

Preconditioned Proximal Point Methods and Notions of Partial Subregularity

Tuomo Valkonen*

*Dept. of Mathematics and Statistics, University of Helsinki, Finland;
and: ModeMat, Escuela Politécnica Nacional, Quito, Ecuador
tuomo.valkonen@iki.fi*

Received: July 3, 2018

Accepted: June 1, 2020

Based on the needs of convergence proofs of preconditioned proximal point methods, we introduce notions of partial strong submonotonicity and partial (metric) subregularity of set-valued maps. We study relationships between these two concepts, neither of which is generally weaker or stronger than the other one. For our algorithmic purposes, the novel submonotonicity turns out to be easier to employ than more conventional error bounds obtained from subregularity. Using strong submonotonicity, we demonstrate the linear convergence of the Primal-Dual Proximal Splitting method to some strictly complementary solutions of example problems from image processing and data science. This is without the conventional assumption that all the objective functions of the involved saddle point problem are strongly convex.

Keywords: Subregularity, submonotonicity, error bounds, partial, proximal point method.

2010 Mathematics Subject Classification: 49J52, 47H05, 49M05, 49M29.

1. Introduction

What is the weakest useful form of regularity of a set-valued map H ? In particular, if $0 \in H(\hat{u})$ for $\hat{u} = (\hat{x}, \hat{y})$ encodes optimality conditions of a saddle point problem

$$\min_{x \in X} \max_{y \in Y} G(x) + \langle Kx, y \rangle - F^*(y), \quad (1)$$

in Hilbert spaces X and Y , what regularity property is useful for showing the fast convergence of optimisation methods? In this work, we study this question in relation to preconditioned proximal point methods [43].

A starting point for the regularity of set-valued maps is to extend the Lipschitz property of single-valued maps. One such approach is the *Aubin*, *pseudo-Lipschitz*, or *Lipschitz-like* property of [6]. When we are interested in the stability of the optimality condition $0 \in H(\hat{u})$, it is more beneficial to study the Aubin property of the inverse H^{-1} . This is called the *metric regularity* of H . In this property, at $\hat{u}$ for $\hat{w} \in H(\hat{u})$, both u and w are allowed to vary in the criterion

$$\kappa \operatorname{dist}(w, H(u)) \geq \operatorname{dist}(u, H^{-1}(w)) \quad (u \in \mathcal{U}, w \in \mathcal{W}),$$

which is assumed to hold for some $\kappa > 0$, and neighbourhoods $\mathcal{U} \ni \hat{u}$ and $\mathcal{W} \ni \hat{w}$.

*Most of this work was completed while at the Department of Mathematical Sciences, University of Liverpool, United Kingdom.

We make standard notation such as “dist” explicit in Section 2. Metric regularity is equivalent to *openness at a linear rate* at $(\hat{u}, \hat{w})$, and holds for smooth maps by the classical Lyusternik-Graves theorem; see, e.g., [25]. It is too strong a property to be satisfied in many applications. *Metric subregularity* [24, 14] allows much more leeway for H by fixing $w = \hat{w}$. In other words, we require

$$\kappa \operatorname{dist}(\hat{w}, H(u)) \geq \operatorname{dist}(u, H^{-1}(\hat{w})) \quad (u \in \mathcal{U}). \quad (2)$$

The counterpart of metric subregularity that relaxes the Aubin property is known as *calmness* or the *upper Lipschitz* property [39]. We refer to the books [7, 34, 42, 13, 25] for further information on these and other related properties. These include the Mordukhovich criterion that allows verifying the Aubin property or metric regularity through coderivative considerations.

Monotonicity is an important basic property satisfied by the convex subdifferential. For general set-valued maps $H : U \rightrightarrows U$ on a Hilbert space U , it states

$$\langle q - q^*, u - u^* \rangle \geq 0 \quad ((u, q), (u^*, q^*) \in \operatorname{graph} H).$$

Basic monotonicity does not help to achieve fast convergence of optimisation methods. This can be achieved with strong monotonicity

$$\langle q - q^*, u - u^* \rangle \geq \gamma \|u - u^*\|^2 \quad ((u, q), (u^*, q^*) \in \operatorname{graph} H).$$

In this work, in the spirit of subregularity, we want to relax strong monotonicity to *strong submonotonicity* at a point. To study the convergence of optimisation methods on subspaces, we will formulate both submonotonicity and subregularity in terms of operators: we introduce *partial (metric) subregularity* and *partial strong submonotonicity*. In this theme, our *main contributions* are:

- (1) In Section 3 we introduce the new concept of *(partial, strong) submonotonicity*. Expanding from our previous work in [43], we apply it to the convergence proofs of abstract preconditioned proximal point methods.
- (2) In Section 4 we introduce partial variants of familiar error bounds and metric subregularity. Based on them we again formulate convergence results for abstract preconditioned proximal point methods.
- (3) In Section 5 we study the relationship between submonotonicity and subregularity.
- (4) In Section 6 we apply the work of the previous sections to obtain improved convergence results for the primal-dual proximal splitting (PDPS) of Chambolle and Pock [8].

Before this, in Section 2, we introduce notation.

There are further properties in the literature, related to our work. *Strong metric subregularity* is a strengthening of metric subregularity. Its many properties are studied in [10], along with introducing q -exponent versions. Particularly worth noting is that strong metric subregularity is invariant with respect to perturbations by smooth functions, while metric subregularity is not. Minding our focus on monotonicity, of particular relevance is that for convex functions f , strong metric subregularity of the subdifferential amounts to local strong convexity.

Indeed, according to [3, 4], the (convex) subdifferential ∂f is *strongly* metrically subregular at $\hat{u}$ for $\hat{w} \in \partial f(\hat{u})$ if and only if there exists $\gamma > 0$ and a neighbourhood $\mathcal{U} \ni \hat{u}$ such that

$$f(u) - f(\hat{u}) \geq \langle \hat{w}, u - \hat{u} \rangle + \gamma \|u - \hat{u}\|^2 \quad (u \in \mathcal{U}).$$

Metric subregularity, on the other hand, amounts of measuring growth away from the entire set $[\partial f]^{-1}(\hat{w})$ instead of the single point $\hat{u}$. This gives the corresponding criterion

$$f(u) - f(\hat{u}) \geq \langle \hat{w}, u - \hat{u} \rangle + \gamma \operatorname{dist}^2(u, [\partial f]^{-1}(\hat{w})) \quad (u \in \mathcal{U}).$$

In Section 5 we derive similar correspondences between partial metric subregularity and partial strong submonotonicity.

Some weaker and partial concepts of regularity have also been considered in the literature. Of particular note is the directional metric subregularity of [19]. The idea here is to study necessary optimality conditions by only requiring metric regularity or subregularity to critical directions instead of all directions. A somewhat similar idea is considered in [37]. We, by contrast, are interested in regularity on entire subspaces where our objective functions behave better than globally. A form of directional metric subregularity more closely resembling our partial subregularity has been studied in [15, 46]. This notion models directionality with subsets, while ours uses operators. Moreover, through the use of three rather than two operators, our definitions in Section 3 split the partiality or directionality on two sides of the defining inequality. As we show in Section 5, this yields a somewhat weaker property, sufficient for our needs.

Related to our objectives, partial calmness of bi-level programs with respect to perturbations of only certain optimality criteria have been studied in [47], see also [23]. We have also ourselves studied partial strong convexity in [45, 44] for the acceleration of optimisation methods on subspaces of strong convexity. Moving from subspaces to manifolds, [30, 28] apply smoothness restricted to submanifolds to prove local linear convergence of optimisation algorithms to such submanifolds.

More generally, local linear convergence can be derived from error bounds, first introduced in [31] for matrix splitting problems, and studied for other methods including the ADMM and the proximal point method, among others in [5, 21, 27, 29]. An alternative approach to the proximal point method is taken in [2] based on Lyusternik-Graves style estimates, while [1] presents an approach based on metric regularity to Newton's method for variational inclusions. In [48] a unified approach to error bounds for generic smooth constrained problems is introduced. In Section 4 we will do something roughly similar, but for nonsmooth optimisation. Robinson stability [38, 20] can also be seen as a stronger parametric version of error bounds, useful in the study of sensitivity of solution mappings. In this work, we are primarily interested in the application of our regularity criteria to the convergence of optimisation methods.

2. Notation

Let X and Y be Hilbert spaces. We denote by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ the corresponding inner products and induced norms. We use $\mathbb{L}(X; Y)$ to denote the space of bounded linear operators between X and Y . We denote the *identity operator* by I .

For $T, S \in \mathbb{L}(X; X)$, we write $T \geq S$ when $T - S$ is positive semidefinite. Also for possibly non-self-adjoint $T \in \mathbb{L}(X; X)$, we introduce the inner product and norm-like notations

$$\langle x, z \rangle_T := \langle Tx, z \rangle, \quad \text{and} \quad \|x\|_T := \sqrt{\langle x, x \rangle_T}. \quad (3)$$

For a set $A \subset X$, we write δ_A for the $\{0, \infty\}$ -valued *indicator function* and define the distance

$$\text{dist}_T(z, A) := \inf_{u \in A} \|z - u\|_T.$$

We also write $\text{dist}_T^2(z, A) := \text{dist}_T(z, A)^2$.

We denote the $\{0, \infty\}$ -valued indicator function of a set $A \subset X$ by δ_A , and write $A \geq 0$ if every element $t \in A$ satisfies $t \geq 0$. Also, we write $\langle A, x \rangle = \{\langle y, x \rangle \mid y \in A\}$ whenever $x \in X$, so that the inequality $\langle A, x \rangle \geq 0$ means $\langle y, x \rangle \geq 0$ for all $y \in A$.

Finally, we write $\mathbb{B}(x, \alpha)$ for the open ball of radius α around $x \in X$, and, for a possibly set-valued $H : X \rightrightarrows Y$, the graph by $\text{graph } H := \{(x, y) \mid y \in H(x), x \in X\}$.

Let $f : X \rightrightarrows \mathbb{R}$. We write ∂f for the convex subdifferential if f convex. If f is locally Lipschitz, we write ∂_C for the *Clarke subdifferential*. We write f^* for the Fenchel conjugate.

3. Submonotonicity and preconditioned proximal point methods

Our objective is to solve the variational inclusion $0 \in H(\hat{u})$, where $H : U \rightrightarrows U$ is a set-valued map on a Hilbert space U . Our strategy is for some iteration-dependent $M_{i+1} \in \mathbb{L}(U; U)$ the preconditioned proximal point method: given an initial iterate u^0 , iteratively solve u^{i+1} from

$$0 \in H(u^{i+1}) + M_{i+1}(u^{i+1} - u^i). \quad (\text{PP})$$

In [43] we developed general convergence theory of such methods and more general methods that include *non-linear* preconditioners for the modelling forward steps. Here, for simplicity, we only consider linear preconditioning. By [22], the primal-dual proximal splitting of [8] is of the form (PP). We will discuss it in Section 6. By [43, 12] many more algorithms are of the form (PP).

We improve the results of [43] to require weaker regularity of H by only asking for convergence to the *set* of roots $H^{-1}(0)$ instead of a specific (unique) root. In this section, we specifically use submonotonicity that we define in Section 3.2 while in Section 4 we look at subregularity and error bounds. Before this, in Section 3.1, we provide the fundamental estimate for such convergence results.

3.1. A convergence estimate

The next lemma is the basis of our convergence results. The *testing operators* Z_{i+1} combined with the preconditioners M_{i+1} form local metrics that measure the rate of convergence.

Lemma 3.1. *Let $H : U \rightrightarrows U$ on a Hilbert space U with $H^{-1}(0)$ non-empty. Also let $M_{i+1}, Z_{i+1} \in \mathbb{L}(U; U)$ with $Z_{i+1}M_{i+1}$ self-adjoint for all $i \in \mathbb{N}$. Suppose (PP) is solvable for the iterates $\{u^{i+1}\}_{i \in \mathbb{N}}$ given $u^0 \in U$.*

If, for all $i \in \mathbb{N}$,

$$\begin{aligned} \frac{1}{2} \|u^{i+1} - u^i\|_{Z_{i+1}M_{i+1}}^2 + \inf_{u^* \in H^{-1}(0)} \left(\frac{1}{2} \|u^{i+1} - u^*\|_{Z_{i+1}M_{i+1}}^2 + \langle H(u^{i+1}), u^{i+1} - u^* \rangle_{Z_{i+1}} \right) \\ \geq \frac{1}{2} \text{dist}_{Z_{i+2}M_{i+2}}^2(u^{i+1}; H^{-1}(0)) \end{aligned} \quad (4)$$

$$\text{then} \quad \frac{1}{2} \text{dist}_{Z_{n+1}M_{n+1}}^2(u^N, H^{-1}(0)) \leq \frac{1}{2} \text{dist}_{Z_1M_1}^2(u^0, H^{-1}(0)) \quad (n \geq 1). \quad (5)$$

Proof. Let $u^* \in H^{-1}(0)$ be arbitrary. Inserting (PP) into (4), we obtain

$$\begin{aligned} \frac{1}{2} \|u^{i+1} - u^i\|_{Z_{i+1}M_{i+1}}^2 + \inf_{u^* \in H^{-1}(0)} \left(\frac{1}{2} \|u^{i+1} - u^*\|_{Z_{i+1}M_{i+1}}^2 - \langle u^{i+1} - u^i, u^{i+1} - u^* \rangle_{Z_{i+1}M_{i+1}} \right) \\ \geq \frac{1}{2} \text{dist}_{Z_{i+2}M_{i+2}}^2(u^{i+1}; H^{-1}(0)). \end{aligned} \quad (6)$$

We recall for general self-adjoint M the three-point formula

$$\langle u^{i+1} - u^i, u^{i+1} - u^* \rangle_M = \frac{1}{2} \|u^{i+1} - u^i\|_M^2 - \frac{1}{2} \|u^i - u^*\|_M^2 + \frac{1}{2} \|u^{i+1} - u^*\|_M^2. \quad (7)$$

Using this with $M = Z_{i+1}M_{i+1}$, we rewrite (6) as

$$\frac{1}{2} \text{dist}_{Z_{i+1}M_{i+1}}^2(u^i; H^{-1}(0)) \geq \frac{1}{2} \text{dist}_{Z_{i+2}M_{i+2}}^2(u^{i+1}; H^{-1}(0)).$$

Summing over $i = 0, \dots, n-1$, we obtain the claim. $\square$

In the present work, we concentrate on $Z_{i+1}M_{i+1} \geq 0$ growing fast enough that (5) yields convergence rates. If such growth is not present, then under some technical assumptions, it is still possible to obtain weak convergence. Since this is not our focus, we merely refer to [43] for arguments applicable when $\text{dist}_{Z_{N+1}M_{N+1}}^2(u^N, H^{-1}(0))$ is replaced by $\|u^N - \hat{u}\|_{Z_{N+1}M_{N+1}}^2$ for some fixed $\hat{u} \in H^{-1}(0)$.

