Jethro Warnett

Gradient Flows in Hilbert Spaces with Applications

Semester Paper supervised by: Prof. Dr. Alessio Figalli

ETH Zürich, Herbstsemester 2021

1 Abstract

The paper presents a collage of results for the theory of gradient flow from the textbook [3] from Haïm Brézis. The main focus is placed on maximal monotone multivalued operators in $\mathbb{R}$-Hilbert spaces and on the inhomogeneous generalised gradient descent evolution equation. In particular we study the case of when the maximal monotone multivalued operator is the sub differential operator of a proper convex lower semi continuous function on a $\mathbb{R}$-Hilbert space. The paper demonstrates at the end how the theory of gradient flow is applied in machine learning with an example in ridge regression.

Contents

2 Introduction

The implementation and use of machine learning algorithms to solve intricate and complex problems in areas such as data analysis has been staggering in recent years. This is not surprising due to the ability of the state-of-the-art machine learning algorithms to efficiently and reliably solve problems, which are impracticable if not infeasible for a human to attempt to resolve. These problems include semantic segmentation [1], language translation [6] or speech recognition [10]. The fundamental algorithm that is used in machine learning (or at least a variant) is called the discrete gradient descent method. The theory behind this algorithm finds some of its origin from the theory of maximal monotone multivalued operators from Haïm Brezis [3]. For a detailed account of the history behind the theory of maximal monotone multivalued operators, we refer interested readers to the paper [2].

This semester paper presents the theory of maximal monotone operators and in particular the theory of gradient flow from the textbook [3] from Haïm Brézis and indicates the connection to machine learning algorithms.

2.1 Semester Paper Structure

In Section 3 we state the theory of maximal monotone multivalued operators in $\mathbb{R}$-Hilbert spaces and study in particular the subdifferential operator of proper convex lower semicontinuous maps.

We start section 4 by introducing the reader to some preliminaries of function spaces over a $\mathbb{R}$-Hilbert space. We proceed by introducing the evolution equation of the generalised gradient descent equation, which is defined by

$$\begin{cases} \partial_t u + Au \ni f & \text{on } [0,T], \\ u = u_0 & \text{on } \{0\}. \end{cases}$$

First we consider the homogeneous case where $f = 0$, and then the inhomogeneous case with $f \in L^1\big(0,T;\mathcal{H}\big)$. We finish the section by investigating the special case, where the maximal multivalued operator A is the subdifferential of a proper convex lower semi continuous function.

In Section 5 we provide a brief explanation of what neural networks are in machine learning and how gradient flow is applied in algorithms to train these networks. We finish the section by presenting an example in ridge regression as described in originally in the paper [8], or for a more recent version in [12]. Alongside this example we conduct and present the results of a numerical experiment of ridge regression.

Throughout the entire paper we assume that $\mathcal{H}$ is a $\mathbb{R}$-Hilbert space. We denote by $\langle\cdot,\,\cdot\rangle_{\mathcal{H}}$ the inner product of $\mathcal{H}$, and by $\|\cdot\|_{\mathcal{H}}$ the induced norm of $\mathcal{H}$, i.e. $\|x\|_{\mathcal{H}} = \sqrt{\langle x,\, x\rangle_{\mathcal{H}}}$ for all $x \in \mathcal{H}$.

3 Maximal Monotone Operators in Real Hilbert Spaces

3.1 Multivalued Operators

In this section we will introduce a class of multivalued operators in $\mathbb{R}$-Hilbert spaces called maximal monotone operators and study some of their properties. The theory of gradient flows is based on the class of these operators. We proof a criterion for these operators to be maximally monotone and use this condition to present a few examples.

Definition 3.1. *A multivalued operator A in $\mathcal{H}$ is a mapping $A : \mathcal{H} \to \mathcal{P}(\mathcal{H})$, where $\mathcal{P}$ denotes the powerset. The domain of A is defined by the set $D(A) := \{x \in \mathcal{H} \mid A(x) \neq \emptyset\}$ and the range of A is defined by $R(A) := \{y \in \mathcal{H} \mid \exists x \in \mathcal{H} \,:\, y \in A(x)\}$. The mapping A is single valued, if all elements in the domain of A are mapped to singletons.*

Throughout this paper we will make the following simplifications: For any $x \in \mathcal{H}$ we write Ax as short hand for $A(x)$, and we identify single valued operators with self-mappings in $\mathcal{H}$, i.e. functions of the form $\mathcal{H} \to \mathcal{H}$. We assume $A\eta$ is any value $\xi \in A\eta$ for $\eta \in D(A)$ unless specified otherwise.

It is useful to relate multivalued operator with their graphs in $\mathcal{H}\times\mathcal{H}$. This is because set inclusion will endow the set of multivalued operators in $\mathcal{H}$ with a partial order.

Definition 3.2. *Let A be a multivalued operator in $\mathcal{H}$. The graph of A is defined by the set $\{(\eta,\xi) \mid \eta \in D(A),\ \xi \in A\eta\}$. We identify the multivalued operator A with its own graph.*

We need multivalued operators to be compatible with the vector space structure of $\mathcal{H}$. This motivates the following definition.

Definition 3.3. *Let A and B be multivalued operators in $\mathcal{H}$ and $\lambda \in \mathbb{R}$ any real value. We construct the following multivalued operators in $\mathcal{H}$:*

- *The sum $A + B$ is defined by $(A+B)x := \{y + z \mid y \in Ax,\ z \in Bx\}$ for all $x \in \mathcal{H}$, with the domain given by the intersection $D(A) \cap D(B)$.*

- *The scalar multiple λA is defined by $\lambda Ax := \{\lambda y \mid y \in Ax\}$ for all $x \in \mathcal{H}$, whose domain coincides with that of A.*
- *The inverse A^{-1} of A is defined by its graph $A^{-1} := \{(y,x) \mid (x,y) \in A\}$. The domain and range of A^{-1} coincide with the ones from A but interchanged, this means we have that $D(A^{-1}) = R(A)$ and $R(A^{-1}) = D(A)$.*

In the Hilbert space $\mathbb{R}$ a mapping $f : \mathbb{R} \to \mathbb{R}$ is monotonically increasing if and only if it satisfies for the following inequality

$$\big(f(x) - f(y)\big) \cdot \big(x - y\big) \geq 0 \qquad \forall x, y \in \mathbb{R}.$$

This notion of monotonicity can be generalised to any $\mathbb{R}$-Hilbert space.

Definition 3.4. *We say a multivalued operator A in $\mathcal{H}$ is monotone, if it fulfils the following inequality*

$$\langle \xi_1 - \xi_2,\, \eta_1 - \eta_2 \rangle_{\mathcal{H}} \geq 0 \qquad \forall (\eta_1, \xi_1), (\eta_2, \xi_2) \in A.$$

Lemma 3.5. *Let A be a multivalued operator in $\mathcal{H}$. Then A is monotone if and only if for all $\lambda > 0$ the following inequality holds true*

$$\|\eta_1 - \eta_1\|_{\mathcal{H}} \leq \|(\mathrm{id}_{\mathcal{H}} + \lambda A)\eta_1 - (\mathrm{id}_{\mathcal{H}} + \lambda A)\eta_2\|_{\mathcal{H}} \qquad \forall \eta_1, \eta_2 \in D(A). \tag{1}$$

Proof. First suppose the inequality (1) holds true. Thus by squaring and restructuring the equations we get the equivalent statement

$$-\lambda^2 \left\|A\eta_1 - A\eta_2\right\|_{\mathcal{H}}^2 \leq 2\lambda \left\langle A\eta_1 - A\eta_2,\, \eta_1 - \eta_2 \right\rangle_{\mathcal{H}},$$

If we divide by λ and take the limit $\lambda \searrow 0$, we get that A is monotone. On the other hand, if A is monotone, we can reverse the arguments to see that the inequality (1) is correct. □

Remark 3.6. *Lemma 3.5 shows how the notion of a monotone multivalued operator can be extended to include Banach spaces. This generalization is investigated in papers such as [5].*

Of particular interest are the maximal elements in the set of monotone multivalued operators in $\mathcal{H}$ that are partially ordered by the set inclusions of their graphs.

Definition 3.7. *A monotone multivalued operator A in $\mathcal{H}$ is maximally monotone, if its graph is inclusion-wise maximal in the set of graphs of monotone multivalued operators in $\mathcal{H}$. This means that if B is a monotone multivalued operator in $\mathcal{H}$ with $A \subseteq B$, then $A = B$.*

Remark 3.8. *The graph of a maximally monotone operator A has an equivalent definition, on which this semester project will rely on many times. Namely for $(x,y) \in \mathcal{H} \times \mathcal{H}$ we have*

$$(x,y) \in A \qquad \Leftrightarrow \qquad \langle y - \xi,\, x - \eta \rangle_{\mathcal{H}} \geq 0 \qquad \forall (\eta, \xi) \in A. \tag{2}$$

With the condition (2) we can build other maximally monotone operators from an existing one.

Example 3.9. *If A is a maximally monotone operator on $\mathcal{H}$, then so to is the scalar multiple λA for all $\lambda > 0$ and the inverse A^{-1}.*

In general the same cannot be said about the finite sum of maximally monotone multivalued operators in $\mathcal{H}$, as the intersection of their domains may be empty. Although there are conditions on the domain that permit the aforementioned sum to be maximally monotone (cf. [3, Chapter 2.6]).

A maximally monotone multivalued operator imposes structural constraints on the image of individual elements.

Lemma 3.10. *Let A be a maximally monotone operator in $\mathcal{H}$ and $\eta \in D(A)$. Then the set $A\eta$ is convex and weakly closed for all $\eta \in D(A)$. In addition for any $\eta_1, \eta_2 \in D(A)$ we have that* $\operatorname{int} A\eta_1 \cap \operatorname{int} A\eta_2 = \emptyset$*, whenever $\eta_1 \neq \eta_2$.*

Proof. "$A\eta$ is convex": For all $\theta \in [0,1]$ and $\xi_1, \xi_2 \in A\eta$ we have $\theta\xi_1 + (1-\theta)\xi_2 \in A\eta$, since the value satisfies the condition (2) by the following computation

$$\begin{aligned}\langle \theta\xi_1 + (1-\theta)\xi_2 - y,\, \eta - x\rangle_{\mathcal{H}} &= \langle \theta(\xi_1 - y) + (1-\theta)(\xi_2 - y),\, \eta - x\rangle_{\mathcal{H}} \\ &= \theta\,\langle \xi_1 - y,\, \eta - x\rangle_{\mathcal{H}} + (1-\theta)\,\langle \xi_2 - y,\, \eta - x\rangle_{\mathcal{H}} \qquad \forall (x,y) \in A. \\ &\geq 0\end{aligned}$$

"$A\eta$ is weakly closed": Let $(\xi_n)_{n\in\mathbb{N}} \subseteq A\eta$ be any sequence such that $\xi_n \rightharpoonup \xi$ for some $\xi \in \mathcal{H}$. Then $\xi \in A\eta$, as the value satisfies the condition (2) by the following calculation

$$\langle \xi - y,\, \eta - x\rangle_{\mathcal{H}} = \lim_{n\to\infty} \langle \xi_n - y,\, \eta - x\rangle_{\mathcal{H}} \geq 0 \qquad \forall (x,y) \in A.$$

"$\operatorname{int} A\eta_1 \cap \operatorname{int} A\eta_2 = \emptyset$": Suppose there exists $\xi \in \mathcal{H}$ and $\varepsilon > 0$, such that every $x \in \mathcal{H}$ with $\|x\|_{\mathcal{H}} < \varepsilon$ satisfies $\xi + x \in A\eta_1 \cap A\eta_2$. Then as A is maximally monotone we see

$$\langle x,\, \eta_1 - \eta_2\rangle_{\mathcal{H}} = \langle (\xi + x) - \xi,\, \eta_1 - \eta_2\rangle_{\mathcal{H}} \geq 0.$$

As this inequality must hold for all elements x in a neighbourhood of 0, we deduce that the identity $\eta_1 = \eta_2$ must stand. □

We will see in Theorem 3.23, in combination with example 3.9, that a maximally monotone operator even imposes structural constraints on its domain and range.

Next we present a criterion for a multivalued operator to be maximally monotone. This proposition will require the use of the following theorem, a proof of which can be found in [3, Theorem 2.1].

Theorem 3.11. *Let C be a closed convex set in $\mathcal{H}$ and A a monotone operator in $\mathcal{H}$. For all $y \in \mathcal{H}$ there exists $x \in C$ such that*

$$\langle \xi + x,\, \eta - x\rangle_{\mathcal{H}} \geq \langle y,\, \eta - x\rangle_{\mathcal{H}} \quad \forall (\eta, \xi) \in A.$$

This theorem lets us state a useful criteria for a multivalued operator to be maximally monotone. In fact we will need these conditions for the examples at the end of this section.

Proposition 3.12. *Let A be a multivalued operator in $\mathcal{H}$. Then the following are equivalent*

(i) A is maximally monotone.

(ii) A is monotone and $R(\mathrm{id}_{\mathcal{H}} + A) = \mathcal{H}$.

Proof. "(i) $\Rightarrow$ (ii)": If we set $C = \mathcal{H}$ in theorem 3.11, then for all $y \in \mathcal{H}$ there exists $x \in \mathcal{H}$ with

$$\langle \xi + x,\, \eta - x\rangle_{\mathcal{H}} \geq \langle y,\, \eta - x\rangle_{\mathcal{H}} \quad \forall (\eta, \xi) \in A.$$

By condition (2) we have that $y - x \in Ax$ and thus $y \in R(\mathrm{id}_{\mathcal{H}} + A)$. This proves $R(\mathrm{id}_{\mathcal{H}} + A) = \mathcal{H}$.

"(ii) $\Rightarrow$ (i)": Let B be a monotone multivalued operator on $\mathcal{H}$ such that $A \subseteq B$. By assumption for all $(\eta, \xi) \in B$ there must exist $z \in D(A)$ satisfying $\eta + \xi \in z + Az \subseteq z + Bz$. Clearly we have $\eta + \xi \in \eta + B\eta$. Then by using the monotonicity of B, we compute

$$0 \leq \langle (\eta + \xi - z) - (\eta + \xi - \eta),\, z - \eta \rangle_{\mathcal{H}} = - \|\eta - z\|_{\mathcal{H}}^2 \leq 0.$$

This shows $\eta = z$ and thus $(\eta, \xi) \in A$. Then A is maximally monotone as $A = B$. □

Of independent interest is that any multivalued monotone operator in $\mathcal{H}$ can be extended to a maximally monotone operator. The extension of the domain can be restricted to at most the closure of the convex hull of the monotone operator.

Corollary 3.13. *Let A be a monotone multivalued operator in $\mathcal{H}$. There exists some maximally monotone operator B in $\mathcal{H}$ with $A \subseteq B$ and $D(B) \subseteq \overline{\operatorname{conv}\, D(A)}$.*

Proof. Let us define the set $\mathscr{D}$ of all monotone multivalued operators B in $\mathcal{H}$ with $A \subseteq B$ and $D(B) \subseteq \overline{\operatorname{conv}\, D(A)}$. The set is endowed with the partial order of the graphs of multivalued operators in $\mathcal{H}$. The set $\mathscr{D}$ is non-empty as it contains A. Thus by Zorn's Lemma, there exists a maximal element B in $\mathscr{D}$.
We claim that B is maximally monotone. Indeed if we set $C = \overline{\operatorname{conv}\, D(A)}$ in theorem 3.11, then for all $y \in \mathcal{H}$ there exists $x \in C$ with

$$\langle \xi + x,\, \eta - x \rangle_{\mathcal{H}} \geq \langle y,\, \eta - x \rangle_{\mathcal{H}} \quad \forall (\eta, \xi) \in B.$$

As B is maximal in $\mathscr{D}$, we have that $y - x \in Bx$ and thus $y \in R(\operatorname{id}_{\mathcal{H}} + B)$. From this we deduce $R(\operatorname{id}_{\mathcal{H}} + B) = \mathcal{H}$. So by proposition 3.12 we have that B is maximally monotone. □

Remark 3.14. *One may wonder if in corollary 3.13, the domain of B can be restricted to a smaller set than $\overline{\operatorname{conv}\, D(A)}$. We will see in theorem 3.23 that this is not true.*

With proposition 3.12 it is possible to count some examples of maximally monotone operators.

Example 3.15. *Let $f : \mathbb{R} \to \mathbb{R}$ be a non-decreasing mapping. Define the multivalued operator A in $\mathbb{R}$ by $A\eta := [f(\eta-), f(\eta+)]$ for all $\eta \in \mathbb{R}$, with $f(\eta\pm) := \lim_{h \searrow 0} f(\eta \pm h)$. Clearly $R(A)$ is a closed interval in $\mathbb{R}$ and thus $\operatorname{id}_{\mathbb{R}} + A$ is a surjective multivalued operator and it is monotone as f is non-decreasing. Thus by proposition 3.12 we know that A is maximally monotone.*

Example 3.16. *There is a mapping $\mathcal{I}$ from maximally monotone operators in $\mathcal{H}$ into maximally monotone operators in the $\mathbb{R}$-Hilbert space[1] $L^2\big(0, T;\, \mathcal{H}\big)$. For any maximally monotone operator A in $\mathcal{H}$ we define the multivalued operator $\mathcal{I}A$ by*

$$\mathcal{I}Af = \{u \in L^2\big(0, T;\, \mathcal{H}\big) \mid u(t) \in Af(t) \text{ for a.e. } t \in [0, T]\} \qquad \forall f \in D(\mathcal{I}A),$$

where the domain of $\mathcal{I}A$ is defined by $\{f \in L^2\big(0, T;\, \mathcal{H}\big) \mid f(t) \in D(A) \text{ for a.e. } t \in [0, T]\}$. To show that $\mathcal{I}A$ is maximally monotone, by proposition 3.12 it suffices to verify that $\mathcal{I}A$ is monotone and $R(\operatorname{id}_{L^2} + \mathcal{I}A) = L^2\big(0, T;\, \mathcal{H}\big)$. Then $\mathcal{I}A$ is indeed monotone, since for all $(f_1, u_1), (f_2, u_2) \in \mathcal{I}A$ we see by the monotonicity of A that

$$\langle u_1 - u_2,\, f_1 - f_2 \rangle_{L^2} = \int_0^T \langle u_1 - u_2,\, f_1 - f_2 \rangle_{\mathcal{H}}\, dt \geq 0.$$

[1] See definition 4.4

Moreover it holds that $R(\mathrm{id}_{L^2} + \mathcal{I}\,A) = L^2\big(0,T;\,\mathcal{H}\big)$, since for any $u \in L^2\big(0,T;\,\mathcal{H}\big)$ there exists for almost all $t \in [0,T]$ some $f(t) \in D(A)$ with $u(t) = f(t) + Af(t)$. We have that $f \in L^2\big(0,T;\,\mathcal{H}\big)$ since by lemma 3.20 we know that $\|f(t)\|_{\mathcal{H}} \leq \|u(t)\|_{\mathcal{H}}$ for almost all $t \in [0,T]$ and thus by the dominated convergence theorem we infer the claim. Furthermore $f \in D(\mathcal{I}\,A)$ and $u = f + \mathcal{I}\,Af$.

Example 3.17. *Let A be a linear operator in $\mathcal{H}$. Then A is maximally monotone if and only if A is positive semi-definite, the range of A is dense in $\mathcal{H}$ and A is closed under monotone extensions.*

By definition the linear operator A is positive semi-definite if and only if its monotone.

So suppose first that A is maximally monotone. To show that the range is dense, suppose that $y \in R(A)^\perp$. So $\langle A\eta - y,\, \eta - 0\rangle_{\mathcal{H}} \geq 0$ for all $\eta \in D(A)$ and thus by condition (2) we have $y = A0 = 0$. The fact that A is closed under monotone extensions holds by definition.

For the converse direction, suppose that A has a dense range and is closed under monotone extensions. Assume that the pair $(x,y) \in \mathcal{H} \times \mathcal{H}$ satisfies the necessary condition (2), that is $\langle A\eta - y,\, \eta - x\rangle_{\mathcal{H}} \geq 0$ for all $\eta \in D(A)$. Then the linear mapping $D(A) + \mathbb{R}x \to \mathcal{H}, \eta + \lambda x \mapsto A\eta + \lambda y$ is a monotonic extension of A. Indeed for $\eta_1, \eta_2 \in D(A)$ and $\lambda_1, \lambda_2 \in \mathbb{R}$ we compute by the linearity of A and the property of (x,y) that

$$\begin{aligned}&\langle (A\eta_1 + \lambda_1 y) - (A\eta_2 + \lambda_2 y),\, (\eta_1 + \lambda_1 x) - (\eta_2 - \lambda_2 x)\rangle_{\mathcal{H}}\\ &\qquad = \langle A(\eta_1 - \eta_2) + (\lambda_1 - \lambda_2)y,\, (\eta_1 - \eta_2) + (\lambda_1 - \lambda_2)x\rangle_{\mathcal{H}}\\ &\qquad \geq 0.\end{aligned}$$

As A is closed under monotonic extensions, we determine that $x \in D(A)$ and $Ax = y$. Hence we see from condition (2) that A is a maximally monotone operator.

Example 3.18. *Let $\mathcal{K}$ be any $\mathbb{R}$-Hilbert space and define the $\mathbb{R}$-Hilbert space $\mathcal{H} = L^2(0,\mathcal{L};\mathcal{K})$ and the linear operator A on $\mathcal{H}$ as the spatial derivative $A = \partial_x$ with domain defined by the set*[2] *$\{f \in W^{1,2}(0,\mathcal{L};\mathcal{K}) \mid f(0) = 0\}$. The operator is positive semi-definite since*

$$\langle Af,\, f\rangle_{L^2} = \int_0^{\mathcal{L}} \langle \partial_x f,\, f\rangle_{\mathcal{K}}\, dx = \frac{1}{2}\,\|f(\mathcal{L})\|_{\mathcal{K}}^2 \geq 0.$$

It can be shown that A has dense range in $\mathcal{H}$. Indeed if $g \in \mathcal{H}$ is such that $g(0) = 0$, then its integral $G(x) := \int_0^x g(y)\,dy$ for $x \in [0,T]$ is in the domain of A and it will map to g, i.e. $AG = g$. It thus suffices to prove that any constant function can be approximated arbitrarily well. But this is clear, by taking the integral of a step function that is sufficiently large near a neighbourhood of 0 and vanishes everywhere else. Thus by example 3.17 we know that A is maximally monotone.

In fact for any $u_0 \in \mathcal{K}$ we know that $A + u_0$ is also maximally monotone. Thus we may exchange the domain of A with the domain that sets u_0 as an initial condition, i.e. we can swap $D(A)$ with the set $\{f \in W^{1,2}(0,\mathcal{L};\mathcal{K}) \mid f(0) = u_0\}$.

3.2 The Resolvent and Yosida Approximation

In this section we will define two single valued operators, namely the resolvent and the Yosida approximation of maximal monotone operators. These constructions are very important tools for the study of maximal monotone operators and gradient flow, on which this paper will rely on many

[2] The set $W^{1,p}(0,\mathcal{L};\mathcal{K})$ is the Sobolev space (cf. definition 4.11)

times. In addition we will show for a maximal monotone multivalued operator, that the closure of its domain is convex and that its resolvent strongly converges to the projection operator of its own domain. We end the section by studying the attributes of the Yosida approximation.

Throughout this section let A be a maximally monotone multivalued operator in $\mathcal{H}$. For a convex set $C \subseteq \mathcal{H}$, we let $\mathrm{Pr}_C : \mathcal{H} \to \overline{C}$ be the projection mapping of $\mathcal{H}$ onto the closure of C.

