Calculus of Variations

Jump processes as generalized gradient flows

Mark A. Peletier[1] · Riccarda Rossi[2] · Giuseppe Savaré[3] · Oliver Tse[4]

Abstract
We have created a functional framework for a class of non-metric gradient systems. The state space is a space of nonnegative measures, and the class of systems includes the Forward Kolmogorov equations for the laws of Markov jump processes on Polish spaces. This framework comprises a definition of a notion of solutions, a method to prove existence, and an archetype uniqueness result. We do this by using only the structure that is provided directly by the dissipation functional, which need not be homogeneous, and we do not appeal to any metric structure.

Mathematics Subject Classification 34K30 · 35A15 · 35D99 · 45K99 · 47G20

Contents

Communicated by Andrea Malchiodi.

✉ Oliver Tse
  o.t.c.tse@tue.nl

  Mark A. Peletier
  M.A.Peletier@tue.nl

  Riccarda Rossi
  riccarda.rossi@unibs.it

  Giuseppe Savaré
  giuseppe.savare@unibocconi.it

[1] Department of Mathematics and Computer Science and Institute for Complex Molecular Systems, TU Eindhoven 5600 MB, Eindhoven, The Netherlands

[2] DIMI, Università degli studi di Brescia, Via Branze 38, 25133 Brescia, Italy

[3] Department of Decision Sciences and BIDSA, Bocconi University, Via Roentgen 1, 20136 Milan, Italy

[4] Department of Mathematics and Computer Science, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands

1 Introduction

The study of dissipative variational evolution equations has seen a tremendous activity in the last two decades. A general class of such systems is that of *generalized gradient flows*, which formally can be written as

$$\dot{\rho} = D_\zeta R^*(\rho, -D_\rho E(\rho)) \tag{1.1}$$

in terms of a *driving functional* E and a *dual dissipation potential* $R^* = R^*(\rho, \zeta)$, where D_ζ and D_ρ denote derivatives with respect to ζ and ρ. The most well-studied of these are classical gradient flows [4], for which $\zeta \mapsto D_\zeta R^*(\rho, \zeta) = \mathbb{K}(\rho)\zeta$ is a linear operator $\mathbb{K}(\rho)$, and rate-independent systems [61], for which $\zeta \mapsto D_\zeta R^*(\rho, \zeta)$ is zero-homogeneous.

However, various models naturally lead to gradient structures that are neither classic nor rate-independent. For these systems, the map $\zeta \mapsto D_\zeta R^*(\rho, \zeta)$ is neither linear nor zero-homogeneous, and in many cases it is not even homogeneous of any order. Some examples are

(1) Models of chemical reactions, where R^* depends exponentially on ζ [6,32,38,46],
(2) The Boltzmann equation, also with exponential R^* [38],
(3) Nonlinear viscosity relations such as the Darcy-Forchheimer equation for porous media flow [39,44],
(4) Effective, upscaled descriptions in materials science, where the effective potential R^* arises through a cell problem, and can have many different types of dependence on ζ [16,28,46,55,65,72–74],
(5) Gradient structures that arise from large-deviation principles for sequences of stochastic processes, in particular jump processes [58,59].

The last example is the inspiration for this paper.

Regardless whether R^* is classic, rate-independent, or otherwise, equation (1.1) typically is only formal, and it is a major mathematical challenge to construct an appropriate functional framework for this equation. Such a functional framework should give the equation a rigorous meaning, and provide the means to prove well-posedness, stability, regularity and approximation results to facilitate the study of the equation.

For classical gradient systems, in which $D_\zeta R^*$ is linear and R^* is quadratic in ζ (therefore also called 'quadratic' gradient systems) and when R^* generates a metric space, a rich framework has been created by Ambrosio, Gigli, and Savaré [4]. For rate-independent systems, in which R^* is 1-homogeneous in ζ, the complementary concepts of 'Global Energetic solutions' and 'Balanced Viscosity solutions' give rise to two different frameworks [19,61,62,64,67].

For the examples (1–5) listed above, however, R^* is not homogeneous in ζ, and neither the rate-independent frameworks nor the metric-space theory apply. Nonetheless, the existence of such models of real-world systems with a formal variational-evolutionary structure suggests that there may exist a functional framework for such equations that relies on this structure. In this paper we build exactly such a framework for an important class of equations of this type, those that describe Markov jump processes. We expect the approach advanced here to be applicable to a broader range of systems.

1.1 Generalized gradient systems for Markov jump processes

Some generalized gradient-flow structures of evolution equations are generated by the large deviations of an underlying, more microscopic stochastic process [1,2,22,46,59,60]. This explains the origin and interpretation of such structures, and it can be used to identify hitherto unknown gradient-flow structures [36,71].

It is the example of Markov *jump* processes that inspires the results of this paper, and we describe this example here; nonetheless, the general setup that starts in Sect. 3.1 has wider application. We think of Markov jump processes as jumping from one 'vertex' to another 'vertex' along an 'edge' of a 'graph'; we place these terms between quotes because the space V of vertices may be finite, countable, or even uncountable, and similarly the space $E := V \times V$ of edges may be finite, countable, or uncountable (see Assumption $(V\pi\kappa)$ below). In this paper, V is a standard Borel space.

The laws of such processes are time-dependent measures $t \mapsto \rho_t \in \mathcal{M}^+(V)$ (with $\mathcal{M}^+(V)$ the space of positive finite Borel measures—see Sect. 2). These laws satisfy the Forward Kolmogorov equation

$$\partial_t \rho_t = Q^* \rho_t, \qquad (Q^* \rho)(\mathrm{d}x) = \int_{y \in V} \rho(\mathrm{d}y)\kappa(y, \mathrm{d}x) - \rho(\mathrm{d}x) \int_{y \in V} \kappa(x, \mathrm{d}y). \qquad (1.2)$$

Here $Q^* : \mathcal{M}(V) \to \mathcal{M}(V)$ is the dual of the infinitesimal generator $Q : B_b(V) \to B_b(V)$ of the process, which for an arbitrary bounded Borel function $\varphi \in B_b(V)$ is given by

$$(Q\varphi)(x) = \int_V [\varphi(y) - \varphi(x)] \kappa(x, \mathrm{d}y). \qquad (1.3)$$

The jump kernel κ in these definitions characterizes the process: $\kappa(x, \cdot) \in \mathcal{M}^+(V)$ is the infinitesimal rate of jumps of a particle from the point x to points in V. Here we address the reversible case, which means that the process has an invariant measure $\pi \in \mathcal{M}^+(V)$, i.e., $Q^*\pi = 0$, and that the joint measure $\pi(\mathrm{d}x)\kappa(x, \mathrm{d}y)$ is symmetric in x and y.

In this paper we consider evolution equations of the form (1.2) for the nonnegative measure ρ, as well as various linear and nonlinear generalizations. We will view them as gradient systems of the form (1.1), and use this gradient structure to study their properties.

The gradient structure for equation (1.2) consists of the state space $\mathcal{M}^+(V)$, a driving functional $\mathcal{E} : \mathcal{M}^+(V) \to [0, +\infty]$, and a dual dissipation potential $\mathscr{R}^* : \mathcal{M}^+(V) \times B_b(E) \to [0, +\infty]$ (where $B_b(E)$ denotes the space of bounded Borel functions on E). We now describe this structure in formal terms, and making it rigorous is one of the aims of this paper.

The functional that drives the evolution is the relative entropy with respect to the invariant measure π, namely

$$\mathcal{E}(\rho) = \mathscr{F}_\phi(\rho|\pi) := \begin{cases} \int_V \phi\big(u(x)\big)\pi(\mathrm{d}x) & \text{if } \rho \ll \pi, \text{ with } u = \dfrac{\mathrm{d}\rho}{\mathrm{d}\pi}, \\ +\infty & \text{otherwise,} \end{cases} \qquad (1.4)$$

where for the example of Markov jump processes the 'energy density' ϕ is given by

$$\phi(s) := s \log s - s + 1. \qquad (1.5)$$

(In the general development below we consider more general functions ϕ, such as those that arise in strongly interacting particle systems; see e.g. [23,45]).

The dissipation potential R^* is best written in terms of an alternative potential $\mathscr{R}^*$,

$$\mathsf{R}^*(\rho, \zeta) := \mathscr{R}^*(\rho, \overline{\nabla}\zeta).$$

Here the 'graph gradient' $\overline{\nabla} : B_b(V) \to B_b(E)$ and its negative dual, the 'graph divergence operator' $\overline{\mathrm{div}} : \mathcal{M}(E) \to \mathcal{M}(V)$, are defined as follows:

$$(\overline{\nabla}\varphi)(x, y) := \varphi(y) - \varphi(x) \qquad \text{for any } \varphi \in B_b(V), \qquad (1.6a)$$

$$(\overline{\mathrm{div}}\, j)(\mathrm{d}x) := \int_{y \in V} \big[j(\mathrm{d}x, \mathrm{d}y) - j(\mathrm{d}y, \mathrm{d}x)\big] \qquad \text{for any } j \in \mathcal{M}(E), \qquad (1.6b)$$

and are linked by

$$\iint_E \overline{\nabla}\varphi(x, y)\, j(\mathrm{d}x, \mathrm{d}y) = -\int_V \varphi(x)\, \overline{\mathrm{div}}\, j(\mathrm{d}x) \quad \text{for every } \varphi \in B_b(V). \qquad (1.7)$$

The dissipation functional $\mathscr{R}^*$ is defined for $\xi \in B_b(E)$ by

$$\mathscr{R}^*(\rho, \xi) := \frac{1}{2} \int_E \Psi^*(\xi(x, y))\, \mathbf{v}_\rho(\mathrm{d}x\,\mathrm{d}y), \tag{1.8}$$

where the function Ψ^* and the 'edge' measure $\mathbf{v}_\rho$ will be fixed in (1.10) below.

With these definitions, the gradient-flow equation (1.1) can be written alternatively as

$$\partial_t \rho_t = -\overline{\mathrm{div}}\Big[\mathrm{D}_\xi \mathscr{R}^*\Big(\rho_t, -\overline{\nabla}\phi'\Big(\frac{\mathrm{d}\rho_t}{\mathrm{d}\pi}\Big)\Big)\Big], \tag{1.9}$$

which can be recognized by observing that

$$\big\langle \mathrm{D}_\zeta \mathsf{R}^*(\rho, \zeta), \tilde{\zeta} \big\rangle = \frac{\mathrm{d}}{\mathrm{d}h} \mathscr{R}^*(\rho, \overline{\nabla}\zeta + h\overline{\nabla}\tilde{\zeta})\Big|_{h=0} = \big\langle \mathrm{D}_\xi \mathscr{R}^*(\rho, \overline{\nabla}\zeta), \overline{\nabla}\tilde{\zeta} \big\rangle = \big\langle -\overline{\mathrm{div}}\mathrm{D}_\xi \mathscr{R}^*(\rho, \overline{\nabla}\zeta), \tilde{\zeta} \big\rangle,$$

and $\mathrm{D}\mathscr{E}(\rho) = \phi'(u)$ (which corresponds to $\log u$ for the logarithmic entropy (1.5)). This $(\overline{\mathrm{div}}, \overline{\nabla})$-duality structure is a common feature in both physical and probabilistic models, and has its origin in the distinction between 'states' and 'processes'; see [70, Sec. 3.3] and [69] for discussions.

For this example of Markov jump processes we consider a class of generalized gradient structures of the type above, given by $\mathscr{E}$ and $\mathscr{R}^*$ (or equivalently by the densities ϕ, Ψ^*, and the measure $\mathbf{v}_\rho$), with the property that equations (1.1) and (1.9) coincide with (1.2). Even for fixed $\mathscr{E}$ there exists a range of choices for Ψ^* and $\mathbf{v}_\rho$ that achieve this (see also the discussion in [35,59]). A simple calculation (see the discussion at the end of Sect. 3.1) shows that, if one chooses for the measure $\mathbf{v}_\rho$ the form

$$\mathbf{v}_\rho(\mathrm{d}x\,\mathrm{d}y) = \alpha(u(x), u(y))\, \pi(\mathrm{d}x)\kappa(x, \mathrm{d}y), \tag{1.10}$$

for a suitable fuction $\alpha : [0, \infty) \times [0, \infty) \to [0, \infty)$, and one introduces the map F : $(0, \infty) \times (0, \infty) \to \mathbb{R}$

$$\mathrm{F}(u, v) := (\Psi^*)'\big[\phi'(v) - \phi'(u)\big]\alpha(u, v) \quad u, v > 0, \tag{1.11}$$

then (1.9) takes the form of the integro-differential equation

$$\partial_t u_t(x) = \int_{y \in V} \mathrm{F}\big(u_t(x), u_t(y)\big) \kappa(x, \mathrm{d}y), \tag{1.12}$$

in terms of the density u_t of ρ_t with respect to π. Therefore, a pair $(\Psi^*, \mathbf{v}_\rho)$ leads to equation (1.2) whenever (Ψ^*, ϕ, α) satisfy the *compatibility property*

$$\mathrm{F}(u, v) = v - u \quad \text{for every } u, v > 0. \tag{1.13}$$

The classical quadratic-energy, quadratic-dissipation choice

$$\Psi^*(\xi) = \tfrac{1}{2}\xi^2, \quad \phi(s) = \tfrac{1}{2}s^2, \quad \alpha(u, v) = 1 \tag{1.14}$$

corresponds to the Dirichlet-form approach to (1.2) in $L^2(V, \pi)$. Here $\mathscr{R}^*(\rho, \boldsymbol{j}) = \mathscr{R}^*(\boldsymbol{j})$ is in fact independent of ρ: if one introduces the symmetric bilinear form

$$[\![u, v]\!] := \frac{1}{2} \iint_E \overline{\nabla}u(x, y)\, \overline{\nabla}v(x, y)\, \boldsymbol{\vartheta}(\mathrm{d}x, \mathrm{d}y), \quad [\![u, u]\!] = \frac{1}{2} \iint_E \Psi(\overline{\nabla}u)\, \mathrm{d}\boldsymbol{\vartheta}, \tag{1.15}$$

with $\boldsymbol{\vartheta}(\mathrm{d}x, \mathrm{d}y) = \pi(\mathrm{d}x)\kappa(x, \mathrm{d}y)$ (cf. (3.5) ahead), then (1.12) can also be formulated as

$$(\dot{u}_t, v)_{L^2(V, \pi)} + [\![u_t, v]\!] = 0 \quad \text{for every } v \in L^2(V, \pi). \tag{1.16}$$

Two other choices have received attention in the recent literature. Both of these are based not on the quadratic energy $\phi(s) = \frac{1}{2}s^2$, but on the Boltzmann entropy functional $\phi(s) = s \log s - s + 1$:

(1) The large-deviation characterization [58,59] leads to the choice

$$\Psi^*(\xi):=4\big(\cosh(\xi/2) - 1\big) \quad \text{and} \quad \alpha(u, v):=\sqrt{uv}. \tag{1.17a}$$

The corresponding primal dissipation potential $\Psi:=(\Psi^*)^*$ is given by

$$\Psi(s):=2s \log\left(\frac{s + \sqrt{s^2 + 4}}{2}\right) - \sqrt{s^2 + 4} + 4.$$

(2) The 'quadratic-dissipation' choice introduced independently by Maas [49], Mielke [52], and Chow, Huang, and Zhou [13] for Markov processes on *finite* graphs,

$$\Psi^*(\xi):=\tfrac{1}{2}\xi^2, \quad \Psi(s) = \tfrac{1}{2}s^2, \quad \text{and} \quad \alpha(u, v):=\frac{u - v}{\log(u) - \log(v)}. \tag{1.17b}$$

Other examples are discussed in Sect. 1.3. With the quadratic choice (1.17b), the gradient system fits into the metric-space structure (see e.g. [4]) and this feature has been used extensively to investigate the properties of general Markov jump processes [25,29–31,49,52]. In this paper, however, we focus on functions Ψ^* that are not homogeneous, as in (1.17a), and such that the corresponding structure is not covered by the usual metric framework. On the other hand, there are various arguments why this structure nonetheless has a certain 'naturalness' (see Sect. 1.4), and these motivate our aim to develop a functional framework based on this structure.

1.2 Challenges

Constructing a 'functional framework' for the gradient-flow equation (1.9) with the choices (1.5) and (1.17a) presents a number of independent challenges.

1.2.1 Definition of a solution

As it stands, the formulation of equation (1.9) and of the functional $\mathcal{R}^*$ of (1.8) presents many difficulties: the definition of $\mathcal{R}^*$ and the measure $\mathbf{v}_\rho$ when ρ is not absolutely continuous with respect to π, the concept of time differentiability for the curve of measures ρ_t, whether ρ_t is necessarily absolutely continuous with respect to π along an evolution, what happens if $d\rho_t/d\pi$ vanishes and ϕ is not differentiable at 0 as in the case of the logarithmic entropy, etcetera. As a result of these difficulties, it is not clear what constitutes a solution of equation (1.9), let alone whether such solutions exist. In addition, a good solution concept should be robust under taking limits, and the formulation (1.9) does not seem to satisfy this requirement either.

For quadratic and rate-independent systems, successful functional frameworks have been constructed on the basis of variational tools such as the Energy-Dissipation balance [4,42, 61,63]. Moreover, these functional frameworks have been shown to be stable under various forms of asymptotic limits [46,54,66,79,80]. This strongly suggests that also the framework proposed here should enjoy this stability with respect to perturbations of $\mathcal{E}$ and $\mathcal{R}$ which, in particular, would allow one to generalize the more classical notion of solutions developed

in Sect. 6. We have chosen not to dwell upon the stability issue to avoid overburdening the exposition; we only provide a 'partial' stability result (with $\mathcal{E}$ and $\mathcal{R}$ fixed) in Theorem 5.10.

In fact, the same large-deviation principle that gives rise to the 'cosh' structure (1.17a) above formally yields the 'EDP' functional (see Appendix A for a formal derivation)

$$\mathscr{L}(\rho, \boldsymbol{j}) := \begin{cases} \displaystyle\int_0^T \left[\mathcal{R}(\rho_t, \boldsymbol{j}_t) + \mathcal{R}^*\left(\rho_t, -\overline{\nabla}\phi'\left(\tfrac{\mathrm{d}\rho_t}{\mathrm{d}\pi}\right)\right)\right]\mathrm{d}t + \mathcal{E}(\rho_T) - \mathcal{E}(\rho_0) \\ \qquad\qquad \text{if } \partial_t\rho_t + \overline{\mathrm{div}}\,\boldsymbol{j}_t = 0 \text{ and } \rho_t \ll \pi \text{ for all } t \in [0, T], \\ +\infty \qquad \text{otherwise.} \end{cases} \tag{1.18}$$

In this formulation, $\mathcal{R}$ is the Legendre dual of $\mathcal{R}^*$ with respect to the ξ variable, which can be written in terms of the Legendre dual $\Psi := \Psi^{**}$ of Ψ^* as

$$\mathcal{R}(\rho, \boldsymbol{j}) := \frac{1}{2}\int_E \Psi\left(2\frac{\mathrm{d}\boldsymbol{j}}{\mathrm{d}\boldsymbol{v}_\rho}\right)\mathrm{d}\boldsymbol{v}_\rho. \tag{1.19}$$

Along smooth curves $\rho_t = u_t\pi$ with strictly positive densities, the functional $\mathscr{L}$ is nonnegative, since

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}(\rho_t) &= \int_V \phi'(u_t)\partial_t u_t\,\mathrm{d}\pi = \int_V \phi'(u_t(x))\partial_t\rho_t(\mathrm{d}x) \\ &= -\int_V \phi'(u_t(x))(\overline{\mathrm{div}}\,\boldsymbol{j}_t)(\mathrm{d}x) \\ &= \iint_E \overline{\nabla}\phi'(u_t)(x, y)\,\boldsymbol{j}_t(\mathrm{d}x\,\mathrm{d}y) \\ &= \iint_E \overline{\nabla}\phi'(u_t)(x, y)\frac{\mathrm{d}\boldsymbol{j}_t}{\mathrm{d}\boldsymbol{v}_{\rho_t}}(x, y)\,\boldsymbol{v}_{\rho_t}(\mathrm{d}x\,\mathrm{d}y) \tag{1.20} \\ &\geq -\frac{1}{2}\iint_E \left[\Psi\left(2\frac{\mathrm{d}\boldsymbol{j}_t}{\mathrm{d}\boldsymbol{v}_{\rho_t}}(x, y)\right) + \Psi^*\left(-\overline{\nabla}\phi'(u_t)(x, y)\right)\right]\boldsymbol{v}_{\rho_t}(\mathrm{d}x\,\mathrm{d}y). \tag{1.21} \end{aligned}$$

After time integration we find that $\mathscr{L}(\rho, \boldsymbol{j})$ is nonnegative for any pair $(\rho, \boldsymbol{j})$.