3.2. Submonotonicity

Suppose $Z_{i+1}H$ is strongly monotone with respect to the operator $Z_{i+1}\Xi_{i+1}$ for some $\Xi_{i+1} \in \mathbb{L}(U; U)$ in the sense that

$$\langle H(u) - H(u'), u - u' \rangle_{Z_{i+1}} \geq \frac{1}{2} \|u - u'\|_{Z_{i+1}\Xi_{i+1}}^2 \quad (u, u' \in U). \quad (8)$$

Then, by taking $u' = u^* \in H^{-1}(0)$, we see (4) to hold if

$$\begin{aligned} \frac{1}{2} \|u^{i+1} - u^i\|_{Z_{i+1}M_{i+1}}^2 + \inf_{u^* \in H^{-1}(0)} \frac{1}{2} \|u^{i+1} - u^*\|_{Z_{i+1}(M_{i+1} + \Xi_{i+1})}^2 \\ \geq \frac{1}{2} \text{dist}_{Z_{i+2}M_{i+2}}^2(u^{i+1}; H^{-1}(0)). \end{aligned}$$

By the definition of $\text{dist}_{Z_{i+2}M_{i+2}}^2$, this holds if we have $Z_{i+1}(M_{i+1} + \Xi_{i+1}) \geq Z_{i+2}M_{i+2}$. This limits the growth of $i \mapsto Z_{i+1}M_{i+1}$ through the available strong monotonicity, and hence the convergence rates (5) can yield.

We do not, however, need the full power of strong monotonicity. Indeed, since $u' = u^* \in H^{-1}(0)$ was arbitrary above, we are led to think we can take the infimum in (8) over $u' \in H^{-1}(0)$. However, we have to be careful to keep this infimisation compatible with $\text{dist}_{Z_{i+2}M_{i+2}}^2(u^{i+1}; H^{-1}(0))$. We therefore introduce the following concept.

Definition 3.2. Let U be a Hilbert space and $N, M, \Xi \in \mathbb{L}(U; U)$ with $M \geq 0$. We say that $T : U \rightrightarrows U$ is (Ξ, N, M) -*partially strongly submonotone* at $(\hat{u}, \hat{w}) \in \text{graph } T$ (or *at $\hat{u}$ for $\hat{w} \in T(\hat{u})$*) if there exists a neighbourhood $\mathcal{U} \ni \hat{u}$ with

$$\inf_{u^* \in T^{-1}(\hat{w})} (\langle w - \hat{w}, u - u^* \rangle_N + \|u - u^*\|_{M-\Xi}^2) \geq \text{dist}_M^2(u, T^{-1}(\hat{w})) \quad (\text{PSM})$$

for all $u \in \mathcal{U}$ and $w \in T(u)$. If $\Xi = M$, we say that T is (N, M) -*strongly submonotone*. If $\Xi = 0$, we say that T is (N, M) -*submonotone*.

The idea is that Ξ is the “part” of M with respect to which T is *strongly* submonotone. If we fix the element $u^* \in T^{-1}(\hat{w})$ in both the infimisations and the dist in (PSM), then M cancels out and the inequality reduces to a variant of the operator-relative strong monotonicity (8). The infimisation is the reason for calling the concept *submonotonicity*.

Remark 3.3. (Submonotonicity from local monotonicity) (Ξ, M, N) -partial strong submonotonicity for any $M \geq 0$ is implied by

$$\langle w - \hat{w}, u - u^* \rangle_N \geq \|u - u^*\|_{\Xi}^2 \quad (u \in \mathcal{U}, w \in T(u), u^* \in T^{-1}(\hat{w})). \quad \square$$

Remark 3.4. (Limited dependence on base point) Submonotonicity only depends on $\hat{u}$ through $\mathcal{U}$. $\square$

Returning to the preconditioned proximal point method (PP), the next result shows how $Z_{i+2}M_{i+2}$ can be made to grow based on partial strong submonotonicity, and therefore how this can help us obtain convergence rates.

Theorem 3.5. *On a Hilbert space U , let $H : U \rightrightarrows U$, and $M_{i+1}, Z_{i+1}, \Xi_{i+1} \in \mathbb{L}(U; U)$ with $Z_{i+1}M_{i+1} \geq 0$ self-adjoint for all $i \in \mathbb{N}$. Suppose $H^{-1}(0)$ is non-empty and (PP) is solvable for the iterates $\{u^{i+1}\}_{i \in \mathbb{N}}$ given $u^0 \in U$.*

If H is $(Z_{i+1}\Xi_{i+1}, 2Z_{i+1}, Z_{i+2}M_{i+2})$ -partially strongly submonotone at some point $(\hat{u}, 0) \in \text{graph } H$ with a uniform neighbourhood $\mathcal{U}$ for all $i \in \mathbb{N}$, and

$$Z_{i+1}(M_{i+1} + \Xi_{i+1}) \geq Z_{i+2}M_{i+2} \quad (i \in \mathbb{N}), \quad (9)$$

then (5) holds provided $\{u^i\}_{i=0}^N \subset \mathcal{U}$ for the neighbourhood $\mathcal{U}$ of partial strong submonotonicity.

Proof. The assumed partial strong submonotonicity for all $u^* \in H^{-1}(0)$ implies

$$\langle H(u^{i+1}), u^{i+1} - u^* \rangle_{Z_{i+1}} + \frac{1}{2} \|u^{i+1} - u^*\|_{Z_{i+2}M_{i+2} - Z_{i+1}\Xi_{i+1}}^2 \geq \frac{1}{2} \text{dist}_{Z_{i+2}M_{i+2}}^2(u^{i+1}; H^{-1}(0)).$$

Using (9) and taking the infimum over $u^* \in H^{-1}(0)$, we obtain (4). Then we just apply Lemma 3.1. $\square$

Remark 3.6. If $Z_{k+1} \equiv I$, then Theorem 3.5 requires $(\Xi_{i+1}, 2I, M_{i+2})$ -partial strong submonotonicity, roughly speaking, submonotonicity with respect to the the part $\Xi_{i+1}/2$ of $M_{i+2}/2$. $\square$

3.3. Transformation of submonotonicity

Before looking at more concrete examples, we provide results to transform submonotonicity with respect to one triple of operators to another triple of operators. The first result is basic:

Proposition 3.7. (Scaling invariance) *On a Hilbert space U , suppose $T : U \rightrightarrows U$ is (Ξ, N, M) -partially strongly submonotone at $(\hat{u}, \hat{w}) \in \text{graph } T$ in a neighbourhood $\mathcal{U}$. Then for any $\alpha > 0$, at the same point $(\hat{u}, \hat{w})$ in the same neighbourhood $\mathcal{U}$,*

- (i) *T is $(\alpha\Xi, \alpha N, \alpha M)$ -partially strongly submonotone.*
- (ii) *T is $(\Xi, N, (1 + \alpha)M)$ -partially strongly submonotone.*

Proof. (i) is obvious. For (ii) using the (Ξ, N, M) -partial strong submonotonicity we obtain

$$\begin{aligned} & \inf_{u^* \in T^{-1}(\hat{w})} (\langle w - \hat{w}, u - u^* \rangle_N + \|u - u^*\|_{(1+\alpha)M-\Xi}^2) \\ & \geq \inf_{u^* \in T^{-1}(\hat{w})} (\langle w - \hat{w}, u - u^* \rangle_N + \|u - u^*\|_{M-\Xi}^2) + \alpha \inf_{u^* \in T^{-1}(\hat{w})} \|u - u^*\|_M^2 \\ & \geq \text{dist}_{(1+\alpha)M}^2(u, T^{-1}(\hat{w})) \quad (u \in \mathcal{U}, w \in T(u)). \end{aligned} \quad \square$$

The following results will require the following concept, which will also be central Section 5:

Definition 3.8. On a Hilbert space U , for linear operators $M \geq M' \in \mathbb{L}(U; U)$, a set $A \subset U$, and a point $\hat{u} \in U$, we write $(M, M') \in \mathcal{P}(A, \hat{u})$ if there exists a neighbourhood $\mathcal{U}' \ni \hat{u}$ such that each $u \in \mathcal{U}'$ has a common projection u^* to the set A with respect to both of the norms $\|\cdot\|_M$ and $\|\cdot\|_{M'}$:

$$\arg \min_{u' \in A} \|u - u'\|_M = u^* = \arg \min_{u' \in A} \|u - u'\|_{M'}.$$

Example 3.9. (Unique solution on a subspace) Let $H : U \rightrightarrows U$ with $U = X \times Y$. Write $u = (x, y)$. Take $M \geq \Xi := \begin{pmatrix} \alpha I & 0 \\ 0 & 0 \end{pmatrix}$ for some $\alpha > 0$. If $u^* = (x^*, y^*) \in H^{-1}(0)$ implies $x^* = \hat{x}$ for some fixed $\hat{x}$, then $(M, M - \Xi) \in \mathcal{P}(H^{-1}(0), \hat{u})$. $\square$

With this, we can show that strong submonotonicity implies partial strong submonotonicity for a large class of M .

Proposition 3.10. *On a Hilbert space U , let $T : U \rightrightarrows U$, and $\Xi, N, M \in \mathbb{L}(U; U)$ with $M \geq \Xi \geq 0$.*

If $(\Xi, M - \Xi) \in \mathcal{P}(T^{-1}(\hat{w}), \hat{u})$, then (N, Ξ) -strong submonotonicity at the point $(\hat{u}, \hat{w}) \in \text{graph } T$ implies (Ξ, N, M) -partial strong submonotonicity at the same point. If $T^{-1}(\hat{w}) = \{\hat{u}\}$ is a singleton, these two properties are equivalent; moreover $(\Xi, M - \Xi) \in \mathcal{P}(T^{-1}(\hat{w}), \hat{u})$ automatically holds.

Proof. The condition (PSM) for (N, Ξ) -strong submonotonicity reads

$$\inf_{u^* \in T^{-1}(\hat{w})} \langle w - \hat{w}, u - u^* \rangle_N \geq \text{dist}_{\Xi}^2(u, T^{-1}(\hat{w})) \quad (u \in \mathcal{U}, w \in T(u)). \quad (10)$$

Since $(\Xi, M - \Xi) \in \mathcal{P}(T^{-1}(\hat{w}), \hat{u})$, we have

$$\text{dist}_{\Xi}^2(u, T^{-1}(\hat{w})) + \text{dist}_{M-\Xi}^2(u, T^{-1}(\hat{w})) \geq \text{dist}_M^2(u, T^{-1}(\hat{w})) \quad (u \in \mathcal{U}').$$

Adding $\text{dist}_{M-\Xi}^2(u, T^{-1}(\widehat{w}))$ on both sides of (10), we therefore obtain

$$\inf_{u^* \in T^{-1}(\widehat{w})} \langle w - \widehat{w}, u - u^* \rangle_N + \text{dist}_{M-\Xi}^2(u, T^{-1}(\widehat{w})) \geq \text{dist}_M^2(u, T^{-1}(\widehat{w}))$$

for $u \in \mathcal{U} \cap \mathcal{U}'$ and $w \in T(u)$. Using $\inf + \inf \leq \inf$ on the left hand side, we (Ξ, N, M) -partial strong submonotonicity following the defining (PSM).

If $T^{-1}(\widehat{w}) = \{\widehat{u}\}$, retracing the steps above using the uniqueness of $u^* = \widehat{u} \in T^{-1}(\widehat{w})$ establishes the equivalence claim. $\square$

The next example shows that if $T^{-1}(\widehat{w})$ is not a singleton, the converse implication may not hold.

Example 3.11. Let $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(u) := \frac{1}{2} \min\{|u + 1|^2, |u - 1|^2\}$. Then $\partial_C f$ is $(\gamma I, I, I)$ -partially strongly submonotone for any $\gamma \in (0, 1)$ at $(\widehat{u}, 0)$ when $\widehat{u} \in \{1, -1\}$, but is not $(I, \gamma I)$ -strongly submonotone.

Proof. Clearly $[\partial f_C]^{-1}(0) = \{-1, 0, 1\}$. We take for simplicity $\widehat{u} = 1$ and $\mathcal{U} \subset (0, \infty)$. The case $\widehat{u} = -1$ is similar. Then $\partial_C f(u) = \{u - 1\}$ for $u \in \mathcal{U}$. For the $(\gamma I, I, I)$ -partial strong submonotonicity, (PSM) consequently expands as

$$\min_{u^* \in \{-1, 0, 1\}} ((u - 1)(u - u^*) + (1 - \gamma)|u - u^*|^2) \geq \min\{|u - 1|^2, |u|^2\} \quad (u \in \mathcal{U}).$$

For $u^* = 1$ the inequality is obvious. For the rest, by Young's inequality, it suffices if

$$\min_{u^* \in \{-1, 0\}} (1 - \gamma - \alpha^{-1})(u - u^*)^2 \geq (1 + \alpha/4)(u - 1)^2 \quad (u \in \mathcal{U})$$

for some $\alpha > 0$. Taking $\gamma \in (0, 1)$, $\alpha > 1/(1 - \gamma)$, and a correspondingly small neighbourhood $\mathcal{U}$ around $\widehat{u} = 1$, we prove the partial strong submonotonicity.

