B} = T_M(z, z^*) + (z, z^*)$

$$\langle (a, a^*) - (z, z^*), (y, y^*) - (z, z^*)\rangle \leq 0.$$

That is

$$\begin{aligned}\langle a^*, y\rangle + \langle a, y^*\rangle - \langle y, y^*\rangle - (\langle z^*, y\rangle + \langle z, y^*\rangle - \langle z, z^*\rangle) + \langle y, y^*\rangle - \langle a, a^*\rangle \\ \leq \langle a^*, z\rangle + \langle a, z^*\rangle - \langle a, a^*\rangle - \langle z, z^*\rangle\end{aligned}$$

holds for all $(y, y^*) \in \mathcal{B}$. Thus, for all $(y, y^*) \in \mathcal{B}$ we have (because $(z, z^*) \in \mathcal{B}$)

$$\begin{aligned}\{\langle a^*, y\rangle + \langle a, y^*\rangle - \langle y, y^*\rangle - \langle a, a^*\rangle\} - \{\mathcal{F}_{\mathcal{B}}(y, y^*) - \langle y, y^*\rangle\} \\ \leq \{\langle a^*, z\rangle + \langle a, z^*\rangle - \langle a, a^*\rangle\} - \langle z, z^*\rangle\end{aligned}$$

As $(y, y^*) \in \mathcal{B}$ and $\mathcal{B}$ is monotone we have $\mathcal{F}_{\mathcal{B}}(y, y^*) = \langle y, y^*\rangle$. Taking the supremum over $(y, y^*) \in \mathcal{B}$

$$\mathcal{F}_{\mathcal{B}}(a, a^*) - \langle a, a^*\rangle \leq \{\langle a^*, z\rangle + \langle a, z^*\rangle - \langle a, a^*\rangle\} - \langle z, z^*\rangle. \tag{23}$$

When $(a, a^*) \in \mathcal{B}$ we have $\mathcal{F}_{\mathcal{B}}(a, a^*) = \langle a, a^*\rangle$ and so

$$\mathcal{F}_{\mathcal{B}}(z, z^*) = \langle z, z^*\rangle \leq \langle a^*, z\rangle + \langle a, z^*\rangle - \langle a, a^*\rangle$$

implying $(a, a^*) \in \mathcal{M}_{\mathcal{B}}(z, z^*) \cap \mathcal{B}$. That yeilds by Proposition 5

$$\left\{(T_M(z, z^*))^\circ + (z, z^*)\right\} \cap \mathcal{B} \subseteq \mathcal{M}_{\mathcal{B}}(z, z^*) \cap \mathcal{B} = \partial\mathcal{F}_{\mathcal{B}}(z, z^*) \cap \mathcal{B} \tag{24}$$

and on combining this with (25) we obtain (21).

When we assume $\mathcal{B}$ is maximal monotone then we may relax the assumption that $(a, a^*) \in \mathcal{B}$. To establish that $(a, a^*) \in \mathcal{M}_{\mathcal{B}}(z, z^*)$ we need to establish that $\mathcal{F}_{\mathcal{B}}(a, a^*) - \langle a, a^*\rangle \geq 0$ in (23). But this is always true for all (a, a^*) when $\mathcal{B}$ is maximal monotone and hence (22) holds in this case. ■

In the final section we will extend to this analysis in order to characterise $(T_M(z, z^*))^\circ$. We next show that Gateau differentiability of $\mathcal{F}_{\mathcal{B}}$ occurs at all $(z, z^*) \in M$.

Lemma 26 *Suppose $T : X \rightrightarrows X^*$ is maximal monotone and $(z, z^*) \in M$. Then*

$$\partial \mathcal{F}_{\mathcal{B}}(z, z^*) - (z, z^*) \subseteq (T_M(z, z^*))^{\circ} \cap \left(\overline{\text{co}}\mathcal{B}^{\dagger} - (z^*, z)\right) \qquad (25)$$
$$= (\overline{\text{co}} T_M(z, z^*))^{\circ} \cap \overline{\text{co}} T_M(z, z^*)^{\dagger}.$$

In particular $\partial \mathcal{F}_{\mathcal{B}}(z, z^*) = \{(z, z^*)\}$ when $(\overline{\text{co}} T_M(z, z^*))^{\circ} \cap \overline{\text{co}} T_M(z, z^*)^{\dagger} = \{0\}$.