The minimum of $\mathscr{L}$ is formally achieved at value zero, at pairs $(\rho, \boldsymbol{j})$ satisfying

$$2\boldsymbol{j}_t = (\Psi^*)'\left(-\overline{\nabla}\phi'\left(\tfrac{\mathrm{d}\rho_t}{\mathrm{d}\pi}\right)\right)\boldsymbol{v}_{\rho_t} \quad\text{and}\quad \partial_t\rho_t + \overline{\mathrm{div}}\,\boldsymbol{j}_t = 0, \tag{1.22}$$

which is an equivalent way of writing the gradient-flow equation (1.9). This can be recognized, as usual for gradient systems, by observing that achieving equality in the inequality (1.21) requires equality in the Legendre duality of Ψ and Ψ^*, which reduces to the equations above.

Remark 1.1 It is worth noticing that the joint convexity of the functional $\mathcal{R}$ of (1.19) (a crucial property for the development of our analysis) is equivalent to the *convexity* of Ψ and *concavity* of the function α.

Remark 1.2 Let us add a comment concerning the choice of the factor $1/2$ in front of Ψ^* in (1.8), and the corresponding factors $1/2$ and 2 in (1.19). The cosh-entropy combination (1.17a) satisfies the linear-equation condition $F(u, v) = v - u$ (equation (1.13)) because of the elementary identity

$$2\sqrt{uv}\,\sinh\left(\frac{1}{2}\log\frac{v}{u}\right) = v - u.$$

The factor $1/2$ inside the sinh can be included in different ways. In [59] it was included explicitly, by writing expressions of the form $DR^*(\rho, -\frac{1}{2}DE(\rho))$; in this paper we follow [46] and include this factor in the definition of $\mathscr{R}^*$.

Remark 1.3 The continuity equation $\partial_t \rho_t + \overline{\mathrm{div}}\, j_t = 0$ is invariant with respect to skew-symmetrization of j, i.e. with respect to the transformation $j \mapsto j^\flat$ with $j^\flat(\mathrm{d}x, \mathrm{d}y) := \frac{1}{2}\big(j(\mathrm{d}x, \mathrm{d}y) - j(\mathrm{d}y, \mathrm{d}x)\big)$. Therefore we could also write the second integral in (1.20) as

$$\iint_E \overline{\nabla}\phi'(u_t)(x, y)\frac{\mathrm{d}j_t^\flat}{\mathrm{d}\mathbf{v}_{\rho_t}}(x, y)\,\mathbf{v}_{\rho_t}(\mathrm{d}x\,\mathrm{d}y)$$

$$\geq -\frac{1}{2}\iint_E \left[\Psi\left(\frac{\mathrm{d}(2j_t^\flat)}{\mathrm{d}\mathbf{v}_{\rho_t}}(x, y)\right) + \Psi^*\left(-\overline{\nabla}\phi'(u_t)(x, y)\right)\right]\mathbf{v}_{\rho_t}(\mathrm{d}x\,\mathrm{d}y).$$

thus replacing $\Psi\left(2\frac{\mathrm{d}j_t}{\mathrm{d}\mathbf{v}_{\rho_t}}(x, y)\right)$ with the lower term $\Psi\left(\frac{\mathrm{d}(2j_t^\flat)}{\mathrm{d}\mathbf{v}_{\rho_t}}(x, y)\right)$, cf. Remark 4.12, and

obtaining a corresponding equation as (1.22) for $(2j_t^\flat)$ instead of $2j_t$. This would lead to a weaker gradient system, since the choice (1.19) forces j_t to be skew-symmetric, whereas the choice of a dissipation involving only $j^\flat$ would not control the symmetric part of j. On the other hand, the evolution equation generated by the gradient system would remain the same.

Since at least formally equation (1.9) is equivalent to the requirement $\mathscr{L}(\rho, j) \leq 0$, we adopt this variational point of view to define solutions to the generalized gradient system $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$. This inequality is in fact the basis for the variational Definition 5.4 below. In order to do this in a rigorous manner, however, we will need

(1) A study of the continuity equation

$$\partial_t \rho_t + \overline{\mathrm{div}}\, j_t = 0, \qquad (1.23)$$

that appears in the definition of the functional $\mathscr{L}$ (Sect. 4.1).
(2) A rigorous definition of the measure $\mathbf{v}_{\rho_t}$ and of the functional $\mathscr{R}$ (Definition 4.9);
(3) A class $\mathcal{A}(0, T)$ of curves "of finite action" in $\mathcal{M}^+(V)$ along which the functional $\mathscr{R}$ has finite integral (equation (4.81));
(4) An appropriate definition of the *Fisher-information* functional (see Definition 5.1)

$$\rho \mapsto \mathscr{D}(\rho) := \mathscr{R}^*\big(\rho, -\overline{\nabla}\phi'(\mathrm{d}\rho/\mathrm{d}\pi)\big); \qquad (1.24)$$

(5) A proof of the lower bound $\mathscr{L} \geq 0$ (Theorem 4.16) via a suitable chain-rule inequality.

1.2.2 Existence of solutions

The first test of a new solution concept is whether solutions exist under reasonable conditions. In this paper we provide two existence results that complement each other, one based on dissipative L^1-theory and the other on the Energy-Dissipation balance. These theories are not completely equivalent, as can be observed for the classical heat equation (cf. [42] for a variational approach), and they reveal different properties of a solution, even when the assumptions for both theories are satisfied.

The first existence proof is based on a reformulation of the equation (1.2) as a differential equation in the Banach space $L^1(V, \pi)$, driven by a continuous dissipative operator. Under general compatibility conditions on ϕ, Ψ, and α, we show that the solution provided by

this abstract approach is also a solution in the variational sense that we discussed above. The proof is presented in Sect. 6 and is quite robust for initial data whose density takes value in a compact interval $[a, b] \subset (0, \infty)$. In order to deal with a more general class of initial data, we will adopt two different viewpoints. A first possibility is to take advantage of the robust stability properties of the $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ Energy-Dissipation balance when the Fisher information $\mathscr{D}$ is lower semicontinuous (cf. Theorem 5.10). A second possibility is to exploit the monotonicity properties of (1.12) when the map F in (1.11) exhibits good behaviour at the boundary of $\mathbb{R}^2_+$ and at infinity (cf. Theorem 6.5). As mentioned above, neither of these viewpoints are completely contained within the other.

Since we believe that the variational formulation reveals a relevant structure of such systems and we expect that it may also be useful in dealing with more singular cases and their stability issues, we also present a more intrinsic approach by adapting the well-established 'JKO-Minimizing-Movement' method to the structure of this equation. This method has been used, e.g., for metric-space gradient flows [4,42], for rate-independent systems [51], for some non-metric systems with formal metric structure [7,48], and also for Lagrangian systems with local transport [33].

This approach relies on the *Dynamical-Variational Transport cost* (DVT) $\mathscr{W}(\tau, \mu, \nu)$, which is the τ-dependent transport cost between two measures $\mu, \nu \in \mathcal{M}^+(V)$ induced by the dissipation potential $\mathscr{R}$ via

$$\mathscr{W}(\tau, \mu, \nu) := \inf \left\{ \int_0^\tau \mathscr{R}(\rho_t, \boldsymbol{j}_t) \, dt \; : \; \partial_t \rho_t + \overline{\mathrm{div}} \, \boldsymbol{j}_t = 0, \; \rho_0 = \mu, \; \text{and} \; \rho_\tau = \nu \right\}. \quad (1.25)$$

In the Minimizing-Movement scheme a single increment with time step $\tau > 0$ is defined by the minimization problem

$$\rho^n \in \operatorname{argmin}_\rho \left(\mathscr{W}(\tau, \rho^{n-1}, \rho) + \mathscr{E}(\rho) \right). \quad (1.26)$$

By concatenating such solutions, constructing appropriate interpolations, and proving a compactness result—all steps similar to the procedure in [4, Part I]—we find a curve $(\rho_t, \boldsymbol{j}_t)_{t \in [0,T]}$ satisfying the continuity equation (1.23) such that

$$\int_0^t \left[\mathscr{R}(\rho_r, \boldsymbol{j}_r) + \mathscr{S}^-(\rho_r) \right] dr + \mathscr{E}(\rho_t) \leq \mathscr{E}(\rho_0) \quad \text{for all } t \in [0, T], \quad (1.27)$$

where $\mathscr{S}^- : \mathrm{D}(\mathscr{E}) \to [0, +\infty)$ is a suitable *relaxed slope* of the energy functional $\mathscr{E}$ with respect to the cost $\mathscr{W}$ (see (7.29)). Under a lower-semicontinuity condition on $\mathscr{D}$ we show that $\mathscr{S}^- \geq \mathscr{D}$. It then follows that ρ is a solution as defined above (see Definition 5.4).

Section 7 is devoted to developing the 'Minimizing-Movement' approach for general DVTs. This requires establishing

(6) Properties of $\mathscr{W}$ that generalize those of the 'metric version' $\mathscr{W}(\tau, \mu, \nu) = \frac{1}{2\tau} d(\mu, \nu)^2$ (Sect. 7.2);

(7) A generalization of the 'Moreau-Yosida approximation' and of the 'De Giorgi variational interpolant' to the non-metric case, and a generalization of their properties (Sects. 7.1 and 7.2);

(8) A compactness result as $\tau \to 0$, based on the properties of $\mathscr{W}$ (Sect. 7.4);

(9) A proof of $\mathscr{S}^- \geq \mathscr{D}$ (Corollary 7.11).

This procedure leads to our existence result, Theorem 7.4, of solutions in the sense of Definition 5.4.

1.2.3 Uniqueness of solutions

We prove uniqueness of variational solutions under suitable convexity conditions of $\mathscr{D}$ and $\mathscr{E}$ (Theorem 5.9), following an idea by Gigli [34].

1.3 Examples

We will use the following two guiding examples to illustrate the results of this paper. Precise assumptions are given in Sect. 3.1. In both examples the state space consists of measures ρ on a standard Borel space $(V, \mathfrak{B})$ endowed with a reference Borel measure π. The kernel $x \mapsto \kappa(x, \cdot)$ is a measurable family of nonnegative measures with uniformly bounded mass, such that the pair (π, κ) satisfies detailed balance (see Sect. 3.1).

Example 1: Linear equations driven by the Boltzmann entropy. This is the example that we have been using in this introduction. The equation is the linear equation (1.2),

$$\partial_t \rho_t(\mathrm{d}x) = \int_{y \in V} \rho(\mathrm{d}y)\kappa(y, \mathrm{d}x) - \rho(\mathrm{d}x) \int_{y \in V} \kappa(x, \mathrm{d}y),$$

which can also be written in terms of the density $u = \mathrm{d}\rho/\mathrm{d}\pi$ as

$$\partial_t u_t(x) = \int_{y \in V} \big[u_t(y) - u_t(x)\big] \kappa(x, \mathrm{d}y),$$

and corresponds to the linear field F of (1.13). Apart from the classical quadratic setting of (1.14), two gradient structures for this equation have recently received attention in the literature, both driven by the Boltzmann entropy (1.5) $\phi(s) = s \log s - s + 1$ as described in (1.17):

(1) The 'cosh' structure: $\Psi^*(\xi) = 4\big(\cosh(\xi/2) - 1\big)$ and $\alpha(u, v) = \sqrt{uv}$;
(2) The 'quadratic' structure: $\Psi^*(\xi) = \frac{1}{2}\xi^2$ and $\alpha(u, v) = (u - v)/\log(u/v)$.

However, the approach of this paper applies to more general combinations (ϕ, Ψ^*, α) that lead to the same equation. Due to the particular structure of (1.11), it is clear that the 1-homogeneity of the linear map F (1.13) and the 0-homogeneity of the term $\phi'(v) - \phi'(u)$ associated with the Boltzmann entropy (1.5) restrict the range of possible α to 1-*homogenous functions* such as the 'mean functions' $\alpha(u, v) = \sqrt{uv}$ (geometric) and $\alpha(u, v) = (u - v)/\log(u/v)$ (logarithmic).

Confining the analysis to concave functions (according to Remark 1.1), we observe that every concave and 1-homogeneous function α can be obtained by the concave generating function $\mathfrak{f} : (0, +\infty) \to (0, +\infty)$

$$\alpha(u, v) = u\mathfrak{f}(v/u) = v\mathfrak{f}(u/v), \quad \mathfrak{f}(r) := \alpha(r, 1), \quad u, v, r > 0. \tag{1.28}$$

The symmetry of α corresponds to the property

$$r\mathfrak{f}(1/r) = \mathfrak{f}(r) \quad \text{for every } r > 0, \tag{1.29}$$

and shows that the function

$$\mathfrak{g}(s) := \frac{\exp(s) - 1}{\mathfrak{f}(\exp(s))} \quad s \in \mathbb{R}, \text{ is odd.} \tag{1.30}$$

The concaveness of $\mathfrak{f}$ also shows that $\mathfrak{g}$ is increasing, so that we can define

$$\Psi^*(\xi) := \int_0^\xi \mathfrak{g}(s) \, \mathrm{d}s = \int_1^{\exp(\xi)} \frac{r - 1}{\mathfrak{f}(r)} \frac{\mathrm{d}r}{r}, \quad \xi \in \mathbb{R}, \tag{1.31}$$

which is convex, even, and superlinear if

$$\alpha(0, 1) = \mathfrak{f}(0) = \lim_{r \to 0} r \mathfrak{f}\left(\frac{1}{r}\right) = 0. \tag{1.32}$$

A natural class of concave and 1-homogeneous weight functions is provided by the *Stolarsky means* $\mathfrak{c}_{p,q}(u, v)$ with appropriate $p, q \in \mathbb{R}$, and any $u, v > 0$ [12, Chapter VI]:

$$\alpha(u, v) = \mathfrak{c}_{p,q}(u, v) := \begin{cases} \left(\frac{p}{q} \frac{v^q - u^q}{v^p - u^p}\right)^{1/(q-p)} & \text{if } p \neq q, \ q \neq 0, \\ \left(\frac{1}{p} \frac{v^p - u^p}{\log(v) - \log(u)}\right)^{1/p} & \text{if } p \neq 0, \ q = 0, \\ e^{-1/p} \left(\frac{v^{v^p}}{u^{u^p}}\right)^{1/(v^p - u^p)} & \text{if } p = q \neq 0, \\ \sqrt{uv} & \text{if } p = q = 0, \end{cases}$$

from which we identify other simpler means, such as the *power means* $\mathfrak{m}_p(u, v) = \mathfrak{c}_{p,2p}(u, v)$ with $p \in [-\infty, 1]$:

$$\mathfrak{m}_p(u, v) = \begin{cases} \left(\frac{1}{2}(u^p + v^p)\right)^{1/p} & \text{if } 0 < p \leq 1 \text{ or } -\infty < p < 0 \text{ and } u, v \neq 0, \\ \sqrt{uv} & \text{if } p = 0, \\ \min(u, v) & \text{if } p = -\infty, \\ 0 & \text{if } p < 0 \text{ and } uv = 0, \end{cases} \tag{1.33}$$

and the generalized logarithmic mean $\mathfrak{l}_p(u, v) = \mathfrak{c}_{1,p+1}(u, v)$, $p \in [-\infty, -1]$.

The power means are obtained from the concave generating functions

$$\mathfrak{f}_p(r) := 2^{-1/p}(r^p + 1)^{1/p} \quad \text{if } p \neq 0, \quad \mathfrak{f}_0(r) = \sqrt{r}, \quad \mathfrak{f}_{-\infty}(r) = \min(r, 1), \quad r > 0. \tag{1.34}$$

We can thus define

$$\Psi_p^*(\xi) := 2^{1/p} \int_1^{\exp \xi} \frac{r - 1}{(r^p + 1)^{1/p}} \frac{dr}{r}, \quad \xi \in \mathbb{R}, \quad p \in (-\infty, 1] \setminus \{0\}, \tag{1.35}$$

with the obvious changes when $p = 0$ (the case $\Psi_0^*(\xi) = 4(\cosh(\xi/2) - 1)$) or $p = -\infty$ (the case $\Psi_{-\infty}^*(\xi) = \exp(|\xi|) - 1 - |\xi|$).

It is interesting to note that the case $p = -1$ (harmonic mean) corresponds to

$$\Psi_{-1}^*(\xi) = \cosh(\xi) - 1. \tag{1.36}$$

We finally note that the arithmetic mean $\alpha(u, v) = \mathfrak{m}_1(u, v) = (u + v)/2$ would yield $\Psi_1^*(\xi) = 4 \log(1/2(1 + e^\xi)) - 2\xi$, which is not superlinear.

Example 2: Nonlinear equations. We consider a combination of ϕ, Ψ^*, and α such that the function F introduced in (1.11) has a continuous extension up to the boundary of $[0, +\infty)^2$ and satisfies a suitable growth and monotonicity condition (see Sect. 6). The resulting integro-differential equation is given by (1.12). Here is a list of some interesting cases (we will neglect all the issues concerning growth and regularity).

(1) A field of the form $F(u, v) = f(v) - f(u)$ with $f : \mathbb{R}_+ \to \mathbb{R}$ monotone corresponds to the equation

$$\partial_t u_t(x) = \int_{y \in V} \big(f(u_t(y)) - f(u_t(x))\big) \kappa(x, dy),$$

and can be classically considered in the framework of the Dirichlet forms, i.e. $\alpha \equiv 1$, $\Psi^*(r) = r^2/2$, with energy ϕ satisfying $\phi' = f$.

(2) The case $F(u, v) = g(v - u)$, with $g : \mathbb{R} \to \mathbb{R}$ monotone and odd, yields the equation

$$\partial_t u_t(x) = \int_{y \in V} g\big(u_t(y) - u_t(x)\big) \kappa(x, dy),$$

and can be obtained with the choices $\alpha \equiv 1$, $\phi(s):=s^2/2$ and $\Psi^*(r):=\int_0^r g(s)\, ds$.

