Weierstraß-Institut
für Angewandte Analysis und Stochastik

Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 2198-5855

An introduction to the analysis of gradients systems

Alexander Mielke

submitted: June 19, 2023

Weierstraß-Institut
Mohrenstr. 39
10117 Berlin
Germany
and
Institut für Mathematik
Humboldt-Universität zu Berlin
Rudower Chaussee 25
12489 Berlin
E-Mail: alexander.mielke@wias-berlin.de

No. 3022
Berlin 2023

2020 *Mathematics Subject Classification.* 35-02, 49-02, 35K55, 47H05.

Key words and phrases. Gradient-flow equations, dissipation potential, kinetic relation, Fenchel equivalences, Fréchet subdifferentials, time-incremental minimization, abstract chain rule, energy-dissipation balance, energy-dissipation principle, minimizing movements, metric slope and metric speed, curves of maximal slope, De Giorgi lemma, evolutionary Gamma-convergence.

The author is grateful to Moritz Gau and Jia-Jie Zhu for several critical and constructive remarks that helped to improve these lecture notes. Of course, this work benefited greatly from fruitful discussion with many collaborators, in particular Thomas Frenzel, Matthias Liero, Mark Peletier, Riccarda Rossi, Giuseppe Savaré, and Artur Stephan.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

An introduction to the analysis of gradients systems

Alexander Mielke

Abstract

The present notes provide an extended version of a small lecture course (of 36 hours) given at the Humboldt-Universität zu Berlin in the Winter Term 2022/23. The material starting in Section 5.4 was added afterwards.

The aim of these notes to give an introductory overview on the analytical approaches for gradient-flow equations in Hilbert spaces, Banach spaces, and metric spaces and to show that on the first entry level these theories have a lot in common. The theories and their specific setups are illustrated by suitable examples and counterexamples.

The merit of the right gradient flow formulation
of a dissipative evolution equation is that
it separates energetics and kinetics:
The energetics endow the state space M
with a functional E,
the kinetics endow the state space with a
(Riemannian) geometry via the metric tensor g.

Felix Otto 2001

Contents

DOI 10.20347/WIAS.PREPRINT.3022 Berlin 2023

1 Introduction

In this section we introduce our notions, provide a series of examples and give motivations concerning the origins of gradients systems.

1.1 Gradients in the finite dimensional case

We first discuss the notion of gradient of a function $\mathcal{F}:\mathbb{R}^d\to\mathbb{R}$. We distinguish the gradient $\operatorname{grad}\mathcal{F}$ and the Fréchet derivative $\mathrm{D}\mathcal{F}$ via

$$\operatorname{grad}\mathcal{F}(u)=\begin{pmatrix}\partial_{u_1}\mathcal{F}(u)\\ \vdots\\ \partial_{u_d}\mathcal{F}(u)\end{pmatrix}\quad\text{and}\quad \mathrm{D}\mathcal{F}(u)=\big(\partial_{u_1}\mathcal{F}(u),\cdots,\partial_{u_d}\mathcal{F}(u)\big)\in\big(\mathbb{R}^d\big)^*.$$

We will use the abbreviation "grad" for general gradients and reserve the symbol "∇" for PDE applications like $\Delta u=\operatorname{div}(\nabla u)$.

For the function $\mathcal{F}(u_1,u_2)=\frac12 u_1^2+\frac12 u_2^2+\frac a4 u_2^4$ we obtain

$$\operatorname{grad}\mathcal{F}(u)=\begin{pmatrix}u_1\\ u_2{+}au_2^3\end{pmatrix}\quad\text{and}\quad \mathrm{D}\mathcal{F}(u)=\big(u_1,u_2{+}au_2^3\big).$$

However we may describe the same function in polar coordinates $x=r(\cos\phi,\sin\phi)$ giving $\widetilde{\mathcal{F}}(r,\phi)=\frac12 r^2+\frac a4 r^4(\sin\phi)^4$. The definition of the gradient of $\mathcal{F}$ in polar coordinates, $\widetilde\nabla\widetilde{\mathcal{F}}$, is no longer given by the vector of partial derivatives but

$$\widetilde{\operatorname{grad}}\,\widetilde{\mathcal{F}}(r,\phi)=\widetilde{\mathbb{K}}(r,\phi)\begin{pmatrix}\partial_r\widetilde{\mathcal{F}}(r,\phi)\\ \partial_\phi\widetilde{\mathcal{F}}(r,\phi)\end{pmatrix}\quad\text{with } \widetilde{\mathbb{K}}(r,\phi)=\begin{pmatrix}1&0\\0&1/r^2\end{pmatrix}.$$

What is the reason for the nontrivial $\widetilde{\mathbb{K}}$? One justification is that we want the gradient-flow equations

$$\dot u=-\operatorname{grad}\mathcal{F}(u)\quad\text{and}\quad\begin{pmatrix}\dot r\\ \dot\phi\end{pmatrix}=-\widetilde{\operatorname{grad}}\,\widetilde{\mathcal{F}}(r,\phi)\tag{1.1}$$

to be the same.

However, more importantly, the right perspective is to consider the space $\mathbb{R}^2$ as a manifold M and $\mathcal{F}:M\to\mathbb{R}$ as a general function. Then, $\mathrm{D}\mathcal{F}(u)$ is the differential of $\mathcal{F}$ at u (in differential geometry written as $\boldsymbol{d}\mathcal{F}(u)$). It is defined via

$$\mathrm{D}\mathcal{F}(u)[v]:=\lim_{h\to0}\frac1h\big(\mathcal{F}(u{+}hv)-\mathcal{F}(u)\big)$$

and thus we have $\mathrm{D}\mathcal{F}(u)\in\mathrm{Lin}(\mathrm{T}_uM;\mathbb{R})=:\mathrm{T}_u^*M$. Here we use the notion of the tangent space T_uM and the co-tangent space T_u^*M at a point $u\in M$. We also use the duality notation

$$\mathrm{D}\mathcal{F}(u)[v]={}_{\mathrm{T}_u^*M}\big\langle\mathrm{D}\mathcal{F}(u),v\big\rangle_{\mathrm{T}_uM},$$

where ${}_{X^*}\langle\cdot,\cdot\rangle_X$ always means a duality pairing between a space X and its dual space X^*.

However, by the definition of the gradient-flow equation $\dot u = -\operatorname{grad}\mathcal{F}(u)$ we see that the gradient has to lie in the tangent space T_uM. Hence, we need a mapping that maps the differential $\mathrm{D}\mathcal{F}(u) \in \mathrm{T}^*_uM$ into the vector $\operatorname{grad}\mathcal{F}(u) \in \mathrm{T}_uM$.

This mapping is generated by a Riemannian structure $\mathbb{G}$. A pair $(M,\mathbb{G})$ is called a Riemannian manifold, if

- M is a manifold and
- $\mathbb{G}(u) : \mathrm{T}_uM \to \mathrm{T}^*_uM$ is symmetric and positive,
- g defined via $g(v,\widetilde v)_u = \langle \mathbb{G}(u)v(u), \widetilde v(u)\rangle$ is a symmetric 2-tensor.

Riemannian structures are used for measuring length of curves and angles between curves, as they define a scalar product on each T_uM. For curves $\gamma : [s_0,s_1] \to M$ one sets

$$\mathrm{length}_{\mathbb{G}}(\gamma) := \int_{s_0}^{s_1} \Big(\big\langle \mathbb{G}(\gamma(s))\gamma'(s), \gamma'(s)\big\rangle\Big)^{1/2} \,\mathrm{d}s.$$

When doing a transformation $u = \Phi(w)$ with $\Phi : N \to M$ the chain rule gives immediately the transformation rule $\widetilde{\mathbb{G}}(w) = \mathrm{D}\Phi(w)^*\mathbb{G}(\Phi(w))\mathrm{D}\Phi(w) : \mathrm{T}_wN \to \mathrm{T}^*_wN$.

Definition 1.1 (Gradient) *The gradient of a function $\mathcal{F}$ in a Riemannian manifold is defined via*

$$\operatorname{grad}_{\mathbb{G}}\mathcal{F}(u) := \mathbb{G}(u)^{-1}\mathrm{D}\mathcal{F}(u) = \mathbb{K}(u)\mathrm{D}\mathcal{F}(u), \tag{1.2}$$

where $\mathbb{K}(u) := \big(\mathbb{G}(u)\big)^{-1} : \mathrm{T}^*_uM \to \mathrm{T}_uM$ *is called the* Onsager operator*.*

For the above example in $\mathbb{R}^2$ we have $\mathbb{G}_{\mathrm{Euclid}} = \begin{pmatrix}1&0\\0&1\end{pmatrix}$ and the transformation $u = \Phi(r,\phi)$ into polar coordinates gives $\widetilde{\mathbb{G}}(r,\phi) = \begin{pmatrix}1&0\\0&r^2\end{pmatrix}$. Thus we find the gradient in polar coordinates in the following form

$$\widetilde{\operatorname{grad}}\,\widetilde{\mathcal{F}}(r,\phi) = \begin{pmatrix}1&0\\0&\frac{1}{r^2}\end{pmatrix}\mathrm{D}\widetilde{\mathcal{F}}(r,\phi) = \begin{pmatrix}\partial_r\widetilde{\mathcal{F}}(r,\phi)\\ \frac{1}{r^2}\,\partial_\phi\widetilde{\mathcal{F}}(r,\phi)\end{pmatrix} = \begin{pmatrix} r + ar^3(\sin\phi)^4\\ ar^2(\sin\phi)^3\cos\phi\end{pmatrix}.$$

With this, one can indeed check that the the two ODEs in (1.1) transform properly into each other.

1.2 Gradient systems and their gradient-flow equations

We still stay in the framework of finite-dimensional manifolds M and define what exactly we mean by the words “gradient system”, “gradient structure”, “gradient flow”, and “gradient-flow equation”.

Definition 1.2 *A* gradient system *is a triple $(M,\mathcal{F},\mathbb{G})$ such that $(M,\mathbb{G})$ is a Riemannian manifold and $\mathcal{F} : M \to \mathbb{R}$ is a C^1 function.*

This gradient system generates the associated gradient-flow equation

$$\dot u = -\operatorname{grad}_{\mathbb{G}}\mathcal{F}(u) = -\mathbb{K}(u)\mathrm{D}\mathcal{F}(u) \in \mathrm{T}_uM \quad\Longleftrightarrow\quad 0 = \mathbb{G}(u)\dot u + \mathrm{D}\mathcal{F}(u) \in \mathrm{T}^*_uM. \tag{1.3}$$

We say that $u : [0,T[\to M$ is a solution for $(M,\mathcal{F},\mathbb{G};u^0)$ if it satisfies (1.3) *with $u(0) = u^0$.*

We see that $\mathcal{F}$ is a Lyapunov function, i.e. along solutions $u : [0,T] \to M$ the function $\mathcal{F}$ is decreasing:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(u(t)) = \langle \mathrm{D}\mathcal{F}(u), \dot u\rangle = -\langle \mathrm{D}\mathcal{F}(u), \mathbb{K}(u)\mathrm{D}\mathcal{F}(u)\rangle$$
$$= -\langle \mathbb{G}(u)\mathrm{grad}_{\mathbb{G}}\mathcal{F}(u), \mathrm{grad}_{\mathbb{G}}\mathcal{F}(u)\rangle \leq 0.$$

The left equation in (1.3) will be called the rate form of the gradient-flow equation, whereas the right equation is called the force-balance form of the gradient-flow equation. Here $\xi = \mathbb{G}(u)\dot u$ is the viscous force induced by the rate. We call

$$\xi_{\mathrm{v}} = \mathbb{G}(u)\dot u \quad \text{or equivalently} \quad \dot u = \mathbb{K}(u)\xi_{\mathrm{v}}$$

the *kinetic relation* encoding the frictional properties of the system. The force $\xi = \mathrm{D}\mathcal{F}(u)$ is the potential restoring force. Of course, kinetic relations can be more general, e.g. by a non-symmetric linear relation or by nonlinear relations, see Section 3.1.

However, from a thermodynamical point of view the case of symmetric and positive definite $\mathbb{G}$ or $\mathbb{K}$ is distinguished as is shown by the fundamental work by Lars Onsager "*Reciprocal relations in irreversible processes*" [Ons31]. His "reciprocal relations" were derived in the context of linearized irreversible thermodynamics and simply mean, in modern language, the *symmetry relation* $\mathbb{G} = \mathbb{G}^*$. In fact, Onsager was awarded the Nobel prize for chemistry in 1968 for exactly this work, see

https://www.nobelprize.org/prizes/chemistry/1968/ceremony-speech/

As Onsager and Machlup state in the follow-up work [OnM53, p. 1507] [formulas slightly adapted]:

> *The tendency of the system to seek equilibrium is measured by the* thermodynamic forces *(=restoring forces)* $\xi = \mathrm{D}\mathcal{S}(\alpha)$ *(eqn. (2-1)), which evidently vanish at* $\alpha = 0$.
>
> *The fluxes (of matter, heat, electricity) are measured by the time derivative* $\dot\alpha$. *The essential physical assumption about the irreversible processes is that they are linear; i.e., that the fluxes depend linearly on the forces that "cause" them:*
>
> $$\mathbb{G}\dot\alpha = \xi \quad \text{(2-2)} \qquad \text{or} \qquad \mathbb{K}\xi = \dot\alpha \quad \text{(2-3)},$$
>
> *where the matrices* $\mathbb{G}$ *and* $\mathbb{K}$ *are mutual reciprocal [inverses].*
>
> *These equations express, for instance, Ohm's law for electric conduction, Fourier's law for heat conduction, Fick's law for diffusion, and the extension of these laws to interacting flows, e.g., anisotropic conduction (heat, electricity), thermoelectric effects, thermal diffusion. For systems for which microscopic reversibility holds (to which this work is confined), we have the* reciprocal relations *[symmetry relations]* $\mathbb{G} = \mathbb{G}_{tr}$ *(eqn. (2-4)), where the subscript* $_{tr}$ *means transpose.*

Under sufficient smoothness, for each $u^0 \in M$ there exists a (local or global) solution $u(t) = S_t(u^0)$ where $S_t : M \to M$ is the *gradient flow* associated with $(M, \mathcal{F}, \mathbb{G})$. Assuming that all solutions exist globally, i.e. for $t \in [0, \infty[$ the gradient flow $(S_t)_{t\geq 0}$ satisfies

$$(1) \quad S_0 = \mathrm{id}_M \text{ and } \forall\, t, r \geq 0 : \quad S_t \circ S_r = S_{t+r},$$
$$(2) \quad u(t) = S_t(u_0) \text{ is a solution for } (M, \mathcal{F}, \mathbb{G}; u^0).$$

Property (1) is called the semigroup property of the family $(S_t)_{t\geq 0}$.

Remark 1.3 (Hamiltonian systems) *The notion of gradient systems is chosen in analogy to Hamiltonian systems* $(M, \mathcal{H}, \Omega)$ *(cf. [AbM78, Arn89]) where* (M, Ω) *is a symplectic manifold with* $\Omega(u) : \mathrm{T}_u M \to \mathrm{T}_u^* M$ *satisfying* $\Omega(u)^* = -\Omega(u)$, $\mathbb{J}(u) = (\Omega(u))^{-1}$ *exists, and* $\mathrm{d}\Omega \equiv 0$ *(in the sense of two-forms). The associated Hamiltonian equations are given by*

$$\dot{u} = \mathbb{J}(u)\mathrm{D}\mathcal{H}(u) \in \mathrm{T}_u M \quad \Longleftrightarrow \quad \Omega(u)\dot{u} = \mathrm{D}\mathcal{H}(u) \in \mathrm{T}_u^* M. \tag{1.4}$$

Along solutions we have $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{H}(u(t)) = \langle \mathrm{D}\mathcal{H}(u), \dot{u} \rangle = \langle \mathrm{D}\mathcal{H}(u), \mathbb{J}(u)\mathrm{D}\mathcal{H}(u) \rangle = 0$*, which means energy conservation.*

Definition 1.4 (Gradient structure) *Given a differential equation* $\dot{u} = \boldsymbol{V}(u)$ *on a manifold* M *we say that the equation has the* gradient structure $(M, \mathcal{F}, \mathbb{G})$ *if* $\boldsymbol{V}(u) = \mathbb{K}(u)\mathrm{D}\mathcal{F}(u)$ *for all* $u \in M$*, i.e. the ODE is the gradient-flow equation associated with* $(M, \mathcal{F}, \mathbb{G})$*.*

Note the two different perspectives:

(I) The GS $(M, \mathcal{F}, \mathbb{G})$ generates the (unique) gradient-flow equation $\dot{u} = -\mathbb{K}(u)\mathrm{D}\mathcal{F}(u)$.

(II) A given ODE can have one or many gradient structure or no at all.

Example 1.5 (Trivial scaling) *If* $\dot{u} = \boldsymbol{V}(u)$ *has the gradient structure* $(M, \mathcal{F}, \mathbb{G})$*, then for all* $\lambda > 0$ *it also has the gradient structure* $(M, \widetilde{\mathcal{F}}, \widetilde{\mathbb{G}}) = (M, \lambda\mathcal{F}, \lambda\mathbb{G})$*. Simply observe that* $\widetilde{\mathbb{K}} = (\lambda\mathbb{G})^{-1} = \frac{1}{\lambda}\mathbb{K}$*, such that* λ *cancels.*

Example 1.6 (Two nontrivial structures) *Let* $M = \mathbb{R}^2$ *and* $\dot{u} = \boldsymbol{V}(u) = \binom{-u_1}{-u_2 - au_2^3}$ *with* $a > 0$*. From above we know that we have the gradient structure*

$$\mathbb{G} = \mathbb{I}_{\mathrm{Eucl}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \textit{and} \quad \mathcal{F}(u) = \frac{1}{2}u_1^2 + \frac{1}{2}u_2^2 + \frac{a}{4}u_2^4.$$

However, there is another gradient structure $(\mathbb{R}^2, \widetilde{\mathcal{F}}, \widetilde{\mathbb{G}})$*, namely*

$$\widetilde{\mathbb{G}} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{1+au_2^2} \end{pmatrix} \quad \textit{and} \quad \mathcal{F}(u) = \frac{1}{2}u_1^2 + \frac{1}{2}u_2^2.$$

Thus, when looking at the ODE we do not know whether the coefficient $a > 0$ *and the nonlinear term* $-au_2^3$ *arises because of a nonquadratic energy (as in* $\mathcal{F}$*) or because of a state-dependent friction law (as in* $\widetilde{\mathbb{G}}$*).*

The last example shows that different gradient structures for an ODE refer to different physics/mechanics behind the model. The gradient structure contains *additional information* that is not contained in the ODE.

The next example is a more recent one and relates to chemical reaction-rate equations.

Example 1.7 (Reaction-rate equations) *We consider three chemical species denoted by* $X_1,\ X_2,$ *and* X_3 *with densities* $c_1,\ c_2$*, and* c_3*, respectively. Hence, the states are* $c = (c_i)_i$ *in the manifold is* $M = \,]0, \infty[^3$*. We consider three reactions*

$$X_1 \rightleftharpoons X_2, \quad X_1 + X_2 \rightleftharpoons 2X_3, \quad a_1X_1 + a_3X_3 \rightleftharpoons b_2X_2$$

which follow the mass-action law, i.e. the reaction rates are proportional to the corresponding monomials. The ODE reads

$$\dot c = \boldsymbol{R}(c) := k_1(c_1{-}c_2)\begin{pmatrix}-1\\1\\0\end{pmatrix} + k_2(c_1c_2{-}c_3^2)\begin{pmatrix}-1\\-1\\2\end{pmatrix} + k_3(c_1^{a_1}c_3^{a_3}{-}c_2^{b_2})\begin{pmatrix}-a_1\\b_2\\-a_3\end{pmatrix}$$

It was observed in [Yon08] and in a more general setting in [Mie11c, MaM20], that the above equation has a gradient structure (because of the detailed-balance condition, see the references above). If we set

$$\mathcal{F}(c) = \sum_{i=1}^{3} \lambda_{\mathrm{B}}(c_i) \quad \textit{and} \quad \mathbb{K}(c) = \sum_{r=1}^{3} k_r \Lambda\big(c^{\boldsymbol{\alpha}^r},\, c^{\boldsymbol{\beta}^r}\big)(\boldsymbol{\alpha}^r{-}\boldsymbol{\beta}^r)\otimes(\boldsymbol{\alpha}^r{-}\boldsymbol{\beta}^r),$$

where $\lambda_{\mathrm{B}}(c) = c\log c - c + 1$ is the Boltzmann function and $\Lambda(r,\rho) = \int_0^1 r^s\rho^{1-s}\,\mathrm{d}s = (r{-}\rho)/\log(r/\rho)$ is the logarithmic mean of r and ρ. The stoichiometric vectors $\boldsymbol{\alpha}^r, \boldsymbol{\beta}^r \in \mathbb{N}_0^3$ are given via

$$\boldsymbol{\alpha}^1 = \begin{pmatrix}1\\0\\0\end{pmatrix},\ \boldsymbol{\beta}^1 = \begin{pmatrix}0\\1\\0\end{pmatrix},\ \boldsymbol{\alpha}^2 = \begin{pmatrix}1\\1\\0\end{pmatrix},\ \boldsymbol{\beta}^2 = \begin{pmatrix}0\\0\\2\end{pmatrix},\ \boldsymbol{\alpha}^3 = \begin{pmatrix}a_1\\0\\a_3\end{pmatrix},\ \boldsymbol{\beta}^3 = \begin{pmatrix}0\\b_2\\0\end{pmatrix}.$$

One nice feature of the above model is that it nicely shows the additive structure of $\mathbb{K}$: it is given as a sum over individual terms corresponding to a single reaction. This additive structure will often reappear, namely whenever there are several distinguishable dissipative processes. Their effect will be additive on the level of $\mathbb{K}$ but not on the level of $\mathbb{G}$. Hence, for modeling it is often more convenient to work with $\mathbb{K}$.

1.3 Gradient structures for partial differential equations

In this part we do mainly formal calculations only, and see this as a motivation for the analysis in the following sections. Nevertheless we are motivated by the philosophy from the smooth, finite-dimensional case discussed in the previous section. But now the function $\mathcal{F}$ may no longer be smooth but may attain the value $+\infty$ outside a dense set. Moreover the operator $\mathbb{K}$ may be unbounded.

As a first example we consider the Allen-Cahn equation, which is a nonlinear parabolic equation, sometimes called reaction-diffusion equation:

$$m\dot u = \alpha\Delta u + \beta\,(u - u^3) \quad \text{in } \Omega, \qquad u = 0 \quad \text{on } \partial\Omega, \tag{1.5}$$

where Ω is a bounded Lipschitz domain in $\mathbb{R}^d$.

We want to show that this equation has the gradient structure $(\mathrm{L}^2(\Omega), \mathcal{F}_{\mathrm{AC}}, m\mathbb{I}_{\mathrm{R}})$, where $\mathbb{I}_R : H \to H^*$ denotes the Riesz isomorphism of a Hilbert space H with its dual space H^*. The Allen-Cahn functional is given by

$$\mathcal{F}_{\mathrm{AC}}(u) = \begin{cases} \int_\Omega \big(\frac{\alpha}{2}|\nabla u|^2 + \frac{\beta}{4}(u^2{-}1)^2\big)\,\mathrm{d}x & \text{for } u \in \mathrm{dom}(\mathcal{F}_{\mathrm{AC}}),\\ \infty & \text{for } u \in \mathrm{L}^2(\Omega)\setminus\mathrm{dom}(\mathcal{F}_{\mathrm{AC}}),\end{cases}$$

where $\mathrm{dom}(\mathcal{F}_{\mathrm{AC}}) = \mathrm{H}^1_0(\Omega)\cap\mathrm{L}^4(\Omega)$. Moreover the differential $\mathrm{D}\mathcal{F}_{\mathrm{AC}}$ is replaced by the variational derivative, which is defined on an even smaller set:

$$\mathrm{D}\mathcal{F}_{\mathrm{AC}}(u) = -\alpha\Delta u - \beta\,(u - u^3) \quad \text{for } u \in \mathrm{dom}(\mathrm{D}\mathcal{F}_{\mathrm{AC}}) := \mathrm{H}^2(\Omega)\cap\mathrm{H}^1_0(\Omega)\cap\mathrm{L}^6(\Omega).$$

This notion of derivative will be made rigorous in terms of the Fréchet subdifferential to be introduced in Section 2. Recalling our choice $\mathbb{G} = m\mathbb{I}_\mathrm{R}$ for the Riemannian metric, we see that the "force-balance formulation" $\mathbb{G}\dot u = -\mathrm{D}\mathcal{F}(u)$ for the given gradient structure indeed yields the Allen-Cahn equation (1.5).

Next we consider the simple linear parabolic equation

$$\dot u = \Delta u \quad \text{in } \Omega, \qquad \nabla u \cdot \nu = 0 \quad \text{on } \partial\Omega, \tag{1.6}$$

where $\Omega \subset \mathbb{R}^d$ is again a bounded Lipschitz domain and ν is the outward unit normal on $\partial\Omega$. We will construct four quite different gradient structures, each of which corresponds to a different application of this equation. Recall that the name for this equation is usually "heat equation"; however, it is sometimes also called "diffusion equation".

Gradient Structure 1: Allen-Cahn type L^2 gradient flow: We consider the GS $(\mathrm{L}^2(\Omega), \mathcal{F}_\mathrm{Dir}, \mathbb{I}_\mathrm{R})$ with the Dirichlet functional defined on $\mathrm{L}^2(\Omega)$, namely

$$\mathcal{F}_\mathrm{Dir}(u) = \begin{cases} \int_\Omega \frac{1}{2}|\nabla u|^2 \,\mathrm{d}x & \text{for } u \in \mathrm{H}^1(\Omega), \\ +\infty & \text{for } u \in \mathrm{L}^2(\Omega) \setminus \mathrm{H}^1(\Omega) \end{cases}$$

Here the differential can be interpreted as a convex subdifferential, which is either empty or a singleton, namely $\mathrm{D}\mathcal{F}_\mathrm{Dir}(u) = -\Delta u$. Thus, we obtain (1.6) as the associated gradient-flow equation. This will be made rigorous in Section 2.

Gradient Structure 2: $\mathrm{H}^{-1}(\Omega)$ gradient flow: We again consider a spatially constant Hilbert-space structure, but now in the space $H := \big(\mathrm{H}^1(\Omega)\big)^* =: \mathrm{H}_0^{-1}(\Omega)$ such that the dual space is $H^* = \mathrm{H}^1(\Omega)$ and we have the Riesz isomorphism $\mathbb{I}_\mathrm{R} : \mathrm{H}_0^{-1}(\Omega) \to \mathrm{H}^1(\Omega)$. On the formal level we consider

$$\mathcal{F}_{\mathrm{L}^2}(u) = \frac{1}{2}\|u\|_{\mathrm{L}^2}^2 = \int_\Omega \frac{1}{2}\, u^2 \,\mathrm{d}x \quad \text{and} \quad \mathbb{K}^{(2)} \begin{cases} \mathrm{H}^1(\Omega) & \to & \mathrm{H}_0^{-1}(\Omega), \\ \xi & \mapsto & -\Delta\xi. \end{cases}$$

Note that the $\mathcal{F}_{\mathrm{L}^2}$ has domain $\mathrm{dom}(\mathcal{F}_{\mathrm{L}^2}) = \mathrm{L}^2(\Omega) \subsetneqq \mathrm{H}_0^{-1}(\Omega)$. Moreover, the differential $\mathrm{D}\mathcal{F}_{\mathrm{L}^2}(u) = u \in H^* = \mathrm{H}^1(\Omega)$ has the even smaller domain $\mathrm{dom}\,\big(\mathrm{D}\mathcal{F}_{\mathrm{L}^2}\big) = \mathrm{H}^1(\Omega)$.

Again we obtain the desired gradient-flow equation

$$\dot u = -\mathbb{K}^{(2)}\mathrm{D}\mathcal{F}_{\mathrm{L}^2}(u) = -\big(-\Delta\big)u = \Delta u.$$

The exact details will be made rigorous in Section 2.

Two more gradient structures will be handled in the two following subsections. They play an important role in the modeling as well as in the initiation of a new branch of mathematics, namely optimal transport for PDEs, see [Ott01, AGS05, Vil09, Pel14, DaS14, San17]. We give some more details here, because the standard parabolic equation $\dot u = \Delta u$ is most often simply called the "heat equation" but sometimes also "diffusion equation". On the level of PDEs there is no distinction, it is simply a parabolic equation. However, on the level of gradient-flow equations the distinction will become apparent.

1.4 Otto's gradient structure for diffusion

The theory of gradient systems received a major push around the year 2000 through the seminal work of Felix Otto in [Ott96, JKO98, Ott01]. It is interesting to note the title and a citation of the latter work:

> *"The geometry of dissipative evolution equations: the porous medium equation"*
>
> *p. 108: ... The merit of the right gradient flow formulation of a dissipative evolution equation is that it separates energetics and kinetics: The energetics endow the state space M with a functional E, the kinetics endow the state space with a (Riemannian) geometry via the metric tensor g.*

This work suggests the following choice of a gradient structure $(\mathrm{Prob}(\Omega), \mathcal{E}_{\text{Bolz}}, \mathbb{K}_{\text{Otto}})$:

$$M = \mathrm{Prob}(\Omega) := \big\{\, u \in \mathrm{L}^1(\Omega) \,\big|\, u \geq 0 \text{ a.e., } \textstyle\int_\Omega u \,\mathrm{d}x = 1 \,\big\} \subset \mathrm{L}^1(\Omega),$$

$$\mathcal{F}(u) = \mathcal{E}_{\text{Bolz}}(u) = \int_\Omega \lambda_{\mathrm{B}}(u(x))\,\mathrm{d}x \quad \text{where } \lambda_{\mathrm{B}}(z) = \begin{cases} z\log z - z + 1 & \text{for } z > 0, \\ 1 & \text{for } z = 0, \\ +\infty & \text{for } z < 0; \end{cases}$$

$$\mathbb{K}_{\text{Otto}}(u)\xi := -\operatorname{div}\big(u\nabla\xi\big).$$

Of course, it was known for a century that the (relative) Boltzmann entropy $\mathcal{E}_{\text{Bolz}}$ is a good Lyapunov function for the diffusion equation $\dot u = \Delta u$. However, introducing the (Riemannian-type) geometrical structure $\mathbb{K}_{\text{Otto}}$ was the key step. In these papers, and in more than one hundred follow-up papers, the geometry is called *Wasserstein geometry* because calculating the corresponding geodesic distance $d_{\mathbb{K}_{\text{Otto}}}$ one obtains the 2-Wasserstein distance W_2 on $\mathrm{Prob}(\Omega)$, see more on that in Section 4.

On the formal level we easily see that the associated gradient-flow equation is indeed the linear diffusion equation, if we use $\mathrm{D}\mathcal{E}_{\text{Bolz}}(u) = \lambda'_{\mathrm{B}}(u) = \log u$ and the classical chain rule $\nabla \log u = \frac{1}{u}\nabla u$:

$$\dot u = -\mathbb{K}_{\text{Otto}}(u)\mathrm{D}\mathcal{E}_{\text{Bolz}}(u) = -\big(-\operatorname{div}\big[u\nabla\lambda'_{\mathrm{B}}(u)\big]\big) = \operatorname{div}(u\frac{1}{u}\nabla u) = \operatorname{div}(\nabla u) = \Delta u.$$

Of course, the works [Ott96, JKO98, Ott01] and the follow-up works provide the rigorous analysis following from this choice of the gradient structure. Because of its big importance in the recent developments for diffusion equation, we define the Otto gradient (unfortunately often called Wasserstein gradient) of a general functional $\mathcal{F}(u) = \int_\Omega \big(F(u(x)) - V(x)u(x)\big)\,\mathrm{d}x$, namely

$$\mathrm{grad}_{\text{Otto}}\mathcal{F}(u) := \mathbb{K}_{\text{Otto}}(u)\mathrm{D}\mathcal{F}(u) = -\operatorname{div}\big(u\nabla\big[F'(u) - V\big]\big). \tag{1.7}$$

Clearly this choice is physically highly relevant (and can be justified in the Onsager-Machlup sense [OnM53] via fluctuation theory for diffusion, see e.g. [DaG87, AD*11, MPR14]), but it leaves the range of linear theory. The energy is nonquadratic and even enforces the positivity of u. Otto's approach to diffusion applies genuinely nonlinear methods to a linear problem, which hence opens the theory to nonlinear applications such as the porous medium equation as in [Ott01]. In particular, this new gradient structure has created a whole new branch of mathematics, namely the treatment of diffusion equations using ideas from optimal transport of probability measures, see [AGS05].

1.5 Gradient structures for the heat equation

On the level of gradient systems there is a strong distinction between the heat and the diffusion equation, which will become clear below. For diffusion a good gradient structure is Otto's gradient structure, but it is not appropriate for heat conduction.

When writing the heat equation in terms of the absolute temperature $\theta > 0$ we need the internal (or heat) energy $e = E(x,\theta)$ and the internal entropy $s = S(x,\theta)$ which are related by the Gibbs relation $E'(x,\theta) = \theta S'(x,\theta)$, where $'$ means ∂_θ. One major point is that E' is called heat capacity and it must be positive, following the intuition that for heating up a body one has to invest energy (e.g. 4.18 Joule for heating up 1 kg of water by 1 Kelvin). By Gibbs' relation also $S'(x,\theta) > 0$.

The fundamental laws of thermodynamics say that the total energy is conserved in a closed system while the total entropy increases. The heat equation reads

$$\dot e + \operatorname{div} \boldsymbol{q} = 0 \text{ in } \Omega, \qquad \boldsymbol{q}\cdot\nu = 0 \text{ on } \partial\Omega. \tag{1.8}$$

Here $e(t,x) = E(x,\theta(t,x))$ and $\boldsymbol{q}(t,x) \in \mathbb{R}^d$ denotes the heat flux that is given by Fourier's law in the form $\boldsymbol{q}(t,x) = -K(x,\theta)\nabla\theta$, where $K(x,\theta) = K(x,\theta)^* > 0$ is the heat conduction matrix (recall Onsager's symmetry) and $\nabla\theta$ is now the classical Euclidean gradient of the function $\theta(t,\cdot) : \Omega \to \mathbb{R}$. The boundary conditions $\boldsymbol{q}\cdot\nu = 0$ say that the body Ω is insulated such that heat cannot leave or enter Ω. Integrating over Ω we find conservation of total energy $t \mapsto \mathcal{E}(x,\theta(t))$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}(x,\theta(t)) = \int_\Omega \frac{\partial}{\partial t}E(x,\theta(t,x))\,\mathrm{d}x \overset{(1.8)}{=} \int_\Omega -\operatorname{div}\boldsymbol{q}\,\mathrm{d}x \overset{\text{Gauß}}{=} -\int_{\partial\Omega} \boldsymbol{q}\cdot\nu\,\mathrm{d}a = 0.$$

We can now try to generate the heat equation as a (anti-) gradient-flow equation for the total entropy

$$\mathcal{S}(\theta) = \int_\Omega S(x,\theta(x))\,\mathrm{d}x \quad \text{with } \mathrm{D}\mathcal{S}(\theta) = S'(x,\theta),$$

where "anti" stands for a functional that increases along solutions.

We now follow [Mie11d] and generalize the idea of Otto by looking for an Onsager operator $\mathbb{H}_{\text{heat}}$ in the form

$$\mathbb{K}_{\text{heat}}(\theta)\xi = -\frac{1}{E'(x,\theta)}\operatorname{div}\Big(\mathbb{A}(x,\theta)\nabla\big(\frac{\xi}{E'(x,\theta)}\big)\Big),$$

where the factor $1/E'(x,\theta)$ was introduced twice in such a way that $\mathbb{K}^*_{\text{heat}}$ is still a symmetric differential operator. This prefactor is essential to handle the term $\dot e = \partial_t(E(x,\theta(t,x))) = E'(x,\theta)\dot\theta$ in the heat equation (1.8).

With this we calculate the anti gradient-flow equation

$$\begin{aligned}\dot\theta &= +\mathbb{K}_{\text{heat}}(\theta)\mathrm{D}\mathcal{S}(\theta) = -\frac{1}{E'(x,\theta)}\operatorname{div}\Big(\mathbb{A}(x,\theta)\nabla\big(\frac{S'(x,\theta)}{E'(x,\theta)}\big)\Big)\\ &\overset{\text{Gibbs}}{=} -\frac{1}{E'(x,\theta)}\operatorname{div}\Big(\mathbb{A}(x,\theta)\nabla\big(\frac{1}{\theta}\big)\Big) \overset{*}{=} \frac{1}{E'(x,\theta)}\operatorname{div}\Big(\frac{1}{\theta^2}\,\mathbb{A}(\theta)\,\nabla\theta\Big),\end{aligned}$$

where in $\overset{*}{=}$ we used $\nabla(\frac{1}{\theta}) = -\frac{1}{\theta^2}\nabla\theta$. Thus, the abstract equation leads to the heat equation

$$\dot e = E'(x,\theta)\dot\theta = -\operatorname{div}\boldsymbol{q} = \operatorname{div}(K\nabla\theta) \quad \text{with } K(x,\theta) = \frac{1}{\theta^2}\mathbb{A}(x,\theta).$$

This approach teaches us, just by formal arguments, that $-\nabla(S'/E') = -\nabla(1/\theta)$ is the correct (nonlinear) term that drives heat conduction. This is indeed important at interfaces, where the jump of $1/\theta$ matters.

To obtain the simple linear heat equation $\dot\theta = \Delta\theta$, we can use $E(\theta) = \theta$ and $S(\theta) = \log(\theta/\theta_\circ)$ and have to choose $\mathbb{A}(\theta) = \theta^2\mathbb{I}$, i.e.

$$\mathbb{K}_{\text{heat}}(\theta)\xi = -\operatorname{div}\big(\theta^2\nabla\xi\big),$$

which is clearly different from $\mathbb{K}_{\text{Otto}}$ because of the power 2 in θ^2.

So far, the analysis for this (Riemannian) geometry has still to be developed.

1.6 Further remarks on modeling with gradient systems

A general approach to modeling with gradient systems is given in the expository work [Pel14]. In particular, it addresses the proper derivation of gradient systems from microscopic stochastic models via so-called large-deviation principles. Thus, proceeds along the path developed in [OnM53].

General development of gradient structures for semiconductor models or energy-reaction-diffusion systems, also with interfaces, can be found in [Mie11c, Mie13, GlM13].

The interplay of Hamiltonian dynamics and gradient systems can be described in term of the framework GENERIC, which is an acronym for General Equation for Non-Equilibrium Reversible Irreversible Coupling, see [GrÖ97, Ött05, Grm10, Mie11b, DPZ13]. This approach was also used to couple classical thermodynamical models to quantum systems in [MiM17, KM*19], where the interaction of the quantum system and its classical environment is modeled by a suitable Onsager operator.

2 Gradient systems with Hilbert-space structure

In this section we provide a mathematical rigorous framework for gradient systems in Hilbert spaces. By this name we do not only mean that the underlying space is a Hilbert space H, but we also use the full nice properties of the Hilbert-space geometry, i.e. we will always assume that $\mathbb{G}$ is independent of the state variable $u \in H$ and equals the Riesz isomorphism $\mathbb{I}_\mathrm{R} : H \to H^*$. Of course, this still allows us to adapt the Hilbert-space norm, if we have an equivalent norm. For example we consider the parabolic PDE

$$c(x)\dot u = \mathrm{div}\big(A(x)\nabla u\big) - \partial_u F(x, u(x)) \quad \text{in } \Omega, \qquad u = 0 \quad \text{on } \partial\Omega,$$

where $c \in \mathrm{L}^\infty(\Omega)$ with $c(x) \geq c_0 > 0$ a.e. and suitable A and F. Then we can choose the gradient structure $(H, \mathcal{F}, \mathbb{G})$ with

$$H = \mathrm{L}^2(\Omega), \quad \mathbb{G}v = \mathbb{I}_{\text{Riesz}} v = cv, \quad \mathcal{F}(u) = \int_\Omega \frac{1}{2}\nabla u(x) \cdot A(x)\nabla u(x) + F(x, u(x))\big)\,\mathrm{d}x$$

for $u \in \mathrm{H}^1_0(\Omega)$ and $+\infty$ otherwise on $\mathrm{L}^2(\Omega)$. Here $\mathrm{H}^1(\Omega)$ is the Sobolev space of functions with square integrable gradient, and $\mathrm{H}^1_0(\Omega)$ is the closed subspace obtained by closing $\mathrm{C}^\infty_\mathrm{c}(\Omega)$ in $\mathrm{H}^1(\Omega)$.

2.1 Differentials and subdifferentials on Banach spaces

For PDEs it is essential to have a suitable notion of differential, because of two important facts:
• (even quadratic) functionals and their differentials need to be defined on dense subsets
• nonsmoothness is important in applications (contact, Coulomb friction, plasticity, ...)

We are now working on general Banach spaces X with dual spaces X^* and dual pairing ${}_{X^*}\langle \cdot, \cdot \rangle_X$. In particular, we avoid the identification $H \sim H^*$ in Banach spaces. As Gâteaux and Fréchet differentials are only useful for continuous functions, we directly define so-called subdifferentials, which are *set-valued mappings*. For a mapping $A : X \to 2^Y = \mathfrak{P}(Y)$ we shortly write $A : X \rightrightarrows Y$, i.e. for all $u \in X$ we have $A(u) \subset Y$, where $A(u) = \emptyset$ is of course possible.

Here we develop a theory in the spirit of Brézis' foundational work, see in particular the existence result in [Bré73, Thm. 3.6, p. 72]. However, the approach there is completely different, because it is based on Yosida regularizations for maximal monotone operators whereas we use time-incremental minimization for gradient systems. Our approach can be adapted easily to Banach spaces and metric spaces.

Definition 2.1 (Subdifferentials) *Let $\mathcal{F} : X \to \mathbb{R}_\infty := \mathbb{R} \cup \{+\infty\}$ be a functional. The* (convex) subdifferential $\partial\mathcal{F} : X \rightrightarrows X^*$ *is defined via* $\partial\mathcal{F}(u) = \emptyset$ *for* $\mathcal{F}(u) = \infty$ *and*

$$\partial\mathcal{F}(u) := \big\{\, \xi \in X^* \,\big|\, \forall\, w \in X : \ \mathcal{F}(w) \geq \mathcal{F}(u) + \langle \xi, w{-}u\rangle_X \,\big\} \subset X^*$$

otherwise. The Fréchet subdifferential $\partial^{\mathrm{F}}\mathcal{F} : X \rightrightarrows X^*$ *is defined via* $\partial^{\mathrm{F}}\mathcal{F}(u) = \emptyset$ *for* $\mathcal{F}(u) = \infty$ *and*

$$\partial^{\mathrm{F}}\mathcal{F}(u) := \big\{\, \xi \in X^* \,\big|\, \mathcal{F}(w) \geq \mathcal{F}(u) + \langle \xi, w{-}u\rangle_X + o(\|w{-}u\|_X) \text{ for } w \to u \,\big\} \subset X^*$$

otherwise. The domains of $\mathcal{F}$*,* $\partial\mathcal{F}$*, and* $\partial^{\mathrm{F}}\mathcal{F}$ *are the subsets of* X *defined via*

$$\begin{aligned} \mathrm{dom}(\mathcal{F}) &= \big\{\, u \in X \,\big|\, \mathcal{F}(u) < \infty \,\big\}, \quad \mathrm{dom}(\partial\mathcal{F}) = \big\{\, u \in X \,\big|\, \partial\mathcal{F}(u) \neq \emptyset \,\big\}, \\ \mathrm{dom}(\partial^{\mathrm{F}}\mathcal{F}) &= \big\{\, u \in X \,\big|\, \partial^{\mathrm{F}}\mathcal{F}(u) \neq \emptyset \,\big\}. \end{aligned}$$

By the definition, we clearly have $\partial\mathcal{F}(u) \subset \partial^{\mathrm{F}}\mathcal{F}(u)$.

Exercise 2.1 *Consider* $X = \mathbb{R}$ *and the following functions:*

$$\mathcal{F}_1(u) = \frac{1}{4}(u^2{-}1)^2, \quad \mathcal{F}_2(u) = -|u| + u^2, \quad \mathcal{F}_3(u) = \min\{0, |u|{-}1, \frac{1}{2}u^2 - 1\}.$$

Calculate $\partial\mathcal{F}$ *and* $\partial^{\mathrm{F}}\mathcal{F}$ *for all three cases.*

Exercise 2.2 *Let* Ω *be a smooth bounded domain in* $\mathbb{R}^d$*.*
(A) As an example we consider the quadratic functional $\mathcal{F} : \mathrm{L}^2(\Omega) \to \mathbb{R}_\infty$ *with*

$$\mathcal{F}(u) = \int_\Omega \frac{1}{2}|\nabla u(x)|^2 \,\mathrm{d}x \quad \text{on } \mathrm{dom}(\mathcal{F}) = \mathrm{H}^1(\Omega).$$

Show that $\mathrm{dom}(\partial\mathcal{F}) = \big\{\, u \in \mathrm{H}^1(\Omega) \big| \Delta u \in \mathrm{L}^2(\Omega),\ \nabla u{\cdot}\nu = 0 \text{ on } \partial\Omega \,\big\}$ *and* $\partial\mathcal{F}(u) = \{-\Delta u\} \subset \mathrm{L}^2(\Omega)$.
(B) Consider exponents p *and* q *with* $1 < p < q$ *and let*

$$X = \mathrm{L}^p(\Omega) \quad \text{and} \quad \mathcal{F}(u) = \int_\Omega \frac{1}{q}|u(x)|^q \,\mathrm{d}x.$$

Calculate $\mathrm{dom}(\mathcal{F})$ *and the differentials* $\partial\mathcal{F}$ *and* $\partial^{\mathrm{F}}\mathcal{F}$*.*

The important property of the Fréchet subdifferential is that there is a *sum rule*. Similar sum rules play an important role many areas of applied analysis: calculus of variations, optimization, abstract evolution equations, and of course in the theory of gradient systems.

Proposition 2.2 (Sum rule for subdifferentials) *If* $\mathcal{F}_1 : X \to \mathbb{R}_\infty$ *is convex and* $\mathcal{F}_2 : X \to \mathbb{R}$ *is Fréchet differentiable (i.e. for all* $u \in X$ *we have* $\mathcal{F}_2(u{+}v) - \mathcal{F}_2(u) - \langle \mathrm{D}\mathcal{F}_2(u), v\rangle = o(\|v\|)$ *for* $v \to 0$*), then*

$$\partial^{\mathrm{F}}\mathcal{F}(u) = \mathrm{D}\mathcal{F}_2(u) + \partial\mathcal{F}_1(u) \subset X^*.$$

Proof. As $\mathcal{F}_2(u) \in \mathbb{R}$ for all $u \in X$ we have $\mathrm{dom}(\mathcal{F}) = \mathrm{dom}(\mathcal{F}_1)$.

Now consider $\xi \in \partial\mathcal{F}_1(u)$. Then, for $w \in X$ we have

$$\begin{aligned}\mathcal{F}(w) &= \mathcal{F}_1(w) + \mathcal{F}_2(w) \geq \mathcal{F}_1(u) + \langle \xi, w{-}u\rangle + \mathrm{D}\mathcal{F}_2(u)[w{-}u] + o(\|w{-}u\|_X)\\ &= \mathcal{F}(u) + \langle \xi{+}\mathrm{D}\mathcal{F}_2(u), w{-}u\rangle + o(\|w{-}u\|_X),\end{aligned}$$

which means that $\mathrm{D}\mathcal{F}_2(u) + \partial\mathcal{F}_1(u) \subset \partial^{\mathrm{F}}\mathcal{F}(u)$.

For the opposite inclusion we assume $\eta \in \partial^{\mathrm{F}}\mathcal{F}(u)$ and obtain

$$\begin{aligned}\mathcal{F}_1(w) &= \mathcal{F}(w) - \mathcal{F}_2(w)\\ &\geq \mathcal{F}(u) + \langle \eta, w{-}u\rangle + o(\|w{-}u\|) - \mathcal{F}_2(u) - \langle \mathrm{D}\mathcal{F}_2(u), w{-}u\rangle + o(\|w{-}u\|)\\ &= \mathcal{F}_1(u) + \big\langle \eta{-}\mathrm{D}\mathcal{F}_2(u), w{-}u\big\rangle + o(\|w{-}u\|).\end{aligned}$$

By convexity of $\mathcal{F}_1$ we have $\mathcal{F}_1(w_\theta) \leq (1{-}\theta)\mathcal{F}_1(w) + \theta\mathcal{F}_1(u)$, where $w_\theta = (1{-}\theta)w + \theta u$, and conclude (by setting $w = w_\theta$ in the above estimate)

$$\begin{aligned}\mathcal{F}_1(w) &\geq \frac{1}{1{-}\theta}\big(\mathcal{F}_1(w_\theta) - \theta\mathcal{F}_1(u)\big)\\ &\overset{\text{above}}{\geq} \frac{1}{1{-}\theta}\big(\mathcal{F}_1(u) + \big\langle \eta{-}\mathrm{D}\mathcal{F}_2(u), w_\theta{-}u\big\rangle + o(\|w_\theta{-}u\|) - \theta\mathcal{F}_1(u)\big)\\ &= \mathcal{F}_1(u) + \big\langle \eta{-}\mathrm{D}\mathcal{F}_2(u), w{-}u\big\rangle + o((1{-}\theta)\|w{-}u\|)\\ &\overset{\theta\to1^-}{\longrightarrow} \mathcal{F}_1(u) + \big\langle \eta{-}\mathrm{D}\mathcal{F}_2(u), w{-}u\big\rangle.\end{aligned}$$

Thus, we conclude $\eta{-}\mathrm{D}\mathcal{F}_2(u) \in \partial\mathcal{F}_1(u)$ which means $\partial^{\mathrm{F}}\mathcal{F}(u) \subset \mathrm{D}\mathcal{F}_2(u){+}\partial\mathcal{F}_1(u)$. ■

Exercise 2.3 (Convex subdifferentials) *Consider a reflexive Banach space X and a functional $\mathcal{F} : X \to \mathbb{R}_\infty$ that is proper, lower semicontinuous, and convex.*

(A) For $\xi \in X^*$ define the functional $\mathcal{G}_\xi : u \mapsto \mathcal{F}(u) - \langle \xi, u\rangle$. Show the sum rule $\partial\mathcal{G}_\xi(u) = -\xi + \partial\mathcal{F}(u)$.

(B) Assume additionally that $\mathcal{F}$ is superlinear, i.e. $\mathcal{F}(u)/(1{+}\|u\|) \to \infty$ for $\|u\| \to \infty$. Show that the subdifferential $\partial\mathcal{F} : X \rightrightarrows X^*$ is surjective, i.e. for each $\xi \in X^*$ there exists $u_\xi \in X$ such that $\xi \in \partial\mathcal{F}(u_\xi)$. (Hint: Minimize a suitable functional.)

2.2 Semiconvexity and closedness of subdifferentials

An important class of functionals will be the following one.

Definition 2.3 (Semiconvexity) *A function $\mathcal{F} : X \to \mathbb{R}_\infty$ is called* λ-convex, *if*

$$\begin{aligned}&\forall\, u_0, u_1 \in X\ \forall\, \theta \in [0,1]:\\ &\qquad \mathcal{F}\big((1{-}\theta)u_0{+}\theta u_1\big) \leq (1{-}\theta)\mathcal{F}(u_0) + \theta\mathcal{F}(u_1) - \frac{\lambda}{2}\theta(1{-}\theta)\|u_1{-}u_0\|_X^2.\end{aligned} \tag{2.1}$$

We simply say that $\mathcal{F}$ is semiconvex *if there exists $\lambda \in \mathbb{R}$ such that $\mathcal{F}$ is λ-convex.*

We will often use the notion of sublevels of $\mathcal{F}$, namely $S_E^{\mathcal{F}} := \{ u \in H \,|\, \mathcal{F}(u) \leq E \}$. It is a classical fact that $\mathcal{F}$ is (weakly) lower semicontinuous if and only if for all $E \in \mathbb{R}$ the sublevels $S_E^{\mathcal{F}}$ are (weakly) closed. (For that reason, in some papers and books, lsc functionals are simply called 'closed'.)

Exercise 2.4 (Convex hulls of sublevels of semiconvex functionals) *Assume that $\mathcal{F} : X \to \mathbb{R}_\infty$ is λ-convex.*

(A) Show that in the case $\lambda \geq 0$ the sublevels $S_E^{\mathcal{F}}$ are convex.

(B) Give an example where $\mathcal{F}$ is (-1)-convex and $S_E^{\mathcal{F}}$ is nonconvex for some $E \in \mathbb{R}$.

(C) Consider a subset A of X such that $A \subset B_R(0) \cap S_E^{\mathcal{F}}$. Show that the convex hull $\mathrm{co}(A)$ lies in $S_{\widetilde{E}}^{\mathcal{F}}$ for a suitable $\widetilde{E}$ depending on λ and R.

Two of the fundamental properties of semiconvex functionals are a simple global characterization of the Fréchet subdifferential and the so-called closedness of the graph of $\partial^{\mathrm{F}}\mathcal{F}$.

Lemma 2.4 (Characterization of Fréchet subdifferential) *Assume that $\mathcal{F} : X \to \mathbb{R}_\infty$ is λ-convex, then the Fréchet subdifferential admits the following global representation: For all $u \in \mathrm{dom}(\mathcal{F})$ we have*

$$\partial^{\mathrm{F}}\mathcal{F}(u) = \Big\{ \xi \in X^* \,\Big|\, \forall\, w \in X:\ \mathcal{F}(w) \geq \mathcal{F}(u) + \langle \xi, w{-}u\rangle + \frac{\lambda}{2}\|w{-}u\|_X^2 \Big\} \tag{2.2}$$

Proof. Set $\boldsymbol{A}(u)$ for the right-hand side in (2.2). As $\|w{-}u\|^2 = o(\|w{-}u\|)$ we immediately have $\boldsymbol{A}(u) \subset \partial^{\mathrm{F}}\mathcal{F}(u)$.

For the opposite inclusion consider $\xi \in \partial^{\mathrm{F}}\mathcal{F}(u)$ and arbitrary $x \in X$. By λ-convexity we have, with $w_\theta = (1{-}\theta)w + \theta u$,

$$\begin{aligned}
\mathcal{F}(w) &\geq \frac{1}{1{-}\theta}\big(\mathcal{F}(w_\theta) - \theta\mathcal{F}(u) + \frac{\lambda}{2}\theta(1{-}\theta)\|w{-}u\|_X^2\big)\\
&\geq \frac{1}{1{-}\theta}\big(\mathcal{F}(u) + \big\langle \xi, w_\theta{-}u\big\rangle + o(\|w_\theta{-}u\|) - \theta\mathcal{F}(u)\big) + \frac{\lambda}{2}\,\theta\|w{-}u\|_X^2\\
&= \mathcal{F}(u) + \langle \xi, w{-}u\rangle + \frac{o((1{-}\theta)\|w{-}u\|)}{1-\theta} + \frac{\lambda}{2}\,\theta\|w{-}u\|_X^2.
\end{aligned}$$

Taking the limit $\theta \to 1^-$ we obtain $\xi \in \boldsymbol{A}(u)$ and conclude $\partial^{\mathrm{F}}\mathcal{F}(u) \subset \boldsymbol{A}(u)$ as desired. ∎

While the above lemma can be seen as a technical tool, the following closedness property is essential for showing existence of solutions via limiting processes. This condition parallels the important concept of "closedness of a graph of a linear operator" (recall the closed-graph theorem).

Definition 2.5 (Closedness of the differential) *A set-valued mapping $\boldsymbol{A} : X \rightrightarrows Y$ is called* (strong-weak) closed *if*

$$\left.\begin{array}{c} u_n \to u \text{ in } X,\ y_n \rightharpoonup y \text{ in } Y \\ y_n \in \boldsymbol{A}(u_n) \end{array}\right\} \implies y \in \boldsymbol{A}(u).$$

Let $\overline{\partial}\mathcal{F} : X \rightrightarrows X^*$ by any (sub-) differential of $\mathcal{F} : X \to \mathbb{R}_\infty$, then $\overline{\partial}\mathcal{F}$ is called (strong-weak) energy closed *(in short E-closed) if*

$$\left.\begin{array}{c} u_n \to u \text{ in } X, \quad \xi_n \rightharpoonup \xi \text{ in } X^* \\ \sup_{n\in\mathbb{N}} \mathcal{F}(u_n) < \infty,\ \ \xi_n \in \overline{\partial}\mathcal{F}(u_n) \end{array}\right\} \implies \xi \in \overline{\partial}\mathcal{F}(u).$$

One can also define (α, β)-closedness for $\alpha, \beta \in \{\text{weak}, \text{strong}\}$. In particular, for quadratic functionals with $\partial^{\mathrm{F}}\mathcal{F}(u) = \{\mathbb{A}u\}$ it may be relevant to define (weak,weak) closedness.

Exercise 2.5 (Closedness) *Consider $X = \mathrm{L}^2(\Omega)$ with $\Omega =]0, \ell[\subset \mathbb{R}^1$ and the functional $\mathcal{F}(u) = \int_\Omega \big(\frac{2}{3}|u(x)|^{3/2} + \frac{1}{2}u(x)^2\big)\,\mathrm{d}x$.*

Show that $\mathcal{F}$ is strong-weak closed but not weak-weak closed.

Obviously, in general a subdifferential is not closed, simply consider $\mathcal{F} : \mathbb{R} \to \mathbb{R};\ u \mapsto \frac{1}{2}u^2 - |u|$ then $\partial^{\mathrm{F}}F(0) = \emptyset$ while $\partial^{\mathrm{F}}\mathcal{F}(u) = u - \mathrm{sign}(u)$ for $u \neq 0$. Hence $\xi_n = 1/n - 1 \in \partial^{\mathrm{F}}\mathcal{F}(1/n)$ and $(u_n, \xi_n) = (\frac{1}{n}, \frac{1}{n}{-}1) \to (0, -1) = (u_*, \xi_*)$, but $\xi_* = -1 \notin \partial^{\mathrm{F}}\mathcal{F}(u_*)$.

However, the situation is much better for semiconvex functionals, where we can take advantage of the global characterization of the Fréchet subdifferential.

Proposition 2.6 (Closedness of $\partial^{\mathrm{F}}\mathcal{F}$) *If $\mathcal{F} : X \to \mathbb{R}_\infty$ is proper, lower semicontinuous, and semiconvex, then its Fréchet subdifferential $\partial^{\mathrm{F}}\mathcal{F}$ is strong-weak [energy???] closed.*

Proof. We consider sequences $(u_n)_n$ in X and $(\xi_n)_n$ in X^* satisfying the properties in the definition of weak-strong closedness. Using the global characterization of Lemma 2.4 we have, for all $n \in \mathbb{N}$ and all $w \in X$, the estimate

$$\mathcal{F}(w) \geq \mathcal{F}(u_n) + \langle \xi_n, w{-}u_n\rangle + \frac{\lambda}{2}\|w{-}u_n\|^2.$$

In this identity we can pass to the limit $n \to \infty$ using strong lsc of $\mathcal{F}$, the weak-strong continuity of the duality product $(v, \eta) \mapsto \langle \eta, v\rangle$, and the strong continuity of the norm. Thus, we find $\mathcal{F}(w) \geq \mathcal{F}(u) + \langle \xi, w{-}u\rangle + \frac{\lambda}{2}\|w{-}u\|^2$.

As $\mathcal{F}$ is proper, we conclude $\mathcal{F}(u) < \infty$, and applying the global characterization (2.2) gives $\xi \in \partial^{\mathrm{F}}\mathcal{F}(u)$ as desired. ∎

Of course, this result is only one of the easy results and there are many other possibilities for establishing closedness of subdifferentials.

2.3 Existence via time-incremental minimization

One of the most versatile methods of showing existence results for evolutionary problems is that of time discretization. Fixing a time horizon $T > 0$ (which will be completely arbitrary here) we choose $N \in \mathbb{N}$ and define the (constant) time step $\tau = T/N > 0$. One of the main advantages of treating gradient-flow equations is that the time-incremental problem can be formulated as a minimization problem. Thus, we are speaking about *time-incremental minimization* or the *minimizing-movement scheme*.

Given a gradient system $(H, \mathcal{F}, \mathbb{I}_{\mathrm{R}})$ on a Hilbert space H and an initial condition $u^0 \in H$, the aim is to find a solution $u \in \mathrm{W}^{1,1}([0,T]; H)$ such that

$$\mathbb{I}_{\mathrm{R}}\dot{u}(t) \in -\partial^{\mathrm{F}}\mathcal{F}(u(t)) \quad \text{for a.a. } t \in [0,T], \qquad u(0) = u^0. \tag{2.3}$$

Note that for Hilbert spaces we have $\mathrm{W}^{1,1}([0,T]; H) = \mathrm{AC}([0,T]; H) \subset \mathrm{C}^0([0,T]; H)$ such that posing the initial condition $u(0) = u^0$ is well defined.

The backward Euler time-discretization (fully implicit) is defined via

$$0 \in \mathbb{I}_{\mathrm{R}} \frac{1}{\tau}(u_k - u_{k-1}) + \partial^{\mathrm{F}} \mathcal{F}(u_k) \quad \text{in } H^*. \tag{2.4}$$

Here $u_0 = u^0$ is the initial condition and u_k is to be found incrementally for $k = 1, \ldots, N$. Recalling that the functional

$$u \mapsto \frac{1}{2\tau} \|u - u_{k-1}\|^2 = \frac{1}{2\tau} \big\langle \mathbb{I}_{\mathrm{R}}(u - u_{k-1}), u - u_{k-1} \big\rangle$$

is Fréchet differentiable with derivative $\mathbb{I}_{\mathrm{R}} \frac{1}{\tau}(u - u_{k-1})$ we see that (2.4) is the Euler-Lagrange equation for the following

> **time-incremental minimization scheme** for $\tau = T/N$:
> set $u_0 = u^0$;
> for $k = 1, \ldots, N$ find u_k as a minimizer of the functional
> $$u \ \mapsto \ \Phi_\tau^{\mathcal{F}}(u_{k-1}; u) := \frac{1}{2\tau} \|u - u_{k-1}\|_H^2 + \mathcal{F}(u). \tag{2.5}$$

In the case that $\mathcal{F}$ is lower semicontinuous and λ-convex on H we easily see that $\Phi_\tau^{\mathcal{F}}(w, \cdot)$ is $(\lambda + \frac{1}{\tau})$-convex. Hence, for sufficiently small $\tau > 0$, the minimizer u_k is unique and minimizing $\Phi_\tau^{\mathcal{F}}(u_{k-1}; \cdot)$ is equivalent to solving the Euler scheme (2.4).