Proof. Assuming $(a^*, a) \in \partial \mathcal{F}_{\mathcal{B}}(z, z^*)$ then

$$\langle (a, a^*), (y, y^*) - (z, z^*) \rangle \leq \mathcal{F}_{\mathcal{B}}(y, y^*) - \mathcal{F}_{\mathcal{B}}(z, z^*) \qquad (26)$$

and placing $(y, y^*) = (z, z^*) + t(h, h^*) \in \mathcal{B}$ for arbitrary member $(h, h^*) \in T_M(z, z^*)$ and $t > 0$ we obtain from (26) that

$$t\langle (a, a^*), (h, h^*) \rangle \leq \langle z + t^* h, z^* + t h^* \rangle - \langle z, z^* \rangle$$
$$= t(\langle h, z \rangle + \langle h^*, z \rangle) + t^2 \langle h, h^* \rangle.$$

Letting $t \downarrow 0$ we obtain

$$\langle (a, a^*), (h, h^*) \rangle \leq \langle (z, z^*), (h, h^*) \rangle$$
$$\text{or} \quad \langle (a, a^*) - (z, z^*), (h, h^*) \rangle \leq 0 \quad \text{for all } (h, h^*) \in T_M(z, z^*),$$

establishing (25) when we take Proposition 5 into account. ■

Example 27 *For the simple example with $f(x) := |x|$ and $M = \operatorname{Graph} \partial f$ at $(0, 0) \in M$ we have $(\overline{\text{co}} T_M(z, z^*))^{\circ} \cap \overline{\text{co}} T_M(z, z^*)^{\dagger} = \mathbf{R} \times \{0\}$ with $\partial \mathcal{P}_{T_M(0,0)}(0, 0) = \mathbf{R} \times \{0\}$ and $\partial \mathcal{F}_{T_M(0,0)}(0, 0) = \{0\} \times \mathbf{R}$. In contrast suppose that $f : \mathbf{R} \to \mathbf{R}$ with $M = \operatorname{Graph} \partial f$ and $T_M(z, z^*)$ is a tangent line of the form $\{(y, y^*) \mid y^* = cy\}$ with $c > 0$. Then a simple calculation gives $\mathcal{F}_{T_M(0,0)}(y, y^*) = \frac{1}{4c}(cy + y^*)^2$ which is clearly Fréchet differentiable.*

Of course in finite dimensions $\partial \mathcal{F}_{\mathcal{B}}(z, z^*) = \{(z, z^*)\}$ is enough to ensure Fréchet differentiability. If we are able to assure Fréchet differentiability of $\mathcal{F}_{\mathcal{B}}$ at $(z, z^*) \in M$ then a strong form of single valuedness holds for T at z when M is strongly Proto-differentiable at (z, z^*). In the next section we supply conditions that ensure the Fréchet differentiability of $\mathcal{F}_{\mathcal{B}}$.

Theorem 28 *Suppose $T : X \rightrightarrows X^*$ is maximal monotone, $M := \operatorname{Graph} T$ and $(z, z^*) \in M$. Suppose in addition that M is strongly Proto-differentiable at (z, z^*) and $\mathcal{F}_{\mathcal{B}}(z, z^*)$ is Fréchet differentiable at (z, z^*). Then $T(z) = \{z^*\}$ with $d^*(z^*, T(y)) = O(\|y - z\|)$ and in addition $T_M(z, z^*)$ is maximal monotone.*

Proof. From Lemma 26 we know that $\nabla \mathcal{F}_{\mathcal{B}}(z, z^*) = (z, z^*)$ and hence by (19) it follows that for all $(y, y^*) \in B_{\delta}(0)$ we have

$$b\text{-}\lim_{t \downarrow 0} \frac{1}{t} \{\partial \mathcal{F}_T((z, z^*) + t(y, y^*)) - (z, z^*)\} = \partial \mathcal{F}_{T_M(z,z^*)}(y, y^*). \qquad (27)$$

Using (20) we have

$$\partial \mathcal{F}_{T_M(z,z^*)}(y, y^*) = \partial \mathcal{F}_{\mathcal{B}(z,z^*)}(y + z, y^* + z^*) - (z, z^*)$$

and so by (27) and the Fréchet differentiability of $\mathcal{F}_{\mathcal{B}}$ at (z, z^*) we have for $\delta > 0$ small

$$\partial \mathcal{F}_T((z, z^*) + t(y, y^*)) - (z, z^*) \subseteq t\left[\partial \mathcal{F}_{\mathcal{B}(z,z^*)}(y + z, y^* + z^*) - (z, z^*)\right] + o(t) B_1(0).$$

As $\mathcal{F}_{\mathcal{B}}$ is Fréchet differentiable at (z, z^*) we have norm to norm upper semi–continuity of $\partial \mathcal{F}_{\mathcal{B}}$ at (z, z^*). Thus

$$d^*\left((z, z^*), \partial \mathcal{F}_{\mathcal{B}(z,z^*)}(y + z, y^* + z^*)\right) = O(\|(y, y^*)\|)$$

and so

$$d^*((z, z^*), \partial \mathcal{F}_T((z, z^*) + t(y, y^*))) = tO(\|(y, y^*)\|) + o(t) = O(\|t(y, y^*)\|).$$

On applying Theorem 7 we deduce that $d^*(z^*, T(y)) = O(\|y - z\|)$, implying $T(z) = z^*$. Now we apply Proposition 16. ■

Remark 29 *In finite dimensions we have the rather strong conclusion that Proto-differentiability of* $M = \operatorname{Graph} T$ *at* $(z, z^*) \in M$ *implies that* $d^*(z^*, T(y)) = O(\|y - z\|)$ *and that* $T_M(z, z^*)$ *is maximal monotone.*

5 Fréchet Differentiability $\mathcal{F}_{\mathcal{B}}$

We now address the issue of when $\mathcal{F}_{\mathcal{B}}(z, z^*)$ is Fréchet differentiable. This requirement appears in some of the previous results and is not immediate in infinite dimensions. We note that from Lemma 26 we have Fréchet differentiability of $\mathcal{F}_{\mathcal{B}}$ at every $(z, z^*) \in M$ when X is finite dimensional. The following homogeneity property is useful.