(3) Consider now the case when F is positively q-homogeneous, with $q \in [0, 1]$. It is then natural to consider a q-homogeneous α and the logarithmic entropy $\phi(r) = r \log r - r + 1$. If the function $h : (0, \infty) \to \mathbb{R}$, $h(r):=F(r, 1)/\alpha(r, 1)$ is increasing, then setting as in (1.35)

$$\Psi^*(\xi):= \int_1^{\exp(\xi)} h(r)\, dr$$

equation (1.12) provides an example of generalized gradient system $(\mathcal{E}, \mathcal{R}, \mathcal{R}^*)$. Simple examples are $F(u, v) = v^q - u^q$, corresponding to the equation

$$\partial_t u_t(x) = \int_{y \in V} \big(u_t^q(y) - u_t^q(x)\big) \kappa(x, dy),$$

with $\alpha(u, v):=m_p(u^q, v^q)$ and $\Psi^*(\xi):=\frac{1}{q}\Psi_p^*(q\xi)$, where Ψ_p^* has been defined in (1.35). In the case $p = 0$ we get $\Psi^*(\xi) = \frac{4}{q}\big(\cosh(q\xi/2) - 1\big)$.

As a last example, we can consider $F(u, v) = \text{sign}(v - u)|v^m - u^m|^{1/m}$, $m > 0$, and $\alpha(u, v) = \min(u, v)$; in this case, the function h given by $h(r) = (r^m - 1)^{1/m}$ when $r \geq 1$, and $h(r) = -(r^{-m} - 1)^{1/m}$ when $r < 1$, satisfies the required monotonicity property.

1.4 Comments

Rationale for studying this structure. We think that the structure of generalized gradient systems $(\mathcal{E}, \mathcal{R}, \mathcal{R}^*)$ is sufficiently rich and interesting to deserve a careful analysis. It provides a genuine extension of the more familiar quadratic gradient-flow structure of Maas, Mielke, and Chow–Huang–Zhou, which better fits into the metric framework of [4]. In Sect. 6 we will also show its connection with the theory of dissipative evolution equations.

Moreover, the specific non-homogeneous structure based on the cosh function (1.17a) has a number of arguments in its favor, which can be summarized in the statement that it is 'natural' in various different ways:

(1) It appears in the characterization of large deviations of Markov processes; see Appendix A or [10,59];
(2) It arises in evolutionary limits of other gradient structures (including quadratic ones) [6,46,53,66];
(3) It 'responds naturally' to external forcing [66, Prop. 4.1];
(4) It can be generalized to nonlinear equations [37,38].

We will explore these claims in more detail in a forthcoming paper. Last but not least, the very fact that non-quadratic, generalized gradient flows may arise in the limit of gradient flows suggests that, allowing for a broad class of dissipation mechanisms is crucial in order to (1) fully exploit the flexibility of the gradient-structure formulation, and (2) explore its robustness with respect to Γ-converging energies and dissipation potentials.

Potential for generalization. In this paper we have chosen to concentrate on the consequences of non-homogeneity of the dissipation potential Ψ for the techniques that are commonly used in gradient-flow theory. Until now, the lack of a sufficiently general rigorous construction of the functional $\mathscr{R}$ and its minimal integral over curves $\mathscr{W}$ have impeded the use of this variational structure in rigorous proofs, and a main aim of this paper is to provide a way forward by constructing a rigorous framework for these objects, while keeping the setup (in particular, the ambient space V) as general as possible.

In order to restrict the length of this paper, we considered only simple driving functionals $\mathscr{E}$, which are of the local variety $\mathscr{E}(\rho) = \int \phi(\mathrm{d}\rho/\mathrm{d}\pi)\mathrm{d}\pi$. Many gradient systems appearing in the literature are driven by more general functionals, that include interaction and other nonlinearities [25,26,40,78], and we expect that the techniques of this paper will be of use in the study of such systems.

As one specific direction of generalization, we note that the Minimizing-Movement construction on which the proof of Theorem 7.4 is based has a scope wider than that of the generalized gradient structure $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ under consideration. In fact, as we show in Sect. 7, Theorem 7.4 yields the existence of (suitably formulated) gradient flows in a general *topological space* endowed with a cost fulfilling suitable properties. While we do not develop this discussion in this paper, at places throughout the paper we hint at this prospective generalization: the 'abstract-level' properties of the DVT cost are addressed in Sect. 4.7, and the whole proof of Theorem 7.4 is carried out under more general conditions than those required on the 'concrete' system set up in Sect. 3.

Challenges for generalization. A well-formed functional framework includes a concept of solutions that behaves well under the taking of limits, and the existence proof is the first test of this. Our existence proof highlights a central challenge here, in the appearance of *two* slope functionals $\mathscr{S}^-$ and $\mathscr{D}$ that both represent rigorous versions of the 'Fisher information' term $\mathscr{R}^*\big(\rho, -\overline{\nabla}\phi'(\mathrm{d}\rho/\mathrm{d}\pi)\big)$. The chain-rule lower-bound inequality holds under general conditions for $\mathscr{D}$ (Theorem 4.16), but the Minimizing-Movement construction leads to the more abstract object $\mathscr{S}^-$. Passing to the limit in the minimizing-movement approach requires connecting the two through the inequality $\mathscr{S}^- \geq \mathscr{D}$. We prove it by first obtaining the inequality $\mathscr{S} \geq \mathscr{D}$, cf. Proposition 7.10, under the condition that a solution to the $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ system exists (for instance, by the approach developed in Sect. 6). We then deduce the inequality $\mathscr{S}^- \geq \mathscr{D}$ under the further condition that $\mathscr{D}$ be lower semicontinuous, which can be in turn proved under a suitable convexity condition (cf. Prop. 5.3). We hope that more effective ways of dealing with these issues will be found in the future.

Comparison with the Weighted Energy-Dissipation method. It would be interesting to develop the analogous variational approach based on studying the limit behaviour as $\varepsilon \downarrow 0$ of the minimizers $(\rho_t, \boldsymbol{j}_t)_{t \geq 0}$ of the Weighted Energy-Dissipation (WED) functional

$$\mathscr{W}_\varepsilon(\rho, \boldsymbol{j}) := \int_0^{+\infty} \mathrm{e}^{-t/\varepsilon}\Big(\mathscr{R}(\rho_t, \boldsymbol{j}_t) + \frac{1}{\varepsilon}\mathscr{E}(\rho_t)\Big)\mathrm{d}t \qquad (1.37)$$

among the solutions to the continuity equation with initial datum ρ_0, see [76]. Indeed, the *intrinsic character* of the WED functional, which only features the dissipation potential $\mathscr{R}$, makes it suitable to the present non-metric framework.

1.5 Notation

The following table collects the notation used throughout the paper.

$\overline{\nabla}, \overline{\mathrm{div}}$	graph gradient and divergence (1.6)
$\alpha(\cdot, \cdot)$	multiplier in flux rate $\mathbf{v}_\rho$ Assumption $(\mathscr{R}^*\Psi\alpha)$
α^∞, α_*	recession function, Legendre transform Sect. 2.3
$\alpha[\,\cdot\,|\,\cdot\,], \hat{\alpha}$	measure map, perspective function Sect. 2.3
$\mathcal{A}(a, b)$	set of curves ρ with finite action (4.33)
$\|\kappa_V\|_\infty$	upper bound on κ Assumption $(V\pi\kappa)$
C_b	space of bdd, ct. functions with supremum norm
$\mathcal{CE}(a, b)$	set of pairs $(\rho, \boldsymbol{j})$ satisfying the continuity equation Def. 4.1
$D_\phi(u, v), D_\phi^\pm(u, v)$	integrands defining the Fisher information $\mathscr{D}$ (4.53)
$\mathscr{D}$	Fisher-information functional Def. 5.1
$E = V \times V$	space of edges Assumption $(V\pi\kappa)$
$\mathscr{E}, \mathrm{D}(\mathscr{E})$	driving entropy functional and its domain (1.4) & Assumption $(\mathscr{E}\phi)$
F	vector field (1.11)
$\boldsymbol{\vartheta}_\rho^\pm$	, ρ-adjusted jump rates (4.18)
$\boldsymbol{\vartheta}$	equilibrium jump rate (3.5)
κ	jump kernel (1.3) & Assumption $(V\pi\kappa)$
$\boldsymbol{\kappa}_\gamma$	$\gamma \otimes \kappa$ (2.33)
$\mathscr{L}$	Energy-Dissipation balance functional (1.18)
$\mathcal{M}(\Omega; \mathbb{R}^m), \mathcal{M}^+(\Omega)$	vector (positive) measures on Ω Sec. 2
$\mathbf{v}_\rho$	edge measure in definition of $\mathscr{R}^*, \mathscr{R}$ (1.8), (1.19), (1.10)
Q, Q^*	generator and dual generator (1.2)
$\mathscr{R}, \mathscr{R}^*$	dual pair of dissipation potentials (1.8), (1.19), Def. 4.9
$\mathbb{R}_+ := [0, \infty)$	[$\boldsymbol{\varsigma}$ symmetry map $(x, y) \mapsto (y, x)$ (3.1)
$\mathscr{S}^-$	relaxed slope (7.29)
Υ	perspective function associated with
Ψ and α	(4.13)
V	space of states Assumption $(V\pi\kappa)$
ϕ	density of $\mathscr{E}$ (1.4) & Assumption $(\mathscr{E}\phi)$
Ψ, Ψ^*	dual pair of dissipation functions Assumption $(\mathscr{R}^*\Psi\alpha)$, Lem. 3.1
$\mathscr{W}$	Dynamic-Variational Transport cost (1.25) & Sec. 4.6
$\mathbb{W}$	$\mathscr{W}$- action (4.89)
x, y	coordinate maps $(x, y) \mapsto x$ and $(x, y) \mapsto y$ (3.1)

2 Preliminary results

2.1 Measure theoretic preliminaries

Let $(Y, \mathfrak{B})$ be a measurable space. When Y is endowed with a (metrizable and separable) topology τ_Y we will often assume that $\mathfrak{B}$ coincides with the Borel σ-algebra $\mathfrak{B}(Y, \tau_Y)$ induced by τ_Y. We recall that $(Y, \mathfrak{B})$ is called a *standard Borel space* if it is isomorphic (as a measurable space) to a Borel subset of a complete and separable metric space; equivalently, one can find a Polish topology τ_Y on Y such that $\mathfrak{B} = \mathfrak{B}(Y, \tau_Y)$.

We will denote by $\mathcal{M}(Y; \mathbb{R}^m)$ the space of σ-additive measures on $\mu : \mathfrak{B} \to \mathbb{R}^m$ of *finite total variation* $\|\mu\|_{TV} := |\mu|(Y) < +\infty$, where for every $B \in \mathfrak{B}$

$$|\mu|(B) := \sup \left\{ \sum_{i=0}^{+\infty} |\mu(B_i)| : B_i \in \mathfrak{B}, \ B_i \text{ pairwise disjoint}, \ B = \bigcup_{i=0}^{+\infty} B_i \right\}.$$

The set function $|\mu| : \mathfrak{B} \to [0, +\infty)$ is a positive finite measure on $\mathfrak{B}$ [3, Thm. 1.6] and $(\mathcal{M}(Y; \mathbb{R}^m), \|\cdot\|_{TV})$ is a Banach space.

In the case $m = 1$, we will simply write $\mathcal{M}(Y)$, and we shall denote the space of *positive finite measures* on $\mathfrak{B}$ by $\mathcal{M}^+(Y)$. For $m > 1$, we will identify any element $\mu \in \mathcal{M}(Y; \mathbb{R}^m)$ with a vector $(\mu^1, \dots, \mu^m)$, with $\mu^i \in \mathcal{M}(Y)$ for all $i = 1, \dots, m$. If $\varphi = (\varphi^1, \dots, \varphi^m) \in B_b(Y; \mathbb{R}^m)$, the set of bounded $\mathbb{R}^m$-valued $\mathfrak{B}$-measurable maps, the duality between $\mu \in \mathcal{M}(Y; \mathbb{R}^m)$ and φ can be expressed by

$$\langle \mu, \varphi \rangle := \int_Y \varphi \cdot \mu(\mathrm{d}x) = \sum_{i=1}^m \int_Y \varphi^i(x) \mu^i(\mathrm{d}x).$$

For every $\mu \in \mathcal{M}(Y; \mathbb{R}^m)$ and $B \in \mathfrak{B}$ we will denote by $\mu \llcorner B$ the restriction of μ to B, i.e. $\mu \llcorner B(A) := \mu(A \cap B)$ for every $A \in \mathfrak{B}$.

Let $(X, \mathfrak{A})$ be another measurable space and let $\mathsf{p} : X \to Y$ a measurable map. For every $\mu \in \mathcal{M}(X; \mathbb{R}^m)$ we will denote by $\mathsf{p}_\sharp \mu$ the push-forward measure obtained by

$$\mathsf{p}_\sharp \mu(B) := \mu(\mathsf{p}^{-1}(B)) \quad \text{for every } B \in \mathfrak{B}. \tag{2.1}$$

For every couple $\mu \in \mathcal{M}(Y; \mathbb{R}^m)$ and $\gamma \in \mathcal{M}^+(Y)$ there exist a unique (up to the modification in a γ-negligible set) γ-integrable map $\frac{\mathrm{d}\mu}{\mathrm{d}\gamma} : Y \to \mathbb{R}^m$, a γ-negligible set $N \in \mathfrak{B}$ and a unique measure $\mu^\perp \in \mathcal{M}(Y; \mathbb{R}^m)$ yielding the *Lebesgue decomposition*

$$\mu = \mu^a + \mu^\perp, \quad \mu^a = \frac{\mathrm{d}\mu}{\mathrm{d}\gamma} \gamma = \mu \llcorner (Y \setminus N), \quad \mu^\perp = \mu \llcorner N, \quad \gamma(N) = 0$$

$$|\mu^\perp| \perp \gamma, \quad |\mu|(Y) = \int_Y \left| \frac{\mathrm{d}\mu}{\mathrm{d}\gamma} \right| \mathrm{d}\gamma + |\mu^\perp|(Y). \tag{2.2}$$

2.2 Convergence of measures

Besides the topology of convergence in total variation (induced by the norm $\|\cdot\|_{TV}$), we will also consider the topology of *setwise convergence*, i.e. the coarsest topology on $\mathcal{M}(Y; \mathbb{R}^m)$ making all the functions

$$\mu \mapsto \mu(B) \quad B \in \mathfrak{B}$$

continuous. For a sequence $(\mu_n)_{n \in \mathbb{N}}$ and a candidate limit μ in $\mathcal{M}(Y; \mathbb{R}^m)$ we have the following equivalent characterizations of the corresponding convergence [9, §4.7(v)]:

(1) Setwise convergence:

$$\lim_{n \to +\infty} \mu_n(B) = \mu(B) \quad \text{for every set } B \in \mathfrak{B}. \tag{2.3}$$

(2) Convergence in duality with $B_b(Y; \mathbb{R}^m)$:

$$\lim_{n \to +\infty} \langle \mu_n, \varphi \rangle = \langle \mu, \varphi \rangle \quad \text{for every } \varphi \in B_b(Y; \mathbb{R}^m). \tag{2.4}$$

(3) Weak topology of the Banach space: the sequence μ_n converges to μ in the weak topology of the Banach space $(\mathcal{M}(Y; \mathbb{R}^m); \| \cdot \|_{TV})$.
(4) Weak L^1-convergence of the densities: there exists a common dominating measure $\gamma \in \mathcal{M}^+(Y)$ such that $\mu_n \ll \gamma$, $\mu \ll \gamma$ and

$$\frac{\mathrm{d}\mu_n}{\mathrm{d}\gamma} \rightharpoonup \frac{\mathrm{d}\mu}{\mathrm{d}\gamma} \quad \text{weakly in } L^1(Y, \gamma; \mathbb{R}^m). \tag{2.5}$$

(5) Alternative form of weak L^1-convergence: (2.5) holds *for every* common dominating measure γ.

We will refer to *setwise convergence* for sequences satisfying one of the equivalent properties above. The above topologies also share the same notion of compact subsets, as stated in the following useful theorem, cf. [9, Theorem 4.7.25], where we shall denote by $\sigma(\mathcal{M}(Y; \mathbb{R}^m); B_b(Y; \mathbb{R}^m))$ the weak topology on $\mathcal{M}(Y; \mathbb{R}^m)$ induced by the duality with $B_b(Y; \mathbb{R}^m)$.

Theorem 2.1 *For every set $\emptyset \neq M \subset \mathcal{M}(Y; \mathbb{R}^m)$ the following properties are equivalent:*

(1) M has a compact closure in the topology of setwise convergence.
(2) M has a compact closure in the topology $\sigma(\mathcal{M}(Y; \mathbb{R}^m); B_b(Y; \mathbb{R}^m))$.
(3) M has a compact closure in the weak topology of $(\mathcal{M}(Y; \mathbb{R}^m); \| \cdot \|_{TV})$.
(4) Every sequence in M has a subsequence converging on every set of $\mathfrak{B}$.
(5) There exists a measure $\gamma \in \mathcal{M}^+(Y)$ such that

$$\forall \varepsilon > 0 \, \exists \delta > 0 : \quad B \in \mathfrak{B}, \, \gamma(B) \leq \delta \quad \Rightarrow \quad \sup_{\mu \in M} \mu(B) \leq \varepsilon. \tag{2.6}$$

(6) There exists a measure $\gamma \in \mathcal{M}^+(Y)$ such that $\mu \ll \gamma$ for every $\mu \in M$ and the set $\{\mathrm{d}\mu/\mathrm{d}\gamma : \mu \in M\}$ has compact closure in the weak topology of $L^1(Y, \gamma; \mathbb{R}^m)$.

We also recall a useful characterization of weak compactness in L^1.

Theorem 2.2 *Let $\gamma \in \mathcal{M}^+(Y)$ and $\emptyset \neq F \subset L^1(Y, \gamma; \mathbb{R}^m)$. The following properties are equivalent:*

(1) F has compact closure in the weak topology of $L^1(Y, \gamma; \mathbb{R}^m)$;
(2) F is bounded in $L^1(Y, \gamma; \mathbb{R}^m)$ and equi-absolutely continuous, i.e.

$$\forall \varepsilon > 0 \, \exists \delta > 0 : \quad B \in \mathfrak{B}, \, \gamma(B) \leq \delta \quad \Rightarrow \quad \sup_{f \in F} \int_B |f| \, \mathrm{d}\gamma \leq \varepsilon. \tag{2.7}$$

(3) There exists a convex and superlinear function $\beta : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\sup_{f \in F} \int_Y \beta(|f|) \, \mathrm{d}\gamma < +\infty. \tag{2.8}$$

The name 'equi-absolute continuity' above derives from the interpretation that the measure $f\gamma$ is absolutely continuous with respect to γ in a uniform manner; 'equi-absolute continuity' is a shortening of Bogachev's terminology 'F has uniformly absolutely continuous integrals' [9, Def. 4.5.2]. A fourth equivalent property is equi-integrability with respect to γ [9, Th. 4.5.3], a fact that we will not use.

