Chapter 8

The proximal point algorithm

In this chapter we consider an algorithm for computing minima of a convex function. Actually, such an algorithm determines a zero of the subdifferential of a convex function. It benefits from the fact that a subdifferential is a maximal monotone operator. Therefore, the main part of the considerations in this chapter is devoted to algorithms which can be used to find zeroes of maximal monotone operators. We exploit the idea of splitting. This idea unifies a substantial part of decomposition methods for convex programming.

For the larger part of the chapter we study the methods and algorithms in Hilbert spaces. The subdifferential of convex functions and their properties play an important role.

The literature concerning the proximal operator and monotone operators is really immense. We refer without making a claim for the best choice [1, 4, 6, 10, 23, 25, 27, 28, 32, 35, 36].

8.1 The proximal operator: preliminaries

Let $\mathcal{X}$ be a Banach space with norm $\|\cdot\|$. In the following we use the **abbreviation**:

$$\Gamma_0(\mathcal{X}) := \{f : \mathcal{X} \longrightarrow \hat{\mathbb{R}} : f \text{ proper, lower semicontinuous, convex}\}.$$

Notice that a function f is called **lower semicontinuous** iff all level sets $N_t := \{x \in \mathcal{X} : f(x) \leq t\}, t \in \mathbb{R}$, are closed. This equivalent to the closedness of the epigraph $\text{epi}(f) := \{(x,t) \in \mathcal{X} \times \mathbb{R} : t \geq f(x)\}$.

Let $\mathcal{X}$ be a Banach space and let $f \in \Gamma_0(\mathcal{X})$. We consider the minimization problem

$$\text{Minimize } f(u) \text{ subject to } u \in \mathcal{X}. \tag{8.1}$$

Even if f is bounded from below, existence und uniqueness of a solution cannot be guaranteed. Moreover, unstable dependence upon the function f can be observed. In order to get a minimization problem with better properties, the idea of regularization is helpfull. Let us consider

$$\text{mor}_{s,f}(x) := \inf_{u \in \mathcal{X}} \left(f(u) + \frac{1}{2s}\|u - x\|^2\right) \tag{8.2}$$

where $x \in \mathcal{X}$ is a given point in $\mathcal{X}$ and $s > 0$. The family $(\text{mor}_{s,f})_{s>0}$ is called the **Moreau envelope** or the **Moreau-Yosida regularization** family. s is a scaling parameter. We set $\text{mor}_f := \text{mor}_{1,f}$.

Notice that $\mathrm{mor}_{s,f}$ is the inf convolution of f and $s^{-1}j$ where $j(x) := \frac{1}{2}\|x\|^2, x \in \mathcal{X}$:

$$\mathrm{mor}_{s,f} = f \,\square\, s^{-1}j \tag{8.3}$$

Lemma 8.1. *Let $\mathcal{X}$ be a reflexive Banach space and let $f \in \Gamma_0(\mathcal{X})$. Let $(mor_{s,f})_{s>0}$ be the Moreau-envelope of f. Then we have:*

(1) $mor_{s,f}$ is convex and lower semicontinuous for all $s > 0$.

(2) $\#\{z \in \mathcal{X} : f(z) + \frac{1}{2s}\|z - x\|^2 = mor_{s,f}(x)\} = 1$.

(3) $dom(mor_{s,f}) = \mathcal{X}$ for all $s > 0$.

(4) $\inf_{u\in\mathcal{X}} f(u) = \inf_{x\in\mathcal{H}} mor_{s,f}(x)$ for all $s > 0$.

Proof:
Ad (1) Consider the function l, defined by $l(x, u) := f(u) + \frac{1}{2s}\|u - x\|^2$. l is jointly convex in x and u. Then $\mathrm{mor}_{s,f}$ is convex since it is the infimum of l with respect to variable u. Since $\mathrm{mor}_{s,f}$ is the infimum of lower semicontinuous functions $\mathrm{mor}_{s,f}$ is lower semicontinuous.
Ad (2) We have $\lim_{\|u\|\to\infty}(f(u) + \frac{1}{2s}\|u - x\|^2) = \infty$. This follows from the fact that due to the duality theorem f is bounded from below by an affine function. Therefore $\mathrm{mor}_{s,f}(x)$ is found by minimizing the mapping $u \longmapsto f(u) + \frac{1}{2s}\|u - x\|^2$ in a ball $\overline{B}_r, r > 0$. Since bounded sets in reflexive Banach spaces are weakly sequential compact and since f is lower semicontinuous we obtain the existence of a minimizer z. The uniqueness of the minimizer z follows from the fact that the mapping $g : \mathcal{X} \ni u \longmapsto f(u) + \frac{1}{2s}\|u - x\|^2 \in \hat{\mathbb{R}}$ is strongly convex with modulus $\frac{1}{2s}$, i.e. $g - \frac{1}{2s}\| \cdot - x\|^2$ is convex.
Ad (3) Follows from (2).
Ad (4) Follows from

$$\inf_{x\in\mathcal{H}} \inf_{u\in\mathcal{X}} (f(u) + \frac{1}{2s}\|u - x\|^2) = \inf_{u\in\mathcal{H}} \inf_{x\in\mathcal{X}} (f(u) + \frac{1}{2s}\|u - x\|^2) = \inf_{u\in\mathcal{X}} f(u).$$

■

Definition 8.2. *Let $\mathcal{X}$ be a reflexive Banach space, let $f \in \Gamma_0(\mathcal{X})$ and let $s > 0$. The mapping*

$$prox_{s,f} : \mathcal{X} \ni x \longmapsto argmin_{u\in\mathcal{X}}(f(u) + \frac{1}{2s}\|u - x\|^2) \in \mathcal{X} \tag{8.4}$$

is called the **proximal operator** *of the function f with scaling parameter s.*
We set $prox_f := prox_{1,f}$. □

Lemma 8.3. *Let $\mathcal{X}$ be a reflexive Banach space with duality mapping J_X, let $x \in \mathcal{X}$ and let $f \in \Gamma_0(\mathcal{X})$. Then the following statements are equivalent:*

(a) $w = prox_{s,f}(x)$.

(b) $\theta \in J_X(w - x) + s\partial f(w)$.

Proof:
We know that w minimizes $u \longmapsto f(u) + \frac{1}{2s}\|u - x\|^2$ if and only if $\theta \in \partial(f + \frac{1}{2s}\| \cdot - x\|^2)$. Since the norm is continuous in each point, especially in the points of dom(f), we may apply the sum-decomposition of subdifferentials and obtain that w minimizes $u \longmapsto f(u) + \frac{1}{2s}\|u - x\|^2$ if and only if $\theta \in \partial f(w) + \frac{1}{s}J_X(w - x)$. Here we have used the fact that $\partial(\frac{1}{2}\| \cdot \|^2) = J_X$; see Theorem 3.33. Now, the equivalence of (a), (b) is clear. ■

Lemma 8.4. *Let $\mathcal{X}$ be a reflexive Banach space and let $f \in \Gamma_0(\mathcal{X})$. Let $x^* \in \mathcal{X}, s > 0$. Then the following statements are equivalent:*

(a) x^* *minimizes* f.

(b) $prox_{s,f}(x^*) = x^*$.

Proof:
Let J_X be the duality map of $\mathcal{X}$.
Ad $(a) \implies (b)$ We have

$$f(x^*) + \frac{1}{2s}\|x^* - x^*\|^2 \le f(u) + \frac{1}{2s}\|u - x^*\|^2 \text{ for all } u \in \mathcal{X}.$$

Hence $\text{prox}_{s,f}x^* = x^*$.
Ad $(b) \implies (a)$ $w := x^*$ minimizes $u \longmapsto f(u) + \frac{1}{2s}\|u - x^*\|^2$ and therefore $\theta \in s\partial f(w) + J_X(w - x^*)$. Since $w = x^*$ we obtain $\theta \in \partial f(x^*)$ due to the fact that $J_{\mathcal{X}}(\theta) = \{\theta\}$. This implies that x^* minimizes f. ∎

Lemma 8.5. *Let $\mathcal{H}$ be a Hilbert space and let $f \in \Gamma_0(\mathcal{H})$ with the Fenchel-conjugate f^*. Then for all $x \in \mathcal{H}$:*

(a) $prox_f(x) + prox_{f^*}(x) = x$.

(b) $mor_f(x) + mor_{f^*}(x) = \frac{1}{2}\|x\|^2$.

Proof:
Ad (a) Let $x \in \mathcal{H}$. Then

$$w = \text{prox}_f(x) \iff x - w \in \partial f(x) \iff x \in \partial f^*(x - w) \iff x - w = \text{prox}_{f^*}(x). \quad (8.5)$$

Ad (b) Let $x \in \mathcal{H}$. We have (see (8.3))

$$\begin{aligned} \text{mor}_f(x) &= \inf_{u \in \mathcal{H}} (f(u) + \frac{1}{2}\|x\|^2 + \frac{1}{2}\|u\|^2 - \langle u|x \rangle) = \frac{1}{2}\|x\|^2 - (f + j)^*(x) \\ &= \frac{1}{2}\|x\|^2 - (f^* \,\square\, j)(x) = \frac{1}{2}\|x\|^2 - \text{mor}_{f^*}(x). \end{aligned}$$

Here we have used the fact that j is selfdual, i.e. $j^* = j$. ∎

Example 8.6. *Consider the function $f : \mathbb{R} \ni u \longmapsto |u| \in \mathbb{R}$. The proximal operator and the Moreau envelope is given as follows:*[1]

$$\begin{aligned} prox_{s,f}(x) &= (1 - \frac{s}{|x|})_+ x, \; x \in \mathbb{R}, \\ mor_{s,f}(x) &= \begin{cases} \frac{1}{2s}|x|^2 & , |x| \le s \\ |x| - \frac{s}{2} & , |x| > s \end{cases} \end{aligned}$$

Notice that $mor_{s,f}$ is differentiable in $x = 0$. □

Lemma 8.7. *Let $\mathcal{X}$ be a Banach space and let $f \in \Gamma_0(\mathcal{X})$. Then $mor_{s,f}$ is locally Lipschitz continuous.*

[1] $a_+ = a$, if $a \ge 0, = 0$ else.

Proof:
This follows from the fact that $\text{mor}_{s,f}$ is convex with $\text{dom}(\text{mor}_{s,f}) = \mathcal{X}$; see Theorem 10.10. ∎

Theorem 8.8. *Let $\mathcal{X}$ be a reflexive Banach space and let $f \in \Gamma_0(\mathcal{X})$. If f is bounded from below then $\lim_{s\downarrow 0} \text{mor}_{s,f}(x) = f(x)$ for all $x \in \mathcal{X}$.*

Proof:
Let $x \in \mathcal{X}$. Clearly, when $f(x) = \infty$ then the result is true since $\lim_{s\downarrow 0} \frac{1}{2s}\|u - x\|^2 = \infty$ for all $u \in \mathcal{X}\backslash\{x\}$.
Let $x \in \text{dom}(f)$ and let $b \leq f(u)$ for all $u \in \mathcal{X}$. Then

$$f(w) + \frac{1}{2s}\|w - x\|^2 \geq b + \frac{1}{2s}\|w - x\|^2 > f(x) \text{ if } \|w - x\|^2 > 2s(f(x) - b).$$

Let $r_s := (2s(f(x) - b))^{\frac{1}{2}}$. Then $\lim_{s\downarrow 0} r_s = 0$ and since $\text{mor}_{s,f}(x) \leq f(x)$, we have

$$\text{mor}_{s,f}(x) = \inf_{w \in \overline{B}_{r_s}(x)} (f(w) + \frac{1}{2s}\|w - x\|^2)$$

and hence

$$\text{mor}_{s,f}(x) \geq \inf_{w \in \overline{B}_{r_s}(x)} f(w).$$

This implies

$$\begin{aligned} f(x) &\geq \limsup_{s\downarrow 0} \text{mor}_{s,f}(x) \geq \liminf_{s\downarrow 0} \text{mor}_{s,f}(x) \\ &\geq \liminf_{s\downarrow 0} \inf_{w \in \overline{B}_{r_s}(x)} f(w) = \liminf_{w \to x} f(w) \\ &\geq f(x) \end{aligned}$$

where we used the fact that f is lower semicontinuous. Thus, the result is proved. ∎

We now briefly describe some basic interpretations of the proximal operator that we will revisit in more detail later. Here, we restrict ourselves to considerations in Hilbert spaces.

Let $\mathcal{H}$ be a Hilbert space and let $f \in \Gamma_0(\mathcal{H})$. The definition of the proximal operator indicates that $\text{prox}_f(x)$ is a point that compromises between minimizing f and being near to x. In prox_{sf}, the parameter s can be interpreted as a relative weight or trade-off parameter between these terms.

When f is the indicator function

$$\delta_C(x) := \begin{cases} 0 & , x \in C \\ \infty & , x \notin C \end{cases}$$

where C is a nonempty closed convex set, the proximal operator of f reduces to the projection onto C, for which we have the denotation

$$P_C(x) := \text{argmin}_{u \in C}\|x - u\|^2$$

Proximal operators can thus be viewed as generalized projections, and this perspective suggests various properties that we expect proximal operators to obey.

The proximal operator of f can also be interpreted as a kind of gradient step for the function f. In particular, we have under the assumption that f is differentiable with gradient ∇f (we will consider this in the next sections more detailed) for the minimizer u^* of $u \longmapsto f(u) + \frac{1}{2s}\|u - x\|^2$ the necessary condition

$$(s\nabla f + I)(u^*) = x \text{ or alternatively } u^* = (s\nabla f + I)^{-1}(x)\,.$$

From this we conclude that

$$\mathrm{prox}_{s,f}(x) \approx x - s\nabla f(x)$$

when s is small. This suggests a close connection between proximal operators and gradient methods, and also hints that the proximal operator may be useful in optimization. It also suggests that the scaling parameter s will play a role similar to a step size in a gradient method.

Finally, the fixed points of the proximal operator of f are precisely the minimizers of f; see Lemma 8.4. This implies a close connection between proximal operators and fixed point theory, and suggests that proximal algorithms can be interpreted as solving optimization problems by finding fixed points of appropriate operators. Since minimizers of f are fixed points of $\mathrm{prox}_{s,f}$, we can minimize f by finding a fixed point of its proximal operator. It turns out that while $\mathrm{prox}_{s,f}$ need not be a contraction (unless f is strongly convex), it does have a different property, firmly nonexpansiveness, a sufficient property for efficient fixed point iteration. Firmly nonexpansive operators are special cases of nonexpansive operators. This immediately suggests the simplest proximal method

$$x^{k+1} := \mathrm{prox}_{s,f}(x^k)$$

which is called **proximal minimization** or the **proximal point algorithm**. In practice, the parameter s will be replaced by a sequence $(s_k)_{k\in\mathbb{N}_0}$ which should have the property $\lim_k s_k = 0$. Then we end up with the iteration

$$x^{k+1} := \mathrm{prox}_{s_k,f}(x^k)\,,\ k \in \mathbb{N}_0\,. \tag{8.6}$$

This is the proximal point algorithm, as introduced by Martinet first [23, 24] and later generalized by Rockafellar [33]. In the realization of the iteration method (8.6) all the methods of fixed point iteration and alternate projection methods may be applied. We will study a few specific methods.

The consideration of the proximal operator may be embedded in a larger context, namely the theory of monotone and maximal monotone operators. Monotone operators are setvalued mappings. The minimization of a proper lower semicontinuous convex function $f : \mathcal{X} \longrightarrow \hat{\mathbb{R}}$ may be formulated as follows:

$$\text{Find } x \in \mathcal{X} \text{ with } \theta \in \partial f(x) \tag{8.7}$$

Since the subdifferential is a maximal monotone operator the problem (8.7) can be generalized.

$$\text{Given a mapping } A : \mathcal{X} \rightrightarrows \mathcal{X}^* \text{ find } x \in \mathcal{X} \text{ with } \theta \in A(x) \tag{8.8}$$

In the next section we present results of the theory of monotone operators. These results are then used to solve the equation (8.8) by an iteration method similar to the proximal point method above.