Definition 3.19. *Let A be a maximally monotone multivalued operator and $\lambda > 0$ arbitrary. Define the resolvent of A (with parameter λ) as the multivalued operator $\mathcal{J}_\lambda := (\mathrm{id}_{\mathcal{H}} + \lambda A)^{-1}$.*

Lemma 3.20. *The resolvent $\mathcal{J}_\lambda$ of a maximally monotone multivalued operator A in $\mathcal{H}$ is a single valued operator with image $D(A)$. Specifically it is a contraction from $\mathcal{H}$ onto $D(A)$.*

Proof. From proposition 3.12 we learn that $D(\mathcal{J}_\lambda) = \mathcal{H}$. By definition of the inverse operator, we have that $R(\mathcal{J}_\lambda) = D(A)$. From lemma 3.5 we learn the resolvent is a contraction. In particular this means that the mapping is single valued. □

We analyze the asymptotic behavior of the resolvent $\mathcal{J}_\lambda$ of A whenever $\lambda \searrow 0$.

Example 3.21. *Let A be defined as in example 3.15 with $f : (-1,1) \to \mathbb{R}$; $x \to \ln(1-x) - \ln(1+x)$. From example 3.15 we know that A is a maximally monotone operator with $D(A) = (-1,1)$. Note for all $x \in (-1,1)$ we have $x + \lambda Ax = x + \lambda f(x) \to x$ when $\lambda \searrow 0$. This shows that $\mathcal{J}_\lambda x \to \mathrm{Pr}_{D(A)}\, x$ for all $x \in \mathbb{R}$. Compare this with figure 1.*

Example 3.22. *Define the Hilbert space $\mathcal{H}$ and maximally monotone multivalued operator A as in example 3.18. We assume that $D(A)$ sets $u_0 \in \mathcal{H}$ as an initial condition for any $u_0 \in \mathcal{H}$. Thus for every $u \in D(A)$ and all $\lambda > 0$ the resolvent $\mathcal{J}_\lambda$ of A has the explicit form*

$$\mathcal{J}_\lambda u(x) = e^{-\frac{x}{\lambda}} u_0 + \frac{1}{\lambda} \int_0^x e^{\frac{y-x}{\lambda}} u(y)\, dy \qquad \forall x \in [0, \mathcal{L}].$$

Indeed the following computation verifies the claim

$$\begin{aligned} &(\mathrm{id}_{L^2} + \lambda A) e^{-\frac{x}{\lambda}} u_0 + \frac{1}{\lambda} \int_0^x e^{\frac{y-x}{\lambda}} u(y)\, dy \\ &\qquad = e^{-\frac{x}{\lambda}} u_0 + \lambda A e^{-\frac{x}{\lambda}} u_0 + \frac{1}{\lambda} e^{\frac{-x}{\lambda}} \int_0^x e^{\frac{y}{\lambda}} u(y)\, dy + A e^{\frac{-x}{\lambda}} \int_0^x e^{\frac{y}{\lambda}} u(y)\, dy \\ &\qquad = e^{-\frac{x}{\lambda}} u_0 - e^{\frac{-x}{\lambda}} u_0 + \frac{1}{\lambda} e^{\frac{-x}{\lambda}} \int_0^x e^{\frac{y}{\lambda}} u(y)\, dy - \frac{1}{\lambda} e^{\frac{-x}{\lambda}} \int_0^x e^{\frac{-x}{\lambda}} u(y)\, dy + u(x) \\ &\qquad = u(x). \end{aligned} \qquad \forall x \in [0, \mathcal{L}].$$

Using integration by parts and the weak derivative[3]*, we see that $\mathcal{J}_\lambda \to \mathrm{id}_{L^2}$ whenever $\lambda \searrow 0$.*

$$\mathcal{J}_\lambda u(x) = e^{-\frac{x}{\lambda}} u_0 + \frac{1}{\lambda} \int_0^x e^{\frac{y-x}{\lambda}} u(y)\, dy = e^{-\frac{x}{\lambda}} u_0 + \left[e^{\frac{y-x}{\lambda}} u(y) \right]_0^x - \int_0^x e^{\frac{y-x}{\lambda}} \partial_y u(y)\, dy.$$

If we take the limit $\lambda \to 0$ we see that

$$\lim_{\lambda \searrow 0} \mathcal{J}_\lambda u(x) = \lim_{\lambda \searrow 0} e^{-\frac{x}{\lambda}} u_0 + \left[e^{\frac{y-x}{\lambda}} u(y) \right]_0^x - \int_0^x e^{\frac{y-x}{\lambda}} \partial_y u(y)\, dy = u(x).$$

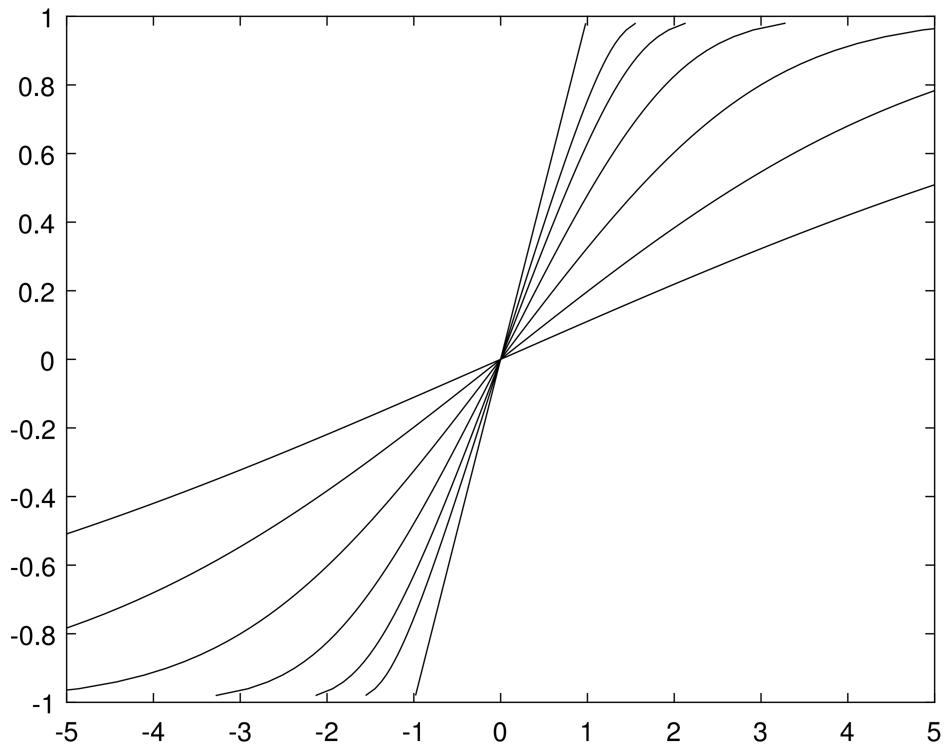

Figure 1: Shows how the resolvent $\mathcal{J}_\lambda$ approximates the projection $\mathrm{Pr}_{D(A)}$ whenever $\lambda \searrow 0$.

The following theorem shows that the examples generalise to all maximal monotone operators.

Theorem 3.23. *The set $\overline{D(A)}$ is convex and $\mathcal{J}_\lambda x \to \mathrm{Pr}_{\overline{D(A)}} x$ as $\lambda \searrow 0$ for all $x \in \mathcal{H}$.*

Proof. Define $C := \overline{\mathrm{conv}\ D(A)}$ and for any $x \in \mathcal{H}$ define $x_\lambda := \mathcal{J}_\lambda x$. First we show that x_λ is bounded as $\lambda \searrow 0$. By definition of the resolvent we have $x \in x_\lambda + \lambda A x_\lambda$, then with the monotonicity of A we get for any $(\eta, \xi) \in A$ the inequality

$$\left\langle \frac{x - x_\lambda}{\lambda} - \xi,\, x_\lambda - \eta \right\rangle_{\mathcal{H}} \geq 0.$$

We multiply with λ and use the Cauchy-Schwartz inequality to prove the boundedness of x_λ by

$$\|x_\lambda\|_{\mathcal{H}}^2 \leq \langle \lambda\xi - x,\, \eta\rangle_{\mathcal{H}} + \langle x_\lambda,\, \eta + x - \lambda\xi\rangle_{\mathcal{H}} \leq \langle \lambda\xi - x,\, \eta\rangle_{\mathcal{H}} + \|x_\lambda\|_{\mathcal{H}} \|\eta + x - \lambda\xi\|_{\mathcal{H}}. \tag{3}$$

Next we show that $x_\lambda \to \mathrm{Pr}_{\overline{D(A)}} x$ as $\lambda \searrow 0$. By lemma 3.20 we know that $(x_\lambda)_{\lambda > 0}$ is contained in $D(A)$. As any bounded sequence in a Hilbert space has a weakly convergent subsequence, there exists a positive null-sequence $(\lambda_n)_{n\in\mathbb{N}}$ such that $x_{\lambda_n} \rightharpoonup x_0$ for some $x_0 \in C$. If we apply the limit to (3), we get

$$\|x_0\|_{\mathcal{H}}^2 \leq \limsup_{n\to\infty} \|x_\lambda\|_{\mathcal{H}}^2 = \langle x_0,\, \eta + x\rangle_{\mathcal{H}} - \langle x,\, \eta\rangle_{\mathcal{H}} \qquad \Rightarrow \qquad \langle x - x_0,\, \eta - x_0\rangle_{\mathcal{H}} \leq 0. \tag{4}$$

Observe the inequality on the right hand side in (4) holds for all $\eta \in D(A)$ and thus for all $\eta \in C$ (by a similar argument as in remark 3.10). Yet as C is closed and convex, this inequality can

[3] See definition 4.14

only hold if $x_0 = \Pr_C x$. Note that this result is independent of the choice of null-sequence, which implies $x_\lambda \rightharpoonup \Pr_C x$ as $\lambda \searrow 0$.
Lastly we must show that $\overline{D(A)}$ is convex. The same argument as above shows that the inequality on the left hand side of (4) holds for all $\eta \in C$. In particular if we set $\eta = x_0$ in this inequality we deduce

$$\limsup_{\lambda \to \infty} \|x_\lambda\|_{\mathcal{H}}^2 = \langle x_0,\, x_0 + x \rangle_{\mathcal{H}} - \langle x,\, x_0 \rangle_{\mathcal{H}} = \|x_0\|_{\mathcal{H}}^2 \tag{5}$$

Then (5) and weak convergence is enough to imply $x_\lambda \to x_0$. Thus if x is in C, then the sequence $(x_\lambda)_{\lambda>0}$ is in $D(A)$ and converges to x. This implies $\overline{D(A)} = C$ and hence $\overline{D(A)}$ is convex. □

Now we introduce the Yosida approximation. A vital tool for studying gradient flow and proper convex lower semi-continuous functions.

Definition 3.24. *Let B be a maximally monotone multivalued operator in $\mathcal{H}$. Define the Yosida approximation of B as the multivalued operator $B_\lambda := \lambda^{-1}(\mathrm{id}_{\mathcal{H}} - \mathcal{J}_\lambda)$.*

Remark 3.25. • *The Yosida Approximation is a single valued operator on $\mathcal{H}$.*

• *The operators A and A_λ are related for all $x \in \mathcal{H}$ by $A_\lambda x \in A\,\mathcal{J}_\lambda x$. Indeed this follows from*

$$A_\lambda x \in A\,\mathcal{J}_\lambda x \qquad \Leftrightarrow \qquad x \in (\mathrm{id}_{\mathcal{H}} + \lambda A)\,\mathcal{J}_\lambda x \qquad \Leftrightarrow \qquad x \in \mathcal{J}_\lambda^{-1}\,\mathcal{J}_\lambda x.$$

• *In case A is a linear operator, then A and A_λ are related for all $\eta \in D(A)$ by $A_\lambda \eta = \mathcal{J}_\lambda A\eta$. The same argument as above holds, but we can replace the membership with an equality.*

Definition 3.26. *Let B be a maximally monotone multivalued operator in $\mathcal{H}$. Define the minimum norm section B_0 of B to be the single valued operator with domain $D(B_0) = D(B)$ and is defined by $B_0\eta := \arg\min_{x \in B\eta} \|x\|_{\mathcal{H}}$ for $\eta \in D(B_0)$. This map is well-defined by lemma 3.10.*

We will need the following lemma a few times during this paper.

Lemma 3.27. *Let $(x_n, y_n)_{n\in\mathbb{N}} \subseteq A$ be a sequence. Then we have $(x, y) \in A$ if at least one of the following conditions is true.*

(i) We observe that $x_n \to x$ and $y_n \to y$ when $n \to +\infty$.

(ii) We can verify $x_n \to x$ and $y_n \rightharpoonup y$ whenever $n \to +\infty$.

(iii) We have that $x_n \rightharpoonup x$, $y_n \rightharpoonup y$ and $|\langle x_n,\, y_n \rangle_{\mathcal{H}}|$ is bounded from above by $|\langle x,\, y \rangle_{\mathcal{H}}|$ as $n \to +\infty$.

Proof. We have that (i) follows from (ii), and (ii) follows from (iii). Indeed the first implication is clear and for the second we have that as $(y_n)_{n\in\mathbb{N}}$ is weakly convergent, it is bounded in norm by some $M > 0$ and thus we deduce with the Cauchy-Schwarz inequality

$$\begin{aligned}
|\langle x_n,\, y_n \rangle_{\mathcal{H}} - \langle x,\, y \rangle_{\mathcal{H}}| &\le |\langle x_n - x,\, y_n \rangle_{\mathcal{H}}| + |\langle x,\, y_n - y \rangle_{\mathcal{H}}| \\
&\le \|y_n\|_{\mathcal{H}} \|x_n - x\|_{\mathcal{H}} + |\langle x,\, y_n - y \rangle_{\mathcal{H}}| \\
&\le M \|x_n - x\|_{\mathcal{H}} + |\langle x,\, y_n - y \rangle_{\mathcal{H}}| \\
&\to 0.
\end{aligned}$$

For the property (iii), it suffices to verify the condition (2). To that end for all $(\eta, \xi) \in A$ we compute

$$\begin{aligned} 0 \leq \limsup_{n\to\infty} \langle y_n - \xi,\, x_n - \eta\rangle_{\mathcal{H}} &= \limsup_{n\to\infty} \langle y_n,\, x_n\rangle_{\mathcal{H}} - \langle \xi,\, x_n\rangle_{\mathcal{H}} - \langle y_n,\, \eta\rangle_{\mathcal{H}} + \langle \xi,\, \eta\rangle_{\mathcal{H}} \\ &\leq \langle y,\, x\rangle_{\mathcal{H}} - \langle \xi,\, x\rangle_{\mathcal{H}} - \langle y,\, \eta\rangle_{\mathcal{H}} + \langle \xi,\, \eta\rangle_{\mathcal{H}} \\ &\leq \langle y - \xi,\, x - \eta\rangle_{\mathcal{H}} \end{aligned} .$$

□

The Yosida Approximation has important properties, which we will use to construct the theory of Gradient Flows.

Proposition 3.28. *Let A be a maximally monotone multivalued operator in $\mathcal{H}$. Then the Yosida approximation of A has the following properties*

(i) The mapping A_λ is λ^{-1}-Lipschitz.

(ii) The function A_λ is maximally monotone.

(iii) We have the homomorphicity property $(A_\mu)_\lambda = A_{\mu+\lambda}$ for $\lambda, \mu > 0$.

(iv) For all $x \in D(A)$ and for any $0 < \mu \leq \lambda$ we have $\|A_\lambda x\|_{\mathcal{H}} \leq \|A_\mu x\|_{\mathcal{H}} \leq \|A_0 x\|_{\mathcal{H}}$. In fact we have the estimate $\|A_\lambda x - A_\mu x\|^2_{\mathcal{H}} \leq \|A_\mu x\|^2_{\mathcal{H}} - \|A_\lambda x\|^2_{\mathcal{H}}$.

(v) For all $x \in D(A)$ we have $A_\lambda x \to A_0 x$ when $\lambda \searrow 0$.

(vi) If $x \notin D(A)$, then $\|A_\lambda x\|_{\mathcal{H}} \nearrow \infty$ when $\lambda \searrow 0$.

Proof. "(i)": As A is monotone we see with the Cauchy-Schwarz inequality and by remark 3.25 for all $\eta_1, \eta_2 \in \mathcal{H}$ that

$$\begin{aligned} \|A_\lambda\eta_1 - A_\lambda\eta_2\|_{\mathcal{H}} \|\eta_1 - \eta_2\|_{\mathcal{H}} &\geq \langle A_\lambda\eta_1 - A_\lambda\eta_2,\, \eta_1 - \eta_2\rangle_{\mathcal{H}} \\ &= \langle A_\lambda\eta_1 - A_\lambda\eta_2,\, \lambda A_\lambda\eta_1 - \lambda A_\lambda\eta_2 + \mathcal{J}_\lambda\eta_1 - \mathcal{J}_\lambda\eta_2\rangle_{\mathcal{H}} \\ &= \lambda \|A_\lambda\eta_1 - A_\lambda\eta_2\|^2_{\mathcal{H}} + \langle A_\lambda\eta_1 - A_\lambda\eta_2,\, \mathcal{J}_\lambda\eta_1 - \mathcal{J}_\lambda\eta_2\rangle_{\mathcal{H}} \\ &\geq \lambda \|A_\lambda\eta_1 - A_\lambda\eta_2\|^2_{\mathcal{H}} . \end{aligned}$$

"(ii)": Suppose for some $(x, y) \in \mathcal{H} \times \mathcal{H}$ we have

$$0 \leq \langle \xi - y,\, \eta - x\rangle_{\mathcal{H}} \qquad \forall(\eta, \xi) \in A_\lambda. \tag{6}$$

By remark 3.25 we have that $D(A_\lambda) = \mathcal{H}$. Thus in (6) we get for all $\eta \in \mathcal{H}$ and any $t > 0$ that

$$0 \leq \big\langle A_\lambda\big((1-t)x + t\eta\big) - y,\, \big((1-t)x + t\eta\big) - x\big\rangle_{\mathcal{H}} = t\, \langle A_\lambda((1-t)x + t\eta) - y,\, \eta - x\rangle_{\mathcal{H}} .$$

Because A_λ is λ^{-1}-Lipschitz by property (i), we can divide by t and take the limit $t \searrow 0$ to compute

$$0 \leq \langle A_\lambda x - y,\, \eta - x\rangle_{\mathcal{H}} .$$

As this inequality must hold for all $\eta \in \mathcal{H}$, we deduce that $A_\lambda x = y$. This proves that A_λ is maximally monotone.

"(iii)": The claim is an immediate consequence of the following equivalence chain

$$(\eta,\xi) \in A_\lambda \quad \Leftrightarrow \quad \mathcal{J}_\lambda \eta = \eta - \lambda\xi \quad \Leftrightarrow \quad \eta \in (\eta - \lambda\xi) + \lambda A(\eta - \lambda\xi) \quad \Leftrightarrow \quad (\eta - \lambda\xi, \xi) \in A.$$

"(iv)": For $0 < \mu < \lambda$ and by remark 3.25 we compute using the monotonicity of A and the Cauchy-Schwarz inequality

$$\begin{aligned}\langle A_0 x - A_{\lambda-\mu} x,\, x - \mathcal{J}_{\lambda-\mu}\, x\rangle_{\mathcal{H}} \geq 0 \quad &\Rightarrow \quad \langle A_0 x - A_{\lambda-\mu} x,\, A_{\lambda-\mu} x\rangle_{\mathcal{H}} \geq 0 \\ &\Rightarrow \quad \|A_0 x\|_{\mathcal{H}} \|A_{\lambda-\mu} x\|_{\mathcal{H}} \geq \langle A_0 x,\, A_{\lambda-\mu} x\rangle_{\mathcal{H}} \geq \|A_{\lambda-\mu} x\|^2_{\mathcal{H}}\,.\end{aligned}$$

We extract the following inequalities which are of interest to us

$$\|A_{\lambda-\mu} x\|_{\mathcal{H}} \leq \|A_0 x\|_{\mathcal{H}}\,, \tag{7}$$

$$\|A_{\lambda-\mu} x\|^2_{\mathcal{H}} \leq \langle A_0 x,\, A_{\lambda-\mu} x\rangle_{\mathcal{H}}\,. \tag{8}$$

Now we show the first bound of (iv). From (7) we can easily see that $\|A_\mu x\|_{\mathcal{H}} \leq \|A_0 x\|_{\mathcal{H}}$. Note that $(A_\mu)_0 = A_\mu$ as A_μ is single valued from remark 3.25. So we may replace A with A_μ in (7) and use property (iii) to get $\|A_\lambda x\|_{\mathcal{H}} \leq \|A_\mu x\|_{\mathcal{H}}$.
For the second bound in (iv), we argue analogue as before and transform (7) into the inequality $\langle A_\lambda x,\, A_\mu x\rangle_{\mathcal{H}} \leq \|A_\lambda x\|^2_{\mathcal{H}}$. Then we get the desired inequality

$$\|A_\lambda x - A_\mu x\|^2_{\mathcal{H}} = \|A_\lambda x\|^2_{\mathcal{H}} + \|A_\mu x\|^2_{\mathcal{H}} - 2\,\langle A_\lambda x,\, A_\mu x\rangle_{\mathcal{H}} \leq \|A_\mu x\|^2_{\mathcal{H}} + \|A_\lambda x\|^2_{\mathcal{H}} - 2\,\|A_\lambda x\|^2_{\mathcal{H}}\,.$$

"(v)": Property (iv) tells us that $\|A_\lambda x\|_{\mathcal{H}}$ is monotonically increasing as $\lambda \searrow 0$ with upper bound $\|A_0 x\|_{\mathcal{H}}$. This fact coupled with the second property (iv) shows that the sequence $(A_\lambda x)_{\lambda>0}$ is Cauchy and hence strongly converges to some $y \in \mathcal{H}$ with $\|y\|_{\mathcal{H}} \leq \|A_0 x\|_{\mathcal{H}}$. Moreover by theorem 3.23 as $x \in D(A)$ we have that $\mathcal{J}_\lambda\, x \to \mathrm{Pr}_{\overline{D(A)}}\, x = x$, and by remark 3.25 we have a sequence $(\mathcal{J}_\lambda\, x, A_\lambda x)_{\lambda>0} \subseteq A$. Thus by lemma 3.27 we see that $(x, y) \in A$. By definition of the minimum norm section and as $\|y\|_{\mathcal{H}} \leq \|A_0 x\|_{\mathcal{H}}$ we must have $y = A_0 x$.

"(vi)": We show the contra-positive. So suppose that the sequence $\|A_\lambda x\|_{\mathcal{H}}$ is bounded as $\lambda \searrow 0$. As an immediate consequence we get $x - \mathcal{J}_\lambda\, x = \lambda A_\lambda x \to 0$. Then by theorem 3.23 we have that $\mathcal{J}_{\lambda_n}\, x \to \mathrm{Pr}_{\overline{D(A)}}\, x = x$. There must exist some positive null-sequence $\lambda_n \to 0$ and $y \in \mathcal{H}$ such that $A_{\lambda_n} x \rightharpoonup y$. Then by lemma 3.27 we have $(x, y) \in A$. In particular we derive that $x \in D(A)$ thus proving the property (vi). ☐

3.3 Proper Convex Lower Semicontinous Potentials in Hilbert Spaces

In this section we analyse an important class of maximal monotone operators, namely the sub differentials of proper convex lower semicontinous functions in a Hilbert Space.

For any mapping $\varphi : \mathcal{H} \to (-\infty, +\infty]$ we define its domain $D(\varphi)$ by the set $\{x \in \mathcal{H} \mid \varphi(x) \neq +\infty\}$.

Definition 3.29. *A mapping $\varphi : \mathcal{H} \to (-\infty, +\infty]$ is called a proper convex function if it satisfies*

(i) The domain of φ is non-empty, i.e. $D(\varphi) \neq \emptyset$.

(ii) For all $x, y \in D(\varphi)$ and all $0 < \lambda < 1$ we have $\varphi(\lambda x + (1-\lambda) y) \leq \lambda\varphi(x) + (1-\lambda)\varphi(y)$.

