Chapter 17
Gradient flows

17.1 ▪ The classical continuous steepest descent

17.1.1 ▪ Existence and uniqueness of global orbits

$\mathcal{H}$ is a real Hilbert space with scalar product and norm denoted by $\langle \cdot, \cdot \rangle$ and $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$, respectively. Let $\Phi : \mathcal{H} \to \mathbf{R}$ be a real-valued function which is continuously differentiable. We often refer to Φ as the potential function. By the Riesz representation theorem, for any $u \in \mathcal{H}$ there exists a unique vector in $\mathcal{H}$, denoted by $\nabla\Phi(u)$, and called the gradient of Φ at u, such that

$$\Phi'(u)(v) = \langle \nabla\Phi(u), v \rangle \quad \forall v \in \mathcal{H}.$$

The differential equation on $\mathcal{H}$ which is governed by the gradient vector field $v \in \mathcal{H} \mapsto -\nabla\Phi(v) \in \mathcal{H}$

$$\text{(SD)} \quad \dot{u}(t) = -\nabla\Phi(u(t)) \tag{17.1}$$

is called the (classical) continuous steepest descent. Throughout the chapter we write either $\frac{dv}{dt}$ or $\dot{v}$ to denote the distributional derivative of any function $v : (0, +\infty) \to \mathcal{H}$. We call the *trajectory* (also the *solution curve* or *orbit*) of the differential equation (SD) any C^1 function $u : [0, +\infty) \to \mathcal{H}$ satisfying (17.1) for all $t \geq 0$. The word "classical" refers to the fact that the potential Φ is continuously differentiable, which allows us to consider classical solutions ($u \in C^1$) of (SD). By contrast, for nonsmooth potential functions, we will have to consider weaker notions of solution.

Descent property. The following property explains the important role played by the steepest descent dynamic in optimization.

Proposition 17.1.1 (descent property). *Let $u \in C^1([0, +\infty); \mathcal{H})$ be a trajectory of* (SD). *Then $t \mapsto \Phi(u(t))$ is a decreasing function, and for all $t \geq 0$*

$$\frac{d}{dt}\Phi(u(t)) = -\|\dot{u}(t)\|^2. \tag{17.2}$$

Thus, as long as the trajectory does not reach a stationary point, the function $t \mapsto \Phi(u(t))$ is decreasing.

PROOF. Taking the scalar product with $\dot{u}(t)$ in (17.1) yields

$$\|\dot{u}(t)\|^2 + \langle \nabla\Phi(u(t)), \dot{u}(t) \rangle = 0. \tag{17.3}$$

Using the classical derivation chain rule

$$\frac{d}{dt}\Phi(u(t)) = \langle \nabla\Phi(u(t)), \dot{u}(t)\rangle,$$

(17.3) gives

$$\frac{d}{dt}\Phi(u(t)) = -\|\dot{u}(t)\|^2.$$

Hence $t \mapsto \Phi(u(t))$ is a nonincreasing function. Indeed, as long as the trajectory moves (does not stop), it is a decreasing function. □

Related notions: We use the term *integral curve*, especially if we are interested in the image in $\mathcal{H}$ of a trajectory rather than in the trajectory itself as a function. The terminology *dynamical system* refers to the evolution of each point of $\mathcal{H}$ by the *semiflow* (or *semigroup*) $(S(t))_{t\geq 0}$ generated by $-\nabla\Phi$. For each $t \in [0,+\infty)$, the operator $S(t)$ associates to each $u_0 \in \mathcal{H}$ the point $S(t;u_0) = u(t)$, where u is the unique orbit starting at u_0, that is, $\dot{u}(t) = -\nabla\Phi(u(t))$; $u(0) = u_0$.

The minus sign in (SD) reflects the fact that we are interested in the minimization of Φ. In fact, in view of the maximization of Φ, one would rather consider the reversing flow $\dot{u}(t) = \nabla\Phi(u(t))$ called steepest ascent, which has the same integral curves as (SD) with a different orientation. The descent property of the trajectories is obtained by increasing the time variable. That's why we consider only the semiflow $\{S(t); t \geq 0\}$, and from now on, unless specified, we consider only orbits which are defined on some positive time interval.

An *equilibrium* for a semiflow $(S(t))_{t\geq 0}$ is a point $z \in \mathcal{H}$ such that $S(t)z = z$ for all $t \geq 0$. An equivalent terminology is *stationary point*. For the semiflow $(S(t))_{t\geq 0}$ generated by $-\nabla\Phi$ the equilibria are the critical points of Φ, i.e., $\mathrm{crit}\Phi = \{v \in \mathcal{H} : \nabla\Phi(v) = 0\}$.

The potential function $\Phi : \mathcal{H} \to \mathbf{R}$ is a *strict Lyapunov function* for the steepest descent differential system (SD). This means that for each trajectory $t \mapsto u(t)$ of (SD) the real-valued mapping $t \mapsto \Phi(u(t))$ is (strictly) decreasing, as long as the trajectory has not reached an equilibrium. This property makes (SD) a *dissipative* system, which, as we shall see, has a great impact on the asymptotic behavior of the trajectories of (SD). In particular, the only periodic trajectories of (SD) are the trajectories which remain at critical points. This makes a great difference with conservative systems (like Hamiltonian systems) which naturally exhibit many periodic trajectories.

Geometrical aspects. The steepest descent direction has a natural geometrical interpretation. Being at $u \in \mathcal{H}$, and $v \in \mathcal{H}$ being normalized $\|v\| = 1$, the directional derivative of Φ at u in the direction v is equal to

$$\frac{d}{dt}\Phi(u + tv)|_{t=0} = \langle \nabla\Phi(u), v\rangle.$$

When u is noncritical, i.e., $\nabla\Phi(u) \neq 0$, it is minimal for

$$v = -\frac{1}{\|\nabla\Phi(u)\|}\nabla\Phi(u).$$

Being at u, by using only first-order local conditions on Φ, the direction $v \in \mathcal{H}$ that provides the greatest decrease of Φ is $-\nabla\Phi(u)$, whence the steepest descent terminology for differential equation (17.1). We could as well consider the differential equation

$$\text{(SD)} \quad \dot{u}(t) = -\alpha(t)\nabla\Phi(u(t)), \tag{17.4}$$

where $\alpha : \mathbf{R}^+ \to \mathbf{R}^+$ is a smooth positive function. Differential systems (17.1) and (17.4) have the same integral curves. One can pass from one system to the other by time reparametrization (e.g., length parametrization), which does not change the portrait of the system.

The flow is often represented by its phase portrait, that is, the picture of its integral curves in $\mathcal{H}$. For gradient flows, the integral curves are in the direction perpendicular to the equipotential surfaces, which helps in visualizing them. This is a consequence of the following result, which plays a key role in differential geometry.

Lemma 17.1.1. *Let $\Phi : \mathcal{H} \to \mathbf{R}$ be a C^1 function. Let*

$$\Sigma_c = \{v \in \mathcal{H} : \Phi(v) = c\}$$

be a level set of Φ for some level $c \in \mathbf{R}$. Assume that $u \in \Sigma_c$ (i.e., $\Phi(u) = c$), and u is not critical (i.e., $\nabla\Phi(u) \neq 0$). Then, in a neighborhood of u, Σ_c is a C^1 surface, with codimension 1, *and $\nabla\Phi(u)$ is orthogonal to Σ_c at u.*

One can also get a mechanical intuition of the gradient flow by considering the graph of Φ, which is a surface in $\mathcal{H}\times\mathbf{R}$. Given an orbit $t \mapsto u(t)$ of (SD), the point $(u(t), \Phi(u(t))$ moves along the graph of Φ just like a drop of water: it slides down the surface in the direction of steepest descent until it reaches the bottom of the surface. Then it stops. In accordance with this mechanical interpretation, there is no inertial effect in (SD). This is due to the fact that, in the equation of mechanics, the acceleration term $\ddot{u}$ has been neglected. (SD) is a first-order differential equation with respect to time t. Due to space limitations, the study of dissipative gradient systems involving inertial features, such as the heavy ball with friction dynamical system, is not considered here, the interested reader can consult [12], [50], and references therein.

Let us show the existence and uniqueness of a global solution of the Cauchy problem for (SD).

Theorem 17.1.1. *Let $\Phi : \mathcal{H} \to \mathbf{R}$ be a real-valued function which satisfies the following:*

(i) *Φ is minorized, i.e.,* $\inf_{\mathcal{H}} \Phi > -\infty$.

(ii) *Φ is continuously differentiable, and $\nabla\Phi : \mathcal{H} \to \mathcal{H}$ is Lipschitz continuous on bounded sets, i.e., for each positive $R > 0$ there exists some constant $L_R \geq 0$ such that*

$$\|\nabla\Phi(v_2) - \nabla\Phi(v_1)\| \leq L_R \|v_2 - v_1\| \quad \forall v_1, v_2 \in B(0, R).$$

Then, for any $u_0 \in \mathcal{H}$, the following properties hold:

(a) global existence: *There exists a unique classical global solution $u \in C^1([0, +\infty); \mathcal{H})$ of the Cauchy problem*

$$\begin{cases} \dot{u}(t) = -\nabla\Phi(u(t)), \\ u(0) = u_0. \end{cases} \tag{17.5}$$

(b) descent property: *$t \mapsto \Phi(u(t))$ is a decreasing function, and, for all $t \geq 0$,*

$$\frac{d}{dt}\Phi(u(t)) = -\|\dot{u}(t)\|^2. \tag{17.6}$$

(c) finite energy property:

$$\int_0^\infty \|\dot{u}(t)\|^2 dt \leq \Phi(u_0) - \inf_{\mathcal{H}} \Phi < +\infty. \tag{17.7}$$

PROOF. The existence of solutions to (SD) is based on the Cauchy–Lipschitz theorem that we recall below (in a local form that fits our study, and in the classical global form).

Theorem 17.1.2 (Cauchy–Lipschitz). (a) (local version) *Let $F : \mathcal{H} \to \mathcal{H}$ be a vector field which is Lipschitz continuous on bounded sets, i.e., for each positive $R > 0$ there exists some constant $L_R \geq 0$ such that*

$$||F(v_2) - F(v_1)|| \leq L_R ||v_2 - v_1|| \quad \forall v_1, v_2 \in B(0,R).$$

Then, for any $u_0 \in \mathcal{H}$, there exists some $T > 0$, such that there exists a unique classical solution $u \in C^1([-T,+T];\mathcal{H})$ of the Cauchy problem

$$\begin{cases} \dot{u}(t) = F(u(t)) \\ u(0) = u_0. \end{cases} \tag{17.8}$$

(b) (global version) *Let $F : \mathcal{H} \to \mathcal{H}$ be a vector field which is globally Lipschitz continuous, i.e., there exists some constant $L \geq 0$ such that*

$$||F(v_2) - F(v_1)|| \leq L ||v_2 - v_1|| \quad \forall v_1, v_2 \in \mathcal{H}.$$

Then, for any $u_0 \in \mathcal{H}$ there exists a unique global classical solution $u \in C^1(\mathbf{R};\mathcal{H})$ of Cauchy problem (17.8).

PROOF OF THEOREM 17.1.1 CONTINUED. By the Cauchy–Lipschitz theorem (local version), there exists some $T > 0$ such that Cauchy problem (17.5) admits a unique local solution $u \in C^1([0,T];\mathcal{H})$. Let

$$T_{max} = \sup\left\{T > 0 : \text{ there exists a solution of (17.5) on } [0,T]\right\}.$$

Thus the maximal solution u of (17.5) belongs to $C^1([0,T_{max}[;\mathcal{H})$. Let us show that $T_{max} = +\infty$. We argue by contradiction and assume that $T_{max} < +\infty$. Let us show that $\lim_{t \to T_{max}} u(t)$ exists. For all $t \in [0,T_{max}[$ we have

$$\dot{u}(t) + \nabla\Phi(u(t)) = 0.$$

By the descent property (17.2)

$$\frac{d}{dt}\Phi(u(t)) = -||\dot{u}(t)||^2. \tag{17.9}$$

By integration of (17.9) on $[0,T]$, with $0 < T < T_{max}$, we obtain

$$\int_0^T ||\dot{u}(t)||^2 dt + \Phi(u(T)) - \Phi(u_0) = 0.$$

By assumption (i), Φ is minorized. Hence

$$\int_0^T ||\dot{u}(t)||^2 dt \leq \Phi(u_0) - \inf_{\mathcal{H}} \Phi < +\infty.$$

This majorization being valid for any $0 < T < T_{max}$, taking the supremum with respect to T yields

$$\int_0^{T_{max}} ||\dot{u}(t)||^2 dt \leq \Phi(u_0) - \inf_{\mathcal{H}} \Phi < +\infty. \tag{17.10}$$

From (17.10) we deduce a property of uniform continuity of $u : [0, T_{max}[\to \mathcal{H}$. For any $0 \le s \le t < T_{max}$, by the Cauchy–Schwarz inequality

$$\begin{aligned} \|u(t) - u(s)\| &\le \int_s^t \|\dot{u}(\tau)\| d\tau \\ &\le \sqrt{t-s} \left(\int_s^t \|\dot{u}(\tau)\|^2 d\tau \right)^{\frac{1}{2}} \\ &\le \sqrt{t-s} \left(\int_0^{T_{max}} \|\dot{u}(\tau)\|^2 d\tau \right)^{\frac{1}{2}} \\ &\le C\sqrt{t-s} \end{aligned}$$

with $C = (\Phi(u_0) - \inf_{\mathcal{H}} \Phi)^{\frac{1}{2}}$ given by (17.10). Thus, $u : [0, T_{max}[\to \mathcal{H}$ is Hölder continuous and hence uniformly continuous from $[0, T_{max}[$ into the complete metric space $\mathcal{H}$. By the classical continuous extension theorem, u admits a unique extension by continuity to $[0, T_{max}]$, that is,

$$\lim_{t \to T_{max}} u(t) := u_{T_{max}} \quad \text{exists.}$$

By applying the Cauchy–Lipschitz theorem (local version) with Cauchy data $u_{T_{max}}$ at initial time $t = T_{max}$, we obtain a solution $w \in C^1([T_{max}, T_1]; \mathcal{H})$ with $T_1 > T_{max}$ of the Cauchy problem:

$$\begin{cases} \dot{w}(t) = -\nabla\Phi(w(t)), \\ w(T_{max}) = u_{T_{max}}. \end{cases}$$

The function $\tilde{u} : [0, T_1] \to \mathcal{H}$ which is equal to u on $[0, T_{max}]$ and equal to w on $[T_{max}, T_1]$ belongs to $C^1([0, T_1]; \mathcal{H})$. Indeed, $\tilde{u}$ and its time derivative are continuous at $t = T_{max}$. This last property follows from the continuity of the vector field $\nabla\Phi$:

$$\lim_{t \to T_{max},\ t < T_{max}} \dot{\tilde{u}}(t) = \lim_{t \to T_{max},\ t > T_{max}} \dot{\tilde{u}}(t) = -\nabla\Phi(u_{T_{max}}).$$

Thus, $\tilde{u}$ satisfies the steepest descent equation on an interval strictly larger than T_{max}, which contradicts the maximality of T_{max}. □

17.1.2 ▪ Asymptotic properties, $t \to +\infty$

Let us examine the asymptotic behavior of the trajectories and provide some first general properties. This is a topic of fundamental importance in optimization. Moreover, it models the evolution of the transition between the initial state and the final equilibrium for many systems in physics, biology, economics, and so forth. We will use the following lemma.

Lemma 17.1.2. *Let* $g : [0, +\infty[\to [0, +\infty[$ *be a continuous function which satisfies* (i) *and* (ii)*:*

(i) $\int_0^{+\infty} g(t)^2 dt < +\infty$.

(ii) g *is Lipschitz continuous on* $[0, +\infty[$.

Then $g(t) \to 0$ *as* $t \to +\infty$.

PROOF. Let us argue by contradiction and suppose that the statement ($g(t) \to 0$ as $t \to +\infty$) is false. Then, there exists some $\epsilon_0 > 0$ and a sequence $t_n \to +\infty$ such that for each $n \in \mathbf{N}$, $g(t_n) \geq \epsilon_0$. After extraction of a subsequence, we can assume that, for each $n \in \mathbf{N}$, $|t_{n+1} - t_n| > 1$. Suppose that g is L-Lipschitz continuous. On the interval $[t_n - \frac{\epsilon_0}{2L}, t_n + \frac{\epsilon_0}{2L}]$, we have

$$\begin{aligned} g(t) &\geq g(t_n) - L|t - t_n| \\ &\geq \epsilon_0 - L|t - t_n| \\ &\geq \frac{\epsilon_0}{2}. \end{aligned}$$

Set $\eta = \inf\left\{\frac{1}{2}, \frac{\epsilon_0}{2L}\right\} > 0$. The intervals $[t_n - \eta, t_n + \eta]$ do not overlap, and, on each of them, g is minorized by $\frac{\epsilon_0}{2}$. From this we infer

$$\begin{aligned} \int_0^{+\infty} g(t)^2 dt &\geq \sum_n \int_{t_n - \eta}^{t_n + \eta} g(t)^2 dt \\ &\geq \sum_n 2\eta \left(\frac{\epsilon_0}{2}\right)^2 = +\infty, \end{aligned}$$

a clear contradiction with $\int_0^{+\infty} g(t)^2 dt < +\infty$. □

Theorem 17.1.3. *Let $\Phi : \mathcal{H} \to \mathbf{R}$ be a real-valued function which satisfies the following:*

(i) *Φ is minorized, i.e., $\inf_{\mathcal{H}} \Phi > -\infty$.*

(ii) *Φ is continuously differentiable, and $\nabla \Phi : \mathcal{H} \to \mathcal{H}$ is Lipschitz continuous on bounded sets.*

Let $u \in C^1([0, +\infty); \mathcal{H})$ be a bounded orbit generated by the gradient flow associated to Φ. Then,

$$\lim_{t \to +\infty} \dot{u}(t) = 0, \qquad \lim_{t \to +\infty} \nabla \Phi(u(t)) = 0.$$

As a consequence, if $u(t_n) \to u_\infty$ for some sequence $t_n \to +\infty$, then $\nabla \Phi(u_\infty) = 0$.

PROOF. Let us show that the function $g : [0, +\infty[\to [0, +\infty[$, which is defined for any $t \geq 0$ by $g(t) = ||\dot{u}(t)||$, satisfies conditions (i) and (ii) of Lemma 17.1.2.

(i) By the finite energy property (see (17.7) in Theorem 17.1.1), we have $\int_0^\infty ||\dot{u}(t)||^2 dt \leq \Phi(u_0) - \inf_{\mathcal{H}} \Phi < +\infty$. Equivalently, $\int_0^{+\infty} g(t)^2 dt < +\infty$, which is item (i).

(ii) Let us now show that g is Lipschitz continuous on $[0, +\infty[$. Since u has been assumed to be bounded, there exists some $R > 0$ such that $||u(t)|| \leq R$ for all $t \geq 0$. Since $\nabla \Phi : \mathcal{H} \to \mathcal{H}$ is Lipschitz continuous on bounded sets, there exists some $L_R > 0$ such that for all $s, t \geq 0$,

$$||\nabla \Phi(u(t)) - \nabla \Phi(u(s))|| \leq L_R ||u(t) - u(s)||.$$

From this, and the definition of the gradient flow, we infer that for all $s, t \geq 0$,

$$\begin{aligned} |g(t) - g(s)| &= |\, ||\dot{u}(t)|| - ||\dot{u}(s)||\, | \\ &\leq ||\dot{u}(t) - \dot{u}(s)|| \\ &\leq ||\nabla \Phi(u(t)) - \nabla \Phi(u(s))|| \\ &\leq L_R ||u(t) - u(s)||. \end{aligned} \tag{17.11}$$

Let us complete the proof by showing that u is Lipschitz continuous on $[0,+\infty[$. Indeed, from

$$u(t)-u(s)=\int_s^t \dot{u}(\tau)d\tau$$

and the definition of the gradient flow, we obtain

$$\begin{aligned}\|u(t)-u(s)\| &\leq \int_s^t \|\dot{u}(\tau)\|d\tau \\ &\leq \int_s^t \|\nabla\Phi(u(\tau))\|d\tau.\end{aligned}$$

Moreover we have, for any $\tau \geq 0$,

$$\begin{aligned}\|\nabla\Phi(u(\tau))\| &\leq \|\nabla\Phi(u_0)\|+L_R\|u(\tau)-u_0\| \\ &\leq \|\nabla\Phi(u_0)\|+L_R(\|u_0\|+R).\end{aligned}$$

Setting $C_R := \|\nabla\Phi(u_0)\|+L_R(\|u_0\|+R)$, we deduce from the two above inequalities that

$$\|u(t)-u(s)\| \leq C_R|t-s|.$$

Returning to (17.11), we obtain that for all $s,t \geq 0$,

$$|g(t)-g(s)| \leq C_R L_R|t-s|.$$

By Lemma 17.1.2 we infer that $g(t)\to 0$ as $t \to +\infty$, which gives $\lim_{t\to+\infty} \dot{u}(t)=0$ and, by the gradient flow definition, $\lim_{t\to+\infty} \nabla\Phi(u(t))=0$. Since $\nabla\Phi$ is continuous, this clearly implies that the strong cluster points of the orbits are critical points of Φ. □

The following classical example from Palis and De Melo [315] shows that without any further geometrical assumption on Φ, bounded orbits of the gradient flow may fail to converge.

Let $\Phi : \mathbf{R}^2 \to \mathbf{R}$ be defined (in polar coordinates) by

$$\Phi(r\cos\theta, r\sin\theta)=\begin{cases}\exp\left(\frac{1}{r^2-1}\right) & \text{if } r<1,\\ 0 \quad \text{if } r=1, & \\ \exp\left(\frac{1}{r^2-1}\right)\sin\left(\frac{1}{r-1}-\theta\right) & \text{if } r>1.\end{cases}$$

Then Φ is C^1 and there exists an orbit of the gradient flow whose ω-limit set is the unit circle S^1. A similar example of such a "Mexican hat" function was given in [2]. Thus, in order to obtain the asymptotic convergence property of the orbits of the gradient flow, we need to make some additional assumptions. The above example suggests that we have to make some geometrical assumptions on Φ which prevent it from wild oscillations.

Because of their dissipative properties (existence of strict Lyapunov functions), gradient flows enjoy remarkable asymptotic convergence properties ($t \to +\infty$). Indeed, we are going to show that, in some important cases, their orbits converge to equilibria which are critical points (global minima in the convex case) of the potential Φ. More precisely, there are two important classes of functions Φ for which it has been established the convergence of the orbits of the gradient flow associated to Φ:

(i) the convex case (and related situations like quasi-convex),

(ii) the analytic case (and related situations like semialgebraic).

In the two next sections we are going to examine successively these situations.

17.2 ▪ The gradient flow associated to a convex potential

In many instances (PDEs, constrained problems, sparse representation in signal/image, optimization problems involving the total variation and BV spaces), one has to consider a potential Φ which is nonsmooth. A natural way to manage this situation is to use a regularization method. We thus reduce to the classical steepest descent. This approach has been particularly successful in the case of convex potentials. This is the situation that we now examine.

17.2.1 ▪ Moreau–Yosida approximation of nonsmooth convex functions

Let us consider a potential function $\Phi : \mathcal{H} \to \mathbf{R} \cup \{+\infty\}$ which is convex, lower semicontinuous, and proper: $\Phi \in \Gamma_0(\mathcal{H})$ for short. For $\Phi \in \Gamma_0(\mathcal{H})$, a natural extension of the notion of gradient is the notion of subdifferential, which we recall below. Given $u \in \operatorname{dom}\Phi$

$$z \in \partial\Phi(u) \Leftrightarrow \Phi(v) \geq \Phi(u) + \langle z, v-u\rangle \quad \forall v \in \mathcal{H}.$$

The classical steepest descent becomes a differential inclusion $-\dot{u}(t) \in \partial\Phi(u(t))$. The lack of continuity of the operator $\partial\Phi$ prevents a direct application of a general existence theorem for differential equations. A classical way to overcome this difficulty is to use the Moreau–Yosida regularization of the nonsmooth potential Φ. This technique is widely used in convex variational analysis, which is why we present a detailed study. The following statements respectively give its definition, regularization, and approximation properties.

Proposition 17.2.1. *Let $\Phi : \mathcal{H} \to \mathbf{R}\cup\{+\infty\}$ be a convex, lower semicontinuous, and proper function. For any $\lambda > 0$, the Moreau–Yosida approximation of index λ of Φ is the function $\Phi_\lambda : \mathcal{H} \to \mathbf{R}$ which is defined for all $u \in \mathcal{H}$ by*

$$\Phi_\lambda(u) = \inf_{v\in\mathcal{H}} \left\{ \Phi(v) + \frac{1}{2\lambda}\|u-v\|^2 \right\}. \tag{17.12}$$

1. *The infimum in* (17.12) *is attained at a unique point $J_\lambda u \in \mathcal{H}$, which satisfies*

$$\Phi_\lambda(u) = \Phi(J_\lambda u) + \frac{1}{2\lambda}\|u - J_\lambda u\|^2; \tag{17.13}$$

$$J_\lambda u + \lambda\partial\Phi(J_\lambda u) \ni u. \tag{17.14}$$

The operator $J_\lambda = (I + \lambda\partial\Phi)^{-1} : \mathcal{H} \to \mathcal{H}$ is everywhere defined and nonexpansive. It is called the resolvent of index λ of $A = \partial\Phi$.

2. *Φ_λ is convex, and continuously differentiable. For each $u \in \mathcal{H}$, its gradient is given by*

$$\nabla\Phi_\lambda(u) = \frac{1}{\lambda}(u - J_\lambda u). \tag{17.15}$$

3. *The operator*

$$A_\lambda = \frac{1}{\lambda}(I - J_\lambda) \tag{17.16}$$

is called the Yosida approximation of index λ of the maximal monotone operator $A = \partial\Phi$. It is Lipschitz continuous with Lipschitz constant $\frac{1}{\lambda}$. Thus, for all $u \in \mathcal{H}$

$$A_\lambda u = \nabla\Phi_\lambda(u), \tag{17.17}$$

$$A_\lambda u \in \partial\Phi(J_\lambda u). \tag{17.18}$$

PROOF. For $u \in \mathcal{H}$ fixed, the function $v \mapsto \Phi(v) + \frac{1}{2\lambda}\|u-v\|^2$ is strictly convex, lower semicontinuous, and coercive (indeed it is strongly convex). Therefore it reaches its minimal value at a unique point $J_\lambda u \in \mathcal{H}$ which satisfies

$$\Phi_\lambda(u) = \inf_{v \in \mathcal{H}} \left\{ \Phi(v) + \frac{1}{2\lambda}\|u-v\|^2 \right\} = \Phi(J_\lambda u) + \frac{1}{2\lambda}\|u - J_\lambda u\|^2,$$

and $\Phi_\lambda : \mathcal{H} \to \mathbf{R}$. By writing the first-order optimality condition, and using the subdifferential additivity rule of Moreau and Rockafellar, Theorem 9.5.4, we obtain

$$\partial\Phi(J_\lambda u) + \frac{1}{\lambda}(J_\lambda u - u) \ni 0, \tag{17.19}$$

which gives (17.14). Then notice that Φ_λ is the epi-sum of the two convex functions Φ and $\frac{1}{2\lambda}\|\cdot\|^2$:

$$\Phi_\lambda = \Phi \#_e \frac{1}{2\lambda}\|\|^2.$$

By Proposition 9.2.2, Φ_λ is a convex function.

Let us now prove that $J_\lambda = (I + \lambda\partial\Phi)^{-1} : \mathcal{H} \to \mathcal{H}$ is nonexpansive. Take $u, v \in \mathcal{H}$. By (17.19),

$$\frac{1}{\lambda}(u - J_\lambda u) \in \partial\Phi(J_\lambda u),$$
$$\frac{1}{\lambda}(v - J_\lambda v) \in \partial\Phi(J_\lambda v),$$

and the monotonicity property of $\partial\Phi$ (see Proposition 17.2.3), we obtain

$$\left\langle \frac{1}{\lambda}(u - J_\lambda u) - \frac{1}{\lambda}(v - J_\lambda v), J_\lambda u - J_\lambda v \right\rangle \geq 0. \tag{17.20}$$

Equivalently

$$\langle J_\lambda u - J_\lambda v, u - v\rangle \geq \|J_\lambda u - J_\lambda v\|^2,$$

which by the Cauchy–Schwarz inequality gives

$$\|J_\lambda u - J_\lambda v\| \leq \|u - v\|.$$

Set $A_\lambda = \frac{1}{\lambda}(I - J_\lambda)$. Reformulating (17.20) with A_λ, we obtain

$$\langle A_\lambda u - A_\lambda v, (u - \lambda A_\lambda u) - (v - \lambda A_\lambda v)\rangle \geq 0.$$

Equivalently

$$\langle A_\lambda u - A_\lambda v, u - v\rangle \geq \lambda\|A_\lambda u - A_\lambda v\|^2,$$

which, by the Cauchy–Schwarz inequality, implies that A_λ is Lipschitz continuous with Lipschitz constant $\frac{1}{\lambda}$:

$$\|A_\lambda u - A_\lambda v\| \leq \frac{1}{\lambda}\|u - v\|.$$

Let us now prove that Φ_λ is differentiable, with $\nabla\Phi_\lambda(u) = A_\lambda u$, for any $u \in \mathcal{H}$. In order to verify the Fréchet differentiability of Φ_λ at $u \in \mathcal{H}$, set, for any $v \in \mathcal{H}$,

$$P_\lambda(v) := \Phi_\lambda(v) - \Phi_\lambda(u) - \langle A_\lambda u, v-u\rangle, \tag{17.21}$$

and prove that $P_\lambda(v) = o(\|v-u\|)$. According to the convexity of Φ_λ, let us first prove that $P_\lambda(v) \geq 0$. By definition (17.21) of P_λ, definition (17.16) of A_λ, and (17.13)

$$\begin{aligned} P_\lambda(v) &= \left(\Phi(J_\lambda v) + \frac{\lambda}{2}\|A_\lambda v\|^2\right) - \left(\Phi(J_\lambda u) + \frac{\lambda}{2}\|A_\lambda u\|^2\right) - \langle A_\lambda(u), v-u\rangle \\ &= (\Phi(J_\lambda v) - \Phi(J_\lambda u)) + \frac{\lambda}{2}\left(\|A_\lambda v\|^2 - \|A_\lambda u\|^2\right) - \langle A_\lambda(u), v-u\rangle. \end{aligned} \tag{17.22}$$

By $A_\lambda u \in \partial\Phi(J_\lambda u)$ (see (17.18)), we have the convex subdifferential inequality

$$\Phi(J_\lambda v) - \Phi(J_\lambda u) \geq \langle A_\lambda u, J_\lambda v - J_\lambda u\rangle. \tag{17.23}$$

Let us successively combine (17.22) and (17.23), then use $J_\lambda v = v - \lambda A_\lambda v$, $J_\lambda u = u - \lambda A_\lambda u$ to obtain

$$\begin{aligned} P_\lambda(v) &\geq \langle A_\lambda u, J_\lambda v - J_\lambda u\rangle + \frac{\lambda}{2}\left(\|A_\lambda v\|^2 - \|A_\lambda u\|^2\right) - \langle A_\lambda u, v-u\rangle \\ &\geq \langle A_\lambda u, v-u\rangle + \lambda\langle A_\lambda u, A_\lambda u - A_\lambda v\rangle + \frac{\lambda}{2}\left(\|A_\lambda v\|^2 - \|A_\lambda u\|^2\right) - \langle A_\lambda u, v-u\rangle. \end{aligned}$$

After simplification

$$\begin{aligned} P_\lambda(v) &\geq \frac{\lambda}{2}\|A_\lambda u\|^2 + \frac{\lambda}{2}\|A_\lambda v\|^2 - \lambda\langle A_\lambda u, A_\lambda v\rangle \\ &\geq \frac{\lambda}{2}\|A_\lambda u - A_\lambda v\|^2 \geq 0. \end{aligned} \tag{17.24}$$

The above argument being valid for any $u, v \in \mathcal{H}$, by reversing the role of u and v, we obtain

$$\Phi_\lambda(u) - \Phi_\lambda(v) - \langle A_\lambda v, u-v\rangle \geq 0. \tag{17.25}$$

Equivalently

$$\Phi_\lambda(v) - \Phi_\lambda(u) - \langle A_\lambda v, v-u\rangle \leq 0. \tag{17.26}$$

Using successively the definition of $P_\lambda(v)$, (17.26), the Cauchy–Schwarz inequality, and the Lipschitz continuity with Lipschitz constant $\frac{1}{\lambda}$ of A_λ

$$\begin{aligned} P_\lambda(v) &= (\Phi_\lambda(v) - \Phi_\lambda(u) - \langle A_\lambda v, v-u\rangle) + \langle A_\lambda v - A_\lambda u, v-u\rangle \\ &\leq \langle A_\lambda v - A_\lambda u, v-u\rangle \\ &\leq \|A_\lambda v - A_\lambda u\|\|v-u\| \\ &\leq \frac{1}{\lambda}\|v-u\|^2. \end{aligned} \tag{17.27}$$

Combining (17.24) and (17.27) gives, for any $v \in \mathcal{H}$,

$$0 \leq \Phi_\lambda(v) - \Phi_\lambda(u) - \langle A_\lambda u, v-u\rangle \leq \frac{1}{\lambda}\|v-u\|^2,$$

which shows that Φ_λ is differentiable, with gradient at u being equal to $A_\lambda u$. Since A_λ is continuous this proves that Φ_λ is continuously differentiable. □

Remark 17.2.1. Just like for classical convolution, it is the regularity of the quadratic kernel $\|\cdot\|^2$ which confers to Φ_λ its regularity property. But, by contrast with classical convolution, in general we cannot expect more regularity than $C^{1,1}$ for Φ_λ, i.e., C^1 with a Lipschitz continuous gradient. Take, for example, $\mathcal{H} = \mathbf{R}$ and Φ equal to the indicator function of $(-\infty, 0]$. Then $\Phi_\lambda(r) = \frac{1}{2\lambda}(r^+)^2$, which is $C^{1,1}$ but not C^2.

We complement Proposition 17.2.1 by showing the following approximation results.

Proposition 17.2.2. *Let $\Phi : \mathcal{H} \to \mathbf{R} \cup \{+\infty\}$ be convex, lower semicontinuous, and proper. Then we have the following:*

(i) Monotone convergence:

$$\Phi_\lambda(u) \uparrow \Phi(u) \quad \textit{as } \lambda \downarrow 0 \ \forall u \in \mathcal{H}.$$

(ii) Convergence of resolvents:

$$J_\lambda u \to u \quad \textit{as } \lambda \to 0 \ \forall u \in \overline{\operatorname{dom}\Phi};$$
$$J_\lambda u \to \operatorname{proj}_{\overline{\operatorname{dom}\Phi}} u \quad \textit{as } \lambda \to 0 \ \forall u \in \mathcal{H}.$$

(iii) Convergence of Yosida approximation: *For all $u \in \operatorname{dom} \partial\Phi$*

$$A_\lambda u \to \partial\Phi(u)^0 \qquad \textit{as } \lambda \to 0; \tag{17.28}$$
$$\|A_\lambda u\| \le \|\partial\Phi(u)^0\| \quad \forall \lambda > 0. \tag{17.29}$$

PROOF. (i) Let us fix $u \in \mathcal{H}$. By definition (17.12) of Φ_λ, for any $v \in \mathcal{H}$,

$$\Phi_\lambda(u) \le \Phi(v) + \frac{1}{2\lambda}\|u - v\|^2. \tag{17.30}$$

Taking $v = u$ in (17.30) we obtain

$$\Phi_\lambda(u) \le \Phi(u). \tag{17.31}$$

As λ decreases, the sequence of functions $v \mapsto \Phi(v) + \frac{1}{2\lambda}\|u - v\|^2$ increases, as well as the sequence of its minimal values. Hence $\lim_{\lambda\to 0} \Phi_\lambda(u)$ exists, which by (17.31) gives

$$\lim_{\lambda\to 0} \Phi_\lambda(u) \le \Phi(u). \tag{17.32}$$

To show the opposite inequality, we successively consider the cases $u \in \operatorname{dom}\Phi$, $u \in \overline{\operatorname{dom}\Phi}$, and $u \notin \overline{\operatorname{dom}\Phi}$.

(a) $u \in \operatorname{dom}\Phi$. By considering a continuous affine minorant of Φ, we obtain the existence of some positive constant c such that

$$\Phi(v) \ge -c(1 + \|v\|) \quad \forall v \in \mathcal{H}. \tag{17.33}$$

By (17.31), the definition of Φ_λ, and (17.33),

$$\begin{aligned}\Phi(u) \ge \Phi_\lambda(u) &= \Phi(J_\lambda u) + \frac{1}{2\lambda}\|u - J_\lambda u\|^2 \\ &\ge -c(1 + \|J_\lambda u\|) + \frac{1}{2\lambda}\|u - J_\lambda u\|^2. \end{aligned} \tag{17.34}$$

By using the triangle inequality in (17.34),

$$\|u-J_\lambda u\|^2-2\lambda c\|u-J_\lambda u\|-2\lambda(c+c\|u\|+\Phi(u))\le 0,$$

which after elementary computation gives

$$\|u-J_\lambda u\|\le 2\lambda c+\sqrt{2\lambda(c+c\|u\|+\Phi(u))}.$$

Since $u\in\operatorname{dom}\Phi$, it follows that $J_\lambda u\to u$ as $\lambda\to 0$. By

$$\Phi_\lambda(u)\ge\Phi(J_\lambda u),$$

and by lower semicontinuity of Φ we deduce that

$$\begin{aligned}\lim_{\lambda\to 0}\Phi_\lambda(u)&\ge\liminf_{\lambda\to 0}\Phi(J_\lambda u)\\&\ge\Phi(u).\end{aligned}\tag{17.35}$$

Combining (17.32) and (17.35) gives $\lim_{\lambda\to 0}\Phi_\lambda(u)=\Phi(u)$ for $u\in\operatorname{dom}\Phi$.

(b) $u\in\overline{\operatorname{dom}\Phi}$. Let us show that $J_\lambda u\to u$ as $\lambda\to 0$ still holds true. Take $v\in\operatorname{dom}\Phi$. By the triangle inequality and the nonexpansive property of J_λ,

$$\begin{aligned}\|J_\lambda u-u\|&\le\|J_\lambda u-J_\lambda v\|+\|J_\lambda v-v\|+\|v-u\|\\&\le 2\|v-u\|+\|J_\lambda v-v\|.\end{aligned}\tag{17.36}$$

Letting $\lambda\to 0$, and since $v\in\operatorname{dom}\Phi$

$$\limsup_{\lambda\to 0}\|J_\lambda u-u\|\le 2\|v-u\|.$$

This being true for any $v\in\operatorname{dom}\Phi$, we can take v arbitrarily close to u, which gives

$$J_\lambda u\to u\quad\forall u\in\overline{\operatorname{dom}\Phi}.\tag{17.37}$$

As in case (a), we complete the argument by using the lower semicontinuity property of Φ.

(c) $u\notin\overline{\operatorname{dom}\Phi}$. Since $J_\lambda u\in\operatorname{dom}\Phi$, we have

$$\|u-J_\lambda u\|\ge\operatorname{dist}(u,\overline{\operatorname{dom}\Phi}):=\gamma>0.$$

On the other hand, by the nonexpansive property, $J_\lambda u$ remains bounded: taking some $u_0\in\operatorname{dom}\Phi$, we have $J_\lambda u_0\to u_0$, and

$$\|J_\lambda u\|\le\|J_\lambda u_0\|+\|u-u_0\|\le M<+\infty.\tag{17.38}$$

By (17.33) and (17.38)

$$\begin{aligned}\Phi_\lambda(u)&\ge -c(1+\|J_\lambda u\|)+\frac{1}{2\lambda}\|u-J_\lambda u\|^2\\&\ge -c(1+M)+\frac{\gamma^2}{2\lambda}.\end{aligned}$$

Thus $\lim_{\lambda\to 0}\Phi_\lambda(u)=+\infty=\Phi(u)$.

(ii) In the process, in (17.37), we have obtained that $J_\lambda u \to u$ for any $u \in \overline{\operatorname{dom}\Phi}$. Let us complete this result by examining the convergence of $(J_\lambda u)$ when $u \notin \overline{\operatorname{dom}\Phi}$. By (17.38), $(J_\lambda u)$ remains bounded. Let ξ be a weak cluster point of the generalized sequence $(J_\lambda u)$. For simplicity we write

$$J_\lambda u \rightharpoonup \xi \quad \text{weakly in } \mathscr{H}.$$

Since $J_\lambda u \in \operatorname{dom}\partial\Phi \subset \overline{\operatorname{dom}\Phi}$, by the weak closedness property of the closed convex set $\overline{\operatorname{dom}\Phi}$, we have $\xi \in \overline{\operatorname{dom}\Phi}$.

By (17.18), we have $\frac{1}{\lambda}(u - J_\lambda u) \in \partial\Phi(J_\lambda u)$. By the subdifferential inequality, for any $v \in \operatorname{dom}\Phi$,

$$\Phi(v) \geq \Phi(J_\lambda u) + \frac{1}{\lambda}\langle u - J_\lambda u, v - J_\lambda u\rangle.$$

Using a continuous affine minorant of Φ (see (17.33)) and developing the above expression we obtain

$$\lambda\Phi(v) \geq -\lambda c(1 + \|J_\lambda u\|) + \langle u - J_\lambda u, v\rangle - \langle J_\lambda u, u\rangle + \|J_\lambda u\|^2. \tag{17.39}$$

Passing to the limit as $\lambda \to 0$, and using the lower semicontinuity for the weak topology of the convex continuous function $z \mapsto \|J_\lambda z\|^2$, we obtain

$$\langle u - \xi, v - \xi\rangle \leq 0 \quad \forall v \in \operatorname{dom}\Phi.$$

This inequality can be extended by continuity to $v \in \overline{\operatorname{dom}\Phi}$, which gives

$$\begin{cases} \langle u - \xi, v - \xi\rangle \leq 0 \quad \forall v \in \overline{\operatorname{dom}\Phi}, \\ \xi \in \overline{\operatorname{dom}\Phi}. \end{cases}$$

This is the obtuse angle condition which characterizes $\xi = \operatorname{proj}_{\overline{\operatorname{dom}\Phi}} u$. By uniqueness of the weak sequential cluster point we deduce that the whole generalized sequence $J_\lambda u$ converges weakly to $\operatorname{proj}_{\overline{\operatorname{dom}\Phi}} u$. Strong convergence is obtained by showing the convergence of the norms. To simplify notation set $p(u) = \operatorname{proj}_{\overline{\operatorname{dom}\Phi}} u$. Returning to (17.39), by taking the lim sup as $\lambda \to 0$, we obtain

$$0 \geq \langle u - p(u), v\rangle - \langle p(u), u\rangle + \limsup \|J_\lambda u\|^2 \quad \forall v \in \operatorname{dom}\Phi.$$

This inequality can be extended by continuity to $v \in \overline{\operatorname{dom}\Phi}$. By taking $v = p(u)$ we obtain

$$0 \geq \langle u - p(u), p(u)\rangle - \langle p(u), u\rangle + \limsup \|J_\lambda u\|^2,$$

that is,

$$\|p(u)\|^2 \geq \limsup \|J_\lambda u\|^2,$$

which completes the proof of (ii).

(iii) Fix $u \in \operatorname{dom}\partial\Phi$. By (17.18), we have $A_\lambda u \in \partial\Phi(J_\lambda u)$. By the monotonicity property of $\partial\Phi$, we deduce that, for any $z \in \partial\Phi(u)$,

$$\langle A_\lambda u - z, J_\lambda u - u\rangle \geq 0.$$

Since $J_\lambda u - u = -\lambda A_\lambda u$, we obtain

$$\|A_\lambda u\|^2 \leq \langle A_\lambda u, z\rangle.$$

Hence, by the Cauchy–Schwarz inequality

$$||A_\lambda u|| \leq ||z|| \quad \forall z \in \partial\Phi(u). \tag{17.40}$$

Since $\partial\Phi(u)$ is a closed convex nonempty set, it has a unique element of minimal norm $\partial\Phi(u)^0$, and (17.40) gives

$$||A_\lambda u|| \leq ||\partial\Phi(u)^0||. \tag{17.41}$$

Hence $(A_\lambda u)$ remains bounded. Let η be a weak sequential cluster point of the generalized sequence $(A_\lambda u)$. For simplicity we write

$$A_\lambda u \rightharpoonup \eta \quad \text{weakly in } \mathscr{H}.$$

By passing to the limit on the subdifferential inequality,

$$\Phi(v) \geq \Phi(J_\lambda u) + \langle A_\lambda u, v - J_\lambda u\rangle \quad \forall v \in \operatorname{dom}\Phi,$$

and using that $J_\lambda u \to u$, we obtain

$$\Phi(v) \geq \Phi(u) + \langle \eta, v - u\rangle \quad \forall v \in \operatorname{dom}\Phi,$$

that is, $\eta \in \partial\Phi(u)$. On the other hand, by (17.41)

$$||\eta|| \leq \liminf_{\lambda\to 0} ||A_\lambda u|| \leq ||\partial\Phi(u)^0||. \tag{17.42}$$

By $\eta \in \partial\Phi(u)$ and (17.42) we conclude that $\eta = \partial\Phi(u)^0$. By uniqueness of its weak sequential cluster point, the whole sequence $(A_\lambda u)$ converges weakly to $\partial\Phi(u)^0$. Moreover, by (17.41)

$$\limsup_{\lambda\to 0} ||A_\lambda u|| \leq ||\partial\Phi(u)^0||.$$

Weak convergence and convergence of the norms imply strong convergence, which completes (iii). □

Remark 17.2.2. Let us mention some related properties of the Moreau–Yosida approximation.

(i) By Proposition 17.2.2, $J_\lambda u \to u$ as $\lambda \to 0$ for all $u \in \overline{\operatorname{dom}\Phi}$. Since $J_\lambda u \in \operatorname{dom}\partial\Phi$, this property implies that the domain of $\partial\Phi$ is dense in the domain of Φ, i.e.,

$$\overline{\operatorname{dom}\partial\Phi} = \overline{\operatorname{dom}\Phi}.$$

(ii) Since $\Phi_\lambda = \Phi \#_e \frac{1}{2\lambda}||\cdot||^2$, the Legendre–Fenchel calculus gives

$$\Phi_\lambda^* = \Phi^* + \frac{\lambda}{2}||\cdot||^2.$$

Thus, for any $\lambda > 0,\ \mu > 0$,

$$\begin{aligned}((\Phi_\lambda)_\mu)^* &= (\Phi_\lambda)^* + \frac{\mu}{2}||\cdot||^2\\ &= \Phi^* + \frac{\lambda}{2}||\cdot||^2 + \frac{\mu}{2}||\cdot||^2\\ &= \Phi^* + \frac{\lambda+\mu}{2}||\cdot||^2\\ &= (\Phi_{\lambda+\mu})^*.\end{aligned}$$

Hence

$$(\Phi_\lambda)_\mu = \Phi_{\lambda+\mu}. \tag{17.43}$$

This relation defines a semiflow $\lambda \mapsto \Phi_\lambda$, which suggests interpreting λ as a time variable. We shall confirm this interpretation when considering the discrete time version of our dynamics. Indeed (see [37, Remark 3.32]), one can prove that

$$\frac{d}{d\lambda}\Phi_\lambda(u) = -\frac{1}{2}\|\nabla\Phi_\lambda(u)\|^2.$$

Thus, $(t,u) \mapsto w(t,u) = \Phi_t(u)$ is a solution of the Hamilton–Jacobi equation

$$\begin{cases} w_t + \frac{1}{2}\|\nabla_u w\|^2 = 0; \\ w(0,u) = \Phi(u). \end{cases} \tag{17.44}$$

It is the viscosity solution of this equation; $w(t,u) = \Phi_t(u)$, with

$$\Phi_t(u) = \inf_{v\in\mathcal{H}} \left\{ \Phi(v) + \frac{1}{2t}\|u-v\|^2 \right\},$$

is known as the Lax or Hopf formula for the viscosity solution of (17.44).