When Y is endowed with a (separable and metrizable) topology τ_Y, we will use the symbol $C_b(Y; \mathbb{R}^m)$ to denote the space of bounded $\mathbb{R}^m$-valued continuous functions on (Y, τ_Y). We will consider the corresponding weak topology $\sigma(\mathcal{M}(Y; \mathbb{R}^m); C_b(Y; \mathbb{R}^m))$ induced by the

duality with $C_b(Y; \mathbb{R}^m)$. Prokhorov's Theorem yields that a subset $M \subset \mathcal{M}(Y; \mathbb{R}^m)$ has compact closure in this topology if it is bounded in the total variation norm and it is equally tight, i.e.

$$\forall \varepsilon > 0 \, \exists \, K \text{ compact in } Y: \quad \sup_{\mu \in M} |\mu|(Y \setminus K) \leq \varepsilon. \tag{2.9}$$

It is obvious that for a sequence $(\mu_n)_{n \in \mathbb{N}}$ convergence in total variation implies setwise convergence (or in duality with bounded measurable functions), and setwise convergence implies weak convergence in duality with bounded continuous functions.

2.3 Convex functionals and concave transformations of measures

We will use the following construction several times. Let $\psi : \mathbb{R}^m \to [0, +\infty]$ be convex and lower semicontinuous and let us denote by $\psi^\infty : \mathbb{R}^m \to [0, +\infty]$ its recession function

$$\psi^\infty(z) := \lim_{t \to +\infty} \frac{\psi(tz)}{t} = \sup_{t > 0} \frac{\psi(tz) - \psi(0)}{t}, \tag{2.10}$$

which is a convex, lower semicontinuous, and positively 1-homogeneous map with $\psi^\infty(0) = 0$. We define the functional $\mathscr{F}_\psi : \mathcal{M}(Y; \mathbb{R}^m) \times \mathcal{M}^+(Y) \mapsto [0, +\infty]$ by

$$\mathscr{F}_\psi(\mu|\nu) := \int_Y \psi\left(\frac{d\mu}{d\nu}\right) d\nu + \int_Y \psi^\infty\left(\frac{d\mu^\perp}{d|\mu^\perp|}\right) d|\mu^\perp|, \quad \text{for } \mu = \frac{d\mu}{d\nu}\nu + \mu^\perp. \tag{2.11}$$

Note that when ψ is superlinear then $\psi^\infty(x) = +\infty$ in $\mathbb{R}^m \setminus \{0\}$. Equivalently,

$$\psi \text{ superlinear}, \quad \mathscr{F}_\psi(\mu|\nu) < \infty \quad \Rightarrow \quad \mu \ll \nu, \quad \mathscr{F}_\psi(\mu|\nu) = \int_Y \psi\left(\frac{d\mu}{d\nu}\right) d\nu. \tag{2.12}$$

We collect in the next Lemma a list of useful properties.

Lemma 2.3

(1) When ψ is also positively 1-homogeneous, then $\psi \equiv \psi^\infty$, $\mathscr{F}_\psi(\cdot|\nu)$ is independent of ν and will also be denoted by $\mathscr{F}_\psi(\cdot)$: it satisfies

$$\mathscr{F}_\psi(\mu) = \int_Y \psi\left(\frac{d\mu}{d\gamma}\right) d\gamma \quad \text{for every } \gamma \in \mathcal{M}^+(Y) \text{ such that } \mu \ll \gamma. \tag{2.13}$$

(2) If $\hat{\psi} : \mathbb{R}^{m+1} \to [0, \infty]$ denotes the 1-homogeneous, convex, perspective function associated with ψ by

$$\hat{\psi}(z, t) := \begin{cases} \psi(z/t)t & \text{if } t > 0, \\ \psi^\infty(z) & \text{if } t = 0, \\ +\infty & \text{if } t < 0, \end{cases} \tag{2.14}$$

then

$$\mathscr{F}_\psi(\mu|\nu) = \mathscr{F}_{\hat{\psi}}(\mu, \nu) \quad \text{for every } (\mu, \nu) \in \mathcal{M}(Y; \mathbb{R}^m) \times \mathcal{M}^+(Y) \tag{2.15}$$

with $\mathscr{F}_{\hat{\psi}}$ defined as in (2.13).

(3) In particular, if $\gamma \in \mathcal{M}^+(Y)$ is a common dominating measure such that $\mu = u\gamma$, $\nu = v\gamma$, and $Y' := \{x \in Y : v(x) > 0\}$ we also have

$$\mathscr{F}_\psi(\mu|\nu) = \int_Y \hat{\psi}(u, v) \, d\gamma = \int_{Y'} \psi(u/v)v \, d\gamma + \int_{Y \setminus Y'} \psi^\infty(u) \, d\gamma. \tag{2.16}$$

(4) The functional $\mathscr{F}_\psi$ is convex; if ψ is also positively 1-homogeneous then

$$\begin{aligned}
\mathscr{F}_\psi(\mu + \mu') &\le \mathscr{F}_\psi(\mu) + \mathscr{F}_\psi(\mu') \\
\mathscr{F}_\psi(\mu + \mu') &= \mathscr{F}_\psi(\mu) + \mathscr{F}_\psi(\mu') \quad \text{if } \mu \perp \mu'.
\end{aligned} \tag{2.17}$$

(5) Jensen's inequality:

$$\hat{\psi}(\mu^a(B), \nu(B)) + \psi^\infty(\mu^\perp(B)) \le \mathscr{F}_\psi(\mu \llcorner B | \nu \llcorner B) \quad \text{for every } B \in \mathfrak{B} \tag{2.18}$$

(with $\mu = \mu^a + \mu^\perp$ the Lebesgue decomposition of μ w.r.t. ν).
(6) If $\psi(0) = 0$ then for every $\mu \in \mathcal{M}(Y, \mathbb{R}^m)$, $\nu, \nu' \in \mathcal{M}^+(Y)$

$$\nu \le \nu' \quad \Rightarrow \quad \mathscr{F}_\psi(\mu | \nu) \ge \mathscr{F}_\psi(\mu | \nu'). \tag{2.19}$$

(7) $\mathscr{F}_\psi$ is sequentially lower semicontinuous in $\mathcal{M}(Y; \mathbb{R}^m) \times \mathcal{M}^+(Y)$ with respect to the topology of setwise convergence.
(8) If $\mathfrak{B}$ is the Borel family induced by a Polish topology τ_Y on Y, $\mathscr{F}_\psi$ is lower semicontinuous with respect to weak convergence (in duality with continuous bounded functions).

Proof The above properties are mostly well known; we give a quick sketch of the proofs of the various claims for the ease of the reader.
(1) Let us set $u := \mathrm{d}\mu/\mathrm{d}\nu$, $u^\perp := \mathrm{d}\mu^\perp/\mathrm{d}|\mu|$ and let $N \in \mathfrak{B}$ ν-negligible such that $\mu^\perp = \mu \llcorner N$. We also set $N' := \{y \in Y \setminus N : |u(y)| > 0\}$; notice that $\nu \llcorner N' \ll |\mu|$. If v is the Lebesgue density of $|\mu|$ w.r.t. γ, since $\psi = \psi^\infty$ is positively 1-homogeneous and $\psi(0) = 0$, we have

$$\begin{aligned}
\mathscr{F}_\psi(\mu | \nu) &= \int_{N'} \psi(u) \, \mathrm{d}\nu + \int_N \psi(u^\perp) \, \mathrm{d}|\mu^\perp| = \int_{N'} \psi(u)/|u| \, \mathrm{d}|\mu| + \int_N \psi(u^\perp) \, \mathrm{d}|\mu^\perp| \\
&= \int_{N'} v\psi(u)/|u| \, \mathrm{d}\gamma + \int_N v\psi(u^\perp) \, \mathrm{d}\gamma = \int_{N'} \psi(uv/|u|) \, \mathrm{d}\gamma + \int_N \psi(u^\perp v) \, \mathrm{d}\gamma \\
&= \int_{N'} \psi(\mathrm{d}\mu/\mathrm{d}\gamma) \, \mathrm{d}\gamma + \int_N \psi(\mathrm{d}\mu/\mathrm{d}\gamma) \, \mathrm{d}\gamma = \int_Y \psi(\mathrm{d}\mu/\mathrm{d}\gamma) \, \mathrm{d}\gamma = \mathscr{F}_\psi(\mu | \gamma),
\end{aligned}$$

where we also used the fact that $|\mu|(Y \setminus (N \cup N')) = 0$, so that $\mathrm{d}\mu/\mathrm{d}\gamma = 0$ γ-a.e. on $Y \setminus (N \cup N')$.
(2) Since $\hat{\psi}$ is 1-homogeneous, we can apply the previous claim and evaluate $\mathscr{F}_{\hat{\psi}}(\mu, \nu)$ by choosing the dominating measure $\gamma := \nu + \mu^\perp$.
(3) It is an immediate consequence of the first two claims.
(4) By (2.15) it is sufficient to consider the 1-homogeneous case. The convexity then follows by the convexity of ψ and by choosing a common dominating measure to represent the integrals. Relations (2.17) are also immediate.
(5) Using (2.15) and selecting a dominating measure γ with $\gamma(B) = 1$, Jensen's inequality applied to the convex functional $\hat{\psi}$ yields

$$\begin{aligned}
\hat{\psi}(\mu(B), \nu(B)) = \hat{\psi}\Big(\int_B \frac{\mathrm{d}\mu}{\mathrm{d}\gamma} \, \mathrm{d}\gamma, \int_B \frac{\mathrm{d}\nu}{\mathrm{d}\gamma} \, \mathrm{d}\gamma \Big) &\le \int_B \hat{\psi}\Big(\frac{\mathrm{d}\mu}{\mathrm{d}\gamma}, \frac{\mathrm{d}\nu}{\mathrm{d}\gamma} \Big) \mathrm{d}\gamma \\
&= \mathscr{F}_{\hat{\psi}}(\mu \llcorner B, \nu \llcorner B).
\end{aligned}$$

Applying now the above inequality to the mutally singular couples (μ^a, ν) and $(\mu^\perp, 0)$ and using the second identity of (2.17) we obtain (2.18).
(6) We apply (2.15) and the first identity of (2.16), observing that if $\psi(0) = 0$ then $\hat{\psi}$ is decreasing with respect to its second argument.
(7) By (2.15) it is not restrictive to assume that Ψ is 1-homogeneous. If $(\mu_n)_n$ is a sequence setwise converging to μ in $\mathcal{M}(Y; \mathbb{R}^m)$ we can find a common dominating measure γ such

that (2.5) holds. The claimed property is then reduced to the weak lower semicontinuity of the functional

$$u \mapsto \int_Y \Psi(u)\,\mathrm{d}\gamma \tag{2.20}$$

in $L^1(Y, \gamma; \mathbb{R}^m)$. Since the functional of (2.20) is convex and strongly lower semicontinuous in $L^1(Y, \gamma; \mathbb{R}^m)$ (thanks to Fatou's Lemma), it is weakly lower semicontinuous as well.
(8) It follows by the same argument of [3, Theorem 2.34], by using a suitable dual formulation which holds also in Polish spaces, where all the finite Borel measures are Radon (see e.g. [47, Theorem 2.7] for positive measures). □

Concave transformation of vector measures

Let us set $\mathbb{R}_+ := [0, +\infty[$, $\mathbb{R}_+^m := (\mathbb{R}_+)^m$, and let $\alpha : \mathbb{R}_+^m \to \mathbb{R}_+$ be a continuous and concave function. It is obvious that α is non-decreasing with respect to each variable. As for (2.10), the recession function α^∞ is defined by

$$\alpha^\infty(z) := \lim_{t \to +\infty} \frac{\alpha(tz)}{t} = \inf_{t>0} \frac{\alpha(tz) - \alpha(0)}{t}, \quad z \in \mathbb{R}_+^m. \tag{2.21}$$

We define the corresponding map $\alpha : \mathcal{M}(Y; \mathbb{R}_+^m) \times \mathcal{M}^+(Y) \to \mathcal{M}^+(Y)$ by

$$\alpha[\mu|\gamma] := \alpha\Big(\frac{\mathrm{d}\mu}{\mathrm{d}\gamma}\Big)\gamma + \alpha^\infty\Big(\frac{\mathrm{d}\mu}{\mathrm{d}|\mu^\perp|}\Big)|\mu^\perp| \quad \mu \in \mathcal{M}(Y; \mathbb{R}_+^m), \ \gamma \in \mathcal{M}_+(Y), \tag{2.22}$$

where as usual $\mu = \frac{\mathrm{d}\mu}{\mathrm{d}\gamma}\gamma + \mu^\perp$ is the Lebesgue decomposition of μ with respect to γ; in what follows, we will use the short-hand $\mu_\gamma := \frac{\mathrm{d}\mu}{\mathrm{d}\gamma}\gamma$. We also mention in advance that, for shorter notation we will write $\alpha[\mu_1, \mu_2|\gamma]$ in place of $\alpha[(\mu_1, \mu_2)|\gamma]$.

Like for $\mathscr{F}$, it is not difficult to check that $\alpha[\mu|\gamma]$ is independent of γ if α is positively 1-homogeneous (and thus coincides with α^∞). If we define the perspective function $\hat{\alpha} : \mathbb{R}_+^{m+1} \to \mathbb{R}_+$

$$\hat{\alpha}(z, t) := \begin{cases} \alpha(z/t)t & \text{if } t > 0, \\ \alpha^\infty(z) & \text{if } t = 0 \end{cases} \tag{2.23}$$

we also get $\alpha[\mu|\gamma] = \hat{\alpha}(\mu, \gamma)$.

We denote by $\alpha_* : \mathbb{R}_+^m \to [-\infty, 0]$ the upper semicontinuous concave conjugate of α

$$\alpha_*(y) := \inf_{x \in \mathbb{R}_+^m} (y \cdot x - \alpha(x)), \quad D(\alpha_*) := \{y \in \mathbb{R}_+^m : \alpha_*(y) > -\infty\}. \tag{2.24}$$

The function α_* provides simple affine upper bounds for α

$$\alpha(x) \le x \cdot y - \alpha_*(y) \quad \text{for every } y \in D(\alpha_*) \tag{2.25}$$

and Fenchel duality yields

$$\alpha(x) = \inf_{y \in \mathbb{R}_+^m} (y \cdot x - \alpha_*(y)) = \inf_{y \in D(\alpha_*)} (y \cdot x - \alpha_*(y)). \tag{2.26}$$

We will also use that

$$\alpha^\infty(z) = \inf_{y \in D(\alpha_*)} y \cdot z. \tag{2.27}$$

Indeed, on the one hand for every $y \in D(\alpha_*)$ and $t > 0$ we have that

$$\alpha^\infty(z) \le \frac{1}{t}\big(\alpha(tz) - \alpha(0)\big) \le \frac{1}{t}\Big(y \cdot (tz) - \alpha(0) - \alpha^*(y)\Big);$$

by the arbitrariness of $t > 0$, we conclude that $\alpha^\infty(z) \le y \cdot z$ for every $y \in D(\alpha_*)$. On the other hand, by (2.26) we have

$$\alpha^\infty(z) = \inf_{t>0} \frac{\alpha(tz) - \alpha(0)}{t} = \inf_{t>0} \inf_{y \in D(\alpha^*)} \frac{y \cdot (tz) - \alpha^*(y) - \alpha(0)}{t}$$

$$= \inf_{y \in D(\alpha^*)} \left(y \cdot z + \inf_{t>0} \frac{-\alpha^*(y) - \alpha(0)}{t} \right)$$

$$= \inf_{y \in D(\alpha^*)} y \cdot z,$$

where we have used that $-\alpha^*(y) - \alpha(0) \ge 0$ since $\alpha(0) = \inf_{y \in D(\alpha^*)}(-\alpha^*(y))$.

For every Borel set $B \subset Y$, Jensen's inequality yields (recall the notation $\mu_\gamma = \frac{d\mu}{d\gamma}\gamma$)

$$\alpha[\mu|\gamma](B) \le \alpha\left(\frac{\mu_\gamma(B)}{\gamma(B)}\right)\gamma(B) + \alpha^\infty(\mu^\perp(B))$$

$$\alpha[\mu|\gamma](B) \le \alpha(\mu(B)) \quad \text{if } \alpha = \alpha^\infty \text{ is 1-homogeneous.}$$

(2.28)

In fact, for every $y, y' \in D(\alpha_*)$,

$$\alpha[\mu|\gamma](B) = \int_B \alpha[\mu|\gamma] \le \int_B \left(y \cdot \frac{d\mu}{d\gamma} - \alpha_*(y) \right) d\gamma + \int_B \left(y' \cdot \frac{d\mu}{d|\mu^\perp|} \right) d|\mu^\perp|$$

$$= y \cdot \mu_\gamma(B) - \alpha_*(y)\gamma(B) + y' \cdot \mu^\perp(B).$$

Taking the infimum with respect to y and y', and recalling (2.26) and (2.27), we find (2.28). Choosing $y = y'$ in the previous formula we also obtain the linear upper bound

$$\alpha[\mu|\gamma] \le y \cdot \mu - \alpha_*(y)\gamma \quad \text{for every } y \in D(\alpha_*).$$

(2.29)

2.4 Disintegration and kernels

Let $(X, \mathfrak{A})$ and $(Y, \mathfrak{B})$ be measurable spaces and let $\big(\kappa(x, \cdot)\big)_{x \in X}$ be a $\mathfrak{A}$-measurable family of measures in $\mathcal{M}^+(Y)$, i.e.

$$\text{for every } B \in \mathfrak{B}, \quad x \mapsto \kappa(x, B) \text{ is } \mathfrak{A}\text{-measurable.}$$

(2.30)

We will set

$$\kappa_Y(x) := \kappa(x, Y), \quad \|\kappa_Y\|_\infty := \sup_{x \in X} |\kappa|(x, Y),$$

(2.31)

and we say that κ is a bounded kernel if $\|\kappa_Y\|_\infty$ is finite. If $\gamma \in \mathcal{M}^+(X)$ and

$$\kappa_Y \text{ is } \gamma\text{-integrable, i.e. } \int_X \kappa(x, Y)\,\gamma(dx) < +\infty,$$

(2.32)

then Fubini's Theorem [18, II, 14] shows that there exists a unique measure $\kappa_\gamma(dx, dy) = \gamma(dx)\kappa(x, dy)$ on $(X \times Y, \mathfrak{A} \otimes \mathfrak{B})$ such that

$$\kappa_\gamma(A \times B) = \int_A \kappa(x, B)\,\gamma(dx) \quad \text{for every } A \in \mathfrak{A}, \ B \in \mathfrak{B}.$$

(2.33)

If $X = Y$, the measure γ is called *invariant* if κ_γ has the same marginals; equivalently

$$y_\sharp \kappa_\gamma(dy) = \int_X \kappa(x, dy)\gamma(dx) = \kappa_Y(y)\gamma(dy),$$

(2.34)

where $\mathsf{y} : E \to V$ denotes the projection on the second component, cf. (3.1) ahead. We say that γ is *reversible* if it satisfies *the detailed balance condition*, i.e. κ_γ is symmetric: $\mathsf{s}_\sharp \kappa_\gamma = \kappa_\gamma$. The concepts of invariance and detailed balance correspond to the analogous concepts in stochastic-process theory; see Sect. 3.1. It is immediate to check that reversibility implies invariance.

If $f : X \times Y \to \mathbb{R}$ is a positive or bounded measurable function, then

$$\text{the map } x \mapsto \kappa f(x) := \int_Y f(x, y) \kappa(x, \mathrm{d}y) \text{ is } \mathfrak{A}\text{-measurable} \qquad (2.35)$$

and

$$\int_{X \times Y} f(x, y)\, \kappa_\gamma(\mathrm{d}x, \mathrm{d}y) = \int_X \left(\int_Y f(x, y)\, \kappa(x, \mathrm{d}y) \right) \gamma(\mathrm{d}x). \qquad (2.36)$$

Conversely, if X, Y are standard Borel spaces, $\kappa \in \mathcal{M}^+(X \times Y)$ (with the product σ-algebra) and the first marginal $\mathsf{p}_\sharp^X \kappa$ of κ is absolutely continuous with respect to $\gamma \in \mathcal{M}^+(V)$, then we may apply the disintegration Theorem [9, Corollary 10.4.15] to find a γ-integrable kernel $(\kappa(x, \cdot))_{x \in X}$ such that $\kappa = \kappa_\gamma$.