Regularization is an important tool to solve problems in a stable way. Problems which are in general not solvable without regularization are ill-posed problems; see Subsection 5.9. These are problems where the continuous dependence of the solution on the data of the problem does not hold.

Let be given an equation

$$Ax = y^0 \tag{8.9}$$

where A is a continuous injective mapping from the Hilbert space $\mathcal{H}$ into the Hibert space $\mathcal{K}$. Suppose that a solution x^0 exists. The equation is difficult to solve when the range of the operator A is dense in $\mathcal{K}$ but not closed. Then the inverse A^{-1} exists on the range of A but is not continuous.

If we want to find a solution of (8.9) by minimizing the defect $f(x) := \|Ax - y^0\|^2, x \in \mathcal{H}$, this minimization is not very stable to errors in the data y^0. Regularization along the lines of the proximal operator is called in the community of ill-poed problems **Tikhonov regularization**. We are then lead to the following problem:

$$\text{Minimize } \|Au - y^0\|^2 + \alpha\|u - x^*\|^2 \text{ subject to } x \in \mathcal{H}. \tag{8.10}$$

Here x^* is a given point in $\mathcal{H}$ and α is the **regularization parameter**. In practice, instead of y^0 we have at hand only an erroneous data vector y^ε with $\|y^\varepsilon - y^0\| \leq \varepsilon$. It is the goal of regularization to find a sequence of regularization parameters $\alpha(\varepsilon)$ such that for the solution $x^{\varepsilon,\alpha}$ of (8.10) with y^ε instead of y^0 holds:

$$\lim_{\varepsilon \to 0} x^{\varepsilon,\alpha(\varepsilon)} = x^0. \tag{8.11}$$

The choice of the regularization parameter strategy $\alpha(\varepsilon)$ depends heavily on the „degree of discontinuity“ of A^{-1}; see [3, 22].

Regularization has become an indispensable part of reconstruction of signals and images, modern machine learning algorithm and pattern recognition. These problems may be formulated in the euclidean space $\mathbb{R}^n$ by an optimization problem like

$$\text{Minimze } l(u) + r(u) \text{ subject to } u \in \mathbb{R}^n. \tag{8.12}$$

Here $l : \mathbb{R}^n \longrightarrow \hat{\mathbb{R}}$ is a **loss function/cost function** and $r : \mathbb{R}^n \longrightarrow \hat{\mathbb{R}}$ is the regularization term which may be decomposed into a sum of terms of different qualities. A loss function is used to measure the degree of a fit for data. In statistics, a loss function is used for parameter estimation, in machine learning a loss function might be a measure for the deviation of a real situation from a training set. An often used loss function is the l_2 norm in $\mathbb{R}^n$:

$$l(u) := \frac{1}{2}\|u - x\|^2 := \frac{1}{2}\sum_{i=1}^n |u_i - x_i|^2, \; u = (u_1, \dots, u_n), x = (x_1, \dots, x_n) \in \mathbb{R}^n. \tag{8.13}$$

Here, x is a reference point choosen a priori.

The regularization term is a function which should induce a quality of the minimzers. Very popular regularization functions are ($u = (u_1, \dots, u_n) \in \mathbb{R}^n$)

$$r(u) := \|u\|_1 := \sum_{i=1}^{n} |u_i| \,,\; r(u) := \|u\|_{\mathrm{TV}} := \sum_{i=1}^{n-1} |u_{i+1} - u_i| \,,\; r(u) := \|u\|_\infty := \max_{i=1,\dots,n} |u_i| \,.$$

Regularization by the function $\|\cdot\|_{\mathrm{TV}}$ called **total variation regularization** has become a standard method for jump or edge preserving reconstruction of signals and images; see for instance [38]. Regularization with the l_1-norm $\|\cdot\|_1$ is more robust to noise and to the presence of outliers than regularization by the l_2-norm; see [26].

Suppose we consider the loss function in (8.13). Then with **L^1-TV regularization** we end up with the following minimization problem:

$$\text{Minimize } \frac{1}{2}\|u - x\|^2 + \|u\|_1 + \|u\|_{\mathrm{TV}} \text{ subject to } u \in \mathbb{R}^n \,. \tag{8.14}$$

In practice, various weights are introduced to bring the regularization terms and the loss function into a balance. As we see, the minimizer of the problem is realized by the proximal operator $\mathrm{prox}_{\|\cdot\|_1 + \|\cdot\|_{\mathrm{TV}}}$. Fortunately, $\mathrm{prox}_{\|\cdot\}|_1 + \|\cdot\|_{\mathrm{TV}}}$ can be decomposed as follows:

$$\mathrm{prox}_{\|\cdot\|_1 + \|\cdot\|_{\mathrm{TV}}} = \mathrm{prox}_{\|\cdot\|_1} \circ \mathrm{prox}_{\mathrm{TV}} \,; \tag{8.15}$$

see [14] and [39] where the decomposition problem is considered from a systematically point of view. In practice, when several weights are introduced this decomposition cannot be used and one has to use a numerical algorithm to compute the proximal operator.

The discrete models above have their equivalents in the continuous case. We sketch this (a little sloppy) in the framework of image processing. Suppose we have an image f in a bounded set $\Omega \subset \mathbb{R}^2$, i.e. a mapping $f : \Omega \longrightarrow \mathbb{R}$. This image f may be corrupted by noise and we try to reconstruct the „clean" image behind.

Let $\|\cdot\|_p$ denote the norm in $L_p(\Omega), 1 \le p \le \infty$, and let $\|\cdot\|_{\mathrm{TV}}$ denote the TV-norm of a sufficient smooth g:

$$\|g\|_{\mathrm{TV}} := \int_\Omega \|\nabla g(\xi)\| d\xi \,.$$

In 1992 Rudin, Osher and Fatemi proposed the so called **ROF-model** – we call it the **ROF2-model** – for **image reconstruction/denoising**; see [34]:

$$\text{Minimize } \mathrm{rof2}(u, \alpha; f) := \frac{1}{2}\|u - f\|_2^2 + \alpha\|u\|_{\mathrm{TV}} \text{ subject to } u \in X_2 \tag{8.16}$$

where X_2 is a subspace of $L_2(\Omega)$ such that $\|u\|_{\mathrm{TV}}$ is well-defined. In the following, the **fidelity term**[2] $\frac{1}{2}\|u - f\|^2$ is replaced by the distance in $L_1(\Omega)$:

$$\text{Minimize } \mathrm{rof1}(u, \alpha; f) := \|u - f\|_1 + \alpha\|u\|_{\mathrm{TV}} \text{ subject to } u \in X_1 \tag{8.17}$$

where X_1 is a subspace of $L_2(\Omega)$ such that $\|u\|_{\mathrm{TV}}$ is well-defined; see [26]. We call this model the **ROF1-model**. Both models are convex problems, the ROF2-model has a

[2] „Image fidelity" refers to the ability of a process/model to render an image accurately, without any visible distortion or information loss. „Quality of an image" is a more subjective preference of images and is much more difficult to evaluate.

uniquely determined (global) minimzer, the ROF1-model may have many minimizer since the function to be minimized is not strictly convex.

To compare the different models test images like $f = \chi_{\overline{B}_r}$ in $\Omega := \overline{B}_R, 0 < r < R$, may be used. The efficiency may then be evaluated along of two lines: how good is the **contrast** 1 versus 0 and how good is the edge $\partial\overline{B}_r$ reconstructed. In general, the ROF1-model show better results. For example, the ROF1-model has the remarkable property that a minimizer u of rof1$(\cdot, \alpha; f)$ is the minimizer of rof1$(\cdot, \alpha; u)$: a denoised image u is considered as „clean".

8.2 Monotone operators

Let us begin with a few introductory remarks. We want to solve the equation

$$f(x) = y \quad (y \in \mathbb{R} \text{ given}) . \tag{8.18}$$

whwere $f : \mathbb{R} \longrightarrow \mathbb{R}$ is a given function. We may formulate the property „f is monotone increasing" as follows:

$$(f(u) - f(v))(u - v) \geq 0 \text{ for all } u, v \in \mathbb{R} . \tag{8.19}$$

Let additional

$$f \text{ be continuous} \tag{8.20}$$

We know that under assumptions (8.19), (8.20) the range of f is $\mathbb{R}$ or a convex interval of $\mathbb{R}$. So, if we want to have solvability of the equation (8.18) for all $y \in \mathbb{R}$ we need an additional assumption. We consider

$$\lim_{x \to \pm\infty} f(x) = \pm\infty , \tag{8.21}$$

a property which is called **coercivity**. Now, we conclude that the equation (8.18) is solvable under the assumptions (8.19), (8.20), (8.21). Under the additional assumption that f is **strictly increasing** which may be formulated by

$$(f(u) - f(v))(u - v) > 0 \text{ for all } u, v \in \mathbb{R}, u \neq v , \tag{8.22}$$

the solution of (8.18) is then uniquely determined. It is the goal to generalize these observations to the case of nonlinear mappings defined on a Banach spaces. We will see that this generalization was successful during the last 60 years.

Let $\mathcal{X}$ be a Banach space and let $A : \mathcal{X} \rightrightarrows \mathcal{X}^*$ be a setvalued mapping. We set

$$\begin{aligned} \text{dom}(A) &:= \{x \in \mathcal{X} : A(x) \neq \emptyset\} \\ \text{gra}(A) &:= \{(x, \lambda) \in \mathcal{X} \times \mathcal{X}^* : x \in \mathcal{X}, \lambda \in A(x)\} \\ \text{ran}(A) &:= \{\lambda \in \mathcal{X}^* : \lambda \in A(x) \text{ for some } x \in \mathcal{X}\} . \end{aligned}$$

We call dom(A) the **effective domain of** A. gra(A) is the **graph of** A and ran(A) is the **range of** A. From the definition of gra(A^{-1}) we read off the definition of the **inverse** of the set-valued mapping A:

$$A^{-1}(\lambda) := \{x \in \mathcal{X} : \lambda \in A(x)\} , \lambda \in \mathcal{X}^* .$$

Notice

$$\mathrm{dom}(A^{-1}) = \mathrm{ran}(A)\,,\ \mathrm{ran}(A^{-1}) = \mathrm{dom}(A)\,.$$

A simple example of a setvalued mapping is the sign-operator on $\mathbb{R}$:

$$A(x) := \mathrm{sign}(x) := \begin{cases} -1 & ,x<0 \\ [-1,1] & ,x=0 \\ 1 & ,x>0 \end{cases}.$$

Definition 8.9. *Let $\mathcal{X}$ be a Banach space and let $A : \mathcal{X} \rightrightarrows \mathcal{X}^*$ be a setvalued mapping.*

(a) A *is called* **monotone** *if*

$$\langle \lambda - \mu, x - y\rangle \geq 0 \textit{ for all } \lambda \in A(x), \mu \in A(y), x, y \in \mathcal{X}\,.$$

(b) A *is called* **strictly monotone** *if*

$$\langle \lambda - \mu, x - y\rangle > 0 \textit{ for all } \lambda \in A(x), \mu \in A(y), x, y \in \mathcal{X}\,,\ x \neq y\,.$$

(c) A *is called* **strongly monotone** *if there exists* $b > 0$ *such that*

$$\langle \lambda - \mu, x - y\rangle \geq b\|x-y\|^2 \textit{ for all } \lambda \in A(x), \mu \in A(y), x, y \in \mathcal{X}\,.$$

(d) A *is called* **maximal monotone** *if* A *is monotone and if its* **graph** *is not properly contained in the graph of any other monotone operator* $A' : \mathcal{X} \rightrightarrows \mathcal{X}^*$, *i.e. if* $A' : \mathcal{X} \rightrightarrows \mathcal{X}^*$ *is a monotone operator with gra*$(A) \subset$ *gra*(A') *then* $A(x) = A'(x)$ *for all* $x \in \mathcal{X}$.

(e) A *is called* **coercive** *if dom*(A) *is bounded or if dom*(A) *is unbounded and*

$$\lim_{\|x\|\to\infty} \frac{\inf\{\langle\lambda, x\rangle : \lambda \in A(x)\}}{\|x\|} = \infty\,.$$

□

By applying Zorn's Lemma it is easy to see that every monotone operator posesses a maximal monotone extension. Notice that in the definition of monotone operators in Hilbert spaces the canonical pairing $\langle\cdot,\cdot\rangle$ is replaced by the inner product $\langle\cdot|\cdot\rangle$. Clearly, the identity on a Hilbert space is a maximal monotone operator.

We will see that subdifferentials of proper lowersemicontinuous convex functions are maximal monotone operators. As a rule, whenever a property is valid for subdifferentials in arbitrary Banach spaces there is some hope that it also holds for all maximal monotone operators. The following considerations demonstrate this fact quite valid.

The sign-operator is maximal monotone. It is the maximal extension of

$$A(x) := \begin{cases} -1 & ,x<0 \\ 0 & ,x=0 \\ 1 & ,x>0 \end{cases}.$$

If $f : \mathbb{R} \longrightarrow \mathbb{R}$ is monotone increasing then we may define a mapping $\overline{f} : \mathbb{R} \rightrightarrows \mathbb{R}$ by

$$\overline{f}(x) := [f(x-), f(x+)] \text{ where } f(x-) := \sup_{t<x} f(t), f(x+) := \inf_{t>x} f(t)\,,\ x \in \mathbb{R}\,.$$

$\overline{f} : \mathbb{R} \rightrightarrows \mathbb{R}$ is a maximal monotone operator.

Lemma 8.10. *Let $\mathcal{H}$ be a Hilbert space and let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a mapping. Then the following statements are equivalent:*

(a) A *is maximal monotone operator.*

(b)

$$\|x - y\| \le \|(x + tu) - (y + tv)\| \text{ for all } (x, u), (y, v) \in gra(A) \text{ and } t \ge 0\,. \quad (8.23)$$

Proof:
Let $(x, u), (y, v) \in \mathcal{H} \times \mathcal{H}, t \ge 0$. For $t = 0$ nothing has to be proved. So, assume $t > 0$. Then

$$\frac{1}{t}(\|(x + tu) - (y + tv)\|^2 - \|x - y\|^2) = 2\langle u - v|x - y\rangle + t\|u - v\|^2\,.$$

Now, the equivalence of (a) and (b) is clear. ■

The property (8.23) may be interpreted as nonexpansivity of the mapping $(I + tA)^{-1}$. Notice that this mapping is single-valued due to this property.

Lemma 8.11. *Let $\mathcal{X}$ be a reflexive Banach space and let $A : \mathcal{X} \rightrightarrows \mathcal{X}^*$ be a monotone operator. Then the followings statements are equivalent:*

(a) A *is a maximal monotone operator.*

(b) *For all* $(x, \lambda) \in \mathcal{X} \times \mathcal{X}^*$ *the condition*

$$\langle \lambda - \mu, x - y\rangle \ge 0 \text{ for all } (y, \mu) \in gra(A) \text{ i.e. } \inf_{(y,\mu)\in gra(A)} \langle \lambda - \mu, x - y\rangle \ge 0 \quad (8.24)$$

implies $\lambda \in A(x)$.