Based on the discrete solution $(u_k)_{k=0,\ldots,N}$ we are now able to define the *piecewise affine interpolant* $\widehat{u}_\tau \in \mathrm{C}^0([0,T]; H)$ and the *piecewise constant interpolant* $\overline{u}_\tau \in \mathrm{L}^\infty([0,T]; H)$ as follows:

$$\widehat{u}_\tau((k+\theta-1)\tau) = (1-\theta)u_{k-1} + \theta u_k \quad \text{for } k \in \{1, \ldots, N\} \text{ and } \theta \in [0,1],$$
$$\overline{u}_\tau(0) = u^0 \text{ and } \overline{u}_\tau(t) = u_k \text{ for } t \in \,](k-1)\tau, k\tau] \text{ and } k \in \{1, \ldots, N\}.$$

These two interpolants are constructed in such a way that the discrete equation (2.4) leads to the relation in the evolutionary form

$$0 \ \in \ \mathbb{I}_{\mathrm{R}} \dot{\widehat{u}}_\tau(t) \ + \ \partial^{\mathrm{F}} \mathcal{F}(\overline{u}_\tau(t)) \quad \text{for all } t \in [0,T] \setminus \{\, k\tau \mid k = 0, ..., N \,\}. \tag{2.6}$$

On each open subinterval $](k-1)\tau, k\tau[$ the terms on the right-hand side are constant and equal the terms in (2.4).

The following theorem shows that in the limit $\tau \to 0^+$ we indeed obtain convergence to a limiting function u and this function indeed is a solution of the gradient-flow equation (2.3). Thus, the following result is not only an existence result, but it is also a convergence result for the time-incremental minimization scheme.

Theorem 2.7 (Existence of a gradient flow for $(H, \mathcal{F}, \mathbb{I}_{\mathrm{R}})$) *Consider the gradient system $(H, \mathcal{F}, \mathbb{I}_{\mathrm{R}})$ where H is a Hilbert space with Riesz isomorphism $\mathbb{I}_{\mathrm{R}}$ and $\mathcal{F} : H \to \mathbb{R}_\infty$ is proper, lower semicontinuous, λ-convex for some $\lambda \in \mathbb{R}$, and has compact sublevels, i.e. for all $E \in \mathbb{R}$ the sets $S_E^{\mathcal{F}} := \big\{\, u \in H \,\big|\, \mathcal{F}(u) \leq E \,\big\}$ are compact in H.*

Then for all $u^0 \in \mathrm{dom}(\mathcal{F})$ the solutions $\widehat{u}_\tau : [0,T] \to H$ obtained from the incremental minimizing scheme converge to the unique solution $u \in \mathrm{W}^{1,2}([0,T]; H)$ of (2.3), i.e.

$$\forall\, t \in [0,T]: \quad \widehat{u}_\tau(t) \to u(t) \ \text{ in } H.$$

Moreover, for any two solutions u_0 and u_1 we have the λ-contractivity estimate

$$\|u_1(t) - u_0(t)\|_H \leq \mathrm{e}^{-\lambda(t-s)} \|u_1(s) - u_0(s)\|_H \quad \text{for } 0 \leq s < t. \tag{2.7}$$

The proof will be given in the next subsection.

Example 2.8 (Nonsmooth energy in $\mathbb{R}^2$) *We consider $\mathbb{R}^2$ equipped with the Hilbert space norm $\|v\|^2 = v_1^2/a + v_2^2/b$, i.e. $\mathbb{K} = \begin{pmatrix} a & 0 \\ 0 & b\end{pmatrix}$. Moreover, we consider the nonsmooth functional $\mathcal{F}(u) = \max\{|u_1|, |u_2|\}$. Clearly, $\mathcal{F}$ is convex but nonsmooth. The subdifferential is a singleton for points not lying on the two diagonals $u_1 = \pm u_2$:*

$$\partial\mathcal{F}(u) = \partial^{\mathrm{F}}\mathcal{F}(u) = \begin{cases} \{\begin{pmatrix}\mathrm{sign}(u_1)\\ 0\end{pmatrix}\} & \text{for } 0<|u_2|<|u_1|, \\ \{\begin{pmatrix}0\\ \mathrm{sign}(u_2)\end{pmatrix}\} & \text{for } 0<|u_1|<|u_2|, \\ \big\{ \mathrm{sign}(u_1)\begin{pmatrix}\theta\\ 1{-}\theta\end{pmatrix} \,\big|\, \theta\in[0,1]\big\} & \text{for } 0\neq u_1=u_2, \\ \big\{ \mathrm{sign}(u_2)\begin{pmatrix}-\theta\\ 1{-}\theta\end{pmatrix} \,\big|\, \theta\in[0,1]\big\} & \text{for } 0\neq u_1=-u_2, \\ \big\{ \begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix} \,\big|\, |\xi_1|{+}|\xi_2|\leq 1\big\} & \text{for } 0=u_1=u_2. \end{cases}$$

As $\mathcal{F}$ is convex, we know that the GFE has exactly one solution for each initial condition.

We now piece together the solutions of the GFE $\dot u \in \mathbb{K}\partial\mathcal{F}(u)$. Without loss of generality we start in u^0 in the triangle $u_1^0 > u_2^0 > 0$. As long as the solution stays in this triangle the subdifferential $\partial\mathcal{F}(u)$ is the singleton $(1,0)^\top$. Hence, we have the velocity $\dot u = -(a,0)^\top$, i.e.

$$u(t) = u^0 - \begin{pmatrix} at \\ 0\end{pmatrix} \quad \text{for } t\in[0,t_1] \text{ with } t_1 := (u_1^0{-}u_2^0)/a.$$

At $t = t_1$ the solution has reached the ray $u_1 = u_2 > 0$, and it has to stay there, i.e.

$$\dot u = \alpha\begin{pmatrix}1\\1\end{pmatrix} \quad \text{and} \quad 0 \in \alpha\begin{pmatrix}1\\1\end{pmatrix} + \begin{pmatrix} a & 0\\ 0 & b\end{pmatrix}\begin{pmatrix}\theta\\ 1{-}\theta\end{pmatrix}.$$

Thus, we find $\theta = b/(a{+}b)$ and $\alpha = -ab/(a{+}b)$ which gives

$$u(t) = \big(u_2^0 - \frac{ab}{a{+}b}(t{-}t_1)\big)\begin{pmatrix}1\\1\end{pmatrix} \text{ for } t_1\leq t\leq t_2 = t_1 + \frac{a{+}b}{ab}u_2^0 \text{ and } u(t)=0 \text{ for } t\geq t_2.$$

Clearly this provides a solution, and by uniqueness it is the only solution.

We emphasize that for this system our existence theory implies the existence of a contractive semiflow, i.e. uniqueness for positive times and Lipschitz continuous dependence on the initial data.

However, this example does not admit any uniqueness or Lipschitz continuity backward in time. Indeed, all solutions starting in the ball $B_R(0)$ reach $u = 0$ in a finite time $t_R > 0$ and then satisfy $u(t) = 0$ for $t \geq t_R$.

Example 2.9 (Allen-Cahn equation) *Here we want to show that the result applies to the Allen-Cahn equation* (1.5) *(also called Chafee-Infante equation) the GFE for $(\mathrm{L}^2(\Omega), \mathcal{F}_{\mathrm{AC}}, m\mathbb{I}_{\mathrm{R}})$, where m is a positive constant.*

Let $\Omega \subset \mathbb{R}^d$ be a smooth bounded domain and $d \leq 3$ such that $\mathrm{H}^1_0(\Omega) \subset \mathrm{L}^6(\Omega)$. Then, using $\alpha, \beta > 0$, it is standard to see that the functional

$$\mathcal{F}_{\mathrm{AC}}(u) = \begin{cases} \int_\Omega \big(\frac{\alpha}{2}|\nabla u|^2 + \frac{\beta}{4}(u^2{-}1)^2\big)\,\mathrm{d}x & \text{for } u\in \mathrm{H}^1_0(\Omega), \\ +\infty & \text{for } u\in \mathrm{L}^2(\Omega)\setminus \mathrm{H}^1_0(\Omega) \end{cases}$$

has domain $\mathrm{dom}(\mathcal{F}_{\mathrm{AC}}) = \mathrm{H}^1_0(\Omega)$ *and Fréchet subdifferential* $\partial^{\mathrm{F}}\mathcal{F}_{\mathrm{AC}}$ *given by*

$$\partial^{\mathrm{F}}\mathcal{F}_{\mathrm{AC}}(u) = \begin{cases} \big\{-\alpha\Delta u - \beta(u{-}u^3)\big\} & \text{for } u \in \mathrm{dom}(\partial^{\mathrm{F}}\mathcal{F}_{\mathrm{AC}}) := \mathrm{H}^2(\Omega)\cap \mathrm{H}^1_0(\Omega), \\ \emptyset & \text{for } u \in \mathrm{L}^2(\Omega)\setminus \mathrm{dom}(\partial^{\mathrm{F}}\mathcal{F}_{\mathrm{AC}}). \end{cases}$$

Moreover, we see that $u \mapsto \mathcal{F}_{\mathrm{AC}}(u) + \frac{\lambda m}{2}\|u{-}u_0\|^2_H$ *is convex if and only if* $\lambda m \geq \beta$. *To see this one uses that* $u \mapsto \frac{\alpha}{2}\|\nabla u\|^2$ *is quadratic and non-negative, and hence convex. Moreover,* $z \mapsto \frac{\beta}{4}(z^2{-}1)^2 + \frac{\lambda m}{2}(z{-}z_0)^2$ *is convex if and only if* $\lambda m \geq \beta$. *Thus, we conclude that* $\mathcal{F}_{\mathrm{AC}}$ *is* $(-\beta/m)$-*convex.*

Thus, we conclude existence of solutions for the Allen-Cahn equation for all initial values $u^0 \in \mathrm{H}^1(\Omega)$ *and obtain Lipschitz-continuous dependence of the solution on the initial data in the sense that*

$$\|u_{\mathrm{AC}}(t) - \widetilde{u}_{\mathrm{AC}}(t)\|_{\mathrm{L}^2} \leq \mathrm{e}^{\beta(t-s)/m}\,\|u_{\mathrm{AC}}(s) - \widetilde{u}_{\mathrm{AC}}(s)\|_{\mathrm{L}^2} \quad \text{for } 0 \leq s < t.$$

2.4 The first convergence proof

The following proof consists of the classical steps for most constructions of the solutions of PDEs. We give the steps in some detail to prepare for the more advanced cases.

Step 0: construction of approximations (here via time discretization),

Step 1: a priori estimates,

Step 2: extraction of convergent subsequences,

Step 3: identification of the equation,

Step 4: uniqueness and convergence of the full sequence.

In particular, we will essentially rely on the gradient structure in two points, namely in (1) by doing energy estimates, in (3) when using the closedness of subdifferentials, and in (4) when using semi-convexity. Of course, very similar steps will appear in later sections.

Proof of Theorem 2.7. We follow the above five steps.

Step 0: Approximants via time discretization. The time discretization with time step $\tau = T/N$ with $N \in \mathbb{N}$ is described above leading to the time-incremental minimization scheme (2.5). We have existence of minimizers because $\mathcal{F}$ is lower semicontinuous and bounded from below by $F_{\min} = \min_{u\in H}\mathcal{F}(u)$. Here we used that $F_{\min} = \inf_{u\in H}\mathcal{F}(u)$ is attained by the one-sided Weierstraß extremal principle exploiting the compactness of the sublevels of $\mathcal{F}$. Similarly u_k as minimizer of $\Phi^{\mathcal{F}}_\tau(u_{k-1};\cdot)$ exists.

Step 1: A priori estimates. As u_k is a minimizer of $\Phi^{\mathcal{F}}_\tau(u_{k-1};\cdot)$ we have

$$\frac{1}{2\tau}\|u_k{-}u_{k-1}\|^2 + \mathcal{F}(u_k) = \Phi^{\mathcal{F}}_\tau(u_{k-1};u_k) \leq \Phi^{\mathcal{F}}_\tau(u_{k-1};u_{k-1}) = \mathcal{F}(u_{k-1}).$$

From this we immediately obtain

$$\begin{aligned} &F_{\min} \leq \mathcal{F}(u_k) \leq \mathcal{F}(u^0) < \infty \text{ for } k \in \{0,\dots,N\} \quad \text{and} \\ &\int_0^T \|\dot{\widehat{u}}_\tau(t)\|^2\,\mathrm{d}t = \sum_{k=1}^N \tau\,\Big\|\frac{1}{\tau}(u_k{-}u_{k-1})\Big\|^2 \leq 2\big(\mathcal{F}(u^0) - F_{\min}\big). \end{aligned} \tag{2.8}$$

The second estimate follows by adding up the incremental estimate for $k = 1, \ldots, N$.

Step 2: Extraction of subsequences. As the sequence $\widehat{u}_\tau$ is bounded in $\mathrm{W}^{1,2}([0,T];H)$ we can extract a subsequence (not relabeled) such that

$$\widehat{u}_\tau \rightharpoonup u \text{ in } \mathrm{L}^2([0,T];H) \quad \text{and} \quad \dot{\widehat{u}}_\tau \rightharpoonup \dot{u} \text{ in } \mathrm{L}^2([0,T];H)$$

for a limit $u \in \mathrm{W}^{1,2}([0,T];H)$.

Moreover, for all $t \in [0,T]$ and $\tau = T/N$ the values $\widehat{u}_\tau(t)$ lie in the compact sublevel $S^{\mathcal{F}}_{\mathcal{F}(u^0)}$. Together with the equi-continuity

$$\|\widehat{u}_\tau(t) - \widehat{u}_\tau(s)\| \leq |t-s|^{1/2} \|\dot{\widehat{u}}_\tau\|_{\mathrm{L}^2([0,T];H)} \leq |t-s|^{1/2} \big(2(\mathcal{F}(u^0) - F_{\min})\big)^{1/2}$$

we can apply the Arzelà-Ascoli theorem and find, after extracting a further sequence (not relabeled), the uniform convergences

$$\widehat{u}_\tau \to u \text{ in } \mathrm{C}^0([0,T];H) \quad \text{and} \quad \overline{u}_\tau \to u \text{ in } \mathrm{L}^\infty([0,T];H).$$

For the second convergence we observe $\widehat{u}_\tau(k\tau) = \overline{u}_\tau(k\tau)$ and that $\overline{u}_\tau$ is piecewise constant. Hence we have $\|\widehat{u}_\tau - \overline{u}_\tau\|_{\mathrm{L}^\infty([0,T];H)} \leq \sqrt{\tau}\, \big(2(\mathcal{F}(u^0) - F_{\min})\big)^{1/2} \to 0$ for $\tau \to 0^+$.

Step 3: Identification of equation. To show that the limit u satisfies the gradient-flow equation we define

$$\xi_\tau = -\mathbb{I}_{\mathrm{R}} \dot{\widehat{u}}_\tau \in \mathrm{L}^2([0,T];H^*) \cong \big(\mathrm{L}^2([0,T];H)\big)^*.$$

By construction we have the following three properties

$$\begin{aligned} &\overline{u}_\tau \to u \text{ in } \mathrm{L}^2([0,T];H), \quad \xi_\tau \rightharpoonup \xi_* := -\mathbb{I}_{\mathrm{R}} \dot{u} \text{ in } \mathrm{L}^2([0,T];H^*), \\ &\sup \mathcal{F}(\overline{u}_\tau(t)) \leq \mathcal{F}(u^0), \quad \xi_\tau(t) \in \partial^{\mathrm{F}} \mathcal{F}(\overline{u}_\tau(t)) \text{ a.e. in } [0,T]. \end{aligned} \tag{2.9}$$

We now would like to apply the closedness property following from Proposition 2.6. For this we set

$$X = \mathrm{L}^2([0,T];H) \quad \text{and} \quad \mathfrak{F}(u(\cdot)) := \int_0^T \mathcal{F}(u(t))\,\mathrm{d}t.$$

It is a simply calculation to show that $\mathfrak{F}$ is still proper, lower semicontinuous, and λ-convex on X if $\mathcal{F}$ is λ-convex on H. A deeper result is the characterization of the Fréchet subdifferential of $\mathfrak{F}$; namely

$$\partial^{\mathrm{F}} \mathfrak{F}(u(\cdot) = \Big\{ \xi \in \mathrm{L}^2([0,T];H^*) \;\Big|\; \xi(t) \in \partial^{\mathrm{F}} \mathcal{F}(u(t)) \text{ a.e. in } [0,T] \Big\},$$

see Exercise 2.6. With this, we can apply Proposition 2.6 to $\mathfrak{F} : X \to \mathbb{R}_\infty$ such that (2.9) implies $\xi_* \in \partial^{\mathrm{F}} \mathfrak{F}(u)$. Thus, using the characterization of $\partial^{\mathrm{F}} \mathfrak{F}(u)$ once again, we have

$$\xi_*(t) = -\mathbb{I}_{\mathrm{R}} \dot{u}(t) \in \partial^{\mathrm{F}} \mathcal{F}(u(t)) \quad \text{for a.a. } t \in [0,T],$$

which is the desired gradient-flow equation (2.3) as $u(0) = u^0$ holds as well.

Step 4: Uniqueness and full convergence. For this we use that $u \mapsto F_\lambda(u) := \mathcal{F}(u) - \frac{\lambda}{2}\|u\|_H^2$ is convex. Clearly we have $\partial F_\lambda(u) = -\lambda \mathbb{I}_{\mathrm{R}} u + \partial^{\mathrm{F}} \mathcal{F}(u)$. Thus, for arbitrary $u_1, u_0 \in \mathrm{dom}(\partial^{\mathrm{F}} \mathcal{F})$ and $\xi_j \in \partial^{\mathrm{F}} \mathcal{F}(u_j)$ we set $\eta_j = \xi_j - \lambda \mathbb{I}_{\mathrm{R}} u_j \in \partial F_\lambda(u_j)$ and obtain

$$\big\langle \xi_1 - \xi_0, u_1 - u_0 \big\rangle = \big\langle \eta_1 - \eta_0, u_1 - u_0 \big\rangle + \lambda \big\langle \mathbb{I}_{\mathrm{R}}(u_1 - u_0), u_1 - u_0 \big\rangle \geq 0 + \lambda \|u_1 - u_0\|^2,$$

by using the monotonicity of subdifferentials of the convex function F_λ.

We now assume that we have two solutions $u_0, u_1 \in \mathrm{W}^{1,2}([0,T];H)$ which implies that $t \mapsto \|u_1(t){-}u_0(t)\|^2$ is absolutely continuous. Because of $-\mathbb{I}_\mathrm{R}\dot u_j \in \partial^\mathrm{F}\mathcal{F}(u_j(t))$ a.e., we have

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u_1(t){-}u_0(t)\|^2 = \langle \mathbb{I}_\mathrm{R}(\dot u_1(t) - \dot u_0(t)), u_1(t) - u_0(t)\rangle \leq -\lambda\|u_1(t){-}u_0(t)\|^2.$$

Applying Grönwall's estimate we obtain the desired Lipschitz continuity (2.7). Assuming $u_1(0) = u_0(0) = u^0$ we obtain uniqueness of solutions.

Having this uniqueness we see that the choice of the subsequences does not matter and the whole sequence $(\widehat u_\tau)$ has to converge without taking any subsequence. ■

We emphasize that semiconvexity was used only at two positions: (i) to show closedness for subdifferential $\partial^\mathrm{F}\mathcal{E}$ and (ii) for the contraction estimate (2.7). Thus, semiconvexity is not really necessary for showing existence of solutions, if we obtain closedness with other methods.

In fact, using the λ-convexity of $\mathcal{F}$ it is possible to show better even quantitative convergence rates. This is important for two reasons: first it shows convergence without assuming the compactness of the sublevels $S_E^\mathcal{F}$, and secondly the convergence rate $\widehat u_\tau(t){-}u(t) = O(\tau^\alpha)$ for $\alpha = 1/2$ or even $\alpha = 1$ is useful in numerical implementations.

Exercise 2.6 (Evolutionary closedness) *Consider a reflexive Banach space X and a proper, lower semicontinuous and λ-convex functional $\mathcal{F} : X \to \mathbb{R}_\infty$ and denote by $\partial^\mathrm{F}\mathcal{F} : X \rightrightarrows X^*$ its subdifferential.*

(a) Define the Banach space $\boldsymbol{X} := \mathrm{L}^2([0,T];X)$ with its dual $\boldsymbol{X}^* = \mathrm{L}^2([0,T];X^*)$ and the functional

$$\mathcal{E} : \begin{cases} \boldsymbol{X} & \to \quad \mathbb{R}_\infty, \\ u(\cdot) & \mapsto \quad \int_0^T \mathcal{F}(u(t))\,\mathrm{d}t. \end{cases}$$

Show that $\mathcal{E}$ is again proper, lsc, and λ-convex.

(b) Show that $\partial^\mathrm{F}\mathcal{E}$ admits the following characterization:

$$\partial^\mathrm{F}\mathcal{E}(u) = \boldsymbol{N}(u) := \Big\{\, \xi \in \boldsymbol{X}^* \;\Big|\; \xi(t) \in \partial^\mathrm{F}\mathcal{F}(u(t)) \text{ a.e. in } [0,T] \,\Big\}.$$

Hint: It is useful to know that Lebesgue points are dense, i.e. for a.a. $t \in [0,T]$ we have $\frac{1}{2\delta}\int_{|t-s|<\delta}\xi(s)\,\mathrm{d}s \to \xi(t)$ (strongly in X^*) as $\delta \to 0^+$.

Exercise 2.7 (Alternative proof of evolutionary closedness)
(A) Given a sequence $\xi_n \rightharpoonup \xi \in \mathrm{L}^1([0,T];Y)$, define, for each $t \in [0,T]$, the accumulation set $\Xi(t) \subset Y$ via

$$\Xi(t) := \overline{\mathrm{co}}(A_\mathrm{w}(t)) \quad \text{where } A_\mathrm{w}(t) := \big\{\, y \in Y \;\big|\; \exists\,(n_k)_k :\ n_k \to \infty \text{ and } \xi_{n_k}(t) \rightharpoonup y \,\big\}.$$

Show that $\xi(t) \in \Xi(t)$ for a.a. $t \in [0,T]$. (Hint: use a version of Mazur's theorem.)

(B) For semiconvex $\mathcal{F}$ show that each $\partial^\mathrm{F}\mathcal{F}(u)$ is a closed and convex set.

(C) Assume that $\mathcal{F}$ is semiconvex and

$$u_n \to u \text{ in } \mathrm{L}^1([0,T];X), \quad \xi_n \rightharpoonup \xi \text{ in } \mathrm{L}^1([0,T];X^*),$$
$$\sup_{n\in\mathbb{N},\ t\in[0,T]} \mathcal{F}(u_n(t)) < \infty, \quad \xi_n(t) \in \partial^\mathrm{F}\mathcal{F}(u_n(t)) \text{ a.e.},$$

and conclude $\xi(t) \in \partial^\mathrm{F}\mathcal{F}(u(t))$, i.e. "evolutionary" closedness.

2.5 Completion of the Hilbert-space gradient flow via Evolutionary Variational Inequalities (EVI)

The previous existence and uniqueness theorem provides a semiflow $(\Sigma_t)_{t\geq 0}$ on $\mathrm{dom}(\mathcal{F})$ defined as follows: For $u^0 \in \mathrm{dom}(\mathcal{F})$ we set $\Sigma_t(u^0) := u(t)$ where $u : [0,\infty[\to H$ is the unique solution of the GFE (2.3). As the problem is autonomous, we obtain the semigroup property $\Sigma_{t+r} = \Sigma_t \circ \Sigma_r$ for all $t, r \geq 0$. Moreover, we have a global λ-contractivity which is completely independent of $\mathcal{F}(u^0)$:

$$\forall\, u^0, u^1 \in \mathrm{dom}(\mathcal{F})\; \forall\, t \geq 0 : \quad \|\Sigma_t(u^1) - \Sigma_t(u^0)\| \leq \mathrm{e}^{-\lambda t}\|u^1 - u^0\|.$$

Hence, we can solve the initial-value problem for more initial data by approximating them with elements from $\mathrm{dom}(\mathcal{F})$. This is possible on the closure of the domain:

$$\mathscr{D} := \overline{\mathrm{dom}(\mathcal{F})}^{H}.$$

Note that for the Allen-Cahn equation we have $H = \mathrm{L}^2(\Omega)$ and $\mathrm{dom}(\mathcal{F}_{\mathrm{AC}}) = \mathrm{H}^1(\Omega)$. Hence, in this case we find $\mathscr{D} = H = \mathrm{L}^2(\Omega)$ which enlarges the class of admissible initial conditions considerably.

For $u^0 \in \mathscr{D}$ we can choose $(u^0_m)_m$ with $u^0_m \in \mathrm{dom}(\mathcal{F})$ and $u^0_m \to u^0$ in H. Then, there is a unique solution $u_m : [0,\infty[\to H$ and we have $\|u_m(t) - u_k(t)\|_H \leq \mathrm{e}^{-\lambda t}\|u^0_m - u^0_k\|_H \to 0$ for $k, m \to \infty$. Hence, on each bounded interval $[0,T]$ we have a Cauchy sequence in $\mathrm{C}^0([0,T];H)$ and we obtain a continuous limit $u : [0,\infty[\to H$ with $u(0) = u^0$ and $\|u_m(t) - u(t)\|_H \leq \mathrm{e}^{-\lambda t}\|u^0_m - u^0\|_H$ for all $t \geq 0$. The question is of course, in what sense this limit u satisfies the GFE (2.3).

The prototypical example is the Allen-Cahn gradient system $(\mathrm{L}^2(\Omega), \mathcal{F}_{\mathrm{AC}}, m\mathbb{I}_{\mathrm{R}})$, where $\mathrm{dom}(\mathcal{F}_{\mathrm{AC}}) = \mathrm{H}^1(\Omega)$ is dense in $H = \mathrm{L}^2(\Omega)$. The following theory will show that initial conditions $u^0 \in \mathscr{D} = \mathrm{L}^2(\Omega)$ can be treated.

We summarize the result in the following theorem. Its proof is based on Evolutionary Variational Inequalities, which are not really needed, but we can prepare in this way to a general idea used later in metric spaces.

Theorem 2.10 (Completed gradient flow for $(H, \mathcal{F}, \mathbb{I}_{\mathrm{R}})$) *Let the GS $(H, \mathcal{F}, \mathbb{I}_{\mathrm{R}})$ be given as in Theorem 2.7. Then, there exists a λ-contractive, continuous semiflow $(S_t)_{t\geq 0}$ on $\mathscr{D}$ ("the gradient flow" associated with the GS $(H, \mathcal{F}, \mathbb{I}_{\mathrm{R}})$), i.e.*

(S1) $S_t : \mathscr{D} \to \mathscr{D},\ S_0 = \mathrm{id}_{\mathscr{D}},\ S_t \circ S_r = S_{t+r}$ *for all* $t, r \geq 0$.

(S2) For all $u^0 \in \mathscr{D}$ the function $[0,\infty[\, \ni t \mapsto S_t(u^0) \in H$ is continuous,

(S3) For all $u^0, u^1 \in \mathscr{D}$ and all $t \geq 0$ we have $\|S_t(u^1) - S_t(u^0)\| \leq \mathrm{e}^{-\lambda t}\|u^1 - u^0\|$,

such that for all u^0 the function $u : [0,\infty[\to H;\ t \mapsto S_t(u^0)$ is a solution of the GFE (2.3)*. In particular, we have $u \in \mathrm{W}^{1,2}_{\mathrm{loc}}(]0,\infty[;H)$ and $]0,\infty[\, \ni t \to \mathcal{F}(u(t)) \in \mathbb{R}$ (finite values) is continuous, decreasing and satisfies $\lim_{t\to 0^+} t\,\mathcal{F}(u(t)) = 0$.*

In $\mathrm{W}^{1,2}_{\mathrm{loc}}(]0,\infty[;H)$ the subscript "$_{\mathrm{loc}}$" means that for all subsets $D \Subset\ ]0,\infty[$ (compactly contained) we have $\mathrm{W}^{1,2}(D;H)$

Exercise 2.8 (Non-integrability of $\dot u$) *Consider $(H, \mathcal{F}, \mathbb{I}_{\mathrm{R}})$ with $\mathcal{F}(u) = \frac{1}{2}\langle \mathbb{A}u, u\rangle$ and $\mathbb{A} = \mathbb{A}^* \geq 0$ is a possibly unbounded self-adjoint operator.*

(A) Show that the gradient flow $(S_t)_{t\geq 0}$ equals the classical strongly continuous semigroup $(\mathrm{e}^{-t\mathbb{A}})_{t\geq 0}$, i.e. $S_t(u^0) = \mathrm{e}^{-t\mathbb{A}}u^0$ for all $u^0 \in \mathscr{D} = H$.

(B) Assume further that $\mathbb{A}$ has compact resolvent and H is infinite dimensional. Show that there exists $u^0 \in \mathscr{D} = H$ such that for $t \mapsto u(t) = \mathrm{e}^{-t\mathbb{A}}u^0$ we have $\dot u \notin \mathrm{L}^1(]0,\tau[;H)$ for any $\tau > 0$.

Before starting the proof we develop a preliminary theory for EVI for Hilbert spaces. The full theory was developed in [AGS05, Sav07, DaS14, MuS20] and will be studied further in Section 4. The major advantage of the EVI formulation is that it is a weak form in the classical sense: all solutions constructed above also solve (EVI)$_\lambda$. Moreover, it does not need any time derivative $\dot u$ nor any subdifferential $\partial^{\mathrm F}\mathcal F$. Thus, taking limits in EVI will be especially simple.

For the solutions $u \in \mathrm W^{1,2}([0,T];H)$ constructed above and arbitrary $w \in H$, we have

$$\frac12 \frac{\mathrm d}{\mathrm dt}\|u(t){-}w\|^2 = \langle \mathbb I_{\mathrm R}\dot u(t), u(t){-}w\rangle = \langle \xi(t), w{-}u(t)\rangle \quad \text{for } \xi(t)\in \partial^{\mathrm F}\mathcal F(u(t))$$
$$\overset{\mathcal F\ \lambda\text{-cvx}}{\leq} \mathcal F(w) - \mathcal F(u(t)) - \frac{\lambda}{2}\,\|u(t){-}w\|^2.$$

This is already the Differential form of the Evolutionary Variational Inequality

$$\text{(DEVI)}_\lambda \quad \left\{ \begin{array}{l} \forall\, w\in \mathrm{dom}(\mathcal F)\ \forall_{\mathrm{a.a.}} t \in [0,\infty[\, : \\ \dfrac12 \dfrac{\mathrm d}{\mathrm dt}\|u(t){-}w\|^2 + \dfrac{\lambda}{2}\|u(t){-}w\|^2 \leq \mathcal F(w) - \mathcal F(u(t)). \end{array}\right. \tag{2.10}$$

We see that DEVI is much weaker, because we do not need to impose the existence of $\dot u$. Instead we only need to impose absolute continuity of $t \mapsto \|u(t){-}w\|^2$.

We can simplify further by applying a Grönwall estimate and using that $t \mapsto \mathcal F(u(t))$ is decreasing. Then, no derivative is needed any more and we can impose conditions for all $s\geq 0$ and all $t>0$. This leads to the final Evolutionary Variational Inequality:

$$\text{(EVI)}_\lambda \quad \left\{ \begin{array}{l} \forall\, w\in \mathrm{dom}(\mathcal F)\ \forall\, s\geq 0\ \forall\, t>s : \\ \dfrac12\,\|u(t){-}w\|^2 \leq \dfrac12\,\mathrm e^{-\lambda(t-s)}\|u(s){-}w\|^2 + M_\lambda(t{-}s)\big(\mathcal F(w) - \mathcal F(u(t))\big). \end{array}\right. \tag{2.11}$$

where $M_\lambda(\tau) = \int_0^\tau \mathrm e^{-\lambda(\tau-s)}\,\mathrm ds$. We emphasize that we need $\mathcal F(u(t)) < \infty$ for $t>0$ but not for $t=0$.

Indeed, starting from (DEVI)$_\lambda$ we define $\rho(t) = \frac12 \mathrm e^{\lambda t}\|u(t){-}w\|^2$ and obtain $\dot\rho \leq \mathrm e^{\lambda t}(\mathcal F(w){-}\mathcal F(u(t))$. Integration over $[s,t]$ we find

$$\rho(t) \leq \rho(s) + \int_s^t \mathrm e^{\lambda r}(\mathcal F(w){-}\mathcal F(u(r))\,\mathrm dr \leq \rho(s) + \int_s^t \mathrm e^{\lambda r}\,\mathrm dr\ \big(\mathcal F(w){-}\mathcal F(u(t))\big).$$

Now, inserting the definition of ρ and multiplying with $\mathrm e^{-\lambda t}$ gives (EVI)$_\lambda$.

The main observation for completing the proof of Theorem 2.10 is that all functions $t\mapsto S_t(u^0)$ with $u^0 \in \mathscr D$ satisfy (EVI)$_\lambda$. In Section 4 we will show that (EVI)$_\lambda$ already characterizes the solutions uniquely, thus (EVI)$_\lambda$ characterizes the gradient flow $(S_t)_{t\geq 0}$ on $\mathscr D = \overline{\mathrm{dom}(\mathcal F)} \subset H$ completely.

Proposition 2.11 (Gradient flow and EVI) *Let the GS $(H,\mathcal F,\mathbb I_{\mathrm R})$ and the gradient flow $S_t : \mathscr D \to \mathscr D$ be given as in Theorem 2.10. Then, for all $u^0\in\mathscr D$ the functions $u : [0,\infty[\, \mapsto S_t(u^0) \in \mathscr D \subset H$ satisfy (EVI)$_\lambda$. Moreover, we have $\mathcal F(u(t))<\infty$ for $t>0$ and $\limsup_{t\to 0^+}\ t\,\mathcal F(u(t)) = 0$.*

Proof. By construction, we already know that for all $u^0 \in \mathrm{dom}(\mathcal F)$ the solutions $u(t) = S_t(u^0)$ satisfy (EVI)$_\lambda$. For all other initial conditions $u^0 \in \mathscr D\setminus \mathrm{dom}(\mathcal F)$ we can choose a sequence u^0_m with $u^0_m \in \mathrm{dom}(\mathcal F)$ and $u^0_m \to u^0$ in H. The corresponding solutions $u_m = S_t(u^0_m)$ satisfy

$$\frac12\|u_m(t){-}w\|^2 \leq \frac12 \mathrm e^{-\lambda(t-s)}\|u_m(s){-}w\|^2 + M_\lambda(t{-}s)\big(\mathcal F(w) - \mathcal F(u_m(t))\big)$$

for $t > s \geq 0$ and $w \in \mathrm{dom}(\mathcal{F})$. By λ-contractivity we have uniform (strong) convergence of u_m to $u : t \mapsto S_t(u^0)$ on all compact subsets of $[0, \infty[$. Thus, we can pass to the limit in the first two terms. In the last term we can use the lower semicontinuity $\mathcal{F}(u(t)) \leq \liminf_{m\to\infty} \mathcal{F}(u_m(t))$ and $M_\lambda(t{-}s) \geq 0$.

This limit passage would even be allowed if $\mathcal{F}(u(t)) = \infty$, however (EVI)$_\lambda$ gives, for all $w \in \mathrm{dom}(\mathcal{F})$ and $s = 0$ the upper bound

$$\mathcal{F}(u(t)) \leq \mathcal{F}(w) + \frac{\mathrm{e}^{-\lambda t}}{2M_\lambda(t)}\,\|u^0{-}w\|^2.$$

As $\mathcal{F}$ is proper, we have shown $\mathcal{F}(u(t)) < t$ for $t > 0$. Moreover, multiplying by $t > 0$ we can take the limsup for $t \to 0^+$ and find

$$\limsup_{t\to 0^+}\; t\,\mathcal{F}(u(t)) \;\leq\; \frac{1}{2}\|u^0{-}w\|^2 \quad \text{for all } w \in \mathrm{dom}(\mathcal{F}).$$

As $\mathrm{dom}(\mathcal{F})$ is dense in $\mathscr{D}$ we find the desired result $\limsup_{t\to 0^+} t\,\mathcal{F}(u(t)) = 0$. ■

To appreciate the last relation concerning the boundedness of $t\mathcal{F}(u(t)$, we consider the example $(\mathrm{L}^2(\Omega), \mathcal{F}_{\mathrm{Dir}}, \mathbb{I}_{\mathrm{R}})$ with $\mathcal{F}_{\mathrm{Dir}}(u) = \frac{1}{2}\|\nabla u\|^2_{\mathrm{L}^2}$ on $\mathrm{dom}(\mathcal{F}_{\mathrm{Dir}}) = \mathrm{H}^1_0(\Omega)$. From linear PDE theory we know the explicit estimate $\|u(t)\|_{\mathrm{H}^1} \leq Ct^{-1/2}\|u(0)\|_{\mathrm{L}^2}$ which corresponds to the statement $t\,\mathcal{F}_{\mathrm{Dir}}(u(t)) \leq \frac{1}{2}C^2\|u(0)\|^2_{\mathrm{L}^2}$. Hence, our general and abstract theory recovers a very similar behavior which is optimal in the sense that the power $\alpha = 1$ cannot be decreased without losing boundedness of $t^\alpha\mathcal{F}(u(t)$.

We are now in the position to study the remaining properties of the completion of the semiflow.

Proof of Theorem 2.10. It remains to show that $u(t) = S_t(u^0)$ is differentiable a.e. and that GFE holds.

We observe that $\widetilde{u} : [0, \infty[\mapsto u(t{+}t_*)$ is a solution of the GFE (2.3) with $\widetilde{u}(0) = u(t_*)$. For this, first note that $\mathcal{F}(\widetilde{u}(0)) = \mathcal{F}(u(t_*)) < \infty$. Hence, there is a unique solution $\widehat{u} \in \mathrm{W}^{1,1}_{\mathrm{loc}}([0, \infty[; H)$ with $\widehat{u}(0) = u(t_*)$. Moreover, by the semigroup property and $\widetilde{u}(t) = u(t{+}t_*)$ we have $\widetilde{u}(t) = u(t{+}t_*) = S_{t+t_*}(u^0) = S_t(S_{t_*}(u^0)) = S_t(u(t_*)) = \widehat{u}(t)$. Thus, we conclude that the solutions obtained in the limit $u_m(\cdot) \to u(\cdot)$ are differentiable, as desired. ■

It is possible to establish many more properties of the gradient flows in Hilbert spaces. E.g. in [Bré73, Thm. 3.1(5+6)] it is shown that for convex functionals $\mathcal{F}$ all solutions u of the GFE (2.3) have the property that the one-sided derivatives

$$\dot u^+(t) = \lim_{h\to 0^+} \frac{1}{h}\big(u(t{+}h) - u(t)\big)$$

exist and can be identified with the "norm-minimal selection" in the subdifferential of $\mathcal{F}$, i.e. for all $t > 0$ one has

$$-\mathbb{I}_{\mathrm{R}}\dot u^+(t) = \partial^0\mathcal{F}(u(t)) := \mathrm{arg\,min}\big\{\, \|\xi\|_{H^*} \,\big|\, \xi \in \partial\mathcal{F}(u(t)) \,\big\}.$$

Moreover, the mapping $]0, \infty[\ni t \mapsto \|\dot u^+(t)\|$ is decreasing.

Indeed the latter property is not so surprising if we use the λ-contractivity (2.7) for the two solutions $t \mapsto u(t)$ and $t \mapsto u(t{+}h)$. After dividing by $h > 0$ we obtain

$$\big\|\frac{1}{h}\big(u(t{+}h) - u(t)\big)\big\| \;\leq\; \mathrm{e}^{-\lambda(t-s)}\big\|\frac{1}{h}\big(u(s{+}h) - u(s)\big)\big\|.$$

Thus, for $\lambda \geq 0$ we obtain that the norms of the difference quotients are decreasing. Clearly, if the limits exist in the strong sense, then they are still decreasing.

3 Generalized gradient systems in Banach spaces

In this section we generalize the theory in a twofold way. First we go from Hilbert spaces to Banach spaces and second we generalize the linear kinetic relation $\xi = \mathbb{G}(u)\dot u$ or $\dot u = \mathbb{K}(u)\xi$ to *nonlinear kinetic relations*, which allows a much larger set of applications.

3.1 Legendre duality and nonlinear kinetic relations

Now the kinetic relation $X \ni \dot u = v \leftrightarrow \xi \in X^*$ cannot be given by a simple linear map such as the Riesz isomorphism $\mathbb{I}_\mathrm{R} : H \to H^*$, because X and X^* are not isomorphic in general.

The typical replacements in general (separable, reflexive) Banach spaces are maximal monotone operators $\boldsymbol{A} : X \rightrightarrows X^*$ and their inverse

$$\boldsymbol{A}^{-1} : X^* \rightrightarrows X;\ \boldsymbol{A}^{-1}(\xi) := \big\{\, v \in X \,\big|\, \xi \in \boldsymbol{A}(v)\,\big\},$$

which is again a maximal monotone operator, because the notion of (maximal) monotonicity is symmetric:

$$\boldsymbol{A} \text{ monotone } \iff \forall\, v_1, v_0 \in X\ \forall\, \xi_1 \in \boldsymbol{A}(v_1),\ \xi_0 \in \boldsymbol{A}(v_0) :\ \langle \xi_1{-}\xi_0, v_1{-}v_0\rangle \geq 0.$$

The corresponding evolution equations are then called *doubly nonlinear equations* (cf. [CoV90, Col92]): $0 \in \boldsymbol{A}(\dot u) + \partial^\mathrm{F}\mathcal{F}(u(t)) - \ell(t)$, where $\ell \in \mathrm{L}^1([0,T]; X^*)$ is a general external forcing.

To obtain a theory of generalized gradient systems we consider a subclass of these kinetic relations that encode a nonlinear version of the Onsager symmetry. This class is given by *subdifferentials of convex potentials*.

To motivate this, we first consider the quadratic functional $\Psi : H \to \mathbb{R};\ v \mapsto \frac12\langle \mathbb{A}v, v\rangle$. Then the differential reads

$$\mathrm{D}\Psi(v) = \mathbb{G}v \text{ with } \mathbb{G} = \frac12(\mathbb{A}{+}\mathbb{A}^*),$$

such that $\mathbb{G}$ automatically enjoys the Onsager symmetry $\mathbb{G}^* = \mathbb{G}$ for $\mathbb{A} \in \mathrm{Lin}(H; H^*)$. Moreover, $\mathbb{G} \geq 0$ is equivalent to $\Psi(v) \geq 0$ for all $v \in H$. However, it is easy to see that $\boldsymbol{A}(v) = \{\mathbb{A}v\}$ defines a (maximal) monotone operator $\boldsymbol{A} : H \rightrightarrows H^*$ if and only if $\frac12(\mathbb{A}{+}\mathbb{A}^*) \geq 0$, i.e. the skew-symmetric part $\frac12(\mathbb{A}{-}\mathbb{A}^*)$ is completely arbitrary.

Secondly, we consider time-incremental minimization schemes in the form

$$u_k \ \text{ minimizes } \ u \mapsto \tau\Psi\big(\frac{1}{\tau}(u{-}u_{k-1})\big) + \mathcal{F}(u)$$

for a convex function $\Psi : X \to \mathbb{R}_\infty$. Assuming the sum rule, the Euler-Lagrange equation reads $0 \in \partial\Psi\big(\frac1\tau(u_k{-}u_{k-1})\big) + \partial^\mathrm{F}\mathcal{F}(u_k)$ which will be interpreted as the backward-Euler (fully implicit) discretization of the evolutionary inclusion $0 \in \partial\Psi(\dot u(t)) + \partial^\mathrm{F}\mathcal{F}(u(t))$.

Definition 3.1 (Dissipation potential) *A function $\Psi : X \to \mathbb{R}_\infty$ is called a* dissipation potential on X*, if Ψ is lower semicontinuous, convex and satisfies $\Psi(v) \geq \Psi(0) = 0$.*

We call the Legendre-Fenchel dual (conjugate) $\Psi^* = \mathfrak{L}\Psi : X^* \to \mathbb{R}_\infty$ the dual dissipation potential for Ψ*. It is defined by*

$$\Psi^*(\xi) = (\mathfrak{L}\Psi)(\xi) := \sup\big\{\, \langle \xi, v\rangle - \Psi(v) \,\big|\, v \in X\,\big\}.$$

To justify the above name “dual dissipation potential”, note $\Psi^* = \mathfrak{L}\Psi$ is automatically convex and lsc. Moreover, $\Psi(0) \geq 0$ implies $\Psi^*(\xi) \geq 0$, whereas $\Psi(v) \geq 0$ for all $v \in X$ implies $\Psi^*(0) = 0$.

On a Hilbert space H we have

$$\Psi(v) = \frac{1}{2}\langle \mathbb{G}v, v\rangle \quad \Longleftrightarrow \quad \Psi^*(\xi) = \frac{1}{2}\langle \xi, \mathbb{K}\xi\rangle \text{ with } \mathbb{K} = \mathbb{G}^{-1}.$$

If $p \in]1, \infty[$ and $(X, \|\cdot\|_X)$ is a Banach space with dual space $(X^*, \|\cdot\|_{X^*})$, then

$$\Psi(v) = \frac{1}{p}\,\|v\|_X^p \quad \Longleftrightarrow \quad \Psi^*(\xi) = \frac{1}{p^*}\,\|\xi\|_{X^*}^{p^*} \text{ with } p^* = \frac{p}{p{-}1}.$$

A trivial but nevertheless important consequence of the definition of $\Psi^* = \mathfrak{L}\Psi$ is the

$$\textit{Fenchel-Young inequality}: \qquad \forall\,(v,\xi) \in X{\times}X^*: \quad \Psi(v) + \Psi^*(\xi) \geq \langle \xi, v\rangle. \tag{3.1}$$

We refer to [Fen49] for the first occurrence in $X = \mathbb{R}^n$ and [BaC17, Prop. 13.15] for a general theory (in Hilbert spaces).

The following important relation is the basis of the term “*duality theory*”.

Lemma 3.2 (Legendre transform is an involution) *Assume that X is reflexive, i.e. $X^{**} = (X^*)^* = X$. Then, $\mathfrak{L}$ maps proper, lsc, convex functions on X onto proper, lsc, convex functions on X^* and vice versa. Moreover, $\Psi^{**} = (\Psi^*)^* = \mathfrak{L}(\mathfrak{L}(\Psi)) = \Psi$, i.e. $\mathfrak{L}$ is an involution.*

Proof. The definition of $\mathfrak{L}$ immediately shows that Ψ^* is again proper, lsc, and convex.

Using the Fenchel-Young inequality (3.1) we easily obtain, for all $v \in X$,

$$(\Psi^*)^*(v) = \sup\big\{\, \langle \xi, v\rangle - \Psi^*(\xi) \,\big|\, \xi \in X^* \,\big\} \leq \sup\big\{\, \Psi(v) \,\big|\, \xi \in X^* \,\big\} = \Psi(v).$$

To show $\Psi(v) \leq \Psi^{**}(v)$ we use that Ψ is convex and lsc. For fixed $v_0 \in X$ and $a_0 < \Psi(v_0)$ there exists $\xi_0 \in X^*$ such that $\Psi(v) \geq a_0 + \langle \xi_0, v{-}v_0\rangle$. This implies $\Psi^*(\xi_0) \leq -a_0 + \langle \xi_0, v_0\rangle$ and hence $\Psi^{**}(v_0) \geq a_0$. As $a_0 < \Psi(v_0)$ was arbitrary, we conclude $\Psi(v_0) \leq \Psi^{**}(v_0)$. ■

With this, the pair (Ψ, Ψ^*) is called a *conjugate pair* as $\Psi^* = \mathfrak{L}\Psi$ and $\Psi = \mathfrak{L}(\Psi^*)$. We will use the word ‘primal dissipation potential’ for Ψ and ‘dual dissipation potential’ for Ψ^*, but of course, the notion of ‘primal’ and ‘dual’ can be interchanged in the case of reflexive spaces X and X^*.

For our theory the most important duality result are the so-called Fenchel equivalences which we formulate explicitly here as a theorem, even though in some textbooks they are considered simple lemmas or exercises. See also [BaC17, Thm. 16.29] for a proof in the Hilbert space setting.

Theorem 3.3 (Fenchel equivalences [Fen49]) *Consider a reflexive Banach space X with dual X^* and consider a conjugate pair (Ψ, Ψ^*) of proper, lsc, and convex functions. Then, for all $(v_0, \xi_0) \in X{\times}X^*$ the following five statements are equivalent:*

(i) v_0 *minimizes the functional* $v \mapsto \Psi(v) - \langle \xi_0, v\rangle$ *(optimality of $v \in X$);*

(ii) $\xi_0 \in \partial\Psi(v_0)$ *(subdifferential inclusion in X^*);*

(iii) $\Psi(v_0) + \Psi^*(\xi_0) \leq \langle \xi_0, v_0\rangle$ *(optimality condition in $\mathbb{R}$);*

(iv) $v_0 \in \partial\Psi^*(\xi_0)$ *(subdifferential inclusion in X);*

(v) ξ_0 *maximizes the functional* $\xi \mapsto \langle \xi, v_0\rangle - \Psi^*(\xi)$ *(optimality of $\xi \in X^*$).*

Here in (iii) *we can either write "$\leq$" or "$=$", because the Fenchel-Young inequality* (3.1) *always gives "$\geq$".*

Proof. Obviously, the equivalences (i) $\Leftrightarrow$ (ii) and (iv) $\Leftrightarrow$ (v) easily follow by the Euler-Lagrange equation for the convex functionals.

Thus, it remains to show (ii) $\Leftrightarrow$ (iii) and (iii) $\Leftrightarrow$ (iv). By duality the two equivalences can be proved in the same way, so we concentrate on the first.

"(ii) $\Rightarrow$ (iii)" Starting from (ii) gives $\Psi(v) \geq \Psi(v_0) + \langle \xi_0, v{-}v_0\rangle$. Inserting this into the definition of $\Psi^*(\xi_0)$ we immediately obtain $\Psi^*(\xi_0) \leq \langle \xi_0, v_0\rangle - \Psi(v_0)$, which is (iii).

"(iii) $\Rightarrow$ (ii)" (iii) is the upper bound $\Psi^*(\xi_0) \leq \langle \xi_0, v_0\rangle - \Psi(v_0)$, which using $\Psi = \Psi^{**}$ implies the lower bound

$$\Psi(v) = \sup\left\{ \langle \xi, v\rangle - \Psi^*(\xi) \,\middle|\, \xi \in X^* \right\} \geq \langle \xi_0, v\rangle - \Psi^*(\xi_0) \geq \langle \xi_0, v{-}v_0\rangle + \Psi(v_0).$$

Hence, we have $\xi_0 \in \partial\Psi(v_0)$ which is (ii). ■

We emphasize that the equivalence (ii) $\Longleftrightarrow$ (iv) shows that the set-valued mapping $X \ni v \rightrightarrows \partial\Psi(v) \subset X^*$ is exactly the inverse mapping (in the sense of set-valued monotone operators) of the set-valued mapping $X^* \ni \xi \rightrightarrows \partial\Psi^*(\xi) \subset X$.

Example 3.4 (Viscoplasticity) *As a nontrivial and mechanically important example we treat viscoplasticity, where $p \in X = \mathrm{L}^2(\Omega;\mathbb{R}^{d\times d}_{\mathrm{sym}})$ denotes the plastic distortion. The viscoplastic dissipation potential depends on the plastic rate $\pi = \dot p$ and takes the form*

$$\Psi(\pi) = \int_\Omega \psi(\pi(x))\,\mathrm{d}x \quad \textit{with } \psi(\pi) = \sigma_{\mathrm{yield}}|\pi| + \frac{\mu}{2}|\pi|^2 \quad \textit{where } |\pi|^2 = \sum_{i,j=1}^d \pi_{ij}^2.$$

The dual variables are the plastic (back-) stresses $\Sigma_\mathrm{p} \in \mathrm{L}^2(\Omega;\mathbb{R}^{d\times d}_{\mathrm{sym}})$. Clearly, we have (cf. Exercise 2.6)

$$\partial\Psi(\pi) = \Big\{ \Sigma_\mathrm{p} \in \mathrm{L}^2(\Omega;\mathbb{R}^{d\times d}_{\mathrm{sym}}) \,\Big|\, \Sigma_\mathrm{p}(x) \in \partial\psi(\pi(x)) \text{ a.e. in } \Omega \Big\}.$$

where the pointwise subdifferential $\partial\psi$ is given by

$$\partial\psi(\pi) = \begin{cases} \sigma_{\mathrm{yield}}\overline{B_1(0)} & \text{for } \pi = 0, \\ \left\{ \frac{\sigma_{\mathrm{yield}}}{|\pi|}\,\pi + \mu\pi \right\} & \text{for } \pi \neq 0. \end{cases}$$

The relation $\Sigma_\mathrm{p} \in \partial\psi(\pi)$ can be inverted explicitly, giving

$$\pi = \mathrm{D}\psi^*(\Sigma_\mathrm{p}) \quad \textit{with } \psi^*(\Sigma_\mathrm{p}) = \frac{1}{2\mu}\big(\max\{0, |\Sigma_\mathrm{p}|{-}\sigma_{\mathrm{yield}}\}\big)^2.$$

Note that this relation shows that $\pi = \dot p = 0$ whenever $|\Sigma_\mathrm{p}| \leq \sigma_{\mathrm{yield}}$, i.e. if the stress does not reach the threshold σ_{yield} for yielding.

In particular, we find the dual dissipation potential depending on the plastic stress:

$$\Psi^*(\Sigma_\mathrm{p}) = \int_\Omega \psi^*(\Sigma_\mathrm{p}(x))\,\mathrm{d}x.$$

Exercise 3.1 (Dissipation functions) *For a differentiable dissipation potential $\Psi : X \to [0,\infty]$ we define the*

$$\text{dissipation function } \mathrm{Diss}_\Psi(v) = \langle \mathrm{D}\Psi(v), v\rangle.$$

(a) Show that $\mathrm{Diss}_\Psi(v) \geq \Psi(v)$ and give an example where Diss_Ψ is non convex.

(b) Discuss the equality $\mathrm{Diss}_\Psi(v) = \Psi(v)$.

(c) Assume that Ψ is differentiable and positively p-homogeneous, i.e. $\Psi(\lambda v) = \lambda^p \Psi(v)$ for all $\lambda > 0$ and $v \in X$, and show $\mathrm{Diss}_\Psi(v) = p\Psi(v)$.

(d) Assume now that Ψ is only radially differentiable, i.e. for all $v \in X$ the function $]0,\infty[\ni \lambda \to \Psi(\lambda v)$ is differentiable. Show that for each $v \in X$ the values of $\langle \xi, v\rangle$ are constant for all $\xi \in \partial\Psi(v)$. Conclude

$$\mathrm{Diss}_\Psi(v) = \Psi(v) + \Psi^*(\xi) \quad \text{for all } \xi \in \partial\Psi(v).$$

Hint: Consider $g(\lambda) = \Psi(\lambda v) + \Psi^*(\xi) - \langle \xi, \lambda v\rangle$.

The function $v \mapsto \mathrm{Diss}_\Psi(v)$ is called (primal) dissipation function, and similarly $\xi \to \mathrm{Diss}_{\Psi^*}(\xi)$ is called dual dissipation function. These functions are often used in modeling, especially when their subdifferentials are single-valued. However, they have to be clearly distinguished from the dissipation potentials. They can be used in the energy-dissipation balances below, see e.g. (3.3)*, but have weaker properties.*

Exercise 3.2 (Duality of properties) *On a reflexive Banach space X consider a pair of Legendre dual functions $\Psi : X \to \mathbb{R}_\infty$ and $\Psi^* : X^* \to \mathbb{R}_\infty$.*

For a general lsc, convex functional $\Phi : Y \to \mathbb{R}_\infty$ consider the properties:

(P1) $\Phi(0) \leq 0$;

(P2) $\Phi(y) \geq 0$ for all y;

(P3) $\Phi(y) \geq c\|y\| - C$ for all y;

(P4) Φ is superlinear, i.e. $\Psi(u)/\|u\| \to \infty$ for $\|u\| \to \infty$;

(P5) $\Phi(y) \leq M$ for all $y \in B_R(0) \subset Y$;

(P6) Φ takes only finite values;

(P7) $v \mapsto \Phi(v) - \langle \eta, v\rangle$ has a unique minimizer;

(P8) $\partial\Phi(w)$ is single-valued.

Try to find implications or equivalences like "if Ψ satisfies (Pn) then Ψ^* satisfies (Pk).

3.2 Generalized gradient systems and the gradient-flow equations

Above we have always used $v \in X$ as a placeholder for the rate $\dot u$. Of course, in general systems we may have *state-dependent kinetic relations*. We start again with the manifold setting M where now the state-dependent (primal) dissipation potential $\mathcal{R}$ is defined on the tangent bundle:

$$\mathcal{R}\colon \mathrm{T}M \to [0,\infty] \text{ such that } \forall\, u \in M :\ \mathcal{R}(u,\cdot)\colon \mathrm{T}_u M \to [0,\infty] \text{ is a dissipation potential.}$$

The dual dissipation potential $\mathcal{R}^* : \mathrm{T}^*M \to [0,\infty]$ is obtained at fixed $u \in M$, i.e.

$$\mathcal{R}^*(u,\xi) := \big(\mathfrak{L}(\mathcal{R}(u,\cdot))\big)(\xi)$$

When we write subdifferentials of R or $\mathcal{R}^*$ we always mean subdifferentials with respect to the second variable in the linear space $\mathrm{T}_u M$ or $\mathrm{T}_u^* M$:

$$\partial\mathcal{R}(u,v) := \partial_v \mathcal{R}(u,v) = \partial\big(\mathcal{R}(u,\cdot)\big)(v) \subset \mathrm{T}_u^* M \quad \text{and}$$
$$\partial\mathcal{R}^*(u,\xi) := \partial_\xi \mathcal{R}^*(u,\xi) = \partial\big(\mathcal{R}^*(u,\cdot)\big)(\xi) \subset \mathrm{T}_u M.$$

Definition 3.5 (Generalized gradient system: ODE case) *A triple $(M,\mathcal{F},\mathcal{R})$ (or equivalently $(M,\mathcal{F},\mathcal{R}^*)$) is called a* generalized gradient system*, if M is a manifold, $\mathcal{F}: M \to \mathbb{R}$ is a differentiable function and $\mathcal{R}: \mathrm{T}M \to [0,\infty]$ (or equivalently $\mathcal{R}^*: \mathrm{T}^*M \to [0,\infty]$) is a (state-dependent) dissipation potential.*

The associated gradient-flow equation is given by

$$0 \in \partial\mathcal{R}(u,\dot u) + \mathrm{D}\mathcal{F}(u) \subset \mathrm{T}_u^* M \quad \Longleftrightarrow \quad \dot u \in \partial\mathcal{R}^*(u,-\mathrm{D}\mathcal{F}(u)) \subset \mathrm{T}_u M.$$

By the Fenchel equivalences, we can also reformulate the gradient-flow equation by the optimality condition, which is a power identity:

$$\mathcal{R}(u,\dot u) + \mathcal{R}^*(u,-\mathrm{D}\mathcal{F}(u)) = -\langle \mathrm{D}\mathcal{F}(u),\dot u\rangle = -\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(u(t)). \tag{3.2}$$

This equation we can integrate and obtain the *energy-dissipation balance*

$$\forall\, 0<s<t: \qquad \mathcal{F}(u(t)) + \int_s^t \big(\mathcal{R}(u,\dot u) + \mathcal{R}^*(u,-\mathrm{D}\mathcal{F}(u))\big)\,\mathrm{d}r = \mathcal{F}(u(s)), \tag{3.3}$$

which simply states that the energy $\mathcal{F}(u(t))$ at the later time t plus the dissipated energy in the time interval $[s,t]$ give exactly the energy $\mathcal{F}(u(s))$ at the earlier time s.

Of course, by the results from Exercise 3.1 we can write the energy-dissipation balance also in one of the following simpler forms

$$\mathcal{F}(u(t)) + \int_s^t \mathrm{Diss}_{\mathcal{R}(u,\cdot)}(\dot u)\,\mathrm{d}r = \mathcal{F}(u(s)) \ \text{ or } \ \mathcal{F}(u(t)) + \int_s^t \mathrm{Diss}_{\mathcal{R}^*(u,\cdot)}(-\mathrm{D}\mathcal{F}(u))\,\mathrm{d}r = \mathcal{F}(u(s)).$$

But there is a major difference between (3.3) and the latter two forms. The formulation involving "$\mathcal{R}\oplus\mathcal{R}^*$" is derived from the optimality condition, and thus we will be able to show that the EDB (3.3) is still equivalent to the full gradient-flow equations. The same is not true for the latter two formulations, which hold along all solutions, but do not characterize the solutions. Thus, to emphasize this fact, we will sometimes insist that (EDB) is always assumed to be in "$\mathcal{R}\oplus\mathcal{R}^*$ form".

We now turn to the case of infinite dimensional evolution equations on a reflexive Banach space X. Of course, for PDEs or abstract evolutionary equations one typically needs several Banach spaces, so here X denotes the space in which rates $\dot u$ typically are located; other spaces associated with the energy will be implicitly defined by $\mathrm{dom}(\mathcal{F})$ or $\mathrm{dom}(\partial^{\mathrm{F}}\mathcal{F})$. For simplicity, we will again use the Fréchet subdifferential $\partial^{\mathrm{F}}\mathcal{F}: X \rightrightarrows X^*$ but other choices might be possible.

Definition 3.6 (Generalized gradient systems: PDE case) *A triple $(X,\mathcal{F},\mathcal{R})$ (or equivalently $(X,\mathcal{F},\mathcal{R}^*)$) is called a* generalized gradient system *on the Banach space X, if $\mathcal{F}: X \to \mathbb{R}_\infty$ is a lower semicontinuous functional and $\mathcal{R}$ is a primal dissipation potential meaning that $\mathcal{R}(u,\cdot)$:*

$X \to [0,\infty]$ *is a dissipation potential for all* $u \in X$ *(or equivalently* $\mathcal{R}^*(u,\cdot) : X^* \to [0,\infty]$*). The associated gradient flow equation is given by*

$$0 \in \partial\mathcal{R}(u(t),\dot u(t)) + \partial^{\mathrm{F}}\mathcal{F}(u(t)) \text{ a.e. in } [0,T] \iff \begin{cases} \dot u(t) \in \partial\mathcal{R}^*(u,-\xi(t)) \\ \text{and } \xi(t) \in \partial^{\mathrm{F}}\mathcal{F}(u(t)) \\ \text{for a.a. } t \in [0,T]. \end{cases} \tag{3.4}$$

Subsequently, we will often use the pair (u,ξ) *to denote the solutions.*

Example 3.7 (Doubly nonlinear diffusion equation) *For* $p,q \in \,]1,\infty[$ *we consider the space* $X = \mathrm{L}^q(\Omega)$*, the energy* $\mathcal{F}$ *with* $\mathcal{F}(u) = \int_\Omega \frac{1}{p}|\nabla u|^p \,\mathrm{d}x$ *for* $u \in \mathrm{W}^{1,p}(\Omega)$ *and* $+\infty$ *otherwise in* X*. Moreover, we consider the dissipation potential* $\mathcal{R}(u,v) = \int_\Omega \frac{1}{q}(2{+}\cos u)|v|^q \,\mathrm{d}x$*. Both differentials* $\partial^{\mathrm{F}}\mathcal{F}$ *and* $\partial\mathcal{R}$ *are single-valued and we obtain the doubly nonlinear diffusion equation*

$$(2{+}\cos u)|\dot u|^{q-2}\dot u = \Delta_p u := \mathrm{div}\big(|\nabla u|^{p-2}\nabla u\big) \text{ in } \Omega, \qquad \nabla u\cdot\nu = 0 \text{ on } \partial\Omega,$$

as the associated gradient-flow equation.