Remark 3.30. *Any proper convex function $\varphi : \mathcal{H} \to (-\infty, +\infty]$ is complemented with a convex domain. Indeed this is an immediate consequence of the definition.*

Definition 3.31. *A function $\varphi : \mathcal{H} \to (-\infty, +\infty]$ is lower semi-continuous (l.s.c) if it satisfies the lower continuity condition, i.e. if it satisfies*

$$\forall x \in D(\varphi)\, \forall \eta > 0\, \exists \delta > 0\, \forall y \in D(\varphi) \quad : \ \|x - y\|_{\mathcal{H}} < \delta \implies f(x) - \varepsilon < f(y).$$

Remark 3.32. *Let $\varphi : \mathcal{H} \to (-\infty, +\infty]$ be a l.s.c. function, $(x_n)_{n\in\mathbb{N}} \in D(\varphi)$ a sequence with limit $x \in D(\varphi)$. As an immediate consequence of the definition we have the following limit behaviour*

$$\varphi(x) \leq \liminf_{n\to\infty} \varphi(x_n).$$

Remark 3.33. *From [11] we know that for every proper lower semi-continuous function $\varphi : \mathcal{H} \to (-\infty, +\infty]$ there exists a family of affine functions $\ell_i(x) := \langle v_i,\, x\rangle_{\mathcal{H}} + b_i$ for $i \in I$, $v_i \in \mathcal{H}$, $b_i \in \mathbb{R}$ and every $x \in \mathcal{H}$, such that*

$$\varphi(x) = \sup_{i\in I} \ell_i(x) \qquad \forall x \in \mathcal{H}.$$

Definition 3.34. *For any proper convex l.s.c function $\varphi : \mathcal{H} \to (-\infty, +\infty]$ we define the sub differential mapping $\partial\varphi$ of φ as the multivalued operator in $\mathcal{H}$ with the graph defined by*

$$\partial\varphi := \{(\eta, \xi) \in D(\varphi) \times \mathcal{H} \mid \forall x \in D(\varphi) \ : \ \varphi(x) \geq \varphi(\eta) + \langle \xi,\, x - \eta\rangle_{\mathcal{H}}\}.$$

Remark 3.35. *From the definition we have that $D(\partial\varphi) \subseteq D(\varphi)$.*

In other literature a proper convex l.s.c mapping $\varphi : \mathcal{H} \to (-\infty, +\infty]$ is also called a potential on $\mathcal{H}$. We can show that the sub differential mapping is a multivalued maximal monotone operator in $\mathcal{H}$. This follows from the next two lemmas.

Lemma 3.36. *The sub differential $\partial\varphi$ of a proper convex l.s.c. function $\varphi : \mathcal{H} \to (-\infty, +\infty]$ is monotone.*

Proof. We have for all $(\eta_1, \xi_1), (\eta_2, \xi_2) \in \partial\varphi$ by definition of the sub differential

$$\begin{aligned} \varphi(\eta_2) &\geq \varphi(\eta_1) + \langle \xi_1,\, \eta_2 - \eta_1\rangle_{\mathcal{H}}\,, \\ \varphi(\eta_1) &\geq \varphi(\eta_2) + \langle \xi_2,\, \eta_1 - \eta_2\rangle_{\mathcal{H}}\,. \end{aligned}$$

We can reformulate these inequalities as

$$\begin{aligned} \langle \xi_1,\, \eta_1 - \eta_2\rangle_{\mathcal{H}} &\geq \quad \varphi(\eta_1) - \varphi(\eta_2), \\ \langle -\xi_2,\, \eta_1 - \eta_2\rangle_{\mathcal{H}} &\geq -(\varphi(\eta_1) - \varphi(\eta_2)). \end{aligned}$$

We prove the claim by taking the sum of these inequalities since

$$\langle \xi_1 - \xi_2,\, \eta_1 - \eta_2\rangle_{\mathcal{H}} \geq 0.$$

□

Lemma 3.37. *Let $\varphi : \mathcal{H} \to (-\infty, +\infty]$ be a proper convex l.s.c. function, pick any $\alpha > 0$ and let $\eta, \xi \in \mathcal{H}$. Then the mapping $\psi : \mathcal{H} \to (-\infty, +\infty]$; $x \mapsto \varphi(x) + \frac{\alpha}{2}\|x - \xi\|^2_{\mathcal{H}}$ attains its minimal value at η if and only if $\alpha(\xi - \eta) \in \partial\varphi(\eta)$.*

Proof. Suppose first that the mapping ψ attains its minimal value at η. Then for all $x \in \mathcal{H}$ and all $0 < t < 1$ we have by the monotonicity of φ that

$$\begin{aligned} t(\varphi(x) - \varphi(\eta)) &\geq \varphi\big((1-t)\eta + tx\big) - \varphi(\eta) \\ &\geq \frac{\alpha}{2} \|\eta - \xi\|_{\mathcal{H}}^2 - \frac{\alpha}{2} \|(1-t)\eta + tx - \xi\|_{\mathcal{H}}^2 \\ &= \frac{\alpha}{2} \|\eta - \xi\|_{\mathcal{H}}^2 - \frac{\alpha}{2} \|t(x-\eta) + \eta - \xi\|_{\mathcal{H}}^2 \\ &\geq -t^2 \frac{\alpha}{2} \|x - \eta\|_{\mathcal{H}}^2 - t\alpha \langle x - \eta, \, \eta - \xi \rangle_{\mathcal{H}}. \end{aligned}$$

If we divide by t and take the limit $t \searrow 0$, we prove $\alpha(\xi - \eta) \in \partial\varphi(\eta)$ by

$$\varphi(x) - \varphi(\eta) \geq -\alpha \langle x - \eta, \, \eta - \xi \rangle_{\mathcal{H}} \qquad \Rightarrow \qquad \varphi(x) \geq \varphi(\eta) + \langle \alpha(\xi - \eta), \, x - \eta \rangle_{\mathcal{H}}.$$

Now assume that the converse holds true, i.e. $\alpha(\xi - \eta) \in \partial\varphi(\eta)$. Then ψ is has a minimum at η by the following computation

$$\begin{aligned} \varphi(x) - \varphi(\eta) &\geq \alpha \langle \xi - \eta, \, x - \eta \rangle_{\mathcal{H}} \\ &\geq \alpha \langle \xi - \eta, \, x - \eta \rangle_{\mathcal{H}} - \frac{\alpha}{2} \|x - \eta\|_{\mathcal{H}}^2 \qquad \forall x \in \mathcal{H}. \\ &= \frac{\alpha}{2} \|\eta - \xi\|_{\mathcal{H}}^2 - \frac{\alpha}{2} \|x - \xi\|_{\mathcal{H}}^2. \end{aligned}$$

□

Corollary 3.38. *The sub differential $\partial\varphi$ of a proper convex l.s.c. function $\varphi : \mathcal{H} \to (-\infty, +\infty]$ is maximally monotone.*

Proof. From lemma 3.36 we know that $\partial\varphi$ is monotone. So by proposition 3.12 it suffices to prove that $R(\mathrm{id}_{\mathcal{H}} + \partial\varphi) = \mathcal{H}$. Note that the function $\psi : \mathcal{H} \to (-\infty, +\infty]$; $x \mapsto \varphi(x) + \frac{1}{2} \|x - \xi\|_{\mathcal{H}}^2$ diverges towards $+\infty$ as $\|x\|_{\mathcal{H}} \to +\infty$ for all $\xi \in \mathcal{H}$. Indeed this is clear if φ is an affine function. Then by remark 3.33 the claim holds for any l.s.c. proper convex function. Therefore ψ has some minimal value $\eta \in D(\varphi)$.
But by lemma 3.37 we know that η is a minimal value of ψ if and only if $\xi - \eta \in \partial\varphi(\eta)$. This means that $\xi \in \eta + \partial\varphi(\eta)$ from which follows the claim. □

We will need an elementary lemma for the next proposition, characterizing the convexity of a differentiable function on $\mathbb{R}$ by its first derivative. We will only state the lemma and not the proof, as it is only tangentially related to the subject.

Lemma 3.39. *Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a differentiable function, then we have that $\partial_t f$ is monotonically increasing if and only if f is convex.*

The upcoming proposition shows that the Yosida approximation of the sub differential operator of a proper convex l.s.c function, is the sub differential of a Fréchet differentiable convex function on $\mathcal{H}$.

Proposition 3.40. *Let φ be proper convex l.s.c. mapping on $\mathcal{H}$ and define $A := \partial\varphi$. Then for all $\lambda > 0$ define the function*

$$\varphi_\lambda(x) = \min_{y \in \mathcal{H}} \frac{1}{2\lambda} \|x - y\|_{\mathcal{H}}^2 + \varphi(y). \tag{9}$$

This function has the following properties *(i) The function φ_λ has an equivalent formulation*

$$\varphi_\lambda(x) = \frac{\lambda}{2} \|A_\lambda x\|_{\mathcal{H}}^2 + \varphi(\mathcal{J}_\lambda x), \tag{10}$$

(ii) The map φ_λ is Fréchet-differentiable with derivative $\partial_x \varphi_\lambda(x) y = \langle A_\lambda x,\, y\rangle_{\mathcal{H}}$. Moreover the sub differential of φ_λ coincides with the Yosida approximation A_λ, i.e $(\partial\varphi_\lambda)_\lambda = A_\lambda$.

(iii) The function φ_λ is convex function with domain $\mathcal{H}$.

(iv) We have that $\varphi_\lambda(x) \nearrow \varphi(x)$ whenever $\lambda \searrow 0$ for all $x \in \mathcal{H}$.

(v) The domain of φ and A are related by $\overline{D(\varphi)} = \overline{D(A)}$.

Proof. First we tackle the property (i). By remark 3.25 we have $\frac{1}{\lambda}(x - \mathcal{J}_\lambda x) = A_\lambda x \in A\,\mathcal{J}_\lambda x$. Then if in lemma 3.37 we set $\alpha = \lambda^{-1}$, $\xi = x$ and $\eta = \mathcal{J}_\lambda x$, we have that the map $\mathcal{H} \to$ (-$\infty, +\infty$]; $y \mapsto \frac{1}{2\lambda}\|x - y\|_{\mathcal{H}}^2 + \varphi(y)$ attains its minimal value at $\mathcal{J}_\lambda x$, thus proving the alternate definition of φ_λ in (10).

Next we verify that φ_λ is Fréchet differentiable. For all $x, y \in D(\varphi)$ note by (10) that

$$\begin{aligned}
\varphi_\lambda(x) - &\varphi_\lambda(y) \\
&= \frac{\lambda}{2}(\|A_\lambda x\|_{\mathcal{H}}^2 - \|A_\lambda y\|_{\mathcal{H}}^2) + \varphi(\mathcal{J}_\lambda x) - \varphi(\mathcal{J}_\lambda y) \\
&\geq \frac{\lambda}{2}(\|A_\lambda x\|_{\mathcal{H}}^2 - \|A_\lambda y\|_{\mathcal{H}}^2) + \langle A_\lambda y,\, \mathcal{J}_\lambda x - \mathcal{J}_\lambda y\rangle_{\mathcal{H}} \\
&= \frac{\lambda}{2}(\|A_\lambda x\|_{\mathcal{H}}^2 - \|A_\lambda y\|_{\mathcal{H}}^2) + \langle A_\lambda y,\, y - \mathcal{J}_\lambda y\rangle_{\mathcal{H}} - \langle A_\lambda y,\, x - \mathcal{J}_\lambda x\rangle_{\mathcal{H}} + \langle A_\lambda y,\, x - y\rangle_{\mathcal{H}} \\
&= \frac{\lambda}{2}\Big(\|A_\lambda x\|_{\mathcal{H}}^2 - \|A_\lambda y\|_{\mathcal{H}}^2 + 2\langle A_\lambda y,\, A_\lambda y - A_\lambda x\rangle_{\mathcal{H}} \Big) + \langle A_\lambda y,\, x - y\rangle_{\mathcal{H}}.
\end{aligned}$$

We reorder the inequality and continue

$$\begin{aligned}
\varphi_\lambda(x) - \varphi_\lambda(y) - &\langle A_\lambda y,\, x - y\rangle_{\mathcal{H}} \\
&\geq \frac{\lambda}{2}\big(\|A_\lambda x\|_{\mathcal{H}}^2 - \|A_\lambda y\|_{\mathcal{H}}^2 + 2\langle A_\lambda y,\, A_\lambda y - A_\lambda x\rangle_{\mathcal{H}} \big) \\
&= \frac{\lambda}{2}\big(\|A_\lambda x - A_\lambda y\|_{\mathcal{H}}^2 + 2\langle A_\lambda x - A_\lambda y,\, A_\lambda y\rangle_{\mathcal{H}} + 2\langle A_\lambda y,\, A_\lambda y - A_\lambda x\rangle_{\mathcal{H}} \big) \qquad (11) \\
&= \frac{\lambda}{2}\|A_\lambda x - A_\lambda y\|_{\mathcal{H}}^2 \\
&\geq 0.
\end{aligned}$$

We may swap x and y and negate to get $\varphi_\lambda(x) - \varphi_\lambda(y) - \langle A_\lambda x,\, x - y\rangle_{\mathcal{H}} \leq 0$. By restructuring this inequality we see

$$\varphi_\lambda(x) - \varphi_\lambda(y) - \langle A_\lambda y,\, x - y\rangle_{\mathcal{H}} \leq \langle A_\lambda x - A_\lambda y,\, x - y\rangle_{\mathcal{H}}. \tag{12}$$

From preposition 3.28 we know that A_λ is λ^{-1}-Lipshitz. Thus by using both (11) and (12) and the Cauchy-Schwarz inequality wee see that

$$|\varphi_\lambda(x) - \varphi_\lambda(y) - \langle A_\lambda y,\, x - y\rangle_{\mathcal{H}}| \leq \langle A_\lambda x - A_\lambda y,\, x - y\rangle_{\mathcal{H}} \leq \frac{1}{\lambda}\|x - y\|_{\mathcal{H}}^2.$$

This shows that φ_λ is Fréchet differentiable in $D(\varphi)$ with derivative $\partial_x \varphi_\lambda(x)y = \langle A_\lambda x,\, y\rangle_{\mathcal{H}}$ for all $x \in D(\varphi)$ and $y \in \mathcal{H}$.

Now we show property (iv). By lemma 3.39 it suffices to show for all $\eta_1, \eta_2 \in D(\varphi)$ that the mapping $(0,1) \to \mathbb{R};\, t \mapsto \partial_t \varphi_\lambda(t\eta_1 + (1-t)\eta_2)$ is monotonically increasing. By proposition 3.28 we know that A_λ is monotone. Thus for all $0 < s \le t < 1$ we have

$$\begin{aligned}
\big(\partial_t\varphi_\lambda(t\eta_1+&(1-t)\eta_2) - \partial_t\varphi_\lambda(s\eta_1 + (1-s)\eta_2)\big)\cdot(t-s)\\
&= \langle A_\lambda(t\eta_1 + (1-t)\eta_2) - A_\lambda(s\eta_1 + (1-s)\eta_2),\, \eta_1 - \eta_2\rangle_{\mathcal{H}} \cdot (t-s)\\
&= \langle A_\lambda(\eta_2 + t(\eta_1-\eta_2)) - A_\lambda(\eta_2 + s(\eta_1-\eta_2)),\, \eta_2 + t(\eta_1-\eta_2) - (\eta_2 + s(\eta_1-\eta_2))\rangle_{\mathcal{H}}\\
&\ge 0.
\end{aligned}$$

Next we show statement (v). From the two function definitions (9) and (10) we see that $\varphi_\lambda(x)$ monotonically increases as λ decreases and that $\varphi(\mathcal{J}_\lambda x) \le \varphi_\lambda(x) \le \varphi(x)$. For any $\eta \in \overline{D(A)}$ we have by theorem 3.23 that $\mathcal{J}_\lambda \eta \to \eta$ as $\lambda \searrow 0$. Then by remark 3.32 we have

$$\varphi(x) \le \liminf_{\lambda\searrow 0} \varphi(\mathcal{J}_\lambda x) \le \liminf_{\lambda\searrow 0} \varphi_\lambda(x) \le \limsup_{\lambda\searrow 0} \varphi_\lambda(x) \le \varphi(x).$$

This shows that $\varphi_\lambda(\eta) \nearrow \varphi(\eta)$ when $\lambda \searrow 0$. If $x \notin \overline{D(A)}$, then $x - \mathcal{J}_\lambda x \to x - \operatorname{Pr}_{D(A)} x \neq 0$. As φ is lower semi continuous, there exists a neighborhood of 0 and some constant $C \in \mathbb{R}$ such that for all $\lambda > 0$ sufficiently small we have $\varphi(\mathcal{J}_\lambda x) \ge C$. In conjunction we observe a divergence

$$\liminf_{\lambda\to 0} \varphi_\lambda(x) = \liminf_{\lambda\to 0} \frac{\lambda}{2}\|A_\lambda x\|^2_{\mathcal{H}} + \varphi(\mathcal{J}_\lambda x) \ge \liminf_{\lambda\to 0}\ \frac{1}{2}\|A_\lambda x\|_{\mathcal{H}}\|x - \mathcal{J}_\lambda x\|_{\mathcal{H}} + C = +\infty.$$

The last property (v) is an immediate consequence, since we have just deduced that $D(\varphi) \subseteq \overline{D(A)}$, and thus $\overline{D(\varphi)} = \overline{D(A)}$. □

4 Gradient Flow

In this section we will introduce the Gradient Flow problem. We begin the section by presenting some results of function spaces over a $\mathbb{R}$-Hilbert space, which we will need for the duration of this chapter. The Gradient Flow problem is divided into the homogeneous gradient flow problem and the inhomogeneous gradient flow problem. We end this section by extensively studying the inhomogenuous gradient flow problem, in the case the operator is the sub differential of a proper convex lower semicontinuous mapping on a $\mathbb{R}$-Hilbert space.

4.1 Preliminaries

In this section we introduce measurable functions in a Hilbert space, which we will need to discuss Gradient Flow. In particular we need to define the p-measures on a $\mathbb{R}$-Hilbert space, the notion of weak derivatives, absolute continuity and bounded variation.

Throughout this section that $I \subseteq \mathbb{R}$ be any interval of $\mathbb{R}$.

We will need some standard results of L^p spaces on a $\mathbb{R}$-Hilbert space. These will not be proven, as they are only tangentially related with the topics of this paper. The proofs can be found in the appendix of the textbook of Brézis [3, Appendix 1 and 2].

Definition 4.1. *We say a function $r : I \to \mathcal{H}$ is simple if it takes only finite many values and the preimages are elements of the Borel set of I. The integral of r is given by*

$$\int_I r \, dm(x) := \sum_{x \in \mathcal{H}} x \cdot m(r^{-1}(x)),$$

where m is the Lebesgue measure. In future we will replace the notation $dm(x)$ with dx.

Definition 4.2. *We call a function $f : I \to$ measurable, if there exists a sequence of simple functions $(f_n)_{n\in\mathbb{N}}$ that converges pointwise almost everywhere to f. Almost everywhere here means that the set of points P for which the condition does not apply is a null set, i.e. it has Lebesgue measure zero or $m(P) = 0$.*

Definition 4.3. *We denote by $L^0(I; \mathcal{H})$ the space of equivalence classes of measurable functions $I \to \mathcal{H}$. We say two measurable functions $f, g : I \to \mathcal{H}$ are equivalent if and only if the difference $f - g$ is equal to 0 almost everywhere. Whenever we refer to an element in $L^0(I; \mathcal{H})$, we will always work with any representative of an equivalence class.*

Definition 4.4. *For any $p \in [1, +\infty]$ let $L^p\big(I; \mathcal{H}\big)$ be the Banach space of all measurable functions $u \in L^0(I; \mathcal{H})$ that attain a finite value of the corresponding norms*

$$\|u\|_p := \left(\int_{\mathbb{R}} \|u\|_{\mathcal{H}}^p \, dt \right)^{\frac{1}{p}} \qquad \text{for } p \in [1, +\infty),$$
$$\|u\|_\infty := \text{ess sup } \|u\|_{\mathcal{H}} \qquad \text{for } p = +\infty.$$

Theorem 4.5 (Lebesgue Dominated Convergence)**.** *Let $(f_n)_{n\in\mathbb{N}} \subseteq L^0(I; \mathcal{H})$ be a sequence of measurable functions such that $f_n \to f$ almost everywhere on I for some mapping $f : I \to \mathbb{R}$. Let $g \in L^1(I; \mathbb{R})$ with $\|f_n\|_{\mathcal{H}} \leq g$ almost everywhere on I for all $n \in \mathbb{N}$. Then we have that $f \in L^0(I; \mathcal{H})$ and*

$$\lim_{n\to+\infty} \int_I \|f_n - f\|_{\mathcal{H}} \, dt = 0,$$

in particular

$$\lim_{n\to+\infty} \int_I f_n \, dt = \int_I f_n \, dt.$$

Theorem 4.6 (Fatou's Lemma)**.** *Let $(f_n)_{n\in\mathbb{N}} \subseteq L^0(I; \mathcal{H})$ be a sequence of measurable functions such that $f_n \rightharpoonup f$ almost everywhere on I for some mapping $f : I \to \mathbb{R}$. If the sequence is bounded in the L^1 norm, i.e.*

$$\sup_{n\in\mathbb{N}} \int_I \|f\|_{\mathcal{H}} \, dt < +\infty,$$

then we have that $f \in L^1(0, T; \mathcal{H})$ with an estimate

$$\int_I \|f\|_{\mathcal{H}} \, dt \leq \liminf_{n\to+\infty} \int_I \|f_n\|_{\mathcal{H}} \, dt.$$

Lemma 4.7. *Let $I = [a, b]$ and let $f \in L^1(a, b; \mathcal{H})$. Then for any $\varepsilon > 0$, there exists a partition $a = x_0 < x_1 < \ldots < x_n = b$ and values $c_1, \ldots, c_n \in \mathcal{H}$ such that for the simple function $g : [a, b] \to \mathcal{H}$, defined by $g = c_j$ on $[x_{j-1}, x_j]$, we have $\|f - g\|_{L^1(a,b;\mathcal{H})} < \varepsilon$.*

Definition 4.8. *Let $I = [a, b]$, then we say a function $f : I \to \mathcal{H}$ is of bounded variation on $[a, b]$, if there exists a constants $C > 0$, such that for every partition $a = x_0 < x_1 < \ldots < x_n = b$ we have*

$$\sum_{j=1}^{n} \|f(x_j) - f(x_{j-1})\|_{\mathcal{H}} \leq C.$$

We denote by $\mathrm{Var(f;\, [a, b])}$ *the infimum of admissible constants C. The set of all functions of bounded variation on $[a, b]$ is denoted by* $\mathrm{BV}(a, b)$. *We write* $V_f(t) := \mathrm{Var(f;\, [a, t])}$ *for all $t \in [a, b]$.*

Definition 4.9. *We say a function $f : I \to \mathcal{H}$ is absolutely continuous if and only if for every $\varepsilon > 0$ there exists $\delta > 0$ such that for any countable set U of open bounded disjoint intervals of I, whose closures are subsets of I, we have that*

$$\sum_{J \in U} m(J) < \delta \qquad \Rightarrow \qquad \sum_{J \in U} \|f(\sup J) - f(\inf J)\|_{\mathcal{H}} < \varepsilon.$$

Remark 4.10. • *One quickly sees from the definition that an absolutely continuous function on a compact interval $[a, b]$ is of bounded variation.*

• *It is known for $I = [a, b]$ and $f \in \mathrm{BV}(a, b)$, then V_f is differentiable almost everywhere on $[a, b]$.*

Definition 4.11. *For $1 \leq p \leq +\infty$, we define the p-Sobolev spaces (of order 1) as the set $W^{1,p}(I;\, \mathcal{H})$ such that $f \in W^{1,p}(I\ \mathcal{H})$ if and only if $f \in L^p\big(I;\, \mathcal{H}\big)$ and there exists $g \in L^p\big(I;\, \mathcal{H}\big)$ such that for some $t_0 \in I$ we have for almost every $t \in I$ the identity*

$$f(t) = \int_{t_0}^{t} g(x)\, dx.$$

Definition 4.12. *For $1 \leq p \leq +\infty$ and $I = [a, b]$, we define the space $\widetilde{W}^{1,p}(a, b; \mathcal{H})$ as every absolutely continuous function $f : [a, b] \to \mathcal{H}$ such that $\partial_t V_f$ is in $L^p(a, b;\, \mathcal{H})$.*

Definition 4.13. *The $\mathbb{R}$-linear space of test functions in $\mathcal{H}$ over I is defined by*

$$\mathscr{D}(I;\, \mathcal{H}) := \{f : I \to \mathcal{H} \mid \operatorname{supp} f \textit{ is compact and } f \textit{ is smooth}\}.$$

Here smoothness is interpreted with the Fréchet derivative.