Proof:
Let $(x, \lambda) \in \mathcal{X} \times \mathcal{X}^*$. We set

$$\tilde{A}(x) := A(x) \cup \{\lambda\},\ \tilde{A}(z) := A(z) \text{ for } z \ne x\,.$$

Then $\tilde{A}$ is an extension of A which is monotone due to (8.24). ■

Lemma 8.12. *Let $\mathcal{X}$ be a reflexive Banach space and let $A : \mathcal{X} \rightrightarrows \mathcal{X}^*$ be a maximal monotone operator. Then A is strong-weak closed, i.e.*

$$(x^n, \lambda^n) \in gra(A), n \in \mathbb{N}, x = \lim_n x^n, \lambda = w - \lim_n \lambda^n \text{ implies } (x, \lambda) \in gra(A)\,.$$

Proof:
Let $(y, \mu) \in \mathrm{gra}(A)$. Then due to the monotonicity

$$\langle \lambda^n - \mu, x^n - y\rangle \ge 0, n \in \mathbb{N}, \text{ and } \lim_n \langle \lambda^n - \mu, x^n - y\rangle = \langle \lambda - \mu, x - y\rangle$$

since $(x^n)_{n\in\mathbb{N}}$ converges strongly and $(\lambda^n)_{n\in\mathbb{N}}$ converges weakly. Since A is a maximal monotone operator we conclude $\lambda \in A(x)$ by Lemma 8.11. ■

Lemma 8.13. *Let $\mathcal{X}$ be a reflexive Banach space and let $A : \mathcal{X} \rightrightarrows \mathcal{X}^*$ be a maximal monotone operator. Then A is weak-strong closed, i.e.*

$$(x^n, \lambda^n) \in gra(A), n \in \mathbb{N}, x = w - \lim_n x^n, \lambda = \lim_n \lambda^n \text{ implies } (x, \lambda) \in gra(A)\,.$$

Proof:
This can be proved similar to Lemma 8.12. ■

Theorem 8.14 (Minty's trick). *Let $\mathcal{X}$ be a reflexive Banach space and let $A : \mathcal{X} \rightrightarrows \mathcal{X}^*$ be a maximal monotone operator. Let $(x^n)_{n\in\mathbb{N}}, (\lambda^n)_{n\in\mathbb{N}}$ be sequences in $\mathcal{X}$ and $\mathcal{X}^*$ respectively and let $x \in \mathcal{X}, \lambda \in \mathcal{X}^*$. Assume*

(1) $u^n \in A(x^n), n \in \mathbb{N}$.

(2) $x = w - \lim_n x^n, u = w - \lim_n u^n$.

(3) $\limsup_n \langle \lambda^n, x^n \rangle \leq \langle \lambda, x \rangle$.

Then $\lambda \in A(x)$ *and* $\lim_n \langle \lambda^n, x^n \rangle = \langle \lambda, x \rangle$.

Proof:
Let $(y, \mu) \in \text{gra}(A)$. Since A is monotone we have

$$0 \leq \langle \lambda^n - \mu, x^n - y \rangle = \langle \lambda^n, x^n \rangle - \langle \lambda^n, y \rangle - \langle \mu, x^n \rangle + \langle \mu | y \rangle .$$

Tacking the lim sup we conclude using the reflexivity of $\mathcal{X}$

$$0 \leq \limsup_n \langle \lambda^n, x^n \rangle - \langle \lambda, y \rangle - \langle \mu | x \rangle + \langle \mu, y \rangle \leq \langle \lambda - \mu, x - y \rangle .$$

Since A is maximal monotone this implies $\lambda \in A(x)$. Moreover,

$$0 \leq \langle \lambda^n - \lambda | x^n - x \rangle = \langle \lambda^n, x^n \rangle - \langle \lambda^n, x \rangle - \langle \lambda, x^n \rangle + \langle \lambda, x \rangle .$$

Tacking the lim inf we obtain $\langle \lambda, x \rangle \leq \liminf_n \langle \lambda^n, x^n \rangle$. Hence, $\lim_n \langle \lambda^n, x^n \rangle = \langle \lambda, x \rangle$. ■

Lemma 8.15. *Let $\mathcal{X}$ be a reflexive Banach space and let $A : \mathcal{X} \rightrightarrows \mathcal{X}^*$ be a maximal monotone operator. Let $x \in \mathcal{X}, \lambda \in \mathcal{X}^*$, let $(x^n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{X}$, and let $(\lambda^n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{X}^*$ such that $\lambda^n \in A(x^n), n \in \mathbb{N}, x = w - \lim_n x^n, \lambda = w - \lim_n \lambda^n$, and*

$$\limsup_{m,n} \langle \lambda^n - \lambda^m, x^n - x^m \rangle \leq 0 . \tag{8.25}$$

Then $(x, \lambda) \in \text{gra}(A)$ *and* $\langle \lambda, x \rangle = \lim_n \langle \lambda^n, x^n \rangle$.

Proof:
We folllow [8].
Due to the monotonicity and (8.25)

$$\lim_{m,n} \langle \lambda^n - \lambda^m, x^n - x^m \rangle = 0 .$$

$(\langle \lambda^{n_k}, x^{n_k} \rangle)_{n\in\mathbb{N}}$ is bounded since each sequence $(x^n)_{n\in\mathbb{N}}, (\lambda^n)_{n\in\mathbb{N}}$ is bounded. Let $(n_k)_{k\in\mathbb{N}}$ be a subsequence such that $L := \lim_k (\langle \lambda^{n_k}, x^{n_k} \rangle)$ exists. Then

$$\begin{aligned} 0 &= \lim_i \lim_k \langle \lambda^{n_i} - \lambda^{n_k}, x^{n_i} - x^{n_k} \rangle \\ &= \lim_i (\langle \lambda^{n_i}, x^{n_i} \rangle - \langle \lambda, x^{n_i} \rangle - \langle \lambda^{n_i}, x \rangle + L) \\ &= 2L - 2\langle \lambda, x \rangle . \end{aligned}$$

Hence $L = \langle \lambda, x \rangle$ and therefore (since L is uniquely determined) $\lim_n \langle \lambda^n, x^n \rangle = \langle \lambda, x \rangle$. Let $(y, \mu) \in \text{gra}(A)$. Then

$$\begin{aligned} \langle \lambda - \mu, x - y \rangle &= \langle \lambda, x \rangle - \langle \lambda, y \rangle - \langle \mu, x \rangle + \langle \mu, y \rangle \\ &= \lim_n \langle \lambda^n, x^n \rangle - \langle \lambda^n, y \rangle - \langle \mu, x^n \rangle + \langle \mu, y \rangle \\ &= \lim_n \langle \lambda^n - \mu, x^n - y \rangle \ \geq \ 0 \end{aligned}$$

This implies $\lambda \in A(x)$ by Lemma 8.11. ∎

Corollary 8.16. *Let $\mathcal{X}$ be a uniformly convex Banach space with duality mapping $J_{\mathcal{X}}$. Let $x \in \mathcal{X}, \lambda \in \mathcal{X}^*$, let $(x^n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{X}$ and let $(\lambda^n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{X}^*$ with $\lambda^n \in J_{\mathcal{X}}(x^n), n \in \mathbb{N}$. Suppose*

$$\limsup_n \langle \lambda^n - \lambda, x^n - x \rangle \leq 0 . \tag{8.26}$$

Then $\lambda \in J_{\mathcal{X}}(x)$, $\langle \lambda, x \rangle = \lim_n \langle \lambda^n, x^n \rangle$, and $x = \lim_n x^n$.

Proof:
$J_{\mathcal{X}}$ is a single-valued maximal monotone operator. Then from Lemma 8.15 we obtain $\langle J_{\mathcal{X}}(x), x \rangle = \lim_n \langle J_{\mathcal{X}}(x^n), x^n \rangle$. Moreover

$$\|x^n\|^2 = \langle J_{\mathcal{X}}(x^n), x^n \rangle, n \in \mathbb{N}, \langle J_{\mathcal{X}}(x), x \rangle = \|x\|^2 .$$

Now we have $\lim_n \|x^n\| = \|x\|$ and this implies in a uniformly convex Banach space $\lim_n x^n = x$. ∎

Let $\mathcal{X}$ be a Banach space. We define for $A : \mathcal{X} \rightrightarrows \mathcal{X}^*, B : \mathcal{X} \rightrightarrows \mathcal{X}^*$ aA $(a \in \mathbb{R}), A + B, \overline{\text{co}}(A)$ as follows:

$$\begin{aligned} (aA)(x) &:= aA(x) , \ x \in \mathcal{X} , \\ (A + B)(x) &:= A(x) + B(x) , \ x \in \mathcal{X} , \\ \overline{\text{co}}(A)(x) &:= \overline{\text{co}(A(x))} , \ x \in \mathcal{X} . \end{aligned}$$

Here we use the convention that the sum $M + N$ of two sets is empty is one of the sets M or N is empty.

Lemma 8.17. *Let $\mathcal{X}$ be a Banach space, $A, B : \mathcal{X} \rightrightarrows \mathcal{X}^*$ be monotone operators, and let $a \geq 0$. Then $aA, A + B$ and $\overline{\text{co}}(A)$ are monotone.*

Proof:
Follows from the definition of monotonicity in a simple way. ∎

Corollary 8.18. *Let $\mathcal{X}$ be a reflexive Banach space and let $A : \mathcal{X} \rightrightarrows \mathcal{X}^*$ be a maximal monotone operator. Then aA $(a > 0), \overline{\text{co}}(A)$ are maximal monotone operators. Moreover, $A(x)$ is convex and closed for all $x \in \text{dom}(A)$.*

Proof:
The assertion concerning aA is obvious. Due to Lemma 8.17 $\overline{\text{co}}(A)$ is a monotone operator. Since $x \longmapsto \overline{\text{co}}(A)(x)$ defines an extension of A we conclude $A = \overline{\text{co}}(A)$ and $\overline{\text{co}}(A)$ is maximal monotone. Since $A = \overline{\text{co}}(A)$ all sets $A(x)$ are convex. The closedness of $A(x)$ for each $x \in \text{dom}(A)$ follows from an easy application of Lemma 8.13. ∎

Corollary 8.19. *Let $\mathcal{X}$ be a reflexive Banach space and let $A : \mathcal{X} \rightrightarrows \mathcal{X}^*$ be a maximal monotone operator. Then $A^{-1}(\lambda)$ is convex and closed for all $\lambda \in ran(A)$.*

Proof:
Let $x, z \in A^{-1}(\lambda)$. Let $(y, \mu) \in \text{gra}(A)$ and let $t \in [0, 1]$. Then due to the monotonicity of A we have

$$\langle \lambda - \mu, tx + (1-t)z - y\rangle = t\langle \lambda - \mu, x - y\rangle + (1-t)\langle \lambda - \mu, z - y\rangle \geq 0\,.$$

Since A is a maximal monotone operator this implies $\lambda \in A(tx+(1-t)z)$, i.e. $tx+(1-t)z \in A^{-1}(\lambda)$. Thus, the convexity of $A^{-1}(\lambda)$ is proved.
Let $\lambda \in \text{ran}(A)$. The closedness of $A^{-1}(\lambda)$ is proved by applying Lemma 8.12. ■

Notice, the sum of maximal monotone operators must not be a maximal monotone operator. A sufficient condition for this property is $\text{int}(\text{dom}(A)) \cap \text{dom}(B) \neq \emptyset$; see [31].

Lemma 8.20. *Let $\mathcal{X}$ be a uniformly convex Banach space with duality mapping J_X and let $A : \mathcal{X} \rightrightarrows \mathcal{X}^*$ be a maximal monotone operator. Then for every $\lambda \in \mathcal{X}^*$ the mapping*

$$g : (0, \infty) \ni t \longmapsto (J_X + tA)^{-1}(\lambda) \in \mathcal{X}$$

is continuous. If $\lambda = J_X(x)$ for some $x \in dom(A)$, then g is also continuous in $t = 0$ and we have $\lim_{t\to 0} g(t) = x$.

Proof:
We follow [21].
Suppose $t_0 > 0$. Let $0 < a < t_0 < b$ and set

$$x_t := (J_{\mathcal{X}} + tA)^{-1}(\lambda)\,,\ t \in (a, b)\,.$$

We have $\lambda = (J_{\mathcal{X}} + tA)(x_t) = J_{\mathcal{X}}(x_{t_0}) + t_0 x_{t_0}$ with $\mu_t \in A(x_t), \mu_{t_0} \in A(x_{t_0})$ and this implies

$$\begin{aligned} 0 &= \langle (J_{\mathcal{X}}(x_t) + t\mu_t) - (J_{\mathcal{X}}(x_{t_0}) + t_0\mu_{t_0}), x_t - x_{t_0}\rangle \\ &= t\langle \mu_t - \mu_{t_0}, x_t - x_{t_0}\rangle + (t - t_0)\langle \mu_{t_0}, x_t - x_{t_0}\rangle + \langle J_{\mathcal{X}}(x_t) - J_{\mathcal{X}}(x_{t_0}), x_t - x_{t_0}\rangle\,. \end{aligned}$$

Hence

$$(\|x_t\| - \|x_{t_0}\|)^2 \leq |t - t_0|\|\mu_{t_0}\|(\|x_{t_0}\|^2 + \|\lambda\|^2)\,.$$

Here we have used (3.4). Now, we obtain $(\|x_t\|)_{t\in(a,b)}$ is bounded. Letting c be an upper bound for $(\|x_t\|)_{t\in(a,b)}$ we obtain from the inequalities above

$$\langle (J_{\mathcal{X}}(x_t) - (J_{\mathcal{X}}(x_{t_0})), x_t - x_{t_0}\rangle \leq (c + \|x_{t_0}\|)|t - t_0|\|\lambda\|\,,\ t \in (a.b)\,.$$

Then by Corollary 8.16 $\lim_{t\to t_0} x_t = x_{t_0}$.
The additional result follows by inspection of the proof above. ■

Definition 8.21. *Let $\mathcal{X}$ be a Banach space and let $A : \mathcal{X} \longrightarrow \mathcal{X}^*$ be a mapping.*[3]
A is called **hemicontinuous** *if for all $x, y \in \mathcal{X}$ the mapping*

$$[0, 1] \ni t \longmapsto \langle A(ty + (1-t)x), y - x\rangle \in \mathbb{R}$$

is continuous in $t = 0$. □

[3]If an operator is not setvalued like $A : \mathcal{X} \longrightarrow \mathcal{X}^*$ then the effective domain of definition is the the whole space.

Theorem 8.22. *Let $\mathcal{X}$ be a reflexive Banach space and let $A : \mathcal{X} \longrightarrow \mathcal{X}^*$ be a monotone hemicontinuous operator. Then A is a maximal monotone operator.*

Proof:
Let $(x, \lambda) \in \mathcal{X} \times \mathcal{X}^*$ and let

$$\langle \lambda - A(u), x - u \rangle \geq 0 \text{ for all } u \in \mathcal{X}\,.$$

We set $u := x \pm tw$ with $w \in \mathcal{X}$ and $t > 0$. Then we obtain

$$\pm t\langle \lambda - A(x \pm tw), -w \rangle \geq 0$$

and hence

$$\langle \lambda - A(x + tw), w \rangle \leq 0\,,\ \langle \lambda - A(x - tw), w \rangle \geq 0$$

Letting $t \to 0$ we obtain $\langle \lambda - A(x), w \rangle = 0$ since A is hemicontinuous. This shows $\lambda = A(x)$ since $w \in \mathcal{X}$ is arbitrary chosen. With Lemma 8.17 we obtain that A is a maximal monotone. ■

Monotone operators are important objects in modern optimization and analysis. Here are core examples for monotone operators:

Subgradients Subdifferentials of convex lower semicontinuous functions are examples of maximal monotone operators; see below. A very important example is the duality mapping which captures many geometric aspects of Banach spaces. For the duality mapping we have already schown its monotonicity; see (5) in Lemma 3.6.

Skew linear operators We will see examples of this class of monotone operators below.

Laplacian operators Using monotone operators we are able to find weak solutions in Sobolev spaces of (nonlinear) elliptic and parabolic partial differential equations. With these type of operators we may describe physical phenomena like friction, internal forces, additional constraints, ...