For course, as in the ODE case we can replace the two formulations in (3.4) by the optimality condition

$$\mathcal{R}(u(t),\dot u(t)) + \mathcal{R}^*(u(t),-\xi(t)) = -\langle \xi(t),\dot u(t)\rangle \text{ and } \xi(t) \in \partial^{\mathrm{F}}\mathcal{F}(u(t)) \text{ a.e. in } [0,T]. \tag{3.5}$$

However, integration of this relation is no longer trivial for two reasons: First we cannot simply assume that "$\mathcal{R}{\oplus}\mathcal{R}^*$" is integrable for solutions (u,ξ), and secondly, the application of the chain rule to $\langle \xi,\dot u\rangle$ may be not valid. These two points will be discussed in the following subsection.

However, at this stage we can see already the main impact of the gradient system on the gradient flow equation. The gradient structure provides an easy way for setting up a

time-incremental minimization scheme for time step $\tau = T/N > 0$: $u_0^\tau = u^0 \in X$ (given initial value) for $k = 1,\dots,N$ find u_k^τ as minimizer of the functional $\Phi_\tau^{\mathcal{F},\mathcal{R}}(u_{k-1};\,\cdot\,) : X \ni u \mapsto \tau\mathcal{R}\big(u_{k-1},\frac{1}{\tau}(u{-}u_{k-1})\big) + \mathcal{F}(u).$

(3.6)

As in the Hilbert-space case the minimizers u_k (for notational convenience we drop the superscript for the upcoming calculation) satisfy the Euler-Lagrange equation

$$0 \in \partial\mathcal{R}\big(u_{k-1},\frac{1}{\tau}(u_k{-}u_{k-1})\big) + \partial^{\mathrm{F}}\mathcal{F}(u_k)$$

or equivalently

$$\exists\, \xi_k \in X^* : \quad \xi_k \in \partial^{\mathrm{F}}\mathcal{F}(u_k) \quad \text{and} \quad -\xi_k \in \partial\mathcal{R}\big(u_{k-1},\frac{1}{\tau}(u_k{-}u_{k-1})\big). \tag{3.7}$$

The last relation can be used to apply the Fenchel equivalences giving

$$\mathcal{R}\big(u_{k-1},\frac{1}{\tau}(u_k{-}u_{k-1})\big) + \mathcal{R}^*(u_{k-1},-\xi_k) = -\big\langle \xi_k,\frac{1}{\tau}(u_k{-}u_{k-1})\big\rangle.$$

Assuming further that $\mathcal{F}$ is λ-convex we can estimate the right-hand side by using $\mathcal{F}(u_{k-1}) \geq \mathcal{F}(u_k) + \langle \xi_k, u_{k-1}-u_k\rangle + \frac{\lambda}{2}\|u_k-u_{k-1}\|^2$. After multiplying by $\tau > 0$ we arrive at a discrete type of energy-dissipation inequality:

$$\tau\Big(\mathcal{R}\big(u_{k-1}, \frac{1}{\tau}(u_k-u_{k-1})\big) + \mathcal{R}^*(u_{k-1}, -\xi_k)\Big) \leq \mathcal{F}(u_{k-1}) - \mathcal{F}(u_k) - \frac{\lambda}{2}\|u_k-u_{k-1}\|^2. \quad (3.8)$$

Reintroducing the superscript τ in u_k^τ and ξ_k^τ again, We can now define the four interpolants $\widehat{u}_\tau$, $\overline{u}_\tau$, and $\underline{u}_\tau$ from $[0,T] \to X$ and $\overline{\xi}_\tau : [0,T] \to X^*$ as follows:

$$\begin{aligned}
&\widehat{u}_\tau((k+\theta-1)\tau) = (1-\theta)u_{k-1}^\tau + \theta u_k^\tau \quad \text{for } k \in \{1,\dots,N\} \text{ and } \theta \in [0,1];\\
&\overline{u}_\tau(0) = u_0^\tau \text{ and } \overline{u}_\tau(t) = u_k^\tau \text{ for } t \in \,]k\tau-\tau, k\tau] \text{ and } k \in \{1,\dots,N\};\\
&\underline{u}_\tau(t) = u_k^\tau \text{ for } t \in [k\tau, k\tau-\tau] \text{ and } k \in \{0,\dots,N-1\} \text{ and } \underline{u}_\tau(T) = u_N^\tau;\\
&\overline{\xi}_\tau(0) = 0 \in X^* \text{ and } \overline{\xi}_\tau(t) = \xi_k^\tau \text{ for } t \in \,](k-1)\tau, k\tau] \text{ and } k \in \{1,\dots,N\}.
\end{aligned} \quad (3.9)$$

Here $\overline{u}_\tau$ and $\overline{\xi}_\tau$ are the left-continuous, piecewise constant interpolants, $\underline{u}_\tau$ is the right-continuous, piecewise constant interpolant, and $\widehat{u}_\tau$ is the continuous piecewise affine interpolant which has the piecewise constant derivative

$$\dot{\widehat{u}}_\tau(t) = \frac{1}{\tau}\big(u_k^\tau - u_{k-1}^\tau\big) \quad \text{for } k \in \{1,\dots,N\} \text{ and } t \in \,]k\tau-\tau, k\tau[.$$

With these definitions we can rewrite the incremental Euler-Lagrange equation (3.7) as an approximate equation on $[0,T]$ as follows

$$\overline{\xi}_\tau(t) \in \partial^{\mathrm{F}}\mathcal{F}(\overline{u}_\tau(t)) \quad \text{and} \quad 0 \in \partial\mathcal{R}\big(\underline{u}_\tau(t), \dot{\widehat{u}}_\tau(t)\big) + \overline{\xi}_\tau(t) \quad \text{for a.a. } t \in [0,T]. \quad (3.10)$$

Note that all four interpolants are needed because our scheme is "semi-implicit", namely implicit in the functional $\mathcal{F}$ and explicit in the state-dependence of $\mathcal{R}$.

We may also consider the discrete energy dissipation inequality (3.8). After summation over $k = 1,\dots,N$ we find

$$\mathcal{F}(\overline{u}_\tau(T)) + \int_0^T\Big(\mathcal{R}\big(\underline{u}_\tau, \dot{\widehat{u}}_\tau\big) + \mathcal{R}^*\big(\underline{u}_\tau, -\overline{\xi}_\tau\big)\Big)\,\mathrm{d}t \leq \mathcal{F}(u^0) - \frac{\tau\lambda}{2}\int_0^T \|\dot{\widehat{u}}_\tau(t)\|^2\,\mathrm{d}t. \quad (3.11)$$

In the following we will show that we can pass to the limit $\tau \to 0^+$ in this discrete energy-dissipation inequality and thus find solutions.

3.3 The energy-dissipation principle

We have already seen in the previous subsection that in the ODE case we obtain the energy-dissipation balance (3.3), i.e. for all solutions u of the gradient-flow equation we have the Energy-Dissipation Inequality

$$\textbf{(EDI)} \qquad \mathcal{F}(u(T)) + \int_0^T \big(\mathcal{R}(u,\dot{u}) + \mathcal{R}^*(u, -\mathrm{D}\mathcal{F}(u))\big)\,\mathrm{d}t \leq \mathcal{F}(u(0)).$$

In fact, (3.3) gives the balance with "$=$" instead of "$\leq$", but we want to make the point that even the estimate is *equivalent* to solving the GFE

$$0 \in \partial\mathcal{R}(u,\dot{u}) + \mathrm{D}\mathcal{F}(u) \qquad \text{a.e. in } [0,T]. \quad (3.12)$$

The argument involves only the Fenchel theory and the chain rule. Indeed, by the chain rule, (EDB) can be rewritten as

$$\int_0^T \big(\mathcal{R}(u,\dot u)+\mathcal{R}^*(u,-\mathrm{D}\mathcal{F}(u)) + \langle \mathrm{D}\mathcal{F}(u),\dot u\rangle\big)\,\mathrm{d}t \leq 0.$$

However, by the Fenchel-Young estimate we know that the integrand is nonnegative. Thus, we conclude that it must be 0 a.e. in $[0,T]$. But this implies the power identity (3.2). But by the Fenchel equivalence this implies the GFE (3.12).

To make this argument also rigorous for the nonsmooth setting in infinite-dimensional Banach spaces, we need a corresponding *abstract chain rule*. At this point we simply give a definition that exactly provides what we need, and in Section 3.4 we then show that this condition can be obtained in the Banach-space setting under suitable conditions such as semiconvexity of $\mathcal{F}$.

Definition 3.8 (Abstract chain rule condition) *We say that a GS $(X,\mathcal{F},\mathcal{R})$ satisfies the (abstract) chain rule, if the following holds:*

$$\left.\begin{array}{l}\textit{If } u\in \mathrm{W}^{1,1}([0,T];X) \textit{ and } \xi\in \mathrm{L}^1([0,T];X^*) \textit{ satisfies } \sup\limits_{[0,T]}\big|\mathcal{F}(u(t))\big|<\infty,\\ \xi(t)\in\partial^{\mathrm{F}}\mathcal{F}(u(t)) \textit{ a.e. in } [0,T] \quad \textit{and} \quad \displaystyle\int_0^T\big(\mathcal{R}(u,\dot u)+\mathcal{R}^*(u,-\xi)\big)\,\mathrm{d}t<\infty,\\ \textit{then } t\mapsto \mathcal{F}(u(t)) \textit{ is absolutely continuous, } \big(t\mapsto\langle\xi(t),\dot u(t)\rangle\big)\in \mathrm{L}^1([0,T]),\\ \textit{and } \dfrac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(u(t)) = \langle \xi(t),\dot u(t)\rangle \quad \textit{a.e. in } [0,T].\end{array}\right\} \tag{3.13}$$

With this we are ready to state a precise version of the so-called energy-dissipation principle, which concerns roughly that solving the gradient-flow equation is equivalent to finding a function satisfying the energy-dissipation inequality (EDI). However, we warn the reader that sometimes the gradient-flow equation as a PDE may have solutions that do not have finite energy (cf. [SSZ12, Rem. 2.8]) and such solutions are not covered by this principle.

Several versions of the *Energy-Dissipation Principle* were used previously, see e.g. [Mie16, Thm. 3.3.1]. The following precise, but still very general version is due to Riccarda Rossi and Artur Stephan, see [MRS22].

Theorem 3.9 (The Energy-Dissipation Princple (EDP)) *Consider the generalized gradient system $(X,\mathcal{F},\mathcal{R})$ on a separable Banach space X that satisfies the abstract chain-rule condition* (3.13)*. Then, for all pairs $(u,\xi)\in \mathrm{W}^{1,1}([0,T];X)\times\mathrm{L}^1([0,T];X^*)$ with $\xi(t)\in\partial^{\mathrm{F}}\mathcal{F}(u(t))$ a.e. in $[0,T]$ the following two statements are equivalent:*

(A) (u,ξ) satisfies $\sup_{t\in[0,T]}\mathcal{F}(u(t))<\infty$ and

$$\text{(EDI)}\qquad \mathcal{F}(u(T))+\int_0^T\big(\mathcal{R}(u,\dot u)+\mathcal{R}^*(u,-\xi)\big)\,\mathrm{d}t\leq \mathcal{F}(u(0))<\infty. \tag{3.14}$$

(B) (u,ξ) satisfies the gradient-flow equation

$$0\in\xi(t)+\partial\mathcal{R}(u(t),\dot u(t)) \quad \textit{and} \quad \xi(t)\in\partial^{\mathrm{F}}\mathcal{F}(u(t)) \quad \textit{for a.a. } t\in[0,T] \tag{3.15}$$

and the energy-dissipation balance in $\mathcal{R}\mathcal{R}^*$ form:

$$\text{(EDB)} \qquad \mathcal{F}(u(t)) + \int_s^t \big(\mathcal{R}(u,\dot u) + \mathcal{R}^*(u,-\xi)\big)\,\mathrm{d}r = \mathcal{F}(u(s)) < \infty \tag{3.16}$$

for $0 \le s < t \le T$.

Proof. *(B) $\Longrightarrow$ (A).* This direction is trivial.

(A) $\Longrightarrow$ (B). We proceed exactly as in the ODE case. We start from (A). Because of $\mathcal{F}(u(0)) < \infty$ and $\mathcal{F}(u(T)) > -\infty$ we conclude that the finiteness of the dissipation integral in the $\mathcal{R}\mathcal{R}^*$ form in Assumption (3.13) is satisfied. Hence, we can apply the assumed abstract chain rule and rewrite $\mathcal{F}(u(T)) - \mathcal{F}(u(0))$ in the form $\int_0^T \langle \xi(t), \dot u(t)\rangle\,\mathrm{d}t$. Combining this with (EDI) we find

$$\int_0^T \Big(\mathcal{R}(u,\dot u) + \mathcal{R}^*(u,-\xi) + \langle \xi, \dot u\rangle\Big)\,\mathrm{d}t = -\delta \le 0,$$

where $\delta \ge 0$ is the gap (RHS minus LHS) in (EDI). By the Young-Fenchel inequality the integrand is nonnegative, hence we conclude $\delta = 0$ which means that (EDI) is in fact (EDB)$_{[0,T]}$ given in (3.16). Moreover, the nonnegative integrand must be 0 a.e. in $[0,T]$, which implies the identity $\mathcal{R}(u,\dot u) + \mathcal{R}^*(u,-\xi) = -\langle \xi, \dot u\rangle$. By the Fenchel equivalences this is equivalent to the GFE (3.15).

By the abstract chain rule we can also integrate the identity $\mathcal{R}(u,\dot u) + \mathcal{R}^*(u,-\xi) = -\langle \xi, \dot u\rangle$ on the subinterval $[s,t] \subset [0,T]$ and thus obtain (EDB)$_{[s,t]}$. ■

To illustrate the EDP we look at a very simple example, namely the Hilbert-space GS $(\mathrm{L}^2(\Omega), \mathcal{F}_{\mathrm{Dir}}, \mathbb{I}_{\mathrm{R}})$ with $\mathcal{F}_{\mathrm{Dir}}(u) = \frac{\alpha}{2}\|\nabla u\|^2_{\mathrm{L}^2}$ and $\mathrm{dom}(\mathcal{F}_{\mathrm{Dir}}) = \mathrm{H}^1_0(\Omega)$. Then we have $\xi = \partial^{\mathrm{F}}\mathcal{F}_{\mathrm{Dir}}(u) = -\alpha\Delta u$ and $\mathcal{R}^*(-\xi) = \frac{1}{2}\|\alpha\Delta u\|^2_{\mathrm{L}^2}$. Thus, (EDI) can be written in the form

$$\begin{aligned} 0 &\ge \frac{\alpha}{2}\|u(T)\|^2_{\mathrm{L}^2} + \int_0^T \big(\frac{1}{2}\|\dot u\|^2_{\mathrm{L}^2} + \frac{1}{2}\|\alpha\Delta u\|^2_{\mathrm{L}^2}\big)\,\mathrm{d}t - \frac{\alpha}{2}\|u(0)\|^2_{\mathrm{L}^2} \\ &= \int_0^T \big(\frac{1}{2}\|\dot u\|^2_{\mathrm{L}^2} + \frac{1}{2}\|\alpha\Delta u\|^2_{\mathrm{L}^2} + \langle -\alpha\Delta u, \dot u\rangle\big)\,\mathrm{d}t \;=\; \int_0^T \frac{1}{2}\big\|\dot u - \alpha\Delta u\big\|^2_{\mathrm{L}^2}\,\mathrm{d}t \end{aligned}$$

The major importance is that all terms in the left-hand side of (EDI) have good lower semicontinuity properties when passing to limits of approximating sequences. Hence starting from the discrete energy-dissipation inequality (3.11) it is reasonable to end up with (EDI) if suitable technical conditions hold, see Section 3.5. To finalize the proof we will then use the abstract chain rule to invoke the energy-dissipation principle to obtain solutions.

3.4 The abstract chain rule

We want $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(u(t)) = \langle \xi(t), \dot u(t)\rangle$ under as general as possible conditions. Can it work for nonsmooth energies?

Example 3.10 (Chain rule for nonsmooth $\mathcal{F}$) *We consider $X = \mathbb{R}^2$ and the nonsmooth, but convex functional $\mathcal{F}(u_1,u_2) = \max\{|u_1|,|u_2|\}$. For $u = (y,y)$ with $y > 0$ we obtain the set-valued subdifferential $\partial\mathcal{F}(u) = \big\{\,(\theta, 1{-}\theta) \in \mathbb{R}^2 \;\big|\; \theta \in [0,1]\,\big\}$.*

Thus, for the curve $u(t) = (y(t), y(t))$ *with* $y(t) > 0$ *we obtain elements in the subdifferential* $\xi(t) = (\theta(t), 1-\theta(t))$*, where* $\theta \in [0,1]$ *is completely arbitrary.*

Moreover, we have $f(t) = \mathcal{F}(u(t)) = y(t)$ *which implies*

$$\dot{y}(t) = \dot{f}(t) \overset{CR}{=} \langle \xi(t), \dot{u}(t)\rangle = \langle \binom{\theta}{1-\theta}, \binom{\dot{y}}{\dot{y}}\rangle = \dot{y}.$$

Indeed the chain rule holds, although $\xi \in \partial^{\mathrm{F}}\mathcal{F}(u(t))$ *is not unique.*

Example 3.11 (Classical Gelfand evolutionary triple) *For solving parabolic equation one often considers a so-called Gelfand triple* $V \overset{\mathrm{d}}{\subset} H \cong H^* \overset{\mathrm{d}}{\subset} V^*$*.*

By approximation the solutions u *are constructed with* $u \in \mathrm{W}^{1,2}([0,T]; V^*) \cap \mathrm{L}^2([0,T]; V)$*. A major step is then to show that this implies* $u \in \mathrm{C}^0([0,T]; H)$*.*

Sometimes one even shows that the mapping $t \mapsto \|u(t)\|_H^2$ *is absolutely continuous with*

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u(t)\|_H^2 = {}_{V^*}\langle \dot{u}(t), u(t)\rangle_V.$$

This is typically used when solving the diffusion equation $\dot{u} = \Delta u$ *with* $V = \mathrm{H}^1_0(\Omega)$*,* $H = \mathrm{L}^2(\Omega)$*, and* $V^* = \mathrm{H}^{-1}(\Omega)$*. Then*

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u\|_{\mathrm{L}^2}^2 = {}_{\mathrm{H}^{-1}}\langle \dot{u}, u\rangle_{\mathrm{H}^1_0} = {}_{\mathrm{H}^{-1}}\langle \Delta u, u\rangle_{\mathrm{H}^1_0} = -\int_\Omega |\nabla u|^2 \,\mathrm{d}x.$$

A more general chain rule was established in [Bré73, Lem. 3.3, p. 73] for general convex functionals on a Hilbert space (literal interpretation with "A" replaced by "$\partial\mathcal{F}$") :

> LEMMA 3.3. Let $u \in \mathrm{W}^{1,2}(0,T;H)$ be such that $u(t) \in \mathrm{dom}(\partial\mathcal{F})$ a.e. in $]0,T[$. Suppose there exists $g \in \mathrm{L}^2(0,T;H)$ such that $g(t) \in \partial\mathcal{F}(u(t))$ a.e. in $]0,T[$. Then the function $t \mapsto \mathcal{F}(u(t))$ is absolutely continuous.
>
> Denote by $\mathscr{T}$ the set of points $t \in \,]0,T[$ such that $u(t) \in \mathrm{dom}(\partial\mathcal{F})$ and that u and $\mathcal{F} \circ u$ are differentiable. Then, for all $t \in \mathscr{T}$ we have
>
> $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(u(t)) = \langle h, \dot{u}(t)\rangle \qquad$ for all $h \in \partial\mathcal{F}(u(t))$.

We will generalize such a result to λ-convex functionals on a Banach space. Our result is based on the theory developed in [MRS13, MiR23] which relies on ideas in [AGS05, Thm. 1.2.5]. Of course, the result in [MiR23, Prop. A.1] is much more general, in particular the condition of λ-convexity is weakened significantly.

The following result will use the *quantitative Young estimate* for the dissipation potential $\mathcal{R}$:

$$\exists\, c_{\mathrm{Y}}, C_{\mathrm{Y}} > 0\ \forall\, u, v \in X\ \forall\, \xi \in X^*: \quad \mathcal{R}(u,v) + \mathcal{R}^*(u,\xi) \geq c_{\mathrm{Y}}\|v\|_X\|\xi\|_{X^*} - C_{\mathrm{Y}}. \tag{3.17}$$

If $\mathcal{R}$ only depends on v through its norm, i.e. $\mathcal{R}(u,v) = \rho(u, \|v\|)$, then one has $c_{\mathrm{Y}} = 1$ and $C_{\mathrm{Y}} = 0$, see the discussion in [MiR23]. Another case where (3.17) holds is given when $\mathcal{R}$ has uniform upper and lower p-growth, namely $c\|v\|^p - C\mathcal{R}(u,v) \leq C\|v\|^p + C$. Then, $\mathcal{R}^*(u,\xi) \geq \widetilde{c}\|\xi\|^{p^*} - \widetilde{C}$ and (3.17) follows.

In the following result the main chain-rule property is totally independent of the dissipation potential $\mathcal{R}$, i.e. it is a property of $(X, \mathcal{F})$ alone. The connection to the gradient systems $(X, \mathcal{F}, \mathcal{R})$ is coming when we want to establish the integrability condition $\int_0^T \|\dot{u}\|_X \,\|\xi(t)\|_{X^*}\,\mathrm{d}t < \infty$ via the quantitative Young estimate (3.17).

Theorem 3.12 (Chain rule in Banach spaces) *On a reflexive Banach space X consider a proper, lsc, and semiconvex functional $\mathcal{F} : X \to \mathbb{R}_\infty$. Then, the following chain rule holds:*

If $u \in \mathrm{W}^{1,1}([0,T];X)$ and $\xi \in \mathrm{L}^1([0,T];X^*)$ satisfies $\sup_{[0,T]} \big|\mathcal{F}(u(t))\big| < \infty$,

$$\xi(t) \in \partial^{\mathrm{F}}\mathcal{F}(u(t)) \text{ a.e. in } [0,T], \quad \text{and} \quad \int_0^T \|\dot u\|_X \, \|\xi(t)\|_{X^*} \,\mathrm{d}t < \infty,$$

then $t \mapsto \mathcal{F}(u(t))$ is absolutely continuous and $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(u(t)) = \langle \xi(t), \dot u\rangle$ a.e.

In particular, if $\mathcal{R}$ satisfies the quantitative Young estimate (3.17)*, then the abstract chain rule* (3.13) *holds.*

Proof. To shorten the presentation we set $\Sigma := \big\{\, t \in [0,T] \,\big|\, \partial^{\mathrm{F}}\mathcal{F}(u(t)) \neq \emptyset \,\big\}$ and abbreviate $f(t) := \mathcal{F}(u(t))$. By assumption Σ is a set of full measure.

Step 1: Absolute integrability of $\widetilde f$ under arc-length parametrization. We first consider the case that $\|\dot u(t)\| = 1$ a.e. in $[0,T]$. Then, we immediately have $\|u(t_1)-u(t_0)\| \le |t_1-t_0|$.

Choosing arbitrary $t_{j-1} < t_j$ in Σ we use λ-convexity of $\mathcal{F}$ to obtain

$$\begin{aligned} f(t_j) - f(t_{j-1}) &\ge \langle \xi(t_{j-1}), u(t_j)-u(t_{j-1})\rangle + \frac{\lambda}{2}\|u(t_j)-u(t_{j-1})\|^2 \\ &\ge -\|\xi(t_{j-1})\|\,(t_j-t_{j-1}) + \frac{\lambda}{2}|t_j-t_{j-1}|^2 \quad \text{and} \\ f(t_{j-1}) - f(t_j) &\ge \ -\|\xi(t_j)\|\,(t_j-t_{j-1}) \ + \frac{\lambda}{2}|t_j-t_{j-1}|^2, \end{aligned}$$

where the second inequality follows from the first by interchanging t_{j-1} and t_j.

For an arbitrary interval $[s,t]$ with $s,t \in \Sigma$ we choose partitions $s = t_0 < t_1 < \cdots < t_N = t$ with $t_j \in \Sigma$ and add up the inequalities which leads to

$$\begin{aligned} \sum_{j=1}^N \big(-\|\xi(t_{j-1})\|(t_j-t_{j-1}) + \frac{\lambda}{2}|t_j-t_{j-1}|^2\big) &\le f(t) - f(s) \\ &\le \sum_{j=1}^N \big(\|\xi(t_j)\|(t_j-t_{j-1}) - \frac{\lambda}{2}|t_j-t_{j-1}|^2\big). \end{aligned} \tag{3.18}$$

By a refined theory of the Riemann integral for L^1 functions (see [DFT05, Sec. 4.4] and [Hah15] for the historic origin) it can be shown that it is always possible to choose a sequence of partitions with fineness tending to 0 such that the limit of Riemann sums equals the Lebesgue integral. (There ξ is defined everywhere, and we can set $\xi(r) = 0$ for $r \notin \Sigma$.) Hence, we conclude

$$-\int_s^t \|\xi(r)\|\,\mathrm{d}r \le f(t) - f(s) \le \int_s^t \|\xi(r)\|\,\mathrm{d}r.$$

Thus, we have established $|f(t) - f(s)| \le \int_s^t \|\xi(r)\|\,\mathrm{d}r$ for all $s,t \in \Sigma \subset [0,T]$. Because of $\xi \in \mathrm{L}^1([0,T];X^*)$ this shows that there is a absolutely continuous function $\widetilde f \in \mathrm{W}^{1,1}([0,T]) \cap \mathrm{C}^0([0,T])$ satisfying $f(t) = \widetilde f(t)$ for all $t \in \Sigma$.

Step 2. $f = \widetilde{f}$ under arclength parametrization. We continue under the same conditions as in Step 1 and show $f(t) = \widetilde{f}(t)$ for all $t \in [0,T]$. By continuity of $t \mapsto u(t) \in X$ and lsc of $\mathcal{F}$ we know that f is lower semicontinuous, which implies $f(t) \leq \widetilde{f}(t)$ for all $t \in [0,T]$.

To show the opposite inequality we restrict to $t \in [0, 2T/3]$ and define for $r \in\,]0, T/3[$ the averages

$$f_r(t) = \frac{1}{r}\int_t^{t+r} f(s)\,\mathrm{d}s = \frac{1}{r}\int_t^{t+r} \widetilde{f}(s)\,\mathrm{d}s \;\to\; \widetilde{f}(t) \quad \text{for } r \to 0^+.$$

(For $t \in [T/3, T]$ one can proceed analogously by taking backward averages $f_r(t) = \frac{1}{r}\int_{t-r}^t f(s)\,\mathrm{d}s$.) Here f_r is well defined, because f is bounded by assumption and lsc, hence (Borel) measurable and integrable. Thus, it suffices to show $f(t) \geq \limsup_{r\to 0^+} f_r(t) = \widetilde{f}(t)$. For this we proceed as above and obtain

$$\begin{aligned} f(t) - f_r(t) &= \frac{1}{r}\int_t^{t+r}\big(f(t) - \widetilde{f}(s)\big)\,\mathrm{d}s \geq \frac{1}{r}\int_t^{t+r}\Big(-\|\xi(s)\|\,|t-s| - \frac{|\lambda|}{2}\,|t-s|^2\Big)\,\mathrm{d}s\Big) \\ &\geq \frac{1}{r}\int_t^{t+r}\Big(-\|\xi(s)\|\,r - \frac{|\lambda|}{2}\,r^2\Big)\,\mathrm{d}s = -\int_t^{t+r}\|\xi(s)\|\,\mathrm{d}s - \frac{|\lambda|}{2}\,r^2 \to 0 \ \text{ for } r \to 0^+. \end{aligned}$$

Thus, we conclude that $t \mapsto f(t) = \mathcal{F}(u(t))$ is equal to the continuous representative $\widetilde{f}$, and the desired absolute continuity of f is shown under the assumption $\|\dot{u}(t)\| = 1$ a.e.

Step 3: Reparametrization. For the general case with $\dot{u} \in \mathrm{L}^1([0,T];X)$ we follow [AGS05, Lem. 1.1.4] and consider the reparametrization

$$\sigma(t) = \int_0^t \|\dot{u}(r)\|\,\mathrm{d}r \ \text{ giving } \sigma : [0,T] \to [0,\ell],$$

where $\ell = \sigma(T) = \int_0^T \|\dot{u}(r)\|\,\mathrm{d}r$. Clearly, $\sigma \in \mathrm{W}^{1,1}([0,T]) \subset \mathrm{C}^0([0,T])$ and $\sigma'(t) \geq 0$. We define the inverse

$$\tau : \left\{ \begin{array}{ccc} [0,\ell] & \to & [0,T], \\ s & \mapsto & \min\big\{\, t \in [0,T] \,\big|\, \sigma(t) = s \,\big\}, \end{array} \right.$$

which is increasing and continuous from the left such that $\sigma(\tau(s)) = s$ for all $s \in [0,\ell]$. Moreover, we have

$$\tau(\sigma(t)) \leq t \quad \text{and} \quad u(\tau(\sigma(t))) = u(t) \quad \text{for all } t \in [0,T]. \tag{3.19}$$

For the second relation note that on intervals $]t_0, t_1[$ where $t_0 = \tau(\sigma(t_0)) = \tau(\sigma(t)) < t$ we have $\dot{u}(t) = 0$ giving $u(t) = u(t_0)$.

With this we define $\widehat{u}(s) = u(\tau(s))$, and for $0 \leq s_0 < s_1 \leq \ell$ we have

$$\|\widehat{u}(s_1) - \widehat{u}(s_0)\| = \|u(\tau(s_1)) - u(\tau(s_0))\| \leq \int_{\tau(s_0)}^{\tau(s_1)} \|\dot{u}(r)\|\,\mathrm{d}r = \sigma(\tau(s_1)) - \sigma(\tau(s_0)) = s_1 - s_0.$$

Thus, $\widehat{u}$ is 1-Lipschitz. Moreover, the reflexivity of X gives $\mathrm{C}^{\mathrm{Lip}}([0,T];X) = \mathrm{W}^{1,\infty}([0,T];X)$, such that the derivative $\widehat{u}'(s)$ exists and $\|\widehat{u}'(s)\| \leq 1$ a.e. in $[0,\ell]$.

Moreover, for $0 \leq t_0 < t_1 \leq T$ we find

$$\|u(t_1) - u(t_0)\| = \|\widehat{u}(\sigma(t_1)) - \widehat{u}(\sigma(t_0))\| \leq \int_{\sigma(t_0)}^{\sigma(t_1)} \|\widehat{u}'(\rho)\|\,\mathrm{d}\rho = \int_{t_0}^{t_1} \|\widehat{u}'(\sigma(r))\|\dot{\sigma}(r)\,\mathrm{d}r.$$

This implies $\|\dot u(t)\| \leq \|\widehat{u}'(\sigma(t))\|\dot\sigma(t)$ for a.a. $t \in [0,T]$. Using $\dot\sigma(t) = \|\dot u(t)\|$ and $\|\widehat{u}'(s)\| \leq 1$ from above we find $\|\widehat{u}'(s)\| = 1$ a.e. in $[0,\ell]$.

Step 4. Absolute continuity of f in the general case. We now apply the reparametrization from the previous step also to f and ξ by setting $\widehat{f}(s) = f(\tau(s))$ and $\widehat{\xi}(s) = \xi(\tau(s))$ and obtain $f(t) = \widehat{f}(\sigma(t)) = \mathcal{F}(u(\tau(\sigma(t))))$ by using $u(\tau(\sigma(t))) = u(t)$ from (3.19). Moreover, we have

$$\int_0^\ell \|\widehat{\xi}(s)\| \, \mathrm{d}s \overset{\mathrm{tr}}{=} \int_0^T \|\widehat{\xi}(\sigma(t))\|\dot\sigma(t) \, \mathrm{d}t = \int_0^T \|\xi(\tau(\sigma(t)))\| \, \|\dot u(t)\| \, \mathrm{d}t \overset{*}{=} \int_0^T \|\xi(t)\| \, \|\dot u(t)\| \, \mathrm{d}t < \infty.$$

(For a justification of the transformation rule in "$\overset{\mathrm{tr}}{=}$" with $s = \sigma(t)$ we refer to [Bog07, Thm. 5.8.30] and note that the absolute continuous function σ satisfies the Lusin property (N).) The last bound is simply the assumption, whereas in $\overset{*}{=}$ we use that $\dot u(t) = 0$ whenever $\tau(\sigma(t)) \neq t$, see the comments after (3.19). Hence we have $\widehat{\xi} \in \mathrm{L}^1([0,\ell])$ as well as $\|\widehat{u}'(s)\| = 1$ a.e. in $[0,\ell]$. Thus, we can apply Step 2 and find $|\widehat{f}(s_1) - \widehat{f}(s_0)| \leq \int_{s_0}^{s_1} \|\widehat{\xi}(s)\| \, \mathrm{d}s$ and conclude via

$$\begin{aligned} |f(t_1) - f(t_0)| = |\widehat{f}(\sigma(t_1)) - \widehat{f}(\sigma(t_0))| &\leq \int_{\sigma(t_0)}^{\sigma(t_1)} \|\widehat{\xi}(s)\| \, \mathrm{d}s \\ = \int_{t_0}^{t_1} \|\widehat{\xi}(\sigma(t))\|\dot\sigma(t) \, \mathrm{d}t &= \int_{t_0}^{t_1} \|\xi(t)\| \, \|\dot u(t)\| \, \mathrm{d}t, \end{aligned}$$

which is the desired absolute integrability of $f : t \mapsto f(t) = \mathcal{F}(u(t))$ as $t \mapsto \|\dot u\| \, \|\xi\|$ lies in $\mathrm{L}^1([0,T])$.

Step 5: Identification of the derivative. As f is differentiable a.e. in $[0,T]$ the set $\mathbb{T} \subset \,]0,T[$ on which $u : [0,T] \to H$ is differentiable, $f : [0,T] \to \mathbb{R}$ is differentiable, and $\partial^{\mathrm{F}}\mathcal{F}(u(t))$ is nonempty is of full measure. Now take $t \in \mathbb{T}$ and choose an arbitrary $\eta \in \partial^{\mathrm{F}}\mathcal{F}(u(t))$. Then, for all $h \in [-t, T-t]$ we have

$$f(t+h) - f(t) \geq \langle \eta, u(t+h) - u(t) \rangle + \frac{\lambda}{2}\big\|u(t+h) - u(t)\big\|^2.$$

Dividing by $h > 0$ and taking the limit $h \to 0^+$ we find $\dot f(t) \geq \langle \eta, \dot u(t) \rangle$. Dividing by $h < 0$ and taking the limit $h \to 0^-$ gives the opposite estimate. Hence we have shown $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(u(t)) = \dot f(t) = \langle \eta, \dot u(t) \rangle$ for all $t \in \mathbb{T}$, and the chain rule is established.

Step 6: Abstract chain rule. Starting from $\int_0^T \big(\mathcal{R}(u,\dot u) + \mathcal{R}^*(u,-\xi)\big) \, \mathrm{d}t < \infty$ and the quantitative Young estimate (3.17) we obtain

$$\int_0^T \|\dot u\|_X \|\xi\|_{X^*} \, \mathrm{d}t \leq \frac{1}{c_{\mathrm{Y}}}\Big(\int_0^T \big(\mathcal{R}(u,\dot u) + \mathcal{R}^*(u,-\xi)\big) \, \mathrm{d}t + C_{\mathrm{Y}}\Big) \; < \; \infty.$$

Thus the above results are applicable and we obtain (3.13). ■

3.5 Existence theory via time-incremental minimization

Our existence theory indeed follows very similar steps as in the Hilbert-space setting. The difference is that we are now using the energy-dissipation principle, i.e. we do not work with the evolutionary equation directly. We rather exploit the favorable structure of the energy-dissipation inequality, which allows us to pass to the limit by arguments of the calculus of variations.

We emphasize that uniqueness of solutions cannot be expected in this general setting. Even if we are able to obtain uniqueness of the incremental minimizers u_k^τ we cannot expect the continuous solutions to be unique because of the doubly nonlinear structure.

We start by collecting a set of sufficient condition that allow us to study a large class of generalized gradient systems. However, the assumptions are restricted for didactic reasons, and we will comment on possible extensions and generalizations in the next subsection.

For the following list of conditions we recall the sublevels $S_E^{\mathcal{F}} = \{ u \in X \mid \mathcal{F}(u) \leq E \}$.

X is a separable reflexive Banach space. (3.20a)

$\mathcal{F} : X \to \mathbb{R}_\infty$ is semiconvex and has compact sublevels. (3.20b)

$$\left.\begin{array}{l} \mathcal{R} : X{\times}X \to [0,\infty] \text{ and } \mathcal{R}^* : X{\times}X^* \to [0,\infty] \text{ are} \\ \text{uniformly superlinear on sublevels, i.e. } \forall\, E \in \mathbb{R} \\ \exists \text{ increasing, convex, superlinear } \psi_E : [0,\infty[\,\to [0,\infty[\text{ such that} \\ \forall\,(u,v,\xi) \in S_E^{\mathcal{F}}{\times}X{\times}X^* : \ \ \mathcal{R}(u,v) \geq \psi_E(\|v\|) \text{ and } \mathcal{R}^*(u,\xi) \geq \psi_E(\|\xi\|). \end{array}\right\} \tag{3.20c}$$

$(X, \mathcal{F}, \mathcal{R})$ satisfies the abstract chain rule condition (3.13). (3.20d)

$$\left.\begin{array}{l} \text{For each level } E \in \mathbb{R} \text{ there exists a modulus of continuity } \omega_E^{\mathcal{R}} \\ \text{such that } \forall\, u_0, u_1 \in S_E^{\mathcal{F}}\ \forall\, v \in X : \\ \big|\mathcal{R}(u_1,v) - \mathcal{R}(u_0,v)\big| \ \leq\ \omega_E^{\mathcal{R}}\big(\|u_1{-}u_0\|\big)\,\big(1 + \mathcal{R}(u_0,v)\big). \end{array}\right\} \tag{3.20e}$$

We emphasize that the semiconvexity condition in (3.20b) is rather strong: First, it allows us to derive the discrete approximation of the energy-dissipation inequality. Secondly, by our results in Section 2.2 it implies the important condition of closedness of the Fréchet subdifferential $\partial^{\mathrm{F}}\mathcal{F} : X \rightrightarrows X^*$. Thirdly, it is a very helpful condition of establishing the chain rule, see Theorem 3.12.

The upcoming existence result is now based on the time-incremental minimization scheme (3.6), the four associated interpolants $\widehat{u}_\tau$, $\underline{u}_\tau$, $\overline{u}_\tau$, and $\overline{\xi}_\tau$ (see (3.9)), and the discrete EDI (3.11). The proof follows similar steps as the existence proof in the Hilbert-space setting, but now in the last step we exploit the Energy-Dissipation Principle, where the quantitative Young estimate (3.20d) is needed to provide the abstract chain rule.

Theorem 3.13 (Existence for $(X, \mathcal{F}, \mathcal{R})$) *Consider a generalized GS $(X, \mathcal{F}, \mathcal{R})$ satisfying the assumptions (3.20). Then, for all $u_0 \in X$ with $\mathcal{F}(u^0) < \infty$ there exists a solution $(u, \xi) \in \mathrm{W}^{1,1}([0,T]; X){\times}\mathrm{L}^1([0,T]; X^*)$ satisfying $u(0) = u^0$, the gradient-flow equation*

$$0 \in \xi(t) + \partial\mathcal{R}(u(t), \dot u(t)) \quad \textit{and} \quad \xi(t) \in \partial^{\mathrm{F}}\mathcal{F}(u(t)) \quad \textit{for a.a. } t \in [0,T], \tag{3.21}$$

and the energy-dissipation balance $\mathcal{F}(u(t)) + \int_s^t \big(\mathcal{R}(u,\dot u){+}\mathcal{R}^*(u,-\xi)\big)\,\mathrm{d}r = \mathcal{F}(u(s))$ for $0 \leq s < t \leq T$.

Before starting the full proof, we provide a few auxiliary results that are useful but are also of independent interest. First we recall that the Legendre transformation $\mathfrak{L}$ is antimonotone, which implies that lower (upper) bounds for $\mathcal{R}$ imply upper (lower) bounds for $\mathcal{R}^*$ and vice versa.

The lower bounds for $\mathcal{R}$ and $\mathcal{R}^*$ in (3.20e) hence imply the upper bounds

$$\mathcal{R}(u,v) \leq \psi_E^*(\|v\|) \quad \text{and} \quad \mathcal{R}^*(u,\xi) \leq \psi_E^*(\|\xi\|),$$

where $\psi^*(\zeta) = \sup\{ z\zeta - \psi(z) \mid z \geq 0 \}$. As ψ is finite everywhere, ψ^* is again increasing, convex, and superlinear. As examples we can keep in mind $\psi(z) = cz^p - C$ for $p > 1$ giving $\psi^*(\zeta) = \widetilde{c}\zeta^{p^*} + C$ or $\psi(z) = (z{+}1)\log(z{+}1) - C$ giving $\psi^*(\zeta) = C + \mathrm{e}^{\zeta-1} - \mathrm{e}^{-1}$. Since $\mathcal{R}$ and $\mathcal{R}^*$ are upper and lower bounded on each ball $B_R(0)$, they are even Lipschitz continuous with bounded subdifferentials.

The continuity of $\mathcal{R}$ in (3.20e) also provides upper and lower bounds of $\mathcal{R}(u_1,\cdot)$ in terms of $\mathcal{R}(u_0,\cdot)$, namely

$$-\omega_{1,0} + (1{-}\omega_{1,0})\mathcal{R}(u_0,v) \leq \mathcal{R}(u_1,v) \leq \omega_{1,0} + (1{+}\omega_{1,0})\mathcal{R}(u_0,v), \tag{3.22}$$

where $\omega_{1,0} = \omega_E^{\mathcal{R}}\big(\|u_1{-}u_0\|\big)$. The upper bound for $\mathcal{R}(u_1,\cdot)$ transforms into a lower bound for $\mathcal{R}^*(u_1,\cdot)$, namely

$$\mathcal{R}^*(u_1,\xi) \geq -\omega_{1,0} + (1{+}\omega_{1,0})\,\mathcal{R}\big(u_0, \frac{1}{1{+}\omega_{1,0}}\,\xi\big). \tag{3.23}$$

If for $w \in \mathrm{W}^{1,1}([0,T];X)$ we have the superlinear bound $B := \int_0^T \psi\big(\|\dot w\|\big)\,\mathrm{d}t < \infty$, we obtain an explicit equicontinuity.

$$\|w(t){-}w(s)\| \leq \frac{1}{\mu}\int_s^t \mu\|\dot w\|\,\mathrm{d}t \leq \frac{1}{\mu}\int_s^t \big(\psi^*(\mu) + \psi(\|\dot w\|)\big)\,\mathrm{d}t \leq \big((t{-}s)\psi^*(\mu) + B\big)/\mu.$$

Taking the infimum over $\mu > 0$, we obtain the desired result, namely

$$\big\|w(t){-}w(s)\big\| \leq \omega_\psi^B\big(|t{-}s|\big) \quad \text{where } \omega_\psi^B(r) := \inf\Big\{ \frac{1}{\mu}\big(r\psi^*(\mu) + B\big) \;\Big|\; \mu > 0 \Big\}. \tag{3.24}$$

For every $B > 0$ the function ω_ψ^B is a modulus of continuity, i.e., $\omega_\psi^B(r) \to 0$ for $r \to 0^+$.

Proof of Theorem 3.13. The proof consists of the typical steps.

Step 0: construction of approximations via time-incremental minimization. We first show that scheme in (3.6) has minimizer $u_k = u_k^\tau$ for all $k = 1,\ldots,N$. For this we use that $\mathcal{F}$ is lsc (because of closed sublevels) and that $u \mapsto \tau\mathcal{R}\big(u_{k-1}, \frac{1}{\tau}(u{-}u_{k-1})\big)$ is continuous and coercive. Hence, $\Phi_\tau^{\mathcal{F},\mathcal{R}}(u_{k-1};\cdot)$ is lsc and coercive. Moreover, the sublevels are compact, as they are contained in a sublevel of $\mathcal{F}$. Hence, by the one-sided Weierstraß extremal principle a minimizer u_k^τ exists, namely

$$\forall\, w \in X: \quad \tau\mathcal{R}\big(u_{k-1}, \frac{1}{\tau}(u_k{-}u_{k-1})\big) + \mathcal{F}(u_k) \leq \tau\mathcal{R}\big(u_{k-1}, \frac{1}{\tau}(w{-}u_{k-1})\big) + \mathcal{F}(w). \tag{3.25}$$

As $\mathcal{R}(u,\cdot)$ is convex and continuous, the sum rule gives $\partial^{\mathrm{F}}\Phi(u_*;u) = \partial_v\mathcal{R}\big(u_*, \frac{1}{\tau}(u{-}u_*)\big) + \partial^{\mathrm{F}}\mathcal{F}(u)$ and we obtain the inclusion $0 \in \partial_v\mathcal{R}\big(u_{k-1}, \frac{1}{\tau}(u_k{-}u_{k-1})\big) + \partial^{\mathrm{F}}\mathcal{F}(u_k)$ or equivalently

$$-\xi_k \in \partial_v\mathcal{R}\big(u_{k-1}, \frac{1}{\tau}(u_k{-}u_{k-1})\big) \quad \text{and} \quad \xi_k \in \partial^{\mathrm{F}}\mathcal{F}(u_k), \quad k = 1,\ldots,N. \tag{3.26}$$

Step 1: a priori estimates. Testing (3.25) with $w = u_{k-1}$ we immediately conclude $\mathcal{F}(u_k) \leq \mathcal{F}(u_{k-1}) \leq \mathcal{F}(u^0) =: F_0 < \infty$. Hence, all u_k lie in the compact sublevel $S_{F_0}^{\mathcal{F}} \Subset X$. The same test of (3.25) also provides a bound on increments, namely

$$\tau\psi\big(\frac{1}{\tau}\|u_k{-}u_{k-1}\|\big) \leq \tau\mathcal{R}\big(u_{k-1}, \frac{1}{\tau}(u_k{-}u_{k-1})\big)\big) \leq \mathcal{F}(u_k) - \mathcal{F}(u_{k-1}). \tag{3.27}$$

To obtain a supremum bound we divide by τ and estimate the energies:

$$\psi\big(\frac{1}{\tau}\|u_k-u_{k-1}\|\big) \leq \frac{1}{\tau}\big(\mathcal{F}(u^0) - F_{\min}\big), \quad \text{where } F_{\min} := \min_X \mathcal{F} > -\infty. \tag{3.28}$$

For an integral bound we sum (3.27) over $k = 1, \ldots, N$ to obtain

$$\int_0^T \psi\big(\|\dot{\widehat{u}}_\tau\|\big)\,\mathrm{d}t = \sum_{k=1}^N \tau\psi\big(\frac{1}{\tau}\|u_k-u_{k-1}\|\big) \leq \Delta_\mathcal{F} := \mathcal{F}(u^0) - F_{\min} < \infty, \tag{3.29}$$

where $\widehat{u}_\tau$ is the piecewise affine interpolant. Hence, (3.24) gives $\|\widehat{u}_\tau(t)-\widehat{u}_\tau(s)\| \leq \omega_\psi^{\Delta_\mathcal{F}}(|t-s|)$ for all $t, s \in [0, T]$.

We also observe that $\dot{\widehat{u}}_\tau$ is piecewise constant such that $\tau\dot{\widehat{u}}_\tau(t) = u_k^\tau-u_{k-1}^\tau$ for $t \in \,]k\tau-\tau, k\tau[$. Thus, we find $\tau\|\dot{\widehat{u}}_\tau\|_{\mathrm{L}^\infty([0,T];X)} = \max\big\{\, \|\widehat{u}_\tau(k\tau)-\widehat{u}_\tau(k\tau-\tau)\| \;\big|\; k = 1, ..., N \,\big\} \leq \omega_\psi^{\Delta_\mathcal{F}}(\tau) \to 0$ for $\tau \to 0$. Later we will need the following estimate:

$$\begin{aligned} E(\tau) &:= \tau \int_0^T \|\dot{\widehat{u}}_\tau\|^2\,\mathrm{d}t \leq \tau\|\dot{\widehat{u}}_\tau\|_{\mathrm{L}^\infty} \int_0^T 1\,\|\dot{\widehat{u}}_\tau\|\,\mathrm{d}t \\ &\leq \omega_\psi^{\Delta_\mathcal{F}}(\tau)\ \int_0^T \Big(\psi^*(1) + \psi\big(\|\dot{\widehat{u}}_\tau\|\big)\Big)\,\mathrm{d}t \leq \omega_\psi^{\Delta_\mathcal{F}}(\tau)\ \big(T\psi^*(1)+\Delta_\mathcal{F}\big), \end{aligned} \tag{3.30}$$

such that $E(\tau) \to 0$ for $\tau \to 0^+$.

To obtain an a priori estimate on the dual variable ξ_k, we proceed as at the end of Section 3.2 where we use again the interpolants $\overline{u}_\tau$, $\underline{u}_\tau$, and $\overline{\xi}_\tau$ and obtained the discrete approximate energy-dissipation inequality (3.11), namely

$$\mathcal{F}(\widehat{u}_\tau(T)) + \int_0^T \Big(\mathcal{R}\big(\underline{u}_\tau, \dot{\widehat{u}}_\tau\big)+\mathcal{R}^*\big(\underline{u}_\tau, -\overline{\xi}_\tau\big)\Big)\,\mathrm{d}t \leq \mathcal{F}(u^0) - \frac{\lambda}{2}\,E(\tau), \quad \overline{\xi}_\tau \in \partial^\mathrm{F}\mathcal{F}(\overline{u}_\tau). \tag{3.31}$$

This immediately implies

$$\int_0^T \Big(\psi\big(\|\dot{\widehat{u}}_\tau\|\big) + \psi\big(\|\overline{\xi}_\tau\|\big)\Big)\,\mathrm{d}t \leq \Delta_\mathcal{F} + \frac{|\lambda|}{2}\,E(\tau) \leq \Delta_\mathcal{F} + 1 < \infty, \tag{3.32}$$

for $0 < \tau \ll 1$. Of course, from $u_k^\tau \in S_{F_0}^\mathcal{F} \Subset X$ we also have the a priori estimates

$$\|\widehat{u}_\tau\|_{\mathrm{L}^\infty([0,T];X)} \leq R, \quad \|\overline{u}_\tau\|_{\mathrm{L}^\infty([0,T];X)} \leq R, \quad \|\underline{u}_\tau\|_{\mathrm{L}^\infty([0,T];X)} \leq R.$$

Step 2: extraction of convergent subsequences. As all $\widehat{u}_\tau$ satisfy the uniform bound (3.29) we obtain equi-continuity via (3.24):

$$\|\widehat{u}_\tau(t) - \widehat{u}_\tau(s)\| \leq \omega_\psi^B\big(|t-s|\big) \quad \text{with } B = \Delta_\mathcal{F}.$$

As the interpolants $\widehat{u}_\tau$, $\overline{u}_\tau$, and $\underline{u}_\tau$ coincide for $t = k\tau$ we conclude

$$\widetilde{\omega}(\tau) := \|\widehat{u}_\tau - \overline{u}_\tau\|_{\mathrm{L}^\infty([0,T];X)} + \|\widehat{u}_\tau - \underline{u}_\tau\|_{\mathrm{L}^\infty([0,T];X)} \leq 2\omega_\psi^B(\tau) \to 0 \quad \text{for } \tau \to 0^+.$$

Moreover, exploiting the compactness of the sublevel $S_{F_0}^\mathcal{F}$ in X, we can apply the Arzelà-Ascoli selection principle to $(\widehat{u}_\tau)_\tau$ and obtain a subsequence (not relabeled) and a limit function $u \in \mathrm{C}^0([0, T]; X)$ such that

$$\widehat{u}_\tau \to u, \quad \overline{u}_\tau \to u, \quad \underline{u}_\tau \to u \quad \text{in } \mathrm{C}^0([0, T]; X).$$

Moreover, since ψ in (3.32) is superlinear, the criterion of de la Vallée Poussin shows that $(\dot{\widehat{u}}_\tau)_\tau$ and $(\overline{\xi}_\tau)_\tau$ are uniformly equi-integrable families in $\mathrm{L}^1([0,T];X)$ and $\mathrm{L}^1([0,T];X^*)$, respectively. Hence there exists a further subsequence (again not relabeled) and limits $v \in \mathrm{L}^1([0,T];X)$ and $\xi \in \mathrm{L}^1([0,T];X^*)$ such that

$$\dot{\widehat{u}}_\tau \rightharpoonup v \ \text{ in } \mathrm{L}^1([0,T];X) \quad \text{and} \quad \overline{\xi}_\tau \rightharpoonup \xi \ \text{ in } \mathrm{L}^1([0,T];X^*).$$

Choosing a test function $\eta \in \mathrm{C}^1_\mathrm{c}(]0,T[;X^*)$ we can pass to the limit $\tau \to 0^+$ in the identity $\int_0^T \langle \eta, \dot{\widehat{u}}_\tau\rangle \,\mathrm{d}t = -\int_0^T \langle \dot\eta, \widehat{u}_\tau\rangle \,\mathrm{d}t$ and find $v = \dot u$. Thus, we have $\widehat{u}_\tau \rightharpoonup u$ in $\mathrm{W}^{1,1}([0,T];X)$ (along the subsequence chosen above).

Step 3: derivation of (EDI). We derive (EDI) by passing to the limit $\tau \to 0^+$ in (3.31).

(3.a) Because of $E(\tau) \to 0$ (cf. (3.30)) the right-hand side in (3.31) converges to the desired limit $\mathcal{F}(u^0) = F_0$.

On the left-hand side we treat the three terms separately and note that it is sufficient to derive a liminf estimate.

(3.b) By $\widehat{u}_\tau(T) \to u(T)$ and lower semicontinuity $\mathcal{F}(u(T)) \leq \liminf_{\tau\to0} \mathcal{F}(\widehat{u}_\tau(T))$.

(3.c) For the rate term $\mathcal{R}$, we use the lower bound (3.22) with $u_0 = u(t)$, $u_1 = \underline{u}_\tau(t)$, and $v = \dot{\widehat{u}}_\tau(t)$. Note that $\underline{u}_\tau(t)$ and $u(t)$ lie in the sublevel $S^\mathcal{F}_{F_0}$, such that we can apply (3.22) with $\omega_{1,0} = \omega^\mathcal{R}_{F_0}\big(\|\underline{u}_\tau(t){-}u(t)\|_{\mathrm{L}^\infty}\big) \leq \widehat\omega(\tau) := \omega^\mathcal{R}_{F_0}\big(\|\underline{u}_\tau{-}u\|_{\mathrm{L}^\infty}\big) \to 0$ for $\tau \to 0$. With this we obtain the estimate

$$\begin{aligned}\liminf_{\tau\to0}\int_0^T \mathcal{R}\big(\underline{u}_\tau(t),\dot{\widehat{u}}_\tau(t)\big)\,\mathrm{d}t &\geq \liminf_{\tau\to0}\int_0^T\Big({-}\widehat\omega(\tau) + \big(1{-}\widehat\omega(\tau)\big)\mathcal{R}\big(u(t),\dot{\widehat{u}}_\tau(t)\big)\Big)\,\mathrm{d}t\\ = \liminf_{\tau\to0}\int_0^T\Big({-}0 + 1\,\mathcal{R}\big(u(t),\dot{\widehat{u}}_\tau(t)\big)\Big)\,\mathrm{d}t &\geq \int_0^T \mathcal{R}\big(u(t),\dot u(t)\big)\,\mathrm{d}t,\end{aligned}$$

where in the last estimate we used the weak lower semicontinuity following from the convexity of the primal dissipation potential $\mathcal{R}(u,\cdot)$.

(3.d) For the $\mathcal{R}^*$ term we proceed analogously now relying on (3.23):

$$\begin{aligned}\liminf_{\tau\to0}\int_0^T \mathcal{R}^*\big(\underline{u}_\tau(t),-\overline{\xi}_\tau(t)\big)\,\mathrm{d}t &\geq \liminf_{\tau\to0}\int_0^T\Big({-}\widehat\omega(\tau) + \big(1{-}\widehat\omega(\tau)\big)\mathcal{R}^*\big(u(t),\frac{-1}{1{-}\widehat\omega(\tau)}\overline{\xi}_\tau(t)\big)\Big)\,\mathrm{d}t\\ = \liminf_{\tau\to0}\int_0^T\Big({-}0 + 1\,\mathcal{R}^*\big(u(t),\frac{-1}{1{-}\widehat\omega(\tau)}\overline{\xi}_\tau(t)\big)\Big)\,\mathrm{d}t &\geq \int_0^T \mathcal{R}^*\big(u(t),-\xi(t)\big)\,\mathrm{d}t.\end{aligned}$$

Combining the results of (3.a-d) we obtain the desired EDI

$$\mathcal{F}(u(T)) + \int_0^T \big(\mathcal{R}(u,\dot u) + \mathcal{R}^*(u,-\xi)\big)\,\mathrm{d}t \leq \mathcal{F}(u^0).$$

Clearly, we still have $u(0) = u^0$ and it remains to identify ξ. We recall that for all τ we have $\overline{\xi}_\tau(t) \in \partial^\mathrm{F}\mathcal{F}(\overline{u}_\tau(t))$ for a.a. $t \in [0,T]$. Since $\overline{u}_\tau \to u$ and $\overline{\xi}_\tau \rightharpoonup \xi$, we can use the strong-weak closedness of $\partial^\mathrm{F}\mathcal{F}$ and obtain $\xi(t) \in \partial^\mathrm{F}\mathcal{F}(u(t))$ for a.a. $t \in [0,T]$.

We remark here that this we need a generalization of the approach in Step 3 (see page 19) the proof of Theorem 2.7, which relies on the result in Exercise 2.6. Instead we can exploit the result of Exercise 2.7, which only needs $\underline{u}_\tau \to u$ in $\mathrm{L}^1([0,T];X)$ (strongly) and $\overline{\xi}_\tau \rightharpoonup \xi$ in $\mathrm{L}^1([0,T];X^*)$ (weakly).

Step 4: derivation of the gradient-flow equation. It remains to apply the Energy-Dissipation Principle from Theorem 3.9, which can be applied because our assumption (3.20d) enforces the abstract chain rule condition (3.13). Thus, we conclude that the constructed pair (u, ξ) satisfies the gradient-flow equation (3.15) and the energy-dissipation balance (3.16). ■

We emphasize that in this case we are not able to show uniqueness. Thus, different choices of the subsequences may lead to different solutions. Hence, we cannot define a "gradient flow" as in Section 2.5.

From the proof we can even learn more by observing that we did several liminf estimates to obtain (EDI). However, later we showed that in fact (EDB) holds. This implies that the liminf estimates must have been "attained" at least along the chosen subsequence. Thus we additionally conclude:

- for $0 \leq s < t \leq T$ we have $\mathcal{F}(u(t)) + \int_s^t \big(\mathcal{R}(..)+\mathcal{R}^*(..)\big)\,\mathrm{d}r = \mathcal{F}(u(s))$.
- $\forall\, t \in [0,T]: \quad \mathcal{F}(\widehat{u}_\tau(t)) \to \mathcal{F}(u(t))$.
- $\int_0^T \mathcal{R}(u, \dot{\widehat{u}}_\tau)\,\mathrm{d}t \to \int_0^T \mathcal{R}(u, \dot{u})\,\mathrm{d}t$.
- $\int_0^T \mathcal{R}(u, \overline{\xi}_\tau)\,\mathrm{d}t \to \int_0^T \mathcal{R}(u, \xi)\,\mathrm{d}t$.
- $\mathcal{R}^*(u(t), -\xi(t)) = \inf\big\{\, \mathcal{R}^*(u(t), -\eta \,\big|\, \eta \in \partial^{\mathrm{F}}\mathcal{F}(u(t))\,\big\}$ for a.a. $[t \in [0,T]$

Hence, under additional strict convexity assumptions on $\mathcal{R}(u,\cdot)$ and $\mathcal{R}^*(u,\cdot)$ it is even possible to show that the strong convergences $\dot{\widehat{u}}_\tau \to \dot{u}$ in $\mathrm{L}^1([0,T];X)$ and $\xi_\tau \to \xi$ in $\mathrm{L}^1([0,T];X^*)$, see [MiR15, Prop. C.3.3] for Visintin's argument from [Vis84].

Example 3.14 (State-dependent dissipation) *We consider the GS* $(\mathrm{L}^q(\Omega), \mathcal{F}, \mathcal{R})$ *with*

$$\mathrm{dom}(\mathcal{F}) = \mathrm{W}_0^{1,p}(\Omega), \;\; \mathcal{F}(u) = \int_\Omega \big(\frac{1}{p}|\nabla u|^p + F(u)\big)\,\mathrm{d}x \quad \textit{and} \quad \mathcal{R}(u,v) = \int_\Omega \frac{a(u)}{q}|v|^q\,\mathrm{d}x,$$

where $\Omega \subset \mathbb{R}^d$ *is a bounded Lipschitz domain,* $p \in \,]d,\infty[$*, and* $q \in \,]1,\infty[$*. The function* $F : \mathbb{R} \to [0,\infty[$ *is* C^1 *and semiconvex and* $a : \mathbb{R} \to \,]0,\infty[$ *is continuous.*

Exercise 3.3 *Formulate the associated gradient-flow equation and check that the assumptions of Theorem 3.13 hold.*

3.6 Extensions

We discuss a few possible extensions that allow us to widen the applicability of the theory.

3.6.1 Time dependent gradient systems

Often one is interested in the case of time-dependent functionals $\mathcal{F} : [0,T]{\times}X \to \mathbb{R}_\infty$. A typical case is $\mathcal{F}(t,u) = \mathcal{E}(u) - \langle \ell(t), u\rangle$ implying that $\partial^{\mathrm{F}}\mathcal{F}(t,u) = \partial^{\mathrm{F}}\mathcal{E}(u) - \ell(t)$, where the convention is now that $\partial^{\mathrm{F}}\mathcal{F}(t,u) = \partial^{\mathrm{F}}\big(\mathcal{F}(t,\cdot)\big)(u)$. The forcing ℓ appears in the associated gradient-flow equation as source term:

$$0 \in \partial\mathcal{R}(u,\dot{u}) + \partial^{\mathrm{F}}\mathcal{F}(t,u) = \partial\mathcal{R}(u,\dot{u}) + \partial^{\mathrm{F}}\mathcal{E}(u) - \ell(t).$$

The above theory can be carried through under suitable technical assumptions such as

$$\mathsf{D} := \mathrm{dom}(\mathcal{F}(t,\cdot)) \text{ is independent of } t \in [0,T], \tag{3.33a}$$
$$\forall\, u \in \mathsf{D}: \quad \mathcal{F}(\cdot,u) \in \mathrm{C}^1([0,T]), \tag{3.33b}$$
$$\exists\, c_\mathrm{F},\ C_\mathrm{F} > 0\ \forall\, u \in \mathsf{D},\ t \in [0,T]: \quad |\partial_t \mathcal{F}(t,u)| \le c_\mathrm{F}\mathcal{F}(t,u) + C_\mathrm{F}, \tag{3.33c}$$
$$(t_n,u_n) \to (t,u) \text{ and } \sup_{n\in\mathbb{N}} \mathcal{F}(t_n,u_n) < \infty \quad \text{imply} \quad \partial_t\mathcal{F}(t_n,u_n) \to \partial_t \mathcal{F}(t,u). \tag{3.33d}$$

With this the chain rule needs to be generalized into

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathcal{F}(t,u(t)) = \langle \xi(t), \dot u(t)\rangle + \partial_t \mathcal{F}(t,u(t))$$

and the energy-dissipation balance takes correspondingly the form

$$\mathcal{F}(t,u(t)) + \int_s^t \big(\mathcal{R}(u,\dot u) + \mathcal{R}^*(u,-\xi)\big)\,\mathrm{d}r = \mathcal{F}(s,u(s)) + \int_s^t \partial_r \mathcal{F}(r,u(r))\,\mathrm{d}r,$$

where the last term can be understood as the work of the time-dependent external forces.