For the $(\gamma I, I)$ -strong submonotonicity we would need

$$\min_{u^* \in \{-1, 0, 1\}} (u - 1)(u - u^*) \geq \gamma \min\{|u - 1|^2, |u|^2\} \quad (u \in \mathcal{U}).$$

The left-hand side is negative for $u < 1$, so no neighbourhood $\mathcal{U}$ of $\widehat{u} = 1$ satisfies the condition. $\square$

3.4. Examples: the basic proximal point method

We next provide examples regarding the basic proximal point method. We return to primal-dual methods in Section 6. We first consider the non-operator case of partial strong submonotonicity, and afterwards use the full power of the operator formulation of partial strong submonotonicity to study convergence on subspaces.

Example 3.12. (Basic proximal point method, submonotonicity) Suppose for some $m > \gamma > 0$ that $H : U \rightrightarrows U$ is $(\gamma I, 2I, mI)$ -partially strongly submonotone at $(\widehat{u}, 0) \in \text{graph } H$ in the neighbourhood $\mathcal{U}$. Let $\{u^{i+1}\}_{i \in \mathbb{N}}$ be generated by the basic proximal point method

$$0 \in H(u^{i+1}) + \tau^{-1}(u^{i+1} - u^i) \tag{11}$$

for some $u^0 \in U$ and a step length parameter $\tau \geq 1/(m - \gamma)$. If $\{u^i\}_{i \geq n} \in \mathcal{U}$ for large enough n , then $\text{dist}^2(u^i; H^{-1}(0)) \rightarrow 0$ at a linear rate.

Proof. We take $M_{i+1} \equiv \tau^{-1}I$, $\Xi_{i+1} := \gamma I$, and $Z_{i+1} := \phi_i I$ for $\phi_{i+1} := \phi_i(1 + \gamma\tau)$, $\phi_0 := 1$. Then (PP) describes (11) and (9) holds as an equality. Moreover, the $(Z_{i+1}\Xi_{i+1}, 2Z_{i+1}, Z_{i+2}M_{i+2})$ -partial strong submonotonicity required by Theorem 3.5 is by Proposition 3.7(i) equivalent to $(\gamma I, 2I, (\tau^{-1} + \gamma)I)$ -partial strong submonotonicity. By Proposition 3.7(ii) this follows from the assumed $(\gamma I, 2I, mI)$ -partial strong submonotonicity when $\tau^{-1} + \gamma \geq m$, that is, $\tau \geq 1/(m - \gamma)$. Thus by Theorem 3.5, $\text{dist}^2(u^N; H^{-1}(0)) \leq (\phi_0/\phi_N) \text{dist}^2(u^0; H^{-1}(0))$ for large enough N . This yields the claimed convergence rate. $\square$

Remark 3.13. (Superlinear convergence) With iteration-dependent step lengths $\tau_k \rightarrow \infty$, it is also possible to obtain superlinear convergence. Under basic strong monotonicity this is proved in [41]; in the “testing” framework, which is also our overall approach to convergence proofs, in [43, Example 2.2] or [12, Theorem 10.1].

Remark 3.14. (Reaching the local neighbourhood) If U is finite-dimensional, the condition $\{u^i\}_{i \geq n} \in \mathcal{U}$ can be guaranteed by standard convergence results [32] for large $n \in \mathbb{N}$ and some $\hat{u} \in H^{-1}(0)$ with the corresponding neighbourhood $\mathcal{U}$. $\square$

A basic application of Example 3.12 is to $H = \partial G$ for a convex function G . As basic building blocks, we next show on $\mathbb{R}$ that the subdifferentials of the indicator of the unit ball, and of the absolute value function are strongly submonotone. None of these subdifferentials are *strongly* monotone in the conventional sense. With $(\hat{x}, \hat{q}) \in \text{graph } \partial G$, we will find a neighbourhood $\mathcal{U} \ni \hat{u}$ such that we have the $(I, \gamma I)$ -strong submonotonicity

$$\inf_{x^* \in [\partial G]^{-1}(\hat{q})} \langle q - \hat{q}, x - x^* \rangle \geq \gamma \inf_{x^* \in [\partial G]^{-1}(\hat{q})} \|x - x^*\|^2 \quad (x \in \mathcal{U}, q \in \partial G(x)). \quad (12)$$

We recall from Proposition 3.10 that $(I, \gamma I)$ -strong submonotonicity, i.e., $(\gamma I, I, \gamma I)$ -partial strong submonotonicity, implies $(\gamma I, I, I)$ -partial strong submonotonicity, and by Proposition 3.7, further $(2\gamma I, 2I, 2mI)$ -partial strong submonotonicity for any $m > 1$.

Lemma 3.15. *Let $G := \delta_{\text{cl } \mathbb{B}(0, \alpha)}$ on $\mathbb{R}^n$. Then ∂G is $(I, \gamma I)$ -strongly submonotone at $(\hat{x}, \hat{q}) \in \text{graph } \partial G$ with*

$$\mathcal{U} := \text{Dom } G \quad \text{and} \quad \gamma := \begin{cases} \|\hat{q}\|/(2\alpha), & \hat{q} \neq 0, \\ \infty, & \hat{q} = 0. \end{cases}$$

Proof. We need to prove (12). If $\hat{q} = 0$, then $[\partial G]^{-1}(\hat{q}) = \text{cl } \mathbb{B}(0, \alpha)$, so (12) trivially holds by the monotonicity of ∂G as a convex subdifferential [40].

Otherwise, if $\hat{q} \neq 0$, necessarily $\hat{q} = \beta \hat{x}$ for some $\beta > 0$, and $\|\hat{x}\| = \alpha$. Moreover, $[\partial G]^{-1}(\hat{q}) = \{\hat{x}\}$, and $\langle -q, \hat{x} - x \rangle \geq G(x) - G(\hat{x}) = 0$. Therefore (12) holds if $\beta \langle \hat{x}, \hat{x} - x \rangle \geq \gamma \|x - \hat{x}\|^2$. In other words $(\beta - \gamma)\|\hat{x}\|^2 \geq \gamma \|x\|^2 + (\beta - 2\gamma)\langle \hat{x}, x \rangle$.

Clearly $x \in \text{Dom } \partial G$ implies $\|x\| \leq \alpha$. Since $\|\hat{x}\| = \alpha$, this condition holds for $\beta \geq 2\gamma$. Since $\hat{q} = \beta \hat{x}$ and $\|\hat{x}\| = \alpha$, the maximal choice is $\gamma = \|\hat{q}\|/(2\alpha)$. $\square$

Lemma 3.16. *Let $G := |\cdot|$ on $\mathbb{R}$. Then ∂G is $(I, \gamma I)$ -strongly submonotone at $(\hat{x}, \hat{q}) \in \text{graph } \partial G$ for any $\gamma > 0$ in the neighbourhood*

$$\mathcal{U} := \begin{cases} [-2\gamma^{-1}, \infty), & \hat{q} = 1 \\ (-\infty, 2\gamma^{-1}], & \hat{q} = -1, \\ ([-1, 1] - \hat{q})\gamma^{-1}, & |\hat{q}| < 1. \end{cases}$$

Proof. We need to prove (12). We have $\hat{q} \in [-1, 1]$. Consider first $\hat{q} = 1$. Then $[\partial G]^{-1}(\hat{q}) = [0, \infty)$, so (12) reduces to

$$\inf_{x^* \geq 0, q \in \partial G(x)} (q - 1)(x - x^*) \geq \gamma \inf_{x^* \geq 0} |x - x^*|^2$$

If $x \geq 0$, this holds for any $\gamma \geq 0$ by the monotonicity of ∂G . Otherwise, if $x < 0$, we have $q = -1$, so need $-2x \geq \gamma|x|^2$, which, because $x < 0$, is to say $2 \geq \gamma|x|$. This is guaranteed by our choice of $\mathcal{U}$.

The case $\hat{q} = -1$ is analogous.

Consider then $|\hat{q}| < 1$. Then $[\partial G]^{-1}(\hat{q}) = \{\hat{x}\} = \{0\}$, so (12) holds if

$$\inf_{q \in \partial G(x)} (q - \hat{q})x \geq \gamma x^2. \quad (13)$$

If $x = 0$, this is clear. If $x > 0$, $\partial G(x) = \{1\}$, so (13) holds if $1 - \hat{q} \geq \gamma x$. This holds if $x \leq (1 - \hat{q})/\gamma$. Similarly, if $x < 0$, we obtain for (13) condition $-1 - \hat{q} \leq \gamma x$. This holds when $x \geq (-1 - \hat{q})/\gamma$. The conditions $x \leq (1 - \hat{q})/\gamma$ and $-1 - \hat{q} \leq \gamma x$ give the expression for $\mathcal{U}$ in the statement of the lemma. $\square$

Example 3.17. Let $F(x, y) := |x| + \delta_{[-1, 1]}(y)$. Then ∂F is $(I, \gamma I)$ -strongly submonotone at $(\hat{x}, \hat{y}), 0$ for any $(\hat{x}, \hat{y}) \in \{0\} \times [-1, 1]$ and $\gamma > 0$. Consequently the iterates of the proximal point method of Example 3.12 for $H = \partial F$ and any $\tau > 0$ converge linearly to $\{0\} \times [-1, 1]$

Proof. Clearly $\{0\} \times [-1, 1] = H^{-1}(0)$. We combine Lemmas 3.16 and 3.15 to show the $(I, \gamma I)$ -strong submonotonicity. By Propositions 3.7 and 3.10 we then obtain $(2\gamma I, 2I, 2mI)$ -partial strong submonotonicity for any $m > 1$. By taking $m > \gamma$ large enough that $\tau \geq 1/(2m - 2\gamma)$, we obtain the claim from Example 3.12 provided the iterates are in the neighbourhood of submonotonicity $\mathcal{U}$. By Remark 3.14 the convergence is, in fact, global. $\square$

3.5. Examples: varying behaviour on subspaces

The next examples demonstrate the word “partial” in the definitions.

Example 3.18. (Partial convergence of a preconditioned proximal point method) Suppose $H : X \times Y \rightrightarrows X \times Y$ is $(\Xi, 2I, M)$ -partially strongly submonotone at $(\hat{u}, 0) \in \text{graph } H$ in the (fixed) neighbourhood $\mathcal{U}$ with $\Xi = \text{diag}(\gamma I, 0)$ for some fixed $\gamma > 0$ and $M = \text{diag}(m_x I, m_y I)$ for any $m_x > \gamma$ and $m_y > 0$. For any $\tau, \sigma_0 > 0$

and initial $u^0 = (x^0, y^0) \in X \times Y$, let $\{u^{i+1} = (x^{i+1}, y^{i+1})\}_{i \in \mathbb{N}}$ be generated by the preconditioned proximal point method

$$0 \in H(u^{i+1}) + M_{i+1}(u^{i+1} - u^i)$$

where $M_{i+1} = \begin{pmatrix} \tau^{-1}I & 0 \\ 0 & \sigma_i^{-1}I \end{pmatrix}$ and $\sigma_{i+1} := (1 + \gamma\tau)\sigma_i$.

If $\{u^i\}_{i \geq n} \in \mathcal{U}$ for large enough n , then $\text{dist}(x^i, \{\hat{x} \mid (\hat{x}, \hat{y}) \in H^{-1}(0)\}) \rightarrow 0$ at a linear rate.

Proof. We take $Z_{i+1} = \phi_i I$ and $\Xi_{i+1} = \text{diag}(\gamma I, 0)$ for $\phi_{i+1} := \phi_i(1 + \gamma\tau)$ and $\phi_0 := 1$. Then $\sigma_{i+1} = \sigma_i \phi_{i+1} \phi_i^{-1}$. We use Theorem 3.5. The condition (9) holds with our choices. Moreover,

$$Z_{i+2}M_{i+2} = \begin{pmatrix} \phi_i(\tau^{-1} + \gamma)I & 0 \\ 0 & \phi_i\sigma_i^{-1}I \end{pmatrix} \quad \text{and} \quad Z_{i+1}\Xi_{i+1} = \begin{pmatrix} \phi_i\gamma I & 0 \\ 0 & 0 \end{pmatrix}.$$

Therefore the required $(Z_{i+1}\Xi_{i+1}, 2Z_{i+1}, Z_{i+2}M_{i+2})$ -partial strong submonotonicity amounts to $(\text{diag}(\phi_i\gamma I, 0), 2\phi_i I, \text{diag}(\phi_i(\tau^{-1} + \gamma)I, \phi_i\sigma_i^{-1}))$ -partial strong submonotonicity!

By the scaling invariance Proposition 3.7, this is implied by our assumptions with $m_x = \tau^{-1} + \gamma$ and $m_y = \sigma_i^{-1}$. Now $\{\phi_i\}_{i \in \mathbb{N}}$ grows exponentially, so (5) established by Theorem 3.5 shows the claim. $\square$

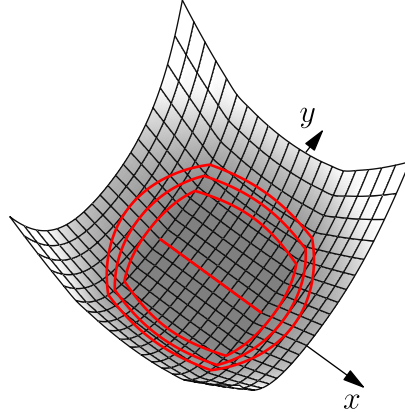


Figure 3.1: The function F from Example 3.19 with some contour lines plotted. The function is submonotone at any $((\hat{x}, 0), 0) \in \text{graph } \partial F$ strongly in the x direction, but non-strongly in the y direction.