We will often apply the above construction in two cases: when $X = Y := V$, the main domain of our evolution problems (see Assumptions $(V\pi\kappa)$ below), and when $X := I = (a, b)$ is an interval of the real line endowed with the Lebesgue measure λ. In this case, we will denote by t the variable in I and by $(\mu_t)_{t \in X}$ a measurable family in $\mathcal{M}(Y)$ parametrized by $t \in I$:

$$\text{if } \int_I \mu_t(Y)\, \mathrm{d}t < +\infty \text{ then we set } \mu_\lambda \in \mathcal{M}(I \times Y), \quad \mu_\lambda(\mathrm{d}t, \mathrm{d}y) = \lambda(\mathrm{d}t)\mu_t(\mathrm{d}y). \quad (2.37)$$

Lemma 2.4 *If $\mu_n \in \mathcal{M}(X)$ is a sequence converging to μ setwise and $(\kappa(x, \cdot))_{x \in X}$ is a bounded measurable kernel in $\mathcal{M}^+(Y)$, then $\kappa_{\mu_n} \to \kappa_\mu$ setwise in $\mathcal{M}(X \times Y, \mathfrak{A} \otimes \mathfrak{B})$.*

If X, Y are Polish spaces and κ also satisfies the weak Feller property, i.e.

$$x \mapsto \kappa(x, \cdot) \text{ is weakly continuous in } \mathcal{M}^+(Y), \qquad (2.38)$$

(where 'weak' means in duality with continuous bounded functions), then for every weakly converging sequence $\mu_n \to \mu$ in $\mathcal{M}(X)$ we have $\kappa_{\mu_n} \to \kappa_\mu$ weakly as well.

Proof If $f : X \times Y \to \mathbb{R}$ is a bounded $\mathfrak{A} \otimes \mathfrak{B}$-measurable map, then by (2.35) also the map κf is bounded and $\mathfrak{A}$-measurable so that

$$\lim_{n \to +\infty} \int_{X \times Y} f\, \mathrm{d}\kappa_{\mu_n} = \lim_{n \to +\infty} \int_X \kappa f\, \mathrm{d}\mu_n = \int_X \kappa f\, \mathrm{d}\mu = \int_{X \times Y} f\, \mathrm{d}\kappa_\mu,$$

showing the setwise convergence. The other statement follows by a similar argument. $\qquad \square$

3 Jump processes, large deviations, and their generalized gradient structures

3.1 The systems of this paper

In the Introduction we described jump processes on V with kernel κ, and showed that the evolution equation $\partial_t \rho_t = Q^* \rho_t$ for the law ρ_t of the process is a generalized gradient flow characterized by a driving functional $\mathscr{E}$ and a dissipation potential $\mathscr{R}^*$.

The mathematical setup of this paper is slightly different. Instead of starting with an evolution equation and proceeding to the generalized gradient system, our mathematical development starts with the generalized gradient system; we then consider the equation to be defined by this system. In this Section, therefore, we describe assumptions that we make on $\mathscr{E}$ and $\mathscr{R}^*$ that will allow us to set up the rigorous functional framework for the evolution equation (1.9).

We first state the assumptions about the sets V of 'vertices' and $E:=V \times V$ of 'edges'. 'Edges' are identified with ordered pairs (x, y) of vertices $x, y \in V$. We will denote by $\mathsf{x}, \mathsf{y} : E \to V$ and $\mathsf{s} : E \to E$ the coordinate and the symmetry maps defined by

$$\mathsf{x}(x, y):=x, \quad \mathsf{y}(x, y):=y, \quad \mathsf{s}(x, y) := (y, x) \quad \text{for every } x, y \in V. \tag{3.1}$$

Assumption ($V\pi\kappa$). We assume that

$$(V, \mathfrak{B}, \pi) \text{ is a standard Borel measure space, } \pi \in \mathcal{M}^+(V), \tag{3.2}$$

$(\kappa(x, \cdot))_{x \in V}$ is bounded kernel in $\mathcal{M}^+(V)$ (see Section 2.4), satisfying the *detailed-balance condition*

$$\int_A \kappa(x, B)\, \pi(\mathrm{d}x) = \int_B \kappa(y, A)\, \pi(\mathrm{d}y) \quad \text{for all } A, B \in \mathfrak{B}, \tag{3.3}$$

and the uniform upper bound

$$\|\kappa_V\|_\infty = \sup_{x \in V} \kappa(x, V) < +\infty. \tag{3.4}$$

The measure $\pi \in \mathcal{M}^+(V)$ often is referred to as the *invariant measure*, and it will be stationary under the evolution generated by the generalized gradient system. By Fubini's Theorem (see Sect. 2.4) we also introduce the measure ϑ on E given by

$$\vartheta(\mathrm{d}x\, \mathrm{d}y) = \kappa_\pi(\mathrm{d}x, \mathrm{d}y) = \pi(\mathrm{d}x)\kappa(x, \mathrm{d}y), \quad \vartheta(A \times B) = \int_A \kappa(x, B)\, \pi(\mathrm{d}x). \tag{3.5}$$

Note that the invariance of the measure π and the detailed balance condition (3.3) can be rephrased in terms of ϑ as

$$\mathsf{x}_\sharp \vartheta = \mathsf{y}_\sharp \vartheta, \quad \mathsf{s}_\# \vartheta = \vartheta. \tag{3.6}$$

Conversely, if we choose a symmetric measure $\vartheta \in \mathcal{M}^+(E)$ such that

$$\mathsf{x}_\sharp \vartheta \ll \pi, \quad \frac{\mathrm{d}(\mathsf{x}_\sharp \vartheta)}{\mathrm{d}\pi} \leq \|\kappa_V\|_\infty < +\infty \quad \pi\text{-a.e.} \tag{3.7}$$

then the disintegration Theorem [9, Corollary 10.4.15] shows the existence of a bounded measurable kernel $(\kappa(x, \cdot))_{x \in X}$ satisfying (3.3) and (3.5).

We next turn to the driving functional, which is given by the construction in (2.11) and (2.12) for a superlinear density $\psi = \phi$ and for the choice $\gamma = \pi$.

Assumption ($\mathscr{E}\phi$). The driving functional $\mathscr{E} : \mathcal{M}^+(V) \to [0, +\infty]$ is of the form

$$\mathscr{E}(\rho) := \mathscr{F}_\phi(\rho|\pi) = \begin{cases} \int_V \phi\left(\dfrac{\mathrm{d}\rho}{\mathrm{d}\pi}\right) \mathrm{d}\pi & \text{if } \rho \ll \pi, \\ +\infty & \text{otherwise,} \end{cases} \qquad (3.8)$$

with

$$\phi \in C([0, +\infty)) \cap C^1((0, +\infty)), \ \ \min \phi = 0, \text{ and } \phi \text{ is convex} \qquad (3.9)$$
$$\text{with superlinear growth at infinity.}$$

Under these assumptions the functional $\mathscr{E}$ is lower semicontinuous on $\mathcal{M}^+(V)$ both with respect to the topology of setwise convergence, and any compatible weak topology (see Lemma 2.3). A central example was already mentioned in the introduction, i.e. the Boltzmann-Shannon entropy function

$$\phi(s) = s \log s - s + 1, \qquad s \geq 0. \qquad (3.10)$$

Finally, we state our assumptions on the dissipation.

Assumption ($\mathscr{R}^*\Psi\alpha$). We assume that the dual dissipation density Ψ^* satisfies

$$\left.\begin{array}{l} \Psi^* : \mathbb{R} \to [0, +\infty) \text{ is convex, differentiable, even, with } \Psi^*(0) = 0, \text{ and} \\ \qquad\qquad \lim_{|\xi| \to \infty} \dfrac{\Psi^*(\xi)}{|\xi|} = +\infty. \end{array}\right\} \qquad (3.11)$$

The flux density map $\alpha : [0, +\infty) \times [0, +\infty) \to [0, +\infty)$, with $\alpha \not\equiv 0$, is continuous, concave, symmetric:

$$\alpha(u_1, u_2) = \alpha(u_2, u_1) \quad \text{for all } u_1, u_2 \in [0, +\infty), \qquad (3.12)$$

and its recession function α^∞ vanishes on the boundary of $\mathbb{R}_+^2$:

$$\text{for every } u_1, u_2 \in \mathbb{R}_+^2 : \quad u_1 u_2 = 0 \implies \alpha^\infty(u_1, u_2) = 0. \qquad (3.13)$$

Note that since α is nonnegative, concave, and not trivially 0, it cannot vanish in the interior of $\mathbb{R}_+^2$, i.e.

$$u_1 u_2 > 0 \implies \alpha(u_1, u_2) > 0. \qquad (3.14)$$

The examples that we gave in the introduction of the cosh-type dissipation (1.17a) and the quadratic dissipation (1.17b) both fit these assumptions; other examples are

$$\alpha(u, v) = 1 \quad \text{and} \quad \alpha(u, v) = u + v.$$

In some cases we will use an additional property, namely that α is positively 1-homogeneous, i.e. $\alpha(\lambda u_1, \lambda u_2) = \lambda \alpha(u_1, u_2)$ for all $\lambda \geq 0$. This 1-homogeneity is automatically satisfied under the compatibility condition (1.13), with the Boltzmann entropy function $\phi(s) = s \log s - s + 1$.

Concaveness of α is a natural assumption in view of the convexity of $\mathscr{R}$ (cf. Remark 1.1 and Lemma 4.10 ahead), while 1-homogeneity will make the definition of $\mathcal{R}$ independent of the choice of a reference measure. It is interesting to observe that the concavity and symmetry conditions, that one has to naturally assume to ensure the aforementioned properties of $\mathscr{R}$, were singled out for the analog of the function α in the construction of the distance yielding the quadratic gradient structure of [49].

The choices for Ψ^* above generate corresponding properties for the Legendre dual Ψ:

Lemma 3.1 *Under Assumption ($\mathscr{R}^*\Psi\alpha$), the function $\Psi : \mathbb{R} \to \mathbb{R}$ is even and satisfies*

$$0 = \Psi(0) < \Psi(s) < +\infty \text{ for all } s \in \mathbb{R} \setminus \{0\}. \tag{3.15a}$$

$$\Psi \text{ is strictly convex, strictly increasing, and superlinear.} \tag{3.15b}$$

Proof The superlinearity of Ψ^* implies that $\Psi(s) < +\infty$ for all $s \in \mathbb{R}$, and similarly the finiteness of Ψ^* on $\mathbb{R}$ implies that Ψ is superlinear. Since Ψ^* is even, Ψ is convex and even, and therefore $\Psi(s) \geq \Psi(0) = \sup_{\xi \in \mathbb{R}}[-\Psi^*(\xi)] = 0$. Furthermore, since for all $p \in \mathbb{R}$, $\operatorname{argmin}_{s \in \mathbb{R}}(\Psi(s) - ps) = \partial\Psi^*(p)$ (see e.g. [77, Thm. 11.8]) and Ψ^* is differentiable at every p, we conclude that $\operatorname{argmin}_s(\Psi(s) - ps) = \{(\Psi^*)'(p)\}$; therefore each point of the graph of Ψ is an exposed point. It follows that Ψ is strictly convex, and $\Psi(s) > 0$ for all $s \neq 0$. $\qquad\square$

As described in the introduction, we use Ψ, Ψ^*, and α to define the dual pair of dissipation potentials $\mathscr{R}$ and $\mathscr{R}^*$, which for a couple of measures $\rho = u\pi \in \mathcal{M}^+(V)$ and $j \in \mathcal{M}(E)$ are formally given by

$$\mathscr{R}(\rho, j) := \frac{1}{2} \int_E \Psi\left(2\frac{dj}{dv_\rho}\right) dv_\rho, \qquad \mathscr{R}^*(\rho, \xi) := \frac{1}{2} \int_E \Psi^*(\xi)\, dv_\rho, \tag{3.16}$$

with

$$v_\rho(dx\, dy) := \alpha\big(u(x), u(y)\big)\, \vartheta(dx\, dy) = \alpha\big(u(x), u(y)\big)\, \pi(dx)\kappa(x, dy). \tag{3.17}$$

This expression for the edge measure v_ρ also is implicitly present in the structure built in [30,49]. The above definitions are made rigorous in Definition 4.9 and in (4.20) below.

The three sets of conditions above, Assumptions ($V\pi\kappa$), ($\mathscr{E}\phi$), and ($\mathscr{R}^*\Psi\alpha$), are the main assumptions of this paper. Under these assumptions, the evolution equation (1.9) may be linear or nonlinear in ρ. The equation coincides with the Forward Kolmogorov equation (1.2) if and only if condition (1.13) is satisfied, as shown below.

Calculation for (1.13)

Let us call $\mathscr{Q}[\rho]$ the right-hand side of (1.9) and let us compute

$$\langle \mathscr{Q}[\rho], \varphi \rangle = \left\langle -\overline{\mathrm{div}}\left[D_\xi \mathscr{R}^*\left(\rho, -\overline{\nabla}\phi'\left(\frac{d\rho}{d\pi}\right)\right)\right], \varphi \right\rangle$$

for every $\varphi \in B_b(V)$ and $\rho \in \mathcal{M}^+(V)$ with $\rho \ll \pi$. With $u = \frac{d\rho}{d\pi}$ we thus obtain

$$\langle \mathscr{Q}[\rho], \varphi \rangle = \left\langle D_\xi \mathscr{R}^*\big(\rho, -\overline{\nabla}\phi'(u)\big), \overline{\nabla}\varphi \right\rangle$$

$$= \frac{1}{2} \iint_E (\Psi^*)'\big(-\overline{\nabla}\phi'(u)(x, y)\big)\, \overline{\nabla}\varphi(x, y) v_\rho(dx, dy). \tag{3.18}$$

Recalling the definitions (3.17) of $\mathbf{v}_\rho$ and (1.11) of F, (3.18) thus becomes

$$
\begin{aligned}
\langle \mathscr{Q}[\rho], \varphi \rangle &= \frac{1}{2} \iint_E \left(\Psi^* \right)' \left(\phi'(u(x)) - \phi'(u(y)) \right) \overline{\nabla} \varphi(x, y) \, \alpha(u(x), u(y)) \boldsymbol{\vartheta}(\mathrm{d}x, \mathrm{d}y) \\
&= \frac{1}{2} \iint_E F(u(x), u(y)) \left(\varphi(x) - \varphi(y) \right) \boldsymbol{\vartheta}(\mathrm{d}x, \mathrm{d}y) \\
&\overset{(*)}{=} \iint_E F(u(x), u(y)) \varphi(x) \, \boldsymbol{\vartheta}(\mathrm{d}x, \mathrm{d}y) \\
&= \int_V \varphi(x) \left(\int_V F(u(x), u(y)) \kappa(x, \mathrm{d}y) \right) \pi(\mathrm{d}x)
\end{aligned}
\tag{3.19}
$$

where for $(*)$ we used the symmetry of $\boldsymbol{\vartheta}$ (i.e. the detailed-balance condition). This calculation justifies (1.12).

In the linear case of (1.2) it is immediate to see that

$$
\begin{aligned}
\langle Q^* \rho, \varphi \rangle = \langle \rho, Q\varphi \rangle &= \iint_E [\varphi(y) - \varphi(x)] \, \kappa(x, \mathrm{d}y) \rho(\mathrm{d}x) \\
&= \frac{1}{2} \iint_E \overline{\nabla} \varphi(x, y) \big[\kappa(x, \mathrm{d}y) \rho(\mathrm{d}x) - \kappa(y, \mathrm{d}x) \rho(\mathrm{d}y) \big] \\
&= \frac{1}{2} \iint_E \overline{\nabla} \varphi(x, y) \big[u(x) - u(y) \big] \boldsymbol{\vartheta}(\mathrm{d}x, \mathrm{d}y),
\end{aligned}
\tag{3.20}
$$

Comparing (3.20) and (3.19) we obtain that F has to fulfill (1.13).

4 Curves in $\mathcal{M}^+(V)$

A major challenge in any rigorous treatment of an equation such as (1.1) is finding a way to deal with the time derivative. The Ambrosio-Gigli-Savaré framework for metric-space gradient systems, for instance, is organized around absolutely continuous curves. These are a natural choice because on the one hand this class admits a 'metric velocity' that generalizes the time derivative, while on the other hand solutions are automatically absolutely continuous by the superlinear growth of the dissipation potential.

For the systems of this paper, a similar role is played by curves such that the 'action' $\int \mathscr{R} \, \mathrm{d}t$ is finite; we show below that the superlinearity of $\mathscr{R}(\rho, \boldsymbol{j})$ in $\boldsymbol{j}$ leads to similarly beneficial properties. In order to exploit this aspect, however, a number of intermediate steps need to be taken:

(a) We define the class $\mathcal{CE}(0, T)$ of solutions $(\rho, \boldsymbol{j})$ of the continuity equation (1.23) (Definition 4.1).
(b) For such solutions, $t \mapsto \rho_t$ is continuous in the total variation distance (Corollary 4.3).
(c) We give a rigorous definition of the functional $\mathscr{R}$ (Definition 4.9), and describe its behaviour on absolutely continuous and singular parts of $(\rho, \boldsymbol{j})$ (Lemma 4.10 and Theorem 4.13).
(d) If the action functional $\int \mathscr{R}$ is finite along a solution $(\rho, \boldsymbol{j})$ of the continuity equation in $[0, T]$, then the property that ρ_t is absolutely continuous with respect to π at some time $t \in [0, T]$ propagates to all the interval $[0, T]$ (Corollary 4.14).
(e) We prove a chain rule for the derivative of convex entropies along curves of finite $\mathscr{R}$-action (Theorem 4.16) and derive an estimate involving $\mathscr{R}$ and a Fisher-information-like term (Corollary 4.20).

(f) If the action $\int \mathscr{R}$ is uniformly bounded along a sequence $(\rho^n, j^n) \in \mathcal{CE}(0, T)$, then the sequence is compact in an appropriate sense (Proposition 4.21).

Once properties (a)–(f) have been established, the next step is to consider finite-action curves that also connect two given values μ, ν, leading to the definition of the Dynamical-Variational Transport (DVT) cost

$$\mathscr{W}(\tau, \mu, \nu) := \inf \left\{ \int_0^\tau \mathscr{R}(\rho_t, j_t) \, dt \; : \; (\rho, j) \in \mathcal{CE}(0, \tau), \; \rho_0 = \mu, \; \rho_\tau = \nu \right\}. \quad (4.1)$$

This definition is in the spirit of the celebrated Benamou-Brenier formula for the Wasserstein distance [8], generalized to a broader family of transport distances [20] and to jump processes [30,49]. However, a major difference with those constructions is that $\mathscr{W}$ also depends on the time variable τ and that $\mathscr{W}(\tau, \cdot, \cdot)$ is not a (power of a) distance, since Ψ is not, in general, positively homogeneous of any order. Indeed, when $\mathscr{R}$ is p-homogeneous in j, for $p \in (1, +\infty)$, we have (see also the discussion at the beginning of Sec. 7.1)

$$\mathscr{W}(\tau, \mu, \nu) = \frac{1}{\tau^{p-1}} \mathscr{W}(1, \mu, \nu) = \frac{1}{p\tau^{p-1}} d_\mathscr{R}^p(\mu, \nu), \quad (4.2)$$

where $d_\mathscr{R}$ is an extended distance and is a central object in the usual Minimizing-Movement construction. In Sect. 7, the DVT cost $\mathscr{W}$ will replace the rescaled p-power of the distance and play a similar role for the Minimizing-Movement approach.