Now the energy $\mathcal{F}(t,u(t))$ is no longer decreasing, but (3.33c) provides the upper bound

$$\mathcal{F}(t,u(t)) + C_\mathrm{F} \le \mathrm{e}^{c_\mathrm{F}(t-s)}\big(\mathcal{F}(s,u(s)) + C_\mathrm{F}\big) \quad \text{for } 0 \le s < t.$$

The construction of solutions still follows the time-incremental minimization scheme (3.6), namely

$$u_k^\tau \ \text{ minimizes } \ u \mapsto \tau\mathcal{R}\big(u_{k-1}^\tau, \frac{1}{\tau}(u - u_{k-1})\big) + \mathcal{F}(k\tau,u).$$

As the minimizer u_k^τ satisfies $\xi_k^\tau \in \partial^\mathrm{F}\mathcal{F}(k\tau,u_k^\tau)$ and $-\xi_k^\tau \in \partial\mathcal{R}\big(u_{k-1}^\tau, \frac{1}{\tau}(u_k^\tau - u_{k-1})\big)$, we can proceed as for (3.8). Using the Fenchel equivalences and the λ-convexity of $\mathcal{F}(t,\cdot)$ we find

$$\begin{aligned}\mathcal{F}(k\tau,u_k) + \tau\Big(\mathcal{R}\big(u_{k-1}, \frac{1}{\tau}(u_k - u_{k-1})\big) + \mathcal{R}^*(u_{k-1},-\xi_k)\Big) &\le \mathcal{F}(k\tau,u_{k-1}) - \frac{\lambda}{2}\|u_k - u_{k-1}\|^2 \\ = \mathcal{F}(k\tau-\tau,u_{k-1}) + \int_{k\tau-\tau}^{k\tau} \partial_t \mathcal{F}(t,u_{k-1})\,\mathrm{d}t - \frac{\lambda}{2}\|u_k - u_{k-1}\|^2.\end{aligned}$$

Using the interpolants as introduced in (3.9) we see that the approximate EDI (3.11) generalizes to

$$\begin{aligned}\mathcal{F}(T,\overline{u}_\tau(T)) + \int_0^T \Big(\mathcal{R}\big(\underline{u}_\tau, \dot{\hat{u}}_\tau\big) + \mathcal{R}^*\big(\underline{u}_\tau, -\overline{\xi}_\tau\big)\Big)\,\mathrm{d}t& \\ \le \mathcal{F}(0,u^0) + \int_0^T \partial_t \mathcal{F}(t,\underline{u}_\tau)\,\mathrm{d}t + \frac{\tau\lambda}{2}\int_0^T \|\dot{\hat{u}}_\tau(t)\|^2\,\mathrm{d}t.&\end{aligned} \tag{3.34}$$

From this, suitable a priori estimates can be derived and the limit passage works as before, where (3.33d) is used for the term involving $\partial_t \mathcal{F}$.

3.6.2 Weakly compact sublevels

A similar theory can be developed if the sublevels of the energy are not compact in the strong topology, but only in the weak topology. The major difference needed then, is that the closedness of the subdifferential has to be imposed in the weak-weak topology. But this is the case of the leading term in the energy is quadratic. For instance consider the Allen-Cahn energy $\mathcal{F}_{\mathrm{AC}}$ on $H = \mathrm{H}_0^1(\Omega)$ with bounded $\Omega \subset \mathbb{R}^d$ and $d \in \{1,2,3\}$:

$$\mathcal{F}_{\mathrm{AC}}(u) = \int_\Omega \big(\frac{\alpha}{2}\,|\nabla u|^2 + \frac{\beta}{4}\,(u^2{-}1)^2\big)\,\mathrm{d}x \quad \text{for } u \in \mathrm{dom}(\mathcal{F}_{\mathrm{AC}}) = H = \mathrm{H}_0^1(\Omega).$$

Then $(u_n,\xi_n) \rightharpoonup (u,\xi)$ in $H{\times}H^* = \mathrm{H}_0^1(\Omega){\times}\mathrm{H}^{-1}(\Omega)$ and $\xi_n = \mathrm{D}\mathcal{F}_{\mathrm{AC}}(u_n) = \{-\alpha\Delta u_n + \beta(u_n^3{-}u_n)\}$. Hence, the embedding $\mathrm{H}_0^1(\Omega) \leq \mathrm{L}^6(\Omega)$ for $p \in [1,p]$ which is compact for $p < 6$, implies boundedness of $\beta(u_n^3{-}u_n)$ and strong convergence to the desired limit $\beta(u^3{-}u)$ in $\mathrm{L}^q(\Omega)$ for all $q \in [1,2[$. Thus, we have the desired closedness $\xi = \mathrm{D}\mathcal{F}_{\mathrm{AC}}(u)$.

3.6.3 Approaches without semiconvexity and variational interpolants

Semiconvexity of the functional $\mathcal{F}$ has proved to be a very useful condition, because it implies closedness of the subdifferential, it helps to establish the abstract chain rule, and it provides a simple approach the discrete EDI. However, for many applications semiconvexity is too strong and it is desirable to avoid this assumption.

For instance, in [MiR23, Prop. A.1] the chain rule is established under a much weaker “uniform Fréchet differentiability”. Also the closedness of the subdifferential can be shown by advanced PDE methods, thus avoiding semiconvexity.

The major problem is the derivation of the approximate discrete EDI, which then provides an a priori estimate for the forces $\overline{\xi}_\tau$. The main idea is to avoid the linear interpolation in the piecewise affine interpolant $\widehat{u}_\tau$, which can only be useful, if the functional $\mathcal{F}$ can be controlled along straight lines. The main new idea is due to Ennio De Giorgi, but he has never published it. It can be found in the works [Amb95, AGS05] of his PhD student Luigi Ambrosio in the context of metric gradient flows, see Section 4.4. For Hilbert-space gradient systems without λ-convexity this idea was developed first in [RoS06] and for generalized gradient systems on Banach spaces in [MRS13, Lem. 6.1].

We construct $(u_k^\tau)_{k=1:N}$ by time-incremental minimization as before and define the *variational (De Giorgi) interpolant* $\widetilde{u}_\tau : [0,T] \to X$ such that for all $k = 0,\dots,N{-}1$ and $\theta \in\,]0,1[$ we have

$$\widetilde{u}_\tau\big(k\tau + \theta\tau\big) \;\text{ minimizes }\; u \mapsto \theta\tau\,\mathcal{R}\big(u_k^\tau, \frac{1}{\theta\tau}(u{-}u_k^\tau)\big) + \mathcal{F}(u).$$

As $\widetilde{u}_\tau(t)$ for $t \in\,]k\tau, k\tau{+}\tau]$ is obtained as a minimizer, there is a $\widetilde{\xi}_\tau(t) \in \partial^{\mathrm{F}}\mathcal{F}(\widetilde{u}_\tau(t))$ with $-\widetilde{\xi}_\tau(t) \in \partial\mathcal{R}\big(u_k^\tau, \frac{1}{t-k\tau}(\widetilde{u}_\tau(t){-}u_k^\tau)\big)$. Under suitable assumptions, it is then possible to show

$$\mathcal{F}(u_k^\tau) + \tau\mathcal{R}\big(u_{k-1}^\tau, \frac{1}{\tau}(u_k^\tau{-}u_{k-1}^\tau)\big) + \int_{k\tau-\tau}^{k\tau} \mathcal{R}^*\big(u_{k-1}^\tau, -\widetilde{\xi}_\tau(t)\big)\,\mathrm{d}t \leq \mathcal{F}(u_{k-1}^\tau)$$

with $\widetilde{\xi}_\tau(t) \in \partial^{\mathrm{F}}\mathcal{F}(\widetilde{u}_\tau(t))$ a.e., which replaces the former discrete EDI (3.8), which was derived using λ-convexity.

But now λ-convexity is no longer needed for obtaining the discrete EDI. It remains to generalize the abstract chain rule to cases without λ-convexity.

4 Metric gradient systems

In this section we generalize the previous theory from Banach spaces to much more general metric spaces, where we mainly follow [AGS05, Ch. 2–4]. It is surprising that the concept of gradient systems can be generalized to spaces without a linear structure. The main reason for this is the variational character encoded in the time-incremental minimization scheme via the energy functional $\mathcal{F}$ and the dissipation potential $\mathcal{R}$. The major idea is to replace the time derivative $\dot u(t) = \lim_{h\to 0} \frac{1}{h}\big(u(t{+}h) - u(t)\big) \in X$ and the forces $\xi \in \partial^{\mathrm{F}}\mathcal{F}(u) \subset X^*$ by appropriate quantities that are still available in metric spaces. In particular, convexity methods are no longer available. In generalizing to metric spaces, we will also drop the assumption of semiconvexity that was very helpful in the Hilbert and Banach space setting. For this we exploit the variational interpolant as introduced by De Giorgi, see Section 4.4.

4.1 Minimizing movements for metric gradient systems

Throughout Section 4 we will work with a complete metric space $(M, \mathcal{D})$, i.e. $\mathcal{D} : M{\times}M \to [0,\infty[$ is a metric satisfying positivity, symmetry and the triangle inequality. Completeness of $(M, \mathcal{D})$ means that all Cauchy sequences have a limit in M. For simplicity, we will always use the topology on M that is induced by the metric, however the general theory needs to be developed by a second weaker topology, where convergence is often denoted by $\stackrel{\sigma}{\rightharpoonup}$, see e.g. [AGS05, Cha. 3]. This is in analogy to Banach spaces where convergence in $\mathcal{D}$ corresponds to norm convergence, whereas $\stackrel{\sigma}{\rightharpoonup}$ indicates weak convergence.

Definition 4.1 (Metric gradient systems and minimizing movements) *A quadruple* $(M, \mathcal{F}, \mathcal{D}, \psi)$ *is called* generalized metric gradient system *if*

- $(M, \mathcal{D})$ *is a complete metric space,*
- $\mathcal{F} : M \to \mathbb{R}_\infty$ *is a proper, lsc functional,*
- $\psi : \mathbb{R} \to [0,\infty]$ *is a dissipation potential.*

Standard metric gradient systems *are given by the special choice* $\psi = \psi_{\mathrm{quadr}} : r \mapsto \frac{1}{2}r^2$*. One then shortly writes* $(M, \mathcal{F}, \mathcal{D}) := (M, \mathcal{F}, \mathcal{D}, \psi_{\mathrm{quadr}})$*.*

The associated minimizing movement scheme (MMS) *is given by*

$$u_k^\tau \quad \textit{minimizes} \quad u \;\mapsto\; \tau_k\, \psi\big(\frac{1}{\tau_k}\mathcal{D}(u_{k-1}^\tau, u)\big) + \mathcal{F}(u),$$

where $\tau_k > 0$ *is a possibly variable time step. A curve* $u : [0,\infty[\; \to M$ *is called* minimizing movement *for the metric GS* $(M, \mathcal{F}, \mathcal{D}, \psi)$ *if it is the limit (pointwise in* t*) of the piecewise constant interpolants* $\overline{u}_\tau : [0,\infty[\; \to M$ *of the MMS even, when varying time steps are allowed. One then writes* $u \in \mathrm{MM}(M, \mathcal{F}, \mathcal{D}, \psi)$*. If* u *is only the limit of some sequence of partitions (with fineness tending to* 0*), then* u *is called a* generalized minimizing movement *and we write* $u \in \mathrm{GMM}(M, \mathcal{F}, \mathcal{D}, \psi)$*.*

Note that the MMS for the standard metric GS $(M, \mathcal{F}, \mathcal{D}) = (M, \mathcal{F}, \mathcal{D}, \psi_{\mathrm{quadr}})$ leads to the standard *minimizing movement scheme (MMS)* given by

$$\text{standard MMS:} \qquad u_k^\tau \quad \text{minimizes} \quad u \;\mapsto\; \frac{1}{2\tau_k}\,\mathcal{D}(u_{k-1}^\tau, u)^2 + \mathcal{F}(u),$$

which is a direct generalization of the time-incremental minimization scheme (2.5) in the Hilbert-space setting.

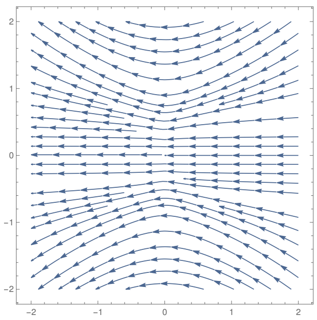


Figure 4.1: The figure shows the streamlines for the gradient-flow equations (4.1). All solutions satisfy $\dot u_1 \leq 1$, and the axis $u_2 = 0$ is invariant leading to the ODE $\dot u_1 = -|u_1|^{1/2}$ having non-unique solutions. All other solutions stay away from $u_2 = 0$ and, hence, are uniquely determined by their initial condition.

The notion of (generalized) minimizing movements can be seen as a solution concept for metric GS. However, these solutions are only defined as limit (or accumulation) points, which is a situation that is not always satisfactory. The point is that it is difficult to derive further properties of solutions, in particular a continuous dependence on a parameter μ. The latter relies on interchanging the two limits $\tau \to 0$ and $\mu_k \to \mu$, which is absolutely nontrivial. See also Example 4.2. Thus, it is desirable to find a formulation of solutions that replaces the gradient-flow and allows a direct study of solution without referring to the limiting process $\tau \to 0$.

Example 4.2 (Missing upper semicontinuity for minimizing movements) *Consider $M = \mathbb{R}^2$ with $\mathcal{D}(u,w) = |u{-}w|_{\textit{Eucl}}$ and $\psi(r) = \frac12 r^2$ such that we are in the Hilbert-space setting $\mathcal{R}(u,v) = \frac12(v_1^2{+}v_2^2)$ of Section 2. We choose the energy functional $\mathcal{F}(u) = \frac23 u_1(u_1^2{+}u_2^4)^{5/4}$ which is smooth on $\mathbb{R}^2 \setminus \{0\}$. The gradient-flow equation reads*

$$\dot u_1 = -\frac{3u_1^2 + 2u_2^4}{2(u_1^2{+}u_2^4)^{3/4}}, \quad \dot u_2 = -\frac{u_1 u_2^3}{(u_1^2{+}u_2^4)^{3/4}}, \tag{4.1}$$

and Figure 4.1 shows the solutions. The vector field is locally Lipschitz continuous on $\mathbb{R}^2 \setminus \{0\}$, and the u_1 axis is invariant because $\mathcal{F}$ is even in u_2.

We consider the solutions starting at the initial points $u^0 = (1,a)^\top$ for small a and denote these solutions by $t \mapsto U^a(t)$. We obviously have a one-parameter family of solutions for $a = 0$, namely

$$U^0_{(\mu)}(t) = \begin{cases} \frac{9}{16}\big(\frac43{-}t\big)^2 & \text{for } t \in [0,\frac43], \\ 0 & \text{for } \frac43 \leq t < \mu, \\ \frac{9}{16}(t{-}\mu)^2 & \text{for } t \geq \mu. \end{cases}$$

For $a \neq 0$ the solutions U^a are unique, as they never hit the non-Lipschitz point $u = (0,0)^\top$. To see this, observe that for $u_1(t) \leq 0$ we have $\dot u_2 \geq 0$, while for $u_1(t) \in [0,1]$ and $u_2 \geq 0$ we have $(u_1^2{+}u_2^4)^{3/4} \geq \max\{u_1^{3/2}, u_2^3\}$ which gives $\dot u_2 \geq -\min\{u_1, u_2^3/u_1^{1/2}\}$ and bounds u_2 away from 0. Hence, we see that taking the limit $0 \neq a \to 0$ (from above or from below) we see that

$$\forall\, t > 0: \quad U^a(t) \to U^0_{(4/3)}(t) \qquad \text{for } 0 \neq a \to 0.$$

Below we will show that for this example, the MMS determines, for each initial datum, a unique minimizing movement in the sense of Definition 4.1. The theory to be developed below will show that the MMS always provides as least one solution (a GMM) for each initial condition $u^0 = (1,a)^\top$. If the solution U^a of the gradient-flow equation is unique, then the GMM is an MM and coincides with this solution.

This is the case for $a \neq 0$. However, also for $a = 0$ we obtain a unique minimizing movement. For $a = 0$ we start on the invariant line $u_2 = 0$ and the following Euler-Lagrange equations show that we always stay there: $u = u_k^\tau$ has to satisfy

$$\frac{1}{\tau}\big(u - u_{k-1}^\tau\big) + \frac{1}{2(u_1^2{+}u_2^4)^{3/4}}\begin{pmatrix}3u_1^2 + 2u_2^4\\ 2u_1u_2^3\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}.$$

By induction over k we see first that $u_{k,2}^\tau = 0$ and then that

$$\frac{1}{\tau}\big(u_{k,1}^\tau{-}u_{k-1,1}^\tau\big) + \frac{3}{2}\sqrt{u_{k,1}^\tau} = 0 \quad\Longrightarrow\quad \sqrt{u_{k,1}^\tau} = \sqrt{\frac{9\tau^2}{16} + u_{k-1,1}^\tau} - \frac{3\tau}{4} \ > \ 0.$$

Since all $u_{k,1}^\tau$ are positive, we have convergence to the only solution $U_{(\mu)}^0$ that is non-negative, i.e. $\mu = \infty$.

The important remark is now that considering the limit $0 \neq a \to 0$ we see that the limit of $U^a = \mathrm{MM}\big(\mathbb{R}^2, \mathcal{F}, |\cdot|_\mathrm{E}, (1,a)^\top\big)$ is not in $\mathrm{GMM}\big(\mathbb{R}^2, \mathcal{F}, |\cdot|_\mathrm{E}, (1,0)^\top\big)$, i.e. the solution set is not upper semi-continuous. The notion of curves of maximal slope encompasses this disadvantage of MM or GMM.

We refer to [FlS20] for a way to modify the MMS to obtain all curves of maximal slope.

4.2 Curves of maximal slope

The solution concept "curves of maximal slopes for the GS $(M, \mathcal{F}, \mathcal{D}, \psi)$" will be tailored exactly to contain all $u \in \mathrm{GMM}(M, \mathcal{F}, \mathcal{D}, \psi)$. Moreover, it is a direct generalization of the solutions concept derived for Banach-space GS $(X, \mathcal{F}, \mathcal{R})$.

The major idea of generalizing the gradient-flow theory from Banach spaces to metric spaces is obtained by looking at special classes in the Banach-space setting. For this we consider generalized GS $(X, \mathcal{F}, \mathcal{R})$ in a Banach space X with dissipation potentials $\mathcal{R}(u,v) = \psi\big(\|v\|_X\big)$ where $\psi : \mathbb{R} \to [0,\infty]$ is a scalar dissipation potential. The dual dissipation potential reads $\mathcal{R}^*(u,\xi) = \psi^*\big(\|\xi\|_{X^*}\big)$. Note that this choice is a special instance of a generalized metric GS, where we choose $M = X$ and $\mathcal{D}(u,w) = \|w{-}u\|_X$. Moreover, the energy dissipation balance (EDB), see e.g. (3.16), now takes the special form

$$\mathcal{F}(u(T)) + \int_0^T \Big(\psi\big(\|\dot u\|_X\big) + \psi^*\big(\|\mathrm{D}\mathcal{F}(u(t))\|_{X^*}\big)\Big)\,\mathrm{d}t = \mathcal{F}(u(0)),$$

where we assumed that $\partial^\mathrm{F}\mathcal{F}(u)$ is the singleton $\big\{\mathrm{D}\mathcal{F}(u)\big\}$.

The main observation is that within this special class, we do not need the vector-valued quantities $\dot u(t) \in X$ and $\mathrm{D}\mathcal{F}(u(t)) \in X^*$, but it is enough to control the real-valued quantities $\|\dot u(t)\|_X \in \mathbb{R}$ and $\big\|\mathrm{D}\mathcal{F}(u(t))\big\|_{X^*} \in \mathbb{R}$. We will see below that there are natural generalizations of these two real-valued quantities in the metric setting, where no linear structure is available.

We first study absolutely continuous curves $\gamma : [0,T] \to M$ in the metric space $(M, \mathcal{D})$.

Definition 4.3 (Absolutely continuous curves) *A curve $\gamma : [0,T] \to M$ is called* absolutely continuous *in $(M, \mathcal{D})$, if there exists a function $g \in \mathrm{L}^1([0,T])$ such that*

$$\mathcal{D}\big(\gamma(t_1), \gamma(t_2)\big) \leq \int_{t_1}^{t_2} g(t)\,\mathrm{d}t \quad \textit{for all } t_1, t_2 \in [0,T] \textit{ with } t_1 < t_2. \tag{4.2}$$

We then write $\gamma \in \mathrm{AC}\big([0,T];(M,\mathcal{D})\big)$ *or shortly* $\gamma \in \mathrm{AC}\big([0,T];M\big)$ *if* $\mathcal{D}$ *is clear from the context. If additionally* $g \in \mathrm{L}^p([0,T])$ *for some* $p \in [1,\infty]$*, we write* $\gamma \in \mathrm{AC}^p\big([0,T];M\big)$.

As in the case of Banach spaces, we also have the embeddings in the Hölder spaces $\mathrm{AC}^p([0,T];M) \subset \mathrm{C}^{1-1/p}([0,T];M)$, which follows via Hölder's inequality:

$$\mathcal{D}\big(\gamma(t_1),\gamma(t_2)\big) \leq \int_{t_1}^{t_2} 1\cdot g(t)\,\mathrm{d}t \leq \Big(\int_{t_1}^{t_2} 1^{p^*}\,\mathrm{d}t\Big)^{1/p^*}\Big(\int_{t_1}^{t_2} g(t)^p\,\mathrm{d}t\Big)^{1/p} \leq \big|t_2{-}t_1\big|^{1/p^*}\|g\|_{\mathrm{L}^p},$$

where $p^* = p/(p{-}1)$. If $(M,\mathcal{D})$ is given by a reflexive Banach space $(X,\|\cdot\|)$, then we have $\mathrm{AC}^p([0,T];X) = \mathrm{W}^{1,p}([0,T];X)$. However, for general Banach spaces we only have the inclusion $\mathrm{W}^{1,p}([0,T];X) \subset \mathrm{AC}^p([0,T];X)$. As an example consider $X = \mathrm{L}^1(\mathbb{R})$ with the standard norm. Now consider the curve $\widehat\gamma : [0,T] \to \mathrm{L}^1(\mathbb{R})$ with

$$\widehat\gamma(t) = \mathbf{1}_{[0,\cosh(t)]} :\ x \mapsto \begin{cases} 1 & \text{for } x \in [0,\cosh(t)], \\ 0 & \text{otherwise.} \end{cases}$$

Clearly, we have $\|\widehat\gamma(t_1) - \widehat\gamma(t_2)\|_{\mathrm{L}^1} = |\cosh(t_2){-}\cosh(t_1)|$ such that $\widehat\gamma \in \mathrm{AC}^p([0,T];\mathrm{L}^1(\mathbb{R}))$ for all p with function $g : t \mapsto \sinh(t)$. However, $\widehat\gamma$ does not lie in $\mathrm{W}^{1,1}([0,T];\mathrm{L}^1(\mathbb{R}))$ because for $h > 0$ the difference quotients $\frac{1}{h}\big(\widehat\gamma(t{+}h){-}\widehat\gamma(t)\big) = \frac{1}{h}\mathbf{1}_{]\cosh(t),\cosh(t+h)]}$ converge to $\sinh(t)\,\boldsymbol{\delta}_t(\cdot)$ (Dirac distribution at $x = t$) in the sense of measures, but do not converge in $\mathrm{L}^1(\mathbb{R})$, even though the difference quotients are bounded.

The following result from [AGS05, Thm. 1.1.2] shows that the metric speed is well-defined a.e. along absolutely continuous curves.

Theorem 4.4 (Metric speed) *For* $p \in [1,\infty]$ *assume* $\gamma \in \mathrm{AC}^p\big([0,T];M\big)$*. Then, the metric speed* $|\dot\gamma|_{\mathcal{D}}(t)$ *defined via*

$$|\dot\gamma|_{\mathcal{D}}(t) := \lim_{|h|\to 0}\frac{1}{|h|}\mathcal{D}\big(\gamma(t),\gamma(t{+}h)\big)$$

exists a.e. in $[0,T]$ *and* $|\dot\gamma|_{\mathcal{D}}(\cdot) \in \mathrm{L}^p([0,T])$.

Moreover, for every g *satisfying* (4.2)*, we have* $|\dot\gamma|_{\mathcal{D}} \leq g$ *a.e.*

Proof. We choose a countable dense set $\big\{\, s_n \in [0,T] \;\big|\; n \in \mathbb{N}\,\big\}$ and define the auxiliary functions

$$\delta_n(t) = \mathcal{D}\big(\gamma(s_n),\gamma(t)\big) \quad \text{for } t \in [0,T].$$

By the inverse triangle inequality we find, for $0 \leq t_1 < t_2 \leq T$,

$$|\delta_n(t_2) - \delta_n(t_1)| \leq \mathcal{D}(\gamma(t_1),\gamma(t_2)) \leq \int_{t_1}^{t_2} g(t)\,\mathrm{d}t, \tag{4.3}$$

Thus, we conclude $\delta_n \in \mathrm{AC}^p([0,T]) = \mathrm{W}^{1,p}([0,T])$, where we use that $X = \mathbb{R}$ is a reflexive Banach space. Thus, δ_n is differentiable a.e., more precisely $\dot\delta_n(t) = \lim_{h\to 0}\frac{1}{h}\big(\delta_n(t{+}h) - \delta_n(t)\big)$ exists for $t \in [0,T] \setminus E_n$ with $\mathcal{L}^1(E_n) = 0$. Clearly, $|\dot\delta_n| \leq g$ a.e.

We now define the function

$$\mu(t) = \sup\big\{\, |\dot\delta_n(t)| \;\big|\; n \in \mathbb{N}\,\big\} \ \text{ for } t \in [0,T] \setminus E \quad \text{and} \quad \mu(t) = 0 \ \text{ for } t \in E := \bigcup_{n\in\mathbb{N}} E_n$$

and observe $\mu \le g \in \mathrm{L}^p([0,T])$. Using (4.3) for all $t \in [0,T] \setminus E$ we find

$$\liminf_{h\to 0} \frac{1}{|h|}\mathcal{D}\big(\gamma(t),\gamma(t{+}h)\big) \ge \sup_{n\in\mathbb{N}} \Big(\liminf_{h\to 0} \frac{1}{|h|}\big|\delta_n(t)-\delta_n(t{+}h)\big|\Big) = \sup_{n\in\mathbb{N}} \big|\dot\delta_n(t)\big| = \mu(t).$$

Moreover, if $s_{n_k} \to t_1$ then $\delta_{n_k}(t_2)-\delta_{n_k}(t_1) \to \mathcal{D}\big(\gamma(t_1),\gamma(t_2)\big)-0$. Together with (4.3) we observe, for $h>0$,

$$\mathcal{D}\big(\gamma(t),\gamma(t{+}h)\big) = \sup_{n\in\mathbb{N}} \big|\delta_n(t)-\delta_n(t{+}h)\big| \le \sup_{n\in\mathbb{N}} \int_t^{t+h} |\dot\delta_n(r)|\,\mathrm{d}r \le \int_t^{t+h} \mu(r)\,\mathrm{d}r.$$

Dividing by $h>0$ and doing the corresponding estimate for $h<0$ we arrive at

$$\limsup_{h\to 0} \frac{1}{|h|}\mathcal{D}\big(\gamma(t),\gamma(t{+}h)\big) \le \limsup_{h\to 0} \frac{1}{|h|}\Big| \int_t^{t+h} \mu(r)\,\mathrm{d}r\Big| \;=\; \mu(t),$$

for a..a. $t \in [0,T]$, namely all right and left Lebesgue points of $\mu \in \mathrm{L}^p([0,T])$.

Together we have shown $\frac{1}{|h|}\mathcal{D}\big(\gamma(t),\gamma(t{+}h)\big) \to \mu(t) \le g(t)$ a.e. ∎

We may return to the above example $\widehat\gamma : [0,T] \to \mathrm{L}^1(\mathbb{R})$ which does not lie in $\mathrm{W}^{1,1}\big([0,T];\mathrm{L}^1(\mathbb{R})\big)$. We can now easily verify that the metric speed in $\mathrm{L}^1(\mathbb{R})$ exists for all $t \in [0,T]$, namely $|\dot{\widehat\gamma}|_{\mathrm{L}^1}(t) = \sinh(t)$.

The second important notion for metric gradient systems is a scalar notion for the differential $\partial^{\mathrm{F}}\mathcal{F} : X \rightrightarrows X^*$ of the energy functional $\mathcal{F} : X \to \mathbb{R}_\infty$. In the following definition of the metric slope $|\partial\mathcal{F}|_{\mathcal{D}}$ we call a point $u \in M$ isolated if there exists a positive r such that $B_r(u) \cap M = \{u\}$ and use the notation $[F]_+ := \max\{F,0\}$ for the positive part.

Definition 4.5 (Metric slope) *Given a metric GS* $(M,\mathcal{F},\mathcal{D},\psi)$ *we define the* (local) metric slope $|\partial\mathcal{F}|_{\mathcal{D}} : M \to [0,\infty]$ of the functional $\mathcal{F}$ *via*

$$|\partial\mathcal{F}|_{\mathcal{D}}(u) := \begin{cases} \infty & \text{for } u \notin \mathrm{dom}(\mathcal{F}), \\ 0 & \text{for isolated } u \in \mathrm{dom}(\mathcal{F}), \\ \displaystyle\limsup_{w\to u} \frac{\big[\mathcal{F}(u)-\mathcal{F}(w)\big]_+}{\mathcal{D}(u,w)} & \text{for nonisolated } u \in \mathrm{dom}(\mathcal{F}). \end{cases} \tag{4.4}$$

For $\lambda \in \mathbb{R}$ *we also define the* global metric λ-slope $|\partial^{\mathrm{gl}}_\lambda\mathcal{F}|_{\mathcal{D}}$ of $\mathcal{F}$ *via*

$$|\partial^{\mathrm{gl}}_\lambda\mathcal{F}|_{\mathcal{D}}(u) := \begin{cases} \infty & \text{for } u \notin \mathrm{dom}(\mathcal{F}), \\ \displaystyle\sup_{w\in M\setminus\{u\}} \Big[\frac{\mathcal{F}(u){-}\mathcal{F}(w)}{\mathcal{D}(u,w)} + \frac{\lambda}{2}\,\mathcal{D}(u,w)\Big]_+ & \text{for } u \in \mathrm{dom}(\mathcal{F}). \end{cases} \tag{4.5}$$

We say that $\mathcal{F}$ has a λ-global metric slope *if* $|\partial^{\mathrm{gl}}_\lambda\mathcal{F}|_{\mathcal{D}} = |\partial\mathcal{F}|_{\mathcal{D}}$*, and we say that* $\mathcal{F}$ has a semiglobal metric slope *if there exists* $\lambda \in \mathbb{R}$ *such that* $|\partial^{\mathrm{gl}}_\lambda\mathcal{F}|_{\mathcal{D}} = |\partial\mathcal{F}|_{\mathcal{D}}$.

From the definitions we easily see that $\lambda_1 < \lambda_2$ implies

$$|\partial\mathcal{F}|_{\mathcal{D}}(u) \le |\partial^{\mathrm{gl}}_{\lambda_1}\mathcal{F}|_{\mathcal{D}}(u) \le |\partial^{\mathrm{gl}}_{\lambda_2}\mathcal{F}|_{\mathcal{D}}(u).$$

The important consequence of the λ-global slopes is that we have the estimate

$$\forall\, u \in \mathrm{dom}(\mathcal{F})\ \forall\, w \in M: \quad \mathcal{F}(w) \geq \mathcal{F}(u) - |\partial\mathcal{F}|_{\mathcal{D}}(u)\, \mathcal{D}(u,w) + \frac{\lambda}{2}\, \mathcal{D}(u,w)^2, \tag{4.6}$$

which is a generalization of the characterization of the Fréchet subdifferential in Lemma 2.4. In particular, $|\partial\mathcal{F}|_{\mathcal{D}}(u)$ is the smallest number such that (4.6) for all $w \in M$.

The notion of semiglobal and λ-global metric slopes will play a similar role as semiconvexity and λ-convexity of functionals in the Banach-space setting. In Sections 2 and 3 we used semiconvexity for three important steps, namely (i) showing strong-weak closedness of the Fréchet subdifferential, (ii) establishing the chain rule, and (iii) deriving a discrete energy-dissipation estimate from the time-incremental minimization scheme. In the metric setting the notion of "semiglobal slopes" will be good enough to how (i') the lower semicontinuity of the metric slope and (ii') a metric chain-rule estimate.

Example 4.6 (Local and semiglobal slopes)

(A) Consider $(M,\mathcal{D}) = (\mathbb{R}, \mathcal{D}_{\mathrm{Eucl}})$ and $\mathcal{F}(u) = a^{\pm}u$ for $\pm u \geq 0$. For $u > 0$ we obviously have $|\partial\mathcal{F}|_{\mathcal{D}}(u) = |a^+|$, while $|\partial\mathcal{F}|_{\mathcal{D}}(u) = |a^-|$ for $u < 0$. The case $u = 0$ is special and we obtain $|\partial\mathcal{F}|_{\mathcal{D}}(u) = \max\{0, a^-, -a^+\}$.

(B) Consider $(M,\mathcal{D}) = (\mathbb{R}, \mathcal{D}_{\mathrm{Eucl}})$ and $\mathcal{F}(u) = \big||u|{-}1\big|$. We easily find the local slope $|\partial\mathcal{F}|_{\mathcal{D}}(u) = 1$ for $u \neq \pm 1$ and $|\partial\mathcal{F}|_{\mathcal{D}}(\pm 1) = 0$, i.e. $|\partial\mathcal{F}|_{\mathcal{D}}$ is lsc but not continuous.

For the λ-global slope we obtain $|\partial^{\mathrm{gl}}_{\lambda}\mathcal{F}|_{\mathcal{D}} = |\partial\mathcal{F}|_{\mathcal{D}}$ for all $\lambda \leq 0$, i.e. $|\partial\mathcal{F}|_{\mathcal{D}}$ is a semiglobal slope.

For $\lambda > 0$ we obtain larger values, e.g. for $u >$ we have $|\partial^{\mathrm{gl}}_{\lambda}\mathcal{F}|_{\mathcal{D}}(u) = 1 + \frac{\lambda}{2}(u{-}1)$ if $\lambda \in \,]0, \frac{2}{u+1}]$ and $|\partial^{\mathrm{gl}}_{\lambda}\mathcal{F}|_{\mathcal{D}}(u) = \frac{u-1}{u+1} + \frac{\lambda}{2}(u{+}1)$ for $\lambda \geq \frac{2}{u+1}$.

(C)If $(M,\mathcal{D})$ is given by a Banach space $(X; \|\cdot\|)$ and $\mathcal{F} : X \to \mathbb{R}_\infty$ is lsc, then for $u \in \mathrm{dom}(\mathcal{F})$ we have

$$|\partial\mathcal{F}|_{\mathcal{D}}(u) = \big\{\, \|\xi\|_{X^*} \,\big|\, \xi \in \partial^{\mathrm{F}}\mathcal{F}(u)\,\big\}.$$

Moreover, if $\mathcal{F}$ is μ-convex, then $|\partial\mathcal{F}|_{\mathcal{D}}$ is a λ-global slope for all $\lambda \geq \mu$.

Exercise 4.1 (Slopes) *(a) $(M,\mathcal{D}) = (\mathbb{R}, \mathcal{D}_{\mathrm{Eucl}})$ consider $\mathcal{F}(u) = \min\{u, 0\}$. Calculate $|\partial\mathcal{F}|_{\mathcal{D}}$ explicitly and show that it is not lsc. Moreover, calculate $|\partial^{\mathrm{gl}}_{\lambda}\mathcal{F}|_{\mathcal{D}}$ for all $\lambda \in \mathbb{R}$.*

(b) Establish the claims in part (C) of Example 4.6.

Note that the proof of the following result is very similar to the corresponding closedness of the Fréchet subdifferential for semiconvex functionals, see Proposition 2.6. Part (A) in Example 4.6 shows that $|\partial\mathcal{F}|_{\mathcal{D}}$ is not lsc in general, see Exercise 4.6

Proposition 4.7 (Lsc of semiglobal metric slopes) *If $\mathcal{F}$ is lsc and has a semiglobal slope $|\partial\mathcal{F}|_{\mathcal{D}}$ on the metric space $(M,\mathcal{D})$, then $|\partial\mathcal{F}|_{\mathcal{D}} : M \to [0,\infty]$ is lower semicontinuous.*

Proof. For a sequence $u_k \to u$ in $(M,\mathcal{D})$, we have to show $\sigma := \liminf_{k\to\infty} |\partial\mathcal{F}|_{\mathcal{D}}(u_k) \geq |\partial\mathcal{F}|_{\mathcal{D}}(u)$. Obviously, the case $\alpha = \infty$ is trivial.

Hence we assume $\sigma < \infty$ which implies $u_k \in \mathrm{dom}(\mathcal{F})$. Thus we have (4.6) for $u = u_k$ for all k. Using the lsc of $\mathcal{F}$ and $\mathcal{D}(u_k, w) \to \mathcal{D}(u,w)$ we immediately find

$$\mathcal{F}(w) \geq \mathcal{F}(u) - \sigma\, \mathcal{D}(u,w) + \frac{\lambda}{2}\, \mathcal{D}(u,w)^2 \quad \text{for all } w \in M.$$

But this implies $|\partial\mathcal{F}|_{\mathcal{D}}(u) \leq \sigma$, which is the desired estimate. ■

We have now all the ingredients to define the metric version of the generalized gradient-flow equation.

Definition 4.8 (Curves of maximal slope) *Given a generalized metric GS $(M, \mathcal{F}, \mathcal{D}, \psi)$ we call a curve $u : [0,T] \to M$ a* ψ-curve of maximal slope *if $u \in \mathrm{AC}\big([0,T];(M,\mathcal{D})\big)$ and for all $t_1, t_2 \in [0,T]$ with $t_1 < t_2$ we have*

$$\mathcal{F}(u(t_2)) + \int_{t_1}^{t_2} \Big(\psi\big(|\dot{u}|_{\mathcal{D}}(t)\big) + \psi^*\big(|\partial\mathcal{F}|_{\mathcal{D}}(u(t))\big)\Big)\,\mathrm{d}t = \mathcal{F}(u(t_1)). \tag{4.7}$$

If $\psi(r) = \frac{1}{p}r^p$ for $p \in \,]1,\infty[$ we shortly say that u is a p-curve of maximal slope of *$(M, \mathcal{F}, \mathcal{D})$. If $\psi = \psi_{\mathrm{quadr}} : r \mapsto \frac{1}{2}r^2$, then u is simply called a* curve of maximal slope *for the standard metric GS $(M, \mathcal{F}, \mathcal{D})$.*

As we have learned in Sections 2 and 3, we know that the above formulations are enough to characterize the solutions of the corresponding gradient-flow equations, if we are in the special case $\mathcal{R}(u,v) = \psi(\|v\|_X)$. Of course, the notion is much more general as will become clear by the following examples.

Example 4.9 (Different instances of curves of maximal slope)

(A) ***Nonuniqueness.*** *We consider $(\mathbb{R}, \mathcal{F}, \mathcal{D}_{\mathrm{Eucl}}, \psi_{\mathrm{quadr}})$ with $F(u) = \frac{1}{2}u^2 - |u|$.*

This system can also be treated as a Hilbert-space GS but, then the subdifferential $\partial^{\mathrm{F}}\mathcal{F}$ is not closed: $\partial^{\mathrm{F}}\mathcal{F}(u) = \{u-1\}$ for $u > 0$ and $\partial^{\mathrm{F}}\mathcal{F}(0) = \emptyset$.

Treating it as a metric GS leads to the metric slope $|\partial\mathcal{F}|_{\mathcal{D}}(u) = \big|1-|u|\big|$ which is even semiglobal with $\lambda = 0$.

We now show that there are two solutions starting at $u^0 = 0$, namely $u(t) = \pm(1-\mathrm{e}^{-t})$. To show that these two solutions are curves of maximal slope, we can simply check that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(u(t)) = -\frac{1}{2}\big(|\dot{u}|_{\mathcal{D}}(t)\big)^2 - \frac{1}{2}\big(|\partial\mathcal{F}|_{\mathcal{D}}(u(t))\big)^2$$

by inserting the explicit solutions.

(B) ***Riemannian manifold.*** *We consider a Riemannian manifold $(M, \mathbb{G})$ with a smooth functional $\mathcal{F} \in \mathrm{C}^1(M)$. For the smooth GS $(M, \mathcal{F}, \mathbb{G})$ we have the associated GFE $\dot{u} = -\mathrm{grad}_{\mathbb{G}}\mathcal{F}(u) = -\mathbb{G}(u)^{-1}\mathrm{D}\mathcal{F}(u)$.*

We now want to switch to the metric picture. For this we define the metric distance

$$\mathcal{D}_{\mathbb{G}}(u_0, u_1) := \inf\Big\{ \int_0^1 \|\dot{\gamma}\|_{\mathbb{G}}\,\mathrm{d}t \;\Big|\; \gamma \in \mathrm{C}^1([0,T];M),\ \gamma(0) = u_0,\ \gamma(1) = u_1 \Big\},$$

where $\|\dot{\gamma}\|^2_{\mathbb{G}} = \langle \mathbb{G}(\gamma)\dot{\gamma}, \dot{\gamma}\rangle$.

Doing some classical calculations in local charts one finds $\mathrm{AC}^p\big([0,T];(M,\mathcal{D}_{\mathbb{G}})\big) = \mathrm{W}^{1,p}([0,T];M)$ and $|\dot{u}|_{\mathcal{D}}(t) = \|\dot{\gamma}(t)\|_{\mathbb{G}}$.

Similarly, the metric slope takes the form $|\partial\mathcal{F}|_{\mathcal{D}}(u) = \|\mathrm{D}\mathcal{F}(u)\|_{\mathbb{G}^{-1}}$.

With this the condition for curves of maximal slope takes the form

$$\begin{aligned}0 &= \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(u(t)) + \frac{1}{2}|\dot u|_{\mathcal{D}}(t)^2 + \frac{1}{2}|\partial\mathcal{F}|_{\mathcal{D}}(u(t))^2\\ &= \langle \mathrm{D}\mathcal{F}(u),\dot u\rangle + \frac{1}{2}\langle \mathbb{G}(u)\dot u,\dot u\rangle + \langle \mathrm{D}\mathcal{F}(u),\mathbb{G}(u)^{-1}\mathrm{D}\mathcal{F}(u)\rangle\\ &= \frac{1}{2}\big\langle \mathbb{G}(u)\big(\dot u - \mathbb{G}(u)^{-1}\mathrm{D}\mathcal{F}(u)\big), \dot u - \mathbb{G}(u)^{-1}\mathrm{D}\mathcal{F}(u)\big\rangle = \frac{1}{2}\big\|\dot u - \mathbb{G}(u)^{-1}\mathrm{D}\mathcal{F}(u)\big\|^2_{\mathbb{G}}.\end{aligned}$$

Thus, we see that for this nice case the metric formulation is equivalent to the classical gradient-flow equation.

(C) ***Wasserstein space and Otto diffusion.*** *We consider a bounded open set $\Omega \subset \mathbb{R}^d$ and denote by $\mathrm{Prob}(\overline{\Omega})$ the space of probability measures, which is a closed convex subset of the signed measures $\mathrm{SM}(\overline{\Omega}) = \big(\mathrm{C}(\overline{\Omega})\big)^*$. On this set the Kantorovich-Wasserstein distances W_p (cf. [AGS05, Vil09]) are defined via*

$$\mathrm{W}_p(\mu_0,\mu_1)^p := \inf\Big\{ \int_{\overline{\Omega}\times\overline{\Omega}} |x-y|^p \Pi(\mathrm{d}x,\mathrm{d}y) \ \Big|\ \Pi \in \mathcal{C}(\mu_0,\mu_1)\Big\}, \quad \textit{where}$$

$\mathcal{C}(\mu_0,\mu_1) := \big\{\Pi\in\mathrm{Prob}(\overline{\Omega}\times\overline{\Omega}) \ \big|\ \forall$ *meas.* $A \subset \overline{\Omega} : \ \Pi(A\times\overline{\Omega}) = \mu_0(A),\ \Pi(\overline{\Omega}\times A) = \mu_1(A)\big\}$. *For all $p \in [1,\infty[$, the pair $(\mathrm{Prob}(\overline{\Omega}),\mathrm{W}_p)$ defines a complete metric space and the convergence is equal to the weak* convergence.*

For $p = 2$ the Wasserstein space $(\mathrm{Prob}(\overline{\Omega}),\mathrm{W}_2)$ is even a geodesic metric space (see Definition 4.21) which has many similarities with a Riemannian manifold with nonsmooth boundaries. In a series of papers around 2000, the corresponding metric theory was developed and summarized in [AGS05]. The metric speed of a curve $\mu \in \mathrm{AC}\big([0,T];(\mathrm{Prob}(\overline{\Omega}),\mathrm{W}_2)\big)$ can be defined as follows. For every such function, there exists a vector field $V \in \mathrm{L}^1([0,T]\times\overline{\Omega};\mathbb{R}^d)$ such that the continuity equation

$$\dot\mu + \mathrm{div}\,\big(V\mu\big) = 0 \quad \textit{holds in } \big(\mathrm{C}^\infty_\mathrm{c}(]0,T[\times\overline{\Omega})\big)^*$$

(where "$_\mathrm{c}$" stands for compactly contained support), and the metric speed takes the form $|\dot\mu|_{\mathrm{W}_2}(t) = \big(\int_{\overline{\Omega}}|V(t,x)|^2\mu(t,\mathrm{d}x)\big)^{1/2}$ for a.a. $t \in [0,T]$.

Similarly one can derive a formula for the metric slope of certain functionals. Choosing a lsc, convex, and superlinear functional $E : [0,\infty[\to [0,\infty]$ and $\varphi \in \mathrm{C}^1(\overline{\Omega})$ on can define

$$\mathcal{F}(\mu) = \begin{cases} \int_\Omega \big(E(\rho(x)) + \varphi(x)\rho(x)\big)\,\mathrm{d}x & \textit{for } \mu = \rho\,\mathrm{d}x \textit{ with } \rho \in \mathrm{L}^1_\geq(\Omega),\\ \infty & \textit{otherwise},\end{cases}$$

where $\mathrm{L}^1_\geq(\Omega)$ denotes the non-negative functions in $\mathrm{L}^1(\Omega)$. Then, $\mathcal{F}$ is lsc on $(\mathrm{Prob}(\overline{\Omega}),\mathrm{W}_2)$ and the metric slope is given by

$$|\partial\mathcal{F}|_{\mathrm{W}_2}(\rho\mathrm{d}x) = \Big(\int_\Omega \big|\nabla(E'(\rho)+\varphi)\big|^2\rho\,\mathrm{d}x\Big)^{1/2},$$

see [AGS05] for more precise statements and the justification of these relations.

Under the assumption that a curve of maximal slope for $(\mathrm{Prob}(\overline{\Omega}),\mathcal{F},\mathrm{W}_2)$ has the form $\mu(t) = \rho(t,\cdot)\,\mathrm{d}x$ with ρ sufficiently smooth and bounded from below, one can show that ρ satisfies a *drift-diffusion equation. We argue as in Example (B):*

$$\begin{aligned} 0 \;&=\; \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(\mu(t)) + \frac{1}{2}|\dot\mu|_{\mathrm{W}_2}(t)^2 + \frac{1}{2}|\partial\mathcal{F}|_{\mathrm{W}_2}(\mu(t))^2 \\ &\overset{\mu=\rho\mathrm{d}x}{=} \int_\Omega \Big((E'(\rho){+}\varphi)\dot\rho + \frac{1}{2}\rho|V|^2 + \frac{1}{2}\rho\big|\nabla(E'(\rho){+}\varphi)\big|^2\Big)\,\mathrm{d}x \\ &\overset{*}{=} \int_\Omega \frac{\rho}{2}\big|V{+}\nabla(E'(\rho){+}\varphi)\big|^2\,\mathrm{d}x, \end{aligned}$$

where for the last identity we inserted the continuity equation and integrated by parts, thus finding a complete square.

This shows that for curves $\mu = \rho\,\mathrm{d}x$ of maximal slope (with ρ sufficiently smooth and positive), the velocity field V in the continuity equation can be identified as $-\nabla(E'(\rho)+\varphi)$. This leads to the nonlinear drift-diffusion equation

$$\dot\rho = -\operatorname{div}(\rho V) = \operatorname{div}\big(\rho\nabla(E'(\rho){+}\varphi)\big) = \operatorname{div}\big(\rho E''(\rho)\nabla\rho + \rho\nabla\varphi\big).$$

In particular, one may consider the Boltzmann entropy with $E(\rho) = \rho\log\rho - \rho + 1$. Then $E''(\rho) = 1/\rho$ and we left with the linear Fokker-Planck equation as the associated gradient-flow equation

$$\dot\rho = \operatorname{div}\Big(\nabla\rho + \rho\nabla\varphi\Big).$$

This link between Wasserstein distance and the linear Fokker-Planck equation was first observed in [Ott96, JKO98]. For that reason the Minimizing Movement Scheme in the case of $\mathcal{D} = \mathrm{W}_2$ is nowadays called the JKO scheme for "Jordan-Kinderlehrer-Otto".

For the entropies $E(\rho) = \big(\rho^m - m\rho + m - 1\big)/(m^2{-}m)$ we have $E''(\rho) = \rho^{m-2}$ and in the case $\varphi = 0$ the associated gradient-flow equation is the porous medium equation

$$\dot\rho = \operatorname{div}\big(\rho^{m-1}\nabla\rho\big) = \frac{1}{m}\Delta\rho^m.$$

Exercise 4.2 (Nontrivial metric space) *We consider $M = \mathbb{R}^k$ with the nontrivial metric $\mathcal{D}_{\mathrm{sq}}(u,w) = \sqrt{|u{-}w|_{\mathrm{Eucl}}}$ that is topologically equivalent to the Euclidean one.*

(a) Show that $\mathrm{AC}\big([0,T];(\mathbb{R}^k,\mathcal{D}_{\mathrm{sq}})\big)$ is trivial in the sense that it only contains constant functions.

(b) For a smooth function $\mathcal{F} \in \mathrm{C}^1(\mathbb{R}^k)$ calculate $|\partial\mathcal{F}|_{\mathcal{D}}$ with $\mathcal{D} = \mathcal{D}_{\mathrm{sq}}$.

(c) Characterize all curves of maximal slope for $(M,\mathcal{F},\mathcal{D}_{\mathrm{sq}},\psi)$ in terms of their initial condition $u(0) = u^0$.

4.3 The metric chain-rule inequality

To prove existence of curves of maximal slopes we will use the minimization scheme in a similar way as for in the Banach-space setting. In this subsection we develop the corresponding replacement of the chain rule formula $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(u(t)) = \langle\xi(t),\dot u(t)\rangle$ with $\xi(t) \in \partial^{\mathrm{F}}\mathcal{F}(u(t))$. As we do not have any linear structure in the metric space $(M,\mathcal{D})$, we use the fact that in the existence proof for the GFE in Banach spaces we do not really need the above chain-rule identity. From the Fenchel-Young inequality we already have an inequality such that it would be sufficient to have the lower estimate $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(u(t)) \geq \langle\xi(t),\dot u(t)\rangle$. It turns out that in the metric setting a corresponding chain-rule inequality can be established.

Definition 4.10 (Abstract metric chain-rule inequality) *We say that the generalized metric GS* $(M, \mathcal{F}, \mathcal{D}, \psi)$ *satisfies the* abstract metric chain-rule inequality *if the following holds.*

If $u \in \mathrm{AC}([0,T];M)$ *satisfies* $\sup_{t\in[0,T]} \mathcal{F}(u(t)) < \infty$
and $\int_0^T \big(\psi(|\dot u|_{\mathcal{D}}(t)) + \psi^*\big(|\partial\mathcal{F}|_{\mathcal{D}}(u(t))\big)\big)\,\mathrm{d}t < \infty$,
then $t \mapsto \mathcal{F}(u(t))$ *is absolutely continuous and*

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(u(t)) \geq -|\partial\mathcal{F}|_{\mathcal{D}}(u(t))\,|\dot u|_{\mathcal{D}}(t) \quad \textit{a.e. in } [0,T]. \tag{4.8}$$

We will see that this inequality is enough for completing the existence proof for curves of maximal slope. The next result demonstrates that the condition that $|\partial\mathcal{F}|_{\mathcal{D}}$ is a semiglobal slope is sufficient for showing that the abstract metric chain-rule inequality holds. Moreover, the proof is almost identical to the corresponding Theorem 3.12 in Banach spaces. In fact, the origin of the proof of the latter result is [AGS05, Thm. 1.2.5], which is almost identical to our next result. Hence, a full proof for the case $|\partial\mathcal{F}|_{\mathcal{D}} = |\partial_0^{\mathrm{gl}}\mathcal{F}|_{\mathcal{D}}$ can be found there. Here we only give a sketch of the proof, by referring back to our proof of Theorem 3.12.

Proposition 4.11 (Metric chain-rule inequality) *Consider a generalized metric GS* $(M, \mathcal{F}, \mathcal{D}, \psi)$ *such that* $|\partial\mathcal{F}|_{\mathcal{D}}$ *is a semiglobal slope. Then, the following holds:*

If $u \in \mathrm{AC}([0,T];M)$ *satisfies* $\sup_{t\in[0,T]} \mathcal{F}(u(t)) < \infty$
and $\int_0^T \big(\psi(|\dot u|_{\mathcal{D}}(t)) + \psi^*\big(|\partial\mathcal{F}|_{\mathcal{D}}(u(t))\big)\big)\,\mathrm{d}t < \infty$,
then $t \mapsto \mathcal{F}(u(t))$ *is absolutely continuous and*

$$\Big|\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(u(t))\Big| \leq |\partial\mathcal{F}|_{\mathcal{D}}(u(t))\,|\dot u|_{\mathcal{D}}(t) \quad \textit{a.e. in } [0,T]. \tag{4.9}$$

In particular, for all dissipation potentials $\psi : \mathbb{R} \to [0,\infty[$ *the abstract metric chain-rule inequality of Definition 4.10 holds.*

Sketch of proof. We follow the proof of Theorem 3.12 and set $f(t) = \mathcal{F}(u(t))$ and $\sigma(t) = |\partial\mathcal{F}|_{\mathcal{D}}(u(t))$. Using that $\mathcal{F}$ has a semiglobal slope, i.e. there exists $\lambda \in \mathbb{R}$ such that $|\partial\mathcal{F}|_{\mathcal{D}} = |\partial_\lambda^{\mathrm{gl}}\mathcal{F}|_{\mathcal{D}}$, and choosing any partition of $[s,t]$ lying in $\Sigma := \big\{\, t \in [0,T] \;\big|\; \sigma(t) < \infty \,\big\}$, we obtain the estimates

$$\begin{aligned}
&\sum_{j=1}^N \Big(-\sigma(t_{j-1})\mathcal{D}\big(u(t_{j-1}), u(t_j)\big) + \frac{\lambda}{2}\mathcal{D}\big(u(t_{j-1}), u(t_j)\big)^2 \Big) \\
&\quad \leq \mathcal{F}(u(t)) - \mathcal{F}(u(s)) = f(t) - f(s) \\
&\quad \leq \sum_{j=1}^N \Big(-\sigma(t_j)\mathcal{D}\big(u(t_{j-1}), u(t_j)\big) + \frac{\lambda}{2}\mathcal{D}\big(u(t_{j-1}), u(t_j)\big)^2 \Big),
\end{aligned}$$

which correspond to (3.18) in the case of arclength parametrization, i.e. $\mathcal{D}(u(s_1), u(s_2)) = |s_2 - s_1|$ for all $s_1, s_2 \in [0,T]$.

As before we can pass to the limit and find the desired estimate

$$\big|\mathcal{F}(u(t)) - \mathcal{F}(u(s))\big| \leq \int_s^t \sigma(r)\,|\dot u|_{\mathcal{D}}(r)\,\mathrm{d}r,$$

which provides the absolute continuity as well as the desired estimate.

Finally, the abstract metric chain-rule inequality follows by applying the Fenchel-Young inequality to the scalar dissipation potential ψ, namely

$$\int_0^T \sigma(r)\,|\dot u|_{\mathcal D}(r)\,\mathrm{d}r \le \int_0^T\Big(\psi\big(\sigma(r)\big)+\psi^*\big(|\dot u|_{\mathcal D}(r)\big)\Big)\,\mathrm{d}r<\infty.$$

As (4.9) implies (4.8), Proposition 4.11 is established. ∎

In [AGS05] the names "chain rule" and "metric chain-rule inequality" are not used as here. There, the same notion is encoded in the term "*strong upper gradient*". For instance, [AGS05, Thm. 1.2.5] states that "if $\mathcal F$ is $\mathcal D$-lsc, then $|\partial_0^{\mathrm{gl}}\mathcal F|_{\mathcal D}$ is a strong upper gradient for $\mathcal F$", which means that the metric chain-rule inequality holds, if we replace $|\partial\mathcal F|_{\mathcal D}$ by $|\partial_0^{\mathrm{gl}}\mathcal F|_{\mathcal D}$ in (4.9).

As for generalized GS $(X,\mathcal F,\mathcal R)$ in Banach spaces we again have an Energy-Dissipation Principle in the following form.

Proposition 4.12 (Metric energy-dissipation principle) *Consider a generalized metric GS $(M,\mathcal F,\mathcal D,\psi)$ that satisfies the abstract metric chain-rule inequality of Definition 4.10. Then, for $u\in \mathrm{AC}([0,T];M)$ the following two statements are equivalent:*

(A) u satisfies the EDI given by $\mathcal F(u(T))+\int_0^T\big(\psi(|\dot u|_{\mathcal D})+\psi^*(|\partial\mathcal F|_{\mathcal D}(u))\big)\,\mathrm{d}t\le\mathcal F(u(0))$.

(B) u is a ψ-curve of maximal slope, i.e. $(\mathrm{EDB})_{[t_0,t_1]}$ holds for $0\le t_0<t_1\le T$.

Proof. $(B)\implies(A)$. This direction is trivial.

$(A)\implies(B)$. We set

$$f(t)=\mathcal F(u(t)),\quad v(t)=|\dot u|_{\mathcal D}(t),\quad\text{and } \sigma(t)=|\partial\mathcal F|_{\mathcal D}(u(t))$$

and observe that the chain-rule inequality implies $\dot f+\sigma v\ge 0$ a.e. Hence, we obtain

$$0\le\int_0^T\big(\dot f+\sigma v\big)\,\mathrm{d}t\overset{\text{FenYou}}{\le}\int_0^T\big(\dot f+\psi(v)+\psi^*(\sigma)\big)\,\mathrm{d}t\overset{\text{(EDI)}}{\le}0.$$

Thus, all inequalities "$\le$" must be equalities "$=$". Moreover the nonnegative integrand $\dot f+\psi(v)+\psi^*(\sigma)\ge\dot f+\sigma v\ge 0$ must vanish a.e. in $[0,T]$. However, integrating $\dot f+\psi(v)+\psi^*(\sigma)=0$ a.e. over $t\in[t_0,t_1]$ given (EDB) on $[t_0,t_1]$. ∎

4.4 De Giorgi's variational interpolant

In the subsequent analysis we will use the following assumptions on ψ:

$$\psi:\mathbb R\to[0,\infty[\text{ is a strictly convex dissipation potential and }\psi\in\mathrm C^1\big([0,\infty[\big).\tag{4.10}$$

The MMS gives global minimizers $u_k=u_k^\tau$, namely

$$\forall\,w\in M:\quad \tau\psi\big(\tfrac{1}{\tau}\mathcal D(u_{k-1},u_k)\big)+\mathcal F(u_k)\le\tau\psi\big(\tfrac{1}{\tau}\mathcal D(u_{k-1},w)\big)+\mathcal F(w).\tag{4.11}$$

From this we can derive a first slope estimate.

Proposition 4.13 (Metric slope estimate) *Assume that ψ satisfies* (4.10) *and let $(u_k)_{k=1,..,N}$ be the sequence obtained via the MMS for the generalized metric GS $(M, \mathcal{F}, \mathcal{D}, \psi)$, then*

$$|\partial\mathcal{F}|_{\mathcal{D}}(u_k) \leq \psi'\big(\frac{1}{\tau}\mathcal{D}(u_{k-1}, u_k)\big) \quad \textit{for } k = 1, \dots, N.$$

Proof. Rearranging the terms in (4.11) for $w \neq u$ we have

$$\frac{\mathcal{F}(u_k) - \mathcal{F}(w)}{\mathcal{D}(u_k, w)} \leq \frac{\psi\big(\frac{1}{\tau}\mathcal{D}(u_{k-1}, w)\big) - \psi\big(\frac{1}{\tau}\mathcal{D}(u_{k-1}, u_k)\big)}{\frac{1}{\tau}\mathcal{D}(u_k, w)}.$$

Using the triangle inequality and the monotonicity of ψ we have $\psi\big(\frac{1}{\tau}\mathcal{D}(u_{k-1}, w)\big) \leq \psi\big(\frac{1}{\tau}\mathcal{D}(u_{k-1}, u_k) + \frac{1}{\tau}\mathcal{D}(u_k, w)\big)$. Now taking the limit $w \to u_k$ we find

$$\begin{aligned}\limsup_{w\to u_k} \frac{\mathcal{F}(u_k)-\mathcal{F}(w)}{\mathcal{D}(u_k, w)} &\leq \limsup_{w\to u_k} \frac{\psi\big(\frac{1}{\tau}\mathcal{D}(u_{k-1}, u_k)+\frac{1}{\tau}\mathcal{D}(u_k, w)\big) - \psi\big(\frac{1}{\tau}\mathcal{D}(u_{k-1}, u_k)\big)}{\frac{1}{\tau}\mathcal{D}(u_k, w)} \\ = \limsup_{H\to 0^+} \frac{\psi\big(\frac{1}{\tau}\mathcal{D}(u_{k-1}, u_k) + H\big) - \psi\big(\frac{1}{\tau}\mathcal{D}(u_{k-1}, u_k)\big)}{H} &= \psi'\big(\frac{1}{\tau}\mathcal{D}(u_{k-1}, u_k)\big).\end{aligned}$$

As the right-hand side is non-negative (as ψ is a dissipation potential), we can take the positive part on both sides and obtain the desired result. ∎

In some sense, the last result can be seen as a generalization of the Euler-Lagrange equations $0 \in \partial\mathcal{R}\big(\frac{1}{\tau}(u_k-u_{k-1})\big) + \partial^{\mathrm{F}}\mathcal{F}(u_k)$ in the Banach-space setting. There we used Fenchel's equivalence and λ-convexity of $\mathcal{F}$ to derive the discrete EDI (with $\xi_k \in \partial^{\mathrm{F}}\mathcal{F}(u_k)$)

$$\tau\Big(\mathcal{R}\big(\frac{1}{\tau}(u_k-u_{k_1})\big) + \mathcal{R}^*(-\xi_k)\Big) \leq -\langle\xi_k, u_k-u_{k-1}\rangle \leq \mathcal{F}(u_{k-1}) - \mathcal{F}(u_k) - \frac{\lambda}{2}\|u_k-u_{k-1}\|^2.$$

The importance of this inequality is the telescoping structure with respect to the energies $\mathcal{F}(u_j)$.