Example 3.19. (Varying behaviour along subspaces) With $u = (x, y) \in \mathbb{R}^2$ and $\mu > 0$, let

$$F(u) := G(x) + \frac{1}{2}(x^2 + y^2)y^2 \quad \text{with} \quad G(x) := \begin{cases} 0, & |x| \leq \mu, \\ x - \mu, & x > \mu, \\ x + \mu, & x < -\mu. \end{cases}$$

This is illustrated in Figure 3.1. Let $\gamma > 0$. Then ∂F is $(\Xi, 2I, M)$ -partially strongly submonotone at any $((\hat{x}, 0), 0) \in \text{graph } \partial f$ in the neighbourhood $\mathcal{U} = (\mu + 2\gamma^{-1})[-1, 1] \times \mathbb{R}$ with $M = \text{diag}(m_x I, m_y I)$ and $\Xi = \text{diag}(\gamma I, 0)$ for any $m_x \geq \gamma$ and $m_y \geq 0$. $\square$

Proof. Let $Q(u) := \frac{1}{2}(x^2 + y^2)y^2$. The function Q is convex as verified by the positivity of the Hessian, hence is ∇Q monotone. Clearly $[\partial F]^{-1}(0) = [-\mu, \mu] \times \{0\}$ with $\nabla Q(u^*) = 0$ for $u^* \in [\partial F]^{-1}(0)$. We use the notation $u = (x, y)$ and $u^* = (x^*, y^*)$. We have $\partial F(u) = \{(w, 0) + \nabla Q(u) \mid w \in \partial G(x)\}$ and, by the monotonicity of ∇Q ,

$$\begin{aligned} \inf_{u^* \in [\partial F]^{-1}(0)} (\langle (w, 0) + \nabla Q(u) - 0, u - u^* \rangle_{2I} + \|u - u^*\|_{M-\Xi}^2) \\ \geq \inf_{u^* \in [\partial F]^{-1}(0)} (2w(x - x^*) + \|u - u^*\|_{M-\Xi}^2). \end{aligned}$$

Consequently, the claimed partial strong submonotonicity at $((\hat{x}, 0), 0) \in \text{graph } \partial F$ holds if

$$\min_{x^* \in [-\mu, \mu]} 2w(x - x^*) + (m_x - \gamma)(x - x^*)^2 + m_y y^2 \geq \min_{x^{**} \in [-\mu, \mu]} m_x(x - x^{**})^2 + m_y y^2 \quad (14)$$

for all $u = (x, y)$ in a neighbourhood $\mathcal{U}$ of $\hat{u} = (\hat{x}, 0)$, and all

$$w \in \partial G(x) = \begin{cases} 0, & |x| < \mu, \\ 1, & x > \mu, \\ [0, 1], & x = \mu, \\ -1, & x < -\mu, \\ [-1, 0], & x = -\mu. \end{cases}$$

If $x \geq \mu$, then $w \geq 1$ and $x > x^*$. With $m_x \geq \gamma$, clearly both sides of the inequality in (14) have the same minimiser $x^* = x^{**} = \mu$. So (14) holds if $2(x - \mu) \geq \gamma(x - \mu)^2$, which is to say with $x \leq \mu + 2\gamma^{-1}$.

The case $x \leq -\mu$ is analogous.

If $|x| < \mu$, we have $w = 0$, so with $m_x \geq \gamma$, both sides of (14) are minimised and are equal with $x^{**} = x^* = x$. $\square$

4. Rates from error bounds and subregularity

We now study an approach alternative to submonotonicity: the error bounds that we discussed in the introduction. Their essence is to prove for some $\kappa > 0$ that

$$\kappa \|u^{i+1} - u^i\| \geq \|u^{i+1} - \hat{u}\|. \quad (\text{EB})$$

One can see how this would improve (4) by allowing $Z_{i+1}M_{i+2}$ to grow faster. However, we generally cannot fix $\hat{u}$, so would take the infimum over $\hat{u} \in H^{-1}(0)$ above. We also want a *partial*, operator-relative, variant, to deal with primal-dual methods, and even block-adapted methods [44, 33].

4.1. Partial error bounds and subregularity

We generalise error bounds to be operator-relative:

Theorem 4.1. *On a Hilbert space U , let $H : U \rightrightarrows U$, and $M_{i+1}, Z_{i+1}, P_{i+1} \in \mathbb{L}(U; U)$ with $Z_{i+1}M_{i+1} \geq 0$ self-adjoint for all $i \in \mathbb{N}$. Suppose $H^{-1}(0)$ is non-empty and (PP) is solvable for the iterates $\{u^{i+1}\}_{i \in \mathbb{N}}$ given $u^0 \in U$. If for some $\delta \in [0, 1]$ the partial error bound*

$$\begin{aligned} \delta \|u^{i+1} - u^i\|_{Z_{i+1}M_{i+1}}^2 + \text{dist}_{Z_{i+2}M_{i+2}-Z_{i+1}P_{i+1}}^2(u^{i+1}, H^{-1}(0)) \\ \geq \text{dist}_{Z_{i+2}M_{i+2}}^2(u^{i+1}, H^{-1}(0)) \end{aligned} \quad (\text{PEB})$$

holds and

$$\begin{aligned} \frac{1-\delta}{2} \|u^{i+1} - u^i\|_{Z_{i+1}M_{i+1}}^2 + \frac{1}{2} \|u^{i+1} - u^*\|_{Z_{i+1}(M_{i+1}+P_{i+1})-Z_{i+2}M_{i+2}}^2 \\ + \langle H(u^{i+1}), u^{i+1} - u^* \rangle_{Z_{i+1}} \geq 0 \end{aligned} \quad (15)$$

for all $i \in \mathbb{N}$ and $u^* \in H^{-1}(0)$, then (5) holds.

Proof. By (15) for all $u^* \in H^{-1}(0)$ it holds

$$\begin{aligned} \frac{\delta}{2} \|u^{i+1} - u^i\|_{Z_{i+1}M_{i+1}}^2 + \frac{1}{2} \|u^{i+1} - u^*\|_{Z_{i+2}M_{i+2}-Z_{i+1}P_{i+1}}^2 \\ \geq \frac{1}{2} \text{dist}_{Z_{i+2}M_{i+2}}^2(u^{i+1}; H^{-1}(0)). \end{aligned}$$

Summing this with (15), and taking the infimum over $u^* \in H^{-1}(0)$, we obtain (4). Then we just apply Lemma 3.1. $\square$

An essential ingredient in proving the basic error bound (EB) is the *metric subregularity of H at $\hat{u}$ for $\hat{w} = 0$* : the existence of a neighbourhood $\mathcal{U} \ni \hat{u}$ and $\kappa > 0$ such that

$$\kappa \text{dist}(\hat{w}, H(u)) \geq \text{dist}(u, H^{-1}(\hat{w})) \quad (u \in \mathcal{U}). \quad (16)$$

We refer to [18, 25, 26, 13, 36] for more on error bounds and metric subregularity. To prove (PEB) we need a partial version.

Definition 4.2. Let U, W be Hilbert spaces. Also let $M, P \in \mathbb{L}(U; U)$ and $N \in \mathbb{L}(W; W)$ with $N \geq 0$, $M \geq 0$, and $M \geq P$. We say that $T : U \rightrightarrows W$ is (P, N, M) -*partially subregular* at $(\hat{u}, \hat{w}) \in \text{graph } T$ if there exists a neighbourhood $\mathcal{U} \ni \hat{u}$ such that

$$\text{dist}_N^2(\hat{w}, T(u)) + \text{dist}_{M-P}^2(u, T^{-1}(\hat{w})) \geq \text{dist}_M^2(u, T^{-1}(\hat{w})) \quad (u \in \mathcal{U}). \quad (\text{PSR})$$

We say that T is (N, M) -*subregular* if $P = M$.

Lemma 4.3. *Suppose $Z_{i+1}M_{i+1} \geq 0$ is self-adjoint and positive definite. Let u^{i+1} be generated by (PP) given u^i . Then*

$$\|u^{i+1} - u^i\|_{Z_{i+1}M_{i+1}}^2 \geq \text{dist}_{Z_{i+1}(Z_{i+1}M_{i+1})^{-1}Z_{i+1}}^2(0, H(u^{i+1})). \quad (17)$$

Proof. Let $q^{i+1} := -M_{i+1}(u^{i+1} - u^i)$. Then $q^{i+1} \in H(u^{i+1})$.

By applying $\langle \cdot, u^{i+1} - u^i \rangle_{Z_{i+1}}$ to (PP), we therefore obtain

$$\|u^{i+1} - u^i\|_{Z_{i+1}M_{i+1}}^2 = -\langle q^{i+1}, u^{i+1} - u^i \rangle_{Z_{i+1}}.$$

By our assumptions $Z_{i+1}M_{i+1}$ is invertible. Therefore we can solve

$$u^{i+1} - u^i = -(Z_{i+1}M_{i+1})^{-1}Z_{i+1}q^{i+1}.$$

It follows $\|u^{i+1} - u^i\|_{Z_{i+1}M_{i+1}}^2 = \|q^{i+1}\|_{Z_{i+1}^*(Z_{i+1}M_{i+1})^{-1}Z_{i+1}}^2$.

This immediately yields the claim. $\square$

As a consequence we obtain:

Lemma 4.4. *Suppose that $Z_{i+1}M_{i+1}$ is self-adjoint and positive definite. Take $P_{i+1} \in \mathbb{L}(U; U)$ and $\delta \in [0, 1]$. Then the partial error bound (PEB) holds if H is*

$$(Z_{i+1}P_{i+1}, \delta Z_{i+1}^*(Z_{i+1}M_{i+1})^{-1}Z_{i+1}, Z_{i+2}M_{i+2})\text{-partially subregular}$$

at some $\hat{u} \in H^{-1}(0)$ in a neighbourhood $\mathcal{U}$ containing u^{i+1} .

Proof. The condition (PSR) now reads

$$\begin{aligned} & \text{dist}_{\delta Z_{i+1}^*(Z_{i+1}M_{i+1})^{-1}Z_{i+1}}^2(0, H(u)) + \text{dist}_{Z_{i+2}M_{i+2} - Z_{i+1}P_{i+1}}^2(u, H^{-1}(0)) \\ & \geq \text{dist}_{Z_{i+2}M_{i+2}}^2(u, H^{-1}(0)) \quad (u \in \mathcal{U}). \end{aligned}$$

Taking $u = u^{i+1}$ and combining with (17) from Lemma 4.3 we obtain (PEB). $\square$

Corollary 4.5. *On a Hilbert space U , let $H : U \rightrightarrows U$, and $M_{i+1}, Z_{i+1}, P_{i+1} \in \mathbb{L}(U; U)$ with $Z_{i+1}M_{i+1}$ self-adjoint for all $i \in \mathbb{N}$. Suppose $H^{-1}(0)$ is non-empty and that (PP) is solvable for the iterates $\{u^{i+1}\}_{i \in \mathbb{N}}$ given $u^0 \in U$. Pick $\delta \in [0, 1]$. If H is $(Z_{i+1}P_{i+1}, \delta Z_{i+1}^*(Z_{i+1}M_{i+1})^{-1}Z_{i+1}, Z_{i+2}M_{i+2})$ -partially subregular at some $(\hat{u}, 0) \in \text{graph } H$ and (15) holds (in particular, if $Z_{i+1}H$ is monotone and*

$$Z_{i+1}(M_{i+1} + P_{i+1}) \geq Z_{i+2}M_{i+2} \quad (i \in \mathbb{N}), \quad (18)$$

then (5) holds provided $\{u^i\}_{i=0}^N \subset \mathcal{U}$ for the neighbourhood $\mathcal{U}$ of partial subregularity.

Proof. We use Lemma 4.4 with Theorem 4.1. $\square$

Example 4.6. (Basic proximal point method, subregularity) Suppose $H : U \rightrightarrows U$ is monotone and $(\pi I, \delta\tau^2 I, (1 + \pi)I)$ -partially subregular at $(\hat{u}, 0) \in \text{graph } H$ in the neighbourhood $\mathcal{U}$ for some $\tau > 0$, $\pi \geq 0$, and $\delta \in [0, 1]$. Let $\{u^{i+1}\}_{i \in \mathbb{N}}$ be generated by the basic proximal point method of Example 3.12 for some initial $u^0 \in U$. If $\{u^i\}_{i \geq N} \subset \mathcal{U}$ for large enough $N \in \mathbb{N}$, then $\text{dist}^2(u^i; H^{-1}(0)) \rightarrow 0$ at a linear rate.

Proof. We take $M_{i+1} := \tau^{-1}I$, $P_{i+1} := \pi I$, and $Z_{i+1} := \phi_i I$ for $\phi_{i+1} := \phi_i(1 + \pi)$ and $\phi_0 := 1$. Then $Z_{i+1}(M_{i+1} + P_{i+1}) = Z_{i+2}M_{i+2}$. Combined with the monotonicity of H , this verifies (15). Observe then that $\delta Z_{i+1}^*(Z_{i+1}M_{i+1})^{-1}Z_{i+1} = \delta\tau^2\phi_i I$. By scaling invariance (detailed in the next Proposition 4.7) we thus verify that $(\pi I, \delta\tau^2 I, (1 + \pi)I)$ -partial subregularity yields the subregularity demanded by Lemma 4.4. Consequently we obtain the claim from Corollary 4.5. $\square$

We do not provide further examples in this instance as they arise from those of Sections 3.4 and 3.5 combined with the relationships between submonotonicity and subregularity that we study in the next section. They depend on the following transformation results.

4.2. Transformation of subregularity

The analogues of Propositions 3.7 and 3.10 hold for subregularity with virtually identical proofs.