Example 8.23. *Let $\mathcal{H}$ be a Hilbert space, let $y \in \mathcal{H}$, and let $T : \mathcal{H} \longrightarrow \mathcal{H}$ be a linear continuous operator. Then the (affine) mapping $A : \mathcal{H} \ni x \longmapsto Tx + y \in \mathcal{H}$ is monotone if the symmetric part $T + T^*$ is nonnegative, i.e.*

$$\langle (T + T^*)z|z \rangle \geq 0 \textit{ for all } z \in \mathcal{H}\,.$$

Actually, A is maximal monotone. This follows by using Theorem 8.22.
The skewsymmetric part $S_T := T - T^*$ is – due to the fact $\langle S_T z|z \rangle = 0$ for all $z \in \mathcal{H}$ – a monotone operator. Actually, S_T is maximal monotone. Again, this follows by using Theorem 8.22. □

Example 8.24. *Consider*[4]

$$A(u) := -div(|\nabla u|^{p-2} \nabla u)$$

[4]This is an example for readers which are familiar with the theory of weak solutions for elliptic equations in Sobolev spaces.

for functions $u \in C_0^\infty(\Omega)$;[5] *with* $\Omega \subset \mathbb{R}^n$. *Let* $W^{1,p}$ *be the space of functions with a distributional derivative of order zero and one in* $L_p(\Omega)$. *This becomes a Banach space endowed with the sum of* L_p*-norms of all derivatives. The functions in* $W^{1,p}$ *with zero trace on the boundary* $\partial\Omega$ *is denoted by* $W_0^{1,p}$. *The dual space of* $W_0^{1,p}$ *is* $W^{-1,q}$ *where* $1/p + 1/q = 1$. *Then* A' *has an extension to the* p*-Laplacian* Δ_p *on the space* $W_0^{1,p}$*:*

$$\Delta_p u := div(|\nabla u|^{p-2}\nabla u)$$

Thus the operator describes (for $f \in W^{-1,p}(\Omega)$*)*

$$-div(|\nabla u|^{p-2}\nabla u) = f, \ \text{in } \Omega, \ u = \theta \ \text{in } \partial\Omega. \tag{8.27}$$

The operator Δ_p *(called also the* p*-harmonic operator) may be used to describe models for fluids. Three cases are of importance:*

$p = 2$ *This is the case of newtonian fluids (air, water,...) and* Δ_2 *is the well known Laplace operator.*

$p > 2$ *The viscosity (of oil,...) is a monotone increasing function of the gradient* ∇u.

$p < 2$ *The viscosity (of blood,...) is a monotone decreasing function of the gradient* ∇u.

Another field of application is located in image processing.

It can be shown that Δ_p *is a maximal monotone operator. To study this some special inequalities are helpful. Expressions like*

$$\langle |\nabla v|^{p-2}\nabla v - |\nabla u|^{p-2}\nabla u | \nabla v - \nabla u\rangle$$

can be treated by using inequalities resulting from

$$\langle \|b\|^{p-2}b - \|a\|^{p-2}a | b - a\rangle \frac{1}{2}(\|b\|^{p-2} + \|a\|^{p-2})\|b-a\|^2 + \frac{1}{2}(\|b\|^{p-2} - \|a\|^{p-2})(\|b\|^2 - \|a\|^2)$$

for vectors in $\mathbb{R}^n$. *Here are two results:*

$p \geq 2$ $\langle |b|^{p-2}b - |a|^{p-2}a | b - a\rangle \geq 2^{-1}(|b|^{p-2} + |a|^{p-2})|b-a| \geq 2^{2-p}|b-a|^p$.

$p \leq 2$ $\langle |b|^{p-2}b - |a|^{p-2}a | b - a\rangle \leq \frac{1}{2}(|b|^{p-2} + |a|^{p-2})|b-a|^2$.

Δ_p *has different properties depending on the numbers* p, n.

$p = 1$ *We have* $\Delta_1 u = div(\frac{\nabla u}{|\nabla u|})$. $\Delta_1 u$ *describes the mean curvature of* u.

$p = 2$ Δ_2 *is the usual Laplace operator.*

$p > n$ *Due to embedding theorems for Sobolev spaces solutions of* (8.27) *show good continuity properties (if* f *is smooth).*

$p = n$ *This is the critical case.*

$p = \infty$ *As* $p \to \infty$ *one encounters the* p*-Laplacian becomes*

$$\Delta_\infty u = \sum_{i,j=1}^{n} \frac{\partial u}{\partial x_i}\frac{\partial u}{\partial x_j}\frac{\partial^2 u}{\partial x_i \partial x_j}.$$

It can be used to describe certain phenomena in image processing.

□

[5] $|\cdot|$ denotes the euclidean norm in $\mathbb{R}^n$

8.3 Solving equations governed by a maximal monotone operator

Let $\mathcal{X}$ be a Banach space and let $A : \mathcal{X} \rightrightarrows \mathcal{X}^*$. The question whether the equation

$$\theta \in A(x) \text{ for some } x \in \mathcal{X} \tag{8.28}$$

has a solution is of interest especially in the case when A is the subdifferential of a function $f \in \Gamma_0(\mathcal{X})$.

Theorem 8.25. *Let $\mathcal{X}$ be a uniformly convex Banach space with duality mapping $J_{\mathcal{X}}$. Let $A : \mathcal{X} \rightrightarrows \mathcal{X}^*$ be a maximal monotone operator. Then the following conditions are equivalent:*

(a) $\theta \in \mathit{ran}(A)$.

(b) There exists an open and bounded set $G \subset \mathcal{X}$ and $x^0 \in G \cap \mathit{dom}(A)$ with[6]

$$\langle \lambda, x - x^0 \rangle \geq 0 \textit{ for all } x \in \partial G \cap \mathit{dom}(A), \lambda \in A(x).$$

(c) There exists an open and bounded set $G \subset \mathcal{X}$ and $x^0 \in G \cap \mathit{dom}(A)$ with

$$s(J_{\mathcal{X}}(x) - J_{\mathcal{X}}(x^0)) \notin A(x) \textit{ for all } s > 0, x \in \partial G \cap \mathit{dom}(A).$$

Proof:
We follow [21].
Ad $(a) \implies (b)$ Let $\theta \in A(x^0)$ with $x^0 \in \mathrm{dom}(A)$ and let G be an open set with $x^0 \in G$. Suppose $x \in \partial G \cap \mathrm{dom}(A), \lambda \in A(x)$. Since A is monotone we have $\langle \lambda, x - x^0 \rangle = \langle \lambda - \theta, x - x^0 \rangle \geq 0$.
Ad $(b) \implies (c)$ We want to show (c) with the same set G and x^0 as in (b). Assume that (c) does not hold. Then there exists $s > 0$ and $x \in \partial G \cap \mathrm{dom}(A)$ with

$$\lambda = s(J_{\mathcal{X}}(x) - J_{\mathcal{X}}(x^0)) \text{ for some } \lambda \in A(x).$$

Since $J_{\mathcal{X}}$ is strongly monotone we obtain by (b) in

$$0 \leq \langle \lambda, x - x^0 \rangle = s\langle J_{\mathcal{X}}(x) - J_{\mathcal{X}}(x^0), x - x^0 \rangle < 0$$

a contradiction.
Ad $(c) \implies (a)$ We observe that the mapping

$$g : [0, \infty) \ni t \longmapsto \langle (J_{\mathcal{X}} + tA)^{-1} J_{\mathcal{X}}(x^0) \in \mathrm{dom}(A) \subset \mathcal{X}$$

is continuous in $(0, \infty)$; see Lemma 8.20. Then (c) says that $g(t) \notin \partial G$ for any $t \in (0, \infty)$. Since $g(0) = x^0 \in G$ we conclude that $g(t) \in G, t \in [0, \infty)$. Due to the boundeness of G there exists $\kappa > 0$ with $\|g(t)\| \leq \kappa, t \in [0, \infty)$. Consider a sequence $(t_n)_{n \in \mathbb{N}}$ in $(0, \infty)$ with $\lim_n t_n = \infty$. Set $x^n := (J_{\mathcal{X}} + t_n A)^{-1} J_{\mathcal{X}}(x^0), n \in \mathbb{N}$. Then we have

$$\|x^n\| \leq \kappa,\ J_{\mathcal{X}}(x^0) \in J_{\mathcal{X}}(x^n) + t_n A(x^n), n \in \mathbb{N}.$$

[6] ∂G is the boundary of the set G

Thus, for some $\lambda^n \in A(x^n), n \in \mathbb{N}$, we have

$$\|\lambda\| = t_n^{-1}\|J_{\mathcal{X}}(x^n) - J_{\mathcal{X}}(x^0)\| \le t_n^{-1}(\kappa + \|x^0\|), n \in \mathbb{N}\,.$$

Since $\mathcal{X}$ is reflexive $(x^n)_{n\in\mathbb{N}}$ has a weak cluster point; let $x' = w - \lim_n x^{n_k}$. Since $\lim_n\langle\lambda^n, x^n - x'\rangle$ (notice $\lim_n \lambda^n = \theta$) we can apply Lemma 8.15 and obtain $x' \in \text{dom}(A)$ and $\theta \in A(x')$. ■

Corollary 8.26. *Let $\mathcal{X}$ be a uniformly convex Banach space and let $A : \mathcal{X} \rightrightarrows \mathcal{X}^*$ be a maximal monotone operator. Consider the condition*

$$\text{There exists } x^0 \in \mathcal{X}, r \ge 0, \text{ such that } \langle\lambda, x - x^0\rangle \text{ for all } (x,\lambda) \in gra(A), \|x - x^0\| \ge r\,. \tag{8.29}$$

Then $\theta \in ran(A)$.

Proof:
(8.29) implies the condition (b) in Theorem 8.25. ■

Corollary 8.27. *Let $\mathcal{X}$ be a uniformly convex Banach space and let $A : \mathcal{X} \rightrightarrows \mathcal{X}^*$ be a maximal monotone operator with a bounded domain of definition dom(A). Then $\theta \in$ ran(A).*

Proof:
The condition (8.29) in Corollary 8.26 is trivially satisfied. ■

Lemma 8.28. *Let $\mathcal{X}$ be a Banach space and let $A : \mathcal{X} \rightrightarrows \mathcal{X}^*$ be a maximal monotone operator. Then A is locally bounded at $x \in$ int(dom(A)), i.e. there exists $r > 0$ and $c > 0$ such that*

$$\sup_{\lambda\in A(y)} \|\lambda\| \le c \text{ for all } y \in \overline{B}_r(y)\,.$$

Proof:
Let $x \in \text{int}(\text{dom}(A))$. Without loss of generality we may assume $x = \theta$ and $\theta \in A(\theta)$. Define

$$f : \mathcal{X} \ni y \longmapsto \sup_{(u,v)\in\text{gra}(A),\|u\|\le 1} \langle y - u, v\rangle \in \hat{\mathbb{R}}\,.$$

Then $f \in \Gamma_0(\mathcal{X})$. Since $\theta \in \text{int}(\text{dom}(A))$, there exists $s > 0$ such that $\overline{B}_s \subset \text{dom}(f)$. Let $y \in \overline{B}_s$ and $w \in A(y)$. Then we have for all $(u,v) \in \text{gra}(A)$ with $\|u\| \le 1$

$$\langle y - u, w - v\rangle \ge 0\,,\ \langle y - u, w\rangle \ge \langle y - u, v\rangle\,,\ \langle y - u, v\rangle \le (\|y\| + 1)\|v\| < \infty\,.$$

This implies $f(y) < \infty$, i.e. $y \in \text{dom}(f)$. This shows $\overline{B}_s \subset \text{int}(\text{dom}(f))$. Then there exists $r > 0$ with $r \le \min(\frac{1}{2}, \frac{1}{2}s)$ such that

$$f(y) \le f(\theta) + 1 \text{ for all } y \in \overline{B}_{2r}\,.$$

Since $(\theta,\theta) \in \text{gra}(A)$ we have $f(\theta) \ge 0$. On the other hand, by the monotonicity of A,

$$\langle u, v\rangle = \langle u - \theta, v - \theta\rangle \ge 0 \text{ for all } (u,v) \in \text{gra}(A)\,.$$

Then we have by the definition of f the inequality $f(\theta) \le 0$. Alltogether, $f(\theta) = 0$. Thus,

$$\langle y, v\rangle \le \langle u, v\rangle + 1 \text{ for all } y \in \overline{B}_{2r}, (u,v) \in \text{gra}(A), \|u\| \le r\,.$$

Taking the supremum with respect to y we obtain for all $(u, v) \in \mathrm{gra}(A), \|u\| \le r$,

$$2r\|v\| \le \|u\| + 1 \le r\|v\| + 1\,,\ \|v\| \le r^{-1}\,.$$

Setting $c := r^{-1}$ the result is proved. ■

Maximal monotone opertors and subdifferentials show very similar properties. This suggests to ask whether the theory of maximal monotone operators and subdifferentials can be considered under a common point of view. This is nearly the case but subdifferential have a property which has not every maximal monotone operator: cyclic monotonicity.

Definition 8.29. *Let $\mathcal{X}$ be a Banach space and let $A : \mathcal{X} \rightrightarrows \mathcal{X}^*$ be a mapping. A is called* **n-cyclically monotone** *if the following holds: $\sum_{i=1}^{n} \langle \lambda^i, x^i - x^{i-1}\rangle \ge 0$ whenever $n \ge 2$ and $x^0, \dots, x^n \in \mathcal{X}, x^0 = x^n$, $\lambda^i \in A(x^i), i = 1, \dots, n$.*
A is called **cyclically monotone** *if A is n-cyclically monotone for every $n \in \mathbb{N}$.* □

Clearly, a 2-cyclically monotone operator is monotone. Subdifferentials are cyclically monotone but not every maximal monotone operator is n-cyclically monotone for $n \ge 3$.

Example 8.30. *Consider $A : \mathbb{R}^2 \ni (x, y) \longmapsto (y, -x) \in \mathbb{R}^2$. A is maximal monotone but not 3-cyclically monotone. To show this consider the vectors $(1, 1), (0, 1), (1, 0), (1, 1)$.* □

Let $\mathcal{X}$ be a Banach space and let $A : \mathcal{X} \rightrightarrows \mathcal{X}^*$ be a monotone operator. Then

$$\langle \lambda - \mu, x - y\rangle \ge 0 \text{ for all } (x, \lambda), (y, \mu) \in \mathrm{gra}(A)\,.$$

This implies

$$\langle \lambda, y\rangle + \mu, x\rangle - \mu, y\rangle \le \langle \lambda, x\rangle \text{ for all } (x, \lambda), (y, \mu) \in \mathrm{gra}(A)\,.$$

From this observation starts the following definition.

Definition 8.31. *Let $\mathcal{X}$ be a Banach space and let $A : \mathcal{X} \rightrightarrows \mathcal{X}^*$ be a monotone operator. The* **Fitzpatrick function** *associated with this operator is the function*

$$F_A : \mathcal{H} \times \mathcal{H} \ni (x, u) \longmapsto \sup_{y,v)\in gra(A)} (\langle v, x\rangle + \langle u, y\rangle - \langle v, y\rangle) \in \hat{\mathbb{R}}\,.$$

□

Fitzpatrick functions have been proved to be an important tool in modern monotone operator theory; see [12, 13, 29]. It connects the theory of convex functions with the theory of (maximal) monotone operators. It can be shown that a maximal monotone operator can be represented by a convex function. Nowadays it is an effective tool to study the existence problem for stochastic differential equations. This is inspired by the fact that stochastic differential equations where the nonlinearity is governed by a subdifferential operator allows existence results; see for instance [27].

In the following we study and use the Fitzpatrick function in a reflexive Banach space $\mathcal{X}$. Then the pairing between the spaces $\mathcal{X} \times \mathcal{X}^*$ and $\mathcal{X}^* \times \mathcal{X}^{**}$ may be identified with the pairing between $\mathcal{X} \times \mathcal{X}^*$ and $\mathcal{X}^* \times \mathcal{X}$. To provide more orientation in the following we prefer for the analysis and the use of the Fitzpatrick function the notion $x^*, y^*, \dots$ for functionals in $\mathcal{X}^*$.

Lemma 8.32. *Let $\mathcal{X}$ be a reflexive Banach space and let $A : \mathcal{X} \rightrightarrows \mathcal{X}^*$ be a monotone operator with dom$(A) \neq \emptyset$. Let F_A be the Fitzpatrick function of A. Then:*

(1) F_A *is proper*

(2) F_A *is convex.*

(3) $F_A \in \Gamma_0(\mathcal{X} \times \mathcal{X}^*)$ *where $\mathcal{X}, \mathcal{X}^*$ are endowed with the strong topologies.*

(4) $F_A(x, x^*) \leq \langle x^*, x\rangle$ *for all* $(x, x^*) \in$ *gra*(A).

(5) If $(y^*, y) \in$ *gra*(A) *then* $(y, y^*) \in \partial F_A(y, y^*)$.