In the metric setting we can also apply Fenchel's equivalence to the scalar relation $\sigma_k := \psi'(v_k)$ where $v_k := \frac{1}{\tau}\mathcal{D}(u_{k-1}, u_k)$. Exploiting the slope estimate in Proposition 4.13 we obtain

$$\begin{aligned}&\tau\Big(\psi\big(\frac{1}{\tau}\mathcal{D}(u_{k-1}, u_k)\big)+\psi^*\big(|\partial\mathcal{F}|_{\mathcal{D}}(u_k)\big)\Big) \overset{\text{Prop.}}{\leq} \tau\big(\psi(v_k)+\psi^*\big(\psi'(v_k)\big)\big) \\ &= \tau\big(\psi(v_k)+\psi^*(\sigma_k)\big) \overset{\text{Fenchel}}{=} \tau\sigma_k v_k = \psi'\big(\frac{1}{\tau}\mathcal{D}(u_{k-1}, u_k)\big)\, \mathcal{D}(u_{k-1}, u_k) \\ &\overset{??}{\leq} |\partial\mathcal{F}|_{\mathcal{D}}(u_k)\, \mathcal{D}(u_{k-1}, u_k).\end{aligned}$$

The last estimate $\overset{??}{\leq}$ would be necessary to exploit the λ-global slope in (4.6) for generating again a discrete energy estimate with a telescoping structure. However, this would mean that one has to show equality in the slope estimate of Proposition 4.13, which is false in general metric spaces, e.g. in the simple case $M = R$, $\mathcal{F}(u) = \frac{\alpha}{u}^2/2$, and $\mathcal{D}(u, w) = \arctan(|u-w|)$.

De Giorgi's variational interpolant will be a way around this problem and, much more importantly, paves the way to solve problems without semiconvexity in Banach spaces or semiglobal slopes in metric spaces. The definition of the interpolant is based on minimization only. We refer to [Amb95, Lem. 2.5] and [AGS05, Def. 3.2.1] for the first occurrences of the variational interpolant.

Definition 4.14 (De Giorgi's variational interpolant) *For a generalized metric GS $(M, \mathcal{F}, \mathcal{D}, \psi)$, a starting point $u^0 \in M$, and a time step $\tau = T/N$ we consider a discrete approximant $(u_k^\tau)_{k=0,..,N}$ obtained via the MMS for $u_0^\tau = u^0$. Then, the variational interpolants $\widetilde{u}_\tau : [0,T] \to M$ are defined via $\widetilde{u}_\tau(j\tau) = u_j^\tau$ for $j = 0, 1, \ldots, N$ and*

$$\widetilde{u}_\tau(k\tau{+}r) \text{ minimizes } w \mapsto \Phi_r(u_k^\tau, w) := r\psi\big(\frac{1}{r}\mathcal{D}(u_k^\tau, w)\big) + \mathcal{F}(w).$$

for $k = 0, \ldots, N{-}1$ and $r \in \,]0, \tau[$.

In general, the variational interpolant will not be continuous in t, but nevertheless it has good properties, because it is created by the intrinsic building blocks of the metric GS. In particular, we will not need the interpolant $\widetilde{u}_\tau$ so often, but can rely on the so-called *value function*

$$\phi(r, u_{k-1}^\tau) := \Phi_r\big(u_{k-1}^\tau, \widetilde{u}_\tau(k\tau{+}r)\big).$$

It will turn out that $r \mapsto \phi(r, u_{k-1}^\tau)$ is absolutely continuous and that the derivative can be expressed by the derivative of $r \mapsto \Phi_r(u_{k-1}^\tau, w)$. For showing this, we introduce the auxiliary function

$$\Psi : [0,\infty[^2 \to [0,\infty]; \quad \Psi(r,a) = \begin{cases} 0 & \text{for } a = 0, \\ r\psi(a/r) & \text{for } r > 0, \\ \infty & \text{for } r = 0 \text{ and } a > 0. \end{cases}$$

The following gives a series of properties of Ψ that will be used in the upcoming analysis. We leave the elementary proof to the reader.

Lemma 4.15 (Properties of Ψ) *Assume that the dissipation potential $\psi : \mathbb{R} \to [0,\infty[$ satisfies* (4.10)*, then $\Psi : [0,\infty[^2 \to [0,\infty]$ is lsc and satisfies the following properties:*

(i) *For all $a \geq 0$ the function $r \mapsto \Psi(r,a)$ is decreasing with $\partial_r \Psi(r,a) = -\psi^*\big(\psi'(a/r)\big)$ for all $r > 0$.*

(ii) *For all $a \geq 0$ the function $r \mapsto \Psi(r,a)$ is strictly convex.*

(iii) *For $0 < r_1 < r_2$ the function $a \mapsto \Psi(r_1,a) - \Psi(r_2,a)$ is strictly increasing.*

With this we are able to provide some first results concerning the value function. For this we introduce a few simplifying notations. We fix a state $u_* \in M$ and define, for $r > 0$,

$$\begin{aligned} \phi(r,u_*) &:= \inf\big\{\, \Phi_r(u_*, w) \,\big|\, w \in M \,\big\}, \\ A(r,u_*) &:= \operatorname{Arg\,min}\big\{\, \Phi_r(\overline{u}, w) \,\big|\, w \in M \,\big\} := \big\{\, u_r \in M \,\big|\, \Phi_r(\overline{u}, u_r) = \phi(r,u_*) \,\big\}, \\ d^+(r,u_*) &:= \sup\big\{\, \mathcal{D}(\overline{u}, u_r) \,\big|\, u_r \in A(r,u_*) \,\big\}, \quad d^-(r,u_*) := \inf\big\{\, \mathcal{D}(\overline{u}, u_r) \,\big|\, u_r \in A(r,u_*) \,\big\}. \end{aligned}$$

Proposition 4.16 (Value function and distances) *Consider a generalized metric GS $(M, \mathcal{F}, \mathcal{D}, \psi)$ such that $\mathcal{F}$ has compact sublevels and ψ satisfies* (4.10)*. Then for all $u_* \in M$ and $r_2 > r_1 > 0$ the functions ϕ, A, and $d^\pm$ are well defined and satisfy*

(a) $\phi(r_2, u_*) \leq \phi(r_1, u_*) \leq \mathcal{F}(u_*)$;

(b) $d^+(r_2, u_*) \geq d^-(r_2, u_*) \geq d^+(r_1, u_*)$;

(c) $\phi(r, u_*) \to \mathcal{F}(u_*)$ *for* $r \to 0^+$;

(d) *If* $u_* \in \overline{\mathrm{dom}(\mathcal{F})}$, *then* $d^+(r, u_*) \to 0$ *for* $r \to 0^+$.

Property (b) *implies that* $d^\pm(\cdot, u_*) :]0, \infty[\to \mathbb{R}$ *are increasing and continuous for* $t \in [0, \tau] \setminus J$, *where* J *is at most countable. Moreover,* $d^+(r, u_*) = d^-(r, u_*)$ *for all* $\in]0, \infty[\setminus J$.

Proof. *Step 1: Wellposedness and attainment.* We first observe that the properties that $\mathcal{F}$ is proper and has compact sublevels guarantee that $A(r, u_*)$ is nonempty and compact. Hence, the infimum in the definition of ϕ is attained. Moreover, by Weierstraß' extreme-value principle the continuous function $w \mapsto \mathcal{D}(u_*, w)$ attains its minimum and maximum on $A(t, u_*)$.

Step 2: Monotonicity (a). Clearly $\Phi_r(u_*, w) = \Psi(r, \mathcal{D}(u_*, w)) + \mathcal{F}(w)$ is decreasing in r. Thus, choosing any $u_j \in A(r_j, u_*)$ we have

$$\mathcal{F}(u_*) = \Phi_{r_1}(u_*, u_*) \geq \phi(r_1, u_*) = \Phi_{r_1}(u_*, u_1) \geq \Phi_{r_2}(u_*, u_1) \geq \Phi_{r_2}(u_*, u_2) = \phi(r_2, u_*),$$

which is the desired monotonicity.

Step 3. Intertwining property. We trivially have $d^+ \geq d^-$ because $\sup \geq \inf$. However, it is absolutely nontrivial that $d^+(r_1, u_*)$ can be estimated from above by $d^-(r_2, u_*)$ for all $r_2 > r_1$. To see this, we have to exploit the special structure of $\Phi_r(u_*, w) = \Psi(r, \mathcal{D}(u_*, w)) + \mathcal{F}(w)$. Again choose arbitrary $u_j \in A(r_j, u_*)$ and set $\mathcal{D}_j := \mathcal{D}(u_*, u_j)$, then we have

$$\begin{aligned}\Psi(r_1, \mathcal{D}_1) + \mathcal{F}(u_1) &= \Phi_{r_1}(u_*, u_1) = \phi(r_1, u_*) \leq \Phi_{r_1}(u_*, u_2) = \Psi(r_1, \mathcal{D}_2) + \mathcal{F}(u_2) \\ &= \Phi_{r_2}(u_*, u_2) + \Psi(r_1, \mathcal{D}_2) - \Psi(r_2, \mathcal{D}_2) \leq \Phi_{r_2}(u_*, u_1) + \Psi(r_1, \mathcal{D}_2) - \Psi(r_2, \mathcal{D}_2) \\ &= \Psi(r_2, \mathcal{D}_1) + \mathcal{F}(u_1) + \Psi(r_1, \mathcal{D}_2) - \Psi(r_2, \mathcal{D}_2).\end{aligned}$$

We observe that $\mathcal{F}(u_1)$ can be eliminated on both ends, and rearranging gives

$$\Psi(r_1, \mathcal{D}_1) - \Psi(r_2, \mathcal{D}_1) \leq \Psi(r_1, \mathcal{D}_2) - \Psi(r_2, \mathcal{D}_2).$$

Now we can exploit Lemma 4.15(iii) and conclude $\mathcal{D}(u_*, u_1) = \mathcal{D}_1 \leq \mathcal{D}_2 = \mathcal{D}(u_*, u_2)$. As $u_j \in A(r_j)$ were arbitrary, we can take the supremum over u_1 and the infimum over u_2 and obtain $d^+(r_1, u_*) \leq d^-(r_2, u_*)$ as desired.

Step 4: $d^+ = d^-$ *whenever one is continuous.* By the last step we know that d^+ and d^- are increasing functions. Hence, they are continuous for all t except for an at most countable jump set J^+ or J^-, respectively. However, if d^+ is continuous at $r_* > 0$, then for $0 < \varepsilon_n \to 0$ we have

$$\begin{aligned}d^+(r_*, u_*) \leftarrow d^+(r_*{-}\varepsilon_n, u_*) &\leq d^-(r^*{-}\varepsilon_n/2, u_*) \leq d^+(r_*, u_*) \\ &\leq d^-(r_*{+}\varepsilon_n/2, u_*) \leq d^+(r_*{+}\varepsilon_n, u_*) \to d^+(r_*, u_*).\end{aligned}$$

As $\varepsilon_n \to 0$ was arbitrary, we conclude that d^- is continuous at $r = r_*$, i.e. $J^- \subset J^+$, as well as $d^+(r_*, u_*) = d^-(r_*, u_*)$. Interchanging "+" and "−" we obtain $J^+ = J^- =: J$, and the final assertion is established.

Step 5: $\phi(r, u_*) \to \mathcal{F}(u_*)$. From Step 1 we have $\phi(r, u_*) \leq \mathcal{F}(u_*)$. Hence, by the monotonicity we have $\phi(r, u_*) \to \phi_* \leq \mathcal{F}(u_*)$ for $r \to 0^+$.

Choose $d^+(r,u_*) \to d_* > 0$ for $r \to 0$, then $\phi(r,u_*) = \Psi(r,\mathcal{D}(u_*,u_r)) + \mathcal{F}(u_r) \geq \Psi(r,d^+(r,u_*)) + \mathcal{F}_{\min} \to \infty$. This means $\phi_* = \mathcal{F}(u_*) = \infty$ and the assertion holds.

If $d^+(r,u_*) \to 0$ for $r \to 0^+$, then there exists $u_r \in A(t,u_*)$ such that $\mathcal{D}(u_*,u_r) \leq d^+(r,u_*) \to 0$, and the lsc of $\mathcal{F}$ implies $\liminf_{r\to 0^+} \mathcal{F}(u_r) \geq \mathcal{F}(u_*)$. Because of $\Psi \geq 0$ we find

$$\phi(r,u_*) = \Phi_r(u_*,u_r) = \Psi(r,\mathcal{D}(u_*,u_r)) + \mathcal{F}(u_r) \geq \mathcal{F}(u_r),$$

which now implies $\phi_* \geq \mathcal{F}(u_*)$. Thus $\phi_* = \mathcal{F}(u_*)$ is established.

Step 6: $d^+(r,u_*) \to 0$. For arbitrary $w \in \mathrm{dom}(\mathcal{F})$ and $u_r \in A(r,u_*)$ we have

$$\phi(r,u_*) = r\psi\big(\frac{1}{r}\mathcal{D}(u_*,u_r)\big) + \mathcal{F}(u_r) \leq r\psi\big(\frac{1}{r}\mathcal{D}(u_*,w)\big) + \mathcal{F}(w).$$

Solving for $\mathcal{D}(u,u_r)$ we use the strict monotonicity of ψ and find

$$\mathcal{D}(u,u_r) \leq r\psi^{-1}\Big(\psi\big(\frac{1}{r}\mathcal{D}(u,w)\big) + \frac{1}{\tau}\big(\mathcal{F}(w) - \mathcal{F}(u_r)\big)\Big).$$

As ψ is convex with $\psi(0) = 0$, the inverse ψ^{-1} is concave with $\psi^{-1}(0) = 0$, and thus subadditive. Hence, we have

$$\mathcal{D}(u,u_r) \leq \underbrace{r\psi^{-1}\Big(\psi\big(\frac{1}{r}\mathcal{D}(u,w)\big)\Big)}_{=\mathcal{D}(u,w)} + r\psi^{-1}\Big(\frac{1}{\tau}\big[\mathcal{F}(w) - \mathcal{F}(u_r)\big]_+\Big). \tag{4.12}$$

As Ψ is superlinear, we have $\psi^{-1}(a) = o(a)_{a\to\infty}$ such that the last term tends to 0 for $r \to 0^+$. This implies $\lim_{r\to 0^+} d^+(r,u_*) \leq \mathcal{D}(u_*,w)$. Since $w \in \mathrm{dom}(\mathcal{F})$ was arbitrary, and $u_* \in \overline{\mathrm{dom}(\mathcal{F})}$ we obtain $\lim_{r\to 0^+} d^+(r,u_*) = 0$. ∎

We are now ready to prove the following discrete energy-dissipation estimate, which first appears in [Amb95, Lem. 2.5] and in a slightly more elaborate version in [AGS05, Thm. 3.1.4]. According to several oral presentations of these authors, the following result should be called "*De Giorgi's lemma*", as it was inspired by his personal communication. Our version is slightly more general, as we treat arbitrary dissipation potentials ψ. We can now show that the value function $r \mapsto \phi(r,u_*)$ is differentiable and satisfies

$$\frac{\mathrm{d}}{\mathrm{d}r}\phi(r,u_*) = -\psi^*\big(\psi'\big(\frac{1}{r}\, d^+(r,u_*)\big)\big) \quad \text{a.e. in }]0,\tau].$$

Theorem 4.17 (De Giorgi's lemma) *Consider a generalized metric GS $(M,\mathcal{F},\mathcal{D},\psi)$ where $\mathcal{F}$ has compact sublevels and ψ satisfies* (4.10)*. Fix $u_* \in \mathrm{dom}(\mathcal{F})$ and $\tau > 0$ and define ϕ, d^+, and the variational interpolant $\widetilde{u} : [0,\tau] \to M$ as above. Then, we have*

$$\phi(\tau,u_*) + \int_0^\tau \psi^*\big(\psi'\big(\frac{1}{r}\, d^+(r,u_*)\big)\big)\,\mathrm{d}r = \mathcal{F}(u_*). \tag{4.13}$$

If additionally the function $r \mapsto \widetilde{u}_\tau(r) \in M$ is measurable, then

$$\tau\psi\big(\frac{1}{\tau}\mathcal{D}(u_*,\widetilde{u}_\tau(\tau))\big) + \int_0^\tau \psi^*\big(|\partial\mathcal{F}|_{\mathcal{D}}(\widetilde{u}_\tau(r))\big)\,\mathrm{d}r \leq \mathcal{F}(u_*). \tag{4.14}$$

Proof. *Step 1:* (4.13) *implies* (4.14). We exploit the metric slope estimate in Proposition 4.13 and the monotonicity of ψ^* giving

$$\psi^*\big(|\partial\mathcal{F}|_{\mathcal{D}}(\widetilde{u}_\tau(r))\big) \leq \psi^*\big(\psi'\big(\frac{\mathcal{D}(u_*,\widetilde{u}_\tau(r))}{r}\big)\big) \leq \psi^*\big(\psi'\big(\frac{d^+(r,u_*)}{r}\big)\big) \ \text{ for } r \in \,]0,\tau]. \tag{4.15}$$

As $r \mapsto \widetilde{u}_\tau(r) \in M$ is measurable, and $u \mapsto \psi^*\big(|\partial\mathcal{F}|_{\mathcal{D}}(u)\big) \in [0,\infty]$ is Borel measurable (as a composition of a continuous and a lsc map), we see that $r \mapsto \psi^*\big(\psi'\big(\frac{1}{r}\,d^+(r,u_*)\big)\big) \geq 0$ is integrable and we obtain the desired estimate (4.14) by integrating (4.15) and exploiting (4.13).

Step 2: Local Lipschitz continuity of $]0,\tau] \ni r \mapsto \phi(r,u_*)$. For $0<r<s$ and all $u_s \in A(s,u_*)$ and $u_r \in A(r,u_*)$ we have

$$\begin{aligned}\Phi_s(u_*,u_r) - \Phi_r(u_*,u_r) \geq \Phi_s(u_*,u_s) - \Phi_r(u_*,u_r) &= \phi(s,u_*) - \phi(r,u_*)\\ &\geq \Phi_s(u_*,u_s) - \Phi_r(u_*,u_s).\end{aligned}$$

In the first and last term the appearance of $\mathcal{F}$ cancels and we are left with the estimate

$$\Psi(s,\mathcal{D}(u_*,u_r)) - \Psi(r,\mathcal{D}(u_*,u_r)) \geq \phi(s,u_*) - \phi(r,u_*) \geq \Psi(s,\mathcal{D}(u_*,u_s)) - \Psi(r,\mathcal{D}(u_*,u_s)).$$

By Lemma 4.15(iii) the mapping $a \mapsto \Psi(s,a) - \Psi(r,a)$ is decreasing hence, we may maximize for $u_r \in A(r,u_*)$ and minimize for $u_s \in A(s,u_*)$ to obtain

$$\begin{aligned}&\Psi(s,d^+(r,u_*)) - \Psi(r,d^+(r,u_*))\\ &\geq \phi(s,u_*) - \phi(r,u_*) \geq \Psi(s,d^-(s,u_*)) - \Psi(r,d^-(s,u_*)).\end{aligned} \tag{4.16}$$

From this we can derive Lipschitz continuity by assuming $0<r_*\leq r<s\leq\tau$, namely

$$\begin{aligned}0 \geq \phi(s,u_*) - \phi(r,u_*) &\geq \int_r^s \partial_t\Psi\big(t,d^-(s,u_*)\big)\,\mathrm{d}t\\ &\overset{\text{(i)}}{=} -\int_r^s \psi^*\big(\psi'\big(\frac{1}{t}d^-(t,u_*)\big)\big)\,\mathrm{d}t \ \geq\ -(s{-}r)\,\psi^*\big(\psi'\big(\frac{1}{r_*}d^+(\tau,u_*)\big)\big) =: -(s{-}r)K_*,\end{aligned}$$

where we used the monotonicity of $\psi^*\circ\psi'$ in the last step and $\overset{\text{(i)}}{=}$ indicates the identity derived in Lemma 4.15(i). Thus, we have Lipschitz continuity with Lipschitz constant K_* on $[r_*,\tau]$.

Step 3: Identification of the derivative. Because of local Lipschitz continuity, we have differentiability a.e. in $]0,\tau]$. To identify the derivative we divide (4.16) by $s-r>0$ and obtain, again using Lemma 4.15(i),

$$\begin{aligned}-\psi^*\big(\psi'\big(\frac{d^+(r,u_*)}{r}\big)\big) &= \lim_{s\to r^+} \frac{\Psi(s,d^+(r,u_*)){-}\Psi(r,d^+(r,u_*))}{s\,-\,r}\\ &\geq \limsup_{s\to r^+} \frac{\phi(s,u_*){-}\phi(r,u_*))}{s\,-\,r} \qquad \text{and}\end{aligned} \tag{4.17}$$

$$\begin{aligned}-\psi^*\big(\psi'\big(\frac{d^-(s,u_*)}{s}\big)\big) &= \lim_{r\to s^-} \frac{\Psi(s,d^-(s,u_*)){-}\Psi(r,d^-(s,u_*))}{s\,-\,r}\\ &\leq \liminf_{r\to s^-} \frac{\phi(s,u_*){-}\phi(r,u_*))}{s\,-\,r}.\end{aligned} \tag{4.18}$$

Denote by $\mathbb{T} \subset]0,\tau]$ the set of points where $r \mapsto \phi(r,u_*)$ is differentiable and where d^+ and d^- are continuous. Together with Propositions 4.16 we know that $\mathbb{T}$ is a set of full measure and that $d^+(t,u_*) = d^-(t,u_*)$ on $\mathbb{T}$. Taking $r = t$ in (4.17) and $s = t$ in (4.18) we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\phi(t,u_*) = -\psi^*\big(\psi'\big(\frac{d^{\pm}(t,u_*)}{t}\big)\big) \quad \text{for all } t \in \mathbb{T}.$$

Step 4: Integral formula on $[0,\tau]$. Step 3 implies, for all $r \in]0,\tau[$, the relation

$$\phi(\tau,u_*) + \int_r^\tau \psi^*\big(\psi'\big(\frac{1}{s}\,d^+(s,u_*)\big)\big)\,\mathrm{d}s = \phi(r,u_*).$$

By Proposition 4.16(c) we have $\phi(r,u_*) \to \mathcal{F}(u_*)$ for $r \to 0^+$, i.e. convergence on the right-hand side. The convergence for $r \to 0^+$ on the left-hand side follows from Beppo Levi's monotone convergence theorem as the integrand is nonnegative. Thus, identity (4.13) is established. ■

4.5 Existence of curves of maximal slopes via MMS

We are now ready to show the existence of ψ-curves of maximal slope. Of course, the construction is based on the MMS and it will follow closely the proof of Theorem 3.13 for Banach-space gradient systems. The major difference is that we do no longer assume any type of λ-convexity (of λ-global slopes) and exploit De Giorgi's variational interpolant instead.

Theorem 4.18 (Existence of ψ-curves of maximal slope) *Consider a generalized metric gradient system $(M,\mathcal{F},\mathcal{D},\psi)$ that additionally satisfies*

$\mathcal{F}$ *has compact sublevels* $S^{\mathcal{F}}_E \subset M$; (4.19a)

$|\partial\mathcal{F}|_{\mathcal{D}} : M \to [0,\infty]$ *is lower semicontinuous*; (4.19b)

$\psi \in \mathrm{C}^1([0,\infty[)$ *and is strictly convex*; (4.19c)

$(M,\mathcal{F},\mathcal{D},\psi)$ *satisfies the metric chain-rule inequality* (4.8). (4.19d)

Then, for all $u^0 \in \mathrm{dom}(\mathcal{F})$ there exists a ψ-curve of maximal slope $u : [0,\infty[\to M$ satisfying $u(0) = u^0$.

Proof. We fix a time $T > 0$ and construct solutions on $[0,T]$ at first. For $N \in \mathbb{N}$ we define the time step $\tau > 0$.

Step 0: Construction of approximants. Because of the compact sublevels of $\mathcal{F}$ (see (4.19a)) we know that $\mathcal{F}(u) \geq \mathcal{F}_{\min}$ for all $u \in M$. Moreover, using $\mathcal{F}(u^0) < \infty$ we know that the MMS produces solutions $(u^\tau_k)_{k=0,..,N}$ lying in the compact sublevel $S^{\mathcal{F}}_{\mathcal{F}(u^0)}$. Moreover, we can construct De Giorgi's variational interpolant $\widetilde{u}_\tau : [0,T] \to M$ and apply De Giorgi's lemma (i.e. Theorem 4.17) on each time interval $[k\tau{-}\tau, k\tau]$ and obtain

$$\mathcal{F}(\widetilde{u}_\tau(k\tau)) + \int_{k\tau-\tau}^{k\tau} \Big(\psi\big(S_\tau(r)\big) + \psi^*\big(G_\tau(r)\big)\Big)\,\mathrm{d}r = \mathcal{F}(\widetilde{u}_\tau(k\tau{-}\tau)) \tag{4.20}$$

for $k = 1, \ldots, N = T/\tau$, where we introduced the functions S_τ and G_τ as follows:

$$S_\tau(t) = \frac{1}{\tau}\mathcal{D}(\widetilde{u}_\tau(k\tau{-}\tau), \widetilde{u}_\tau(k\tau)) \qquad \text{for } t \in \,]k\tau{-}\tau, k\tau],$$
$$G_\tau(t) = \psi'\big(\frac{1}{r}\, d^+(t, \widetilde{u}_\tau(k\tau{-}\tau))\big) \qquad \text{for } t = k\tau{-}\tau + r \in \,]k\tau{-}\tau, k\tau].$$

We note that it is tempting to replace $G_\tau(t)$ by the smaller value $|\partial\mathcal{F}|_\mathcal{D}(\widetilde{u}_\tau(t))$ (cf. the slope estimate in Proposition 4.13), however we refrain from doing so because then we would need to show measurability (which is possible but technical). It is better to keep G_τ as defined, which is automatically measurable and apply the slope estimate later (see Step 3).

Step 1: A priori estimates. Clearly, summing (4.20) over $k = 1, \ldots, N$ leads to a telescope sum and we find

$$\int_0^T \big(\psi(S_\tau(t)) + \psi^*(G_\tau(t))\big)\,\mathrm{d}t = \mathcal{F}(\widetilde{u}_\tau(0)){-}\mathcal{F}(\widetilde{u}_\tau(T)) \leq \mathcal{F}(u^0){-}\mathcal{F}_{\min} =: \Delta_\mathcal{F} < \infty. \quad (4.21)$$

This provides superlinear a priori estimates for S_τ and G_τ.

We also want to derive a "kind of equi-continuity" of the sequence $(\widetilde{u}_\tau)_\tau$. Of course, we cannot expect the individual $\widetilde{u}_\tau$ for fixed $\tau = T/N$ to be continuous but it should be close to a continuous function. We will show that there exists a modulus of continuity $\widetilde{\omega}$ such that

$$\mathcal{D}\big(\widetilde{u}_\tau(s), \widetilde{u}_\tau(t)\big) \leq \widetilde{\omega}\big(\tau + |t{-}s|\big) \quad \text{for all } s, t \in [0, T] \text{ and all } \tau = T/N. \quad (4.22)$$

For this we first quantify the convergence $d^+(r, u_*) \to 0$ in Proposition 4.16(d), i.e. we show that variational interpolants $\widetilde{u}_\tau$ are close to the nodal points $\widetilde{u}_\tau(k\tau)$. Setting $u = w = \widetilde{u}_\tau(k\tau)$ in (4.12), for $k = 0, \ldots, N{-}1$ and $r \in \,]0, \tau[$ we find

$$\begin{aligned}\mathcal{D}\big(\widetilde{u}_\tau(k\tau), \widetilde{u}_\tau(k\tau{+}r)\big) &\leq r\,\psi^{-1}\Big(\frac{1}{r}\big(\mathcal{F}(\widetilde{u}_\tau(k\tau)) - \mathcal{F}(\widetilde{u}_\tau(k\tau{+}r))\big)\Big)\\ &\leq r\,\psi^{-1}\big(\frac{1}{r}\,\Delta_\mathcal{F}\big) \;=:\; \widehat{\omega}(r) = o(1)_{r\to 0^+}.\end{aligned}$$

Here we used that ψ^{-1} is increasing and growing less than linear, because ψ is superlinear. Hence $\widehat{\omega}$ is an modulus of continuity.

We define the function $\underline{t}_\tau : [0, T] \to [0, T]$ via $\underline{t}_\tau(s) = \max\{\, k\tau \mid k\tau \leq s \,\} = \tau\lfloor s/\tau\rfloor$. With this, we obtain, for $0 \leq r < s \leq T$, the estimate

$$\begin{aligned}\mathcal{D}\big(\widetilde{u}_\tau(r), \widetilde{u}_\tau(s)\big) &\leq \mathcal{D}\big(\widetilde{u}_\tau(r), \widetilde{u}_\tau(\underline{t}_\tau(r))\big) + \mathcal{D}\big(\widetilde{u}_\tau(\underline{t}_\tau(r)), \widetilde{u}_\tau(\underline{t}_\tau(s))\big) + \mathcal{D}\big(\widetilde{u}_\tau(\underline{t}_\tau(s)), \widetilde{u}_\tau(s)\big)\\ &\leq \widehat{\omega}\big(r{-}\underline{t}_\tau(r)\big) + \sum_{k=\lfloor r/\tau\rfloor}^{\lfloor s/\tau\rfloor-1} \tau\,\frac{1}{\tau}\,\mathcal{D}\big(\widetilde{u}_\tau(k\tau), \widetilde{u}_\tau(k\tau{+}\tau)\big) + \widehat{\omega}\big(s{-}\underline{t}_\tau(s)\big)\\ &\leq \widehat{\omega}(\tau) + \int_{\underline{t}_\tau(r)}^{\underline{t}_\tau(s)} S_\tau(t)\,\mathrm{d}t + \widehat{\omega}(\tau). \end{aligned} \quad (4.23)$$

We proceed as in the Banach-space case (cf. Section 3.5) by estimating $S_\tau \leq \frac{1}{\mu}\,\mu S_\tau \leq \frac{1}{\mu}\big(\psi(S_\tau){+}\psi^*(\mu)\big)$ and obtain

$$\mathcal{D}\big(\widetilde{u}_\tau(r), \widetilde{u}_\tau(s)\big) \leq 2\widehat{\omega}(\tau) + \int_{[r-\tau]_+}^{s} S_\tau(t)\,\mathrm{d}t \leq 2\widehat{\omega}(\tau) + \omega_\psi^{\Delta_\mathcal{F}}\big(s{-}r + \tau\big),$$

where ω_ψ^B is defined in (3.24) and $\Delta_\mathcal{F}$ in (4.21). Hence, (4.22) is established with $\widetilde\omega = 2\widehat\omega + \omega_\psi^{\Delta_\mathcal{F}}$.

Step 2: Extraction of converging subsequences. Since ψ and ψ^* are superlinear, the a priori estimate (4.21) and the criterion of de la Vallée-Poussin guarantee that the sequences $(S_\tau)_\tau$ and $(G_\tau)_\tau$ are equi-integrable and there exists a subsequence (not relabeled) such that

$$S_\tau \rightharpoonup S_0 \quad \text{and } G_\tau \rightharpoonup G_0 \quad \text{in } \mathrm{L}^1([0,T]).$$

Moreover, the equi-continuity (4.22) allows us to employ the generalized Arzelà-Ascoli theorem, such that along a further subsequence (not relabeled) we have pointwise convergence to a continuous limit function $u : [0,T] \to M$, namely

$$\forall\, t \in [0,T] : \quad \widetilde u_\tau(t) \to u(t) \ \text{ as } \tau \to 0^+.$$

Because of $\widetilde u_\tau(0) = u^0$, we also have $u(0) = u^0$. By passing to the limit in (4.23) we obtain

$$\forall\, s,t \in [0,T] \text{ with } s<t : \quad \mathcal{D}(u(s),u(t)) \le \int_s^t S_0(t)\,\mathrm{d}t, \tag{4.24}$$

which shows $u \in \mathrm{AC}([0,T];M)$.

Step 3: Derivation of (EDI). We return to (4.21) in the form

$$\mathcal{F}(\widetilde u_\tau(T)) + \int_0^T \psi(S_\tau(t))\,\mathrm{d}t + \int_0^T \psi^*(G_\tau(t))\,\mathrm{d}t = \mathcal{F}(\widetilde u_\tau(0)),$$

and calculate the liminf for $\tau \to 0^+$ for the three terms on the left-hand side.

From $\widetilde u_\tau(T) \to u(T)$ and the lsc of $\mathcal{F}$ we have $\liminf_{\tau\to 0^+} \mathcal{F}(\widetilde u_\tau(T)) \ge \mathcal{F}(u(T))$.

For the second term we observe that the mapping $\alpha \mapsto \int_0^T \psi(\alpha(t))\,\mathrm{d}t$ is convex and strongly lsc on $\mathrm{L}^1([0,T])$. Hence, the mapping is also weakly lsc and $S_\tau \rightharpoonup S_0$ implies $\liminf_{\tau\to 0^+} \int_0^T \psi(S_\tau(t))\,\mathrm{d}t \ge \int_0^T \psi(S_0(t))\,\mathrm{d}t \ge \int_0^T \psi\big(|\dot u|_\mathcal{D}(t)\big)\,\mathrm{d}t$. For the last estimate we used that $\psi : [0,\infty[\to [0,\infty[$ is increasing and the characterization of the metric speed in Theorem 4.4, i.e. $|\dot u|_\mathcal{D} \le S_0$ because of (4.24).

For the third term we fix $t \in [0,T]$ and exploit the slope estimate in Proposition 4.13 as well as the lsc of the slope $|\partial\mathcal{F}|_\mathcal{D}$, see assumption (4.19b). Using that $\psi^* : [0,\infty[\to [0,\infty[$ is continuous and increasing and that $\widetilde u_\tau(t) \to u(t)$ we have

$$\psi^*\big(|\partial\mathcal{F}|_\mathcal{D}(u(t))\big) \le \liminf_{\tau\to 0^+} \psi^*\big(|\partial\mathcal{F}|_\mathcal{D}(\widetilde u_\tau(t))\big) \le \liminf_{\tau\to 0^+} \psi^*\big(G_\tau(t)\big).$$

Thus, Fatou's lemma yields $\liminf_{\tau\to 0^+} \int_0^T \psi^*\big(G_\tau(t)\big)\,\mathrm{d}t \ge \int_0^T \psi^*\big(|\partial\mathcal{F}|_\mathcal{D}(u(t))\big)\,\mathrm{d}t$.

In summary, we find the EDI

$$\mathcal{F}(u(T)) + \int_0^T \Big(\psi\big(|\dot u|_\mathcal{D}(t)\big) + \psi^*\big(|\partial\mathcal{F}|_\mathcal{D}(u(t))\big)\Big)\,\mathrm{d}t \le \mathcal{F}(u(0)).$$

Step 4: Derivation of (EDB). As we have assumed the abstract metric chain-rule inequality in (4.19d) we can apply the metric energy-dissipation principle from Proposition 4.12. Hence, u is a ψ-curve of maximal slope. ∎

As in Section 3 one can infer additional convergences (along the chosen subsequence), if we assume strict convexity of ψ and ψ^*:

$$\begin{aligned}\forall\, t\in[0,T]:\quad & \widetilde u_\tau(t)\to u(t) \text{ and } \mathcal F(\widetilde u_\tau(t))\to\mathcal F(u(t)).\\ \forall_{\text{a.a.}} t\in[0,T]:\quad & \frac{1}{\tau}\mathcal D\big(\widetilde u_\tau(\underline t_\tau(t)),\widetilde u_\tau(\underline t_\tau(t){+}\tau)\big)\to |\dot u|_{\mathcal D}(t)\\ & \text{and } |\partial\mathcal F|_{\mathcal D}(\widetilde u_\tau(t))\to|\partial\mathcal F|_{\mathcal D}(u(t)).\end{aligned}$$

We emphasize that there is no easy way of showing uniqueness in this general setting. Example 4.9(A) provides a case where all assumption of the above existence theorem are satisfied, but uniqueness fails. Moreover, the following example shows that one may have even uncountably many solutions for a given initial point u^0.

Example 4.19 (Non-uniqueness for curves of maximal slope) *Consider the gradient system $(\mathbb R^2,\mathcal F,|\cdot|_1,\psi_{\rm quadr})$ with $\mathcal F(u)=u_1+u_2$ and $|(v_1,v_2)|_1=|v_1|+|v_2|$. A curve $u:[0,T]\to\mathbb R^2$ is a curve of maximal slope if and only if $u\in \mathrm W^{1,\infty}([0,T];\mathbb R^2)$ with*

$$\dot u_1(t),\dot u_2(t)\in[-1,0] \;\text{ and }\; \dot u_1(t)+\dot u_2(t)=-1 \quad \text{a.e. in } [0,T].$$

Thus, all the curves $u(t)=u^0-t(1{-}\theta,\theta)+g\sin(\omega t)(1,-1)$ with $\theta\in[0,1]$ and $|g\omega|\leq\min\{\theta,1{-}\theta\}$ are curves of maximal slope starting at u^0.

4.6 Metric evolutionary variational inequalities (EVI)

We recall that in the case of Hilbert spaces (see Section 2.5) the evolutionary variational inequality (EVI)$_\lambda$ did only use the norms $\|u{-}w\|$ and no time derivatives $\dot u$ or subdifferentials $\partial^{\rm F}\mathcal F(u)$ appear. Hence, we can easily define the corresponding EVI notion for metric GS. We emphasize that this theory is restricted to the quadratic dissipation function $\psi=\psi_{\text{quadr}}:\delta\mapsto\delta^2/2$, thus use the short-hand $(M,\mathcal F,\mathcal D):=(M,\mathcal F,\mathcal D,\psi_{\text{quadr}})$.

Definition 4.20 (Metric EVI$_\lambda$ solutions) *We consider a metric GS $(M,\mathcal F,\mathcal D)$. Then, we call $u:[0,T]\to M$ an (EVI)$_\lambda$ solution, if*

$$\begin{aligned}&\forall\, s,t\in[0,T] \text{ with } s<t\ \forall\, w\in\mathrm{dom}(\mathcal F):\\ &\frac12\mathcal D(u(t),w)^2\leq\frac12\mathrm e^{-\lambda(t-s)}\mathcal D(u(s),w)^2+M_\lambda(t{-}s)\big(\mathcal F(w)-\mathcal F(u(t))\big),\end{aligned}$$

where $M_\lambda(r)=\int_0^r\mathrm e^{-\lambda(r-s)}\,\mathrm ds$.

We will see below that it is possible to derive uniqueness for EVI solutions, however it is very difficult to establish existence. Except for the Hilbert-space case discussed in Section 2, there is no direct way of showing that curves of maximal slope (with $\psi=\psi_{\text{quadr}}$) are also EVI solutions if $\mathcal F$ satisfies a suitable λ-convexity condition.

Instead, there is an independent existence theory for EVI solutions based on rather strong assumptions on the metric space $(M,\mathcal D)$ and on the functional $\mathcal F$. We refer to [AGS05, Cha. 4] and [Sav07, DaS14, MuS22] because the general existence theory is ongoing research.

The major new assumption is that of the existence of geodesic curves.

Definition 4.21 (Geodesic metric spaces) *In a metric space* $(M, \mathcal{D})$ *a curve* $\gamma : [0,1] \to M$ *is called a* (constant speed) geodesic *if*

$$\forall\, r, s \in [0,1] : \quad \mathcal{D}(\gamma(r), \gamma(s)) = |s{-}r|\, \mathcal{D}(\gamma(0), \gamma(1)).$$

In this case we say that the geodesic γ *connects the points* $\gamma(0)$ *and* $\gamma(1)$ *and write* $\mathrm{Geod}\big(\gamma(0), \gamma(1)\big)$ *for the set of all such geodesics.*

The metric space $(M, \mathcal{D})$ *is called a* geodesic space*, if for all* $u_0, u_1 \in M$ *there exists a geodesic connecting* u_0 *and* u_1*.*

A function $\mathcal{F} : M \to \mathbb{R}_\infty$ *is called* geodesically λ-convex *if*

$$\forall\, u_0, u_1 \in \mathrm{dom}(\mathcal{F})\ \exists\, \gamma \in \mathrm{Geod}(u_0, u_1)\ \forall\, s \in [0,1] :$$
$$\mathcal{F}(\gamma(s)) \le (1{-}s)\mathcal{F}(\gamma(0)) + s\mathcal{F}(\gamma(1)) - \frac{\lambda}{2}\, s(1{-}s)\, \mathcal{D}(\gamma(0), \gamma(1))^2.$$

With these conditions we are able to state the following simplified version of the existence result in [AGS05, Thm. 4.0.4]. Again, the construction uses the MMS and, because of uniqueness, the constructed solutions are minimizing movements in the sense of Definition 4.1. In this case we also have a true gradient flow $(S_t)_{t\ge 0}$ on $\overline{\mathrm{dom}(\mathcal{F})}$, similar to Theorem 2.10 for Hilbert spaces.

Theorem 4.22 (Existence of EVI solutions) *Consider the metric GS* $(M, \mathcal{F}, \mathcal{D})$ *(with* $\psi = \psi_{\text{quadr}}$*) with the following properties*

$$(M, \mathcal{D}) \text{ is a geodesic space}, \tag{4.25a}$$
$$\forall\, u_* \in M : \quad u \mapsto \frac{1}{2}\mathcal{D}(u_*, u)^2 \text{ is geodesically } 1\text{-convex}, \tag{4.25b}$$
$$\exists\, \lambda \in \mathbb{R} : \quad \mathcal{F}; M \to \mathbb{R}_\infty \text{ is geodesically } \lambda\text{-convex}. \tag{4.25c}$$

Then, for all $u^0 \in \mathscr{D} := \overline{\mathrm{dom}(\mathcal{F})}$ *there exists a unique (EVI)*$_\lambda$ *solution* $u : [0, \infty[\to M$ *which satisfies* $u(0) = u^0$ *and* $u \in \mathrm{MM}(M, \mathcal{F}, \mathcal{D})$*.*

Moreover, the mapping $S_t : \mathscr{D} \to \mathscr{D}$ *defined by the unique solutions via* $S_t(u(0)) := u(t)$ *is a* λ*-contractive, continuous semigroup, namely*

(S1) $S_t : \mathscr{D} \to \mathscr{D}, \quad S_0 = \mathrm{id}_{\mathscr{D}}, \quad S_t \circ S_r = S_{t+r}$ *for all* $r, t \ge 0$.

(S2) For all u^0 *the function* $[0, \infty[\, \ni t \mapsto S_t(u^0)$ *is continuous.*

(S3) For all $u_0, u_1 \in \mathscr{D}$ *we have* $\mathcal{D}\big(S_t(u_0), S_t(u_1)\big) \le \mathrm{e}^{-\lambda t}\mathcal{D}(u_0, u_1)$.

The critical condition in the above theorem is that of the geodesic 1-convexity of $u \mapsto \frac{1}{2}\mathcal{D}(u_*, u)^2$ in (4.25b). This condition is satisfied in Hilbert spaces, but it does not hold for many geodesic spaces. In particular, it does not hold for the Wasserstein space $(\mathrm{Prob}(\overline{\Omega}), \mathrm{W}_2)$ from Example 4.9(C). Thus, in [AGS05, Thm. 4.0.4] condition (4.25b) is replaced by a weaker one.

In [MuS20, Ch. 3+4] the question is addressed how EVI solutions and curves of maximal slope are related. From [MuS20, Thm. 3.5, cf. (3.17)] one easily sees that every EVI solution is a curve of maximal slope. The reverse statement that a curve of maximal slope is also an EVI solution (and hence unique) is more desirable, but it is known only under strong additional conditions, see [MuS20, Thm. 4.2]. In particular, one needs an independent existence result for EVI solutions.

For the proof of the above existence result we refer to [AGS05, Cha. 4]. Here we provide the analysis that is necessary for establishing the λ-contractivity. For this we derive a few properties (P.n) for all EVI solutions u.

(P.1) Finite energy: For all $t \gneqq 0$ we have $\mathcal{F}(u(t)) < \infty$.

We insert $s = 0$ and $w \in \mathrm{dom}(\mathcal{F}) \neq \emptyset$ into (EVI)$_\lambda$ and obtain after dropping $\frac{1}{2}\mathcal{D}(u(t), w)^2$ the estimate

$$\mathcal{F}(u(t)) \leq \mathcal{F}(w) + \frac{\mathrm{e}^{-\lambda t}}{2M_\lambda(t)}\,\mathcal{D}(u(0), w)^2 < \infty.$$

(P.2): $t \mapsto \mathcal{F}(u(t))$ is decreasing.

For $0 < s < t$ we insert $w = u(s)$ into (EVI)$_\lambda$ and obtain

$$\mathcal{F}(u(t)) \leq \mathcal{F}(u(s)) + \frac{\mathrm{e}^{-\lambda(t-s)}}{2M_\lambda(t)}\big(\,0\; - \frac{1}{2}\mathcal{D}(u(t), u(s))^2\big) \leq \mathcal{F}(u(s)).$$

If $u(0) \in \mathrm{dom}(\mathcal{F})$ we can also do this for $s = 0$, whereas in the case $\mathcal{F}(u(0)) = \infty$ we have $\infty = \mathcal{F}(u(0)) > \mathcal{F}(u(s)) \geq \mathcal{F}(u(t))$ for $0 < s < t$.

(P.3) Local Hölder continuity: $u \in C^{1/2}_{\mathrm{loc}}(]0, \infty[; M)$.

Choose $[t_0, T] \Subset\,]0, \infty[$ (compactly contained), then for $t_0 \leq s < t \leq T$ and $w = u(s)$ in (EVI)$_\lambda$ we find

$$\mathcal{D}(u(s), u(t))^2 \leq 2M_\lambda(t{-}s)\big(\mathcal{F}(u(s)) - \mathcal{F}(u(t))\big) \leq C_{t_0,T,\lambda}\,|t{-}s|\,\big(\mathcal{F}(u(t_0)) - \mathcal{F}(u(T))\big).$$

This implies $\mathcal{D}(u(s), u(t)) \leq \widetilde{C}_{t_0,T,\lambda}\,|t{-}s|^{1/2}$ as desired.

(P.4) Local absolute continuity: $u \in \mathrm{AC}^2_{\mathrm{loc}}(]0, \infty[; M)$.

For $[t_0, T] \Subset\,]0, \infty[$ as above and $N \in N$ we define $\tau_N = (T{-}t_0)/N$ and the partition $t^N_k = t_0 + k\tau_N$ for $k = 0, 1, ..., N$. Now (EVI)$_\lambda$ gives

$$\mathcal{F}(u(t^N_k)) + \frac{1}{2M_\lambda(\tau_N)}\mathcal{D}\big(u(t^N_{k-1}), u(t^N_k)\big)^2 \leq \mathcal{F}(u(t^N_{k-1})).$$

When adding over $k = 1, ..., N$ we can exploit the telescope sum and obtain

$$\frac{\tau_N}{2M_\lambda(\tau_N)}\sum_{k=1}^{N}\tau_N\Big(\frac{1}{\tau_N}\mathcal{D}\big(u(t^N_{k-1}), u(t^N_k)\big)\Big)^2 \leq \mathcal{F}(u(t_0)) - \mathcal{F}(u(T)) =: \Delta.$$

Defining the piecewise constant function S^N via $S^N(t) = \frac{1}{\tau_N}\mathcal{D}\big(u(t^N_{k-1}), u(t^N_k)\big)$ for $t \in\,]t^N_{k-1}, t^N_k]$, we have the L^2 bound $\int_{t_0}^T S^N(t)^2\,\mathrm{d}t \leq \Delta$. Thus, after extracting a subsequence (not relabeled) we may assume $S^N \rightharpoonup S_0$ in $\mathrm{L}^2([t_0, T])$.

For arbitrary $r, s \in [t_0, T]$ with $r < s$ we choose $l(N), m(N) \in \{0, 1, ..., N\}$ such that $\widetilde{r}_N := t^N_{l(N)} \to r$ and $\widetilde{s}_N := t^N_{m(N)} \to s$. Using the triangle inequality we obtain

$$\begin{aligned}\mathcal{D}(u(r), u(s)) &\leq \mathcal{D}(u(r), u(\widetilde{r}_N)) + \Big(\sum_{k=l(N)+1}^{m(N)}\mathcal{D}\big(u(t^N_{k-1}), u(t^N_k)\big)\Big) + \mathcal{D}(u(\widetilde{s}_N), u(s))\\ &\leq C\big|r - \widetilde{r}_N\big|^{1/2} + \int_{t_0}^T \mathbf{1}_{[\widetilde{r}_N,\widetilde{s}_N]}(t)\,S^N(t)\,\mathrm{d}t + C\big|s - \widetilde{s}_N\big|^{1/2},\end{aligned}$$

where we used the Hölder continuity (P.3) and the definition of S^N. We can now pass to the limit $N \to \infty$ on the right-hand side and arrive at $\mathcal{D}(u(r), u(s)) \leq \int_r^s S_0(t)\,\mathrm{d}t$ which implies $u \in \mathrm{AC}^2([t_0, T]; M)$ with $|\dot u|_{\mathcal{D}} \leq S_0 \in \mathrm{L}^2([t_0, T])$ a.e. in $[t_0, T]$, see Theorem 4.4.

Proposition 4.23 (λ-contractivity for solutions of (EVI)$_\lambda$) *For $(M, \mathcal{F}, \mathcal{D})$ consider two (EVI)$_\lambda$ solutions $u, \widetilde{u} : [0, \infty[\to M$. Then, we have*

$$\mathcal{D}\big(u(t), \widetilde{u}(t)\big) \leq \mathrm{e}^{-\lambda(t-s)}\, \mathcal{D}\big(u(s), \widetilde{u}(s)\big) \quad \textit{for } 0 \leq s < t. \tag{4.26}$$

Proof. Step 1: First two applications of EVI. We insert $w = \widetilde{u}(t)$ into the (EVI)$_\lambda$ for u and $\widetilde{w} = u(t)$ into the (EVI)$_\lambda$ for $\widetilde{u}$. Adding the two inequalities we see that all terms involving $\mathcal{F}$ cancel, and we obtain

$$\mathcal{D}\big(u(t), \widetilde{u}(t)\big)^2 = \big(\frac{1}{2} + \frac{1}{2}\big)\mathcal{D}\big(u(t), \widetilde{u}(t)\big)^2 \leq \mathrm{e}^{-\lambda(t-s)}\Big(\frac{1}{2}\mathcal{D}\big(u(s), \widetilde{u}(t)\big)^2 + \frac{1}{2}\mathcal{D}\big(\widetilde{u}(s), u(t)\big)^2\Big).$$

Note that on the right-hand side the four different points $u(s),\ u(t),\ \widetilde{u}(s)$, and $\widetilde{u}(t)$ appear.

Step 2: Third and fourth application of EVI. We again use (EVI)$_\lambda$ for u but now with $w = \widetilde{u}(s)$ and (EVI)$_\lambda$ for $\widetilde{u}$ with $\widetilde{w} = u(s)$. Thus we can estimate the terms on the right-hand side and arrive at

$$\begin{aligned}\mathcal{D}\big(u(t), \widetilde{u}(t)\big)^2 \leq \mathrm{e}^{-\lambda(t-s)}\Big(&\mathrm{e}^{-\lambda(t-s)}\big(\frac{1}{2} + \frac{1}{2}\big)\mathcal{D}\big(u(s), \widetilde{u}(s)\big)^2 \\ &+ M_\tau(t{-}s)\big(\mathcal{F}(u(s)){-}\mathcal{F}(u(t)) + \mathcal{F}(\widetilde{u}(s)) - \mathcal{F}(\widetilde{u}(t))\big)\Big).\end{aligned} \tag{4.27}$$

Step 3: Absolute continuity of $[t_0, T] \ni t \to \delta(t) = \mathcal{D}\big(u(t), \widetilde{u}(t)\big)$. For $r, s \in [t_0, T]$ the triangle inequality gives

$$\begin{aligned}\big|\delta(r) - \delta(s)\big| &= \big|\mathcal{D}\big(u(r), \widetilde{u}(r)\big) - \mathcal{D}\big(u(s), \widetilde{u}(s)\big)\big| \\ &\leq \big|\mathcal{D}\big(u(r), \widetilde{u}(r)\big) - \mathcal{D}\big(u(s), \widetilde{u}(r)\big)\big| + \big|\mathcal{D}\big(u(s), \widetilde{u}(r)\big) - \mathcal{D}\big(u(s), \widetilde{u}(s)\big)\big| \\ &\leq \mathcal{D}\big(u(r), u(s)\big) + \mathcal{D}\big(\widetilde{u}(r), \widetilde{u}(s)\big) \leq \int_r^s \big(|\dot u|_{\mathcal{D}}(t) + |\dot{\widetilde{u}}|_{\mathcal{D}}(t)\big)\,\mathrm{d}t.\end{aligned}$$

Hence, $u, \widetilde{u} \in \mathrm{AC}^2([t_0, T]; M)$ implies $\delta \in \mathrm{AC}^2([t_0, T]; \mathbb{R}) = \mathrm{W}^{1,2}([t_0, T])$.

Step 4: Conclusion. We set $\rho(t) = \mathrm{e}^{2\lambda t}\delta(t)^2$, then the product rule and Step 3 give $\rho \in \mathrm{W}^{1,2}([t_0, T])$. Moreover, by the definition of ρ, the estimate (4.27) turns into

$$\rho(t) - \rho(s) \leq \mathrm{e}^{\lambda(t+s)} M_\lambda(t{-}s)\big(\mathcal{F}(u(s)){-}\mathcal{F}(u(t)) + \mathcal{F}(\widetilde{u}(s)) - \mathcal{F}(\widetilde{u}(t))\big),$$

for $t_0 \leq s < t \leq T$.

Now assume that $s = s_*$ is a point of differentiability of ρ, which is true on a set of full measure. Then, dividing by $t - s_* > 0$ and taking the limit $t \to s_*^+$ gives

$$\begin{aligned}\dot\rho(s_*) &= \lim_{t \to s_*^+} \frac{\rho(t) - \rho(s_*)}{t - s_*} \\ &\leq \limsup_{t \to s_*^+} \Big(B_\lambda(t, s_*)\,\big(\mathcal{F}(u(s_*)){-}\mathcal{F}(u(t)) + \mathcal{F}(\widetilde{u}(s_*)) - \mathcal{F}(\widetilde{u}(t))\big)\Big)\end{aligned}$$

with $B_\lambda(t,s_*) = \mathrm{e}^{\lambda(t+s_*)} M_\lambda(t-s_*)/(t-s_*) \to \mathrm{e}^{2\lambda s_*}$ for $t \to s_*$.

Using the Hölder continuity (P.3) and lsc of $\mathcal{F}$ we have

$$\limsup_{t\to s_*} \big(\mathcal{F}(u(s_*)) - \mathcal{F}(u(t))\big) = \mathcal{F}(u(s_*)) - \liminf_{t\to s_*} \mathcal{F}(u(t)) \le \mathcal{F}(u(s_*)) - \mathcal{F}(u(s_*)) = 0,$$

and similarly for $\widetilde{u}$. Hence, we conclude $\dot\rho(s_*) \le 0$. Because ρ is absolutely continuous, we have the monotonicity $\rho(t) \le \rho(s)$ for $s < t$, which is the desired estimate (4.26) when recalling the definition $\rho(t) = \mathrm{e}^{2\lambda t}\mathcal{D}\big(u(t),\widetilde{u}(t)\big)^2$. ■

5 Evolutionary Γ-convergence for gradient systems

In this section we study families of gradient systems $(X,\mathcal{F}_\varepsilon,\mathcal{R}_\varepsilon)$ or $(M,\mathcal{F}_\varepsilon,\mathcal{D}_\varepsilon,\psi_\varepsilon)$ where $\varepsilon \in [0,1]$. The typical question one is interested are the following:

Q1 Assume we have solutions $u_\varepsilon : [0,T] \to X$ for $(X,\mathcal{F}_\varepsilon,\mathcal{R}_\varepsilon)$ with $u_\varepsilon(0) \rightsquigarrow u^0$. Is it possible to find a subsequence (not relabeled) and a limit function $u : [0,T] \to X$ such that $u_\varepsilon(t) \rightsquigarrow u_0(t)$ for all $t \in [0,T]$.

Q2 Is their a notion of convergence for the energies $\mathcal{F}_\varepsilon \overset{\mathrm{energ}}{\rightsquigarrow} \mathcal{F}_0$ and for dissipation potentials $\mathcal{R}_\varepsilon \overset{\text{diss}}{\rightsquigarrow} \mathcal{R}_0$ such that u_0 is a solution of the **effective gradient system** $(X,\mathcal{F}_0,\mathcal{R}_0)$.

Q3 There are cases, where limits $\mathcal{F}_0$ and $\mathcal{R}_0$ as in Q2 exists, but they produce the wrong solutions! Is there a direct way to construct the correct effective GS $(X,\mathcal{F}_{\text{eff}},\mathcal{R}_{\text{eff}})$ is the sense that $(\mathcal{F}_\varepsilon,\mathcal{R}_\varepsilon) \overset{\text{GS}}{\rightsquigarrow} (\mathcal{F}_{\text{eff}},\mathcal{R}_{\text{eff}})$.

In light of our examples in Section 1 question Q3 cannot be answered by studying the solutions u_ε of the gradient-flow equations $0 \in \partial\mathcal{R}_\varepsilon(u,\dot u) + \partial^{\mathrm{F}}\mathcal{F}_\varepsilon(u)$ and then showing that the limits u_0 of sequences u_ε solve the *effective* evolution equation $\dot u = \boldsymbol{V}_{\text{eff}}(u)$. Of course, it is always a major achievement to find the effective evolution equation, but it does not answer the question whether the effective equation has a gradient structure. Moreover, if it has a gradient structure it may have many of them. Hence, it is of independent interest, in particular in the sense of physical modeling, to show how the gradient structure passes to the limit.

Of course, we are not so interested to study the case of “continuous dependence on parameters” as is studied in the theory of ODEs. If $V : [0,1]\times[0,T]\times\mathbb{R}^n \to \mathbb{R}^n$ is continuous and globally Lipschitz in $u \in \mathbb{R}^n$., then the unique solution $u_\varepsilon : [0,T] \to \mathbb{R}^n$ of $\dot u_\varepsilon(t) = V(\varepsilon,t,u_\varepsilon(t))$ depends continuously on $\varepsilon \in [0,1]$ and $t \in [0,T]$. If we follow this approach in the setting of classical gradient systems $(M,\mathcal{F}_\varepsilon,\mathbb{G}_\varepsilon)$ on a finite-dimensional manifold M, then we need assumptions on the energy $\mathcal{F} : [0,1] \to M \to \mathbb{R}$ as well as on the Riemannian tensor $\mathbb{G}_\varepsilon(u) : \mathrm{T}_u M \to \mathrm{T}^*_u M$ may depend on ε. The gradient-flow equation reads

$$\dot u = -\mathbb{G}_\varepsilon(u)\mathrm{D}\mathcal{F}_\varepsilon(u) =: V(\varepsilon,u).$$

Thus, to apply the above-mentioned continuous dependence result for ODEs, we need $\mathbb{G} \in \mathrm{C}^0\big([0,1];\, \mathrm{C}^{\mathrm{Lip}}(M;F_2(M))\big)$ and $\mathcal{F} \in \mathrm{C}^0\big([0,1];\mathrm{C}^{1,\mathrm{Lip}}(M)\big)$.

Such results are not relevant for PDEs because the vector fields are not not smooth and only defined on dense subsets. There the question of “singular limits” is studied (cf. [FeN09]), for instance PDEs of the form

$$\dot u_\varepsilon = \operatorname{div}\big(A(\tfrac{1}{\varepsilon}x)\nabla u\big) - b(\tfrac{1}{\varepsilon}x)u_\varepsilon, \quad x\in\Omega, \quad u_\varepsilon|_{\partial\Omega}=0.$$
$$\lambda_\varepsilon \dot w_\varepsilon = \varepsilon\partial_x^2(w_\varepsilon) + \frac{1}{\varepsilon}\big(w_\varepsilon - w_\varepsilon^3\big), \quad x\in\Omega, \quad w_\varepsilon|_{\partial\Omega}=1.$$

We refer to [SaS04, Ser11, Bra14, MMP21, Mie16, MuS22] for general approaches in evolutionary Γ-convergence.

5.1 Γ-convergence for (static) functionals

To study limits of functionals we define a notion of convergence in the spirit of question Q3 above, but now in the static case. If the “problem” associated with a GS $(X,\mathcal{F},\mathcal{R})$ is the solution of the gradient-flow equation, then the “problem” associated with a static functional $\mathcal{J}$ is to find its minimizer. Of course, we have seen that these problems are strongly linked by the time-incremental minimization sometimes also called minimizing movement scheme. Thus, for a family $(\mathcal{J}_\varepsilon)_{\varepsilon>0}$ of functionals $\mathcal{J}_\varepsilon : M\to\mathbb{R}_\infty$, we ask the (static) question:

Question: What is a good notion of convergence $\mathcal{J}_\varepsilon \rightsquigarrow \mathcal{J}_0$ such that any limit u_0 of (a subsequence of) minimizers u_ε of $\mathcal{J}_\varepsilon$ is automatically a minimizer of $\mathcal{J}_0$.

Again, we are not so much interested in the case $\mathcal{J}_\varepsilon \to \mathcal{J}_0$ in $\mathrm{C}^1_{\mathrm{loc}}(X)$, which is of course sufficient to show convergence in the associated Euler-Lagrange equations.