Proposition 4.7. *Suppose $T : U \rightrightarrows U$ is (P, N, M) -partially subregular at $(\hat{u}, \hat{w}) \in \text{graph } T$ in a neighbourhood $\mathcal{U}$. Then for any $\alpha > 0$, at the same point $(\hat{u}, \hat{w})$ in the same neighbourhood $\mathcal{U}$,*

- (i) *T is $(\alpha P, \alpha N, \alpha M)$ -partially subregular.*
- (ii) *T is $(P, N, (1 + \alpha)M)$ -partially subregular.*

Proposition 4.8. *Let $T : U \rightrightarrows U$, and $P, N, M \in \mathbb{L}(U; U)$ with $M \geq P \geq 0$.*

If $(P, M - P) \in \mathcal{P}(T^{-1}(\hat{w}), \hat{u})$, then (N, P) -subregularity at $(\hat{u}, \hat{w}) \in \text{graph } T$ implies (P, N, M) -partial subregularity at the same point.

If $T^{-1}(\hat{w}) = \{\hat{u}\}$ is a singleton, these two properties are equivalent.

The next lemma helps scale some of the factors of partial subregularity in more general cases.

Lemma 4.9. *Let $T : U \rightrightarrows U$, and $P, N, M \in \mathbb{L}(U; U)$ with $M \geq 0$, and $M \geq P$. Also pick $\alpha \geq 0$, and, if $\alpha \in (0, 1)$, suppose that $(M, M - P) \in \mathcal{P}(T^{-1}(\hat{w}), \hat{u})$. Then T is (P, N, M) -partially subregular at $(\hat{u}, \hat{w}) \in \text{graph } T$ if it is $(\alpha P, \alpha N, M)$ -partially subregular at this point.*

Conversely, if $(M, M - \alpha P) \in \mathcal{P}(T^{-1}(\hat{w}), \hat{u})$ when $\alpha > 1$, then T is $(\alpha P, \alpha N, M)$ -partially subregular at $(\hat{u}, \hat{w}) \in \text{graph } T$ if it is (P, N, M) -partially subregular at this point.

In particular, if $(M, M - \max\{\alpha, 1\}P) \in \mathcal{P}(T^{-1}(\hat{w}), \hat{u})$, then the relationship is “if and only if”.

Proof. If T is $(\alpha P, \alpha N, M)$ -partially subregular, in some neighbourhood $\mathcal{U}$ of $\hat{u}$ for every $u^* \in T^{-1}(\hat{w})$ and $u \in \mathcal{U}$ it holds

$$\text{dist}_{\alpha N}^2(\hat{w}, T(u)) + \|u - u^*\|_{M - \alpha P}^2 \geq \text{dist}_M^2(u, T^{-1}(\hat{w})).$$

After multiplying by α^{-1} , we rearrange this as

$$\text{dist}_N^2(\hat{w}, T(u)) + \|u - u^*\|_{M - P}^2 \geq \text{dist}_{\alpha^{-1}M}^2(u, T^{-1}(\hat{w})) + \|u - u^*\|_{M - \alpha^{-1}M}^2.$$

If $\alpha < 1$, we now let u^* be the common projection of u to the closed set $T^{-1}(\hat{w})$ in the norms $\|\cdot\|_M$ and $\|\cdot\|_{M - P}$. Otherwise, if $\alpha \geq 1$, we just take the infimum over $u^* \in T^{-1}(\hat{w})$ on both sides. This proves (P, N, M) -partial subregularity.

For the converse implication, we have to prove the relationship in the other direction. This amounts to applying the first claim with α replaced by α^{-1} to

$$(P', N', M) = (\alpha P, \alpha N, M).$$

If $\alpha \in (0, 1]$, the projection condition is not required in this direction as $\alpha^{-1} \geq 1$.

If $\alpha > 1$ then we need the projection to be the same with respect to M and $M - P' = M - \alpha P$. This gives the projection condition in the converse claim.

Finally, the “if and only if” claim just combines the two implications. $\square$

5. Relationships between subregularity and submonotonicity

Having introduced the distinct concepts of partial strong submonotonicity and partial subregularity motivated by algorithmic needs, we now study some basic theoretical relationships between these concepts. Throughout, we assume that U and W are Hilbert spaces.

5.1. General implications between submonotonicity and subregularity

The following lemmas generalise [3, Theorem 3.3] from convex subdifferentials to general set-valued maps and partial subregularity.

Lemma 5.1. *With U and W Hilbert spaces, let $T : U \rightrightarrows U$, and $\Xi, N, M \in \mathbb{L}(U; U)$ with $M \geq 0$. Pick $\alpha > 0$. Suppose the following structural conditions hold:*

- (i) $N = A^*B$ for some $A, B \in \mathbb{L}(W; U)$,
- (ii) $\Xi_\alpha := \alpha^{-1}\Xi - \alpha^{-2}A^*A/4 \leq \min\{1, \alpha^{-1}\}M$, and
- (iii) $(M, M - \Xi_\alpha) \in \mathcal{P}(T^{-1}(\hat{w}), \hat{u})$ when $\alpha \in (0, 1)$.

*Then T is (Ξ_α, B^*B, M) -partially subregular at $(\hat{u}, \hat{w}) \in \text{graph } T$ if it is (Ξ, N, M) -partially strongly submonotone at this point.*

*If (iii) does not hold, we still have $(\alpha\Xi_\alpha, \alpha B^*B, M)$ -partial subregularity.*

Proof. By partial strong submonotonicity, for some neighbourhood $\mathcal{U}$ of $\hat{u}$, for all $u^* \in T^{-1}(\hat{w})$, $u \in \mathcal{U}$, and $w \in T(u)$ it holds

$$\langle w - \hat{w}, u - u^* \rangle_N + \|u - u^*\|_{M-\Xi}^2 \geq \text{dist}_M^2(u, T^{-1}(\hat{w})).$$

By Young's inequality and i, for any $\alpha > 0$ therefore

$$\|w - \hat{w}\|_{\alpha B^*B}^2 + \|u - u^*\|_{M-\Xi+A^*A/(4\alpha)}^2 \geq \text{dist}_M^2(u, T^{-1}(\hat{w})).$$

Since $\Xi - A^*A/(4\alpha) = \alpha\Xi_\alpha$, and $M \geq \alpha\Xi_\alpha$ by ii, we obtain $(\alpha\Xi_\alpha, \alpha B^*B, M)$ -partial subregularity after taking the infimum over $u^* \in T^{-1}(\hat{w})$ and $w \in T(u)$. By ii also $M \geq \Xi_\alpha$. Minding iii, we may therefore apply Lemma 4.9, which yields the claimed (Ξ_α, B^*B, M) -partial subregularity. $\square$

For the converse relationship, we essentially have to assume that T is a convex subdifferential.

Definition 5.2. Let $T : U \rightrightarrows U$, $N, \Gamma \in \mathbb{L}(U; U)$, and $(\hat{u}, \hat{w}) \in \text{graph } T$. We say that $\tilde{\mathcal{G}} : U \times U \rightarrow \overline{\mathbb{R}}$ is a (N, Γ) -gap function for T at $\hat{w}$ if for all $u^* \in T^{-1}(\hat{w})$ we have $\langle T(u), u - u^* \rangle_N \geq \tilde{\mathcal{G}}(u; u^*) + \|u - u^*\|_\Gamma^2$ with $\partial\tilde{\mathcal{G}}(\cdot; u^*) = NT$, and $\tilde{\mathcal{G}}(u^*; u^*) = 0$. $\square$

In the next result, we stress that $\mathcal{U}$ is open; otherwise in the present work, we do not strictly require this, and $\mathcal{U}$ could indeed be any subset of U .

Lemma 5.3. *Let $T : U \rightrightarrows U$ have closed values, and $\Xi, N, M, P \in \mathbb{L}(U; U)$ with $M \geq 0$, and $M \geq P$. Suppose the following structural conditions hold:*

- (i) $N = A^*B$ for some $A, B \in \mathbb{L}(U; U)$.
- (ii) A is invertible with $M \leq mA^*A$ for some $m > 0$.
- (iii) T admits for some $\Gamma \in \mathbb{L}(U; U)$ an (N, Γ) -gap function $\tilde{\mathcal{G}}$ at $\hat{w}$.

(iv) $M - \Xi + \Gamma$ is self-adjoint and positive semi-definite.

(v) $M \leq \alpha P$ for some $0 < \alpha < 1/(16m)$.

(vi) $(M, M - \Xi_\alpha) \in \mathcal{P}(T^{-1}(\hat{w}), \hat{u})$ when $\alpha > 1$.

Then T is (Ξ, N, M) -partially strongly submonotone at $(\hat{u}, \hat{w}) \in \text{graph } T$ provided it is (P, B^*B, M) -partially subregular at this point. In both properties, we require the neighbourhood $\mathcal{U}$ to be open.

We may replace (vi) and (P, B^*B, M) -partial subregularity by $(\alpha P, \alpha B^*B, M)$ -partial subregularity.

Remark 5.4. The condition (v) typically forces $P > 0$, so we require “full” subregularity in the sense of non-singularity of P . Together with the condition $M \geq P$, it may happen that only $M = P$ is possible. $\square$

The proof follows ideas from [3, 4], adding the necessary extra work for partial regularity.

Proof. Suppose, to reach a contradiction, that the claimed partial strong submonotonicity does not hold in any neighbourhood of $\hat{u}$. Then for any $r > 0$, we can find some $u^* \in T^{-1}(\hat{w})$ and

$$u' \in \tilde{\mathbb{B}}_M(\hat{u}, r) := \{u \in U \mid \|\hat{u} - u\|_M < r, \|\hat{u} - u\| < r\} \quad (19)$$

$$\text{with } \inf_{w \in T(u')} (\langle w - \hat{w}, u' - u^* \rangle_N + \|u' - u^*\|_{M-\Xi}^2) < \text{dist}_M^2(u', T^{-1}(\hat{w})).$$

$$\text{This implies } f(u') < \epsilon := \text{dist}_M^2(u', T^{-1}(\hat{w})) - \|u' - u^*\|_{M-\Xi+\Gamma}^2 \quad (20)$$

$$\text{for } f(u) := \tilde{\mathcal{G}}(u; u^*) - \langle \hat{w}, u - u^* \rangle_N. \quad (21)$$

Since $N\hat{w} \in \partial \tilde{\mathcal{G}}(u^*; u^*)$, clearly u^* is a minimiser of f . Moreover $f(u^*) = 0$ because $\tilde{\mathcal{G}}(u^*; u^*) = 0$. Thus $f \geq 0$ and $f(u') \leq f(\hat{u}) + \epsilon$. By Ekeland's variational principle [16], given $\lambda > 0$ (to be specified later), there now exists $v_\lambda \in U$ with

$$\|u' - v_\lambda\|_M \leq \lambda, \quad (22a)$$

$$f(v_\lambda) \leq f(u'), \quad \text{and} \quad (22b)$$

$$f(x) > f(v_\lambda) - (\epsilon/\lambda)\|x - v_\lambda\|_M \quad (x \neq v_\lambda). \quad (22c)$$

It follows $f(x) + (\epsilon/\lambda)\|x - v_\lambda\|_M > f(v_\lambda) + (\epsilon/\lambda)\|v_\lambda - v_\lambda\|_M \quad (x \neq v_\lambda)$.

Thus v_λ is a minimiser of the convex function $x \mapsto f(x) + (\epsilon/\lambda)\|x - v_\lambda\|_M$. The necessary first-order optimality conditions therefore state

$$N(T(v_\lambda) - \hat{w}) + (\epsilon/\lambda)Mq \ni 0 \quad \text{for some } \|q\| \leq 1.$$

Since $N = A^*B$, using ii we obtain

$$\text{dist}_{B^*B}^2(\hat{w}, T(v_\lambda)) \leq m(\epsilon/\lambda)^2. \quad (23)$$

Under the assumption that T is (P, B^*B, M) -partially subregular at $(\hat{u}, \hat{w}) \in \text{graph } T$, we obtain by (vi) and Lemma 4.9 corresponding $(\alpha P, \alpha B^*B, M)$ -partial subregularity. This implies

$$\text{dist}_M^2(v_\lambda, T^{-1}(\hat{w})) - \|v_\lambda - u^*\|_{M-\alpha P}^2 \leq \text{dist}_{\alpha B^*B}^2(\hat{w}, T(v_\lambda)), \quad (24)$$

provided $v_\lambda \in \mathcal{U}$ for $\mathcal{U}$ the neighbourhood of partial subregularity. We may without loss of generality assume that $\mathcal{U} = \tilde{\mathbb{B}}_M(\hat{u}, R)$ for some $R > 0$.

Recalling the definition of ϵ in (20), using (iv) and Young's inequality, we estimate

$$\begin{aligned} \epsilon &= \text{dist}_M^2(u', T^{-1}(\hat{w})) - \|u' - u^*\|_{M-\Xi+\Gamma}^2 \leq \text{dist}_M^2(u', T^{-1}(\hat{w})) \\ &\leq 2[\text{dist}_M^2(v_\lambda, T^{-1}(\hat{w})) - \|v_\lambda - u^*\|_{M-\alpha P}^2] + \|v_\lambda - u^*\|_{2(M-\alpha P)}^2 + \|u' - v_\lambda\|_{2M}^2. \end{aligned} \quad (25)$$

Further using (24), (23), (22a) and (v), we obtain

$$\epsilon \leq 2\alpha \text{dist}_{B^*B}^2(\hat{w}, T(v_\lambda)) + 2\lambda^2 \leq 2\alpha m(\epsilon/\lambda)^2 + 2\lambda^2. \quad (26)$$

By (20) and $f \geq 0$, necessarily $\epsilon > 0$. Choosing $\lambda := \sqrt{m\alpha\theta^{-1}\epsilon}$ for some $\theta > 0$, (26) therefore shows

$$1 \leq 2\theta + 2\alpha m\theta^{-1}. \quad (27)$$

We recall that for (24) to hold we still need to ensure $v_\lambda \in \mathcal{U} = \tilde{\mathbb{B}}(\hat{u}, R)$. By (19) and (22a), we have

$$\|v_\lambda - \hat{u}\|_M \leq \|v_\lambda - u'\|_M + \|u' - \hat{u}\|_M \leq \lambda + r.$$

Since $r > 0$ was arbitrary, it suffices to ensure $\lambda < R$. This will hold and we will contradict (27) if we choose θ to satisfy

$$1 - \sqrt{1 - 16\alpha m} < 4\theta < 1 + \sqrt{1 - 16\alpha m}, \quad \text{and} \quad m\alpha\epsilon < \theta R^2. \quad (28)$$

The expression under the square root is positive by (v). Therefore the first part of (28) holds for some $\theta > 0$. From the first line of (25), we have $\epsilon \leq \|u' - \hat{u}\|_M^2 \leq r^2$. Therefore, the second part of (28) can also be made to hold by taking $r > 0$ small enough. This gives the desired contradiction. $\square$

5.2. Concrete results in special cases

We now specialise the implications above to some simple cases.