Proof:
Ad (1) If $(x, x^*) \in \text{gra}(A)$ then $F_A(x, x^*) \leq \langle x^*, x\rangle < \infty$.
Ad (2) F_A is the supremum of a family of affine continuous functions.
Ad (c) it is easy to see that level sets of F_A are closed.
Ad (4) See the observation above Definition 8.31. ■

Theorem 8.33. *Let $\mathcal{X}$ be a reflexive Banach space and let $A : \mathcal{X} \rightrightarrows \mathcal{X}^*$ be a maximal monotone operator with dom$(A) \neq \emptyset$. Let F_A be the associated Fitzpatrick function of A. Then*

$$F_A(x, x^*) \geq \langle x^*, x\rangle \text{ for all } (x, x^*) \in \mathcal{X} \times \mathcal{X}^*$$

and the following statements for $(x.x^*) \in \mathcal{X} \times \mathcal{X}^*$ are equivalent:

(a) $(x, x^*) \in$ *gra*(A).

(b) $F_A(x, x^*) = \langle x^*, x\rangle$.

(c) There exists $(u, u^*) \in$ *dom*(∂F_A) *with* $\langle u^* - x^*, u - x\rangle$.

(d) $(x^*, x) \in \partial F_A(x, x^*)$.

Proof:
Let $(x, x^*) \in \mathcal{X} \times \mathcal{X}^*$. Notice that

$$F_A(x, x^*) = \langle x^*, x\rangle - \inf_{(u,u^*)\in\text{gra}(A)} \langle x^* - u^*, x - u\rangle . \tag{8.30}$$

Assumption: $F_A(x, x^+) < \langle x^*, x\rangle$. Then $\inf_{(u,u^*)\in\text{gra}(A)}\langle x^* - u^*, x - u\rangle > 0$ and hence, $(x, x^*) \in \text{gra}(A)$. Using (8.30) we obtain $F_A(x, x^*) \geq \langle x^*, x\rangle$ contradicting the assumption above.
It is not difficult to show that $(b) \implies (a) \implies (e) \implies (d) \implies (a)$ holds.
Using the Fenchel-Young equality

$$(x^*, x) \in \partial F_A(x, x^*) \iff F_A(x, x^*) + F_B^+(x^*, x) = \langle (x, x^+), (x^*, x)\rangle$$

we obtain $(a)(\text{and}(b)) \implies (c)$.
The prove of the implication $(c) \implies (a)$ starts from the equality

$$F_A^*(x^*, x) = \langle x, x^*\rangle - \inf_{(u,u^*)\in\mathcal{X}\times\mathcal{X}^*} \langle x^* - u^*, x - u\rangle + F_A(u, u^*) - \langle u^*, u\rangle .$$

This implies

$$\inf_{(u,u^*)\in\mathcal{X}\times\mathcal{X}^*} \langle x^* - u^*, x - u\rangle + F_A(u,u^*) - \langle u^*, u\rangle$$

and hence,

$$\langle x^* - u^*, x - u\rangle \geq \langle u^*, u\rangle - F_A(u,u^*) = 0 \text{ for all } (u,u^*) \in \text{gra}(A)\,.$$

This shows $(x, x^*) \in \text{gra}(A)$. ■

Definition 8.34. *Let $\mathcal{X}$ be a Banach space and let $A : \mathcal{X} \rightrightarrows \mathcal{X}^*$ be a mapping. A is called* **nonexpansive** *if*

$$\|\lambda - \mu\| \leq \|x - y\| \textit{ for all } (x,\lambda),(y,\mu) \in \textit{gra}(A)\,.$$

□

Definition 8.35. *Let $\mathcal{H}$ be a Hilbert space, let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a mapping and let $s > 0$. Then:*

$$R(A,s) : \mathcal{H} \rightrightarrows \mathcal{H}\,,\ R(A,s)(x) := (I + sA)^{-1}(x)\,,$$

is called the **resolvent** *of A with parameter s.*
We set $C(A,s) := 2R(A,s) - I$ and call $C(A,s)$ the **Caley operator** *of A with parameter $s > 0$.* □

Corollary 8.36. *Let $\mathcal{H}$ be a Hilbert space and let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a monotone operator. Then $R(A,s), C(A,s)$ are single-valued nonexpansive mappings for all $s \geq 0$.*

Proof:
For $s = 0$ nothing has to proved. Let $s > 0$ and let $x, y, u, v \in \mathcal{H}$ with

$$x \in u + sA(u)\,,\ y \in v + sA(v)\,.$$

Then we get $x - y \in u - v + sA(u) - A(v)$ and therefore by using the monotonicity $\|u - v\|^2 \leq \langle x - y | u - v\rangle$. So when $x = y$, we must have $u = v$. This shows that R is a single-valued mapping and consequentely C too.
We have

$$\begin{aligned} C(A,s) &= \|(2u - x) - (2v - y)\|^2 = \|2(u - v) - (x - y)\|^2 \\ &= 4\|u - v\|^2 - 4\langle u - v | x - y\rangle + \|x - y\|^2 \\ &\leq \|x - y\|^2 \end{aligned}$$

$R(A,\cdot)$ is nonexpansive since it is the average of I and $C(A,\cdot)$:

$$R(A,\cdot) = \frac{1}{2}I + \frac{1}{2}(2R(A,\cdot) - I)\,.$$

■

Theorem 8.37 (Minty, 1963)**.** *Let $\mathcal{H}$ be a Hilbert space and let Let $A : \mathcal{H} \rightrightarrows \mathcal{H}^*$ be a mapping. Then the following conditions are equivalent:*

(a) *A is maximal monotone.*

(b) A *is monotone and* $ran(I + sA) = \mathcal{H}$ *for all* $s > 0$.

(c) *We have for all* $s > 0$*:* $R(A,s)$ *is nonexpansive,* $dom(R(A,s)) = \mathcal{H}, ran(R(A,s)) = \mathcal{H}$.

Proof:
Ad $(a) \implies (b)$ It is enough to prove this for $s = 1$ since sA is maximal monoton for all $s > 0$. Given $z^0 \in \mathcal{H}$, we want to show that z^0 belongs to $\mathrm{ran}(I + A)$.
We define $B : \mathcal{H} \rightrightarrows \mathcal{H}$ by $B(x) := A(x) - \{z^0\}$. Then B is a maximal monotone operator. Define with the Fitzpatrick function F_B of B

$$F : \mathcal{H} \times \mathcal{H} \longrightarrow \hat{\mathbb{R}}\,, \; F(x,u) := F_B(x,u) + \frac{1}{2}\|x\|^2 + \frac{1}{2}\|u\|^2\,, \; (x,u) \in \mathcal{H} \times \mathcal{H}\,.$$

Since F_B is due to $F_B \in \mathrm{Gamma}_0(\mathcal{X} \times \mathcal{X}^*)$ bounded from below by a continuous affine function, F is coercive. Then F has a minimizer, say (y,v), and we have $(\theta,\theta) \in \partial F(y,v)$. Thus, $(\theta,\theta) \in \partial F_B(y,v) + (y,v)$ and $(-y,-v) \in \partial F_B(y,v)$. Then

$$\langle(-y,-v)|(b,w) - (y,v)\rangle \le F_B(b,w) - F_B(y,v) \text{ for all } (b,w) \in \mathrm{gra}(B)\,,$$

and by Lemma 8.32

$$\langle(-y,-v)|(b,w) - (y,v)\rangle \le \langle b|w\rangle - \langle y|v\rangle \text{ for all } (b,w) \in \mathrm{gra}(B)\,.$$

This implies

$$0 \le \langle b|w\rangle - \langle(y|v)\rangle + \langle y|b\rangle + \langle v|w\rangle - \|y\|^2 - \|v\|^2 \text{ for all } (b,w) \in \mathrm{gra}(B)\,, \tag{8.31}$$

and hence

$$\langle b+v, w+y\rangle = \langle b,w\rangle + \langle y,b\rangle + \langle v,w\rangle + \langle y,v\rangle \ge \|y+v\|^2 \ge 0 \text{ for all } (b,w) \in \mathrm{gra}(B)\,.$$

We obtain $(-y,-v) \in \mathrm{gra}(B)$ since B is a maximal monotone operator. This and (8.31) implies $0 \le -2\langle y,v\rangle - \|y\|^2 - \|v\|^2$. Then we have $y = -v$ and $(v,-v) \in \mathrm{gra}(B)$ and hence $-v \in B(v) = A(v) - \{z^0\}$. Therefore $z^0 \in A(v) + v$, which implies $z^0 \in \mathrm{ran}(I + A)$.
Ad $(b) \implies (a)$ Let $A' : \mathcal{H} \rightrightarrows \mathcal{H}$ be a monotone extension of A. Let $v \in A'(y)$. Choose $x \in \mathrm{dom}(A)$ such that $y + v \in x + A(x)$. Then $y + v \in x + A'(x), y + v \in y + A'(y)$. From Lemma 8.10 we conclude $x = y$ which implies $y + v \in y + A(y)$. Therefore, $v \in A(y)$. Thus, we have shown $A = A'$.
Ad $(a),(b) \implies (c)$ Let $s > 0$. Then sA is a maximal monotone operator and we have $\mathrm{ran}(I + sA) = \mathcal{H}$. From Lemma 8.10 we obtain the assertion in (c).
$(c) \implies (b)$ This is a consequence of Lemma 8.10 too. ■

Example 8.38. *Consider*[7] *(see Example 8.24) in the Hilbert space* $\mathcal{H} := L_2(\Omega)$ *the Laplace operator:*

$$dom(A) := H_0^1(\Omega) \cap H^2(\Omega)\,,\; Ax := -\Delta x, x \in dom(A)\,.$$

Obviously, A *is linear.* A *is mnotone since*

$$\langle Ax|x\rangle = -\int_\Omega x\Delta x d\xi = \int_\Omega \|\nabla x\| d\xi \ge 0, x \in dom(A)\,.$$

[7] This is an example for readers which are familiar with the thora of weak solutions for elliptic equations in Sobolev spaces.

A is maximal monotone due to Theorem 8.37 if

$$\Delta x + x = f$$

has a solution $x \in dom(A)$ for each $f \in \mathcal{H}$. The problem consists in the regularity: one has to show that the weak solution of $-\Delta x + x = f$ which can be ensured to exists in $H_0^1(\Omega)$ has the property $x \in H^2(\Omega)$. This is the case under the assumptions that Ω is bounded and its boundary is sufficiently smooth. □

Let $\mathcal{H}$ be a Hilbert space and let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximal monotone operator. Then acording to (c) in Theorem 8.37 the inclusion

$$y \in x + sA(x) \tag{8.32}$$

has a solution in $\mathcal{H}$ for each $y \in \mathcal{H}$ and each $s > 0$. Now, we are interested in a solution of

$$y \in A(x) \tag{8.33}$$

Theorem 8.39. *Let $\mathcal{H}$ be a Hilbert space and let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximal monotone operator which is coercive. Then $ran(A) = \mathcal{H}$.*

Proof:
We know $\mathrm{ran}(I + sA) = \mathcal{H}$ for all $s > 0$. Let $y \in \mathcal{H}$. Since due to Theorem 8.14

$$\mathrm{ran}(I + sA) = \mathcal{H} \text{ and } sI + A = s(I + 1/s\, A)\,,\ s > 0\,,$$

we have for each $s > 0$

$$sx_s + w_s = y\,,\ \text{ with } x_s \in \mathcal{H}, w_s \in A(x_s)\,. \tag{8.34}$$

Assume that there exists a sequence $(s_n)_{n\in\mathbb{N}}$ with $\lim_n s_n = 0$ and $(x_n := x_{s_n})_{n\in\mathbb{N}}$ is not bounded. Then we obtain

$$s_n\langle x_n|x_n\rangle + \langle w_n|x_n\rangle = \langle y|x_n\rangle\,,\ n \in \mathbb{N}\,,$$

with $w_n \in A(x_n)$ and we conclude that $\mathrm{dom}(A)$ is not bounded. Without loss of generality we may assume that $\lim_n \|x_n\| = \infty$. Then we obtain

$$s_n\|x_n\| + \frac{\langle w_n|x_n\rangle}{\|x_n\|} = \langle y|\frac{x_n}{\|x_n\|}\rangle\,,\ n \in \mathbb{N}\,.$$

This is a contradiction since the sum on the left side is unbounded whereas – due to the coercivity – the term on the right side is bounded.
Let $(s_n)_{n\in\mathbb{N}}$ be a sequence with $\lim_n s_n = 0$. Then $(x_n := x_{s_n})_{n\in\mathbb{N}}$ is bounded and as a consequence of (8.34) $(w_n := w_{s_n})_{n\in\mathbb{N}}$ is bounded. Without loss of generality we may assume that $(x_n)_{n\in\mathbb{N}}$, $(w_n)_{n\in\mathbb{N}}$ converge weakly; notice that closed balls in a Hilbert space are weakly sequential compact. Let $x := w - \lim_n x_n$, $w := w - \lim_n w_n$. We have $s_n x_n + w_n = y$ with $w_n \in A(x_n)$, $n \in \mathbb{N}$, and we conclude that $\lim_n w_n = y$. This shows $w \in A(x)$ due to Lemma 8.13. ■

Theorem 8.40 (Browder, 1965). *Let $\mathcal{X}$ be a reflexive Banach space and let $A : \mathcal{H} \rightrightarrows \mathcal{H}^*$ be a a coercive maximal monotone operator. Then*

$$ran(A) = \mathcal{X}^* .$$

Proof:
We do not give the prove; see [9]. ■

Remark 8.41. *The proof of $(a) \implies (b)$ in Theorem 8.37 is the harder part of the Theorem. The proof for the solvability of the equation $(I+A)(x) = y$ can be produced also along the following five steps: finite dimensional approximations (Galerkin method) of the equation, solvability in the finite dimensional case by using the fixed point theorem of Brouwer, a priori boundeness of the resulting sequence of the finite dimensional problems, weak convergence of this sequence of the finite dimensional solutions, use of the Minty trick when passing to the limit for the finite dimensional approximations. We refer to [4, 7, 9, 40].*

A different proof of the part $(a) \implies (b)$ my be based on the Min-Max Theorem; see [7]. □

Remark 8.42. *The assumption concerning the reflexivity in Theorem 8.40 is essential. Without reflexivity one can prove under additional assumptions that ran$(A($ is dense in $\mathcal{X}^*$. This follows from the Theorem of Bishop-Phelps [5]. If A is the subdifferential of a proper lower semicontinous function then ran(A) is dense in $\mathcal{X}^*$ without reflexivity. A counterexample for the density of ran(A) in $\mathcal{X}^*$ in the general case is given in [16].* □

Theorem 8.43. *Let $\mathcal{X}$ be a reflexive Banach space and let $A : \mathcal{X} \rightrightarrows \mathcal{X}^*$ be a maximal monotone operator. Then $\overline{dom(A)}, \overline{ran(A)}$ are convex.*

Proof:

■

Theorem 8.44 (Rockafellar). *Let $\mathcal{X}$ be a Banach space and let $A : \mathcal{X} \rightrightarrows \mathcal{X}^*$ be a maximal cyclically monotone operator with dom$(A) \neq \emptyset$. . Then there exists $f \in \Gamma_0(\mathcal{X})$ with $A = \partial f$.*

Proof:

■

8.4 Computing zeros of maximal monotone operators

Let $\mathcal{X}$ be a Banach space and let $A : \mathcal{X} \rightrightarrows \mathcal{X}^*$ be a mapping. Then we consider the (generalized) equation

$$\text{Find } x \in \mathcal{X} \text{ with } \theta \in A(x) . \tag{8.35}$$

We study this problem in the case when $\mathcal{X}$ is a Hilbert space for maximal monotone operators A. As we know, the solution set of the equation $x \in A^{-1}(\theta)$ is convex and closed.

Each zero of A is a fixed point of $R(A,s)$ (and $C(A,s)$) for all $s > 0$. Therefore, one can solve the equation $\theta \in A(x)$ by finding fixed points of $R(A,\cdot)$ or $C(A,\cdot)$ and hence, the fixed point iteration is appropriate to compute a zero of the maximal operator A:

$$x^{k+1} := R(A, s_k)(x^k) = (I + s_k A)^{-1}(x^k)\,,\ k \in \mathbb{N}_0\,. \tag{8.36}$$

Here x^0 is a given starting point and we **assume** throughout in the following that $s_k, k \in \mathbb{N}_0$, are positive numbers. These numbers play the role of regularization parameters.

