An Introduction to Gradient Flows

Oliver Tse

Eindhoven University of Technology

Updated on August 15, 2024

Lecture notes from a course given at uOttawa in Summer 2023

Abstract. *Gradient flows*, or *curves of maximal slope*, are fundamental objects in the modeling of dissipative evolution equations and can take many forms depending on the state spaces on which they are modeled. Yet, they all have one thing in common: they dissipate some energy via some dissipative mechanism. This lecture introduces variational formulations of gradient flows in various scenarios leading up to gradient flows in metric spaces, including the important example of the Wasserstein space.

Contents

1 Gradient flows: The finite-dimensional case

1.1 Gradient flows in Euclidean space

We begin with introducing gradient flows in $\mathbb{R}^d$, $d \geq 1$, which remains true also for arbitrary Hilbert spaces but we restrict ourselves to finite-dimensional ones at this point.

Let $\mathcal{F}\colon \mathbb{R}^d \to \mathbb{R}$ be a sufficiently smooth function—henceforth called a *driving functional*—and $\mathbf{x}_0 \in \mathbb{R}^d$ be a given point. An $\mathbb{R}^d$-gradient flow is a curve $\mathbf{x} : [0,\infty) \to \mathbb{R}^d$ with starting point $\mathbf{x}_0$ that minimizes the driving functional $\mathcal{F}$ as quickly as possible. More precisely,

Definition 1.1 (Gradient flows in $\mathbb{R}^d$) The *gradient flow* of a driving functional $\mathcal{F} : \mathbb{R}^d \to \mathbb{R}$ is the family of maps $\mathsf{S}_t\colon \mathbb{R}^d \to \mathbb{R}^d$, $t \in [0,\infty)$ characterized by the following properties:

(a) For every $\mathbf{x}_0 \in \mathbb{R}^d$, $\mathsf{S}_0(\mathbf{x}_0) = \mathbf{x}_0$;

(b) The curve $\mathbf{x}_t \coloneqq \mathsf{S}_t(\mathbf{x}_0)$, $t \in (0,\infty)$ is the (unique) solution to the *Cauchy problem*

$$\dot{\mathbf{x}}(t) = -\nabla\mathcal{F}(\mathbf{x}(t)) \quad \text{for } t > 0, \qquad \mathbf{x}(0) = \mathbf{x}_0. \tag{1.1}$$

Here, we distinguish between the gradient $\nabla\mathcal{F}$ and the Fréchet derivative $\mathrm{D}\mathcal{F}$ via

$$\mathrm{D}\mathcal{F}(\mathbf{x}) = (\partial_1\mathcal{F}(\mathbf{x}),\ldots,\partial_d\mathcal{F}(\mathbf{x})) \in (\mathbb{R}^d)^*, \qquad \nabla\mathcal{F}(\mathbf{x}) = (\mathrm{D}\mathcal{F}(\mathbf{x}))^\top \in \mathbb{R}^d.$$

In literature, the quantity $\dot{\mathbf{x}}$ is the *velocity* of the curve $\mathbf{x}$ and $-\nabla\mathcal{F}(\mathbf{x})$ is commonly known as the *restoring force* of the driving functional $\mathcal{F}$ at $\mathbf{x}$. Note that to describe the gradient flow equation (1.1), we require several ingredients:

(1) A *state space* $\mathfrak{X}$, which in this case is $\mathbb{R}^d$;

(2) A *driving functional* $\mathcal{F} : \mathfrak{X} \to \mathbb{R} \cup \{+\infty\}$, i.e., possibly taking the value $+\infty$;

(3) A notion of *temporal derivative* $\dot{\mathbf{x}}$, a notion of *gradient* $\nabla\mathcal{F}$;

(4) A map between the restoring force $-\nabla\mathcal{F}$ and velocity $\dot{\mathbf{x}}$. This map is called a *kinetic relation* (or *force-to-velocity* map) and encodes *frictional* properties of the system.

Remark 1.2 Recall that a unique global solution to (1.1) exists if $\nabla\mathcal{F}$ is Lipschitz continuous due to the Picard–Lindelöf theorem (or Cauchy–Lipschitz theorem). We will see below that the assumption can be relaxed due to the variational structure of (1.1).

Example 1.3 Consider the ODE

$$\dot{\mathbf{x}}(t) = V(\mathbf{x}(t)), \qquad V(\mathbf{x}) = (-x_1, x_2 - \alpha x_2^3)^\top, \quad \alpha > 0.$$

Setting $\mathcal{F}(\mathbf{x}) \coloneqq \frac{1}{2}(x_1^2 + x_2^2) + \frac{\alpha}{4}x_2^4$, we find $\nabla\mathcal{F}(\mathbf{x}) = (x_1, x_2 + \alpha x_2^3)^\top = -V(\mathbf{x})$, and hence

$$\dot{\mathbf{x}}(t) = -\nabla\mathcal{F}(\mathbf{x}(t)),$$

i.e. the ODE can be expressed as a gradient flow equation.

Example 1.4 The *damped harmonic oscillator*, i.e.

$$\dot{\mathbf{x}}(t) = V(\mathbf{x}(t)), \qquad V(\mathbf{x}) = (x_2, -x_1 - \gamma x_2)^\top,$$

does not possess a gradient structure, but does have a more general structure—the so-called *GENERIC* structure—which is currently a field of active research.

Applying the chain rule, we obtain the *utmost* important property of a gradient flow:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(\mathbf{x}(t)) &= \mathrm{D}\mathcal{F}(\mathbf{x}(t))[\dot{\mathbf{x}}(t)] = \nabla\mathcal{F}(\mathbf{x}(t)) \cdot \dot{\mathbf{x}}(t) \\ &= -|\dot{\mathbf{x}}(t)|^2 = -|\nabla\mathcal{F}(\mathbf{x}(t))|^2 = -|\nabla\mathcal{F}(\mathbf{x}(t))||\dot{\mathbf{x}}(t)| \\ &= -\frac{1}{2}|\dot{\mathbf{x}}(t)|^2 - \frac{1}{2}|\nabla\mathcal{F}(\mathbf{x}(t))|^2 \le 0,\end{aligned}$$

where $\nabla\mathcal{F}(\mathbf{x}) = 0$ if and only if $\mathbf{x}$ is a stationary point of $\mathcal{F}$. In particular, $t \mapsto \mathcal{F}(\mathbf{x}(t))$ is a strictly decreasing function until $\mathbf{x}$ hits a stationary point. The final form on the right-hand side of the previous equation will play a fundamental role in the study of gradient flows and will provide us with the means to formulate (1.1) as a variational problem.

Variational formulation From the previous equalities, one deduces that a solution of (1.1) can be characterized by the following scalar conditions for every $t > 0$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(\mathbf{x}(t)) = -|\nabla\mathcal{F}(\mathbf{x}(t))||\dot{\mathbf{x}}(t)|, \tag{1.2}$$

$$|\dot{\mathbf{x}}(t)| = |\nabla\mathcal{F}(\mathbf{x}(t))|, \tag{1.3}$$

where (1.2) forces the direction of the velocity $\dot{\mathbf{x}}$ to be opposite to that of the gradient $\nabla\mathcal{F}(\mathbf{x})$, while the modulus of $\dot{\mathbf{x}}$ is determined by (1.3). In fact, conditions (1.2) and (1.3) are equivalent, via Young's inequality[1], to the single equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(\mathbf{x}(t)) = -\frac{1}{2}|\dot{\mathbf{x}}(t)|^2 - \frac{1}{2}|\nabla\mathcal{F}(\mathbf{x}(t))|^2 \qquad \text{for every } t > 0. \tag{1.4}$$

While (1.1) only makes sense in the Hilbertian setting, the formulations (1.2,1.3) and (1.4) are of purely metric nature and can be extended to more general metric spaces provided we understand $|\dot{\mathbf{x}}|$ as the *metric speed* of a curve $\mathbf{x}$ and $|\nabla\mathcal{F}(\mathbf{x})|$ as the *metric slope*.

Additionally, (1.4) gives us a way to formulate (1.1) as a variational problem. Integrating (1.4) over any interval $[s,t] \subset [0,\infty)$, we obtain the so-called *energy-dissipation balance*

$$\mathscr{L}(\mathbf{x},[s,t]) := \int_s^t \left\{\frac{1}{2}|\dot{\mathbf{x}}(r)|^2 + \frac{1}{2}|\nabla\mathcal{F}(\mathbf{x}(r))|^2\right\}\mathrm{d}r + \mathcal{F}(\mathbf{x}(t)) - \mathcal{F}(\mathbf{x}(s)) = 0, \tag{1.5}$$

[1] *Young's inequality:* For any $a,b \in \mathbb{R}^d$, $\langle a,b\rangle \le \frac{1}{2}|a|^2 + \frac{1}{2}|b|^2$. Equality holds if and only if $a = b$.

where $\mathscr{L}$ is called the *energy-dissipation functional*. Clearly, any solution $\mathbf{x}$ of (1.1) will give $\mathscr{L}(\mathbf{x}, [s,t]) = 0$ for every $[s,t] \subset [0,\infty)$ via (1.4). Conversely, if $\mathbf{x}$ satisfies (1.5) and the chain rule for $\mathcal{F}$ along the curve $\mathbf{x}$ holds, i.e.

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(\mathbf{x}(t)) = \nabla\mathcal{F}(\mathbf{x}(t)) \cdot \dot{\mathbf{x}}(t) \qquad \text{for every } t > 0, \tag{1.6}$$

then $\mathscr{L}(\mathbf{x}, [s,t])$ can be expressed as

$$\mathscr{L}(\mathbf{x}, [s,t]) = \int_s^t \left\{ \frac{1}{2}|\dot{\mathbf{x}}(r)|^2 + \frac{1}{2}|-\nabla\mathcal{F}(\mathbf{x}(r))|^2 - (-\nabla\mathcal{F}(\mathbf{x}(r))) \cdot \dot{\mathbf{x}}(r) \right\} \mathrm{d}r = 0.$$

Due to Young's inequality, the integrand in the identity above is nonnegative, and thus

$$\frac{1}{2}|\dot{\mathbf{x}}(r)|^2 + \frac{1}{2}|-\nabla\mathcal{F}(\mathbf{x}(r))|^2 - (-\nabla\mathcal{F}(\mathbf{x}(r))) \cdot \dot{\mathbf{x}}(r) = 0 \quad \Longleftrightarrow \quad \dot{\mathbf{x}}(r) = -\nabla\mathcal{F}(\mathbf{x}(r)),$$

for almost every $r \in [0,\infty)$ (since $[s,t]$ was arbitrary). Since the map $t \mapsto -\nabla\mathcal{F}(\mathbf{x}(t))$ is continuous, the equality holds true also for all $r \in [0,\infty)$, i.e. $\mathbf{x}$ is a solution of (1.1).

More importantly, the observation that $\mathscr{L}(\mathbf{x}, [s,t]) \geq 0$ for every curve $\mathbf{x}$ satisfying the chain rule (1.6) leads us to a variational characterization of gradient flows:

$$\mathbf{x} \text{ solves } (1.1) \quad \Longleftrightarrow \quad \mathbf{x} \text{ is a minimizer of the } \mathscr{L}(\,\cdot\,, [s,t]) \text{ for every } [s,t] \subset [0,\infty).$$

In particular, this means that the Cauchy problem (1.1) can be analyzed using tools from the Calculus of Variations, e.g. the direct method, Γ-convergence, etc, which brings us to an alternative definition of a gradient flow:

Definition 1.5 The gradient flow of a driving functional $\mathcal{F} : \mathbb{R}^d \to \mathbb{R}$ is the family of maps $\mathsf{S}_t \colon \mathbb{R}^d \to \mathbb{R}^d$, $t \in [0,\infty)$ characterized by the following properties:

(a) For every $\mathbf{x}_0 \in \mathbb{R}^d$, $\mathsf{S}_0(\mathbf{x}_0) = \mathbf{x}_0$;

(b) The curve $\mathbf{x}_t := \mathsf{S}_t(\mathbf{x}_0)$, $t \in (0,\infty)$ satisfies the energy-dissipation balance

$$\mathscr{L}(\mathbf{x}, [s,t]) = 0 \qquad \text{for all } [s,t] \subset [0,\infty),$$

or equivalently, the *energy-dissipation principle*

$$\mathbf{x} \in \operatorname{argmin}\Big\{ \mathscr{L}(\mathbf{y}, [s,t]) : \mathbf{y} \in \mathcal{C}([0,\infty); \mathbb{R}^d) \Big\} \qquad \text{for all } [s,t] \subset [0,\infty).$$

Stability estimates While the variational formulation provides an alternative approach to proving the existence of a solution to (1.1), it does *not* readily provide the uniqueness of solutions unless $\mathscr{L}$ can be shown to be strictly convex—a nontrivial endeavour. Instead, we can often rely on stability estimates.

Assuming that $\nabla\mathcal{F}$ is Lipschitz continuous with Lipschitz constant $c_{\mathcal{F}} > 0$, and considering two solutions $\mathbf{x}$, $\mathbf{y}$ with initial data $\mathbf{x}_0$, $\mathbf{y}_0$ respectively, we obtain

$$\begin{aligned}
\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}|\mathbf{y}(t) - \mathbf{x}(t)|^2 &= (\mathbf{y}(t) - \mathbf{x}(t)) \cdot (\dot{\mathbf{y}}(t) - \dot{\mathbf{x}}(t)) \\
&= -(\mathbf{y}(t) - \mathbf{x}(t)) \cdot (\nabla\mathcal{F}(\mathbf{y}(t)) - \nabla\mathcal{F}(\mathbf{x}(t))) \\
&\leq c_{\mathcal{F}}|\mathbf{y}(t) - \mathbf{x}(t)|^2.
\end{aligned}$$

An application of Gronwall's inequality then yields the stability estimate

$$|\mathbf{y}(t) - \mathbf{x}(t)| \le e^{c_{\mathcal{F}} t} |\mathbf{y}_0 - \mathbf{x}_0| \qquad \text{for all } t \ge 0,$$

which consequently provides the uniqueness of solutions to (1.1).

It turns out that the computation above can be done also for less regular driving functionals $\mathcal{F}$, and in certain cases (e.g. when $\mathcal{F}$ is convex) can be used to provide yet another variational characterization of gradient flows $\rightsquigarrow$ *Evolution Variational Inequalities.*

1.2 Gradient flows in Riemannian manifolds

A more general perspective of (1.1) is necessary when the state space $\mathfrak{X}$ is a Riemannian manifold, i.e. $\mathfrak{X} = (M, g)$ where M is a smooth manifold with a symmetric and coercive family of bilinear forms $g_{\mathbf{x}} : T_{\mathbf{x}}M \times T_{\mathbf{x}}M \to \mathbb{R}$, $\mathbf{x} \in M$ (called the *metric tensor*), where $T_{\mathbf{x}}M$ is the tangent space of M at $\mathbf{x}$. For a curve $\gamma \colon [a, b] \to M$, its length is given by

$$\mathrm{Length}_g^{[a,b]}(\gamma) \coloneqq \int_a^b \sqrt{g_{\gamma(t)}(\dot\gamma(t), \dot\gamma(t))}\, \mathrm{d}t.$$

Since $g_{\mathbf{x}}$ is symmetric and coercive for each $\mathbf{x} \in M$, it defines a symmetric and positive definite linear map $\mathbb{G}(\mathbf{x}) \colon T_{\mathbf{x}}M \to T^*_{\mathbf{x}}M$ via $\langle \mathbb{G}(\mathbf{x})u, v\rangle \coloneqq g_{\mathbf{x}}(u, v)$, $u, v \in T_{\mathbf{x}}M$, where $T^*_{\mathbf{x}}M$ is the cotangent space of M at $\mathbf{x}$, and $\langle \cdot, \cdot \rangle$ denotes the duality pairing between the tangent space $T_{\mathbf{x}}M$ at $x \in M$ and its dual $T^*_{\mathbf{x}}M$.

Given a driving functional $\mathcal{F} : M \to \mathbb{R}$, the differential $\mathrm{D}\mathcal{F}(\mathbf{x})$ of $\mathcal{F}$ at $\mathbf{x}$ is defined by

$$\mathrm{D}\mathcal{F}(\mathbf{x})[v] \coloneqq \lim_{t \to 0} \frac{1}{h} \Big\{ \mathcal{F}(\gamma(t)) - \mathcal{F}(\gamma(0)) \Big\}, \qquad \forall v \in T_{\mathbf{x}}M,$$

where γ is the (unique) *geodesic curve*—curve of minimal length—emitting from $\gamma(0) = \mathbf{x}$ with initial tangent vector $\dot\gamma(0) = v$. In particular, $\mathrm{D}\mathcal{F}(\mathbf{x}) \in T^*_{\mathbf{x}}M$ is an element of the cotangent space $T^*_{\mathbf{x}}M$. The invertibility of $\mathbb{G}(\mathbf{x})$ then allows us to uniquely associate a cotangent vector with a tangent vector.

Definition 1.6 (Gradient) The gradient of a function $\mathcal{F} : M \to \mathbb{R}$ on a Riemannian manifold (M, g) is defined via

$$\nabla_g \mathcal{F}(\mathbf{x}) \coloneqq \mathbb{K}(\mathbf{x}) \mathrm{D}\mathcal{F}(\mathbf{x}),$$

where $\mathbb{K}(\mathbf{x}) \coloneqq \mathbb{G}(\mathbf{x})^{-1} \colon T_{\mathbf{x}}M^* \to T_{\mathbf{x}}M$ is called the *Onsager operator.*

The gradient flow (1.1) in this setting then reads

$$\dot{\mathbf{x}}(t) = -\nabla_g \mathcal{F}(\mathbf{x}) = -\mathbb{K}(\mathbf{x}(t))\mathrm{D}\mathcal{F}(\mathbf{x}(t)) \in T_{\mathbf{x}(t)}M, \tag{1.7}$$

or equivalently,

$$0 = \mathbb{G}(\mathbf{x}(t))\dot{\mathbf{x}}(t) + \mathrm{D}\mathcal{F}(\mathbf{x}(t)) \in T^*_{\mathbf{x}(t)}M.$$

A gradient flow in (M, g) is then defined similarly to the $\mathbb{R}^d$ case.

Definition 1.7 (Gradient flows in Riemannian manifolds) The gradient flow associated with a *gradient structure* $(M, \mathcal{F}, \mathbb{G})$ is the family of maps $\mathsf{S}_t \colon M \to M$, $t \in [0, \infty)$ characterized by the following properties:

(a) For every $\mathbf{x}_0 \in M$, $\mathsf{S}_0(\mathbf{x}_0) = \mathbf{x}_0$;

(b) The curve $\mathbf{x}_t \coloneqq \mathsf{S}_t(\mathbf{x}_0)$, $t \in (0, \infty)$ is the (unique) solution to the Cauchy problem

$$\dot{\mathbf{x}}(t) = -\mathbb{K}(\mathbf{x}(t))\mathrm{D}\mathcal{F}(\mathbf{x}(t)) \in T_{\mathbf{x}(t)}M \quad \text{for } t > 0, \qquad \mathbf{x}(0) = \mathbf{x}_0. \tag{1.8}$$

As before, one applies the chain rule to obtain the gradient flow property

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(\mathbf{x}(t)) &= \mathrm{D}\mathcal{F}(\mathbf{x}(t))[\dot{\mathbf{x}}(t)] = -\mathrm{D}\mathcal{F}(\mathbf{x}(t))[\nabla_g \mathcal{F}(\mathbf{x}(t))] \\
&= -\langle \mathbb{G}(\mathbf{x}(t))\dot{\mathbf{x}}(t), \dot{\mathbf{x}}(t)\rangle =: |\dot{\mathbf{x}}(t)|^2_{\mathbb{G}(\mathbf{x}(t))} \\
&= -\langle \mathrm{D}\mathcal{F}(\mathbf{x}(t)), \mathbb{K}(\mathbf{x}(t))\mathrm{D}\mathcal{F}(\mathbf{x}(t))\rangle =: |\mathrm{D}\mathcal{F}(\mathbf{x}(t))|^2_{\mathbb{K}(\mathbf{x}(t))} \\
&= -\frac{1}{2}|\dot{\mathbf{x}}(t)|^2_{\mathbb{G}(\mathbf{x}(t))} - \frac{1}{2}|\mathrm{D}\mathcal{F}(\mathbf{x}(t))|^2_{\mathbb{K}(\mathbf{x}(t))} \leq 0.
\end{aligned}$$

Remark 1.8 The symmetry and positive definiteness of $\mathbb{G}$ are fundamental from the point of view of thermodynamics, as was shown by Lars Onsager in his work on *"Reciprocal relations in irreversible processes,"* (1931), which won him the Nobel prize in 1963. His *"reciprocal relations"* were derived in the context of linearized irreversible thermodynamics and simply mean, in modern language, the symmetry relation $\mathbb{G}^\top = \mathbb{G}$.

Note that an equation $\dot{\mathbf{x}}(t) = V(\mathbf{x}(t))$ may have multiple gradient structures or none at all. If it has at least one, then the question remains: *Which one do we take?*

Example 1.9 Let $M = \mathbb{R}^2$ and as in Example 1.3, consider

$$\dot{\mathbf{x}}(t) = V(\mathbf{x}(t)), \qquad V(\mathbf{x}) = (-x_1, x_2 - \alpha x_2^3)^\top, \quad \alpha > 0.$$

Then we know from the previous section that we have a gradient structure

$$\mathbb{G}(\mathbf{x}) = I_d, \qquad \mathcal{F}(\mathbf{x}) = \frac{1}{2}(x_1^2 + x_2^2) + \frac{\alpha}{4}x_2^4.$$

Alternatively, we could consider the gradient structure

$$\widetilde{\mathbb{G}}(\mathbf{x}) = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{1+\alpha x_2^2} \end{bmatrix}, \qquad \widetilde{\mathcal{F}}(\mathbf{x}) = \frac{1}{2}(x_1^2 + x_2^2).$$

Therefore, when looking at the ODE, one does not know whether the nonlinear term $-\alpha x_2^3$, $\alpha > 0$, arises because of a non-quadratic energy $\mathcal{F}$ or because of a state-dependent friction law $\widetilde{\mathbb{G}}$. The choice of a gradient structure is thus a choice of modelling and usually contains *more* information than the ODE itself.