Take $M = I$ and let Ξ be a projection to a subspace. If T is $(\Xi, \kappa I, I)$ -partially strongly submonotone, the next proposition yields $(2\Xi - I, \kappa^2 I, I)$ -partial subregularity, which reads

$$\kappa^2 \text{dist}_I^2(\hat{w}, T(u)) + 2 \text{dist}_{I-\Xi}^2(u, T^{-1}(\hat{w})) \geq \text{dist}_I^2(u, T^{-1}(\hat{w})) \quad (u \in \mathcal{U}).$$

If $\hat{u} \in T^{-1}(\hat{w})$ were unique, this would be the same as

$$\kappa^2 \text{dist}_I^2(\hat{w}, T(u)) + \text{dist}_{I-\Xi}^2(u, T^{-1}(\hat{w})) \geq \text{dist}_\Xi^2(u, T^{-1}(\hat{w})).$$

This is a stronger property than $(\Xi, \kappa^2 I, I)$ -partial subregularity, for which (PSR) in the case of unique $\hat{u}$ would read

$$\kappa^2 \text{dist}_I^2(\hat{w}, T(u)) \geq \text{dist}_\Xi^2(u, T^{-1}(\hat{w})).$$

Of course, if $\Xi = I$, the two properties are the same.

Proposition 5.5. *Let $T : U \rightrightarrows U$ have closed values, and $M, \Xi \in \mathbb{L}(U; U)$ with M positive definite and self-adjoint. Let $(\hat{u}, \hat{w}) \in \text{graph } T$, $\kappa > 0$.*

Then T is $(2\Xi - M, \kappa^2 M, M)$ -partially subregular if it is $(\Xi, \kappa M, M)$ -partially strongly submonotone, $\Xi \leq M$, and $(M, M - \Xi) \in \mathcal{P}(T^{-1}(\hat{w}), \hat{u})$.

Conversely, T is $(\Xi, \kappa M, M)$ -partially strongly submonotone if it is $(M, \rho M, M)$ -partially subregular for some $\rho > (\kappa/4)^2$ in an open neighbourhood $\mathcal{U} \ni \hat{u}$, and admits an (M, Γ) -gap function at $\hat{w}$ with $0 \leq M - \Xi + \Gamma$ self-adjoint.

Proof. For the ‘non-converse’ claim, we use Lemma 5.1 with $A = M^{1/2}$ and $B = (\kappa M)^{1/2}$. Taking $\alpha^{-1} = 2$, we have $\Xi_\alpha = 2\Xi - A^*A = 2\Xi - M$. For condition (ii) we then need our assumption $\Xi \leq M$. Since $\alpha \in (0, 1)$, for iii we need $(M, M - \Xi_\alpha) \in \mathcal{P}(T^{-1}(\hat{w}), \hat{u})$, which follows from us assuming $(M, M - \Xi) \in \mathcal{P}(T^{-1}(\hat{w}), \hat{u})$. Since $B^*B = \kappa M$, we see from the lemma that $(\Xi, \kappa M, M)$ -submonotonicity implies $(2\Xi - M, \kappa^2 M, M)$ -partial subregularity.

For the ‘converse’ claim, we use Lemma 5.3, where (i) and (ii) with $m = c^{-1}\kappa^{-1}$ and $A = (c\kappa M)^{1/2}$, and $B = (c^{-1}\kappa M)^{1/2}$ are easily verified for any $c > 0$. By assumption $M - \Xi + \Gamma$ is self-adjoint and positive semi-definite, so (iv) holds. For (v) we then require $\alpha \leq c\kappa/16$ and $M \leq \alpha P$. We also need $P \leq M$. We can take $P = M$ and $\alpha = 1$ if we take $c > 16\kappa^{-1}$. Then $B^*B = \rho M$ for $\rho = \kappa c^{-1} > (\kappa/4)^2$. With these choices $\alpha \leq 1$, so condition (iii) also holds. Now Lemma 5.3 shows that $(M, \rho M, M)$ subregularity implies $(\Xi, \kappa M, M)$ -submonotonicity. $\square$

The following generalises [3, Theorem 3.3] from the setting of convex functions. Since we formulate subregularity using squared norms, the conversion also squares κ .

Corollary 5.6. *Let $T : U \rightrightarrows U$ have closed values, and $M \in \mathbb{L}(U; U)$ be positive definite and self-adjoint. Let $(\hat{u}, \hat{w}) \in \text{graph } T$, and $\kappa > 0$.*

Then T is $(\kappa^2 M, M)$ -subregular if it is $(\kappa M, M)$ -strongly submonotone.

Conversely, if T also admits an $(M, 0)$ -gap function at $\hat{w}$, then T is $(\kappa M, M)$ -strongly submonotone if it is $(\rho M, M)$ -subregular for some $\rho > (\kappa/4)^2$ in an open neighbourhood $\mathcal{U} \ni \hat{u}$.

Proof. We apply Proposition 5.5 with $\Xi = M$ and $\Gamma = 0$. Clearly we have $M \geq \Xi$, $(M, M - \Xi) \in \mathcal{P}(T^{-1}(\hat{w}), \hat{u})$, and $M - \Xi + \Gamma$ is self-adjoint and positive semi-definite. $\square$

Corollary 5.7. *Let $f : U \rightarrow \overline{\mathbb{R}}$ be convex, proper, and lower semicontinuous, and $M \in \mathbb{L}(U; U)$ be positive definite and self-adjoint. Let $(\hat{u}, \hat{w}) \in \text{graph } \partial f$, and $\kappa > 0$.*

Then ∂f is $(\kappa^2 I, I)$ -subregular if it is $(\kappa I, I)$ -strongly submonotone.

Conversely, ∂f is $(\kappa I, I)$ -strongly submonotone if it is $(\rho I, I)$ -subregular for some $\rho > (\kappa/4)^2$ in an open neighbourhood $\mathcal{U} \ni \hat{u}$.

Proof. Since f is convex, ∂f admits the $(I, 0)$ -gap function $\tilde{\mathcal{G}}(u; u^*) := f(u) - f(u^*)$. Now we use Corollary 5.6 with $M = I$. $\square$

Remark 5.8. Each of the results in this section easily extends to submonotonicity of the form $(\Xi, \kappa N, M)$ for $N = M^{1/2}B$, and some $\kappa > 0$, $B, \Xi \in \mathbb{L}(U; U)$ if the corresponding subregularity is of the form $(P, \kappa' B^*B, M)$ for some P and $\kappa' > 0$. $\square$

5.3. Examples

The next example demonstrates that the ‘non-converse’ direction of Proposition 5.5 cannot in the general case be improved to yield subregularity of similar “partiality” Ξ as the underlying submonotonicity.

Example 5.9. (Partial submonotonicity without corresponding subregularity)

On $\mathbb{R}^2$, for $u = (u_1, u_2)$ and some $\gamma \in [0, 1]$, let us define

$$T(u) := \begin{cases} \left(\gamma u_1, \sqrt{(1 - \gamma^2)u_1^2 - u_2^2} \right), & (1 - \gamma^2)u_1^2 \geq u_2^2 \text{ and } u_2 \geq 0, \\ \emptyset, & \text{otherwise.} \end{cases}$$

Let $\Xi := \begin{pmatrix} \gamma & 0 \\ 0 & 0 \end{pmatrix}$. Then T is (Ξ, I, I) -partially strongly monotone at the point $(\hat{u}, \hat{w}) := (0, 0) \in \text{graph } T$. However, T is not (Ξ, I, I) -partially subregular.

Proof. Clearly $T^{-1}(0) = 0$. With $N := M := I$, (PSM) expands as

$$\left[\gamma u_1^2 + u_2 \sqrt{(1 - \gamma^2)u_1^2 - u_2^2} \right] + (1 - \gamma)u_1^2 + u_2^2 \geq u_1^2 + u_2^2 \quad ((1 - \gamma^2)u_1^2 \geq u_2^2, u_2 \geq 0).$$

This clearly holds.

Since $T^{-1}(0) = \{0\}$ is a singleton, the projection condition of Proposition 5.5 clearly holds. We also have $I \geq \Xi$. Referring to the proposition we therefore obtain $(2\Xi - I, I, I)$ -partial subregularity, for which (PSR) expands as

$$[\gamma^2 u_1^2 + ((1 - \gamma^2)u_1^2 - u_2^2)] + 2[(1 - \gamma)u_1^2 + u_2^2] \geq u_1^2 + u_2^2 \quad ((1 - \gamma^2)u_1^2 \geq u_2^2, u_2 \geq 0).$$

This clearly holds. On the other hand (PSR) for (Ξ, I, I) -partial subregularity would require

$$[\gamma^2 u_1^2 + ((1 - \gamma^2)u_1^2 - u_2^2)] + [(1 - \gamma)u_1^2 + u_2^2] \geq u_1^2 + u_2^2 \quad ((1 - \gamma^2)u_1^2 \geq u_2^2, u_2 \geq 0).$$

This can be simplified as $(1 - \gamma)u_1^2 \geq u_2^2$. Taking $u_2^2 = (1 - \gamma^2)u_1^2$, we see that the condition cannot hold for $\gamma \in (0, 1)$. $\square$

Finally, having already shown the converse, we show that subregularity is also not a weaker property than strong submonotonicity.

Example 5.10. (Subregularity without *strong* submonotonicity) On $\mathbb{R}^2$, taking $u = (u_1, u_2)$, let

$$T(u) = \begin{cases} (u_2, u_1), & u_1, u_2 \geq 0, \\ \emptyset, & \text{otherwise.} \end{cases}$$

Then T is (I, I, I) -partially subregular at $(\hat{u}, \hat{w}) = 0 \in \text{graph } T$, but (Ξ, I, I) -partially strongly submonotone with $0 \leq \Xi \leq I$ only for $\Xi = 0$.

Proof. Clearly $0 = \hat{u} \in T^{-1}(0)$ is unique. Therefore, since $\langle T(u), u \rangle = 2u_1 u_2$ for $u_1, u_2 \geq 0$, (PSM) for $M = N = I$ and arbitrary $0 \leq \Xi \leq I$ reads

$$2u_1 u_2 \geq \|(u_1, u_2)\|_{\Xi}^2 \quad (u_1, u_2 \geq 0).$$

This clearly holds for $\Xi = 0$. Generally, for $\Xi \leq \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, it is necessary that $2 \geq b + c$ and $a, d = 0$. The only positive semi-definite choice that satisfies this is $\Xi = 0$.

On the other hand, (PSR) for (I, I, I) -partially subregularity with $\mathcal{U} = U$ reads

$$\|T(u_1, u_2)\|^2 \geq \|(u_1, u_2)\|^2 \quad (u_1, u_2 \geq 0).$$

Since $\|T(u)\|^2 = u_2^2 + u_1^2$, this immediately holds. $\square$

6. Primal-dual methods

We now apply the work of the previous sections to primal-dual methods. We concentrate on submonotonicity, as it appears in general to be easier to prove directly than subregularity. If in a specific application subregularity is known, it can using the results of Section 5 be converted into submonotonicity. We start with general results on saddle-point problems, follow with the primal-dual proximal splitting method (PDPS) of [8], and finish with examples regarding specific applications.

6.1. Saddle-point problems

Many problems in data science and image processing can be written in the form

$$\min_x G(x) + F(Kx), \quad (29)$$

for $G : X \rightarrow \overline{\mathbb{R}}$ and $F : Y \rightarrow \overline{\mathbb{R}}$ be convex, proper and lower semicontinuous, and $K \in \mathbb{L}(X; Y)$. In image processing one would often work in the Banach space of functions of bounded variation, but after discretisation, if necessary, we can assume that the spaces X and Y are Hilbert (and finite-dimensional).

It can be difficult to apply an optimisation algorithm directly to (29): F is typically nonsmooth, so gradient steps are out of the question. Due to the coupling effects of K , also a proximal step for $F \circ K$ is not feasible. This is why we are interested in the equivalent saddle point problem. For F^* the convex conjugate of F , this can be written

$$\min_x \max_y G(x) + \langle Kx, y \rangle - F^*(y). \quad (30)$$

The first-order necessary optimality conditions for (30) can be written

$$-K^*\hat{y} \in \partial G(\hat{x}), \quad \text{and} \quad K\hat{x} \in \partial F^*(\hat{y}). \quad (31)$$

Setting $U := X \times Y$ and introducing the variable splitting notation $u = (x, y)$, $\hat{u} = (\hat{x}, \hat{y})$, etc., this can succinctly be written as $0 \in H(\hat{u})$ in terms of the operator

$$H(u) := \begin{pmatrix} \partial G(x) + K^*y \\ \partial F^*(y) - Kx \end{pmatrix}. \quad (32)$$

From now on, we will not explicitly assume G and F^* to be convex functions, indeed ∂G and ∂F^* can be replaced by any set-valued operators. In particular, they can be nonconvex functions, and ∂ can be replaced by the Clarke, limiting, or other subdifferential of nonconvex functions. In this case, it should be shown that (31) characterises critical points to (30) or (29).