17.2.2 ▪ Gradient flow for a convex lower semicontinuous potential on a Hilbert space: Existence and uniqueness results

Let us consider a potential function $\Phi \in \Gamma_0(\mathcal{H})$, i.e., $\Phi : \mathcal{H} \to \mathbf{R} \cup \{+\infty\}$ is convex, lower semicontinuous, and proper. This is an important situation that is preparatory to the study of gradient flows associated with general nonsmooth potentials. For $\Phi \in \Gamma_0(\mathcal{H})$, a natural extension of the notion of gradient is the notion of subdifferential. The operator $\partial\Phi : \mathcal{H} \to 2^{\mathcal{H}}$ is multivalued, with domain $\operatorname{dom}\partial\Phi = \{u \in \operatorname{dom}\Phi : \partial\Phi(u) \neq \emptyset\}$ which is a dense subset of the domain of Φ. Thus we are led to consider the differential inclusion

$$\text{(GSD)} \quad \dot{u}(t) + \partial\Phi(u(t)) \ni 0,$$

called the generalized steepest descent. When Φ is differentiable, $\partial\Phi = \nabla\Phi$, and we recover the classical (SD) equation.

As a model situation, take $\mathcal{H} = L^2(\Omega)$, where Ω is a regular bounded open set in $\mathbf{R}^n$, and

$$\Phi(v) = \begin{cases} \frac{1}{2}\int_\Omega \|\nabla v(x)\|^2 dx & \text{if } v \in H_0^1(\Omega), \\ +\infty & \text{if } v \in L^2(\Omega),\ v \notin H_0^1(\Omega). \end{cases}$$

Then $A = \partial\Phi : L^2(\Omega) \to L^2(\Omega)$ is the (minus) Laplace operator

$$\begin{cases} \operatorname{dom} A = H^2(\Omega) \cap H_0^1(\Omega), \\ A(v) = -\Delta v \quad \text{for } v \in \operatorname{dom}(A), \end{cases}$$

and (GSD) is the heat equation (with Dirichlet boundary condition)

$$\frac{\partial u}{\partial t} - \Delta u = 0.$$

Let us now return to (GSD) for a general potential $\Phi \in \Gamma_0(\mathcal{H})$ and note that (GSD) is a particular instance of evolution equations governed by maximal monotone operators. Let us recall some basic notions concerning maximal monotone operators that will be useful for our study; see [66], [85], [135], [364] for an extended presentation. It is convenient to identify an operator $A : \mathcal{H} \to 2^{\mathcal{H}}$ with its graph. We equivalently write $z \in A(u)$ or $(u,z) \in A$.

Definition 17.2.1. *Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be an operator.*

(i) *A is monotone if for all $u_1,\ u_2 \in \operatorname{dom} A$, for all $z_1 \in Au_1,\ z_2 \in Au_2$,*

$$\langle z_2 - z_1, u_2 - u_1 \rangle \geq 0. \tag{17.45}$$

(ii) *A is maximal monotone if it is monotone, and there is no proper monotone extension of A, i.e.,*

$$\langle \bar{z} - z, \bar{u} - u \rangle \geq 0 \quad \forall (u,z) \in A \ \Rightarrow \bar{z} \in A(\bar{u}).$$

Maximality holds in the class of monotone operators with respect to the inclusion relation on the graphs: it is not possible to extend the graph of a maximal monotone operator A into the graph of a monotone operator which is strictly larger than A. Subdifferentials of closed convex functions provide an important subclass of maximal monotone operators.

Proposition 17.2.3. *Let $\Phi : \mathcal{H} \to \mathbf{R} \cup \{+\infty\}$ be a convex, lower semicontinuous, and proper function. Then, its subdifferential $A = \partial\Phi : \mathcal{H} \to 2^{\mathcal{H}}$ is a maximal monotone operator.*

PROOF. Let $z_1 \in \partial\Phi(u_1)$, $z_2 \in \partial\Phi(u_2)$. Adding the subdifferential inequalities

$$\Phi(u_2) \geq \Phi(u_1) + \langle z_1, u_2 - u_1 \rangle,$$

$$\Phi(u_1) \geq \Phi(u_2) + \langle z_2, u_1 - u_2 \rangle,$$

gives (17.45), and the monotonicity of $A = \partial\Phi$.

The maximal monotonicity of $A = \partial\Phi$ is a direct consequence of the fact that the operator $I + A$ is surjective, i.e., $R(I + A) = \mathcal{H}$; see Proposition 17.2.1(1). Suppose that there exists B monotone, $B \supset A$, and $z \in Bu$. Since $R(I + A) = \mathcal{H}$ there exists some $\tilde{u} \in \operatorname{dom} A$ such that

$$\tilde{u} + A(\tilde{u}) \ni u + z. \tag{17.46}$$

Since $B \supset A$, we have $B(\tilde{u}) \supset A(\tilde{u})$, and (17.46) gives

$$\tilde{u} + B(\tilde{u}) \ni u + z.$$

On the other hand, since $z \in Bu$

$$u + B(u) \ni u + z.$$

Comparing these two last relations, by the contraction property of $(I + B)^{-1}$ (a direct consequence of the monotonicity of B), we obtain $u = \tilde{u}$. Returning to (17.46), we obtain $z \in A(u)$, which proves that $A = \partial\Phi$ is maximal monotone. $\square$

In the above argument, the maximal monotonicity of $A = \partial\Phi$ has been obtained as a consequence of the monotonicity of $\partial\Phi$ and of the surjectivity of $I + \partial\Phi$. Indeed, this is a particular instance of the following theorem, due to Minty (see [292]), which provides a very useful characterization of maximal monotone operators.

Theorem 17.2.1 (Minty). *Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be a monotone operator. Then the following equivalence holds:*

$$A \text{ is maximal monotone} \Longleftrightarrow R(I+A)=\mathcal{H}.$$

PROOF. A detailed proof of this nontrivial result can be found in [66, Theorem 5, Chapter 6, Section 7], [85, Theorem 21.1], [135]. □

A maximal monotone operator, and in particular the subdifferential of a function $\Phi \in \Gamma_0(\mathcal{H})$, is demiclosed, a property that is useful, and is described below.

Proposition 17.2.4. *Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be a maximal monotone operator. Then A is demiclosed, i.e.,*

$$(u_n \to u \ \text{ strongly},\ z_n \to z \ \text{ weakly},\ z_n \in A(u_n)\ \forall\, n \in \mathbf{N}) \Longrightarrow (z \in A(u)).$$

PROOF. By monotonicity of A, for any $w \in A(v)$, for all $n \in \mathbf{N}$

$$\langle z_n - w, u_n - v\rangle \geq 0.$$

Thanks to the respective strong and weak convergence properties of $u_n - v$ and $z_n - w$, passing to the limit gives

$$\langle z - w, u - v\rangle \geq 0.$$

This being true for any $(v,w) \in A$, by maximal monotonicity of A, we conclude that $z \in Au$. □

Note that in accordance with the theory of evolution equations governed by maximal monotone operators, in general we cannot expect to obtain classical global solutions of (GSD). Take, for example, $\mathcal{H} = \mathbf{R}$ and $\Phi(v) = v^+$. Given Cauchy data $u_0 > 0$, it can be easily seen that the unique solution of (GSD) is the function

$$u(t) = \begin{cases} u_0 - t & \text{for } 0 \leq t \leq u_0, \\ 0 & \text{for } t \geq u_0. \end{cases}$$

Intuitively, the drop of water $(u(t), \Phi(u(t))) \in \mathbf{R}^2$ slides along the oblique line $y = x$ at constant speed, and at time $t = u_0$ it reaches the origin. Then, it stops. The resulting global solution u is not differentiable at $t = u_0$. It is only Lipschitz continuous on $\mathbf{R}$.

In order to define the notion of strong solution for (GSD), let us first recall some notions concerning vector-valued functions of real variables. (See [135, Appendix] for more details.)

Definition 17.2.2. *Given $T \in \mathbf{R}^+$, a function $f : [0,T] \longrightarrow \mathcal{H}$ is said to be absolutely continuous if one of the following equivalent properties holds:*

(i) *There exists an integrable function $g : [0,T] \longrightarrow \mathcal{H}$ such that*

$$f(t) = f(0) + \int_0^t g(s)\,ds \qquad \forall t \in [0,T].$$

(ii) *f is continuous and its distributional derivative belongs to the Lebesgue space*

$$L^1([0,T];\mathcal{H}).$$

(iii) *For every $\epsilon > 0$, there exists some $\eta > 0$ such that for any finite family of intervals $I_k = (a_k, b_k)$*

$$I_k \cap I_j = \emptyset \text{ for } i \neq j \quad \text{and} \quad \sum_k |b_k - a_k| \leq \eta \Longrightarrow \sum_k \|f(b_k) - f(a_k)\| \leq \epsilon.$$

Moreover, an absolutely continuous function is differentiable almost everywhere, its pointwise derivative coincides with its distributional derivative (a.e.), and one can recover the function from its derivative $f' = g$ using the integration formula (i).

Definition 17.2.3. *We say that $u(\cdot)$ is a strong global solution of (GSD) if* (i) *and* (ii) *hold:*

(i) *$u : [0, +\infty) \to \mathcal{H}$ is continuous and is absolutely continuous on each interval $[0, T]$, $T < \infty$;*

(ii) *for almost all $t > 0$, $u(t) \in \operatorname{dom} \partial\Phi$, and*

$$-\dot{u}(t) \in \partial\Phi(u(t)).$$

Theorem 17.2.2. *Let $\Phi : \mathcal{H} \to \mathbf{R} \cup \{+\infty\}$ be a convex, lower semicontinuous, and proper function. Supposed that Φ is minorized, i.e., $\inf_{\mathcal{H}} \Phi > -\infty$. Then, for any $u_0 \in \operatorname{dom} \partial\Phi$, there exists a unique strong global solution $u : [0, +\infty) \to \mathcal{H}$ of the Cauchy problem*

$$\begin{cases} \dot{u}(t) + \partial\Phi(u(t)) \ni 0, \\ u(0) = u_0. \end{cases} \tag{17.47}$$

Moreover the following properties hold:

(i) *$u(t) \in \operatorname{dom} \partial\Phi$ for all $t \geq 0$.*

(ii) *$\dot{u} \in L^2(0, +\infty; \mathcal{H}) \cap L^\infty(0, +\infty; \mathcal{H})$; in particular u is Lipschitz continuous on $[0, +\infty)$.*

(iii) *For each $t \geq 0$, u has a right derivative, and*

$$\frac{d^+ u}{dt}(t) = -\partial\Phi(u(t))^0,$$

where $\partial\Phi(u(t))^0$ is the element of minimal norm of $\partial\Phi(u(t))$.

(iv) *$t \mapsto \|\frac{d^+ u}{dt}(t)\|$ is nonincreasing.*

(v) *$t \mapsto \Phi(u(t))$ is nonincreasing, absolutely continuous on each bounded interval $[0, T]$, and*

$$\frac{d}{dt}\Phi(u(t)) = -\|\dot{u}(t)\|^2 \quad \text{for almost all } t > 0. \tag{17.48}$$

PROOF. A. Uniqueness is a direct consequence of the monotonicity property of $\partial\Phi$. Suppose u and v are two strong global solutions of (GSD), associated with the respective Cauchy data u_0 and v_0. Since $-\dot{u}(t) \in \partial\Phi(u(t))$ and $-\dot{v}(t) \in \partial\Phi(v(t))$, by monotonicity of $\partial\Phi$

$$\langle -\dot{u}(t) + \dot{v}(t), u(t) - v(t) \rangle \geq 0. \tag{17.49}$$

One can easily verify that the function $t \mapsto \|u(t) - v(t)\|^2$ is absolutely continuous on bounded intervals. By (17.49), it satisfies for almost every $t \geq 0$

$$\frac{d}{dt}\|u(t) - v(t)\|^2 \leq 0.$$

As a consequence, $t \mapsto \|u(t)-v(t)\|$ is a decreasing function, and

$$\|u(t)-v(t)\| \le \|u(0)-v(0)\| \quad \forall t \ge 0. \tag{17.50}$$

For a given Cauchy data we have $u(0)=v(0)$, which gives uniqueness of the solution.

B. *Existence*. (a) For each $\lambda>0$, consider the Cauchy problem obtained by replacing $A=\partial\Phi$ by its Yosida approximation $A_\lambda = \nabla\Phi_\lambda$. By Proposition 17.2.1, the operator $\nabla\Phi_\lambda : \mathcal{H} \to \mathcal{H}$ is Lipschitz continuous with Lipschitz constant $\frac{1}{\lambda}$. Therefore, according to the Cauchy–Lipschitz theorem 17.1.2 (global version), there is a unique global classical solution $u_\lambda : [0,+\infty) \to \mathcal{H}$ of the Cauchy problem

$$\begin{cases} \dot u_\lambda(t) + \nabla\Phi_\lambda(u_\lambda(t)) = 0, \\ u_\lambda(0) = u_0. \end{cases} \tag{17.51}$$

(b) Let us establish estimations on the sequence (u_λ). Energy estimate (17.7) gives

$$\int_0^\infty \|\dot u_\lambda(t)\|^2 dt \le \Phi_\lambda(u_0) - \inf_{\mathcal{H}} \Phi_\lambda.$$

Since $\Phi_\lambda \le \Phi$, and $\inf_{\mathcal{H}} \Phi_\lambda = \inf_{\mathcal{H}} \Phi$ (a direct consequence of the definition of Φ_λ), it follows that

$$\int_0^\infty \|\dot u_\lambda(t)\|^2 dt \le \Phi(u_0) - \inf_{\mathcal{H}} \Phi.$$

Due to the autonomous nature of the differential equation in (17.51), for any positive real number $h>0$, $t \mapsto u_\lambda(t)$ and $t \mapsto u_\lambda(t+h)$ are solutions of this differential equation. Using the argument developed in the proof of uniqueness, we obtain that $t \mapsto \|u_\lambda(t+h)-u_\lambda(t)\|$ is a decreasing function. Hence, for any $0 \le s \le t$

$$\|u_\lambda(t+h)-u_\lambda(t)\| \le \|u_\lambda(s+h)-u_\lambda(s)\|.$$

Dividing by $h>0$ and letting $h \to 0$ gives

$$\|\dot u_\lambda(t)\| \le \|\dot u_\lambda(s)\|. \tag{17.52}$$

Hence $t \mapsto \|\dot u_\lambda(t)\|$ is a decreasing function. In particular, for any $t \ge 0$

$$\|\dot u_\lambda(t)\| \le \|\dot u_\lambda(0)\| = \|\nabla\Phi_\lambda(u_0)\|.$$

By (17.29), and $u_0 \in \operatorname{dom}\partial\Phi$, we have $\|\nabla\Phi_\lambda(u_0)\| \le \|\partial\Phi(u_0)^0\|$. Hence

$$\sup_{\lambda>0,\ t>0} \|\dot u_\lambda(t)\| \le \|\partial\Phi(u_0)^0\|. \tag{17.53}$$

(c) Let us show that for all $T>0$, (u_λ) converges uniformly on $[0,T]$ as $\lambda \to 0$. To that end, fix arbitrary $T>0$, and prove that (u_λ) is a Cauchy sequence in the Banach space $\mathcal{C}([0,T];\mathcal{H})$, equipped with the supremum norm. Take $\lambda>0$, $\mu>0$, and consider the corresponding solutions u_λ, u_μ of the Cauchy problems

$$\begin{cases} \dot u_\lambda(t) + \nabla\Phi_\lambda(u_\lambda(t)) = 0, \quad u_\lambda(0)=u_0, \\ \dot u_\mu(t) + \nabla\Phi_\mu(u_\mu(t)) = 0, \quad u_\mu(0)=u_0. \end{cases}$$

Set $h(t) := \frac{1}{2}\|u_\lambda(t) - u_\mu(t)\|^2$. We have (recall that $A_\lambda = \nabla\Phi_\lambda, A_\mu = \nabla\Phi_\mu$)

$$\begin{aligned} \dot{h}(t) &= \langle u_\lambda(t) - u_\mu(t), \dot{u}_\lambda(t) - \dot{u}_\mu(t)\rangle \\ &= \langle u_\lambda(t) - u_\mu(t), -A_\lambda u_\lambda(t) + A_\mu u_\mu(t)\rangle. \end{aligned} \tag{17.54}$$

From now on, to simplify notation, we omit the variable t and write u_λ for $u_\lambda(t)$.

By (17.15) and (17.16) we have $u_\lambda = J_\lambda u_\lambda + \lambda A_\lambda u_\lambda$ and $u_\mu = J_\mu u_\mu + \mu A_\mu u_\mu$. Replacing u_λ and u_μ by these expressions in (17.54) gives

$$\dot{h}(t) + \langle J_\lambda u_\lambda - J_\mu u_\mu, A_\lambda u_\lambda - A_\mu u_\mu\rangle + \langle A_\lambda u_\lambda - A_\mu u_\mu, \lambda A_\lambda u_\lambda - \mu A_\mu u_\mu\rangle \leq 0. \tag{17.55}$$

By (17.18) we have $A_\lambda u_\lambda \in \partial\Phi(J_\lambda u_\lambda)$, and $A_\mu u_\mu \in \partial\Phi(J_\mu u_\mu)$. Hence, by monotonicity of $\partial\Phi$,

$$\langle J_\lambda u_\lambda - J_\mu u_\mu, A_\lambda u_\lambda - A_\mu u_\mu\rangle \geq 0$$

and (17.55) gives

$$\dot{h}(t) + \langle A_\lambda u_\lambda - A_\mu u_\mu, \lambda A_\lambda u_\lambda - \mu A_\mu u_\mu\rangle \leq 0. \tag{17.56}$$

Let us consider the last expression in (17.56), develop it, and apply the Cauchy–Schwarz inequality:

$$\begin{aligned} \langle A_\lambda u_\lambda - A_\mu u_\mu, \lambda A_\lambda u_\lambda - \mu A_\mu u_\mu\rangle &= \lambda\|A_\lambda u_\lambda\|^2 + \mu\|A_\mu u_\mu\|^2 - (\lambda+\mu)\langle A_\lambda u_\lambda, A_\mu u_\mu\rangle \\ &\geq \lambda\|A_\lambda u_\lambda\|^2 + \mu\|A_\mu u_\mu\|^2 - (\lambda+\mu)\|A_\lambda u_\lambda\|\|A_\mu u_\mu\|. \end{aligned} \tag{17.57}$$

By adding the two elementary inequalities

$$\lambda\|A_\lambda u_\lambda\|\,\|A_\mu u_\mu\| \leq \lambda\|A_\lambda u_\lambda\|^2 + \frac{\lambda}{4}\|A_\mu u_\mu\|^2,$$
$$\mu\|A_\lambda u_\lambda\|\,\|A_\mu u_\mu\| \leq \mu\|A_\mu u_\mu\|^2 + \frac{\mu}{4}\|A_\lambda u_\lambda\|^2,$$

we obtain

$$\lambda\|A_\lambda u_\lambda\|^2 + \mu\|A_\mu u_\mu\|^2 - (\lambda+\mu)\|A_\lambda u_\lambda\|\|A_\mu u_\mu\| \geq -\frac{\lambda}{4}\|A_\mu u_\mu\|^2 - \frac{\mu}{4}\|A_\lambda u_\lambda\|^2. \tag{17.58}$$

Combining (17.56), (17.57), and (17.58) we obtain

$$\dot{h}(t) \leq \frac{\lambda}{4}\|A_\mu u_\mu\|^2 + \frac{\mu}{4}\|A_\lambda u_\lambda\|^2.$$

From $\dot{u}_\lambda = -A_\lambda u_\lambda$, $\dot{u}_\mu = -A_\mu u_\mu$, and (17.53) we deduce that

$$\dot{h}(t) \leq \frac{\lambda+\mu}{4}\|\partial\Phi(u_0)^0\|^2. \tag{17.59}$$

By integration of (17.59), and definition of h, we obtain

$$\|u_\lambda(t) - u_\mu(t)\| \leq \sqrt{\frac{\lambda+\mu}{2}}\|\partial\Phi(u_0)^0\|\sqrt{t} \quad \forall t \geq 0 \tag{17.60}$$

and

$$\|u_\lambda - u_\mu\|_{L^\infty(0,T;H)} \le \sqrt{\frac{\lambda+\mu}{2}}\|\partial\Phi(u_0)^0\|\sqrt{T}.$$

Thus (u_λ) is a Cauchy sequence in the Banach space $\mathscr{C}([0,T];\mathscr{H})$, equipped with the supremum norm, and hence converges uniformly. The argument being valid for all $T > 0$, set

$$u_\lambda \to u \quad L^\infty(0,T;\mathscr{H}) \quad \forall T \ge 0 \text{ as } \lambda \to 0.$$

Passing to the limit on (17.60) gives

$$\|u_\lambda(t) - u(t)\| \le \sqrt{\frac{\lambda t}{2}}\|\partial\Phi(u_0)^0\| \quad \forall t \ge 0.$$

Moreover, by (17.53)

$$\begin{aligned}\|J_\lambda u_\lambda(t) - u_\lambda(t)\| &= \lambda\|A_\lambda u_\lambda\| = \lambda\|\dot{u}_\lambda(t)\| \\ &\le \lambda\|\partial\Phi(u_0)^0\|.\end{aligned}$$

Hence

$$J_\lambda u_\lambda \to u \quad L^\infty(0,T;\mathscr{H}) \quad \forall T \ge 0 \text{ as } \lambda \to 0. \tag{17.61}$$

Since $(\dot{u}_\lambda)$ is bounded in $L^2(0,\infty;\mathscr{H})$, we also deduce that

$$\dot{u}_\lambda \rightharpoonup \dot{u} \quad \text{weak-}L^2(0,T;\mathscr{H}) \quad \forall T \ge 0 \text{ as } \lambda \to 0.$$

(d) Let us now pass to the limit on the approximate equations and prove that u is a strong solution of the Cauchy problem (17.47). Note first that

$$-\dot{u}_\lambda(t) = \nabla\Phi_\lambda(u_\lambda(t)) \in \partial\Phi(J_\lambda u_\lambda(t)). \tag{17.62}$$

Then reformulate (17.62) in a variational way by using the Legendre–Fenchel conjugate:

$$\begin{gathered}-\dot{u}_\lambda(t) \in \partial\Phi(J_\lambda u_\lambda(t)) \\ \Updownarrow \\ \Phi(J_\lambda u_\lambda(t)) + \Phi^*(-\dot{u}_\lambda(t)) + \langle J_\lambda u_\lambda(t), \dot{u}_\lambda(t)\rangle = 0.\end{gathered}$$

Since this last expression is always nonnegative, there is no loss of information by integrating it:

$$\int_0^T \left(\Phi(J_\lambda u_\lambda(t)) + \Phi^*(-\dot{u}_\lambda(t)) + \langle J_\lambda u_\lambda(t), \dot{u}_\lambda(t)\rangle\right) dt = 0.$$

Equivalently

$$I_\Phi(J_\lambda u_\lambda) + I_{\Phi^*}(-\dot{u}_\lambda) + \langle J_\lambda u_\lambda, \dot{u}_\lambda\rangle_{L^2(0,T;H)} = 0, \tag{17.63}$$

where the integral functionals I_Φ and I_{Φ^*} on $L^2(0,T;\mathscr{H})$ are respectively defined by $I_\Phi(v) = \int_0^T \Phi(v(t))\,dt$ and $I_{\Phi^*}(v) = \int_0^T \Phi^*(v(t))\,dt$. By passing to the lower limit on (17.63) we obtain

$$\liminf_{\lambda\to 0} I_\Phi(J_\lambda u_\lambda) + \liminf_{\lambda\to 0} I_{\Phi^*}(-\dot{u}_\lambda) + \liminf_{\lambda\to 0}\langle J_\lambda u_\lambda, \dot{u}_\lambda\rangle_{L^2(0,T;H)} \le 0.$$

Functionals I_Φ and I_{Φ^*} are lower semicontinuous on $L^2(0,T;\mathscr{H})$ (Fatou's lemma) and convex. Hence they are lower semicontinuous for the weak topology of $L^2(0,T;\mathscr{H})$. On the other hand, the scalar product $\langle J_\lambda u_\lambda, \dot u_\lambda\rangle_{L^2(0,T;H)}$ involves two sequences which respectively converge strongly and weakly in $L^2(0,T;\mathscr{H})$. As a consequence

$$I_\Phi(u)+I_{\Phi^*}(-\dot u)+\langle u,\dot u\rangle_{L^2(0,T;\mathscr{H})}\le 0.$$

Equivalently

$$\int_0^T \left(\Phi(u(t))+\Phi^*(-\dot u(t))+\langle u(t),\dot u(t)\rangle\right)dt\le 0. \tag{17.64}$$

By definition of the Legendre–Fenchel conjugate, in (17.64) the integrand is nonnegative. As a consequence

$$\Phi(u(t))+\Phi^*(-\dot u(t))+\langle u(t),\dot u(t)\rangle=0 \text{ for almost all } t>0.$$

This is the Legendre–Fenchel extremality condition, which gives

$$-\dot u(t)\in\partial\Phi(u(t)) \quad\text{for almost all } t>0. \tag{17.65}$$

Thus, we have proved the central part of Theorem 17.2.2: for any $u_0\in \mathrm{dom}\,\partial\Phi$, there exists a unique strong global solution $u:[0,+\infty)\to\mathscr{H}$ of the Cauchy problem

$$\begin{cases}\dot u(t)+\partial\Phi(u(t))\ni 0,\\ u(0)=u_0.\end{cases}$$

(e) Let us prove some further regularity properties of the solution u. By (17.53), for any $\lambda>0$, for any $0\le s\le t<\infty$,

$$\|u_\lambda(t)-u_\lambda(s)\|\le|t-s|\,\|\partial\Phi(u_0)^0\|.$$

By letting $\lambda\to 0$, we obtain, for any $0\le s\le t<\infty$,

$$\|u(t)-u(s)\|\le|t-s|\,\|\partial\Phi(u_0)^0\|. \tag{17.66}$$

Hence, u is globally Lipschitz continuous on $[0,+\infty)$ and

$$\|\dot u(t)\|\le\|\partial\Phi(u_0)^0\| \text{ for almost all } t>0.$$

Let's use again the autonomous nature of the (GSD) dynamic. Take $t_0\ge 0$ such that $u(t_0)\in\mathrm{dom}\,\partial\Phi$, u is differentiable at t_0, and $-\dot u(t_0)\in\partial\Phi(u(t_0))$, which is verified for almost all t_0. The orbit $t\mapsto u(t_0+t)$ is the strong solution of (GSD) with Cauchy data $u(t_0)$ at time $t=0$. By (17.66), for any $h>0$

$$\|u(t_0+h)-u(t_0)\|\le|h|\,\|\partial\Phi(u(t_0))^0\|.$$

Letting $h\to 0$

$$\|\dot u(t_0)\|\le\ \|\partial\Phi(u(t_0))^0\|.$$

Since $-\dot u(t_0)\in\partial\Phi(u(t_0))$, we deduce that $-\dot u(t_0)=\partial\Phi(u(t_0))^0$. Thus

$$-\dot u(t)=\partial\Phi(u(t))^0 \text{ for almost all } t>0. \tag{17.67}$$

(f) Let us now prove that for all $t \geq 0$, $u(t) \in \operatorname{dom} \partial\Phi$, $\frac{d^+u}{dt}(t)$ exists, and

$$\frac{d^+u}{dt}(t) = -\partial\Phi(u(t))^0.$$

To that end, let us use again (17.53), and $-\dot{u}_\lambda(t) = \nabla\Phi_\lambda(u_\lambda(t))$, to obtain

$$\|\nabla\Phi_\lambda(u_\lambda(t))\| \leq \|\partial\Phi(u_0)^0\|. \tag{17.68}$$

Let us fix $t \geq 0$, and let $\eta \in \mathcal{H}$ be a weak cluster point of the net $(\nabla\Phi_\lambda(u_\lambda(t)))_\lambda$, say,

$$\nabla\Phi_\lambda(u_\lambda(t)) \rightharpoonup \eta.$$

Since $\nabla\Phi_\lambda(u_\lambda(t)) \in \partial\Phi(J_\lambda u_\lambda(t))$, by passing to the limit on the subdifferential inequality

$$\forall \xi \in \mathcal{H} \quad \Phi(\xi) \geq \Phi(J_\lambda u_\lambda(t)) + \langle \nabla\Phi_\lambda(u_\lambda(t)), \xi - J_\lambda u_\lambda(t)\rangle$$

we obtain

$$\forall \xi \in \mathcal{H} \quad \Phi(\xi) \geq \Phi(u(t)) + \langle \eta, \xi - u(t)\rangle.$$

Hence $\eta \in \partial\Phi(u(t))$, which proves that for all $t \geq 0$, $u(t) \in \operatorname{dom}\partial\Phi$. Moreover, by (17.68) and the lower semicontinuity of the norm for the weak topology,

$$\|\eta\| \leq \|\partial\Phi(u_0)^0\|.$$

Since $\eta \in \partial\Phi(u(t))$, it follows that

$$\|\partial\Phi(u(t))^0\| \leq \|\partial\Phi(u_0)^0\|.$$

Again using the autonomous nature of (GSD) we infer that

$$t \mapsto \|\partial\Phi(u(t))^0\| \text{ is a nonincreasing function.} \tag{17.69}$$

From this we infer that $t \mapsto \partial\Phi(u(t))^0$ is right-continuous. Fix $t_0 \geq 0$, and let $t_n \to t_0$, $t_n > t_0$. By (17.69),

$$\|\partial\Phi(u(t_n))^0\| \leq \|\partial\Phi(u(t_0))^0\|. \tag{17.70}$$

Let η be a weak cluster point of $(\partial\Phi(u(t_n))^0)_n$. By the strong $\times$ weak closedness property of the maximal monotone operator $\partial\Phi$ (see Proposition 17.2.4), we have $\eta \in \partial\Phi(u(t_0))$. By (17.70) and the lower semicontinuity property of the norm with respect to the weak topology, we have $\|\eta\| \leq \|\partial\Phi(u(t_0))^0\|$. Hence $\eta = \partial\Phi(u(t_0))^0$, which by uniqueness of the cluster point gives the weak convergence

$$\partial\Phi(u(t_n))^0 \rightharpoonup \partial\Phi(u(t_0))^0.$$

On the other hand, by (17.70), we have

$$\limsup_n \|\partial\Phi(u(t_n))^0\| \leq \|\partial\Phi(u(t_0))^0\|.$$

Hence $\partial\Phi(u(t_n))^0$ converges strongly in $\mathcal{H}$ to $\partial\Phi(u(t_0))^0$, which proves the right-continuity of $t \mapsto \partial\Phi(u(t))^0$. From this, we can easily deduce that for all $t \geq 0$, $u(t) \in \operatorname{dom}\partial\Phi$, and $\frac{d^+u}{dt}(t)$ exists. Since u is Lipschitz continuous, it is absolutely continuous on each bounded interval, and

$$\forall t \geq 0, \ \forall h > 0 \quad u(t+h) - u(t) = \int_t^{t+h} \dot{u}(\tau) d\tau.$$

By (17.67) we have $-\dot{u}(t) = \partial\Phi(u(t))^0$ for almost all $t > 0$. Hence

$$\forall t \geq 0, \ \forall h > 0 \quad u(t+h) - u(t) = -\int_t^{t+h} \partial\Phi(u(\tau))^0 d\tau.$$

Dividing by $h > 0$, letting $h \to 0$, and taking account of the right-continuity of $\tau \mapsto \partial\Phi(u(\tau))^0$, we obtain

$$\frac{1}{h}(u(t+h) - u(t)) = -\frac{1}{h}\int_t^{t+h} \partial\Phi(u(\tau))^0(\tau)\, d\tau \longrightarrow -\partial\Phi(u(t))^0,$$

which proves that for all $t \geq 0$, $u(t) \in \operatorname{dom} \partial\Phi$, $\frac{d^+u}{dt}(t)$ exists, and

$$\frac{d^+u}{dt}(t) = -\partial\Phi(u(t))^0. \tag{17.71}$$

Combining (17.69) with (17.71) we obtain

$$t \mapsto \left\| \frac{d^+u}{dt}(t) \right\| \text{ is nonincreasing.}$$

(g) Let us complete the proof of Theorem 17.2.2 by showing that $t \mapsto \Phi(u(t))$ is nonincreasing, absolutely continuous on each bounded interval $[0, T]$, and

$$\frac{d}{dt}\Phi(u(t)) = -\|\dot{u}(t)\|^2 \quad \text{for almost all } t > 0. \tag{17.72}$$

Indeed this is a direct consequence of Proposition 17.2.5 below. By taking $h = \dot{u}$ in (17.73) we obtain (17.72). $\square$

The following result is due to Brézis [135, Lemma 3.3]. This is a remarkable derivation chain rule in a nonsmooth context.

Proposition 17.2.5. *Let $u \in W^{1,2}([0, T]; \mathcal{H})$. Suppose that $u(t) \in \operatorname{dom} \partial\Phi$ for almost all $t \in [0, T]$ and that there exists some $h \in L^2([0, T]; \mathcal{H})$ such that $h(t) \in \partial\Phi(u(t))$ for almost all $t \in [0, T]$. Then, the function $t \mapsto \Phi(u(t))$ is absolutely continuous on $[0, T]$, and*

$$\frac{d}{dt}\Phi(u(t)) = \langle h(t), \dot{u}(t) \rangle \quad \textit{for almost all } t \in [0, T]. \tag{17.73}$$

PROOF. By integration between $t_1, t_2 \in [0, T]$ of the classical derivation chain rule

$$\frac{d}{dt}\Phi_\lambda(u) = \langle A_\lambda u, \dot{u} \rangle,$$

we obtain

$$\Phi_\lambda(u(t_2)) - \Phi_\lambda(u(t_1)) = \int_{t_1}^{t_2} \langle A_\lambda u(\tau), \dot{u}(\tau)\rangle d\tau.$$

In order to pass to the limit as $\lambda \to 0$ we apply the dominated convergence theorem. We have

$$\|A_\lambda u(\tau)\| \leq \|\partial\Phi(u(\tau))^0\| \leq \|h(\tau)\|.$$

As a consequence, $|\langle A_\lambda u, \dot{u}\rangle| \leq \|h\|\,\|\dot{u}\|$, which belongs to $L^1([0,T];\mathcal{H})$. On the other hand, $A_\lambda u \to \partial\Phi(u)^0$, and $\Phi_\lambda(u) \to \Phi(u)$. Hence

$$\Phi(u(t_2)) - \Phi(u(t_1)) = \int_{t_1}^{t_2} \langle \partial\Phi(u(\tau))^0, \dot{u}(\tau)\rangle d\tau.$$

As a consequence, the function $t \mapsto \Phi(u(t))$ is absolutely continuous on $[0,T]$. By $h(t) \in \partial\Phi(u(t))$, the convex subdifferential inequality gives, for almost all $t \in [0,T]$, for all $v \in \mathcal{H}$,

$$\Phi(v) - \Phi(u(t)) \geq \langle h(t), v - u(t)\rangle.$$

Taking successively $v = u(t+\epsilon)$, $v = u(t-\epsilon)$, dividing by ϵ, and letting $\epsilon \to 0$, for almost all $t \in [0,T]$ we obtain

$$\frac{d}{dt}\Phi(u(t)) = \langle h(t), \dot{u}(t)\rangle,$$

which completes the proof. □

Let us show some further convergence properties of the net (u_λ). Our approach relies on a variational argument first developed in [38] and [74] and which uses the following elementary result.

Lemma 17.2.1. *Let $(a_{n,1})_{n\in\mathbf{N}}, \dots, (a_{n,l})_{n\in\mathbf{N}}$ be a finite family of real sequences which satisfy*

$$\sum_{k=1}^{l} a_{n,k} \leq 0 \quad \textit{for each } n \in \mathbf{N};$$
$$a_k \leq \liminf_n a_{n,k} \quad \textit{for each } k = 1,2,\dots,l;$$
$$\sum_{k=1}^{l} a_k = 0.$$

Then $a_{n,k} \to a_k$ for each $k = 1,2,\dots,l$.

Proposition 17.2.6. *Let u be the solution of Cauchy problem* (17.47) *and (u_λ) the sequence of solutions of problems* (17.51) *obtained by Moreau–Yosida approximation. For any $T > 0$, as $\lambda \to 0$,*

(i) *$u_\lambda \to u$ uniformly on $[0,T]$,*

(ii) *$\dot{u}_\lambda \to \dot{u}$ strongly in $L^2([0,T];\mathcal{H})$,*

(iii) *$\Phi_\lambda(u_\lambda) \to \Phi(u)$ uniformly on $[0,T]$,*

(iv) *$\Phi^*(-\dot{u}_\lambda) \to \Phi^*(-\dot{u})$ strongly in $L^1([0,T])$.*

PROOF. In the proof of Theorem 17.2.2 we showed that (u_λ) converges uniformly to u on $[0,T]$ and that $\dot u_\lambda \rightharpoonup \dot u$ weakly in $L^2([0,T];\mathcal{H})$ for all $T>0$. Taking the scalar product by $\dot u_\lambda(t)$ in

$$\dot u_\lambda(t) + \nabla\Phi_\lambda(u_\lambda(t)) = 0$$

and integrating on $[0,T]$ gives the energy estimate

$$\int_0^T \|\dot u_\lambda(t)\|^2 dt + \Phi_\lambda(u_\lambda(T)) - \Phi_\lambda(u_0) = 0. \tag{17.74}$$

Let us apply Lemma 17.2.1 to (17.74).

By lower semicontinuity of $v \mapsto \|v\|^2_{L^2(0,T;\mathcal{H})}$ for the weak topology of $L^2(0,T;H)$

$$\int_0^T \|\dot u(t)\|^2 dt \le \liminf_\lambda \int_0^T \|\dot u_\lambda(t)\|^2 dt. \tag{17.75}$$

From $\Phi_\lambda(u_\lambda) \ge \Phi(J_\lambda u_\lambda)$, by uniform convergence of $J_\lambda u_\lambda$ to u on $[0,T]$ (see (17.61)), and lower semicontinuity of Φ, we infer

$$\Phi(u(T)) \le \liminf_\lambda \Phi_\lambda(u_\lambda(T)). \tag{17.76}$$

Moreover

$$\Phi(u_0) = \lim_\lambda \Phi_\lambda(u_0). \tag{17.77}$$

On the other hand, by 17.48, $t \mapsto \Phi(u(t))$ is absolutely continuous on $[0,T]$ and satisfies

$$\frac{d}{dt}\Phi(u(t)) = -\|\dot u(t)\|^2 \quad \text{for almost all } t>0. \tag{17.78}$$

Integration of (17.78) on $[0,T]$ gives

$$\int_0^T \|\dot u(t)\|^2 dt + \Phi(u(T)) - \Phi(u_0) = 0. \tag{17.79}$$

Comparing (17.74) with (17.79), taking account of (17.75), (17.76), (17.77), gives, by Lemma 17.2.1,

$$\int_0^T \|\dot u_\lambda(t)\|^2 dt \to \int_0^T \|\dot u(t)\|^2 dt,$$
$$\Phi_\lambda(u_\lambda(T)) \to \Phi(u(T)).$$

Weak convergence and convergence of the norms imply strong convergence. Thus

$$\dot u_\lambda \to \dot u \quad \text{strongly in } L^2([0,T];\mathcal{H}).$$

The convergence property $\Phi_\lambda(u_\lambda(T)) \to \Phi(u(T))$ holds for all $T>0$. Moreover, by (17.53)

$$\left|\frac{d}{dt}\Phi_\lambda(u_\lambda(t))\right| = \|\dot u_\lambda(t)\|^2 \le \|\partial\Phi(u_0)^0\|^2 \quad \text{for almost all } t>0.$$

As a consequence, by Ascoli's theorem, the sequence of functions $(\Phi_\lambda(u_\lambda))$ is relatively compact for the topology of the uniform convergence on $[0,T]$. Thus

$$\Phi_\lambda(u_\lambda) \to \Phi(u) \quad \text{uniformly on } [0,T].$$

By the Fenchel extremality relation, for almost all $t > 0$

$$\Phi_\lambda(u_\lambda(t)) + {\Phi_\lambda}^*(-\dot{u}_\lambda(t)) + \langle u_\lambda(t), \dot{u}_\lambda(t)\rangle = 0.$$

By uniform convergence of u_λ to u and of $\Phi_\lambda(u_\lambda)$ to $\Phi(u)$ and convergence of $\dot{u}_\lambda$ to $\dot{u}$ in $L^2([0,T];\mathscr{H})$,

$$ {\Phi_\lambda}^*(-\dot{u}_\lambda) = -\Phi_\lambda(u_\lambda) - \langle u_\lambda, \dot{u}_\lambda\rangle \to -\Phi(u) - \langle u, \dot{u}\rangle \quad \text{in } L^1([0,T]). \tag{17.80}$$

By the Fenchel extremality relation, for almost all $t > 0$

$$\Phi(u(t)) + \Phi^*(-\dot{u}(t)) + \langle u(t), \dot{u}(t)\rangle = 0. \tag{17.81}$$

Comparing (17.80) and (17.81) gives

$$ {\Phi_\lambda}^*(-\dot{u}_\lambda) \to \Phi^*(-\dot{u}) \quad \text{in } L^1([0,T]).$$

Since ${\Phi_\lambda}^*(-\dot{u}_\lambda) = \Phi^*(-\dot{u}_\lambda) + \frac{\lambda}{2}\|\dot{u}_\lambda\|^2$ we deduce that

$$\Phi^*(-\dot{u}_\lambda) \to \Phi^*(-\dot{u}) \quad \text{in } L^1([0,T]),$$

which completes the proof of Proposition 17.2.6. □

17.2.3 ▪ Regularizing effect

In this section, $\Phi : \mathscr{H} \to \mathbf{R} \cup \{+\infty\}$ is a convex, lower semicontinuous, and proper function. Following Theorem 17.2.2, for any initial data $u_0 \in \operatorname{dom} \partial\Phi$, we denote by $u(\cdot, u_0) : [0, +\infty) \to \mathscr{H}$ the solution of the Cauchy problem

$$\begin{cases} \dot{u}(t) + \partial\Phi(u(t)) \ni 0, \\ u(0) = u_0. \end{cases} \tag{17.82}$$

By monotonicity of $\partial\Phi$, (see (17.50), proof of uniqueness), for any u_0, $\hat{u}_0 \in \operatorname{dom}\partial\Phi$,

$$\|u(t,u_0) - u(t,\hat{u}_0)\| \le \|u_0 - \hat{u}_0\| \quad \forall t \ge 0. \tag{17.83}$$

For any $t \ge 0$, let us consider the operator $S(t) : u_0 \in \operatorname{dom}\partial\Phi \mapsto S(t)u_0 = u(t,u_0) \in \operatorname{dom}\partial\Phi$. By (17.83), $S(t)$ is a nonexpansive mapping from $\operatorname{dom}\partial\Phi$ into the complete metric space $\overline{\operatorname{dom}\partial\Phi} = \overline{\operatorname{dom}\Phi}$. By a classical uniform continuity argument, $S(t)$ can be uniquely extended into an operator, still denoted by $S(t) : \overline{\operatorname{dom}\Phi} \to \overline{\operatorname{dom}\Phi}$, which satisfies

$$\begin{aligned} & S(t)u_0 = u(t,u_0) \quad \text{for } u_0 \in \operatorname{dom}\partial\Phi; \\ & S(t)u_0 = \lim_n u(t,u_{0n}) \quad \text{for } u_0 \in \overline{\operatorname{dom}\Phi},\ u_{0n} \to u_0,\ u_{0n} \in \operatorname{dom}\partial\Phi; \\ & \|S(t)u_0 - S(t)\hat{u}_0\| \le \|u_0 - \hat{u}_0\| \quad \text{for } u_0,\ \hat{u}_0 \overline{\operatorname{dom}\Phi}. \end{aligned} \tag{17.84}$$

Let us verify the following semigroup (semiflow) properties of the family of operators $(S(t))_{t\ge0}$:

$$S(t) \circ S(s) = S(t+s) \quad \forall s,t \ge 0; \tag{17.85}$$

$$\lim_{t\to0^+} S(t)u_0 = u_0 \quad \forall u_0 \in \overline{\operatorname{dom}\Phi}. \tag{17.86}$$

Given $u_0 \in \overline{\operatorname{dom}\Phi}$, let $u_{0n} \to u_0$, $u_{0n} \in \operatorname{dom}\partial\Phi$. By uniqueness of the solution of the Cauchy problem (17.82), for each $n \in \mathbf{N}$, for all $s, t \geq 0$,

$$S(t)\circ S(s)u_{0n} = S(s+t)u_{0n}. \tag{17.87}$$

Passing to the limit on (17.87), as $n \to +\infty$, we obtain the semigroup property (17.85). Moreover, by the triangle inequality, and nonexpansiveness of $S(t)$, we have

$$\begin{aligned} ||S(t)u_0 - u_0|| &\leq ||S(t)u_0 - S(t)u_{0n}|| + ||S(t)u_{0n} - u_{0n}|| + ||u_{0n} - u_0||; \\ &\leq 2||u_{0n} - u_0|| + ||S(t)u_{0n} - u_{0n}||. \end{aligned}$$

As a consequence

$$\limsup_t ||S(t)u_0 - u_0|| \leq 2||u_{0n} - u_0||.$$

This being true for all $n \in \mathbf{N}$, we obtain (17.86). We call $(S(t))_{t\geq 0}$ the semigroup of contractions generated by $A = \partial\Phi$. Let us prove that for $u_0 \in \overline{\operatorname{dom}\Phi}$, the mapping $t \mapsto S(t)u_0$ can still be interpreted as the strong solution of (17.82). This result, which is due to Brézis, is called the regularization effect. It has important consequences in PDE's evolution problems.

Theorem 17.2.3. *Let $\Phi : \mathcal{H} \to \mathbf{R} \cup \{+\infty\}$ be convex, lower semicontinuous, and proper. Suppose that Φ is minorized, i.e., $\inf_{\mathcal{H}} \Phi > -\infty$. Let $u_0 \in \overline{\operatorname{dom}\Phi}$. Then, there exists a unique strong global solution $u : [0, +\infty) \to \mathcal{H}$ of the Cauchy problem*

$$\begin{cases} \dot{u}(t) + \partial\Phi(u(t)) \ni 0, \\ u(0) = u_0, \end{cases} \tag{17.88}$$

which is given by $u(t) = S(t, u_0)$. The Cauchy problem (17.88) *is satisfied in the following sense:*

(i) $u \in C([0, +\infty); \mathcal{H})$;

(ii) $u(t) \in \operatorname{dom}\partial\Phi$ *for all* $t > 0$;

(iii) *u is Lipschitz continuous on $[\delta, +\infty)$ for any $\delta > 0$;*

(iv) (17.88) *is satisfied for almost all $t > 0$.*

Moreover

(v) $\sqrt{t}\,\dot{u} \in L^2(0, T; \mathcal{H})$ *for all* $T > 0$;

(vi) *for each $t > 0$,*

$$||\partial\Phi(u(t))^0|| \leq ||\partial\Phi(v)^0|| + \frac{1}{t}||u_0 - v|| \quad \forall v \in \operatorname{dom}\partial\Phi.$$

(vii) *For each $t > 0$, u has a right derivative, and*

$$\frac{d^+u}{dt}(t) = -\partial\Phi(u(t))^0, \tag{17.89}$$

where $\partial\Phi(u(t))^0$ is the element of minimal norm of $\partial\Phi(u(t))$.

(viii) $t \mapsto \|\frac{d^+u}{dt}(t)\|$ *is nonincreasing.*

(ix) $t \mapsto \Phi(u(t))$ *is nonincreasing, absolutely continuous on each bounded interval* $[\delta, T]$, $\delta > 0$, *and*

$$\frac{d}{dt}\Phi(u(t)) = -\|\dot{u}(t)\|^2 \quad \text{for almost all } \ t > 0. \tag{17.90}$$

PROOF. Let $u_0 \in \overline{\operatorname{dom}\Phi}$. According to the Cauchy–Lipschitz theorem, Theorem 17.1.2, for any $\lambda > 0$, there is a unique global classical solution $u_\lambda : [0, +\infty) \to \mathcal{H}$ of the Cauchy problem

$$\begin{cases} \dot{u}_\lambda(t) + \nabla\Phi_\lambda(u_\lambda(t)) = 0, \\ u_\lambda(0) = u_0. \end{cases} \tag{17.91}$$

Let us denote respectively by $S(t)$ and $S_\lambda(t)$ the semigroups generated by $A = \partial\Phi$ and $A_\lambda = \nabla\Phi_\lambda$. Let us show that

$$u_\lambda(t) = S_\lambda(t)u_0 \to S(t)u_0 \ \text{ as } \lambda \to 0. \tag{17.92}$$

Let $u_{0n} \to u_0$ with $u_{0n} \in \operatorname{dom}\partial\Phi$ for each $n \in \mathbf{N}$. By the triangle inequality and nonexpansiveness of $S(t)$ and $S_\lambda(t)$

$$\begin{aligned} \|S_\lambda(t)u_0 - S(t)u_0\| &\le \|S_\lambda(t)u_0 - S_\lambda(t)u_{0n}\| + \|S_\lambda(t)u_{0n} - S(t)u_{0n}\| + \|S(t)u_{0n} - S(t)u_0\| \\ &\le 2\|u_{0n} - u_0\| + \|S_\lambda(t)u_{0n} - S(t)u_{0n}\|. \end{aligned}$$

By Proposition 17.2.6 and $u_{0n} \in \operatorname{dom}\partial\Phi$, we have $S_\lambda(t)u_{0n} \to S(t)u_{0n}$ as $\lambda \to 0$. Hence,

$$\limsup_{\lambda\to 0} \|S_\lambda(t)u_0 - S(t)u_0\| \le 2\|u_{0n} - u_0\|.$$

Letting $n \to +\infty$ gives the result.