For the rigorous construction of $\mathscr{W}$,

(g) we show that minimizers of (4.1) exist (Corollary 4.22);
(h) we establish properties of $\mathscr{W}$ that generalize those of the metric-space version (4.2) (Theorem 4.26).

Finally,

(i) we close the loop by showing that from a given functional $\mathscr{W}$ integrals of the form $\int_a^b \mathscr{R}$ can be reconstructed (Proposition 4.27).

Throughout this section we adopt **Assumptions ($V\pi\kappa$) and ($\mathscr{R}^*\Psi\alpha$)**.

4.1 The continuity equation

We now introduce the formulation of the continuity equation we will work with. Hereafter, for a given function $\mu : I \to \mathcal{M}(V)$, or $\mu : I \to \mathcal{M}(E)$, with $I = [a, b] \subset \mathbb{R}$, we shall often write μ_t in place of $\mu(t)$ for a given $t \in I$ and denote the time-dependent function μ by $(\mu_t)_{t \in I}$. We will write λ for the Lebesgue measure on I. The following definition mimics those given in [4, Sec. 8.1] and [21, Def. 4.2].

Definition 4.1 (Solutions (ρ, j) of the continuity equation) Let $I = [a, b]$ be a closed interval of $\mathbb{R}$. We denote by $\mathcal{CE}(I)$ the set of pairs (ρ, j) given by

- a family of time-dependent measures $\rho = (\rho_t)_{t \in I} \subset \mathcal{M}^+(V)$, and
- a measurable family $(j_t)_{t \in I} \subset \mathcal{M}(E)$ with $\int_0^T |j_t|(E) \, dt < +\infty$, satisfying the continuity equation

$$\dot\rho + \overline{\mathrm{div}} \, j = 0 \quad \text{in } I \times V, \quad (4.3)$$

in the following sense:

$$\int_V \varphi \, d\rho_{t_2} - \int_V \varphi \, d\rho_{t_1} = \iint_{J \times E} \overline{\nabla}\varphi \, d\boldsymbol{j}_\lambda \quad \text{for all } \varphi \in B_b(V), \ J = [t_1, t_2] \subset I. \quad (4.4)$$

where $\boldsymbol{j}_\lambda(dt, dx, dy) := \lambda(dt)\boldsymbol{j}_t(dx, dy)$.

Given $\rho_0, \rho_1 \in \mathcal{M}^+(V)$, we will use the notation

$$\mathcal{CE}(I; \rho_0, \rho_1) := \left\{(\rho, \boldsymbol{j}) \in \mathcal{CE}(I) : \rho(a) = \rho_0, \ \rho(b) = \rho_1\right\}.$$

Remark 4.2 The requirement (4.4) shows in particular that $t \mapsto \rho_t$ is continuous with respect to the total variation metric. Choosing $\varphi \equiv 1$ in (4.4), one immediately finds that

$$\text{the total mass } \rho_t(V) \text{ is constant in } I. \quad (4.5)$$

By the disintegration theorem, it is equivalent to assign the measurable family $(\boldsymbol{j}_t)_{t \in I}$ in $\mathcal{M}(E)$ or the measure $\boldsymbol{j}_\lambda$ in $\mathcal{M}(I \times E)$.

We can in fact prove a more refined property. The proof of the Corollary below is postponed to Appendix B.

Corollary 4.3 *If $(\rho, \boldsymbol{j}) \in \mathcal{CE}(0, T)$, then there exist a* common *dominating measure $\gamma \in \mathcal{M}^+(V)$ (i.e., $\rho_t \ll \gamma$ for all $t \in [a, b]$), and an absolutely continuous map $\tilde{u} : [a, b] \to L^1(V, \gamma)$ such that $\rho_t = \tilde{u}_t \gamma \ll \gamma$ for every $t \in [a, b]$.*

The interpretation of the continuity equation in Definition 4.1—in duality with all bounded measurable functions—is quite strong, and in particular much stronger than the more common continuity in duality with *continuous* and bounded functions. However, this continuity equation can be recovered starting from a much weaker formulation. The following result illustrates this; it is a translation of [4, Lemma 8.1.2] (cf. also [21, Lemma 4.1]) to the present setting. The proof adapts the argument for [4, Lemma 8.1.2] and is given in Appendix B.

Lemma 4.4 (Continuous representative) *Let $(\rho_t)_{t \in I} \subset \mathcal{M}^+(V)$ and $(\boldsymbol{j}_t)_{t \in I}$ be measurable families that are integrable with respect to λ and let τ be any separable and metrizable topology inducing $\mathcal{B}$. If*

$$-\int_0^T \eta'(t)\left(\int_V \zeta(x)\rho_t(dx)\right) dt = \int_0^T \eta(t)\left(\iint_E \overline{\nabla}\zeta(x, y)\boldsymbol{j}_t(dx\,dy)\right) dt, \quad (4.6)$$

holds for every $\eta \in C_c^\infty((a, b))$ and $\zeta \in C_b(V, \tau)$, then there exists a unique curve $I \ni t \mapsto \tilde{\rho}_t \in \mathcal{M}^+(V)$ such that $\tilde{\rho}_t = \rho_t$ for λ-a.e. $t \in I$. The curve $\tilde{\rho}$ is continuous in the total-variation norm with estimate

$$\|\tilde{\rho}_{t_2} - \tilde{\rho}_{t_1}\|_{TV} \le 2\int_{t_1}^{t_2} |\boldsymbol{j}_t|(E)\, dt \quad \text{for all } t_1 \le t_2, \quad (4.7)$$

and satisfies

$$\int_V \varphi(t_2, \cdot)\, d\tilde{\rho}_{t_2} - \int_V \varphi(t_1, \cdot)\, d\tilde{\rho}_{t_1} = \int_{t_1}^{t_2}\int_V \partial_t\varphi\, d\tilde{\rho}_t\, dt + \int_{J \times E} \overline{\nabla}\varphi\, d\boldsymbol{j}_\lambda \quad (4.8)$$

for all $\varphi \in C^1(I; B_b(V))$ and $J = [t_1, t_2] \subset T$.

Remark 4.5 In (4.4) we can always replace $\boldsymbol{j}$ with the positive measure $\boldsymbol{j}^+ := (\boldsymbol{j} - s_\#\boldsymbol{j})_+ = (2\boldsymbol{j}^\flat)_+$, since $\overline{\text{div}}\,\boldsymbol{j} = \overline{\text{div}}\,\boldsymbol{j}^+$ (see Lemma B.1); therefore we can assume without loss of generality that $\boldsymbol{j}$ is a positive measure.

As another immediate consequence of (4.4), the concatenation of two solutions of the continuity equation is again a solution; the result below also contains a statement about time rescaling of the solutions, whose proof follows from trivially adapting that of [4, Lemma 8.1.3] and is thus omitted.

Lemma 4.6 (Concatenation and time rescaling)

(1) Let $(\rho^i, j^i) \in \mathcal{CE}(0, T_i)$, $i = 1, 2$, with $\rho_{T_1}^1 = \rho_0^2$. Define $(\rho_t, j_t)_{t \in [0, T_1 + T_2]}$ by

$$\rho_t := \begin{cases} \rho_t^1 & \text{if } t \in [0, T_1], \\ \rho_{t-T_1}^2 & \text{if } t \in [T_1, T_1 + T_2], \end{cases} \qquad j_t := \begin{cases} j_t^1 & \text{if } t \in [0, T_1], \\ j_{t-T_1}^2 & \text{if } t \in [T_1, T_1 + T_2]. \end{cases}$$

Then, $(\rho, j) \in \mathcal{CE}(0, T_1 + T_2)$.

(2) Let $t : [0, \hat{T}] \to [0, T]$ be strictly increasing and absolutely continuous, with inverse $s : [0, T] \to [0, \hat{T}]$. Then, $(\rho, j) \in \mathcal{CE}(0, T)$ if and only if $\hat{\rho} := \rho \circ t$ and $\hat{j} := t'(j \circ t)$ fulfill $(\hat{\rho}, \hat{j}) \in \mathcal{CE}(0, \hat{T})$.

4.2 Definition of the dissipation potential $\mathscr{R}$

In this section we give a rigorous definition of the dissipation potential $\mathscr{R}$, following the formal descriptions above. In the special case when ρ and j are absolutely continuous, i.e.

$$\rho = u\pi \ll \pi \quad \text{and} \quad 2j = w\boldsymbol{\vartheta} \ll \boldsymbol{\vartheta}, \tag{4.9}$$

we set

$$E' := \{(x, y) \in E : \alpha(u(x), u(y)) > 0\}, \tag{4.10}$$

and in this case we can define the functional $\mathscr{R}$ by the direct formula

$$\mathscr{R}(\rho, j) = \begin{cases} \dfrac{1}{2} \displaystyle\int_{E'} \Psi\left(\dfrac{w(x, y)}{\alpha(u(x), u(y))}\right) \alpha(u(x), u(y)) \, \boldsymbol{\vartheta}(\mathrm{d}x, \mathrm{d}y) & \text{if } |j|(E \setminus E') = 0, \\ +\infty & \text{if } |j|(E \setminus E') > 0. \end{cases} \tag{4.11}$$

Recalling the definition of the perspective function $\hat{\Psi}$ (2.14), we can also write (4.11) in the equivalent and more compact form

$$\mathscr{R}(\rho, j) = \frac{1}{2} \iint_E \hat{\Psi}\big(w(x, y), \alpha(u(x), u(y))\big) \, \boldsymbol{\vartheta}(\mathrm{d}x, \mathrm{d}y), \quad 2j = w\boldsymbol{\vartheta}. \tag{4.12}$$

so that it is natural to introduce the function $\Upsilon : [0, +\infty) \times [0, +\infty) \times \mathbb{R} \to [0, +\infty]$,

$$\Upsilon(u, v, w) := \hat{\Psi}(w, \alpha(u, v)), \tag{4.13}$$

observing that

$$\mathscr{R}(\rho, j) = \frac{1}{2} \iint_E \Upsilon(u(x), u(y), w(x, y)) \, \boldsymbol{\vartheta}(\mathrm{d}x, \mathrm{d}y) \quad \text{for } 2j = w\boldsymbol{\vartheta}. \tag{4.14}$$

Lemma 4.7 The function $\Upsilon : [0, +\infty) \times [0, +\infty) \times \mathbb{R} \to [0, +\infty]$ defined above is convex and lower semicontinuous, with recession functional

$$\Upsilon^\infty(u, v, w) = \hat{\Psi}(w, \alpha^\infty(u, v)) = \begin{cases} \Psi\left(\dfrac{w}{\alpha^\infty(u, v)}\right) \alpha^\infty(u, v) & \text{if } \alpha^\infty(u, v) > 0 \\ 0 & \text{if } w = 0 \\ +\infty & \text{if } w \neq 0 \text{ and } \alpha^\infty(u, v) = 0. \end{cases} \tag{4.15}$$

For any $u, v \in [0, \infty)$ with $\alpha^\infty(u, v) > 0$, the map $w \mapsto \Upsilon(u, v, w)$ is strictly convex.
    If α is positively 1-homogeneous then Υ is positively 1-homogeneous as well.

Proof Note that Υ may be equivalently represented in the form

$$\Upsilon(u, v, w) = \sup_{\xi \in \mathbb{R}} \left\{ \xi w - \alpha(u, v) \Psi^*(\xi) \right\} =: \sup_{\xi \in \mathbb{R}} f_\xi(u, v, w). \tag{4.16}$$

The convexity of f_ξ for each $\xi \in \mathbb{R}$ readily follows from its linearity in w and the convexity of $-\alpha$ in (u, v). Therefore, Υ is convex and lower semicontinuous as the pointwise supremum of a family of convex continuous functions.

The characterization (4.15) of Υ^∞ follows from observing that $\Upsilon(0, 0, 0) = \hat{\Psi}(0, 0) = 0$ and using the 1-homogeneity of $\hat{\Psi}$:

$$\lim_{t \to +\infty} t^{-1} \Upsilon(tu, tv, tw) = \lim_{t \to +\infty} t^{-1} \hat{\Psi}\left(tw, \alpha(tu, tv) \right) = \lim_{t \to +\infty} \hat{\Psi}\left(w, t^{-1} \alpha(tu, tv) \right)$$
$$= \hat{\Psi}\left(w, \alpha^\infty(u, v) \right),$$

where the last equality follows from the continuity of $r \mapsto \hat{\Psi}(w, r)$ for all $w \in \mathbb{R}$.

The strict convexity of $w \mapsto \Upsilon(u, v, w)$ for any $u, v \in [0, \infty)$ with $\alpha^\infty(u, v) > 0$ follows directly from the strict convexity of Ψ (cf. Lemma 3.1). $\qquad \square$

The choice (4.14) provides a rigorous definition of $\mathscr{R}$ for couples of measures (ρ, j) that are absolutely continuous with respect to π and ϑ. In order to extend $\mathscr{R}$ to pairs (ρ, j) that are not absolutely continuous, it is useful to interpret the measure

$$\nu_\rho(\mathrm{d}x, \mathrm{d}y) := \alpha(u(x), u(y)) \vartheta(\mathrm{d}x, \mathrm{d}y) \tag{4.17}$$

in the integral of (4.11) in terms of a suitable concave transformation as in (2.22) of two couplings generated by ρ. We therefore introduce the measures

$$\vartheta_\rho^-(\mathrm{d}x\,\mathrm{d}y) := \rho(\mathrm{d}x)\kappa(x, \mathrm{d}y), \qquad \vartheta_\rho^+(\mathrm{d}x\,\mathrm{d}y) := \rho(\mathrm{d}y)\kappa(y, \mathrm{d}x) = s_\# \vartheta_\rho^-(\mathrm{d}x\,\mathrm{d}y), \tag{4.18}$$

observing that

$$\rho = u\pi \ll \pi \quad \Longrightarrow \quad \vartheta_\rho^\pm \ll \vartheta, \quad \frac{\mathrm{d}\vartheta_\rho^-}{\mathrm{d}\vartheta}(x, y) = u(x), \quad \frac{\mathrm{d}\vartheta_\rho^+}{\mathrm{d}\vartheta}(x, y) = u(y). \tag{4.19}$$

We thus obtain that (4.17), (4.11) and (4.14) can be equivalently written as

$$\nu_\rho = \alpha[\vartheta_\rho^-, \vartheta_\rho^+ | \vartheta], \quad \mathscr{R}(\rho, j) = \frac{1}{2} \mathscr{F}_\Psi(2j | \nu_\rho), \tag{4.20}$$

where $\alpha[\vartheta_\rho^-, \vartheta_\rho^+ | \vartheta]$ stands for $\alpha[(\vartheta_\rho^-, \vartheta_\rho^+) | \vartheta]$, and the functional $\mathscr{F}_\psi(\cdot | \cdot)$ is from (2.11), and also

$$\mathscr{R}(\rho, j) = \frac{1}{2} \mathscr{F}_\Upsilon(\vartheta_\rho^-, \vartheta_\rho^+, 2j | \vartheta), \tag{4.21}$$

again writing for shorter notation $\mathscr{F}_\Upsilon(\vartheta_\rho^-, \vartheta_\rho^+, 2j | \vartheta)$ in place of $\mathscr{F}_\Upsilon((\vartheta_\rho^-, \vartheta_\rho^+, 2j) | \vartheta)$.

Therefore we can use the same expressions (4.20) and (4.21) to extend the functional $\mathscr{R}$ to measures ρ and j that need not be absolutely continuous with respect to π and ϑ; the next lemma shows that they provide equivalent characterizations. We introduce the functions $u^\pm : E \to \mathbb{R}$, adopting the notation

$$u^- := u \circ \mathsf{x} \quad \text{and} \quad u^+ := u \circ \mathsf{y},$$
$$\text{or equivalently} \quad u^-(x, y) := u(x), \quad u^+(x, y) := u(y). \tag{4.22}$$

(Recall that x and y denote the coordinate maps from E to V).

Lemma 4.8 *For every $\rho \in \mathcal{M}^+(V)$ and $\boldsymbol{j} \in \mathcal{M}(E)$ we have*

$$\mathscr{F}_\Upsilon(\boldsymbol{\vartheta}_\rho^-, \boldsymbol{\vartheta}_\rho^+, 2\boldsymbol{j}|\boldsymbol{\vartheta}) = \mathscr{F}_\Psi(2\boldsymbol{j}|\boldsymbol{v}_\rho). \tag{4.23}$$

If $\rho = \rho^a + \rho^\perp$ and $\boldsymbol{j} = \boldsymbol{j}^a + \boldsymbol{j}^\perp$ are the Lebesgue decompositions of ρ and $\boldsymbol{j}$ with respect to π and $\boldsymbol{\vartheta}$, respectively, we have

$$\mathscr{F}_\Upsilon(\boldsymbol{\vartheta}_\rho^-, \boldsymbol{\vartheta}_\rho^+, 2\boldsymbol{j}|\boldsymbol{\vartheta}) = \mathscr{F}_\Upsilon(\boldsymbol{\vartheta}_{\rho^a}^-, \boldsymbol{\vartheta}_{\rho^a}^+, 2\boldsymbol{j}^a|\boldsymbol{\vartheta}) + \mathscr{F}_{\Upsilon^\infty}(\boldsymbol{\vartheta}_{\rho^\perp}^-, \boldsymbol{\vartheta}_{\rho^\perp}^+, 2\boldsymbol{j}^\perp). \tag{4.24}$$

Proof Let us consider the Lebesgue decomposition $\rho = \rho^a + \rho^\perp$, $\rho^a = u\pi$, and a corresponding partition of V in two disjoint Borel sets R, P such that $\rho^a = \rho \, \llcorner \, R$, $\rho^\perp = \rho \, \llcorner \, P$ and $\pi(P) = 0$, which yields

$$\boldsymbol{\vartheta}_\rho^\pm = \boldsymbol{\vartheta}_{\rho^a}^\pm + \boldsymbol{\vartheta}_{\rho^\perp}^\pm, \quad \boldsymbol{\vartheta}_{\rho^a}^\pm \ll \boldsymbol{\vartheta}, \quad \boldsymbol{\vartheta}_{\rho^\perp}^-:=\boldsymbol{\vartheta}_\rho^- \, \llcorner \, P \times V, \quad \boldsymbol{\vartheta}_{\rho^\perp}^+:=\boldsymbol{\vartheta}_\rho^+ \, \llcorner \, V \times P. \tag{4.25}$$

Since $\boldsymbol{\vartheta}(P \times V) = \boldsymbol{\vartheta}(V \times P) \le \|\kappa_V\|_\infty \pi(P) = 0$, $\boldsymbol{\vartheta}_{\rho^\perp}^\pm$ are singular with respect to $\boldsymbol{\vartheta}$.