Example 1.10 (Reaction-rate equations) Consider three chemical species X_i, $i = 1, 2, 3$ with concentrations $\mathbf{c} = (c_1, c_2, c_3)$, respectively, which lives in the manifold $M = (0, \infty)^3$. If the reactions

$$X_1 \rightleftharpoons X_2, \qquad X_1 + X_2 \rightleftharpoons 2X_3, \qquad a_1 X_1 + a_3 X_3 \rightleftharpoons b_2 X_2,$$

occur following the mass-action law, i.e. the reaction rates are proportional to the corresponding monomials, then the ODE reads $\dot{\mathbf{c}}(t) = V(\mathbf{c}(t))$ with

$$V(\mathbf{c}) = k_1(c_1 - c_2)\begin{pmatrix}-1\\1\\0\end{pmatrix} + k_2(c_1c_2 - c_3^2)\begin{pmatrix}-1\\-1\\2\end{pmatrix} + k_3(c_1^{a_1}c_3^{a_3} - c_2^{b_2})\begin{pmatrix}-a_1\\b_2\\-a_3\end{pmatrix}.$$

The equation above has a gradient structure with

$$\mathcal{F}(\mathbf{c}) = \sum_{i=1}^{3}\lambda_B(c_i) \quad \text{and} \quad \mathbb{K}(\mathbf{c}) = \sum_{i=1}^{3} k_i \Lambda_{\log}(\mathbf{c}^{\boldsymbol{\alpha}^i}, \mathbf{c}^{\boldsymbol{\beta}^i})(\boldsymbol{\alpha}^i - \boldsymbol{\beta}^i)\otimes(\boldsymbol{\alpha}^i - \boldsymbol{\beta}^i),$$

where $\lambda_B(c) = c\log c - c + 1$ is the Boltzmann entropy function and

$$\Lambda_{\log}(a,b) = \frac{a-b}{\log a - \log b} \quad \text{is the logarithmic mean of } a \text{ and } b.$$

The *stoichiometric* vectors $\boldsymbol{\alpha}^i, \boldsymbol{\beta}^i \in \mathbb{N}_0^3$ are given by

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

In many situations, it is easier to prescribe the Onsager operator $\mathbb{K}$ than the metric tensor $\mathbb{G}$. This example is one such situation, where we also see the additive structure of the individual driving functionals.

Variational formulation As in the previous section, we would like to formulate the Cauchy problem (1.8) as a variational problem. To do so, we will need a generalization of Young's inequality for general convex functions.

Definition 1.11 (Legendre-Fenchel conjugate) Let $\Psi\colon \mathfrak{Z} \to \mathbb{R}\cup\{+\infty\}$ be a proper[2] functional on a Banach space $\mathfrak{Z}$. The *Legendre-Fenchel conjugate* Ψ^* of Ψ is defined by

$$\mathfrak{Z}^* \ni \xi \mapsto \Psi^*(\xi) := \sup\Big\{\langle \xi, \mathbf{s}\rangle - \Psi(\mathbf{s}) : \mathbf{s}\in\mathfrak{Z}\Big\}.$$

If Ψ is convex and lower semicontinuous, the pair (Ψ, Ψ^*) is called a *conjugate pair.*

[2] *proper* simply means that the domain of Ψ, i.e. $\mathrm{dom}\Psi := \{\mathbf{s}\in\mathfrak{Z} : \Psi(\mathbf{s}) < \infty\}$, is nonempty.

The *subdifferential* $\partial\Psi$ of a convex function $\Psi : \mathfrak{Z} \to \mathbb{R} \cup \{+\infty\}$ is given by

$$\partial\Psi(\mathbf{s}) := \Big\{\xi \in \mathfrak{X}^* : \Psi(\mathbf{r}) \geq \Psi(\mathbf{s}) + \langle \xi, \mathbf{r} - \mathbf{s}\rangle \quad \forall\, \mathbf{r} \in \mathfrak{Z}\Big\} \subset \mathfrak{Z}^*, \qquad \mathbf{s} \in \mathrm{dom}\Psi,$$

and $\partial\Psi(\mathbf{s}) = \emptyset$ otherwise. If Ψ is differentiable at $\mathbf{s}$, then $\partial\Psi(\mathbf{s}) = \{D\Psi(\mathbf{s})\}$. In fact, Ψ is differentiable at $\mathbf{s}$ if and only if $\partial\Psi(\mathbf{s})$ is a singleton.

Remark 1.12 (a) Ψ^* is convex and lower semicontinuous, even if Ψ is not. *(Exercise)*

(b) A trivial consequence of the definition of Ψ^* is the *Fenchel–Young inequality*

$$\Psi(\mathbf{s}) + \Psi^*(\xi) \geq \langle \xi, \mathbf{s}\rangle \qquad \forall\, (\mathbf{s}, \xi) \in \mathfrak{Z} \times \mathfrak{Z}^*. \tag{1.9}$$

(c) *Involution property:* If $\mathfrak{X}$ is reflexive and Ψ is convex and lower semicontinuous, then $\Psi^{**} = (\Psi^*)^* = \Psi$.

Proposition 1.13. *Let $\mathfrak{Z}$ be a reflexive Banach space and (Ψ, Ψ^*) be a conjugate pair of proper, convex and lower semicontinuous functionals. Then, for every $(\mathbf{s}, \xi) \in \mathfrak{Z} \times \mathfrak{Z}^*$, the following statements are equivalent:*

(i) $\mathbf{s}$ *minimizes the functional* $\mathbf{r} \mapsto \Psi(\mathbf{r}) - \langle \xi, \mathbf{r}\rangle$ *(optimality of* $\mathbf{s} \in \mathfrak{X}$*)*

(ii) $\xi \in \partial\Psi(\mathbf{s})$ *(subdifferential inclusion in* $\mathfrak{X}^*$*)*

(iii) $\Psi(\mathbf{s}) + \Psi^*(\xi) = \langle \xi, \mathbf{s}\rangle$ *(optimality condition in* $\mathbb{R}$*)*

(iv) $\mathbf{s} \in \partial\Psi^*(\xi)$ *(subdifferential inclusion in* $\mathfrak{X}$*)*

(v) ξ *maximizes the functional* $\eta \mapsto \langle \eta, \mathbf{s}\rangle - \Psi^*(\eta)$ *(optimality of* ξ *in* $\mathfrak{X}^*$*)*

Note that "$=$" in (iii) may be replaced with "$\leq$" instead since the Young–Fenchel inequality always gives "$\geq$".

Proof. *(i) ⇒ (ii):* Let $\mathbf{s}$ be a minimizer of $\mathbf{r} \mapsto \Psi(\mathbf{r}) - \langle \xi, \mathbf{r}\rangle$, then

$$\Psi(\mathbf{r}) \geq \Psi(\mathbf{s}) + \langle \xi, \mathbf{r} - \mathbf{s}\rangle \quad \forall\, \mathbf{r} \in \mathfrak{Z},$$

and hence, $\xi \in \partial\Psi(\mathbf{s})$.

(ii) ⇒ (iii): Let $\xi \in \partial\Psi(\mathbf{s})$, i.e., $\Psi(\mathbf{r}) \geq \Psi(\mathbf{s}) + \langle \xi, \mathbf{r} - \mathbf{s}\rangle$ for all $\mathbf{r} \in \mathfrak{Z}$. Then

$$\langle \xi, \mathbf{r}\rangle - \Psi(\mathbf{r}) \leq \langle \xi, \mathbf{s}\rangle - \Psi(\mathbf{s}) \quad \forall\, \mathbf{r} \in \mathfrak{Z}.$$

Taking the supremum over $\mathbf{r}$ then yields, $\Psi^*(\xi) \leq \langle \xi, \mathbf{s}\rangle - \Psi(\mathbf{s})$, which yields the equality due to the Fenchel–Young inequality (1.9).

(iii) ⇒ (i): Let $\Psi(\mathbf{s}) + \Psi^*(\xi) = \langle \xi, \mathbf{s}\rangle$. Then, $\langle \xi, \mathbf{s}\rangle - \Psi(\mathbf{s}) \geq \langle \xi, \mathbf{r}\rangle - \Psi(\mathbf{r})$ for all $\mathbf{r} \in \mathfrak{Z}$ by the definition of $\Psi^*(\xi)$. Rearranging the terms give $\Psi(\mathbf{s}) - \langle \xi, \mathbf{s}\rangle \leq \Psi(\mathbf{r}) - \langle \xi, \mathbf{r}\rangle$ for all $\mathbf{r} \in \mathfrak{Z}$, which shows that $\mathbf{s}$ is a minimizer.

The implications *(iii) ⇒ (iv) ⇒ (v) ⇒ (iii)* can be shown in a similar fashion. □

Example 1.14 (a) The function $\mathbb{R} \ni \mathbf{s} \mapsto |\mathbf{s}|$ is convex with Legendre–Fenchel conjugate $|\cdot|^*(\xi) = 0$ if $|\xi| \leq 1$ and $+\infty$ otherwise. The subdifferential of $|\cdot|$ is given by

$$(\partial|\cdot|)(\mathbf{s}) = \begin{cases} \{-1\} & \text{if } \mathbf{s} < 0, \\ [-1, 1] & \text{if } \mathbf{s} = 0, \\ \{1\} & \text{if } \mathbf{s} > 0. \end{cases}$$

(b) On a Hilbert space $\mathfrak{Z}$ with $\mathbb{G}\colon \mathfrak{Z} \to \mathfrak{Z}^*$ symmetric and positive definite,

$$\Psi(\mathbf{s}) = \frac{1}{2}\langle \mathbb{G}\mathbf{s}, \mathbf{s}\rangle \quad \iff \quad \Psi^*(\xi) = \frac{1}{2}\langle \xi, \mathbb{K}\xi\rangle, \quad \mathbb{K} = (\mathbb{G})^{-1}.$$

Since $\Psi(\mathbf{s})$ and $\Psi^*(\xi)$ are differentiable with derivatives $\mathbb{G}\mathbf{s}$ and $\mathbb{K}\xi$ respectively, we have that $\partial\Psi(\mathbf{s}) = \{\mathbb{G}\mathbf{s}\}$ and $\partial\Psi^*(\xi) = \{\mathbb{K}\xi\}$.

(c) Let $\mathfrak{Z}$ be a Banach space with norm $\|\cdot\|_{\mathfrak{Z}}$. Its Legendre-Fenchel conjugate is

$$\|\cdot\|_{\mathfrak{Z}}^*(\xi) = \begin{cases} 0 & \text{if } \|\xi\|_{\mathfrak{Z}^*} \leq 1, \\ +\infty & \text{otherwise.} \end{cases}$$

Consequently, we have that

$$(\partial\|\cdot\|_{\mathfrak{Z}})(\mathbf{s}) = \Big\{\xi \in \mathfrak{Z}^* : \langle \xi, \mathbf{s}\rangle = \|\mathbf{s}\|_{\mathfrak{Z}}, \ \ \|\xi\|_{\mathfrak{Z}^*} \leq 1\Big\}.$$

(d) If $p \in (1, \infty)$ and $\mathfrak{Z}$ is a Banach space, then

$$\Psi(\mathbf{s}) = \frac{1}{p}\|\mathbf{s}\|_{\mathfrak{Z}}^p \quad \iff \quad \Psi^*(\xi) = \frac{1}{q}\|\xi\|_{\mathfrak{Z}^*}^q, \quad q = \frac{p}{p-1}.$$

We are now ready to give a variational formulation to (1.8). For each $\mathbf{x} \in M$, we define the conjugate pair $(\Psi(\mathbf{x}, \cdot), \Psi^*(\mathbf{x}, \cdot))$ as

$$\Psi(\mathbf{x}, \mathbf{s}) := \frac{1}{2}\langle \mathbb{G}(\mathbf{x})\mathbf{s}, \mathbf{s}\rangle, \qquad \Psi^*(\mathbf{x}, \xi) := \frac{1}{2}\langle \xi, \mathbb{K}(\mathbf{x})\xi\rangle, \qquad (\mathbf{s}, \xi) \in T_{\mathbf{x}}M \times T_{\mathbf{x}}^*M.$$

We call Ψ the *dissipation potential* and Ψ^* the *dual dissipation potential*.

In terms of the dissipation potentials the differential identity

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(\mathbf{x}(t)) = -\frac{1}{2}|\dot{\mathbf{x}}(t)|_{\mathbb{G}(\mathbf{x}(t))}^2 - \frac{1}{2}|\mathrm{D}\mathcal{F}(\mathbf{x}(t))|_{\mathbb{K}(\mathbf{x}(t))}^2$$

can be expressed as (after integrating over an interval $[s, t] \subset [0, \infty]$)

$$\mathscr{L}(\mathbf{x}, [s, t]) := \int_s^t \Psi(\mathbf{x}(r), \dot{\mathbf{x}}(r)) + \Psi^*(\mathbf{x}(r), -\mathrm{D}\mathcal{F}(\mathbf{x}(r)))\,\mathrm{d}r + \mathcal{F}(\mathbf{x}(t)) - \mathcal{F}(\mathbf{x}(s)) = 0,$$

i.e. we recover the energy-dissipation balance counterpart in Riemannian manifolds. In the same spirit as in the previous section, we obtain the following characterization:

$$\mathbf{x} \text{ solves (1.1)} \quad \iff \quad \mathbf{x} \text{ is a minimizer of the } \mathscr{L}(\,\cdot\,, [s, t]) \text{ for every } [s, t] \subset [0, \infty).$$

1.3 Gradient flows of λ-convex driving functionals

In the Riemannian setup, we generalized the dissipation mechanism. In this section, we generalize by considering less regular driving functionals $\mathcal{F}$ on Banach spaces, which we can do due to the variational nature of gradient flows.

Definition 1.15 (λ-convex functionals and their subdifferentials) A functional $\mathcal{F} : \mathfrak{X} \to \mathbb{R} \cup \{+\infty\}$ on a Banach space is said to be λ-*convex* if the functional

$$\mathfrak{X} \ni \mathbf{x} \mapsto \mathcal{F}_\lambda(\mathbf{x}) := \mathcal{F}(\mathbf{x}) - \frac{\lambda}{2}\|\mathbf{x}\|_{\mathfrak{X}}^2 \quad \text{is convex.}$$

The *subdifferential* of a λ-convex functional $\mathcal{F}$ is defined by

$$\partial_\lambda \mathcal{F}(\mathbf{x}) := \left\{ p \in \mathfrak{X}^* : \mathcal{F}(\mathbf{y}) \geq \mathcal{F}(\mathbf{x}) + \langle p, \mathbf{y} - \mathbf{x} \rangle + \frac{\lambda}{2}\|\mathbf{y} - \mathbf{x}\|_{\mathfrak{X}}^2 \quad \text{for all } \mathbf{y} \in \mathfrak{X} \right\} \subset \mathfrak{X}^*.$$

We define $\partial_\lambda^\circ \mathcal{F}(\mathbf{x})$ as the element of minimal norm in $\partial \mathcal{F}(\mathbf{x})$ and the the *(global) slope* as

$$|\partial_\lambda \mathcal{F}|(\mathbf{x}) := \inf\left\{ \|p\|_{\mathfrak{X}^*} : p \in \partial \mathcal{F}(\mathbf{x}) \right\} = \|\partial_\lambda^\circ \mathcal{F}(\mathbf{x})\|_{\mathfrak{X}^*}.$$

Remark 1.16 Recall that every convex function is continuous in its domain and is locally Lipschitz continuous in the interior of its domain.

Remark 1.17 In the Banach space setting, the following are well-known equivalent characterizations of λ-convex functions:

(i) For any $\mathbf{x}_0, \mathbf{x}_1 \in \mathfrak{X}$ and all $\theta \in [0,1]$:

$$\mathcal{F}_\lambda((1-\theta)\mathbf{x}_0 + \theta\mathbf{x}_1) \leq (1-\theta)\mathcal{F}_\lambda(\mathbf{x}_0) + \theta\mathcal{F}_\lambda(x_1) - \frac{\lambda}{2}\theta(1-\theta)\|\mathbf{x}_0 - \mathbf{x}_1\|_{\mathfrak{X}}^2.$$

(ii) *λ-monotonicity of $\partial\mathcal{F}$:* $\langle q - p, \mathbf{y} - \mathbf{x} \rangle \geq \lambda\|\mathbf{y} - \mathbf{x}\|_{\mathfrak{X}}^2$ for all $p \in \partial\mathcal{F}_\lambda(\mathbf{x})$, $q \in \partial\mathcal{F}_\lambda(\mathbf{y})$.

(iii) *Subgradient inequality:* For any $p \in \partial\mathcal{F}_\lambda(\mathbf{x})$, $q \in \partial\mathcal{F}_\lambda(\mathbf{y})$,

$$\langle q, \mathbf{y} - \mathbf{x} \rangle - \frac{\lambda}{2}\|\mathbf{y} - \mathbf{x}\|_{\mathfrak{X}}^2 \geq \mathcal{F}_\lambda(\mathbf{y}) - \mathcal{F}_\lambda(\mathbf{x}) \geq \langle p, \mathbf{y} - \mathbf{x} \rangle + \frac{\lambda}{2}\|\mathbf{y} - \mathbf{x}\|_{\mathfrak{X}}^2$$

Exercise 1.1 (a) Show that the map $\mathbb{R}^d \ni \mathbf{x} \to \frac{1}{2}|\mathbf{y} - \mathbf{x}|^2$ is λ-convex. What is λ?

(b) Determine the correspondence between the subdifferentials $\partial_\lambda \mathcal{F}$ and $\partial \mathcal{F}_\lambda$.

Since we will be considering less regular driving functionals $\mathcal{F}$, we cannot expect the solution $\mathbf{x}$ to the Cauchy problem to be continuously differentiable, thus requiring us to consider less regular curves in our solution concept. It turns out that the appropriate class of curves is that of absolutely continuous curves:

Definition 1.18 (Absolutely continuous curves) Let $(\mathfrak{X}, d)$ be a complete metric space. A curve $\mathbf{x}\colon (a,b) \to \mathfrak{X}$ is said to be *p-absolutely continuous* if there exists a p-integrable function $m \in L^p((a,b))$, $p \in [1, +\infty]$ such that

$$d(\mathbf{x}(t), \mathbf{x}(s)) \leq \int_s^t m(r)\,\mathrm{d}r \qquad \text{for every } (s,t) \subset (a,b). \tag{1.10}$$

We denote the space of all p-absolutely continuous curves by $\mathcal{AC}^p((a,b);\mathfrak{X})$.

In the case $p = 1$, we are simply dealing with absolutely continuous curves denote the corresponding space with $\mathcal{AC}((a,b);\mathfrak{X})$.

Among all the possible choices of $m \in L^1((a,b))$ in (1.10), there exists a minimal one, which is provided by the following proposition.

Proposition 1.19 (Metric derivative)**.** *For any curve* $\mathbf{x} \in \mathcal{AC}^p((a,b);\mathfrak{X})$*, the limit*

$$|\dot{\mathbf{x}}|(t) := \lim_{h\to 0} \frac{d(\mathbf{x}(t+h), \mathbf{x}(t))}{|h|} \quad \textit{exists for almost every } t \in (a,b).$$

Moreover, the function $t \mapsto |\dot{\mathbf{x}}|(t)$ *belongs to* $L^p((a,b))$*, is an admissible integrand in* (1.10)*, and is minimal in the following sense: For any function* $m \in L^p((a,b))$ *satisfying* (1.10)*,*

$$|\dot{\mathbf{x}}|(t) \leq m(t) \quad \textit{for almost every } t \in (a,b).$$

Now suppose that $\mathcal{F} : \mathbb{R}^d \to \mathbb{R}$ is λ-convex (i.e. possibly non-differentiable). Instead of the gradient flow equation (1.1), one then looks for an *absolutely continuous* curve $\mathbf{x}\colon [0,\infty) \to \mathbb{R}^d$ satisfying the *differential inclusion*

$$\dot{\mathbf{x}}(t) \in -\partial_\lambda \mathcal{F}(\mathbf{x}(t)) \quad \text{for almost every } t > 0, \qquad \mathbf{x}(0) = \mathbf{x}_0. \tag{1.11}$$

Let us mention a few nice properties of the differential inclusion (1.11) before investigating the existence of solutions to the inclusion.

Lemma 1.20. *Let* $\mathcal{F}$ *be* λ*-convex,* $\lambda \in \mathbb{R}$ *and* $\mathbf{x}$ *be a solution of* (1.11) *with* $\mathbf{x}(0) = \mathbf{x}_0 \in \mathbb{R}^d$*. Then the following inequalities hold:*

(a) (Evolution-variational inequality) For any $\mathbf{y} \in \mathbb{R}^d$ *and almost every* $t \in (0,\infty)$*:*

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}|\mathbf{x}(t) - \mathbf{y}|^2 + \frac{\lambda}{2}|\mathbf{x}(t) - \mathbf{y}|^2 \leq \mathcal{F}(\mathbf{y}) - \mathcal{F}(\mathbf{x}(t)). \tag{EVI$_\lambda$}$$

(b) (Contraction inequality) Let $\mathbf{y}$ *be a solution of* (1.11) *with* $\mathbf{y}(0) = \mathbf{y}_0 \in \mathbb{R}^d$*. Then*

$$|\mathbf{y}(t) - \mathbf{x}(t)| \leq |\mathbf{y}(0) - \mathbf{x}(0)|e^{-\lambda t} \qquad \textit{for every } t \geq 0. \tag{Cont$_\lambda$}$$

In particular, the Cauchy problem (1.11) *has a unique solution.*

Proof. (a) Formally, we simply compute (rigorous justification will be given later on)

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}|\mathbf{x}(t)-\mathbf{y}|^2 = \langle \dot{\mathbf{x}}(t), \mathbf{x}(t)-\mathbf{y}\rangle = \langle -\dot{\mathbf{x}}(t), \mathbf{y}-\mathbf{x}(t)\rangle.$$

Since $-\dot{\mathbf{x}}(t) \in \partial\mathcal{F}(\mathbf{x}(t))$, we obtain the assertion via the subgradient inequality.