The proof of Theorem 17.2.3 will consist in establishing estimates on the net (u_λ), and then pass to the limit as $\lambda \to 0$. The key idea is to establish an energy estimate on $(\dot{u}_\lambda)$ in a weighted L^2 space, with a weight which vanishes at zero. Doing so we can cancel the singularities at the origin (occurring from the nonsmooth data u_0). Thus, let us multiply (17.91) by $t\,\dot{u}_\lambda(t)$ and integrate from 0 to T:

$$\int_0^T t\|\dot{u}_\lambda(t)\|^2 dt + \int_0^T t\frac{d}{dt}\Phi_\lambda(u_\lambda(t))dt = 0.$$

Integrating by parts

$$\int_0^T t\|\dot{u}_\lambda(t)\|^2 dt + T\Phi_\lambda(u_\lambda(T)) = \int_0^T \Phi_\lambda(u_\lambda(t))dt. \tag{17.93}$$

To exploit this estimate, we establish majorization/minimization of each of its constituents. By (17.52), $t \mapsto \|\dot{u}_\lambda(t)\|$ is a nonincreasing function. Hence

$$\int_0^T t\|\dot{u}_\lambda(t)\|^2 dt \ge \frac{T^2}{2}\|\dot{u}_\lambda(T)\|^2 = \frac{T^2}{2}\|A_\lambda u_\lambda(T)\|^2. \tag{17.94}$$

Take $v \in \operatorname{dom} \partial\Phi$. By $-\dot{u}_\lambda(t) = \nabla\Phi_\lambda(u_\lambda(t))$, the subdifferential inequality for Φ_λ at $u_\lambda(t)$ gives

$$\begin{aligned}\Phi_\lambda(v) &\geq \Phi_\lambda(u_\lambda(t)) + \langle -\dot{u}_\lambda(t), v - u_\lambda(t)\rangle \\ &\geq \Phi_\lambda(u_\lambda(t)) + \frac{1}{2}\frac{d}{dt}\|u_\lambda(t) - v\|^2.\end{aligned}$$

Integrating from from 0 to T

$$T\Phi_\lambda(v) \geq \int_0^T \Phi_\lambda(u_\lambda(t))dt + \frac{1}{2}\|u_\lambda(T) - v\|^2 - \frac{1}{2}\|u_0 - v\|^2. \tag{17.95}$$

The subdifferential inequality for Φ_λ at v gives

$$\Phi_\lambda(u_\lambda(T)) \geq \Phi_\lambda(v) + \langle A_\lambda v, u_\lambda(T) - v\rangle,$$

which, after multiplication by T, gives

$$T\Phi_\lambda(u_\lambda(T)) \geq T\Phi_\lambda(v) + T\langle A_\lambda v, u_\lambda(T) - v\rangle. \tag{17.96}$$

Combining (17.93) with (17.94), (17.95), (17.96) gives

$$\frac{T^2}{2}\|A_\lambda u_\lambda(T)\|^2 + T\Phi_\lambda(v) + T\langle A_\lambda v, u_\lambda(T) - v\rangle \leq T\Phi_\lambda(v) + \frac{1}{2}\|u_0 - v\|^2 - \frac{1}{2}\|u_\lambda(T) - v\|^2.$$

After simplification

$$\frac{T^2}{2}\|A_\lambda u_\lambda(T)\|^2 \leq \frac{1}{2}\|u_0 - v\|^2 - T\langle A_\lambda v, u_\lambda(T) - v\rangle - \frac{1}{2}\|u_\lambda(T) - v\|^2. \tag{17.97}$$

By using the elementary majorization

$$-\frac{1}{2}\|u_\lambda(T) - v\|^2 - T\langle A_\lambda v, u_\lambda(T) - v\rangle \leq \frac{1}{2}\|TA_\lambda v\|^2$$

in (17.97) we obtain

$$\frac{T^2}{2}\|A_\lambda u_\lambda(T)\|^2 \leq \frac{1}{2}\|u_0 - v\|^2 + \frac{T^2}{2}\|A_\lambda v\|^2.$$

Equivalently

$$\|A_\lambda u_\lambda(T)\|^2 \leq \frac{1}{T^2}\|u_0 - v\|^2 + \|A_\lambda v\|^2,$$

which implies

$$\|A_\lambda u_\lambda(T)\| \leq \|A_\lambda v\| + \frac{1}{T}\|u_0 - v\|.$$

Since $v \in \operatorname{dom} \partial\Phi$, we have $\|A_\lambda v\| \leq \|A^0 v\|$. Hence

$$\|A_\lambda u_\lambda(T)\| \leq \|A^0 v\| + \frac{1}{T}\|u_0 - v\|. \tag{17.98}$$

As a consequence, for each $t > 0$ the net $(A_\lambda u_\lambda(t))_{\lambda \to 0}$ is bounded. Let us fix $t > 0$. After extraction of a subnet (we keep the same notation) we have

$$A_\lambda u_\lambda(t) \rightharpoonup \eta \quad \text{weakly in } \mathcal{H}.$$

By the subdifferential inequality for Φ_λ at $u_\lambda(t)$, for any $\xi \in \mathcal{H}$

$$\Phi_\lambda(\xi) \geq \Phi_\lambda(u_\lambda(t)) + \langle A_\lambda u_\lambda(t), \xi - u_\lambda(t)\rangle.$$

By the nonincreasing property of $\lambda \mapsto \Phi_\lambda v$ we deduce that, for any $0 < \lambda < \lambda_0$,

$$\Phi(\xi) \geq \Phi_\lambda(\xi) \geq \Phi_{\lambda_0}(u_\lambda(t)) + \langle A_\lambda u_\lambda(t), \xi - u_\lambda(t)\rangle. \tag{17.99}$$

By (17.92), $u_\lambda(t) = S_\lambda(t)u_0 \to S(t)u_0$ as $\lambda \to 0$. By passing to the limit on (17.99), as $\lambda \to 0$ (the scalar product involves two sequences which converge respectively weakly and strongly),

$$\Phi(\xi) \geq \Phi_{\lambda_0}(S(t)u_0) + \langle \eta, \xi - S(t)u_0\rangle.$$

This being true for any $\lambda_0 > 0$, letting $\lambda_0 \to 0$ gives, for any $\xi \in \mathcal{H}$,

$$\Phi(\xi) \geq \Phi(S(t)u_0) + \langle \eta, \xi - S(t)u_0\rangle.$$

Therefore, for all $t > 0$,

$$S(t)u_0 \in \operatorname{dom}\partial\Phi \quad \text{and} \quad \eta \in \partial\Phi(S(t)u_0).$$

As a consequence, by (17.84), for any $\delta > 0$, $t \mapsto S(t)u_0$ is the strong solution on $[\delta, +\infty)$ of the evolution equation $\dot{u}(t) + \partial\Phi(u(t)) \ni 0$ with Cauchy data $u(\delta) = S(\delta)u_0$. This implies that $u(t) = S(t)u_0$ satisfies properties (i), (ii), (iii), and (iv) of Theorem 17.2.3. Moreover, passing to the limit on (17.98), by the lower semicontinuity of the norm for the weak topology,

$$\|\eta\| \leq \|A^0 v\| + \frac{1}{t}\|u_0 - v\|.$$

Since $\eta \in \partial\Phi(u(t))$ we deduce that

$$\|\partial\Phi(u(t))^0\| \leq \|\partial\Phi(v)^0\| + \frac{1}{t}\|u_0 - v\| \quad \forall v \in \operatorname{dom}\partial\Phi.$$

Let us now return to (17.93) and use (17.95), (17.96). We obtain

$$\int_0^T t\|\dot{u}_\lambda(t)\|^2 dt + T\Phi_\lambda(v) + T\langle A_\lambda v, u_\lambda(T) - v\rangle \leq T\Phi_\lambda(v) + \frac{1}{2}\|u_0 - v\|^2 - \frac{1}{2}\|u_\lambda(T) - v\|^2.$$

After simplification and using the same majorization as above we obtain

$$\begin{aligned}\int_0^T t\|\dot{u}_\lambda(t)\|^2 dt &\leq \frac{1}{2}\|u_0 - v\|^2 - \frac{1}{2}\|u_\lambda(T) - v\|^2 - T\langle A_\lambda v, u_\lambda(T) - v\rangle \\ &\leq \frac{1}{2}\|u_0 - v\|^2 + \frac{T^2}{2}\|A_\lambda v\|^2.\end{aligned}$$

Passing to the limit, as $\lambda \to 0$, according to an argument of lower semicontinuity for the weak topology

$$\int_0^T t\|\dot{u}(t)\|^2 dt \leq \frac{1}{2}\|u_0 - v\|^2 + \frac{T^2}{2}\|A^0 v\|^2 \quad \forall v \in \operatorname{dom}\partial\Phi.$$

Hence $\sqrt{t}\,\dot{u} \in L^2(0, T; \mathcal{H})$ for all $T > 0$.

Items (vii), (viii), and (ix) are direct consequences of Theorem 17.2.2 and of the fact that $u(t) = S(t)u_0$ is a strong solution of (GSD) on $[\delta, +\infty)$ for any $\delta > 0$. This completes the proof of Theorem 17.2.3. $\square$

17.2.4 ▪ The exponential formula

As shown in the previous section, the Cauchy problem

$$\begin{cases} \dot{u}(t) + \partial\Phi(u(t)) \ni 0, \\ u(0) = u_0, \ u_0 \in \overline{\operatorname{dom} \partial\Phi} \end{cases} \tag{17.100}$$

possesses a unique strong global solution u in $\mathbf{C}([0,+\infty),\mathcal{H})$ which may be written in terms of a semigroup: $u(t) = \mathcal{S}(t)u_0$. By analogy with the solution $u(t) = e^{-tA}u_0$ of the linear Cauchy problem $\dot{u} + Au = 0$, $u(0) = u_0$ when $\mathcal{H}$ is a finite dimensional space, it is convenient to introduce the notation $u(t) = e^{-t\partial\Phi}u_0$, which is consistent with the semigroup property. According to the definition of the matrix exponential, the solution of $\dot{u} + Au = 0$, $u(0) = u_0$ can also be expressed as the limit $u(t) = \lim_{n\to+\infty}\left(I + \frac{t}{n}A\right)^{-n}u_0$. In this section, we establish an analogous formula for the global strong solution of (17.100).

Theorem 17.2.4 (exponential formula). *Let $\Phi : \mathcal{H} \to \mathbf{R} \cup \{+\infty\}$ be a convex function satisfying the conditions of Theorem 17.2.3 and $u_0 \in \overline{\operatorname{dom} \partial\Phi}$. Then for each $t \in [0,+\infty)$ the limit*

$$u(t) := \lim_{n\to+\infty}\left(I + \frac{t}{n}\partial\Phi\right)^{-n} u_0 \tag{17.101}$$

exists and is uniform on bounded intervals of $[0,+\infty)$. Furthermore, u is the unique strong global solution of the Cauchy problem (17.100).

For proving Theorem (17.2.4), it is convenient to introduce the notion of backward implicit discrete scheme. Given a sequence $(\lambda_k)_{k\in\mathbf{N}}$ of positive numbers, we call a proximal sequence with step size $(\lambda_k)_{k\in\mathbf{N}}$, associated with $\partial\Phi$ (or more generally with a maximal monotone operator A), the sequence $(x_k)_{k\in\mathbf{N}}$ in $\mathcal{H}$ defined by

$$\begin{cases} \dfrac{x_k - x_{k-1}}{\lambda_k} \in -\partial\Phi(x_k) \text{ for } k \geq 1, \\ x_0 \text{ given in } \operatorname{dom}\partial\Phi. \end{cases}$$

Such a sequence $(x_k)_{k\in\mathbf{N}}$ is well defined and satisfies the formula $x_k = J_{\lambda_k}x_{k-1} = (I + \lambda_k\partial\Phi)^{-1}x_{k-1}$ for all $k \geq 1$. Consequently if the step size λ_k is constant equal to λ, then

$$x_k = J_\lambda^k x_0 = \left(I + \lambda\partial\Phi\right)^{-k} x_0.$$

It should be noted that $x_k \in \operatorname{dom}\partial\Phi$ for all $k \geq 1$.

These sequences play an important role in the discrete (or difference) approximation of gradient flow problems in the general context of monotone operators. More precisely let $t_0 = 0 < t_1 < \cdots t_{k-1} < t_k < \cdots t_n = T$ be a finite partition of $[0,T]$, set $\lambda_k = t_k - t_{k-1} > 0$ for $k = 1,\ldots,n$, and consider the proximal sequence associated with $(\lambda_k)_{k=1,\ldots,N}$ and $\partial\Phi$. Then, under some additional conditions, the step function $u_n := \sum_{k=1}^n \mathbf{1}_{]t_{k-1},tk]}x_k$ gives rise to the notion of (backward) DS-approximate solution of the corresponding Cauchy problem in $[0,T]$ (see [259, 260] and Section 17.6). We do not address this issue in this section.

Among other things, the following lemma provides an estimation for the distance between two proximal sequences $(x_k)_{k\in\mathbf{N}}$ and $(\hat{x}_l)_{l\in\mathbf{N}}$ with step size λ_k and $\hat{\lambda}_l$, respectively, associated with $\partial\Phi$ (or a maximal monotone operator A). We use the notation

$$\sigma_k = \sum_{i=1}^{k} \lambda_i, \ \tau_k = \sum_{i=1}^{k} \lambda_i^2 \quad \text{(similarly for } \hat{\sigma}_k \text{ and } \hat{\tau}_k)$$

for $k \geq 1$ and set $\sigma_0 = \tau_0 = 0$.

Lemma 17.2.2 (Kobayashi inequality). *Let $(x_k)_{k\in\mathbf{N}}$ and $(\hat{x}_l)_{l\in\mathbf{N}}$ be two proximal sequences as above. Then for every $v \in \operatorname{dom}\partial\Phi$ and all $(k,l)\in\mathbf{N}^2$,*

$$\|x_k - \hat{x}_l\| \leq \|x_0 - v\| + \|\hat{x}_0 - v\| + \|\partial\Phi(v)^0\| \sqrt{(\sigma_k - \hat{\sigma}_l)^2 + \tau_k + \tau_l}. \tag{17.102}$$

PROOF. To shorten notation we set $a_{k,l} = \sqrt{(\sigma_k - \hat{\sigma}_l)^2 + \tau_k + \tau_l}$, $y_k = \frac{x_{k-1}-x_k}{\lambda_k}$, and we replace $\partial\Phi$ by a maximal monotone operator A. We reason by induction on (k,l).

We begin by establishing (17.102) for $(k,0)$ and $(0,l)$. Since J_{λ_1} is nonexpansive, and noticing that $v = J_{\lambda_1}(v + \lambda_1 Av^0)$ and $x_1 = J_{\lambda_1} x_0$, we have

$$\begin{aligned}\|x_1 - v\| &\leq \|x_0 - v - \lambda_1 Av^0\| \\ &\leq \|x_0 - v\| + \lambda_1 \|Av^0\|.\end{aligned}$$

Hence, by induction

$$\|x_k - v\| \leq \|x_0 - v\| + \sigma_k \|Av^0\|.$$

Consequently

$$\begin{aligned}\|x_k - \hat{x}_0\| &\leq \|x_k - v\| + \|v - \hat{x}_0\| \\ &\leq \|x_0 - v\| + \sigma_k\|Av^0\| + \|v - \hat{x}_0\| \\ &\leq \|x_0 - v\| + \|v - \hat{x}_0\| + a_{k,0}\|Av^0\|\end{aligned}$$

because $\sigma_k \leq a_{k,0}$. The sequences $(x_k)_{k\in\mathbf{N}}$ and $(\hat{x}_l)_{l\in\mathbf{N}}$ playing a symmetrical role, inequality (17.102) also holds for $(0,l)$.

To continue the proof, we need the following technical lemma.

Lemma 17.2.3. *Let (x,y) and $(\hat{x},\hat{y})$ be two elements of A, and $\lambda \geq 0$, $\mu \geq 0$ in $\mathbf{R}$. Then we have the following inequality:*

$$(\lambda+\mu)\|x - \hat{x}\| \leq \lambda\|\hat{x} + \mu\hat{y} - x\| + \mu\|x + \lambda y - \hat{x}\|. \tag{17.103}$$

PROOF OF LEMMA 17.2.3. It suffices to follow the calculation

$$\begin{aligned}(\lambda+\mu)\|x-\hat{x}\|^2 &= \lambda\langle x-\hat{x}, x-\hat{x}\rangle + \mu\langle x-\hat{x}, x-\hat{x}\rangle \\ &= \lambda\langle x-\hat{x}-\mu\hat{y}, x-\hat{x}\rangle + \mu\langle x-\hat{x}+\lambda y, x-\hat{x}\rangle + \lambda\mu\langle \hat{y}-y, x-\hat{x}\rangle \\ &\leq \lambda\langle x-\hat{x}-\mu\hat{y}, x-\hat{x}\rangle + \mu\langle x-\hat{x}+\lambda y, x-\hat{x}\rangle \\ &\leq \Big(\lambda\|\hat{x}+\mu\hat{y}-x\| + \mu\|x+\lambda y-\hat{x}\|\Big)\|x-\hat{x}\|,\end{aligned}$$

where we have used the monotonicity of A to claim that $\lambda\mu\langle \hat{y}-y, x-\hat{x}\rangle \leq 0$.

PROOF OF LEMMA 17.2.2 CONTINUED. Assume that (17.102) holds for $(k-1,l)$ and $(k,l-1)$. Applying Lemma 17.2.3 we infer

$$(\lambda_k+\hat{\lambda}_l)\|x_k-\hat{x}_l\| \le \lambda_k\|\hat{x}_l+\hat{\lambda}_l\hat{y}_l-x_k\|+\hat{\lambda}_l\|x_k+\lambda_k y_k-\hat{x}_l\|. \tag{17.104}$$

Hence, setting $\alpha_{k,l}:=\frac{\hat{\lambda}_l}{\lambda_k+\hat{\lambda}_l}$ and $\beta_{k,l}:=\frac{\lambda_k}{\lambda_k+\hat{\lambda}_l}$, from (17.104) we infer

$$\|x_k-\hat{x}_l\| \le \alpha_{k,l}\|x_{k-1}-\hat{x}_l\|+\beta_{k,l}\|\hat{x}_{l-1}-x_k\|.$$

Using the induction hypothesis we obtain

$$\begin{aligned}\|x_k-\hat{x}_l\| &\le \alpha_{k,l}\Big[\|x_0-v\|+\|\hat{x}_0-v\|+a_{k-1,l}\|Au^0\|\Big]\\ &\quad+\beta_{k,l}\Big[\|x_0-v\|+\|\hat{x}_0-v\|+a_{k,l-1}\|Au^0\|\Big]\\ &=\|x_0-v\|+\|\hat{x}_0-v\|+(\alpha_{k,l}a_{k-1,l}+\beta_{k,l}a_{k,l-1})\|Au^0\|.\end{aligned}$$

To conclude, it suffices to check that $\alpha_{k,l}a_{k-1,l}+\beta_{k,l}a_{k,l-1}\le a_{k,l}$ (see [319, Proposition 2.12]). □

PROOF OF THEOREM 17.2.4. The proof proceeds in three steps.

Step 1. Existence and Lipschitz continuity of the limit (17.101). Since $J^n_{\frac{t}{n}}$ is nonexpansive, the limit (17.101) exists iff it exists for $u_0\in\operatorname{dom}\partial\Phi$. Hence, in what follows, we assume that $u_0\in\operatorname{dom}\partial\Phi$. For each fixed $T>0$, let us define the function $u_n:[0,T]\to\mathcal{H}$ by $u_n(t)=(I+\frac{t}{n}\partial\Phi)^{-n}u_0$.

Let $(m,n)\in\mathbf{N}^*\times\mathbf{N}^*$, $s>0$, $t>0$ in $[0,T]$, and consider the two proximal sequences $(x_k)_{k\in\mathbf{N}}$ and $(\hat{x}_l)_{l\in\mathbf{N}}$ associated with the two step sizes $\lambda_k=s/m$ and $\hat{\lambda}_l=t/n$, respectively, and with initial condition $x_0=\hat{x}_0=u_0$. Noticing that $\hat{x}_n=u_n(t)$ and $x_m=u_m(s)$, from Lemma 17.2.2 (take $v=u_0$) we infer

$$\|u_n(t)-u_m(s)\|\le\|\partial\Phi(u_0)^0\|\sqrt{(t-s)^2+\frac{s^2}{m}+\frac{t^2}{n}}. \tag{17.105}$$

Taking $t=s$, (17.105) yields that $(u_n)_{n\in\mathbf{N}}$ is a Cauchy sequence in $\mathbf{C}([0,T],\mathcal{H})$. Consequently it uniformly converges to some function u (for $t=s=0$, $u_n(t)=u_m(t)=u_0$). Taking $m=n$ and going to the limit on n in (17.105), we conclude that u is a $\|\partial\Phi(u_0)^0\|$-Lipschitz function in $\mathbf{C}([\delta,T],\mathcal{H})$ for every $\delta>0$. Note that $u(t)\in\overline{\operatorname{dom}\partial\Phi}$, since $u_n(t)$ belongs to $\operatorname{dom}\partial\Phi$. Actually we will see in the last step that $u(t)$ belongs to $\operatorname{dom}\partial\Phi$.

Step 2. We prove that $\dot{u}(t)\in-\partial\Phi(u(t))$ for a.e. t in $[0,+\infty)$.

For any $\lambda>0$, we look at the unique global classical solution $u_\lambda:[0,+\infty)\to\mathcal{H}$ of the Cauchy problem (17.91) considered in the previous section. From Theorem 17.2.3, for all $t>0$, $u_\lambda(t)$ converges to $\bar{u}(t)$ when $\lambda\to 0$, where $\bar{u}$ satisfies $\dot{\bar{u}}(t)\in-\partial\Phi(\bar{u}(t))$ for a.e. t in $[0,+\infty)$. Therefore, in order to complete the proof, it suffices to establish that for all $t>0$, $u_\lambda(t)\to u(t)$ when $\lambda\to 0$.

For each fixed $t>0$ and each λ intended to go to 0, $t>\lambda>0$, choose m the integer part of $\frac{t}{\lambda}$, thus satisfying $t=\lambda m+\delta$ with $0\le\delta<\lambda$, and write

$$\|u_\lambda(t)-u(t)\|\le\|u_\lambda(t)-u_\lambda(m\lambda)\|+\|u_\lambda(\lambda m)-J^m_\lambda u_0\|+\|J^m_\lambda u_0-u(m\lambda)\|+\|u(m\lambda)-u(t)\|. \tag{17.106}$$

We are going to estimate each of the four terms of the right-hand side of (17.106). We know that u_λ and u are Lipschitz continuous with a constant Lipschitz equal to $||\partial\Phi(u_0)^0||$. For the function u this property holds in each interval $[\delta, T]$ with $\delta > 0$ and has been established in Step 1. Thus, since $m\lambda > 0$,

$$\begin{aligned} ||u_\lambda(t)-u_\lambda(m\lambda)|| &\leq |t-m\lambda|||\partial\Phi(u_0)^0|| \leq \lambda||\partial\Phi(u_0)^0||, \\ ||u(m\lambda)-u(t)|| &\leq |t-m\lambda|||\partial\Phi(u_0)^0|| \leq \lambda||\partial\Phi(u_0)^0||. \end{aligned} \tag{17.107}$$

On the other hand, with the notation of the first step, and from (17.105),

$$\begin{aligned} ||J_\lambda^m u_0 - u(m\lambda)|| &= ||J^m_{\frac{t-\delta}{m}} u_0 - u(m\lambda)|| \\ &= ||u_m(t-\delta)-u(t-\delta)|| \\ &\leq ||u_m(t-\delta)-u_m(t)|| + ||u_m(t)-u(t)|| + ||u(t)-u(t-\delta)|| \\ &\leq \lambda||\partial\Phi(u_0)^0|| + ||u_m(t)-u(t)|| + ||\partial\Phi(u_0)^0||\sqrt{\delta^2 + \frac{t^2+(t-\delta)^2}{m}}. \end{aligned}$$

Noticing that $\lambda \to 0 \Longrightarrow m \to +\infty$, from the first step and the estimate above we infer that

$$\lim_{\lambda\to 0} ||J_\lambda^m u_0 - u(m\lambda)|| = 0. \tag{17.108}$$

The second term is more complex to estimate and requires the following lemma.

Lemma 17.2.4 (Chernoff). *Let $J : \mathcal{H} \to \mathcal{H}$ be a nonexpansive operator, $v_0 \in \mathcal{H}$, and v the strong global solution of the Cauchy problem in $[0,+\infty)$,*

$$\begin{cases} \dot v(t) = -\frac{1}{\lambda}(I-J)v(t), \\ v(0) = v_0. \end{cases} \tag{17.109}$$

Then, for all $n \in \mathbf{N}$ and all $t \geq 0$,

$$||v(t)-J^n v_0|| \leq ||\dot v(0)||\sqrt{\lambda t + (n\lambda - t)^2}. \tag{17.110}$$

PROOF OF LEMMA 17.2.4. According to Theorem 17.1.2, (17.109) is well-posed in the sense that there exists a unique solution v. It is enough to consider the case $\lambda = 1$ (otherwise apply the inequality to the function v_λ defined by $v_\lambda(t) = v(\lambda t)$). We establish (17.110) inductively. For $n = 0$, due to the fact that $t \mapsto ||\dot v(t)||$ is decreasing (a consequence of $v \mapsto (I-J)$ monotone), we infer

$$||v(t)-v_0|| = \left\|\int_0^t \dot v(s)\, ds\right\| \leq \int_0^t ||\dot v(s)||\, ds \leq t||\dot v(0)|| \leq ||\dot v(0)||\sqrt{t+t^2},$$

and (17.110) holds. Assume that (17.110) holds for $n-1$. From

$$e^t \dot v(t) + e^t v(t) = e^t J v(t)$$

we obtain

$$v(t) = v_0 e^{-t} + \int_0^t e^{s-t} J v(s)\, ds.$$

Hence

$$\begin{aligned}\|v(t)-J^n v_0\| &= \left\|(v_0-J^n v_0)e^{-t}+\int_0^t e^{s-t}\Big(Jv(s)-J^n v_0\Big)\,ds\right\| \\ &\le e^{-t}\left(\|v_0-J^n v_0\|+\int_0^t e^{s}\|v(s)-J^{n-1}v_0\|\,ds\right),\end{aligned}$$

where we have used the fact that J is nonexpansive. Noticing that

$$\|v_0-J^n v_0\| \le \sum_{k=1}^{n}\|J^{k-1}v_0-J^k v_0\| \le n\|v_0-Jv_0\| = n\|\dot v(0)\|,$$

we infer that

$$\|v(t)-J^n v_0\| \le e^{-t}\left(n\|\dot v(0)\|+\int_0^t e^{s}\|v(s)-J^{n-1}v_0\|\,ds\right).$$

We complete the proof by using the induction hypothesis and the inequality

$$n+\int_0^t e^s\sqrt{s+\big((n-1)-s\big)^2}\,ds \le e^t\sqrt{t+(n-t)^2}$$

obtained in an elementary way ($h(t) := n+\int_0^t e^s\sqrt{s+((n-1)-s)^2}\,ds - e^t\sqrt{t+(n-t)^2}$ satisfies $h(0)=0$ and $h'(t)\le 0$).

CONCLUSION OF THE PROOF OF STEP 2. We go back to the estimate of the second term $\|u_\lambda(\lambda m)-J_\lambda^m u_0\|$ of (17.106). Applying Lemma 17.2.4 with $J := J_\lambda$, thus $v=u_\lambda$, and using the notation of Proposition 17.2.1, we obtain

$$\|u_\lambda(\lambda m)-J_\lambda^m u_0\| \le \|A_\lambda(u_0)\|\lambda\sqrt{m} \le \|A_\lambda(u_0)\|\sqrt{\lambda t}.$$

But from Proposition 17.2.2, $\|A_\lambda(u_0)\| \le \|\partial\Phi(u_0)^0\|$ so that

$$\|u_\lambda(\lambda m)-J_\lambda^m u_0\| \le \|\partial\Phi(u_0)^0\|\sqrt{\lambda t}. \tag{17.111}$$

Collecting (17.107), (17.108), and (17.111), inequality (17.106) yields $\lim_{\lambda\to 0}\|u_\lambda(t)-u(t)\| = 0$, which completes the proof. □

Remark 17.2.3. When $\partial\Phi$ is single valued and Lipschitz continuous with some constant $L>0$, the proof of Step 2 above, i.e., $\dot u(t) = -\nabla\Phi(u(t))$ for a.e. t in $[0,+\infty)$, can be achieved in a more direct and independent way.

For any $t\ge 0$ let us define the operator $\mathcal{T}(t) : u_0 \in \overline{\operatorname{dom}\partial\Phi} \mapsto \mathcal{T}(t)u_0 := u(t) \in \overline{\operatorname{dom}\partial\Phi}$, where u is the limit obtained in the first step above. We first prove that $\mathcal{T}$ satisfies the semigroup property: $\mathcal{T}(s+t)=\mathcal{T}(s)\mathcal{T}(t)$. Indeed for $m\in\mathbf{N}$ we have

$$\mathcal{T}(mt) = \lim_{n\to+\infty} J^n_{\frac{mt}{n}} = \lim_{p\to+\infty} J^{mp}_{\frac{t}{p}} = \lim_{p\to+\infty}\big[J^p_{\frac{t}{p}}\big]^m = \big[\mathcal{T}(t)\big]^m,$$

where, in the third equality, we took into account the continuity of the resolvent operator. (Recall that the resolvent is nonexpansive.) Now, let p, q, p', q' be positive integers. Then, from the above,

$$\mathcal{T}\left(\frac{p}{q}+\frac{p'}{q'}\right)=\mathcal{T}\left(\frac{pq'+p'q}{qq'}\right)=\mathcal{T}\left(\frac{1}{qq'}\right)^{pq'+p'q}$$
$$=\mathcal{T}\left(\frac{1}{qq'}\right)^{pq'}\mathcal{T}\left(\frac{1}{qq'}\right)^{p'q}=\mathcal{T}\left(\frac{p}{q}\right)T\left(\frac{p'}{q'}\right).$$

Hence $\mathcal{T}(s+t)=\mathcal{T}(s)\mathcal{T}(t)$ holds if t and s are rational. The claim follows in view of the continuity in t.

Since u is a Lipschitz continuous function on $[\delta,T]$ for all $\delta>0$ and all $T>0$, u is absolutely continuous on each interval $[\delta,T]$ and thus is a.e. differentiable on $(0,+\infty)$. In what follows t is fixed in $(0,+\infty)$ in such a way that

$$\lim_{h\to 0}\frac{u(t+h)-u(t)}{h}$$

exists. According to the semigroup property we have

$$\frac{u(t+h)-u(t)}{h}=\frac{\mathcal{T}(h)\mathcal{T}(t)u_0-\mathcal{T}(t)u_0}{h}=\lim_{n\to 0}\frac{J^n_{\frac{h}{n}}\mathcal{T}(t)u_0-\mathcal{T}(t)u_0}{h}.$$

Consider the proximal sequence associated with $\nabla\Phi$, with constant size $\frac{h}{n}$, and initialized in $\mathcal{T}(t)u_0=u(t)$:

$$\begin{cases}\dfrac{x_k-x_{k-1}}{h/n}\in-\nabla\Phi(x_k) & \text{for } k\geq 1,\\ x_0=u(t).\end{cases}$$

The difference quotient $\frac{J^n_{\frac{h}{n}}\mathcal{T}(t)u_0-\mathcal{T}(t)u_0}{h}$ may be written in terms of the proximal sequence above. Indeed one has

$$\begin{aligned}\frac{J^n_{\frac{h}{n}}\mathcal{T}(t)u_0-\mathcal{T}(t)u_0}{h}&=\frac{1}{n}\frac{x_n-x_0}{h/n}\\&=\frac{1}{n}\sum_{k=1}^{n}\frac{x_k-x_{k-1}}{h/n}\\&=-\frac{1}{n}\sum_{k=1}^{n}y_k,\end{aligned}$$

where $y_k=\nabla\Phi(x_k)$. Thus

$$\frac{u(t+h)-u(t)}{h}=\lim_{n\to+\infty}-\frac{1}{n}\sum_{k=1}^{n}y_k. \tag{17.112}$$

On the other hand, since $\nabla\Phi$ is Lipschitz continuous, for $k=1,\dots,n$ we have

$$||y_k-\nabla\Phi(u(t))||\leq L||x_k-u(t)||. \tag{17.113}$$

We claim that

$$||x_k-u(t)||\leq k\frac{h}{n}||\nabla\Phi(u(t))||.$$

Indeed, using the fact that the resolvent is nonexpansive, we have for $i = 1, \dots, k$

$$\begin{aligned}
\|x_i - u(t)\| &\leq \left\| x_i + \frac{h}{n}\nabla\Phi(x_i) - \left(u(t) + \frac{h}{n}\nabla\Phi(u(t))\right)\right\| \\
&= \left\| x_i + x_{i-1} - x_i - \left(u(t) + \frac{h}{n}\nabla\Phi(u(t))\right)\right\| \\
&= \left\| x_{i-1} - u(t) - \frac{h}{n}\nabla\Phi(u(t))\right\| \\
&\leq \|x_{i-1} - u(t)\| + \frac{h}{n}\|\nabla\Phi(u(t))\|.
\end{aligned}$$

The claim follows by summing these inequalities for $i = 1, \dots, k$. Thus (17.113) yields

$$\|y_k - \nabla\Phi(u(t))\| \leq Lk\frac{h}{n}\|\nabla\Phi(u(t))\|. \tag{17.114}$$

Now, according to (17.112) and from (17.114), for all $(x, y) \in \nabla\Phi$ we have

$$\begin{aligned}
&\left\langle -\frac{u(t+h) - u(t)}{h} - y, u(t) - x \right\rangle \\
&= \lim_{n\to+\infty} \left\langle \nabla\Phi(u(t)) - y + \left(\frac{1}{n}\sum_{k=1}^{n} y_k - \nabla\Phi(u(t))\right), u(t) - x \right\rangle \\
&\geq \left\langle \nabla\Phi(u(t)) - y, u(t) - x \right\rangle - \sup_{n\in\mathbf{N}^*} \left| \left\langle \frac{1}{n}\sum_{k=1}^{n} y_k - \nabla\Phi(u(t)), u(t) - x \right\rangle \right| \\
&\geq \left\langle \nabla\Phi(u(t)) - y, u(t) - x \right\rangle - L|h|\|\nabla\Phi(u(t))\|\,\|u(t) - x\| \\
&\geq -L|h|\|\nabla\Phi(u(t))\|\,\|u(t) - x\|.
\end{aligned}$$

Letting $h \to 0$ we finally infer that for all $(x, y) \in \nabla\Phi$

$$\left\langle -\dot{u}(t) - y, u(t) - x \right\rangle \geq 0.$$

From the maximality of $\nabla\Phi$ we deduce that $-\dot{u}(t) = \nabla\Phi(u(t))$, which completes the proof.

17.2.5 ▪ The nonautonomous case: Time-dependent convex potentials

In this section, we consider quasi-autonomous evolution equations

$$\begin{cases} \dot{u}(t) + \partial\Phi(u(t)) \ni f(t), \\ u(0) = u_0. \end{cases}$$

The function $\Phi : \mathscr{H} \to \mathbf{R} \cup \{+\infty\}$ is convex, lower semicontinuous, and proper. It can be interpreted as the internal potential energy of the system, while the second element $f \in L^2(0, T; \mathscr{H})$ represents the external action. This is a particular case (take $\Phi(t, v) = \Phi(v) - \langle f(t), v\rangle$) of the general nonautonomous evolution equation

$$\begin{cases} \dot{u}(t) + \partial\Phi(t, u(t)) \ni 0, \\ u(0) = u_0, \end{cases} \tag{17.115}$$

where the operator $\partial\Phi(t,\cdot)$ is the subdifferential of a convex, lower semicontinuous, and proper function $\Phi(t,\cdot):\mathcal{H}\to\mathbf{R}\cup\{+\infty\}$, which varies with t.

By using techniques similar to those used in the autonomous case ($f=0$), the following results were obtained (see [134], [135] for further details).

Theorem 17.2.5. *Let $\Phi:\mathcal{H}\to\mathbf{R}\cup\{+\infty\}$ be a convex, lower semicontinuous, and proper function. Then, for every $f\in L^2(0,T;\mathcal{H})$, and $u_0\in\overline{\mathrm{dom}\Phi}$, there is a unique strong solution $u\in C([0,T];\mathcal{H})$ of the Cauchy problem*

$$\begin{cases} \dot{u}(t)+\partial\Phi(u(t))\ni f(t), \\ u(0)=u_0. \end{cases}$$

Moreover the following properties hold:

(i) $u(t)\in\mathrm{dom}\,\partial\Phi$ *a.e.* $t\in(0,T)$.

(ii) $\sqrt{t}\,\dot{u}\in L^2(0,T;\mathcal{H})$.

(iii) *For a.e.* $t\in(0,T)$, *u is derivable at t, and*

$$\dot{u}(t)+(\partial\Phi(u(t))-f(t))^0=0,$$

where $(\partial\Phi(u(t))-f(t))^0$ is the element of minimal norm of $\partial\Phi(u(t))-f(t)$.

(iv) *$t\mapsto\Phi(u(t))$ is absolutely continuous on each interval $[\delta,T]$, and*

$$\frac{d}{dt}\Phi(u(t))+\|\dot{u}(t)\|^2=\langle f(t),\dot{u}(t)\rangle \quad \textit{for almost all } t>0.$$

(v) *If $u_0\in\mathrm{dom}\,\Phi$, then $\dot{u}\in L^2(0,T;\mathcal{H})$, and $t\mapsto\Phi(u(t))$ is continuous on $[0,T]$.*

When the second member is absolutely continuous, we can obtain additional regularity results that make quasi-autonomous and autonomous cases very similar.

Theorem 17.2.6. *Let $\Phi:\mathcal{H}\to\mathbf{R}\cup\{+\infty\}$ be a convex, lower semicontinuous, and proper function. Let $f\in W^{1,1}(0,T;\mathcal{H})$ (i.e., f is absolutely continuous). Then, for any $u_0\in\overline{\mathrm{dom}\,\Phi}$, there exists a unique strong solution $u:[0,T]\to\mathcal{H}$ of the Cauchy problem*

$$\begin{cases} \dot{u}(t)+\partial\Phi(u(t))\ni f(t), \\ u(0)=u_0. \end{cases}$$

Moreover the following properties hold:

(i) $u(t)\in\mathrm{dom}\,\partial\Phi$ *for all* $t\in\,]0,T]$.

(ii) $t\|\dot{u}(t)\|\in L^\infty[0,T]$.

(iii) *For each $t\geq 0$, u has a right derivative, and*

$$\frac{d^+u}{dt}(t)+(\partial\Phi(u(t))-f(t))^0=0,$$

where $\partial\Phi(u(t))^0$ is the element of minimal norm of $\partial\Phi(u(t))$.

(iv) *For each* $t \geq 0$, $t \mapsto \Phi(u(t))$ *has a right derivative, and*

$$\frac{d^+}{dt}\Phi(u(t)) + \left\|\frac{d^+u}{dt}(t)\right\|^2 = \langle f(t), \dot{u}(t)\rangle \quad \forall t \in]0,T].$$

Remark 17.2.4. Evolution equations governed by time-dependent subdifferential operators are used to model a wide range of situations. A well-known example is the sweeping process introduced by Moreau [298], which plays an important role in unilateral mechanics, economics, and control; see [265] for a recent survey. The equation has the form

$$\begin{cases} \dot{u}(t) + N_{C(t)}(u(t)) \ni f(t), \\ u(0) = u_0, \end{cases}$$

where $C(t)$ is a time-dependent (moving) convex set in $\mathcal{H}$, and $N_{C(t)}(u)$ is the normal cone to $C(t)$ at $u \in C(t)$. Since $N_{C(t)}$ is the subdifferential of the indicator function of $C(t)$ we are in the framework of (17.115).

For further examples and results concerning evolution equations governed by time-dependent subdifferential operators, see [46], [47], [48], [253], [254], [263], [313].

17.2.6 ▪ Gradient flow for a convex potential: Asymptotic analysis, $t \to +\infty$

Let $\Phi : \mathcal{H} \to \mathbf{R} \cup \{+\infty\}$ be a convex, lower semicontinuous, and proper function, which is minorized, i.e., $\inf_{\mathcal{H}} \Phi > -\infty$. Following Theorem 17.2.3, given $u_0 \in \overline{\mathrm{dom}\,\Phi}$, there exists a unique strong global solution $u : [0,+\infty) \to \mathcal{H}$ of the Cauchy problem

$$\begin{cases} \dot{u}(t) + \partial\Phi(u(t)) \ni 0, \\ u(0) = u_0. \end{cases}$$

We are interested in the asymptotic behavior of $u(t)$ and $\Phi(u(t))$ as $t \to +\infty$.

Minimizing property. The following minimization property holds, even if the set of minimizers of Φ is empty.

Proposition 17.2.7 (minimizing property). (i) $t \mapsto \Phi(u(t))$ *is a decreasing function and*

$$\lim_{t\to+\infty} \Phi(u(t)) = \inf_{\mathcal{H}} \Phi.$$

(ii) *The following estimate holds: for any any* $v \in \mathrm{dom}\,\Phi$, *for any* $t > 0$,

$$\Phi(u(t)) \leq \Phi(v) + \frac{1}{2t}\|u_0 - v\|^2.$$

As a consequence, if $S = \arg\min \Phi \neq \emptyset$

$$\Phi(u(t)) \leq \inf_{\mathcal{H}} \Phi + \frac{1}{2t}\,\mathrm{dist}(u_0, S)^2.$$

PROOF. Because of the effect of regularization, when the asymptotic behavior is studied, it is not restrictive to assume that $u_0 \in \mathrm{dom}\,\partial\Phi$. For any $v \in \mathrm{dom}\,\Phi$, by $-\dot{u}(t) \in \partial\Phi(u(t))$, we have the subdifferential inequality

$$\Phi(v) \geq \Phi(u(t)) + \langle -\dot{u}(t), v - u(t)\rangle.$$

Equivalently,

$$\Phi(v) \geq \Phi(u(t)) + \frac{1}{2}\frac{d}{dt}||u(t)-v||^2.$$

After integration from 0 to T

$$0 \geq \int_0^T (\Phi(u(t)) - \Phi(v))dt + \frac{1}{2}||u(T)-v||^2 - \frac{1}{2}||u_0 - v||^2.$$

Since $t \mapsto \Phi(u(t))$ is nonincreasing

$$0 \geq T(\Phi(u(T)) - \Phi(v)) + \frac{1}{2}||u(T)-v||^2 - \frac{1}{2}||u_0 - v||^2.$$

As a consequence, for any $t > 0$,

$$\Phi(u(t)) \leq \Phi(v) + \frac{1}{2t}||u_0 - v||^2.$$

Passing to the limit, as $t \to +\infty$, gives

$$\lim_{t\to+\infty} \Phi(u(t)) \leq \Phi(v).$$

This inequality being valid for any $v \in \operatorname{dom}\Phi$, we obtain

$$\lim_{t\to+\infty} \Phi(u(t)) \leq \inf_{\mathcal{H}} \Phi.$$

The opposite inequality is trivially satisfied, which gives the result. □

Velocity goes to zero. The following result holds, even if the set of minimizers of Φ is empty.

Proposition 17.2.8. (i) $t \mapsto ||\frac{d^+u}{dt}(t)||$ *is nonincreasing and*

$$\lim_{t\to+\infty} \left\|\frac{d^+u}{dt}(t)\right\| = 0.$$

(ii) *The following estimate holds:*

$$\left\|\frac{d^+u}{dt}(t)\right\| \leq \frac{C}{\sqrt{t}}. \tag{17.116}$$

PROOF. Because of the regularization effect, when the asymptotic behavior is studied, it is not restrictive to assume that $u_0 \in \operatorname{dom}\partial\Phi$. By integration of the energy estimate (17.90), for any $t > 0$

$$\int_0^t ||\dot{u}(\tau)||^2 d\tau \leq \Phi(u_0) - \inf_{\mathcal{H}} \Phi.$$

Since for each $t > 0$, u has a right derivative, and $t \mapsto ||\frac{d^+u}{dt}(t)||$ is nonincreasing, we deduce that, for each $t > 0$,

$$t\left\|\frac{d^+u}{dt}(t)\right\|^2 \leq \Phi(u_0) - \inf_{\mathcal{H}} \Phi,$$

which gives the result. □

Weak convergence results. In order to prove the weak convergence of the orbits of (GSD) we use the classical Opial's lemma. We recall it in its continuous form and give a short proof of it.

Lemma 17.2.5. *Let S be a nonempty subset of $\mathcal{H}$ and $u : [0,+\infty) \to H$ a map. Assume that*

(i) *for every $z \in S$, $\lim_{t\to+\infty} \|u(t)-z\|$ exists;*

(ii) *every sequential weak cluster point of the map u belongs to S.*

Then

$$w - \lim_{t\to+\infty} u(t) = u_\infty \quad \text{exists for some element } u_\infty \in S.$$

PROOF. By (i) and $S \neq \emptyset$, the trajectory u is asymptotically bounded in $\mathcal{H}$. In order to obtain its weak convergence, we just need to prove that the trajectory has a unique sequential weak cluster point. Let $u(t_n^1) \rightharpoonup z^1$ and $u(t_n^2) \rightharpoonup z^2$, with $t_n^1 \to +\infty$, and $t_n^2 \to +\infty$. By (ii), $z^1 \in S$, and $z^2 \in S$. By (i), it follows that $\lim_{t\to+\infty} \|u(t)-z^1\|$ and $\lim_{t\to+\infty} \|u(t)-z^2\|$ exist. Hence, $\lim_{t\to+\infty}(\|u(t)-z^1\|^2 - \|u(t)-z^2\|^2)$ exists. Developing and simplifying this last expression, we deduce that

$$\lim_{t\to+\infty} \left\langle u(t), z^2 - z^1 \right\rangle \text{ exists.}$$

Hence

$$\lim_{n\to+\infty} \left\langle u(t_n^1), z^2 - z^1 \right\rangle = \lim_{n\to+\infty} \left\langle u(t_n^2), z^2 - z^1 \right\rangle,$$

which gives $\|z^2 - z^1\|^2 = 0$, and hence $z^2 = z^1$. □

Theorem 17.2.7 (Bruck [142]). *Let $\Phi : \mathcal{H} \to \mathbf{R} \cup \{+\infty\}$ be a convex, lower semicontinuous, and proper function, and suppose that $S = \arg\min \Phi \neq \emptyset$. Let $u : [0,+\infty) \to \mathcal{H}$ be a strong global trajectory of the generalized steepest descent* (GSD). *Then $u(t)$ converges weakly to some $u_\infty \in S$, as $t \to +\infty$.*

PROOF. Let us apply Opial's lemma with $S = \arg\min \Phi$, which has been supposed to be nonempty. Let $z = w - \lim u(t_n)$ be a weak sequential cluster point of u with $t_n \to +\infty$. By Proposition 17.2.7 and the lower semicontinuity of Φ with respect to the weak topology of $\mathcal{H}$ we infer

$$\begin{aligned}\inf_{\mathcal{H}} \Phi &= \lim_{t\to+\infty} \Phi(u(t)) \\ &= \lim \Phi(u(t_n)) \\ &\geq \Phi(z),\end{aligned}$$

which yields $z \in S$ and proves item (ii) of Opial's lemma. Let us now show that for any $z \in S$, $t \mapsto \|u(t)-z\|$ is a nonincreasing function, which will prove item (i). Set $h_z(t) = \frac{1}{2}\|u(t)-z\|^2$, which is absolutely continuous on the bounded intervals. For almost all $t > 0$

$$\dot{h}_z(t) = \langle u(t) - z, \dot{u}(t) \rangle . \tag{17.117}$$

Since u is a strong global trajectory of the generalized steepest descent (GSD), we have $-\dot{u}(t) \in \partial \Phi(u(t))$. Hence, we have the subdifferential inequality

$$\Phi(z) \geq \Phi(u(t)) + \langle -\dot{u}(t), z - u(t) \rangle .$$

Since $\Phi(z) \le \Phi(u(t))$ we infer

$$\langle \dot{u}(t), u(t) - z \rangle \le 0,$$

which, with (17.117), gives $\dot{h}_z(t) \le 0$. As a consequence, the function $t \mapsto \|u(t) - z\|$ is nonincreasing. Since it is nonnegative, it converges as $t \to +\infty$. □

Corollary 17.2.1. *Concerning the asymptotic behavior of the orbits of the generalized steepest descent* (GSD)*, there are two types of situations:*

(i) $S = \arg\min \Phi \neq \emptyset$*, in which case every orbit* $u(t)$ *is bounded, and converges weakly to some* $u_\infty \in S$*, as* $t \to +\infty$*.*

(ii) $S = \arg\min \Phi = \emptyset$*, in which case every orbit* $u(t)$ *verifies* $\lim_{t \to +\infty} \|u(t)\| = +\infty$.