Remark 8.45. *The iteration scheme in* (8.36) *can be reformulated as*

$$\frac{x^{k+1} - x^k}{s_k} \in -A(x^{k+1})\,,\ k \in \mathbb{N}_0\,.$$

This inclusion can be seen as an implicit discretization of the differential inclusion

$$u'(t) \in -A(u(t)) \text{ a.e. on } (0,\infty)\,,\ u(0) = x^0\,. \tag{8.37}$$

(A solution is an absolute continuous mapping from $(0,\infty)$ into $\mathcal{H}$.) In this interpretation s_k is a discretization parameter and the discretization scheme is called the **backward Euler scheme**. □

Let $(x^k)_{k\in\mathbb{N}}$ be the sequence produced by the iteration (8.36). Associated to this sequence is the sequence $(y^k)_{k\in\mathbb{N}}$ of **„velocities“**:

$$y^k := \frac{x^{k+1} - x^k}{s_k}\,,\ k \in \mathbb{N}\,. \tag{8.38}$$

Lemma 8.46. *From the iteration scheme 8.36 we conclude:*

(1) $y^k \in -A(x^{k+1}), k \in \mathbb{N}_0$.

(2) The sequence $(\|y^k\|)_{k\in\mathbb{N}}$ *is decreasing.*

(3) For all $(x,y) \in gra(A)$

$$\|x^k - x\|^2 \geq \|x^k - x^{k+1}\|^2 + \|x^{k+1} - x\|^2 - 2s_k\langle y|x^k - x\rangle\,,\ k \in \mathbb{N}_0\,.$$

(4) For all $x \in A^{-1}(\theta)$

$$\|x^{k+1} - x\|^2 + s_k^2\|y^k\|^2 \leq \|x^k - x\|^2\,,\ k \in \mathbb{N}_0\,.$$

(5) For all $x \in A^{-1}(\theta)$

$$\|x^{k+1} - x\|^2 + 2\sum_{i=1}^{k} s_i^2\|y^i\| \leq \|x^0 - x\|^2\,,\ k \in \mathbb{N}_0\,.$$

Proof:
Ad (1) follows from the iteration scheme.
Ad (2) The inequality $\langle y^k - y^{k-1} | x^k - x^{k-1} \leq 0$ implies $\langle y^k - y^{k-1} | y^k \rangle \leq 0$ and therefore $\|y^k\| \leq \|y^{k-1}\|$.
Ad (3) Observe that

$$\|x^{k-1} - x\|^2 = \|x^{k-1} - x^k\|^2 + \|x^k - x\|^2 + 2\langle x^{k-1} - x^k | x^k - x\rangle$$

and $\langle x^{k-1} - x^k | x^k - x\rangle \geq s_k \langle y | x^k - x\rangle$.
Ad (4) We have for $k \in \mathbb{N}_0$ (see (3))

$$\|x^k - x\|^2 = \|x^k - x^{k+1}\|^2 + \|x^{k+1} - x\|^2 .$$

Ad (5) Follows from (4). ■

Lemma 8.47. *The sequence* $(x^k)_{k\in\mathbb{N}_0}$ *is Fejer monotone with respect to* $A^{-1}(\theta)$, *i.e.*

$$\|x^{k+1} - x\| \leq \|x^k - x\|, n \in \mathbb{N}, \text{ for all } x \in A^{-1}(\theta) . \tag{8.39}$$

Proof:
Follows from (5) in Lemma 8.46. ■

Corollary 8.48. *The following statements are equivalent:*

(a) $\theta \in A(x)$ *for some* $x \in \mathcal{H}$.

(b) The sequence $(x^k)_{k\in\mathbb{N}}$ *produced by the iteration* (8.36) *is bounded.*

Proof:
$(a) \implies (b)$ Let $\theta \in A(x)$. By Lemma 8.47 we know that $(\|x^n - x\|^2_{n\in\mathbb{N}})$ is bounded. Then $(x^k)_{k\in\mathbb{N}}$ is bounded.
$(b) \implies (a)$ We follow [30]. Suppose that $x^k \in \overline{B}_r, k \in \mathbb{N}_0$, for some $r > 0$. Consider the monotone operator

$$A' := A + \partial h \text{ with } h = \delta_{\overline{B}_{2r}} .$$

We know

$$\partial h(x) = \{\theta\}, \text{ if } \|z\| < 2r, = \overline{B}_{2r} \text{ if } \|z\| = 2r, = \emptyset \text{ if } \|z\| > 2r .$$

Then $A'(x) = A(x), x \in B_{2r}$, T' is a maximal monotone operator since A' is the sum of two maximal monotone operators with $\mathrm{dom}(A) \cap \mathrm{int}(\mathrm{dom})(\partial h) \neq \emptyset$. Therefore we may assume $x^{k+1} = (I + s_k A')^{-1}(x^k)$ where $(I + s_k A')^{-1}$ is single-valued too. But the domain of A' is bounded and by Corollary 8.27 there exists some $x \in \mathrm{dom}(A')$ with $\theta \in A(x)$. Since $A'(x) = A(x)$ we have a solution of $\theta \in A(x)$. ■

Theorem 8.49. *Let* $\mathcal{H}$ *be a Hilbert spacea, let* $A : \mathcal{H} \rightrightarrows \mathcal{H}$ *be a maximal monotone operator and let* $(s_k)_{k\in\mathbb{N}}$ *be a sequence with* $\sum_{k=0}^{\infty} s_k^2 = \infty$. *Suppose that the sequence* $(x^k)_{k\in\mathbb{N}_0}$ *produced by the iteration* (8.36) *is bounded then it converges weakly to a* $x \in \mathcal{H}$ *with* $\theta \in A(x)$.

Proof:
$(x^k)_{k\in\mathbb{N}_0}$ is bounded. Let $(x^{n_k})_{k\in\mathbb{N}_0}$ be a weakly convergent subsequence; $x := w\text{-}\lim_k x^{n_k}$. Due to (5) in Lemma 8.46 the series $\sum_{i=1}^{k} s_i^2 \|y^i\|^2$ is convergent. Since the sequence

$(s_k^2)_{k\in\mathbb{N}}$ is not summable there must hold $\liminf_k \|y^k\|^2 = 0$. Since $(\|y^k\|^2)_{k\in\mathbb{N}}$ is convergent we must have $\lim_k y^k = \theta$. Now, we conclude $y^{n_k} \in A(x^{n_k+1}, k \in \mathbb{N}_0$, and since $x := w - \lim_k x^{n_k}$ we obtain $\theta \in A(x)$.
Let x' be another cluster point of $(x^k)_{k\in\mathbb{N}_0}$; $x' = w - \lim_l x^{n_l}$. With the same argument as above, $\theta \in A(x')$. Then the sequences $(\|x^n - x\|^2_{n\in\mathbb{N}})$, $(\|x^n - x'\|^2_{n\in\mathbb{N}})$ are convergent due to Lemma 8.47. Due to

$$\langle x^k, x - x'\rangle = \frac{1}{2}(\|x^k - x\|^2 - \|x^k - x'\|^2 + \|x\|^2 - \|x'\|^2)$$

$(\langle x^k, x - x'\rangle)_{k\in\mathbb{N}}$ is convergent. Then

$$\langle x|x - x'\rangle = \lim_k \langle x^{n_k}|x - x'\rangle = \lim_l \langle x^{n_l}|x - x'\rangle = \langle x'|x - x'\rangle\,.$$

This implies $\|x - x'\| = 0$ and therefore $x = x'$. Alltogether, this shows that the sequence $(x^k)_{k\in\mathbb{N}_0}$ converges to x. ■

The question whether the algoritm above converges strongly has been decidered by Güler [17]. He introduced an example for which the convergence generated by the algorithm (8.36) converges weakly, but not strongly.

Kamimura and Takahashi [19, 20] and Solodov and Svaiter [37] modified the algorithm (8.36) in such a way that the iteration converges strongly.

Algorithm 8.1 Proximal algorithm of Solodov and Svaiter

Given a maximal monotone operator $A : \mathcal{H} \rightrightarrows \mathcal{H}$ where $\mathcal{H}$ is a Hilbert space. This algorithm computes a sequence $(x^n)_{n\in\mathbb{N}}$ which converges strongly to a zero of A.

(1) Choose $x^0 \in \mathcal{H}$ and set $n := 0$.

(2) Find $(y^n, v^n) \in \text{gra}(A)$ with $\theta = v^n + \frac{1}{r_n}(y^n - x^n)$.

(3) Set $H_n := \{z \in \mathcal{H} : \langle z - y^n|v^n\rangle \le 0\}$.

(4) Set $W_n := \{z \in \mathcal{H} : \langle z - x^n|x^0 - x^n\rangle \le 0\}$.

(5) Set $x^{n+1} := P_{H_n \cap W_n}$.

(6) Set $n := n + 1$ and go to line (2).

8.5 The proximal operator

Lemma 8.50. *Let $\mathcal{H}$ be a Hilbert spoace and let $f \in \Gamma_0(\mathcal{H})$. Let $x, w \in \mathcal{H}$. Then the following conditions are equivalent:*

(a) $w = prox_f(x)$.

(b) $\langle y - w|x - w\rangle + f(w) \le f(y)$ *for all* $y \in \mathcal{H}$.

Proof:
Ad $(a) \implies (b)$ Let $y \in \mathcal{H}$. Set $w_t := ty + (1-t)w, t \in (0,1)$. Then

$$\begin{aligned} f(w) &\leq f(w_t) + \frac{1}{2}\|w_t - x\|^2 - \frac{1}{2}\|x - w\|^2 \\ &\leq tf(y) + (1-t)f(w) - t\langle x - w|y - w\rangle + \frac{t^2}{2}\|y - w\|^2 . \end{aligned}$$

Then

$$\langle x - w|y - w\rangle \leq f(y) + \frac{t^2}{2}\|y - w\|^2 .$$

Taking $t \to 0$ (b) follows.
Ad $(b) \implies (a)$ Let $y \in \mathcal{H}$. We obtain from (b)

$$\begin{aligned} f(w) + \frac{1}{2}\|x - w\|^2 &\leq f(y) + \frac{1}{2}\|x - w\|^2 + \langle x - w|w - y\rangle + \frac{1}{2}\|w - y\|^2 \\ &= f(y) + \frac{1}{2}\|x - y\|^2 . \end{aligned}$$

This implies $w = \mathrm{prox}_f(x)$. ■

Lemma 8.51. *Let $\mathcal{H}$ be a Hilbert spoace and let $f \in \Gamma_0(\mathcal{H})$. Then $prox_{s,f}, I - prox_{s,f}$ are firmly nonexpansive.*

Proof:
It is enough to prove the result for $s = 1$. Let $x, y \in \mathcal{H}$ and $w = \mathrm{prox}_f(x), v = \mathrm{prox}_f(y)$. Then

$$\langle v - w|x - w\rangle + f(w) \leq f(v) , \ \langle w - v|y - v\rangle + f(v) \leq f(w) .$$

This implies $w, v \in \mathrm{dom}(f)$ and

$$0 \leq \langle w - v|x - w - (y - v)\rangle .$$

■

Theorem 8.52. *Let $\mathcal{H}$ be a Hilbert space, let $f \in \Gamma_0(\mathcal{X})$ and let $s > 0$. Then the mapping $mor_{s,f} : \mathcal{H} \longrightarrow \mathcal{H}$ is Fréchet differentiable and we have*

$$\nabla mor_{s,f}(x) = \frac{1}{s}(I - prox_{s,f})(x), x \in \mathcal{H} .$$

Proof:
Let $x, y \in \mathcal{H}, w := \mathrm{prox}_{s,f}(x), v := \mathrm{prox}_{s,f}(y)$. Then with the help of Lemma 8.50

$$\begin{aligned} \mathrm{mor}_{s,f}(y) - \mathrm{mor}_{s,f}(x) &= f(v) - f(w) + \frac{1}{2s}(\|y - v\|^2 + \|x - w\|^2) \\ &\geq (2\langle v - w|x - w\rangle + \|y - v\|^2 - \|x - w\|^2)\frac{1}{2s} \\ &= (\|y - v - x + w\|^2 + 2\langle y - x|x - w\rangle)\frac{1}{2s} \\ &\geq \frac{1}{s}\langle y - x|x - w\rangle . \end{aligned}$$

In the same way one shows

$$\mathrm{mor}_{s,f}(y) - \mathrm{mor}_{s,f}(x) \le \frac{1}{s}\langle y - x | y - w\rangle .$$

Since the operator $\mathrm{prox}_{s,f}$ is firmly nonexpansive we obtain

$$\begin{aligned} 0 &\le \mathrm{mor}_{s,f}(y) - \mathrm{mor}_{s,f}(x) - \frac{1}{s}\langle y - x | x - w\rangle \\ &\le \frac{1}{s}\langle y - x | (y - v) - (x - w)\rangle \\ &\le \frac{1}{s}(\|y - x\|^2 - \|v - w\|^2) \le \frac{1}{s}\|y - x\|^2 . \end{aligned}$$

This implies

$$\lim_{y \to x} \frac{\mathrm{mor}_{s,f}(y) - \mathrm{mor}_{s,f}(x) - \langle y - x | (x - w)/s\rangle}{\|y - x\|} = 0 .$$

■

Lemma 8.53. *Let $\mathcal{H}$ be a Hilbert space, let $f \in \Gamma_0(\mathcal{H})$ and let $s > 0$. consider $f_s : \mathcal{H} \ni u \longmapsto (sf + j)(u) \in \hat{\mathbb{R}}$. Then f_s^* is Fréchet differentiable and $\mathrm{prox}_{s,f} = \nabla f_s^*$.*

Proof:
Since $\mathrm{prox}_{s,f} = R(s, \partial f)$ we have to show that

$$\mathrm{prox}_{s,f} = R(s, \partial f) = (\partial f_s)^{-1} .$$

Since f_s is strongly convex, f_s^* is Fréchet differentiable. ■

Lemma 8.54. *Let $\mathcal{H}$ be a Hilbert space and let $g \in \Gamma_0(\mathcal{H})$ be Fréchet differentiable. Then for $x, y \in dom(g)$ the following conditions are equivalent:*

(a) $\|\nabla g(x) - \nabla g(y)\| \le \|x - y\|$.

(b) $g(x) \ge g(y) + \langle \nabla g(y) | x - y\rangle + \frac{1}{2}\|\nabla g(x) - \nabla g(y)\|^2$.

(c) $\langle \nabla g(x) - \nabla g(y) | x - y\rangle \ge \|\nabla g(x) - \nabla g(y)\|^2$.

Proof:
The implication $(a) \implies (b)$ is the harder part of the result. Thus, we prove only this implication.
The function $h : \mathcal{H} \ni u \longmapsto \frac{1}{2}\|u\|^2 + g(u) \in \hat{\mathbb{R}}$ is convey. Then

$$\frac{1}{2}\|x - y\|^2 \ge g(x) - g(y) - \langle \nabla | x - y\rangle .$$

Fix y and set $d(z) := g(z) - g(y) - \langle \nabla | z - y\rangle, z \in \mathcal{H}$. Since g is convex, so is d. Since $\nabla d(z) = \nabla g(z) - \nabla g(y)$ we have

$$\|\nabla f(z) - \nabla d(x)\| \le \|z - x\| , \ \frac{1}{2}\|z - x\|^2 \ge d(z) - d(x) - \langle \nabla d(z) | z - x\rangle .$$

Choose $z : x - \nabla g(x) + \nabla g(y)$. Then

$$0 \le d(z) \le d(x) - \frac{1}{2}\|\nabla g(x) - \nabla g(y)\|^2 .$$

■

Theorem 8.55. *Let $\mathcal{H}$ be a Hilbert space, let $f \in \Gamma_0(\mathcal{H})$, let $\beta > 0$ and set $h := f^* - \beta^{-1}j$. The the following are equivalent.*

(a) f *is Fréchet differentiable and* ∇f *is Lipschitz continuous with Lipschitz constant* β.