Motivated by integration by parts, we can define the derivative of a measurable function weakly.

Definition 4.14. *We call a function $f \in L^0(I;\, \mathcal{H})$ weakly differentiable with a weak derivative $\partial_t f \in L^0(I;\, \mathcal{H})$ if and only if*

$$\langle \partial_t f,\, \phi \rangle = -\, \langle f,\, \partial_t \phi \rangle \qquad \forall \phi \in \mathscr{D}(I;\, \mathcal{H}).$$

Lemma 4.15. *If $f \in L^0(I;\, \mathcal{H})$ is weakly differentiable, then its weak derivative is unique.*

Proof. It suffices to verify that if $\partial_t f = 0$ almost everywhere on I, then $f = 0$ almost everywhere on I. Indeed for all $\phi \in \mathscr{D}(I;\, \mathcal{H})$ we have that

$$0 = \langle \partial_t f,\, \phi \rangle_{L^2} = -\, \langle f,\, \partial_t \phi \rangle_{L^2}\,.$$

The upper equation can only hold if and only if $f = 0$. □

We will use the following lemma a couple of times during this paper.

Lemma 4.16. *Let $(f_n)_{n\in\mathbb{N}} \subseteq L^2(I;\, \mathcal{H})$ be a sequence of weakly differentiable functions, such that $f_n \to f$ in $L^2(I;\, \mathcal{H})$ for some $f \in L^2(I;\, \mathcal{H})$ and the sequence $(\|\partial_t f\|_{L^2(I;\mathcal{H})})_{n\in\mathbb{N}}$ is bounded. Then f is weakly differentiable and $\partial_t f_n \rightharpoonup \partial_t f$ in $L^2(I;\, \mathcal{H})$.*

Proof. As $(\|\partial_t f\|_{L^2(I;\mathcal{H})})_{n\in\mathbb{N}}$ is bounded we have up to passing to some subsequence that $\partial_t f_n \rightharpoonup g$ in $L^2(I;\, \mathcal{H})$ for some $g \in L^2(I;\, \mathcal{H})$. For any $\phi \in \mathscr{D}(I;\, rH)$ we compute

$$\langle g,\, \phi\rangle_{\mathcal{H}} = \lim_{n\to+\infty} \langle \partial_t f_n,\, \phi\rangle_{\mathcal{H}} = \lim_{n\to+\infty} \langle f_n,\, -\partial_t\phi\rangle_{\mathcal{H}} = \langle f,\, -\partial_t\phi\rangle_{\mathcal{H}}\,.$$

As the weak derivative is unique, it is not necessary to pass to some weakly convergent subsequence and instead we have that $\partial_t f_n \rightharpoonup \partial_t f$ in $L^2(I;\, \mathcal{H})$. ☐

If $I = [a, b]$, then the Sobolev space $W^{1,p}(a, b;\, \mathcal{H})$, the space $\widetilde{W}^{1,p}(a, b; \mathcal{H})$ and the weak derivative share the following relation.

Proposition 4.17. *Let $I = [a, b]$, let $1 \leq p \leq +\infty$ and let $f : I \to \mathcal{H}$ be a function. Then the following are equivalent.*

(i) *The map f is in $W^{1,p}(a, b;\, \mathcal{H})$.*

(ii) *The function f is in $\widetilde{W}^{1,p}(a, b;\, \mathcal{H})$ and f is differentiable almost everywhere on (a, b).*

(iii) *For every $x \in \mathcal{H}$ the mapping $[a, b] \to \mathbb{R};\, t \mapsto \langle x,\, f(t)\rangle_{\mathcal{H}}$ is absolutely continuous, the function f is weakly differentiable and the weak derivative $\partial_t f$ is in $L^p(a, b;\, \mathcal{H})$.*

4.2 Homogeneous Gradient Flow

This section will focus on the ordinary differential equation

$$\begin{cases} \partial_t u + Au \ni 0 & \text{on } I, \\ u = u_0 & \text{on } \{0\} \end{cases} \tag{hGSD}$$

with $I = (0, +\infty)$ or $I = (0, T]$ for some $T > 0$. We call this ODE the homogenous generalised steepest descent hGSD equation, for some maximal monotone multivalued operator A on $\mathcal{H}$ and initial value $u_0 \in \mathcal{H}$. We call the equation hGSD a homogenous steepest descent equation if we have an identity instead of an inclusion on I. Note that the derivative ∂_t in hGSD should be interpreted as a weak derivative. We say that a function $u : I \to \mathcal{H}$ is a solution to the hGSD equation with operator A and initial value u_0 on I if it satisfies the associated equations[4]. We prove that there always exists a solution to the hGSD, provided the initial value is in the domain of the operator. At the end of this section we will be able to generalize the solutions, by picking their initial value to be in the closure of the domain. We do this by introducing the semi-group of contractions generated by a maximally monotone multivalued operator.

For the remainder of this section, we set A to be a maximal monotone multivalued operator on $\mathcal{H}$.

We begin by proving the existence of a unique solution to the hGSD equation when the initial value is in the domain of the associated maximal monotone multivalued operator.

[4] Note we implicitly must have $u(t) \in D(A)$ for all $t \in I$.

Theorem 4.18. *For any $u_0 \in D(A)$ there exists a solution u in $C^0(0, +\infty; \mathcal{H})$ to the hGSD equation with operator A and initial value u_0 on $[0, +\infty)$. The solution satisfies the following properties*

(i) The weak derivative $\partial_t u$ is in $L^\infty\big(0, +\infty; \mathcal{H}\big)$ with upper bound $\|\partial_t u\|_\infty \leq \|A_0 u_0\|_{\mathcal{H}}$.

(ii) The right derivative $\partial_t^+ u$ exists and is monotonically decreasing in its norm, right continuous and satisfies the equation $\partial_t^+ u + A_0 u = 0$.

The proof of this theorem uses a theorem from Brézis [3, Theorem 1.6].

Theorem 4.19. *(Special Version) Let E be a Banach space and $C \subseteq E$ a convex closed subset with a 1-Lipschitz mapping $J : C \to E$. For any $\lambda > 0$ and $T > 0$ let u be a solution of the ordinary differential equation*

$$\partial_t u + \frac{1}{\lambda}(u - Ju) = 0 \qquad \text{on } [0, T].$$

Then the mapping $\|\partial_t u\|_{\mathcal{H}}$ is monotonically decreasing.

proof of Theorem 4.18. We will show the existence of a solution by approximating the hGSD with the Yosida approximation[5] A_λ of A for any $\lambda > 0$. Then the proposition 3.28 states that the map A_λ is λ^{-1}-Lipschitz. Thus by the Picard-Lindelöf Theorem (cf. [4, Theorem 7.3]) for any $T > 0$ the hGSD with operator A_λ and initial value u_0 on $[0, T]$ possesses a unique solution u_λ in $C^1(0, +\infty; \mathcal{H})$. From theorem 3.23 we know that $\overline{D(A)}$ is convex. Then we can use use proposition 3.28 and theorem 4.19 with $J = \mathcal{J}_\lambda$, $C = \overline{D(A)}$ and $E = \mathcal{H}$ to deduce the following stability property

$$\|\partial_t u_\lambda\|_{\mathcal{H}} = \|A_\lambda u_\lambda\|_{\mathcal{H}} \leq \|A_\lambda u_0\|_{\mathcal{H}} \leq \|A_0 u_0\|_{\mathcal{H}}. \tag{13}$$

We show that $(u_\lambda)_{\lambda > 0}$ is a Cauchy sequence in $C^0\big(0, T; \mathcal{H}\big)$. Indeed from the hGSD equation we get for all $\lambda, \mu > 0$

$$\partial_t u_\lambda + A_\lambda u_\lambda - \partial_t u_\mu - A_\mu u_\mu = 0. \tag{14}$$

Note we have an identity instead of an inclusion, since A_λ and A_μ are single valued. By taking the the inner product of (14) with $u_\lambda - u_\mu$ we compute an upper bound

$$\begin{aligned}
\partial_t \frac{1}{2} \|u_\lambda - u_\mu\|_{\mathcal{H}}^2 &= - \langle A_\lambda u_\lambda - A_\mu u_\mu,\, u_\lambda - u_\mu \rangle_{\mathcal{H}} \\
&= -\lambda \langle A_\lambda u_\lambda - A_\mu u_\mu,\, A_\lambda u_\lambda \rangle_{\mathcal{H}} + \mu \langle A_\lambda u_\lambda - A_\mu u_\mu,\, A_\mu u_\mu \rangle_{\mathcal{H}} \\
&\qquad - \langle A_\lambda u_\lambda - A_\mu u_\mu,\, \mathcal{J}_\lambda u_\lambda - \mathcal{J}_\mu u_\mu \rangle_{\mathcal{H}} \\
&\leq -\lambda \|A_\lambda u_\lambda\|_{\mathcal{H}}^2 - \mu \|A_\mu u_\mu\|_{\mathcal{H}}^2 + (\lambda + \mu) \langle A_\lambda u_\lambda,\, A_\mu u_\mu \rangle_{\mathcal{H}} \\
&\leq (\lambda + \mu) \langle A_\lambda u_\lambda,\, A_\mu u_\mu \rangle_{\mathcal{H}} \\
&\leq (\lambda + \mu) \|A_\lambda u_\lambda\|_{\mathcal{H}} \|A_\mu u_\mu\|_{\mathcal{H}} \\
&\leq (\lambda + \mu) \|A_0 u_0\|_{\mathcal{H}}^2 .
\end{aligned} \tag{15}$$

[5] see definition 3.24

Note we used remark 3.25 in the first inequality paired with the monotonicity of A, the Cauchy-Schwarz inequality in the penultimate inequality and we used (13) in the last inequality. By integrating (15) we get

$$\|u_\lambda(t) - u_\mu(t)\|_{\mathcal{H}} \leq \sqrt{2\,(\lambda+\mu)\,t}\,\|A_0 u_0\|_{\mathcal{H}} \qquad \forall t \in [0,T].$$

Then $(u_\lambda)_{\lambda>0}$ is a Cauchy-sequence in the Banach space $C^0\big(0,T;\,\mathcal{H}\big)$ with a limit $u \in C^0\big(0,T;\,\mathcal{H}\big)$ and convergence estimate

$$\|u_\lambda(t) - u(t)\|_{\mathcal{H}} \leq \sqrt{\lambda\,t}\,\|A_0 u_0\|_{\mathcal{H}} \qquad \forall t \in [0,T]. \tag{16}$$

To show that u is the desired solution, we must prove it's range is in $D(A)$, its weak derivative exists and satisfies the stated hGSD equation and u also fulfils properties (i) and (ii).

We begin by showing u has a range in $D(A)$. Then by equation (13) and (16) the sequence $(\mathcal{J}_\lambda\, u_\lambda)_{\lambda>0}$ converges uniformly towards u in $C^0\big(0,T;\,\mathcal{H}\big)$ since

$$\begin{aligned}\|u - \mathcal{J}_\lambda\, u_\lambda\|_{\mathcal{H}} &\leq \|u - u_\lambda\|_{\mathcal{H}} + \|u_\lambda - \mathcal{J}_\lambda\, u_\lambda\|_{\mathcal{H}} \\ &= \|u - u_\lambda\|_{\mathcal{H}} + \lambda\,\|A_\lambda u_\lambda\|_{\mathcal{H}} \qquad\qquad \text{on } [0,T]. \\ &\leq (\sqrt{\lambda\,t} + \lambda)\,\|A_0 u_0\|_{\mathcal{H}}\end{aligned} \tag{17}$$

If we focus on a fixed time $t \in [0,T]$, then we can deduce that $u(t) \in D(A)$. Indeed by equation (13) we have that $\|A_\lambda u_\lambda\|_{\mathcal{H}}$ is bounded from above by $\|A_0 u_0\|_{\mathcal{H}}$ as $\lambda \searrow 0$. Thus there exists a weakly convergent subsequence $A_{\lambda_n} u_{\lambda_n} \rightharpoonup y$ for some $y \in \mathcal{H}$. Since $\mathcal{J}_{\lambda_n} u_{\lambda_n} \to u$ by (17) and $A_{\lambda_n} u_{\lambda_n} \rightharpoonup y$ we have from lemma 3.27 that $(u, y) \in A$, moreover of particular interest to us is $u(t) \in D(A)$ and $\|y\|_{\mathcal{H}} \leq \|A_0 u_0\|_{\mathcal{H}}$. We will need at the very end of this proof a consequence of this result, namely that

$$\|A_0 u\|_{\mathcal{H}} \leq \|A_0 u_0\|_{\mathcal{H}} \qquad \text{on } [0,+\infty). \tag{18}$$

Next we show that u is weakly differentiable and satisfies both the hGSD equation and property (i). By (13) we know that the sequence $\|\partial_t u_\lambda\|_{\mathcal{H}}$ is bounded by $\|A_0 u_0\|_{\mathcal{H}}$ as $\lambda \searrow 0$. Thus there exists some weakly converging subsequence $\partial_t u_{\lambda_n} \rightharpoonup v$ in $L^\infty\big(0,T;\,\mathcal{H}\big)$ for some $v \in L^\infty\big(0,T;\,\mathcal{H}\big)$. Note that $L^\infty\big(0,T;\,\mathcal{H}\big) \subseteq L^2\big(0,T;\,\mathcal{H}\big)$. From lemma 4.16 we have that v is the weak derivative of u with $\partial_t u_\lambda \rightharpoonup v$ in $L^2\big(0,T;\,\mathcal{H}\big)$ whenever $\lambda \searrow 0$ and $\|v\|_\infty \leq \|A_0 u_0\|_{\mathcal{H}}$. This in turn proves property (i). If we use the operator $\mathcal{I}$ from example 3.16 we get that $(u_\lambda, -\partial_t u_\lambda) \in \mathcal{I}\,A$ for all $\lambda > 0$. We know that $u_\lambda \to u$ in $C^0\big(0,T;\,\mathcal{H}\big)$, and hence also $u_\lambda \to u$ in $L^2\big(0,T;\,\mathcal{H}\big)$. And we showed $\partial_t u_\lambda \rightharpoonup \partial_t u$ in $L^2\big(0,T;\,\mathcal{H}\big)$. Then by lemma 3.27 we have that $(u, -\partial_t u) \in \mathcal{I}\,A$ or equivalently $\partial_t u + Au \ni 0$.

It remains to justify the claim (ii). We begin by showing that the function $A_0 u$ is right continuous at the point 0. Then from equation (13) as $\|A_\lambda u_\lambda\|_{\mathcal{H}}$ bounded from above by $\|A_0 u_0\|_{\mathcal{H}}$ as $\lambda \to 0$, there exists a weakly converging subsequence $A_0 u(t_n) \rightharpoonup \xi$ for some $\xi \in \mathcal{H}$. As u is continuous we have the converging sequence $u(t_n) \to u(0)$ and $(u(t_n), A_0 u(t_n)) \in A$. So by lemma 3.27 we know $(u_0, \xi) \in A$ and $\|\xi\|_{\mathcal{H}} \leq \|A_0 x\|_{\mathcal{H}}$ and thus by definition of the minimum norm section $\xi = A_0 u_0$. As we have considered any convergent subsequence, we see that u is right continuous at 0 with limit $A_0 u(t) \to A_0 u_0$ as $t \searrow 0$.

We define the set $E := \{t \in (0,+\infty) \mid u \text{ is differentiable in } t \text{ and } \partial_t u(t) \in Au(t)\}$. As u is Lipschitz, it has bounded variation. From proposition 4.17 we know that the complement of E is a null-set. For all $t_0, h > 0$ we have that $\|u(t_0) - u(t_0 + h)\|_{\mathcal{H}} \leq h \|A_0 u(t_0)\|_{\mathcal{H}}$. Thus if $t_0 \in E$ we have that $\|\partial_t u(t_0)\|_{\mathcal{H}} \leq \|A_0 u(t_0)\|_{\mathcal{H}}$, but by definition of the minimum norm section we have that $\partial_t(t_0) + A_0 u(t_0) = 0$. If we integrate on $(0, t)$ we get that $u(t) - u(0) = \int_0^t A_0 u \, dt$. As $A_0 u$ is right continuous at 0, we quickly deduce from the integral that u is right differentiable at 0 with $\partial_t^+ u(0) + A_0 u_0 = 0$. Lastly note that $[0, +\infty) \to \mathcal{H}$ $t \mapsto u(t_0 + t)$ is a solution to the hGSD equation with operator A and initial value $u(t_0)$ on $[0, +\infty)$. Thus we have by equation (18) that $\left\|\partial_t^+ u(t_0 + t)\right\|_{\mathcal{H}} = \|A_0 u(t_0 + t)\|_{\mathcal{H}} \leq \|A_0 u(t_0)\|_{\mathcal{H}} = \left\|\partial_t^+ u(t_0)\right\|_{\mathcal{H}}$. This shows that $\partial_t^+ u$ is monotonically decreasing in the norm and thus ends the proof of the theorem. □

Corollary 4.20. *Let $v_0 \in D(A)$ be another initial value. Then for the solutions u and v to the hGSD equation with operator A and initial value u_0 and v_0 respectively on the interval $[0, +\infty)$, we have that the function $\|u - v\|_{\mathcal{H}}$ is monotonically decreasing. As a consequence the solutions to the hGSD equations are unique.*

Proof. We show that the function $\|u - v\|_{\mathcal{H}}$ is monotonically decreasing by computing the weak derivative and using the monotonicity of A

$$\partial_t \frac{1}{2} \|u - v\|_{\mathcal{H}}^2 = \langle \partial_t u - \partial_t v, \, u - v \rangle_{\mathcal{H}} = - \langle Au - Av, \, u - v \rangle_{\mathcal{H}} \leq 0.$$

In case $u_0 = v_0$, we must have that $u = v$ on all of $[0, +\infty)$, which shows uniqueness. □

We will show a bit later, that the initial values of the solutions to the hGSD equation can in fact be chosen in the closure of the domain of the maximally monotone operator. But first we show some examples of solutions to the hGSD equation.

Example 4.21. *Define the multivalued operator A on $\mathbb{R}$ as in example 3.15 with the function $f : \mathbb{R} \to \mathbb{R}$ defined as the heavyside function*

$$f(t) := \begin{cases} 1 & \text{if } t \geq 0, \\ 0 & \text{if } t < 0 \end{cases} \qquad \forall t \in \mathbb{R}.$$

For $u_0 \in \mathbb{R}$ the unique solution of the hGSD equation with operator A and initial value u_0 on $[0, +\infty)$ is given by

- *The constant mapping $u = u_0$, provided if $u_0 \leq 0$.*
- *The affine map $u = u_0 - t$ on $[0, u_0]$ and constant map $u = 0$ on (u_0, ∞), provided if $u_0 > 0$.*

Example 4.22. *Let $\mathcal{H}$ be a finite dimensional $\mathbb{R}$-Hilbert space and A a positive semi-definite linear operator on $\mathcal{H}$. By example 3.17 the operator A is maximally monotone. Then the unique solution to hGSD equation with operator A and initial value u_0 on $[0, +\infty)$ is given by $u = \exp(-At)u_0$. As the eigenvalues of A are real and non-negative, the entries of the solution u with respect to any finite basis, converge exponentially to some constant value. In case A is positive definite, the entries converge to 0.*

Example 4.23. *Use the same $\mathbb{R}$-Hilbert space $\mathcal{H} = W^{1,2}(0, \mathcal{L}; \mathcal{K})$ and maximally monotone operator $A = \partial_x$ as in example 3.18. Set any value $u_0 \in \mathcal{K}$ as the initial condition in the domain $D(A)$ and let $g \in D(A)$ be arbitrary. We can represent the solution u to the hGSD equation* *with maximally monotone mulitvalued operator A and initial value u_0 on $[0,+\infty)$ as a function $u \in W^{1,2}(\Omega;\mathcal{K})$ with $\Omega := [0,+\infty) \times [0,\mathcal{L}]$. Then the hGSD equation reads as the PDE*

$$\begin{cases} \partial_t u + \partial_x u = 0 & \text{on } (0,+\infty) \times [0,\mathcal{L}], \\ u = g & \text{on } \{0\} \times [0,\mathcal{L}]. \end{cases}$$

This is however just the transport equation and has solution

$$u(t,x) = \begin{cases} g(x-t) & \text{if } x \leq t, \\ u_0 & \text{otherwise} \end{cases} \qquad \forall (t,x) \in \Omega.$$

The solutions of the *hGSD* equation can be described by flows of a multivalued stable vector field on the domain of a maximally monotone operator. In fact as the solutions are unique, we may assume that the vector field is single valued. This motivates the definition of a family of solution operators, which shares a semi-group structure (or more accurately an abelian monoid structure).

Definition 4.24. *For all $t \geq 0$ define the mapping $\tilde{S}(t) : D(A) \to \overline{D(A)}$ that sends $u_0 \in D(A)$ to $u(t)$, where u is the solution to the hGSD equation with operator A and initial value u_0 on $[0,+\infty)$. Let $S(t) : \overline{D(A)} \to \overline{D(A)}$ be the continuous extension of the map $\tilde{S}(t)$ to the closure of the domain of A. The family of maps $(S(t))_{t\geq 0}$ is called the semi-group of contractions generated by A.*

Corollary 4.25. *The semi-group $(S_t)_{t\geq 0}$ of contractions generated by A satisfies*

(i) At 0 we have $S(0) = \mathrm{id}_{\overline{D(A)}}$.

(ii) For all $t_1, t_2 \geq 0$ we have the homomorphic relation $S(t_1 + t_2) = S(t_1)S(t_2)$.

(iii) The family $(S(t))_{t\geq 0}$ is right continuous at 0, i.e. we observe $\|S(t)u_0 - u_0\|_{\mathcal{H}} \to 0$ whenever $t \to 0$ for all $u_0 \in \overline{D(A)}$.

(iv) The family of maps are contractions, that is to say $\|S(t)u_0 - S(t)v_0\|_{\mathcal{H}} \leq \|u_0 - v_0\|_{\mathcal{H}}$ for all $u_0, v_0 \in \overline{D(A)}$ and $t \geq 0$.

Proof. Claims (i) and (ii) follow directly from the definition. The property (iii) is a consequence of the right differentiability of the solutions as stated in theorem 4.18. The attribute (iv) results directly from corollary 4.20. □

Corollary 4.26. *Let $u_0 \in \overline{D(A)}$ be any initial value. Let $(S(t))_{t\geq 0}$ be the semi-group of contractions generated by A. Define $u(t) := S(t)u_0$ for all $t \geq 0$. Let u_λ be the solution to the hGSD equation with operator A_λ and initial value u_0 on $[0,+\infty)$ for any $\lambda > 0$. Then for all $t > 0$ we have that $u_\lambda(t) \to u(t)$ whenever $\lambda \searrow 0$.*

Proof. Note that the claim is immediately true, if $u_0 \in D(A)$ as u is the pointwise limit of u_λ when $\lambda \searrow 0$, which we have shown in the proof of theorem 4.18. For the general case pick any value $v_0 \in D(A)$ and let v and v_λ be the solutions to the hGSD with operators A and A_λ and respectively and initial value v_0. Then we have by corollary 4.20 that

$$\begin{aligned} \lim_{\lambda\searrow 0} \|u_\lambda(t) - u(t)\|_{\mathcal{H}} &\leq \lim_{\lambda\searrow 0} \|u_\lambda(t) - v_\lambda(t)\|_{\mathcal{H}} + \|v_\lambda(t) - v(t)\|_{\mathcal{H}} + \|v(t) - u(t)\|_{\mathcal{H}} \\ &\leq \lim_{\lambda\searrow 0} 2\|u_0 - v_0\|_{\mathcal{H}} + \|v_\lambda(t) - v(t)\|_{\mathcal{H}} \\ &\leq 2\|u_0 - v_0\|_{\mathcal{H}}. \end{aligned}$$

The claim immediately follows as v_0 can be chosen arbitrarily close to u_0. □

4.3 Inhomogeneous Gradient Flow

This section will focus on the inhomogeneous generalisation of the hGSD equation. Namely

$$\begin{cases} \partial_t u + Au \ni f & \text{on } (0,T), \\ u = u_0 & \text{on } \{0\} \end{cases} \tag{GSD}$$

which we call the inhomogenous generalised steepest descent GSD equation, for some maximal monotone multivalued operator A on $\mathcal{H}$, inhomogeneous term $f \in L^1\big(0,T;\,\mathcal{H}\big)$ and initial value $u_0 \in \mathcal{H}$. Unlike the hGSD equation, we cannot directly prove the existence of a solution to the GSD equation. Instead we must differentiate between a strong and weak solution.