PROOF. Item (i) is just the Bruck theorem. To prove (ii), let us prove the reverse implication. Suppose that there exists an orbit u and a sequence $t_n \to +\infty$ such that $\sup_n \|u(t_n)\| < +\infty$. Let us show that this property implies $S = \arg\min \Phi \neq \emptyset$. Indeed, since $\sup_n \|u(t_n)\| < +\infty$, the orbit u admits at least a weak asymptotic sequential cluster point $\bar{u}$. By Proposition 17.2.7, we know that $\lim_{t \to +\infty} \Phi(u(t)) = \inf_{\mathcal{H}} \Phi$. Hence by the lower semicontinuity property of Φ for the weak topology, we infer $\Phi(\bar{u}) \le \lim_{t \to +\infty} \Phi(u(t)) = \inf_{\mathcal{H}} \Phi$. As a consequence $\bar{u} \in S$, and $S = \arg\min \Phi \neq \emptyset$. □

Weak versus strong convergence. Let us state Baillon's result [71, Proposition 1].

Theorem 17.2.8. *There exists a closed convex proper function* $\Phi : \mathcal{H} = l^2(\mathbf{N}) \to \mathbf{R} \cup \{+\infty\}$*, with* $\partial \Phi^{-1}(0) \neq \emptyset$*, such that the semigroup* $S(t)$ *generated by the maximal monotone operator* $A = \partial \Phi$ *satisfies the following property: there exists some* $a \in \overline{\operatorname{dom} \Phi}$ *such that* $S(t)a$ *does not converge strongly to an element of* $\partial \Phi^{-1}(0)$*.*

Bruck's theorem, Theorem 17.2.7, states that for all $u_0 \in \overline{\operatorname{dom} \Phi}$, $S(t)u_0$ converges weakly to an element of $\partial \Phi^{-1}(0)$. Thus, Baillon's counterexample is a constructive example of a closed convex proper function $\Phi : \mathcal{H} \to \mathbf{R} \cup \{+\infty\}$ and of a trajectory of the semigroup generated by $\partial \Phi$ which converges weakly and not strongly. Baillon's thesis [72] contains an extended version of [71] with a counterexample involving a convex function Φ of class $\mathscr{C}^1$. Let us state it below in a precise form.

Proposition 17.2.9. *There exists a closed convex proper function* $\Phi : \mathcal{H} = l^2(\mathbf{N}) \to \mathbf{R} \cup \{+\infty\}$*, with* $S = \partial \Phi^{-1}(0) \neq \emptyset$*, such that the semigroup* $S_\lambda(t)$ *generated by the Yosida approximation* $A_\lambda = \nabla \Phi_\lambda$ *of the maximal monotone operator* $A = \partial \Phi$ *satisfies the following property: there exists some* $a \in \operatorname{dom} \partial \Phi$*, and* $\lambda_0 > 0$*, such that for any* $0 < \lambda < \lambda_0$*,* $S_\lambda(t)a$ *does not converge strongly to an element of* $\partial \Phi^{-1}(0)$*.*

There are some important situations where the the orbits of the gradient flow associated to a convex potential Φ converge strongly:

(i) Φ inf-compact,

(ii) Φ strongly convex,

(iii) Φ is an even function.

Item (i) is a clear consequence of the fact that, on the bounded subsets of the lower level sets of Φ, weak and strong convergence coincide. Let us successively examine item (ii) and (iii).

The strongly convex case.

Proposition 17.2.10. *Suppose that $\Phi : \mathscr{H} \to \mathbf{R}$ is a strongly convex lower semicontinuous function. Let u be an orbit of the generalized steepest descent associated to Φ.*

(i) *Then, u converges strongly to the unique minimizer of Φ.*

(ii) *Suppose, moreover, that Φ is differentiable with $\nabla\Phi$ Lipschitz continuous on bounded sets. Then u has a finite length, i.e., $\int_0^{+\infty} \|\dot{u}(t)\| dt < +\infty$.*

PROOF. (i) For a strongly convex lower semicontinuous function (cf. (17.119) below), the set of minimizers is nonvoid and reduced to a single element. Moreover, any minimizing sequence converges strongly to the unique minimizer. By the minimization property Proposition 17.2.7, we infer the strong convergence of u.

(ii) Let us now suppose that Φ is differentiable with $\nabla\Phi$ Lipschitz continuous on bounded sets. Let us show that u (a classical orbit) has a finite length. As a direct consequence, this will give us another proof of the strong convergence property. The proof relies on the use of a Łojasiewicz inequality satisfied by a strongly convex function and is preparatory to the next section.

Let z be the unique minimizer of Φ (we have $\nabla\Phi(z) = 0$). Since u is bounded, there exists some $R > 0$ such that the whole orbit u is contained in the ball $\mathbf{B}(z, R)$. Let $L_R > 0$ be the Lipschitz constant of $\nabla\Phi$ on $\mathbf{B}(z, R)$. The classical derivation chain rule gives, for any $v \in \mathscr{H}$,

$$\begin{aligned}\Phi(v) - \Phi(z) &= \int_0^1 \langle \nabla\Phi(z + t(v - z)), v - z\rangle \, dt \\ &= \int_0^1 \langle \nabla\Phi(z + t(v - z)) - \nabla\Phi(z), v - z\rangle \, dt.\end{aligned}$$

By the Cauchy–Schwarz inequality, and integration on $[0, 1]$, we deduce that for any $v \in \mathbf{B}(z, R)$

$$|\Phi(v) - \Phi(z)| \leq \frac{L_R}{2} \|v - z\|^2. \tag{17.118}$$

On the other hand, by the strong convexity assumption on Φ, there exists some positive constant α such that, for any $v \in \mathscr{H}$,

$$\langle \nabla\Phi(v) - \nabla\Phi(z), v - z\rangle \geq \alpha \|v - z\|^2. \tag{17.119}$$

Since $\nabla\Phi(z) = 0$, by the Cauchy–Schwarz inequality, we deduce that

$$\|v - z\| \leq \frac{1}{\alpha} \|\nabla\Phi(v)\|. \tag{17.120}$$

Combining (17.118) and (17.120) we obtain

$$|\Phi(v) - \Phi(z)| \leq \frac{L_R}{2\alpha^2} \|\nabla\Phi(v)\|^2.$$

Equivalently, for any $v \in \mathbf{B}(z,R)$, $v \neq z$,

$$\frac{\|\nabla\Phi(v)\|}{(\Phi(v)-\Phi(z))^{\frac{1}{2}}} \geq \alpha\left(\frac{2}{L_R}\right)^{\frac{1}{2}}. \tag{17.121}$$

Let us introduce $h : [0,+\infty[\to [0,+\infty[$

$$h(t) := 2\,(\Phi(u(t))-\Phi(z))^{\frac{1}{2}}$$

and show that h is a strict Lyapunov function. Recalling that $\Phi(u(\cdot))$ is a C^1 function, the classical derivation chain rule gives (as long as $u(t) \neq z$)

$$\dot{h}(t) := \frac{\langle \nabla\Phi(u(t)), \dot{u}(t)\rangle}{(\Phi(u(t))-\Phi(z))^{\frac{1}{2}}}.$$

Using the gradient flow equation, we have

$$\dot{h}(t) + \frac{\|\nabla\Phi(u(t))\|^2}{(\Phi(u(t))-\Phi(z))^{\frac{1}{2}}} \leq 0.$$

Equivalently

$$\dot{h}(t) + \frac{\|\nabla\Phi(u(t))\|}{(\Phi(u(t))-\Phi(z))^{\frac{1}{2}}}\|\dot{u}(t)\| \leq 0.$$

Since $u(t) \in \mathbf{B}(z,R)$, by using the Łojasiewicz-type inequality (17.121), we obtain

$$\dot{h}(t) + \alpha\left(\frac{2}{L_R}\right)^{\frac{1}{2}}\|\dot{u}(t)\| \leq 0.$$

By integration of this inequality, and $h \geq 0$, we obtain $\int_0^{+\infty}\|\dot{u}(t)\|dt < +\infty$, which gives the finite length property of the orbits in the strongly convex case. □

The case Φ even. Strong convergence in the case Φ even is a rather unexpected result from Bruck [142]. A detailed proof is given below.

Proposition 17.2.11. *Let $\Phi : \mathcal{H} \to \mathbf{R}\cup\{+\infty\}$ be a convex, lower semicontinuous, and proper function, which is supposed to be even, i.e., $\Phi(-v) = \Phi(v)$ for all $v \in \mathcal{H}$. Let $u : [0,+\infty) \to \mathcal{H}$ be a strong global trajectory of the generalized steepest descent* (GSD)*. Then $u(t)$ converges strongly to some $u_\infty \in S = \operatorname{arg\,min}\Phi$, as $t \to +\infty$.*

PROOF. First note that for an even function, $\partial\Phi(-v) = -\partial\Phi(v)$, which implies $\partial\Phi(0) \ni 0$. Hence $S = \operatorname{arg\,min}\Phi$ is nonempty and contains the origin. By the Bruck theorem, $u(t)$ converges weakly to some $u_\infty \in S = \operatorname{arg\,min}\Phi$, as $t \to +\infty$. Let us show that there is strong convergence. Let us fix some $t_0 > 0$ and work on the interval $[0,t_0]$. For any $t \in [0,t_0]$, set

$$k(t) = \|u(t)-u(t_0)\|^2 - 2\|u(t)\|^2.$$

Derivation of k gives

$$\begin{aligned}\dot{k}(t) &= 2\,\langle u(t)-u(t_0), \dot{u}(t)\rangle - 4\,\langle u(t), \dot{u}(t)\rangle \\ &= -2\,\langle u(t)+u(t_0), \dot{u}(t)\rangle.\end{aligned} \tag{17.122}$$

On the other hand, using successively the nonincreasing property of $t \mapsto \Phi(u(t))$, the fact that Φ is even, and the convex subdifferential inequality associated to $-\dot{u}(t) \in \partial\Phi(u(t))$, we obtain

$$\begin{aligned}\Phi(u(t)) &\geq \Phi(u(t_0))\\ &\geq \Phi(-u(t_0))\\ &\geq \Phi(u(t)) + \langle -u(t_0) - u(t), -\dot{u}(t)\rangle .\end{aligned}$$

Hence

$$\langle u(t_0) + u(t), \dot{u}(t)\rangle \leq 0. \tag{17.123}$$

Combining (17.122) and (17.123) gives $\dot{k}(t) \geq 0$. Hence k is nondecreasing on $[0, t_0]$. As a consequence, for any $t \in [0, t_0]$, $k(t) \leq k(t_0)$. Equivalently

$$||u(t) - u(t_0)||^2 - 2||u(t)||^2 \leq -2||u(t_0)||^2,$$

that is, for any $0 < t < t_0$,

$$||u(t) - u(t_0)||^2 \leq 2||u(t)||^2 - 2||u(t_0)||^2. \tag{17.124}$$

We know that for any $z \in S$, $\lim_{t\to+\infty} ||u(t) - z||$ exists (see the proof of the Bruck theorem). In particular, since $0 \in S$, we infer that

$$\lim_{n\to+\infty} ||u(t)||^2 \text{ exists.} \tag{17.125}$$

From (17.124) and (17.125) we deduce that

$$\lim_{t,s\to+\infty,\ t<s} ||u(t) - u(s)||^2 = 0.$$

Thus, the Cauchy criteria at infinity is satisfied, which implies the strong convergence of u. □

17.2.7 ▪ Gradient-projection dynamics

Let us particularize the previous results to the case $\Phi = \Psi + \delta_C$, where $\Psi : \mathscr{H} \to \mathbf{R}$ is a convex differentiable function, and δ_C is the indicator function of a closed convex set $C \subset \mathscr{H}$. Minimizing Φ on $\mathscr{H}$ is equivalent to the constrained minimization problem

$$\min\{\Psi(v) : \ v \in C\}. \tag{17.126}$$

The Moreau–Rockafellar subdifferential additivity rule applies, giving $\partial\Phi = \partial\Psi + N_C$, where N_C is the (outward) normal cone mapping to C. In what follows, when playing with the normal cone mapping, we will make frequent use of the Moreau decomposition theorem [297]:

Theorem 17.2.9 (Moreau). *Let T be a closed convex cone of a real Hilbert space $\mathscr{H}$ and N be the polar cone, i.e., $N = \{v \in \mathscr{H} : \langle v, \xi\rangle \leq 0$ for all $\xi \in T\}$. Then, for all $v \in \mathscr{H}$ there exists a unique decomposition*

$$\begin{gathered} v = v_T + v_N, \quad v_T \in T,\ v_N \in N;\\ \langle v_T, v_N\rangle = 0. \end{gathered}$$

Moreover, $v_T = \mathrm{proj}_T(v)$ and $v_N = \mathrm{proj}_N(v)$.

1. Let us first examine the generalized steepest descent, when applied to $\Phi = \Psi + \delta_C$. It is the subdifferential inclusion

$$\dot{u}(t) + N_C(u(t)) + \nabla\Psi(u(t)) \ni 0.$$

The lazy property (see Theorem 17.2.2(iii)) gives that for each $t \geq 0$, u has a right derivative, and

$$\frac{d^+u}{dt}(t) = -\partial\Phi(u(t))^0,$$

where $\partial\Phi(u(t))^0$ is the element of minimal norm of $\partial\Phi(u(t))$. Using the Moreau theorem, an elementary computation gives

$$\begin{aligned} -\partial\Phi(u(t))^0 &= (-\nabla\Psi(u(t)) - N_C(u(t)))^0 \\ &= -\nabla\Psi(u(t))) - \operatorname{proj}_{N(u(t))}(-\nabla\Psi(u(t))) \\ &= \operatorname{proj}_{T(u(t))}(-\nabla\Psi(u(t))). \end{aligned}$$

Combining this result with Theorems 17.2.2 and 17.2.7 gives the next proposition.

Proposition 17.2.12. *Let $\Phi = \Psi + \delta_C$, where $\Psi : \mathscr{H} \to \mathbf{R}$ is a convex differentiable function, and δ_C is the indicator function of a closed convex set $C \subset \mathscr{H}$. Then, for any $u_0 \in C$, there exists a unique strong global solution $u : [0, +\infty) \to \mathscr{H}$ of the generalized steepest descent* (GSD)

$$\begin{cases} \dot{u}(t) + N_C(u(t)) + \nabla\Psi(u(t)) \ni 0, \\ u(0) = u_0. \end{cases}$$

The following properties are satisfied:

(i) *For any $t > 0$, u has a right derivative at t which satisfies*

$$\frac{d^+u}{dt}(t) = \operatorname{proj}_{T(u(t))}(-\nabla\Psi(u(t))).$$

(ii) $\Psi(u(t))$ *decreases to* $\inf_C \Psi$ *as t increases to $+\infty$.*

(iii) *If $S = \operatorname{argmin}_C \Psi$ is nonvoid, then $u(t)$ converges weakly to some $u_\infty \in S$, as $t \to +\infty$.*

In the above system, the trajectories satisfy a completely inelastic shock law at the boundary. When reaching the boundary of the constraint C, the normal component of the velocity vector is set to zero, and the trajectory restarts tangentially to the boundary. This is especially interesting for modeling in unilateral mechanics or PDEs.

2. Indeed, from the perspective of optimization, the previous system has a major drawback: the orbits ignore the constraints until they meet the boundary of C. Moreover, the vector field which governs the dynamic is discontinuous (at the boundary of the constraint). The following dynamic, which was first considered by Antipin [32] and Bolte [102], gives a positive answer to these questions. First note that the optimality condition for (17.126)

$$\nabla\Psi(u) + N_C(u) \ni 0 \tag{17.127}$$

can be equivalently formulated as

$$u - \operatorname{proj}_C(u - \mu\nabla\Psi(u)) = 0, \tag{17.128}$$

where μ is a positive parameter (arbitrarily chosen). To obtain it, just write (17.127) as

$$u + N_C(u) \ni u - \mu\nabla\Psi(u)$$

and use that the resolvent $(I + N_C)^{-1}$ of the normal cone mapping is precisely the projection operator on C. This transformation is widely used in convex optimization. It transforms a variational inequality into a fixed point problem governed by an operator which is hopefully nonexpansive. The dynamic whose stationary points are the solutions of (17.128) is given by

$$\dot{u}(t) + u(t) - \mathrm{proj}_C\,(u(t) - \mu\nabla\Psi(u(t))) = 0$$

and is called the relaxed gradient-projection dynamic. The dynamic is now governed by a Lipschitz continuous vector field, and the orbits are classical solutions, i.e., continuously differentiable. Its properties are summarized in the following proposition.

Proposition 17.2.13. *Let $\Psi : \mathcal{H} \to \mathbf{R}$ be a convex differentiable function whose gradient is Lipschitz continuous on bounded sets. Let C be a closed convex set in $\mathcal{H}$, and suppose that Ψ is bounded from below on C. Then, for any $u_0 \in \mathcal{H}$, there exists a unique classical global solution $u : [0, +\infty[\to \mathcal{H}$ of the Cauchy problem for the relaxed gradient-projection dynamic*

$$\begin{cases} \dot{u}(t) + u(t) - \mathrm{proj}_C\,(u(t) - \mu\nabla\Psi(u(t))) = 0; \\ u(0) = u_0. \end{cases} \tag{17.129}$$

The following asymptotic properties are satisfied:

(i) *If $S = \arg\min_C \Psi$ is nonvoid, then $u(t)$ converges weakly to some $u_\infty \in S$, as $t \to +\infty$.*

(ii) *If moreover $u_0 \in C$, then $u(t) \in C$ for all $t \geq 0$, $\Psi(u(t))$ decreases to $\inf_C \Psi$ as t increases to $+\infty$, and*
$$\mu\frac{d}{dt}\Psi(u(t)) + \|\dot{u}(t)\|^2 \leq 0.$$

PROOF. (i) We just give the main lines of the proof; see [102] for further details. The proof is based on using the following Lyapunov function : given $z \in S = \arg\min_C \Psi$, set

$$E(t,z) = \frac{1}{2}\|u(t) - z\|^2 + \mu[\Psi(u(t)) - \Psi(z) - \langle \nabla\Psi(z), u(t) - z\rangle].$$

Time derivation of $E(\cdot, z)$ gives (for short we write $E(t) = E(t,z)$)

$$\frac{d}{dt}E(t) = \langle u(t) - z, \dot{u}(t)\rangle + \mu\,\langle \nabla\Psi(u(t)), \dot{u}(t)\rangle - \mu\,\langle \nabla\Psi(z), \dot{u}(t)\rangle\,. \tag{17.130}$$

On the one hand, the optimality condition for $z \in S = \arg\min_C \Psi$ gives $-\nabla\Psi(z) \in N_C(z)$, i.e.,
$$\langle \nabla\Psi(z), z - \xi\rangle \leq 0 \quad \forall \xi \in C.$$
Taking $\xi = \dot{u}(t) + u(t)$, which belongs to C (by (17.129)), we obtain

$$\langle \nabla\Psi(z), z - \dot{u}(t) - u(t)\rangle \leq 0.$$

Equivalently

$$-\langle \nabla\Psi(z), \dot{u}(t)\rangle \leq \langle \nabla\Psi(z), u(t) - z\rangle\,. \tag{17.131}$$

On the other hand, the obtuse angle condition for $\dot{u}(t)+u(t)=\mathrm{proj}_C\,(u(t)-\mu\nabla\Psi(u(t)))$ gives

$$\langle u(t)-\mu\nabla\Psi(u(t))-(\dot{u}(t)+u(t)),\xi-(\dot{u}(t)+u(t))\rangle\leq 0\quad\forall\xi\in C.$$

Equivalently

$$\langle\mu\nabla\Psi(u(t))+\dot{u}(t),\dot{u}(t)+u(t)-\xi\rangle\leq 0\quad\forall\xi\in C.$$

Taking $\xi=z$ gives

$$\langle u(t)-z,\dot{u}(t)\rangle+\mu\,\langle\nabla\Psi(u(t)),\dot{u}(t)\rangle\leq-\|\dot{u}(t)\|^2-\mu\,\langle\nabla\Psi(u(t)),u(t)-z\rangle\,. \tag{17.132}$$

Combining (17.130) with (17.131) and (17.132) gives

$$\frac{d}{dt}E(t)\leq-\|\dot{u}(t)\|^2-\mu\,\langle\nabla\Psi(u(t))-\nabla\Psi(z),u(t)-z\rangle\,.$$

This clearly implies that $E(\cdot)$ is a nonincreasing function, from which one classically infers the global existence of trajectories and the energy estimate $\dot{u}\in L^2(0,+\infty)$.

By using the Lyapunov function $E(\cdot)$, let us show how one can adapt the Opial argument and show the asymptotic weak convergence property. Let $z_1=w-\lim u(t_n)$ and $z_2=w-\lim u(s_n)$ be two weak sequential cluster points of u with $t_n\to+\infty$ and $s_n\to+\infty$. An elementary algebraic computation gives

$$E(t,z_2)-E(t,z_1)=\langle u(t),z_1-z_2\rangle+\mu\,\langle\nabla\Psi(z_1)-\nabla\Psi(z_2),u(t)\rangle+C(z_1,z_2),$$

where $C(z_1,z_2)$ is independent of t. Hence

$$\lim_{t\to\infty}[\langle u(t),z_1-z_2\rangle+\mu\,\langle\nabla\Psi(z_1)-\nabla\Psi(z_2),u(t)\rangle]\text{ exists.}$$

Replacing t successively by t_n and s_n and passing to the limit gives

$$\langle z_1,z_1-z_2\rangle+\mu\,\langle\nabla\Psi(z_1)-\nabla\Psi(z_2),z_1\rangle=\langle z_2,z_1-z_2\rangle+\mu\,\langle\nabla\Psi(z_1)-\nabla\Psi(z_2),z_2\rangle\,.$$

Equivalently

$$\|z_1-z_2\|^2+\mu\,\langle\nabla\Psi(z_1)-\nabla\Psi(z_2),z_1-z_2\rangle=0.$$

Since $\langle\nabla\Psi(z_1)-\nabla\Psi(z_2),z_1-z_2\rangle\geq 0$ (by monotonicity of $\nabla\Psi$), we obtain $z_1=z_2$. Then, one completes the proof as in the Bruck theorem.

(ii) When $u_0\in C$, the relaxed gradient projection dynamics enjoys some supplementary properties. In that case, by writing the equation under the form

$$\dot{u}(t)+u(t)=f(t)$$

with $f(t)\in C$, and by integrating this linear differential equation, we obtain

$$u(t)=e^{-t}u_0+e^{-t}\int_0^t f(s)e^s\,ds.$$

Thus, $u(t)$ is equal to a barycenter of elements of C, which, by convexity of C, implies $u(t)\in C$. Indeed, the same argument shows that if u_0 belongs to the interior of C, then $u(t)$ also remains in the interior of C. The orbit possibly reaches the boundary of C only at the limit, as $t\to\infty$. $\square$

17.2.8 ▪ First examples

Let us first describe some direct applications in the case of a smooth potential. We shall further examine applications to PDEs, which in general require working with a nonsmooth potential.

Steepest descent for linear least squares problems. Let $\mathcal{H}$ and $\mathcal{Y}$ be two Hilbert spaces equipped with the scalar products $\langle , \rangle_{\mathcal{H}}$, and $\langle , \rangle_{\mathcal{Y}}$, respectively. Let $A : \mathcal{H} \to \mathcal{Y}$ be a linear continuous operator from $\mathcal{H}$ into $\mathcal{Y}$. We denote by ${}^tA : \mathcal{Y} \to \mathcal{H}$ the transpose (adjoint) of A:

$$\forall v \in \mathcal{H},\ \forall y \in \mathcal{Y} \quad \langle Av, y\rangle_{\mathcal{Y}} = \langle v,\ {}^tAy\rangle_{\mathcal{H}}.$$

Let us give $b \in \mathcal{Y}$. The function $\Phi_A : \mathcal{H} \to \mathbf{R}^+$

$$\Phi_A(v) = \frac{1}{2}\|Av - b\|^2_{\mathcal{Y}}$$

is convex and differentiable. For any $v \in \mathcal{H}$, its gradient at v is given by $\nabla\Phi_A(v) = {}^tA(Av - b)$. The operator $v \mapsto {}^tA(Av - b)$ is affine, continuous, and hence Lipschitz continuous on $\mathcal{H}$. The conditions of Theorem 17.1.1 are satisfied. Hence, for any Cauchy data $u_0 \in \mathcal{H}$, there exists a unique global classical solution $u \in C^1([0, +\infty); \mathcal{H})$ of the Cauchy problem

$$\begin{cases} \dot{u}(t) + {}^tAA(u(t)) - {}^tAb = 0, \\ u(0) = u_0. \end{cases} \tag{17.133}$$

Let us now discuss the asymptotic behavior of the trajectories of (17.133). Since $\Phi_A : \mathcal{H} \to \mathbf{R}^+$ is a convex function, the critical points of Φ_A are the solutions of the convex minimization problem

$$\min_{v \in \mathcal{H}} \left\{ \|Av - b\|^2_{\mathcal{Y}} \right\}. \tag{17.134}$$

The following properties are direct consequences of the general results concerning the asymptotic behavior of trajectories of (SD) associated to a convex potential. Let $u \in C^1([0, +\infty); \mathcal{H})$ be the solution trajectory of (17.133). Then, as $t \to +\infty$,

(i) $\Phi_A(u(t)) = \frac{1}{2}\|Au(t) - b\|^2_{\mathcal{Y}}$ decreases to the infimal value $\inf_{v \in \mathcal{H}} \left\{ \|Av - b\|^2_{\mathcal{Y}} \right\}$;

(ii) assume moreover that the solution set S of (17.134) is nonempty; then $u(t)$ converges weakly in $\mathcal{H}$ to some $u_\infty \in S$.

The condition $S \neq \emptyset$ is satisfied in each of the following situations:

(a) The range of A (denoted by $R(A)$) is a closed subspace of $\mathcal{Y}$. In that case, (17.134) is equivalent to finding the projection of b on the closed subspace $R(A)$. Denoting by z an element such that $Az = \mathrm{proj}_{R(A)} b$, the solution set is equal to $S = z + \ker A$.

Let us give an example where the condition $R(A)$ is closed and fails to be satisfied: take $\mathcal{H} = l^2(\mathbf{N})$, the Hilbert space of sequences which are square summable. Given (α_n) a sequence of real numbers that satisfies $0 < \alpha_n \leq 1$ and $\sum_{n \in \mathbf{N}} (\alpha_n)^2 < +\infty$, let $A : \mathcal{H} \to \mathcal{H}$ be defined by

$$(\xi_n) \mapsto (\alpha_n \xi_n).$$

One can easily verify that A is linear continuous, its range is a dense subspace of $\mathcal{H} = l^2(\mathbf{N})$ (it contains the sequences with compact support), but its range is not equal to the whole space ((α_n) does not belong to the range). Hence, the range of A is not closed.

(b) If $b \in R(A)$, then clearly $S \neq \emptyset$, the infimal value is zero, and it is attained on the affine subset $z + \ker A$, where z is an element such that $Az = b$. In that case, let us show that each trajectory of (SD) converges strongly to an element of S. Indeed, with $w(t) = u(t) - z$, (SD) can be reformulated as

$$\dot{w}(t) + {}^tAA(w(t)) = 0.$$

This is the steepest descent equation associated with the convex even potential $\Psi(v) = \frac{1}{2}\|Av\|^2_{\mathscr{Y}}$. By Proposition 17.2.11, $w(t)$ converges strongly, as $t \to +\infty$, and so does $u(t) = w(t) + z$.

Steepest descent for coupled systems. With the same notation as in the preceding example, take $A : \mathscr{H} \to \mathscr{Y}$ as a linear continuous operator. Let us give $f, g : \mathscr{H} \to \mathbf{R}$ convex, of class C^1. We suppose that f and g are bounded from below on $\mathscr{H}$ and that ∇f and ∇g are Lipschitz continuous on bounded sets. Set $\Phi : \mathscr{H} \times \mathscr{H} \to \mathbf{R}$

$$\Phi(v, w) = f(v) + g(w) + \frac{1}{2}\|A(v - w)\|^2_{\mathscr{Y}}.$$

Let us equip $\mathscr{H} \times \mathscr{H}$ with the Hilbertian product structure

$$\langle\langle (v, w), (\xi, \eta)\rangle\rangle = \langle v, \xi\rangle + \langle w, \eta\rangle.$$

One can easily verify that Φ is convex and continuously differentiable. The operator $(v, w) \mapsto \nabla\Phi(v, w) = (\nabla f(v) + {}^tAA(v - w), \nabla g(w) + {}^tAA(w - v))$ is Lipschitz continuous on bounded subsets of $\mathscr{H} \times \mathscr{H}$. The conditions of Theorem 17.1.1 are satisfied. Thus, for any Cauchy data $(v_0, w_0) \in \mathscr{H} \times \mathscr{H}$, there exists a unique classical solution $(v(\cdot), w(\cdot)) \in C^1([0, +\infty); \mathscr{H} \times \mathscr{H})$ of

$$\begin{cases} \dot{v}(t) + \nabla f(v(t)) + {}^tAA(v(t) - w(t)) = 0, \\ \dot{w}(t) + \nabla g(w(t)) + {}^tAA(w(t) - v(t)) = 0, \\ v(0) = v_0, \quad w(0) = w_0. \end{cases} \tag{17.135}$$

Let us discuss the asymptotic behavior of the trajectories of (17.135). Since $\Phi : \mathscr{H} \times \mathscr{H} \to \mathbf{R}$ is a convex function, the critical points of Φ are the solutions of the convex minimization problem

$$\min_{(v,w) \in \mathscr{H} \times \mathscr{H}} \left\{ f(v) + g(w) + \frac{1}{2}\|A(v - w)\|^2_{\mathscr{Y}} \right\}. \tag{17.136}$$

The following properties are direct consequences of the general results concerning the asymptotic behavior of trajectories of (SD) associated to a convex potential. Let

$$(v(\cdot), w(\cdot)) \in C^1([0, +\infty); \mathscr{H})$$

be the solution trajectory of (17.135). Then, as $t \to +\infty$,

(i) $\Phi(v(t), w(t)) = f(v(t)) + g(w(t)) + \frac{1}{2}\|A(v(t) - w(t))\|^2_{\mathscr{Y}}$ decreases to the infimal value $\inf_{(v,w) \in \mathscr{H} \times \mathscr{H}} \{f(v) + g(w) + \frac{1}{2}\|A(v - w)\|^2_{\mathscr{Y}}\}$;

(ii) assume moreover that the solution set S of (17.136) is nonempty; then $(v(t), w(t))$ converges weakly in $\mathscr{H} \times \mathscr{H}$ to some $(v_\infty, w_\infty) \in S$.

17.2.9 ▪ Applications to PDEs

Analysis of the generalized steepest descent, which was developed in the previous sections, is valid for arbitrary convex lower semicontinuous potentials. Thus, in order to apply it to a specific potential Φ, it suffices to calculate the corresponding subdifferential operator $\partial\Phi$. However, when working in functional spaces, solving this problem of convex analysis is not immediate. The full description of the operator $\partial\Phi$ requires using subtle techniques. Just to mention a few of them, we will use the theory of maximal monotone operators, Fenchel duality, Sobolev spaces, and the regularity theory for elliptic PDEs.

1. The linear heat equation.

Dirichlet boundary condition. Let Ω be a bounded open set in $\mathbf{R}^N$. Take $\mathcal{H} = L^2(\Omega)$, and define $\Phi : L^2(\Omega) \to \mathbf{R} \cup \{+\infty\}$ by

$$\Phi(v) = \begin{cases} \frac{1}{2}\int_\Omega \|\nabla v(x)\|^2 dx & \text{if } v \in H_0^1(\Omega), \\ +\infty & \text{if } v \in L^2(\Omega),\ v \notin H_0^1(\Omega). \end{cases}$$

Clearly, Φ is a convex and proper function, whose domain is $H_0^1(\Omega)$. Let us verify that Φ is lower semicontinuous for the topology of $L^2(\Omega)$. Equivalently, let us show that for any $\gamma \in \mathbf{R}$, the sublevel set $lev_\gamma \Phi$ is closed for the topology of $L^2(\Omega)$. Let (v_n) be a sequence which satisfies, $v_n \in H_0^1(\Omega)$, $v_n \to v$ in $L^2(\Omega)$, and $\Phi(v_n) \leq \gamma$. Hence $\int_\Omega \|\nabla v_n(x)\|^2 dx \leq 2\gamma$. Thus, (v_n) and its first distributional derivatives are bounded in $L^2(\Omega)$. As a consequence, the sequence (v_n) is bounded in $H_0^1(\Omega)$. Hence, v_n converges weakly to v in $H_0^1(\Omega)$. Since Φ is convex continuous on $H_0^1(\Omega)$, it is lower semicontinuous for the weak topology of $H_0^1(\Omega)$. Thus, $\Phi(v) \leq \liminf \Phi(v_n) \leq \gamma$.

Theorem 17.2.10. (a) *The subdifferential $A = \partial\Phi$ is equal to*

$$\begin{cases} \operatorname{dom} A = \left\{ v \in H_0^1(\Omega) :\ \Delta v \in L^2(\Omega) \right\}, \\ A(v) = -\Delta v \quad \textit{for } v \in \operatorname{dom}(A), \end{cases}$$

(b) *When Ω is regular,* $\operatorname{dom} A = H^2(\Omega) \cap H_0^1(\Omega)$.

Because of the importance of this result, and to illustrate the different strategies of demonstration, we give two different proofs (which can be extended to more involved situations).

FIRST PROOF. We use the characterization of the subdifferential via the Fenchel conjugate and the extremality relation. Let us recall (see Proposition 9.5.1) that

$$f \in \partial\Phi(u) \Leftrightarrow \Phi(u) + \Phi^*(f) - \langle f, u \rangle = 0.$$

Thus, the problem has been converted into the computation of $\Phi^*(f)$. Indeed, in Chapter 9, Section 9.8, as an illustration of the Fenchel duality calculus, we proved that

$$\inf_{v \in H_0^1(\Omega)} \left\{ \frac{1}{2}\int_\Omega \|\nabla v(x)\|^2 dx - \int_\Omega f(x)v(x)dx \right\}$$
$$= -\inf\left\{ \frac{1}{2}\int_\Omega \|y(x)\|^2 dx :\ y \in L^2(\Omega)^N,\ \operatorname{div} y = f \right\}.$$

Equivalently

$$\Phi^*(f) = \inf\left\{\frac{1}{2}\int_\Omega \|y(x)\|^2 dx : \ y \in L^2(\Omega)^N, \ \mathrm{div} y = f\right\}.$$

Clearly, in the above expression, the infimum is achieved. As a consequence, the extremality relation which characterizes $f \in \partial\Phi(u)$ can be equivalently formulated as follows: there exists $y \in L^2(\Omega)^N$ such that $\mathrm{div} y = f$, and

$$\frac{1}{2}\int_\Omega \|\nabla u(x)\|^2 dx + \frac{1}{2}\int_\Omega \|y(x)\|^2 dx - \int_\Omega f(x)u(x)dx = 0. \tag{17.137}$$

By $\mathrm{div} y = f$, and the definition of the derivation in the sense of distributions, we have that for any test function $v \in \mathscr{D}(\Omega)$

$$\int_\Omega f(x)v(x)dx = -\int_\Omega y(x).\nabla v(x)dx.$$

Since $y \in L^2(\Omega)^N$, this equality can be extended by continuity to any $v \in H_0^1(\Omega)$ and in particular to u. Hence,

$$\int_\Omega f(x)u(x)dx = -\int_\Omega y(x).\nabla u(x)dx. \tag{17.138}$$

Combining (17.137) and (17.138) we obtain

$$\frac{1}{2}\int_\Omega \|\nabla u(x)\|^2 dx + \frac{1}{2}\int_\Omega \|y(x)\|^2 dx + \int_\Omega y(x).\nabla u(x)dx = 0.$$

Equivalently

$$\frac{1}{2}\int_\Omega \|\nabla u(x) + y(x)\|^2 dx = 0.$$

Hence $y = -\nabla u$. From $\mathrm{div} y = f$, we finally infer $-\Delta u = f$.

Note that the above argument only involves equivalent relations, which gives the claim.

(b) When Ω is regular ($\partial\Omega$ of class C^2), we use the Agmon–Douglis–Nirenberg theorem (see [8], [137, Theorem IX.32]) which gives the $H^2(\Omega)$ regularity of the solution of the Poisson equation

$$\begin{cases} -\Delta u = f, \\ u \in H_0^1(\Omega). \end{cases}$$

Thus $\mathrm{dom} A = H^2(\Omega) \cap H_0^1(\Omega)$. □

SECOND PROOF. Let us introduce the operator A acting on $\mathscr{H} = L^2(\Omega)$ which is defined by

$$\begin{cases} \mathrm{dom} A = \left\{v \in H_0^1(\Omega) : \ \Delta v \in L^2(\Omega)\right\}, \\ A(v) = -\Delta v \quad \text{ for } v \in \mathrm{dom} A. \end{cases}$$

We are going to prove that A is maximal monotone, and $A \subset \partial\Phi$. Since $\partial\Phi$ is monotone (indeed, it is maximal monotone), this will imply $A = \partial\Phi$. Clearly, A is linear, and for any $v \in \mathrm{dom} A$

$$\int_\Omega -\Delta v.(v)dx = \int_\Omega \|\nabla v(x)\|^2 dx.$$

Hence, A is monotone. Similarly, for any $u \in \operatorname{dom} A$ and $v \in \operatorname{dom} \Phi$

$$\begin{aligned}\int_\Omega -\Delta u.(v-u)dx &= \int_\Omega \langle \nabla u, \nabla v - \nabla u\rangle\, dx.\\ &\leq \Phi(v) - \Phi(u).\end{aligned}$$

Thus, $A \subset \partial\Phi$. By Minty's theorem, Theorem 17.2.1, it just remains to prove that $R(I+A) = \mathscr{H}$. Equivalently, for any $f \in L^2(\Omega)$, we have to prove the existence of a solution to the Dirichlet problem

$$\begin{cases} u - \Delta u = f & \text{on } \Omega,\\ u \in H_0^1(\Omega). \end{cases}$$

This is a classical result (see Chapter 6, Theorem 6.1.1), which was obtained by applying the Lax–Milgram theorem. □

Let us now combine the above results and the general properties of the gradient flow associated to a convex lower semicontinuous potential. For simplicity of the statements, we suppose Ω is a bounded regular open set in $\mathbf{R}^n$, whose topological boundary is denoted by $\partial\Omega$.

Theorem 17.2.11. *Let $h \in L^2(\Omega)$. Then, the following properties hold:*

(a) *For any $u_0 \in L^2(\Omega)$, there exists a unique solution u of the heat equation*

$$\begin{cases} \frac{\partial u}{\partial t} - \Delta u = h, & a.e. \text{ on } \Omega\times]0,+\infty[,\\ u(x,t) = 0 & a.e. \text{ on } \partial\Omega\times]0,+\infty[,\\ u(x,0) = u_0 & a.e. \text{ on } \Omega \end{cases}$$

satisfying $u \in C([0,+\infty[;L^2(\Omega))$, $u(\cdot,t) \in H^2(\Omega)$ for any $t > 0$, and $\sqrt{t}\frac{\partial u}{\partial t} \in L^2([0,T];L^2(\Omega))$ for any $T > 0$.

(b) *As $t \to +\infty$, $u(t)$ converges strongly in $H_0^1(\Omega)$ to the unique solution w of the Dirichlet problem*

$$\begin{cases} -\Delta w = h & \text{on } \Omega,\\ w = 0 & \text{on } \partial\Omega. \end{cases}$$

PROOF. Let us consider the functional $\Phi : \mathscr{H} = L^2(\Omega) \to \mathbf{R} \cup \{+\infty\}$ which is defined by

$$\Phi(v) = \begin{cases} \frac{1}{2}\int_\Omega \|\nabla v(x)\|^2 dx - \int_\Omega h(x)v(x)dx & \text{if } v \in H_0^1(\Omega),\\ +\infty & \text{if } v \in L^2(\Omega),\ v \notin H_0^1(\Omega). \end{cases}$$

The functional $v \mapsto \int_\Omega h(x)v(x)dx$ is linear continuous on $L^2(\Omega)$. Its gradient is constant and equal to h. By the additivity rule for the subdifferential of the sum of two convex lower semicontinuous functions (Theorem 9.5.4), and Theorem 17.2.10, the subdifferential $A = \partial\Phi$ is equal to

$$\begin{cases} \operatorname{dom} A = \left\{v \in H_0^1(\Omega) :\ \Delta v \in L^2(\Omega)\right\},\\ A(v) = -\Delta v - h \quad \text{for } v \in \operatorname{dom} A. \end{cases}$$

The existence of a strong global solution follows from Theorem 17.2.3 (regularizing effect) and the fact that the domain of Φ, which is equal to $H_0^1(\Omega)$, is dense in $L^2(\Omega)$. The asymptotic convergence result follows from Bruck's theorem, Theorem 17.2.7, and the fact that the weak convergence in $\mathcal{H} = L^2(\Omega)$ and the convergence of the potential energies imply the strong convergence in $H_0^1(\Omega)$. □

Neumann boundary condition. Let Ω be a bounded open set in $\mathbf{R}^N$ with smooth boundary $\partial\Omega$. Let $h \in L^2(\Omega)$. Take $\mathcal{H} = L^2(\Omega)$, and define $\Phi : L^2(\Omega) \to \mathbf{R} \cup \{+\infty\}$ by

$$\Phi(v) = \begin{cases} \frac{1}{2}\int_\Omega \|\nabla v(x)\|^2 dx - \int_\Omega h(x)v(x)dx & \text{if } v \in H^1(\Omega), \\ +\infty & \text{if } v \in L^2(\Omega),\ v \notin H^1(\Omega). \end{cases}$$

Clearly, Φ is a convex and proper function, whose domain is $H^1(\Omega)$. By a similar argument to that used in the Dirichlet case, Φ is lower semicontinuous on $\mathcal{H} = L^2(\Omega)$, and its subdifferential $A = \partial\Phi$ is equal to

$$\begin{cases} \operatorname{dom} A = \left\{ v \in H^2(\Omega) : \frac{\partial u}{\partial n} = 0 \text{ on } \Omega \right\}, \\ A(v) = -\Delta v - h \quad \text{for } v \in \operatorname{dom} A. \end{cases}$$

Note that the Neumann boundary condition $\frac{\partial u}{\partial n} = 0$ on $\partial\Omega$ naturally occurs, when expressing an extremality relation on the space $H^1(\Omega)$. This phenomena has been studied in detail in Section 6.2. Thus, by applying Theorem 17.2.3 (regularizing effect), and Bruck's theorem, Theorem 17.2.7 (asymptotic behavior), we obtain the following result.

Theorem 17.2.12. (a) *For any $u_0 \in L^2(\Omega)$ and $h \in L^2(\Omega)$, there exists a unique solution u of*

$$\begin{cases} \frac{\partial u}{\partial t} - \Delta u = h & a.e. \text{ on } \Omega \times]0, +\infty[, \\ \frac{\partial u}{\partial n} = 0 & a.e. \text{ on } \partial\Omega \times]0, +\infty[, \\ u(x, 0) = u_0 & a.e. \text{ on } \Omega \end{cases}$$

satisfying $u \in C([0, +\infty[; L^2(\Omega))$, $u(\cdot, t) \in H^2(\Omega)$ for any $t > 0$, and $\sqrt{t}\frac{\partial u}{\partial t} \in L^2([0, T]; L^2(\Omega))$ for any $T > 0$.

(b) *Suppose that the solution set S of the Neumann problem*

$$\begin{cases} -\Delta w = h & \text{on } \Omega, \\ \frac{\partial w}{\partial n} = 0 & \text{on } \partial\Omega. \end{cases} \tag{17.139}$$

is nonvoid. Then, as $t \to +\infty$, $u(t)$ converges strongly in $H^1(\Omega)$ to a solution $w \in S$.

Recall that the above Neumann problem is semicoercive. A necessary and sufficient condition for the existence of solutions to the stationary problem (17.139) is that $\int_\Omega h(x)dx = 0$, in which case all the solutions differ by an additive constant (see Section 6.2). Note that if $S = \emptyset$, then for any initial data $u_0 \in L^2(\Omega)$, the corresponding orbit u verifies $\lim_{t\to+\infty} \|u(t)\|_{L^2(\Omega)} = +\infty$.

2. The Stefan problem. The following problem is named after the physicist J. Stefan, who introduced it around 1890, in relation to problems of ice formation. It describes the temperature distribution in a medium undergoing a phase change, for example, ice passing to water. The Stefan problem is a free boundary problem, and the determination of the (time evolving) interface between the two phases is the central part of the problem. (As soon as it is known, we just need to solve the heat problem in both phases.) The following gradient flow approach to the Stefan problem was introduced by Brezis in [134].

The operator which governs the Stefan problem is identified to the subdifferential of an integral functional with respect to the $H^{-1}(\Omega)$ metric. It can be formulated as

$$\frac{\partial u}{\partial t} - \Delta\beta(u) = f \tag{17.140}$$

with given boundary conditions and Cauchy data. The operator β is a maximal monotone graph from $\mathbf{R}$ onto $\mathbf{R}$, and it carries the physical information on the phase change. Let us fix the functional setting. Let Ω be a bounded open set in $\mathbf{R}^N$. Let $\mathscr{H} = H^{-1}(\Omega)$ be the topological dual of $H_0^1(\Omega)$; see Section 5.2. We know that the (minus) Laplace–Dirichlet operator $-\Delta$ is an isomorphism between $H_0^1(\Omega)$ and $H^{-1}(\Omega)$. For our purpose, it is convenient to introduce the scalar product on $H^{-1}(\Omega)$ which is induced by this isomorphism and by the scalar product on $H_0^1(\Omega)$,

$$\langle u, v\rangle_{H_0^1(\Omega)} = \int_\Omega \nabla u(x)\cdot\nabla v(x)dx,$$

which, by the Poincaré inequality, induces a norm which is equivalent to the classical one. For any $f, g \in H^{-1}(\Omega)$ we set

$$\begin{aligned}
\langle f, g\rangle_{H^{-1}(\Omega)} &= \left\langle (-\Delta)^{-1} f, (-\Delta)^{-1} g\right\rangle_{H_0^1(\Omega)} \\
&= \sum_{i=1}^N \int_\Omega \frac{\partial}{\partial x_i}\left((-\Delta)^{-1} f\right) \frac{\partial}{\partial x_i}\left((-\Delta)^{-1} g\right) dx \\
&= \left\langle (-\Delta)^{-1} f, g\right\rangle_{\left(H_0^1(\Omega), H^{-1}(\Omega)\right)}.
\end{aligned}$$

This last expression is the duality bracket between $g \in H^{-1}(\Omega) = H_0^1(\Omega)^*$ and $(-\Delta)^{-1} f \in H_0^1(\Omega)$. Equipped with this scalar product, $\mathscr{H} = H^{-1}(\Omega)$ is a Hilbert space with a norm that is equivalent to the classic. Let us give $j : \mathbf{R} \to \mathbf{R} \cup \{+\infty\}$, a convex, lower semicontinuous, proper function which is strongly coercive, i.e.,

$$\lim_{|r|\to+\infty} \frac{j(r)}{|r|} = +\infty. \tag{17.141}$$

Let $\beta = \partial j$ be the subdifferential of j. It is a maximal monotone graph from $\mathbf{R}$ into $\mathbf{R}$ that satisfies $R(\beta) = \mathbf{R}$. To obtain this last property, note that for any $s \in \mathbf{R}$, the convex lower semicontinuous function $r \mapsto j(r) - rs$ is coercive (a consequence of (17.141)) and hence attains its minimum at a point $\bar{r}$. Writing the optimality condition gives $\beta(\bar{r}) \ni s$, whence the result.