Lemma 30 *Suppose* $T : X \rightrightarrows X^*$ *is maximal monotone,* $M := \operatorname{Graph} T$ *and* $(z, z^*) \in M$*. Then for all* $(h, h^*) \in X \times X^*$ *we have*

$$\mathcal{F}_{T_M(z,z^*)}(t(h, h^*)) = t^2 \mathcal{F}_{T_M(z,z^*)}(h, h^*) \tag{28}$$

$$\text{and} \quad \mathcal{P}_{T_M(z,z^*)}(t(h, h^*)) = t^2 \mathcal{P}_{T_M(z,z^*)}(h, h^*) \tag{29}$$

Proof. We use the fact that T_M is a cone. Let ε, $t > 0$ and $(x_i, x_i^*) \in T_M(z, z^*)$, $\lambda_i \geq 0$ for $i = 1, \ldots, N$ satisfy $\sum_{i=1}^N \lambda_i (x_i, x_i^*, 1) = (h, h^*, 1)$ and

$$p_{T_M(z,z^*)}(h, h^*) + \frac{\varepsilon}{t^2} \geq \sum_{i=1}^N \lambda_i \langle x_i, x_i^* \rangle.$$

Then it follows that $(tx_i, tx_i^*) \in T_M(z, z^*)$ and $\sum_{i=1}^N \lambda_i (tx_i, tx_i^*) = t(h, h^*)$. Consequently

$$\sum_{i=1}^N \lambda_i \langle tx_i, tx_i^* \rangle \geq p_{T_M(z,z^*)}(th, th^*).$$

Hence

$$t^2 p_{T_M(z,z^*)}(h, h^*) + \varepsilon \geq \sum_{i=1}^N \lambda_i \langle tx_i, tx_i^* \rangle \geq p_{T_M(z,z^*)}(th, th^*).$$

As $\varepsilon > 0$ is arbitrary we have

$$p_{T_M(z,z^*)}(t(h, h^*)) = t^2 p_{T_M(z,z^*)}(h, h^*).$$

The reverse inequality follows immediately on replacing (h, h^*) by $\frac{1}{t}(h, h^*)$ and t by $\frac{1}{t}$. On taking closures we get (29).
For (28) we use duality. Let $(h, h^*) \in \operatorname{dom} \partial \mathcal{F}_{T_M(z,z^*)}$. Then for $(y, y^*)\,(y, y^*) \in \partial \mathcal{F}_{T_M(z,z^*)}(h, h^*)$ we have

$$\mathcal{P}_{T_M(z,z^*)}(y, y^*) + \mathcal{F}_{T_M(z,z^*)}(h, h^*) = \langle (y, y^*), (h, h^*) \rangle$$

and so for all $t > 0$ we have

$$\mathcal{P}_{T_M(z,z^*)}(ty, ty^*) + t^2 \mathcal{F}_{T_M(z,z^*)}(h, h^*) = \langle t(y, y^*), t(h, h^*) \rangle. \tag{30}$$

Using the Fenchel inequality

$$\mathcal{P}_{T_M(z,z^*)}(ty, ty^*) + \mathcal{F}_{T_M(z,z^*)}(th, th^*) \geq \langle t(y, y^*), t(h, h^*) \rangle \tag{31}$$

and subtracting (30) and (31) we have

$$\mathcal{F}_{T_M(z,z^*)}(th, th^*) \geq t^2 \mathcal{F}_{T_M(z,z^*)}(h, h^*).$$

Now replace (h,h^*) with $\gamma(h,h^*)$ and t by $\frac{1}{\gamma}$ for $\gamma>0$. We then have $\gamma^2\mathcal{F}_{T_M(z,z^*)}(h,h^*)\geq\mathcal{F}_{T_M(z,z^*)}(\gamma h,\gamma h^*)$ implying equality. As $\operatorname{dom}\partial\mathcal{F}_{T_M(z,z^*)}$ is dense in $\operatorname{dom}\mathcal{F}_{T_M(z,z^*)}$ the result follows from the results on page 168 of [8] where it is shown that one may recover a closed convex functions from its values at points of subdifferentiability. ■

Note that the last results imply that $\operatorname{dom}\mathcal{P}_{T_M(z,z^*)}$ and $\operatorname{dom}\mathcal{F}_{T_M(z,z^*)}$ are convex cones. The next result is singles out an important property that figures in Fréchet differentiability of $\mathcal{F}_{\mathcal{B}}(z,z^*)$.