(b) $\beta j - f$ *is convex.*

(c) $f^* - \beta^{-1}j$ *is convex.*

(d) $h \in \Gamma_0(\mathcal{H})$ *and* $f = mor_{\beta^{-1}h^*}$.

(e) $h \in \Gamma_0(\mathcal{H})$ *and* $\nabla f = \beta(I - prox_{\beta^{-1}h^*})$.

(f) f *is Fréchet differentiable and*

$$\|\nabla f(x) - \nabla f(y)\|^2 \leq \beta\langle x - y|\nabla f(x) - \nabla f(y)\rangle\,,\ x, y \in \mathcal{H}. \tag{8.40}$$

Proof:
We follow [4].
Ad $(a) \implies (b)$ Let $x, y \in \mathcal{H}$. By the Cauchy-Schwarz inequality

$$\begin{aligned}\langle x - y|\beta x - \nabla f(x) - \beta y + \nabla f(y)\rangle &= \beta\|x - y\|^2 - \langle x - y|\nabla f(x) - \nabla f(x)\rangle \\ &\geq \|x - y\|(\beta\|x - y\| - \|\nabla f(x) - \nabla f(x)\|) \geq 0.\end{aligned}$$

Hence, $\nabla(\beta j - f) = \beta I - \nabla f$ is a monotone operator and it follows that $\beta j - f$ is convex.
Ad $(b) \implies (c)$ Set $g := \beta j - f$. Then $g \in \Gamma_0(\mathcal{H})$ and therefore $g = g^{**}$. Accordingly,

$$f = \beta j - g = \beta j - g^{**} = \beta j - \sup_{u \in \mathcal{H}}(\langle \cdot|u\rangle - g^*(u)) = \inf_{u \in \mathcal{H}}(\beta j - \langle \cdot|u\rangle + g^*(u)). \tag{8.41}$$

Hence,

$$\begin{aligned} f^* &= \sup_{u \in \mathcal{H}}(\beta j - \langle \cdot|u\rangle + g^*(u)) \\ &= \sup_{u \in \mathcal{H}}(\beta^{-1}j(\cdot + u) - g^*(u)) \\ &= \beta^{-1}j + \sup_{u \in \mathcal{H}}(\beta^{-1}(\langle \cdot|u\rangle + j(u)) - g^+(u)),\end{aligned}$$

The last term is a supremum of affine functions and hence convex. Thus, h is convex.
Ad $(c) \implies (b)$ Since $f \in \Gamma_0(\mathcal{H})$ and h is convex, we have $h, h^* \in \Gamma_0(\mathcal{H})$ and

$$f = f^{**} = (h + \beta^{-1}j)^* = h^* \,\square\, \beta j = \mathrm{mor}_{\beta^{-1},h^*} = \beta j - \beta(I - \mathrm{prox}_{\beta^{-1},h}.$$

Ad $(d) \implies (e)$ Obvious.
Ad $(e) \implies (f)$ We know, $\mathrm{prox}_{\beta,h}$ is firmly nonexpansive. Hence, the assertion in (f) follows.
Ad $(f) \implies (a)$ Apply the Cauch-Schwarz inequality. ■

Remark 8.56. *The implication $(a) \implies (b)$ is is the Baillon-Haddad theorem; see [2].* □

Lemma 8.57. *Let $\mathcal{X}$ be a Banach space and let $f \in \Gamma_0(\mathcal{X})$. Then the subdifferential $\partial f : \mathcal{X} \rightrightarrows \mathcal{X}^*$ is a monotone operator.*

Proof:
Let $(x,\lambda),(y,\mu) \in \text{gra}(\partial f)$. By definition

$$f(y) \geq f(x) + \langle \lambda, y-x\rangle \,,\; f(x) \geq f(y) + \langle \mu, x-y\rangle$$

with $f(x), f(y) \in \mathbb{R}$. Adding up these inequalities, we obtain

$$f(y) + f(x) \geq f(x) + \langle \lambda, y-x\rangle + f(y) + \langle \mu, x-y\rangle$$

from which we conclude $\langle \lambda - \mu, x-y\rangle \geq 0$. ■

Theorem 8.58. *Let $\mathcal{X}$ be a Banach space and let $f \in \Gamma_0(\mathcal{X})$. Then the subdifferential $\partial f : \mathcal{X} \rightrightarrows \mathcal{X}^*$ is a maximal monotone operator with dom$(\partial f) \subset$ dom(f).*

Proof:
We know from Lemma 8.57 that $\partial f : \mathcal{X} \rightrightarrows \mathcal{X}^*$ is a monotone operator. To prove maximal monotonicity we want to argue with Lemma 8.11.
Let us suppose $(x^0, \lambda^0) \in \mathcal{X} \times \mathcal{X}^*$ is such that

$$\langle \lambda - \lambda^0, x - x^0\rangle \geq 0 \text{ for all } \lambda \in \partial f(x), x \in \mathcal{X}, \tag{8.42}$$

holds true. We want to prove that $\lambda^0 \in \partial f(x^0)$.
Define $f_0 : \mathcal{X} \longrightarrow \hat{\mathbb{R}}, j : \mathcal{X} \longrightarrow \mathbb{R}$ as follows: $f_0(x) := f(x + x^0) - \langle \lambda^0, x\rangle, j(x) := \frac{1}{2}\|x\|^2$.
Applying the duality result in Theorem 10.62 to f_0 and j we conclude that there exists $\mu \in \mathcal{X}^*$ such that

$$\inf_{x \in \mathcal{X}} (f_0(x) + j(x)) = -f_0^*(\mu) - \frac{1}{2}\|\mu\|^2 \tag{8.43}$$

and both sides in the equation above are finite. Then there exists a sequence $(y_n)_{n\in\mathbb{N}}$ such that for every $n \in \mathbb{N}$

$$\begin{aligned} 1/n^2 &\geq f_0(y_n) + \frac{1}{2}\|y_n\|^2 + f_0^*(\mu) + \frac{1}{2}\|\mu\|^2 \\ &\geq \langle \mu, y_n\rangle + \frac{1}{2}\|y_n\|^2 + \frac{1}{2}\|\mu\|^2 \\ &\geq \frac{1}{2}(\|y_n\| - \|\mu\|)^2 \;\geq\; 0 \end{aligned} \tag{8.44}$$

where the second inequality follows from the Fenchel-Young inequality (10.34). This implies

$$f_0(y_n) + f_0^*(\mu) - \langle \mu, y_n\rangle \leq 1/n^2, n \in \mathbb{N}.$$

Hence, $\mu \in \partial_{1/n^2} f_0(y_n)$ and by Theorem 10.47 (variational principle of Ekeland) it follows that there exist sequences $(z_n)_{n\in\mathbb{N}}$ in $\mathcal{X}$ and $(\mu_n)_{n\in\mathbb{N}}$ in $\mathcal{X}^*$ such that

$$\mu_n \in \partial f_0(z_n)\,,\; \|\mu_n - \mu\| \leq 1/n\,,\; \|z_n - y_n\| \leq 1/n\,.$$

Using the assumptions (8.42), (8.44) we obtain

$$\lim_n \|y_n\| = \|\mu\|\,,\; \lim_n \langle \mu, y_n\rangle = -\|\mu\|^2 \tag{8.45}$$

which, combined with the inequalities above yields $\mu = \theta$. Therefore, $\lim_n y_n = \theta$. Since f_0 is lower semicontinuous, $x = \theta$ minimizes $x \longmapsto f_0(x) + \frac{1}{2}\|x\|^2$ and from (8.43) we have $f_0(\theta) + f_0^*(\theta) = 0$. Therefore $\theta \in \partial f_0(\theta)$ which is equivalent to $\lambda_0 \in \partial f(x^0)$. ■

Example 8.59. *Let $\mathcal{H}$ be a Hilbert space, let $T : \mathcal{H} \longrightarrow \mathcal{H}$ be a linear continuous operator and let $f, g^* : \mathcal{H} \longrightarrow \hat{\mathbb{R}}$ be convex lower semicontinuous functions. Then the mapping (in a vector-matrix-notation)*

$$F_{T,f,g} : \mathcal{H} \times \mathcal{H} \ni (x, z) \longmapsto \begin{pmatrix} \theta & T^* \\ T & \theta \end{pmatrix} \begin{pmatrix} x \\ z \end{pmatrix} + \begin{pmatrix} \partial f(x) \\ \partial g^*(z) \end{pmatrix} \in \mathcal{H} \times \mathcal{H}$$

is monotone. Actually, $F_{T,f,g}$ is maximal monotone. □

Corollary 8.60. *Let $\mathcal{H}$ be a Hilbert space and let $f \in \Gamma_0(\mathcal{H})$. Then we have*

$$R(\partial f, s) = prox_{s,f}(x)\,,\ x \in \mathcal{H}, s > 0\,.$$

Proof:

■

Let us collect some additional results concerning the proximal operator. Let $\mathcal{H}$ be a Hilbert space and let $f \in \Gamma_0(\mathcal{H})$. We know already that the subdifferential $\partial f : \mathcal{H} \rightrightarrows \mathcal{H}$ is a maxmal monotone operator. Moreover,

- $\partial f(x) \neq \emptyset$ for all $x \in \mathrm{dom}(f)$ such that f is continous in x.
- $\partial f(x)$ is a closed convex subset of $\mathcal{H}$.
- $\mathrm{prox}_{s,f} = (I + \partial f)^{-1}(\theta) = R(A, s)(\theta)\,,\ s > 0$.

Let $\mathcal{H}$ be a Hilbert space and let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximal monotone operator. Then we know from the results above that the mapping

$$R(A, s) : \mathcal{H} \longrightarrow \mathcal{H}\,,\ x \longmapsto (sI + A)^{-1}(x)$$

is well defined for each $s > 0$. In the specific case that the maximal monotone operator is given as the subdifferential of a function $f \in \Gamma_0(\mathcal{H})$ the resolvent of $A := \partial f$ is given as

$$R(\partial f, s)(x) = \mathrm{prox}_{s,f}(x) := \mathrm{argmin}_{u \in \mathcal{H}}(f(u) + \frac{1}{2s}\|x - u\|^2)\,.$$

Theorem 8.61. *Let $\mathcal{X}$ be a reflexive Banach space with duality map $J_{\mathcal{X}}$ and let $f \in \Gamma_0(\mathcal{X})$. Then*

$$prox_{s,f}(x) = (J_{\mathcal{X}} + s\partial f)^{-1}(\theta) \tag{8.46}$$

Proof:
Let $x \in \mathcal{X}$. Then $w = f_s(x)$ if and only if $\theta \in (\partial(f + \frac{1}{2s}\|x - \cdot\|^2))(w)$. Since the norm is continuos in each point of $\mathrm{dom}(f)$ we may apply the identity for sums of subdifferentials and obtain

$$\theta \in \partial f(w) + \frac{1}{s}\partial(\frac{1}{2}\|\cdot\|^2)(w) \text{ and hence } w = f_s(x) \in (J_{\mathcal{X}} + s\partial f)^{-1}(\theta)\,.$$

■

Theorem 8.62. *Let $\mathcal{H}$ be a Hilbert space and let $f \in \Gamma_0(\mathcal{H})$. Then*

(a) $prox_{s,f}(x) = (I + s\partial f)^{-1}(\theta)$.

(b) $prox_{s,f} : \mathcal{H} \longrightarrow \mathcal{H}$ *is firmly nonexpansive for each* $s > 0$, *that is*

$$\|prox_{s,f}(x) - prox_{s,f}(y)\|^2 \le \langle prox_{s,f}(x) - prox_{s,f}(y) | x - y\rangle \text{ for all } x, y \in \mathcal{H}. \quad (8.47)$$

(c) $prox_{s,f} : \mathcal{H} \longrightarrow \mathcal{H}$ *is nonexpansive for each* $s > 0$, *that is*

$$\|prox_{s,f}(x) - prox_{s,f}(y)\| \le \|x - y\| \text{ for all } x, y \in \mathcal{H}. \quad (8.48)$$

Proof:

■

8.6 Minimization by the proximal point method

8.7 Pseudomonotone operators

[18]

Definition 8.63. *Let $\mathcal{X}$ be a Banach space and let $A : \mathcal{X} \rightrightarrows \mathcal{X}^*$ be a setvalued mapping.*

(a) A *is called* **pseudomonotone** *if for every* $(x, \lambda), (y, \mu) \in gra(A)$ *the following implication holds:*

$$\langle \lambda - \mu, x - y\rangle \ge 0 \implies \langle \mu, y - x\rangle \ge 0 .$$

(b) A *is called* **quasimonotone** *if for every* $(x, \lambda), (y, \mu) \in gra(A)$ *the following implication holds:*

$$\langle \lambda - \mu, x - y\rangle > 0 \implies \langle \mu, y - x\rangle \ge 0 .$$

□

It is clear that every monotone operator is pseudomonotone, and every pseudomonotone operator is quasimonotone.

There is an important difference between monotone and pseudomonotone operators. If A and B are monotone operators then the sum $A + B$ is monotone. For pseudomonotone operators this does not hold in general.

8.8 Splitting methods

Let $\mathcal{H}$ be a Hilbert space and let $F : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximal monotone operator. Again, we want to solve the inclusion

$$\theta \in F(x) \text{ for some } x \in \text{dom}(F). \quad (8.49)$$

The main idea which we exploit in this section is a decomposition of F into a sum of two maximal monotone operators $A, B : \mathcal{H} \rightrightarrows \mathcal{H}$, called **operator splitting**. Then we have to consider the inclusion

$$\theta \in (A + B)(x) \text{ for some } x \in \text{dom}(A) \cap \text{dom}(B). \quad (8.50)$$

We introduce the **resolvent operators**

$$R(s, A) := (I + sA)^{-1}, \; R(s, B) := (I + sB)^{-1},$$

and the **Caley operators**

$$C(s, A) := rR(s, A) - I, \; C(s; B) := 2R(s, B).$$

Here s is a scaling/regularization parameter. The key observation for the following developments are the following fact:

Lemma 8.64. *Let $\mathcal{H}$ be a Hilbert space and let $A, B : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximal monotone operators. Then $R(s, A), R(s, B), C(s, A), C(s, B)$ are nonexpansive mappings and we have the equivalence of the following conditions:*

(a) $\theta \in (A + B)(x)$

(b) $x = R(s, B)(z)$ *with* $z = C(s, A) \circ C(s, B)(z)$.

Proof:
The assertions concerning the nonexpansivity of the mentioned mappings see **??**.
Ad $(b) \implies (a)$ We write

$$x = R(s, B)(z), \; z' := 2x - z, \; x' = R(s, A)(z'), \; z = 2x' - z'.$$

and conclude $x = x'$. Then $2x = z' + z$ and

$$2x = z' + z \in 2x + s(A(x) + B(x)).$$

Ad $(a) \implies (b)$ Consider the conputational steps above in reverse order. ■

Let $x^0 \in \mathcal{H}$ be a starting point. Here are the „big two“of splitting methods:

Peaceman-Rachford

$$x^{k+1} := C(s, A) \circ C(s, B)(x^k), \; k \in \mathbb{N}_0$$

This is undamped method. In general, the itaration does not converge; see [].

Douglas-Rachford

$$x^{k+1} := \frac{1}{2}(I + C(s, A) \circ C(s, B))(x^k), \; k \in \mathbb{N}_0$$

Here $\alpha = \frac{1}{2}$ is a damping parameter. Clearly, other damping parameters $\alpha \in (0, 1)$ can be applied. This method converges strongly when the inclusion $\theta \in (A + B)(x)$ is solvable; see [].

The Douglas-Rachford method can be realized as follows:

Algorithm 8.2 Douglas-Rachford splitting

Given maximal monotone operators $A, B : \mathcal{H} \rightrightarrows \mathcal{H}$ where $\mathcal{H}$ is a Hilbert space. This algorithm computes a sequence $(x^n)_{n\in\mathbb{N}}$ which converges strongly to a zero of $A + B$. The sequence $(z^{n+1} - z^{n+1/2})_{n\in\mathbb{N}}$ can be considered as sequence of residuals and $(x^n)_{n\in\mathbb{N}}$ is the sequence of the sum of residuals.