We consider a complete metric space $(M,\mathcal{D})$ and functionals $\mathcal{J}_\varepsilon : X\to\mathbb{R}_\infty$. In a metric space “$u_k \to u$” will always denote convergence in the metric; if M is a Banach space X then $u_k\to u$ and $v_k \rightharpoonup v$ denote strong and weak convergence, respectively. We first introduce more classical notions of convergence of functionals, namely the *pointwise convergence* $\mathcal{J}_\varepsilon \xrightarrow{\text{pw}} \mathcal{J}_0$ and *continuous convergence* (also weak in Banach spaces) defined via

$$\mathcal{J}_\varepsilon \xrightarrow{\text{pw}} \mathcal{J}_0, \quad \text{if } \mathcal{J}_\varepsilon(u)\to\mathcal{J}_0(u) \text{ for all } u\in M; \tag{5.1a}$$
$$\mathcal{J}_\varepsilon \xrightarrow{\text{cc}} \mathcal{J}_0, \quad \text{if } u_\varepsilon\to u \implies \mathcal{J}_\varepsilon(u_\varepsilon)\to\mathcal{J}_0(u). \tag{5.1b}$$

In the context of minimization of functionals, the concept of Γ*-convergence* is more natural, see Theorem 5.6. This convergence was originally called *variational convergence* or *epi-graph convergence* (cf. [DeF75, DeG77, Att84]), but nowadays the term Γ-convergence is more common and we refer to [Dal93, Bra02, Bra06, Bra14] for further details.

Definition 5.1 (Γ and Mosco convergence) *Let $(M,\mathcal{D})$ be a complete metric space. We say that $\mathcal{J}_\varepsilon$* Γ-converges to $\mathcal{J}_0$ *and write $\mathcal{J}_\varepsilon \xrightarrow{\Gamma} \mathcal{J}_0$ or $\mathcal{J}_0 = \Gamma\text{-}\lim_{\varepsilon\to0}\mathcal{J}_\varepsilon$, if $(\Gamma.\mathrm{inf})$ and $(\Gamma.\mathrm{sup})$ hold:*

$(\Gamma.\mathrm{inf})$ $\quad u_\varepsilon\to u \implies \mathcal{J}_0(u)\le \liminf_{\varepsilon\to0}\mathcal{J}_\varepsilon(u_\varepsilon)$ *(liminf estimate)*

$(\Gamma.\mathrm{sup})$ $\quad \forall\,\widehat u\ \exists\,(\widehat u_\varepsilon)_\varepsilon\colon \widehat u_\varepsilon\to\widehat u$ *and* $\mathcal{J}_0(\widehat u) = \limsup_{\varepsilon\to0}\mathcal{J}_\varepsilon(\widehat u_\varepsilon)$ *(limsup estimate)*

If $(M,\mathcal{D})$ is a Banach space $(X;\|\cdot\|)$ we say that $\mathcal{J}_\varepsilon$ (sequentially) weakly Γ-converges to $\mathcal{J}_0$

and write $\mathcal{J}_\varepsilon \xrightharpoonup{\Gamma} \mathcal{J}_0$ *and* $\mathcal{J}_0 = \Gamma_{\mathrm{w}}\text{-}\lim_{\varepsilon\to 0} \mathcal{J}_\varepsilon$*, if* $(\Gamma.\mathrm{inf})$ *and* $(\Gamma.\mathrm{sup})$ *when "$\to$" is replaced by "$\rightharpoonup$". If* $\mathcal{J}_\varepsilon \xrightarrow{\Gamma} \mathcal{J}_0$ *and* $\mathcal{J}_\varepsilon \xrightharpoonup{\Gamma} \mathcal{J}_0$ *hold, then we say that* $\mathcal{J}_\varepsilon$ Mosco-converges to $\mathcal{J}_0$ *and write* $\mathcal{J}_\varepsilon \xrightarrow{\mathsf{M}} \mathcal{J}_0$ *or* $\mathcal{J}_0 = \mathrm{M}\text{-}\lim_{\varepsilon\to 0} \mathcal{J}_\varepsilon$*. In this case, it suffices to show* $(\Gamma.\mathrm{inf})$ *for weak convergence and* $(\Gamma.\mathrm{sup})$ *for strong convergence.*

We will see in Lemma 5.3 that there are simple quadratic functionals for which weak and strong Γ-limits exist, but they are different.

The conditions $(\Gamma.\mathrm{sup})$ is often replaced by the so-called *existence of recovery sequences*:

$(\Gamma.\mathrm{rec})$ $\quad \forall\, \widehat u\; \exists\, (\widehat u_\varepsilon)_\varepsilon\colon \widehat u_\varepsilon \to \widehat u$ and $\mathcal{J}_\varepsilon(\widehat u_\varepsilon) \to \mathcal{J}_0(\widehat u)$. (recovery sequence)

Of course, $(\Gamma.\mathrm{rec})$ implies $(\Gamma.\mathrm{sup})$. Moreover, assuming that $(\Gamma.\mathrm{inf})$ holds, $(\Gamma.\mathrm{rec})$ follows from $(\Gamma.\mathrm{sup})$. The sequence $(\widehat u_\varepsilon)_\varepsilon$ is called recovery sequence as it recovers the correct energy $\mathcal{J}_0(\widehat u)$. Moreover, one sees in several examples that $\widehat u_\varepsilon$ has to "recover" the correct microscopic structure which makes the energy $\mathcal{J}_\varepsilon(\widehat u_\varepsilon)$ small enough to reach (recover) the lowest possible value for $\mathcal{J}_0(\widehat u)$.

We emphasize that the definition of Γ-convergence is asymmetric and fits to "minimization". For "liminf" we impose a condition for *all* sequences, while for "limsup" we only need *one* sequence. This way we lose the linearity for Γ-convergence. If $\Gamma\text{-}\lim \mathcal{J}_\varepsilon$, $\Gamma\text{-}\lim \mathcal{G}_\varepsilon$, and $\Gamma\text{-}\lim(\mathcal{G}_\varepsilon{+}\mathcal{J}_\varepsilon)$ exist we do *not* have $\Gamma\text{-}\lim(\mathcal{G}_\varepsilon{+}\mathcal{J}_\varepsilon) = \big(\Gamma\text{-}\lim \mathcal{G}_\varepsilon\big) + \big(\Gamma\text{-}\lim \mathcal{J}_\varepsilon\big)$ in general.

Example 5.2 *(A) Consider* $X = \mathbb{R}^1$ *and* $\mathcal{J}_\varepsilon(u) = \frac12 u^2 - \cos(u/\varepsilon)$*. We claim*

$$\mathcal{J}_\varepsilon \xrightarrow{\Gamma} \mathcal{J}_0 \quad \text{with } \mathcal{J}_0(u) = \frac12 u^2 - 1.$$

To show $(\Gamma.\mathrm{inf})$ *we use* $\cos\alpha \le 1$ *and obtain* $\mathcal{J}_\varepsilon(u) \ge \mathcal{J}_0(u)$ *for all* u*. As* $\mathcal{J}_0$ *is continuous, the result follows. To show* $(\Gamma.\mathrm{sup})$ *we start from an arbitrary* $\widehat u \in \mathbb{R}$ *and look for a close-by* $\widehat u_\varepsilon$ *such that* $\mathcal{J}_\varepsilon(\widehat u_\varepsilon)$ *is close to* $\mathcal{J}_0(\widehat u)$*. This means that we want to have* $\cos(\widehat u_\varepsilon)$ *close to* 1*. Thus, we choose* $\widehat u_\varepsilon = 2\pi\varepsilon\lfloor \widehat u/(2\pi\varepsilon)\rfloor$*, where the floor function* $\lfloor\cdot\rfloor$ *rounds down to the nearest integer. Obviously, we have* $\widehat u_\varepsilon \to \widehat u$ *and* $\mathcal{J}_\varepsilon(\widehat u_\varepsilon) = \mathcal{J}_0(\widehat u_\varepsilon) \to \mathcal{J}_0(\widehat u)$ *as desired.*

(B) For an arbitrary $\lambda \in \mathbb{R}$*, we set* $\mathcal{G}_\varepsilon = \lambda \mathcal{J}_\varepsilon$*. With an analogous argument we obtain*

$$\mathcal{G}_\varepsilon \xrightarrow{\Gamma} \mathcal{G}_0 \quad \text{with } \mathcal{G}_0(0) = \frac{\lambda}{2} u^2 - |\lambda|$$

In the case $\lambda < 0$ *one chooses* $\widehat u_\varepsilon = \pi\varepsilon\big(2\lfloor \widehat u/(2\pi\varepsilon)\rfloor + 1\big)$ *to find* $\cos(\widehat u_\varepsilon/\varepsilon) = -1$*.*

(C) We see that linearity is destroyed, in particular we have

$$0 = \Gamma\text{-}\lim_{\varepsilon\to 0}(\mathcal{J}_\varepsilon - \mathcal{J}_\varepsilon) \neq \big(\Gamma\text{-}\lim_{\varepsilon\to 0} \mathcal{J}_\varepsilon\big) + \Gamma\text{-}\lim_{\varepsilon\to 0}\big({-}\mathcal{J}_\varepsilon\big) = -1 + (-1) = -2.$$

The following lemma presents a simple quadratic example in which the weak and the strong Γ-limits exist but they are different. We define

$$\mathcal{F}_\varepsilon(w) = \int_\Omega \frac12 w(x)\cdot \mathbb{A}\big(\frac1\varepsilon x\big) w(x)\,\mathrm{d}x \quad \text{for } w \in X = \mathrm{L}^2(\Omega;\mathbb{R}^m),$$

where $\Omega \subset \mathbb{R}^d$ is a bounded Lipschitz domain and $\mathbb{A} \in \mathrm{L}^\infty(\mathbb{R}^d;\mathbb{R}^{m\times m}_{\mathsf{sym}})$ is 1-periodic, i.e. $\mathbb{A}(y{+}n) = \mathbb{A}(y)$ for all $y \in \mathbb{R}^d$ and all $n \in \mathbb{Z}^d$. Moreover, we assume that $\mathbb{A}$ is uniformly positive definite, i.e. $\underline{a}|w|^2 \le w\cdot\mathbb{A}(y)w \le \overline{a}|w|^2$ for $\overline{a} > \underline{a} > 0$. The main tools is the Riemann lemma stating that the sequence $A_\varepsilon : x \mapsto \mathbb{A}(\frac1\varepsilon x)$ satisfies $A_\varepsilon \overset{*}{\rightharpoonup} \mathbb{A}_{\mathsf{arith}}$, see [Dal93, Exa. 6.6] for more general results of this type.

Lemma 5.3 *Define the arithmetic and harmonic mean of* $\mathbb{A}$ *via*

$$\mathbb{A}_{arith} := \int_{[0,1]^d} \mathbb{A}(y)\,\mathrm{d}y \quad \text{and} \quad \mathbb{A}_{harm} := \Big(\int_{[0,1]^d} \mathbb{A}(y)^{-1}\,\mathrm{d}y\Big)^{-1}$$

and the two functionals

$$\mathcal{F}_{arith}(w) = \int_\Omega \frac{1}{2} w(x)\cdot \mathbb{A}_{arith} w(x)\,\mathrm{d}x \ \text{ and } \ \mathcal{F}_{harm}(w) = \int_\Omega \frac{1}{2} w(x)\cdot \mathbb{A}_{harm} w(x)\,\mathrm{d}x.$$

In $X = \mathrm{L}^2(\Omega;\mathbb{R}^m)$ *we have* $\mathcal{F}_\varepsilon \xrightharpoonup{\Gamma} \mathcal{F}_{harm}$ *and* $\mathcal{F}_\varepsilon \xrightarrow{\mathrm{cc}} \mathcal{F}_{arith}$*, which implies* $\mathcal{F}_\varepsilon \xrightarrow{\Gamma} \mathcal{F}_{arith}$*.*

Proof. We first prove $\mathcal{F}_\varepsilon \xrightharpoonup{\Gamma} \mathcal{F}_{\mathrm{harm}}$. For the liminf estimate assume $w_\varepsilon \rightharpoonup w$ in $\mathrm{L}^2(\Omega)$. Writing $\mathbb{A}_\varepsilon(x) = \mathbb{A}(\frac{1}{\varepsilon}x)$ we have

$$\begin{aligned} \mathcal{F}_\varepsilon(w_\varepsilon) &= \frac{1}{2}\int_\Omega w_\varepsilon \cdot \mathbb{A}_\varepsilon w_\varepsilon \,\mathrm{d}x = &(5.2)\\ &= \frac{1}{2}\int_\Omega \underbrace{(w_\varepsilon - \mathbb{A}_\varepsilon^{-1}\mathbb{A}_{\mathrm{harm}} w)\cdot \mathbb{A}_\varepsilon (w_\varepsilon - \mathbb{A}_\varepsilon^{-1}\mathbb{A}_{\mathrm{harm}} w)}_{\geq 0} \\ &\qquad + 2\underbrace{w_\varepsilon}_{\rightharpoonup w}\cdot \mathbb{A}_{\mathrm{harm}} w - \mathbb{A}_{\mathrm{harm}} w \cdot \underbrace{\mathbb{A}_\varepsilon^{-1}}_{\stackrel{*}{\rightharpoonup} \mathbb{A}_{\mathrm{harm}}^{-1}} \mathbb{A}_{\mathrm{harm}} w \,\mathrm{d}x. &(5.3) \end{aligned}$$

Dropping the nonnegative term, the limit $\varepsilon \to 0$ leads to the desired lower estimate $\liminf_\varepsilon \mathcal{F}_\varepsilon(w_\varepsilon) \geq \frac{1}{2}\int_\Omega 0 + 2w\cdot\mathbb{A}_{\mathrm{harm}} w - w\cdot \mathbb{A}_{\mathrm{harm}} w \,\mathrm{d}x = \mathcal{F}_0(w)$.

For the limsup-estimate we use the same reformulation of $\mathcal{F}_\varepsilon$ as in (5.2). For a given $\widehat{w}$ we choose $\widehat{w}_\varepsilon = \mathbb{A}_\varepsilon^{-1}\mathbb{A}_{\mathrm{harm}}\widehat{w}$. Since by construction the first term in the integral is 0 we find $\mathcal{F}_\varepsilon(\widehat{w}_\varepsilon) = \frac{1}{2}\int_\Omega 0 + 2\mathbb{A}_\varepsilon^{-1}\mathbb{A}_{\mathrm{harm}}\widehat{w}\cdot\mathbb{A}_{\mathrm{harm}}\widehat{w} - \mathbb{A}_{\mathrm{harm}}\widehat{w}\cdot\mathbb{A}_\varepsilon^{-1}\mathbb{A}_{\mathrm{harm}}\widehat{w}\,\mathrm{d}x \to \mathcal{F}_{\mathrm{harm}}(\widehat{w})$.

For strong continuous convergence take any $w_\varepsilon \to w$ in $\mathrm{L}^2(\Omega)$ and write

$$\begin{aligned} \mathcal{F}_\varepsilon(w_\varepsilon) &= \frac{1}{2}\int_\Omega w\cdot \underbrace{\mathbb{A}_\varepsilon w}_{\rightharpoonup \mathbb{A}_{\mathrm{arith}} w} - 2w\cdot\mathbb{A}_\varepsilon \underbrace{(w - w_\varepsilon)}_{\to 0} + \underbrace{(w - w_\varepsilon)}_{\to 0}\cdot \mathbb{A}_\varepsilon (w - w_\varepsilon)\,\mathrm{d}x &(5.4)\\ &\to \mathcal{F}_{\mathrm{arith}}(w). &(5.5) \end{aligned}$$

This proves the strong continuous and hence the strong Γ-convergence. ■

Clearly, continuous convergence is much stronger than Γ-convergence. We have the following relations.

Lemma 5.4 (Properties of Γ-limits) *On the complete metric space* $(M,\mathcal{D})$ *consider the functionals* $\mathcal{J}_\varepsilon,\ \mathcal{K}_\varepsilon : M \to \mathbb{R}_\infty$.

(a) $\mathcal{J}_\varepsilon \xrightarrow{\Gamma} \mathcal{J}_0 \implies \mathcal{J}_0 : M \to \mathbb{R}_\infty$ *is lsc.*

(b) $\mathcal{J}_\varepsilon \xrightarrow{\Gamma} \mathcal{J}_0$ *and* $\mathcal{K}_\varepsilon \xrightarrow{\mathrm{cc}} \mathcal{K}_0 \implies \mathcal{J}_\varepsilon + \mathcal{K}_\varepsilon \xrightarrow{\Gamma} \mathcal{J}_0 + \mathcal{K}_0$

(c) $\mathcal{J}_\varepsilon \xrightarrow{\mathrm{cc}} \mathcal{J}_0 \implies \mathcal{J}_\varepsilon \xrightarrow{\Gamma} \mathcal{J}_0$

Proof. *Part (a):* We use an argument that is standardly used for constructing recovery sequences. For $u_n \to u$ we have to show $\mathcal{J}_0(u) \leq \liminf_{n\to\infty} \mathcal{J}_0(u_n)$. As $\mathcal{J}_0 = \Gamma\text{-}\lim \mathcal{J}_\varepsilon$ we find $(\widehat{u}^n_\varepsilon)_\varepsilon$ with $\widehat{u}^n_\varepsilon \to u_n$ and $\mathcal{J}_\varepsilon(\widehat{u}^n_\varepsilon) \to \mathcal{J}_0(u_n)$. Thus for each n we can find $\varepsilon_n > 0$ such that

$$\varepsilon_n \in]0, 1/n[, \quad \mathcal{D}(\widehat{u}^n_{\varepsilon_n}, u_n) \leq 1/n, \quad \mathcal{J}_{\varepsilon_n}(\widehat{u}^n_{\varepsilon_n}) \leq \mathcal{J}_0(u_n) + 1/n.$$

Setting $\widetilde{u}_{\varepsilon_n} := \widehat{u}^n_{\varepsilon_n}$ we have $\varepsilon_n \to 0$ and $\mathcal{D}(\widetilde{u}_{\varepsilon_n}, u) \leq \mathcal{D}(\widetilde{u}_{\varepsilon_n}, u_n) + \mathcal{D}(u_n, u) \to 0$. Setting $\widetilde{u}_\varepsilon = u$ for $\varepsilon \notin \{ \varepsilon_n \mid n \in \mathbb{N} \}$, we have $\widetilde{u}_\varepsilon \to u$ and obtain

$$\liminf_{n\to\infty} \mathcal{J}_0(u_n) \geq \liminf_{n\to\infty} \mathcal{J}_{\varepsilon_n}(\widetilde{u}_{\varepsilon_n}) \geq \liminf_{\varepsilon\to 0^+} \mathcal{J}_\varepsilon(\widetilde{u}_\varepsilon) \geq \mathcal{J}_0(u),$$

where the last estimate follows from $(\Gamma.\inf)$ and $\widetilde{u}_\varepsilon \to u$.

Part (b): This follows easily as convergent sequences can be chosen as needed for $\mathcal{J}_\varepsilon$, and continuous convergence for $\mathcal{K}_\varepsilon$ gives the result.

Part (c): This is trivial because the liminf estimate is a limit. As recovery sequence one can take any convergent sequence, e.g. the constant sequence with $\widehat{u}_\varepsilon = \widehat{u}$. ■

The following properties of sequences of functionals will be useful in the formulation of the following results. Recall the sublevels $S^{\mathcal{F}}_E := \{ u \in M \mid \mathcal{F}(u) \leq E \}$.

Definition 5.5 (Uniform properties) *On a complete metric space $(M, \mathcal{D})$ consider a family $(\mathcal{J}_\varepsilon)_\varepsilon$ of functionals $\mathcal{J}_\varepsilon : M \to \mathbb{R}_\infty$.*

(i) The family is called equi-coercive*, if*

$$\forall\, E \in \mathbb{R}\; \exists\, R > 0,\; u_* \in M\; \forall\, \varepsilon : \quad S^{\mathcal{J}_\varepsilon}_E \subset B_R(u_*).$$

(ii) The family is called equi-compact*, if (where "$\Subset$" means compactly contained)*

$$\forall\, E \in \mathbb{R}\; \exists\, K \Subset M\; \forall\, \varepsilon : \quad S^{\mathcal{J}_\varepsilon}_E \subset K.$$

(iii) If $(M, \mathcal{D})$ is a Banach space $(X, \|\cdot\|)$, we call the family equi-superlinear*, if there exists a superlinear function $\varphi : [0, \infty[\to \mathbb{R}$ such that*

$$\forall\, u \in X\; \forall\, \varepsilon : \quad \mathcal{J}_\varepsilon(u) \geq \varphi\big(\|u\|\big).$$

Warning: In many papers and textbooks our notion of "equi-compactness" is simply called "equi-coercivity". We distinguish these two concepts, which is quite useful for gradient systems where different functionals like $\mathcal{F}_\varepsilon$ and $\mathcal{R}_\varepsilon$ or $\mathcal{D}_\varepsilon$ are considered on the same space. Moreover, it allows us to avoid switching between weak and strong topologies in Banach spaces, where equi-coercivity implies weak equi-compactness.

The origin for the definition of Γ-convergence, which is clearer in the original name "variational convergence", is the following convergence of minimizers, see [Dal93, Bra02].

Theorem 5.6 (Convergence of minimizers) *In a complete metric space $(M, \mathcal{D})$ assume $\mathcal{J}_\varepsilon \xrightarrow{\Gamma} \mathcal{J}_0$ with $\inf \mathcal{J}_0 =: \alpha_0 \in \mathbb{R}$.*

(a) If $u_\varepsilon \to u_0$ and $\liminf_{\varepsilon\to 0^+} \mathcal{J}_\varepsilon(u_\varepsilon) = \alpha_0$, then u_0 is a minimizer of $\mathcal{J}_0$.

(b) If the family $(\mathcal{J}_\varepsilon)_\varepsilon$ is equi-compact, then $\alpha_\varepsilon = \inf \mathcal{J}_\varepsilon$ satisfies $\alpha_\varepsilon \to \alpha_0$. Moreover, every sequence $(u_\varepsilon)_{\varepsilon>0}$ with $\mathcal{J}_\varepsilon(u_\varepsilon) \to \alpha_0$ has a convergent subsequence $u_{\varepsilon_k} \to u_0$ and each such limit u_0 is a minimizer of $\mathcal{J}_0$. In particular, if (u_{ε_k}) is a sequence of minimizers for $\mathcal{J}_{\varepsilon_k}$, then all accumulation points u_0 of this sequence are minimizers of $\mathcal{J}_0$.

Proof. Part (a). By the $(\Gamma.\inf)$ we have $\mathcal{J}_0(u_0) \leq \liminf_{\varepsilon\to 0} \mathcal{J}_\varepsilon(u_\varepsilon) = \alpha_0$. However, with $\alpha_0 = \inf \mathcal{J}_0 \leq \mathcal{J}_0(u_0)$ we conclude $\alpha_0 = \mathcal{J}_0(u_0)$, i.e. u_0 is a minimizer.

Part (b). By Lemma 5.4 we know that $\mathcal{J}_0$ is lsc and the equi-compactness implies that the sublevels of $\mathcal{J}_0$ are compact. Hence $\mathcal{J}_0$ has a minimizer u_0 with $\mathcal{J}_0(u_0) = \alpha_0$.

By $(\Gamma.\sup)$ there exists a recovery sequence $\widehat{u}_\varepsilon \to u_0$ with $\mathcal{J}_\varepsilon(\widehat{u}_\varepsilon) \to \mathcal{J}_0(u_0) = \alpha_0$. Using $\alpha_\varepsilon = \inf \mathcal{J}_\varepsilon \leq \mathcal{J}_\varepsilon(\widehat{u}_\varepsilon)$ we find $\limsup_{\varepsilon\to 0} \alpha_\varepsilon \leq \alpha_0$. Moreover, we can choose a subsequence $\varepsilon_k \to 0$ such that $\liminf_{\varepsilon\to 0} \alpha_\varepsilon = \lim_{k\to\infty} \alpha_{\varepsilon_k}$. In addition, there exist u_k with $\mathcal{J}_{\varepsilon_k}(u_k) \leq \alpha_{\varepsilon_k} + \varepsilon_k$, and the equi-compactness guarantees the existence of a convergent subsequence $u_{k(l)} \to u_*$ for $l \to \infty$. Now $(\Gamma.\inf)$ implies

$$\alpha_0 \leq \mathcal{J}_0(u_*) \leq \liminf_{l\to\infty} \mathcal{J}_{\varepsilon_{k(l)}}(u_{k(l)}) \leq \liminf_{l\to\infty}(\alpha_{\varepsilon_{k(l)}} + \varepsilon_{k(l)}) = \liminf_{\varepsilon\to 0} \alpha_\varepsilon \leq \limsup_{\varepsilon\to 0} \alpha_\varepsilon \leq \alpha_0.$$

Hence, $\alpha_\varepsilon \to \alpha_0$ is established.

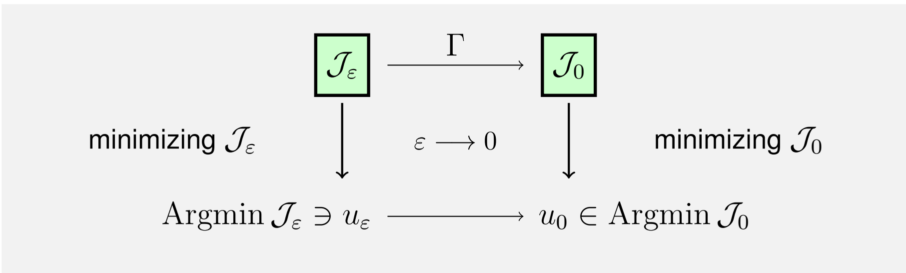
If a sequence u_ε satisfies $\mathcal{J}_\varepsilon(u_\varepsilon) \to \alpha_0 \in \mathbb{R}$, it lies in a compact set, because of equicompactness. By (a) all accumulation points are minimizers.

The last statement is a consequence of the previous assertion and the convergence $\alpha_\varepsilon \to \alpha_0$. ■

Example 5.7 *For $(M, \mathcal{D}) = (\mathbb{R}, \mathcal{D}_{\mathrm{Eucl}})$ the sequence $\mathcal{J}_\varepsilon(u) = \cos(\varepsilon u)$ satisfies $\mathcal{J}_\varepsilon \xrightarrow{\Gamma} \mathcal{J}_0$ with $\mathcal{J}_0 \equiv 1$. Indeed, for u in the compact interval $[-R, R]$ we have $|\varepsilon u| \leq \varepsilon R$ and find $0 \leq 1 - \mathcal{J}_\varepsilon(u) \leq \frac{1}{2}\varepsilon^2 R^2$, which gives uniform convergence on compact sets to $\mathcal{J}_0 : u \mapsto 1$.*

However, for all $\varepsilon > 0$ we have $\alpha_\varepsilon = \inf_{\mathbb{R}} \mathcal{J}_\varepsilon = -1$, whereas for $\varepsilon = 0$ we have $\alpha_0 \inf_{\mathbb{R}} \mathcal{J}_0 = 1$.

The main result of the above theorem is that solving a minimization problem for $\mathcal{J}_\varepsilon$ can be interchanged with passing to the limit $\varepsilon \to 0$. This can be depicted by the following commuting diagram:



We make this more explicit in the Banach space setting by considering an equi-superlinear family $\mathcal{F}_\varepsilon$ with $\mathcal{F}_\varepsilon \xrightarrow{\Gamma} \mathcal{F}_0$. Then, for all $\ell \in X^*$ we set $\mathcal{J}_\varepsilon^\ell = \mathcal{F}_\varepsilon - \langle \ell, \cdot \rangle$ and observe $\mathcal{J}_\varepsilon^\ell \xrightarrow{\Gamma} \mathcal{J}_0^\ell$, cf. Lemma 5.4(b). We define the LimSup for a family $\big(A_\varepsilon\big)_\varepsilon$ of sets $A_\varepsilon \subset X$ via

$$\operatorname*{LimSup}_{\varepsilon\to 0} A_\varepsilon := \Big\{\, u \in X \;\Big|\; \exists\, (\varepsilon_k, u_k)_{k\in\mathbb{N}} : \;\; 0 < \varepsilon_k \to 0,\; u_k \in A_{\varepsilon_k},\; u_k \to u \,\Big\}$$

Thus, the theory of Γ-convergence leads to the following result on the upper semicontinuity of minimizers.

Corollary 5.8 (Upper semicontinuity of the sets of minimizers) *If $\big(\mathcal{F}_\varepsilon\big)_\varepsilon$ is equi-superlinear on a Banach space X. Then, we have*

$$\mathcal{F}_\varepsilon \xrightarrow{\Gamma} \mathcal{F}_0 \quad \Longrightarrow \quad \forall\, \ell \in X^* : \quad \underset{\varepsilon\to 0}{\mathrm{LimSup}}\, \mathrm{Argmin}\big(\mathcal{F}_\varepsilon{-}\ell\big) \,\subset\, \mathrm{Argmin}\big(\mathcal{F}_0{-}\ell\big).$$

The following useful result seems to be folklore, but it is not easy to locate a specific reference. Hence, we give a full proof.

Proposition 5.9 (Γ-convergence versus Mosco convergence) *Assume that X and $\boldsymbol{Z}$ are reflexive Banach spaces such that $\boldsymbol{Z}$ is compactly embedded in X, written $\boldsymbol{Z} \Subset X$. Moreover, assume that the functionals $\mathcal{J}_\varepsilon$ are equi-coercive in $\boldsymbol{Z}$, i.e.*

$$\forall\, J > 0\, \exists\, R > 0\, \forall\, \varepsilon > 0,\; u \in X: \quad \mathcal{J}_\varepsilon(u) \le J \;\Rightarrow\; \|u\|_{\boldsymbol{Z}} \le R: \tag{5.6}$$

Then, $\mathcal{J}_\varepsilon \xrightarrow{\mathsf{M}} \mathcal{J}_0$ in X is equivalent to $\mathcal{J}_\varepsilon \xrightarrow{\Gamma} \mathcal{J}_0$ in $\boldsymbol{Z}$.

Proof. The equi-coercivity is meant such that all $\mathcal{J}_\varepsilon$ take the value $+\infty$ on $X \setminus \boldsymbol{Z}$.

"$\Rightarrow$" We start from $\mathcal{J}_\varepsilon \xrightarrow{\mathsf{M}} \mathcal{J}_0$ in X. If $u_\varepsilon \rightharpoonup u$ in $\boldsymbol{Z}$, then this also holds in X. Hence, the liminf estimate follows. To construct a recovery sequence $\widehat{u}_\varepsilon \rightharpoonup \widehat{u}$ in $\boldsymbol{Z}$ for arbitrary $\widehat{u} \in \boldsymbol{Z}$, we first assume $\mathcal{J}_0(\widehat{u}) < \infty$. We choose the recovery sequence $\widehat{u}_\varepsilon$ guaranteed by $\mathcal{J}_\varepsilon \xrightarrow{\mathsf{M}} \mathcal{J}_0$ in X, i.e. we know $\widehat{u}_\varepsilon \to \widehat{u}$ in X. The equi-coercivity (5.6) and $\mathcal{J}_\varepsilon(\widehat{u}_\varepsilon) \to \mathcal{J}_0(\widehat{u}) < \infty$ imply $\|\widehat{u}_\varepsilon\|_{\boldsymbol{Z}} \le R$. Hence, $\widehat{u}_\varepsilon \rightharpoonup \widehat{u}$ in $\boldsymbol{Z}$ by reflexivity of $\boldsymbol{Z}$. If $\mathcal{J}_0(\widehat{u}) = \infty$, we choose $\widehat{u}_\varepsilon = \widehat{u}$ giving $\widehat{u}_\varepsilon \to \widehat{u}$ in $\boldsymbol{Z}$. Hence, the liminf estimate yields $\infty = \mathcal{J}_0(\widehat{u}) \le \liminf_{\varepsilon\to 0} \mathcal{J}_\varepsilon(\widehat{u})$, which shows that we have a recovery sequence in $\boldsymbol{Z}$.

"$\Leftarrow$" Given $\mathcal{J}_\varepsilon \xrightarrow{\Gamma} \mathcal{J}_0$ in $\boldsymbol{Z}$, we take any sequence $u_\varepsilon \rightharpoonup u$ in X. If we have $\liminf_{\varepsilon\to 0} \|u_\varepsilon\|_{\boldsymbol{Z}} = \infty$, then the equi-coercivity implies $\mathcal{J}_\varepsilon(u_\varepsilon) \to \infty$ and the liminf estimate holds. If for some subsequence $\|u_{\varepsilon_k}\|_{\boldsymbol{Z}} \le C$, then $u_{\varepsilon_k} \rightharpoonup u$ in $\boldsymbol{Z}$, and the liminf estimate follows from that of $\mathcal{J}_\varepsilon \xrightarrow{\Gamma} \mathcal{J}_0$ in $\boldsymbol{Z}$. For the construction of recovery sequences, we can choose $\widehat{u}_\varepsilon = \widehat{u}$ if $\widehat{u} \in X \setminus \boldsymbol{Z}$. If $\widehat{u} \in \boldsymbol{Z}$ we choose a recovery sequence $\widehat{u}_\varepsilon \rightharpoonup \widehat{u}$ in $\boldsymbol{Z}$. By the compact embedding we have $\widehat{u}_\varepsilon \to \widehat{u}$ in X and the proof is finished. ∎

The following result will be very useful for studying the evolutionary Γ-convergence for gradient systems $(X, \mathcal{F}_\varepsilon, \Psi_\varepsilon)$ in Banach spaces, because there we need Γ-convergence for Ψ_ε and for Ψ^*_ε. The connection between Legendre transform and Γ-convergence is nontrivial because it involves the duality product $X{\times}X^* \ni (v, \xi) \mapsto \langle \xi, v\rangle$ which is only weak-strongly or strong-weakly continuous and moreover it is order reversing because of $-\Psi_\varepsilon$, hence "$\inf$" and "$\sup$" are interchanged.

Theorem 5.10 ([Att84, pp. 271]) *Let X be a separable, reflexive Banach space and assume that all $\Psi_\varepsilon : X \to [0, \infty]$ are dissipation potentials (namely lsc, convex and $\Psi_\varepsilon(0) = 0$). Then,*

$$\Psi_\varepsilon \xrightarrow{\Gamma} \Psi \qquad \Longleftrightarrow \qquad \Psi^*_\varepsilon \xrightarrow{\Gamma} \Psi^*\,.$$

The proof uses techniques from [Mos71], where the following equivalence was shown:

$$\Psi_\varepsilon \xrightarrow{\mathsf{M}} \Psi \quad \Longleftrightarrow \quad \Psi^*_\varepsilon \xrightarrow{\mathsf{M}} \Psi^*, \tag{5.7}$$

which is a direct consequence of the above theorem, but holds under weaker assumptions.

Sketch of proof. The following four implications imply the desired result.

(1) $(\Gamma_{\rm w}.\inf)$ for $\Psi_\varepsilon \quad \Longrightarrow \quad (\Gamma_{\rm s}.\sup)$ for Ψ_ε^*

(2) $(\Gamma_{\rm w}.\sup)$ for $\Psi_\varepsilon \quad \Longrightarrow \quad (\Gamma_{\rm s}.\inf)$ for Ψ_ε^*

(3) $(\Gamma_{\rm s}.\inf)$ for $\Psi_\varepsilon^* \quad \Longrightarrow \quad (\Gamma_{\rm w}.\sup)$ for Ψ_ε

(4) $(\Gamma_{\rm s}.\sup)$ for $\Psi_\varepsilon^* \quad \Longrightarrow \quad (\Gamma_{\rm w}.\inf)$ for Ψ_ε

The simpler directions are from "$\sup$" to "$\inf$", because we don't have to show existence of a converging sequence. We give the proof of (2) and observe that (4) is analogous.

Part (2): We consider an arbitrary sequence $\xi_\varepsilon \to \xi$.

For $\delta > 0$ we find $\widehat{v}_0$ such that $\Psi^*(\xi) \leq \delta + \langle \xi, \widehat{v}_0\rangle - \Psi(\widehat{v}_0)$. By $(\Gamma_{\rm w}.\sup)$ we find a recovery sequence $\widehat{v}_\varepsilon \rightharpoonup \widehat{v}_0$ and $\Psi(\widehat{v}_0) \geq \limsup_{\varepsilon\to 0} \Psi_\varepsilon(\widehat{v}_\varepsilon)$. With this, we have

$$\begin{aligned}\liminf_{\varepsilon\to 0} \Psi_\varepsilon^*(\xi_\varepsilon) &\overset{\text{Legr}}{\geq} \liminf_{\varepsilon\to 0} \big(\langle \xi_\varepsilon, \widehat{v}_\varepsilon\rangle - \Psi_\varepsilon(\widehat{v}_\varepsilon)\big)\\ &\overset{*}{=} \langle \xi, \widehat{v}_0\rangle - \limsup_{\varepsilon\to 0} \Psi_\varepsilon(\widehat{v}_\varepsilon) \geq \langle \xi, \widehat{v}_0\rangle - \Psi(\widehat{v}_0) \geq \Psi^*(\xi) - \delta.\end{aligned}$$

In $\overset{*}{=}$ we use the weak-strong continuity of $(v,\xi) \mapsto \langle \xi, v\rangle$. As $\delta > 0$ was arbitrary, we have $\liminf_{\varepsilon\to 0} \Psi_\varepsilon^*(\xi_\varepsilon) \geq \Psi^*(\xi)$ as desired for $(\Gamma_{\rm s}.\inf)$.

Part (1): We show this under the additional assumption that the family $\big(\Psi_\varepsilon\big)_\varepsilon$ is equi-superlinear. In this case, the constant recovery sequence $\widehat{\xi}_\varepsilon = \widehat{\xi}$ always works.

We first observe $v \mapsto \Psi_\varepsilon(v) - \langle \xi, v\rangle$ has at least one minimizer v_ε because Ψ_ε is superlinear, lsc, and convex. By equi-superlinearity we find $\|v_\varepsilon\| \leq C < \infty$ for all $\varepsilon > 0$. We first choose a subsequence $\varepsilon_k \to 0$ such that $\Psi_{\varepsilon_k}(\xi) \to \limsup_{\varepsilon\to 0} \Psi_\varepsilon(\xi)$. Next, we can extract a further subsequence such that $v_{\varepsilon_{k(l)}} \rightharpoonup v_*$ for $l \to \infty$ and conclude

$$\begin{aligned}&-\Psi^*(\xi) \overset{\text{Legr}}{=} \inf_{v\in X} \big(\Psi(v) - \langle \xi, v\rangle\big) \leq \Psi(v_*) - \langle \xi, v_*\rangle \overset{(\Gamma_{\rm w}.\inf)}{\leq} \liminf_{l\to\infty} \big(\Psi(v_{\varepsilon_{k(l)}}) - \langle \xi, v_{\varepsilon_{k(l)}}\rangle\big)\\ &= \lim_{l\to\infty} \big(-\Psi_{\varepsilon_{k(l)}}^*(\xi)\big) = -\limsup_{\varepsilon\to 0} \Psi_\varepsilon(\xi),\end{aligned}$$

which is the desired estimate of $(\Gamma_{\rm s}.\sup)$.

Part (3): This is much more difficult and we refer the reader to [Att84, pp. 271]. ■

Lemma 5.3 provides an interesting example for the application of Theorem 5.10. In fact, we have $\mathcal{F}_\varepsilon^*(\xi) = \frac{1}{2}\int_\Omega \xi\cdot\mathbb{A}_\varepsilon^{-1}\xi \,\mathrm{d}x$. Thus, the strong convergence for $\mathcal{F}_\varepsilon^*$ leads to the effective matrix arith$(\mathbb{A}^{-1})$ $=$ harm$(\mathbb{A})^{-1}$.

Another important tool of convex analysis is the weak-strong closedness of the graphs of the subdifferentials $\partial^{\rm F}\mathcal{F}_\varepsilon : X \rightrightarrows X^*$ in the limit $\varepsilon \to 0$. The following result is a variant of [Att84, Thm. 3.66], and it again relies strongly on semi-convexity.

Proposition 5.11 (Strong-weak closedness for subdifferentials for Γ-limits) *Assume that all $\mathcal{F}_\varepsilon : X \to \mathbb{R}_\infty$ are proper and lsc and $\mathcal{F}_\varepsilon \xrightarrow{\Gamma} \mathcal{F}_0$ in the reflexive Banach space X. Moreover, assume that*

$\big(\mathcal{F}_\varepsilon\big)_\varepsilon$ *is equi-semiconvex, i.e. there exists* $\lambda \in \mathbb{R}$ *such that all* $\mathcal{F}_\varepsilon$ *are* λ*-convex. Then, we have*

$$\left.\begin{array}{l} u_\varepsilon \to u \text{ in } X, \quad \xi_\varepsilon \rightharpoonup \xi \text{ in } X^* \\ \forall\, \varepsilon > 0:\ \xi_\varepsilon \in \partial^{\mathrm{F}} \mathcal{F}_\varepsilon(u_\varepsilon) \end{array}\right\} \quad \Longrightarrow \quad \mathcal{F}_\varepsilon(u_\varepsilon) \to \mathcal{F}_0(u) \text{ and } \xi \in \partial^{\mathrm{F}} \mathcal{F}_0(u). \tag{5.8}$$

Proof. The λ-convexity of $\mathcal{F}_\varepsilon$ gives

$$\mathcal{F}_\varepsilon(w) \geq \mathcal{F}_\varepsilon(u_\varepsilon) + \langle \xi_\varepsilon, w{-}u_\varepsilon\rangle + \frac{\lambda}{2}\|w{-}u_\varepsilon\|^2 \quad \text{for all } w \in X. \tag{5.9}$$

Choosing an arbitrary $\widehat{w} \in X$ the limsup condition $(\Gamma.\sup)$ provides a recovery sequence $\widehat{w}_\varepsilon \to \widehat{w}$ with $\mathcal{F}_\varepsilon(\widehat{w}_\varepsilon) \to \mathcal{F}_0(\widehat{w})$. Inserting $w = \widehat{w}_\varepsilon$ into (5.9) and passing to the limit $\varepsilon \to 0$ we can exploit the strong convergence $\widehat{w}_\varepsilon - u_\varepsilon \to \widehat{w} - u$ and the weak convergence $\xi_\varepsilon \rightharpoonup \xi$. Setting $\overline{F}_0 = \limsup_{\varepsilon\to 0} \mathcal{F}_\varepsilon(u_\varepsilon)$ we find

$$\mathcal{F}_0(\widehat{w}) \geq \overline{F}_0 + \langle \xi, \widehat{w}{-}u\rangle + \frac{\lambda}{2}\|\widehat{w}{-}u\|^2 \quad \text{for all } \widehat{w} \in X. \tag{5.10}$$

Using $(\Gamma.\inf)$ we have $\mathcal{F}_0(u) \leq \liminf_{\varepsilon\to 0} \mathcal{F}_\varepsilon(u_\varepsilon)$, while (5.10) with $\widehat{w} = u$ gives $\mathcal{F}_0(u) \geq \overline{F}_0 = \limsup \mathcal{F}_\varepsilon(u_\varepsilon)$, which provides the desired convergence $\mathcal{F}_\varepsilon(u_\varepsilon) \to \mathcal{F}_0(u)$.

Moreover, replacing $\overline{F}_0$ in (5.10) by $\mathcal{F}_0(u)$ we conclude $\xi \in \partial^{\mathrm{F}} \mathcal{F}_0(u)$ as desired. ∎

5.2 Evolutionary Γ-convergence via EVI

We consider metric GS $(M, \mathcal{F}_\varepsilon, \mathcal{D}_\varepsilon)$ (recall that this notation implies $\psi = \psi_{\mathrm{quadr}}$) and the associated EVI formulation which is ideal to pass to the limit $\varepsilon \to 0$ because the formulation only contains the functionals $\mathcal{F}_\varepsilon$ and $\mathcal{D}_\varepsilon$, but no derivatives like $|\dot{u}_\varepsilon|_{\mathcal{D}_\varepsilon}$ or $|\partial \mathcal{F}_\varepsilon|_{\mathcal{D}_\varepsilon}$ appear.

The following result is a variant of [DaS14, Thm. 2.17], where the more restrictive case $\mathcal{D}_\varepsilon = \mathcal{D}$ is treated, see also [MuS22].

Theorem 5.12 (Evolutionary Γ-convergence via EVI) *Consider a complete metric space* $(M, \mathcal{D})$ *and the metric GS* $(M, \mathcal{F}_\varepsilon, \mathcal{D}_\varepsilon)$ *with the following properties:*

$$\exists\, C \geq 1\ \forall\, \varepsilon > 0: \quad \mathcal{D} \leq \mathcal{D}_\varepsilon \leq C\mathcal{D}; \tag{5.11a}$$

$$\mathcal{D}_\varepsilon \xrightarrow{\mathrm{cc}} \mathcal{D}_0 \ \text{ in } (M, \mathcal{D}); \tag{5.11b}$$

$$\big(\mathcal{F}_\varepsilon\big)_\varepsilon \ \text{ is equi-compact in } (M, \mathcal{D}); \tag{5.11c}$$

$$\mathcal{F}_\varepsilon \xrightarrow{\Gamma} \mathcal{F}_0 \ \text{ in } (M, \mathcal{D}), \text{ where } \mathcal{F}_0 \text{ is proper}; \tag{5.11d}$$

$$\exists\, \lambda \in \mathbb{R}\ \forall\, \varepsilon > 0: \ (M, \mathcal{F}_\varepsilon, \mathcal{D}_\varepsilon) \text{ has an (EVI)}_\lambda \text{ semiflow}$$
$$S^\varepsilon_t : \mathscr{D}_\varepsilon \to \mathscr{D}_\varepsilon := \overline{\mathrm{dom}(\mathcal{F}_\varepsilon)}. \tag{5.11e}$$

Then, (EVI)$_\lambda$ *for* $(M, \mathcal{F}_0, \mathcal{D}_0)$ *has for each* $u^0_0 \in \mathscr{D}_0$ *a unique solution* $t \mapsto u_0(t) =: S^0_t(u^0_0)$. *Moreover,* $S^0_t : \mathscr{D}_0 \to \mathscr{D}_0$ *is a* λ*-contractive semiflow, and we have convergence of solutions as follows*

$$\mathscr{D}_\varepsilon \ni u^0_\varepsilon \to u^0_0 \in \mathscr{D}_0 \quad \Longrightarrow \quad \forall\, t > 0: \ \left\{ \begin{array}{c} S^\varepsilon_t(u^0_\varepsilon) \to S^0_t(u^0_0) \ \text{ and} \\ \mathcal{F}_\varepsilon\big(S^\varepsilon_t(u^0_\varepsilon)\big) \to \mathcal{F}_0\big(S^0_t(u^0_0)\big). \end{array}\right.$$

Proof. The proof follows closely the general strategy of the existence proofs.

Step 0: Approximating sequences. Here $u_\varepsilon : [0,T] \to M$ are given as EVI solutions.

Step 1: A priori estimate for finite energies. We start with $u_0^0 \in \mathrm{dom}(\mathcal{F}_0)$ and use $(\Gamma.\sup)$ to construct a recovery sequence $u_\varepsilon^0 \to u_0^0$ with $\mathcal{F}_\varepsilon(u_\varepsilon^0) \to \mathcal{F}_0(u_0^0)$. Moreover, $\mathcal{F}_0$ is lsc and has compact sublevels, hence by equi-compactness (5.11c) $\inf \mathcal{F}_0 =: \alpha_0 \in \mathbb{R}$ and $\inf \mathcal{F}_\varepsilon =: \alpha_\varepsilon \to \alpha_0$, see Theorem 5.6. Thus, for $\varepsilon \in]0, \varepsilon_*[$ we have

$$\mathcal{F}_\varepsilon(u_\varepsilon^0) \leq \mathcal{F}_0(u_0^0) + 1, \quad \alpha_\varepsilon \geq \alpha_0 - 1, \quad \mathcal{F}_\varepsilon(u_\varepsilon^0) - \alpha_\varepsilon \leq \mathcal{F}_0(u_0^0) - \alpha_0 + 2 =: \Delta_{\mathcal{F}}.$$

Our a priori estimates for EVI solutions in Section 4.6 provide

$$\int_0^T \frac{1}{2} |\dot{u}_\varepsilon|_{\mathcal{D}}(t)^2 \,\mathrm{d}t \leq \int_0^T \frac{1}{2} |\dot{u}_\varepsilon|_{\mathcal{D}_\varepsilon}(t)^2 \,\mathrm{d}t \leq \mathcal{F}_\varepsilon(u_\varepsilon^0) - \mathcal{F}_\varepsilon(u_\varepsilon(T)) \leq \Delta_{\mathcal{F}},$$

where we used the first estimate in (5.11a), which implies $|\dot{u}|_{\mathcal{D}} \leq |\dot{u}|_{\mathcal{D}_\varepsilon}$ a.e. Moreover, using $\mathcal{F}_\varepsilon(u_\varepsilon(t)) \leq \mathcal{F}_\varepsilon(u_\varepsilon^0) \leq \mathcal{F}_0(u_0^0) + 1$ and the equi-compactness (5.11c) show that there is a compact set $K \Subset M$ such that $u_\varepsilon(t) \in K$ for all $t \in [0,T]$ and all $\varepsilon \in]0, \varepsilon_*]$.

Step 2: Extraction of converging subsequences. With the results of Step 1 we have the equi-continuity $\mathcal{D}\big(u_\varepsilon(s), u_\varepsilon(t)\big) \leq C|t-s|^{1/2}$ and we can apply the Arzelà-Ascoli theorem to obtain a uniformly converging subsequence (not relabeled) in $[0,T]$, where $T > 0$ was arbitrary, hence we have

$$\forall\, t \geq 0 : \quad u_\varepsilon(t) \to u(t) \ \text{ in } (M, \mathcal{D}).$$

Step 3: Limit passage $\varepsilon \to 0$. For all $\varepsilon > 0$ we have (EVI)$_\lambda$:

$$\begin{aligned}&\forall\, w_\varepsilon \in \mathrm{dom}(\mathcal{F}_\varepsilon)\ \forall\, 0 \leq s < t :\\ &\frac{1}{2}\mathcal{D}_\varepsilon\big(w_\varepsilon, u_\varepsilon(t)\big)^2 \leq \frac{\mathrm{e}^{-\lambda(t-s)}}{2}\mathcal{D}_\varepsilon\big(w_\varepsilon, u_\varepsilon(s)\big)^2 + M_\lambda(t-s)\big(\mathcal{F}_\varepsilon(w_\varepsilon) - \mathcal{F}_\varepsilon(u_\varepsilon(t))\big).\end{aligned} \tag{5.12}$$

We emphasize here that λ is independent of ε.

For given $w \in \mathrm{dom}(\mathcal{F}_0)$ $(\Gamma.sup)$ from (5.11d) provides a recovery sequence $\widehat{w}_\varepsilon \to w$ with $\mathcal{F}_\varepsilon(\widehat{w}_\varepsilon) \to \mathcal{F}_0(w)$. Inserting $w_\varepsilon = \widehat{w}_\varepsilon$ into (5.12) we can pass to the limit $\varepsilon \to 0^+$, where we use $M_\lambda(t-s) > 0$, the continuous convergence (5.11b) for the distance, and $(\Gamma.\inf)$ for $\mathcal{F}_\varepsilon(u_\varepsilon(t))$:

$$\begin{aligned}&\forall\, w \in \mathrm{dom}(\mathcal{F}_0)\ \forall\, 0 \leq s < t :\\ &\frac{1}{2}\mathcal{D}_0\big(w, u(t)\big)^2 \leq \frac{\mathrm{e}^{-\lambda(t-s)}}{2}\mathcal{D}_0\big(w, u(s)\big)^2 + M_\lambda(t-s)\big(\mathcal{F}_0(w) - \mathcal{F}_0(u(t))\big).\end{aligned}$$

Thus, $u : [0, \infty[\to M$ solves (EVI)$_\lambda$ for the GS $(M, \mathcal{F}_0, \mathcal{D}_0)$.

As the EVI solutions are unique, we conclude that the convergence does hold for the whole family, i.e. without the extraction of a subsequence.

Step 4: Convergence of general initial data. From Section 4.6 we know that the induced semigroups $(S_t^\varepsilon)_{t\geq 0}$ are λ-contractions in $(M, \mathcal{D}_\varepsilon)$ wherever they are defined. In particular, we can extend the domain to its closure $\mathscr{D}_\varepsilon := \overline{\mathrm{dom}(\mathcal{F}_\varepsilon)}$. This also holds for the case $\varepsilon = 0$. Assume now

$$\mathscr{D}_\varepsilon \ni u_\varepsilon^0 \to u_0^0 \in \mathscr{D}_0 \quad \text{and} \quad u_\varepsilon(t) = S_t^\varepsilon(u_\varepsilon^0).$$

For arbitrary $\delta > 0$ we choose $\widehat{u}_0^0 \in \mathrm{dom}(\mathcal{F}_0)$ with $\mathcal{D}(u_0^0, \widehat{u}_0^0) < \delta$ and a recovery sequence $\widehat{u}_\varepsilon^0 \to \widehat{u}_0^0$ with $\mathcal{F}_\varepsilon(\widehat{u}_\varepsilon^0) \to \mathcal{F}_0(\widehat{u}_0^0)$. Then,

$$\mathcal{D}(u_\varepsilon^0, \widehat{u}_\varepsilon^0) \leq \mathcal{D}(u_\varepsilon^0, u_0^0) + \mathcal{D}(u_0^0, \widehat{u}_0^0) + \mathcal{D}(\widehat{u}_0^0, \widehat{u}_\varepsilon^0) < 2\delta \text{ for } \varepsilon \in \,]0, \varepsilon_1[.$$

With this and setting $\widehat{\Delta}_\varepsilon^\delta(t) := \mathcal{D}(S_t^\varepsilon(\widehat{u}_\varepsilon^0), S_t^0(\widehat{u}_0^0)\big)$, we can estimate for all $t \geq 0$ as follows:

$$\begin{aligned}\mathcal{D}\big(u_\varepsilon(t), u_0(t)\big) \;&\leq\; \mathcal{D}\big(u_\varepsilon(t), S_t^\varepsilon(\widehat{u}_\varepsilon^0)\big) + \widehat{\Delta}_\varepsilon^\delta(t) + \mathcal{D}(S_t^0(\widehat{u}_0^0), u_0(t)\big) \\ &\overset{(5.11a)}{\leq} \widehat{\Delta}_\varepsilon^\delta(t) + C\,\mathrm{e}^{\lambda t}\,\mathcal{D}\big(u_\varepsilon^0, \widehat{u}_\varepsilon^0\big) + C\,\mathrm{e}^{-\lambda t}\,\mathcal{D}(\widehat{u}_0^0, u_0^0\big) \;\leq\; \widehat{\Delta}_\varepsilon^\delta(t) + C\,\mathrm{e}^{-\lambda t}\,\big(2\delta + \delta\big).\end{aligned}$$

Because Step 2 shows the uniform convergence of $\widehat{\Delta}_\varepsilon^\delta \to 0$ on all $[0, T]$ for $\delta > 0$ fixed and $\varepsilon \to 0$, we obtain uniform convergence of $u_\varepsilon \to u = u_0$ on $[0, T]$ by first making δ small and then ε.

Step 5: Energy convergence. We refer to Step 4 in the proof of [DaS14, Thm. 2.17]. ∎

As in the static case we have a commuting diagram. Passing to the "right limit" in $(\mathcal{F}_\varepsilon, \mathcal{D}_\varepsilon, u_\varepsilon^0)$ for $\varepsilon \to 0$ (horizontal direction) can be interchanged by solving (EVI)$_\lambda$ (vertical direction).

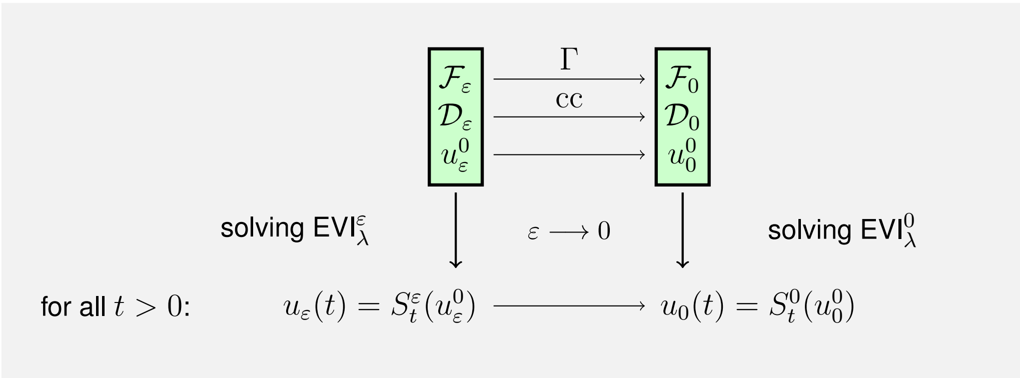


Figure 5.1: Commuting diagram for Γ-convergence of EVI solutions.

We consider two applications, where the first treats a linear parabolic equation and shows that homogenization can be treated with the above result. In the second example we treat a simple ODE in $M = H = \mathbb{R}^1$ and show that solutions do not converge, because of the solutions of the gradient-flow equation cannot be EVI solutions.

Example 5.13 (Homogenization of an Allen-Cahn equation) *Consider the Hilbert space $M = H = \mathrm{L}^2(\Omega)$ with $\Omega = \,]0, \ell[\, \subset \mathbb{R}^1$. For 1-periodic functions $a,\ b,\ B,\ c \in \mathrm{L}^\infty(\mathbb{R})$ satisfying $a(y) \geq \underline{a} > 0$, $B(y) \geq \underline{B} > 0$ and $c(y) \geq \underline{c} > 0$ for all $y \in \mathbb{R}$ we set $a_\varepsilon(x) = a(x/\varepsilon)$ and similarly b_ε, B_ε, and c_ε. With this, we define the energy functional*

$$\mathcal{F}_\varepsilon(u) = \begin{cases} \displaystyle\int_\Omega \big(\frac{a_\varepsilon}{2}\,u_x^2 + \frac{b_\varepsilon}{2}\,u^2 + \frac{B_\varepsilon}{4}\,u^4\big)\,\mathrm{d}x & \textit{for } u \in \mathrm{H}^1(\Omega), \\ \infty & \textit{otherwise.} \end{cases}$$

and the distances $\mathcal{D}_\varepsilon$ and $\mathcal{D}$ via

$$\mathcal{D}_\varepsilon(u, w)^2 = \int_\Omega c_\varepsilon |u{-}w|^2\,\mathrm{d}x, \quad \mathcal{D}_0(u, w)^2 = \int_\Omega c_{\mathrm{arith}} |u{-}w|^2\,\mathrm{d}x, \mathcal{D}(u, w)^2 = \int_\Omega \underline{c}\,|u{-}w|^2\,\mathrm{d}x,$$

DOI 10.20347/WIAS.PREPRINT.3022 Berlin 2023

where the subscripts "arith*" and "*harm*" denote the arithmetic and harmonic mean as in Lemma 5.3. Clearly, the assumptions* (5.11a) *and* (5.11b) *for* $\mathcal{D}_\varepsilon$ *and* $\mathcal{D}$ *are satisfied with* $C = \|c\|/\underline{c}$.

Using $\underline{a}$, $\underline{B} > 0$ *we obtain equi-coercivity of* $\big(\mathcal{F}_\varepsilon\big)_{\varepsilon>0}$ *in* $\mathrm{H}^1(\Omega)$*, which implies equi-compactness in* $H = \mathrm{L}^2(\Omega)$*, i.e.* (5.11c) *is also satisfied. Moreover, using the results from the Section 5.1 it is not difficult to show that we have the* Γ*-convergence* (5.11d)*, namely*

$$\mathcal{F}_\varepsilon \xrightarrow{\Gamma} \mathcal{F}_0 : u \mapsto \begin{cases} \int_\Omega \big(\frac{a_{\mathrm{harm}}}{2}\, u_x^2 + \frac{b_{\mathrm{arith}}}{2}\, u^2 + \frac{B_{\mathrm{arith}}}{4}\, u^4\big)\,\mathrm{d}x & \text{for } u \in \mathrm{H}^1(\Omega), \\ \infty & \text{otherwise.} \end{cases}$$

Finally, we can use the existence results of Section 2.5 to show that EVI solutions exist. The importance is of course that we find one $\lambda \in \mathbb{R}$ *that works for all* $\varepsilon > 0$*. In our Hilbert-space case the only non-convexity of* $\mathcal{F}_\varepsilon$ *can arise from the quadratic term* $b_\varepsilon u^2/2$ *which can be negative if* $\operatorname{ess\,inf} b < 0$*. Indeed, choosing* $\lambda = \operatorname{ess\,inf}\big\{\, b(y)/c(y) \,\big|\, y \in \mathbb{R}\,\big\}$ *we see that* $u \mapsto \mathcal{F}_\varepsilon(u) - \frac{\lambda}{2}\mathcal{D}_\varepsilon(0,u)^2$ *is convex. Hence, we obtain an EVI*$_\lambda$ *semiflow for* $\big(\mathrm{L}^2(\Omega), \mathcal{F}_\varepsilon, \mathcal{D}_\varepsilon\big)$.

Because we have $\mathrm{dom}(\mathcal{F}_\varepsilon) = \mathrm{H}^1(\Omega)$ *for all* $\varepsilon \geq 0$*, we have* $\mathscr{D}_\varepsilon = \mathrm{L}^2(\Omega)$*, and the convergence Theorem 5.12 shows that for all sequence* $u_\varepsilon^0 \to u_0^0$ *the solutions* $u_\varepsilon : [0,\infty[\, \to \mathrm{L}^2(\Omega)$ *of the Allen-Cahn equation*

$$c_\varepsilon \dot u_\varepsilon = \partial_x\big(a_\varepsilon \partial_x u_\varepsilon\big) - b_\varepsilon u_\varepsilon - B_\varepsilon u_\varepsilon^3 \text{ in } \Omega, \quad \partial_x u_\varepsilon(t,x) = 0 \text{ on } \partial\Omega, \quad u_\varepsilon(0,\cdot) = u_\varepsilon^0$$

converge to the unique solution $u_0 : [0,\infty[\, \to \mathrm{L}^2(\Omega)$ *of the homogenized Allen-Cahn equation*

$$c_{\mathrm{arith}} \dot u_0 = \partial_x\big(a_{\mathrm{harm}} \partial_x u_0\big) - b_{\mathrm{arith}} u_0 - B_{\mathrm{arith}} u_0^3 \text{ in } \Omega, \quad \partial_x u_0(t,x) = 0 \text{ on } \partial\Omega, \quad u_0(0,\cdot) = u_0^0$$

in the sense that $u_\varepsilon(t) \to u_0(t)$ *in* $\mathrm{L}^2(\Omega)$ *locally uniformly in* $[0,\infty[$.

The next example is the opposite, because it describes a situation where the interchanging of limiting process $\varepsilon \to 0$ and solving the gradient-flow equation does not work.