(b) Similarly, we obtain

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}|\mathbf{y}(t)-\mathbf{x}(t)|^2 = (\mathbf{y}(t)-\mathbf{x}(t))\cdot(\dot{\mathbf{y}}(t)-\dot{\mathbf{x}}(t)) \qquad \text{for almost every } t>0.$$

The λ-monotonicity then yields

$$\langle \dot{\mathbf{y}}-\dot{\mathbf{x}}, \mathbf{y}-\mathbf{x}\rangle \le -\lambda|\mathbf{y}-\mathbf{x}|^2 \quad \Rightarrow \quad f'(t) \le -2\lambda f(t).$$

We then conclude, from Gronwall's inequality, $f(t) \le e^{-2\lambda t} f(0)$.

The uniqueness property is easy: If $\mathbf{y}(0)=\mathbf{x}(0)$, then $\mathbf{y}(t)=\mathbf{x}(t)$ for all $t>0$. □

Remark 1.21 Notice that when $\lambda > 0$, the previous statement shows the contractive nature of the gradient flow. In fact, if $\mathcal{F}$ is λ-convex with $\lambda > 0$, then it admits a unique minimizer $\bar{\mathbf{x}}$. Since $0 \in \partial_\lambda \mathcal{F}(\bar{\mathbf{x}})$, it is also a solution of the differential inclusion (1.11). Hence, the stability estimate in Lemma 1.20 provides the exponential convergence $|\mathbf{x}(t)-\bar{\mathbf{x}}| \le |\mathbf{x}(0)-\bar{\mathbf{x}}|e^{-\lambda t}$.

Another interesting property of λ-convex functionals is that the differential inclusion (1.11) turns into an identity, i.e. in this case, we do have a differential equation instead.

Proposition 1.22. *Let $\mathcal{F}$ be λ-convex, $\lambda \in \mathbb{R}$ and $\mathbf{x}$ be a solution of* (1.11)*. Then for all $t_0 > 0$ for which both $t \mapsto \mathbf{x}(t)$ and $t \mapsto \mathcal{F}(\mathbf{x}(t))$ are differentiable, the differential equation*

$$\dot{\mathbf{x}}(t_0) = -\partial_\lambda^\circ \mathcal{F}(\mathbf{x}(t_0)).$$

In particular, the equation above holds for almost every $t \in (0,\infty)$.

Proof. Let $t_0 \in (0,\infty)$ be a point of differentiability and $p \in \partial\mathcal{F}(\mathbf{x}(t_0))$. Then,

$$\mathcal{F}(\mathbf{x}(t)) \ge \mathcal{F}(\mathbf{x}(t_0)) + p\cdot(\mathbf{x}(t)-\mathbf{x}(t_0)) + \frac{\lambda}{2}|\mathbf{x}(t)-\mathbf{x}(t_0)|^2 \qquad \text{for every } t>0,$$

and becomes an equality for $t=t_0$. Therefore, the quantity

$$f(t) := \mathcal{F}(\mathbf{x}(t)) - \mathcal{F}(\mathbf{x}(t_0)) - \langle p, \mathbf{x}(t)-\mathbf{x}(t_0)\rangle - \frac{\lambda}{2}|\mathbf{x}(t)-\mathbf{x}(t_0)|^2$$

is minimal when $t=t_0$. Differentiating f at $t=t_0$, which is possible due to the assumption on $t=t_0$, we obtain

$$0 = f'(t_0) = \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(\mathbf{x}(t))\bigg|_{t=t_0} - p\cdot\dot{\mathbf{x}}(t_0) \quad \Longleftrightarrow \quad c(t_0) := \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(\mathbf{x}(t))\bigg|_{t=t_0} = p\cdot\dot{\mathbf{x}}(t_0)$$

Since this holds for every $p \in \partial_\lambda \mathcal{F}(\mathbf{x}(t_0))$, we have that $\partial_\lambda \mathcal{F}(\mathbf{x}(t_0))$ is contained in the hyperplane $H(t_0) := \{p \in \mathbb{R}^d : p \cdot \dot{\mathbf{x}}(t_0) = c(t_0)\}$. However, since $\dot{\mathbf{x}}(t) \in \partial_\lambda \mathcal{F}(\mathbf{x}(t))$ holds for almost every $t > 0$ and in particular for $t = t_0$, we have that $c(t_0) = |\dot{\mathbf{x}}(t_0)|^2$, and therefore

$$(p - \dot{\mathbf{x}}(t_0)) \cdot \dot{\mathbf{x}}(t_0) = 0 \qquad \text{for all } p \in \partial_\lambda \mathcal{F}(\mathbf{x}(t_0)),$$

i.e. we have deduced that $\dot{\mathbf{x}}(t_0)$ is the element in $\partial_\lambda \mathcal{F}(\mathbf{x}(t_0))$ with minimal norm. Indeed, if $\mathbf{v} \in \partial_\lambda \mathcal{F}(\mathbf{x}(t_0))$ is a minimizer of the problem

$$\min \left\{ \frac{1}{2}|p|^2 : p \in \partial_\lambda \mathcal{F}(\mathbf{x}(t_0)) \right\}.$$

Then $\mathbf{v}$ satisfies for all $t \in [0,1]$ and $p \in \partial_\lambda \mathcal{F}(\mathbf{x}(t_0))$:

$$\frac{1}{2}|\mathbf{v}| \le \frac{1}{2}|\mathbf{v} + t(p - \mathbf{v})|^2 = \frac{1}{2}|\mathbf{v}|^2 + t\,(p - \mathbf{v}) \cdot \mathbf{v} + \frac{t^2}{2}|p - \mathbf{v}|^2.$$

Therefore, dividing by t and sending $t \to 0$, we recover the variational inequality

$$(p - \mathbf{v}) \cdot \mathbf{v} \ge 0 \qquad \forall p \in \partial_\lambda \mathcal{F}(\mathbf{x}(t_0)).$$

Now since $\mathbf{x}$ is absolutely continuous and $\mathcal{F}$ is locally Lipschitz, we have that $\mathbf{x}$ and $\mathcal{F} \circ \mathbf{x}$ are differentiable almost everywhere. Hence, the differential equation holds for almost every $t > 0$ as asserted. □

Definition 1.23 (Gradient flows for λ-convex driving functionals) The gradient flow associated with a λ-convex functional $\mathcal{F}$ is the family of maps $\mathsf{S}_t \colon \mathbb{R}^d \to \mathbb{R}^d$, $t \in [0, \infty)$ characterized by the following properties:

(a) For every $\mathbf{x}_0 \in \mathbb{R}^d$, $\mathsf{S}_0(\mathbf{x}_0) = \mathbf{x}_0$;

(b) The curve $\mathbf{x}_t := \mathsf{S}_t(\mathbf{x}_0)$, $t \in (0, \infty)$ is the (unique) solution to the *Cauchy problem*

$$\dot{\mathbf{x}}(t) = -\partial_\lambda \mathcal{F}^\circ(\mathbf{x}(t)) \qquad \text{holds for almost every } t > 0, \qquad \mathbf{x}(0) = \mathbf{x}_0. \tag{1.12}$$

As mentioned earlier and as the following proposition suggests, (EVI_λ) provides us with yet another variational characterization of gradient flow solutions

Proposition 1.24. *If* $\mathbf{x} \colon [0, \infty) \to \mathbb{R}^d$ *is an absolutely continuous curve satisfying* (EVI_λ) *for every* $\mathbf{y} \in \mathbb{R}^d$*, then* $\mathbf{x}$ *is a solution of* (1.11)*. In particular, it satisfies* (1.12)*.*

Proof. Applying the chain rule, we obtain for every $\mathbf{y} \in \mathbb{R}^d$ and $t > 0$:

$$\langle \dot{\mathbf{x}}(t), \mathbf{x}(t) - \mathbf{y} \rangle = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}|\mathbf{x}(t) - \mathbf{y}|^2 \le \mathcal{F}(\mathbf{y}) - \mathcal{F}(\mathbf{x}(t)) - \frac{\lambda}{2}|\mathbf{x}(t) - \mathbf{y}|^2.$$

Thus implying that $\dot{\mathbf{x}}(t) \in \partial_\lambda \mathcal{F}(\mathbf{x}(t))$ by the definition of the subdifferential of $\mathcal{F}$. □

Hence, we obtain the following definition of a gradient flow.

Definition 1.25 The gradient flow associated with a λ-convex functional $\mathcal{F}$ is the family of maps $\mathsf{S}_t\colon \mathbb{R}^d \to \mathbb{R}^d$, $t \in [0,\infty)$ characterized by the following properties:

(a) For every $\mathbf{x}_0 \in \mathbb{R}^d$, $\mathsf{S}_0(\mathbf{x}_0) = \mathbf{x}_0$;

(b) The curve $\mathbf{x}_t := \mathsf{S}_t(\mathbf{x}_0)$, $t \in (0,\infty)$ is an absolutely continuous curve satisfying (EVI$_\lambda$).

To arrive at a variational formulation based on the energy-dissipation principle, we will need yet another ingredient, which will be discussed in the next section.

1.4 Towards metric space descriptions of gradient flows

For gradient flows in metric spaces $(\mathfrak{X}, d)$ we need to move away from the formulation based on the Cauchy problem since the equation cannot be meaningfully posed. In this section, e will formulate two complementary formulations of gradient flows. One is based on the energy-dissipation principle, and the other is based on the evolution-variational inequality.

If one looks carefully at the definition of gradient flows based on the evolution-variational inequality (cf. Definition 1.25), one easily notices that this definition is also appropriate for metric spaces. However, we will first require the notion of λ-convex functionals in the metric space setting, which we define only for *geodesically complete metric spaces.*

Definition 1.26 (Geodesic curves and spaces) Let $(\mathfrak{X}, \mathsf{d})$ be a metric space. A curve in $\mathfrak{X}$ $\gamma\colon [0,1] \to \mathfrak{X}$ is a (constant speed) *geodesic* if

$$\mathsf{d}(\gamma_s, \gamma_t) = (t-s)\,\mathsf{d}(\gamma_0, \gamma_1) \quad \text{for any } [s,t] \subset [0,1].$$

The space $\mathfrak{X}$ is said to be *geodesically complete* if, for each $\mathbf{x}, \mathbf{y} \in \mathfrak{X}$, we find a constant speed geodesic γ for which $\gamma_0 = \mathbf{x}$ and $\gamma_1 = \mathbf{y}$. We simply call $\mathfrak{X}$ a *geodesic space.*

Definition 1.27 (Geodesically λ-convex functionals) A functional $\mathcal{F} : \mathfrak{X} \to \mathbb{R} \cup \{+\infty\}$ on a geodesic space $(\mathfrak{X}, \mathsf{d})$ is said to be *geodesically λ-convex* if for any geodesic $\gamma\colon [0,1] \to \mathfrak{X}$,

$$\mathcal{F}(\gamma_\theta) \le (1-\theta)\mathcal{F}(\gamma_0) + \theta\mathcal{F}(\gamma_1) - \frac{\lambda}{2}\theta(1-\theta)\mathsf{d}^2(\gamma_0, \gamma_1) \qquad \forall\, \theta \in [0,1].$$

In a geodesic space, the evolution-variational inequality characterization of gradient flows then read:

Definition 1.28 Let $(\mathfrak{X}, \mathsf{d})$ be a geodesic space. The gradient flow associated with a (geodesically) λ-convex functional $\mathcal{F} : \mathfrak{X} \to \mathbb{R} \cup \{+\infty\}$ is the family of maps $\mathsf{S}_t : \mathfrak{X} \to \mathfrak{X}$, $t \in [0,\infty)$ characterized by the following properties:

(a) For every $\mathbf{x}_0 \in \mathfrak{X}$, $\mathsf{S}_0(\mathbf{x}_0) = \mathbf{x}_0$;

(b) The curve $\mathbf{x}_t := \mathsf{S}_t(\mathbf{x}_0)$, $t \in (0,\infty)$ is an absolutely continuous curve satisfying

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{d}^2(\mathbf{x}(t), \mathbf{y}) + \frac{\lambda}{2}\mathsf{d}^2(\mathbf{x}(t), \mathbf{y}) \le \mathcal{F}(\mathbf{y}) - \mathcal{F}(\mathbf{x}(t)). \tag{EVI$_\lambda$}$$

While this definition is convenient, finding (or constructing) gradient flow solutions based on the evolution-variational inequality is a nontrivial endeavor. For this reason, we will focus mainly on the variational formulation based on the energy-dissipation principle.

Recall that the energy-dissipation functional in $\mathbb{R}^d$ or in a Riemannian manifold (M, g) involves three objects: the modulus of the temporal derivative $|\dot{\mathbf{x}}|$, the driving functional $\mathcal{F}$ and the modulus of its gradient $|\nabla_g \mathcal{F}|$ in some meaningful way. In a metric space, $|\dot{\mathbf{x}}|$ is replaced by the metric derivative, which is well-defined for absolutely continuous curves. Unfortunately, many definitions of the modulus of the 'gradient of $\mathcal{F}$' are possible in the metric space setting.

Definition 1.29 (Upper gradients) Let $\mathcal{F} : \mathfrak{X} \to \mathbb{R} \cup \{+\infty\}$ be a proper functional. A function $g : \mathfrak{X} \to [0, +\infty]$ is called a *(strong) upper gradient* for $\mathcal{F}$ if for every absolutely continuous curve $\mathbf{x} \in \mathcal{AC}((a, b); \mathfrak{X})$, the function $g \circ \mathbf{x}$ is Borel measurable and

$$|\mathcal{F}(\mathbf{x}(t)) - \mathcal{F}(\mathbf{x}(s))| \le \int_s^t g(\mathbf{x}(r))|\dot{\mathbf{x}}|(r)\,\mathrm{d}r \qquad \text{for every } (s, t) \subset (a, b).$$

In particular, if the integrand $g \circ \mathbf{x}\,|\dot{\mathbf{x}}| \in L^1((a, b))$, then $\mathcal{F} \circ \mathbf{x}$ is absolutely continuous and

$$|(\mathcal{F} \circ \mathbf{x})'|(t) \le g(\mathbf{x}(t))|\dot{\mathbf{x}}|(t) \quad \text{for almost every } t \in (a, b). \tag{1.13}$$

Among the many choices for an upper gradient for $\mathcal{F}$, we make most use of the so-called λ-*global (descending) slope* of $\mathcal{F}$ defined by

$$|\partial_\lambda \mathcal{F}|(\mathbf{x}) := \sup_{\mathbf{y} \ne \mathbf{x}} \left[\frac{\mathcal{F}(\mathbf{x}) - \mathcal{F}(\mathbf{y})}{d(\mathbf{x}, \mathbf{y})} + \frac{\lambda}{2}\mathsf{d}(\mathbf{x}, \mathbf{y})\right]^+.$$

Remark 1.30 From the definition of the λ-global slope $|\partial_\lambda \mathcal{F}|$, we have that

$$\mathcal{F}(\mathbf{y}) \ge \mathcal{F}(\mathbf{x}) - |\partial_\lambda \mathcal{F}|(\mathbf{x})\,\mathsf{d}(\mathbf{x}, \mathbf{y}) + \frac{\lambda}{2}\mathsf{d}^2(\mathbf{x}, \mathbf{y}) \qquad \forall\, \mathbf{y} \in \mathfrak{X},$$

which generalizes the Banach space subdifferential of λ-convex functionals $\mathcal{F}$.

Another choice for an upper gradient for $\mathcal{F}$ is the *local slope*

$$|\partial \mathcal{F}|(\mathbf{x}) := \limsup_{\mathbf{y} \to \mathbf{x}} \left[\frac{\mathcal{F}(\mathbf{x}) - \mathcal{F}(\mathbf{y})}{d(\mathbf{x}, \mathbf{y})}\right]^+,$$

which, in general, does not coincide with the λ-global slope $|\partial_\lambda \mathcal{F}|$. However, if $\mathcal{F}$ is (geodesically) λ-convex, then this is true and therefore also an upper gradient for $\mathcal{F}$.

Exercise 1.2 Show that if $\mathcal{F}$ is (geodesically) λ-convex, then

$$|\partial \mathcal{F}|(\mathbf{x}) = |\partial_\lambda \mathcal{F}|(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \operatorname{dom} \mathcal{F}.$$

Definition 1.31 (Curves of maximal slope) Let $\mathcal{F}$ be a proper functional with a (strong) upper gradient g. An absolutely continuous curve $\mathbf{x} \in \mathcal{AC}((a,b);\mathfrak{X})$ is said to be *p-curve of maximal slope*, $p \in (1, +\infty)$ for the functional $\mathcal{F}$ with respect to g if

$$(\mathcal{F} \circ \mathbf{x})'(t) \leq -\frac{1}{p}|\dot{\mathbf{x}}|^p(t) - \frac{1}{q}g^q(\mathbf{x}(t)) \quad \text{for almost every } t \in (a,b), \tag{1.14}$$

where $q = p/(p-1)$ is the conjugate exponent of p.

Combining (1.13) and (1.14), we obtain

$$\frac{1}{p}|\dot{\mathbf{x}}|^p(t) + \frac{1}{q}g^q(\mathbf{x}(t)) \leq -(\mathcal{F} \circ \mathbf{x})'(t) \leq |(\mathcal{F} \circ \mathbf{x})'|(t) \leq g(\mathbf{x}(t))|\dot{\mathbf{x}}|(t),$$

which implies the equality

$$\frac{1}{p}|\dot{\mathbf{x}}|^p(t) + \frac{1}{q}g^q(\mathbf{x}(t)) + (\mathcal{F} \circ \mathbf{x})'(t) = 0 \quad \text{for almost every } t \in (a,b),$$

and we recover the energy-dissipation balance after integrating over intervals $[s,t]$.

Based on the discussion above, if $\mathcal{F}$ is (geodesically) λ-convex, then we can take $g = |\partial_\lambda \mathcal{F}|$ and $p = 2$ to obtain

$$\mathscr{L}(\mathbf{x}, [s,t]) \coloneqq \int_s^t \left\{ \frac{1}{2}|\dot{\mathbf{x}}|^2(r) + \frac{1}{2}|\partial_\lambda \mathcal{F}|^2(\mathbf{x}(r)) \right\} \mathrm{d}r + \mathcal{F}(\mathbf{x}(t)) - \mathcal{F}(\mathbf{x}(s)) = 0.$$

In particular, curves of maximal slope give rise to the gradient flow solutions that we were looking for—gradient flow solutions based on the energy-dissipation principle.

2 Existence of Curves of Maximal Slope

This section aims to construct a curve of maximal slope that satisfies the energy-dissipation balance. More precisely, we want to solve the following problem:

Problem: Given a (geodesically) λ-convex functional $\mathcal{F}\colon \mathfrak{X} \to \mathbb{R} \cup \{+\infty\}$ on a geodesic space $(\mathfrak{X}, \mathsf{d})$ and an initial datum $\mathbf{x}_0 \in \operatorname{dom} \mathcal{F}$, find a curve $\mathbf{x}$ of maximal slope for $\mathcal{F}$ such that $\mathbf{x}(0) = \mathbf{x}_0$. More precisely, we are looking for a curve $\mathbf{x} \in \mathcal{AC}^2((0, +\infty); \mathfrak{X})$ satisfying

$$\mathscr{L}(\mathbf{x}, [s,t]) \coloneqq \int_s^t \left\{ \frac{1}{2}|\dot{\mathbf{x}}|^2(r) + \frac{1}{2}|\partial_\lambda \mathcal{F}|^2(\mathbf{x}(r)) \right\} \mathrm{d}r + \mathcal{F}(\mathbf{x}(t)) - \mathcal{F}(\mathbf{x}(s)) = 0.$$

for every interval $[s,t] \subset [0,\infty)$.

Strategy: The construction of such a curve relies on two main tools:

(1) A minimizing-movement scheme: Provides an admissible curve satisfying $\mathscr{L} \leq 0$.