For the remainder of this section, let A be a maximal monotone multivalued operator on $\mathcal{H}$ and let $f \in L^1\big(0,T;\,\mathcal{H}\big)$ be any inhomogeneous term.

Definition 4.27. *We call a function $u \in C^0\big(0,T;\,\mathcal{H}\big)$ a strong solution to the GSD equation with maximally monotone operator A, inhomogeneous term f and initial value u_0, if u solves the stated ordinary differential equation, is absolutely continuous on all compact subsets of $(0,T)$ and the range of u on $(0,T)$ is contained in $D(A)$.*

Definition 4.28. *We call a function $u \in C^0\big(0,T;\,\mathcal{H}\big)$ a weak solution to the GSD equation with maximally monotone operator A, inhomogeneous term f and initial value u_0, if there exists two sequences $(u_n)_{n\in\mathbb{N}} \subseteq C^0\big(0,T;\,\mathcal{H}\big)$ and $(f_n)_{n\in\mathbb{N}} \subseteq L^1\big(0,T;\,\mathcal{H}\big)$, such that $u_n \to u$ in $C^0\big(0,T;\,\mathcal{H}\big)$, $f_n \to f$ in $L^1\big(0,T;\,\mathcal{H}\big)$ and u_n is a strong solution to GSD with operator A, inhomogeneous function f_n and initial value $u_n(0)$ for all $n \geq 1$.*

We can show an example of a strong solution to the GSD equation.

Example 4.29. *Let A be a positive sem-definite linear operator with dense range in $\mathcal{H}$ as in example 3.17. As $f \in L^1\big(0,T;\,\mathcal{H}\big)$ we deduce via variation of constants that a general solution to the GSD equation is given by*

$$u(t) = \exp(At)\left(u_0 + \int_0^t \exp(-As) f(s)\, ds\right) \qquad \forall t \in [0,T).$$

The following lemma is fundamental for the study of GSD equations.

Lemma 4.30. *Let $g, r \in L^1\big(0,T;\,\mathcal{H}\big)$ and let u, v be strong solutions to the GSD equation with operator A, inhomogeneous functions g, r and initial values $u(0), v(0)$ respectively. Then for any $0 \leq s \leq t \leq T$ and $(\eta, \xi) \in A$ the following inequalities hold*

$$\|u(t) - v(t)\|_{\mathcal{H}} \leq \|u(s) - v(s)\|_{\mathcal{H}} + \int_s^t \|g - r\|_{\mathcal{H}}\, dt, \tag{19}$$

$$\langle u(t) - u(s),\, u(s) - \eta\rangle_{\mathcal{H}} \leq \frac{1}{2}\|u(t) - \eta\|_{\mathcal{H}}^2 - \frac{1}{2}\|u(s) - \eta\|_{\mathcal{H}}^2 \leq \int_s^t \langle r - \xi,\, u - \eta\rangle_{\mathcal{H}}\, dt. \tag{20}$$

The proof of this lemma depends on the well-known lemma from Gronwall with many applications in ordinary differential equations. A proof of which can be found in Brézis [3, Lemma A.5].

Lemma 4.31. *(Gronwall) Let $m \in L^1\big(0,T;\,\mathbb{R}\big)$ be a non-negative function and $a \geq 0$. If the continuous function $u \in C^0\big(0,T;\,\mathbb{R}\big)$ satisfies*

$$\frac{1}{2}u(t)^2 \leq \frac{1}{2}a^2 + \int_0^t m(s)u(s)\,ds \qquad \forall t \in [0,T],$$

then we have the inequality

$$|u(t)| \leq a + \int_0^t m(s)\,ds \qquad \forall t \in [0,T].$$

proof of lemma 4.30. As A is maximally monotone and by the GSD equation, we compute

$$\partial_t \frac{1}{2}\left\|u-v\right\|_{\mathcal{H}}^2 = \langle \partial_t u - \partial_t v,\, u-v\rangle_{\mathcal{H}} = \langle g-r,\, u-v\rangle_{\mathcal{H}} - \langle Au-Av,\, u-v\rangle_{\mathcal{H}} \leq \langle g-r,\, u-v\rangle_{\mathcal{H}}.$$

By taking the integral on (s,t) and applying the Cauchy-Schwarz inequality we get

$$\begin{aligned}\frac{1}{2}\left\|u(t)-v(t)\right\|_{\mathcal{H}}^2 - \frac{1}{2}\left\|u(s)-v(s)\right\|_{\mathcal{H}}^2 &\leq \int_s^t \langle g-r,\, u-v\rangle_{\mathcal{H}}\,dx \\ &\leq \int_s^t \left\|g-r\right\|_{\mathcal{H}}\left\|u-v\right\|_{\mathcal{H}}\,dx.\end{aligned} \tag{21}$$

Both estimates follow directly from (21). Namely the bound (19) is a result of Gronwall's lemma 4.31, and the estimate (20) is given by the constant solution $v = \eta$ and $r = \xi$ for any $(\eta,\xi) \in A$. □

The inequalities in lemma 4.30 remain stable under convergent sequences in $C^0\big(0,T;\,\mathcal{H}\big)\times L^1\big(0,T;\,\mathcal{H}\big)$. This stability attribute justifies the following result.

Corollary 4.32. *The bounds (19) and (20) in Lemma 4.30 hold for weak solutions of the GSD equation.*

We proceed by proving that the GSD equation will always have a weak solution, provided the initial value u_0 is in the closure of the domain of A.

Theorem 4.33. *There exists a unique weak solution for the GSD equation with operator A, inhomogeneous function f and any initial value $u_0 \in \overline{D(A)}$.*

Proof. Note that for all $\theta \in \mathcal{H}$ the multivalued operator defined by $A-\theta$ remains maximally monotone. Then let $(S^{(\theta)}(t))_{t\geq 0}$ be the semi-group of contractions generated by $A-\theta$.
By lemma 4.7 there exists a sequence $(f_n)_{n\in\mathbb{N}} \subseteq L^1\big(0,T;\,\mathcal{H}\big)$ of step functions with $f_n \to f$ in $L^1\big(0,T;\,\mathcal{H}\big)$. For $n \geq 1$ let f_n be defined by a partition $0 = t_0 < \ldots < t_m = T$ and values $c_1,\ldots,c_m \in \mathcal{H}$, such that $f_n = c_k$ on almost all of $[t_{k-1},t_k]$ for $1 \leq k \leq m$. Then the strong solution u_n to the GSD equation with operator A, inhomogeneous function f_n and initial value u_0 is defined inductively by

$$u_n := S^{(c_1)}(t)u_0 \;\text{ on }\; [0,t_1] \quad\text{and}\quad u_n := S^{(c_k)}(t-t_{k-1})u(x_{k-1}) \;\text{ on }\; [t_{k-1},t_k] \qquad \forall 1 \leq k \leq m.$$

In addition by property (i) in theorem 4.18 we have that u_n is Lipschitz and thus absolutely continuous. By the estimate (19) from lemma 4.30, the sequence $(u_n)_{n\in\mathbb{N}}$ forms a Cauchy sequence in $C^0\big(0,T;\,\mathcal{H}\big)$, as we have the uniform estimate

$$\left\|u_n-u_m\right\|_{\mathcal{H}} \leq \int_0^T \left\|f_n-f_m\right\|_{\mathcal{H}}\,dt.$$

Then $u_n \to u$ in $C^0\big(0,T;\,\mathcal{H}\big)$ for some $u \in C^0\big(0,T;\,\mathcal{H}\big)$, namely the desired weak solution. □

As in theorem 4.18, we can determine the right differentiability and Lipschitz continuity of the weak solutions of the GSD equation. We will do this in theorem 4.36. For the statement we will require the following definition and lemma

Definition 4.34. *A point $t_0 \in [0, T]$ is called a right Lebesgue point of $g \in L^1\big(0, T;\, \mathcal{H}\big)$, if we have*

$$\limsup_{t \searrow t_0} \frac{1}{t - t_0} \int_{t_0}^{t} \|g - g(t_0)\|_{\mathcal{H}} \, dt = 0.$$

In that case we set

$$M_r g(t_0) := \lim_{t \searrow t_0} \frac{1}{t - t_0} \int_{t_0}^{t} g \, dt.$$

Lemma 4.35. *Let u be a weak solution of the GSD equation, let $C > 0$ be a constant and let $t_0 \in [0, T]$ be some time and $\alpha, \beta \in \mathcal{H}$ be some values. If a sequence $(t_n)_{n \in \mathbb{N}} \subseteq [0, T] \setminus \{t_0\}$ that converges to t_0 satisfies*

$$\frac{u(t_n) - u(t_0)}{t_n - t_0} \rightharpoonup \alpha \quad \text{and} \quad \frac{1}{t_n - t_0} \int_{t_0}^{t_n} f \, dt \rightharpoonup \beta \quad \text{and} \quad \frac{1}{|t_n - t_0 *|} \int_{t_0}^{t_n} \|f\|_{\mathcal{H}} \, dt < C$$

when $n \to +\infty$, then $u(t_0) \in D(A)$ and $\beta - \alpha \in Au(t_0)$.

Proof. We use the bound (20) from lemma 4.30. Then for all $(\eta, \xi) \in A$ and all $0 \le s \le t \le T$ we have

$$\langle u(t) - u(s),\, u(s) - \eta \rangle_{\mathcal{H}} \le \int_0^T \langle f - \xi,\, u - \eta \rangle_{\mathcal{H}} \, dt.$$

Depending on the order we replace the terms s, t with t_n, t_0 and divide with $\pm(t_0 - t_n)$. We differentiate between two cases, namely if $t_0 < t_n$ then

$$\left\langle \frac{u(t_n) - u(t_0)}{t_n - t_0},\, u(t_0) - \eta \right\rangle_{\mathcal{H}} \le \frac{1}{t_n - t_0} \int_{t_0}^{t_n} \langle f - \xi,\, u - \eta \rangle_{\mathcal{H}} \, dt$$

and if $t_n < t_0$ then

$$\left\langle \frac{u(t_0) - u(t_n)}{t_0 - t_n},\, u(t_n) - \eta \right\rangle_{\mathcal{H}} \le \frac{1}{t_0 - t_n} \int_{t_n}^{t_0} \langle f - \xi,\, u - \eta \rangle_{\mathcal{H}} \, dt.$$

As u is continuous, in either case if we take the limit $n \to +\infty$ we see that

$$\langle \alpha,\, u(t_0) - \eta \rangle_{\mathcal{H}} \le \langle \beta - \xi,\, u(t_0) - \eta \rangle_{\mathcal{H}} \qquad \Rightarrow \qquad 0 \le \langle \beta - \alpha - \xi,\, u(t_0) - \eta \rangle_{\mathcal{H}}.$$

As $(\eta, \xi) \in A$ is arbitrary, by condition (2) we must have $(u(t_0),\, \beta - \alpha) \in A$, proving the claim. □

We are now ready to state criteria for when the weak solution of the GSD is right differentiable.

Theorem 4.36. *Let u be a weak solution of the GSD equation. Suppose that $t_0 \in [0, T)$ is a right Lebesgue point of f. Then the following are equivalent*

(i) *The point $u(t_0)$ is contained in $D(A)$.*

(ii) We have $\liminf_{t\searrow t_0} \frac{1}{t-t_0} \|u(t)-u(t_0)\|_{\mathcal{H}} < +\infty$.

(iii) The function u *is right differentiable at* t_0.

In this case we also have a formula for the right derivative

$$\partial_t^+ u(t_0) = M_r f(t_0) - \operatorname{Pr}_{Au(t_0)} M_r f(t_0).$$

Proof. "$(i) \Rightarrow (iii)$": As t_0 is a right Lebesgue point of f, there exists $C \geq 0$ such that for $n \geq 1$ we have

$$\limsup_{h\searrow 0} \frac{1}{h} \int_{t_0}^{t_0+h} f\, dt = M_r f(t_0), \qquad \limsup_{h\searrow 0} \frac{1}{|h|} \int_{t_0}^{t_0+h} |f|\, dt < C.$$

The constant function $v = u(t_0)$ is a solution to the GSD equation with operator A, the constant inhomogeneous term $g = \operatorname{Pr}_{Au(t_0)} M_r f(t_0)$ and initial value $u(t_0)$. If we use equation (19), then we get an estimate for all $h > 0$ by

$$\|u(t_0+h) - u(t_0)\|_{\mathcal{H}} \leq \int_{t_0}^{t_0+h} \left\|f(s) - \operatorname{Pr}_{Au(t_0)} M_r f(t_0)\right\|_{\mathcal{H}} ds.$$

If we divide by h and take limit we see by the triangle inequality and the definition of M_r that

$$\begin{aligned}
&\limsup_{h\searrow 0} \frac{\|u(t_0+h)-u(t_0)\|_{\mathcal{H}}}{h} \\
&\quad \leq \limsup_{h\searrow 0} \frac{1}{h} \int_{t_0}^{t_0+h} \left\|f(s) - \operatorname{Pr}_{Au(t_0)} M_r f(t_0)\right\|_{\mathcal{H}} ds \\
&\quad = \limsup_{h\searrow 0} \frac{1}{h} \int_{t_0}^{t_0+h} \left\|f(s) - M_r f(t_0) + M_r f(t_0) - \operatorname{Pr}_{Au(t_0)} M_r f(t_0)\right\|_{\mathcal{H}} ds \\
&\quad \leq \left\|M_r f(t_0) - \operatorname{Pr}_{Au(t_0)} M_r f(t_0)\right\|_{\mathcal{H}}.
\end{aligned}$$

There hence exists a null-sequence $h_n \searrow 0$ such that $h_n^{-1}(u(t_0+h_n) - u(t_0))$ weakly converges to some $\alpha \in \mathcal{H}$ with

$$\|\alpha\|_{\mathcal{H}} \leq \left\|M_r f(t_0) - \operatorname{Pr}_{Au(t_0)} M_r f(t_0)\right\|_{\mathcal{H}}. \tag{22}$$

The conditions of lemma 4.35 hold and thus $M_r f(t_0) - \alpha \in Au(t_0)$. But by the inequality in (22) we must have that $\alpha = M_r f(t_0) - \operatorname{Pr}_{Au(t_0)} M_r f(t_0)$ and thus u is right differentiable at t_0.

"$(iii) \Rightarrow (ii)$": Immediate.

"$(ii) \Rightarrow (i)$": The proof is very similar to the proof of "(i)" $\Rightarrow$ "(iii)". ☐

Proposition 4.37. *Let* $g \in \mathrm{BV}(0,T;\mathcal{H})$ *and* u *a weak solution to the GSD equation with operator* A*, inhomogeneous term* g *and initial value* u_0*. Then the following properties are equivalent*

(i) The initial value u_0 *is in the domain of* A*, i.e.* $u_0 \in D(A)$.

(ii) The solution u *is Lipschitz continuous.*

In that case the range of u is contained in the domain of A, the map u is right differentiable in $[0,T)$ with right derivative

$$\partial_t^+ u(t_0) = M_r g(t_0) - \operatorname{Pr}_{Au(t_0)} M_r g(t_0) \qquad \forall t_0 \in [0,T). \tag{23}$$

If $g \in W^{1,1}(0,T;\,\mathcal{H})$ then for any $0 \le s \le t < T$ the following inequality holds true

$$\left\|\partial_t^+ u(t)\right\|_{\mathcal{H}} \le \left\|\partial_t^+ u(s)\right\|_{\mathcal{H}} + \int_s^t \|\partial_t g\|_{\mathcal{H}}\, ds. \tag{24}$$

Proof. "$(i) \Rightarrow (ii)$": If $u_0 \in D(A)$ we know from theorem 4.36 that u is right differentiable in 0. If we define $r(t) := g(t+h)$ and $v(t) := u = (t+h)$ for all $t \in [0, T-h]$, we have by equation (19) from lemma 4.30 that

$$\|u(t+h) - u(t)\|_{\mathcal{H}} \le \|u(h) - u_0\|_{\mathcal{H}} + \int_0^t \|g(s+h) - g(s)\|_{\mathcal{H}}\, ds \qquad \forall t \in (0, T-h). \tag{25}$$

Since u is right differentiable at 0 and as g is of bounded variation on $[0,T]$, there exists a constant $C > 0$ such that for h sufficiently small, we have that

$$\|u(t+h) - u(t)\|_{\mathcal{H}} \le Ch.$$

"$(ii) \Rightarrow (i)$": Since g is of bounded variation on $[0,T]$ we know that $g \in L^1\big(0,T;\,\mathcal{H}\big)$ and almost every point in $[0,T)$ is a right Lebesgue point of g. As u is Lipschitz continuous, the condition (ii) of theorem 4.36 is satisfied at every point of $[0,T)$ and thus the range of u is in the domain of A on $[0,T)$ and u is right differentiable on $[0,T)$ with right derivative

$$\partial_t^+(t_0)u = M_r g(t_0) - \operatorname{Pr}_{Au(t_0)} M_r g(t_0) \qquad \forall t_0 \in [0,T).$$

Since u is Lipschitz continuous and g is of bounded variation on $[0,T]$, we have have that both $\left\|\partial_t^+ u\right\|_{\mathcal{H}}$ and $g(t)$ are bounded from above when $t \nearrow T$. But then by definition of the right derivative of u in (23) we see that $M_r g(t) - \partial_t^+ u(t) = \operatorname{Pr}_{Au(t)} M_r g(t) \in Au(t)$ is also bounded in norm when $t \nearrow T$. There hence exists a convergent subsequence $M_r g(t_n) - \partial_t^+ u(t_n) \rightharpoonup y$ for some $y \in \mathcal{H}$. Since $u(t_n) \to u(T)$ and the range of u is contained in the domain of A on $[0,T)$, we see by lemma 3.25 that $u(T) \in D(A)$.

If $g \in W^{1,1}(0,T;\,\mathcal{H})$, then if we divide equation (25) by h and take the limit $h \searrow 0$ we verify the stated equation (24), that is for all $0 \le s \le t < T$ we have

$$\left\|\partial_t^+ u(t)\right\|_{\mathcal{H}} \le \left\|\partial_t^+ u(s)\right\|_{\mathcal{H}} + \int_s^t \|\partial_t g\|_{\mathcal{H}}\, ds.$$

□

4.4 Gradient Flow for Proper Convex Lower Semicontinous Potentials in Hilbert Spaces

In this section we study both the *hGSD* equation and the *GSD* equation in the case that the maximal monotone multivalued operator A in $\mathcal{H}$ is the sub differential of a proper convex lower semicontinuous function on $\mathcal{H}$. First we deduce some properties and estimates of the hGSD equation. Afterwards we present a condition for a weak solution of the GSD equation to be a strong solution and explore the regularity of the solutions based on this condition. Finally we end this section by studying the asymptotic property of strong solutions $u(t)$ of the GSD equation when $t \nearrow +\infty$ under certain criteria.

Throughout this section let $\varphi : \mathcal{H} \to (-\infty, +\infty]$ be a proper convex l.s.c function on $\mathcal{H}$. Define the sub differential of φ as the maximal monotone multivalued operator $A := \partial\varphi$. Let $u_0 \in \overline{D(A)}$ be arbitrary and define $u(t) := S(t)u_0$ for all $t \geq 0$, where $(S(t))_{t\geq 0}$ is the semi-group of contractions generated by A and let $T > 0$ be any positive value, namely the length of our time interval.

The theorem 4.18 guarantees the existence of a solution to the hGSD equation. However we are able to exploit the properties of sub differential operators to gain many properties and estimates of this solution. These are featured in the following theorem.

Theorem 4.38. *The solution u induced by the semi group of contractions generated by A with initial value u_0 has the following properties*

(i) For all $t > 0$ we have the stability estimate

$$\|A_0 u(t)\|_{\mathcal{H}} \leq \|A_0 v\|_{\mathcal{H}} + \frac{1}{t}\|u_0 - v\|_{\mathcal{H}} \qquad \forall v \in D(A). \tag{26}$$

(ii) The function $\varphi(u) = \varphi \circ u$ is in $L^1_{\text{loc}}(0, +\infty; \mathcal{H})$. On any closed subsets of $(0, +\infty)$ the map $\varphi(u)$ is Lipschitz, convex and right differentiable with right derivative

$$\partial_t^+ \varphi(u) = -\left\|\partial_t^+ u\right\|^2_{\mathcal{H}}.$$

(iii) The weak derivative of u satisfies $\sqrt{t}\,\partial_t u \in L^2_{\text{loc}}(0, +\infty; \mathcal{H})$.

(iv) We have $u_0 \in D(\varphi)$ if and only if $\partial_t u \in L^2_{\text{loc}}(0, +\infty; \mathcal{H})$.