Let us define $\Phi : H^{-1}(\Omega) \to \mathbf{R} \cup \{+\infty\}$ by

$$\Phi(v) = \begin{cases} \int_\Omega j(v(x))dx & \text{if } v \in H^{-1}(\Omega)\cap L^1(\Omega),\ j(v) \in L^1(\Omega), \\ +\infty & \text{otherwise.} \end{cases} \tag{17.142}$$

The following result, which describes the subdifferential of Φ on the space $H^{-1}(\Omega)$, makes the link with formulation (17.140) of the Stefan problem.

Theorem 17.2.13. (a) *The function $\Phi : H^{-1}(\Omega) \to \mathbf{R} \cup \{+\infty\}$ is convex, lower semicontinuous, and proper on $H^{-1}(\Omega)$.*

(b) *The subdifferential $A = \partial\Phi$ is equal to*

$$\begin{cases} \operatorname{dom} A = \left\{ v \in H^{-1}(\Omega) \cap L^1(\Omega) : \exists w \in H_0^1(\Omega) \text{ such that } w(x) \in \beta(v(x)) \text{ a.e. on } \Omega \right\}, \\ A(v) = \left\{ -\Delta w : \quad w \in H_0^1(\Omega) \text{ and } w(x) \in \beta(v(x)) \text{ a.e. on } \Omega \right\}. \end{cases} \tag{17.143}$$

PROOF. (a) Let (u_n) be a sequence such that $u_n \in H^{-1}(\Omega) \cap L^1(\Omega)$, $u_n \to u$ in $H^{-1}(\Omega)$ and

$$\int_\Omega j(u_n(x))dx \le \lambda.$$

By the De La Vallée–Poussin theorem, Theorem 2.4.4, and the Dunford–Pettis theorem, Theorem 2.4.5, the sequence (u_n) is $\sigma(L^1, L^\infty)$ sequentially relatively compact. Thus, we can extract a subsequence (u_{n_k}) such that $u_{n_k} \to \tilde{u}$ weakly in $L^1(\Omega)$. Since $u_n \to u$ in $H^{-1}(\Omega)$, we have $u = \tilde{u}$ and the whole sequence $u_n \to u$ weakly in $L^1(\Omega)$. By Fatou's lemma, the integral functional $v \mapsto \int_\Omega j(v(x))dx$ is lower semicontinuous on $L^1(\Omega)$. Being convex, it is lower semicontinuous for the weak topology $\sigma(L^1, L^\infty)$. Hence

$$\int_\Omega j(u(x))dx \le \liminf_n \int_\Omega j(u_n(x))dx \le \lambda,$$

which proves that the lower-level sets of Φ are closed. Hence Φ is lower semicontinuous on $H^{-1}(\Omega)$.

(b) To show that $A = \partial\Phi$, where A is the operator described in (17.143), we prove that $A \subset \partial\Phi$, and A is maximal monotone. The following lemma from [134] will play a key role. In particular, it allows us to interpret the duality pairing between $F \in H^{-1}(\Omega)$ and $w \in H_0^1(\Omega)$ as an integral $\int_\Omega F(x)w(x)dx$ when $Fw \in L^1(\Omega)$.

Lemma 17.2.6. *Let $F \in H^{-1}(\Omega) \cap L^1(\Omega)$, and let $w \in H_0^1(\Omega)$. Let $g \in L^1(\Omega)$, and let h be measurable with*

$$F(x)w(x) \ge h(x) \ge g(x) \ \text{ a.e. } x \in \Omega. \tag{17.144}$$

Then $h \in L^1(\Omega)$, and

$$\langle w, F \rangle_{(H_0^1(\Omega), H^{-1}(\Omega))} \ge \int_\Omega h(x)dx.$$

PROOF. For each $n \in \mathbf{N}$, set

$$w_n = \begin{cases} n & \text{if } w \ge n, \\ w & \text{if } |w| \le n, \\ -n & \text{if } w \le -n. \end{cases}$$

Set $h_n = h\frac{w_n}{w}$, and $g_n = g\frac{w_n}{w}$. Multiplying (17.144) by the nonnegative function $\frac{w_n}{w}$, we obtain

$$F(x)w_n(x) \ge h_n(x) \ge g_n(x) \quad \text{a.e. } x \in \Omega,$$

and hence

$$0 \le h_n(x) - g_n(x) \le F(x)w_n(x) - g_n(x) \quad \text{a.e. } x \in \Omega. \tag{17.145}$$

Since $F \in L^1(\Omega)$ and $w_n \in L^\infty(\Omega)$, we have $Fw_n \in L^1(\Omega)$. After integration of (17.145), we obtain

$$\int_\Omega (h_n - g_n)dx \le \int_\Omega F w_n dx - \int_\Omega g_n dx. \tag{17.146}$$

Then note that

$$\begin{aligned}\int_\Omega F w_n dx &= \langle w_n, F\rangle_{(L^\infty(\Omega), L^1(\Omega))} \\ &= \langle w_n, F\rangle_{(H_0^1(\Omega), H^{-1}(\Omega))}.\end{aligned}$$

By Fatou's lemma, $h_n - g_n$ nonnegative, and $h_n - g_n \to h - g$ a.e. on Ω, we deduce that

$$\int_\Omega (h-g)dx \le \liminf_n \int_\Omega (h_n - g_n)dx.$$

Since contractions operate on $H_0^1(\Omega)$ (see Section 5.8), we have $w_n \to w$ in $H_0^1(\Omega)$. Moreover, $g_n \to g$ in $L^1(\Omega)$. As a consequence, by passing to the limit on (17.146), we obtain

$$0 \le \int_\Omega (h-g)dx \le \langle w, F\rangle_{(H_0^1(\Omega), H^{-1}(\Omega))} - \int_\Omega g dx. \tag{17.147}$$

Hence, $h - g \in L^1(\Omega)$, which implies $h \in L^1(\Omega)$, and after simplification of (17.147)

$$\int_\Omega h(x)dx \le \langle w, F\rangle_{(H_0^1(\Omega), H^{-1}(\Omega))},$$

which completes the proof of Lemma 17.2.6. □

PROOF OF THEOREM 17.2.13 CONTINUED. Let us prove that $A \subset \partial\Phi$. Let $f \in Au$, i.e., $u \in H^{-1}(\Omega) \cap L^1(\Omega)$, $f = -\Delta w$, with $w \in H_0^1(\Omega)$, and $w(x) \in \beta(u(x))$ a.e. on Ω. Let $v \in \operatorname{dom}\Phi$, i.e., $v \in H^{-1}(\Omega) \cap L^1(\Omega)$, $j(v) \in L^1(\Omega)$. For a.e. $x \in \Omega$, by the convex subdifferential inequality,

$$j(v(x)) - j(u(x)) \ge w(x)(v(x) - u(x)).$$

Equivalently

$$w(x)(u(x) - v(x)) \ge j(u(x)) - j(v(x)).$$

Let us apply Lemma 17.2.6 with $F = u - v \in H^{-1}(\Omega) \cap L^1(\Omega)$, $w \in H_0^1(\Omega)$, and $h = j(u) - j(v)$. Noticing that $j(r) \ge -C(1 + |r|)$ for some positive constant C, we have

$$j(u(x)) - j(v(x)) \ge g(x) := -C(1 + |u(x)|) - j(v(x))$$

with $g \in L^1(\Omega)$. Thus, the conditions of Lemma 17.2.6 are satisfied. We conclude that $j(u) \in L^1(\Omega)$, and

$$\begin{aligned}\int_\Omega j(v)dx - \int_\Omega j(u)dx &\ge \langle w, v - u\rangle_{(H_0^1(\Omega), H^{-1}(\Omega))} \\ &= \left\langle (-\Delta)^{-1} f, v - u\right\rangle_{(H_0^1(\Omega), H^{-1}(\Omega))} \\ &= \langle f, v - u\rangle_{H^{-1}(\Omega)}.\end{aligned}$$

Hence, $f \in \partial\Phi(u)$.

Let us complete the proof of Theorem 17.2.13 by showing that A is maximal monotone. Since $A \subset \partial\Phi$, we have that A is monotone. Thus, by Minty's theorem, Theorem 17.2.1, we just need to prove that $R(I+A)=\mathscr{H}$. Equivalently, for any given $f \in H^{-1}(\Omega)$, we have to find $u \in H^{-1}(\Omega)\cap L^1(\Omega)$, and $w \in H_0^1(\Omega)$ such that

$$\begin{cases} w(x) \in \beta(u(x)) \ \text{ a.e. on } \Omega, \\ u - \Delta w = f. \end{cases} \tag{17.148}$$

Let us rewrite (17.148) as a semilinear equation. Set $\gamma = \beta^{-1}$. Since $R(\beta)=\mathbf{R}$, we have $\operatorname{dom}\gamma = \mathbf{R}$. Thus (17.148) is equivalent to finding $w \in H_0^1(\Omega)$ such that

$$\gamma(w) - \Delta w \ni f \tag{17.149}$$

is satisfied in the following sense: there exists some $z \in H^{-1}(\Omega)\cap L^1(\Omega)$ such that $z(x) \in \gamma(w(x))$ a.e. on Ω and $z - \Delta w = f$. Solving this nonlinear equation is not immediate. The difficulty is that the second member of the equation belongs to $H^{-1}(\Omega)$. By contrast, for $f \in L^2(\Omega)$, the methods developed in Section 6.2.4 provide existence and uniqueness of a solution to (17.149) for an arbitrary monotone graph γ. To solve (17.149) with $f \in H^{-1}(\Omega)$, we use the fact that $\operatorname{dom}\gamma = \mathbf{R}$. Without loss of generality, we can assume that $0 \in \gamma(0)$. (This amounts to replacing f by $f-\gamma(0)$, which still belongs to $H^{-1}(\Omega)$.) Let us approximate (17.149) by using the Yosida approximation γ_λ of the monotone graph γ. For any $\lambda > 0$, there exists a unique solution $w_\lambda \in H_0^1(\Omega)$ of

$$\gamma_\lambda(w_\lambda) - \Delta w_\lambda = f. \tag{17.150}$$

This can be achieved by a variational argument. Let us consider the convex minimization problem

$$\min\left\{ \int_\Omega (j^*)_\lambda(w(x))dx + \frac{1}{2}\int_\Omega |\nabla w(x)|^2 dx - \langle w, f\rangle_{(H_0^1(\Omega),H^{-1}(\Omega))} : \ w \in H_0^1(\Omega)\right\}, \tag{17.151}$$

where $\gamma = (\partial j)^{-1} = \partial j^*$ and $(j^*)_\lambda(r) = \int_0^r \gamma_\lambda(s)ds$. Since (17.151) is a strongly convex minimization problem, it has a unique solution w_λ. To write the corresponding optimality condition, we can notice that the convex integral functional $v \mapsto \int_\Omega (j^*)_\lambda(w(x))dx$ is continuous on $L^2(\Omega)$ (because $(j^*)_\lambda(r) \le C(1+|r|^2)$) and hence on $H_0^1(\Omega)$. The additivity rule (Theorem 9.5.4) for the subdifferential of a sum of convex functions gives (17.150).

Multiplying (17.150) by w_λ, and integrating on Ω, we obtain

$$\sup_\lambda \|w_\lambda\|_{H_0^1(\Omega)} < +\infty \tag{17.152}$$

and

$$\sup_\lambda \int_\Omega \gamma_\lambda(w_\lambda) w_\lambda dx < +\infty. \tag{17.153}$$

By (17.152) and the Rellich–Kondrakov theorem, we can find a sequence $\lambda_n \to 0$ such that

$$w_{\lambda_n} \rightharpoonup w \quad \text{weakly in} \quad H_0^1(\Omega); \tag{17.154}$$

$$w_{\lambda_n}(x) \to w(x) \quad \text{a.e. on} \ \ \Omega;$$

$$(I+\lambda_n\gamma)^{-1} w_{\lambda_n}(x) \to w(x) \quad \text{a.e. on} \ \ \Omega. \tag{17.155}$$

This last result comes from the contraction property of the resolvents (Proposition 17.2.1)

$$|(I+\lambda_n\gamma)^{-1}w_{\lambda_n}(x)-(I+\lambda_n\gamma)^{-1}w(x)| \le |w_{\lambda_n}(x)-w(x)|$$

and the approximation property of the resolvents (Proposition 17.2.2)

$$(I+\lambda_n\gamma)^{-1}r \to r \quad \text{as } n\to\infty\ \forall r\in\overline{\operatorname{dom}\gamma}=\mathbf{R}.$$

In order to pass to the limit on (17.150) in the distribution sense, let us show that $(\gamma_\lambda(w_\lambda))$ is $\sigma(L^1,L^\infty)$ sequentially relatively compact. By definition of the Yosida approximation, we have

$$\begin{aligned}\gamma_\lambda(w_\lambda)(I+\lambda\gamma)^{-1}w_\lambda &= \gamma_\lambda(w_\lambda)(w_\lambda-\lambda\gamma_\lambda(w_\lambda))\\ &= \gamma_\lambda(w_\lambda)w_\lambda-\lambda|\gamma_\lambda(w_\lambda)|^2\\ &\le \gamma_\lambda(w_\lambda)w_\lambda.\end{aligned}$$

From (17.153) we deduce that

$$\sup_\lambda \int_\Omega \gamma_\lambda(w_\lambda)(I+\lambda\gamma)^{-1}w_\lambda dx < +\infty. \tag{17.156}$$

Since $\gamma_\lambda(w_\lambda)\in\gamma\big((I+\lambda\gamma)^{-1}w_\lambda\big)$, and $\gamma=(\partial j)^{-1}=\partial j^*$, by the Fenchel extremality relation

$$j^*((I+\lambda\gamma)^{-1}w_\lambda)+j(\gamma_\lambda(w_\lambda))=\gamma_\lambda(w_\lambda)(I+\lambda\gamma)^{-1}w_\lambda.$$

From (17.156), and j^* minorized (a consequence of $0\in\gamma(0)$), we obtain

$$\sup_\lambda \int_\Omega j(\gamma_\lambda(w_\lambda))dx < +\infty.$$

Since j is strongly coercive (17.141), using again the De La Vallée–Poussin theorem, Theorem 2.4.4, and the Dunford–Pettis theorem, Theorem 2.4.5, we obtain that the net $(\gamma_\lambda(w_\lambda))$ is $\sigma(L^1,L^\infty)$ sequentially relatively compact. Thus, we can extract a sequence (still denoted λ_n) such that $\lambda_n\to 0$ and find some $z\in L^1(\Omega)$ such that

$$\gamma_{\lambda_n}(w_{\lambda_n}) \rightharpoonup z \quad \text{weakly in} \quad L^1(\Omega). \tag{17.157}$$

From (17.154) and (17.157), by passing to the limit on (17.150) in the distribution sense, we obtain

$$z-\Delta w=f$$

with $z\in H^{-1}(\Omega)\cap L^1(\Omega)$. To complete the proof of Theorem 17.2.13, we just need to prove that $z(x)\in\gamma(w(x))$ a.e. on Ω. It is sufficient to prove that, for every $N\in\mathbf{N}$,

$$z(x)\in\gamma(w(x)) \text{ a.e. on } \Omega_N := \{x\in\Omega : |w(x)|\le N\}.$$

By (17.155), $v_n(x):=(I+\lambda_n\gamma)^{-1}w_{\lambda_n}(x)\to w(x)$ a.e. on Ω. Hence, by Egorov's theorem, since $|\Omega|<+\infty$, for any $\epsilon>0$ there is a measurable set $E\subset\Omega_N$ such that $|E|<\epsilon$, and $v_n\to w$ uniformly on $\Omega_N\setminus E$. Thus, setting $z_n:=\gamma_{\lambda_n}(w_{\lambda_n})$, we are reduced to the following situation:

$$\begin{aligned}&z_n(x)\in\gamma(v_n(x)) \quad \text{a.e. on } \Omega,\\ &v_n\to w \quad \text{uniformly on } \Omega;\\ &z_n\rightharpoonup z \quad \text{weakly in} \quad L^1(\Omega),\\ &w \text{ is bounded.}\end{aligned}$$

We can now use a monotonicity argument involving spaces in duality (namely, $L^1(\Omega)$ and $L^\infty(\Omega)$). Let $\tilde{v} \in L^\infty(\Omega)$ and $\tilde{f} \in L^1(\Omega)$ be such that $\tilde{f}(x) \in \gamma(\tilde{v}(x))$ a.e. on Ω. By the monotonicity of γ, and after integration on Ω, we obtain

$$\int_\Omega (\tilde{f}(x) - z_n(x))(\tilde{v}(x) - v_n(x))dx \geq 0.$$

By passing to the limit we obtain

$$\int_\Omega (\tilde{f}(x) - z(x))(\tilde{v}(x) - w(x))dx \geq 0. \tag{17.158}$$

Take $\tilde{v} := (I+\gamma)^{-1}(w+z)$. Since (v_n) is uniformly bounded, $z_n(x) \in \gamma(v_n(x))$, and $\operatorname{dom}\gamma = \mathbf{R}$, the sequence (z_n) remains uniformly bounded, and hence $z \in L^\infty(\Omega)$. As a consequence $w+z \in L^\infty(\Omega)$, and $\tilde{v} := (I+\gamma)^{-1}(w+z) \in L^\infty(\Omega)$. We have $\tilde{v} + \gamma(\tilde{v}) \ni w+z$. By taking in (17.158) $\tilde{f} = w + z - \tilde{v}$, which satisfies $\tilde{f} \in L^1(\Omega)$ and $\tilde{f}(x) \in \gamma(\tilde{v}(x))$ a.e. on Ω, we obtain

$$\int_\Omega |\tilde{v}(x) - w(x)|^2 dx \leq 0.$$

Hence $\tilde{v}(x) = w$, which in turn implies $\tilde{f} = w + z - \tilde{v} = z$. Since $\tilde{f}(x) \in \gamma(\tilde{v}(x))$ a.e. on Ω, we finally obtain $z(x) \in \gamma(w(x))$ a.e. on Ω. $\square$

As a direct consequence of Theorem 17.2.13, and the theory of gradient flow, we obtain the following result. For simplicity of the statement, we suppose that β is a monotone and continuous function from $\mathbf{R}$ onto $\mathbf{R}$.

Theorem 17.2.14. *Let $\beta : \mathbf{R} \to \mathbf{R}$ be a continuous monotone function such that* $\operatorname{dom}\beta = \mathbf{R}$ *and $R(\beta) = \mathbf{R}$. Let $h \in H^{-1}(\Omega)$.*
(a) *For any $u_0 \in H^{-1}(\Omega)$, there exists a unique solution u of the Stefan problem*

$$\begin{cases} \frac{\partial u}{\partial t} - \Delta\beta(u) = h & \text{on } \Omega\times]0,+\infty[, \\ \beta(u)(x,t) = 0 & \text{on } \partial\Omega\times]0,+\infty[, \\ u(x,0) = u_0 & \text{on } \Omega \end{cases} \tag{17.159}$$

satisfying $u \in C([0,+\infty[;H^{-1}(\Omega))$, $\beta(u(x,t)) \in H_0^1(\Omega)$ for all $t > 0$, $u(x,t) \in L^1(\Omega)$ for all $t > 0$, $\sqrt{t}\frac{\partial u}{\partial t} \in L^2([0,T];H^{-1}(\Omega))$ for any $T > 0$.

(b) *As $t \to +\infty$, $\beta(u(x,t))$ converges strongly in $H_0^1(\Omega)$ to the solution $w_\infty \in H_0^1(\Omega)$ of the Dirichlet problem*

$$\begin{cases} -\Delta w_\infty = h & \text{on } \Omega, \\ w_\infty = 0 & \text{on } \partial\Omega. \end{cases} \tag{17.160}$$

Moreover, if the solution set $S = (\partial\Phi)^{-1}(h) \neq \emptyset$, then $u(x,t)$ converges weakly in $H^{-1}(\Omega)$ to an element of S.

PROOF. Take $\mathscr{H} = H^{-1}(\Omega)$, and consider the gradient flow associated to the potential $v \mapsto \Psi(v) = \Phi(v) - \langle v,h\rangle_{(H_0^1(\Omega),H^{-1}(\Omega))}$, where $\Phi : H^{-1}(\Omega) \to \mathbf{R}\cup\{+\infty\}$ is the convex lower semicontinuous functional defined in (17.142) (with $\beta = \partial j$),

$$\Phi(v) = \begin{cases} \int_\Omega j(v(x))dx & \text{if } v \in H^{-1}(\Omega)\cap L^1(\Omega),\ j(v) \in L^1(\Omega), \\ +\infty & \text{otherwise.} \end{cases} \tag{17.161}$$

Note that $L^\infty(\Omega) \subset \operatorname{dom}\Phi$ (a consequence of the continuity of β and j). Thus, the domain of Ψ is dense in $H^{-1}(\Omega)$. By the regularizing effect (cf. Theorem 17.2.3) and Theorem 17.2.13, which describes the subdifferential of Φ on the space $H^{-1}(\Omega)$, we deduce that for any $u_0 \in H^{-1}(\Omega)$, there exists a unique strong solution u of the Stefan problem (17.159). Moreover, $u(t) \in \operatorname{dom}\Psi = \operatorname{dom}\Phi$ for all $t > 0$. That's the existence part (a) of the above theorem.

Concerning the asymptotic behavior, note that

$$-\Delta\beta(u) = h - \frac{\partial u}{\partial t}.$$

Since $\frac{\partial u}{\partial t}(t)$ converges strongly to zero in $H^{-1}(\Omega)$ as $t \to +\infty$ (see Proposition 17.2.8) we deduce that $\beta(u(x,t))$ converges strongly in $H_0^1(\Omega)$ to the solution of the Dirichlet problem (17.160). The weak convergence of $u(t)$ in $H^{-1}(\Omega)$ follows from Theorem 17.2.7. □

Remark 17.2.5. 1. Depending on the choice of the monotone graph β, (17.140) provides different PDEs. After transformation, the free boundary Stefan problem (ice-water) corresponds to

$$\beta(r) = \begin{cases} \alpha r & \text{for } r \le 0, \\ 0 & \text{for } 0 < r < k, \\ \beta(r-k) & \text{for } r \ge k, \end{cases}$$

where α and β are positive diffusion coefficients, and k is related to the latent heat (the amount of heat energy required to change the phase of an unit mass of a substance). The case $\beta(r) = |r|^{p-1}r$ has also received a lot of attention, because of its role in porous media; see [36], [314], [356].

2. The above approach to the Stefan problem illustrates the flexibility of the gradient flow methods. By playing with the potential and the metric (the scalar product may also vary with time), one can get different types of PDEs; see [49], [134], [169], [188], and references therein.

3. $L^1(\Omega)$ plays a central role in the above analysis. Indeed, one can develop a different approach to the Stefan problem, which is based on the maximal accretivity property of the operator $-\Delta\beta$ in $L^1(\Omega)$; see [94].

4. The Stefan problem has a natural link with other phase transition models. The solution of the Cahn–Hilliard phase separation equation for a binary mixture is reasonably comparable with the solution of a Stefan problem; see [318].

17.3 ▪ Gradient flow associated to a tame function. Kurdyka–Łojasiewicz theory

17.3.1 ▪ The analytic case, Łojasiewicz inequality

Let us first recall some classical notation and basic definitions concerning real-valued analytic functions of several variables.

Definition 17.3.1. *Let $K \in \mathbf{N}^N$ be an N-dimensional multi-index, $K = (k_1, \dots, k_N)$.*

(a) *We set $|K| = \sum_{i=1}^N k_i$, $K! = \prod k_i!$, and for each $x \in \mathbf{R}^N$, $x^K = {x_1}^{k_1} \times \cdots \times {x_N}^{k_N}$.*

(b) *Given* $\Phi : \mathbf{R}^N \to \mathbf{R}$, *and* $a \in \mathbf{R}^N$, *we set*

$$D^K\Phi(a) = \frac{\partial^{|K|}\Phi}{\partial x_1^{k_1} \cdots \partial x_N^{k_N}}(a).$$

Definition 17.3.2. *Let* U *be an open subset of* $\mathbf{R}^N$. *A function* $\Phi : U \subset \mathbf{R}^N \to \mathbf{R}$ *is said to be analytic if locally it is given by a convergent power series. An analytic functions is infinitely differentiable and is equal to its Taylor series in some neighborhood of each point of its domain. More precisely, for each* $a \in U$, *there exists an open set* $\mathcal{O}$, $a \in \mathcal{O} \subset U$ *such that for all* $x \in \mathcal{O}$

$$\Phi(x) = \sum_{K \in \mathbf{N}^N} \frac{D^K\Phi(a)}{K!}(x-a)^K.$$

The following well-known result is due to Łojasiewicz [278]. It is known as the Łojasiewicz inequality. It plays a central role in the proof of the convergence of the orbits of the gradient flow associated to a real-analytic potential. Its further extensions are at the core of the modern semialgebraic and semianalytic geometry.

Theorem 17.3.1 (Łojasiewicz inequality). *Let* U *be an open set in* $\mathbf{R}^N$, $\Phi : U \subset \mathbf{R}^N \to \mathbf{R}$ *a real-analytic function, and* $\bar{u} \in U$ *a critical point of* Φ. *Then, there exists* $\theta \in [\frac{1}{2}, 1)$, $C > 0$, *and a neighborhood* W *of* $\bar{u}$ *such that*

$$\forall v \in W \quad |\Phi(v) - \Phi(\bar{u})|^\theta \leq C\|\nabla\Phi(v)\|.$$

Remark 17.3.1. The Łojasiewicz inequality is trivially satisfied at any point $\bar{u}$ which is not critical.

Remark 17.3.2. An elegant proof of the Łojasiewicz inequality can be found in [267]. In dimension $N = 1$, the proof is elementary, as shown below. By analyticity of Φ, there exists a sequence (a_k) of real numbers, $p_0 \geq 2$, and $a_{p_0} \neq 0$ such that for all v in a neighborhood of $\bar{u}$

$$\Phi(v) - \Phi(\bar{u}) = \sum_{k=p_0}^{+\infty} a_k (v - \bar{u})^k.$$

Differentiating term by term, we obtain

$$\Phi'(v) = \sum_{k=p_0}^{+\infty} k a_k (v - \bar{u})^{k-1}.$$

Taking $\theta \in \mathbf{R}_*^+$, and $v \neq \bar{u}$ close to $\bar{u}$,

$$\frac{|\Phi(v) - \Phi(\bar{u})|^\theta}{|\Phi'(v)|} \approx \frac{1}{p_0 |a_{p_0}|^{1-\theta}} |v - \bar{u}|^{p_0(\theta-1)+1}.$$

By taking $1 > \theta > 1 - \frac{1}{p_0}$ and v sufficiently close to $\bar{u}$, we obtain

$$|\Phi(v) - \Phi(\bar{u})|^\theta \leq |\Phi'(v)|.$$

Remark 17.3.3. As a consequence of the Łojasiewicz inequality, we obtain that the critical values of a real-analytic function are isolated. Indeed, if $\bar{u} \in U$ is a critical point of Φ, and v is another critical point in the neighborhood W of $\bar{u}$ for which the inequality holds, then $\Phi(v) = \Phi(\bar{u})$.

Let us stress that, in general, the critical points of a real-analytic function are not isolated. For example, the function $\Phi : U \subset \mathbf{R}^2 \to \mathbf{R}$ which is defined by $\Phi(x) = \left(||x||^2 - 1\right)^2$ is a polynomial (hence analytic) function whose critical set is $S^1 \cup 0$. (The elements of S^1 are critical points which are not isolated.) By contrast the critical values of Φ are 0 and 1, which are isolated in $\mathbf{R}$.

Remark 17.3.4. It is convenient to reformulate the Łojasiewicz inequality as follows. Take $\alpha(s) = cs^{1-\theta}$ (with constant c ad hoc). Then for all v in W that satisfies $\Phi(v) > \Phi(\bar{u})$, we have

$$\alpha'(\Phi(v) - \Phi(\bar{u}))||\nabla\Phi(v)|| \geq 1. \tag{17.162}$$

Equivalently

$$||\nabla(\alpha \circ (\Phi - \Phi(\bar{u})))(v)|| \geq 1.$$

The function α is called a desingularizing function (this terminology is justified by the above property). Note that α is an increasing concave function, and $\alpha(0) = 0$. This formulation is quite useful for the geometrical understanding of the Łojasiewicz inequality and further generalizations.

Theorem 17.3.2 (Łojasiewicz [279]). *Let $\Phi : U \subset \mathbf{R}^N \to \mathbf{R}$ be a real-analytic function. Then any bounded trajectory of the steepest descent dynamical system*

$$\text{(SD)} \quad \dot{u}(t) + \nabla\Phi(u(t)) = 0$$

has a finite length and converges to a critical point of Φ.

PROOF. Φ is a real-analytic function. In particular, it is continuously differentiable, and its gradient is Lipschitz continuous on bounded sets. Hence, we can apply Theorems 17.1.1 and 17.1.3, which concern the classical continuous steepest descent. Note that in these theorems, we make the assumption "Φ is minorized." Indeed, in the proof, we only use the weaker property "Φ is minorized along the orbit." In our situation, this last property is satisfied: it is a consequence of the fact that the orbit u is bounded, Φ is continuous, and we are in a finite dimensional setting. As a consequence, any orbit u of (SD) satisfies the following:

(i) $t \mapsto \Phi(u(t))$ is a decreasing function, and for all $t \geq 0$, $\frac{d}{dt}\Phi(u(t)) = -||\dot{u}(t)||^2$.

(ii) $\int_0^\infty ||\dot{u}(t)||^2 dt < +\infty$.

(iii) $\lim_{t\to+\infty} \dot{u}(t) = 0, \quad \lim_{t\to+\infty} \nabla\Phi(u(t)) = 0$.

As a key property providing asymptotic convergence, we are going to show that in the analytic case, the orbit has a finite length, i.e.,

$$\int_0^{+\infty} ||\dot{u}(t)||dt < +\infty. \tag{17.163}$$

Indeed, this last property implies that for any $0 \le s \le t < +\infty$

$$\begin{aligned}\|u(t)-u(s)\| &\le \int_s^t \|\dot u(\tau)\| d\tau \\ &\le \int_0^t \|\dot u(\tau)\| d\tau - \int_0^s \|\dot u(\tau)\| d\tau.\end{aligned}$$

By the finite length property (17.163), we immediately infer that $\lim_{s,t\to+\infty} \|u(t)-u(s)\| = 0$. This Cauchy property implies the convergence of $u(t)$ as $t \to +\infty$. By continuity of $\nabla\Phi$, and item (iii) above, we will obtain that the limit is a critical point of Φ.

Thus the only point we need to prove is (17.163). The orbit has been assumed to be bounded in $\mathbf{R}^N$. Let $u(t_n) \to u_\infty$ for some sequence $t_n \to +\infty$. Then $\nabla\Phi(u_\infty) = 0$, i.e., u_∞ is a critical point of Φ. By item (i) above, and continuity of Φ,

$$\lim_{t\to+\infty} \Phi(u(t)) = \inf_{t\ge 0} \Phi(u(t)) = \Phi(u_\infty). \tag{17.164}$$

We are going to consider two cases:

Case 1: *There exists some $T > 0$ such that* $\Phi(u(T)) = \Phi(u_\infty)$. By the nonincreasing property of $t \mapsto \Phi(u(t))$ (item (i)) above, and (17.164) we deduce that for any $T \le t < +\infty$

$$\Phi(u_\infty) \le \Phi(u(t)) \le \Phi(u(T)) = \Phi(u_\infty).$$

Hence $\Phi(u(\cdot))$ is constant for $T \le t < +\infty$. From item (i), and $\frac{d}{dt}\Phi(u(t)) = -\|\dot u(t)\|^2$, we deduce that $\dot u(t) = 0$ for $T \le t < +\infty$. As a consequence

$$\int_0^{+\infty} \|\dot u(t)\| dt = \int_0^T \|\dot u(t)\| dt \le \sqrt{T}\Big(\int_0^T \|\dot u(t)\|^2\, dt\Big)^{1/2} < +\infty.$$

Case 2: *For all $t > 0$, we have* $\Phi(u(t)) > \Phi(u_\infty)$. By the Łojasiewicz inequality (Theorem 17.3.1) there exists a neighborhood W of u_∞ and a desingularizing function α (see Remark 17.3.4) such that for all v in W that satisfies $\Phi(v) > \Phi(\bar u)$, we have

$$\alpha'(\Phi(v) - \Phi(u_\infty))\|\nabla\Phi(v)\| \ge 1. \tag{17.165}$$

We consider two steps:

Step 2.1. Suppose that for t large enough, say, $t \ge T$ for some $T > 0$, we have $u(t) \in W$. In that case, let us show that the function

$$h(t) := \alpha(\Phi(u(t)) - \Phi(u_\infty))$$

is a Lyapunov function. Indeed, by using successively the classical derivation chain rule, and the (SD) equation, we have

$$\begin{aligned}\dot h(t) &= \alpha'(\Phi(u(t)) - \Phi(u_\infty))\,\langle \nabla\Phi(u(t)), \dot u(t)\rangle \\ &= -\alpha'(\Phi(u(t)) - \Phi(u_\infty))\,\|\,\nabla\Phi(u(t))\,\|^2 .\end{aligned}$$

Equivalently

$$\dot h(t) + \big[\alpha'(\Phi(u(t)) - \Phi(u_\infty))\,\|\,\nabla\Phi(u(t))\,\|\big]\,\|\,\nabla\Phi(u(t))\,\| = 0. \tag{17.166}$$

Since $u(t) \in W$ for $t \ge T$, and $\Phi(u(t)) > \Phi(u_\infty)$, by the Łojasiewicz inequality (17.165), we have, for all $t \ge T$,

$$\alpha'(\Phi(u(t)) - \Phi(u_\infty))\|\nabla\Phi(u(t))\| \ge 1. \tag{17.167}$$

From (17.166) and (17.167) we infer

$$\dot{h}(t) + \| \nabla\Phi(u(t)) \| \le 0. \tag{17.168}$$

By (SD) we have $\nabla\Phi(u(t)) = -\dot{u}(t)$. Hence, (17.168) gives

$$\dot{h}(t) + \| \dot{u}(t) \| \le 0.$$

Since h is nonnegative, integration of this inequality from T to $t > T$ gives

$$\int_T^t \|\dot{u}(s)\| ds \le h(T) < +\infty.$$

This majorization being valid for any $t > T$, we finally obtain (17.163).

Step 2.2. Let us show that one can always find some $T > 0$ such that $u(t) \in W$ for $t \ge T$ and thus reduce to the situation examined in the previous Step 2.1.

Set $R > 0$ such that the ball $B(u_\infty, R) \subset W$. Since $u(t_n) \to u_\infty$, and α is continuous at 0 ($\alpha(0) = 0$), there exists some integer N such that

$$\begin{aligned} \|u(t_N) - u_\infty\| &< \frac{R}{2}, \\ \alpha(\Phi(u(t_N)) - \Phi(u_\infty)) &< \frac{R}{2}. \end{aligned} \tag{17.169}$$

Let us show that $T = t_N$ satisfies the desired property. Let us argue by contradiction and suppose that there exists some $t_0 \ge t_N$ such that $u(t_0) \notin B(u_\infty, R)$. Set

$$J = \min\{t \ge t_N : \|u(t) - u_\infty\| = R\}.$$

For all $t \in [t_N, J]$ we have $u(t) \in B(u_\infty, R)$. Thus we can apply the Łojasiewicz inequality (17.165), and by the same argument as in Step 2.1, we obtain that for any $t \in [t_N, J]$

$$\dot{h}(t) + \| \dot{u}(t) \| \le 0.$$

By integration of this inequality from t_N to J, we obtain

$$\begin{aligned} \|u(J) - u(t_N)\| &\le \int_{t_N}^J \|\dot{u}(\tau)\| d\tau \\ &\le h(t_N) - h(J) \\ &\le h(t_N). \end{aligned}$$

By (17.169), we have $h((t_N) = \alpha(\Phi(u(t_N)) - \Phi(u_\infty)) < \frac{R}{2}$. Hence $\|u(J) - u(t_N)\| < \frac{R}{2}$. By the triangle inequality, we deduce that

$$\begin{aligned} R = \|u(J) - u_\infty\| &\le \|u(J) - u(t_N)\| + \|u(t_N) - u_\infty\| \\ &< \frac{R}{2} + \frac{R}{2} = R, \end{aligned}$$

a clear contradiction. $\square$

In the case of an analytic potential, besides the finite length property, the gradient flow enjoys an other remarkable asymptotic convergence property. The following directional convergence property was first conjectured by Thom [350]: denoting by $\bar{u}$ the limit of the orbit $u(\cdot)$,

"when $t \to +\infty$, the secants $\frac{u(t)-\bar{u}}{\|u(t)-\bar{u}\|}$ converge towards a fixed direction of the unit sphere."

Indeed, the above so-called Thom conjecture for the gradient orbits of real-analytic functions holds true (see Kurdyka, Mostowski, and Parusinski [268] for the proof). Let us state it precisely.

Theorem 17.3.3. *Let $\Phi : U \subset \mathbf{R}^N \to \mathbf{R}$ be a real analytic function, and let $t \mapsto u(t)$ be an orbit of the associated gradient flow, which converges to a critical point $\bar{u}$ of Φ.*

Then the directional convergence property holds: there exists $d \in S^{N-1}$ such that

$$\lim_{t \to +\infty} \frac{u(t)-\bar{u}}{\|u(t)-\bar{u}\|} = d.$$

Remark 17.3.5. Thom's conjecture fails for convex functions. In a recent article [191], Daniilidis, Ley, and Sabourau showed that there exists a function $f : \mathbf{R}^2 \to \mathbf{R}$ convex, $\mathscr{C}^\infty$, and a trajectory of (SD) which turns infinitely many times around its limit.

17.3.2 ▪ The finite dimensional nonsmooth case. Kurdyka–Łojasiewicz inequality

In the above analysis of the asymptotic analysis of the gradient flow for a real-analytic function, the basic ingredient is the Łojasiewicz inequality. To cover many interesting applications in various field, we need to extend the theory to a larger class of functions, possibly including nonsmooth functions. To achieve this goal, we first need to extend the Łojasiewicz inequality so as to consider nonsmooth functions. This requires extending the concept of gradient in the concept of subdifferential. Then, we must show that there is a large class of functions that satisfy this extended Łojasiewicz inequality. It is a broad program that has been successfully developed over the past 10 years, thanks to the contribution of mathematicians from different domains (real algebraic geometry, variational nonsmooth analysis, optimization, signal, PDEs) and is still ongoing. An interesting survey of the geometrical and variational aspects can be found in [190], [247] and of the optimization aspects in [40], [41].

We wish to delineate some key concepts and results in this section, which is devoted to the finite dimensional case. Extension to the infinite dimensional setting and applications to PDEs will not be discussed. The interested reader can consult Haraux and Jendoubi [236].

Tools from variational analysis. As a standing assumption, in this section, $f : \mathbf{R}^N \to \mathbf{R} \cup \{+\infty\}$ is a proper lower semicontinuous function. The notion of subdifferential plays a central role in the following developments. We give some basic definitions and properties. For an extended survey of the methods of nonsmooth analysis for variational problems, see Clarke [177], Mordukhovich [299], or Rockafellar and Wets [330].

Definition 17.3.3. *For each $x \in \operatorname{dom} f$, the* Fréchet subdifferential *of f at x, written $\hat{\partial} f(x)$, is the set of vectors $v \in \mathbf{R}^N$ which satisfy*

$$\hat{\partial} f(x) := \left\{ x^* \in \mathbf{R}^N : \quad \liminf_{\substack{y \neq x \\ y \to x}} \frac{1}{||x-y||}[f(y)-f(x)-\langle x^*, y-x\rangle] \geq 0 \right\}.$$

When $x \notin \operatorname{dom} f$, we set $\hat{\partial} f(x) = \emptyset$.

The limiting processes used in the dynamical and algorithmic context necessitate the introduction of the more stable notion of *limiting subdifferential* [299] (or simply subdifferential) of f.

Definition 17.3.4. *The subdifferential of f at $x \in \operatorname{dom} f$, written $\partial f(x)$, is defined as follows:*

$$\partial f(x) := \{x^* \in \mathbf{R}^N : \exists x_k \to x,\ f(x_k) \to f(x),\ x_k^* \in \hat{\partial} f(x_k) \to x^*\}.$$

It is straightforward to check from the definition the following closedness property of ∂f: Let $(x^k, v^k)_{k \in \mathbf{N}}$ be a sequence in $\mathbf{R}^N \times \mathbf{R}^N$ such that $(x^k, v^k) \in \operatorname{graph}(\partial f)$ for all $k \in \mathbf{N}$. If (x^k, v^k) converges to (x, v), and $f(x^k)$ converges to $f(x)$, then $(x, v) \in \operatorname{graph}(\partial f)$.

These generalized notions of differentiation give birth to generalized notions of critical point. A necessary (but not sufficient) condition for $x \in \mathbf{R}^N$ to be a minimizer of f is

$$\partial f(x) \ni 0. \tag{17.170}$$

A point that satisfies (17.170) is called *limiting critical* or simply critical.

Kurdyka–Łojasiewicz inequality. We can now introduce the Kurdyka–Łojasiewicz inequality, which is an extension to the nonsmooth setting of the Łojasiewicz inequality. Indeed, it is the reformulation (17.162) (see Remark 17.3.4) of the Łojasiewicz inequality via a desingularizing function, which fits well such a nonsmooth extension.

Definition 17.3.5. *Let $f : \mathbf{R}^N \to \mathbf{R} \cup \{+\infty\}$ be a proper lower semicontinuous function.*

(a) *We say that f has the* (KL) *property at $\bar{x} \in \operatorname{dom} \partial f$ if there exists $\eta \in]0, +\infty]$, a neighborhood U of $\bar{x}$, and $\varphi : [0, \eta[\to \mathbf{R}_+$ (desingularizing function) which verifies*

- *$\varphi(0) = 0$; $\varphi : [0, \eta[\to \mathbf{R}_+$ is continuous; $\varphi \in \mathscr{C}^1(]0, \eta[)$;*
- *φ is increasing: $\varphi'(s) > 0$ for all $s \in]0, \eta[$;*
- *φ is concave;*

and such that for all x in $U \cap [f(\bar{x}) < f < f(\bar{x}) + \eta]$, the (KL) *inequality holds:*

$$\text{(KL)}\quad \varphi'(f(x) - f(\bar{x}))\ \operatorname{dist}(0, \partial f(x)) \geq 1. \tag{17.171}$$

(b) *Proper lower semicontinuous functions that satisfy the* (KL) *inequality at each point of* dom ∂f *are called* (KL) *functions.*

Remark 17.3.6. The (KL) inequality has a rich story. The general concept as defined above has been gradually emerging. The following are some important steps:

- Łojasiewicz in [278] (1963) introduced the concept for real analytic functions with $\varphi(s) = s^{1-\theta}$, $\theta \in [\frac{1}{2}, 1)$.

• Kurdyka in [266] (1998) extended the concept to differentiable functions definable in an o-minimal structure (semialgebraic, subanalytic), whence the terminology.

• Bolte et al. in [105] (2007) gave the first extension to nonsmooth functions by considering Clarke subgradients of nonsmooth functions definable in an o-minimal structure.

• Attouch et al. in [40] (2010) introduced the above (KL) formulation; see also [41].

Semialgebraic sets and functions. Functions which can be defined by a finite number of polynomial equalities or inequalities are called semialgebraic. They provide a rich class of functions satisfying the (KL) inequality.

Definition 17.3.6. (a) *A subset S of $\mathbf{R}^N$ is a* real semialgebraic set *if there exists a finite number of real polynomial functions $P_{ij}, Q_{ij} : \mathbf{R}^N \to \mathbf{R}$ such that*

$$S = \bigcup_{j=1}^{p} \bigcap_{i=1}^{q} \{ x \in \mathbf{R}^n : \; P_{ij}(x) = 0, \, Q_{ij}(x) < 0 \}.$$

(b) *A function $f : \mathbf{R}^N \to \mathbf{R} \cup \{+\infty\}$ (respectively, a point-to-set mapping $F : \mathbf{R}^N \to \mathbf{R}^m$) is called semialgebraic if its graph $\{(x,\lambda) \in \mathbf{R}^{N+1} : f(x) = \lambda\}$ (respectively, $\{(x,y) \in \mathbf{R}^{N+m} : y \in F(x)\}$) is a semialgebraic subset of $\mathbf{R}^{N+1}$ (respectively, $\mathbf{R}^{N+m}$).*

One easily sees that the class of semialgebraic sets is stable under the operation of finite union, finite intersection, Cartesian product, or complementation and that polynomial functions are, of course, semialgebraic functions.

The high flexibility of the concept of semialgebraic sets is captured by the following fundamental theorem, known as the Tarski–Seidenberg principle.

Theorem 17.3.4 (Tarski–Seidenberg). *Let A be a semialgebraic subset of $\mathbf{R}^{N+1}$; then its canonical projection on $\mathbf{R}^N$, namely,*

$$\left\{ (x_1, \dots, x_N) \in \mathbf{R}^N : \exists z \in \mathbf{R}, (x_1, \dots, x_N, z) \in A \right\},$$

is a semialgebraic subset of $\mathbf{R}^N$.

Let us illustrate the power of this theorem by proving that max functions associated to polynomial functions are semialgebraic. Let S be a nonempty semialgebraic subset of $\mathbf{R}^m$ and $g : \mathbf{R}^N \times \mathbf{R}^m \to \mathbf{R}$ a real polynomial function. Set $f(x) = \sup\{g(x,y) : y \in S\}$. (Note that f can assume infinite values.) Let us prove that f is semialgebraic.

Using the definition and the stability with respect to finite intersection, we see that the set

$$\begin{aligned} &\{(x,\lambda,y) \in \mathbf{R}^N \times \mathbf{R} \times S : g(x,y) > \lambda\} \\ &= \{(x,\lambda,y) \in \mathbf{R}^N \times \mathbf{R} \times \mathbf{R}^m : g(x,y) > \lambda\} \bigcap \left(\mathbf{R}^N \times \mathbf{R} \times S\right) \end{aligned}$$

is semialgebraic. For (x,λ,y) in $\mathbf{R}^N \times \mathbf{R} \times \mathbf{R}^m$, define the projection $\Pi(x,\lambda,y) = (x,\lambda)$ and use Π to project the above set on $\mathbf{R}^N \times \mathbf{R}$. One obtains the following semialgebraic set:

$$\{(x,\lambda) \in \mathbf{R}^N \times \mathbf{R} : \exists y \in S, \; g(x,y) > \lambda\}.$$

The complement of this set is

$$\{(x,\lambda) \in \mathbf{R}^N \times \mathbf{R} : \forall y \in S, \; g(x,y) \leq \lambda\} = \operatorname{epi} f.$$

Hence epi f is semialgebraic. Similarly, hypo $f := \{(x,\mu) \in \mathbf{R}^N \times \mathbf{R} : f(x) \geq \mu\}$ is semialgebraic. Hence graph$f = \text{epi}\, f \cap \text{hypo}\, f$ is semialgebraic. Clearly, this result also holds when replacing sup by inf.

As a byproduct of these stability results, we recover the following standard result which is useful in optimization when using for example a penalization method.

Lemma 17.3.1. *Let S be a nonempty semialgebraic subset of $\mathbf{R}^m$; then the function*

$$\mathbf{R}^m \ni x \mapsto \text{dist}(x,S)^2$$

is semialgebraic.