(2) The chain rule: Establishes $\mathscr{L} \geq 0$ for an admissible class of curves;

2.1 The generalized minimizing-movement scheme

In this section, we introduce a time-discrete approach for constructing curves satisfying the inequality $\mathscr{L} \leq 0$—the so-called

Minimizing-movement scheme: For a given a time step $\tau \coloneqq T/N$, $N \geq 1$ and associated partitioning of the time interval $[0,T]$

$$\mathsf{P}_\tau = \{I_\tau^n,\ n = 1, \ldots, N\}, \qquad I_\tau^n = (\tau(n-1), \tau n] \eqqcolon (t_{n-1}, t_n],$$

we construct a discrete approximation by recursively solving the minimization problem

$$\text{Find } \mathbf{X}_\tau^n \in \mathfrak{X}: \qquad \Phi_\tau(\mathbf{X}_\tau^{n-1}; \mathbf{X}_\tau^n) \leq \Phi_\tau(\mathbf{X}_\tau^{n-1}; \mathbf{y}) \quad \text{for all } \mathbf{y} \in \mathfrak{X}, \tag{MM}$$

where

$$\Phi_\tau(\mathbf{x}; \mathbf{y}) \coloneqq \frac{1}{2\tau}\mathsf{d}^2(\mathbf{y}, \mathbf{x}) + \mathcal{F}(\mathbf{y}), \qquad \mathbf{x}, \mathbf{y} \in \mathfrak{X}.$$

The multi-valued operator, which provides all the minimizers of (MM) for a given $\mathbf{X}^n \in \mathfrak{X}$ is called the *resolvent operator* and denoted by

$$J_\tau[\mathbf{x}] \coloneqq \operatorname*{argmin}_{\mathbf{y} \in \mathfrak{X}} \Phi_\tau(\mathbf{y}; \mathbf{x}) \subset \mathfrak{X}.$$

Thus the family $\{\mathbf{X}_\tau^n\}_{n=0,\ldots,N}$ solves the minimizing-movement scheme (MM) if and only if

$$\mathbf{X}_\tau^n \in J_\tau[\mathbf{X}_\tau^{n-1}] \quad \text{for all } n = 1, \ldots, N.$$

Piecewise constant approximation By patching the discrete solutions $\{\mathbf{X}_\tau^n\}_{n=0,\dots,N}$ to (MM) together, we obtain an approximate curve. The *piecewise constant curve* is obtained when we set

$$\overline{\mathbf{X}}_\tau(0) \coloneqq \mathbf{X}_\tau^0, \qquad \overline{\mathbf{X}}_\tau(t) \coloneqq \mathbf{X}_\tau^n \quad \text{for } t \in I_\tau^n, \ \ n = 1, \dots, N.$$

The curve $\overline{\mathbf{X}}_\tau$ is also called the *discrete solution* corresponding to the partition P_τ.

Definition 2.1 (Generalized Minimizing Movements) A curve $\mathbf{x}\colon [0,T] \to \mathfrak{X}$ is called a *Generalized Minimizing Movement* (GMM) starting from $\mathbf{x}_0 \in \mathfrak{X}$ if there exists a sequence of time steps $\tau_N \to 0$ and a corresponding discrete solution $\overline{\mathbf{X}}_{\tau_N}$, such that

$$\lim_{N\to\infty} \mathcal{F}(\mathbf{X}_{\tau_N}^0) = \mathcal{F}(\mathbf{x}_0), \qquad \limsup_{N\to\infty} \mathsf{d}(\mathbf{X}_{\tau_N}^0, \mathbf{x}_0) < +\infty$$

$$\overline{\mathbf{X}}_{\tau_N}(t) \to \mathbf{x}(t) \quad \text{for all } t \in [0,T] \text{ w.r.t. some } \mathsf{d}\text{-compatible topology,}$$

in the sense that $(\mathbf{x},\mathbf{y}) \mapsto \mathsf{d}(\mathbf{x},\mathbf{y})$ is σ-lower semicontinuous, i.e.

$$(\mathbf{x}_n, \mathbf{y}_n) \overset{\sigma}{\rightharpoonup} (\mathbf{x},\mathbf{y}) \quad \Longrightarrow \quad \mathsf{d}(\mathbf{x},\mathbf{y}) \le \liminf_{n\to\infty} \mathsf{d}(\mathbf{x}_n,\mathbf{y}_n).$$

We denote by $\mathrm{GMM}(\mathbf{x}_0)$ the collection of all GMMs starting from $\mathbf{x}_0 \in \mathfrak{X}$.

We will provide two instances where any $\mathbf{x} \in \mathrm{GMM}(\mathbf{x}_0)$ is also a curve of maximal slope in the sense of Definition 1.31 based on two sets of basic assumptions on the driving functional $\mathcal{F}$. The first result is obtained when $\mathcal{F}\colon \mathfrak{X} \to \mathbb{R} \cup \{+\infty\}$ is assumed to be

$$\mathsf{d}\text{-}\textit{lower semicontinuous and } \mathsf{d}\text{-}\textit{coercive.} \tag{$A_\mathcal{F}^1$}$$

In many situations, however, the existence of a minimizer for $\Phi_\tau(\mathbf{x};\cdot)$ can only be ensured under a weaker topology σ that is compatible with d. This leads us to another set of assumptions, where we assume $\mathcal{F}$ to be

$$\begin{cases} \sigma\text{-}\textit{lower semicontinuous on } \mathsf{d}\text{-}\textit{bounded sets:} \\ \text{If } \mathbf{x}_k \overset{\sigma}{\rightharpoonup} \mathbf{x} \text{ with } \sup_{k,\ell} \mathsf{d}(\mathbf{x}_k,\mathbf{x}_\ell) < +\infty, \text{ then } \ \mathcal{F}(\mathbf{x}) \le \liminf_{k\to\infty} \mathcal{F}(\mathbf{x}_k); \\ \sigma\text{-}\textit{coercive on } \mathsf{d}\text{-}\textit{bounded sets:} \\ \text{If } (\mathbf{x}_k)_k \text{ with } \sup_k \mathcal{F}(\mathbf{x}_k) < \infty \text{ and } \sup_{k,\ell} \mathsf{d}(\mathbf{x}_k,\mathbf{x}_\ell) < +\infty, \\ \text{then } (\mathbf{x}_k)_k \text{ has an accumulation point in } \mathfrak{X}. \end{cases} \tag{$A_\mathcal{F}^2$}$$

In either case, we will always assume $\mathcal{F}$ is (geodesically) λ-convex and d-locally bounded from below, i.e. there exists $\tau_* > 0$ and $\mathbf{x}_* \in \mathfrak{X}$ for which

$$\mathcal{F}_{\tau_*}(\mathbf{x}_*) \coloneqq \inf_{\mathbf{y}\in\mathfrak{X}} \Phi_{\tau_*}(\mathbf{x}_*;\mathbf{y}) > -\infty. \tag{$A_\mathcal{F}^\lambda$}$$

Remark 2.2 (a) Recall that $\mathcal{F}$ is σ-coercive if for all $c \in \mathbb{R}$, the sublevel set

$$L_\mathcal{F}(c) \coloneqq \Big\{\mathbf{x} \in \mathfrak{X} : \mathcal{F}(\mathbf{x}) \le c\Big\} \subset \mathfrak{X} \quad \text{is relatively } \sigma\text{-sequentially compact,}$$

i.e. any sequence $(\mathbf{x}_n)_n \subset L_\mathcal{F}(c)$ for some $c \in \mathbb{R}$ has a converging subsequence.

(b) Clearly, ($A^1_{\mathcal{F}}$) implies ($A^2_{\mathcal{F}}$) with the topology σ induced by d.

Exercise 2.1 Let $\Omega \subset \mathbb{R}^d$ be a bounded domain. Consider the Dirichlet energies

$$H^1_0(\Omega) \ni u \mapsto \mathcal{F}(u) = \int_\Omega |\nabla u(x)|^2 \,\mathrm{d}x,$$
$$L^2(\Omega) \ni u \mapsto \widehat{\mathcal{F}}(u) = \begin{cases} \mathcal{F}(u) & \text{if } u \in H^1_0(\Omega) \\ +\infty & \text{otherwise.} \end{cases}$$

Determine which of the two assumptions ($A^1_{\mathcal{F}}$), ($A^2_{\mathcal{F}}$) is satisfied.

The main result we would like to prove under the weaker assumption ($A^2_{\mathcal{F}}$) requires a relaxation of the λ-global slope of $\mathcal{F}$ w.r.t. to the weak topology σ.

Definition 2.3 (Relaxed slope) We define the *relaxed slope* as

$$|\partial^- \mathcal{F}|(\mathbf{x}) := \inf\left\{ \liminf_{k\to\infty} |\partial\mathcal{F}|(\mathbf{x}_k) \;:\; \mathbf{x}_k \overset{\sigma}{\rightharpoonup} \mathbf{x}, \quad \sup_k \big\{\mathsf{d}(\mathbf{x}_k,\mathbf{x}), \mathcal{F}(\mathbf{x}_k)\big\} < +\infty \right\},$$

i.e. $|\partial^-\mathcal{F}|$ is the sequential σ-lower semicontinuous envelope of $|\partial\mathcal{F}|$.

Theorem 2.4 (Existence of curves of maximal slope $A^2_{\mathcal{F}}$). *Let $\mathcal{F}\colon \mathfrak{X} \to \mathbb{R}\cup\{+\infty\}$ satisfy ($A^\lambda_{\mathcal{F}}$) and ($A^2_{\mathcal{F}}$). Then every curve $\mathbf{x} \in GMM(\mathbf{x}_0)$ with $\mathbf{x}_0 \in dom\,\mathcal{F}$ satisfies the energy-dissipation inequality*

$$\int_s^t \left\{ \frac{1}{2}|\dot{\mathbf{x}}|^2(r) + \frac{1}{2}|\partial^-\mathcal{F}|^2(\mathbf{x}(r)) \right\} \mathrm{d}r + \mathcal{F}(\mathbf{x}(t)) \le \mathcal{F}(\mathbf{x}(s)) \qquad \forall\,[s,t]\subset[0,T]. \tag{2.1}$$

If in addition the relaxed slope

$$\mathfrak{X} \ni \mathbf{x} \mapsto |\partial^-\mathcal{F}|(\mathbf{x}) \quad \textit{is a (strong) upper gradient for } \mathcal{F},$$

then $\mathbf{x}$ is a curve of maximal slope for $\mathcal{F}$ w.r.t. $|\partial^-\mathcal{F}|$ and equality in (2.1) holds.

Corollary 2.5 (Existence of curves of maximal slope $A^1_{\mathcal{F}}$). *Let $\mathcal{F}\colon \mathfrak{X} \to \mathbb{R}\cup\{+\infty\}$ satisfy ($A^\lambda_{\mathcal{F}}$) and ($A^1_{\mathcal{F}}$). Then every curve $\mathbf{x} \in GMM(\mathbf{x}_0)$ with $\mathbf{x}_0 \in dom\,\mathcal{F}$ is a curve of maximal slope for $\mathcal{F}$ w.r.t. $|\partial\mathcal{F}|$ and satisfies the energy-dissipation balance*

$$\int_s^t \left\{ \frac{1}{2}|\dot{\mathbf{x}}|^2(r) + \frac{1}{2}|\partial\mathcal{F}|^2(\mathbf{x}(r)) \right\} \mathrm{d}r + \mathcal{F}(\mathbf{x}(t)) = \mathcal{F}(\mathbf{x}(s)) \qquad \forall\,[s,t]\subset[0,T].$$

The proof of Corollary 2.5 follows from the following result.

Lemma 2.6. *Let $\mathcal{F} : \mathfrak{X} \to \mathbb{R}\cup\{+\infty\}$ be d-lower semicontinuous. Then, the λ-global slope $|\partial_\lambda\mathcal{F}|$ is d-lower semicontinuous and a (strong) upper gradient for $\mathcal{F}$.*

Proof. We first notice that $\mathbf{x} \mapsto \partial_\lambda\mathcal{F}(\mathbf{x})$ is d-lower semicontinuous. Indeed, if $\mathbf{y} \neq \mathbf{x}$ and $\mathbf{x}_n \to \mathbf{x}$, when $\mathbf{y} \neq \mathbf{x}_n$ for $n \gg 1$ sufficiently large, and therefore

$$\liminf_{n\to\infty} |\partial_\lambda\mathcal{F}|(\mathbf{x}_n) \ge \liminf_{n\to\infty} \left[\frac{\mathcal{F}(\mathbf{x}_n) - \mathcal{F}(\mathbf{y})}{d(\mathbf{x}_n,\mathbf{y})} + \frac{\lambda}{2}\mathsf{d}(\mathbf{x}_n,\mathbf{y}) \right]^+ \ge \left[\frac{\mathcal{F}(\mathbf{x}) - \mathcal{F}(\mathbf{y})}{d(\mathbf{x},\mathbf{y})} + \frac{\lambda}{2}\mathsf{d}(\mathbf{x},\mathbf{y}) \right]^+.$$

The lower semicontinuity follows by taking the supremum w.r.t. $\mathbf{y}$.

The second part of the statement is a little too technical for this course and is a minor adaptation of [1, Theorem 1.2.5]. $\square$

Proof of Corollary 2.5. Let σ be the topology induced by the distance d. From Theorem 2.4 we find that every curve $\mathbf{x} \in \mathrm{GMM}(\mathbf{x}_0)$ satisfies (2.1). However, since $|\partial\mathcal{F}| = |\partial_\lambda \mathcal{F}|$ is d-lower semicontinuous and a (strong) upper gradient, $|\partial^-\mathcal{F}| = |\partial\mathcal{F}|$ and, therefore, the energy-dissipation balance holds as asserted. □

The rest of this section is devoted to proving Theorem 2.4.

2.2 Solvability of discrete problem

We begin by proving the solvability of (MM) for each time increment.

Lemma 2.7. *Let $\mathcal{F}\colon \mathfrak{X} \to \mathbb{R}\cup\{+\infty\}$ satisfy ($A^\lambda_{\mathcal{F}}$) and $\tau < \tau_*$. Then*

$$\mathcal{F}_\tau(\mathbf{x}) \geq \mathcal{F}_{\tau_*}(\mathbf{x}_*) - \frac{1}{\tau_* - \tau}\mathsf{d}^2(\mathbf{x}, \mathbf{x}_*) > -\infty \qquad \forall\, \mathbf{x} \in \mathfrak{X}, \quad \textit{and}$$

$$\mathsf{d}^2(\mathbf{y}, \mathbf{x}) \leq \frac{4\tau\tau_*}{\tau_* - \tau}\left(\Phi_\tau(\mathbf{x}; \mathbf{y}) - \mathcal{F}_{\tau_*}(\mathbf{x}_*) + \frac{1}{\tau_* - \tau}\mathsf{d}^2(\mathbf{x}, \mathbf{x}_*)\right) \qquad \forall\, \mathbf{x}, \mathbf{y} \in \mathfrak{X}.$$

In particular, the sublevel sets of $\mathbf{y} \mapsto \Phi_\tau(\mathbf{x}; \mathbf{y})$ are bounded in $\mathfrak{X}$.

Proof. An application of Young's inequality gives

$$\begin{aligned}\mathsf{d}^2(\mathbf{y}, \mathbf{x}_*) &\leq (\mathsf{d}(\mathbf{y}, \mathbf{x}) + \mathsf{d}(\mathbf{x}, \mathbf{x}_*))^2 = \mathsf{d}^2(\mathbf{y}, \mathbf{x}) + 2\mathsf{d}(\mathbf{y}, \mathbf{x})\mathsf{d}(\mathbf{x}, \mathbf{x}_*) + \mathsf{d}^2(\mathbf{x}, \mathbf{x}_*)\\ &\leq (1+\epsilon)\mathsf{d}^2(\mathbf{y}, \mathbf{x}) + (1+\epsilon^{-1})\mathsf{d}^2(\mathbf{x}, \mathbf{x}_*).\end{aligned}$$

Choosing $\epsilon = (\tau_* - \tau)/(\tau_* + \tau)$, we obtain

$$\frac{1}{2\tau_*}\mathsf{d}^2(\mathbf{y}, \mathbf{x}_*) \leq \frac{1}{\tau_* + \tau}\mathsf{d}^2(\mathbf{y}, \mathbf{x}) + \frac{1}{\tau_* - \tau}\mathsf{d}^2(\mathbf{x}, \mathbf{x}_*),$$

thus implying that

$$\begin{aligned}\Phi_\tau(\mathbf{x}; \mathbf{y}) = \frac{1}{2\tau}\mathsf{d}^2(\mathbf{y}, \mathbf{x}) + \mathcal{F}(\mathbf{y}) &= \frac{\tau_* - \tau}{2\tau(\tau_* + \tau)}\mathsf{d}^2(\mathbf{y}, \mathbf{x}) + \frac{1}{\tau_* + \tau}\mathsf{d}^2(\mathbf{y}, \mathbf{x}) + \mathcal{F}(\mathbf{y})\\ &\geq \frac{\tau_* - \tau}{2\tau(\tau_* + \tau)}\mathsf{d}^2(\mathbf{y}, \mathbf{x}) + \frac{1}{2\tau_*}\mathsf{d}^2(\mathbf{y}, \mathbf{x}_*) + \mathcal{F}(\mathbf{y}) - \frac{1}{\tau_* - \tau}\mathsf{d}^2(\mathbf{x}, \mathbf{x}_*)\\ &\geq \frac{\tau_* - \tau}{2\tau(\tau_* + \tau)}\mathsf{d}^2(\mathbf{y}, \mathbf{x}) + \mathcal{F}_{\tau_*}(\mathbf{x}_*) - \frac{1}{\tau_* - \tau}\mathsf{d}^2(\mathbf{x}, \mathbf{x}_*) \qquad \forall\, \mathbf{y} \in \mathfrak{X}.\end{aligned}$$

Reordering the terms and using the fact that $\tau < \tau_*$ gives the asserted estimates. □

Applying the Direct Method of the Calculus of Variations, we obtain

Theorem 2.8. *Let $\mathcal{F}\colon \mathfrak{X} \to \mathbb{R}\cup\{+\infty\}$ satisfy ($A^\lambda_{\mathcal{F}}$) and $\tau < \tau_*$. If $\mathcal{F}$ satisfies additionally ($A^1_{\mathcal{F}}$) or ($A^2_{\mathcal{F}}$), then for any $\mathbf{x} \in \mathfrak{X}$,*

the functional $\mathbf{y} \mapsto \Phi_\tau(\mathbf{x}; \mathbf{y})$ admits a minimizer in $\mathfrak{X}$.

In particular, $J_\tau[\mathbf{x}] \neq \emptyset$ for every $\mathbf{x} \in \mathfrak{X}$ and (MM) has a solution

Proof. We only prove the statement for ($A^2_{\mathcal{F}}$) since the result under ($A^1_{\mathcal{F}}$) is easier and follows from a similar argument.

Since $\mathcal{F}$ satisfies ($A^\lambda_{\mathcal{F}}$), Lemma 2.7 shows that $\Phi_\tau(\mathbf{x};\mathbf{y}) \geq \mathcal{F}_\tau(\mathbf{x}) > -\infty$, and is therefore bounded from below. Now take a minimizing sequence $(\mathbf{y}_k)_k \subset \mathfrak{X}$ such that

$$\Phi_\tau(\mathbf{x};\mathbf{y}_k) \to \inf\Big\{\Phi_\tau(\mathbf{x};\mathbf{y}) : \mathbf{y} \in \mathfrak{X}\Big\} =: m > -\infty.$$

Moreover, we have that $\mathcal{F}(\mathbf{y}_k) \leq \Phi_\tau(\mathbf{x};\mathbf{y}_k) \leq c$ for some $c \in \mathbb{R}$ and $\sup_{k,m} \mathsf{d}(\mathbf{y}_k,\mathbf{y}_m) < +\infty$, i.e. the sequence $(\mathbf{y}_k)_k$ is d-bounded. Since $\mathcal{F}$ is σ-coercive on d-bounded sets, the minimizing sequence $(\mathbf{y}_k)_k$ admits a σ-converging subsequence $(\mathbf{y}_{k_\ell})_\ell$ with limit $\bar{\mathbf{y}} \in \mathfrak{X}$ such that $\mathbf{y}_{k_\ell} \stackrel{\sigma}{\rightharpoonup} \bar{\mathbf{y}}$. The σ-lower semicontinuity of d and $\mathcal{F}$ on d-bounded sets then yields

$$\begin{aligned}\Phi_\tau(\mathbf{x};\bar{\mathbf{y}}) &= \frac{1}{2\tau}\mathsf{d}^2(\bar{\mathbf{y}},\mathbf{x}) + \mathcal{F}(\bar{\mathbf{y}}) \leq \frac{1}{2\tau}\liminf_{\ell\to\infty}\mathsf{d}^2(\mathbf{y}_{k_\ell},\mathbf{x}) + \liminf_{\ell\to\infty}\mathcal{F}(\mathbf{y}_{k_\ell})\\ &= \liminf_{\ell\to\infty}\Phi_\tau(\mathbf{x};\mathbf{y}_{k_\ell}) = m,\end{aligned}$$

thus concluding that $\Phi_\tau(\mathbf{x};\bar{\mathbf{y}}) = m$, i.e. $\bar{\mathbf{y}}$ is a minimizer of $\mathbf{y} \mapsto \Phi_\tau(\mathbf{x};\mathbf{y})$. □

2.3 A priori estimates

Theorem 2.8 guarantees that every step in the minimizing-movement scheme (MM) can be solved to obtain a discrete solution $\overline{\mathbf{X}}_\tau$. In the following, we deduce a priori estimates for the discrete solutions that will be essential for extracting a GMM.