Proposition 31 *Suppose $T:X\rightrightarrows X^*$ is maximal monotone, $M:=\operatorname{Graph}T$ and $(z,z^*)\in M$. Suppose $(z,z^*)\in M$ and $\mathcal{B}(z,z^*):=T_M(z,z^*)+(z,z^*)$. If $\nabla\mathcal{F}_{\mathcal{B}}(z,z^*)=(z,z^*)$ then for all $(v,v^*)\in T_M(z,z^*)$ we have $\langle v,v^*\rangle>0$ for all $(v,v^*)\neq 0$. When $\nabla\mathcal{F}_{\mathcal{B}}(z,z^*)=(z,z^*)$ is a Fréchet derivative we have $\inf_{(v,v^*)\in T_M(z,z^*)/(0,0)}\frac{1}{\|(v,v^*)\|^2}\langle v,v^*\rangle>0$.*

Proof. Consider the harder case of $\mathcal{F}_{\mathcal{B}}(z,z^*)$ being Fréchet differentiable. By definition we need to show for all $(h,h^*)\in X\times X^*$ that

$$\lim_{\substack{t\downarrow 0\\(h_t,h_t^*)\to(h,h^*)}}\frac{1}{t}\left(\mathcal{F}_{\mathcal{B}}\left((z,z^*)+t\left(h_t,h_t^*\right)\right)-\mathcal{F}_{\mathcal{B}}(z,z^*)\right)=\langle(z,z^*),(h,h^*)\rangle.$$

Now

$$\begin{aligned}&\frac{1}{t}\left(\mathcal{F}_{\mathcal{B}}\left((z,z^*)+t\left(h,h^*\right)\right)-\mathcal{F}_{\mathcal{B}}(z,z^*)\right)\\&=\frac{1}{t}\left\{\langle z+th,z^*+th^*\rangle-\inf_{(x,x^*)\in\mathcal{B}}\{\langle z+th-x,z^*+th^*-x^*\rangle\}-\langle z,z^*\rangle\right\}\\&=\langle(z,z^*),(h,h^*)\rangle+t\langle h,h^*\rangle-\frac{1}{t}\left\{\inf_{(x,x^*)\in\mathcal{B}}\{\langle z+th-x,z^*+th^*-x^*\rangle\}\right\}.\end{aligned}\tag{32}$$

Consequently we require the following to converge to zero as $t\downarrow 0$ for any $(h,h^*)\in X\times X^*$;

$$\begin{aligned}&-\frac{1}{t}\left\{\inf_{(x,x^*)\in\mathcal{B}}\{\langle z+th-x,z^*+th^*-x^*\rangle\}\right\}+t\langle h,h^*\rangle\\&=\sup_{(x,x^*)\in\mathcal{B}}\left\{\langle x-z,h^*\rangle+\langle h,x^*-z^*\rangle-\frac{1}{t}\langle x-z,x^*-z^*\rangle\right\}.\end{aligned}$$

Now if $\inf_{(v,v^*)\in T_M(z,z^*)/(0,0)}\frac{1}{\|(v,v^*)\|^2}\langle v,v^*\rangle=0$ then there exists a sequence $t_n>0$ with $t_n\to 0$ and $(v_n,v_n^*)\in T_M(z,z^*)$ with $\|(v_n,v_n^*)\|=1$ while

$$1\geq\frac{2}{t_n}\langle v_n,v_n^*\rangle\geq\frac{1}{2}.$$

We may now choose $(h_n,h_n^*)\in X\times X^*$ such that $\langle v_n,h_n^*\rangle+\langle h_n,v_n^*\rangle=1$ and then

$$\begin{aligned}&\sup_{(x,x^*)\in\mathcal{B}}\left\{\langle x-z,h_n^*\rangle+\langle h_n,x^*-z^*\rangle-\frac{1}{t}\langle x-z,x^*-z^*\rangle\right\}\\&\geq\langle v_n,h_n^*\rangle+\langle h_n,v_n^*\rangle-\frac{1}{t_n}\langle v_n,v_n^*\rangle\geq\frac{1}{t_n}\langle v_n,v_n^*\rangle\geq\frac{1}{4}>0\end{aligned}$$

a contradiction. ■

We now give our basic condition for Fréchet differentiability. We need to impose a condition on $T_M(z,z^*)$ in order to characterise Fréchet differentiability of $\mathcal{F}_{\mathcal{B}}(z,z^*)$. When T is Proto-diffentiable at (z,z^*) this condition implies $T_M(z,z^*)$ is contained in a connected component of the set $\{(v,v^*)\mid\langle v,v^*\rangle>0\}$.