(v) If $u_0 \in D(\varphi)$ then we have $\varphi(u(t)) \nearrow \varphi(u_0)$ when $t \searrow 0$. In addition we have the following relations

$$\varphi(u_0) - \varphi(u(T)) = \int_0^T \|\partial_t\, u\|^2_{\mathcal{H}}\, dt, \tag{27}$$

$$\frac{\|u(T) - u_0\|_{\mathcal{H}}}{\sqrt{T}} \leq \sqrt{\varphi(u_0) - \varphi(u(T))}, \tag{28}$$

$$\left\|\partial_t^+\, u(T)\right\|_{\mathcal{H}} \leq \sqrt{\frac{\varphi(u_0) - \varphi(u(T))}{T}}. \tag{29}$$

Proof. We start of by proving the property (i). Along the way we develop inequalities and identities that we will use to prove the other properties. From proposition 3.40 we know the Yosida approximation A_λ is the sub differential of a Fréchet differentiable proper convex map φ_λ on $\mathcal{H}$. By definition of the sub differential we have

$$\varphi_\lambda(u) - \varphi_\lambda(v) \geq \langle A_\lambda v,\, u - v\rangle_{\mathcal{H}} \qquad \forall u, v \in \mathcal{H}. \tag{30}$$

The sum of an affine mapping and a proper convex l.s.c mapping is still a proper convex l.s.c function. Thus by (30) define for any fixed $v \in D(A)$ the non-negative proper convex l.s.c function ρ_v^λ on $\mathcal{H}$ by

$$\rho_\lambda^v(u) := \varphi_\lambda(u) - \varphi_\lambda(v) - \langle A_\lambda v,\, u - v\rangle_{\mathcal{H}} \qquad \forall u \in \mathcal{H}$$

that attains it's minimal value 0 at v, with the sub differential and by proposition 3.40 the Fréchet derivative $B_\lambda^v u := \partial\rho_\lambda^v(u) = A_\lambda u - A_\lambda v$. Let u_λ be the solution to the hGSD equation with operator A_λ and initial value u_0 on $[0, +\infty)$. Note that the hGSD equation with operator A_λ and initial value u_0 can be restated as

$$\partial_t u_\lambda + B_\lambda^v u_\lambda = -A_\lambda v. \tag{31}$$

We are now in a position to proof the estimate (26). By definition of the sub differential we have

$$\rho_\lambda^v(v) - \rho_\lambda^v(u_\lambda) \geq \langle B_\lambda^v u_\lambda,\, v - u_\lambda\rangle_{\mathcal{H}}\,.$$

As $\rho_\lambda^v(v) = 0$ and from (31) we get

$$\rho_\lambda^v(u_\lambda) \leq \langle \partial_t u_\lambda + A_\lambda v,\, v - u_\lambda\rangle_{\mathcal{H}}\,.$$

We take the integral and see that

$$\begin{aligned}\int_0^T \rho_\lambda^v(u_\lambda)\,dt &\leq \int_0^T \langle \partial_t u_\lambda,\, v - u_\lambda\rangle_{\mathcal{H}}\,dt + \int_0^T \langle A_\lambda v,\, v - u_\lambda\rangle_{\mathcal{H}}\,dt \\ &= \frac{1}{2}\left\|u_0 - v\right\|_{\mathcal{H}}^2 - \frac{1}{2}\left\|u_\lambda(T) - v\right\|_{\mathcal{H}}^2 - \int_0^T \langle A_\lambda v,\, u_\lambda - v\rangle_{\mathcal{H}}\,dt.\end{aligned} \tag{32}$$

On the other hand, if we take the inner product of (31) with $t\,\partial_t u_\lambda$ we deduce from the Fréchet derivative from proposition 3.40 that

$$\begin{aligned} t\left\|\partial_t u_\lambda\right\|_{\mathcal{H}}^2 + t\,\partial_t\big(\rho_\lambda^v(u_\lambda)\big) &= t\left\|\partial_t u_\lambda\right\|_{\mathcal{H}}^2 + t\,\langle B_\lambda^v u_\lambda,\, \partial_t u_\lambda\rangle_{\mathcal{H}} \\ &= -t\,\langle A_\lambda v,\, \partial_t u_\lambda\rangle_{\mathcal{H}} \\ &= -t\,\langle A_\lambda v,\, \partial_t(u_\lambda - v)\rangle_{\mathcal{H}}\,.\end{aligned}$$

Again we take the integral and get by using integration by parts

$$\begin{aligned} \int_0^T t\left\|\partial_t u_\lambda\right\|_{\mathcal{H}}^2\,dt + T\,\rho_\lambda^v(u_\lambda(T)) - \int_0^T \rho_\lambda^v(u_\lambda)\,dt \\ &= \int_0^T t\left\|\partial_t u_\lambda\right\|_{\mathcal{H}}^2 + t\,\partial_t \rho_\lambda^v(u_\lambda)\,dt \\ &= \int_0^T -t\,\langle A_\lambda v,\, \partial_t(u_\lambda - v)\rangle_{\mathcal{H}} \\ &= -T\,\langle A_\lambda v,\, u_\lambda(T) - v\rangle_{\mathcal{H}} + \int_0^T \langle A_\lambda v,\, u_\lambda - v\rangle_{\mathcal{H}}\,dt.\end{aligned}$$

If we use (32) and the fact that $T\rho_\lambda^v(u_\lambda(T))$ is non-negative

$$\begin{aligned}\int_0^T t\,\|\partial_t u_\lambda\|_{\mathcal{H}}^2\,dt &= -T\,\langle A_\lambda v,\, u_\lambda(T)-v\rangle_{\mathcal{H}} - T\,\rho_\lambda^v(u_\lambda(T))\\ &\qquad + \int_0^T \langle A_\lambda v,\, u_\lambda - v\rangle_{\mathcal{H}}\,dt + \int_0^T \rho_\lambda^v(u_\lambda)\,dt\\ &\le -T\,\langle A_\lambda v,\, u_\lambda(T)-v\rangle_{\mathcal{H}} - \frac{1}{2}\|u_\lambda(T)-v\|_{\mathcal{H}}^2 + \frac{1}{2}\|u_0-v\|_{\mathcal{H}}^2\\ &\le \frac{1}{2}T^2\|A_\lambda v\|_{\mathcal{H}}^2 + \frac{1}{2}\|u_0-v\|_{\mathcal{H}}^2,\end{aligned} \tag{33}$$

where in the last inequality we used

$$0 \le \frac{1}{2}\|TA_\lambda v + u_\lambda(T) - v\|_{\mathcal{H}}^2 = \frac{1}{2}T^2\|A_\lambda v\|_{\mathcal{H}} + T\,\langle A_\lambda v,\, u_\lambda(T)-v\rangle_{\mathcal{H}} + \frac{1}{2}\|u_\lambda(T)-v\|_{\mathcal{H}}^2.$$

We know from theorem 4.18 that $\|\partial_t u_\lambda\|_{\mathcal{H}}$ is a monotonically decreasing mapping. This can be used in (33) to show

$$\frac{1}{2}T^2\|\partial_t u_\lambda(T)\|_{\mathcal{H}}^2 \le \int_0^T t\,\|\partial_t u_\lambda\|_{\mathcal{H}}^2\,dt \le \frac{1}{2}T^2\|A_\lambda v\|_{\mathcal{H}}^2 + \frac{1}{2}\|u_0-v\|_{\mathcal{H}}^2.$$

By multiplying by $2/T^2$, taking the square root, applying the triangle inequality and using proposition 3.28 we get

$$\|A_\lambda u_\lambda(T)\|_{\mathcal{H}} = \|\partial_t u_\lambda(T)\|_{\mathcal{H}} \le \|A_\lambda v\|_{\mathcal{H}} + \frac{1}{T}\|u_0-v\|_{\mathcal{H}} \le \|A_0 v\|_{\mathcal{H}} + \frac{1}{T}\|u_0-v\|_{\mathcal{H}}. \tag{34}$$

The formula shows that $\|A_\lambda u_\lambda(T)\|_{\mathcal{H}}$ is bounded from above when $\lambda \searrow 0$. Then there is a weakly convergent subsequence $A_{\lambda_n}u_{\lambda_n}(T) \rightharpoonup y$ for some $y \in \mathcal{H}$. Note we have from corollary 4.26 that $u_\lambda(T) \to u(T)$ in $C^0\big(0,T;\,\mathcal{H}\big)$. In theorem 4.18 we proved in the equation (17) that $\mathcal{J}_\lambda u_\lambda(T) \to u(T)$ in $C^0\big(0,T;\,\mathcal{H}\big)$. Then we have a sequence $(\mathcal{J}_{\lambda_n}u_{\lambda_n}(T), A_{\lambda_n}u_{\lambda_n}(T)) \subseteq A$. So by lemma 3.27 we have that $(u(T),y) \in A$. Moreover this means that the bound (34) holds in the limit and hence we can show (26) by

$$\|A_0 u(T)\|_{\mathcal{H}} \le \|y\|_{\mathcal{H}} \le \|A_0 v\|_{\mathcal{H}} + \frac{1}{T}\|u_0-v\|_{\mathcal{H}}.$$

Next we tackle property (ii). We show that $\varphi\circ u$ is in $L^1_{\mathrm{loc}}(0,+\infty;\,\mathcal{H})$ by looking at equation (32). Indeed we can reformulate the inequality

$$\int_0^T \varphi_\lambda(u_\lambda)\,dt \le \frac{1}{2}\|u_0-v\|_{\mathcal{H}}^2 - \frac{1}{2}\|u_\lambda(T)-v\|_{\mathcal{H}}^2 + T\varphi_\lambda(v) \le \frac{1}{2}\|u_0-v\|_{\mathcal{H}}^2 + T\varphi_\lambda(v).$$

From proposition 3.40 and the lower semi continuity of φ we know that $\varphi_\lambda(u_\lambda) \to \varphi(u)$ and $\varphi_\lambda(v) \to \varphi(v)$ whenever $\lambda \to 0$. So from Fatous lemma we get that $L^1_{\mathrm{loc}}(0,+\infty;\,\mathcal{H})$. Next we verify that $\varphi(u)$ satisfies the other properties on $[s,+\infty)$ for any $s>0$. Take any $t \ge s$ and $h>0$ and note by definition of the sub differential and the characterisation of the right derivative in theorem 4.18 that

$$\varphi(u(t+h)) - \varphi(u(t)) \ge \langle A_0 u(t),\, u(t+h)-u(t)\rangle_{\mathcal{H}} = -\left\langle \partial_t^+ u(t),\, u(t+h)-u(t)\right\rangle_{\mathcal{H}}$$

and

$$\varphi(u(t)) - \varphi(u(t+h)) \geq \langle A_0 u(t+h),\, u(t) - u(t+h)\rangle_{\mathcal{H}} = -\left\langle \partial_t^+ u(t+h),\, u(t) - u(t+h)\right\rangle_{\mathcal{H}}.$$

From this we devise a bound

$$\begin{aligned}
&|\varphi(u(t+h)) - \varphi(u(t))| \\
&\quad \leq \max\left\{ \left|\left\langle \partial_t^+ u(t),\, u(t+h) - u(t)\right\rangle_{\mathcal{H}}\right|, \left|\left\langle \partial_t^+ u(t+h),\, u(t) - u(t+h)\right\rangle_{\mathcal{H}}\right| \right\} \\
&\quad \leq \max\left\{ \left\|\partial_t^+ u(t)\right\|_{\mathcal{H}}, \left\|\partial_t^+ u(t+h)\right\|_{\mathcal{H}} \right\} \|u(t+h) - u(t)\|_{\mathcal{H}},
\end{aligned} \tag{35}$$

as-well as the following inequality

$$-\left\langle \partial_t^+ u(t),\, u(t+h) - u(t)\right\rangle_{\mathcal{H}} \leq \varphi(u(t+h)) - \varphi(u(t)) \leq -\left\langle \partial_t^+ u(t+h),\, u(t+h) - u(t)\right\rangle_{\mathcal{H}}. \tag{36}$$

Since $\left\|\partial_t^+ u\right\|_{\mathcal{H}}$ is monotonically decreasing as stated in theorem 4.18 we see that

$$\left\|\partial_t^+ u(t+h)\right\|_{\mathcal{H}} \leq \left\|\partial_t^+ u(t)\right\|_{\mathcal{H}} \leq \left\|\partial_t^+ u(s)\right\|_{\mathcal{H}}$$

and thus

$$\|u(t+h) - u(t)\|_{\mathcal{H}} \leq h \left\|\partial_t^+ u(s)\right\|_{\mathcal{H}}.$$

We apply this to (35) in order to show that $\varphi(u)$ is Lipschitz

$$|\varphi(u(t+h)) - \varphi(u(t))| \leq h \left\|\partial_t^+ u(s)\right\|_{\mathcal{H}}^2.$$

On the other hand we can divide (36) by h and since $\partial_t^+ u$ is right continuous by theorem 4.18-, by taking the limit $h \searrow 0$ we observe that $\varphi(u)$ is right differentiable as

$$-\left\|\partial_t^+ u(t)\right\|_{\mathcal{H}}^2 \leq \partial_t^+ \varphi(u(t)) = \lim_{h \searrow 0} \frac{\varphi(u(t+h)) - \varphi(u(t))}{h} \leq -\left\|\partial_t^+ u(t)\right\|_{\mathcal{H}}^2.$$

Since $\left\|\partial_t^+ u\right\|_{\mathcal{H}}$ is monotonically decreasing, so too is $\varphi(u)$ monotonically decreasing and also convex.
The property (iii) is an immediate consequence of (33) by applying Fatou's lemma again.
Now we consider property (iv). First assume that $u_0 \in D(\varphi)$, then we have the identity

$$\varphi_\lambda(u_0) - \varphi_\lambda(u_\lambda(t)) = -\int_0^t \partial_t \varphi_\lambda(u_\lambda)\, ds = -\int_0^t \langle A_\lambda u_\lambda,\, \partial_t u_\lambda\rangle_{\mathcal{H}}\, ds = \int_0^t \|\partial_t u_\lambda\|_{\mathcal{H}}^2\, ds. \tag{37}$$

From the attributes of the function φ_λ as stated in proposition 3.40 (namely equation (10) and property (iv)) we see that

$$\varphi(\mathcal{J}_\lambda u_\lambda) + \int_0^t \|\partial_t u_\lambda\|_{\mathcal{H}}^2\, ds = \varphi(\mathcal{J}_\lambda u_\lambda) - \varphi_\lambda(u_\lambda(t)) + \varphi_\lambda(u_0) \leq \varphi_\lambda(u_0) \leq \varphi(u_0).$$

Since $\mathcal{J}_\lambda u_\lambda \to u$ when $\lambda \searrow 0$ and φ is l.s.c we see that $\partial_t u \in L^2\big(0, T;\, \mathcal{H}\big)$ with

$$\varphi(u) + \int_0^t \|\partial_t u(s)\|_{\mathcal{H}}^2\, ds \leq \varphi(u_0).$$

Now suppose that $\partial_t u \in L^2\big(0, T;\, \mathcal{H}\big)$. Then analogue to (37) we have for $0 < s < t$ that

$$\varphi(u(t)) = \varphi(u(s)) + \int_s^t \|\partial_t u\|_{\mathcal{H}}^2 \, dx.$$

Note that $\varphi(u(s))$ remains bounded even if $s = 0$. By definition of the domain of a function on a $\mathbb{R}$-Hilbert space, this implies that $u_0 \in D(\varphi)$. This proofs the stated the equivalence relation.

For the property (v) we have already established the equation (27) by an analogue computation as in (37). From this identity we see that $\varphi(u(t)) \nearrow \varphi(u_0)$ when $t \searrow 0$. We can verify the equation (28) by using the identity (27) and the Hölder inequality

$$\|u(T) - u_0\|_{\mathcal{H}} \leq \int_0^T \|\partial_t u\|_{\mathcal{H}} \, dt \leq \left(\int_0^t 1 \, dt\right)^{\frac{1}{2}} \left(\int_0^T \|\partial_t u\|_{\mathcal{H}}^2 \, dt\right)^{\frac{1}{2}} \leq \sqrt{T} \cdot \sqrt{\varphi(u_0) - \varphi(u(t))},$$

and we prove equation (29) by again employing the equation (27) and using the fact that $\left\|\partial_t^+ u\right\|_{\mathcal{H}}$ is monotonically decreasing as stated in theorem 4.18 to show

$$T \left\|\partial_t^+ u(T)\right\|_{\mathcal{H}}^2 \leq \int_0^T \left\|\partial_t^+ u\right\|_{\mathcal{H}}^2 \, dt = \varphi(u_0) - \varphi(u(T)).$$

□

We will state two criteria, one being stronger than the other, on the inhomogeneous term f and proper convex l.s.c function φ such that a weak solution of the GSD equation will be a strong solution. Each of these conditions will have certain effects on the attributes of the solution u.

For the weaker criterion we will need the following lemma.

Lemma 4.39. *Let $u \in W^{1,2}(0, T;\, \mathcal{H})$ and $g \in L^2\big(0, T;\, \mathcal{H}\big)$ with $u(t) \in D(A)$ and $g(t) \in Au(t)$ for almost all $t \in (0, T)$. Then $\varphi(u)$ is absolutely continuous on $[0, T]$. If we define the set P consisting of all points $t \in [0, T]$ such that $u(t) \in D(A)$ and u and $\varphi(u)$ are differentiable in t, then*

$$\partial_t \varphi(u(t)) = \langle h,\, \partial_t u(t)\rangle_{\mathcal{H}} \qquad \forall t \in P \;\; \forall h \in Au(t). \tag{38}$$

Proof. For all $\lambda > 0$ define the function $u_\lambda := A_\lambda u$. Almost everywhere on $[0, T]$ we have by proposition 3.28 and the definition of the minimum norm section, the estimate

$$\|u_\lambda\|_{\mathcal{H}} = \|A_\lambda u\|_{\mathcal{H}} \leq \|A_0 u\|_{\mathcal{H}} \leq \|g\|_{\mathcal{H}}.$$

As g is in $L^2\big(0, T;\, \mathcal{H}\big)$ we have by the dominated convergence theorem that $u_\lambda \to A_0 u$ in $L^2\big(0, T;\, \mathcal{H}\big)$. Next we show that the function $\varphi(u)$ is absolutely continuous. We know that $\partial_t \varphi_\lambda(u) = \langle A_\lambda u,\, \partial_t u\rangle_{\mathcal{H}}$ and thus

$$\varphi_\lambda(u(t)) - \varphi_\lambda(u(s)) = \int_s^t \langle A_\lambda u,\, \partial_t u\rangle_{\mathcal{H}} \, dx.$$

By proposition 3.28 and proposition 3.40 we can take the limit $\lambda \searrow 0$ to show that

$$\varphi(u(t)) - \varphi(u(s)) = \int_s^t \langle A_0 u,\, \partial_t u\rangle_{\mathcal{H}} \, dx.$$

This shows that the function $\varphi(u)$ is absolutely continuous. Finally take $t \in P$ and $h \in Au(t)$. Then note for all $\varepsilon > 0$ sufficiently small we have by definition of the sub differential

$$\varphi(u(t \pm \varepsilon)) - \varphi(u(t)) \geq \langle h,\, u(t \pm \varepsilon) - u(t)\rangle_{\mathcal{H}}.$$

If we divide by ε and take the limit $\varepsilon \searrow 0$ we find that the desired equation (38) is correct. □

For the criteria we will also require three supplementary results from Brézis, namely in respective order [3, Proposition 2.16], [3, Proposition 2.17] and [3, Corollary 2.3]. We will only present the claims but not prove them, as they are tangentially related to the subject.

Proposition 4.40. *(Special Version) For all $u \in L^2\big(0,T;\,\mathcal{H}\big)$ define the function*

$$\Phi(u) := \begin{cases} \int_0^T \varphi(u)\,dt & \text{if } \varphi(u) \in L^1\big(0,T;\,\mathcal{H}\big), \\ +\infty & \text{otherwise.} \end{cases}$$

Then Φ is a proper convex l.s.c function on $L^2\big(0,T;\,\mathcal{H}\big)$ and $\partial\Phi$ coincides with the maximal monotone operator $\mathcal{I}\,A$ on $L^2\big(0,T;\,\mathcal{H}\big)$, which we defined in example 3.16.

The weaker criterion demands that the inhomogeneous term is in $L^2\big(0,T;\,\mathcal{H}\big)$ and that the proper convex l.s.c function φ attains a minimal value.

Proposition 4.41. *(Special Version) Let B be a maximally monotone multivalued operator on $\mathcal{H}$ and define its resolvent as $\mathcal{J}_\lambda^B := (\mathrm{id}_\mathcal{H} + \lambda B)^{-1}$. Suppose that there exists a constant $C \in \mathbb{R}$ such that $\varphi(\mathcal{J}_\lambda^B\,x) \le \varphi(x) + C\lambda$ for all $x \in \mathcal{H}$ and all $\lambda > 0$. Then the operator $B + A$ is maximally monotone and we have the estimate*

$$\|B_0 x\|_\mathcal{H} \le \|(B+A)_0 x\|_\mathcal{H} + \sqrt{C} \qquad \forall x \in D(A) \cap D(B). \tag{39}$$

Proposition 4.42. *Let B be a maximally monotone multivalued operator on $\mathcal{H}$. Then $R(B) = \mathcal{H}$ (i.e. B is surjective) if*

$$\lim_{\substack{x \in D(B) \\ \|x\|_\mathcal{H} \to +\infty}} \|B_0 x\|_\mathcal{H} = +\infty.$$

Theorem 4.43. *If $f \in L^2\big(0,T;\,\mathcal{H}\big)$ and if φ has minimal value 0, then if we define the set $K := \{x \in \mathcal{H} \mid \varphi(x) = 0\}$ we have the following properties*

(i) Weak solutions u of the GSD equation with operator A, inhomogeneous term f and initial value u_0 are strong solutions.

(ii) Almost everywhere on $(0,T)$ we have the identity

$$\|\partial_t u\|_\mathcal{H}^2 + \partial_t \varphi(u) = \langle f,\, \partial_t u\rangle_\mathcal{H}.$$

(iii) If $u_0 \in D(\varphi)$ then $\varphi(u) = \varphi \circ u$ is absolutely continuous and $\partial_t u \in L^2\big(0,T;\,\mathcal{H}\big)$ with

$$\|\partial_t u\|_{L^2(0,T;\mathcal{H})} \le \|f\|_{L^2(0,T;\mathcal{H})} + \sqrt{\varphi(u_0)}. \tag{40}$$

(iv) The weak derivative of u satisfies $\sqrt{t}\,\partial_t u \in L^2\big(0,T;\,\mathcal{H}\big)$ with an estimate

$$\|\sqrt{t}\,\partial_t u\|_{L^2(0,T;\mathcal{H})} \le \|\sqrt{t}\,f\|_{L^2(0,T;\mathcal{H})} + \frac{1}{\sqrt{2}}\|f\|_{L^1(0,T;\,\mathcal{H})} + \frac{1}{\sqrt{2}}\,\mathrm{dist}(u_0, K). \tag{41}$$

(v) In general the map $\varphi(u) = \varphi \circ u$ is in $L^1\big(0,T;\,\mathcal{H}\big)$ and $\varphi(u)$ is absolutely continuous on all compact subsets of $(0,T]$.

(vi) We have the estimates

$$\|\partial_t u\|_{L^2(s,T;\mathcal{H})} \le \|f\|_{L^2(0,T;\mathcal{H})} + \frac{1}{\sqrt{2s}}\|f\|_{L^1(0,s;\,\mathcal{H})} + \frac{1}{\sqrt{2s}}\operatorname{dist}(u_0, K) \quad \forall s \in (0,T). \tag{42}$$

Note in (42) *we take the integral of subintervals of* $[0,T]$.

Proof. First we prove the property (i) and the equation (40) of property (iii). Thus assume that $u_0 \in D(\varphi)$. Define the operator B in $L^2\big(0,T;\,\mathcal{H}\big)$ the same way as the operator in example 3.18, where $D(B)$ sets u_0 as an initial condition. For any $\lambda > 0$ we denote the resolvent of B by $\mathcal{J}_\lambda^B := (\mathrm{id}_{\mathcal{H}} + \lambda B)^{-1}$. From example 3.22 we know that $u_\lambda := \mathcal{J}_\lambda^B u$ is expressed by

$$u_\lambda(t) = e^{-\frac{t}{\lambda}}u_0 + \frac{1}{\lambda}\int_0^t e^{\frac{s-t}{\lambda}}u(s)\,ds.$$

Define a function $\Phi : L^2\big(0,T;\,\mathcal{H}\big) \to (-\infty, +\infty]$ by

$$\Phi(u) = \begin{cases} \int_0^T \varphi(u)\,dt & \text{if } \varphi(u) \in L^1\big(0,T;\,\mathcal{H}\big), \\ +\infty & \text{otherwise} \end{cases} \qquad \forall u \in L^2\big(0,T;\,\mathcal{H}\big).$$

From proposition 4.40 we know hat Φ is a proper convex l.s.c. map. As φ is a proper convex l.s.c. function we have that

$$\varphi(u_\lambda(t)) \le e^{-\frac{t}{\lambda}}\varphi(u_0) + \frac{1}{\lambda}\int_0^t e^{\frac{s-t}{\lambda}}\varphi(u(s))\,ds. \tag{43}$$

Indeed we define a probability measure μ_t supported in $[0,t]$ by

$$\mu_t = e^{-\frac{t}{\lambda}}\delta_0 + \frac{1}{\lambda}e^{\frac{s-t}{\lambda}}\chi_{[0,t]}\,ds,$$

where $\chi_{[0,t]}$ is the indicator function with support $[0,t]$ and δ_0 is the Dirac delta function. Then we are able to verify (43) by the Jensen inequality

$$\begin{aligned} \varphi(u_\lambda(t)) &= \varphi\left(e^{-\frac{t}{\lambda}}u_0 + \frac{1}{\lambda}\int_0^t e^{\frac{s-t}{\lambda}}u(s)\,ds\right) \\ &= \varphi\left(\int_{[0,t]} u\,d\mu_t\right) \\ &\le \int_{[0,t]} \varphi(u)\,d\mu_t \\ &= e^{-\frac{t}{\lambda}}\varphi(u_0) + \frac{1}{\lambda}\int_0^t e^{\frac{s-t}{\lambda}}\varphi(u(s))\,ds. \end{aligned}$$

If we integrate the inequality (43) over $(0,T)$ we get

$$\begin{aligned} \Phi(u_\lambda) &\le \lambda(1 - e^{-\frac{T}{\lambda}})\varphi(u_0) + \frac{1}{\lambda}\int_0^T\int_0^t e^{\frac{s-t}{\lambda}}\varphi(u(s))\,ds\,dt \\ &= \lambda(1 - e^{-\frac{T}{\lambda}})\varphi(u_0) + \frac{1}{\lambda}\int_0^T\int_s^T e^{\frac{s-t}{\lambda}}\varphi(u(s))\,dt\,ds \\ &= \lambda(1 - e^{-\frac{T}{\lambda}})\varphi(u_0) + \frac{1}{\lambda}\int_0^T e^{\frac{s}{\lambda}}\int_s^T e^{\frac{-t}{\lambda}}\,dt\,\varphi(u(s))\,ds \\ &= \lambda(1 - e^{-\frac{T}{\lambda}})\varphi(u_0) + \int_0^T (1 - e^{\frac{s-T}{\lambda}})\varphi(u(s))\,ds \\ &\le \lambda\varphi(u_0) + \Phi(u). \end{aligned}$$

From proposition 4.41 we get that $B + \partial\Phi$ is a maximally monotone operator on $L^2\big(0, T;\, \mathcal{H}\big)$ as-well as the bound (40) by

$$\|\partial_t u\|_{L^2(0,T;\mathcal{H})} = \|B_0 u\|_{L^2(0,T;\mathcal{H})} \le \|(\partial\Phi + B)_0 u\|_{L^2(0,T;\mathcal{H})} + \sqrt{\varphi(u_0)} = \|f\|_{L^2(0,T;\mathcal{H})} + \sqrt{\varphi(u_0)}.$$

Proposition 4.40 states that we can replace $\partial\Phi$ by $\mathcal{I}\,A$. By the Poincaré inequality (adapt the proof in [4, Proposition 8.13] to Banach spaces) we have for some $C > 0$ that

$$\|u\|_{L^2(0,T;\mathcal{H})} \le C\|\partial_t u\|_{L^2(0,T;\mathcal{H})} \le C\|(\mathcal{I}\,A + B)u\|_{L^2(0,T;\mathcal{H})} + C\sqrt{\varphi(u_0)}.$$

However this means we can apply proposition 4.42 to see that $R(\mathcal{I}\,A + B) = L^2\big(0, T;\, \mathcal{H}\big)$, which proves the claim (i).