Example 5.14 (The wiggly-energy problem) *The following model was introduced by [Jam96, ACJ96], but it goes back to much earlier [Pra28, Tom29] explaining the emergence of dry friction from a molecular origin. A treatment of this problem using EDP-convergence, as is discussed in the following section, can be found in [DFM19].*

We consider the Hilbert-space gradient system

$$M = H = \mathbb{R}^1, \quad \mathcal{F}_\varepsilon(u) = \frac{1}{2}\, u^2 + \varepsilon^\alpha \cos(u/\varepsilon) \text{ with } \alpha > 0, \quad \mathcal{R}_\varepsilon(v) = \frac{1}{2}\, v^2.$$

We see that the dissipation does not depend on ε *at all and is given by the Euclidean distance* $\mathcal{D}_\varepsilon = \mathcal{D}_{\mathrm{Eucl}}$*, in particular we have* $\mathcal{D}_\varepsilon = \mathcal{D}_{\mathrm{Eucl}} \xrightarrow{\mathrm{cc}} \mathcal{D}_{\mathrm{Eucl}}$*. For the energy, the condition* $\alpha > 0$ *gives*

$$\mathcal{F}_\varepsilon \xrightarrow{\mathrm{cc}} \mathcal{F}_0 : u \mapsto \frac{1}{2}\, u^2 \quad \Longrightarrow \quad \mathcal{F}_\varepsilon \xrightarrow{\mathsf{M}} \mathcal{F}_0.$$

In this simple case, we can study the gradient-flow equation directly:

$$\dot u_\varepsilon = -\mathrm{D}\mathcal{F}_\varepsilon(u_\varepsilon) = -u_\varepsilon + \varepsilon^{\alpha-1} \sin\big(u_\varepsilon/\varepsilon\big), \quad u_\varepsilon(0) = u_\varepsilon^0.$$

For $\alpha > 1$ the right-hand side in the ODE convergences uniformly to $-u$ and hence, we can expect $u_\varepsilon(t) \to u_0^0\,\mathrm{e}^{-t}$ for all $t \geq 0$, if $u_\varepsilon^0 \to u_0^0$.

For $\alpha \in\,]0,1[$ the situation is different. We see that $u \mapsto -\mathrm{D}\mathcal{F}_\varepsilon(u)$ has many zeros, indeed their spacing around u_ε^0 is roughly $\pi\varepsilon$, if $\varepsilon^{\alpha-1} \gg |u_\varepsilon^0|$. Thus, the solutions $t \mapsto u_\varepsilon(t)$ get stuck between two zeros. Indeed if $\mathrm{D}\mathcal{F}_\varepsilon(\underline{u}_\varepsilon) = 0 = \mathrm{D}\mathcal{F}_\varepsilon(\overline{u}_\varepsilon)$ and $u_\varepsilon^0 \in [\underline{u}_\varepsilon, \overline{u}_\varepsilon]$, then we have $u_\varepsilon(t) \in [\underline{u}_\varepsilon, \overline{u}_\varepsilon]$ for all $t \geq 0$. By the fast oscillations of $u \mapsto \sin(u/\varepsilon)$ we can always find $\underline{u}_\varepsilon$ and $\overline{u}_\varepsilon$ with $\overline{u}_\varepsilon - \underline{u}_\varepsilon \leq 4\varepsilon\pi$. Thus, we conclude

$$\alpha \in\,]0,1[\text{ and } u_\varepsilon^0 \to u_0^0 \quad \Longrightarrow \quad u_\varepsilon(t) \to u_0^0 \text{ for all } t \geq 0.$$

The constant limits u_0 of the solutions u_ε are certainly not the solutions of the limiting gradient system $(\mathbb{R}^1, \mathcal{F}_0, \mathcal{D}_{\mathrm{Eucl}})$.

For $\alpha = 1$ a similar problem occurs: solutions u_ε starting with $u_\varepsilon^0 \to u_0^0 \in [-1,1]$ get stuck and satisfy $u_\varepsilon(t) \to u_0^0$. For $u_0^0 > 1$, first decay and reach $u = +1$ in finite time, namely $u_\varepsilon(t) \to u_0(t) = \cosh\big(\max\{\mathrm{Arcosh}(u_0^0){-}t, 0\}\big)$.

One can easily check that all assumptions in (5.11) *are satisfied except for* (5.11e). *This implies that for the case $\alpha \in\,]0,1]$ there is no $\lambda \in \mathbb{R}$, such that the evolutionary variational inequality (EVI)$_\lambda$ has a solution for all $\varepsilon \in\,]0,1]$. Indeed, from the general existence result in Theorem 4.22, we know that geodesic λ-convexity is a sufficient condition for existence. In this simple example we have $\mathcal{D} = \mathcal{D}_{\mathrm{Eucl}}$, hence geodesic λ-convexity is Hilbert-space λ-convexity, which means $\mathrm{D}^2\mathcal{F}_\varepsilon(u) = 1 - \varepsilon^{\alpha-2}\cos(u/\varepsilon) \geq \lambda$ for all $u \in \mathbb{R}$. Clearly, equi-semiconvexity only holds for $\alpha \geq 2$.*

5.3 Evolutionary Γ-convergence using the energy-dissipation balance

The approach to evolutionary Γ-convergence in the previous section is restrictive because of two major assumptions, namely (i) it applies only to classical gradient systems, i.e. $\psi = \psi_{\mathrm{quadr}}$ (but of course it allows metric gradient systems), and (ii) it needs equi-λ-convexity.

The following result uses the energy-dissipation balance and hence is more flexible. Of course, the result is weaker which is seen in two aspects. First, we will not have uniqueness of solutions and hence can only establish convergence along subsequences. Nevertheless, one can show that all accumulation points of families of solutions solve the limiting problem. Second, we have to impose a stronger condition on the convergence of the initial condition, i.e. they need to be *well prepared*:

$$\text{well-preparedness of initial conditions: } u_\varepsilon^0 \to u_0^0 \text{ and } \mathcal{F}_\varepsilon(u_\varepsilon^0) \to \mathcal{F}_0(u_0^0) < \infty.$$

Thus, we need the sequence of initial conditions $(u_\varepsilon^0)_{\varepsilon>0}$ to be a recovery sequence for $u_0^0 \in \mathrm{dom}(\mathcal{F}_0)$. While the restriction to recovery sequences is not too severe, the restriction to finite energy is significant as we see in the Allen-Cahn equation where $\mathrm{dom}(\mathcal{F}_\varepsilon) = \mathrm{H}^1(\Omega)$ is significantly smaller than the whole space $H = \mathrm{L}^2(\Omega)$.

The proof of following convergence result is only a small variant of the existence result provided in Theorem 3.13, but now we can start directly from the EDB for $\varepsilon > 0$ and pass to the limit in the four terms. Results of this type were originally developed in [MRS13, Thm. 4.8], where still the stronger condition $\mathcal{R}_\varepsilon \xrightarrow{\mathrm{M}} \mathcal{R}_0$ was imposed. Only in [LiR18] it was shown that the weaker condition $\mathcal{R}_\varepsilon \xrightarrow{\Gamma} \mathcal{R}_0$ is sufficient.

Theorem 5.15 (Evolutionary Γ-convergence using EDB) *On a reflexive Banach space we consider a family $(X, \mathcal{F}_\varepsilon, \mathcal{R}_\varepsilon)_{\varepsilon\geq 0}$ of gradient systems. If we assume*

$$\big(\mathcal{R}_\varepsilon\big)_{\varepsilon\geq 0} \text{ and } \big(\mathcal{R}^*_\varepsilon\big)_{\varepsilon\geq 0} \text{ are state-independent and equi-superlinear;} \tag{5.13a}$$

$$\mathcal{R}_\varepsilon \xrightarrow{\Gamma} \mathcal{R}_0 \quad (\text{or equivalently } \mathcal{R}^*_\varepsilon \xrightharpoonup{\Gamma} \mathcal{R}^*_0); \tag{5.13b}$$

$$\big(\mathcal{F}_\varepsilon\big)_{\varepsilon\geq 0} \text{ is equi-compact;} \tag{5.13c}$$

$$\mathcal{F}_\varepsilon \xrightarrow{\Gamma} \mathcal{F}_0, \quad \text{where } \mathcal{F}_0 : X \to \mathbb{R}_\infty \text{ is proper;} \tag{5.13d}$$

the subdifferentials $\big(\partial^{\mathrm{F}}\mathcal{F}_\varepsilon(\cdot)\big)_{\varepsilon\geq 0}$ are "closed for $\varepsilon \to 0$", i.e.

$$\left.\begin{array}{c} u_\varepsilon \to u_0,\ \xi_\varepsilon \rightharpoonup \xi_0, \\ \forall\, \varepsilon > 0 :\ \xi_\varepsilon \in \partial^{\mathrm{F}}\mathcal{F}_\varepsilon(u_\varepsilon) \end{array}\right\} \quad \Longrightarrow \quad \xi_0 \in \partial^{\mathrm{F}}\mathcal{F}_0(u_0). \tag{5.13e}$$

$$(X, \mathcal{F}_0, \mathcal{R}_0) \text{ satisfies the abstract chain rule (3.13);} \tag{5.13f}$$

then the following holds. If $\big(u_\varepsilon\big)_{\varepsilon>0}$ is a family of EDB solutions $u_\varepsilon : [0,T] \to X$ for $(X, \mathcal{F}_\varepsilon, \mathcal{R}_\varepsilon)$ with

$$\text{well-prepared initial conditions, i.e. } u_\varepsilon(0) \to u^0_0 \text{ and } \mathcal{F}_\varepsilon(u_\varepsilon(0)) \to \mathcal{F}_0(u^0_0), \tag{5.14}$$

then there exists a subsequence $\varepsilon_k \to 0$ and an EDB solution $u_0 : [0,T] \to X$ for $(X, \mathcal{F}_0, \mathcal{R}_0)$ with $u_0(0) = u^0_0$ such that

$$u_{\varepsilon_k}(t) \to u_0(t) \quad \text{in } X \quad \text{for all } t \in [0,T]; \tag{5.15a}$$

$$\dot u_{\varepsilon_k} \rightharpoonup \dot u_0 \quad \text{in } \mathrm{L}^1([0,T];X); \tag{5.15b}$$

$$\mathcal{F}_{\varepsilon_k}(u_{\varepsilon_k}(t)) \to \mathcal{F}_0(u_0(t)) \quad \text{for all } t \in [0,T]. \tag{5.15c}$$

In the proof we will also show the convergences

$$\int_0^T \mathcal{R}_{\varepsilon_k}\big(\dot u_{\varepsilon_k}(t)\big)\,\mathrm{d}t \to \int_0^T \mathcal{R}_0\big(\dot u_0(t)\big)\,\mathrm{d}t \quad \text{and} \quad \int_0^T \mathcal{R}^*_{\varepsilon_k}\big({-\xi_{\varepsilon_k}(t)}\big)\,\mathrm{d}t \to \int_0^T \mathcal{R}^*_0\big({-\xi_0(t)}\big)\,\mathrm{d}t,$$

which may be used to improve the convergences of $\dot u_\varepsilon$ and ξ_ε.

Before giving the proof of the above theorem we provide two auxiliary results. We leave the proof of the first lemma as an exercise.

Lemma 5.16 (Γ-convergence of integral functionals) *Consider an equi-superlinear family $\big(\mathcal{G}_\varepsilon\big)_{\varepsilon\geq 0}$ of proper lsc functionals $\mathcal{G}_\varepsilon : Y \to \mathbb{R}_\infty$ on a reflexive Banach space Y satisfying $\mathcal{G}_\varepsilon \xrightharpoonup{\Gamma} \mathcal{G}_0$. On $Z = \mathrm{L}^1([0,T];Y)$ define $\mathcal{J}_\varepsilon(u(\cdot)) := \int_0^T \mathcal{G}_\varepsilon(u(t))\,\mathrm{d}t$ for $T > 0$ and all $\varepsilon \in [0,1]$. Then, we have $\mathcal{J}_\varepsilon \xrightharpoonup{\Gamma} \mathcal{J}_0$.*

The next result should be seen as a simple generalization of the closedness result derived in Proposition 2.6. It shows that equi-semiconvexity for the family $\big(\mathcal{F}_\varepsilon\big)_{\varepsilon\geq 0}$ is sufficient to establish the condition "closed for $\varepsilon \to 0$" imposed abstractly in (5.13e). However, equi-semiconvexity is not necessary which is easily seen in Example 5.14. There the wiggly energy $\mathcal{F}_\varepsilon(u) = \frac12 u^2 + \varepsilon^\alpha \cos(u/\varepsilon)$ is equi-semiconvex for $\alpha \geq 2$, but the subdifferentials $\mathrm{D}\mathcal{F}_\varepsilon(u) = u - \varepsilon^{\alpha-1}\sin(u/\varepsilon)$ are "closed for $\varepsilon \to 0$" whenever $\alpha > 1$.

Proposition 5.17 (Closedness for $\varepsilon \to 0$) *On a reflexive Banach space X we consider a family $(\mathcal{F}_\varepsilon)_{\varepsilon\geq 0}$ that is equi-semiconvex, i.e.*

$$\exists\,\lambda \in \mathbb{R}\;\forall\,\varepsilon \in [0,1]: \quad \mathcal{F}_\varepsilon \text{ is } \lambda\text{-convex on } X.$$

If $\mathcal{F}_\varepsilon \xrightarrow{\Gamma} \mathcal{F}_0$, then we have the following closedness for $\varepsilon \to 0$:

$$\left.\begin{array}{c} u_\varepsilon \to u_0,\ \xi_\varepsilon \rightharpoonup \xi_0, \\ \forall\,\varepsilon>0:\ \xi_\varepsilon \in \partial^{\mathrm{F}}\mathcal{F}_\varepsilon(u_\varepsilon) \end{array}\right\} \quad\Longrightarrow\quad \mathcal{F}_\varepsilon(u_\varepsilon) \to \mathcal{F}_0(u_0) \text{ and } \xi_0 \in \partial^{\mathrm{F}}\mathcal{F}_0(u_0). \tag{5.16}$$

Proof. We follow the proof of Proposition 2.6 but need to use the Γ-convergence $\mathcal{F}_\varepsilon \xrightarrow{\Gamma} \mathcal{F}_0$.

Using the global characterization of the Fréchet subdifferential in Lemma 2.4 we have

$$\forall\,\varepsilon \in [0,1]\;\forall\,w_\varepsilon \in W: \quad \mathcal{F}_\varepsilon(w_\varepsilon) \geq \mathcal{F}_\varepsilon(u_\varepsilon) + \langle \xi_\varepsilon, w_\varepsilon{-}u_\varepsilon\rangle + \frac{\lambda}{2}\,\|w_\varepsilon{-}u_\varepsilon\|^2. \tag{5.17}$$

For all $\widehat{w} \in X$ $(\Gamma.\sup)$ provides a recovery sequence $\widehat{w}_\varepsilon \to \widehat{w}$ with $\mathcal{F}_\varepsilon(\widehat{w}_\varepsilon) \to \mathcal{F}_0(\widehat{w})$. Setting $F^* = \limsup_{\varepsilon\to 0}\mathcal{F}_\varepsilon(u_\varepsilon)$ and inserting $w_\varepsilon = \widehat{w}_\varepsilon$ we can pass to the limit in (5.17) and arrive at

$$\mathcal{F}_0(\widehat{w}) \geq F^* + \langle \xi_0, \widehat{w}{-}u_0\rangle + \frac{\lambda}{2}\,\|\widehat{w}{-}u_0\|^2.$$

Choosing $\widehat{w} = u_0$ we find $\limsup_{\varepsilon\to 0}\mathcal{F}_\varepsilon(u_\varepsilon) = F^* \leq \mathcal{F}_0(u_0)$. By $(\Gamma.\inf)$ we also have $\mathcal{F}_0(u_0) \leq \liminf_{\varepsilon\to 0}\mathcal{F}_\varepsilon(u_\varepsilon)$, which implies $\mathcal{F}_\varepsilon(u_\varepsilon) \to \mathcal{F}_0(u_0)$. Replacing F^* by $\mathcal{F}_0(u_0)$ is the last displayed formula gives $\xi_0 \in \partial^{\mathrm{F}}\mathcal{F}_0(u_0)$ as desired. ∎

Proof of Theorem 5.15.

Step 0: approximating sequences. Here the given solutions $u_\varepsilon : [0,T] \to X$ serve as the approximations for the desired limiting solution $u_0 : [0,T] \to X$.

Step 1: a priori estimates. For $\varepsilon > 0$ we have EDB solutions, i.e.

$$\mathcal{F}_\varepsilon(u_\varepsilon(T)) + \int_0^T \Big(\mathcal{R}_\varepsilon\big(\dot{u}_\varepsilon(t)\big) + \mathcal{R}_\varepsilon^*\big({-}\xi_\varepsilon(t)\big)\Big)\,\mathrm{d}t = \mathcal{F}_\varepsilon(u_\varepsilon(0)).$$

As in Step 1 of the proof of Theorem 5.12 we have

$$\mathcal{F}_\varepsilon(u_\varepsilon(0)) \leq \mathcal{F}_0(u_0^0){+}1 \text{ and } \mathcal{F}_\varepsilon(u_\varepsilon(0)) - \mathcal{F}_\varepsilon(u_\varepsilon(T)) \leq \mathcal{F}_0(u_0^0) - \inf\mathcal{F}_0 + 2 =: \Delta_{\mathcal{F}}. \tag{5.18}$$

According to assumption (5.13a) there exists a superlinear function $\psi : \mathbb{R} \to \mathbb{R}$ such that

$$\int_0^T \psi\big(\|\dot{u}_\varepsilon(t)\|_X\big)\,\mathrm{d}t \leq \Delta_{\mathcal{F}} \quad\text{and}\quad \int_0^T \psi\big(\|\xi_\varepsilon(t)\|_{X^*}\big)\,\mathrm{d}t \leq \Delta_{\mathcal{F}}.$$

As in Step 1 of the proof of Theorem 3.13 we obtain the equi-continuity

$$\big\|u_\varepsilon(t) - u_\varepsilon(s)\big\|_X \;\leq\; \omega_\psi^{\Delta_{\mathcal{F}}}\big(|t{-}s|\big) \quad\text{for all } t,s \in [0,T] \text{ and } \varepsilon > 0.$$

Finally, using $\mathcal{F}_\varepsilon(u_\varepsilon(t)) \leq \mathcal{F}_\varepsilon(u_\varepsilon(0)) \leq \mathcal{F}_0(u_0^0) + 1$ and the equi-compactness assumed in (5.13c), we find a compact set

$$K \Subset X \quad\text{such that}\quad u_\varepsilon(t) \in K \quad\text{for all } t,s \in [0,T] \text{ and } \varepsilon > 0.$$

Step 2: extraction of convergent subsequences. According to Step 1 we can apply Arzelà-Ascoli's selection principle and find a subsequence $\varepsilon_k \to 0$ and a continuous limit function $u_0 : [0,T] \to X$ such that $u_{\varepsilon_k} \to u_0$ in $\mathrm{C}^0([0,T];X)$. Moreover, using the superlinear bounds in the reflexive Banach spaces X and X^* we can choose a further subsequence (not relabeled) such that

$$\dot u_{\varepsilon_k} \rightharpoonup \dot u_0 \ \text{ in } \mathrm{L}^1([0,T];X) \quad \text{and} \quad \xi_{\varepsilon_k} \rightharpoonup \xi_0 \ \text{ in } \mathrm{L}^1([0,T];X^*).$$

Step 3: limit passage in (EDB)$_\varepsilon$, derivation of (EDI). For passing to the limit in (5.18) we will first derive an energy-dissipation inequality (EDI), i.e. it will be enough to derive liminf estimates on the left-hand side, but we need a limsup estimate for the right-hand side.

We first consider the two energy terms. For the right-hand side we simply use the well-preparedness (5.14) to obtain the desired convergence. For the first term on the left-hand side we use $(\Gamma.\inf)$ from $\mathcal{F}_\varepsilon \xrightarrow{\Gamma} \mathcal{F}_0$ and the pointwise convergence from Step 2 and conclude $\mathcal{F}_0(T)) \le \liminf_{\varepsilon\to 0} \mathcal{F}_\varepsilon(u_\varepsilon(T))$.

The two dissipation terms involving $\mathcal{R}_\varepsilon$ and $\mathcal{R}_\varepsilon^*$, respectively, can be treated separately. Using $\xi_{\varepsilon_k} \rightharpoonup \xi_0$ in $\mathrm{L}^1([0,T];X^*)$ and $\mathcal{R}_\varepsilon^* \xrightharpoonup{\Gamma} \mathcal{R}_0^*$ from (5.13b) we can apply Lemma 5.16 and find

$$\int_0^T \mathcal{R}_0^*\big(-\xi_0(t)\big)\,\mathrm{d}t \ \le\ \liminf_{k\to\infty} \int_0^T \mathcal{R}_{\varepsilon_k}^*\big(-\xi_{\varepsilon_k}(t)\big)\,\mathrm{d}t.$$

For the remaining term the convergence $\dot u_{\varepsilon_k} \rightharpoonup \dot u_0$ is not enough, because we only have $\mathcal{R}_\varepsilon \xrightarrow{\Gamma} \mathcal{R}_0$ which does not allow for weak convergence. To compensate for that we do a time discretization via $\tau = T/N$ and $N \in \mathbb{N}$. Defining $\widehat u_\tau : [0,T] \to X$ to be the piecewise affine interpolant of $u_0 : [0,T] \to X$ satisfying $\widehat u_\tau(j\tau) = u_0(j\tau)$ for $j = 0, ..., N$, we obtain

$$\begin{aligned}\int_0^T \mathcal{R}_0\big(\dot u_0(t)\big)\,\mathrm{d}t &= \sum_{j=1}^N \int_{k\tau-\tau}^{k\tau} \mathcal{R}_0(\dot u_0(t))\,\mathrm{d}t \overset{\text{Jensen}}{\ge} \sum_{j=1}^N \tau\,\mathcal{R}_0\Big(\frac{1}{\tau}{\textstyle\int_{j\tau-\tau}^{j\tau}} \dot u_0(t)\,\mathrm{d}t\Big) \\ &= \sum_{j=1}^N \tau\,\mathcal{R}_0\Big(\frac{1}{\tau}\big(u_0(j\tau) - u_0(j\tau{-}\tau)\big)\Big) = \sum_{j=1}^N \tau\,\mathcal{R}_0\Big(\frac{1}{\tau}\,{\textstyle\int_{j\tau-\tau}^{j\tau}} \dot{\widehat u}_\tau(t)\,\mathrm{d}t\Big) = \int_0^T \mathcal{R}_0\big(\dot{\widehat u}_\tau(t)\big)\,\mathrm{d}t.\end{aligned}$$

Doing the same time discretization for $\varepsilon > 0$ we obtain the lower estimate

$$\begin{aligned}\liminf_{k\to\infty} \int_0^T \mathcal{R}_{\varepsilon_k}\big(\dot u_{\varepsilon_k}(t)\big)\,\mathrm{d}t &\ge \liminf_{k\to\infty} \sum_{j=1}^N \tau\,\mathcal{R}_{\varepsilon_k}\Big(\frac{1}{\tau}\big(u_{\varepsilon_k}(j\tau) - u_{\varepsilon_k}(j\tau{-}\tau)\big)\Big) \\ &\overset{*}{\ge} \sum_{j=1}^N \tau\,\mathcal{R}_0\Big(\frac{1}{\tau}\big(u_0(j\tau) - u_0(j\tau{-}\tau)\big)\Big) = \int_0^T \mathcal{R}_0\big(\dot{\widehat u}_\tau(t)\big)\,\mathrm{d}t,\end{aligned}$$

where $\overset{*}{\ge}$ uses the liminf estimate of $\mathcal{R}_\varepsilon \xrightarrow{\Gamma} \mathcal{R}_0$ and the strong convergence $u_{\varepsilon_k}(t) \to u_0(t)$ established in Step 2.

In the last estimate we can now perform the limit $\tau = T/N \to 0$ and use $\dot{\widehat u}_\tau \to \dot u_0$ strongly in $\mathrm{L}^1([0,T];X)$, which implies, after extracting a subsequence $\tau_n \to 0$, the convergence $\dot{\widehat u}_{\tau_n}(t) \to \dot u_0(t)$ a.e. in $[0,T]$. Thus, using Fatou's lemma and the lsc of $\mathcal{R}_0$ we have

$$\liminf_{\tau_n\to 0} \int_0^T \mathcal{R}_0\big(\dot{\widehat u}_{\tau_n}(t)\big)\,\mathrm{d}t \overset{\text{Fatou}}{\ge} \int_0^T \liminf_{\tau_n\to 0} \mathcal{R}_0\big(\dot{\widehat u}_{\tau_n}(t)\big)\,\mathrm{d}t \overset{\text{lsc}}{\ge} \int_0^T \mathcal{R}_0\big(\dot u_0(t)\big)\,\mathrm{d}t.$$

With this we have shown $\liminf_{k\to\infty}\int_0^T \mathcal{R}_{\varepsilon_k}\big(\dot u_{\varepsilon_k}(t)\big)\;\mathrm{d}t \;\geq\; \int_0^T \mathcal{R}_0\big(\dot u_0(t)\big)\;\mathrm{d}t$, and the energy-dissipation inequality

$$\mathcal{F}_0(u_0(T)) + \int_0^T \Big(\mathcal{R}_0\big(\dot u_0(t)\big) + \mathcal{R}_0^*\big(-\xi_0(t)\big)\Big)\,\mathrm{d}t \;\leq\; \mathcal{F}_0(u_0(0))$$

is established.

It remains to identify ξ_0, for this we use the closedness condition (5.13e) and argue as in Exercise 2.7 and obtain $\xi_0(t) \in \partial^{\mathrm{F}}\mathcal{F}_0(u_0(t))$ a.e. in $[0,T]$.

Step 4: Derivation of (EDB) for $\varepsilon = 0$. With the abstract chain rule assumed to hold in (5.13f) we can apply the energy-dissipation principle from Theorem 3.9 and conclude that u_0 is indeed a EDB solution.

Moreover, the chain rule implies that the liminf estimates in Step 3 were indeed limits such that $\mathcal{F}_{\varepsilon_k}(u_{\varepsilon_k}(T)) \;\to\; \mathcal{F}_0(u_0(T))$ holds. However, we could have performed the liminf estimates on any subinterval $[0,t_*]$ with $t_* \in\,]0,T[$, from which we obtain $\mathcal{F}_{\varepsilon_k}(u_{\varepsilon_k}(t_*)) \;\to\; \mathcal{F}_0(u_0(t_*))$ for all $t_* \in [0,T]$. ■

5.4 EDP-convergence for gradient systems

In the above two section we studied the convergence of (a subsequence of) the solutions u_ε of a gradient system $(X,\mathcal{F}_\varepsilon,\mathcal{R}_\varepsilon)$ to a solution u_0 of the limiting gradient system $(X,\mathcal{F}_0,\mathcal{R}_{\mathrm{eff}})$ where $\mathcal{F}_0 = \Gamma\text{-}\lim\mathcal{F}_\varepsilon$ and $\mathcal{R}_{\mathrm{eff}} = \Gamma\text{-}\lim\mathcal{R}_\varepsilon$ in suitable topologies. It is important to note here that the two Γ-limits are not independent, because both have to be considered in the same Banach space X. The choice of X is dictated by the family of dissipation potentials $\big(\mathcal{R}_\varepsilon\big)_\varepsilon$. Then, the family $\big(\mathcal{F}_\varepsilon\big)_\varepsilon$ has to be considered in the same space X, and not in a so-called "energy space" which is often constructed as the smallest space in which $\big(\mathcal{F}_\varepsilon\big)_\varepsilon$ is weakly equi-compact.

However, there are situations in which the interaction between energy and dissipation is even stronger. Recovery sequences for the energy may not be compatible with recovery sequences for the dissipation. In such cases, the effective dissipation $\mathcal{R}_{\mathrm{eff}}$ cannot be obtained by looking at the family $\big(\mathcal{R}_\varepsilon\big)_\varepsilon$ alone, but one needs to consider the family of pairs $\big((\mathcal{F}_\varepsilon,\mathcal{R}_\varepsilon)\big)_\varepsilon$. Such a definition is the so-called *EDP-convergence* which was first defined in [LM*17] and made more precise in [MMP21]. We also refer to [DFM19] for a treatment of the wiggly-energy model of Example 5.14, to [MiS20, MPS21] for applications in reaction systems with slow and fast reactions, and to [Fre19, FrM21, Ste21, FrL21, PeS22] for reaction-diffusion systems.

The name of EDP-convergence derives from *convergence in the sense of the energy-dissipation principle*, because this notion of convergence is strongly linked to the EDP as formulated in Theorem 3.9 or Proposition 4.12. We give the main ideas and a few examples by using the Banach space formulation, but a similar theory can be obtained in the metric setting.

Considering the family $(X,\mathcal{F}_\varepsilon,\mathcal{R}_\varepsilon)_{\varepsilon>0}$ of gradient systems and a time horizon $T>0$, we define the dissipation functionals

$$\mathfrak{D}_\varepsilon(u) := \int_0^T \Big(\mathcal{R}_\varepsilon(u(t),\dot u(t)) + \mathcal{R}_\varepsilon^*\big(u(t), -\mathrm{D}\mathcal{F}_\varepsilon(u(t))\big)\Big)\,\mathrm{d}t. \tag{5.19}$$

Here we wrote the so-called "slope term of the dissipation" in terms of $\mathcal{R}_\varepsilon^*$ and a single-valued single-valued subdifferential $\mathrm{D}\mathcal{F}_\varepsilon$. However, in general one can replace this term by the more correct definition

$$\mathcal{S}_\varepsilon(u) := \mathrm{lsc}\big(\widetilde{\mathcal{S}}_\varepsilon\big) : u \mapsto \inf\Big\{ \inf_{u_k\to u} \widetilde{\mathcal{S}}_\varepsilon(u_k) \,\Big|\, u_k \to u \Big\}, \text{ where}$$
$$\widetilde{\mathcal{S}}_\varepsilon(u) := \begin{cases} \inf\{ \mathcal{R}^*_\varepsilon(u,-\xi) \,|\, \xi \in \partial^{\mathrm{F}}\mathcal{F}(u) \} & \text{for } \partial^{\mathrm{F}}\mathcal{F}(u) \neq \emptyset, \\ \infty & \text{otherwise.} \end{cases}$$

With this, the proper definition of $\mathfrak{D}_\varepsilon$ is given by $\mathfrak{D}_\varepsilon(u) = \int_0^T \big(\mathcal{R}_\varepsilon(u,\dot u)+\mathcal{S}_\varepsilon(u)\big)\,\mathrm{d}t$, however we will continue to use the $(\mathcal{R},\mathcal{R}^*)$ form to emphasize the special duality character encoded into $\mathfrak{D}_\varepsilon$.

The main point of the definition of $\mathfrak{D}_\varepsilon$ is that it is a functional on curves $u:[0,T]\to X$, unlike to $\mathcal{F}_\varepsilon$, which are functionals on the state space X. The idea is now to use classical Γ-convergence for the functionals $\mathfrak{D}_\varepsilon$ as well, but now on a space of curves, let us say $\mathrm{L}^2([0,T];X)$. To reflect the idea of gradient flows with well-prepared initial conditions we adapt the topology of $\mathrm{L}^2([0,T];X)$ by asking the families of functions $(u_\varepsilon)_{\varepsilon>0}$ additionally have uniformly bounded energy.

Definition 5.18 (Energy-bounded Γ-convergence of $\mathfrak{D}_\varepsilon$) *Given a Banach space X, a family $(\mathcal{F}_\varepsilon)_{\varepsilon>0}$ of energies $\mathcal{F}_\varepsilon : X \to \mathbb{R}_\infty$, and a family $(\mathfrak{D}_\varepsilon)_{\varepsilon\geq 0}$ of functionals $\mathfrak{D}_\varepsilon : \mathrm{L}^2([0,T];X) \to \mathbb{R}_\infty$ we say that $\mathfrak{D}_\varepsilon$* Γ-converges to $\mathfrak{D}_0$ with bounded energies, *and shortly write $\mathfrak{D}_\varepsilon \xrightarrow{\Gamma_{\mathrm{E}}} \mathfrak{D}_0$ or $\mathfrak{D}_0 = \Gamma_{\mathrm{E}}\text{-}\lim_{\varepsilon\to 0}\mathfrak{D}_\varepsilon$, if the following holds:*

Energy-bounded liminf estimate: (5.20a)

$$\left.\begin{array}{r} u_\varepsilon \to u_0 \text{ in } \mathrm{L}^2([0,T];X) \text{ and} \\ \sup_{\varepsilon\geq 0,\, t\in[0,T]} \mathcal{F}_\varepsilon(u_\varepsilon(t)) \leq F_* < \infty \end{array}\right\} \implies \liminf_{\varepsilon\to 0^+} \mathfrak{D}_\varepsilon(u_\varepsilon) \geq \mathfrak{D}_0(u_0),$$

Energy-bounded limsup estimate: (5.20b)

$$\begin{aligned} &\forall\, \widehat{u}_0 \in \mathrm{L}^2([0,T];X) \text{ with } \sup_{t\in[0,T]} \mathcal{F}_0(\widehat{u}_0(t)) \leq F_0 < \infty \\ &\quad \exists\, F_* \in \mathbb{R}\ \exists\, \big(\widehat{u}_\varepsilon\big)_{\varepsilon>0} \text{ with } \sup_{\varepsilon>0,\, t\in[0,T]} \mathcal{F}_\varepsilon(\widehat{u}_\varepsilon(t)) \leq F_* : \\ &\qquad\qquad \widehat{u}_\varepsilon \to \widehat{u}_0 \text{ in } \mathrm{L}^2([0,T];X) \text{ and } \limsup_{\varepsilon\to 0^+} \mathfrak{D}_\varepsilon(\widehat{u}_\varepsilon) \leq \mathfrak{D}_0(\widehat{u}_0). \end{aligned}$$

In particular applications, the choice $\mathrm{L}^2([0,T];X)$ for the space of curves can be replaced by other function spaces and the condition of energy boundedness can be dropped or amended by other conditions. The choice of a good notion of Γ-convergence should be seen as a problem-specific task or a modeling issue.

Using the above notion we can now define the simplest notion of EDP-convergence, and we refer to [DFM19, MMP21] for the more advance notions of “EDP-convergence with tilting” (in short tilt-EDP convergence) and “contact EDP-convergence with tilting” (in short “contact-EDP convergence).

Definition 5.19 (EDP-convergence of gradient system) *A family $\big((X,\mathcal{F}_\varepsilon,\mathcal{R}_\varepsilon)\big)_{\varepsilon>0}$ of gradient systems is said to* converge in the sense of the energy-dissipation principle *(in short “to EDP-converge”), if there exists an effective gradient system $(X,\mathcal{F}_{\mathrm{eff}},\mathcal{R}_{\mathrm{eff}})$ such that for all $T>0$ the following holds:*

$$\mathcal{F}_\varepsilon \xrightarrow{\Gamma} \mathcal{F}_{\mathrm{eff}} \text{ in } X \quad \text{and} \quad \mathfrak{D}_\varepsilon \xrightarrow{\Gamma_{\mathrm{E}}} \mathfrak{D}_0 \text{ in } \mathrm{L}^2([0,T];X)$$

where $\mathfrak{D}_\varepsilon$ is defined in (5.19) *and $\mathfrak{D}_0$ has the form*

$$\mathfrak{D}_0(u) = \int_0^T \Big(\mathcal{R}_{\mathrm{eff}}\big(u(t),\dot u(t)\big) + \mathcal{R}^*_{\mathrm{eff}}\big(u(t),-\mathrm{D}\mathcal{F}_{\mathrm{eff}}(u(t))\big)\Big)\,\mathrm{d}t.$$

We then shortly write $(X,\mathcal{F}_\varepsilon,\mathcal{R}_\varepsilon) \xrightarrow{\mathrm{EDP}} (X,\mathcal{F}_{\mathrm{eff}},\mathcal{R}_{\mathrm{eff}})$ for $\varepsilon\to 0^+$.

We observe that EDP-convergence of gradient systems has similar properties as Γ-convergence of functionals:

(I) The notion is independent of the concept of "solution", which in the case of classical functionals means minimizer (after adding a linear loading $-\langle \ell, \cdot \rangle$) and in the case of gradient systems systems means solutions of the gradient-flow equation (after adding an initial condition $u(0) = u^0$.

(II) Nevertheless, under suitable technical assumptions EDP-convergence of gradient systems implies the convergence of solutions if the initial conditions are well-prepared, see Proposition 5.20.

(III) The EDP limit $(X, \mathcal{E}_{\text{eff}}, \mathcal{R}_{\text{eff}})$ is uniquely determined by the family $(X, \mathcal{F}_\varepsilon, \mathcal{R}_\varepsilon)$, see Remark 5.21. Asking only the liminf estimate for $\mathfrak{D}_0$ will be enough to find some gradient structure that produces the correct gradient-flow equation, but the uniqueness of the gradient structure is lost, see Remark 5.24.

(IV) The involvement of general curves $u \in \mathrm{L}^2([0,T]; X)$ in the definition of EDP-convergence can be understood in the sense of fluctuation theory and the associated large-deviation principle, which provide a thermodynamical justification of the theory of gradient systems, see e.g. the discussion in [Pel14, Chap. 4] and [AD*11, MPR14, MP*17].

The next result shows that EDP-converge implies convergence of the solutions if suitable conditions are met. This result corresponds to Theorem 5.6 and Corollary 5.8 for the case of (static) Γ-convergence of functionals.

Proposition 5.20 (EDP-convergence implies convergence of solutions) *Assume* $(X, \mathcal{F}_\varepsilon, \mathcal{R}_\varepsilon) \overset{\text{EDP}}{\longrightarrow} (X, \mathcal{F}_{\text{eff}}, \mathcal{R}_{\text{eff}})$ *for* $\varepsilon \to 0^+$ *and that* $(X, \mathcal{F}_{\text{eff}}, \mathcal{R}_{\text{eff}})$ *satisfies the abstract chain rule* (3.13). *Moreover, assume that for a sequence* $(\varepsilon_k)_{k\in\mathbb{N}}$ *with* $\varepsilon_k \to 0^+$ *there are EDB solutions* $u_{\varepsilon_k} : [0,T] \to X$ *for* $(X, \mathcal{F}_{\varepsilon_k}, \mathcal{R}_{\varepsilon_k})$ *satisfying*

$$u_{\varepsilon_k} \to u \ \text{ in } \mathrm{L}^2([0,T]; X), \quad \forall\, t \in [0,T]\colon u_{\varepsilon_k}(t) \to u(t) \ \text{ in } X, \ \text{ and } \mathcal{F}_{\varepsilon_k}(u_{\varepsilon_k}(0)) \to \mathcal{F}_0(u(0)). \tag{5.21}$$

If additionally $u \in \mathrm{AC}([0,T]; X)$*, then it is an EDB solution for* $(X, \mathcal{F}_{\text{eff}}, \mathcal{R}_{\text{eff}})$*, and for* $\varepsilon_k \to 0^+$ *we have*

$$\mathcal{F}_{\varepsilon_k}(u_{\varepsilon_k}(t)) \to \mathcal{F}_{\text{eff}}(u(t)) \text{ for all } t \in [0,T] \quad \text{and} \quad \mathfrak{D}_{\varepsilon_k}(u_{\varepsilon_k}) \to \mathfrak{D}_0(u).$$

Proof. To simplify notation, we write ε in place of ε_k.

The argument uses the lsc property of the energy-dissipation balance as in previous sections. As u_ε is an EDB solution we have

$$\mathcal{F}_\varepsilon(u_\varepsilon(T)) + \mathfrak{D}_\varepsilon(u_\varepsilon) = \mathcal{F}_\varepsilon(u_\varepsilon(T)) + \int_0^T \big(\mathcal{R}_\varepsilon(u_\varepsilon, \dot u_\varepsilon) + \mathcal{R}^*_\varepsilon(u_\varepsilon, -\mathrm{D}\mathcal{F}_\varepsilon(u_\varepsilon))\big)\,\mathrm{d}t = \mathcal{F}_\varepsilon(u_\varepsilon(0)).$$

We pass to the limit $\varepsilon \to 0^+$ in this relation. By the assumption of the well-preparedness of the initial conditions $u_\varepsilon(0)$ we have convergence on the right-hand side.

On the left-hand side we use $\mathcal{F}_\varepsilon \overset{\Gamma}{\to} \mathcal{F}_{\text{eff}}$ and pointwise convergence $u_\varepsilon(T) \to u(T)$ to obtain $\mathcal{F}_{\text{eff}}(u(T)) \le \liminf_{\varepsilon\to 0^+} \mathcal{F}_\varepsilon(u_\varepsilon(T))$.

To treat the term $\mathfrak{D}_\varepsilon(u_\varepsilon)$ we observe that by the well-preparedness we have $\mathcal{F}_\varepsilon(u_\varepsilon(0)) \leq \mathcal{F}_{\text{eff}}(u(0)) +1 =: E_*$ for sufficiently small ε. Hence, the EDB solutions u_ε satisfy $\mathcal{F}_\varepsilon(u_\varepsilon(t)) \leq \mathcal{F}_\varepsilon(u_\varepsilon(0)) \leq E_*$. Thus, we can use the energy-bounded liminf estimate and obtain $\mathfrak{D}_0(u) \leq \liminf_{\varepsilon\to 0^+} \mathfrak{D}_\varepsilon(u_\varepsilon)$.

Using the duality structure of $\mathfrak{D}_0$ in terms of $\mathcal{F}_{\text{eff}}$ and $\mathcal{R}_{\text{eff}}$ we see that u satisfies the energy-dissipation inequality

$$\mathcal{F}_{\text{eff}}(u(T)) + \int_0^T \big(\mathcal{R}_{\text{eff}}(u,\dot u) + \mathcal{R}^*_{\text{eff}}(u, -\mathrm{D}\mathcal{F}_{\text{eff}}(u))\big)\,\mathrm{d}t \leq \mathcal{F}_{\text{eff}}(u(0)).$$

Since u is absolutely continuous and $(X, \mathcal{F}_{\text{eff}}, \mathcal{R}_{\text{eff}})$ satisfies the abstract chain rule, we can apply the energy-dissipation principle from Theorem (3.9) and conclude that u is an EDB solution.

As u is an EDB solution we know that EDI is in fact an EDB which implies that the liminf estimates are indeed limits providing an equality. This proves $\mathcal{F}_\varepsilon(u_\varepsilon(T)) \to \mathcal{F}_{\text{eff}}(u(T))$ and $\mathfrak{D}_\varepsilon(u_\varepsilon) \to \mathfrak{D}_0(u)$. Since T can be replaced by any $T' \in \,]0,T[$ the assertion is established. ■

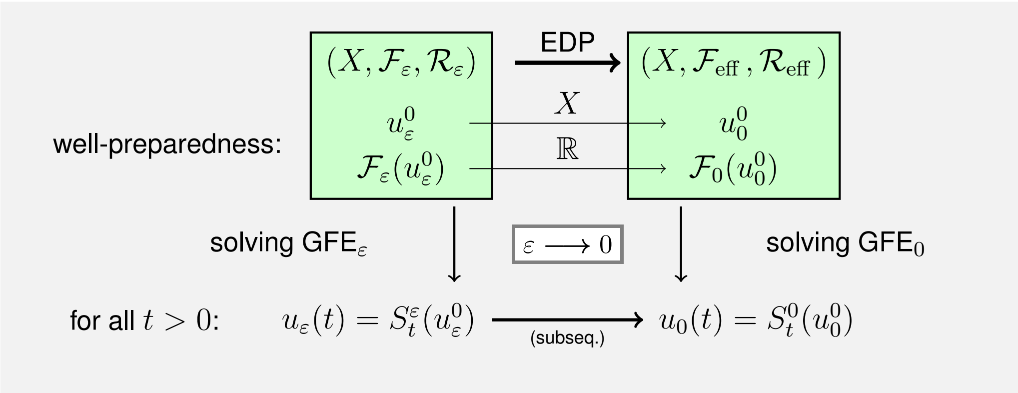


Figure 5.2: Commuting diagram for EDP-convergence and EDB solutions under suitable technical conditions.

Figure (5.2) shows the corresponding commuting diagram that can be established if we have enough compactness on the solutions $(u_\varepsilon)_{\varepsilon>0}$ to extract subsequences satisfying the assumptions in (5.21).

Exercise 5.1 *Discuss what additional conditions are needed such that the evolutionary Γ-convergence in Theorem 5.15 can be turned into a result on EDP-convergence.*

Remark 5.21 (On the uniqueness of EDP-limits) *Assuming that there are two gradient structures we first observe that $\mathcal{E}_{\text{eff}}$ and $\mathfrak{D}_0$ as Γ-limits are uniquely determined. Hence, if there exist two effective dissipation potentials $\mathcal{R}_{\text{eff}}$ and $\overline{\mathcal{R}}_{\text{eff}}$ generating $\mathfrak{D}_0$ we must have*

$$\mathcal{R}_{\text{eff}}(u,\dot u) + \mathcal{R}^*_{\text{eff}}\big(u, -\mathrm{D}\mathcal{F}_{\text{eff}}(u)\big) = \overline{\mathcal{R}}_{\text{eff}}(u,\dot u) + \overline{\mathcal{R}}^*_{\text{eff}}\big(u, -\mathrm{D}\mathcal{F}_{\text{eff}}(u)\big) =: \mathcal{M}(u,\dot u). \quad (5.22)$$

Setting $\dot u = 0$ we have $\mathcal{R}_{\text{eff}}(u,0) = 0 = \overline{\mathcal{R}}_{\text{eff}}(u,0)$ and find immediately $\mathcal{R}^*_{\text{eff}}\big(u, -\mathrm{D}\mathcal{F}_{\text{eff}}(u)\big) = \mathcal{M}(u,0) = \overline{\mathcal{R}}^*_{\text{eff}}\big(u, -\mathrm{D}\mathcal{F}_{\text{eff}}(u)\big)$. Subtracting this identity from (5.22) *and assuming $\mathcal{M}(u,0) < \infty$, we obtain $\mathcal{R}_{\text{eff}}(u,\dot u) = \overline{\mathcal{R}}_{\text{eff}}(u,\dot u)$, which is the desired uniqueness.*

The main advantage in the definition of EDP convergence is that it can be applied in degenerate cases, where $\mathcal{R}_\varepsilon$ and $\mathcal{R}^*_\varepsilon$ are not uniformly coercive, but may degenerate for $\varepsilon \to 0^+$. Moreover, keeping the two terms $\mathcal{R}_\varepsilon(u_\varepsilon, \dot u_\varepsilon)$ and $\mathcal{R}^*_\varepsilon(u_\varepsilon, -\mathrm{D}\mathcal{F}_\varepsilon(u_\varepsilon))$ together we allow for the option that "microscopic information of $\mathcal{F}_\varepsilon$ may move into $\mathcal{R}_{\mathrm{eff}}$". We may define

$$\mathfrak{D}^{\text{rate}}_\varepsilon(u) = \int_0^T \mathcal{R}_\varepsilon(u,\dot u)\,\mathrm{d}t \quad \text{and} \quad \mathfrak{D}^{\text{slope}}_\varepsilon(u) = \int_0^T \mathcal{R}^*_\varepsilon(u, -\mathrm{D}\mathcal{F}_\varepsilon(u))\,\mathrm{d}t,$$

such that $\mathfrak{D}_\varepsilon = \mathfrak{D}^{\text{rate}}_\varepsilon + \mathfrak{D}^{\text{slope}}_\varepsilon$.

Keeping the sum $\mathfrak{D}^{\text{rate}}_\varepsilon + \mathfrak{D}^{\text{slope}}_\varepsilon$ together is the main difference to the theory developed in [SaS04, Ser11] where along EDB solutions $u_\varepsilon \to u$ the two independent liminf estimates

$$\int_0^T \mathcal{R}_{\mathrm{eff}}(u,\dot u)\,\mathrm{d}t \le \liminf_{\varepsilon\to 0^+} \mathfrak{D}^{\text{rate}}_\varepsilon(u_\varepsilon) \quad \text{and} \quad \int_0^T \mathcal{R}^*_{\mathrm{eff}}(u, -\mathrm{D}\mathcal{F}_{\mathrm{eff}}(u))\,\mathrm{d}t \le \liminf_{\varepsilon\to 0^+} \mathfrak{D}^{\text{slope}}_\varepsilon(u_\varepsilon) \tag{5.23}$$

are supposed. We emphasize that here the estimates are on EDB solutions and not on general curves. For a discussion of this concept we refer to [Bra14, Sec. 11.2] and [Mie16, Sec. 3.3.3]. Note also that our assumption for Theorem 5.15 are such that in Step 3 of the proof we can establish the two estimates in (5.23).

We discuss now three simple ODE examples of EDP-convergence and refer to [Fre19, MiS20, MPS21, FrM21, Ste21, FrL21, PeS22] for further applications including PDEs.

Example 5.22 (Two binary reactions generate one ternary reaction) *In [Mie23] as reaction system with four species with density vector $\boldsymbol{c} = (c_1, c_2, c_3, c_4) \in \boldsymbol{C} := [0,\infty[^4$ is considered that react by two binary reaction pairs $X_1 + X_2 \rightleftharpoons X_4$ and $X_1 + X_4 \rightleftharpoons X_3$. The point is that X_4 is considered as an intermediate product that exists only with a much lower equilibrium density $c_4^*(\varepsilon) = \varepsilon^2 w^*$, while the other equilibrium densities c_i^* for $i = 1, 2, 3$ are independent of ε.*

The gradient system $(\boldsymbol{C}, \mathcal{F}_\varepsilon, \mathcal{R}^*_\varepsilon)$ is given by

$$\mathcal{F}_\varepsilon(\boldsymbol{c}) = \lambda_{\mathrm{B}}\big(\frac{c_4}{\varepsilon^2 w^*}\big)\varepsilon^2 w^* + \sum_{i=1}^3 \lambda_{\mathrm{B}}\big(\frac{c_i}{c_i^*}\big)c_i^* \quad \textit{and}$$

$$\mathcal{R}^*_\varepsilon(\boldsymbol{c};\boldsymbol{\xi}) = \frac{\overline{\kappa}_1}{\varepsilon}(c_1c_2c_4)^{1/2}\mathsf{C}^*\big(\xi_1{+}\xi_2{-}\xi_4\big) + \frac{\overline{\kappa}_2}{\varepsilon}(c_1c_3c_4)^{1/2}\mathsf{C}^*\big(\xi_1{-}\xi_3{+}\xi_4\big).$$

The associated gradient-flow equation is the following reaction-rate equation

$$\dot{\boldsymbol{c}} = \frac{\overline{\kappa}_1}{\varepsilon}\Big(\frac{A_1}{\varepsilon}\,c_4 - \frac{\varepsilon}{A_1}\,c_1c_2\Big)\begin{pmatrix}1\\1\\0\\-1\end{pmatrix} + \frac{\overline{\kappa}_2}{\varepsilon}\Big(\frac{\varepsilon}{A_2}\,c_3 - \frac{A_2}{\varepsilon}\,c_1c_4\Big)\begin{pmatrix}1\\0\\-1\\1\end{pmatrix},$$

where $A_1 := (c_1^*c_2^*/w^*)^{1/2}$ and $A_2 = \big(c_3^*/(c_1^*w^*)\big)^{1/2}$. We see that that setting $c_4 = \varepsilon^2 w$ leads to a right-hand side that is independent of ε, but then we have $\varepsilon^2\dot w$ on the left-hand side.

Doing the formal limit $\varepsilon \to 0^+$ (which can be justified rigorously, see [Bot03]) we arrive at

$$\begin{pmatrix}\dot c_1\\ \dot c_2\\ \dot c_3\\ 0\end{pmatrix} = \overline{\kappa}_1\Big(A_1 w - \frac{1}{A_1}\,c_1c_2\Big)\begin{pmatrix}1\\1\\0\\-1\end{pmatrix} + \overline{\kappa}_2\Big(\frac{1}{A_2}\,c_3 - A_2\,c_1 w\Big)\begin{pmatrix}1\\0\\-1\\1\end{pmatrix}.$$

From the last equation we can calculate w explicitly as via $(\overline{\kappa}_1A_1{+}\overline{\kappa}_2A_2)w = \frac{\overline{\kappa}_1}{A_1}c_1c_2 + \frac{\overline{\kappa}_2}{A_2}c_3$. Note that the relation for w guarantees that the two terms in front of the stoichiometric vectors must be *equal, such that we are left with one reaction only having the form*

$$\begin{pmatrix} \dot c_1 \\ \dot c_2 \\ \dot c_3 \end{pmatrix} = \overline\kappa_{\text{eff}}(c_1)\Big(\frac{A_1}{A_2}\,c_3 - \frac{A_2}{A_1}\,c_1^2c_2\Big)\begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix} \quad \textit{with } \overline\kappa_{\text{eff}}(c_1) = \frac{\overline\kappa_1\,\overline\kappa_1}{\overline\kappa_1 A_1 + \overline\kappa_2 A_2 c_1}.$$

Thus, the effective reaction for $\varepsilon \to 0$ *is the ternary reaction pair* $2X_1 + X_2 \rightleftharpoons X_3$.

So far, the analysis was on the gradient-flow equation only. The EDP-limit $(\boldsymbol{C}, \mathcal{F}_{\text{eff}}, \mathcal{R}_{\text{eff}})$ *is shown to exist in [Mie23], where*

$$\mathcal{F}_\varepsilon \xrightarrow{\Gamma} \mathcal{F}_{\text{eff}} : \boldsymbol{c} \mapsto \begin{cases} \sum_{i=1}^3 \lambda_{\mathrm{B}}(c_i/c_i^*)c_i^* & \text{for } c_4 = 0, \\ \infty & \text{for } c_4 > 0, \end{cases} \quad \textit{and}$$

$$\mathcal{R}^*_{\text{eff}}(\boldsymbol{c};\boldsymbol{\xi}) = \overline\kappa_{\text{eff}}(c_1)\big(c_1^2c_2c_3\big)^{1/2}\mathsf{C}^*\big(2\xi_1+\xi_2-\xi_3\big).$$

In the next example we return to the wiggly-energy model that was already discussed in Example 5.14. We now follow the analysis in [DFM19, MMP21] where *contact EDP-convergence with tilting* to the gradient system $(\mathbb{R}, \mathcal{F}_{\text{eff}}, \mathcal{R}_{\text{eff}})$ was established (cf. [MMP21, Def. 2.14]). Here we establish the weaker notion of EDP-convergence to the gradient system $(\mathbb{R}, \mathcal{F}_{\text{eff}}, \overline{\mathcal{R}}_{\text{eff}})$.

Example 5.23 (EDP-convergence for the wiggly-energy model) *We consider a variant of the wiggly-energy problem studied in Example 5.14, namely* $(\mathbb{R}, \mathcal{F}_\varepsilon, \mathcal{R}_\varepsilon)$ *with*

$$\mathcal{F}_\varepsilon(u) = \phi(u) - \frac{A\varepsilon}{2\pi}\cos(2\pi u/\varepsilon) \quad \textit{and} \quad \mathcal{R}_\varepsilon(v) = \frac12 v^2,$$

where A *is a positive constant. Obviously, we have*

$$\mathcal{F}_\varepsilon \to \mathcal{F}_{\text{eff}} : u \mapsto \phi(u) \quad \textit{and} \quad \mathcal{R}_\varepsilon \to \mathcal{R}_0 : v \mapsto \frac12 v^2.$$

However, the Γ*-limit* $\mathfrak{D}_0$ *of*

$$\mathfrak{D}_\varepsilon : u \mapsto \int_0^T \Big(\frac12\,\dot u^2 + \frac12\big(\phi'(u) + A\sin(2\pi u/\varepsilon)\big)^2\Big)\,\mathrm{d}t$$

is nontrivial, see [DFM19], and has the form $\mathfrak{D}_0(u) = \int_0^T M(\dot u, \phi'(u))\,\mathrm{d}t$ *with*

$$M(v,\xi) = \inf\Big\{ \int_0^1 \Big(\frac{v^2}{2} z'(s)^2 + \frac12\big(\xi{+}A\sin(2\pi z(s))\big)^2\Big)\mathrm{d}s \;\Big|\; z \in \mathrm{H}^1(]0,1[),\ z(1) = z(0){+}1 \Big\}.$$

From the definitions we easily see the symmetries $M(-v,\xi) = M(v,\xi) = M(v,-\xi)$.

Moreover, [DFM19, Lem. 4.3] provides the following expansion for $v \approx 0$*:*

$$\begin{aligned} &M(v,\xi) = M_0(\xi) + M_1(\xi)|v| + O(|v|^{3/2}) \quad \textit{with } M_0(\xi) = \frac12 \min\{|\xi|{-}A, 0\}^2 \\ &\textit{and } M_1(\xi) = \int_0^1 \Big(\big(\xi{+}A\sin(2\pi y)\big)^2 - 2M_0(\xi)\Big)^{1/2}\mathrm{d}y. \end{aligned} \tag{5.24}$$

Here M_1 *can be evaluated explicitly (see also Figure 5.3) giving*

$$M_1(\xi) = \begin{cases} \frac{2}{\pi}\big(\sqrt{A^2{-}\xi^2} + \xi\arcsin(\xi/A)\big) & \text{for } |\xi| \le A, \\ \frac{2}{\pi}\big(\sqrt{|\xi|{-}A} + |\xi|\arcsin\big(\sqrt{A/|\xi|}\big)\big) & \text{for } |\xi| \ge A. \end{cases}$$

We first observe that we have the estimate $M(v,\xi) \geq \xi v$ for all $v, \xi \in \mathbb{R}$, which is a remainder of the Fenchel-Young inequality. To see this, we observe

$$\int_0^1 \Big(\frac{v^2}{2} z'(s)^2 + \frac{1}{2}\big(\xi{+}A\sin(2\pi z(s))^2\Big)\,\mathrm{d}s \geq \int_0^1 v z'(s)\big(\xi{+}A\sin(2\pi z(s))\big)\,\mathrm{d}s \tag{5.25}$$

$$= v\xi \int_0^1 z'(s)\,\mathrm{d}s + vA\int_0^1 z'(s)\sin(2\pi z(s))\,\mathrm{d}s$$

$$= v\xi\big(z(1){-}z(0)\big) + \frac{vA}{2\pi}\big(\cos(2\pi z(0)) - \cos(2\pi z(1))\big) = v\xi,$$

where we used the boundary condition $z(1) = z(0){+}1$ for the last identity. Taking the infimum over z gives $M(v,\xi) \geq \xi v$ as desired.

Moreover, we can discuss the equality $M(v,\xi) = \xi v$ explicitly. For $v = 0$ we have $M(0,\xi) = 0$ if and only if $\xi \in [-A, A]$ by the form of M_0. For $v > 0$, we see that the equality $M(v,\xi) = \xi v$ implies equality a.e. for the integrand in (5.25)*, i.e. $vz'(s) = \xi + A\sin(2\pi z(s)) > 0$ and hence $\xi > A$. With this we find*

$$1 = \int_{s=0}^1 \mathrm{d}s = \int_{s=0}^1 \frac{vz'(s)\ \mathrm{d}s}{\xi{+}A\sin(2\pi z(s))} = \int_{z=z(0)}^{z(0)+1} \frac{v\ \mathrm{d}z}{\xi + A\sin(2\pi z)} = \frac{v}{\sqrt{\xi^2{-}A^2}}.$$

With the similar argument for $v < 0$, we obtain $0 \neq |v| = \sqrt{\xi^2{-}A^2}$. As a result we have shown that

$$\forall\, \xi \in \mathbb{R}: \quad \min\big\{\, M(v,\xi) - \xi v \,\big|\, v \in \mathbb{R} \,\big\} = 0. \tag{5.26}$$

We now define the effective dissipation potential $\overline{\mathcal{R}}_{\mathrm{eff}}$ via

$$\overline{\mathcal{R}}_{\mathrm{eff}}(u,v) := M(v,\phi'(u)) - M(0,\phi'(u)).$$

By definition we have $\overline{\mathcal{R}}_{\mathrm{eff}}(u,0) = 0$, and the results in [DFM19, Prop. 4.11] show $\overline{\mathcal{R}}_{\mathrm{eff}}(u,v) = \overline{\mathcal{R}}_{\mathrm{eff}}(u,-v) \geq 0$ and the convexity of $\overline{\mathcal{R}}_{\mathrm{eff}}(u,\,\cdot\,)$.

It remains to show the representation

$$M(v,\phi'(u)) = \overline{\mathcal{R}}_{\mathrm{eff}}(u,v) + \overline{\mathcal{R}}^*_{\mathrm{eff}}\big(u, -\mathrm{D}\mathcal{F}_{\mathrm{eff}}(u)\big). \tag{5.27}$$

Using $\mathrm{D}\mathcal{F}_{\mathrm{eff}}(u) = \phi'(u)$ we obtain

$$\overline{\mathcal{R}}^*_{\mathrm{eff}}\big(u,-\phi'(u)\big) = \sup_{v\in\mathbb{R}}\big(-\phi'(u)v - \overline{\mathcal{R}}_{\mathrm{eff}}(u,v)\big) = \sup_{v\in\mathbb{R}}\big(-\phi'(u)v{-}M(v,\phi'(u)){+}M(0,\phi'(u))\big)$$

$$= \sup_{v\in\mathbb{R}}\big(-\phi'(u)v{-}M(v,-\phi'(u))\big) + M(0,\phi'(u)) = 0 + M(u,\phi'(u)).$$

Using the definition of $\overline{\mathcal{R}}_{\mathrm{eff}}$ this implies (5.27)*, and the desired EDP-convergence for the wiggly-energy model is established, i.e. we have $(\mathbb{R}, \mathcal{F}_\varepsilon, \mathcal{R}) \xrightarrow{\mathrm{EDP}} (\mathbb{R}, \mathcal{F}_{\mathrm{eff}}, \overline{\mathcal{R}}_{\mathrm{eff}})$.*

However, following the argumentation in [DFM19, MMP21] the derived effective dissipation potential $\overline{\mathcal{R}}_{\mathrm{eff}}$ is somehow artificial, because $\overline{\mathcal{R}}_{\mathrm{eff}}$ depends on the force $\xi = \phi'(u)$ via $\overline{\mathcal{R}}_{\mathrm{eff}}(u,v) = M(v,\phi'(u)) - M(0,\phi'(u))$. The notion of contact EDP-convergence appears tilting (see [MMP21, Def. 2.14]) is more more natural and leads to the effective dissipation potential $\mathcal{R}_{\mathrm{eff}}$ with

$$\mathcal{R}_{\mathrm{eff}}(v) = A^2\mathsf{R}(v/A) \ \ \textit{with}\ \mathsf{R}(w) = \frac{1}{2}\Big(|w|\sqrt{1{+}w^2} + \log\big(|w|{+}\sqrt{1{+}w^2}\big)\Big),$$

which is independent of u and hence of the force $\phi'(u)$. But the limit $\mathfrak{D}_0$ of the dissipation integrals $\mathfrak{D}_\varepsilon$ coincides with $\int_0^T \big(\mathcal{R}_{\mathrm{eff}}(\dot u){+}\mathcal{R}^*_{\mathrm{eff}}(-\phi'(u))\big)\,\mathrm{d}t$ only along solutions u of the effective gradient-flow equation $\dot u = \partial_\xi \overline{\mathcal{R}}^*_{\mathrm{eff}}(u, -\mathrm{D}\mathcal{F}_{\mathrm{eff}}(u)) = \partial_\xi \mathcal{R}^*_{\mathrm{eff}}(-\mathrm{D}\mathcal{F}_{\mathrm{eff}}(u))$.