Theorem 32 *Suppose $T:X\rightrightarrows X^*$ is maximal monotone, $M:=\operatorname{Graph}T$ and $(z,z^*)\in M$. Then $\nabla\mathcal{F}_{\mathcal{B}}(z,z^*)=(z,z^*)$ is a Fréchet derivative iff $\inf_{(v,v^*)\in T_M(z,z^*)/(0,0)}\frac{1}{\|(v,v^*)\|^2}\langle v,v^*\rangle>0$.*

Proof. Because of Proposition 31 we only need show that $\delta := \inf_{(v,v^*)\in T_M(z,z^*)/(0,0)} \frac{1}{\|(v,v^*)\|^2}\langle v,v^*\rangle > 0$ implies $\nabla\mathcal{F}_{\mathcal{B}}(z,z^*) = (z,z^*)$ is a Fréchet derivative.

First we show that this assumption implies $(h,h^*) \mapsto \mathcal{F}_{T_M(z,z^*)}(h,h^*)$ is bounded above in a neighbourhood of $(0,0)$. Now we note that as T_M is a cone then for fixed $(v,v^*)\in T_M(z,z^*)$ and any $\gamma > 0$ we have $\gamma(v,v^*)\in T_M(z,z^*)$ and so

$$\frac{d}{d\gamma}\left\{\gamma\langle (v,v^*),(h,h^*)\rangle - \gamma^2\langle v,v^*\rangle\right\} = 0$$

implies $\gamma = \frac{\langle (v,v^*),(h,h^*)\rangle}{2\langle v,v^*\rangle}$ and so we have

$$\begin{aligned}\mathcal{F}_{T_M(z,z^*)}(h,h^*) &= \sup_{(v,v^*)\in T_M(z,z^*)} \frac{1}{4}\frac{\langle (v,v^*),(h,h^*)\rangle^2}{\langle v,v^*\rangle} \\ &\leq \frac{1}{4}\sup_{(v,v^*)\in T_M(z,z^*)} \frac{\|(h,h^*)\|^2}{\left(\langle v,v^*\rangle / \|(v,v^*)\|^2\right)} \leq \frac{1}{4\delta}\|(h,h^*)\|^2 < +\infty. \qquad (33)\end{aligned}$$

From (32) we have

$$\begin{aligned}&\frac{1}{t}(\mathcal{F}_{\mathcal{B}}((z,z^*) + t(h,h^*)) - \mathcal{F}_{\mathcal{B}}(z,z^*)) \\ &\quad= \langle (z,z^*),(h,h^*)\rangle + \sup_{(x,x^*)\in\mathcal{B}}\left\{\langle x-z,h^*\rangle + \langle h,x^*-z^*\rangle - \frac{1}{t}\langle x-z,x^*-z^*\rangle\right\} \\ &\quad= \langle (z,z^*),(h,h^*)\rangle + \sup_{(v,v^*)\in T_M(z,z^*)}\left\{\langle (v,v^*),(h,h^*)\rangle - \frac{1}{t}\langle v,v^*\rangle\right\} \\ &\quad= \langle (z,z^*),(h,h^*)\rangle + \frac{1}{t}\sup_{(v,v^*)\in T_M(z,z^*)}\{\langle (v,v^*),t(h,h^*)\rangle - \langle v,v^*\rangle\} \\ &\quad= \langle (z,z^*),(h,h^*)\rangle + \frac{1}{t}\mathcal{F}_{T_M(z,z^*)}(t(h,h^*)).\end{aligned}$$

Now consider

$$\begin{aligned}&\frac{1}{t}\left(\mathcal{F}_{\mathcal{B}}((z,z^*) + t(h,h^*)) + \mathcal{F}_{\mathcal{B}}((z,z^*) - t(h,h^*)) - 2\mathcal{F}_{\mathcal{B}}(z,z^*)\right) \\ &= \frac{1}{t}\left(\mathcal{F}_{T_M(z,z^*)}(t(h,h^*)) + \mathcal{F}_{T_M(z,z^*)}(t(-h,-h^*))\right) \\ &= t\left(\mathcal{F}_{T_M(z,z^*)}(h,h^*) + \mathcal{F}_{T_M(z,z^*)}(-h,-h^*)\right) \leq t\max\left\{\mathcal{F}_{T_M(z,z^*)}(h,h^*),\mathcal{F}_{T_M(z,z^*)}(-h,-h^*)\right\}. \qquad (34)\end{aligned}$$

We note that as $(h,h^*)\mapsto \mathcal{F}_{T_M(z,z^*)}(h,h^*)$ is bounded above in a neighbourhood of $(0,0)$ (and consequently is globally finite being positively homogeneous of degree two) then for all $\|(h,h^*)\|\leq 1$ we have from (33) that

$$\mathcal{F}_{T_M(z,z^*)}(\pm h,\pm h^*) \leq \frac{1}{4\delta}.$$

Thus from (34) we have

$$\mathcal{F}_{\mathcal{B}}((z,z^*) + t(h,h^*)) + \mathcal{F}_{\mathcal{B}}((z,z^*) - t(h,h^*)) - 2\mathcal{F}_{\mathcal{B}}(z,z^*) \leq t^2\frac{1}{2\delta} \leq \varepsilon t$$

for $t \leq \delta := 2\delta\varepsilon$. Proposition 1.23 of [19] now applies and the Fréchet differentiability follows. ■