Recall that $\mathcal{F}_\tau(\mathbf{x}) := \inf_{\mathbf{y}\in\mathfrak{X}}\Phi_\tau(\mathbf{x};\mathbf{y})$, and for $\mathbf{x}_\tau \in J_\tau[\mathbf{x}]$, we set

$$\mathcal{F}_\tau(\mathbf{x}) := \inf_{\mathbf{y}\in\mathfrak{X}}\Phi_\tau(\mathbf{x},\mathbf{y}) = \Phi_\tau(\mathbf{x};\mathbf{x}_\tau).$$

We further define the Borel map

$$\mathfrak{X} \ni \mathbf{x} \mapsto \mathscr{S}(\mathbf{x}) := \begin{cases}\displaystyle\limsup_{\tau\to 0}\frac{\mathcal{F}(\mathbf{x}) - \mathcal{F}_\tau(\mathbf{x})}{\tau} & \text{for } x \in \operatorname{dom}\mathcal{F}\\ +\infty & \text{otherwise}\end{cases} \in [0,+\infty].$$

Lemma 2.9. *We have that*

$$\mathscr{S}(\mathbf{x}) = \frac{1}{2}|\partial\mathcal{F}|^2(\mathbf{x}) \qquad \forall\, \mathbf{x} \in \mathit{dom}\,\mathcal{F}.$$

Proof. One part of the proof is easy, as

$$\begin{aligned}\mathscr{S}(\mathbf{x}) &= \limsup_{\tau\to 0}\frac{\mathcal{F}(\mathbf{x}) - \mathcal{F}_\tau(\mathbf{x})}{\tau} = \limsup_{\tau\to 0}\left[\frac{\mathcal{F}(\mathbf{x}) - \mathcal{F}(\mathbf{x}_\tau)}{\mathsf{d}(\mathbf{x}_\tau,\mathbf{x})}\frac{\mathsf{d}(\mathbf{x}_\tau,\mathbf{x})}{\tau} - \frac{1}{2}\frac{\mathsf{d}^2(\mathbf{x}_\tau,\mathbf{x})}{\tau^2}\right]\\ &\leq \limsup_{\tau\to 0}\frac{1}{2}\left[\frac{\mathcal{F}(\mathbf{x}) - \mathcal{F}(\mathbf{x}_\tau)}{\mathsf{d}(\mathbf{x}_\tau,\mathbf{x})}\right]^2 \leq \frac{1}{2}|\partial\mathcal{F}|^2(\mathbf{x}).\end{aligned}$$

We refer to [1, Lemma 3.1.5] for the other part of the proof. □

Lemma 2.10. *For* $\mathbf{x} \in dom\,\mathcal{F}$ *and every* $\mathbf{x}_\tau \in J_\tau[\mathbf{x}]$,

$$\mathrm{d}(\mathbf{x}_{\tau_0}, \mathbf{x}) \le \mathrm{d}(\mathbf{x}_{\tau_1}, \mathbf{x}), \qquad \mathcal{F}_{\tau_1}(\mathbf{x}) \le \mathcal{F}_{\tau_0}(\mathbf{x}) \le \mathcal{F}(\mathbf{x}) \qquad \forall\, 0 < \tau_0 < \tau_1 \tag{2.2a}$$

$$\lim_{\tau\to 0} \mathrm{d}(\mathbf{x}_\tau, \mathbf{x}) = 0, \qquad \lim_{\tau\to 0} \mathcal{F}_\tau(\mathbf{x}) = \mathcal{F}(\mathbf{x}) \tag{2.2b}$$

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\mathcal{F}_\tau(\mathbf{x}) \le -\mathscr{S}(\mathbf{x}_\tau) \qquad \textit{for almost every } \tau \in (0, \tau_*) \tag{2.2c}$$

In particular, the following energy-dissipation-type inequality holds:

$$\frac{1}{2\tau}\mathsf{d}^2(\mathbf{x}_\tau, \mathbf{x}) + \int_0^\tau \mathscr{S}(\mathbf{x}_r)\,\mathrm{d}r + \mathcal{F}(\mathbf{x}_\tau) - \mathcal{F}(\mathbf{x}) \le 0 \qquad \textit{for every } \tau \in (0, \tau_*). \tag{2.3}$$

Proof. Let $\tau > 0$, $\mathbf{x} \in \mathrm{dom}\,\mathcal{F}$ and $\mathbf{x}_\tau \in J_\tau[\mathbf{x}]$. Then, for every $\tau_0, \tau_1 \in [0, \tau_*)$ with $\tau_0 < \tau_1$,

$$\begin{aligned}\Phi_{\tau_0}(\mathbf{x}; \mathbf{x}_{\tau_0}) &\le \Phi_{\tau_0}(\mathbf{x}; \mathbf{x}_{\tau_1}) = \frac{1}{2\tau_0}\mathrm{d}^2(\mathbf{x}_{\tau_1}, \mathbf{x}) + \mathcal{F}(\mathbf{x}_{\tau_1}) \\ &= \frac{1}{2}\left(\frac{1}{\tau_0} - \frac{1}{\tau_1}\right)\mathrm{d}^2(\mathbf{x}_{\tau_1}, \mathbf{x}) + \Phi_{\tau_1}(\mathbf{x}, \mathbf{x}_{\tau_1}) \le \frac{1}{2}\left(\frac{1}{\tau_0} - \frac{1}{\tau_1}\right)\mathrm{d}^2(\mathbf{x}_{\tau_1}, \mathbf{x}) + \Phi_{\tau_1}(\mathbf{x}, \mathbf{x}_{\tau_0}),\end{aligned}$$

thus implying the first inequality in (2.2a). The second inequality easily follows from

$$\mathcal{F}_{\tau_1}(\mathbf{x}) - \mathcal{F}_{\tau_0}(\mathbf{x}) \le \Phi_{\tau_1}(\mathbf{x}; \mathbf{x}_{\tau_0}) - \Phi_{\tau_0}(\mathbf{x}; \mathbf{x}_{\tau_0}) = -\frac{\tau_1 - \tau_0}{2\tau_0\tau_1}\mathsf{d}^2(\mathbf{x}_{\tau_0}, \mathbf{x}) \le 0.$$

Using Lemma 2.7 with $\mathbf{y} = \mathbf{x}_\tau$, we obtain

$$\mathsf{d}^2(\mathbf{x}_\tau, \mathbf{x}) \le \frac{4\tau\tau_*}{\tau_* - \tau}\left(\mathcal{F}(\mathbf{x}) - \mathcal{F}_{\tau_*}(\mathbf{x}_*) + \frac{1}{\tau_* - \tau}\mathsf{d}^2(\mathbf{x}, \mathbf{x}_*)\right) \quad \text{for every } \tau \in (0, \tau_*).$$

In particular, we obtain $\mathsf{d}(\mathbf{x}_\tau, \mathbf{x}) \to 0$ as $\tau \to 0$. Moreover, since $\mathcal{F}_\tau(\mathbf{x}) \le \mathcal{F}(\mathbf{x})$, we get

$$\mathcal{F}(\mathbf{x}) \ge \limsup_{\tau\to 0} \mathcal{F}_\tau(\mathbf{x}) \ge \liminf_{\tau\to 0} \mathcal{F}_\tau(\mathbf{x}) \ge \liminf_{\tau\to 0} \mathcal{F}(\mathbf{x}_\tau).$$

If $\mathcal{F}$ satisfies ($A^1_{\mathcal{F}}$), then the d-lowersemicontinuity yields

$$\liminf_{\tau\to 0} \mathcal{F}(\mathbf{x}_\tau) \ge \mathcal{F}(\mathbf{x}).$$

The case when $\mathcal{F}$ satisfies ($A^2_{\mathcal{F}}$) requires a little more attention. Since $\sup_{\tau>0} \mathcal{F}(\mathbf{x}_\tau) \le \mathcal{F}(\mathbf{x})$ and $\sup_{\tau>0} \mathsf{d}(\mathbf{x}_\tau, \mathbf{x}) < +\infty$, and $\mathcal{F}$ is σ-coercive on d-bounded sets, we can extract a σ-converging subsequence $(\mathbf{x}_{\tau_k})_{k\ge 1}$ and some $\widetilde{\mathbf{x}} \in \mathfrak{X}$ such that $\mathbf{x}_{\tau_k} \overset{\sigma}{\rightharpoonup} \widetilde{\mathbf{x}}$. Since $\mathsf{d}(\mathbf{x}_{\tau_k}, \mathbf{x}) \to 0$, we have that $\widetilde{\mathbf{x}} = \mathbf{x}$. Owing the the σ-lower semicontinuity of $\mathcal{F}$ on d-bounded sets, we get

$$\liminf_{\tau_k\to 0} \mathcal{F}(\mathbf{x}_{\tau_k}) \ge \mathcal{F}(\mathbf{x}).$$

In either case, we obtain $\lim_{\tau\to 0} \mathcal{F}_\tau(\mathbf{x}) = \mathcal{F}(\mathbf{x})$, thus proving (2.2b).

We now prove (2.2c). Since the map $(0,\tau_*) \ni \tau \mapsto \mathcal{F}_\tau(\mathbf{x})$ is non-increasing, it is almost everywhere differentiable on $(0,\tau_*)$ by Lebesgue's theorem. Fixing a point of differentiability $\tau \in (0,\tau_*)$, we obtain for $0 < h \ll 1$ and $\mathbf{x}_\tau \in J_\tau[\mathbf{x}]$:

$$\begin{aligned}\frac{\mathcal{F}_{\tau+h}(\mathbf{x}) - \mathcal{F}_\tau(\mathbf{x})}{h} &= \frac{1}{h}\inf_{\mathbf{y}\in\mathfrak{X}}\left\{\frac{1}{2(\tau+h)}\mathsf{d}^2(\mathbf{y},\mathbf{x}) + \mathcal{F}(\mathbf{y}) - \frac{1}{2\tau}\mathsf{d}^2(\mathbf{x}_\tau,\mathbf{x}) - \mathcal{F}(\mathbf{x}_\tau)\right\}\\ &\le \frac{1}{h}\inf_{\mathbf{y}\in\mathfrak{X}}\left\{\frac{1}{2h}\mathsf{d}^2(\mathbf{y},\mathbf{x}_\tau) + \mathcal{F}(\mathbf{y}) - \mathcal{F}(\mathbf{x}_\tau)\right\},\end{aligned}$$

where we used the fact that

$$\frac{1}{2(\tau+h)}\mathsf{d}^2(\mathbf{y},\mathbf{x}) - \frac{1}{2\tau}\mathsf{d}^2(\mathbf{x}_\tau,\mathbf{x}) \le \frac{1}{2h}\mathsf{d}^2(\mathbf{y},\mathbf{x}_\tau).$$

Hence, for almost every $\tau \in (0,\tau_*)$, we find

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\mathcal{F}_\tau(\mathbf{x}) \le \liminf_{h\to 0}\frac{\mathcal{F}_h(\mathbf{x}_\tau) - \mathcal{F}(\mathbf{x}_\tau)}{h} = -\limsup_{h\to 0}\frac{\mathcal{F}(\mathbf{x}_\tau) - \mathcal{F}_h(\mathbf{x}_\tau)}{h} = -\mathscr{S}(\mathbf{x}_\tau).$$

Finally, integrating the previous inequality gives (2.3), therewith completing the proof. $\square$

From the discussion above, we realize that in order to recover the slope in the minimizing-movement scheme, we will need to consider another form of interpolation, the so-called variational interpolation, introduced by De Giorgi.

Definition 2.11 (Variational interpolantion) Let $(\mathbf{X}_\tau^n)_n$ be a solution of the minimizing-movement scheme (MM). A variational interpolant $\widetilde{\mathbf{X}}_\tau : [0,T] \to \mathfrak{X}$ is defined as follows: For each $r \in (0,\tau]$ and I_τ^n, we set

$$\widetilde{\mathbf{X}}_\tau(t) \in J_r[\mathbf{X}_\tau^{n-1}], \qquad t = t_{n-1} + r \in I_\tau^n.$$

We also introduce the real-valued function

$$|\mathbf{V}_\tau|(t) \coloneqq \frac{\mathsf{d}(\mathbf{X}_\tau^n,\mathbf{X}_\tau^{n-1})}{\tau}, \qquad t \in I_\tau^n.$$

An application of Lemma 2.10 then yields

Lemma 2.12. *Let $(\mathbf{X}_\tau^n)_n$ be a solution of the minimizing-movement scheme (MM), and let $|\mathbf{V}_\tau|$ and $\mathbf{G}_\tau$ be defined as above. Then, for $n = 1,\dots,N$,*

$$\int_{I_\tau^n}\left\{\frac{1}{2}|\mathbf{V}_\tau|^2(t) + \frac{1}{2}|\partial\mathcal{F}|^2(\widetilde{\mathbf{X}}_\tau(t))\right\}\mathrm{d}t + \mathcal{F}(\overline{\mathbf{X}}_\tau(t_n)) \le \mathcal{F}(\overline{\mathbf{X}}_\tau(t_{n-1})).$$

Moreover, there are constants c_0, c_1, $c_2 > 0$ independent of $\tau \in (0,\tau_*/4)$ such that

$$\sup_{n\ge 1}\mathsf{d}(\mathbf{X}_\tau^n,\mathbf{x}_*) \le c_0, \qquad \sup_{n\ge 1}\int_0^{t_n}|\mathbf{V}_\tau|^2(t)\,\mathrm{d}t \le c_1, \qquad \sup_{t\in[0,T]}\mathsf{d}(\widetilde{\mathbf{X}}_\tau(t),\overline{\mathbf{X}}_\tau(t)) \le c_2\sqrt{\tau}.$$

Proof. The first part of the statement is easily obtained by setting $\mathbf{x} = \mathbf{X}_\tau^{n-1}$ and $\mathbf{x}_\tau = \mathbf{X}_\tau^n$ in Lemma 2.10.

As for the second part, we will make use of Gronwall's inequality. For each $n \geq 1$,

$$\begin{aligned}\frac{1}{2}\mathsf{d}^2(\mathbf{X}_\tau^n, \mathbf{x}_*) - \frac{1}{2}\mathsf{d}^2(\mathbf{X}_\tau^0, \mathbf{x}_*) &\leq \frac{1}{2}\sum_{j=1}^n \big(\mathsf{d}(\mathbf{X}_\tau^j, \mathbf{x}_*) + \mathsf{d}(\mathbf{X}_\tau^{j-1}, \mathbf{x}_*)\big)\, \mathsf{d}(\mathbf{X}_\tau^j, \mathbf{X}_\tau^{j-1}) \\ &\leq \sum_{j=1}^n \mathsf{d}(\mathbf{X}_\tau^j, \mathbf{x}_*)\, \mathsf{d}(\mathbf{X}_\tau^j, \mathbf{X}_\tau^{j-1}) \\ &\leq \frac{1}{2\varepsilon}\sum_{j=1}^n \tau \mathsf{d}^2(\mathbf{X}_\tau^j, \mathbf{x}_*) + \frac{\varepsilon}{2}\int_0^{t_n} |\mathbf{V}_\tau|^2(t)\,\mathrm{d}t,\end{aligned}$$

with $\varepsilon > 0$ to be chosen below. The last term on the right-hand side can be estimated by

$$\begin{aligned}\frac{1}{2}\int_0^{t_n} |\mathbf{V}_\tau|^2(t)\,\mathrm{d}t &\leq \sum_{j=1}^n \big(\mathcal{F}(\mathbf{X}_\tau^{j-1}) - \mathcal{F}(\mathbf{X}_\tau^j)\big) = \mathcal{F}(\mathbf{X}_\tau^0) - \mathcal{F}(\mathbf{X}_\tau^n) \\ &\leq \mathcal{F}(\mathbf{x}_0) - \mathcal{F}_{\tau_*}(\mathbf{x}_*) + \frac{1}{\tau_* - \tau}\mathsf{d}^2(\mathbf{X}_\tau^n, \mathbf{x}_*),\end{aligned} \tag{2.4}$$

where we used Lemma 2.7, thus implying that

$$\left(\frac{1}{2} - \frac{\varepsilon}{\tau_* - \tau}\right)\mathsf{d}^2(\mathbf{X}_\tau^n, \mathbf{x}_*) \leq \frac{1}{2}\mathsf{d}^2(\mathbf{x}_0, \mathbf{x}_*) + \varepsilon\big[\mathcal{F}(\mathbf{x}_0) - \mathcal{F}_{\tau_*}(\mathbf{x}_*)\big] + \frac{1}{2\varepsilon}\sum_{j=1}^n \tau \mathsf{d}^2(\mathbf{X}_\tau^j, \mathbf{x}_*)$$

Choosing $\varepsilon = (\tau_* - \tau)/4 \geq \tau_*/8$, we arrive at

$$\mathsf{d}^2(\mathbf{X}_\tau^n, \mathbf{x}_*) \leq A + \tau B \sum_{j=1}^n \mathsf{d}^2(\mathbf{X}_\tau^j, \mathbf{x}_*),$$

with $A = 2\,\mathsf{d}^2(\mathbf{x}_0, \mathbf{x}_*) + \tau_*\big[\mathcal{F}(\mathbf{x}_0) - \mathcal{F}_{\tau_*}(\mathbf{x}_*)\big]$ and $B = 4/\tau_*$. An application of the discrete Gronwall inequality (cf. Lemma 2.13 below) then gives

$$\mathsf{d}^2(\mathbf{X}_\tau^n, \mathbf{x}_*) \leq \alpha \exp(\beta\, t_n) \leq \alpha e^{\beta T} =: c_0 \qquad \forall\, n \geq 1,$$

with α and β as given in Lemma 2.13, thus yielding the first estimate. The second estimate is obtained by simply inserting the previous estimate into (2.4). As for the final estimate, we obtain for $t = t_n + r \in I_\tau^n$, $r \in (0, \tau]$,

$$\begin{aligned}\mathsf{d}(\widetilde{\mathbf{X}}_\tau(t), \overline{\mathbf{X}}_\tau(t)) &\leq \mathsf{d}(\widetilde{\mathbf{X}}_\tau(t), \mathbf{X}_\tau^{n-1}) + \mathsf{d}(\mathbf{X}_\tau^n, \mathbf{X}_\tau^{n-1}) \leq 2\mathsf{d}(\mathbf{X}_\tau^n, \mathbf{X}_\tau^{n-1}) \\ &\leq 2\sqrt{\tau}\sqrt{\tau\sum_{j=1}^n |\mathbf{V}_\tau|^2(t_j)} \leq c_2\sqrt{\tau}, \qquad c_2 := 2\sqrt{c_1}.\end{aligned}$$

where the second inequality follows from (2.2a) since $\widetilde{\mathbf{X}}_\tau(t) \in J_r[\mathbf{X}_\tau^{n-1}]$. ∎

Lemma 2.13 (Discrete Gronwall inequality). *Let A, $B \in [0,+\infty)$ and $(a_n)_{n\geq 1} \subset [0,+\infty)$ satisfy the discrete inequality*

$$a_n \leq A + \tau B \sum_{j=1}^{n} a_j \qquad \forall\, n \geq 1, \qquad \text{such that} \quad m := \tau B < 1.$$

Then, setting $\alpha := A/(1-m)$, $\beta := B/(1-m)$ and $\tau_0 = 0$, we have

$$a_n \leq \alpha \exp(\tau \beta\, n) \qquad \forall\, n \geq 1.$$

2.4 Compactness result

With the a priori estimates obtained above, we can now extract a subsequence that converges. To do so, we will require an Arzelà–Ascoli-type result found in [1, Proposition 3.3.1].

Proposition 2.14 (Refined Arzelà–Ascoli theorem). *Let $T > 0$, $\mathcal{K} \subset \mathfrak{X}$ be a σ-sequentially compact set, and let $\mathbf{x}_n\colon [0,T] \to \mathfrak{X}$ be a sequence of curves such that*

$$\mathbf{x}_n(t) \in \mathcal{K} \qquad \forall\, n \in \mathbb{N}, \quad t \in [0,T],$$
$$\limsup_{n\to\infty} \mathsf{d}(\mathbf{x}_n(t), \mathbf{x}_n(s)) \leq \omega(t,s) \qquad \forall\, s,t \in [0,T],$$

for some symmetric function $\omega\colon [0,T]\times[0,T] \to [0,+\infty)$ satisfying

$$\lim_{(s,t)\to(r,r)} \omega(s,t) = 0 \qquad \text{for almost every } r \in [0,T].$$

Then, there exists a subsequence $(\mathbf{x}_{n_k})_k$ and a limit curve $\mathbf{x}\colon [0,T] \to \mathfrak{X}$ such that

$$\mathbf{x}_{n_k}(t) \overset{\sigma}{\rightharpoonup} \mathbf{x}(t) \quad \text{as } k \to \infty \;\; \text{for every } t \in [0,T],$$

and $\mathbf{x}$ is d-continuous almost everywhere in $[0,T]$.

Lemma 2.15. *Let $\mathcal{F}\colon \mathfrak{X} \to \mathbb{R}\cup\{+\infty\}$ satisfy ($A^{\lambda}_{\mathcal{F}}$) and ($A^{2}_{\mathcal{F}}$). Further, let $(\overline{\mathbf{X}}_\tau)_{\tau\in(0,\tau_*/2)}$ be a sequence of discrete solutions obtained from the minimizing-movement scheme (MM). Then, there exists a subsequence $(\overline{\mathbf{X}}_{\tau_k})_k$ and a limit curve $\mathbf{x} \in \mathcal{AC}^2((0,T);\mathfrak{X})$ such that*

$$\overline{\mathbf{X}}_{\tau_k}(t) \overset{\sigma}{\rightharpoonup} \mathbf{x}(t), \qquad \widetilde{\mathbf{X}}_{\tau_k}(t) \overset{\sigma}{\rightharpoonup} \mathbf{x}(t) \quad \text{as } k \to \infty \;\; \text{for every } t \in [0,T],$$
$$\exists\, |\mathbf{v}| \in L^2((0,T)) : \quad |\mathbf{V}_{\tau_k}| \rightharpoonup |\mathbf{v}| \quad \text{weakly in } L^2((0,T)) \text{ as } k \to \infty,$$

where $|\dot{\mathbf{x}}|(t) \leq |v|(t)$ for almost every $t \in (0,T)$.

Proof. From Lemma 2.12 we have that

$$\mathsf{d}(\overline{\mathbf{X}}_\tau(t), \mathbf{x}_*) \leq c_0, \qquad \mathcal{F}(\overline{\mathbf{X}}_\tau(t)) \leq \mathcal{F}(\mathbf{x}_0) \qquad \text{for all } \tau \in (0,\tau_*/2) \text{ and } t \in [0,T]$$

Since $\mathcal{F}$ satisfies ($A^{2}_{\mathcal{F}}$), the sequence $(\overline{\mathbf{X}}_\tau(t))_{\tau\in(0,\tau_*),\, t\in[0,T]}$ is contained in a σ-compact set.