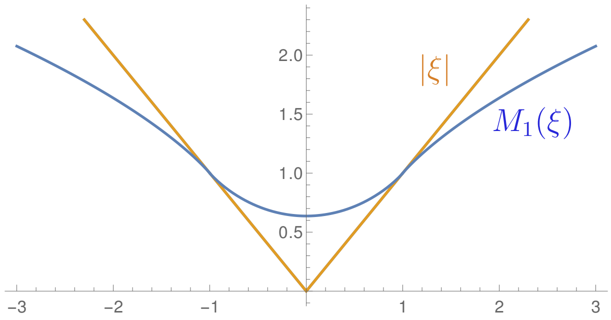


Figure 5.3: The function $\xi_1 \mapsto M_1(\xi)$ (blue) from (5.24) is plotted for the case $A = 1$ together with $\xi \mapsto |x|$ (orange).

Remark 5.24 (Nonuniqueness when using liminf only) *Often it is argued that for obtaining the effective gradient-flow equation it is not necessary to establish any limsup estimate for* $\mathfrak{D}_0$*. In particular, in the Sandier-Serfaty theory [SaS04, Ser11] only the liminf estimates* (5.23) *are requested. This is indeed true, but one has to be aware that by this approach we lose the uniqueness of the gradient structure. If we only impose the liminf estimates we may have two different gradient structures, which both generate the same effective equation.*

As an example consider the wiggly-energy model consider in the previous example. We claim that we can find $\widetilde{\mathcal{R}}$ *such that* $\widetilde{\mathcal{R}}(u,v) + \widetilde{\mathcal{R}}^*(u,-\mathrm{D}\mathcal{F}_{\text{eff}}(u)) \lneqq M(u,v)$*. Clearly, then the liminf estimate holds trivially, but the limsup estimate is false. To find an example the ideas is to make* $\mathcal{R}$ *smaller in some region where it does not increase the slope term* $\mathcal{R}^*(u,-\mathrm{D}\mathcal{F}_{\text{eff}}(u))$*. To be more precise, we choose* $\theta \in \mathrm{C}^0(\mathbb{R};[0,1])$ *with* $\theta(\xi) = 1$ *for* $|\xi| \geq A$ *and* $\theta(\xi) \in\,]0,1[$ *for* $|\xi| < A$ *and set*

$$\widetilde{\mathcal{R}}(u,v) = \theta(\phi'(u))\overline{\mathcal{R}}_{\text{eff}}(u,v) + \big(1-\theta(\phi'(u))\big)|\phi'(u)|\,|v| \;\geq |\phi'(u)|\,|v|.$$

For the last estimate we used $\overline{\mathcal{R}}_{\text{eff}}(u,v) \geq M_1(\phi'(u))|v|$ *and* $M_1(\xi) \geq |\xi|$ *with* M_1 *defined in* (5.24)*, see also Figure 5.3.*

Using the convexity of $\overline{\mathcal{R}}_{\text{eff}}(u,\cdot)$ *we have* $\widetilde{\mathcal{R}}(u,v) \lneqq \overline{\mathcal{R}}_{\text{eff}}(u,v)$*. Because* $\overline{\mathcal{R}}_{\text{eff}}(u,\cdot) = \widetilde{\mathcal{R}}(u,\cdot)$ *for* $|\phi'(u)| \geq A$ *we also have* $\widetilde{\mathcal{R}}^*(u,-\mathrm{D}\mathcal{F}_{\text{eff}}(u)) = \overline{\mathcal{R}}^*_{\text{eff}}(u,-\mathrm{D}\mathcal{F}_{\text{eff}}(u))$ *in that range. However, for* $|\phi'(u)| < A$ *we easily obtain* $\widetilde{\mathcal{R}}^*(u,-\mathrm{D}\mathcal{F}_{\text{eff}}(u)) = 0 = \overline{\mathcal{R}}^*_{\text{eff}}(u,-\mathrm{D}\mathcal{F}_{\text{eff}}(u))$*.*

Thus, we see that $\mathfrak{D}_0$ *generated by* $\overline{\mathcal{R}}_{\text{eff}} \oplus \overline{\mathcal{R}}^*_{\text{eff}}$ *as well as* $\widetilde{\mathfrak{D}}$ *generated by* $\widetilde{\mathcal{R}} \oplus \widetilde{\mathcal{R}}^*$ *satisfy the liminf estimate for the family* $\mathfrak{D}_\varepsilon$*.*

6 Rate-independent systems

6.1 Introduction to rate independence

A very special case of gradient systems is obtained in the so-called rate-independent case. This is a very degenerate model class where $\mathcal{R}(u,\cdot) : X \to [0,\infty]$ is positively homogeneous of degree 1 (shortly: one-homogeneous), i.e.

$$\forall\,\lambda > 0\;\forall\,u,v \in X: \quad \mathcal{R}(u,\lambda v) = \lambda\mathcal{R}(u,v).$$

This case is only interesting if the energy depends on $t \in [0,T]$, i.e. we consider $\mathcal{F} : [0,T]\times X \to \mathbb{R}_\infty$ where the dependence $t \mapsto \mathcal{F}(t,u)$ for fixed u describes an external loading like in a Banach space with $\mathcal{F}(t,u) = \mathcal{E}(u) - \langle \ell(t), u\rangle$. The gradient-flow equation reads

$$0 \in \partial\mathcal{R}(u(t),\dot u(t)) + \partial^{\mathrm{F}}\mathcal{F}(t,u(t)) \ \in X^*. \tag{6.1}$$

The term "*rate independence*" stems from the fact that for a (smooth and) strictly increasing transformation $\phi : [0,S] \to [0,T]$ of the loading in the form $\widetilde{\mathcal{F}}(s,u) = \mathcal{F}(\phi(s),u)$ a solution $u : [0,T] \to X$ for $(X,\mathcal{F},\mathcal{R})$ transforms to a solution $\widetilde{u} : s \mapsto u(\phi(s))$ for $(X,\widetilde{\mathcal{F}},\mathcal{R})$, and vice versa. The reason for this is that $v \mapsto \partial\mathcal{R}(u,v)$ is positively 0-homogeneous, i.e. $\partial\mathcal{R}(u,\lambda v) = \partial\mathcal{R}(u,v)$. Indeed the following result shows that the subdifferential of a one-homogeneous function has very special properties.

Lemma 6.1 (Subdifferential of one-homogeneous functionals) *Consider a lsc, positively one-homogeneous functional $\Psi : X \to \mathbb{R}_\infty$, then the subdifferential $\partial\Psi$ satisfies*

$$\forall\, v \in X : \quad \partial\Psi(v) = \big\{\, \xi \in \partial\Psi(0) \subset X^* \,\big|\, \langle \xi, v\rangle = \Psi(v)\,\big\}.$$

This formula shows that rate independence of $\partial\Psi(v)$ in the sense that the subdifferential does not depend on the length of v but only on the direction.

Of course, we see that our existence theory developed in previous sections does not apply, because $v \mapsto \mathcal{R}(u,v)$ is not superlinear. Hence, a special theory needs to be develop but nevertheless many similarities to the superlinear case remain. In the metric setting the rate-independent case corresponds to the choice $\psi = \psi_{\mathrm{id}} : r \mapsto r$.

We refer to the surveys [Mie05, Mie11a] and the monograph [MiR15] for the full theory which was developed in parallel in the works starting with [MiT99, MTL02, MiT04] using the name "rate-independent systems" and the works [FrM98, DaT10, FrL03, DFT05] using the name "quasistatic evolution". In the following we give a very short introduction into the theory with the single goal to show the connections of this theory with the general theory of gradient systems.

As a simple example we consider the case $M = X = \mathbb{R}^1$ with the energy $\mathcal{F}(t,u) = \frac{a}{2}u^2 - u\lambda t$, where $a > 0$ and $\lambda \in R$, and the dissipation potential $\mathcal{R}(u,v) = 2|v| + v$ satisfying $\partial\mathcal{R}(u,0) = [-1,3]$. The differential form (6.1) of the system takes the form

$$0 \in 2\,\mathrm{Sign}(\dot u) + \dot u + au - \lambda t, \tag{6.2}$$

where $v \mapsto \mathrm{Sign}(v) \subset \mathbb{R}$ is the set-valued signum function obtained as subdifferential of $v \mapsto |v|$. E.g. starting with $u(0) = 0$ we obtain the solution

$$u(t) = \begin{cases} \max\big\{0, (\lambda t{-}3)/a\big\} & \text{for } \lambda \geq 0, \\ \min\big\{0, (\lambda t{+}1)/a\big\} & \text{for } \lambda \leq 0. \end{cases}$$

6.2 Energetic solutions

The concept of energetic solutions plays the role of curves of maximal slope in the metric setting, but there are two major differences. First, the solutions are no longer absolutely continuous, i.e. they are allowed to have jumps with respect to the time variable $t \in [0,T]$. Second, we can allow the dissipation distance $\mathcal{D} : M{\times}M \to [0,\infty]$ to be an extended quasi-distance, i.e. $\mathcal{D}$ doesn't have to be symmetric and it may take the value ∞. Hence, we have to be careful about the order of arguments when writing the triangle inequality for $\mathcal{D}$. We emphasize that in the following we will always use the order "$\mathcal{D}(u_{\text{old}}, u_{\text{new}})$", where 'old' and 'new' refer to the ordering of the time variable $t \in [0,T]$, because $\mathcal{D}$ is considered to be a dissipation distance which associates with an arrow of time.

To simplify our exposition here, we assume that there is another true metric $D : M{\times}M \to [0,\infty[$ satisfying $D(u,w) \leq \mathcal{D}(u,w)$.

Definition 6.2 (Energetic rate-independent system) *A triple* $(M, \mathcal{F}, \mathcal{D})$ *is called an* energetic rate-independent system (ERIS) *with metric* $D : M{\times}M \to [0,\infty[$*, if*

(E.1) (M,D) *is a complete metric space;*

(E.2) $\mathcal{F} : [0,T]{\times}M \to \mathbb{R}_\infty$ *is lsc on* (M,D) *with domain* $\operatorname{dom}\mathcal{F} = [0,T]{\times}F_{\text{dom}} \neq \emptyset$*;*

(E.3) $\exists\, C_{\rm E}, c_{\rm E} > 0\ \forall\, u \in F_{\text{dom}} :\ \mathcal{F}(\cdot,u) \in {\rm C}^1([0,T])$ *and*
$|\partial_t \mathcal{F}(t,u)| \leq C_{\rm E}\big(\mathcal{F}(t,u){+}c_{\rm E}\big)$ *for all* $t \in [0,T]$*;*

(E.4) $\mathcal{D} : M{\times}M \to [0,\infty]$ *is lsc on on* (M,D) *and* $D(u,w) \leq \mathcal{D}(u,w)$ *for all* $u, w \in M$*;*

(E.5) $\forall\, u_1,u_2,u_3 \in M :\quad \mathcal{D}(u_1,u_3) \leq \mathcal{D}(u_1,u_2) + \mathcal{D}(u_2,u_3)$ *and* $\mathcal{D}(u_1,u_1) = 0$.

Below we will define *energetic solutions* (also called quasistatic evolutions) as natural limit of the time-incremental minimization scheme. We emphasize that the rate-independent case associates with the scalar dissipation function $\psi_{\text{ri}}(r) = r$, whence the metric construction

$$\tau\, \psi_{\text{ri}}\big(\frac{1}{\tau}\mathcal{D}(u_{k-1},u)\big) \;=\; \mathcal{D}(u_{k-1},u)$$

in Definition 4.1 simplifies considerably. In particular, the time step τ disappears completely, which can be seen again as a manifestation of rate independence. Thus, defining a partition $0 = t_0 < t_1 < \cdots < t_{N-1} < t_N = T$ we obtain

rate-independent time-incremental minimization scheme (TIMS)

$$u_k \quad \text{minimizes} \quad u \;\mapsto\; \mathcal{D}(u_{k-1},u) + \mathcal{F}(t_k,u). \tag{6.3}$$

We again emphasize that the time step $\tau_k = t_k{-}t_{k-1}$ does not show up because of rate independence. This fact can be used in material modeling for the study of microstructures in nonlinear plasticity [OrR99, CHM02, CoT05], in shape memory alloys [MTL02, BaK11, DeK13], or in crack propagation [DaT02, DFT05, DaZ07, DR*21].

The following result shows that one easily obtains useful information from this minimization scheme even without having a subdifferentials.

Proposition 6.3 (TIMS for ERIS) *Assume that* $(u_k)_{k=1,..,N}$ *solve the TIMS for the ERIS* $(M,\mathcal{F},\mathcal{D})$*, then we have, for all* $k \in \{1,\ldots,N\}$*,*

(i) $\mathcal{F}(t_k,u_k) + \mathcal{D}(u_{k-1},u_k) \leq \mathcal{F}(t_k,u_{k-1}) = \mathcal{F}(t_{k-1},u_{k-1}) + \int_{t_{k-1}}^{t_k} \partial_s\mathcal{F}(s,u_{k-1})\,{\rm d}s.$

(ii) $\mathcal{F}(T,u_N) + \sum_{m=1}^N \mathcal{D}(u_{m-1},u_m) \leq \mathcal{F}(0,u_0) + \int_0^T \partial_s\mathcal{F}(s,\underline{u}(s))\,{\rm d}s.$

(iii) u_k *minimizes the functional* $w \mapsto \mathcal{D}(u_k,w) + \mathcal{F}(t_k,w)$*.*

(iv) $\mathcal{F}(t_k,u_k) + \sum_{j=1}^k \mathcal{D}(u_{j-1},u_j) \leq {\rm e}^{C_{\rm E}t_k}\big(\mathcal{F}(0,u_0) + c_{\rm E}\big) - c_{\rm E}.$

Assertion (ii) uses the right-continuous interpolant $\underline{u} : [0,T] \to M$*, see* (3.9).

Proof. (i) is a simple consequence of (6.3) when comparing with $u = u_k$ and $u = u_{k-1}$.

(ii) then follows by summing over $k = 1$ to N.

To obtain (iii) we use the triangle inequality for $\mathcal{D}$ in $\overset{\triangle}{\leq}$ and obtain

$$\begin{aligned}\mathcal{D}(u_k, u_k) + \mathcal{F}(t_k, u_k) &\leq 0 + \mathcal{F}(t_k, u_k) + \mathcal{D}(u_{k-1}, u_k) - \mathcal{D}(u_{k-1}, u_k) \\ &\overset{(6.3)}{\leq} \mathcal{F}(t_k, w) + \mathcal{D}(u_{k-1}, w) - \mathcal{D}(u_{k-1}, u_k) \overset{\triangle}{\leq} \mathcal{D}(u_k, w) + \mathcal{F}(t_k, w),\end{aligned}$$

which is the desirable result.

For (iv) we abbreviate $f_k = \mathcal{F}(t_k, u_k) + c_\mathrm{E}$ and $d_k = \mathcal{D}(u_{k-1}, u_k)$ and find

$$\begin{aligned}f_k + d_k &\overset{\text{(i)}}{\leq} f_{k-1} + \int_{t_{k-1}}^{t_k} C_\mathrm{E}\big(\mathcal{F}(s, u_{k-1}) + c_\mathrm{E}\big)\,\mathrm{d}s \\ &\overset{**}{\leq} f_{k-1} + \int_{t_{k-1}}^{t_k} C_\mathrm{E}\mathrm{e}^{C_\mathrm{E}(s-t_{k-1})} f_{k-1}\,\mathrm{d}s = \mathrm{e}^{C_\mathrm{E}(t_k-t_{k-1})} f_{k-1},\end{aligned} \tag{6.4}$$

where $\overset{**}{\leq}$ exploits that (E.3) combined with Grönwall's lemma leads to the estimate $\mathcal{F}(s, u_{k-1}) + c_\mathrm{E} \leq \mathrm{e}^{C_\mathrm{E}|t-s|}(\mathcal{F}(t, u_{k-1}) + c_\mathrm{E})$. Using $d_k \geq 0$ we first obtain $f_k \leq \mathrm{e}^{C_\mathrm{E} t_k} f_0$.

With this we return to (6.4) and estimate as follows:

$$\begin{aligned}f_N + \sum_{j=1}^{N} d_N &= \sum_{j=1}^{N} (f_k + d_k) - \sum_{k=1}^{N-1} f_k \leq \sum_{k=1}^{N} f_{k-1}\,\mathrm{e}^{C_\mathrm{E}(t_k-t_{k-1})} - \sum_{k=1}^{N-1} f_k \\ &= f_0\,\mathrm{e}^{C_\mathrm{E}(t_1-t_0)} + \sum_{k=2}^{N} f_{k-1}\big(\mathrm{e}^{C_\mathrm{E}(t_k-t_{k-1})} - 1\big) \\ &\leq f_0\,\mathrm{e}^{C_\mathrm{E} t_1} + \sum_{k=2}^{N} f_0\,\mathrm{e}^{C_\mathrm{E} t_{k-1}}\big(\mathrm{e}^{C_\mathrm{E}(t_k-t_{k-1})} - 1\big) = f_0\,\mathrm{e}^{C_\mathrm{E} t_N}.\end{aligned}$$

Noting that N can be replaced by any $k \in \{1, .., N\}$ assertion (iv) is established. ■

In the above we recognize that (ii) is a discrete energy balance in the spirit of (3.11) or (4.20); however, it is unclear whether a term involving $\mathcal{R}^*$ or ψ^* is missing. We will see that this is not the case, because of the special structure of $\psi = \psi_\mathsf{id}$, leading to the dual function $\psi^*_\mathsf{id}(\zeta) = 0$ for $\zeta \in [0, 1]$ and $\psi^*_\mathsf{id}(\zeta) = \infty$ for $\zeta > 1$.

An important observation is the so-called *global stability* satisfied by u_k as is shown in (iii). We define the *set of globally stable states*

$$\mathcal{S}(t) := \big\{\, u \in M \,\big|\, \mathcal{F}(t, u) < \infty \text{ and } \forall\, w \in M:\ \mathcal{F}(t, u) \leq \mathcal{F}(t, w) + \mathcal{D}(u, w) \,\big\}$$

and call its elements the *(globally) stable states*. This stability has the simple interpretation that it is energetically not favorable to move from u to another point w if the dissipated energy $\mathcal{D}(u, w)$ is taken into account. In the toy example (6.2) we have $\mathcal{S}(t) = \big[(\lambda t{-}3)/a, (\lambda t{+}1)/a\big]$.

To compare this concept with the metric theory we recall the notion of global metric slope (4.5) from the classical metric theory and introduce the same object also for the extended quasi-metric $\mathcal{D}$, where we have to be careful about the order of the arguments:

$$|\partial_0^{\mathrm{gl}} \mathcal{F}(t, \cdot)|_\mathcal{D}(u) := \begin{cases} \infty & \text{for } u \notin \mathrm{dom}(\mathcal{F}(t, \cdot)), \\ \sup\Big\{ \dfrac{\big[\mathcal{F}(t, u) - \mathcal{F}(t, w)\big]_+}{\mathcal{D}(u, w)} \,\Big|\, w \in M \Big\} & \text{for } u \notin \mathrm{dom}(\mathcal{F}(t, \cdot)). \end{cases} \tag{6.5}$$

By simply comparing the definitions we clearly obtain the equivalence

$$u \in \mathcal{S}(t) \quad \Longleftrightarrow \quad |\partial_0^{\mathrm{gl}} \mathcal{F}(t,\cdot)|_{\mathcal{D}}(u) \leq 1. \tag{6.6}$$

For the dissipated energy we also need an adaptation as follows. For arbitrary curves $u : [0,T] \to M$ defined pointwise but assuming no continuity or measurability, we define for all $s,t \in [0,T]$ with $s < t$ the *variation dissipation*

$$\mathrm{Var}_{\mathcal{D}}(u,[s,t]) := \sup \Big\{ \sum_{k=1}^{N} \mathcal{D}\big(u(t_{k-1}),u(t_k)\big) \,\Big|\, N \in \mathbb{N},\ s \leq t_0 < t_1 < \cdots < t_N \leq t \Big\}.$$

By our assumption $D \leq \mathcal{D}$ every curve with $\mathrm{Var}_{\mathcal{D}}(u,[0,T]) < \infty$ also satisfies $\mathrm{Var}_D(u,[0,T]) < \infty$ in the complete metric space (M,D). This implies that such a u can have at most countably many jump points and that left and right limits

$$u(t^-) := \lim_{h \to 0^+} u(t-h) \quad \text{and} \quad u(t^+) := \lim_{h \to 0^+} u(t+h)$$

exist for all $t \in [0,T]$ (by definition one sets $u(0^-) = u(0)$ and $u(T^+) = u(T)$).

We are now ready to give a precise definition of a suitable notion of solutions for ERIS.

Definition 6.4 (Energetic solutions [Mie05, Def. 3.1]) *A curve $u : [0,T] \to M$ is called an* energetic solution *for the ERIS $(M,\mathcal{F},\mathcal{D})$ if the global stability (S) and the energy equality (E) hold:*

(S) $u(t) \in \mathcal{S}(t)$ *for all* $t \in [0,T]$,

(E) $\mathcal{F}(T;u(T)) + \mathrm{Var}_{\mathcal{D}}(u,[0,T]) = \mathcal{F}(0,u(0)) + \int_0^T \partial_s \mathcal{F}(s,u(s))\,\mathrm{d}s.$

We emphasize that the solutions are defined pointwise and that the condition of global stability is asked for *all* $t \in [0,T]$. Moreover, the energy balance (E) is posed only for the whole time interval $[0,T]$. However, using the chain rule from below it follows that it is valid on all subintervals, i.e. for all $r,t \in [0,T]$ with $r < t$ we have

$$\mathcal{F}(t;u(t)) + \mathrm{Var}_{\mathcal{D}}(u,[r,t]) = \mathcal{F}(r,u(r)) + \int_r^t \partial_s \mathcal{F}(s,u(s))\,\mathrm{d}s.$$

It is even possible to consider the limits $r \nearrow s$ and $t \searrow s$ to obtain the jump conditions

$$\mathcal{F}(s,u(s^+)) + \mathcal{D}(u(s),u(s^+)) = \mathcal{F}(s,u(s)) \ \text{ and } \ \mathcal{F}(s,u(s)) + \mathcal{D}(u(s^-),u(s)) = \mathcal{F}(s,u(s^-)).$$

Recall that it is possible that the three states $u(s^-)$, $u(s)$, and $u(s^+)$ may be mutually different.

Finally, we remark that it is tempting to rewrite (S) and (E) in $\mathcal{R}{\oplus}\mathcal{R}^*$ form:

$$\mathcal{F}(T;u(T)) + \int_0^T \Big(\psi_{\mathrm{id}}\big(|\dot u|_{\mathcal{D}}(t)\big) + \psi_{\mathrm{id}}^*\big(|\partial_0^{\mathrm{gl}} \mathcal{F}(t,\cdot)|_{\mathcal{D}}(u)\big) \Big)\,\mathrm{d}t = \mathcal{F}(0,u(0)) + \int_0^T \partial_s \mathcal{F}(s,u(s))\,\mathrm{d}s.$$

Since ψ_{id}^* only takes the value 0 and ∞, the finiteness of the left integral encodes the condition (S) at least almost everywhere. However, the major difficulty is to define the metric speed at jump points taking care of the possibly three different values $u(s^-)$, $u(s)$, and $u(s^+)$. Hence, it turns out that it is much easier and truly necessary to use the exact and pointwise formulation (S)&(E) from Definition 6.4.

6.3 Existence of energetic solutions

The following existence result follows exactly along the lines of the existence theory for curves of maximal slope. We will repeat the main arguments to show the analogies as well as the differences. The first major difference is that we cannot appeal to the Arzelá-Ascoli theorem because of the missing superlinearity. However, a metric version of Helly's selection theorem as derived in [MaM05, Thm. 3.2].

A second difference is more formal than mathematical. It was already observed in [MTL02, Thm. 2.5] that the global stability (S) implies a "lower energy estimate" which is the corresponding version of the metric chain-rule inequality, see (4.8). We will see that the proof is considerably simpler than that of Proposition 4.11, because the stability condition is equivalent to the property that the global slope is bounded by 1, see (6.6).

The essential new condition is the so-called "*closedness of the stable sets* in (6.7b), which can be seen as a replacement of the lower semicontinuity of the (global) slope. This condition is nontrivial here because we allow $\mathcal{D}$ to be non-continuous and take the value $+\infty$, see the discussion in Section 6.4.

Theorem 6.5 (Existence of energetic solutions) *Let the ERIS $(M, \mathcal{F}, \mathcal{D})$ satisfy the conditions* (E.1) *to* (E.5)*. Moreover, assume the following properties:*

compactness of sublevels:

$$\forall\, E > 0\, \forall\, t \in [0, T] : \quad S_E^{\mathcal{F}(t,\cdot)} = \big\{\, u \in M \;\big|\; \mathcal{F}(t, u) \leq E \,\big\} \text{ is compact}, \tag{6.7a}$$

closedness of the stable sets:

$$t_i \to t,\ u_i \to u,\ u_i \in \mathcal{S}(t_i) \quad \Longrightarrow \quad u \in \mathcal{S}(t), \tag{6.7b}$$

conditional continuity of the power $\partial_t\mathcal{F}$:

$$t_i \to t,\ u_i \to u,\ \sup_{i\in\mathbb{N}} \mathcal{F}(t_i, u_i) < \infty \quad \Longrightarrow \quad \partial_t\mathcal{F}(t_i, u_i) \to \partial_t\mathcal{F}(t, u). \tag{6.7c}$$

Then, for all $u_0 \in \mathcal{S}(0)$ there exists an energetic solution $u : [0, T] \to M$ for the ERIS $(M, \mathcal{F}, \mathcal{D})$ with $u(0) = u_0$. In particular, every accumulation point in the sense of pointwise convergence of a sequence of piecewise interpolants for the time-incremental minimization scheme (6.3) *is an energetic solution.*

Before going into the proof of the existence theorem, we will shortly discuss the version of the metric chain-rule inequality for ERIS. An important point is now that the solutions are not continuous, hence we can only derive an integrated version. Moreover, we need to generalize the theory to time-dependent energies. To see the analogy we observe that integrating the differential metric chain-rule inequality (4.8) over $t \in [r, s]$ we find

$$\mathcal{F}(u(s)) + \int_r^s |\dot u|_{\mathcal{D}}(t)\, |\partial\mathcal{F}|_{\mathcal{D}}(u(t))\, \mathrm{d}t \geq \mathcal{F}(u(r))$$

For stable states we have $|\partial\mathcal{F}|_{\mathcal{D}}(u(t)) \leq |\partial_{\mathrm{gl}}\mathcal{F}|_{\mathcal{D}}(u(t)) \leq 1$, such that $\int_r^s |\dot u|_{\mathcal{D}}(t)\, \mathrm{d}t = \mathrm{Var}_{\mathcal{D}}(u, [r, s])$ remains, where the last identity holds for absolutely continuous curves. Thus, the chain-rule inequality (6.8) appears naturally in the context of ERIS. Because of the global slope condition the proof is considerably simpler than that of Proposition 4.11.

Proposition 6.6 (Rate-indep. chain-rule inequality) *Consider the ERIS* $(M, \mathcal{F}, \mathcal{D})$ *satisfying the conditions* (E.1)–(E.5) *as well as* (6.7c). *If the curve* $u : [0, T] \to M$ *satisfies* $u(t) \in \mathcal{S}(t)$ *for all* $t \in [r, s[$ *and* $\sup_{t\in[r,s]} \mathcal{F}(t, u(t)) < \infty$*, then we have the chain-rule inequality*

$$\mathcal{F}(s, u(s)) + \mathrm{Var}_{\mathcal{D}}(u, [r, s]) \geq \mathcal{F}(r, u(r)) + \int_r^s \partial_t \mathcal{F}(t, u(t))\,\mathrm{d}t. \tag{6.8}$$

Proof. By assumption $t \to \mathcal{F}(t, u(t))$ is bounded. Using (E.3) also the power $\partial_t \mathcal{F}(t, u(t))$ is bounded such that the right-hand side in (6.8) is finite. Hence, the assertion holds if $\mathrm{Var}_{\mathcal{D}}(u, [r, s]) = \infty$. Thus, we can assume $\mathrm{Var}_{\mathcal{D}}(u, [r, s]) < \infty$ from now on.

We choose an arbitrary partition $r = t_0 < t_1 < \cdots < t_N = s$ and set $u_j = u(t_j)$, $f_j = \mathcal{F}(t_j, u_j)$ and $d_j = \mathcal{D}(u_{j-1}, u_j)$. For $j = 0, ..., N{-}1$, we have $u_j \in \mathcal{S}(t_j)$ which implies $f_j \leq \mathcal{F}(t_j, u_{j+1}) + d_{j+1}$. Hence, we have

$$f_{j+1} + d_{j+1} - f_j = f_{j+1} - \mathcal{F}(t_j, u_{j+1}) = \int_{t_j}^{t_{j+1}} \partial_t \mathcal{F}(t, u_{j+1})\,\mathrm{d}t \ \text{ for } j = 0, 1, ..., N{-}1.$$

Summing of these j and using the left-continuous interpolant $\overline{u}$ (cf. (3.9)) we find

$$\mathcal{F}(s, u(s)) + \mathrm{Var}_{\mathcal{D}}(u, [r, s]) - \mathcal{F}(r, u(r)) \geq f_N + \sum_{j=0}^{j-1} d_{j+1} - f_0 \geq \int_r^s \partial_t \mathcal{F}(t, \overline{u}(t))\,\mathrm{d}t. \tag{6.9}$$

Finally we choose the sequence of partitions by setting $\tau_N = (s{-}r)/N$ and $t_j^N = r{+}j\tau_N$. This gives the piecewise constant interpolants $\overline{u}_N : [0, T] \to M$. As $\mathrm{Var}_{\mathcal{D}}(u, [r, s]) < \infty$, we have $\overline{u}_N(t) \to u(t)$ for all $t \in [r, s]$ except for the jump points of u, which are at most countable. Moreover, (E.3) and the boundedness of $t \mapsto \mathcal{F}(t, u(t))$ implies $|\partial \mathcal{F}(t, \overline{u}_N(t))| \leq C$. Together with the assumed continuity of the power (6.7c) we can pass to the limit in the right-hand side of (6.9) and obtain the desired lower energy estimate. ■

Proof of Theorem 6.5. We follow the same five steps as in the existence proof for curves of maximal slope, see Theorem 4.18.

Step 0: Construction of approximants. We choose an arbitrary sequence of partitions $0 = t_0^N < t_1^N < \cdots < t_{N_1}^N < t_N^N = T$ whose fineness $\phi_N := \max\big\{\, t_j^N{-}t_{j-1}^N \;\big|\; j = 1, ..., N \,\big\}$ tends to 0 for $N \to \infty$.

The time-incremental minimization problem (6.3) is solvable in each step, because $\mathcal{D}(u_{k-1}^N, \cdot)$ and $\mathcal{F}(t_k, \cdot)$ are lsc on M and $\mathcal{F}(t_k, \cdot)$ has compact sublevels by (6.7a). By Proposition 6.3 the right-continuous interpolants $\underline{u}^N : [0, T] \to M$ satisfy the discrete a priori estimate

$$\mathcal{F}(T, \underline{u}^N(T)) + \mathrm{Var}_{\mathcal{D}}(\underline{u}^N, [0, T]) \leq \mathcal{F}(0, u_0) + \int_0^T \partial_t \mathcal{F}(t, \underline{u}_N(t))\,\mathrm{d}t, \tag{6.10}$$

where we use the identity $\mathrm{Var}_{\mathcal{D}}(\underline{u}^N, [0, T]) = \sum_{j=1}^N \mathcal{D}\big(\underline{u}^N(t_{j-1}), \underline{u}^N(t_j)\big)$ which holds for piecewise constant interpolants.

Step 1: A priori estimates. Proposition 6.3 provides the a priori estimates

$$\forall\, N \in \mathbb{N}\, \forall\, t \in [0, T]\colon\ \mathcal{F}(t, \underline{u}_N(t)) + \mathrm{Var}_{\mathcal{D}}(\underline{u}^N, [0, T]) \leq \mathrm{e}^{C_\mathrm{E} T}\big(\mathcal{F}(0, u_0){+}c_\mathrm{E}\big) - c_\mathrm{E} =: C_*.$$

Using (E.3) we obtain $\mathcal{F}(0,\underline{u}_N(t))+c_{\rm E} \le {\rm e}^{C_{\rm E}t}\big(\mathcal{F}(t,\underline{u}_N(t))+c_{\rm E}\big) \le {\rm e}^{C_{\rm E}T}C_* + c_{\rm E} = C_{**}$. Thus, we have

$$\forall\, N\in\mathbb{N}\,\forall\, t\in[0,T]\colon\ \underline{u}_N(t)\in S^{\mathcal{F}(0,\cdot)}_{C_{**}} \Subset M,$$

where we used the compactness of sublevels from (6.7a).

Step 2: Extraction of a converging subsequence. The a priori estimates from Step 1 allows us to apply the abstract version of Helly's selection principle (see [MaM05, Thm. 3.2] or [MiR15, Thm. B.5.13]). This implies that there exists a subsequence $\big(\underline{u}_{N_l}\big)_{l\in\mathbb{N}}$ and a limit function $u:[0,T]\to M$ such that we have the pointwise convergence

$$\forall\, t\in[0,T]:\quad \underline{u}_{N_l}(t)\to u(t)\quad \text{in } (M,D).$$

In particular, from $\underline{u}_N(0)=u_0$ we conclude $u(0)=u_0$ as desired.

Step 3: Derivation of the upper energy estimate. To pass to the limit $N_l\to\infty$ in (6.10) we first observe that the lsc of $\mathcal{F}(T,\cdot)$ gives $\mathcal{F}(T,u(T)) \le \liminf_{l\to\infty}\mathcal{F}(T,\underline{u}_{N_l}(T))$. For the second term on the left-hand side we deduce lsc from the lsc of $\mathcal{D}$ as follows. follows.

For arbitrary partitions $0=t_1<t_1<\cdots<t_N=T$ we have

$$\sum_{j=1}^N \mathcal{D}\big(u(t_{j-1}),u(t_j)\big) \overset{\mathcal{D}\ \text{lsc}}{\le} \liminf_{l\to\infty}\sum_{j=1}^N \mathcal{D}\big(\underline{u}_{N_l}(t_{j-1}),\underline{u}_{N_l}(t_j)\big) \le \liminf_{l\to\infty}\mathrm{Var}_{\mathcal{D}}(\underline{u}_{N_l},[0,T]) \le C_*.$$

Taking now the supremum over all partitions on the left-hand side gives $\mathrm{Var}_{\mathcal{D}}(u,[0,T]) \le \liminf_{l\to\infty}\mathrm{Var}_{\mathcal{D}}(\underline{u}_{N_l},[0,T])$ as desired.

For the power integral on the right-hand side in (6.10) we can pass to the limit (not liminf) by the same arguments as at the end of the proof of Proposition 6.6, i.e. we use (E.3) and (6.7c) once again. In summary, we have shown that the limiting curve $u:[0,T]\to M$ satisfies the upper energy estimate

$$\mathcal{F}(T,u(T)) + \mathrm{Var}_{\mathcal{D}}(u,[0,T]) \le \mathcal{F}(0,u(0)) + \int_0^T \partial_t\mathcal{F}(t,u(t))\,\mathrm{d}t. \tag{6.11}$$

Step 4: Derivation of energetic solutions. By Proposition 6.3(iii) we have the discrete global stability $\underline{u}_N(t^N_j)\in\mathcal{S}(t^N_j)$. Now fix a $t\in[0,T]$ such that $\underline{u}_{N_l}(t)\to u(t)$. By the construction of the piecewise constant interpolants we have $\underline{u}_N(t) = \underline{u}_N(t^N_{j_N(t)})$ for $t-\phi_N < t^N_{j_N(t)} \le t$, where ϕ_N is the fineness of the partition. Hence, $\widetilde{t}_l = t^{N_l}_{j_{N_l}(t)} \to t$, $\underline{u}_{N_l}(t)\to u(t)$, and $\underline{u}_{N_l}(t)\in\mathcal{S}(t)$, which implies $u(t)\in\mathcal{S}(t)$ by the closedness assumption (6.7b). Since $t\in[0,T]$ was arbitrary, we have established the global stability condition **(S)**.

Moreover, we have shown now all the conditions that are necessary for Proposition 6.6, and we obtain the the lower energy estimate (6.8). Together with the upper estimate in (6.11), we have established the energy balance **(E)**, and hence $u:[0,T]\to M$ is an energetic solution. ■

6.4 Closedness of the stable sets

The crucial and nontrivial condition for showing existence of energetic solutions is the closedness of the stable sets, namely condition (6.7b). This difficulty is comparable to the difficulty of showing closedness of the subdifferentials in rate-dependent gradient system in Banach spaces or to showing lsc of the metric slope.

The first case is the easiest case, namely when $\mathcal{D}$ is continuous.

Lemma 6.7 (Closedness of $\mathcal{S}$ via continuity) *Assume that the ERIS $(M, \mathcal{F}, \mathcal{D})$ satisfies* (E.1)–(E.5) *and that $\mathcal{D} : M{\times}M \to [0, \infty[$ is continuous, then the closedness condition* (6.7b) *holds.*

Proof. From $u_i \in \mathcal{S}(t_i)$ we have

$$\forall\, w \in M : \quad \mathcal{F}(t_i, u_i) \leq \mathcal{F}(t_i, w) + \mathcal{D}(u_i, w).$$

We simply pass to the limit $i \to \infty$ using $t_i \to t,\ u_i \to u$, (E.3), and the lsc of $\mathcal{F}(t, \cdot)$. This we obtain

$$\mathcal{F}(t, u) \leq \liminf_{i\to\infty} \mathcal{F}(t_i, u_i) \leq \lim_{i\to\infty} \big(\mathcal{F}(t_i, w) + \mathcal{D}(u_i, w)\big) = \mathcal{F}(t, u) + \mathcal{D}(u, w),$$

which is the desired result. ■

A typical application of this theory are models used for hysteresis in ferromagnetic materials, see [MiR15, Sec. 4.4]. A simplistic version is given by

$$M = \mathrm{L}^1(\Omega; \mathbb{R}^d), \quad \mathcal{D}(u, w) = \rho\|u{-}w\|_{\mathrm{L}^1}, \quad \mathcal{F}(t, u) = \int_\Omega \Big(\frac{\kappa}{2}|\nabla u|^2 {+} F(u) {-} H(t){\cdot}u\Big)\,\mathrm{d}x,$$

where $u : \Omega \to \mathbb{R}^d$ plays the role of the magnetization and $H(t) : \Omega \to \mathbb{R}^d$ is a time-dependent, applied field.

However, in many applications the continuity of $\mathcal{D}$ is too strong. In some cases a unidirectionality condition is desirable, which leads to

$$\mathcal{D}_{\mathrm{unidir}}(u, w) = \begin{cases} \int_\Omega \big(w(x) - u(x)\big)\,\mathrm{d}x & \text{if } w \geq u \text{ a.e. in } \Omega, \\ \infty & \text{else.} \end{cases}$$

Typical applications of this idea are in damage processes ([Tho10, KnS12, KRZ13]) or crack propagation [FrL03, DFT05, DaZ07, DaT10], not allowing for any healing.

In such cases the theory of "*mutual recovery sequences*" can be helpful. The MRS condition (introduced in [MRS08] as JRS) reads as follows:

$$\begin{aligned} &\text{for } (t_j, u_j) \to (t_*, u_*) \text{ with } u_j \in \mathcal{S}(t_j) \text{ and } \widehat{u} \in M \\ &\text{there exists } \big(\widehat{u}_j\big)_{j\in\mathbb{N}} \text{ such that } \widehat{u}_j \to \widehat{u} \text{ and} \\ &\qquad \limsup_{j\to\infty} \big(\mathcal{F}(t_j, \widehat{u}_j) {+} \mathcal{D}(u_j, \widehat{u}_j) {-} \mathcal{F}(t_j, u_j)\big) \leq \mathcal{F}(t_*, \widehat{u}) {+} \mathcal{D}(u_*, \widehat{u}) {-} \mathcal{F}(t_*, u_*). \end{aligned} \tag{6.12}$$

In the theory of crack propagation this condition is established via the so-called "jump transfer lemma", see [FrL03, DFT05].

Lemma 6.8 (Closedness of $\mathcal{S}$ via MRS) *Assume that the ERIS $(M, \mathcal{F}, \mathcal{D})$ satisfies* (E.1)–(E.5) *and* (6.12)*, then the closedness condition* (6.7b) *holds.*

Proof. We consider t_j, t_*, u_j, and u_* as in (6.7b). The closedness is established if we can show $u_* \in \mathcal{S}(t_*)$.

For an arbitrary test state $\widehat{u}$ we choose $\big(\widehat{u}_j\big)_{j\in\mathbb{N}}$ as provided in (6.7b). Then, we have

$$\mathcal{F}(t_*,\widehat{u})+\mathcal{D}(u_*,\widehat{u})-\mathcal{F}(t_*,u_*) \geq \limsup_{j\to\infty} \big(\mathcal{F}(t_j,\widehat{u}_j)+\mathcal{D}(u_j,\widehat{u}_j)-\mathcal{F}(t_j,u_j)\big) \geq 0$$

where the last estimate follows via $u_i \in \mathcal{S}(t_i)$. Rearranging the terms gives $u_* \in \mathcal{S}(t_*)$. ■

The usefulness of this condition is already seen in classical linearized elastoplasticity, where we have

$$\mathcal{D}(u,w) = \big\|u{-}w\big\|_{\mathrm{L}^1(\Omega)} \quad \text{and} \quad \mathcal{F}(t,u) = \frac{1}{2}\langle \mathbb{A}u, u\rangle_{\mathrm{L}^2(\Omega)} - \langle \ell(t), u\rangle.$$

Here $\mathbb{A} = \mathbb{A}^*$ is bounded and positive definite operator on $\mathrm{L}^2(\Omega)$. Since $\mathrm{L}^2(\Omega)$ does not compactly embed into $\mathrm{L}^1(\Omega)$ the construction of solutions has to be based on the weak topology, in $\mathrm{L}^2(\Omega)$, but $\mathcal{D}$ is only lsc but not continuous.

Nevertheless, the construction of a recovery sequence works because we can use cancellations in the terms appearing in the limsup condition in (6.12). For a sequence $u_j \rightharpoonup u$ in L^2 and a fixed $\widehat{u} \in \mathrm{L}^2$ we define

$$\widehat{u}_j = \widehat{u} + u_j - u.$$

Clearly, we have $\mathcal{D}(u_j,\widehat{u}_j) = \mathcal{D}(u,\widehat{u})$, i.e. the two weakly converging sequences cancel each other. Similarly, using the quadratic structure of $\mathcal{F}(t,\cdot)$ we have

$$\begin{aligned}\mathcal{F}(t_j,\widehat{u}_j) - \mathcal{F}(t_j,u_j) &= \frac{1}{2}\langle \mathbb{A}(\widehat{u}{-}u), \widehat{u}{+}2u_j{-}u\rangle - \langle \ell(t_j), \widehat{u}{-}u\rangle \\ &\to \frac{1}{2}\langle \mathbb{A}(\widehat{u}{-}u), \widehat{u}{+}u\rangle - \langle \ell(t_*), \widehat{u}{-}u\rangle = \mathcal{F}(t_*,\widehat{u}) - \mathcal{F}(t_*,u).\end{aligned}$$

This shows that the construction of mutual recovery sequences as in (6.12) works for this case.

Acknowledgments. The author is grateful to Moritz Gau and Jia-Jie Zhu for several critical and constructive remarks that helped to improve these lecture notes. Of course, this work benefited greatly from fruitful discussion with many collaborators, in particular Thomas Frenzel, Matthias Liero, Mark Peletier, Riccarda Rossi, Giuseppe Savaré, and Artur Stephan.

References

[AbM78] R. Abraham and J. E. Marsden, *Foundations of mechanics*, Benjamin/Cummings Publishing Co. Inc. Advanced Book Program, Reading, Mass., 1978, Second edition, revised and enlarged, With the assistance of Tudor Raţiu and Richard Cushman.

[ACJ96] R. Abeyaratne, C.-H. Chu, and R. James: *Kinetics of materials with wiggly energies: theory and application to the evolution of twinning microstructures in a Cu-Al-Ni shape memory alloy*. Phil. Mag. A **73** (1996) 457–497.

[AD*11] S. Adams, N. Dirr, M. A. Peletier, and J. Zimmer: *From a large-deviations principle to the Wasserstein gradient flow: a new micro-macro passage*. Comm. Math. Phys. **307**:3 (2011) 791–815.

[AGS05] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2005.

[Amb95] L. Ambrosio: *Minimizing movements*. Rend. Accad. Naz. Sci. XL Mem. Mat. Appl. (5) **19** (1995) 191–246.

[Arn89] V. I. Arnol′d, *Mathematical methods of classical mechanics*, Graduate Texts in Mathematics, vol. 60, Springer-Verlag, New York, 1989, Translated from the 1974 Russian original by K. Vogtmann and A. Weinstein, Corrected reprint of the second (1989) edition.

[Att84] H. Attouch, *Variational convergence of functions and operators*, Pitman Advanced Publishing Program, Pitman, 1984.

[BaC17] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in hilbert spaces*, 2nd edn. ed., Springer, 2017.

[BaK11] S. Bartels and M. Kruží k: *An efficient approach to the numerical solution of rate-independent probolems with nonconvex energies*. Multisclae Model. Simul. **9**:3 (2011) 1276–1300.

[Bog07] V. I. Bogachev, *Measure theory. volume 1*, Springer, 2007.

[Bot03] D. Bothe: *Instantaneous limits of reversible chemical reactions in presence of macroscopic convection*. J. Diff. Eqns. **193**:1 (2003) 27–48.

[Bra02] A. Braides, *Γ-convergence for beginners*, Oxford University Press, 2002.

[Bra06] ______, *A handbook of Γ-convergence*, Handbook of Differential Equations. Stationary Partial Differential Equations. Volume 3 (M. Chipot and P. Quittner, eds.), Elsevier, 2006, pp. 101–213.

[Bra14] ______, *Local minimization, variational evolution and gamma-convergence*, Lect. Notes Math. Vol. 2094, Springer, 2014.

[Bré73] H. Brézis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland Publishing Co., Amsterdam, 1973.

[CHM02] C. Carstensen, K. Hackl, and A. Mielke: *Non–convex potentials and microstructures in finite–strain plasticity*. Proc. Royal Soc. London Ser. A **458**:2018 (2002) 299–317.

[Col92] P. Colli: *On some doubly nonlinear evolution equations in Banach spaces*. Japan J. Indust. Appl. Math. **9** (1992) 181–203.

[CoT05] S. Conti and F. Theil: *Single-slip elastoplastic microstructures*. Arch. Rational Mech. Anal. **178** (2005) 125–148.

[CoV90] P. Colli and A. Visintin: *On a class of doubly nonlinear evolution equations*. Comm. Partial Differ. Eqns. **15**:5 (1990) 737–756.

[DaG87] D. A. Dawson and J. Gärtner: *Large deviations from the McKean-Vlasov limit for weakly interacting diffusions*. Stochastics **20**:4 (1987) 247–308.

[Dal93] G. Dal Maso, *An introduction to Γ-convergence*, Birkhäuser Boston Inc., Boston, MA, 1993.

[DaS14] S. Daneri and G. Savaré, *Lecture notes on gradient flows and optimal transport*, Optimal Transportation. Theory and Applications (Y. Ollivier, H. Pajot, and C. Villani, eds.), Cambridge Univ. Press, 2014, pp. 100–144.

[DaT02] G. Dal Maso and R. Toader: *A model for quasi–static growth of brittle fractures: existence and approximation results*. Arch. Rational Mech. Anal. **162** (2002) 101–135.

[DaT10] ______: *Quasistatic crack growth in elasto-plastic materials: the two-dimensional case*. Arch. Rational Mech. Anal. **196**:3 (2010) 867–906.

[DaZ07] G. Dal Maso and C. Zanini: *Quasi-static crack growth for a cohesive zone model with prescribed crack path.*. Proc. R. Soc. Edinb., Sect. A, Math. **137**:2 (2007) 253–279.

[DeF75] E. DeGiorgi and T. Franzoni: *Su un tipo di convergenza variazionale*. Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8) **58**:6 (1975) 842–850.

[DeG77] E. DeGiorgi: *Γ-convergenza e G-convergenza.*. Boll. Unione Mat. Ital., V. Ser. **A 14** (1977) 213–220.

[DeK13] A. DeSimone and M. Kružík: *Domain patterns and hysteresis in phase-transforming solids: analysis and numerical simulations of a sharp interface dissipative model via phase-field approximation*. Netw. Heterog. Media **8**:2 (2013) 481–499.

[DFM19] P. Dondl, T. Frenzel, and A. Mielke: *A gradient system with a wiggly energy and relaxed EDP-convergence*. ESAIM Control Optim. Calc. Var. **25** (2019) 68/1–45.

[DFT05] G. Dal Maso, G. Francfort, and R. Toader: *Quasistatic crack growth in nonlinear elasticity*. Arch. Rational Mech. Anal. **176** (2005) 165–225.

[DPZ13] M. H. Duong, M. A. Peletier, and J. Zimmer: *GENERIC formalism of a Vlasov-Fokker-Planck equation and connection to large-deviation principles*. Nonlinearity **26**:11 (2013) 2951–2971.

[DR*21] G. Dal Maso, R. Rossi, G. Savar, and R. Toader: *Visco-energetic solutions for a model of crack growth in brittle materials*. arXiv (2021) arXiv:2105.00046.

[Fen49] W. Fenchel: *On conjugate convex functions*. Canadian J. Math. **1** (1949) 73–77.

[FeN09] E. Feireisl and A. Novotný, *Singular limits in thermodynamics of viscous fluids*, Birkhäuser, 2009.

[FlS20] F. Fleissner and G. Savaré: *Reverse approximation of gradient flows as minimizing movements: a conjecture by De Giorgi*. Ann. Sc. Norm. Super. Pisa, Cl. Sci. (5) **20**:2 (2020) 677–720.

[Fre19] T. Frenzel, *On the derivation of effective gradient systems via EDP-convergence*, Ph.D. thesis, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät, 2019, Defense date 26.6.2019, publication date 10.6.2020, epub HU Berlin.

[FrL03] G. A. Francfort and C. J. Larsen: *Existence and convergence for quasi-static evolution of brittle fracture*. Comm. Pure Applied Math. **56** (2003) 1495–1500.

[FrL21] T. Frenzel and M. Liero: *Effective diffusion in thin structures via generalized gradient systems and EDP-convergence*. Discr. Cont. Dynam. Systems Ser. S **14**:1 (2021) 395–425.

[FrM98] G. Francfort and J.-J. Marigo: *Revisiting brittle fracture as an energy minimization problem*. J. Mech. Phys. Solids **46** (1998) 1319–1342.

[FrM21] T. Frenzel and A. Mielke: *Deriving the kinetic flux relation for nonlinear diffusion through a membrane using edp-convergence*. In preparation (2021) .

[GlM13] A. Glitzky and A. Mielke: *A gradient structure for systems coupling reaction-diffusion effects in bulk and interfaces*. Z. angew. Math. Phys. (ZAMP) **64** (2013) 29–52.

[Grm10] M. Grmela: *Why GENERIC?*. J. Non-Newtonian Fluid Mech. **165** (2010) 980–986.

[GrÖ97] M. Grmela and H. C. Öttinger: *Dynamics and thermodynamics of complex fluids. I. Development of a general formalism. II. Illustrations of a general formalism*. Phys. Rev. E (3) **56**:6 (1997) 6620–6655.

[Hah15] H. Hahn: *Über eine Verallgemeinerung der Riemannschen Inetraldefinition*. Monatshefte Math. Physik **26** (1915) 3–18.

[Jam96] R. D. James, *Hysteresis in phase transformations*, ICIAM 95 (Hamburg, 1995), Math. Res., vol. 87, Akademie Verlag, Berlin, 1996, pp. 135–154.

[JKO98] R. Jordan, D. Kinderlehrer, and F. Otto: *The variational formulation of the Fokker-Planck equation*. SIAM J. Math. Analysis **29**:1 (1998) 1–17.

[KM*19] M. Kantner, A. Mielke, M. Mittnenzweig, and N. Rotundo, *Mathematical modeling of semiconductors: from quantum mechanics to devices*, Topics in Applied Analysis and Optimisation (J. Rodrigues and M. Hintermüller, eds.), CIM Series in Mathematical Sciences, Springer, 2019, pp. 269–293.

[KnS12] D. Knees and A. Schröder: *Global spatial regularity for elasticity models with cracks, contact and other nonsmooth constraints*. Math. Methods Appl. Sci. (MMAS) **35**:15 (2012) 1859.

[KRZ13] D. Knees, R. Rossi, and C. Zanini: *A vanishing viscosity approach to a rate-independent damage model*. Math. Models Meth. Appl. Sci. (M^3AS) **23**:4 (2013) 565–616.

[LiR18] M. Liero and S. Reichelt: *Homogenization of Cahn–Hilliard-type equations via evolutionary Γ-convergence*. Nonl. Diff. Eqns. Appl. (NoDEA) **25**:1 (2018) Art. 6 (31 pp.).

[LM*17] M. Liero, A. Mielke, M. A. Peletier, and D. R. M. Renger: *On microscopic origins of generalized gradient structures*. Discr. Cont. Dynam. Systems Ser. S **10**:1 (2017) 1–35.

[MaM05] A. Mainik and A. Mielke: *Existence results for energetic models for rate–independent systems*. Calc. Var. Part. Diff. Eqns. **22** (2005) 73–99.

[MaM20] J. Maas and A. Mielke: *Modeling of chemical reaction systems with detailed balance using gradient structures*. J. Stat. Physics **181** (2020) 2257–2303.

[Mie05] A. Mielke, *Evolution in rate-independent systems (Ch. 6)*, Handbook of Differential Equations, Evolutionary Equations, vol. 2 (C. Dafermos and E. Feireisl, eds.), Elsevier B.V., Amsterdam, 2005, pp. 461–559.

[Mie11a] A. Mielke, *Differential, energetic, and metric formulations for rate-independent processes*, Nonlinear PDE's and Applications (L. Ambrosio and G. Savaré, eds.), Springer, 2011, (C.I.M.E. Summer School, Cetraro, Italy 2008, Lect. Notes Math. Vol. 2028), pp. 87–170.

[Mie11b] ______: *Formulation of thermoelastic dissipative material behavior using GENERIC*. Contin. Mech. Thermodyn. **23**:3 (2011) 233–256.

[Mie11c] ______: *A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems*. Nonlinearity **24** (2011) 1329–1346.

[Mie11d] ______: *On thermodynamically consistent models and gradient structures for thermoplasticity*. GAMM Mitt. **34**:1 (2011) 51–58.

[Mie13] ______: *Thermomechanical modeling of energy-reaction-diffusion systems, including bulk-interface interactions*. Discr. Cont. Dynam. Systems Ser. S **6**:2 (2013) 479–499.

[Mie16] ______, *On evolutionary Γ-convergence for gradient systems (Ch. 3)*, Macroscopic and Large Scale Phenomena: Coarse Graining, Mean Field Limits and Ergodicity (A. Muntean, J. Rademacher, and A. Zagaris, eds.), Lecture Notes in Applied Math. Mechanics Vol. 3, Springer, 2016, Proc. of Summer School in Twente University, June 2012, pp. 187–249.

[Mie23] ______: *Non-equilibrium steady states as saddle points and EDP-convergence for slow-fast gradient systems*. J. Math. Physics (2023) , Submitted. WIAS preprint 2998, arXiv:2303.07175.

[MiM17] M. Mittnenzweig and A. Mielke: *An entropic gradient structure for Lindblad equations and couplings of quantum systems to macroscopic models*. J. Stat. Physics **167**:2 (2017) 205–233.

[MiR15] A. Mielke and T. Roubíček, *Rate-independent systems: Theory and application*, Applied Mathematical Sciences, Vol. 193, Springer New York, 2015.

[MiR23] A. Mielke and R. Rossi: *Balanced-Viscosity solutions to infinite-dimensional multi-rate systems*. Arch. Rational Mech. Anal. (2023) , In press, arXiv:2112.01794, WIAS preprint 2902.

[MiS20] A. Mielke and A. Stephan: *Coarse graining via EDP-convergence for linear fast-slow reaction systems*. Math. Models Meth. Appl. Sci. (M^3AS) **30**:9 (2020) 1765–1807, (In the published version, Lemma 3.4 is wrong. See arXiv:1911.06234v2 for the correction.).

[MiT99] A. Mielke and F. Theil, *A mathematical model for rate-independent phase transformations with hysteresis*, Proceedings of the Workshop on "Models of Continuum Mechanics in Analysis and Engineering" (Aachen) (H.-D. Alber, R. Balean, and R. Farwig, eds.), Shaker-Verlag, 1999, pp. 117–129.

[MiT04] ______: *On rate–independent hysteresis models*. Nonl. Diff. Eqns. Appl. (NoDEA) **11** (2004) 151–189, (Accepted July 2001).

[MMP21] A. Mielke, A. Montefusco, and M. A. Peletier: *Exploring families of energy-dissipation landscapes via tilting — three types of EDP convergence*. Contin. Mech. Thermodyn. **33** (2021) 611–637.

[Mos71] U. Mosco: *Continuity of the Young-Fenchel transform*. J. Math. Anal. Appl. **35** (1971) 518–535.

[MP*17] A. Mielke, R. I. A. Patterson, M. A. Peletier, and D. R. M. Renger: *Non-equilibrium thermodynamical principles for chemical reactions with mass-action kinetics*. SIAM J. Appl. Math. **77**:4 (2017) 1562–1585.

[MPR14] A. Mielke, M. A. Peletier, and D. R. M. Renger: *On the relation between gradient flows and the large-deviation principle, with applications to Markov chains and diffusion*. Potential Analysis **41**:4 (2014) 1293–1327.

[MPS21] A. Mielke, M. A. Peletier, and A. Stephan: *EDP-convergence for nonlinear fast-slow reaction systems with detailed balance*. Nonlinearity **34**:8 (2021) 5762–5798.

[MRS08] A. Mielke, T. Roubíček, and U. Stefanelli: *Γ-limits and relaxations for rate-independent evolutionary problems*. Calc. Var. Part. Diff. Eqns. **31** (2008) 387–416.

[MRS13] A. Mielke, R. Rossi, and G. Savaré: *Nonsmooth analysis of doubly nonlinear evolution equations*. Calc. Var. Part. Diff. Eqns. **46**:1-2 (2013) 253–310.

[MRS22] A. Mielke, R. Rossi, and A. Stephan: *Split-step algorithm for gradient systems with two dissipation potentials*. In preparation (2022) .

[MTL02] A. Mielke, F. Theil, and V. I. Levitas: *A variational formulation of rate–independent phase transformations using an extremum principle*. Arch. Rational Mech. Anal. **162** (2002) 137–177.

[MuS20] M. Muratori and G. Savaré: *Gradient flows and evolution variational inequalities in metric spaces. I: structural properties*. J. Funct. Analysis **278**:4 (2020) 108347/1–67.

[MuS22] ______: *Gradient flows and evolution variational inequalities in metric spaces. II: variational convergence and III: generation results*. In preparation (2022) .

[OnM53] L. Onsager and S. Machlup: *Fluctuations and irreversible processes*. Phys. Rev. **91**:6 (1953) 1505–1512.

[Ons31] L. Onsager: *Reciprocal relations in irreversible processes, I+II*. Physical Review **37** (1931) 405–426, (part II, 38:2265–2279).

[OrR99] M. Ortiz and E. Repetto: *Nonconvex energy minimization and dislocation structures in ductile single crystals*. J. Mech. Phys. Solids **47**:2 (1999) 397–462.

[Ott96] F. Otto, *Double degenerate diffusion equations as steepest descent*, Preprint no. 480, SFB 256, University of Bonn, 1996.

[Ott01] ______: *The geometry of dissipative evolution equations: the porous medium equation*. Comm. Partial Diff. Eqns. **26** (2001) 101–174.

[Ött05] H. C. Öttinger, *Beyond equilibrium thermodynamics*, John Wiley, New Jersey, 2005.

[Pel14] M. A. Peletier, *Variational modelling: Energies, gradient flows, and large deviations*, arXiv:1402.1990, 2014.

[PeS22] M. A. Peletier and A. Schlichting: *Cosh gradient systems and tilting*. Preprint (2022) , arXiv:2203.05435.

[PR*22] M. A. Peletier, R. Rossi, G. Savaré, and O. Tse: *Jump processes as generalized gradient flows*. Calc. Var. Part. Diff. Eqns. **61**:1 (2022) 33/1–85.

[Pra28] L. Prandtl: *Gedankenmodel zur kinetischen Theorie der festen Körper*. Z. angew. Math. Mech. (ZAMM) **8** (1928) 85–106.

[RoS06] R. Rossi and G. Savaré: *Gradient flows of non convex functionals in Hilbert spaces and applications*. ESAIM Control Optim. Calc. Var. **12** (2006) 564–614.

[San17] F. Santambrogio: {*Euclidean, metric, Wasserstein*} *gradient flows: an overview*. Bull. Math. Sci. **7**:1 (2017) 87–154.

[SaS04] E. Sandier and S. Serfaty: *Gamma-convergence of gradient flows with applications to Ginzburg-Landau*. Comm. Pure Appl. Math. **LVII** (2004) 1627–1672.

[Sav07] G. Savaré: *Gradient flows and diffusion semigroups in metric spaces under lower curvature bounds*. C. R. Math. Acad. Sci. Paris **345**:3 (2007) 151–154.

[Ser11] S. Serfaty: *Gamma-convergence of gradient flows on Hilbert spaces and metric spaces and applications*. Discr. Cont. Dynam. Systems Ser. A **31**:4 (2011) 1427–1451.

[SSZ12] G. Schimperna, A. Segatti, and S. Zelik: *Asymptotic uniform boundedness of energy solutions to the Penrose-Fife model*. J. Evol. Equ. **12** (2012) 863–890.

[Ste21] A. Stephan: *EDP-convergence for a linear reaction-diffusion system with fast reversible reaction*. Calc. Var. Part. Diff. Eqns. **60**:6 (2021) 226/35 pp.

[Ste22] U. Stefanelli: *A new minimizing-movement scheme for curves of maximal slope*. ESAIM Control Optim. Calc. Var. **28**:59 (2022) 1–29.

[Tho10] M. Thomas, *Rate-independent damage processes in nonlinearly elastic materials*, Ph.D. thesis, Institut für Mathematik, Humboldt-Universität zu Berlin, February 2010.

[Tom29] G. A. Tomlinson: *A molecular theory of friction*. Philos. Mag. **7** (1929) 905–939.

[Vil09] C. Villani, *Optimal transport. Old and new*, Berlin: Springer, 2009.

[Vis84] A. Visintin: *Strong convergence results related to strict convexity*. Comm. Partial Diff. Eqns. **9**:5 (1984) 439–466.

[Yon08] W.-A. Yong: *An interesting class of partial differential equations*. J. Math. Phys. **49** (2008) 033503, 21.