For some primal and dual step length $\tau_i, \sigma_{i+1} > 0$ and testing parameters $\psi_i, \psi_{k+1} > 0$, we take

$$Z_{i+1} := \begin{pmatrix} \phi_i \tau_i I & 0 \\ 0 & \psi_{i+1} \sigma_{i+1} I \end{pmatrix}. \quad (33)$$

For some $\gamma, \rho \geq 0$ we also introduce

$$\Xi_{i+1}^0 := \begin{pmatrix} 0 & 2\tau_i K^* \\ -2\sigma_{i+1} K & 0 \end{pmatrix}, \quad \text{and} \quad \Xi_{i+1} := \Xi_{i+1}^0 + \begin{pmatrix} \tau_i \gamma & 0 \\ 0 & \sigma_{i+1} \rho \end{pmatrix}. \quad (34)$$

We want to use Theorem 3.5 to prove the convergence of (PP) for H defined in (32) for some linear preconditioner $M_{k+1} \in \mathbb{L}(U; U)$. For that purpose we need to prove the $(Z_{i+1}\Xi_{i+1}, 2Z_{i+1}, Z_{i+2}M_{i+2})$ -partial strong submonotonicity for H at $(\hat{u}, 0) \in \text{graph } H$. Dividing (PSM) by 2, this amounts to showing

$$\begin{aligned} \inf_{u^* \in H^{-1}(0)} & \left(\langle w, u - u^* \rangle_{Z_{i+1}} + \frac{1}{2} \|u - u^*\|_{Z_{i+2}M_{i+2} - Z_{i+1}\Xi_{i+1}}^2 \right) \\ & \geq \frac{1}{2} \text{dist}_{Z_{i+2}M_{i+2}}^2(u, H^{-1}(0)) \end{aligned} \quad (35)$$

for all $u \in \mathcal{U}$ and $w \in H(u)$ in some neighbourhood $\mathcal{U}$ of $\hat{u}$.

By (32), for some $q \in \partial G(u)$, $z \in \partial F^*(y)$, $q^* := -K^*y^* \in \partial G(x^*)$, and $z^* := Kx^* \in \partial F^*(y^*)$ we have

$$\langle w, u - u^* \rangle_{Z_{i+1}} = \phi_i \tau_i \langle q - q^*, x - x^* \rangle + \psi_{i+1} \sigma_{i+1} \langle z - z^*, y - y^* \rangle + \frac{1}{2} \|u - u^*\|_{Z_{i+1}\Xi_{i+1}^0}^2$$

Recalling (34), to show (35), it remains to prove

$$\begin{aligned} \inf_{u^* \in H^{-1}(0)} & \left(\phi_i \tau_i [\langle q - q^*, x - x^* \rangle - \frac{\gamma}{2} \|x - x^*\|^2] + \psi_{i+1} \sigma_{i+1} [\langle z - z^*, y - y^* \rangle - \frac{\rho}{2} \|y - y^*\|^2] \right. \\ & \left. + \frac{1}{2} \|u - u^*\|_{Z_{i+2}M_{i+2}}^2 \right) \geq \inf_{u^{**} \in H^{-1}(0)} \frac{1}{2} \|u - u^{**}\|_{Z_{i+2}M_{i+2}}^2. \end{aligned}$$

Splitting the infimum on the left hand side over the three terms, we obtain the “marginalised” condition:

Proposition 6.1. *Let the operator H defined in (32), and Ξ_{i+1} in (34) for some $\gamma \geq 0$ and $\rho \geq 0$. For any $u^* = (x^*, y^*) \in H^{-1}(0)$, let $q^* := -K^*y^* \in \partial G(x^*)$, and $z^* := Kx^* \in \partial F^*(y^*)$. Suppose for $u = (x, y) \in \mathcal{U}$ in some neighbourhood $\mathcal{U}$ of $\hat{u}$ that*

$$\inf_{u^* \in H^{-1}(0)} [\langle \partial G(x) - q^*, x - x^* \rangle - \frac{\gamma}{2} \|x - x^*\|^2] \geq 0, \quad (36a)$$

$$\text{and} \quad \inf_{u^* \in H^{-1}(0)} [\langle \partial F^*(y) - z^*, y - y^* \rangle - \frac{\rho}{2} \|y - y^*\|^2] \geq 0. \quad (36b)$$

Then H is $(Z_{i+1}\Xi_{i+1}, 2Z_{i+1}, Z_{i+2}M_{i+2})$ -partially strongly submonotone at $(\hat{u}, 0) \in \text{graph } H$.

The conditions (36) amount to $(\gamma I, 2I, 0)$ -partial strong submonotonicity, which is simply local strong monotonicity at every respective x^* for q^* or y^* for z^* with $(x^*, y^*) \in H^{-1}(0)$. It is a much stronger requirement than “proper” submonotonicity, but weaker than strong monotonicity. A challenging alternative is to try to prove the $(Z_{i+1}\Xi_{i+1}, 2Z_{i+1}, Z_{i+2}M_{i+2})$ -partial strong submonotonicity of H directly.

6.2. The primal-dual proximal splitting

The primal-dual proximal splitting (PDPS) of [8] consists of iterating the system

$$\begin{aligned} x^{i+1} &:= (I + \tau_i \partial G)^{-1}(x^i - \tau_i K^* y^i), \\ \bar{x}^{i+1} &:= \omega_i(x^{i+1} - x^i) + x^{i+1}, \\ y^{i+1} &:= (I + \sigma_{i+1} \partial F^*)^{-1}(y^i + \sigma_{i+1} K \bar{x}^{i+1}). \end{aligned}$$

Setting $\omega_i := \psi_{i+1}^{-1} \sigma_{i+1}^{-1} \phi_i \tau_i$, according to [22], see also [45, 43], this can also be obtained from (PP) with

$$M_{i+1} = \begin{pmatrix} \tau_i^{-1} I & -K^* \\ -\omega_i K & \sigma_{i+1}^{-1} I \end{pmatrix} \quad \text{so that} \quad Z_{i+1} M_{i+1} = \begin{pmatrix} \phi_i I & -\phi_i \tau_i K^* \\ -\phi_i \tau_i K & \psi_{i+1} I \end{pmatrix}. \quad (37)$$

The role of the top-right off-diagonal term in M_{i+1} is to decouple the update of the primal variable x^{i+1} to not depend on y^{i+1} , and so make the algorithm computable. The bottom-left off-diagonal is then chosen to ensure the self-adjointness of $Z_{i+1} M_{i+1}$.¹

Theorem 6.2. *Take H as in (32), Z_{i+1} as in (33), and M_{i+1} as in (37). Also let Ξ_{i+1} be as in (34) for some $\gamma, \rho \geq 0$. Suppose the testing and step length parameters τ_i, σ_i, ϕ_i , and ψ_i satisfy for some $\delta \in (0, 1)$,*

$$\phi_{i+1} \leq \phi_i(1 + \gamma \tau_i), \quad \psi_{i+2} \leq \psi_{i+1}(1 + \rho \sigma_{i+1}), \quad \text{and} \quad (38a)$$

$$\phi_i \tau_i = \psi_i \sigma_i, \quad \psi_i^{-1} \psi_{i+1} \geq (1 - \delta) \tau_i \sigma_i \|K\|^2. \quad (38b)$$

Suppose H is $(Z_{i+1} \Xi_{i+1}, 2Z_{i+1}, Z_{i+2} M_{i+2})$ -partially strongly submonotone at the point $(\hat{u}, 0) \in \text{graph } H$ with the neighbourhood of submonotonicity satisfying $\{u^i\}_{i=0}^N \subset \mathcal{U}$. Then condition (9) of Theorem 3.5 holds, and

$$Z_{i+1} M_{i+1} \geq \begin{pmatrix} \delta \phi_i I & 0 \\ 0 & 0 \end{pmatrix}. \quad (39)$$

In particular, we have the convergence rate estimate

$$\text{dist}^2(x^N; \hat{X}) \leq \phi_N^{-1} \delta^{-1} \text{dist}_{Z_1 M_1}^2(u^0; H^{-1}(0)) \quad (40)$$

where $\hat{X} := \{x^* \mid (x^*, y^*) \in H^{-1}(0)\}$.

Proof. It is easy to see that $Z_{i+1}(M_{i+1} + \Xi_{i+1}) \geq Z_{i+2} M_{i+2}$ if we enforce the first three conditions of (38) [43]. To ensure (39), using Young's inequality in (37), we derive the condition $\psi_{i+1} \geq (1 - \delta) \phi_i \tau_i^2 \|K\|^2$. Using the third condition of (38), this can alternatively be written as the last condition of (38). Finally, Theorem 3.5 shows the basic estimate (5). Using (39) in (5), we derive (40). $\square$

We now need to satisfy (38) and, minding (40), to derive convergence rates, try to make ϕ_i grow as fast as possible. The latter analysis is also based on the rules (38), and proceeds exactly as in [43, 12] (where ρ and γ arise from mere (strong) monotonicity).

¹The PDPS is sometimes called the Primal-Dual Hybrid Gradient Method, Modified (PDHGM) [17]. The word “modified” refers to this self-adjoint part. The “unmodified” PDHG would have zero in this corner [49, 22].

Example 6.3. (Weak convergence) If $\rho = 0$ and $\gamma = 0$, by (38) ϕ_i and ψ_{i+1} remain constant, so we obtain no convergence rates, merely weak convergence of subsequences of iterates; see [43, 12]. By (38) the step lengths $\tau_i = \tau$ and $\sigma_i = \sigma$ need to be constant and satisfy $1 \geq (1 - \delta)\tau\sigma\|K\|^2$. It is also possible [8, 43] to derive $O(1/N)$ convergence of an ergodic duality gap, which we have avoided introducing here. Note that $\omega_i \equiv 1$.

Example 6.4. (Acceleration to $O(1/N^2)$) If $\rho = 0$ and $\gamma > 0$, we still have $\psi_{i+1} = \psi_0$ constant. We can initialise $1 \geq (1 - \delta)\tau_0\sigma_0\|K\|^2$ and update $\sigma_{i+1} := \sigma_i\tau_i/\tau_{i+1}$. Taking $\tau_i = \phi_i^{-1/2}$, and $\phi_{i+1} := \phi_i(1 + \tilde{\gamma}\tau_i)$ for some $\tilde{\gamma} \in (0, \gamma]$, we can prove that ϕ_i is of the order $\Theta(N^2)$. Therefore we obtain from (40) $O(1/N^2)$ convergence rates for the primal variable x . Note that $\omega_i = \psi_{i+1}^{-1}\sigma_{i+1}^{-1}\phi_i\tau_i = \phi_{i+1}^{-1}\tau_{i+1}^{-1}\phi_i\tau_i = \tau_{i+1}/\tau_i < 1$.

Example 6.5. (Linear rates) If both $\rho > 0$ and $\gamma > 0$, take $\sigma_{i+1} = \sigma$ and $\tau_i = \tau$ as constants satisfying

$$\theta := 1 + \min\{\rho\sigma, \gamma\tau\} \geq (1 - \delta)\tau\sigma\|K\|^2.$$

Then with $\psi_0 = \phi_0\tau\sigma^{-1}$, $\phi_{i+1} := \theta\phi_i$, and $\psi_{i+1} := \theta\psi_i$ we verify (38). Note that $c > 1$, so $\{\phi_i\}_{i \in \mathbb{N}}$ grows exponentially. Thus we obtain linear convergence of the primal variables from (39) and (40).² Note that $\omega_i = \psi_{i+1}^{-1}\sigma_{i+1}^{-1}\phi_i\tau_i = 1$ as in Example 6.3.

Remark 6.6. (Global and local convergence) Convergence results based on submonotonicity are generally local due to the neighbourhood $\mathcal{U}$. With G and F^* convex, H is monotone ($\gamma = 0$ and $\rho = 0$). Therefore the weak convergence results of Example 6.3 are global. This is an *a priori* convergence property. In finite dimensions, the convergence is strong, so eventually the iterates will be in the neighbourhood $\mathcal{U}$ (for some $\hat{u} \in H^{-1}(0)$) of partial strong submonotonicity. Then the algorithm switches to faster convergence. This is only known *a posteriori* as it depends on $\hat{u}$.

6.3. Application examples

We now look at regularity and convergence rate results for a few prototypical applications. We will show that linear convergence of the PDPS can in specific cases be obtained by the unaccelerated primal-dual algorithm without full strong convexity. These results improve known convergence results for this algorithm. These results are comparable to those obtained in [28, 30] for other classes of algorithms based on a related theory of smooth submanifolds. Throughout, we rely on the marginalisation of submonotonicity in Proposition 6.1, so seek to verify (36).

Example 6.7. (Lasso) For some data matrix and vector $K \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$, and regularisation parameter $\alpha > 0$, consider the Lasso or regularised regression problem

$$\min_x \frac{1}{2}\|b - Kx\|^2 + \alpha\|x\|_1.$$

This can be written in the saddle point form (30) with $G(x) = \alpha\|x\|_1$, and $F^*(y) = \frac{1}{2}\|y\|^2 - \langle b, y \rangle$. Let $(\hat{x}, \hat{y}) \in H^{-1}(0)$ for H as in (32). The PDPS converges linearly

²It is possible to improve (39) for linear convergence of the dual variables.

whenever the primal solution $\hat{x} = 0$ and is unique, and any corresponding dual solution is strictly complementary ($|[K^*\hat{y}]_k| < 1$ for all coordinates $k = 1, \dots, n$).