Let us also consider the Lebesgue decomposition $\boldsymbol{j} = \boldsymbol{j}^a + \boldsymbol{j}^\perp$ of $\boldsymbol{j}$ with respect to $\boldsymbol{\vartheta}$. We can select a measure $\varsigma \in \mathcal{M}^+(E)$ such that $\boldsymbol{\vartheta}_{\rho^\perp}^\pm = z^\pm \varsigma \ll \varsigma$, $\boldsymbol{j}^\perp \ll \varsigma$ and $\varsigma \perp \boldsymbol{\vartheta}$, obtaining

$$\boldsymbol{v}_\rho = \alpha[\boldsymbol{\vartheta}_\rho^-, \boldsymbol{\vartheta}_\rho^+|\boldsymbol{\vartheta}] = \boldsymbol{v}_\rho^1 + \boldsymbol{v}_\rho^2, \quad \boldsymbol{v}_\rho^1:=\alpha(u^-, u^+)\boldsymbol{\vartheta}, \quad \boldsymbol{v}_\rho^2:=\alpha^\infty(z^-, z^+)\varsigma. \tag{4.26}$$

Since $\boldsymbol{j} \ll \boldsymbol{\vartheta} + \varsigma$, we can decompose

$$2\boldsymbol{j} = w\boldsymbol{\vartheta} + w'\varsigma, \tag{4.27}$$

and by the additivity property (2.17) we obtain

$$\begin{aligned}
\mathscr{F}_\Psi(2\boldsymbol{j}|\boldsymbol{v}_\rho) &= \mathscr{F}_{\hat{\Psi}}(2\boldsymbol{j}, \boldsymbol{v}_\rho) = \mathscr{F}_{\hat{\Psi}}(w\boldsymbol{\vartheta}, \boldsymbol{v}_\rho^1) + \mathscr{F}_{\hat{\Psi}}(w'\varsigma, \boldsymbol{v}_\rho^2) \\
&\overset{(*)}{=} \iint_E \Upsilon(u(x), u(y), w(x, y)) \, \boldsymbol{\vartheta}(\mathrm{d}x, \mathrm{d}y) \\
&+ \iint_E \Upsilon^\infty(z^-(x, y), z^+(x, y), w'(x, y)) \, \varsigma(\mathrm{d}x, \mathrm{d}y) \\
&= \mathscr{F}_\Upsilon(\boldsymbol{\vartheta}_{\rho^a}^-, \boldsymbol{\vartheta}_{\rho^a}^+, 2\boldsymbol{j}^a|\boldsymbol{\vartheta}) + \mathscr{F}_{\Upsilon^\infty}(\boldsymbol{\vartheta}_{\rho^\perp}^-, \boldsymbol{\vartheta}_{\rho^\perp}^+, 2\boldsymbol{j}^\perp) = \mathscr{F}_\Upsilon(\boldsymbol{\vartheta}_\rho^-, \boldsymbol{\vartheta}_\rho^+, 2\boldsymbol{j}|\boldsymbol{\vartheta}).
\end{aligned} \tag{4.28}$$

Indeed, identity (*) follows from the fact that, since $\hat{\Psi}$ is 1-homogeneous,

$$\mathscr{F}_{\hat{\Psi}}(w\boldsymbol{\vartheta}, \boldsymbol{v}_\rho^1) = \iint_E \hat{\Psi}\left(\frac{\mathrm{d}(w\boldsymbol{\vartheta}, \boldsymbol{v}_\rho^1)}{\mathrm{d}\gamma}\right) \mathrm{d}\gamma$$

for every $\gamma \in \mathcal{M}^+(E)$ such that $w\boldsymbol{\vartheta} \ll \gamma$ and $\boldsymbol{v}_\rho^1 \ll \gamma$, cf. (2.13). Then, it suffices to observe that $w\boldsymbol{\vartheta} \ll \boldsymbol{\vartheta}$ and $\boldsymbol{v}_\rho^1 \ll \boldsymbol{\vartheta}$ with $\frac{\mathrm{d}\boldsymbol{v}_\rho^1}{\mathrm{d}\boldsymbol{\vartheta}} = \alpha(u^-, u^+)$. The same argument applies to $\mathscr{F}_{\hat{\Psi}}(w'\varsigma, \boldsymbol{v}_\rho^2)$, cf. also Lemma 2.3(3). □

Definition 4.9 The *dissipation potential* $\mathscr{R} : \mathcal{M}^+(V) \times \mathcal{M}(E) \to [0, +\infty]$ is defined by

$$\mathscr{R}(\rho, \boldsymbol{j}):=\frac{1}{2}\mathscr{F}_\Upsilon(\boldsymbol{\vartheta}_\rho^-, \boldsymbol{\vartheta}_\rho^+, 2\boldsymbol{j}|\boldsymbol{\vartheta}) = \frac{1}{2}\mathscr{F}_\Psi(2\boldsymbol{j}|\boldsymbol{v}_\rho). \tag{4.29}$$

where $\boldsymbol{\vartheta}_\rho^\pm$ are defined by (4.18). If α is 1-homogeneous, then $\mathscr{R}(\rho, \boldsymbol{j})$ is independent of $\boldsymbol{\vartheta}$.

Lemma 4.10 *Let* $\rho = \rho^a + \rho^\perp \in \mathcal{M}^+(V)$ *and* $\boldsymbol{j} = \boldsymbol{j}^a + \boldsymbol{j}^\perp \in \mathcal{M}(E)$, *with* $\rho^a = u\pi$, $2\boldsymbol{j}^a = w\boldsymbol{\vartheta}$, *and* $\rho^\perp$, $\boldsymbol{j}^\perp$ *as in Lemma* 4.8, *satisfy* $\mathcal{R}(\rho, \boldsymbol{j}) < +\infty$, *and let* $P \in \mathcal{B}(V)$ *be a* π-*negligible set such that* $\rho^\perp = \rho \llcorner P$.

(1) We have $|\boldsymbol{j}|(P \times (V \setminus P)) = |\boldsymbol{j}|((V \setminus P) \times P) = 0$, $\boldsymbol{j}^\perp = \boldsymbol{j} \llcorner (P \times P)$, *and*

$$\mathcal{R}(\rho, \boldsymbol{j}) = \mathcal{R}(\rho^a, \boldsymbol{j}^a) + \frac{1}{2}\mathcal{F}_{\Upsilon^\infty}(\boldsymbol{\vartheta}^-_{\rho^\perp}, \boldsymbol{\vartheta}^+_{\rho^\perp}, 2\boldsymbol{j}^\perp). \tag{4.30}$$

In particular, if α *is* 1-*homogeneous we have the decomposition*

$$\mathcal{R}(\rho, \boldsymbol{j}) = \mathcal{R}(\rho^a, \boldsymbol{j}^a) + \mathcal{R}(\rho^\perp, \boldsymbol{j}^\perp). \tag{4.31}$$

(2) If $\rho \ll \pi$ *or* α *is sub-linear, i.e.* $\alpha^\infty \equiv 0$, *or* $\kappa(x, \cdot) \ll \pi$ *for every* $x \in V$, *then* $\boldsymbol{j} \ll \boldsymbol{\vartheta}$ *and* $\boldsymbol{j}^\perp \equiv 0$. *In any of these three cases,* $\mathcal{R}(\rho, \boldsymbol{j}) = \mathcal{R}(\rho^a, \boldsymbol{j})$, *and setting* E' *as in* (4.10) *we have* $w = 0$ $\boldsymbol{\vartheta}$-*a.e. on* $E \setminus E'$, *and* (4.11) *holds.*

(3) Furthermore, $\mathcal{R}$ *is convex and lower semicontinuous with respect to setwise convergence in* $(\rho, \boldsymbol{j})$. *If* κ *satisfies the weak Feller property, then* $\mathcal{R}$ *is also lower semicontinuous with respect to weak convergence in duality with continuous bounded functions.*

Proof *(1)* Equation (4.30) is an immediate consequence of (4.24).

To prove the properties of $\boldsymbol{j}$, set $R = V \setminus P$ for convenience. By using the decompositions $\boldsymbol{j} = w\boldsymbol{\vartheta} + w'\boldsymbol{\varsigma}$ and $\boldsymbol{\vartheta}^\pm_\rho = \boldsymbol{\vartheta}^\pm_{\rho^a} + \boldsymbol{\vartheta}^\pm_{\rho^\perp} = \boldsymbol{\vartheta}^\pm_{\rho^a} + z^\pm\boldsymbol{\varsigma}$ introduced in the proof of the previous Lemma, the definition (4.25) implies that $\boldsymbol{\vartheta}^+_{\rho^\perp}(P \times R) = 0$, so that $z^+ = 0$ $\boldsymbol{\varsigma}$-a.e. in $P \times R$; analogously $z^- = 0$ $\boldsymbol{\varsigma}$-a.e. in $R \times P$. By (3.13) we find that $\alpha^\infty(z^-, z^+) = 0$, $\boldsymbol{\varsigma}$-a.e. in $(P \times R) \cup (R \times P)$ and therefore $w' = 0$ as well, since $\Upsilon^\infty(z^-, z^+, w') < +\infty$ $\boldsymbol{\varsigma}$-a.e (see (4.28)). We eventually deduce that $\boldsymbol{j}^\perp = \boldsymbol{j} \llcorner P \times P$.

(2) When $\rho \ll \pi$ we can choose $P = \emptyset$ so that $\boldsymbol{j}^\perp = \boldsymbol{j} \llcorner P = 0$. When α is sub-linear then $\boldsymbol{v}_\rho \ll \boldsymbol{\vartheta}$ so that $\boldsymbol{j} \ll \boldsymbol{\vartheta}$ since Ψ is superlinear.

If $\kappa(x, \cdot) \ll \pi$ for every $x \in V$, then $\mathsf{y}_\sharp\boldsymbol{\vartheta}^-_{\rho^\perp} \ll \pi$ and $\mathsf{x}_\sharp\boldsymbol{\vartheta}^+_{\rho^\perp} \ll \pi$, so that $\boldsymbol{\vartheta}^\pm_{\rho^\perp}(P \times P) = 0$, since P is π-negligible. We deduce that $\boldsymbol{j}^\perp(P \times P) = 0$ as well.

(3) The convexity of $\mathcal{R}$ follows by the convexity of the functional $\mathcal{F}_\Upsilon$. The lower semicontinuity follows by combining Lemma 2.4 with Lemma 2.3. □

Corollary 4.11 *Let* $\pi_1, \pi_2 \in \mathcal{M}^+(V)$ *be mutually singular measures satisfying the detailed balance condition with respect to* κ, *and let* $\boldsymbol{\vartheta}_i = \boldsymbol{\kappa}_{\pi_i}$ *be the corresponding symmetric measures in* $\mathcal{M}^+(E)$ *(see Sect.* 2.4*). For every pair* $(\rho, \boldsymbol{j})$ *with* $\rho = \rho_1 + \rho_2$, $\boldsymbol{j} = \boldsymbol{j}_1 + \boldsymbol{j}_2$ *for* $\rho_i \ll \pi_i$ *and* $\boldsymbol{j}_i \ll \boldsymbol{\vartheta}_i$, *we have*

$$\mathcal{R}(\rho, \boldsymbol{j}) = \mathcal{R}_1(\rho_1, \boldsymbol{j}_1) + \mathcal{R}_2(\rho_2, \boldsymbol{j}_2), \tag{4.32}$$

where $\mathcal{R}_i$ *is the dissipation functional induced by* $\boldsymbol{\vartheta}_i$. *When* α *is* 1-*homogeneous,* $\mathcal{R}_i = \mathcal{R}$.

4.3 Curves with finite $\mathcal{R}$-action

In this section, we study the properties of curves with finite $\mathcal{R}$-action, i.e., elements of

$$\mathcal{A}(a, b) := \left\{(\rho, \boldsymbol{j}) \in \mathcal{CE}(a, b) : \int_a^b \mathcal{R}(\rho_t, \boldsymbol{j}_t)\, \mathrm{d}t < +\infty\right\}. \tag{4.33}$$

The finiteness of the $\mathcal{R}$-action leads to the following remarkable property: A curve $(\rho, \boldsymbol{j})$ with finite $\mathcal{R}$-action can be separated into two mutually singular curves $(\rho^a, \boldsymbol{j}^a)$, $(\rho^\perp, \boldsymbol{j}^\perp) \in$

$\mathcal{A}(a, b)$ that evolve independently, and contribute independently to $\mathcal{R}$. Consequently, finite $\mathcal{R}$-action preserves π-absolute continuity of ρ: if $\rho_t \ll \pi$ at any t, then $\rho_t \ll \pi$ at all t. These properties and others are proved in Theorem 4.13 and Corollary 4.14 below.

Remark 4.12 If $(\rho, \boldsymbol{j}) \in \mathcal{A}(a, b)$ then the 'skew-symmetrization' $\boldsymbol{j}^{\flat} = (\boldsymbol{j} - \mathsf{s}_{\#}\boldsymbol{j})/2$ of $\boldsymbol{j}$ gives rise to a pair $(\rho, \boldsymbol{j}^{\flat}) \in \mathcal{A}(a, b)$ as well, and it has lower $\mathcal{R}$-action:

$$\int_a^b \mathcal{R}(\rho_t, \boldsymbol{j}_t^{\flat}) \, \mathrm{d}t \le \int_a^b \mathcal{R}(\rho_t, \boldsymbol{j}_t) \, \mathrm{d}t.$$

This follows from the convexity of $w \mapsto \Upsilon(u_1, u_2, w)$, the symmetry of $(u_1, u_2) \mapsto \Upsilon(u_1, u_2, w)$, and the invariance of the continuity equation (4.3) under the 'skew-symmetrization' $\boldsymbol{j} \mapsto \boldsymbol{j}^{\flat}$ (cf. also the calculations in the proof of Corollary 4.20).

As a result, we can often assume without loss of generality that a flux $\boldsymbol{j}$ is skew-symmetric, i.e. that $\mathsf{s}_{\#}\boldsymbol{j} = -\boldsymbol{j}$.

Theorem 4.13 *Let $(\rho, \boldsymbol{j}) \in \mathcal{A}(a, b)$ and let us consider the Lebesgue decompositions $\rho_t = \rho_t^a + \rho_t^{\perp}$ and $\boldsymbol{j}_t = \boldsymbol{j}_t^a + \boldsymbol{j}_t^{\perp}$ of ρ_t with respect to π and of $\boldsymbol{j}_t$ with respect to $\boldsymbol{\vartheta}$.*

(1) We have $(\rho^a, \boldsymbol{j}^a) \in \mathcal{A}(a, b)$ with

$$\int_a^b \mathcal{R}(\rho_t^a, \boldsymbol{j}_t^a) \, \mathrm{d}t \le \int_a^b \mathcal{R}(\rho_t, \boldsymbol{j}_t) \, \mathrm{d}t. \tag{4.34}$$

In particular $t \mapsto \rho_t^a(V)$ and $t \mapsto \rho_t^{\perp}(V)$ are constant.

(2) If α is 1-homogeneous then also $(\rho^{\perp}, \boldsymbol{j}^{\perp}) \in \mathcal{A}(a, b)$ and

$$\int_a^b \mathcal{R}(\rho_t^a, \boldsymbol{j}_t^a) \, \mathrm{d}t + \int_a^b \mathcal{R}(\rho_t^{\perp}, \boldsymbol{j}_t^{\perp}) \, \mathrm{d}t = \int_a^b \mathcal{R}(\rho_t, \boldsymbol{j}_t) \, \mathrm{d}t. \tag{4.35}$$

(3) If α is sub-linear or $\kappa(x, \cdot) \ll \pi$ for every $x \in V$, then $\rho_t^{\perp}$ is constant in $[a, b]$ and $\boldsymbol{j}^{\perp} \equiv 0$.

Proof *(1)* Let $\gamma \in \mathcal{M}^+(V)$ be a dominating measure for the curve ρ according to Corollary 4.3 and let us denote by $\gamma = \gamma^a + \gamma^{\perp}$ the Lebesgue decomposition of γ with respect to π; we also denote by $P \in \mathcal{B}(V)$ a π-negligible Borel set such that $\gamma^{\perp} = \gamma \llcorner P$. Setting $R := V \setminus P$, since $\rho_t \ll \gamma$ we thus obtain $\rho_t^a = \rho_t \llcorner R$, $\rho_t^{\perp} = \rho_t \llcorner P$. By Lemma 4.10 for λ-a.e. $t \in (a, b)$ we obtain $\boldsymbol{j}_t^{\perp} = \boldsymbol{j} \llcorner (P \times P)$ and $\boldsymbol{j}_t^a = \boldsymbol{j} \llcorner (R \times R)$ with $|\boldsymbol{j}_t|(R \times P) = |\boldsymbol{j}_t|(P \times R) = 0$. For every function $\varphi \in \mathrm{B}_b$ we have $\overline{\nabla}(\varphi \chi_R) \equiv 0$ on $P \times P$ so that we get

$$\int_V \varphi \, \mathrm{d}\rho_{t_2}^a - \int_V \varphi \, \mathrm{d}\rho_{t_1}^a = \int_R \varphi \, \mathrm{d}\rho_{t_2} - \int_R \varphi \, \mathrm{d}\rho_{t_1} = \int_{t_1}^{t_2} \iint_E \overline{\nabla}(\varphi \chi_R) \, \mathrm{d}(\boldsymbol{j}_t^a + \boldsymbol{j}_t^{\perp}) \, \mathrm{d}t$$

$$= \int_{t_1}^{t_2} \iint_{R \times R} \overline{\nabla}(\varphi \chi_R) \, \mathrm{d}\boldsymbol{j}_t^a \, \mathrm{d}t = \int_{t_1}^{t_2} \iint_E \overline{\nabla}\varphi \, \mathrm{d}\boldsymbol{j}_t^a \, \mathrm{d}t,$$

showing that $(\rho^a, \boldsymbol{j}^a)$ belongs to $\mathcal{CE}(a, b)$. Estimate (4.34) follows by (4.30). From Lemma 4.4 we deduce that $\rho_t^a(V)$ and $\rho_t^{\perp}(V)$ are constant.

(2) This follows by the linearity of the continuity equation and (4.31).

(3) If α is sub-linear or $\kappa(x, \cdot) \ll \pi$ for every $x \in V$, then Lemma 4.10 shows that $\boldsymbol{j}^{\perp} \equiv 0$. Since by linearity $(\rho^{\perp}, \boldsymbol{j}^{\perp}) \in \mathcal{CE}(a, b)$, we deduce that $\rho_t^{\perp}$ is constant. $\square$

Corollary 4.14 *Let $(\rho, \boldsymbol{j}) \in \mathcal{A}(a, b)$. If there exists $t_0 \in [a, b]$ such that $\rho_{t_0} \ll \pi$, then we have $\rho_t \ll \pi$ for every $t \in [a, b]$, $\boldsymbol{j}^{\perp} \equiv 0$, and $\overline{\mathrm{div}}\, \boldsymbol{j}_t \ll \pi$ for λ-a.e. $t \in (a, b)$. In particular,*

there exists an absolutely continuous and a.e. differentiable map $u : [a, b] \to L^1(V, \pi)$ and a map $w \in L^1(E, \lambda \otimes \vartheta)$ such that

$$2j_\lambda = w\lambda \otimes \vartheta, \quad \partial_t u_t(x) = \frac{1}{2} \int_V \left(w_t(y, x) - w_t(x, y) \right) \kappa(x, \mathrm{d}y) \quad \textit{for a.e. } t \in (a, b). \tag{4.36}$$

Moreover there exists a measurable map $\xi : (a, b) \times E \to \mathbb{R}$ such that $w = \xi\alpha(u^-, u^+)$ $\lambda \otimes \vartheta$-a.e. and

$$\mathscr{R}(\rho_t, j_t) = \frac{1}{2} \iint_E \Psi(\xi_t(x, y))\alpha(u_t(x), u_t(y)) \, \vartheta(\mathrm{d}x, \mathrm{d}y) \quad \textit{for a.e. } t \in (a, b). \tag{4.37}$$

If w is skew-symmetric, then ξ is skew-symmetric as well and (4.36) *reads as*

$$\partial_t u_t(x) = \int_V w_t(y, x) \, \kappa(x, \mathrm{d}y) = \int_V \xi_t(y, x)\alpha(u_t(x), u_t(y)) \, \kappa(x, \mathrm{d}y) \quad \textit{a.e. in } (a, b). \tag{4.38}$$

Remark 4.15 Relations (4.36) and (4.38) hold both in the sense of a.e. differentiability of maps with values in $L^1(V, \pi)$ and pointwise a.e. with respect to $x \in V$: more precisely, there exists a set $U \subset V$ of full π-measure such that for every $x \in U$ the map $t \mapsto u_t(x)$ is absolutely continuous and equations (4.36) and (4.38) hold for every $x \in U$, a.e. with respect to $t \in (0, T)$.