Next we show the attribute (ii). Let $(u_n)_{n\ge1}$ be strong solutions of the GSD equation with operator A, inhomogeneous term f and initial value $w_n \in D(\varphi)$ such that $w_n \to u_0$. Then $u_n \to u$ in $C^0\big(0, T;\, \mathcal{H}\big)$. Note that the property (i) holds for u_n. In particular we observe that $\partial_t u_n \in L^2\big(0, T;\, \mathcal{H}\big)$. Moreover as u_n is continuous we deduce that $u_n \in W^{1,2}(0, T;\, \mathcal{H})$. Also there is $g(t) \in Au(t)$ for almost all $t \in [0, T]$ with $\partial_t u + g = f$ and so $g = f - \partial_t u \in L^2\big(0, T;\, \mathcal{H}\big)$. We are able to apply lemma 4.39 to get

$$\|\partial_t u_n\|^2_{\mathcal{H}} + \partial_t\varphi(u_n) = \langle f,\, \partial_t u_n\rangle_{\mathcal{H}}\,. \tag{44}$$

From the attribute (ii) we see that $\varphi(u)$ is absolutely continuous. As we have already verified the identity (40), we can thus conclude the proof of property (iii).

Next we tackle (iv). If we apply the Cauchy-Schwarz inequality, multiply with t and integrate over $(0, T)$ we get

$$\int_0^T \|\partial_t u_n\|^2_{\mathcal{H}}\, t\, dt + T\varphi(u_n(T)) - \int_0^T \varphi(u_n)\, dt \le \int_0^T \|f\|_{\mathcal{H}} \cdot \|\partial_t u_n\|_{\mathcal{H}}\, t\, dt.$$

Using the Hölder inequality we see that

$$\begin{aligned}\|\sqrt{t}\,\partial_t u_n\|^2_{L^2(0,T;\mathcal{H})} - \|\sqrt{\varphi(u_n)}\|^2_{L^2(0,T;\mathcal{H})} &\le \|\sqrt{t}\,\partial_t u_n\|^2_{L^2(0,T;\mathcal{H})} + T\varphi(u_n(T)) - \|\sqrt{\varphi(u_n)}\|^2_{L^2(0,T;\mathcal{H})} \\ &\le \|\sqrt{t}\,f\|_{L^2(0,T;\mathcal{H})}\|\sqrt{t}\,\partial_t u_n\|_{L^2(0,T;\mathcal{H})}.\end{aligned}$$

From this we deduce

$$\|\sqrt{t}\,\partial_t u_n\|_{L^2(0,T;\mathcal{H})} \le \|\sqrt{t}\,f\|_{L^2(0,T;\mathcal{H})} + \|\sqrt{\varphi(u_n)}\|_{L^2(0,T;\mathcal{H})}. \tag{45}$$

Indeed this follows from an elementary calculation with $a, b, c \ge 0$ by

$$a(a+b) \le a^2 - b^2 \le ac \qquad \Rightarrow \qquad a + b \le c \qquad \Rightarrow \qquad a \le c - b \le c + b.$$

By definition of the sub differential A we have almost everywhere on $(0, T)$ the inequality

$$\varphi(v) - \varphi(u_n) \ge \langle Au_n,\, v - u_n\rangle_{\mathcal{H}} = \langle f - \partial_t u_n,\, v - u_n\rangle_{\mathcal{H}}\,.$$

In case $v \in K$ we derive the following bound by using the Cauchy-Schwarz inequality

$$\begin{aligned}-\varphi(u_n) &\ge \langle f - \partial_t u_n,\, v - u_n\rangle_{\mathcal{H}} \\ &\Rightarrow \qquad \varphi(u_n) + \langle \partial_t(u_n - v),\, u_n - v\rangle_{\mathcal{H}} \le \langle f,\, u_n - v\rangle_{\mathcal{H}} \le \|f\|_{\mathcal{H}}\,\|u_n - v\|_{\mathcal{H}} \\ &\Rightarrow \qquad \varphi(u_n) + \frac{1}{2}\partial_t\,\|u_n - v\|^2_{\mathcal{H}} \le \|f\|_{\mathcal{H}}\,\|u_n - v\|_{\mathcal{H}}\,.\end{aligned}$$

Then if we take the integral over $(0,t)$ we get

$$\begin{aligned}\frac{1}{2}\left(\|u_n - v\|_{\mathcal{H}}^2 + 2\int_0^t \varphi(u_n)\,ds\right) &\le \frac{1}{2}\|w_n - v\|_{\mathcal{H}}^2 + \int_0^t \|f\|_{\mathcal{H}}\,\|u_n - v\|_{\mathcal{H}}\,ds \\ &= \frac{1}{2}\|w_n - v\|_{\mathcal{H}}^2 + \int_0^t \|f\|_{\mathcal{H}}\left(\|u_n - v\|_{\mathcal{H}}^2\right)^{\frac{1}{2}}\,ds \\ &\le \frac{1}{2}\|w_n - v\|_{\mathcal{H}}^2 + \int_0^t \|f\|_{\mathcal{H}}\left(\|u_n - v\|_{\mathcal{H}}^2 + 2\int_0^t \varphi(u_n)\,dx\right)^{\frac{1}{2}}\,ds.\end{aligned}$$

Using the Gronwall lemma (cf. lemma 4.31)

$$\sqrt{2}\,\|\sqrt{\varphi(u_n)}\|_{L^2(0,T;\mathcal{H})} \le \left(\|u_n - v\|_{\mathcal{H}}^2 + 2\int_0^t \varphi(u_n)\,dx\right)^{\frac{1}{2}} \le \|w_n - v\|_{\mathcal{H}} + \int_0^t \|f\|_{\mathcal{H}}\,dt.$$

In particular

$$\|\sqrt{\varphi(u_n)}\|_{L^2(0,T;\mathcal{H})} \le \frac{1}{\sqrt{2}}\|f\|_{L^1(0,T;\,\mathcal{H})} + \frac{1}{\sqrt{2}}\|w_n - v\|_{\mathcal{H}}. \tag{46}$$

We combine (45) and (46) to get

$$\|\sqrt{t}\,\partial_t u_n\|_{L^2(0,T;\mathcal{H})} \le \|\sqrt{t}\,f\|_{L^2(0,T;\mathcal{H})} + \frac{1}{\sqrt{2}}\|f\|_{L^1(0,T;\,\mathcal{H})} + \frac{1}{\sqrt{2}}\|w_n - v\|_{\mathcal{H}}.$$

By lemma 4.16 if we take the limit $n \to \infty$ we see that (41) holds true, which proves (iv).
From (46) we see that $\varphi(u) \in L^1\big(0,T;\,\mathcal{H}\big)$. Then the absolute continuity claim of (v) is a consequence of the attribute (iii). Thus we have shown (v).
For the attribute (vi), let $s \in (0,T)$ and apply the mean value theorem on (46). Then there exists $t_n \in (0,s)$ with

$$\varphi(u_n(t_n)) = \frac{1}{s}\int_0^s \varphi(u_n)\,dt \le \frac{1}{2s}\left(\|w_n - v\|_{\mathcal{H}} + \|f\|_{L^1(0,s;\,\mathcal{H})}\right)^2. \tag{47}$$

If we integrate (44) on (t_n, T) we get

$$\begin{aligned}\|\partial_t u_n\|_{L^2(t_n,T;\mathcal{H})}^2 &= \varphi(u_n(t_n)) - \varphi(u_n(T)) + \int_{t_n}^T \|f\|_{\mathcal{H}}\,\|\partial_t u_n\|_{\mathcal{H}}\,dt \\ &\le \varphi(u_n(t_n)) + \int_{t_n}^T \|f\|_{\mathcal{H}}\,\|\partial_t u_n\|_{\mathcal{H}}\,dt.\end{aligned}$$

Then applying (47) to this inequality we derive

$$\begin{aligned}\|\partial_t u_n\|_{L^2(s,T;\mathcal{H})} &\le \|\partial_t u_n\|_{L^2(t_n,T;\mathcal{H})} \\ &\le \|f\|_{L^2(t_n,T;\mathcal{H})} + \sqrt{\varphi(u_n(t_n))} \\ &\le \|f\|_{L^2(0,T;\mathcal{H})}^2 + \frac{1}{\sqrt{2s}}\|f\|_{L^1(0,s;\,\mathcal{H})} + \frac{1}{\sqrt{2s}}\|w_n - v\|_{\mathcal{H}}.\end{aligned}$$

By lemma 4.16 if we take the limit $n \to \infty$ we get (42). This concludes the proof of the Theorem. □

For the stronger criterion we require that f is absolutely continuous and that both f and $\partial_t f$ are in $L^2\big(0,T;\,\mathcal{H}\big)$.

Theorem 4.44. *Let $f \in W^{1,2}(0,T;\,\mathcal{H})$ and assume that φ has minimal value 0 and define the set $K := \{x \in \mathcal{H} \mid \varphi(x) = 0\}$. Then we have the following properties.*

(i) All weak solutions of the GSD equation with operator A, inhomogeneous term f and initial value $u_0 \in \overline{D(A)}$ are strong solutions.

(ii) The range of u is in $D(A)$ on $(0,T]$, u is right differentiable on $(0,T)$, $t\partial_t u \in L^\infty\big(0,T;\,\mathcal{H}\big)$ and the following estimate holds true

$$\begin{aligned}\left\|\partial_t^+ u(t)\right\|_{\mathcal{H}} \leq \|f\|_{\mathcal{H}} + \frac{1}{t}\operatorname{dist}(u_0,K) + \int_0^t \|\partial_t f\|_{\mathcal{H}} \frac{s^2}{t^2}\,ds \\ + \frac{\sqrt{2}}{t}\left(\int_0^t \|\partial_t f\|_{\mathcal{H}}\, s\,ds\right)^{\frac{1}{2}}\left(\operatorname{dist}(u_0,K) + \int_0^t \|f\|_{\mathcal{H}}\,ds\right)^{\frac{1}{2}}.\end{aligned} \tag{48}$$

Proof. From theorem 4.43 we know that weak solutions are strong solutions. By definition of strong solutions we have that the range of u is in $D(A)$ on $(0,T)$. From proposition 4.37 we know that the range of u is in $D(A)$ on $(0,T]$, u is right differentiable in $(0,T)$ and u is Lipschitz on all compact subsets of $(0,T]$. We must only verify the bound in equation (48).
First we assume that $u_0 \in D(\varphi)$. Since u is continuous we know that $u \in L^2\big(0,T;\,\mathcal{H}\big)$ and by theorem 4.43 we get $\partial_t u \in L^2\big(0,T;\,\mathcal{H}\big)$. But then also $f - \partial_t u \in L^2\big(0,T;\,\mathcal{H}\big)$, which means that for all $v \in \mathcal{H}$ we can apply lemma 4.39 to get

$$\|\partial_t u\|_{\mathcal{H}}^2 + \partial_t\varphi(u) = \langle f,\, \partial_t u\rangle_{\mathcal{H}} = \langle f,\, \partial_t(u-v)\rangle_{\mathcal{H}}\,.$$

Since $\partial_t u \in L^2\big(0,T;\,\mathcal{H}\big)$ as stated in theorem 4.43, if we multiply this equation by t and integrate over $(0,T)$ we get

$$\begin{aligned}&\|\sqrt{t}\,\partial_t u\|_{L^2(0,T;\mathcal{H})}^2 + T\varphi(u(T)) \\ &\qquad = \|\varphi(u)\|_{L^1(0,T;\,\mathcal{H})} + \int_0^T \langle t\,f,\, \partial_t u - \partial_t v\rangle_{\mathcal{H}}\,dt \\ &\qquad = \|\varphi(u)\|_{L^1(0,T;\,\mathcal{H})} + T\,\langle f(T),\, u(T)-v\rangle_{\mathcal{H}} - \int_0^T \langle f + t\,\partial_t f,\, u-v\rangle_{\mathcal{H}}\,dt.\end{aligned} \tag{49}$$

By definition of the sub differential we have for almost everywhere $(0,T)$ that

$$\varphi(v) - \varphi(u) \geq \langle Au,\, v-u\rangle_{\mathcal{H}} = \langle f - \partial_t u,\, v-u\rangle_{\mathcal{H}}\,.$$

If $v \in K$ then by restructuring the equation and integrating over $(0,T)$ we get that

$$\|\varphi(u)\|_{L^1(0,T;\,\mathcal{H})} \leq \int_0^T \langle f,\, u-v\rangle_{\mathcal{H}}\,dt + \frac{1}{2}\left\|u_0 - v^2\right\|_{\mathcal{H}} - \frac{1}{2}\|u(T)-v\|_{\mathcal{H}}^2\,. \tag{50}$$

If we use this estimate (50) in the equation (49) we get by the Cauchy Schwarz inequality

$$
\begin{aligned}
\|\sqrt{t}\,\partial_t u\|^2_{L^2(0,T;\mathcal{H})} &\leq \|\sqrt{t}\,\partial_t u\|_{L^2(0,T;\mathcal{H})} + T\varphi(u(T)) \\
&= \|\varphi(u)\|_{L^1(0,T;\,\mathcal{H})} + T\,\langle f(T),\, u(T)-v\rangle_{\mathcal{H}} - \int_0^T \langle f + t\,\partial_t f,\, u-v\rangle_{\mathcal{H}}\, dt \\
&\leq \frac{1}{2}\left\|u_0 - v^2\right\|_{\mathcal{H}} - \frac{1}{2}\|u(T)-v\|^2_{\mathcal{H}} + T\,\langle f(T),\, u(T)-v\rangle_{\mathcal{H}} \\
&\qquad + \int_0^T \langle f,\, u-v\rangle_{\mathcal{H}}\, dt - \int_0^T \langle f + t\,\partial_t f,\, u-v\rangle_{\mathcal{H}}\, dt \\
&\leq \frac{1}{2}T^2\,\|f(T)\|^2_{\mathcal{H}} + \frac{1}{2}\|u_0 - v\|^2_{\mathcal{H}} - \int_0^T \langle u-v,\, t\,\partial_t f\rangle_{\mathcal{H}}\, dt \\
&\leq \frac{1}{2}T^2\,\|f(T)\|^2_{\mathcal{H}} + \frac{1}{2}\|u_0 - v\|^2_{\mathcal{H}} + \int_0^T \|u-v\|_{\mathcal{H}}\,\|t\,\partial_t f\|_{\mathcal{H}}\, dt,
\end{aligned} \tag{51}
$$

where in the penultimate inequality we added

$$
\frac{1}{2}\|Tf(T) - (u(T)-v)\|^2_{\mathcal{H}} = \frac{1}{2}T^2\,\|f(T)\|^2_{\mathcal{H}} - T\,\langle f(T),\, u(T)-v\rangle_{\mathcal{H}} + \frac{1}{2}\|u(T)-v\|^2_{\mathcal{H}}\,.
$$

Note we have $0 \in \partial\varphi(v)$, since for any $x \in \mathcal{H}$ we have

$$
\varphi(x) - \varphi(v) = \varphi(x) \geq 0 = \langle 0,\, x-v\rangle_{\mathcal{H}}\,.
$$

Then equation (19) in lemma 4.30 states for the constant solution v that

$$
\|u-v\|_{\mathcal{H}} \leq \|u_0 - v\|_{\mathcal{H}} + \int_0^T \|f\|_{\mathcal{H}}\, dt.
$$

We apply this bound in (51) and derive

$$
\begin{aligned}
&\|\sqrt{t}\,\partial_t u\|^2_{L^2(0,T;\mathcal{H})} \\
&\quad\leq \frac{1}{2}T^2\,\|f(T)\|^2_{\mathcal{H}} + \frac{1}{2}\|u_0 - v\|^2_{\mathcal{H}} + \big(\,\|u_0 - v\|_{\mathcal{H}} + \|f\|_{L^1(0,T;\,\mathcal{H})}\big)\|t\partial_t f\|_{L^1(0,T;\,\mathcal{H})}.
\end{aligned} \tag{52}
$$

From equation (24) in proposition 4.37 we know

$$
\left\|\partial_t^+ u(T)\right\|_{\mathcal{H}} \leq \left\|\partial_t^+ u(t)\right\|_{\mathcal{H}} + \|\partial_t f\|_{L^1(t,T;\,\mathcal{H})}. \tag{53}
$$

In addition from Hölders inequality we see that

$$
\|t\,\partial_t u\|_{L^1(0,T;\,\mathcal{H})} \leq \|\sqrt{t}\|_{L^2(0,T;\mathcal{H})}\|\sqrt{t}\,\partial_t u\|_{L^2(0,T;\mathcal{H})} = \frac{T}{\sqrt{2}}\|\sqrt{t}\,\partial_t u\|_{L^2(0,T;\mathcal{H})}.
$$

Then if we integrate (53) on $(0,T)$ and apply this bound we deduce

$$
\begin{aligned}
\frac{T^2}{2}\left\|\partial_t^+ u(T)\right\|_{\mathcal{H}} &= \int_0^T \left\|\partial_t^+ u(T)\right\|_{\mathcal{H}} t\, dt \\
&\leq \int_0^T \left\|\partial_t^+ u\right\|_{\mathcal{H}} t\, dt + \int_0^T \int_t^T \|\partial_t f(s)\|_{\mathcal{H}}\, t\, ds\, dt \\
&\leq \frac{T}{\sqrt{2}}\|\sqrt{t}\,\partial_t u\|_{L^2(0,T;\mathcal{H})} + \int_0^T \|\partial_t f(s)\|_{\mathcal{H}} \int_0^s t\, dt\, ds \\
&\leq \frac{T}{\sqrt{2}}\|\sqrt{t}\,\partial_t u\|_{L^2(0,T;\mathcal{H})} + \int_0^T \|\partial_t f\|_{\mathcal{H}}\, \frac{s^2}{2}\, ds.
\end{aligned}
$$

Then if we multiply with $2/T^2$ we get

$$\left\|\partial_t^+ u(T)\right\|_{\mathcal{H}} \leq \frac{\sqrt{2}}{T} \|\sqrt{t}\,\partial_t u\|_{L^2(0,T;\mathcal{H})} + \int_0^T \|\partial_t f\|_{\mathcal{H}} \frac{t^2}{T^2}\, dt.$$

If we insert the estimate (52) we deduce the bound (48).

For the general case $u_0 \in \overline{D(\varphi)}$, take any sequence $(w_n)_{n\geq 1} \subseteq D(\varphi)$ with $w_n \to u_0$ as $n \to \infty$. Let u_n be the solution to the *GSD* with operator A, inhomogeneous term f and initial value w_n. From equation (19) in lemma 4.30 we know that $u_n \to u$ in $C^0\big(0, T;\, \mathcal{H}\big)$. Indeed we have that $\|u_n - u\|_{\mathcal{H}}$ is uniformly bounded from above by $\|w_n - u_0\|_{\mathcal{H}}$. Then the desired inequality (48) follows if we take the limit $n \to \infty$ in the following equation

$$\begin{aligned}\|A_0 u_n - f\|_{\mathcal{H}} = \left\|\partial_t^+ u_n\right\|_{\mathcal{H}} \leq \|f\|_{\mathcal{H}} + \frac{1}{t}\operatorname{dist}(w_n, K) + \int_0^t \|\partial_t f\|_{\mathcal{H}} \frac{s^2}{t^2}\, ds \\ + \frac{\sqrt{2}}{t}\left(\int_0^t s\,\|\partial_t f\|_{\mathcal{H}}\, ds\right)^{\frac{1}{2}} \big(\operatorname{dist}(w_n, K) + \|f\|_{L^1(0,t;\,\mathcal{H})}\big)^{\frac{1}{2}}.\end{aligned}$$

□

Theorem 4.45. *Let $f : [0, +\infty) \to \mathcal{H}$ be a function such that it is absolutely continuous on all compact subsets of $(0, +\infty)$, the weak derivative satisfies $\partial_t f \in L^1\big(0, T;\, \mathcal{H}\big)$ and the map has the limit $f(t) \to w$ when $t \to \infty$ such that $w \in R(A)$. If u is a strong solution to the GSD equation with operator A, inhomogeneous term f and initial value u_0, then the following attributes hold true*

(i) The right derivatie vanishes at infinity, i.e. $\left\|\partial_t^+ u\right\|_{\mathcal{H}} \to 0$ when $t \to +\infty$.

(ii) If $t\,\partial_t f \in L^1\big(0, T;\, \mathcal{H}\big)$, then we have the asymptotic behavior $\left\|\partial_t^+ u\right\|_{\mathcal{H}} = O(1/t)$.