Corollary 33 *Suppose $T : X \rightrightarrows X^*$ is maximal monotone, $M := \operatorname{Graph} T$, $(z,z^*)\in M$ and M is strongly Proto–differentiable at (z,z^*) with Proto-derivative $T_M(z,z^*)$. Suppose in addition that $\inf_{(v,v^*)\in T_M(z,z^*)/(0,0)}\frac{1}{\|(v,v^*)\|^2}\langle v,v^*\rangle > 0$. Then $T(z) = \{z^*\}$ with $d^*(z^*,T(y)) = O(\|y-z\|)$ and also $T_M(z,z^*)$ is maximal monotone.*

Proof. This follows from Theorem 28 and Theorem 32. ■

6 A Characterisation of the normal cone $T_M(z,z^*)^\circ$

We finish this study by demonstrating that the normal cone $T_M(z,z^*)^\circ$ at $(z,z^*) \in M$ has a rather nice characterisation in terms of Fitzpatrick functions. When these results are married with those of Section 3 we have the machinery to actually calculate such normal cones, at least at points of strong Proto-differentiability when we have at hand the Fitzpatrick function. The calculation of the Fitzpatrick function is not as intractable as one might first conjecture. A number of examples may be found in [7].

First we note that in view of Lemma 3 the inclusion in (22) may be restated in the following form:

$$T_M(z,z^*)^\circ \subseteq \{(v,v^*) \mid \langle v,v^*\rangle \le 0\}.$$

We may strengthen this inclusion in the following form.

Lemma 34 *Suppose $T : X \rightrightarrows X^*$ is maximal monotone, $M := \operatorname{Graph} T$, $(z,z^*) \in M$. Then*

$$T_M(z,z^*)^\circ \subseteq \left\{(v,v^*) \mid \mathcal{F}_{T_M(z,z^*)}(v,v^*) \le 0\right\}$$

and when $T_M(z,z^*)$ is maximal we have

$$\left\{(v,v^*) \mid \mathcal{F}_{T_M(z,z^*)}(v,v^*) \le 0\right\} \subseteq \{(v,v^*) \mid \langle v,v^*\rangle \le 0\}.$$

Proof. Note that as $T_M(z,z^*)$ is monotone and $(0,0) \in T_M(z,z^*)$ we have $\langle w,w^*\rangle \ge 0$ for all $(w,w^*) \in T_M(z,z^*)$. Thus

$$\begin{aligned}\mathcal{F}_{T_M(z,z^*)}(v,v^*) &= \sup_{(w,w^*)\in T_M(z,z^*)} \{\langle (v,v^*),(w,w^*)\rangle - \langle w,w^*\rangle\} \\ &\le \sup_{(w,w^*)\in T_M(z,z^*)} \{\langle (v,v^*),(w,w^*)\rangle\} \le 0 \quad ,\forall (v,v^*) \in T_M(z,z^*)^\circ.\end{aligned}$$

When $T_M(z,z^*)$ is maximal we have $\langle v,v^*\rangle \le \mathcal{F}_{T_M(z,z^*)}(v,v^*)$ providing the last inclusion. ■

We now require an alternative expression for the zero level-set of the Fitzpatrick function $\mathcal{F}_{T_M(z,z^*)}$.

Proposition 35 *Suppose $T : X \rightrightarrows X^*$ is maximal monotone, $M := \operatorname{Graph} T$, $(z,z^*) \in M$. Then*

$$\partial \mathcal{P}_{T_M(z,z^*)}(0,0) = \left\{(v,v^*) \mid \mathcal{F}_{T_M(z,z^*)}(v,v^*) \le 0\right\}.$$

Proof. Take $(w,w^*) \in \partial \mathcal{P}_{T_M(z,z^*)}(0,0)$ then

$$\mathcal{P}_{T_M(z,z^*)}(y,y^*) - \mathcal{P}_{T_M(z,z^*)}(0,0) \ge \langle (w,w^*),(y,y^*)\rangle. \tag{35}$$

Now $(0,0) \in T_M(z,z^*)$ and monotonicity implies $\langle v,v^*\rangle \ge 0$ for all $(v,v^*) \in T_M(z,z^*)$ and so $\mathcal{P}_{T_M(z,z^*)}(y,y^*) \ge 0$. As $(0,0) \in \operatorname{co} T_M(z,z^*) \subseteq \operatorname{dom} \mathcal{P}_{T_M(z,z^*)}$ we may take $t > 0$ and note that the positive homegenuity gives

$$0 = \liminf_{t\downarrow 0} t^2 \mathcal{P}_{T_M(z,z^*)}(0,0) = \liminf_{t\downarrow 0} \mathcal{P}_{T_M(z,z^*)}(t\times 0, t\times 0) \ge \mathcal{P}_{T_M(z,z^*)}(0,0) \ge 0. \tag{36}$$