Proof. We use Proposition 6.1 with Theorem 6.2 and Example 6.5. We therefore need to verify (36). Since F^* is strongly convex, we can take $\rho = 1$ in (36b). To prove (36a) for $G(x) = \sum_{k=1}^n |x_k|$, we proceed as in Lemma 3.16, however (36a) is much more restrictive than the $(I, \gamma I)$ -strong submonotonicity shown in the lemma. Let $(x^*, y^*) \in H^{-1}(0)$. Write $q^* := (q_1^*, \dots, q_n^*) = -K^*y^* \in \partial G(x^*)$ and $x^* = (x_1^*, \dots, x_n^*)$. Of course $q_k^* \in \partial|\cdot|(0)$. We need to prove

$$\langle q_k - q_k^*, x_k - x_k^* \rangle \geq \frac{\gamma}{2} |x_k - x_k^*|^2 \quad (q_k \in \partial|\cdot|(x_k); k = 1, \dots, n). \quad (41)$$

If $x_k^* \neq 0$ for some k , we are only able to take $\gamma = 0$ using the monotonicity of $\partial|\cdot|$. However, by assumption, $x_k^* = 0$ for all k . Fix k and let $\gamma > 0$ be arbitrary. If $x_k > 0$, $q_k = 1 \geq q_k^*$. Under strict complementarity the inequality is strict, so (41) for this k holds in a small enough neighbourhood of $x_k^* = 0$. The case $x_k < 0$ is analogous. This proves, via Proposition 6.1, the submonotonicity of H required by Theorem 6.2. Finally, Example 6.5 proves (38) with testing parameters that yield linear convergence. By Remark 6.6 the convergence is global. $\square$

Example 6.8. (Denoising-type problems) Consider then for $K \in \mathbb{R}^{m \times n}$ and vector $b \in \mathbb{R}^n$ a problem of the form

$$\min_x \frac{1}{2} \|b - x\|^2 + \alpha \|Kx\|_1,$$

With K a discretised gradient operator this includes in particular total variation denoising (anisotropic, for simplicity) and with K representing decomposition into a wavelet basis, wavelet-based denoising. In the saddle point form the problem is

$$\min_x \max_y \frac{1}{2} \|b - x\|^2 + \langle Kx, y \rangle + \delta_{[-\alpha, \alpha]^m}(y).$$

Since the indicator function $F^* = \delta_{[-\alpha, \alpha]^m}$ is not strongly convex, but $G(x) = \frac{1}{2} \|b - x\|^2$ is strongly convex, standard results [8] give the $O(1/N^2)$ rate for the accelerated PDPS.

Let $(\hat{x}, \hat{y}) \in H^{-1}(0)$ for H as in (32). Using subregularity the PDPS, in fact, converges linearly whenever $[K\hat{x}]_k \neq 0$ for all $k = 1, \dots, m$. In the TV denoising case, this means that the method converges linearly when the solution image does not contain any flat regions, while in the wavelet denoising case this means that all wavelet coefficients of the solution have to be non-zero.

Proof. We again use Proposition 6.1 with Theorem 6.2 and Example 6.5. We therefore need to verify (36). Since G is strongly convex, we can take $\gamma = 1$ in (36a). To prove (36b), we proceed as in Lemma 3.15, however (36b) is much more restrictive than the $(I, \gamma I)$ -strong submonotonicity shown in the lemma. Let $(x^*, y^*) \in H^{-1}(0)$. Observe that x^* is unique by the strong convexity of G . Write $z^* = z^* = (z_1^*, \dots, z_m^*) := Kx^* \in \partial F^*(y^*)$. Then $z_k^* \in \partial \delta_{[-\alpha, \alpha]}(y_k^*)$. We need to prove

$$\langle z_k - z_k^*, y_k - y_k^* \rangle \geq \frac{\rho}{2} |y_k - y_k^*|^2 \quad (z_k \in \partial \delta_{[-\alpha, \alpha]}(y_k); k = 1, \dots, m). \quad (42)$$

If $z_k^* = 0$ for some k , we are only able to take $\rho = 0$ using the monotonicity of ∂F^* . Otherwise, let $\rho > 0$ be arbitrary and k fixed. If $z_k^* > 0$, we have $y_k^* = \alpha$. Then $y_k \leq y_k^*$. It is clear that (42) holds for $y_k = y_k^*$ where as for $y_k < y_k^*$, we have $z_k = 0$. Thus the left hand side of (42) is positive, and the inequality holds for y_k in a small enough neighbourhood around $y_k^* = \alpha$. The case $z_k^* < 0$ is similar. This proves, via Proposition 6.1, the submonotonicity of H required by Theorem 6.2. Finally, Example 6.5 proves (38) with testing parameters that yield linear convergence. By Remark 6.6 the convergence is global. $\square$

7. Conclusions

We have studied notions of partial strong submonotonicity and partial subregularity motivated by convergence proofs of optimisation methods. We have showed that these concepts can be used to show improved linear convergence rates, where conventional strong convexity or strong monotonicity is not present. To facilitate the verification of partial submonotonicity, in particular for saddle point problems, in further research we would like to develop a characterisation of partial submonotonicity based on nonsmooth derivatives, similar in spirit to the Mordukhovich criterion for the Aubin property. The latter we have studied in context of the PDPS for nonlinear K in [11]. Considering the non-trivial characterisations of metric subregularity in [18], this may or may not be a fruitful path. For local monotonicity and hypomonotonicity, which do not involve the infimisation present in submonotonicity, coderivative characterisations are considered in [35, 9].

Acknowledgements. This research has been supported by the EPSRC First Grant EP/P021298/1, “PARTIAL Analysis of Relations in Tasks of Inversion for Algorithmic Leverage”. The author would also like to thank Christian Clason for pointing out the papers [3, 4].

A data statement for the EPSRC. This is a theory paper that does not rely on any data or program code.

References

- [1] S. Adly, R. Cibulka, H. V. Ngai: *Newton’s method for solving inclusions using set-valued approximations*, SIAM J. Optimization 25 (2015) 159–184.
- [2] F. J. Aragón Artacho, M. Gaydu: *A Lyusternik-Graves theorem for the proximal point method*, Comp. Optimization Appl. 52 (2012) 785–803.
- [3] F. J. Aragón Artacho, M. H. Geoffroy: *Characterization of metric regularity of subdifferentials*, J. Convex Analysis 15 (2008) 365–380.
- [4] F. J. Aragón Artacho, M. H. Geoffroy: *Metric subregularity of the convex subdifferential in Banach spaces*, J. Nonlinear Convex Analysis 15 (2014) 35–47.
- [5] T. Aspelmeier, C. Charitha, D. R. Luke: *Local linear convergence of the ADMM/Douglas-Rachford algorithms without strong convexity and application to statistical imaging*, SIAM J. Imaging Sciences 9 (2016) 842–868.
- [6] J.-P. Aubin: *Lipschitz behavior of solutions to convex minimization problems*, Math. Oper. Res. 9 (1984) 87–111.
- [7] J. Aubin, H. Frankowska: *Set-Valued Analysis*, Springer, Berlin (1990).

- [8] A. Chambolle, T. Pock: *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision 40 (2011) 120–145.
- [9] N. H. Chieu, G. M. Lee, B. S. Mordukhovich, T. T. A. Nghia: *Coderivative characterizations of maximal monotonicity for set-valued mappings*, J. Convex Analysis 23 (2016) 461–480.
- [10] R. Cibulka, A. Dontchev, A. Kruger: *Strong metric subregularity of mappings in variational analysis and optimization*, J Math. Analysis Appl. 457 (2018) 1247–1282.
- [11] C. Clason, T. Valkonen: *Stability of saddle points via explicit coderivatives of pointwise subdifferentials*, Set-Valued Var. Analysis 25 (2017) 69–112.
- [12] C. Clason, T. Valkonen: *Introduction to nonsmooth analysis and optimization*, work in progress, arXiv: 2001.00216 (2020).
- [13] A. Dontchev, R. Rockafellar: *Implicit Functions and Solution Mappings: A View from Variational Analysis*, Springer Series in Operations Research and Financial Engineering, Springer, New York (2014).
- [14] A. L. Dontchev, R. T. Rockafellar: *Regularity and conditioning of solution mappings in variational analysis*, Set-Valued Var. Analysis 12 (2004) 79–109.
- [15] M. Durea, M. Pantîruc, R. Strugariu: *A new type of directional regularity for mappings and applications to optimization*, SIAM J. Optimization 27 (2017) 1204–1229.
- [16] I. Ekeland: *On the variational principle*, J. Math. Analysis Appl. 47 (1974) 324–353.
- [17] E. Esser, X. Zhang, T. F. Chan: *A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science*, SIAM J. Imaging Sciences 3 (2010) 1015–1046.
- [18] H. Gfrerer: *First order and second order characterizations of metric subregularity and calmness of constraint set mappings*, SIAM J. Optimization 21 (2011) 1439–1474.
- [19] H. Gfrerer: *On directional metric regularity, subregularity and optimality conditions for nonsmooth mathematical programs*, Set-Valued Var. Analysis 21 (2013) 151–176.
- [20] H. Gfrerer, B. S. Mordukhovich: *Robinson stability of parametric constraint systems via variational analysis*, SIAM J. Optimization 27 (2017) 438–465.
- [21] D. Han, X. Yuan: *Local linear convergence of the alternating direction method of multipliers for quadratic programs*, SIAM J. Num. Analysis 51 (2013) 3446–3457.
- [22] B. He, X. Yuan: *Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective*, SIAM J. Imaging Sciences 5 (2012) 119–149.
- [23] R. Henrion, T. Surowiec: *On calmness conditions in convex bilevel programming*, Appl. Analysis 90 (2011) 951–970.
- [24] A. D. Ioffe: *Regular points of Lipschitz functions*, Trans. Amer. Math. Soc. 251 (1979) 61–69.
- [25] A. D. Ioffe: *Variational Analysis of Regular Mappings: Theory and Applications*, Springer Monographs in Mathematics, Springer, Berlin (2017).
- [26] A. Y. Kruger: *Error bounds and metric subregularity*, Optimization 64 (2015) 49–79.
- [27] D. Leventhal: *Metric subregularity and the proximal point method*, J. Math. Analysis Appl. 360 (2009) 681–688.
- [28] A. S. Lewis, S. Zhang: *Partial smoothness, tilt stability, and generalized hessians*, SIAM J. Optimization 23 (2013) 74–94.

- [29] G. Li, B. S. Mordukhovich: *Hölder metric subregularity with applications to proximal point method*, SIAM J. Optimization 22 (2012) 1655–1684.
- [30] J. Liang, J. Fadili, G. Peyré: *Local linear convergence of forward-backward under partial smoothness*, Adv. Neural Information Processing Systems 27 (2014) 1970–1978.
- [31] Z.-Q. Luo, P. Tseng: *Error bound and convergence analysis of matrix splitting algorithms for the affine variational inequality problem*, SIAM J. Optimization 2 (1992) 43–54.
- [32] B. Martinet: *Brève communication. Régularisation d'inéquations variationnelles par approximations successives*, ESAIM: Math. Modelling Numer. Analysis 4 (1970) 154–158.
- [33] S. Mazurenko, J. Jauhainen, T. Valkonen: *Primal-dual block-proximal splitting for a class of non-convex problems*, arXiv: 1911.06284 (2019).
- [34] B. S. Mordukhovich: *Variational Analysis and Generalized Differentiation. I: Basic Theory*, Grundlehren der mathematischen Wissenschaften 330, Springer, Berlin (2006).
- [35] B. S. Mordukhovich, T. T. A. Nghia: *Local monotonicity and full stability for parametric variational systems*, SIAM J. Optimization 26 (2016) 1032–1059.
- [36] H. V. Ngai, M. Théra: *Error bounds in metric spaces and application to the perturbation stability of metric regularity*, SIAM J. Optimization 19 (2008) 1–20.
- [37] H. V. Ngai, M. Théra: *Directional metric regularity of multifunctions*, Math. Oper. Res. 40 (2015) 969–991.
- [38] S. M. Robinson: *Stability theory for systems of inequalities. II: Differentiable non-linear systems*, SIAM J. Numer. Analysis 13 (1976) 497–513.
- [39] S. M. Robinson: *Some continuity properties of polyhedral multifunctions*, in: *Mathematical Programming at Oberwolfach*, H. König, B. Korte, K. Ritter (eds), Mathematical Programming Studies 14, Springer, Berlin (1981) 206–214.
- [40] R. T. Rockafellar: *Convex Analysis*, Princeton University Press, Princeton (1972).
- [41] R. T. Rockafellar: *Monotone operators and the proximal point algorithm*, SIAM J. Optimization 14 (1976) 877–898.
- [42] R. T. Rockafellar, R. J.-B. Wets: *Variational Analysis*, Springer, Berlin (1998).
- [43] T. Valkonen: *Testing and non-linear preconditioning of the proximal point method*, Appl. Math. Optimization, doi.org/10.1007/s00245-018-9541-6 (2018).
- [44] T. Valkonen: *Block-proximal methods with spatially adapted acceleration*, Electronic Trans. Numer. Analysis 51 (2019) 15–49.
- [45] T. Valkonen, T. Pock: *Acceleration of the PDHGM on partially strongly convex functions*, J. Math. Imaging Vision 59 (2017) 394–414.
- [46] H. Van Ngai, A. Y. Kruger, M. Théra: *Slopes of multifunctions and extensions of metric regularity*, Vietnam J. Mathematics 40 (2012) 355–369.
- [47] J. J. Ye and D. L. Zhu: *Optimality conditions for bilevel programming problems*, Optimization 33 (1995) 9–27.
- [48] Z. Zhou, A. M.-C. So: *A unified approach to error bounds for structured convex optimization problems*, Math. Programming 165 (2017) 689–728.
- [49] M. Zhu, T. Chan: *An efficient primal-dual hybrid gradient algorithm for total variation image restoration*, CAM Report 08-34, UCLA, Los Angeles (2008).