Now let $[s,t] \subset [0,T]$, then we find $s(\tau) \le s \le t \le t(\tau)$ in the partition P_τ such that $s(\tau) \to s$ and $t(\tau) \to t$ as $\tau \to 0$. From the estimate

$$\mathsf{d}(\overline{\mathbf{X}}_\tau(t), \overline{\mathbf{X}}_\tau(s)) = \mathsf{d}(\overline{\mathbf{X}}_\tau(t(\tau)), \overline{\mathbf{X}}_\tau(s(\tau))) \le \int_{s(\tau)}^{t(\tau)} |\mathbf{V}_\tau(r)| \,\mathrm{d}r.$$

Since $(|\mathbf{V}_\tau|)_\tau \subset L^2((0,T))$ is bounded, we can extract a (not relabelled) subsequence and some $|\mathbf{v}| \in L^2((0,T))$ such that $|\mathbf{V}_\tau| \rightharpoonup |\mathbf{v}|$ weakly in $L^2((0,T))$. We then have that

$$\limsup_{\tau\to 0} \mathsf{d}(\overline{\mathbf{X}}_\tau(t), \overline{\mathbf{X}}_\tau(s)) \le \limsup_{\tau\to 0} \int_0^T 1_{[s(\tau),t(\tau)]}(r) |\mathbf{V}_\tau|(r) \,\mathrm{d}r = \int_s^t |\mathbf{v}(r)| \,\mathrm{d}r =: \omega(s,t),$$

where we used the fact that $1_{[s(\tau),t(\tau)]} \to 1_{[s,t]}$ pointwise almost everywhere as $\tau \to 0$. It is easy to see that ω satisfies the property in Proposition 2.14 since

$$\omega(s,t) \le \|\mathbf{v}\|_{L^2((0,T))} \sqrt{|t-s|} \;\longrightarrow\; 0 \qquad \text{as } (s,t) \to (r,r).$$

Consequently, Proposition 2.14 provides a subsequence $(\overline{\mathbf{X}}_{\tau_k})_k$ and a curve $\mathbf{x}$ such that

$$\overline{\mathbf{X}}_{\tau_k}(t) \overset{\sigma}{\rightharpoonup} \mathbf{x}(t) \quad \text{as } k \to \infty \;\text{ for every } t \in [0,T].$$

Moreover, since σ is d-compatible, we have that

$$\mathsf{d}(\mathbf{x}(t), \mathbf{x}(s)) \le \liminf_{\tau\to 0} \mathsf{d}(\overline{\mathbf{X}}_\tau(t), \overline{\mathbf{X}}_\tau(s)) \le \int_s^t |\mathbf{v}|(r) \,\mathrm{d}r.$$

Since $|\mathbf{v}| \in L^2((0,T))$, $\mathbf{x}$ is in $\mathcal{AC}^2((0,T);\mathfrak{X})$ with $|\dot{\mathbf{x}}| \le |\mathbf{v}|$ almost everywhere.

In a similar fashion, we obtain from Lemma 2.12 the following inequalities for $(\widetilde{\mathbf{X}}_\tau)_\tau$:

$$\begin{aligned}
\mathsf{d}(\widetilde{\mathbf{X}}_\tau(t), \mathbf{x}_*) \le \mathsf{d}(\widetilde{\mathbf{X}}_\tau(t), \overline{\mathbf{X}}_\tau(t)) + \mathsf{d}(\overline{\mathbf{X}}_\tau(t), \mathbf{x}_*) \le c_2\sqrt{\tau_*} + c_0,& \\
\mathcal{F}(\widetilde{\mathbf{X}}_\tau(t)) \le \mathcal{F}(\mathbf{X}_\tau^{n-1}) \le \mathcal{F}(\mathbf{x}_0), \qquad t = t_{n-1} + r \in I_\tau^n, \;\; r \in (0,\tau],& \\
\limsup_{\tau\to 0} \mathsf{d}(\widetilde{\mathbf{X}}_\tau(t), \widetilde{\mathbf{X}}_\tau(s)) \le \limsup_{\tau\to 0} \mathsf{d}(\overline{\mathbf{X}}_\tau(t), \overline{\mathbf{X}}_\tau(s)) \le \int_s^t |v|(r) \,\mathrm{d}r,&
\end{aligned}$$

and therewith a subsequence $(\widetilde{\mathbf{X}}_{\tau_\ell})_\ell$ and a curve $\widetilde{\mathbf{x}}$ such that

$$\widetilde{\mathbf{X}}_{\tau_k}(t) \overset{\sigma}{\rightharpoonup} \widetilde{\mathbf{x}}(t) \quad \text{as } \ell \to \infty \;\text{ for every } t \in [0,T].$$

However, upon extracting subsequences we have that

$$\mathsf{d}(\widetilde{\mathbf{x}}(t), \mathbf{x}(t)) \le \liminf_{j\to 0} \mathsf{d}(\widetilde{\mathbf{X}}_{\tau_j}(t), \overline{\mathbf{X}}_{\tau_j}(t)) = 0,$$

i.e. $\widetilde{\mathbf{x}} = \mathbf{x}$, thereby concluding the proof. □

We now have all the ingredients to prove the first part of Theorem 2.4.

2.5 Energy-dissipation inequality

Let the assumptions of Theorem 2.4 hold and $t \subset [0,T]$. As before, we find $t \le t(\tau)$ in the partition P_τ such that $t(\tau) \to t$ as $\tau \to 0$. Morever, take the subsequences $(\overline{\mathbf{X}}_{\tau_j})_j$ and $(\widetilde{\mathbf{X}}_{\tau_j})_j$ as introduced in the proof of Lemma 2.15. From Lemma 2.12, we then obtain

$$\int_0^{t(\tau_j)} \left\{ \frac{1}{2}|\mathbf{V}_{\tau_j}|^2(r) + \frac{1}{2}|\partial\mathcal{F}|^2(\widetilde{\mathbf{X}}_\tau(r)) \right\} \mathrm{d}r + \mathcal{F}(\overline{\mathbf{X}}_\tau(t)) \le \mathcal{F}(\mathbf{x}_0).$$

The assumption $(A^2_{\mathcal{F}})$ on $\mathcal{F}$ then gives

$$\mathcal{F}(\mathbf{x}(t)) \le \liminf_{j\to\infty} \mathcal{F}(\overline{\mathbf{X}}_\tau(t)) \qquad \text{for every } t \in [0,T].$$

Furthermore, the lower semicontinuity of the norm yields

$$\int_0^t |\dot{\mathbf{x}}|^2(r)\,\mathrm{d}r \le \int_0^t |\mathbf{v}|^2(r)\,\mathrm{d}r \le \liminf_{j\to\infty} \int_0^t |\mathbf{V}_{\tau_j}|^2(r)\,\mathrm{d}r \le \liminf_{j\to\infty} \int_0^{t(\tau_j)} |\mathbf{V}_{\tau_j}|^2(r)\,\mathrm{d}r.$$

Finally, we apply Fatou's lemma to obtain

$$\begin{aligned}\int_0^t |\partial^-\mathcal{F}|^2(\mathbf{x}(r))\,\mathrm{d}r &\le \int_0^t \liminf_{j\to\infty} |\partial\mathcal{F}|^2(\widetilde{\mathbf{X}}_\tau(r))\,\mathrm{d}r \\ &\le \liminf_{j\to\infty} \int_0^t |\partial\mathcal{F}|^2(\widetilde{\mathbf{X}}_\tau(r))\,\mathrm{d}r \le \liminf_{j\to\infty} \int_0^{t(\tau_j)} |\partial\mathcal{F}|^2(\widetilde{\mathbf{X}}_\tau(r))\,\mathrm{d}r.\end{aligned}$$

Altogether, we obtain

$$\int_0^t \left\{ \frac{1}{2}|\dot{\mathbf{x}}|^2(r) + \frac{1}{2}|\partial^-\mathcal{F}|^2(\mathbf{x}(r)) \right\} \mathrm{d}r + \mathcal{F}(\mathbf{x}(t)) \le \mathcal{F}(\mathbf{x}_0).$$

The result for arbitrary intervals $[s,t] \subset [0,T]$ can easily be deduced from this.

To conclude the proof, we make use of Theorem 2.16 below on a chain rule inequality to deduce the equality whenever $\mathbf{x} \mapsto |\partial^-\mathcal{F}|(\mathbf{x})$ is a (strong) upper gradient.

2.6 The chain rule inequality

We say that a curve $\mathbf{x} \in \mathcal{AC}^2([0,T];\mathfrak{X})$ is $(\mathfrak{X},\mathsf{d},\mathcal{F})$-*admissible* if

$$\int_0^T \left\{ \frac{1}{2}|\dot{\mathbf{x}}|^2(r) + \frac{1}{2}|\partial^-\mathcal{F}|^2(\mathbf{x}(r)) \right\} \mathrm{d}r < \infty \qquad \text{and} \qquad \sup_{t\in[0,T]} |\mathcal{F}(\mathbf{x}(t))| < \infty.$$

Clearly, if $\mathbf{x}$ is admissible, then $\mathscr{L}(\mathbf{x};[s,t]) < \infty$ for every $[s,t] \subset [0,T]$.

Theorem 2.16 (Chain rule inequality). *Let $\mathcal{F}\colon \mathfrak{X} \to \mathbb{R} \cup \{+\infty\}$ satisfy $(A^\lambda_{\mathcal{F}})$ and $(A^2_{\mathcal{F}})$. Suppose in addition that $\mathbf{x} \mapsto |\partial^-\mathcal{F}|(\mathbf{x})$ is a (strong) upper gradient for $\mathcal{F}$. Then for every admissible curve $\mathbf{x} \in \mathcal{AC}^2((0,T);\mathfrak{X})$, the map $t \mapsto \mathcal{F}(\mathbf{x}(t))$ is absolutely continuous and*

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(\mathbf{x}(t))\right| \le |\partial^-\mathcal{F}|(\mathbf{x}(t))|\dot{\mathbf{x}}|(t) \qquad \textit{for almost every } t \in (0,T).$$

Proof. By assumption, $|\partial^- \mathcal{F}|$ is an upper gradient, and therefore

$$|\mathcal{F}(\mathbf{x}(t)) - \mathcal{F}(\mathbf{x}(s))| \leq \int_s^t |\partial^- \mathcal{F}|(\mathbf{x}(r))|\dot{\mathbf{x}}|(r)\,\mathrm{d}r \qquad \text{for every } (s,t) \subset (0,T).$$

Since the curve $\mathbf{x}$ is admissible, we also have that

$$|\partial_\lambda \mathcal{F}|(\mathbf{x}(t))|\dot{\mathbf{x}}|(t) \leq \frac{1}{2}|\dot{\mathbf{x}}|^2(t) + \frac{1}{2}|\partial^- \mathcal{F}|^2(\mathbf{x}(t)) \qquad \text{for almost every } t \in (0,T),$$

thus implying that the function $t \mapsto |\partial^- \mathcal{F}|(\mathbf{x}(t))|\dot{\mathbf{x}}|(t)$ is in $L^1((0,T))$. Hence, the mapping $t \mapsto \mathcal{F}(\mathbf{x}(t))$ is absolutely continuous as asserted.

Since $t \mapsto \mathcal{F}(\mathbf{x}(t))$ is absolutely continuous, it is differentiable almost everywhere with its derivative satisfying the estimate

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(\mathbf{x}(t))\right| \leq |\partial^- \mathcal{F}|(\mathbf{x}(t))|\dot{\mathbf{x}}|(t) \qquad \text{for almost every } t \in (0,T),$$

thus implying the chain rule inequality. □

Under the assumptions of Theorem 2.16, we then obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(\mathbf{x}(t)) \geq -|\partial^- \mathcal{F}|(\mathbf{x}(t))|\dot{\mathbf{x}}|(t) \geq -\frac{1}{2}|\dot{\mathbf{x}}|^2(t) - \frac{1}{2}|\partial^- \mathcal{F}|^2(\mathbf{x}(t))$$

for almost every $t \in (0,T)$, which, after integrating over arbitrary intervals $[s,t] \subset [0,T]$ and rearranging the terms, yields

$$\mathscr{L}(\mathbf{x},[s,t]) = \int_s^t \left\{\frac{1}{2}|\dot{\mathbf{x}}|^2(r) + \frac{1}{2}|\partial^- \mathcal{F}|^2(\mathbf{x}(r))\right\}\mathrm{d}r + \mathcal{F}(\mathbf{x}(t)) - \mathcal{F}(\mathbf{x}(s)) \geq 0$$

for all $(\mathfrak{X}, \mathsf{d}, \mathcal{F})$-admissible curves $\mathbf{x} \in \mathcal{AC}([0,T];\mathfrak{X})$.

3 Gradient Flows in $W_0^{-1,2}(\Omega)$

Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with Lipschitz boundary throughout this assignment. Consider a *uniformly elliptic* symmetric matrix-valued function $\mathbb{A} \in L^\infty(\Omega; \mathbb{R}^{d\times d})$, i.e.

$$\exists\, C_{\mathbb{A}} > c_{\mathbb{A}} > 0: \qquad c_{\mathbb{A}}|\xi|^2 \le (\mathbb{A}(x)\,\xi, \xi) \le C_{\mathbb{A}}|\xi|^2 \qquad \forall\, x \in \Omega,\ \xi \in \mathbb{R}^d. \tag{3.1}$$

In this assignment, we want to construct a gradient flow solution to

$$\partial_t \rho = \operatorname{div}(\mathbb{A}\nabla\rho), \qquad \rho_0 = \bar\rho \in L^2(\Omega)$$

in the Hilbert space $W_0^{-1,2}(\Omega) := (W_0^{1,2}(\Omega))^*$ with respect to an appropriate norm.

Preliminary We begin by defining the Riesz operator

$$I_{\mathrm{R}}\colon W_0^{-1,2}(\Omega) \to W_0^{1,2}(\Omega), \;\; \eta \mapsto I_{\mathrm{R}}[\varphi] = u,$$

where $u \in W_0^{1,2}(\Omega)$ is the unique weak solution to the variational problem

$$\int_\Omega (\mathbb{A}(x)\nabla u(x), \nabla v(x))\,\mathrm{d}x = \eta[v] \qquad \forall\, v \in W_0^{1,2}(\Omega). \tag{3.2}$$

The Riesz operator provides an isometry between $W_0^{-1,2}(\Omega)$ and $W_0^{1,2}(\Omega)$ with

$$\langle \eta, \zeta\rangle_{-1,\mathbb{A}} = \langle u, v\rangle_{1,\mathbb{A}} = \int_\Omega (\mathbb{A}(x)\nabla u(x), \nabla v(x))\,\mathrm{d}x,$$

where u and $v \in W_0^{1,2}(\Omega)$ solves (3.2) for η and $\zeta \in W_0^{-1,2}(\Omega)$ respectively.

Now let $\mathfrak{X} := (W_0^{-1,2}(\Omega), \|\,\|_{-1,\mathbb{A}})$ be the Hilbert space equipped the norm

$$\|\eta\|_{-1,\mathbb{A}}^2 := \langle \eta, \eta\rangle_{-1,\mathbb{A}} = \int_\Omega (\mathbb{A}(x)\nabla I_{\mathrm{R}}[\eta](x), \nabla I_{\mathrm{R}}[\eta](x))\,\mathrm{d}x =: \|I_{\mathrm{R}}[\eta]\|_{1,\mathbb{A}}^2.$$

Then $\mathfrak{X}^*$ is the Hilbert space $(W_0^{1,2}(\Omega), \|\,\|_{1,\mathbb{A}})$. Observe that since $\mathbb{A}$ is uniformly elliptic (cf. (3.1)), the $\|\,\|_{-1,\mathbb{A}}$-norm is equivalent to the $\|\,\|_{-1,\mathrm{I}_d}$-norm with

$$\sqrt{c_{\mathbb{A}}}\|\eta\|_{-1,\mathbb{A}} \le \|\eta\|_{-1,\mathrm{I}_d} \le \sqrt{C_{\mathbb{A}}}\|\eta\|_{-1,\mathbb{A}} \qquad \forall\, \eta \in \mathfrak{X}.$$

Exercise Let $\mathcal{F}\colon \mathfrak{X} \to \mathbb{R} \cup \{+\infty\}$ be defined by

$$\mathfrak{X} \ni \rho \mapsto \mathcal{F}(\rho) := \begin{cases} \dfrac{1}{2}\displaystyle\int_\Omega |\rho(x)|^2\,\mathrm{d}x & \text{if } \rho \in L^2(\Omega) \\ +\infty & \text{otherwise.} \end{cases}$$

(a) Show that $\mathcal{F}$ satisfies ($A_{\mathcal{F}}^2$), where σ is the weak-$*$ topology.[3]

[3] A sequence $(\rho_n)_n \subset \mathfrak{X}$ converges weakly-$*$ to $\rho \in \mathfrak{X}$ if $\rho_n[v] \to \rho[v]$ for every $v \in \mathfrak{X}^*$.

(b) Show that $\mathcal{F}$ satisfies ($A^\lambda_\mathcal{F}$) and that the local slope takes the form

$$\mathfrak{X} \ni \rho \mapsto |\partial\mathcal{F}|(\rho) = \begin{cases} \|\rho\|_{1,\mathbb{A}} & \text{for } \rho \in \mathfrak{X}^*, \\ +\infty & \text{otherwise.} \end{cases}$$

Moreover, show that $|\partial\mathcal{F}| = |\partial^-\mathcal{F}|$.

Hint: To calculate the slope, use the fact that $\mathfrak{X}^*$ is dense in $\operatorname{dom}\mathcal{F}$.

(c) Deduce the existence of a curve $\rho \in \mathcal{AC}^2((0,T);\mathfrak{X})$ satisfying

$$\int_s^t \frac{1}{2}|\dot\rho|^2_{\mathfrak{X}}(r) + \frac{1}{2}\|\rho(r)\|^2_{1,\mathbb{A}}\,\mathrm{d}r + \mathcal{F}(\rho(t)) \le \mathcal{F}(\rho(s)) \qquad \forall\,[s,t] \subset [0,T].$$

(d) Show that $\mathcal{F}$ is $\|\,\|_{-1,\mathbb{A}}$-lower semicontinuous and conclude that the curve obtained in (c) is a curve of maximal slope for the driving functional $\mathcal{F}$ w.r.t. $|\partial\mathcal{F}|$.

Proof. (a) *σ-lower semicontinuity on d-bounded sets:* Let $(\rho_n)_n \subset \operatorname{dom}\mathcal{F} = L^2(\Omega)$ with $\rho_n \rightharpoonup^* \rho$ in $\mathfrak{X}$. Suppose that $\liminf_{n\to\infty}\mathcal{F}(\rho_n) < +\infty$. Otherwise, there is nothing to prove. In this case, we find a (not relabelled) subsequence for which

$$\lim_{n\to\infty}\mathcal{F}(\rho_n) = \liminf_{n\to\infty}\mathcal{F}(\rho_n) \qquad \text{and} \qquad \sup_n \mathcal{F}(\rho_n) < +\infty.$$

Since $(\rho_n)_n$ is bounded in $L^2(\Omega)$ and $L^2(\Omega)$ is reflexive, there exists a subsequence $(\rho_{n_k})_k \subset L^2(\Omega)$ such that $\rho_{n_k} \rightharpoonup \bar\rho$ weakly in $L^2(\Omega)$ for some $\bar\rho \in L^2(\Omega)$. Consequently,

$$\langle I_\mathrm{R}[\rho_{n_k}] - I_\mathrm{R}[\bar\rho], v\rangle_{1,\mathbb{A}} = \int_\Omega (\rho_{n_k} - \bar\rho)\,v\,\mathrm{d}x \to 0 \qquad \forall\, v \in \mathfrak{X}^*,$$

i.e. $I_\mathrm{R}[\rho_{n_k}] \rightharpoonup I_\mathrm{R}[\bar\rho]$ weakly in $\mathfrak{X}^*$. But since $I_\mathrm{R}[\rho_{n_k}] \rightharpoonup I_\mathrm{R}[\rho]$ in $\mathfrak{X}^*$, we have $\bar\rho = \rho$, thus implying $\rho_{n_k} \rightharpoonup \rho$ in $L^2(\Omega)$. Since this argument holds for arbitrary subsequences, we deduce that $\rho_n \rightharpoonup \rho$ weakly in $L^2(\Omega)$. The assertion then follows from the L^2-weak lower semicontinuity of $\mathcal{F}$.

σ-coercivity on d-bounded sets: In the same spirit, a bounded sequence $(\rho_n)_n \in \mathfrak{X}$ with $\sup_n \mathcal{F}(\rho_n) < +\infty$ has a subsequence $(\rho_{n_k})_k \subset L^2(\Omega)$ and some $\rho \in L^2(\Omega)$ such that $\rho_{n_k} \rightharpoonup \rho$ weakly in $L^2(\Omega)$. In particular, $\rho_{n_k} \rightharpoonup^* \rho$ weakly-$*$ in $\mathfrak{X}$.

(b) Clearly, $\mathcal{F}$ is convex, and therefore λ-convex for $\lambda \in (-\infty, 0]$. Since $\mathcal{F}$ is 0-convex,

$$|\partial\mathcal{F}|(\rho) = \limsup_{\mu\to\rho}\left[\frac{\mathcal{F}(\rho) - \mathcal{F}(\mu)}{\|\rho-\mu\|_{-1,\mathbb{A}}}\right]^+ = \limsup_{h\to 0}\left[\frac{\mathcal{F}(\rho) - \mathcal{F}(\rho+h)}{\|h\|_{-1,\mathbb{A}}}\right]^+,$$

where

$$\frac{\mathcal{F}(\rho) - \mathcal{F}(\rho+h)}{\|h\|_{-1,\mathbb{A}}} = \frac{1}{\|h\|_{-1,\mathbb{A}}}\int_\Omega \rho\,h\,\mathrm{d}x + \frac{1}{2}\frac{\|h\|^2_{L^2}}{\|h\|_{-1,\mathbb{A}}}.$$

Owing to the density of $\mathfrak{X}^* \hookrightarrow L^2(\Omega)$, it suffices to consider $h \in \mathfrak{X}^*$.