Thus $\mathcal{P}_{T_M(z,z^*)}(0,0) = 0$ and (35) gives for all $(y,y^*) \in T_M(z,z^*)$ that

$$\langle y,y^*\rangle = \mathcal{P}_{T_M(z,z^*)}(y,y^*) \ge \langle (w,w^*),(y,y^*)\rangle.$$

Consequently

$$0 \ge \sup_{(y,y^*)\in T_M(z,z^*)} \{\langle (w,w^*),(y,y^*)\rangle - \langle y,y^*\rangle\} = \mathcal{F}_{T_M(z,z^*)}(w,w^*)$$

$$\text{and so} \quad \partial \mathcal{P}_{T_M(z,z^*)}(0,0) \subseteq \left\{(w,w^*) \mid \mathcal{F}_{T_M(z,z^*)}(w,w^*) \le 0\right\}.$$

Now take (w,w^*) with $\mathcal{F}_{T_M(z,z^*)}(w,w^*) \le 0$. Consequently

$$\begin{aligned} \langle y,y^*\rangle &\ge \langle (w,w^*),(y,y^*)\rangle \qquad \text{for all } (y,y^*) \in T_M(z,z^*) \\ \text{implying} \quad \sum_i \lambda_i \langle y_i, y_i^*\rangle &\ge \langle (w,w^*),(y,y^*)\rangle \qquad \text{for all } (y,y^*) \in \operatorname{co} T_M(z,z^*). \end{aligned}$$

It follows that $p_{T_M(z,z^*)}(y,y^*) \ge \langle (w,w^*),(y,y^*)\rangle$ and, on taking closures, we have via continuity of $(y,y^*) \mapsto \langle (w,w^*),(y,y^*)\rangle$ that

$$\begin{aligned} \mathcal{P}_{T_M(z,z^*)}(y,y^*) &\ge \langle (w,w^*),(y,y^*)\rangle \\ \text{or} \quad \mathcal{P}_{T_M(z,z^*)}(y,y^*) - \mathcal{P}_{T_M(z,z^*)}(0,0) &\ge \langle (w,w^*),(y,y^*)-(0,0)\rangle, \end{aligned}$$

giving $(w,w^*) \in \partial \mathcal{P}_{T_M(z,z^*)}(0,0)$. ■

We are now able to characterise $T_M(z,z^*)^\circ$. We note that at points where M is strongly Proto-differentiable we may perform a calculation that follows directly from the original Penot or Fitzpatrick function of T itself.

Theorem 36 *Suppose $T : X \rightrightarrows X^*$ is maximal monotone, $M := \operatorname{Graph} T$, $(z,z^*) \in M$. Then*

$$T_M(z,z^*)^\circ = \partial \mathcal{P}_{T_M(z,z^*)}(0,0) = \left\{(v,v^*) \mid \mathcal{F}_{T_M(z,z^*)}(v,v^*) \le 0\right\}. \tag{37}$$

Moreover suppose in addition that M is strongly Proto-differentiable at (z,z^*) then

$$T_M(z,z^*)^\circ = \partial \mathcal{P}''_T(z,z^*)(0,0) = \{(v,v^*) \mid \mathcal{F}''_T(z,z^*)(v,v^*) \le 0\}. \tag{38}$$

Proof. We only need show (37) as (38) then follows immediately from Theorem 22. From Lemma 34 and Proposition 35 we have the left hand side of (37) contained in the right hand side. Now take $(w,w^*) \in \partial \mathcal{P}_{T_M(z,z^*)}(0,0)$ then by duality $(0,0) \in \partial \mathcal{F}_{T_M(z,z^*)}(w,w^*)$ and so for all (y,y^*)

$$\mathcal{F}_{T_M(z,z^*)}(y,y^*) - \mathcal{F}_{T_M(z,z^*)}(w,w^*) \ge 0.$$

Thus, by definition

$$\mathcal{F}_{T_M(z,z^*)}(y,y^*) \ge \langle (w,w^*),(v,v^*)\rangle - \langle v,v^*\rangle \qquad \text{for all } (v,v^*) \in T_M(z,z^*).$$

Let $(y,y^*) \in \partial \mathcal{P}_{T_M(z,z^*)}(0,0)$. Then $\mathcal{F}_{T_M(z,z^*)}(y,y^*) \le 0$ and so

$$\langle v,v^*\rangle \ge \langle (w,w^*),(v,v^*)\rangle \qquad \text{for all } (v,v^*) \in T_M(z,z^*).$$