(1) Choose $x^0 \in \mathcal{H}$ and set $n := 0$.

(2) Set $z^{n+1/2} := R(s, B)(x^n)$.

(3) Set $x^{n+1/2} := 2z^{k+1/2} - x^n$.

(4) Set $z^{n+1} := R(s, A)(x^{n+1/2}$.

(5) Set $x^{n+1} := x^n + z^{n+1} - z^{n+1/2}$.

(6) Set $n := n + 1$ and go to line (2).

As we see, in the realization of the Douglas-Rachford in Algorithm 8.8 the operators A, B are handeled separately.

In the case where both A and B in the Douglas-Rachford splitting method are single-valued, the algorithm 8.8 reduces to the original Douglas-Rachford scheme in [11] for heat conduction problems. It turns out that the algorithm 8.8 is the root of a number of other effective methods.

[?] [?] [?] [15][?]

Many methods for specific problems can be extracted from the Douglas-Rachford algorithm. Here is a list of methods which may considered under Douglas-Rachford approach.

Alternating direction method Here we want to solve in the Hilbert space $\mathcal{H}$ the minimization problem

$$\text{Minimize } (f + g)(u) \text{ subject to } u \in \mathcal{H} \tag{8.51}$$

where $f, g \in \Gamma_0(\mathcal{H})$. This can be done by solving the inclusion

$$\theta \in (\partial f + \partial g)(x) \text{ for some } x \in \text{dom}(f) \cap \text{dom}(g)\,, \tag{8.52}$$

i.e. $A = \partial f, B := \partial g$. Then $R(s, A) = \text{prox}_{s,f}, R(s, B) = \text{prox}_{s,g}$.

Constrained optimization Here the we want to solve in the Hilbert space $\mathcal{H}$ the the contraint optimization problem

$$\text{Minimize } f(u) \text{ subject to } u \in C \tag{8.53}$$

where $f \in \Gamma_0(\mathcal{H})$ and C is a nonempty closed convex subset of $\mathcal{H}$. This can be done by solving the inclusion

$$\theta \in (\partial f + \partial\delta_C)(x) \text{ for some } x \in \text{dom}(f) \cap \text{dom}(g)\,, \tag{8.54}$$

i.e. $A = \partial f, B := \partial\delta_C$. Then $R(s, A) = \text{prox}_{s,f}, R(s, B) = P_C$.

Dykstra's algorithm We want to compute a common point of two nonempty closed convex sets C, D in the Hilbert space $\mathcal{H}$. we consider

$$\text{Minimize } \theta \text{ subject to } u \in C \text{ and } u \in D\,. \tag{8.55}$$

Then we have to realize the Douglas-Rachford algorithm with $A := \partial\delta_C$ and $B := \partial\delta_D$. Then $R(s, A) = P_C, R(s, B) = P_D$ and the algoritm becomes Dykstra's alternating method.

Let us now analyze the convergence of the Douglas-Rachford algorithm. Clearly, $x^+ \in \mathcal{H}$ solves the problem (8.50) if and only if

$$x^* = R(s, A)(x^* - s)$$

We will measure of the accuracy
We will show that the convergence order

8.9 Exercises

1.) Let $\mathcal{X}$ be a Banach space and let $f : \mathcal{X} \longrightarrow \mathcal{X}$ be strongly convex, i.e. $f - \frac{c}{2}\|\cdot\|^2$ is convex for some $c > 0$. Show that the subdifferential of f is strongly monotone.

2.) Let Q be a positive definite matrix in $\mathbb{R}^{n,n}$. Show that for each $y \in \mathbb{R}^n$ the mapping

$$A : \mathbb{R}^n \ni x \longmapsto Qx + y \in \mathbb{R}^n$$

is a strongly monotone operator.

3.) Let Q be a positive semidefinite matrix in $\mathbb{R}^{n,n}$. Show that for each $y \in \mathbb{R}^n$ the mapping

$$A : \mathbb{R}^n \ni x \longmapsto Qx + y \in \mathbb{R}^n$$

is a cocoercive operator with parameter $c > 0$ if the largest eigenvalue of A is not bigger than c^{-1}.

4.) Show that cocoercivity implies Lipschitz continuity.

5.) Show that Lipschitz continuity of a mapping does not imply cocoercivity.

6.) Find a continuous solution of the differentail inclusion

$$-x' \in \text{sign}(x)\,,\; x(0) = 1$$

7.) Let $A \in \mathbb{R}^{n,n}$. Show $C(A, s) = R(A, s)R(A, -s)$.

8.) Consider

$$\tilde{f} : \mathbb{R} \ni x \longmapsto \begin{cases} 2x & x < 2 \\ 3x & x \geq 2 \end{cases} \in \mathbb{R}\,,\; f : \mathbb{R} \ni x \longmapsto \begin{cases} 2x & x < 2 \\ [4, 6] & x = 2 \\ 3x & x > 2 \end{cases} \in \mathbb{R}\,.$$

Show hat f defines a monotone operator on $\mathbb{R}$ and that $\tilde{f}$ is its maximal monotone extension.

9.) Compute the Yosida approximation of $\tilde{f}$ in the exercise above.

10.) Let $\mathcal{H}$ be a Hilbert space and let $f, g \in \Gamma_0(\mathcal{H})$. Find a condition on f, g which ensures $\text{prox}_{s,f}(x) = \text{prox}_{sg}(x)$ for all $s > 0, x \in \mathcal{H}$.

11.) Let $\mathcal{H}$ be the euclidean space $\mathbb{R}^n$. We consider on $\mathcal{H}$ the functions

$$f((x_1, \ldots, x_n)) := \sum_{i=1}^{n} |x_i|\,,\; g((x_1, \ldots, x_n)) := \sum_{i=1}^{n-1} |x_i - x_{i+1}|\,.$$

Verify $\text{prox}_{f+g} = \text{prox}_f \circ \text{prox}_g$.

12.) Let $\mathcal{H}$ be the euclidean space $\mathbb{R}^n$, let $y \in \mathbb{R}^n, b \in \mathbb{R}$, and let $f : \mathbb{R}^n \ni x \longmapsto \langle y|x\rangle + b \in \mathbb{R}$. Compute $\text{prox}_{s,f}$.

13.) Let $\mathcal{H}$ be the euclidean space $\mathbb{R}^n$ let f be the l_1-norm. Prove $\text{prox}_{s,f} = I - P_{[-s,s]^n}$.

14.) Let $\mathcal{H}$ be a Hilbert space and let $f, g \in \Gamma_0(\mathcal{H})$. Consider $F : \mathcal{H}^2 \ni (u, v) \longmapsto f(u) + g(v) \in \mathbb{R}$ and show

$$\text{prox}_{s,f}(u, v) = (\text{prox}_{s,f}(u), \text{prox}_{sg}(v))\,,\; u.v \in \mathcal{H}\,.$$

15.) Let A be the rotation in the euclidean space $\mathbb{R}^2$ by a an angle $\phi \in [-\pi/2, \pi/2]$. Show that A is a monotone operator on $\mathbb{R}^2$.

16.) Let A be the rotation in the euclidean space $\mathbb{R}^2$ by a an angle $\phi \in [-\pi/2, \pi/2]$. Show that A is a cyclically monotone operator on $\mathbb{R}^2$.

17.) Consider in the euclidean space $\mathbb{R}^n$ the function $f : \mathbb{R}^n \ni x \longmapsto \langle x|Ax\rangle + \langle b|x\rangle + c \in \mathbb{R}$ with $A \in \mathbb{R}^{n,n}, b \in \mathbb{R}^n, c \in \mathbb{R}$. Compute $\text{prox}_{s,f}(x), x \in \mathbb{R}^n$.

18.) Let $\mathcal{H}$ be a Hilbert space, let C be a nonempty closed convex subset of $\mathcal{H}$ and let $F : C \longrightarrow C$ be nonexpansive. Show that $A := I - T$ is monotone. When is A maximal monotone?

19.) Let $\mathcal{X}$ be a Banach space and let $f \in \Gamma_0(\mathcal{X})$. Show:

(1) $\text{prox}_h(x) = s^{-1}(\text{prox}_{s^2 f^*}(sx + z) - z), x \in \mathcal{X}$, if $h(x) := f(sx + z), x \in \mathcal{X}$.

(2) $\text{prox}_{sf^*}(\lambda) = \lambda - s\ \text{prox}_{s^{-1}f}(s^{-1}\lambda), \lambda \in \mathcal{X}^*$.

20.) Let $f : \mathbb{R} \longrightarrow \mathbb{R}$. Compute $\text{prox}_{s,f}$ for

$$f(t) := \frac{1}{2}t^2\,,\; f(t) := |t|\,,\; f(t) := \delta_{[-1,1]}\,.$$

21.) Let $\mathcal{X}$ be a Banach space and let $f \in \Gamma_0(\mathcal{X})$. Show: $\text{mor}^*_{s,f}(\lambda) = f^*(\lambda) + \frac{s}{2}\|\lambda\|^2, \lambda \in \mathcal{X}^*, s > 0$.

22.) Let $\mathcal{H}$ be a Hilbert space, let $C \subset \mathcal{H}$ be a nonempty closed convex set, and let $F : C \longrightarrow C$be a nonexpansive operator. Show that $A := I - T$ defines a monotone operator. When is A a maximal monotone operator ?

23.) Consider $f : \mathbb{R} \longrightarrow \mathbb{R}$, defined by

$$f(x := \begin{cases} -x^2 & \text{if } x < 0 \\ x^2 + 2 & \text{if } x > 0 \end{cases}\,.$$

Show that f is not maximal monotone. What is the the maximal monotone extension?

24.) Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be monotone increasing and continuous. Show: f is maximal monotone.

25.) Let $\mathcal{X}$ be a Banach space and let $f \in \Gamma_0(\mathcal{X})$. Show for $s >$

$$\mathrm{mor}_{s,f} = \frac{1}{2s}\|x\|^2 - \frac{1}{s}(\frac{1}{2}\|x\|^2 + sf^*(x))\,,\ x \in \mathcal{X}\,.$$

Bibliography

[1] F.J. Artacho, J.M. Borqein, V. Martin-Marquez, and L. Yao. Applications of convex analysis within mathematics. *Math. Programming*, 148:49–88, 2014.

[2] J.-B. Baillon and G. Haddad. Quelques propriétes des opérateurs angle-bornés et n-cycliquement monotones. *Israel J. Math.*, 26:137–150, 1977.

[3] J. Baumeister. *Stable Solution of Inverse Problems.* Vieweg, Braunschweig, 1987.

[4] H.H. Bauschke and P.L. Combettes. The baillon-haddad theorem revisited. *Convex analysis*, 17:781–787, 2010.

[5] E. Bishop and R.R. Phelps. The support functionals of a convex set. *Proc. Symp. Pure Math.*, 7:27–35, 1963.

[6] J.M. Borwein. Fifty years of maximal monotonicity. *Optimization Letters*, 4:473–490, 2010.

[7] H. Brézis. *Opérateurs maximaux monotone et semi-groups de contractions dans les espacede Hilbert.* North-Holland, Amsterdam, 1973.

[8] H. Brézis, M.G. Crandall, and A. Pazy. Perturbations of nonlinear maximal monotone sets in Banach spaces. *Comm. Pure Appl. Math.*, 23:123–144, 1970.

[9] F. Browder. Nonlinear operators and nonlinear evolution equations in Banach spaces. *Proc. Symp. Pure Math.*, 18:xx–xx, 1965.

[10] C. L. Byrne. *Iterative optimization in inverse problems.* Chapman & Hall, Boca Raton, 2014.

[11] J. Douglas and H.H. Rachford. On the numericl solution of the heat conduction problem in 2 and 3 space variables. *Trans. Amer. Math. Soc.*, 82:421–439, 1956.

[12] S. Fitzpatrick. Representing monotone operators by convex functions. *Proc. Centre Math. Austral. Nat. Univ.*, 20:59–65, 1988.

[13] S. Fitzpatrick and R.R. Phelps. Bounded approximations to monotone operators on Banach spaces. *Ann. Inst. Henri Poincaré, Analyse non linéaire*, 9:573–595, 1992.

[14] J. Friedman, T. Hastie, H. Höfling, and R. Tibshirani. Pathwise coordinate optimization. *The Annals of Applied Statistics*, 1:302–332, 2007.

[15] M. Fukushima. The primal douglas-rachford splitting algorithm for a class of monotone mappings with applications to the traffic equilibrium problem. *Math. Programming*, 72:1–15, 1996.

[16] J.-P. Gossez. On the range of a coercive maximal monotone operator in a nonreflexive Banach space. *Proc. Amer. Math. Soc.*, 15:88–92, 1972.

[17] O. Güler. On the convergence of the proximal point algorithm for convex minimization. *SIAM J. Control Optim.*, 29:403–419, 1991.

[18] N. Hadjisavvas, S. Schaible, and N.-C. Wong. Pseudomonotone operators: A survey of the theory and its applications. *J. Optim. Theora Appl.*, 152:1–20, 2012.

[19] S. Kamimura and W. Takahashi. Approximating solutions of maximal monotone operators in Hilbert space. *J. Approx. Theory*, 106:226–240, 2000.

[20] S. Kamimura and W. Takahashi. Weak and strong convergence of solutions to accretitive operator inclusions and applications. *Set-valued Analysis*, 8:361–374, 2000.

[21] A. Kartsatos. On the connection between the existence of zeros and the asymptotic behavior of resolvents of maximal monotone operators in reflexive Banach spaces. *Trans. American Math. Soc.*, 350:3967–3987, 1998.

[22] A. Kirsch. *An Introduction to the Mathematical Theory of Inverse Problems*. Springer, New York, 1996.

[23] B. Martinet. Regularisation d'inéquations variationnelles par approximations successives. *Rev. Francaise Informat. Recherche Opérationnelle*, 4:154–158, 1970.

[24] B. Martinet. Determination approchée d'un point fixe d'une application pseudocontractante. *Acad. Sci. Paris*, 274:163–165, 1972.

[25] J. Moreau. Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France*, 93:273–299, 1965.

[26] M. Nikolova. Minimizers of cost-functions involving nonsmooth data-fidelity terms. application to the processing of outliers. *SIAM J. Numer. Anal.*, 40:965–994, 2002.

[27] E. Pardoux and A. Rascanu. *Stochastic differential equations, backward SDE's, patial differential equations*. Springer, New York, 2014.

[28] R.R. Phelps. Lectures on maximal monotone operators. *Extracta Mathematicae*, 12:193–230, 1997.

[29] A. Rascanu and E. Rotenstein. The Fitzpatrick function – a bridge beween convex analysis and multivalued stochastic differential equations, 2009. arXiv preprint arXiv:0809.4447.

[30] R.T. Rockafellar. Local boundeness of nonlinear monotone operators. *Michigan Math. J.*, 16:397–407, 1969.

[31] R.T. Rockafellar. On the maximality of sums of nonlinear monotone operators. *Trans. Amer. Math. Soc.*, 149:75–88, 1970.

[32] R.T. Rockafellar. Augmented lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.*, 1:97–116, 1976.

[33] R.T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM J. Control and Optimization*, 14:877–898, 1976.

[34] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[35] M. Ruzicka. *Nichtlineare Funktionalanalysis*. Springer, Berlin, 2007.

[36] E.K. Ryu and S. Boyd. Primer on monotone operator methods. *Appl. Comput. Math.*, 15:3–43, 2016.

[37] M.V. Solodov and B.F. Svaiter. An inexact hybrid generalized proximal point algorithm and some new results on the results on the theory of Bregman functions. *Math. Oper. Res.*, 25:214–230, 2000.

[38] M. Storath, A. Weinmann, and M. Unser. Exact algorithms for $\mathfrak{l}^1$-tv regularization of real-valued or circle-valued signals. *SIAM J. Sci. Comput.*, 38:A614–A630, 2016.

[39] Y. Yu. On decomposing the proximal map. Advances of neural information, 2013. 9 pages.

[40] E.H. Zarantonello. *Contributions to nonlinear functional analysis*. Academic press, New York, 1971.