We begin by noticing that for every $h \in \mathfrak{X}^*$,

$$\|h\|_{L^2}^2 = \langle I_{\mathrm{R}}[h], h\rangle_{1,\mathbb{A}} \leq \|I_{\mathrm{R}}[h]\|_{1,\mathbb{A}}\|h\|_{1,\mathbb{A}} = \|h\|_{-1,\mathbb{A}}\|h\|_{1,\mathbb{A}}.$$

Hence, the latter term vanishes as $\|h\|_{1,\mathbb{A}} \to 0$. Therefore, we have that

$$|\partial\mathcal{F}|(\rho) = \limsup_{h\to 0}\left\{\frac{1}{\|h\|_{-1,\mathbb{A}}}\int_\Omega \rho\, h \,\mathrm{d}x \;:\; h \in \mathfrak{X}^*\right\}.$$

For any $\rho \in \mathfrak{X}^*$, we find

$$h[\rho] = \langle I_{\mathrm{R}}[h], \rho\rangle_{1,\mathbb{A}} \leq \|\rho\|_{1,\mathbb{A}}\|h\|_{-1,\mathbb{A}} \qquad \forall\, h \in \mathfrak{X}.$$

Making the special choice $h_\rho := \langle \rho, \cdot\,\rangle_{1,\mathbb{A}}$, we have that $h_\rho \in \mathfrak{X}$ and

$$h_\rho[\rho] = \|\rho\|_{1,\mathbb{A}}^2.$$

Together with the previous discussion, we then obtain the desired quantity.

σ-lower semicontinuity on d*-bounded sets:* Let $\rho \in \operatorname{dom}\mathcal{F}$ and $\rho_n \rightharpoonup^* \rho$ weakly-$*$ in $\mathfrak{X}$ with $\sup_n \mathcal{F}(\rho_n) < +\infty$. As in (a), we can find a (not relabelled) subsequence:

$$\lim_{n\to\infty}|\partial\mathcal{F}|(\rho_n) = \liminf_{n\to\infty}|\partial\mathcal{F}|(\rho_n) \qquad \text{and} \qquad \sup_n |\partial\mathcal{F}|(\rho_n) < +\infty.$$

Since $(\rho_n)_n$ is bounded in $\mathfrak{X}^*$ and $\mathfrak{X}^*$ is reflexive, we obtain a subsequence such that $\rho_{n_k} \rightharpoonup \eta$ weakly in $\mathfrak{X}^*$ for some $\eta \in \mathfrak{X}^*$. Since $\rho_n \rightharpoonup^* \rho$ weakly-$*$ in $\mathfrak{X}$, we have that $\eta = \rho$, i.e. $\rho_{n_k} \rightharpoonup \rho$ weakly in $\mathfrak{X}^*$. Moreover, since the argument holds for all subsequences of $(\rho_n)_n$, we have that $\rho_n \rightharpoonup \rho$ weakly in $\mathfrak{X}^*$. The weak lower semicontinuity of $\|\ \|_{1,\mathbb{A}}$ then yields the weak lower semicontinuity of $|\partial\mathcal{F}|$.

(c) We simply apply the first part of Theorem 2.4.

(d) d*-lower semicontinuity:* This follows from (a).

Applying Exercise 1.2 and Lemma 2.6 we have that $|\partial\mathcal{F}| = |\partial_\lambda\mathcal{F}|$ is an upper gradient. We then conclude with the second part of Theorem 2.4.

□

Bonus Exercise Now consider a family $(\mathbb{A}_\varepsilon)_\varepsilon \subset L^\infty(\Omega;\mathbb{R}^{d\times d})$ of uniformly elliptic symmetric matrix-valued functions satisfying

$$\exists\, C_{\mathbb{A}} > c_{\mathbb{A}} > 0 : \qquad c_{\mathbb{A}}|\xi|^2 \leq (\mathbb{A}_\varepsilon(x)\,\xi, \xi) \leq C_{\mathbb{A}}|\xi|^2 \qquad \forall\, x \in \Omega,\ \xi \in \mathbb{R}^d,\ \varepsilon > 0,$$

and such that $\mathbb{A}_\varepsilon \rightharpoonup^* \mathbb{A}_0$ weakly-$*$ in $L^\infty(\Omega;\mathbb{R}^{d\times d})$ for some uniformly elliptic symmetric $\mathbb{A}_0 \in L^\infty(\Omega;\mathbb{R}^{d\times d})$ satisfying the same inequality as above. In the following, we set

$$\mathfrak{X}_\varepsilon := (W_0^{-1,2}(\Omega), \|\ \|_{-1,\mathbb{A}_\varepsilon}), \qquad \overline{\mathfrak{X}} := (W_0^{-1,2}(\Omega), \|\ \|_{-1,\mathrm{I}_d}).$$

Further, let $\rho^\varepsilon \in \mathcal{AC}^2((0,T),\mathfrak{X}_\varepsilon)$, $\varepsilon > 0$, be a curve of maximal slope for the driving functional $\mathcal{F}$ w.r.t. the $\mathfrak{X}_\varepsilon$-local slope

$$\mathfrak{X}_\varepsilon \ni \rho \mapsto |\partial\mathcal{F}|_{\mathfrak{X}_\varepsilon}(\rho) = \begin{cases} \|\rho\|_{1,\mathbb{A}_\varepsilon} & \text{for } \rho \in \mathfrak{X}_\varepsilon, \\ +\infty & \text{otherwise.}\end{cases}$$

(a) Prove the existence of a subsequence $(\rho^{\varepsilon_n})_{n\geq 1} \subset \mathcal{AC}^2((0,T);\overline{\mathfrak{X}})$ and a limit curve $\rho : [0,T] \to \overline{\mathfrak{X}}$ satisfying

$$\rho^{\varepsilon_n}(t) \rightharpoonup^* \rho(t) \quad \text{weakly-$*$ in } \overline{\mathfrak{X}} \text{ for every } t \in [0,T].$$

(b) Show that for any sequence $(\mu^\varepsilon)_\varepsilon$ with $\mu^\varepsilon \rightharpoonup^* \mu$ weakly-$*$ in $\overline{\mathfrak{X}}$,

$$|\partial\mathcal{F}|^2_{\mathfrak{X}_0}(\mu) \leq \liminf_{\varepsilon\to 0} |\partial\mathcal{F}|^2_{\mathfrak{X}_\varepsilon}(\mu^\varepsilon).$$

(c) Deduce the existence of a curve of maximal slope in $\mathcal{AC}^2((0,T);\mathfrak{X}_0)$ for the driving functional $\mathcal{F}$ w.r.t. the $\mathfrak{X}_0$-local slope $|\partial\mathcal{F}|_{\mathfrak{X}_0}$.

4 Gradient Flows in Wasserstein Space

In this section, we are concerned with the formulation of the partial differential equation

$$\partial_t \rho_t = \Delta \rho_t + \operatorname{div}(\rho_t \nabla V) = \operatorname{div}\big(\nabla \rho_t + \rho_t \nabla V\big) \tag{4.1}$$

as a gradient flow in the space of probability measures $\mathcal{P}(\mathbb{R}^d)$ for the driving functional

$$\mathcal{F}(\rho) := \begin{cases} \displaystyle\int_{\mathbb{R}^d} u(x) \log u(x)\,\mathrm{d}x + \int_{\mathbb{R}^d} V(x)\, u(x)\,\mathrm{d}x & \text{if } \rho \ll \mathcal{L}^d,\ u = \dfrac{\mathrm{d}\rho}{\mathrm{d}\mathcal{L}^d}, \\ +\infty & \text{otherwise,} \end{cases} \tag{4.2}$$

where the first term in the functional is called the *Boltzmann entropy*, and the second term the *external energy* for a given *external potential* $V\colon \mathbb{R}^d \to \mathbb{R}$. Here, $\mathcal{L}^d$ denotes the Lebesgue measure on $\mathbb{R}^d$. Henceforth, we call (4.1) the *diffusion equation*.

Motivation Formally, the L^2-variation of $\mathcal{F}$ is given by

$$\mathrm{D}\mathcal{F}(\rho) = \log u + V.$$

Hence, for every $\varphi \in \mathcal{C}_c(\mathbb{R}^d)$, we can write right-hand side of the equation as

$$\int_{\mathbb{R}^d} \varphi\ \mathrm{d}\Big[\Delta\rho + \operatorname{div}(\rho\nabla V)\Big] = -\int_{\mathbb{R}^d} \nabla\varphi \cdot\ \mathrm{d}\Big[\nabla\rho + \rho\nabla V\Big] = \int_{\mathbb{R}^d} \nabla\varphi\cdot\nabla\big(-\mathrm{D}\mathcal{F}(\rho)\big)\,\mathrm{d}\rho,$$

which provides an indication of what the Onsager operator should look like. Indeed, taking

$$\Psi^*(\rho,\xi) = \frac{1}{2}\int_{\mathbb{R}^d} |\nabla\xi|^2\,\mathrm{d}\rho, \qquad \xi \in W^{1,2}(\mathbb{R}^d,\rho),$$

we see that

$$\mathrm{D}_2\Psi^*(\rho,\xi)[\varphi] = \int_{\mathbb{R}^d} \nabla\varphi\cdot\nabla\xi\,\mathrm{d}\rho,$$

and therefore,

$$\partial_t\rho[\varphi] = D_2\Psi^*(\rho, -\mathrm{D}\mathcal{F}(\rho))[\varphi] \qquad \forall\,\varphi \in \mathcal{C}_c(\mathbb{R}^d).$$

Supposing that $\mathrm{D}\mathcal{F}(\rho) \in W^{1,2}(\mathbb{R}^d,\rho)$, we then obtain

$$\Psi(\rho,\partial_t\rho) + \Psi^*(\rho,-\mathrm{D}\mathcal{F}(\rho)) = \langle \partial_t\rho, -\mathrm{D}\mathcal{F}(\rho)\rangle,$$

as suggested by Proposition 1.13.

4.1 The Wasserstein space $\mathbb{W}_2$

Let $\mathcal{P}_p(\mathbb{R}^d)$, $p \in [1,+\infty)$ be the space of probability measures with finite p-moment, i.e.

$$\mathcal{P}_p(\mathbb{R}^d) = \left\{ \mu \in \mathcal{P}(\mathbb{R}^d)\ :\ \int_{\mathbb{R}^d} |x|^p\,\mu(dx) < +\infty \right\}.$$

Definition 4.1 (p-Wasserstein space) The p-Wasserstein space $\mathbb{W}_p(\mathbb{R}^d)$ is the space $\mathcal{P}_p(\mathbb{R}^d)$ equipped with the p-Wasserstein metric d_p defined by

$$\mathsf{d}_p^p(\mu,\nu) = \inf\left\{ \iint_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p\,\gamma(dxdy) \;:\; \gamma \in \Gamma(\mu,\nu)\right\},$$

where $\Gamma(\mu,\nu)$ is the set of *transport plans*, i.e.

$$\Gamma(\mu,\nu) = \left\{\gamma \in \mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d) \;:\; \pi^0_\#\gamma = \mu,\; \pi^1_\#\gamma = \nu\right\}.$$

Here, $\pi^0(x,y) = x$ and $\pi^1(x,y) = y$ are the projections of $\mathbb{R}^d\times\mathbb{R}^d$ onto its factors, and $\#$ is the push-forward operation, i.e. for any measurable map $f\colon (X,\mathcal{B}_X) \to (Y,\mathcal{B}_Y)$ between measurable spaces, and $\gamma \in \mathcal{P}(X)$, $f_\#\gamma \in \mathcal{P}(Y)$ defines a probability measure on Y via

$$(f_\#\gamma)(A) = \gamma\big(f^{-1}(A)\big) \qquad \text{for all } A \in \mathcal{B}_Y,$$

with $f^{-1}(A) \subset X$ being the pre-image of A.

In the following, we simply denote $\mathbb{W}_p = \mathbb{W}_p(\mathbb{R}^d)$ and $\Gamma_0(\mu,\nu)$ the set of *optimal* plans.

$\mathbb{W}_p$ as a geodesic space.

(1) $\mathbb{W}_p$ is a geodesic space, i.e. for every $\mu,\nu \in \mathbb{W}_p$, there exists a curve $\rho : [0,T] \to \mathbb{W}_p$ such that $\rho_0 = \mu$, $\rho_1 = \nu$ and

$$\mathsf{d}_p(\rho_t,\rho_s) = (t-s)\,\mathsf{d}_p(\mu,\nu) \qquad \forall\,[s,t] \subset [0,1].$$

(2) (Brenier) If $\mu,\nu \in \mathbb{W}_p$, $p \in (1,+\infty)$, and μ has Lebesgue density ($\mu \ll \mathcal{L}^d$), then there is a unique μ-measurable map $\mathsf{T}\colon \mathbb{R}^d \to \mathbb{R}^d$ such that $\mathsf{T}_\#\mu = \nu$ and

$$\mathsf{d}_p^p(\mu,\nu) = \int_{\mathbb{R}^d} |x - \mathsf{T}(x)|^p\,\mu(dx),$$

i.e. the transport plan $\gamma \coloneqq (id\times\mathsf{T})_\#\mu$ is optimal. The map T is called the *transport map* and is the gradient of a convex function $\psi\colon \mathbb{R}^d \to \mathbb{R}$, i.e. $\mathsf{T} = \nabla\psi$. In particular, T is differentiable almost everywhere due to Aleksandrov's result (cf. Proposition 4.2).

(3) (McCann's displacement interpolation) If $\gamma \in \Gamma_0(\mu,\nu)$ is an optimal transport plan, then the map $\pi^\theta\colon \mathbb{R}^d\times\mathbb{R}^d \to \mathbb{R}^d$, $\pi^\theta \coloneqq (1-\theta)x + \theta y$, defines a (constant speed) geodesic curve

$$[0,1] \ni \theta \mapsto \rho_\theta \coloneqq \pi^\theta_\#\gamma, \qquad \rho_0 = \mu,\quad \rho_1 = \nu.$$

In this case when $\mu \ll \mathcal{L}^d$ and T is an optimal transport map, the map $\mathsf{T}_\theta\colon \mathbb{R}^d \to \mathbb{R}^d$, π^θ takes the form $\pi^\theta = \mathsf{T}_\theta(x) \coloneqq (1-\theta)x + \theta\mathsf{T}(x)$.

Proposition 4.2 (Aleksandrov)**.** *Let $\psi\colon \mathbb{R}^d \to \mathbb{R}$ be a convex function. Then ψ, $\nabla\psi$ are differentiable almost everywhere. Moreover, $\nabla^2\psi(x)$ is symmetric and positive definite for almost every $x \in \mathbb{R}^d$.*

The narrow topology in $\mathcal{P}(\mathbb{R}^d)$. A sequence $(\mu_n)_{n\geq 1} \subset \mathcal{P}(\mathbb{R}^d)$ of probability measures is said to converge *narrowly* to $\mu \in \mathcal{P}(\mathbb{R}^d)$ as $n \to \infty$ if

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} \varphi(x)\,\mu_n(dx) \longrightarrow \lim_{n\to\infty} \int_{\mathbb{R}^d} \varphi(x)\,\mu(dx) \qquad \forall\,\varphi \in \mathcal{C}_b(\mathbb{R}^d),$$

where $\mathcal{C}_b(\mathbb{R}^d)$ is the space of continuous and bounded functions on $\mathbb{R}^d$.

Remark 4.3 The narrow topology coincides with the weak-$*$ topology of $(\mathcal{C}_b(\mathbb{R}^d))^*$ since $\mathcal{P}(\mathbb{R}^d)$ can be identified with a convex subset of the unitary ball of $(\mathcal{C}_b(\mathbb{R}^d))^*$.

The following theorem provides a characterization of relative compact sets under the narrow topology in $\mathcal{P}(\mathbb{R}^d)$.

Proposition 4.4 (Prokhorov). *A set $\mathcal{K} \subset \mathcal{P}(\mathbb{R}^d)$ is **tight**, i.e.*

$$\forall\,\varepsilon > 0 \ \ \exists\,K_\varepsilon \subset \mathbb{R}^d \textit{ compact, such that } \ \ \mu(\mathbb{R}^d \setminus K_\varepsilon) \leq \varepsilon \quad \forall\,\mu \in \mathcal{K}, \tag{4.3}$$

if and only if $\mathcal{K}$ is relatively compact in $\mathcal{P}(\mathbb{R}^d)$.

In practice, one usually encounters another criterion for tightness.

Theorem 4.5 (Integral condition for tightness). *A set $\mathcal{K} \subset \mathcal{P}(\mathbb{R}^d)$ is tight (cf. (4.3)) if and only if there exists a function $\Phi\colon \mathbb{R}^d \to [0,+\infty]$, whose sublevel sets $L_\Phi(c) = \{x \in \mathbb{R}^d \,:\, \Phi(x) \leq c\}$, $c \in \mathbb{R}$ are compact in $\mathbb{R}^d$, such that*

$$\sup_{\mu\in\mathcal{K}} \int_{\mathbb{R}^d} \Phi(x)\,\mu(dx) < +\infty. \tag{4.4}$$

Proof. Let $(\varepsilon_n)_{n\geq 1}$ be a sequence with $\sum_{n=1}^\infty \varepsilon_n < +\infty$ and $K_n := K_{\varepsilon_n} \subset \mathbb{R}^d$ is an increasing sequence of compact sets satisfying (4.3). Then the function

$$\Phi(x) := \inf\Big\{n \geq 1 \ :\ x \in K_n\Big\} = \sum_{n=1}^\infty \mathbb{1}_{\mathbb{R}^d\setminus K_n}(x), \qquad x \in \mathbb{R}^d,$$

satisfies

$$\int_{\mathbb{R}^d} \Phi(x)\,\mu(dx) = \int_{\mathbb{R}^d} \sum_{n=1}^\infty \mathbb{1}_{\mathbb{R}^d\setminus K_n}(x)\,\mu(dx) \leq \sum_{n=1}^\infty \varepsilon_n < +\infty,$$

where the inequality follows from the monotone convergence theorem.

Now let Φ be a function satisfying the assumption with

$$\sup_{\mu\in\mathcal{K}} \int_{\mathbb{R}^d} \Phi(x)\,\mu(dx) =: c_0 < +\infty.$$

Let $\varepsilon > 0$ be arbitrary and set $K_\varepsilon := L_\Phi(c_0/\varepsilon)$, which is compact since any sublevel set of Φ is compact. We then have that

$$\mu(\mathbb{R}^d \setminus K_\varepsilon) = \mu\big(\{x \in \mathbb{R}^d \,:\, \Phi(x) > c_0/\varepsilon\}\big) \leq \frac{\varepsilon}{c_0} \int_{\mathbb{R}^d} \Phi(x)\,\mu(dx) \leq \varepsilon \qquad \forall\,\mu \in \mathcal{K}.$$

therewith concluding the proof. □

Remark 4.6 Proposition 4.5 tells us that sequences $(\mu_n)_{n\geq 1} \subset \mathbb{W}_p$ satisfying a uniform p-moment bound

$$\sup_{n\geq 1} \int_{\mathbb{R}^d} |x|^p \, \mu_n(dx) < +\infty.$$

are in fact relatively compact w.r.t. the narrow topology. Indeed, since the sublevel sets of $\mathbf{x} \mapsto |x|^p$ are compact in $\mathbb{R}^d$, Proposition 4.5 directly implies the relative compactness of $(\mu_n)_{n\geq 1}$ under the narrow topology.

Another important feature of the narrow topology is the following.

Proposition 4.7. *Let $(\mu_n)_{n\geq 1}, (\nu_n)_{n\geq 1} \subset \mathcal{P}_p(\mathbb{R}^d)$ be narrowly converging sequences with narrow limits μ and $\nu \in \mathcal{P}(\mathbb{R}^d)$ respectively, and $\gamma_n \in \Gamma_0(\mu_n, \nu_n)$ be a sequence of optimal plans with*

$$\sup_{n\geq 1} \iint_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p \, \gamma_n(dxdy) < +\infty.$$

Then, $(\gamma_n)_{n\geq 1}$ is narrowly relatively compact in $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ and any narrow limit γ belongs to $\Gamma_0(\mu, \nu)$ with

$$\mathsf{d}_p(\mu, \nu) \leq \liminf_{n\to\infty} \mathsf{d}_p(\mu_n, \nu_n),$$

i.e. the narrow topology is d_p-compatible.

We conclude this section with the equivalence of narrow convergence and d_p-convergence for sets with uniformly integrable p-moments.

Proposition 4.8 (Convergence, compactness and completeness)**.** *$\mathbb{W}_p$ is a complete separable metric space. A set $\mathcal{K} \subset \mathcal{P}_p(\mathbb{R}^d)$ is d_p-relatively compact if and only if it is tight and has uniformly integrable p-moments.*

In particular, for a given sequence $(\mu_n)_{n\geq 1} \subset \mathcal{P}_p(\mathbb{R}^d)$,

$$\lim_{n\to\infty} \mathsf{d}_p(\mu_n, \mu) \qquad \Longleftrightarrow \qquad \begin{cases} \mu_n \text{ converges narrowly to } \mu, \\ (\mu_n)_{n\geq 1} \text{ has uniformly integrable } p\text{-moments.} \end{cases}$$

4.2 Absolutely continuous curves in $\mathbb{W}_2$

Notice that (4.1) can be expressed as

$$\partial_t \rho_t + \operatorname{div}(\rho_t \mathbf{v}_t) = 0, \tag{CE}$$
$$\mathbf{v}_t = -\nabla \mathrm{D}\mathcal{F}(\rho_t), \tag{KR}$$

where (CE) is the *continuity equation* and (KR) is called the *kinetic relation.*

The continuity equation will play an essential role in understanding the geometry of $\mathbb{W}_2$.