As $T_M(z,z^*)$ is a cone we may take $t > 0$ and have $t(v,v^*) \in T_M(z,z^*)$. Consequently

$$t^2\langle v,v^*\rangle \ge t\langle (w,w^*),(v,v^*)\rangle \qquad \text{for all } (v,v^*) \in T_M(z,z^*)$$

and on letting $t \downarrow 0$ we obtain

$$0 = \lim_{t\downarrow 0} t\langle v,v^*\rangle \ge \langle (w,w^*),(v,v^*)\rangle \qquad \text{for all } (v,v^*) \in T_M(z,z^*)$$

or $(w,w^*) \in T_M(z,z^*)^\circ$. ■

7 Conclusion

Much of the literature on representative functions has concerned itself with providing shorter proofs for the preservation of maximal monotonicity under various operations such as sums and compositions. We have shown that the Fitzpatrick function is not only a useful tool in studying these issues but is also useful in studying the differentiability properties of maximal monotone operators. The characterisation of the Proto- normal cone to the graph of a maximal monotone operators, via representative functions, provides a very concise and elegant formula for these important differential objects.

References

[1] H. Attouch and R.J.B. Wets (1983), *A Convergence for Bivariate Functions Aimed at the Convergence of Saddle Values,* Lecture Notes in Mathematics, Vol. 979 Mathematical Theories of Optimization, Ed. Cecconi and Zolezzi, pp 1–42.

[2] G. Beer, (1993), *Topologies on Closed and Closed Convex Sets*, Mathematics and its Applications Vol. 268, Kluwer Acad. Publ.

[3] J.M. Borwein (2006), *Maximal monotonicity via convex analysis*, Journal of Convex Analysis, Vol. 13, no. 3/4, pp 561-586.

[4] J. Borwein and M. Fabian, (1993) *On Second Order Gateaux Differentiability*, Nonlinear Analysis Theory, Method and Applications, Vol. 20, no. 12, pp 1373-1382.

[5] J.M. Borwein (2006), *Maximality of sums of two maximal monotone operators in general Banach space*, Proc. AMS, 135 (2007), pp 3917-3924.

[6] S. Fitzpatrick (1988), *Representing monotone operators by convex functions,* Workshop/Miniconference on Functional Analysis and Optimization (Camberra 1988), Proceedings Center Math. Anal. Austral. Nat. Univ., 20 Austral. Nat Univ. Canberra, pp 59–65.

[7] Bauschke, Heinz H.; McLaren, D. Alexander; Sendov, Hristo S. Fitzpatrick functions: inequalities, examples, and remarks on a problem by S. Fitzpatrick. J. Convex Anal. 13 (2006), no. 3-4, 499–523.

[8] R. B. Holmes (1975), *Geometric Functional Analysis and its Applications*, Springer–Verlag Graduate Texts in Mathematics, Vol. 24.

[9] Krauss, Eckehard (1985) *A representation of arbitrary maximal monotone operators via subgradients of skew-symmetric saddle functions.* Nonlinear Anal. 9, no. 12, 1381–1399.

[10] B. S. Mordukhovich (1984), Nonsmooth analysis with nonconvex generalized differentials and adjoint mappings, *Dokl. Akad. Nauk BSSR* **28**, 976–979.

[11] B. S. Mordukhovich (2005), *Variational Analysis and Generalized Differentiation, I: Basic Theory, II: Applications*, Comprehensive Studies in Mathematics **330**, **331**, Springer.

[12] B.S Mordukhovich, (1994), *Generalized differential calculus for nonsmooth and set-valued mappings.* J. Math. Anal. Appl. 183, pp 250–288.

[13] B.S Mordukhovich and J.V Outrata (2001) *On second-order subdifferentials and their applications.* SIAM J. Optim. 12, pp 139–169.

[14] Do, Ngoc. Chi, (1992) *Generalized Second–Order Derivatives of a Convex Functions in a Reflexive Banach Space,* Transactions of the American Mathematical Society, Vol. 334, no. 1, pp 281-301.

[15] J-P. Penot and C. Zalinescu (2004), *Continuity of the Legendre–Fenchel Transformation for Some Variational Convergences,* Optimization, Vol. 53, no. 5-6, pp 549–562.

[16] J-P. Penot and C. Zalinescu (2005), *On the Convergence of Maximal Monotone Operators*, proceedings of the American Mathematical Society, Vol. 134, no. 7, pp 1937–1946.

[17] J-P. Penot (2004), *The relevance of convex analysis for the study of monotonicity,* Nonlinear Anal., Vol. 58, no. 7-8, pp 855–871.

[18] J-P. Penot and C. Zalinescu (2005), *Some problems about the representation of monotone operators by convex functions*, ANZIAM J., Vol. 47 , no. 1, pp 1–20.

[19] R.R Phelps, (1998), *Convex Functions Montone Operators and Differentiability*, Lecture Notes in Mathematics, no. 1364, Springer-Verlag.

[20] R. T. Rockafellar (1974), *Conjugate Duality and Optimization*, SIAM publ.

[21] R. T. Rockafellar (1970), *Convex Analysis*, Princeton University Press.

[22] Singer, Ivan (1979), *A Fenchel-Rockafellar type duality theorem for maximization*. Bull. Austral. Math. Soc. 20 (1979), no. 2, 193–198.