PROOF. It suffices to consider the polynomial function $g(x,y) = ||x-y||^2$ for x, y in $\mathbf{R}^m$ and to use the definition of the distance function. □

Remark 17.3.7. The fact that the composition of semialgebraic mappings gives a semialgebraic mapping or that the image (respectively, the preimage) of a semialgebraic set by a semialgebraic mapping is a semialgebraic set is also a consequence of the Tarski–Seidenberg principle. See [93, 101] for these and many other consequences of this principle.

Remark 17.3.8. Numerical analysis provides numerous examples of semialgebraic objects [283]: for example, the cone of the positive semidefinite matrices, Stiefel manifolds (spheres, orthogonal group [205]), and matrices with fixed rank.

The following result makes the link between semialgebraic structures and the (KL) inequality.

Theorem 17.3.5. *Let $f : \mathbf{R}^N \to \mathbf{R} \cup \{+\infty\}$ be a proper lower semicontinuous function. Then the following implication holds:*

$$f \text{ semialgebraic} \Rightarrow f \text{ satisfies (KL) inequality,}$$

with $\varphi(s) = cs^{1-\theta}$, for the same $\theta \in [0,1) \cap \mathbf{Q}$ and $c > 0$.

o-minimal structures and (KL) inequality. o-minimal structures correspond to an axiomatization of the geometrical properties of semialgebraic sets, particularly of the stability under projection (Tarski–Seidenberg). Let us cite the important contributions to this theory of Van den Dries [201], Coste [180], and Shiota [332]. This construction allows us to define new classes of sets and functions (semilinear, semialgebraic, subanalytic), for which the (KL) inequality is still valid. Clearly, this considerably enlarges the range of application of this theory. Let us give its main lines.

Definition 17.3.7. *Let $\mathcal{O} = \{\mathcal{O}_n\}_{n \in \mathbf{N}}$, where $\mathcal{O}_n$ is a collection of subsets of $\mathbf{R}^n$. $\mathcal{O}$ is an o-minimal structure iff the following hold:*

(i) *Each $\mathcal{O}_n$ is a boolean algebra: $\emptyset \in \mathcal{O}_n$, A, B in $\mathcal{O}_n \Rightarrow A \cup B, A \cap B, \mathbf{R}^n \setminus A \in \mathcal{O}_n$.*

(ii) *For all A in $\mathcal{O}_n$, $A \times \mathbf{R}$ and $\mathbf{R} \times A$ belong to $\mathcal{O}_{n+1}$.*

(iii) *For all A in $\mathcal{O}_{n+1}$, $\Pi(A) := \{(x_1, \dots, x_n) \in \mathbf{R}^n : (x_1, \dots, x_n, x_{n+1}) \in A\} \in \mathcal{O}_n$.*

(iv) *For all $i \neq j$ in $\{1, \dots, n\}$, $\{(x_1, \dots, x_n) \in \mathbf{R}^n : x_i = x_j\} \in \mathcal{O}_n$.*

(v) *The set* $\{(x_1, x_2) \in \mathbf{R}^2 : x_1 < x_2\}$ *belongs to* $\mathcal{O}_2$.

(vi) *The elements of* $\mathcal{O}_1$ *are exactly finite unions of intervals.*

Definition 17.3.8. *Definable sets and functions are defined as follows:*

- *A is definable in $\mathcal{O}$ iff A belongs to $\mathcal{O}$.*
- *$f : \mathbf{R}^N \to \mathbf{R} \cup \{+\infty\}$ is definable iff its graph is a definable subset of* $\mathbf{R}^N \times \mathbf{R}$.

The following important result from Bolte et al. [105] makes the link between functions definable in an o-minimal structure and (KL) inequality.

Theorem 17.3.6. *Let $f : \mathbf{R}^N \to \mathbf{R} \cup \{+\infty\}$ be lower semicontinuous, definable in an o-minimal structure $\mathcal{O}$. Then, f has the* (KL) *property at each point of* $\operatorname{dom} \partial f$. *Moreover, the desingularizing function φ is definable in $\mathcal{O}$.*

Remark 17.3.9. Let us mention some further examples of functions satisfying (KL):

- Uniform convexity: for all $x, y \in \mathbf{R}^N$, $x^* \in \partial f(x)$,

$$f(y) \geq f(x) + \langle x^*, y - x\rangle + K\|y - x\|^p, \quad p \geq 1,$$

$$\Rightarrow f \text{ satisfies (KL) with } \phi(s) = c s^{1/p}.$$

- Linearly regular intersection of F_i, transversality, [283]: we have

$$f(x) := \frac{1}{2} \sum_i \operatorname{dist}(x, F_i)^2 \quad \text{satisfies (KL).}$$

- Metric regularity: $F : \mathbf{R}^N \to \mathbf{R}^m$ is metrically regular at $\bar{x} \in \mathbf{R}^N$ if there exists a neighborhood V of $\bar{x}$ in $\mathbf{R}^N$, a neighborhood W of $F(\bar{x})$ in $\mathbf{R}^m$, and $k > 0$,

$$x \in V,\ y \in W \ \Rightarrow\ \operatorname{dist}(x, F^{-1}(y)) \leq k\ \operatorname{dist}(y, F(x)).$$

  If F is metrically regular, then [40]

$$f(x) = \frac{1}{2} \operatorname{dist}^2(F(x), C) \quad \text{satisfies (KL), } C \subset \mathbf{R}^m \text{ closed convex, } \phi(s) = c\sqrt{s}.$$

For existence of a smooth convex $f : \mathbf{R}^2 \to \mathbf{R}$ which does not satisfy (KL) see Bolte et al. [106] and Daniilidis, Ley, and Sabourau [191].

Asymptotic behavior of gradient flows and (KL) inequality. The study of the gradient flow associated to a general lower semicontinuous function is a broad subject, still in progress. Let us briefly describe the results obtained in [103] by Bolte, Daniilidis, and Lewis. They concern the subgradient dynamical system associated to a nonsmooth subanalytic function and the use of the (KL) inequality in the study of asymptotic behavior.

We make the following assumptions:

(H1) Φ is either lower semicontinuous convex or lower-C^2 with $\operatorname{dom} \Phi = \mathbf{R}^n$.

(H2) Φ is somewhere finite ($\operatorname{dom} f \neq \emptyset$) and bounded from below.

(H3) Φ is a subanalytic function.

We recall (see [330, Definition 10.29], for example) that a function f is called lower-C^2 if for every $x_0 \in \operatorname{dom} f$ there exist a neighborhood U of x_0, a compact topological space S, and a jointly continuous function $F : U \times S \to \mathbf{R}$ satisfying $f(x) = \max_{s\in S} F(x,s)$, for all $x \in U$, and such that the (partial) derivatives $\nabla_x F$ and $\nabla_x{}^2 F$ exist and are jointly continuous.

The main results of [103] are summarized in the following.

Theorem 17.3.7. (a) *Under the assumptions* (H1) *and* (H2), *for every* $u_0 \in \mathbf{R}^N$ *such that* $\partial\Phi(u_0) \neq \emptyset$, *there exists a unique trajectory* $u : [0,+\infty[\to \mathbf{R}^N$ *of*

$$\begin{cases} \dot u(t) + \partial\Phi(u(t)) \ni 0, \\ u(0) = u_0. \end{cases}$$

In addition, the function $\Phi \circ u$ *is absolutely continuous, and for almost all* $t > 0$

(i) $\frac{d}{dt}(\Phi \circ u)(t) = \langle \dot u(t), v\rangle$ *for all* $v \in \partial\Phi(u(t))$,

(ii) $\|\dot u(t)\| = m_\Phi(u(t))$, $\frac{d}{dt}(\Phi \circ u)(t) = -[m_\Phi(u(t))]^2$,

where $m_\Phi(x) = \inf\{\|v\| : v \in \partial\Phi(x)\}$ *is called the nonsmooth slope of* Φ *at* x *(*$\partial\Phi$ *is the limiting subdifferential of* Φ*).*

(b) *Assume that* Φ *satisfies* (H1)–(H2)–(H3). *Then any bounded maximal orbit of the gradient flow associated to* Φ *has a finite length and converges to some critical point of* Φ.

17.4 ▪ Sequences of gradient flow problems

17.4.1 ▪ Graph-convergence of operators

Let us recall the classical notion of set convergence introduced in Remark 12.1.2, namely, the Kuratowski–Painlevé convergence for sequence of sets: let $(A_n)_{n\in\mathbf{N}}$ be a sequence of subsets of a metric space (X,d) or more generally of a topological space. The lower limit of the sequence $(A_n)_{n\in\mathbf{N}}$ is the subset of X denoted by $\liminf_{n\to+\infty} A_n$ and defined by

$$\liminf_{n\to+\infty} A_n = \left\{ x \in X : \exists x_n \to x,\ x_n \in A_n\ \forall n \in \mathbf{N} \right\}.$$

The upper limit of the sequence $(A_n)_{n\in\mathbf{N}}$ is the subset of X denoted by $\limsup_{n\to+\infty} A_n$ and defined by

$$\limsup_{n\to+\infty} A_n = \left\{ x \in X : \exists (n_k)_{k\in\mathbf{N}},\ \exists (x_k)_{k\in\mathbf{N}}\ \forall k, x_k \in A_{n_k},\ x_k \to x \right\}.$$

The sets $\liminf_{n\to+\infty} A_n$ and $\limsup_{n\to+\infty} A_n$ are clearly two closed subsets of (X,d) satisfying

$$\liminf_{n\to+\infty} A_n \subset \limsup_{n\to+\infty} A_n.$$

The sequence $(A_n)_{n\in\mathbf{N}}$ is said to be convergent if the following equality holds:

$$\liminf_{n\to+\infty} A_n = \limsup_{n\to+\infty} A_n.$$

The common value A is called the limit of $(A_n)_{n\in\mathbf{N}}$ in the Kuratowski–Painlevé sense and denoted by $K - \lim_{n\to+\infty} A_n$. Therefore, by definition $A := K - \lim_{n\to+\infty} A_n$ iff

$$\limsup_{n\to+\infty} A_n \subset A \subset \liminf_{n\to+\infty} A_n,$$

so that $x \in A = K - \lim_{n\to+\infty} A_n$ iff the two following assertions hold:

$$\forall x \in A,\ \exists (x_n)_{n\in\mathbf{N}} \quad \text{such that } \forall n \in \mathbf{N}, x_n \in A_n \text{ and } x_n \to x;$$

$$\forall (n_k)_{k\in\mathbf{N}},\ \forall (x_k)_{k\in\mathbf{N}} \quad \text{such that } \forall k \in \mathbf{N},\ x_k \in A_{n_k},\ x_k \to x \Longrightarrow x \in A.$$

In general such a convergence is not topological. One can show that there exists a topology which governs the Kuratowski–Painlevé convergence iff the space (X,d) is locally compact. For a complete study and comparison of various types of convergence and their associated topologies, namely, Vietoris, Fell, Wijsmann, Attouch–Wets, and Mosco-convergence, see [87, 88].

From now on $(V, ||.||)$ is a Banach space and V^* is its topological dual space whose dual norm is denoted by $||.||_*$ and we recall that for $(u,u^*) \in V \times V^*$ we write $\langle u^*, u\rangle$ for $u^*(u)$. Given a multivalued operator $A : V \to 2^{V^*}$, for any $v \in V$ we write Av instead of $A(v)$. Let us recall some basic definitions which were addressed in Section 17.2.2 in the Hilbertian setting:

$\operatorname{dom} A = \{v \in X : Av \neq \emptyset\}$ denotes the domain of A;
$G(A) := \{(v,v^*) \in V \times V^* : v^* \in Av\}$ denotes the graph of A;
$R(A) := \{v^* \in V^* : \exists v \in V \text{ s.t. } v^* \in Av\}$ denotes the range of A.

We define the inverse operator $A^{-1} : V^* \to V$ of A by

$$A^{-1}(v^*) = \{v \in V : v^* \in Av\}.$$

Definition 17.4.1. *An operator $A : V \to 2^{V^*}$ is said to be monotone if $\langle u^* - v^*, u - v\rangle \geq 0$ whenever $(u,u^*) \in G(A)$ and $(v,v^*) \in G(A)$. It is maximal monotone if it is monotone and if its graph is maximal among all the monotone operators mapping V to V^* when $V \times V^*$ is ordered by inclusion. An element (u,u^*) of $V \times V^*$ is said to be monotonically related to a monotone operator A provided*

$$\langle u^* - v^*, u - v\rangle \geq 0 \quad \forall (v,v^*) \in G(A).$$

A useful form of the definition of maximality for a monotone operator A is the following condition, whose proof follows straightforwardly from the foregoing definition (see also Definition 17.2.1).

Proposition 17.4.1. *Let $A : V \to 2^{V^*}$ be a monotone operator. Then A is maximal monotone iff whenever (u,u^*) is monotonically related to A, then $u \in \operatorname{dom} A$ and $u^* \in Au$.*

Given a sequence of operators, one can consider the lim inf and lim sup of the sequence of their graphs as subsets of $V \times V^*$. This leads to the following definition.

Definition 17.4.2. *A sequence $(A_n)_{n\in\mathbf{N}}$ of operators mapping V to V^* is said to be graph convergent to an operator $A : V \to 2^{V^*}$ if the sequence $(G(A_n))_{n\in\mathbf{N}}$ converges to the graph $G(A)$ of A in the sense of Kuratowski and Painlevé when $V \times V^*$ is endowed with the product norm.*

From now on we systematically identify the operators with their graphs so that we write A instead of $G(A)$ and $A = G\text{–}\lim A_n$ or $A_n \xrightarrow{G} A$ instead of $G(A) = K\text{–}\lim_{n\to+\infty} G(A_n)$. When considering sequences of maximal monotone operators, the definition of the graph convergence is reduced to the following.

Proposition 17.4.2. *Let $(A_n, A)_{n\in\mathbf{N}}$ be a sequence of maximal monotone operators mapping V to V^*. Then we have*

$$A = G - \lim_{n\to+\infty} A_n \iff A \subset \liminf_{n\to+\infty} A_n. \tag{17.172}$$

PROOF. The only implication we have to establish is

$$A \subset \liminf_{n\to+\infty} A_n \Longrightarrow A = G - \lim_{n\to+\infty} A_n,$$

the converse being trivial. Thus, it remains to show that $\limsup_{n\to+\infty} A_n \subset A$ is automatically satisfied. Let $(u, u^*) \in \limsup_{n\to+\infty} A_n$; then there exists a subsequence $(n_k)_{k\in\mathbf{N}}$ of integers and $(u_k, u_k^*) \in A_{n_k}$ such that $(u_k, u_k^*) \to (u, u^*)$ in $V \times V^*$ whenever $k \to +\infty$.

On the other hand, since $A \subset \liminf_{n\to+\infty} A_n$, for all $(v, v^*) \in A$, there exists $(v_n, v_n^*) \in A_n$ such that $(v_n, v_n^*) \to (v, v^*)$ in $V \times V^*$. Going to the limit on

$$\langle u_k^* - v_{n_k}^*, u_k - v_{n_k}\rangle \geq 0$$

(recall that A_{n_k} is monotone), we infer

$$\langle u^* - v^*, u - v\rangle \geq 0 \quad \forall (v, v^*) \in A.$$

Therefore (u, u^*) is monotonically related to A and, according to Proposition 17.4.1, $(u, u^*) \in A$, which completes the proof. □

For various examples of maximal monotone operators, see [321]. In the subsection below we consider the most basic class of maximal monotone operators, namely, the class of subdifferentials of convex functions.

17.4.2 ▪ Mosco-convergence of convex potentials and graph-convergence of their subdifferential operators (Attouch theorem)

Given a convex proper function $\Phi : V \to \mathbf{R} \cup \{+\infty\}$, let us recall (see Sections 9.5 and Section 17.2.2 in the Hilbertian setting) that the subdifferential mapping $\partial\Phi : V \to 2^{V^*}$ is defined for all u in $\operatorname{dom}\Phi$ by

$$\partial\Phi(u) := \{u^* \in V^* : \Phi(v) \geq \Phi(u) + \langle u^*, v - u\rangle \ \forall v \in V\},$$

while $\partial\Phi(u) = \emptyset$ if $u \in V \setminus \operatorname{dom}\Phi$. It may also be empty at points of $\operatorname{dom}\Phi$ as shown in [321, Example 2.7] or [320, Example 3.8].

It is readily seen that $\partial\Phi$ is monotone but it is not obvious that $\partial\Phi$ is maximal (see Proposition 17.2.3 in the Hilbertian setting and Theorem 17.4.1 below). It is not even obvious that $\partial\Phi$ is not trivial, i.e., that $\operatorname{dom}\partial\Phi \neq \emptyset$ as stated in Proposition 9.5.2 under a continuity condition. The lemma below is due to Brønsted and Rockafellar [141] and strengthens Proposition 9.5.2 by showing that if Φ is a convex, proper, lower semicontinuous function, then the domain of $\partial\Phi$ is dense in the domain of Φ (see also Remark 17.2.2). This lemma is also a key argument in the proof of Rockafellar and Attouch's theorems (Theorems 17.4.1 and 17.4.4 below). Before stating this important lemma we have to define the ε-differential, a useful notion in many parts of convex analysis.

Definition 17.4.3. *Let $\Phi : V \to \mathbf{R} \cup \{+\infty\}$ be a convex, proper, lower semicontinuous function. For any $\varepsilon > 0$, the ε-subdifferential mapping $\partial_\varepsilon\Phi : V \to V^*$ of Φ is defined for all* *u in* $\operatorname{dom}\Phi$ *by*

$$\partial_\varepsilon\Phi(u) := \{u^* \in V : \Phi(v) \geq \Phi(u) + \langle u^*, v-u\rangle - \varepsilon \ \forall v \in V\},$$

while $\partial_\varepsilon\Phi(u) = \emptyset$ *if* $u \in V \setminus \operatorname{dom}\Phi$.

Remark 17.4.1. It follows from the definition that $\partial_\varepsilon\Phi(u)$ is a weak* closed convex set. Moreover, using the convexity of epiΦ and the Hahn–Banach separation theorem (Theorem 9.1.1) in $V \times \mathbf{R}$, one can prove that $\partial_\varepsilon\Phi(u)$ is nonempty for every $u \in \operatorname{dom}\Phi$ (see [320, Proposition 3.14]).

The following equivalence follows straightforwardly from the definition

$$0 \in \partial_\varepsilon\Phi(u) \iff \Phi(u) \leq \Phi(v) + \varepsilon \quad \forall v \in V,$$

i.e., u is a ε-minimizer of Φ. Moreover, from the definition of the Legendre–Fenchel conjugate of Φ and by reproducing the proof of Proposition 9.5.1, it readily follows that

$$u^* \in \partial_\varepsilon\Phi(u) \iff 0 \leq \Phi^*(u^*) + \Phi(u) - \langle u^*, u\rangle \leq \varepsilon. \tag{17.173}$$

Lemma 17.4.1 (Brønsted and Rockafellar). *Let* $\Phi : V \to \mathbf{R}\cup\{+\infty\}$ *be a convex, proper, lower semicontinuous function. Then given any* $u_0 \in \operatorname{dom}\Phi$, $\varepsilon > 0$, $\lambda > 0$, *and any* $u_0^* \in \partial_\varepsilon\Phi(u_0)$, *there exists* $u \in \operatorname{dom}\partial\Phi$ *and* $u^* \in V^*$ *such that*

$$u^* \in \partial\Phi(u), \quad \|u-u_0\| \leq \lambda \quad \text{and} \quad \|u^* - u_0^*\|_* \leq \frac{\varepsilon}{\lambda}.$$

In particular (take $\lambda = \sqrt{\varepsilon}$*), the domain of* $\partial\Phi$ *is dense in* $\operatorname{dom}\Phi$.

PROOF. The proof consists in applying the Ekeland variational principle to a suitable function. Consider the convex, proper, lower semicontinuous function $\Psi : V \to \mathbf{R}\cup\{+\infty\}$ defined by

$$\Psi(v) = \Phi(v) - \langle u_0^*, v\rangle.$$

According to Remark 17.4.1, u_0 is a ε-minimizer of Ψ. From Theorem 3.4.5, it follows that there exists $u \in \operatorname{dom}\Phi$ such that

$$\|u-u_0\| \leq \lambda;$$

$$\Psi(u) < \Psi(v) + \frac{\varepsilon}{\lambda}\|v-u\| \quad \forall v \neq u.$$

The second statement says that u is a minimizer of $v \mapsto \Psi(v) + \frac{\varepsilon}{\lambda}\|v-u\|$ so that, using the optimality condition Proposition 9.5.3, we have

$$0 \in \partial\Big(\Psi + \frac{\varepsilon}{\lambda}\|.-u\|\Big)(u).$$

Recalling the definition of Ψ, and from the classical rule about additivity of subdifferentials (cf. Theorem 9.5.4), we infer

$$u_0^* \in \partial\Phi(u) + \frac{\varepsilon}{\lambda}\partial\|.-u\|(u).$$

Equivalently, there exists $u^* \in \partial\Phi(u)$ such that $u_0^* - u^* \in \frac{\varepsilon}{\lambda}\partial\|.-u\|(u)$. The subdifferential inequality then yields

$$\frac{\varepsilon}{\lambda}\|v-u\| \geq \langle u_0^* - u^*, v-u\rangle$$

for all $v \in V$, thus $\|u_0^* - u^*\|_* \leq \frac{\varepsilon}{\lambda}$. This completes the proof. $\square$

In the case when $(V, ||.||)$ is a reflexive Banach space, by using the Moreau–Yosida regularization of Φ, the second assertion of Lemma 17.4.1, i.e., the density of $\operatorname{dom}\partial\Phi$ in $\operatorname{dom}\Phi$, may be specified as stated below in Proposition 17.4.3. (See also Theorem 9.5.3 when V is not assumed to be reflexive, or Section 17.2.1 when V is a Hilbert space.) The formulation of the statement resumes some definitions and results of Section 17.2.1 and needs some preparations.

Lemma 17.4.2 (property of the duality map). *Denote by H the subdifferential of the convex continuous function $v \mapsto \frac{1}{2}||v||^2$, also called the duality map from V into 2^{V^*}. Then H is characterized by*

$$u^* \in H(u) \iff ||u^*||_* = ||u|| \text{ and } \langle u^*, u\rangle = ||u||^2. \tag{17.174}$$

If the norms $||.||$ and $||.||_*$ are strictly convex, then H is one to one and sequentially continuous from V onto V^*, when V and V^* are equipped with their strong convergence and weak convergence, respectively. Furthermore, if the dual space $(V^*, ||.||_*)$ satisfies the property

"the weak convergence and the convergence of the norms imply the strong convergence,"

then H is strongly continuous from V onto V^*.

PROOF. From Proposition 9.5.1 $v^* \mapsto \frac{1}{2}||v^*||_*^2$ is the Legendre–Fenchel conjugate of $v \mapsto \frac{1}{2}||v||^2$, so that according to Corollary 9.3.1

$$u^* \in H(u) \iff \frac{1}{2}||u^*||_*^2 + \frac{1}{2}||u||^2 = \langle u^*, u\rangle. \tag{17.175}$$

Hence, from $\langle u^*, u\rangle \le ||u^*||_*||u||$, (17.175) yields

$$||u^*||_*^2 + ||u||^2 \le 2||u^*||_*||u||,$$

from which we deduce $||u^*||_* = ||u||$. Finally (17.175) gives (17.174).

According to Theorem 9.5.1, H^{-1} is the subdifferential of $v^* \mapsto \frac{1}{2}||v^*||_*^2$, and thus

$$u \in H^{-1}(u^*) \iff ||u^*||_* = ||u|| \text{ and } \langle u^*, u\rangle = ||u^*||_*^2.$$

If the norms $||.||$ and $||.||_*$ are strictly convex, from (17.174) we deduce that H is a one-to-one mapping from V onto V^*. Indeed, assume that there exist u_1^* and u_2^*, two elements of $H(u)$ with $u_1^* \neq u_2^*$, and take $\lambda \in]0,1[$. Since $H(u)$ is a convex subset of V^*, $\lambda u_1^* + (1-\lambda)u_2^* \in H(u)$ and from (17.174) and the strict convexity of $||.||_*$ we infer

$$||u|| = ||\lambda u_1^* + (1-\lambda)u_2^*||_* < \lambda||u_1^*||_* + (1-\lambda)||u_2^*||_* = ||u||,$$

a contradiction. The same argument, using the strict convexity of $||.||$, shows that H^{-1} is univalent.

We are going to establish the continuity of H. Let $(u_n)_{n\in\mathbf{N}}$ be a sequence converging strongly to u in V and set $u_n^* := H(u_n)$. Then from (17.174)

$$||u_n^*||_* = ||u_n||, \tag{17.176}$$

$$\langle u_n^*, u_n\rangle = ||u_n||^2. \tag{17.177}$$

From (17.176) and the fact that $u_n \to u$ in V we infer $\sup_{n\in\mathbf{N}}||u_n^*||_* < +\infty$, so that there exists a subsequence of $(u_n^*)_{n\in\mathbf{N}}$ (not relabeled) weakly converging to some v^* in V^*.

We claim that $v^* = H(u)$. Indeed, going to the limit in (17.177) and (17.176) and from the lower semicontinuity of the norm we obtain

$$\langle v^*, u\rangle = ||u||^2 \tag{17.178}$$

and $||v^*||_* \leq ||u||$. But (17.178) yields $||v^*|| \geq ||u||$ so that $||v^*||_* = ||u||$. Thus from the characterization (17.174) of $H(u)$, we conclude that $v^* = H(u)$, and that the whole sequence $(H(u_n))_{n\in\mathbf{N}}$ converges weakly to $H(u)$.

For establishing the strong continuity under the additional assumption on $(V^*, ||.||_*)$, it is enough to prove $||H(u_n)||_* \to ||H(u)||_*$ which is a straightforward consequence of (17.174),

$$||H(u_n)||_* = ||u_n|| \to ||u|| = ||H(u)||_*,$$

which completes the proof. □

The next proposition resumes some results and definitions established in Propositions 17.2.1 and 17.2.2 previously stated in the Hilbertian setting.

Proposition 17.4.3. *Let $(V, ||.||)$ be a reflexive Banach space, and assume that the norms $||.||$ and $||.||_*$ are strictly convex. Let $\Phi : V \to \mathbf{R}\cup\{+\infty\}$ be a convex, proper, lower semicontinuous function, and for very $\lambda > 0$ consider its Moreau–Yosida regularization, i.e., the function Φ_λ defined for all $u \in V$ by*

$$\Phi_\lambda(u) = \inf_{v\in V}\left\{\Phi(v) + \frac{1}{2\lambda}||v-u||^2\right\},$$

and denote by $J_\lambda u$ the unique point where the infimum is achieved. Then, for every $u \in \operatorname{dom}\Phi$

$$\frac{1}{\lambda}H(u - J_\lambda u) \in \partial\Phi(J_\lambda u), \quad \textit{in particular } J_\lambda u \in \operatorname{dom}\partial\Phi; \tag{17.179}$$

$$J_\lambda u \to u \quad \textit{strongly in } V \textit{ when } \lambda \textit{ goes to zero}; \tag{17.180}$$

$$\Phi(J_\lambda u) \to \Phi(u) \quad \textit{when } \lambda \textit{ goes to zero}. \tag{17.181}$$

PROOF. As noted in Proposition 17.2.1, the function $v \mapsto \Phi(v) + \frac{1}{2\lambda}||v-u||^2$ is convex, proper, lower semicontinuous, and coercive. Thus, from Theorem 3.3.4 it reaches its minimum at a unique point denoted by $J_\lambda u$. (Uniqueness follows from the strict convexity of the norm $||.||$.) According to the optimality condition Proposition 9.5.3, we have

$$0 \in \partial\Big(\Phi + \frac{1}{2\lambda}||.-u||^2\Big)(J_\lambda u).$$

Therefore, from the classical rule about additivity of subdifferentials (cf. Theorem 9.5.4) and since from Lemma 17.4.2 H is a one to one mapping from V onto V^*, we infer that $J_\lambda u$ satisfies the extremality condition

$$\frac{1}{\lambda}H(u - J_\lambda u) \in \partial\Phi(J_\lambda u);$$

in particular we obtain (17.179).

On the other hand, from Theorem 9.3.1, Φ admits an affine continuous minorant, so that for all $v \in V$, $\Phi(v) \geq -\alpha(||v|| + 1)$ for some constant $\alpha \geq 0$. Hence, given $u \in \operatorname{dom}\Phi$

we infer

$$\begin{aligned}\Phi(u) &= \Phi(J_\lambda u) + \frac{1}{2\lambda}\|J_\lambda u - u\|^2 \\ &\geq -\alpha\|J_\lambda u - u\| + \frac{1}{2\lambda}\|J_\lambda u - u\|^2 - \alpha(\|u\|+1) \\ &\geq \left(\frac{1}{2\lambda} - \frac{1}{2}\right)\|J_\lambda u - u\|^2 - \frac{\alpha^2}{2} - \alpha(\|u\|+1),\end{aligned}$$

so that, for λ small enough,

$$\|J_\lambda u - u\|^2 \leq \frac{2\lambda}{1-\lambda}\Big(\Phi(u) + \frac{\alpha^2}{2} + \alpha(\|u\|+1)\Big),$$

from which we derive (17.180).

By using the inequality $\Phi(u) \geq \Phi(J_\lambda u)$, the lower semicontinuity of Φ, and (17.180), we obtain

$$\Phi(u) \geq \limsup_{\lambda\to 0} \Phi(J_\lambda u) \geq \liminf_{\lambda\to 0} \Phi(J_\lambda u) \geq \Phi(u),$$

which ensures (17.181). □

Definition 17.4.4. *Under the assumptions and notation of Proposition* 17.4.3*, the operator* $J_\lambda : V \to V$ *is called the resolvent of index* λ *of* $\partial\Phi$*. We sometimes write* J_λ^Φ *to highlight the dependence on* Φ*.*

Theorem 17.4.1 (Rockafellar). *Let* $(V, \|.\|)$ *be a general Banach space and* $\Phi : V \to \mathbf{R} \cup \{+\infty\}$ *a convex proper lower semicontinuous function. Then, its subdifferential* $\partial\Phi : V \to 2^{V^*}$ *is a maximal monotone operator.*

PROOF. We follow the ideas of Simon's proof [334]. It relies on Lemma 17.4.3 below, whose proof is based on the Brønsted–Rockafellar lemma.

Lemma 17.4.3. *Let* $\Phi : V \to \mathbf{R} \cup \{+\infty\}$ *be a convex proper lower semicontinuous function and suppose that* $u_0 \in V$ *(not necessarily in* $\operatorname{dom}\Phi$*) satisfies* $\inf_{v\in V}\Phi(v) < \Phi(u_0)$*. Then there exists* $u \in \operatorname{dom}\partial\Phi$ *and* $u^* \in \partial\Phi(u)$ *such that*

$$\Phi(u) < \Phi(u_0) \quad \textit{and} \quad \langle u^*, u_0 - u\rangle > 0.$$

PROOF OF LEMMA 17.4.3. By using the subdifferential inequality, the existence of $u \in \operatorname{dom}\partial\Phi$ and $u^* \in \partial\Phi(u)$ satisfying $\langle u^*, u_0 - u\rangle > 0$ readily yields $\Phi(u) < \Phi(u_0)$. Fix $r \in \mathbf{R}$ satisfying $\inf_V \Phi < r < \Phi(u_0)$ and set

$$M := \sup_{v\in V,\ v\neq u_0} \frac{r - \Phi(v)}{\|u_0 - v\|}.$$

First step. We prove that $0 < M < +\infty$. To see that $M > 0$, pick any $v \in V$ such that $\Phi(v) < r$. We have $v \neq u_0$ and $M \geq \frac{r-\Phi(v)}{\|u_0-v\|} > 0$. We are going to prove that $M < +\infty$. For all v satisfying $\Phi(v) > r$ we have $\frac{r-\Phi(v)}{\|u_0-v\|} < 0$. Consider the closed set $E := \{v \in V : \Phi(v) \leq r\}$. Since $u_0 \notin E$, we have $\operatorname{dist}(u_0, E) > 0$. On the other hand, from Theorem 9.3.1, the function Φ admits an affine continuous minorant $w^* + \beta$, where $w^* \in V^*$ and $\beta \in \mathbf{R}$. Therefore for all $v \in E$

$$\begin{aligned} r - \Phi(v) &\leq r - \langle w^*, v \rangle - \beta \\ &= \Big(r - \langle w^*, u_0 \rangle - \beta \Big) + \langle w^*, u_0 - v \rangle \\ &\leq |r - \langle w^*, u_0 \rangle - \beta| + \|w^*\|_* \|u_0 - v\|. \end{aligned}$$

Hence

$$\begin{aligned} \frac{r - \Phi(v)}{\|u_0 - v\|} &\leq \frac{|r - \langle w^*, u_0 \rangle - \beta|}{\|u_0 - v\|} + \|w^*\|_* \\ &\leq \frac{|r - \langle w^*, u_0 \rangle - \beta|}{\operatorname{dist}(u_0, E)} + \|w^*\|_*. \end{aligned}$$

In either case, there is an upper bound for $\frac{r - \Phi(v)}{\|u_0 - v\|}$, so that $M < +\infty$.

Second step. End of the proof. Consider the function $g : V \to \mathbf{R} \cup \{+\infty\}$ defined by $g(v) = \Phi(v) + M\|u_0 - v\|$. By definition of M, for all $\varepsilon > 0$ (chosen such that $0 < \varepsilon < M$), there exists $v_\varepsilon \in V$, $v_\varepsilon \neq u_0$, such that

$$0 < M \leq \frac{r - \Phi(v_\varepsilon)}{\|u_0 - v_\varepsilon\|} + \varepsilon.$$

Hence

$$g(v_\varepsilon) \leq r + \varepsilon \|u_0 - v_\varepsilon\|. \tag{17.182}$$

On the other hand, $r \leq \inf_V g$. Indeed, for $v = u_0$, $g(v) = \Phi(u_0) > r$, and for $v \neq u_0$

$$\Phi(v) + M\|u_0 - v\| \geq \Phi(v) + \frac{r - \Phi(v)}{\|u_0 - v\|}\|u_0 - v\| = r.$$

From (17.182) we deduce that

$$g(v_\varepsilon) \leq \inf_V g + \varepsilon \|u_0 - v_\varepsilon\|,$$

so that (see Remark 17.4.1)

$$0 \in \partial_{\varepsilon \|u_0 - v_\varepsilon\|} g(v_\varepsilon).$$

We are in a position to apply the Brønsted–Rockafellar lemma, Lemma 17.4.1: choosing λ satisfying $\varepsilon < \lambda < M$, there exist $u \in \operatorname{dom} g = \operatorname{dom} \Phi$, $w^* \in \partial g(u)$ such that

$$\|u - v_\varepsilon\| \leq \frac{\varepsilon \|u_0 - v_\varepsilon\|}{\lambda}; \tag{17.183}$$

$$\|w^*\|_* \leq \lambda. \tag{17.184}$$

From (17.183) we infer that $\|u - u_0\| > 0$. Indeed

$$\begin{aligned} \|u - u_0\| &\geq \|u_0 - v_\varepsilon\| - \|u - v_\varepsilon\| \\ &\geq \|u_0 - v_\varepsilon\| \Big(1 - \frac{\varepsilon}{\lambda} \Big) > 0 \end{aligned}$$

by the choice of λ. On the other hand, from the classical rule about additivity of subdifferentials (cf. Theorem 9.5.4),

$$\partial g(u) = \partial\Phi(u) + M\partial\|.-u_0\|(u).$$

Consequently there exists $u^* \in \partial\Phi(u)$ and $z^* \in M\partial\|.-u_0\|(u)$ such that $w^* = u^* + z^*$. The subdifferential inequality

$$M\|v-u_0\| \geq M\|u-u_0\| + \langle z^*, v-u\rangle$$

related to $z^* \in M\partial\|.-u_0\|(u)$ applied for $v = u_0$ gives $\langle z^*, u-u_0\rangle \geq M\|u-u_0\|$. Therefore, from (17.184), by the choice of λ, and since $\|u_0-u\| > 0$,

$$\begin{aligned}
\langle u^*, u_0-u\rangle &= \langle w^*-z^*, u_0-u\rangle \\
&= \langle w^*, u_0-u\rangle + \langle z^*, u-u_0\rangle \\
&\geq \big(-\|w^*\|_* + M\big)\|u-u_0\| \\
&\geq (-\lambda+M)\|u-u_0\| > 0,
\end{aligned}$$

which completes the proof of Lemma 17.4.3.

PROOF OF THEOREM 17.4.1 CONTINUED. We argue by contraposition, taking into account Proposition 17.4.1. Suppose $u \in V$, $u^* \in V^*$, and $u^* \notin \partial\Phi(u)$. Thus $0 \notin \partial(\Phi - u^*)(u)$. Indeed

$$\begin{aligned}
0 \in \partial(\Phi-u^*)(u) &\iff (\Phi-u^*)(v) \geq (\Phi-u^*)(u) \quad \forall v \in V \\
&\iff \Phi(v)-\Phi(u) \geq \langle u^*, v-u\rangle \quad \forall v \in V \\
&\iff u^* \in \partial\Phi(u).
\end{aligned}$$

Thus u satisfies $\inf_{v\in V}(\Phi-u^*) < (\Phi-u^*)(u)$. By Lemma 17.4.3, there exist $z \in \operatorname{dom}(\Phi - u^*) = \operatorname{dom}\Phi$ and $z^* \in \partial(\Phi-u^*)(z)$ such that $\langle z^*, z-u\rangle < 0$. But it is easily seen that $z^* \in \partial(\Phi - u^*)(z)$ is equivalent to $z^* + u^* \in \partial\Phi(z)$. Consequently there exists $w^* \in \partial\Phi(z)$ such that $z^* = w^* - u^*$, so that $\langle w^* - u^*, z-u\rangle < 0$. $\square$

In order to establish the characterization by Rockafellar of subdifferentials among the maximal monotone operators we introduce the notion of cyclic monotonicity. We call the finite chain of V any finite family $u_0, \dots, u_{l-1}, u_l$ of elements in V with $l \geq 1$, and we call the closed chain any finite family $u_0, \dots, u_{l-1}, u_l = u_0$ of elements in V with $l \geq 1$. We write $u_0 \smile u_l$ and $u_0 \smile u_l = u_0$, respectively, for such chains.

Definition 17.4.5. *An operator $A : V \to 2^{V^*}$ is said to be cyclically monotone if*

$$\sum_{k=0}^{l-1} \langle u_k^*, u_k - u_{k+1}\rangle \geq 0$$

for every closed chain $u_0 \smile u_l = u_0$ in $\operatorname{dom} A$ *and every $u_k^* \in Au_k$.*

Note that taking $l = 2$, i.e., a closed chain $u_0 \smile u_2 = u_0$, we obtain $\langle u_0^* - u_1^*, u_0 - u_1\rangle \geq 0$ for every (u_0, u_0^*) and (u_1, u_1^*) in A, which is the definition of a monotone operator. Therefore this notion generalizes the notion of monotonicity for operators $A : V \to 2^{V^*}$.

Proposition 17.4.4. *Let $\Phi : V \to \mathbf{R} \cup \{+\infty\}$ be a convex, proper, lower semicontinuous function. Then for all finite chain $u_0 \smile u_l = u$ with $u_k \in \mathrm{dom}(\partial \Phi)$ and $(u_k, u_k^*) \in \partial \Phi$ for $k = 0, \ldots, l-1$, we have*

$$\Phi(u) \geq \Phi(u_0) + \sum_{k=0}^{l-1} \langle u_k^*, u_{k+1} - u_k \rangle. \tag{17.185}$$

Moreover its subdifferential operator $\partial \Phi$ is cyclically monotone.

PROOF. Take a finite chain $u_0 \smile u_l = u$ with $u_k \in \mathrm{dom}\, \partial \Phi$ and u_k^* with $(u_k, u_k^*) \in \partial \Phi$ for $k = 0, \ldots, l-1$. According to the definition of $\partial \Phi$, for $k = 0, \ldots, l-1$ we have

$$\Phi(u_{k+1}) \geq \Phi(u_k) + \langle u_k^*, u_{k+1} - u_k \rangle.$$

Summing these l inequalities and using the fact that $u_l = u$ we obtain

$$\Phi(u) \geq \Phi(u_0) + \sum_{k=0}^{l-1} \langle u_k^*, u_{k+1} - u_k \rangle.$$

Taking now $u = u_0$, i.e., a closed chain $u_0 \smile u_l = u_0$, the above inequality yields

$$\sum_{k=0}^{l-1} \langle u_k^*, u_{k+1} - u_k \rangle \leq 0$$

so that $\partial \Phi$ is cyclically monotone. □

From now on, unless otherwise specified, the Banach space $(V, \|.\|)$ is assumed to be reflexive. The following theorem stated without proof is due to Rockafellar [329]. It expresses that among the maximal monotone operators, the subdifferentials are the only operators which are cyclically monotone.

Theorem 17.4.2 (Rockafellar). *Let $A : V \to 2^{V^*}$ be a maximal monotone operator with* $\mathrm{dom} A \neq \emptyset$. *Then A is the subdifferential of a convex, proper, lower semicontinuous function $\Phi : V \to \mathbf{R} \cup \{+\infty\}$ iff A is cyclically monotone. In that case the following "integration" formula holds: $A = \partial \Phi$ with for every $u \in V$*

$$\Phi(u) = \sup \Bigg\{ C_0 + \sum_{k=0}^{l-1} \langle u_k^*, u_{k+1} - u_k \rangle : u_0 \smile u_l = u \\ \text{with } u_k \in \mathrm{dom} A \text{ for } k = 0, \ldots, l-1,\ l \in \mathbf{N}^* \Bigg\}, \tag{17.186}$$

where the primitive Φ is defined up to an arbitrary constant C_0 (with $C_0 = \Phi(u_0)$).

Corollary 17.4.1. *The class of subdifferentials is closed in the class of maximal monotone operators. In other words, if $A = G - \lim \partial \Phi_n$, where A is a maximal monotone operator, then there exists a convex proper lower semicontinuous function Φ such that $A = \partial \Phi$.*

PROOF. From Theorem 17.4.2, it suffices to establish that A is cyclically monotone. Let $u_0 \smile u_l = u_0$ be a closed chain in $\mathrm{dom} A$ and consider u_k^*, $k = 0, \ldots, l$, with $u_k^* \in A u_k$.

According to the fact that $A = G - \lim \partial \Phi_n$, for each $k = 0, \dots, l$ there exists $(u_k^n, u_k^{*n}) \in \partial \Phi_n$ satisfying $(u_k^n, u_k^{*n}) \to (u_k, u_k^*)$ in $V \times V^*$. Since for all $n \in \mathbf{N}$, $\partial \Phi_n$ is cyclically monotone we have

$$\sum_{k=0}^{l-1} \langle u_k^{*n}, u_k^n - u_{k+1}^n \rangle \geq 0.$$

Letting $n \to +\infty$ we infer

$$\sum_{k=0}^{l-1} \langle u_k^*, u_k - u_{k+1} \rangle \geq 0,$$

which proves the thesis. $\square$

A natural question now arises. What type of variational convergence should equip the class of convex functions so that the mapping $\Phi \mapsto \partial \Phi$ is continuous when the class of maximal monotone operators is equipped with the graph convergence? The appropriate notion is the convergence in the sense of Mosco defined below.

The Banach space $(V, \|.\|)$ being endowed with two convergences, we have two notions of Γ-convergence. Given a sequence $(\Phi_n)_{n \in \mathbf{N}}$ of functionals $\Phi_n : V \to \mathbf{R} \cup \{+\infty\}$, according to Definition 12.1.1, we denote by $\Gamma_w - \lim \Phi_n$ and $\Gamma_s - \lim \Phi_n$ the Γ-limits associated with the weak and the strong convergence in V, respectively, when they exist.

Definition 17.4.6 (Mosco-convergence). *Let $(V, \|.\|)$ be a Banach space, a sequence $(\Phi_n)_{n \in \mathbf{N}}$ of extended real-valued functions $\Phi_n : V \longrightarrow \mathbf{R} \cup \{+\infty\}$, and $\Phi : V \longrightarrow \mathbf{R} \cup \{+\infty\}$. The sequence $(\Phi_n)_{n \in \mathbf{N}}$ Mosco-converges to Φ and we write $\Phi_n \xrightarrow{M} \Phi$ if*

$$\Phi = \Gamma_w - \Phi_n = \Gamma_s - \Phi_n.$$

The argument which naturally led us to introduce the Mosco-convergence notion yields the bicontinuity of the Fenchel duality transformation in the context of convex functions. We state this more precisely in the following theorem.

Theorem 17.4.3. *Let $(V, \|.\|)$ be a reflexive Banach space, and $(\Phi_n)_{n \in \mathbf{N}}$, Φ a sequence of convex, proper, lower semicontinuous functions from V into $\mathbf{R} \cup \{+\infty\}$. The following statements are equivalent:*

(i) $\Phi_n \xrightarrow{M} \Phi$ *on* V;

(ii) $\Phi_n^* \xrightarrow{M} \Phi^*$ *on* V^*.

PROOF. Since the functions Φ_n and Φ are convex and lower semicontinuous, $(\Phi_n^*)^* = \Phi_n$ and $(\Phi^*)^* = \Phi$, so that it suffices to establish (i) $\Longrightarrow$ (ii). From hypothesis (i) and according to the notation of Section 12.1 we have

$$\Gamma_s - \limsup_{n \to +\infty} \Phi_n \leq \Phi \leq \Gamma_w - \liminf_{n \to +\infty} \Phi_n.$$

By Fenchel conjugation, these inequality are reversed, i.e.,

$$\left(\Gamma_w - \liminf_{n \to +\infty} \Phi_n\right)^* \leq \Phi^* \leq \left(\Gamma_s - \limsup_{n \to +\infty} \Phi_n\right)^*. \tag{17.187}$$

Let us recall that a sequence $(\Phi_n)_{n \in \mathbf{N}}$ is said to be uniformly proper if there exists a bounded sequence $(u_0^n)_{n \in \mathbf{N}}$ in V such that $\sup_{n \in \mathbf{N}} \Phi_n(u_0^n) < +\infty$. Since $\Phi_n \xrightarrow{M} \Phi$, the sequence $(\Phi_n)_{n\in\mathbf{N}}$ is automatically uniformly proper. Indeed, fix $u_0 \in \operatorname{dom}\Phi$. From the fact that $\Phi = \Gamma_s - \Phi_n$, there exists u_0^n strongly converging to u_0 in V such that $\Phi(u_0) = \lim_{n\to+\infty} \Phi_n(u_0^n)$, so that $(\Phi_n(u_0^n))_{n\in\mathbf{N}}$ is bounded. Note that by using the same arguments, one may prove that $(\Phi_n^*)_{n\in\mathbf{N}}$ is uniformly proper too. To complete the proof, the following result, stated without proof, is essential. (For a proof consult [37, Theorem 3.7].)

Lemma 17.4.4. *Let $(\Phi_n)_{n\in\mathbf{N}}$, $\Phi_n : V \to \mathbf{R}\cup\{+\infty\}$ be a sequence of convex, lower semicontinuous, and uniformly proper functions. Then*

$$\left(\Gamma_w - \liminf_{n\to+\infty} \Phi_n\right)^* = \Gamma_s - \limsup_{n\to+\infty} \Phi_n^*.$$

PROOF OF THEOREM 17.4.3 CONTINUED. From Lemma 17.4.4 and noticing that $\Phi_n = \Phi_n^{**}$ and that for any function g the inequality $g^{**} \le g$ holds, (17.187) yields

$$\begin{aligned}\Gamma_s - \limsup \Phi_n^* \le \Phi^* &\le \left(\Gamma_s - \limsup_{n\to+\infty} \Phi_n\right)^* \\ &= \left(\Gamma_s - \limsup_{n\to+\infty} \Phi_n^{**}\right)^* \\ &= \left(\Gamma_w - \liminf_{n\to+\infty} \Phi_n^*\right)^{**} \le \Gamma_w - \liminf_{n\to+\infty} \Phi_n^* \qquad \square\end{aligned}$$

The following proposition, whose proof is straightforward, states an equivalent formulation that is interesting from a practical point of view.

Proposition 17.4.5. *Let $(V, \|.\|)$ be a reflexive Banach space and $(\Phi_n)_{n\in\mathbf{N}}$, Φ a sequence of convex, proper, lower semicontinuous functions from V into $\mathbf{R}\cup\{+\infty\}$. The following statements are equivalent:*

(i) $\Phi_n \xrightarrow{M} \Phi$;

(ii) *for all $v \in V$, $\exists v_n \to v$ such that $\Phi_n(v_n) \to \Phi(v)$;*
for all $v \in V$, $\forall v_n \rightharpoonup v$, $\Phi(v) \le \liminf_{n\to+\infty} \Phi_n(v_n)$;

(iii) *for all $v \in V$, $\exists v_n \to v$ such that $\Phi_n(v_n) \to \Phi(v)$,*
for all $v^* \in V^*$, $\exists v_n^* \to v^*$ such that $\Phi_n^*(v_n) \to \Phi^*(v)$.