Definition 4.9 (Continuity equation) A density-velocity pair $(\rho, \mathbf{v})$, where $\rho\colon [a,b] \to \mathbb{W}_2$, $\mathbf{v}\colon (a,b)\times\mathbb{R}^d \to \mathbb{R}^d$, is said to satisfy the continuity equation (CE) if for every $[s,t] \subset [a,b]$,

$$\int_{\mathbb{R}^d} \varphi(x)\, \rho_t(dx) - \int_{\mathbb{R}^d} \varphi(x)\, \rho_s(dx) = \int_s^t \int_{\mathbb{R}^d} \nabla\varphi(x)\cdot \mathbf{v}_t(x)\, \rho_t(dx) \qquad \forall\, \varphi \in \mathcal{C}_c^\infty(\mathbb{R}^d).$$

A special type of vector field is one that is an element of the tangent bundle over $\mathbb{W}_2$.

Definition 4.10 (Tangent bundle over $\mathbb{W}_2$) Let $\mu \in \mathcal{P}_p(\mathbb{R}^d)$. We define the $\mathbb{W}_2$-tangent space in μ as

$$\mathrm{Tan}_\mu \mathbb{W}_2 := \overline{\left\{\nabla\varphi \,:\, \varphi \in \mathcal{C}_c^\infty(\mathbb{R}^d)\right\}}^{L^2(\mu;\mathbb{R}^d)}.$$

The tangent bundle $\mathrm{Tan}\,\mathbb{W}_2$ is then the disjoint union in μ of the tangent spaces $\mathrm{Tan}_\mu \mathbb{W}_2$.

The following theorem characterizes $\mathbb{W}_2$-absolutely continuous curves with density-velocity pairs $(\rho, \mathbf{v})$ satisfying the continuity equation (CE).

Proposition 4.11. *Let $(\rho, \mathbf{v})$ be a density-velocity pair satisfying the continuity equation* (CE)*, where $\mathbf{v}_t \in L^2(\rho_t; \mathbb{R}^d)$ for a.e. $t \in (a, b)$ and*

$$\int_a^b \|\mathbf{v}_t\|^2_{L^2(\rho_t)}\,\mathrm{d}t < +\infty.$$

Then ρ is a 2-absolutely continuous curve in $\mathbb{W}_2$ with $|\dot\rho|(t) \le \|\mathbf{v}_t\|_{L^2(\rho_t)}$ for a.e. $t \in (a, b)$. If, in addition, $\mathbf{v}_t \in Tan_{\rho_t}\mathbb{W}_2$ for a.e. $t \in (a, b)$, then also $\|\mathbf{v}_t\|_{L^2(\rho_t)} \le |\dot\rho|(t)$.

Conversely, let $\rho\colon [a, b] \to \mathbb{W}_2$ be a 2-absolutely continuous curve in $\mathbb{W}_2$. Then for almost every $t \in (a, b)$, there exists a velocity field $\mathbf{v}_t \in Tan_{\rho_t}\mathbb{W}_2$ such that

(i) the continuity equation (CE) *is satisfied in the sense of Definition 4.9.*

(ii) $\|\mathbf{v}_t\|_{L^2(\rho_t)} = |\dot\rho|(t)$ for a.e. $t \in (a, b)$, where $|\dot\rho|$ is the $\mathbb{W}_2$-metric derivative of ρ.

Due to Proposition 4.11, a geodesic curve ρ between $\mu, \nu \in \mathbb{W}_2$ can be characterized as

$$\rho \in \operatorname{argmin}\left\{\int_0^1\int_{\mathbb{R}^d} |\mathbf{v}_t(x)|^2 \rho_t(dx)\,\mathrm{d}t \;:\; (\rho, \mathbf{v}) \text{ satisfies (CE)},\; \rho_0 = \mu,\; \rho_1 = \nu\right\}.$$

In particular, the 2-Wasserstein metric d_2 takes the form

$$\mathsf{d}_2^2(\mu, \nu) = \min\left\{\int_0^1\int_{\mathbb{R}^d} |\mathbf{v}_t(x)|^2 \rho_t(dx)\,\mathrm{d}t \;:\; (\rho, \mathbf{v}) \text{ satisfies (CE)},\; \rho_0 = \mu,\; \rho_1 = \nu\right\}.$$

This formulation of d_2 is commonly called the *dynamical* Benamou–Brenier formulation.

4.3 Properties of functionals on $\mathbb{W}_2$

Let us now check to see if our driving functional $\mathcal{F}$ satisfies ($A^\lambda_{\mathcal{F}}$) and ($A^2_{\mathcal{F}}$). In the following, we denote the narrow topology in $\mathcal{P}(\mathbb{R}^d)$ by σ, which we know to be d_2-compatible due to Proposition 4.7.

We consider two classes of functionals, namely the so-called *internal energy*

$$\mathcal{P}(\mathbb{R}^d) \ni \rho \mapsto \mathcal{E}_\phi(\rho) := \begin{cases} \displaystyle\int_{\mathbb{R}^d} \phi(u(x))\,\mathrm{d}x & \text{if } \rho \ll \mathcal{L}^d,\;\; u = \dfrac{\mathrm{d}\rho}{\mathrm{d}\mathcal{L}^d}, \\ +\infty & \text{otherwise,} \end{cases}$$

where $\phi : \mathbb{R} \to \mathbb{R}$ is a superlinear convex function, and the so-called *external energy*

$$\mathcal{P}(\mathbb{R}^d) \ni \rho \mapsto \mathcal{V}(\rho) := \int_{\mathbb{R}^d} V(x)\,\rho(dx),$$

for some sufficiently regular function $V\colon \mathbb{R}^d \to \mathbb{R}$, also called the *potential* function.

Geodesic λ-convexity in $\mathbb{W}_2$. Our first result is on the λ-convexity of the external energy under a λ-convexity assumption on the potential function V.

Theorem 4.12. *Let $V\colon \mathbb{R}^d \to \mathbb{R}$ be a λ-convex function, i.e. there exists some $\lambda \in \mathbb{R}$:*

$$V((1-\theta)x + \theta y) \leq (1-\theta)V(x) + \theta V(y) - \frac{\lambda}{2}\theta(1-\theta)|x-y|^2 \qquad \forall\, \theta \in [0,1].$$

Then, $\mathcal{V}\colon \mathbb{W}_2 \to \mathbb{R}$ is geodesic λ-convex.

Proof. Let $\mu, \nu \in \mathbb{W}_2$ be arbitrary and $\rho_\theta := \pi^\theta_\# \gamma$, $\gamma \in \Gamma_0(\mu,\nu)$ be a geodesic connecting μ and ν, where $\pi^\theta(x,y) = (1-\theta)x + \theta y$. Then

$$\begin{aligned}
\mathcal{V}(\rho_\theta) &= \iint_{\mathbb{R}^d\times\mathbb{R}^d} V((1-\theta)x+\theta y)\,\gamma(dxdy) \\
&\leq \iint_{\mathbb{R}^d\times\mathbb{R}^d} (1-\theta)V(x) + \theta V(y) - \frac{\lambda}{2}\theta(1-\theta)|x-y|^2\,\gamma(dxdy) \\
&= (1-\theta)\mathcal{V}(\mu) + \theta\mathcal{V}(\nu) - \frac{\lambda}{2}\theta(1-\theta)\,\mathsf{d}_2^2(\mu,\nu),
\end{aligned}$$

where the last equality follows from the optimality of $\gamma \in \Gamma_0(\mu,\nu)$. □

The more interesting result is on the geodesic λ-convexity of the internal energy $\mathcal{E}_\phi$, which relies on the change-of-variables formula for measures with densities:

Proposition 4.13 (Change-of-variables formula)**.** *Let $\mathsf{T}\colon \mathbb{R}^d \to \mathbb{R}^d$ be an injective Lipschitz map with $\det(D\mathsf{T}(x)) > 0$ for almost every $x \in \mathbb{R}^d$. If μ has Lebesgue-density u, i.e. $u = \mathrm{d}\mu/\mathrm{d}\mathcal{L}^d$ and $\nu := \mathsf{T}_\#\mu$. Then, $\nu \ll \mathcal{L}^d$ and has density*

$$v(y) := \frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^d}(y) = \frac{u}{\det(D\mathsf{T})} \circ \mathsf{T}^{-1}(y) \qquad \text{for almost every } y \in \mathbb{R}^d.$$

In this case, for any Borel function $\phi\colon \mathbb{R} \to \mathbb{R}\cup\{+\infty\}$ with $\phi(0) = 0$, we have that

$$\int_{\mathbb{R}^d} \phi\big(v(y)\big)\,\mathrm{d}y = \int_{\mathbb{R}^d} \phi\left(\frac{u(x)}{\det(D\mathsf{T})(x)}\right)\det(D\mathsf{T})(x)\,\mathrm{d}x.$$

Another ingredient we will need is the following result on the determinant of matrices.

Lemma 4.14. *Let A, B be two $d\times d$ symmetric and positive definite matrices. Then the map $[0,1] \ni \theta \mapsto g(\theta) := \det((1-\theta)A + \theta B)^{1/d}$ is concave.*

Theorem 4.15. *Let $s \mapsto s^d\phi(s^{-d})$ be convex and decreasing with $\phi(0) = 0$. Then the internal energy $\mathcal{E}_\phi$ is geodesically convex in $\mathbb{W}_2$.*

Proof. Let $\mu, \nu \in \operatorname{dom}\mathcal{E}_\phi$. Hence, $\mu,\nu \ll \mathcal{L}^d$ with densities $u = \mathrm{d}\mu/\mathcal{L}^d$ and $v = \mathrm{d}\nu/\mathcal{L}^d$ respectively. Consider the (unique) transport map T between μ and ν, and correspondingly the unique geodesic $\rho_\theta := (\mathsf{T}_\theta)_\#\mu$, where $\mathsf{T}_\theta(x) = (1-\theta)x + \theta\mathsf{T}(x)$ for μ-a.e. $x \in \mathbb{R}^d$. Since

T is differentiable almost everywhere with $\det(D\mathsf{T}) > 0$, T_θ inherits the same properties, thus implying that ρ_θ has Lebesgue-density

$$u_\theta = \frac{u}{\det(D\mathsf{T}_\theta)} \circ \mathsf{T}_\theta^{-1}.$$

It then follows that

$$\mathcal{E}_\phi(\rho_\theta) = \int_{\mathbb{R}^d} \phi\big(u_\theta(y)\big)\,\mathrm{d}y = \int_{\mathbb{R}^d} \phi\left(\frac{u(x)}{\det(D\mathsf{T}_\theta(x))}\right) \det(D\mathsf{T}_\theta(x))\,\mathrm{d}x.$$

Since I_d and $\mathsf{T}(x)$ are symmetric and positive definite for μ-a.e. $x \in \mathbb{R}^d$, the map

$$[0,1] \ni \theta \mapsto \det(D\mathsf{T}_\theta(x)) \qquad \text{is concave (cf. Lemma 4.14).}$$

Since the integrand above can be seen as a composition of the convex and nonincreasing map $s \mapsto s^d\phi(u(x)s^{-d})$ and a concave map, the resulting map is convex in $[0,1]$ for μ-a.e. $x \in \mathbb{R}^d$, and therefore

$$\phi\left(\frac{u(x)}{\det(D\mathsf{T}_\theta(x))}\right) \det(D\mathsf{T}_\theta(x)) \le (1-\theta)\,\phi\big(u(x)\big) + \theta\,\phi\left(\frac{u(x)}{\det(D\mathsf{T}(x))}\right) \det(D\mathsf{T}(x)).$$

Integrating over $x \in \mathbb{R}^d$ and using the fact that $\nu = \mathsf{T}_\#\mu$ yields

$$\mathcal{E}_\phi(\rho_\theta) \le (1-\theta)\,\mathcal{E}_\phi(\mu) + \theta\,\mathcal{E}_\phi(\nu) \qquad \forall\,\theta \in [0,1],$$

thus showing that $\mathcal{E}_\phi$ is geodesically convex. □

From Theorem 4.12 and Theorem 4.15 we deduce that the driving functional $\mathcal{F}$ is geodesically λ-convex with λ given in Theorem 4.12.

Example 4.16 Internal energies satisfying the assumptions of Theorem 4.15 include the Boltzmann entropy $\phi(s) = s\,log(s)$ and the power laws $\phi(s) = s^p/(p-1)$, $p \ge 1 - 1/d$.

Assumption $(A_{\mathcal{F}}^\lambda)$ The first part of $(A_{\mathcal{F}}^\lambda)$ is easily satisfied for $\mathcal{F}$ due to the fact that $\mathcal{E}_\phi$ with $\phi(s) = s\log(s)$ is geodesically convex and $\mathcal{V}$ is geodesically λ-convex whenever V is λ-convex, resulting in the following statement.

Lemma 4.17. *Let $V\colon \mathbb{R}^d \to \mathbb{R}$ be λ-convex. Then, $\mathcal{F}$ is geodesically λ-convex.*

As for the second part of the statement, we will consider an assumption on the external potential V that appears frequently in Physics and other fields, including Data Science.

Lemma 4.18. *Let $V\colon \mathbb{R}^d \to \mathbb{R}$ be such that $e^{-V} \in L^1(\mathbb{R}^d)$ and define the measure*

$$\sigma := \frac{e^{-V}}{\|e^{-V}\|_{L^1}} \mathcal{L}^d \in \mathcal{P}(\mathbb{R}^d).$$

Then for every $\rho \in dom\,\mathcal{F}$,

$$\mathcal{F}(\rho) = \int_{\mathbb{R}^d} \phi_B\left(\frac{\mathrm{d}\rho}{\mathrm{d}\sigma}\right) \mathrm{d}\sigma - \log\|e^{-V}\|_{L^1} \ge -\log\|e^{-V}\|_{L^1} > -\infty,$$

where $\phi_B(s) := s\log(s) - s + 1 \ge 0$ for $s \in [0,\infty)$.

Remark 4.19 The functional

$$\mathcal{P}(\mathbb{R}^d) \ni \rho \mapsto \mathrm{Ent}(\rho|\sigma) := \begin{cases} \displaystyle\int_{\mathbb{R}^d} \phi_B\left(\frac{\mathrm{d}\rho}{\mathrm{d}\sigma}\right) \mathrm{d}\sigma & \text{if } \rho \ll \sigma, \\ +\infty & \text{otherwise,} \end{cases}$$

is known as the *relative entropy* (or the *Kullback–Leibler divergence*) of ρ w.r.t. σ.

Proof of Lemma 4.18. Let $\rho \in \operatorname{dom} \mathcal{F}$ with $\rho = u\mathcal{L}^d$ and set $v = e^{-V}/\|e^{-V}\|_{L^1}$. Then,

$$\begin{aligned} \mathcal{F}(\rho) &= \int_{\mathbb{R}^d} \log(u(x))\, u(x)\, \mathrm{d}x - \int_{\mathbb{R}^d} \log(v(x))\, u(x)\, \mathrm{d}x - \log \|e^{-V}\|_{L^1}, \\ &= \int_{\mathbb{R}^d} \log\left(\frac{u(x)}{v(x)}\right) \frac{u(x)}{v(x)}\, v(x)\, \mathrm{d}x - \log \|e^{-V}\|_{L^1} \\ &= \int_{\mathbb{R}^d} \phi_B\left(\frac{u(x)}{v(x)}\right) v(x)\, \mathrm{d}x - \log \|e^{-V}\|_{L^1}, \end{aligned}$$

which is as required. □

Assumption $(A^2_{\mathcal{F}})$ We now turn to show that $\mathcal{F}$ satisfies $(A^2_{\mathcal{F}})$.

Lemma 4.20. *Let $V \colon \mathbb{R}^d \to \mathbb{R}$ be lower semicontinuous. Then, the driving functional $\mathcal{F}$ is narrowly lower semicontinuous.*

In particular, $\mathcal{F}$ is d_2-lower semicontinuous, and $|\partial\mathcal{F}|$ is a strong upper gradient.

Proof. The first part of the proof follows from the standard results from the Calculus of Variations. The d_2-lower semicontinuity of $\mathcal{F}$ follows from the fact that convergence in d_2 implies narrow convergence (cf. Proposition 4.8), and the strong upper gradient property of $|\partial\mathcal{F}|$ follows from Lemma 2.6. □

Finally, for the narrow coercivity, we will need the following characterisation of weak-L^1 relative compactness.

Proposition 4.21 (Dunford-Pettis/De la Vallée Poussin)**.** *Let $\sigma \in \mathcal{P}(\mathbb{R}^d)$ and $\mathcal{K} \subset L^1(\sigma)$ be a bounded set. Then the following are equivalent:*

(1) $\mathcal{K}$ is relatively compact in the weak-L^1 topology.

(2) $\mathcal{K}$ is uniformly integrable, i.e. $\forall\, \varepsilon > 0$, $\exists\, \delta > 0$ such that

$$\sup_{f \in \mathcal{K}} \int_A |f(x)|\, \sigma(dx) < \varepsilon \qquad \forall \text{ measurable } A \subset \mathbb{R}^d \text{ with } |A| < \delta.$$

(3) there exists a non-decreasing convex function $\phi\colon [0,\infty) \to [0,\infty)$ that is superlinear at infinity, i.e. $\lim_{s\to\infty} \phi(s)/s = +\infty$, such that

$$\sup_{f \in \mathcal{K}} \int_{\mathbb{R}^d} \phi\big(|f(x)|\big)\, \sigma(dx) < +\infty.$$

The previous proposition then gives

Lemma 4.22. *Let $V\colon \mathbb{R}^d \to \mathbb{R}$ be such that $e^{-V} \in L^1(\mathbb{R}^d)$. Then, $\mathcal{F}$ is narrowly coercive.*

Proof. Let $c \in \mathbb{R}$ and $L_{\mathcal{F}}(c)$ be a sublevel set of $\mathcal{F}$. From Lemma 4.18, we have that

$$\mathcal{F}(\rho) = \int_{\mathbb{R}^d} \phi_B\left(\frac{\mathrm{d}\rho}{\mathrm{d}\sigma}\right) \mathrm{d}\sigma - \log \|e^{-V}\|_{L^1}.$$

Hence,

$$\int_{\mathbb{R}^d} \tilde{\phi}_B\left(\frac{\mathrm{d}\rho}{\mathrm{d}\sigma}\right) \mathrm{d}\sigma \le \int_{\mathbb{R}^d} \phi_B\left(\frac{\mathrm{d}\rho}{\mathrm{d}\sigma}\right) \mathrm{d}\sigma \le c + \log \|e^{-V}\|_{L^1} \qquad \forall\, \rho \in L_{\mathcal{F}}(c),$$

where $\tilde{\phi}_B(s) := \phi_B(s) 1_{[1,\infty)}(s)$, $s \in [0,\infty)$. However, since $\tilde{\phi}_B$ satisfies property *(iii)* of Proposition 4.21, we have that $L_{\mathcal{F}}(c)$ is relatively compact w.r.t. the weak-L^1 topology, thus also implying the relative compactness of $L_{\mathcal{F}}(c)$ w.r.t. the narrow topology. □

4.4 Summary

With the ingredients above, we conclude from Theorem 2.4 the following statement:

Theorem 4.23. *Let $V\colon \mathbb{R}^d \to \mathbb{R}$ be a λ-convex lower semicontinuous external potential with $e^{-V} \in L^1(\mathbb{R}^d)$. Then for every $\bar{\rho} \in \operatorname{dom}\mathcal{F}$, there exists a density-velocity pair $(\rho, \mathbf{v})$ with $\rho_0 = \bar{\rho}$ satisfying the continuity equation (CE) and the energy-dissipation balance*

$$\int_s^t \left\{ \frac{1}{2}\|\mathbf{v}_r\|^2_{L^2(\rho_r)} + \frac{1}{2}|\partial^- \mathcal{F}|^2(\rho_r) \right\} \mathrm{d}r + \mathcal{F}(\rho_t) = \mathcal{F}(\rho_s) \qquad \forall\, [s,t] \subset [0,T], \tag{4.5}$$

where the relaxed local slope is a strong upper gradient given by

$$|\partial^- \mathcal{F}|(\rho) = |\partial \mathcal{F}|(\rho) = \begin{cases} \left\| \dfrac{\nabla u}{u} + \nabla V \right\|_{L^2(\rho)} & \text{if } \rho = u\mathcal{L}^d \text{ with } \sqrt{u} \in W^{1,2}(\mathbb{R}^d), \\ +\infty & \text{otherwise.} \end{cases}$$

In particular, we obtain a gradient flow solution of the diffusion equation (4.1).

The End.

References

[1] L. Ambrosio, N. Gigli and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*. Birkhäuser Verlag, 2005.

[2] S. Daneri and G. Savaré, *Lecture notes on gradient flows and optimal transport*. Summer School “Optimal transportation: Theory and applications” in Grenoble, 2009.

[3] A. Mielke, *Gradients systems and their gradient-flow equations*. Lecture notes, 2023.

[4] M. Peletier, *Variational modelling: Energies, gradient flows, and large deviations*, 2014.

[5] F. Santambrogio, {*Euclidean, metric, and Wasserstein*} *gradient flows: an overview*. Bull. Math. Sci. 7:87–154, 2017.