Proof. We start by proving property (i). Observe for any $v \in A^{-1}w$ and $x \in \mathcal{H}$ we have by definition of the sub differential

$$\varphi(x) \geq \varphi(v) + \langle w,\, x - v\rangle_{\mathcal{H}} \qquad \Rightarrow \qquad \varphi(x) - \langle w,\, x\rangle_{\mathcal{H}} \geq \varphi(v) - \langle w,\, v\rangle_{\mathcal{H}}\,.$$

Therefore the function $\psi : \mathcal{H} \to$ (-$\infty, +\infty$] characterized by

$$\psi(x) := \varphi(x) - \langle w,\, x\rangle_{\mathcal{H}} - \min_{u\in D(\varphi)}\{\varphi(u) - \langle w,\, u\rangle_{\mathcal{H}}\} \qquad \forall x \in \mathcal{H},$$

is well-defined. Clearly the map ψ attains a minimal value 0 in any point of $A^{-1}w$. Thus define the set $K = \{v \in \mathcal{H} \mid w \in Av\}$. Let B be the sub differential operator of ψ, i.e. $B := \partial\psi$. Then the GSD equation can be restated as

$$\partial_t^+ u + Bu \ni f - w.$$

If we use equation (48) from theorem 4.44 we get

$$\begin{aligned}\left\|\partial_t^+ u\right\|_{\mathcal{H}} \leq \|f - w\|_{\mathcal{H}} + \frac{1}{t}\operatorname{dist}(u_0, K) + \int_0^t \|\partial_t f(s)\|_{\mathcal{H}} \frac{s^2}{t^2}\, ds \\ + \frac{\sqrt{2}}{t}\left(\int_0^t \|\partial_t f\|_{\mathcal{H}}\, s\, ds\right)^{\frac{1}{2}} \left(\operatorname{dist}(u_0, K) + \int_0^t \|f - w\|_{\mathcal{H}}\, ds\right)^{\frac{1}{2}}.\end{aligned} \tag{54}$$

We now proceed by showing that the individual terms in the above estimate tend to zero as t diverges towards infinity. This is clear for the first two terms. The third term converges to zero by the Lebesgue dominated convergence theorem. Indeed we have

$$\|\partial_t f(s)\|_{\mathcal{H}} \frac{s^2}{t^2} \leq \|\partial_t f(s)\|_{\mathcal{H}} \qquad \forall 0 < s \leq t,$$

and for all $s > 0$ we have a point wise null-sequence

$$\lim_{t\to+\infty} \|\partial_t f(s)\|_{\mathcal{H}} \frac{s^2}{t^2}.$$

For the fourth term in our bound (54) we analyze two factors. For the first factor we have exactly analogue as before by the Lebesgue dominated convergence theorem

$$\lim_{t\to+\infty} \sqrt{\frac{2}{t}} \left(\int_0^t \|\partial_t f\|_{\mathcal{H}} \, s \, ds \right)^{\frac{1}{2}} = 0.$$

For the remaining factor, we compute using Fubini's Theorem

$$\begin{aligned} \int_0^t \int_s^{+\infty} \|\partial_t f(x)\|_{\mathcal{H}} \, dx \, ds &= \int_0^t \int_0^x \|\partial_t f(x)\|_{\mathcal{H}} \, ds \, dx + \int_t^{+\infty} \int_0^t \|\partial_t f(x)\|_{\mathcal{H}} \, ds \, dx \\ &= \int_0^t \|\partial_t f\|_{\mathcal{H}} \, s \, ds + t \int_t^{+\infty} \|\partial_t f\|_{\mathcal{H}} \, dx. \end{aligned}$$

Observe the following identity

$$f(t) - w = f(-(-t)) - w = \int_{-\infty}^{-t} \partial_t\big(f(-s)\big) \, ds = -\int_{-\infty}^{-t} \partial_t f(-s) \, ds = -\int_t^{+\infty} \partial_t f(s) \, ds.$$

We use this to deduce again with the Lebesgue dominated convergence theorem

$$\lim_{t\to+\infty} \frac{1}{t} \int_0^t \|f - w\|_{\mathcal{H}} \, ds \leq \lim_{t\to+\infty} \int_t^{+\infty} \|\partial_t f\|_{\mathcal{H}} \, ds + \int_0^t \|\partial_t f\|_{\mathcal{H}} \frac{s}{t} \, ds = 0.$$

Therefore every term in the bound (54) converges to 0, which proves the claim (i).

Lastly we tackle the attribute (ii). We argue similarly as for the property (i), namely we want to show that each term in the estimate (54) is $O(1/t)$. Then for the first term, as $t\,\partial_t f \in L^1\big(0, T;\, \mathcal{H}\big)$ we have for some $C > 0$ that

$$\|f - w\|_{\mathcal{H}} \leq \int_t^{+\infty} \|\partial_t f\|_{\mathcal{H}} \, ds \leq \int_t^{+\infty} \|\partial_t f\|_{\mathcal{H}} \frac{s}{t} \, ds \leq \frac{C}{t}.$$

and for the third term in (54) we have

$$\int_0^t \|\partial_t f\|_{\mathcal{H}} \frac{s^2}{t^2} \, ds \leq \int_0^t \|\partial_t f\|_{\mathcal{H}} \frac{s}{t} \, ds \leq \frac{C}{t}.$$

For the last term in (54) it suffices to show the second integral is bounded. Indeed we verify

$$\int_0^t \|f - w\|_{\mathcal{H}} \, ds = \int_0^t \int_s^{+\infty} \|\partial_t f\|_{\mathcal{H}} \, dx \, ds = t \int_t^{+\infty} \|\partial_t f\|_{\mathcal{H}} \, dx + \int_0^t \|\partial_t f\|_{\mathcal{H}} \, s \, ds \leq 2C.$$

This proves the clam (ii) and thus concludes the proof. $\square$

5 Application Machine Learning

In this section we will discuss the application of Gradient Flow in Machine Learning. We first explain what Machine Learning is, the connection to the theory of gradient flow and then finish with an example and numerical experiment. This section is based almost entirely on the lecture notes of the course "Advancced Topics in Machine Learning" from the University of Bern [7].

5.1 Gradient Flow in Training of Neural Networks

In many applications of machine learning, we are dealing with an unknown probability distribution μ on $\mathcal{X}\times\mathcal{Y}$ with a set of features $\mathcal{X}$ and a set of targets $\mathcal{Y}$. Given some loss function $L:\mathcal{Y}\times\mathcal{Y}\to\mathbb{R}$ we want to solve the following optimization problem, referred to as Bayes Risk,

$$\underset{f:\mathcal{X}\to\mathcal{Y}}{\text{minimize}} \qquad \int_{\mathcal{X}\times\mathcal{Y}} L(f(x),y)\,d\mu(x,y). \tag{BR}$$

In practice the only information we have of the distribution μ is a finite dataset $D\subseteq\mathcal{X}\times\mathcal{Y}$ drawn independently under the distribution μ. We approximate μ by the empirical risk as a probability measure μ_{emp} on $\mathcal{X}\times\mathcal{Y}$ defined by

$$\mu_{\text{emp}} := \frac{1}{|D|}\sum_{(x,y)\in D}\delta_{(x,y)}.$$

With this probability measure our optimization problem becomes

$$\underset{f:\mathcal{X}\to\mathcal{Y}}{\text{minimize}} \qquad \int_{\mathcal{X}\times\mathcal{Y}} L(f(x),y)\,d\mu_{\text{emp}}(x,y). \tag{eBR}$$

In many applications we must approximate the optimal solution by functions that are parametrisable by some weight. A highly versatile family of such functions are called neural networks. The set of all neural networks N is given inductively by

(i) For any map $f:\mathbb{R}^n\times\mathbb{R}^k\to\mathbb{R}^m$ and any weight $\theta\in\mathbb{R}^k$, the set of all neural networks N contains the function $\mathbb{R}^n\to\mathbb{R}^m$; $x\mapsto f(x;\theta):=f(x,\theta)$. We say $f(\bullet;\theta)$ is a layer with weight θ and $f(\bullet;\theta)_j$ is a node of the layer for $1\le j\le m$.

(ii) Let $g(\bullet;\theta):\mathbb{R}^n\to\mathbb{R}^m$ be some neural network in N weighted by $\theta\in\mathbb{R}^k$ and $h:\mathbb{R}^m\to\mathbb{R}^r$ be some layer in N weighted by $\pi\in\mathbb{R}^s$. Then the composition $f(\bullet;\theta,\pi):=h(g(\bullet;\theta);\pi)$ is another neural network in N weighted by (θ,π). For the particular neural network $f(\bullet;\theta,\pi)$, we call h the output layer. Similarly we call the first layer in the composition of a neural network the input layer.

A neural network is commonly graphically represented as seen in figure 2. The reason the family of functions are called neural networks is that the graph resembles a biological neural network.

In practice a neural network is an alternating sequence between affine layers of the form $f(x;A,b)=Ax+b$ for a weight $(A,b)\in\mathbb{R}^{m\times n}\times\mathbb{R}^m$ and some choice of non-linear layers $h:\mathbb{R}^n\to\mathbb{R}^m$, such as the sigmoid function $s:\mathbb{R}\to\mathbb{R}$; $x\mapsto(1+\exp(-x))^{-1}$ or the rectified linear unit function $\text{ReLu}:\mathbb{R}\to\mathbb{R}$; $x\mapsto\max\{0,x\}$ applied entry-wise to the input. These non-linear mappings are called activation functions.

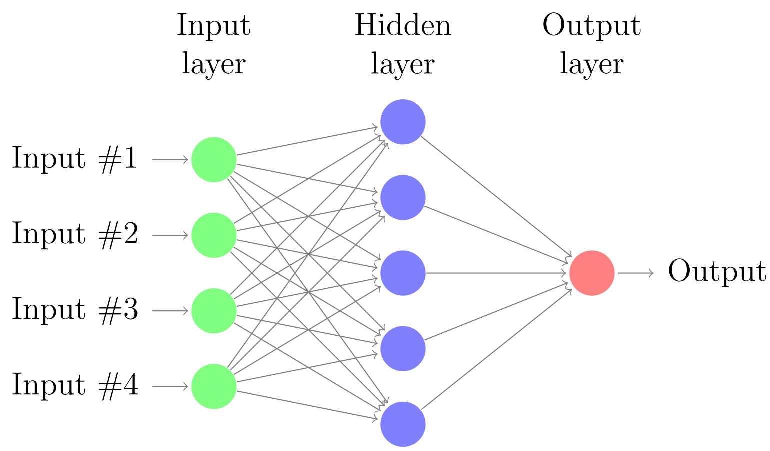

Figure 2: A common graphical representation of a neural network.

The approximation of the ideal function with some neural network is prone to generalization errors. To combat this, we often add a regularization function $R : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$. As an example we may demand that the output data to be sparse. In this case we could choose $R(x, \theta) = |x_1| + \ldots + |x_n|$ for $(x, \theta) \in \mathbb{R}^n \times \mathbb{R}^m$. Our minimization problem thus becomes

$$\underset{\theta \in \mathbb{R}^k}{\text{minimize}} \qquad \int_{\mathcal{X}\times\mathcal{Y}} L(f(x;\theta), y) + R(f(x;\theta), \theta)\, d\mu_{\text{emp}}(x, y). \tag{Reg-eBR}$$

We extract two function definitions

$$\varphi(\theta) := \int_{\mathcal{X}\times\mathcal{Y}} L(f(x;\theta), y)\, d\mu_{\text{emp}}(x, y) \quad \text{and} \quad \psi(\theta) := \int_{\mathcal{X}\times\mathcal{Y}} R(f(x;\theta), \theta)\, d\mu_{\text{emp}}(x, y).$$

and restate state our minimization problem as the general problem

$$\underset{\theta \in \mathbb{R}^k}{\text{minimize}} \qquad \varphi(\theta) + \psi(\theta).$$

Usually the only assumption we place on φ is that it is Lipschitz continuous. However for many applications we may assume that φ is a proper convex differentiable function and that ψ is a proper convex function (cf. [12, Section 3.1]). In this case we know from [11] that $\theta^* \in \mathbb{R}^k$ is an optimal solution to the minimization problem if and only if

$$0 \in \partial f(\theta^*) + \partial g(\theta^*).$$

Provided that both φ and ψ are l.s.c, the operator $A := \partial f + \partial g$ is maximally monotone by corollary 3.38. We find the optimal solution by solving the evolution equation hGSD with operator A and any initial value $\theta_0 \in \mathbb{R}^k$. In particular we solve

$$\begin{cases} \partial_t \theta + A\theta \ni 0 & \text{on } [0, +\infty), \\ \theta = \theta_0 & \text{on } \{0\}. \end{cases}$$

In practice it is far more efficient and suffices to approximate the solution by the discrete gradient descent algorithm, instead of solving the actual evolution equation. The generalized discrete gradient descent method is given by the forward Euler method

$$\begin{cases} \theta_t - (\mathrm{id}_{\mathbb{R}^k} - \alpha A)\theta_{t-1} \ni 0 & \text{on } \mathbb{N}, \\ \theta_t = \theta_0 & \text{on } \{0\}. \end{cases} \tag{gDGM}$$

for some user defined step size $\alpha > 0$. The theory of gradient flows applies to non-convex functions as-well, however fewer properties are known and the asymptotic limit may change depending on the initial value.

5.2 Example Ridge Regression

The paper [12, Section 3.2] from M. Signoretto et al mentions two examples of how the gradient descent method can be applied, where the minimization function is convex. Namely Ridge Regression [8] and Group Lasso [9]. We will present an example of Ridge Regression as it is very simple, a smooth and convex problem, and it highlights important aspects of machine learning.

Ridge Regression is a method for solving the standard model for linear regression. Suppose that there exists some unknown affine linear transformation $h : \mathbb{R}^k \to \mathbb{R}$. We are given a finite dataset $(X_1, Y_1), \dots, (X_n, Y_n) \in \mathbb{R}^k \times \mathbb{R}$ such that $Y_i = h(X_i) + \varepsilon_i$ for all $1 \leq i \leq n$, where the variables $\varepsilon_1, \dots, \varepsilon_n$ are unknown and are chosen i.i.d from some probability distribution.

Naturally we will try to approximate the linear functional h with a single layer affine neural network defined by $f(x; a, b) := a^t x + b$ for any $x, a \in \mathbb{R}^k$ and $b \in \mathbb{R}$.

We define our loss function L as the squared error

$$L(x, y) := \frac{1}{2}(x - y)^2 \qquad \forall x, y \in \mathbb{R}.$$

We want to control the growth of the weights a, b with some user defined parameter $\lambda > 0$ by the regularization function

$$R(a, b) := \frac{\lambda}{2}(a^t a + b^2) \qquad \forall (a, b) \in \mathbb{R}^k \times \mathbb{R}.$$

Our problem statement Reg-eBR thus becomes

$$\underset{(a,b)\in\mathbb{R}^n\times\mathbb{R}}{\text{minimize}} \qquad \frac{\lambda}{2}(a^t a + b^2) + \frac{1}{2n}\sum_{i=1}^{n}(Y_i - f(X_i; a, b))^2.$$

We use matrix notation to simplify the above expression. To that end define $X \in \mathbb{R}^{n\times k}$ by $X := (X_1, \dots, X_n)^t$, define $Y \in \mathbb{R}^n$ by $Y := (Y_1, \dots, Y_n)$ and define the vector the constant one vector $e := (1, \dots, 1) \in \mathbb{R}^n$. We compute

$$\begin{aligned}
\frac{1}{2n}\sum_{i=1}^{n}(Y_i - f(X_i; a, b))^2 &= \frac{1}{2n}(Y - Xa - be)^t(Y - Xa - be) \\
&= \frac{1}{2n}\begin{pmatrix} a \\ b \end{pmatrix}^t \begin{pmatrix} X^t X & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} - \frac{1}{n}Y^t \begin{pmatrix} X & e \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} + \frac{1}{2n}Y^t Y.
\end{aligned}$$

Then the Reg-eBR equation can be simplified to

$$\underset{(a,b)\in\mathbb{R}^n\times\mathbb{R}}{\text{minimize}} \qquad \begin{pmatrix} a \\ b \end{pmatrix}^t \left(\frac{\lambda}{2}\,\mathrm{id}_{\mathbb{R}^{k+1}} + \frac{1}{2n}\begin{pmatrix} X^t X & 0 \\ 0 & 1 \end{pmatrix}\right)\begin{pmatrix} a \\ b \end{pmatrix} - \frac{1}{n}Y^t \begin{pmatrix} X & e \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} + \frac{1}{2n}Y^t Y.$$

We see that we must minimize a quadratic polynomial, which happens to be a proper convex smooth problem. Thus our theory is applicable.

To deduce the discrete gradient descent method we must simply compute the derivative of a and b

$$\begin{aligned}\Delta(a,b) :=& \partial_{(a,b)} \begin{pmatrix} a \\ b \end{pmatrix}^t \left(\frac{\lambda}{2} \operatorname{id}_{\mathbb{R}^{k+1}} + \frac{1}{2n} \begin{pmatrix} X^t X & 0 \\ 0 & 1 \end{pmatrix} \right) \begin{pmatrix} a \\ b \end{pmatrix} - \frac{1}{n} Y^t \begin{pmatrix} X & e \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} + \frac{1}{2n} Y^t Y \\ =& \left(\lambda \operatorname{id}_{\mathbb{R}^{k+1}} + \frac{1}{n} \begin{pmatrix} X^t X & 0 \\ 0 & 1 \end{pmatrix} \right) \begin{pmatrix} a \\ b \end{pmatrix} - \frac{1}{n} \begin{pmatrix} X^t \\ e^t \end{pmatrix} Y.\end{aligned}$$

The discrete gradient descent method is therefore given by

$$\begin{cases} (a_t, b_t) - (a_{t-1}, b_{t-1}) + \alpha \Delta(a_{t-1}, b_{t-1}) = 0 & \text{on } \mathbb{N}, \\ (a_t, b_t) = (a_0, b_0) & \text{on } \{0\}. \end{cases}$$

5.2.1 Numerical Experiment

We end the paper by presenting a concrete example of ridge regression[6]. Given some dimension k, some fixed $a_{\text{real}} \in \mathbb{R}^k$ and $b_{\text{real}} \in \mathbb{R}$, we will use the discrete gradient descent method to approximate the linear functional $h_{\text{real}}(x) := a_{\text{real}}^t x + b_{\text{real}}$. We choose n points $X_1, \dots, X_n$ independently under a uniform distribution from an open subset $\Omega \subseteq \mathbb{R}^k$. For some user defined $\sigma > 0$, we draw n values $\varepsilon_1, \dots, \varepsilon_n$ independently from the normal distribution $\mathcal{N}(0, \sigma^2)$ and define $Y_i := h_{\text{real}}(X_i) + \varepsilon_i$ for $1 \leq i \leq n$. We choose the initial weights (a_0, b_0) uniformly randomly from an open set $\Theta \subseteq \mathbb{R}^k \times \mathbb{R}$.

This numerical experiment is conducted using MatLab and the standard random number generator provided by MatLab.

For our experiment we set $k = 2$, $a_{\text{real}} = (1, -1)$, $b_{\text{real}} = 2$, $\Omega = [-10, 10]^2$ and $\Theta = [-10, 10]^3$. With these parameters we let the discrete gradient descent algorithm run for a 1000 iterations. We repeat the experiment 10 times and plot the results in the following figures:

In figure 3 we show how the parameter a_t evolves depending on its initial weight. We see that for each starting value the graphs more or less converge in a straight line to the value a_{real}.

In figure 4 we present how the parameter b_t changes depending on its initial weight. Again we see that for each starting value the graphs converge to the value b_{real}.

Finally in figure 5 we present how the parameter a_b develops depending on its initial weight. For each starting value the graphs decay linearly to 0.

[6] all of the code that was used in this experimenter is found in the following GitHub repository

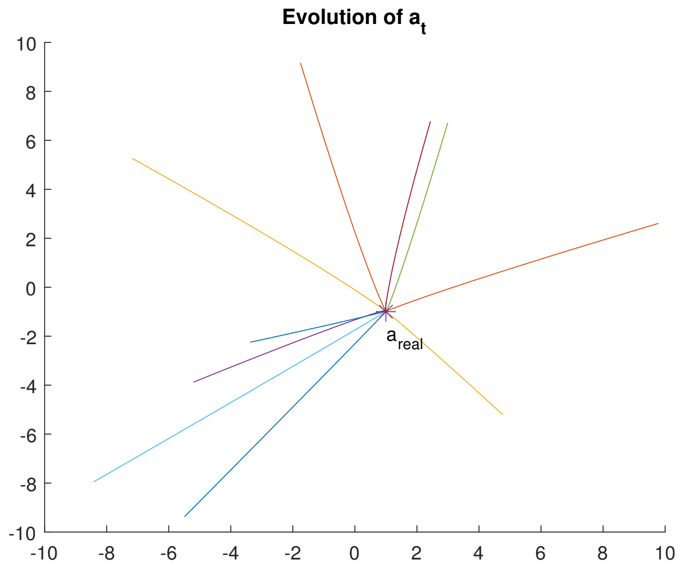

Figure 3: The evolution of a_t in the discrete gradient descent method for different initial values.

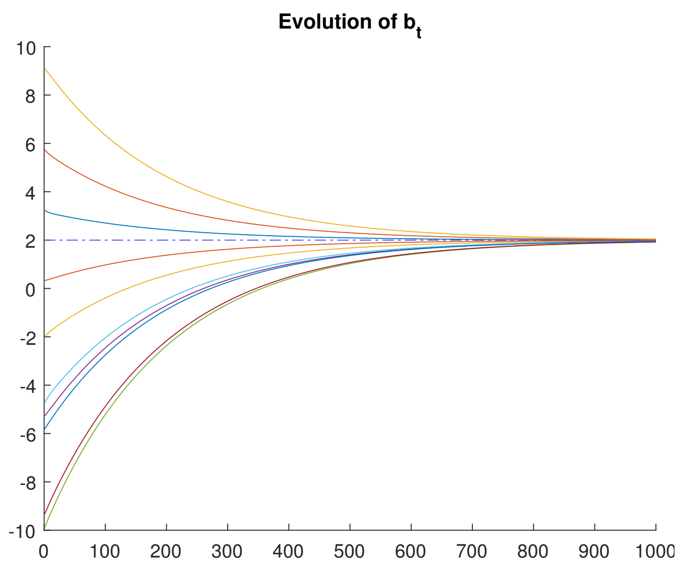

Figure 4: The evolution of b_t in the discrete gradient descent method for different initial values.

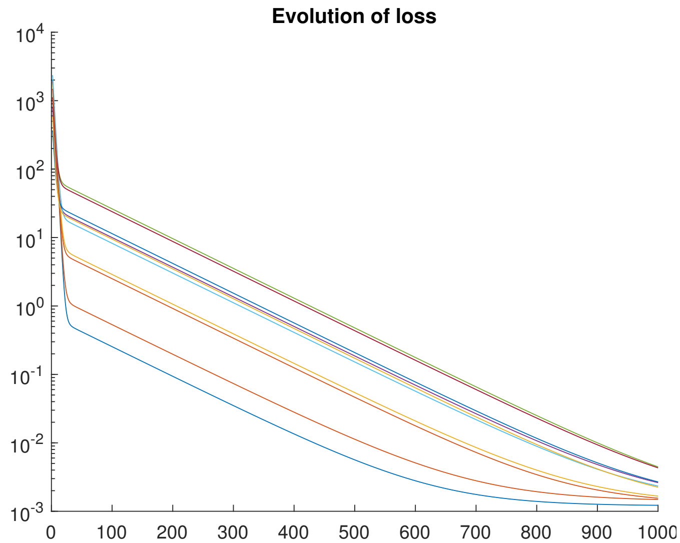

Figure 5: The evolution of the loss function in the discrete gradient descent method for different initial values. Note that the scale is logarithmic in the y axis.

References

[1] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *CoRR*, abs/1511.00561, 2015.

[2] Jonathan M Borwein. Fifty years of maximal monotonicity. *Optimization Letters*, 4(4):473–490, 2010.

[3] Haim Brezis. *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*. Elsevier, 1973.

[4] Haim Brezis and Haim Brézis. *Functional analysis, Sobolev spaces and partial differential equations*, volume 2. Springer, 2011.

[5] M. G. Crandall and T. M. Liggett. Generation of semi-groups of nonlinear transformations on general banach spaces. *American Journal of Mathematics*, 93(2):265–298, 1971.

[6] Sergey Edunov, Myle Ott, Michael Auli, and David Grangier. Understanding back-translation at scale. *arXiv preprint arXiv:1808.09381*, 2018.

[7] P. Favaro. Lecture slides in advanced topics in machine learning. "https://www.cvg.unibe.ch/teaching/course/ATML", February 2021.

[8] Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

[9] Laurent Jacob, Guillaume Obozinski, and Jean-Philippe Vert. Group lasso with overlap and graph lasso. In *Proceedings of the 26th annual international conference on machine learning*, pages 433–440, 2009.

[10] Daniel S Park, Yu Zhang, Ye Jia, Wei Han, Chung-Cheng Chiu, Bo Li, Yonghui Wu, and Quoc V Le. Improved noisy student training for automatic speech recognition. *arXiv preprint arXiv:2005.09629*, 2020.

[11] Ralph Tyrell Rockafellar. *Convex analysis*. Princeton university press, 2015.

[12] Marco Signoretto, Quoc Tran Dinh, Lieven De Lathauwer, and Johan AK Suykens. Learning with tensors: a framework based on convex optimization and spectral regularization. *Machine Learning*, 94(3):303–351, 2014.