We state now the main result of this section.

Theorem 17.4.4 (Attouch). *Let $(V, \|.\|)$ be a reflexive Banach space and $(\Phi_n)_{n\in\mathbf{N}}$, Φ a sequence of convex, proper, lower semicontinuous functions from V into $\mathbf{R}\cup\{+\infty\}$. The following statements are equivalent:*

(i) $\Phi_n \xrightarrow{M} \Phi$;

(ii) *$\partial\Phi = G - \lim_{n\to+\infty} \partial\Phi_n$ and the following normalization condition (NC) holds:*
$\exists(\bar{u}, \bar{u}^*) \in \partial\Phi$, $\exists(u^n, u^{n*}) \in \partial\Phi_n$ s.t. $u^n \to \bar{u}$ in V, $u^{n*} \to \bar{u}^*$ in V^*, $\Phi_n(u^n) \to \Phi(\bar{u})$.

PROOF. *Proof of* (i) $\Longrightarrow$ (ii). The proof of this implication relies on the Brønsted–Rockafellar lemma, Lemma 17.4.1. Let $(u, u^*) \in \partial\Phi$. From Proposition 17.4.5 there exist a sequence $(u_n)_{n\in\mathbf{N}}$ in V and a sequence $(u_n^*)_{n\in\mathbf{N}}$ in V^* such that $u_n \to u$ in V, $u_n^* \to u^*$ in V^*, and

$$\begin{aligned}&\lim_{n\to+\infty} \Phi_n(u_n) = \Phi(u);\\ &\lim_{n\to+\infty} \Phi_n^*(u_n^*) = \Phi^*(u^*).\end{aligned} \tag{17.188}$$

Set $\varepsilon_n := \Phi_n(u_n)+\Phi_n^*(u_n^*)-\langle u_n^*, u_n\rangle$; then from (17.173) of Remark 17.4.1 we have $\varepsilon_n \geq 0$ and we infer that $u_n^* \in \partial_{\varepsilon_n}\Phi_n(u_n)$. On the other hand, from (17.188), and since $(u,u^*)\in \partial\Phi$, ε_n goes to $\Phi(u)+\Phi^*(u^*)-\langle u^*, u\rangle = 0$ when n goes to $+\infty$. Hence, according to Lemma 17.4.1, there exists $(v_n, v_n^*)\in\partial\Phi_n$ such that

$$\begin{aligned}&\|v_n - u_n\| \leq \sqrt{\varepsilon_n};\\ &\|v_n^* - u_n^*\|_* \leq \sqrt{\varepsilon_n}.\end{aligned}$$

This proves that $v_n \to u$ in V and $v_n^* \to u^*$ in V^*. This being true for any $(u,u^*)\in \partial\Phi$, from Proposition 17.4.2, and since from Theorem 17.4.1 $\partial\Phi_n$ and $\partial\Phi$ are maximal monotone operators, we conclude that $\partial\Phi = G-\lim_{n\to+\infty}\partial\Phi_n$.

Given any $(u,u^*)\in\partial\Phi$, we claim that the normalization condition (NC) is automatically satisfied by the sequences $(v_n)_{n\in\mathbf{N}}$ and $(v_n^*)_{n\in\mathbf{N}}$ previously considered. To prove this, we only have to establish that $\Phi_n(v_n)\to\Phi(u)$. Since $v_n^*\in\partial\Phi_n(v_n)$, we have

$$\Phi_n(u_n) \geq \Phi_n(v_n) + \langle v_n^*, u_n - v_n\rangle;$$

thus, letting $n\to+\infty$, we infer $\Phi(u)\geq \limsup_{n\to+\infty}\Phi_n(v_n)$. On the other hand, since $u_n^*\in\partial_{\varepsilon_n}\Phi_n(u_n)$, we have

$$\Phi_n(v_n) \geq \Phi_n(u_n) + \langle u_n^*, v_n - u_n\rangle - \varepsilon_n,$$

from which we infer $\liminf_{n\to+\infty}\Phi_n(v_n)\geq\Phi(u)$, which proves the thesis.

Proof of (ii) $\Longrightarrow$ (i). The proof relies on integration formula (17.186) of Theorem 17.4.2.

First step. For every $v\in V$ and every sequence $(v_n)_{n\in\mathbf{N}}$ satisfying $v_n \rightharpoonup v$ in V we establish $\Phi(v)\leq\liminf_{n\to+\infty}\Phi_n(v_n)$.

Let us connect u given by the normalization condition (NC) and v by any finite chain

$$u \smile v : u_0 := u, \ldots, u_{l-1}, u_l := v,$$

where u_k in $\operatorname{dom}(\partial\Phi)$ for $k=0,\ldots,l-1$, and consider in V^* a chain

$$u_0^* := u^*, \ldots, u_{l-1}^*$$

satisfying $(u_k, u_k^*)\in\partial\Phi$ for $k=0,\ldots,l-1$. Since $\partial\Phi = G-\lim_{n\to+\infty}\partial\Phi_n$, for each $k=0,\ldots,l-1$ there exists $(u_k^n, u_k^{n*})\in\partial\Phi_n$ such that

$$\begin{aligned}&u_k^n \to u_k \quad \text{in } V,\\ &u_k^{n*}\to u_k^* \quad \text{in } V^*\end{aligned} \tag{17.189}$$

(for $k=0$ such a condition is given by (NC)).

Let us connect u^n and v_n by the following chain, where $(u_k^n)_{k=1,\ldots,l-1}$ is defined above:

$$u^n \smile v_n : u_0^n := u^n, \ldots, u_{l-1}^n, v_n.$$

According to (17.185) we have

$$\Phi_n(v_n) \geq \Phi_n(u^n) + \sum_{k=0}^{l-1} \langle u_k^{n*}, u_{k+1}^n - u_k^n \rangle, \tag{17.190}$$

where from (17.189), $u_{k+1}^n - u_k^n \to u_{k+1} - u_k$ for $k = 0, \ldots, l-2$, and for $k = l-1$, $u_l^n - u_{l-1}^n = v_n - u_{l-1}^n \rightharpoonup v - u_{l-1}$. Going to the limit in (17.190) and using the normalization condition (NC) we obtain

$$\liminf_{n \to +\infty} \Phi_n(v_n) \geq \Phi(u) + \sum_{k=0}^{l-1} \langle u_k^*, u_{k+1} - u_k \rangle$$

for every finite chain $u \smile v$ with $u_k \in \operatorname{dom} \partial\Phi$ for $k = 0, \ldots, l-1$. Taking the supremum with respect to all such finite chains $u \smile v$, from integration formula (17.186) of Theorem 17.4.2, we infer

$$\begin{aligned} &\liminf_{n \to +\infty} \Phi_n(v_n) \\ &\quad \geq \sup \left\{ \Phi(u) + \sum_{k=0}^{l-1} \langle u_k^*, u_{k+1} - u_k \rangle : u \smile v,\ u_k \in \operatorname{dom} \partial\Phi \text{ for } k = 0, \ldots, l-1,\ l \in \mathbf{N}^* \right\} \\ &\quad = \Phi(v), \end{aligned}$$

which proves the thesis.

Second step. For every $v \in V$ we prove the existence of a sequence $(v_n)_{n \in \mathbf{N}}$ strongly converging to v in V and satisfying $\Phi(v) = \lim_{n \to +\infty} \Phi_n(v_n)$.

If $v \notin \operatorname{dom} \Phi$ there is nothing to prove since for every sequence $(v_n)_{n \in \mathbf{N}}$, inequality $\Phi(v) \geq \limsup_{n \to +\infty} \Phi_n(v_n)$ is automatically satisfied. On the other hand, according to Proposition 17.4.3, choosing a sequence $(\lambda_k)_{k \in \mathbf{N}}$ of positive numbers going to zero, for every $v \in \operatorname{dom} \Phi$ the sequence $(J_{\lambda_k} v)_{k \in \mathbf{N}}$ belongs to $\operatorname{dom} \partial\Phi$, strongly converges to v in V, and satisfies $\Phi(J_{\lambda_k} v) \to \Phi(v)$. Consequently, for proving the statement, by using the diagonalization lemma, Lemma 11.1.1, in $V \times \mathbf{R}$, it is enough to establish that for every $v \in \operatorname{dom} \partial\Phi$, there exists a sequence $(v_n)_{n \in \mathbf{N}}$ satisfying $v_n \to v$ and $\Phi(v) = \lim_{n \to +\infty} \Phi_n(v_n)$.

Given $v \in \operatorname{dom} \partial\Phi$, let us fix $w^* \in \partial\Phi(v)$. By assumption (ii) there exists an approximating sequence $(v_n, w^{n*}) \in \partial\Phi_n$ such that $v_n \to v$ in V and $w^{n*} \to w^*$ in V^*. We claim that the sequence $(v_n)_{n \in \mathbf{N}}$ satisfies the thesis. For u, u^n, and u^{n*} given by the normalization condition (NC), consider any finite chain

$$v \smile u : u_0 := v, \ldots, u_{l-1}, u_l := u,$$

the approximating chain in V

$$v_n \smile u^n : u_0^n := v_n, \ldots, u_{l-1}^n, u_l^n := u^n,$$

and the chain in V^*

$$u_0^n := w^{n*}, \ldots, u_{l-1}^{n*}, u_l^{n*} := u^{n*}$$

satisfying $(u_k^n, u_k^{n*}) \in \partial\Phi_n$ and $u_k^n \to u_k$ in V, $u_k^{n^*} \to u_k^*$ in V^* for $k = 0, \ldots, l$. From Proposition 17.4.4 we have

$$\Phi_n(u^n) \geq \Phi_n(v_n) + \sum_{k=0}^{l-1} \langle u_k^{n*}, u_{k+1}^n - u_k^n \rangle. \tag{17.191}$$

Going to the limit as n goes to $+\infty$ (recall that from (NC), $\Phi_n(u^n) \to \Phi(u)$), we infer

$$\begin{aligned}\Phi(u) &\geq \limsup_{n\to+\infty} \Phi_n(v_n) + \sum_{k=0}^{l-1} \langle u_k^*, u_{k+1} - u_k \rangle \\ &= \limsup_{n\to+\infty} \Phi_n(v_n) + \left(\Phi(v) + \sum_{k=0}^{l-1} \langle u_k^*, u_{k+1} - u_k \rangle \right) - \Phi(v)\end{aligned}$$

for any finite chain $v \smile u$ with $u_k \in \operatorname{dom} \partial\Phi$ for $k = 0, \ldots, l-1$. Taking the supremum with respect to all such finite chains $u \smile v$, from integration formula (17.186) we deduce

$$\Phi(u) \geq \limsup_{n\to+\infty} \Phi_n(v_n) + \Phi(u) - \Phi(v),$$

which, together with the first step, proves the thesis. $\square$

We assume now that $(V, ||.||)$ along with its dual $(V^*, ||.||_*)$ satisfies the following additional property $(\mathscr{R})$: the norms $||.||$ and $||.||_*$ are strictly convex, and the weak convergence and the convergence of the norms imply the strong convergence. For example, L^p-spaces $(1 < p < +\infty)$ and $W^{k,p}$-spaces $(k \in \mathbf{N},\ 1 < p < +\infty)$ satisfy $(\mathscr{R})$. Under assumption $(\mathscr{R})$, Theorem 17.4.4 may be completed by the convergence of the resolvents introduced in Definition 17.4.4.

Corollary 17.4.2. *Let $(V, ||.||)$ be a reflexive Banach space and $(\Phi_n)_{n\in\mathbf{N}}$, Φ a sequence of convex, proper, lower semicontinuous functions from V into $\mathbf{R}\cup\{+\infty\}$. Assume that $(V, ||.||)$ along with $(V^*, ||.||_*)$ satisfies $(\mathscr{R})$. Then the following statements are equivalent:*

(i) $\Phi_n \xrightarrow{M} \Phi$;

(ii) *for all $\lambda > 0$ and all $u \in V$, $J_\lambda^{\Phi_n} u \to J_\lambda^{\Phi} u$ and the normalization condition* (NC) *holds;*

(iii) $\exists \lambda_0 > 0$, *for all $u \in V$, $J_{\lambda_0}^{\Phi_n} u \to J_{\lambda_0}^{\Phi} u$ and the normalization condition* (NC) *holds;*

(iv) $\partial\Phi = G - \lim_{n\to+\infty} \partial\Phi_n$ *and the normalization condition* (NC) *holds.*

PROOF. Since (iv) $\Longrightarrow$ (i) is established in Theorem 17.4.4, it only remains to prove (i) $\Longrightarrow$ (ii) and (iii) $\Longrightarrow$ (iv).

Proof of (i) $\Longrightarrow$ (ii). We proceed in four steps. In what follows, u is a fixed element of V, and λ is any positive number in $\mathbf{R}$.

Step 1. We claim that $(J_\lambda^{\Phi_n} u)_{n\in\mathbf{N}}$ is bounded in V. Take $v_0 \in \operatorname{dom}\Phi$. Since $\Phi_n \xrightarrow{M} \Phi$, there exists a sequence $(v_n)_{n\in\mathbf{N}}$ such that $v_n \to v$ in V and $\Phi_n(v_n) \to \Phi(v)$. Assume for the moment that there exists $\alpha > 0$ such that for all $n \in \mathbf{N}$

$$\Phi_n \geq -\alpha(||.|| + 1). \tag{17.192}$$

From the definition of $J_\lambda^{\Phi_n}$ it follows that

$$\begin{aligned}\Phi_n(v_n) + \frac{1}{2\lambda}||v_n - u||^2 &\geq \Phi_n(J_\lambda^{\Phi_n} u) + \frac{1}{2\lambda}||v_n - J_\lambda^{\Phi_n} u||^2 \\ &\geq -\alpha||J_\lambda^{\Phi_n} u - v_n|| + \frac{1}{2\lambda}||J_\lambda^{\Phi_n} u - v_n||^2 - \alpha(||v_n|| + 1).\end{aligned}$$

From the boundedness of $(v_n)_{n\in\mathbf{N}}$ and $(\Phi_n(v_n))_{n\in\mathbf{N}}$, we infer that $\sup_{n\in\mathbf{N}} \|J_\lambda^{\Phi_n} u - v_n\| < +\infty$, from which we derive that $(J_\lambda^{\Phi_n} u)_{n\in\mathbf{N}}$ is bounded. Lemma 17.4.5 below states that the uniform bound (17.192) is automatically fulfilled by the sequence $(\Phi_n)_{n\in\mathbf{N}}$. For a proof see [37, Lemma 3.8].

Lemma 17.4.5. *Let $(\Phi_n)_{n\in\mathbf{N}}$ be a sequence of convex, lower semicontinuous, and uniformly proper functions from V into $\mathbf{R}\cup\{+\infty\}$. Assume that $\Gamma_w - \liminf_{n\to+\infty} \Phi_n < +\infty$. Then there exists $\alpha > 0$ such that for all $n \in \mathbf{N}$, $\Phi_n \geq -\alpha(\|.\|+1)$.*

Step 2. We establish that $J_\lambda^{\Phi_n} u \rightharpoonup J_\lambda^{\Phi} u$. From Step 1 and the reflexivity of the space $(V, \|.\|)$, there exist a subsequence of $(J_\lambda^{\Phi_n} u)_{n\in\mathbf{N}}$, not relabeled, and $w \in V$ such that $J_\lambda^{\Phi_n} u \rightharpoonup w$ in V. We claim that $w = J_\lambda^{\Phi} u$. To see this, it suffices to use the variational formulation of $J_\lambda^{\Phi_n} u$ and the Mosco-convergence of Φ_n to Φ. The thesis then follows from the variational property of the Γ-convergence, Theorem 12.1.1(i).

Step 3. We establish that $J_\lambda^{\Phi_n} u \to J_\lambda^{\Phi} u$. Since $(V, \|.\|)$ fulfills the property $(\mathscr{R})$, it is enough to show that $\|J_\lambda^{\Phi_n} u\| \to \|J_\lambda^{\Phi} u\|$. From hypothesis (ii) there exists a sequence $(v_n)_{n\in\mathbf{N}}$ satisfying $v_n \to J_\lambda^{\Phi} u$ and $\Phi_n(v_n) \to \Phi(J_\lambda^{\Phi} u)$. From

$$\Phi_n(J_\lambda^{\Phi_n} u) + \frac{1}{2\lambda}\|u - J_\lambda^{\Phi_n} u\|^2 \leq \Phi_n(v_n) + \frac{1}{2\lambda}\|u - v_n\|^2 \tag{17.193}$$

we infer

$$\begin{aligned}\limsup_{n\to+\infty} \frac{1}{2\lambda}\|u - J_\lambda^{\Phi_n} u\|^2 &\leq -\liminf_{n\to+\infty} \Phi_n(J_\lambda^{\Phi_n} u) + \Phi(J_\lambda^{\Phi} u) + \frac{1}{2\lambda}\|u - J_\lambda^{\Phi} u\|^2 \\ &\leq \frac{1}{2\lambda}\|u - J_\lambda^{\Phi} u\|^2.\end{aligned}$$

(We have used $\Phi(J_\lambda^{\Phi} u) \leq \liminf_{n\to+\infty} \Phi_n(J_\lambda^{\Phi_n} u)$ since from the previous step $J_\lambda^{\Phi_n} u \rightharpoonup J_\lambda^{\Phi} u$.) But, by lower semicontinuity

$$\frac{1}{2\lambda}\|u - J_\lambda^{\Phi} u\|^2 \leq \liminf_{n\to+\infty} \frac{1}{2\lambda}\|u - J_\lambda^{\Phi_n} u\|^2$$

so that

$$\lim_{n\to+\infty} \|u - J_\lambda^{\Phi_n} u\|^2 = \|u - J_\lambda^{\Phi} u\|^2,$$

which proves the thesis.

Step 4. We prove the normalization condition (NC). Taking $(J_\lambda^{\Phi} u, \frac{1}{\lambda} H(u - J_\lambda^{\Phi} u))$ (recall that H is one to one from V onto V^*) which from Proposition 17.4.3 belongs to $\partial\Phi$, and setting $u_n := J_\lambda^{\Phi_n} u$ and $u_n^* := \frac{1}{\lambda} H(u - u_n)$, we have $(u_n, u_n^*) \in \partial\Phi^n$ and from Step 3 $u_n \to J_\lambda^{\Phi} u$. Furthermore, since $(V^*, \|.\|_*)$ fulfills $(\mathscr{R})$, from Lemma 17.4.2, the map H is strongly continuous from V onto V^*. Hence $u_n^* \to \frac{1}{\lambda} H(u - J_\lambda^{\Phi} u)$. It remains to establish that $\Phi_n(u_n) \to \Phi(J_\lambda^{\Phi} u)$ which readily follows from (17.193). Indeed we already have $\Phi(J_\lambda^{\Phi} u) \leq \liminf_{n\to+\infty} \Phi_n(J_\lambda^{\Phi_n} u)$, and (17.193) yields $\limsup_{n\to+\infty} \Phi_n(J_\lambda^{\Phi_n} u) \leq \Phi(J_\lambda^{\Phi} u)$.

Proof of (iii) $\Longrightarrow$ (iv). To simplify the notation, we write $\lambda > 0$ instead of λ_0. Consider $(u,u^*) \in \partial\Phi$. With the notation of Lemma 17.4.2, set $v := u + \lambda H^{-1}(u^*)$, then $\frac{1}{\lambda}H(v - u) = u^* \in \partial\Phi(u)$. According to Proposition 17.4.3 we have

$$\frac{1}{\lambda}H(v - J_\lambda^\Phi v) \in \partial\Phi(J_\lambda^\Phi v)$$

so that by uniqueness $J_\lambda^\Phi v = u$. Let us set $u_n := J_\lambda^{\Phi_n} v$ and $u_n^* := \frac{1}{\lambda}H(v - J_\lambda^{\Phi_n} v)$. From Proposition 17.4.3, $(u_n, u_n^*) \in \partial\Phi_n$, and from hypothesis (iii), $u_n \to J_\lambda^\Phi v = u$ when n goes to $+\infty$. It remains to establish $\frac{1}{\lambda}H(v - J_\lambda^{\Phi_n} v) \to \frac{1}{\lambda}H(v - J_\lambda^\Phi v)$ in V^*. Since $(V, \|.\|)$ fulfills $(\mathscr{R})$, this assertion follows directly from Lemma 17.4.2, which states the strong continuity of $H : V \to V^*$. This completes the proof. □

By rephrasing Corollary 17.4.2 in the context of Hilbert spaces, we obtain the following corollary.

Corollary 17.4.3. *Let $(\Phi_n)_{n\in\mathbf{N}}$, Φ be a sequence of convex, proper, lower semicontinuous functions from a Hilbert space $(\mathscr{H}, \|.\|)$ into $\mathbf{R} \cup \{+\infty\}$. Then the following statements are equivalent:*

(i) $\Phi_n \xrightarrow{M} \Phi$;

(ii) *for all $\lambda > 0$ and all $u \in \mathscr{H}$, $\|(I + \lambda\partial\Phi_n)^{-1}u - (I + \lambda\partial\Phi)^{-1}u\| \to 0$ and the normalization condition* (NC) *holds;*

(iii) $\exists \lambda_0 > 0$, *for all $u \in \mathscr{H}$, $\|(I + \lambda_0\partial\Phi_n)^{-1}u - (I + \lambda_0\partial\Phi)^{-1}u\| \to 0$ and the normalization condition* (NC) *holds;*

(iv) $\partial\Phi = G - \lim_{n\to+\infty} \partial\Phi_n$ *and the normalization condition* (NC) *holds.*

PROOF. In this context, H is the identity map I. Therefore from Proposition 17.4.3, $J_\lambda^\Phi u$ is the unique element of $\mathscr{H}$ satisfying $\frac{1}{\lambda}(u - J_\lambda^\Phi u) \in \partial\Phi(J_\lambda^\Phi u)$ or equivalently $J_\lambda^\Phi u = (I + \lambda\partial\Phi)^{-1}u$. The same calculation holds for $J_\lambda^{\Phi_n} u$. □

17.4.3 ▪ The weak version of the approximation theorem

As noted in Chapter 12, a large number of problems arising from mechanics, physics, economics, or approximation methods in numerical analysis are modeled by means of minimization of functionals depending on some parameter, here formally denoted by n. For instance, we write these functionals Φ_n for Φ_ε when ε is a small parameter associated to a thickness, a stiffness in mechanics, or a size of small discontinuities. Furthermore, in many cases, for instance, in linear elasticity or in thermic or theoretical physics, these functionals are convex. Therefore, for the evolution equations governed by the subdifferentials $\partial\Phi_n$, the theory of approximation of semigroups of operators is powerful for questions of convergence in transient boundary value problems. Theorem 17.4.6 below is a straightforward consequence of the previous corollary and the theory of the convergence of semigroups generated by maximal monotones operators initiated by Trotter in the Hilbertian setting and generalized by many authors (Attouch, Barbu, Brézis, Benilan, Crandall, Kato, Liggett, Pazy, and many others). In the next section, we will establish a strong version by a direct method without resorting to the theory of semigroups.

Before stating the first approximation theorem, Theorem 17.4.6, we need to complete Definition 17.2.3 and Theorem 17.2.2 of Section 17.2.2 when considering nonhomogeneous Cauchy problems in finite time. Let $(\mathscr{H}, \|.\|)$ be a Hilbert space, $\Phi : \mathscr{H} \to \mathbf{R} \cup \{+\infty\}$ be a convex, proper, lower semicontinuous function, and f in $L^1(0,T;\mathscr{H})$. In what follows we look at the gradient flow equation

$$\text{(GF)} \quad \frac{du}{dt} + \partial\Phi(u) \ni f.$$

Definition 17.4.7. *We say that u is a strong solution of* (GF) *if* (i) *and* (ii) *hold:*

(i) $u \in \mathbf{C}([0,T];\mathscr{H})$ *and is absolutely continuous on* $[0,T]$*;*

(ii) *for almost all* $t \in (0,T)$, $u(t) \in \operatorname{dom}\partial\Phi$, *and* $\frac{du}{dt}(t) + \partial\Phi(u(t)) \ni f(t)$.

We say that u is a weak solution of (GF) *if there exist a sequence* $(f_n)_{n\in\mathbf{N}}$ *in* $L^1(0,T;\mathscr{H})$, $f_n \to f$ *in* $L^1(0,T;\mathscr{H})$, *and* $u_n \in \mathbf{C}([0,T];\mathscr{H})$ *such that* u_n *is a strong solution of* $\frac{du_n}{dt} + \partial\Phi(u_n) \ni f_n$ *and satisfies* $u_n \to u$ *in* $(\mathbf{C}([0,T];\mathscr{H}), \|.\|_\infty)$.

Theorem 17.4.5. *Let* $u^0 \in \overline{\operatorname{dom}\partial\Phi}$*; then there exists a unique weak solution of the Cauchy problem*

$$(\mathscr{P}) \quad \begin{cases} \dfrac{du}{dt} + \partial\Phi(u) \ni f, \\ u(0) = u^0. \end{cases}$$

PROOF. We begin by stating the following elementary lemma. For a proof we refer the reader to [135, Lemma 3.1].

Lemma 17.4.6. *Let f and g in $L^1(0,T;\mathscr{H})$ and assume that u and v are weak solutions of*

$$\frac{du}{dt} + \partial\Phi(u) \ni f \text{ and } \frac{du}{dt} + \partial\Phi(u) \ni g,$$

respectively. Then for all (s,t), $0 \le s \le t \le T$,

$$\|u(t) - v(t)\| \le \|u(s) - u(t)\| + \int_s^t \|f(\xi) - g(\xi)\|\, d\xi. \tag{17.194}$$

Uniqueness follows directly from (17.194). For proving existence, we proceed in two steps.

Step 1. We assume that f is a step function defined on the subdivision $0 = a_0 < a_1 \le \cdots a_{i-1} < a_i \cdots a_n = T$ by $f(.) = f_i$ on $[a_{i-1}, a_i[$ for $i = 1, \ldots, n$ and we prove that there exists a strong solution of $(\mathscr{P})$.

Consider $\Phi_i := \Phi - \langle f_i, .\rangle$ and on each interval $[a_{i-1}, a_i]$ define by induction for $i = 1, \ldots, n$, the Cauchy problems

$$(\mathscr{P}_i) \quad \begin{cases} \dfrac{du^i}{dt} + \partial\Phi_i(u^i) \ni 0, \\ u^i(a_{i-1}) = u^{i-1}(a_{i-1}) \end{cases}$$

with $u^0(0) := u^0$. A careful analysis of the proof of Theorem 17.2.2 shows that the hypothesis $\inf_{\mathscr{H}} \Phi > -\infty$ is not necessary to establish existence (and uniqueness) of a strong solution of Problem (17.47) as long as we restrict it to a finite interval. Therefore applying Theorem 17.2.2 to each Cauchy problem $(\mathscr{P}_i)$ in $[a_{i-1}, a_i]$, taking into account the previous remark and reasoning by induction, we infer that there exists a unique strong solution u^i of $(\mathscr{P}_i)$ in $[a_{i-1}, a_i]$. The function u defined in $[0, T]$ by $u := \sum_{i=1}^n \mathbf{1}_{[a_i, a_{i-1}[} u_i$ is clearly a strong solution of $(\mathscr{P})$.

Step 2. We consider the general case: $f \in L^1(0, T; \mathscr{H})$. There exists a sequence $(f_n)_{n\in\mathbf{N}}$ of step functions in $L^1(0, T; \mathscr{H})$ such that $f_n \to f$ in $L^1(0, T, \mathscr{H})$. Denote by u_n the strong solution of

$$\begin{cases} \dfrac{du_n}{dt} + \partial\Phi(u_n) \ni f_n, \\ u_n(0) = u^0 \end{cases}$$

obtained in Step 1. From Lemma 17.4.6 we infer

$$\|u_n(t) - u_m(t)\| \le \int_0^t \|f_n(\xi) - f_m(\xi)\|\, d\xi$$

so that $(u_n)_{n\in\mathbf{N}}$ is a Cauchy sequence in $\mathbf{C}([0, T]; \mathscr{H})$. Thus $(u_n)_{n\in\mathbf{N}}$ uniformly converges to some function u in $\mathbf{C}([0, T]; \mathscr{H})$, which, by definition, is a weak solution of $(\mathscr{P})$. □

Theorem 17.4.6 (Approximation 1). *Let $(\Phi_n)_{n\in\mathbf{N}}$, Φ be a sequence of convex, proper, lower semicontinuous functions from a Hilbert space $(\mathscr{H}, \|.\|)$ into $\mathbf{R} \cup \{+\infty\}$ such that $\overline{\operatorname{dom}\partial\Phi} \subset \overline{\operatorname{dom}\partial\Phi_n}$ for all $n \in \mathbf{N}$. Let $(f_n)_{n\in\mathbf{N}}$, f, be a sequence in $L^1(0, T; \mathscr{H})$, $u_n^0 \in \overline{\operatorname{dom}\partial\Phi_n}$, $u^0 \in \overline{\operatorname{dom}\partial\Phi}$, and consider u_n and u the unique weak solutions of the Cauchy problems*

$$(\mathscr{P}_n) \quad \begin{cases} \dfrac{du_n}{dt} + \partial\Phi_n(u_n) \ni f_n, \\ u_n(0) = u_n^0, \end{cases} \qquad (\mathscr{P}) \quad \begin{cases} \dfrac{du}{dt} + \partial\Phi(u) \ni f, \\ u(0) = u^0, \end{cases} \quad \textit{respectively.}$$

Assuming that $f_n \to f$ in $L^1(0, T; \mathscr{H})$, $u_n^0 \to u^0$ strongly in $\mathscr{H}$, and $\Phi_n \xrightarrow{M} \Phi$, then $u_n \to u$ uniformly on $[0, T]$, i.e., $u_n \to u$ in the normed space $(\mathbf{C}(0, T; \mathscr{H}), \|.\|_\infty)$.

PROOF (*sketch*).

Step 1. We assume $f_n = f = 0$. In the context of subdifferential operators and in the Hilbertian setting, rephrasing the result obtained by Brezis and Pazy (or Trotter for linear operators), we have the following relation between the convergence of the resolvents and the convergence of the corresponding semigroups.

Lemma 17.4.7. *Let $(\Phi_n)_{n\in\mathbf{N}}$, Φ be a sequence of convex, proper, lower semicontinuous functions from a Hilbert space $(\mathscr{H}, \|.\|)$ into $\mathbf{R} \cup \{+\infty\}$, and denote by $(S^{\partial\Phi_n}(t))_{t\ge 0}$ and $(S^{\partial\Phi}(t))_{t\ge 0}$ the contraction semigroups generated by $\partial\Phi_n$ and $\partial\Phi$, respectively. Assume that $\overline{\operatorname{dom}\partial\Phi} \subset \overline{\operatorname{dom}\partial\Phi_n}$ for all $n \in \mathbf{N}$ and that for all $\lambda > 0$ and all $v \in \overline{\operatorname{dom}\partial\Phi}$, $\|(I + \lambda\partial\Phi_n)^{-1}v - (I + \lambda\partial\Phi)^{-1}v\| \to 0$. Then, for every $v \in \overline{\operatorname{dom}\partial\Phi}$, $S^{\partial\Phi_n}(t)v \to S^{\partial\Phi}(t)v$ uniformly on all compact subsets of $[0, +\infty)$.*

For a proof (under less restrictive conditions) see [140, Theorems 3.1, 4.1, and Corollary 4.2] or [135, Theorem 4.2]. Recall that the contraction semigroup $(S^A(t))_{t\geq 0}$ generated by a maximal operator A is given by the "exponential formula"

$$\forall u_0 \in \overline{\operatorname{dom} A} \quad S^A(t)u_0 = \lim_{k\to+\infty} \left[\left(I + \frac{t}{k}A\right)^{-1}\right]^k u_0.$$

We conclude the step by combining Corollary 17.4.3, Lemma 17.4.7, and the standard formulas $u_n(t) = S^{\Phi_n}(t)u_n^0$, $u(t) = S^{\Phi}(t)u^0$.

Step 2: *General case.* Let g be an arbitrary step function in $L^1(0,T;\mathcal{H})$ and consider the two evolution equations

$$\begin{cases} \dfrac{dw_n}{dt} + \partial\Phi_n(w_n) \ni g, \\ w_n(0) = u_n^0, \end{cases} \qquad \begin{cases} \dfrac{dw}{dt} + \partial\Phi(w) \ni g, \\ w(0) = u^0. \end{cases}$$

The result of Step 1 applied for each interval of $[0,T]$ where g is constant and for the subdifferential operators $\partial\big(\Phi^n - \langle g, .\rangle\big)$ and $\partial\big(\Phi - \langle g, .\rangle\big)$ gives

$$||w_n(t - w(t)|| \to 0 \quad \text{uniformly on } [0,T]. \tag{17.195}$$

From (17.194) of Lemma 17.4.6, for all $0 \leq s \leq t \leq T$ one has

$$\begin{aligned} ||u_n(t) - w_n(t)|| &\leq ||u_n(s) - w_n(s)|| + \int_s^t ||f(\xi) - g(\xi)||d\xi, \\ ||u(t) - w(t)|| &\leq ||u(s) - w(s)|| + \int_s^t ||f(\xi) - g(\xi)||d\xi. \end{aligned} \tag{17.196}$$

Applying (17.196) with $s = 0$, we obtain

$$\begin{aligned} ||u_n(t) - w_n(t)|| &\leq ||f_n - g||_{L^1(0,T;\mathcal{H})}, \\ ||u(t) - w(t)|| &\leq ||f - g||_{L^1(0,T;\mathcal{H})}. \end{aligned} \tag{17.197}$$

From $||u_n(t) - u(t)|| \leq ||u_n(t) - w_n(t)|| + ||w_n(t) - w(t)|| + ||u(t) - w(t)||$ and (17.195), (17.197) we deduce

$$\limsup_{n\to+\infty} \sup_{t\in[0,T]} ||u_n(t) - u(t)|| \leq 2||f - g||_{L^1(0,T;\mathcal{H})}.$$

This completes the proof since $||f - g||_{L^1(0,T;\mathcal{H})}$ can be chosen arbitrarily small. □

17.4.4 ▪ A strong version of the approximation theorem

Under slightly less general assumptions, but not too restrictive with regard to the applications, we propose a direct proof of the previous approximation theorem. This approach has the advantage of not requiring the approximation lemma, Lemma 17.4.7; furthermore the conclusion is more accurate. The proof consists in taking into account the fact that the operators are subdifferentials.

Theorem 17.4.7 (Approximation 2). *Let $(\Phi_n)_{n\in\mathbf{N}}$, Φ be a sequence of convex, proper, lower semicontinuous functions from a Hilbert space $(\mathcal{H},||.||)$ into $\mathbf{R}\cup\{+\infty\}$, $(f_n)_{n\in\mathbf{N}}$, f a sequence in $L^2(0,T;\mathcal{H})$, $u_n^0 \in \overline{\operatorname{dom}\partial\Phi_n}$, and $u^0 \in \overline{\operatorname{dom}\partial\Phi}$. Let u_n and u be the solutions of the Cauchy problems*

$$(\mathcal{P}_n) \quad \begin{cases} \dfrac{du_n}{dt} + \partial\Phi_n(u_n) \ni f_n, \\ u_n(0) = u_n^0, \end{cases} \qquad (\mathcal{P}) \quad \begin{cases} \dfrac{du}{dt} + \partial\Phi(u) \ni f, \\ u(0) = u^0, \end{cases}$$

and assume that

(i) $\sup_{n\in\mathbf{N}} \Phi_n(u_n^0) < +\infty$;

(ii) $f_n \to f$ *strongly in* $L^2(0,T;\mathcal{H})$;

(iii) $u_n^0 \to u^0$ *strongly in* $\mathcal{H}$;

(iv) $\Phi_n \xrightarrow{M} \Phi$.

Then $u_n \to u$ in $(\mathbf{C}(0,T;\mathcal{H}),||.||_\infty)$ and $\frac{du_n}{dt} \rightharpoonup \frac{du}{dt}$ in $L^2(0,T;\mathcal{H})$. If moreover $\Phi_n(u_n^0) \to \Phi(u^0)$, then $\frac{du_n}{dt} \to \frac{du}{dt}$ in $L^2(0,T;\mathcal{H})$.

PROOF. For existence and uniqueness of the strong solutions of $(\mathcal{P}_n)$ and $(\mathcal{P})$ see Theorem 17.2.5. (See also [135, Theorem 3.6], where it is noted that, in this situation, there is no difference between strong and weak solutions.) For the sake of simplicity, we assume that Φ_n and Φ are Gâteaux differentiable so that $\partial\Phi_n$ and $\partial\Phi$ are reduced to the singletons $\{\nabla\Phi_n\}$ and $\{\nabla\Phi\}$. The proof proceeds in four steps and we follow the ideas of Attouch in [38].

Step 1. We establish

$$\sup_{n\in\mathbf{N}} \left\| \frac{du_n}{dt} \right\|_{L^2(0,T;\mathcal{H})} < +\infty; \tag{17.198}$$

$$\sup_{n\in\mathbf{N}} ||u_n||_{\mathbf{C}(0,T;\mathcal{H})} < +\infty. \tag{17.199}$$

From $(\mathcal{P}_n)$ we deduce that for a.e. $t \in (0,T)$,

$$\left\| \frac{du_n}{dt}(t) \right\|^2 + \left\langle \nabla\Phi_n(u_n(t)), \frac{du_n}{dt}(t) \right\rangle = \left\langle f_n, \frac{du_n}{dt}(t) \right\rangle.$$

By integrating this equality on $(0,T)$, we obtain

$$\int_0^T \left\| \frac{du_n}{dt}(t) \right\|^2 dt + \int_0^T \left\langle \nabla\Phi_n(u_n(t)), \frac{du_n}{dt}(t) \right\rangle dt = \int_0^T \left\langle f_n(t), \frac{du_n}{dt}(t) \right\rangle dt$$

from which we deduce (note that according to Proposition 17.2.5, for a.e. t in $(0,T)$, $\frac{d}{dt}\Phi_n(u_n(t)) = \langle \nabla\Phi_n(u_n(t)), \frac{du_n}{dt}(t) \rangle$)

$$\int_0^T \left\| \frac{du_n}{dt}(t) \right\|^2 dt = -\Phi_n(u_n(T)) + \Phi_n(u_n^0) + \int_0^T \left\langle f_n(t), \frac{du_n}{dt}(t) \right\rangle dt. \tag{17.200}$$

Since $\Phi_n \xrightarrow{M} \Phi$, from Lemma 17.4.5, there exists $\alpha > 0$ such that for all $n \in \mathbf{N}$, $\Phi_n \geq -\alpha(\|.\| + 1)$. Therefore (17.200) yields

$$\int_0^T \left\| \frac{du_n}{dt}(t) \right\|^2 dt \leq \Phi_n(u_n^0) + \alpha(\|u_n(T)\| + 1) + \|f_n\|_{L^2(0,T;\mathscr{H})} \left(\int_0^T \left\| \frac{du_n}{dt}(t) \right\|^2 dt \right)^{1/2}. \tag{17.201}$$

From

$$u_n(T) = u_n^0 + \int_0^T \frac{du_n}{dt}(t)\, dt$$

we infer

$$\|u_n(T)\| \leq \|u_n^0\| + T^{1/2} \left(\int_0^T \left\| \frac{du_n}{dt}(t) \right\|^2 dt \right)^{1/2}. \tag{17.202}$$

Combining (17.201) and (17.202) we finally obtain

$$\int_0^T \left\| \frac{du_n}{dt}(t) \right\|^2 dt \leq \Phi_n(u_n^0) + \alpha(\|u_n^0\| + 1) + \left(\alpha T^{1/2} + \|f_n\|_{L^2(0,T;\mathscr{H})} \right) \left(\int_0^T \left\| \frac{du_n}{dt}(t) \right\|^2 dt \right)^{1/2}$$

and estimate (17.198) follows from the equiboundedness of $(u_n^0)_{n \in \mathbf{N}}$ in $\mathscr{H}$, that of $(f_n)_{n \in \mathbf{N}}$ in $L^2(0,T;\mathscr{H})$, and the equiboundedness of $(\Phi_n(u_n^0))_{n \in \mathbf{N}}$. Estimate (17.199) follows from (17.198) and from

$$\|u_n(t)\| \leq \|u_n^0\| + T^{1/2} \left(\int_0^T \left\| \frac{du_n}{dt}(t) \right\|^2 dt \right)^{1/2} \quad \forall t \in (0,T).$$

Step 2. *Compactness.* We prove that there exists $u \in L^2(0,T;\mathscr{H})$ and a subsequence of $(u_n)_{n \in \mathbf{N}}$ (not relabeled) such that

$$\frac{du_n}{dt} \rightharpoonup \frac{du}{dt} \quad \text{in } L^2(0,T;\mathscr{H}); \tag{17.203}$$

$$u_n \rightharpoonup u \quad \text{in } L^2(0,T;\mathscr{H}). \tag{17.204}$$

Since from (17.198), $(\frac{du_n}{dt})_{n \in \mathbf{N}}$ is bounded in $L^2(0,T;\mathscr{H})$, there exist a subsequence (not relabeled) and $g \in L^2(0,T;\mathscr{H})$ such that

$$\frac{du_n}{dt} \rightharpoonup g \quad \text{in } L^2(0,T;\mathscr{H}).$$

For all $t \in [0,T]$ set

$$u(t) := u^0 + \int_0^t g(s)\, ds.$$

Thus $\frac{du}{dt} = g$ and $u(0) = u^0$. On the other hand, v_n defined by

$$v_n(t) := u^0 + \int_0^t \frac{du_n}{dt}(s)\, ds$$

is bounded in $L^2(0,T;\mathscr{H})$ (this is a consequence of (17.198)). Thus for a further subsequence that we do not relabel, there exists $v \in L^2(0,T;\mathscr{H})$ such that

$$v_n \rightharpoonup v \quad \text{in } L^2(0,T;\mathscr{H}).$$

Consider the map $L : L^2(0,T;\mathscr{H}) \to L^2(0,T;\mathscr{H})$ defined by

$$L(h)(t) = u^0 + \int_0^t h(s)\,ds.$$

Clearly L is a (strongly) continuous affine function so that its graph is strongly closed, thus weakly closed in $L^2(0,T;\mathscr{H}) \times L^2(0,T;\mathscr{H})$. Therefore, from

$$\left(\frac{du_n}{dt}, v_n = L\left(\frac{du_n}{dt}\right)\right) \rightharpoonup (g, v) \quad \text{in } L^2(0,T;\mathscr{H}) \times L^2(0,T;\mathscr{H})$$

we deduce that $v = L(g)$, i.e., $v = u$. To sum up we have proved for a subsequence that

$$\frac{du_n}{dt} \rightharpoonup \frac{du}{dt} \quad \text{in } L^2(0,T;\mathscr{H});$$

$$u_n = v_n - u^0 + u_n^0 \rightharpoonup u \quad \text{in } L^2(0,T;\mathscr{H}).$$

Step 3. We prove that u is the solution of $(\mathscr{P})$. We will need the following lemma.

Lemma 17.4.8. *Let $(\psi_n)_{n\in\mathbf{N}}$, ψ be a sequence of convex, proper, lower semicontinuous functions from a separable Hilbert space $(\mathscr{H}, \|.\|)$ into $\mathbf{R}\cup\{+\infty\}$ such that $\psi_n \xrightarrow{M} \psi$ and consider $(\Psi_n)_{n\in\mathbf{N}}$, Ψ from $L^2(0,T;\mathscr{H}) \to \mathbf{R}\cup\{+\infty\}$ defined by*

$$\Psi_n(v) := \int_0^T \psi_n(v(t))\,dt; \qquad \Psi(v) := \int_0^T \psi(v(t))\,dt.$$

Then $\Psi_n \xrightarrow{M} \Psi$.

For a proof, see [38, Corollary 1.17]. Note that since from Lemma 17.4.5, $\psi_n + \alpha(\|.\|+1)$ and $\psi + \alpha(\|.\|+1)$ are nonnegative, the integrals entering the definition of Ψ_n and Ψ are well defined.

According to the Fenchel extremality condition $(\mathscr{P}_n)$ is equivalent to

$$\Phi_n(u_n(t)) + \Phi_n^*\left(f_n(t) - \frac{du_n}{dt}(t)\right) + \left\langle \frac{du_n}{dt}(t) - f_n(t), u_n(t)\right\rangle = 0$$

for a.e. $t \in (0,T)$, which is also equivalent to

$$\int_0^T \left[\Phi_n(u_n(t)) + \Phi_n^*\left(f_n(t) - \frac{du_n}{dt}(t)\right) + \left\langle \frac{du_n}{dt}(t) - f_n(t), u_n(t)\right\rangle\right] dt = 0.$$

The above equivalence is due to the Fenchel inequality which asserts that the inequality $\Phi_n(u_n(t)) + \Phi_n^*\big(f_n(t) - \frac{du_n}{dt}(t)\big) + \langle \frac{du_n}{dt}(t) - f_n(t), u_n(t)\rangle \geq 0$ for a.e. $t \in (0,T)$ is always true. Therefore, $(\mathscr{P}_n)$ is equivalent to

$$\int_0^T \left[\Phi_n(u_n(t)) + \Phi^*\left(f_n(t) - \frac{du_n}{dt}(t)\right) + \frac{d}{dt}\frac{1}{2}\|u_n(t)\|^2 - \langle f_n(t), u_n(t)\rangle\right] dt = 0,$$

or, equivalently,

$$\int_0^T \left[\Phi_n(u_n(t)) + \Phi_n^*\left(f_n(t) - \frac{du_n}{dt}(t)\right)\right] dt + \frac{1}{2}\left(\|u_n(T)\|^2 - \|u_n^0\|^2\right) - \int_0^T \langle f_n(t), u_n(t)\rangle\,dt = 0. \tag{17.205}$$

From

$$u_n(T) = u_n^0 + \int_0^T \frac{du_n}{dt}(t)\,dt$$

and (17.203), we infer that $u_n(T) \rightharpoonup u(T)$ in $\mathcal{H}$. Going to the limit in (17.205), from (17.203), (17.204), the strong convergence of f_n to f in $L^2(0,T;\mathcal{H})$ and Lemma 17.4.8, the lower semicontinuity of $v \mapsto \|v\|$ in $\mathcal{H}$, and the strong convergence u_n^0 to u_0 in $\mathcal{H}$, we obtain

$$\int_0^T \left[\Phi(u(t)) + \Phi^*\left(f(t) - \frac{du}{dt}(t)\right)\right] dt + \frac{1}{2}\|u(T)\|^2 - \frac{1}{2}\|u^0\|^2 - \int_0^T \langle f(t), u(t)\rangle\, dt \le 0.$$

Equivalently

$$\int_0^T \left[\Phi(u(t)) + \Phi^*\left(f(t) - \frac{du}{dt}(t)\right) + \left\langle \frac{du}{dt}(t) - f(t), u(t)\right\rangle\right] dt \le 0.$$

Noticing that, according to the Fenchel inequality, $\Phi(u(t)) + \Phi^*\big(f(t) - \frac{du}{dt}(t)\big) + \langle \frac{du}{dt}(t) - f(t), u(t)\rangle \ge 0$, we deduce that for a.e. $t \in (0,T)$, $\Phi(u(t)) + \Phi^*\big(f(t) - \frac{du}{dt}(t)\big) + \langle \frac{du}{dt}(t) - f(t), u(t)\rangle = 0$. Thus

$$\begin{cases} \dfrac{du}{dt} + \nabla\Phi(u) = f, \\ u(0) = u^0. \end{cases}$$

Note that according to the uniqueness of the solution of $(\mathcal{P})$, the whole sequence $(u_n)_{n\in\mathbf{N}}$ satisfies (17.203) and (17.204).

Step 4. We establish

$$u_n \to u \quad \text{in } \mathbf{C}(0,T;\mathcal{H}); \tag{17.206}$$

$$\frac{du_n}{dt} \to \frac{du}{dt} \quad \text{in } L^2(0,T;\mathcal{H}) \text{ under the hypothesis } \Phi_n(u_n^0) \to \Phi(u^0). \tag{17.207}$$

To prove (17.206), we will apply the Ascoli compactness theorem in $(\mathbf{C}(0,T;\mathcal{H}), \|.\|_\infty)$ to the sequence $(u_n)_{n\in\mathbf{N}}$. From (17.199) we already have $\sup_{n\in\mathbf{N}} \|u_n\|_{\mathbf{C}(0,T;\mathcal{H})} < +\infty$. The uniform equicontinuity of the sequence $(u_n)_{n\in\mathbf{N}}$ is a straightforward consequence of (17.198) and

$$\begin{aligned} \|u_n(t) - u_n(s)\| &\le \int_s^t \left\|\frac{du_n}{dt}(\xi)\right\| d\xi \\ &\le |t-s|^{1/2} \left\|\frac{du_n}{dt}\right\|_{L^2(0,T;\mathcal{H})}. \end{aligned}$$

It remains to establish that $u_n(t) \to u(t)$ in $\mathcal{H}$ for every $t \in [0,T]$. Noticing that (17.203) holds in $L^2(0,T';\mathcal{H})$ for all $0 \le T' \le T$, from

$$u_n(T') = u_n^0 + \int_0^{T'} \frac{du_n}{dt}(t)\,dt$$

we infer that $u_n(t) \rightharpoonup u(t)$ in $\mathcal{H}$ for all $t \in [0,T]$. In order to prove the strong convergence $u_n(t) \to u(t)$, we are going to establish that $\|u_n(t)\| \to \|u(t)\|$. To see that, the idea consists in extracting the maximum information from (17.205). For each term of (17.205) we have indeed obtained

$$a := \int_0^T \Phi(u(t))\,dt \le \liminf_{n\to+\infty} \int_0^T \Phi(u_n(t))\,dt;$$

$$b := \int_0^T \Phi^*\left(f(t) - \frac{du}{dt}(t)\right) dt \le \liminf_{n\to+\infty} \int_0^T \Phi^*\left(f_n(t) - \frac{du_n}{dt}(t)\right) dt;$$

$$c := \frac{1}{2}\|u(T)\|^2 - \frac{1}{2}\|u^0\|^2 \le \liminf_{n\to+\infty} \frac{1}{2}\|u_n(T)\|^2 - \frac{1}{2}\|u^0\|^2;$$

$$d := -\int_0^T \langle f(t), u(t)\rangle = \lim_{n\to+\infty} -\int_0^T \langle f_n(t), u_n(t)\rangle$$

with $a+b+c+d=0$. Therefore, denoting by a_n, b_n, c_n, and d_n each of the four terms of (17.205) we have obtained

$$\begin{aligned}
&a \le \liminf_{n\to+\infty} a_n;\\
&b \le \liminf_{n\to+\infty} b_n;\\
&c \le \liminf_{n\to+\infty} c_n;\\
&d = \lim_{n\to+\infty} d_n;\\
&a+b+c+d = a_n + b_n + c_n + d_n = 0,
\end{aligned}$$

from which we easily infer, using Lemma 17.2.1, that $a = \lim_{n\to+\infty} a_n$, $b = \lim_{n\to+\infty} b_n$, and $c = \lim_{n\to+\infty} c_b$. In particular $\|u_n(T)\| \to \|u(T)\|$. This being true for each $0 \le T' \le T$ (by reasoning on on $L^2(0,T';\mathcal{H})$), we infer that $\|u_n(t)\| \to \|u(t)\|$ for all $t \in [0,T]$.

To prove (17.207), it suffices to establish

$$\int_0^T \left\|\frac{du_n}{dt}(t)\right\|^2 dt \to \int_0^T \left\|\frac{du}{dt}(t)\right\|^2 dt,$$

which follows directly from (17.200). Indeed going to the limit on

$$\int_0^T \left\|\frac{du_n}{dt}(t)\right\|^2 dt = -\Phi_n(u_n(T)) + \Phi_n(u_n^0) + \int_0^T \left\langle f_n(t), \frac{du_n}{dt}(t)\right\rangle dt$$

we deduce, since $\Phi_n \xrightarrow{M} \Phi$,

$$\begin{aligned}
\limsup_{n\to+\infty} \int_0^T \left\|\frac{du_n}{dt}(t)\right\|^2 dt &= -\liminf_{n\to+\infty} \Phi_n(u_n(T)) + \Phi(u^0) + \int_0^T \left\langle f(t), \frac{du}{dt}(t)\right\rangle dt\\
&\le -\Phi(u(T)) + \Phi(u^0) + \int_0^T \left\langle f(t), \frac{du}{dt}(t)\right\rangle dt\\
&= \int_0^T \left\|\frac{du}{dt}(t)\right\|^2 dt
\end{aligned}$$

and then the conclusion follows from the lower semicontinuity of the norm in $L^2(0,T,\mathcal{H})$. □

17.4.5 ▪ Application to diffusion in random media

We go back to the notation and definitions of Section 12.4 in the specific case $m = 1$, $p = 2$. We write ε to denote a sequence $(\varepsilon_n)_{n\in\mathbf{N}}$ of positive numbers ε_n going to zero when $n \to +\infty$, and we briefly write $\varepsilon \to 0$ instead of $\lim_{n\to+\infty} \varepsilon_n = 0$.

Given $\alpha > 0$ and $\beta > 0$, we denote by $\mathrm{Conv}_{\alpha,\beta}$ the class of functions $g : \mathbf{R}^N \times \mathbf{R}^N \to \mathbf{R}$, $(x,\xi) \mapsto g(x,\xi)$, measurable in x, convex with respect to ξ, and satisfying the growth condition (12.5), i.e., $\alpha|\xi|^2 \le g(x,\xi) \le \beta(1+|\xi|^2)$ for all $(x,\xi) \in \mathbf{R}^N \times \mathbf{R}^N$. Using the subdifferential inequality together with (12.5), it is easy to prove that the functions g automatically satisfy the local Lipschitz condition (12.6) for some $L \ge 0$ depending only on α and β. Therefore, with the notation of Section 12.4.3, $\mathrm{Conv}_{\alpha,\beta} \subset \mathscr{I}_{\alpha,\beta,L}$. In what follows, $\mathrm{Conv}_{\alpha,\beta}$ is endowed with the trace σ-algebra which equips $\mathscr{I}_{\alpha,\beta,L}$.

Let $(\Sigma, \mathscr{A}, \mathbf{P})$ be a probability space. In this section, we are given a random convex integrand $f : \Sigma \times \mathbf{R}^N \times \mathbf{R}^N \to \mathbf{R}$, i.e., a $\big(\mathscr{A} \otimes \mathscr{B}(\mathbf{R}^N) \otimes \mathscr{B}(\mathbf{R}^N), \mathscr{B}(\mathbf{R})\big)$ measurable function such that for every $\omega \in \Sigma$, the function $f(\omega,.,.)$ belongs to the class $\mathrm{Conv}_{\alpha,\beta}$. It is noted that since $\mathrm{Conv}_{\alpha,\beta}$ is equipped with the trace σ-algebra of the one that equips $\mathscr{I}_{\alpha,\beta,L}$, all the results of Section 12.4 remain valid when replacing the class $\mathscr{I}_{\alpha,\beta,L}$ by the class $\mathrm{Conv}_{\alpha,\beta}$.

Given the group $(T_z)_{z\in\mathbf{Z}^N}$, defined for all g in $\mathrm{Conv}_{\alpha,\beta}$ by $T_z g(x,.) = g(x+z,.)$, in the context of the discrete dynamical system $(\Sigma, \mathscr{A}, \mathbf{P}, (T_z)_{z\in\mathbf{Z}^N})$, we assume that f is periodic in law, i.e., that the law $f\#\mathbf{P}$ of f is invariant with respect to the group $(T_z)_{z\in\mathbf{Z}^N}$. (See Section 12.4 for precise definitions and examples.)

By combining the abstract results of the previous section with Section 12.4 we intend to analyze the asymptotic behavior in $C(0,T;H_0^1(\Omega))$ of the solution $u_\varepsilon(\omega)$ of the random Cauchy problem when $\varepsilon \to 0$:

$$\begin{cases} \dfrac{du_\varepsilon(\omega)}{dt} + A_\varepsilon(\omega)(u_\varepsilon(\omega)) \ni g_\varepsilon(\omega), \\ u_\varepsilon(\omega,0) = u_\varepsilon^0(\omega), \end{cases} \tag{17.208}$$

where the random operator $A_\varepsilon(\omega) : L^2(\Omega) \to 2^{L^2(\Omega)}$ is defined for every $\omega \in \Sigma$ by

$$\mathrm{dom} A_\varepsilon(\omega) = \left\{ v \in H_0^1(\Omega) : \exists \sigma \in \partial_\xi f\left(\omega, \frac{\cdot}{\varepsilon}, \nabla v\right),\ \mathrm{div}\,\sigma \in L^2(\Omega) \right\}$$

and, for all $v \in \mathrm{dom} A_\varepsilon(\omega)$,

$$A_\varepsilon(\omega) v = -\,\mathrm{div}\,\partial_\xi f\left(\omega, \frac{\cdot}{\varepsilon}, \nabla v\right).$$

We assume that $\omega \mapsto u_\varepsilon^0(\omega)$ and $\omega \mapsto g_\varepsilon(\omega)$ are two $\big(\mathscr{A}, \mathscr{B}(L^2(\Omega))\big)$ and

$$\big(\mathscr{A}, \mathscr{B}(L^2(0,T;L^2(\Omega)))\big)$$

measurable functions respectively, and that $u_\varepsilon^0(\omega) \in \overline{\mathrm{dom} A_\varepsilon(\omega)}$. To shorten the notation we will write the evolution problem (17.208) as follows:

$$(\mathscr{P}_\varepsilon(\omega)) \quad \begin{cases} \dfrac{du_\varepsilon(\omega)}{dt} - \mathrm{div}\,\partial_\xi f\left(\omega, \dfrac{\cdot}{\varepsilon}, \nabla u_\varepsilon\right) \ni g_\varepsilon(\omega), \\ u_\varepsilon(\omega,0) = u_\varepsilon^0(\omega). \end{cases}$$

It is easy to check that $A_\varepsilon(\omega)$ is nothing but the subdifferential of the random functional $F_\varepsilon(\omega,.)$ considered in Section 12.4.5 and defined by

$$F_\varepsilon : \Omega \times L^2(\Omega) \longrightarrow \mathbf{R}^+ \cup \{+\infty\},$$

$$F_\varepsilon(\omega,u) = \begin{cases} \displaystyle\int_\Omega f\left(\omega, \frac{x}{\varepsilon}, \nabla u\right) dx & \text{if } u \in W_0^{1,2}(\Omega), \\ +\infty & \text{otherwise.} \end{cases}$$

The functional $F_\varepsilon(\omega,.)$ models a random energy concerning various steady state situations and, with the notation of Section 12.3.2, the equilibrium configuration is given by the field $\overline{u}_\varepsilon$ solution of the random problem

$$\inf\left\{ F_\varepsilon(\omega,u) - \int_\Omega L(u) : u \in L^2(\Omega) \right\},$$

where the small parameter ε accounts for the size of small and randomly distributed heterogeneities. The Cauchy problem $(\mathscr{P}_\varepsilon(\omega))$ then models the corresponding diffusion. From Theorem 12.4.7 for $\mathbf{P}$-almost all ω in Σ the sequence of functional $(F_\varepsilon(\omega,.))_{\varepsilon>0}$ Γ-converges to the random integral functional $F^{hom}(\omega,)$ defined in $L^2(\Omega)$ by

$$F^{hom}(\omega,u) = \begin{cases} \displaystyle\int_\Omega f^{hom}(\omega, \nabla u)\, dx & \text{if } u \in W_0^{1,2}(\Omega), \\ +\infty & \text{otherwise,} \end{cases}$$

when $L^2(\Omega)$ is equipped with the norm topology. Recall that, from Proposition 12.4.3, the density f^{hom} is given, for $\mathbf{P}$-a.e. $\omega \in \Sigma$, by

$$\begin{aligned} f^{hom}(\omega,a) &= \lim_{n\to+\infty} \inf\left\{ \frac{1}{n^N} \int_{nY} f(\omega,y,a+\nabla u(y))\, dy : u \in W_0^{1,2}(Y) \right\} \\ &= \inf_{n\in\mathbf{N}^*} \mathbf{E}^{\mathscr{F}} \inf\left\{ \frac{1}{n^N} \int_{nY} f(\omega,y,a+\nabla u(y))\, dy : u \in W_0^{1,2}(Y) \right\}, \end{aligned}$$

where $\mathbf{E}^{\mathscr{F}}$ denotes the conditional expectation with respect to the σ-algebra of invariant sets of $\mathscr{A}$ by the group $(T_z)_{z\in\mathbf{Z}^N}$. If f is ergodic, then f^{hom} is deterministic and given for $\mathbf{P}$-a.e. $\omega \in \Sigma$ by

$$\begin{aligned} f^{hom}(a) &= \lim_{n\to+\infty} \inf\left\{ \frac{1}{n^N} \int_{nY} f(\omega,y,a+\nabla u(y))\, dy : u \in W_0^{1,2}(Y) \right\} \\ &= \inf_{n\in\mathbf{N}^*} \mathbf{E} \inf\left\{ \frac{1}{n^N} \int_{nY} f(\omega,y,a+\nabla u(y))\, dy : u \in W_0^{1,2}(Y) \right\}. \end{aligned}$$

Given $\omega \mapsto u^0(\omega)$ and $\omega \mapsto g(\omega)$ two $\left(\mathscr{A}, \mathscr{B}(L^2(\Omega))\right)$ and $\left(\mathscr{A}, \mathscr{B}(L^2(0,T;L^2(\Omega)))\right)$ measurable functions, respectively, and assuming that $u^0(\omega) \in \overline{\mathrm{dom} A(\omega)}$, we consider the Cauchy problem

$$(\mathscr{P}^{hom}(\omega)) \quad \begin{cases} \dfrac{du(\omega)}{dt} + A(\omega)(u) \ni g(\omega), \\ u(\omega,0) = u^0(\omega), \end{cases} \tag{17.209}$$

where the random operator $A(\omega): L^2(\Omega) \to 2^{L^2(\Omega)}$ is defined for every $\omega \in \Sigma$ by

$$\mathrm{dom} A(\omega) = \{v \in H_0^1(\Omega) : \exists \sigma \in \partial f^{hom}(\omega, \nabla v),\ \ \mathrm{div}\, \sigma \in L^2(\Omega)\}$$

and, for all $v \in \mathrm{dom} A(\omega)$,

$$A(\omega)v = -\,\mathrm{div}\, \partial f^{hom}(\omega, \nabla v).$$

It is easily seen that $A(\omega)$ is the subdifferential of the random functional $F^{hom}(\omega, .)$. To shorten the notation we write the evolution problem (17.208) as follows:

$$(\mathscr{P}^{hom}(\omega)) \quad \begin{cases} \dfrac{du(\omega)}{dt} - \mathrm{div}\, \partial f^{hom}(\omega, \nabla u_\varepsilon) \ni g(\omega), \\ u(\omega, 0) = u^0(\omega). \end{cases}$$

The next theorem states that for $\mathbf{P}$-a.e. $\omega \in \Sigma$, $(\mathscr{P}^{hom}(\omega))$ is the limit evolution problem of $(\mathscr{P}_\varepsilon(\omega))$ in the sense of Theorem 17.4.7. It is referred to as the homogenized evolution problem. Proposition 17.4.6 expresses the limit operator $A(\omega) = -\,\mathrm{div}\, \partial f^{hom}(\omega, \nabla .)$ in terms of almost sure graph limit.

Theorem 17.4.8. *Assume that for* $\mathbf{P}$*-a.e.* $\omega \in \Sigma$, $u_\varepsilon^0(\omega)$ *strongly converges to* $u^0(\omega)$ *in* $L^2(\Omega)$ *and that* $g_\varepsilon(\omega)$ *strongly converges to* $g(\omega)$ *in* $L^2([0,T]; L^2(\Omega))$*. Suppose further that the unique solution* $u_\varepsilon(\omega)$ *of* $(\mathscr{P}_\varepsilon(\omega))$ *satisfies* $\sup_{\varepsilon>0} F_\varepsilon(u_\varepsilon^0(\omega)) < +\infty$ *for* $\mathbf{P}$*-a.e.* $\omega \in \Sigma$*. Then for* $\mathbf{P}$*-a.e.* $\omega \in \Sigma$, $u_\varepsilon(\omega)$ *uniformly converges in* $\mathbf{C}(0,T; L^2(\Omega))$ *to the unique solution* $u(\omega)$ *of the evolution problem*

$$(\mathscr{P}^{hom}(\omega)) \quad \begin{cases} \dfrac{du(\omega)}{dt} - \mathrm{div}\, \partial f^{hom}(\omega, \nabla u) \ni g(\omega), \\ u(\omega, 0) = u^0(\omega). \end{cases} \tag{17.210}$$

Furthermore, for $\mathbf{P}$*-a.e.* $\omega \in \Sigma$,

$$\frac{du_\varepsilon(\omega)}{dt} \rightharpoonup \frac{du(\omega)}{dt} \quad \text{in } L^2(0,T; L^2(\Omega))$$

and, if $F_\varepsilon(\omega, u_\varepsilon^0) \to F^{hom}(\omega, u^0)$*, then*

$$\frac{du_\varepsilon(\omega)}{dt} \to \frac{du(\omega)}{dt} \quad \text{in } L^2(0,T; L^2(\Omega)).$$

If f *is ergodic, then* f^{hom} *is deterministic. If further* u^0 *and* g *are deterministic, then the homogenized evolution problem is deterministic and given by*

$$(\mathscr{P}^{hom}) \quad \begin{cases} \dfrac{du}{dt} - \mathrm{div}\, \partial f^{hom}(\nabla u) \ni g, \\ u(0) = u^0. \end{cases} \tag{17.211}$$

PROOF. We claim that condition (iv) of Theorem 17.4.7 is satisfied. Indeed, from Theorem 12.4.7, for $\mathbf{P}$-almost all ω in Σ, the sequence of functional $(F_\varepsilon(\omega, .))_{\varepsilon>0}$ Γ-converges to the random integral functional $F^{hom}(\omega,.)$ when $L^2(\Omega)$ is equipped with the norm topology. From the lower bound condition $f(\omega,x,\xi)\geq\alpha|\xi|^2$ we deduce that every sequence $(u_\varepsilon)_{\varepsilon>0}$ of bounded energy, i.e., satisfying $\sup_{\varepsilon>0}F_\varepsilon(\omega,u_\varepsilon)<+\infty$, which weakly converges to some u in $L^2(\Omega)$, weakly converges to u in $H^1_0(\Omega)$, then strongly converges to u in $L^2(\Omega)$. Therefore the sequence $(F_\varepsilon(\omega,.))_{\varepsilon>0}$ Mosco-converges to the random integral functional $F^{hom}(\omega,.)$ and the claim then follows. Since the three other conditions are fulfilled, the conclusion follows from Theorem 17.4.7. □

For every $a\in\mathbf{R}^N$, every $n\in\mathbf{N}^*$, and all $\omega\in\Sigma$, let us set

$$f_n(\omega,a):=\inf\left\{\frac{1}{n^N}\int_{nY}f(\omega,y,a+\nabla u(y))\,dy : u\in W_0^{1,2}(Y)\right\}\tag{17.212}$$

and consider the function $f_n(\omega,.):\mathbf{R}^N\to\mathbf{R}$, $a\mapsto f_n(\omega,a)$. From Proposition 12.4.3, for all fixed $a\in\mathbf{R}^N$, $\lim_{n\to+\infty}f_n(\omega,a)=f^{hom}(\omega,a)$ for $\mathbf{P}$-a.e. $\omega\in\Sigma$. In the proposition below, we establish that $f^{hom}(\omega,.)$ is a Mosco limit so that we can express ∂f^{hom} as a graph limit.

Proposition 17.4.6. *Under the assumptions of Theorem* 17.4.8*, the following assertions hold for* $\mathbf{P}$*-a.e.* $\omega\in\Sigma$*:*

(i) $f_n(\omega,.)\xrightarrow{M}f^{hom}(\omega,.)$.

(ii) $\partial f_n(\omega,.)\xrightarrow{G}\partial f^{hom}(\omega,.)$, *where* $\partial f_n(\omega,.)$ *is characterized by*

$$\partial f_n(\omega,a)=\left\{\frac{1}{n^N}\int_{nY}\sigma\,dy:\operatorname{div}\sigma=0,\ \sigma(y)\in\partial_\xi f(\omega,y,q+a)\right.$$

a.e. in

$$\left. nY,\ q\in\nabla H^1_0(nY)\right\}.$$

(iii) *Assume that for a.e.* $x\in\mathbf{R}^N$, $f(\omega,x,.)$ *is strictly convex and Gâteaux differentiable and that its Fenchel conjugate is such that* $\langle\xi_1^*-\xi_2^*,\xi^1-\xi^2\rangle\geq\gamma|\xi_1-\xi_2|^2$ *for some* $\gamma>0$ *and for all* $(\xi_1,\xi_2)\in\mathbf{R}^N\times\mathbf{R}^N$ *and all* $(\xi_1^*,\xi_2^*)\in\partial_\xi f^*(\omega,x,\xi_1)\times\partial_\xi f^*(\omega,x,\xi_2)$. *Then* $f_n(\omega,.)$ *and* $f^{hom}(\omega,.)$ *are Gâteaux differentiable and for all* $a\in\mathbf{R}^N$,

$$\nabla f^{hom}(\omega,a)=\lim_{n\to+\infty}\nabla f_n(\omega,a)\quad\textit{for } \mathbf{P}\textit{-a.e. } \omega\in\Sigma,$$

where

$$\nabla f_n(\omega,a)=\frac{1}{n^N}\int_{nY}\nabla_\xi f(\omega,y,\nabla u_{a,n}(\omega)(y)+a)\,dy,$$

and $u_{a,n}(\omega)$ *is the unique solution of the random Dirichlet problem*

$$\begin{cases}\operatorname{div}\left(\nabla_\xi f(\omega,.,a+\nabla v(.))\right)=0 & \textit{a.e. in } nY;\\ v=0 \quad \textit{on } \partial nY.\end{cases}$$

PROOF. In the proof, we fix ω in a set of full probability, for which the conclusions of Theorem 17.4.8 hold. It is easy to prove that $f_n(\omega,.)$, $f^{hom}(\omega,.):\mathbf{R}^N\to\mathbf{R}$ are convex and satisfy the growth conditions fulfilled by f, i.e.,

$$\alpha|.|^2\leq f^n(\omega,.),\qquad f^{hom}(\omega,.)\leq\beta(1+|.|^2).$$

Consequently they satisfy the local equi-Lipschitz condition (12.6). This implies that $f_n(\omega,.)$ almost surely Gamma-converges to $f^{hom}(\omega,.)$. (This is the general feature of sequences of equilower semicontinuous functionals.) Indeed let (a_n) be a sequence in $\mathbf{R}^N$ converging to a. From (12.6) we infer

$$f_n(\omega,a_n) \geq f_n(\omega,a) - L|a_n - a|(1+|a_n|+|a|),$$

from which we deduce, since $f_n(\omega,.)$ almost surely converges to $f^{hom}(\omega,.)$, that

$$\liminf_{n\to+\infty} f_n(\omega,a_n) \geq f^{hom}(\omega,a).$$

The upper bound in the definition of the Gamma-convergence is trivially satisfied by taking the constant sequence $(a)_{n\in\mathbf{N}}$ as a recovery sequence. Assertion (i) follows since the Gamma-convergence and the Mosco-convergence coincide in finite dimensional spaces.

Assertion (ii) follows from Theorem 17.4.4. To express the subdifferential $\partial f_n(\omega,.)$, it suffices to remark that $f_n(\omega,.)$ is the epigraphical sum defined in $\mathbf{R}^N$ by

$$f_n(\omega,a) = G\#\delta_K(j(a)),$$

where

$$G(v) = \frac{1}{n^N}\int_{nY} f(\omega,y,v)\,dy \quad \forall v \in L^2(nY,\mathbf{R}^N),$$

δ_K is the indicator function of $K = \{v \in L^2(nY,\mathbf{R}^N) : \exists u \in H_0^1(nY),\ v = \nabla u\}$, and j is the canonical embedding from $\mathbf{R}^N$ to $L^2(nY,\mathbf{R}^N)$, then to use standard subdifferential calculus rules in convex analysis. More precisely

$$\partial G\#\delta_K(j(a)) = j^T \circ \bigcup_{v\in L^2(nY,\mathbf{R}^N)} \Big(\partial G(v+j(a)) \cap \partial\delta_K(-v)\Big),$$

where j^T is the transposed operator of the embedding j given by

$$j^T(v) = \int_{nY} v\,dy.$$

Furthermore $v^* \in \partial\delta_K(v) \iff \langle v^*,\nabla\varphi\rangle_{L^2(nY,\mathbf{R}^N)} = 0$ for all $\varphi \in H_0^1(nY)$, i.e., $\operatorname{div} v^* = 0$ a.e. in nY.

Let us prove (iii). The fact that $f_n(\omega,.)$ is Gâteaux differentiable comes from the formula of its subdifferential operator expressed in (ii). Indeed, under the hypotheses of (iii), $\partial f_n(\omega,a)$ is reduced to

$$\frac{1}{n^N}\int_{nY} \nabla_\xi f(\omega,y,\nabla u_a(y)+a)\,dy,$$

where u_a is the unique minimizer in $H_0^1(nY)$ of (17.212), then satisfies the random Dirichlet problem

$$\begin{cases} \operatorname{div}\big(\nabla_\xi f(\omega,.,a+\nabla u_{a,n}(.))\big) = 0 & \text{a.e. in } nY; \\ u_{a,n} = 0 \quad \text{on } \partial nY. \end{cases}$$

In order to simplify the notation, we do not indicate the dependence on ω and n for the minimizer u_a.

Let $(a,a^*) \in \partial f^{hom}(\omega,.)$. Since $\nabla f_n(\omega,.) \overset{G}{\to} \partial f^{hom}(\omega,.)$, there exists $a_n \in \mathbf{R}^N$ such that $a_n \to a$ and $\nabla f_n(\omega,a_n) \to a^*$. We first claim that

$$|\nabla f_n(\omega,a_n) - \nabla f_n(\omega,a)| \leq \frac{1}{\gamma}|a_n - a|. \tag{17.213}$$

Indeed from Jensen's inequality we have

$$|\nabla f_n(\omega,a_n) - \nabla f_n(\omega,a)|^2 \leq \frac{1}{n^N}\int_{nY} |\nabla_\xi f(\omega,y,\nabla u_{a_n}(y)+a_n) - \nabla_\xi f(\omega,y,\nabla u_a(y)+a)|^2\, dy. \tag{17.214}$$

On the other hand, from the hypothesis of (ii)

$$\begin{aligned}
&\gamma \int_{nY} |\nabla_\xi f(\omega,y,\nabla u_{a_n}(y)+a_n) - \nabla_\xi f(\omega,y,\nabla u_a(y)+a)|^2 dy \\
&\leq \int_{nY} \left\langle \nabla_\xi f(\omega,y,\nabla u_{a_n}(y)+a_n) - \nabla_\xi f(\omega,y,\nabla u_a(y)+a), \nabla u_{a_n}(y)+a_n - \nabla u_a(y) - a\right\rangle_{\mathbf{R}^N} dy \\
&= \int_{nY} \left\langle \nabla_\xi f(\omega,y,\nabla u_{a_n}(y)+a_n) - \nabla_\xi f(\omega,y,\nabla u_a(y)+a), a_n - a\right\rangle_{\mathbf{R}^N} dy
\end{aligned}$$

from which we deduce

$$\frac{1}{n^N}\int_{nY} |\nabla_\xi f(\omega,y,\nabla u_{a_n}(y)+a_n) - \nabla_\xi f(\omega,y,\nabla u_a(y)+a)|^2\, dy \leq \frac{1}{\gamma^2}|a_n - a|^2. \tag{17.215}$$

Combining (17.214) and (17.215) yields (17.213).

From (17.213), we infer that $a^* = \lim_{n\to+\infty} \nabla f_n(\omega,a)$. Thus $\partial f^{hom}(\omega,.)$ is made up of a single point, which concludes the proof. $\square$

Combining Proposition 12.3.4, Theorem 17.4.8, and Proposition 17.4.6, we deduce the following convergence result of the evolution problem:

$$(\mathscr{P}_\varepsilon) \quad \begin{cases} \dfrac{du_\varepsilon}{dt} - \operatorname{div} \partial_\xi f\left(\dfrac{\cdot}{\varepsilon}, \nabla u_\varepsilon\right) \ni g_\varepsilon, \\ u_\varepsilon(0) = u_\varepsilon^0, \end{cases}$$

when $x \mapsto f(x,\xi)$ is Y-periodic.

Corollary 17.4.4 (periodic case). *Assume that u_ε^0 strongly converges to u^0 in $L^2(\Omega)$ and that g_ε strongly converges to g in $L^2([0,T];L^2(\Omega))$. Suppose further that the unique solution u_ε of $(\mathscr{P}_\varepsilon)$ satisfies $\sup_{\varepsilon>0} F_\varepsilon(u_\varepsilon^0) < +\infty$. Then u_ε uniformly converges in $\mathbf{C}(0,T;L^2(\Omega))$ to the unique solution u of the evolution problem*

$$(\mathscr{P}^{hom}(\omega)) \quad \begin{cases} \dfrac{du}{dt} - \operatorname{div} \partial f^{hom}(\nabla u) \ni g, \\ u(0) = u^0, \end{cases} \tag{17.216}$$

where f^{hom} is given by

$$f^{hom}(a) = \inf\left\{ \int_Y f(y, a + \nabla u(y))\, dy : u \in W_\#^{1,p}(Y) \right\}.$$

Furthermore,

$$\frac{du_\varepsilon}{dt} \rightharpoonup \frac{du}{dt} \quad \text{in } L^2(0,T;L^2(\Omega))$$

and, if $F_\varepsilon(u_\varepsilon^0) \to F^{hom}(u^0)$,

$$\frac{du_\varepsilon}{dt} \to \frac{du}{dt} \quad \text{in } L^2(0,T;L^2(\Omega)).$$

Assume that for a.e. $x \in \mathbf{R}^N$, $f(x,.)$ *is strictly convex and Gâteaux-differentiable and that its Fenchel conjugate is such that* $\langle \xi_1^* - \xi_2^*, \xi^1 - \xi^2 \rangle \geq \gamma |\xi_1 - \xi_2|^2$ *for some* $\gamma > 0$ *and for all* $(\xi_1, \xi_2) \in \mathbf{R}^N \times \mathbf{R}^N$ *and all* $(\xi_1^*, \xi_2^*) \in \partial_\xi f^*(x,\xi_1) \times \partial_\xi f^*(x,\xi_2)$. *Then* f^{hom} *is Gâteaux differentiable and for all* $a \in \mathbf{R}^N$,

$$\nabla f^{hom}(a) = \int_Y \nabla_\xi f(y, \nabla u_a(\omega)(y) + a)\, dy,$$

where u_a *is the solution of the Dirichlet problem*

$$\begin{cases} \operatorname{div}\left(\nabla_\xi f(.,a+\nabla v(.))\right) = 0 & \text{a.e. in } Y; \\ v \in W_\#^{1,2}(Y)/\mathbf{R}. \end{cases}$$

Example 17.4.1. We choose to illustrate Theorem 17.4.8 by considering a diffusion through a composite made up of small balls distributed at random following a Poisson point process with intensity $\lambda \mathscr{L}_N$, $\lambda > 0$, included in a homogeneous material. We carry on with Example 12.4.2, where the convex density g_- represents, for instance, a thermal or electrical conductivity of the balls whose centers are randomly distributed with a frequency λ per unit of volume, while the convex density g_+ represents a conductivity outside the balls. The random integrand is then given by

$$f(\omega,x,\xi) = \begin{cases} g_-(\xi) & \text{if } x \in \bigcup_{i\in\mathbf{N}} B(\omega_i, r), \\ g_+(\xi) & \text{otherwise,} \end{cases}$$

or, in an equivalent way, by

$$f(\omega,x,\xi) := g_+(\xi) + \left(g_-(\xi) - g_+(\xi)\right)\min(1, \mathscr{N}(\omega, B(x,r))), \tag{17.217}$$

where $\mathscr{N}$ is the Poisson point process satisfying for all $A \in \mathscr{B}_b(\mathbf{R}^3)$, $\mathscr{N}(\omega,A) = \#(A\cap\Omega)$ and $\mathbf{E}(\mathscr{N}(.,A)) = \lambda \mathscr{L}_N(A)$. Starting from formula (17.217), one can easily see that f is ergodic. We consider the standard quadratic case, i.e., $g_-(\xi) = \frac{\alpha}{2}|\xi|^2$ and $g_+(\xi) = \frac{\beta}{2}|\xi|^2$. It is easy to show that $g_-^*(\xi) = \frac{1}{2\alpha}|\xi^*|^2$ and $g_+^*(\xi) = \frac{1}{2\beta}|\xi^*|^2$ and that condition (iii) of Proposition 17.4.6 is fulfilled with $\gamma = \min\{\frac{1}{\alpha}, \frac{1}{\beta}\}$.

Finally we assume that for $\mathbf{P}$-a.e. $\omega \in \Sigma$, $u_\varepsilon^0(\omega)$ strongly converges in $L^2(\Omega)$ to some $u^0(\omega)$ and that the source $g_\varepsilon(\omega)$ strongly converges to a function $g(\omega)$ in $L^2([0,T];L^2(\Omega))$. Then applying Theorem 17.4.8 together with Proposition 17.4.6, we infer that the unique solution $u_\varepsilon(\omega)$ of the random evolution problem

$$(\mathscr{P}_\varepsilon(\omega)) \quad \begin{cases} \dfrac{du_\varepsilon(\omega)}{dt} - \operatorname{div} \nabla_\xi f\left(\omega, \dfrac{\cdot}{\varepsilon}, \nabla u_\varepsilon\right) = g_\varepsilon(\omega), \\ u_\varepsilon(\omega,0) = u_\varepsilon^0(\omega), \end{cases}$$

$\mathbf{P}$-almost surely uniformly converges in $\mathbf{C}(0,T;L^2(\Omega))$ to the unique solution $u(\omega)$ of the evolution problem

$$(\mathscr{P}^{hom}) \quad \begin{cases} \dfrac{du(\omega)}{dt} - \operatorname{div} \nabla f^{hom}(\nabla u(\omega)) = g(\omega), \\ u(\omega,0) = u^0(\omega). \end{cases} \tag{17.218}$$

Furthermore, for $\mathbf{P}$-a.e. $\omega \in \Sigma$, $\frac{du_\varepsilon(\omega)}{dt} \rightharpoonup \frac{du(\omega)}{dt}$ in $L^2(0,T;L^2(\Omega))$.

The limit deterministic operator $\operatorname{div} \nabla f^{hom}(\nabla u(\omega))$ can be calculated following the process below:

- Solve the random Dirichlet problem

$$\begin{cases} \operatorname{div}\big(\nabla_\xi f(\omega,.,a+\nabla v(.))\big) = 0 \quad \text{a.e. in } nY; \\ v = 0 \quad \text{on } \partial nY, \end{cases}$$

  whose $u_{a,n}(\omega)$ is the unique solution.

- Compute $\nabla f_n(\omega,a) = \frac{1}{n^N}\int_{nY} \nabla_\xi f(\omega,y,\nabla u_{a,n}(\omega)(y)+a)\,dy$; then, for $\mathbf{P}$-a.e. ω in Σ,

$$\nabla f^{hom}(\omega,a) = \lim_{n\to+\infty} \nabla f_n(\omega,a).$$

17.5 ▪ Steepest descent and gradient flow on general metric spaces

Evolution equations in general describe the changing of a physical (or economic, or social) system with respect to the time; in many situations the state of the system is the main variable entering in the evaluation of a cost functional Φ whose values tend to become as low as possible in a unit of time. Then we say that the system evolves through the *maximal slope* or the *steepest descent* of the cost functional Φ and that the evolution occurs through the *gradient flow* of Φ.

The theory of gradient flows has received great attention from the mathematical community in the recent years, mainly because of several links with the mass transportation theory presented in Section 11.5. This made it possible to write several partial differential equations of evolution type as gradient flows of functionals defined in some spaces of measures.

We recall here some notions and results in the direction of variationally driven evolutions in metric spaces. In particular, we start by presenting the general theory of steepest descent and gradient flow on general metric spaces, introduced by De Giorgi in [194] in order to study evolution problems with an underlying variational structure. The theory was later developed in the monograph [27], to which we refer for further details. The framework of the theory is very general and applies both to quasi-static evolutions as well as to gradient flows, under rather mild assumptions.

When the state of the system under consideration is a vector $u(t)$ of the Euclidean space $\mathbf{R}^N$ or more generally of a Hilbert space $\mathscr{H}$, and the cost functional Φ is smooth, the evolution by maximal slope is described by the differential equation

$$\dot{u}(t) = -\nabla\Phi\big(u(t)\big). \tag{17.219}$$

In fact, multiplying by $\dot{u}$ the equation above, we obtain

$$\frac{d}{dt}\Phi\big(u(t)\big) = -|\dot{u}(t)|^2,$$

which shows that $\Phi\big(u(t)\big)$ decreases. The two scalar equalities

$$\begin{cases} \dfrac{d}{dt}\Phi\big(u(t)\big) = -|\nabla\Phi\big(u(t)\big)||\dot{u}(t)|, \\ |\dot{u}(t)| = |\nabla\Phi\big(u(t)\big)| \end{cases}$$

then show that the decreasing rate of $\Phi\big(u(t)\big)$ is maximal. It is interesting to note that we can equivalently write the differential equation (17.219) by the inequality

$$\frac{d}{dt}\Phi\big(u(t)\big) \leq -\frac{1}{2}|\dot{u}(t)|^2 - \frac{1}{2}|\nabla\Phi\big(u(t)\big)|^2, \tag{17.220}$$

where only the quantities $|\nabla\Phi|$ and $|\dot{u}|$ appear.

When the state variable of the system belongs to a metric space, as, for instance, occurs in the case of shape optimization problems, the concept of differentiability and smoothness are no longer available, and the description of evolution by maximal slope and the related gradient flow have to be defined in a more general way.

Our general framework deals with a complete metric space (X,d), an initial condition $u_0 \in X$, and a functional $\Phi : X \to]-\infty,+\infty]$. In the following, we set

$$\operatorname{dom}\Phi = \big\{u \in X \;:\; \Phi(u) < +\infty\big\}$$

and we always assume that Φ is *proper*, that is, $\operatorname{dom}\Phi \neq \emptyset$, which means that Φ is not constantly equal to $+\infty$.

An important concept in this framework is the one of *metric derivative*. For a function $u : [0,T] \to X$ we call metric derivative at the point t_0 the quantity

$$|\dot{u}|(t_0) = \limsup_{t\to t_0} \frac{d\big(u(t),u(t_0)\big)}{|t-t_0|}.$$

The function u is said to belong to the class $AC^p(0,T;X)$, with $p \in [1,\infty]$, if the metric derivative $|\dot{u}|$ belongs to $L^p(0,T)$. In this case it can be shown that the lim sup above is actually a limit for a.e. $t_0 \in [0,T]$.

Definition 17.5.1. *A function $G : X \to [0,+\infty]$ is called an upper gradient of Φ if for every curve $u \in AC^1(0,T;X)$ with $(G\circ u)|\dot{u}| \in L^1(0,T)$ the function $\Phi\circ u$ belongs to $W^{1,1}(0,T)$ and*

$$\left|\frac{d}{dt}\Phi\big(u(t)\big)\right| \leq G\big(u(t)\big)|\dot{u}|(t) \qquad \text{for a.e. } t \in [0,T].$$

Remark 17.5.1. In [27] a weaker definition of upper gradient is also considered, which requires only that the function $\Phi\circ u$ belongs to $BV(0,T)$.

Definition 17.5.2. *A locally absolutely continuous curve $u : [0,T] \to X$ is said a curve of maximal slope (or of steepest descent) for the functional Φ with respect to a functional G if*

(i) *G is an upper gradient of Φ;*

(ii) *$\Phi\big(u(t)\big)$ is decreasing;*

(iii) $\dfrac{d}{dt}\Phi\big(u(t)\big) \leq -\dfrac{1}{2}|\dot{u}|^2(t) - \dfrac{1}{2}G^2\big(u(t)\big)$ *for a.e. $t \in [0,T]$.*

In view of Definitions 17.5.1 and 17.5.2 above it is crucial to identify some canonical functional that is a good candidate to be an upper gradient of a given cost Φ.

Definition 17.5.3. *The local slope $|\partial\Phi|$ of Φ at a point $u_0 \in \operatorname{dom}\Phi$ is defined by*

$$|\partial\Phi|(u_0) = \limsup_{u\to u_0} \frac{\big(\Phi(u_0)-\Phi(u)\big)^+}{d(u,u_0)},$$

where $(\cdot)^+$ denotes the positive part function. If τ is another topology on X we also define the τ-relaxed slope $|\partial_\tau^-\Phi|$ of Φ as

$$|\partial_\tau^-\Phi|(u_0) = \inf\left\{\liminf_n |\partial\Phi|(u_n) \,:\, u_n \to_\tau u,\ u_n \text{ bounded in } X,\ \sup_n \Phi(u_n) < +\infty\right\}.$$

We are now in a position to consider the general problem of existence of steepest descent curves in a complete metric space X; in Section 17.6 we see that under quite mild assumptions on Φ these curves exist and they can be considered as the natural generalizations of the evolution PDEs for smooth cost functionals in Hilbert spaces.

17.6 ▪ Minimizing movements and the implicit Euler scheme

The framework we consider in this section is the same as that of Section 17.5, that is, (X,d) a complete metric space, an initial condition $u_0 \in X$, and a functional $\Phi : X \to]-\infty,+\infty]$ which we assume *proper*, that is, not constantly equal to $+\infty$.

For every fixed $\varepsilon > 0$ the *implicit Euler scheme* of time step ε and initial condition u_0 consists in constructing a function $u_\varepsilon(t) = w([t/\varepsilon])$, where $[\cdot]$ stands for the integer part function, in the following recursive way:

$$w(0) = u_0, \qquad w(n+1) \in \arg\min\left\{\Phi(v) + \frac{d^2\big(v, w(n)\big)}{2\varepsilon}\right\}. \tag{17.221}$$

Definition 17.6.1. *We say that $u : [0,T] \to X$ is a minimizing movement associated to the functional Φ, to the topology τ, and to the initial condition u_0, and we write $u \in MM(\Phi,\tau,u_0)$ if*

$$u_\varepsilon(t) \to_\tau u(t) \qquad \forall t \in [0,T].$$

If the limit above occurs only for a sequence $(\varepsilon_n)_{n\in\mathbf{N}}$ (independent of t), we say that $u : [0,T] \to X$ is a generalized minimizing movement and we write $u \in GMM(\Phi,\tau,u_0)$.

Let $\Phi : X \to]-\infty,+\infty]$ be a proper functional and let τ be a topology on X. We assume the following:

(i) τ is weaker that d and

$$u_n \to_\tau u,\ v_n \to_\tau v \ \Rightarrow\ d(u,v) \le \liminf_n d(u_n,v_n);$$

(ii) Φ is sequentially τ-lower semicontinuous, that is,

$$u_n \to_\tau u \ \Rightarrow\ \Phi(u) \le \liminf_n \Phi(u_n);$$

(iii) Φ is τ-coercive, that is, for every $c \in \mathbf{R}$ the sublevel $\{\Phi \le c\}$ is sequentially τ-compact.

A crucial existence theorem for minimizing movements is the following. (We refer to Proposition 2.2.3 of [27] for the proof.)

Theorem 17.6.1. *Under the assumptions above, for every initial condition $u_0 \in \operatorname{dom}\Phi$ there exists a generalized minimizing movement $u \in GMM(\Phi, \tau, u_0)$. Moreover, we have that $GMM(\Phi, \tau, u_0) \subset AC^2(0, T; X)$.*

Remark 17.6.1. For a fixed $p > 1$, a recursive construction similar to (17.221) can be done by setting

$$w(0) = u_0, \qquad w(n+1) \in \operatorname{arg\,min}\left\{\Phi(v) + \frac{d^p\big(v, w(n)\big)}{p\varepsilon^{p-1}}\right\}.$$

Then a result similar to the existence theorem, Theorem 17.6.1, holds, in the sense that $GMM(\Phi, \tau, u_0)$ is nonempty and $GMM(\Phi, \tau, u_0) \subset AC^p(0, T; X)$.

The case $p = 1$ corresponds to the recursive definition

$$w(0) = u_0, \qquad w(n+1) \in \operatorname{arg\,min}\left\{\Phi(v) + d\big(v, w(n)\big)\right\}$$

and is used to describe the quasi-static evolution problems. We refer to [282] for a general presentation of quasi-static evolution problems and rate-independent processes.

The link between minimizing movements and curves of maximal slope is given by the following result (see Theorem 2.3.3 of [27]).

Theorem 17.6.2. *Let us assume the conditions* (i), (ii), (iii) *above on Φ and τ; assume in addition that*

$$\text{the mapping } |\partial_\tau^- \Phi| \text{ is an upper gradient of } \Phi. \tag{17.222}$$

Then, for every initial condition $u_0 \in \operatorname{dom}\Phi$, every curve $u \in GMM(\Phi, \tau, u_0)$ is a curve of maximal slope for Φ with respect to $|\partial_\tau^- \Phi|$. Moreover, the energy identity

$$\frac{1}{2}\int_0^s |\dot{u}|^2\,dt + \frac{1}{2}\int_0^s |\partial_\tau^- \Phi|^2\big(u(t)\big)\,dt + \Phi\big(u(s)\big) = \Phi(u_0)$$

holds for every $s > 0$.

Notice that from the energy identity above we obtain an equality in condition (iii) of Definition 17.5.2 of steepest descent curves:

$$\frac{d}{dt}\Phi\big(u(t)\big) + \frac{1}{2}|\dot{u}|^2(t) + \frac{1}{2}|\partial_\tau^- \Phi|^2\big(u(t)\big) = 0 \quad \text{for a.e. } t \in [0, T].$$

The problem is now reduced to proving condition (17.222). The simplest situation in which this can be done is when X is a Hilbert space, τ its weak topology, and $\Phi : X \to\ ]-\infty, +\infty]$ a proper, convex, lower semicontinuous cost functional. In this case we have $|\partial_\tau^- v| = |\partial\Phi|$ and both coincide with the element of minimal norm of the subdifferential of Φ, which is an upper gradient of Φ. This case is analyzed in detail in Section 17.2.

An interesting generalization of the concept of convexity in metric spaces is the notion of *geodesic convexity*.

Definition 17.6.2. *Let $\lambda \in \mathbf{R}$ be fixed. A functional $\Phi : X \to]-\infty, +\infty]$ is called λ-geodesically convex if for every $u_0, u_1 \in X$, there exists a curve $\gamma : [0,1] \to X$ such that*

(i) $\gamma(0) = u_0$ *and* $\gamma(1) = u_1$;

(ii) γ *is a constant speed geodesic, that is,*

$$d\big(\gamma(s), \gamma(t)\big) = (t-s)d(u_0, u_1) \qquad \text{for every } 0 \le s \le t \le 1;$$

(iii) Φ *is λ-convex along γ, that is,*

$$\Phi\big(\gamma(t)\big) \le (1-t)\Phi(u_0) + t\Phi(u_1) - \lambda \frac{t(1-t)}{2} d^2(u_0, u_1) \qquad \text{for every } t \in [0,1].$$

Note that in a Banach space, since the geodesics are the line segments, when $\lambda = 0$ we recover the usual convexity. For λ-geodesically convex functionals the following result holds (see Corollary 2.4.12 of [27]).

Theorem 17.6.3. *Let us assume the conditions* (i), (ii), (iii) *on Φ and τ. Assume in addition that*

(a) Φ *is λ-geodesically convex for some $\lambda \in \mathbf{R}$;*

(b) $|\partial \Phi| = |\partial_\tau^-|$, *that is, the map $u \mapsto |\partial \Phi|(u)$ is sequentially τ-lsc on d-bounded subsets of sublevels of Φ.*

Then the map $|\partial \Phi|$ is an upper gradient of Φ and so Theorem 17.6.2 *applies.*

The class of λ-geodesically convex cost functionals includes some very interesting cases coming from mass transportation theory, in which the metric space X is the space $\mathbf{P}(\Omega)$ of probabilities on Ω, metrized by the Wasserstein distance introduced in Section 11.5. We refer the interested reader to the book [27] and the references therein.