Proof The first part of the statement is an immediate consequence of Theorem 4.13, which yields $\rho_t^\perp(V) = 0$ for every $t \in [a, b]$. We can thus write $2j = w(\lambda \otimes \vartheta)$ for some measurable map $w : (a, b) \times E \to \mathbb{R}$. Moreover $\overline{\mathrm{div}}\, j \ll \lambda \otimes \pi$, since $\mathsf{s}_\sharp j \ll \mathsf{s}_\sharp(\lambda \otimes \vartheta) = \lambda \otimes \vartheta$, and therefore

$$2j^\flat = j - \mathsf{s}_\sharp j \ll \lambda \otimes \vartheta \quad \Longrightarrow \quad \overline{\mathrm{div}}\, j = \mathsf{x}_\sharp(2j^\flat) \ll \mathsf{x}_\sharp(\lambda \otimes \vartheta) \ll \lambda \otimes \pi. \tag{4.39}$$

Setting $z_t = \mathrm{d}(\overline{\mathrm{div}}\, j_t)/\mathrm{d}\pi$ we get for a.e. $t \in (a, b)$

$$\partial_t u_t = -z_t,$$

$$-2 \int_V \varphi \, z_t \, \mathrm{d}\pi = \iint_E (\varphi(y) - \varphi(x))w_t(x, y)\vartheta(\mathrm{d}x, \mathrm{d}y)$$

$$= \iint_E \varphi(x)(w_t(y, x) - w_t(x, y))\vartheta(\mathrm{d}x, \mathrm{d}y)$$

$$= \int_V \varphi(x)\left(\int_V (w_t(y, x) - w_t(x, y))\kappa(x, \mathrm{d}y) \right)\pi(\mathrm{d}x),$$

The existence of ξ and formula (4.37) follow from Lemma 4.10(2). $\qquad\square$

4.4 Chain rule for convex entropies

Let us now consider a continuous convex function $\beta : \mathbb{R}_+ \to \mathbb{R}_+$ that is differentiable in $(0, +\infty)$. The main choice for β will be the function ϕ that appears in the definition of the driving functional $\mathscr{E}$ (see Assumption ($\mathscr{E}\phi$)), and the example of the Boltzmann-Shannon entropy function (3.10) illustrates why we only assume differentiability away from zero.

By setting $\beta'(0) = \lim_{r \downarrow 0} \beta'(r) \in [-\infty, +\infty)$, we define the function $A_\beta : \mathbb{R}_+ \times \mathbb{R}_+ \to [-\infty, +\infty]$ by

$$A_\beta(u, v) := \begin{cases} \beta'(v) - \beta'(u) & \text{if } u, v \in \mathbb{R}_+ \times \mathbb{R}_+ \setminus \{(0, 0)\}, \\ 0 & \text{if } u = v = 0. \end{cases} \tag{4.40}$$

Note that A_β is continuous (with extended real values) in $\mathbb{R}_+ \times \mathbb{R}_+ \setminus \{(0, 0)\}$ and is finite and continuous whenever $\beta'(0) > -\infty$. When $\beta'(0) = -\infty$ we have $A_\beta(0, v) = -A_\beta(u, 0) = +\infty$ for every $u, v > 0$.

In the following we will adopt the convention

$$|\pm\infty| = +\infty, \quad a \cdot (+\infty) := \begin{cases} +\infty & \text{if } a > 0, \\ 0 & \text{if } a = 0, \\ -\infty & \text{if } a < 0 \end{cases} \quad a \cdot (-\infty) = -a \cdot (+\infty), \quad (4.41)$$

for every $a \in [-\infty, +\infty]$ and, using this convention, we define the extended valued function $B_\beta : \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R} \to [-\infty, +\infty]$ by

$$B_\beta(u, v, w) := A_\beta(u, v)w. \quad (4.42)$$

We want to study the differentiability properties of the functional $\mathscr{F}_\beta(\cdot|\pi)$ along solutions $(\rho, j) \in \mathcal{CE}(I)$ of the continuity equation. Note that if β is superlinear and $\mathscr{F}_\beta$ is finite at a time $t_0 \in I$, then Corollary 4.14 shows that $\rho_t \ll \pi$ for every $t \in I$. If β has linear growth then

$$\mathscr{F}_\beta(\rho_t|\pi) = \int_V \beta(u_t) \, d\pi + \beta^\infty(1)\rho^\perp(V), \quad \rho_t = u_t\pi + \rho_t^\perp, \quad (4.43)$$

where we have used that $t \mapsto \rho_t^\perp(V)$ is constant. Thus, we are reduced to studying $\mathscr{F}_\beta$ along (ρ^a, j^a), which is still a solution of the continuity equation. The absolute continuity property of ρ_t with respect to π is therefore quite a natural assumption in the next result.

Theorem 4.16 (Chain rule I) *Let $(\rho, j) \in \mathcal{A}(a, b)$ with $\rho_t = u_t\pi \ll \pi$ and let $2j^\flat = j - s_\sharp j = w^\flat \lambda \otimes \vartheta$ as in Corollary 4.14 satisfy*

$$\int_V \beta(u_a) \, d\pi < +\infty, \quad \int_a^b \iint_E \left(B_\beta(u_t(x), u_t(y), w_t^\flat(x, y)) \right)_+ \vartheta(dx, dy) \, dt < +\infty \quad (4.44)$$

Then the map $t \mapsto \int_V \beta(u_t) \, d\pi$ is absolutely continuous in $[a, b]$, the map $B_\beta(u^-, u^+, w^\flat)$ is $\lambda \otimes \vartheta$-integrable and

$$\frac{d}{dt} \int_V \beta(u_t) \, d\pi = \frac{1}{2} \iint_E B_\beta(u_t(x), u_t(y), w_t^\flat(x, y))\vartheta(dx, dy) \quad \text{for a.e. } t \in (a, b). \quad (4.45)$$

Remark 4.17 At first sight condition (4.44) on the positive part of B_β is remarkable: we only require the positive part of B_β to be integrable, but in the assertion we obtain integrability of the negative part as well. This integrability arises from the combination of the upper bound on $\int_V \beta(u_a) \, d\pi$ in (4.44) with the lower bound $\beta \geq 0$.

Proof *Step 1: Chain rule for an approximation.* Define for $k \in \mathbb{N}$ an approximation β_k of β as follows: Let $\beta_k'(\sigma) := \max\{-k, \min\{\beta'(\sigma), k\}\}$ be the truncation of β' to the interval $[-k, k]$. Due to the assumptions on β, we may assume that β achieves a minimum at the point $s_0 \in [0, +\infty)$. Now set $\beta_k(s) := \beta(s_0) + \int_{s_0}^s \beta_k'(\sigma) \, d\sigma$. Note that β_k is differentiable and globally Lipschitz, and converges monotonically to $\beta(s)$ for all $s \geq 0$ as $k \to \infty$.

For each $k \in \mathbb{N}$ and $t \in [a, b]$ we define

$$S_k(t) := \int_V \beta_k(u_t) \, d\pi, \quad S(t) := \int_V \beta(u_t) \, d\pi.$$

By convexity and Lipschitz continuity of β_k, we have that

$$\beta_k(u_t(x)) - \beta_k(u_s(x)) \leq \beta_k'(u_t(x))(u_t(x) - u_s(x)) \leq k|u_t(x) - u_s(x)| .$$

Hence, we deduce by Corollary 4.14 that for every $a \leq s < t \leq b$

$$S_k(t) - S_k(s) = \int_V \big[\beta_k(u_t(x)) - \beta_k(u_s(x))\big]\pi(\mathrm{d}x)$$

$$\leq k\|u_t - u_s\|_{L^1(V;\pi)} \leq k \int_s^t \|\partial_r u_r\|_{L^1(V;\pi)} \,\mathrm{d}r.$$

We conclude that the function $t \mapsto S_k(t)$ is absolutely continuous. Let us pick a point $t \in (a, b)$ of differentiability for $t \mapsto S_k(t)$: it easy to check that

$$\frac{\mathrm{d}}{\mathrm{d}t} S_k(t) = \int_V \beta_k'(u_t)\,\partial_t u_t\,\mathrm{d}\pi = \frac{1}{2}\iint_E \overline{\nabla}\beta_k'(u_t)w_t^\flat\,\mathrm{d}\boldsymbol{\vartheta} ,$$

which by integrating over time yields

$$S_k(t) - S_k(s) = \frac{1}{2}\int_s^t \iint_E \overline{\nabla}\beta_k'(u_r)w_r^\flat\,\mathrm{d}\boldsymbol{\vartheta}\,\mathrm{d}r \qquad \text{for all } a \leq s \leq t \leq b. \tag{4.46}$$

Step 2: The limit $k \to \infty$ Since $0 \leq \beta_k'' \leq \beta''$ we have

$$0 \leq A_{\beta_k}(u, v) = \beta_k'(v) - \beta_k'(u) \leq \beta'(v) - \beta'(u) = A_\beta(u, v) \quad \text{whenever } 0 \leq u \leq v \tag{4.47}$$

and

$$|\beta_k'(v) - \beta_k'(u)| \leq |A_\beta(u, v)| \quad \text{for every } u, v \in \mathbb{R}_+. \tag{4.48}$$

We can thus estimate the right-hand side in (4.46)

$$(B_k)_+ = \big(\overline{\nabla}\beta_k'(u)\,w^\flat\big)_+ \leq \big(A_\beta(u^-, u^+)w^\flat\big)_+ = B_+ \tag{4.49}$$

where we have used the short-hand notation

$$B_k(r, x, y) = B_{\beta_k}(u_r(x), u_r(y), w_r^\flat(x, y)), \quad B(r, x, y) := B_\beta(u_r(x), u_r(y), w_r^\flat(x, y)). \tag{4.50}$$

Assumption (4.44) implies that the right-hand side in (4.49) is an element of $L^1([a, b] \times E; \lambda \otimes \boldsymbol{\vartheta})$, so that in particular $B_+ \in \mathbb{R}$ for $(\lambda \otimes \boldsymbol{\vartheta})$-a.e. (t, x, y).

Moreover, (4.46) yields

$$\int_a^b \iint_E (B_k)_-\,\mathrm{d}\boldsymbol{\vartheta}_\lambda = \int_a^b \iint_E (B_k)_+\,\mathrm{d}\boldsymbol{\vartheta}_\lambda + S_k(a) - S_k(b)$$

$$\leq \int_a^b \iint_E (B)_+\,\mathrm{d}\boldsymbol{\vartheta}_\lambda + S(a) < +\infty. \tag{4.51}$$

Note that the sequence $k \mapsto (B_k)_-$ is definitely 0 or is monotonically increasing to B_-. Beppo Levi's Monotone Convergence Theorem and the uniform estimate (4.51) then yields that $B_- \in L^1((a, b) \times E, \lambda \otimes \boldsymbol{\vartheta})$, thus showing that $B_\beta(u^-, u^+, w^\flat)$ is $(\lambda \otimes \boldsymbol{\vartheta})$-integrable as well.

We can thus pass to the limit in (4.46) as $k \to +\infty$ and we have

$$\lim_{k \to +\infty} \overline{\nabla}\beta_k'(u)\,w^\flat = B \quad \lambda \otimes \boldsymbol{\vartheta}\text{-a.e. in } (a, b) \times E. \tag{4.52}$$

The identity (4.52) is obvious if $\beta'(0)$ is finite, and if $\beta'(0) = -\infty$ then it follows by the upper bound (4.49) and the fact that the right-hand side of (4.49) is finite almost everywhere.

The Dominated Convergence Theorem then implies that

$$\int_s^t \iint_E \overline{\nabla}\beta_k'(u_r)\, w_r^\flat \, d\vartheta \, dr \quad \longrightarrow \quad \int_s^t \iint_E B \, d\vartheta \, dr \quad \text{as } k \to \infty.$$

By the monotone convergence theorem $S(t) = \lim_{k \to +\infty} S_k(t) \in [0, +\infty]$ for all $t \in [a, b]$ and the limit is finite for $t = 0$. For all $t \in [a, b]$, therefore,

$$S(t) = S(a) + \frac{1}{2}\int_a^t \iint_E B \, d\vartheta \, dr,$$

which shows that S is absolutely continuous and (4.45) holds.   □

We now introduce three functions associated with the (general) continuous convex function $\beta : \mathbb{R}_+ \to \mathbb{R}_+$, differentiable in $(0, +\infty)$, that we have considered so far, and whose main example will be the entropy density ϕ from (3.9). Recalling the definition (4.40), the convention (4.41), and setting $\Psi^*(\pm\infty):=+\infty$, let us now introduce the functions $D_\beta^+, D_\beta^-, D_\beta : \mathbb{R}_+^2 \to [0, +\infty]$

$$D_\beta^-(u, v):=\Psi^*(A_\beta(u, v))\alpha(u, v)$$

$$\begin{cases} \Psi^*(A_\beta(u, v))\alpha(u, v) & \text{if } \alpha(u, v) > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{4.53a}$$

$$D_\beta^+(u, v):= \begin{cases} \Psi^*(A_\beta(u, v))\alpha(u, v) & \text{if } \alpha(u, v) > 0, \\ 0 & \text{if } u = v = 0, \\ +\infty & \text{otherwise, i.e. if } \alpha(u, v) = 0,\ u \neq v, \end{cases} \tag{4.53b}$$

$$D_\beta(\cdot, \cdot):=\text{the lower semicontinuous envelope of } D_\beta^+ \text{ in } \mathbb{R}_+^2. \tag{4.53c}$$

The function D_ϕ corresponding to the choice $\beta = \phi$ shall feature in the (rigorous) definition of the *Fisher information* functional $\mathscr{D}$, cf. (5.1) ahead. Nonetheless, it is significant to introduce the functions D_ϕ^- and D_ϕ^+ as well, cf. Remarks 5.8 and 7.12 ahead.

Example 4.18 (**The functions** $D_\phi^\pm$ **and** D_ϕ **in the quadratic and in the** cosh **case**) In the two examples of the linear equation (1.2), with Boltzmann entropy function ϕ, and with quadratic and cosh-type potentials Ψ^* (see (1.17a) and (1.17b)), the functions $D_\phi^\pm$ and D_ϕ take the following forms:

(1) If $\Psi^*(s) = s^2/2$ and, accordingly, $\alpha(u, v) = (u - v)/(\log(u) - \log(v))$ for all $u, v > 0$ (with $\alpha(u, v) = 0$ otherwise), then

$$D_\phi^-(u, v) = \begin{cases} \frac{1}{2}(\log(u) - \log(v))(u - v) & \text{if } u,\ v > 0, \\ 0 & \text{if } u = 0 \text{ or } v = 0, \end{cases}$$

$$D_\phi(u, v) = D_\phi^+(u, v) = \begin{cases} \frac{1}{2}(\log(u) - \log(v))(u - v) & \text{if } u,\ v > 0, \\ 0 & \text{if } u = v = 0, \\ +\infty & \text{if } u = 0 \text{ and } v \neq 0,\ \text{or vice versa.} \end{cases}$$

For this example D_ϕ^+ and D_ϕ are convex, and all three functions are lower semicontinuous.

(2) For the case $\Psi^*(s) = 4\big(\cosh(s/2) - 1\big)$ and $\alpha(u, v) = \sqrt{uv}$ for all $u, v \geq 0$, one finds

$$D_\phi^-(u, v) = \begin{cases} 2\big(\sqrt{u} - \sqrt{v}\big)^2 & \text{if } u, v > 0, \\ 0 & \text{if } u = 0 \text{ or } v = 0, \end{cases}$$

$$D_\phi(u, v) = 2\big(\sqrt{u} - \sqrt{v}\big)^2 \quad \text{for all } u, v \geq 0,$$

$$D_\phi^+(u, v) = \begin{cases} 2\big(\sqrt{u} - \sqrt{v}\big)^2 & \text{if } u, v > 0 \text{ or } u = v = 0, \\ +\infty & \text{if } u = 0 \text{ and } v \neq 0, \text{ or vice versa.} \end{cases}$$

For this example, D_ϕ^+ and D_ϕ again are convex, but only D_ϕ^- and D_ϕ are lower semicontinuous.

Other examples of functions D_ϕ in the case of power means (1.33) with the Boltzmann entropy function ϕ are discussed in Appendix E.

We collect a number of general properties of D_β and $D_\beta^\pm$.

Lemma 4.19

(1) $D_\beta^- \leq D_\beta \leq D_\beta^+$;

(2) D_β^- *and* D_β *are lower semicontinuous;*

(3) For every $u, v \in \mathbb{R}_+$ *and* $w \in \mathbb{R}$ *we have*

$$\big|B_\beta(u, v, w)\big| \leq \Upsilon(u, v, w) + D_\beta^-(u, v). \tag{4.54}$$

(4) Moreover, when the right-hand side of (4.54) *is finite, then the equality*

$$-B_\beta(u, v, w) = \Upsilon(u, v, w) + D_\beta^-(u, v) \tag{4.55}$$

is equivalent to the condition

$$\alpha(u, v) = w = 0 \quad or \quad \Big[\alpha(u, v) > 0, \ A_\beta(u, v) \in \mathbb{R}, \ -w = (\Psi^*)'\big(A_\beta(u, v)\big)\alpha(u, v)\Big]. \tag{4.56}$$

Proof It is not difficult to check that D_β^- is lower semicontinuous: such a property is trivial where α vanishes, and in all the other cases it is sufficient to use the positivity and the continuity of Ψ^* in $[-\infty, +\infty]$, the continuity of A_β in $\mathbb{R}_+^2 \setminus \{(0, 0)\}$, and the continuity and the positivity of α. It is also obvious that $D_\beta^- \leq D_\beta^+$, and therefore $D_\beta^- \leq D_\beta \leq D_\beta^+$.

For the inequality (4.54), let us distinguish the various cases:

- If $w = 0$ or $u = v = 0$, then $B_\beta(u, v, w) = 0$ so that (4.54) is trivially satisfied. We can thus assume $w \neq 0$ and $u + v > 0$.
- When $\alpha(u, v) = 0$ then $\Upsilon(u, v, w) = +\infty$ so that (4.54) is trivially satisfied as well. We can thus assume $\alpha(u, v) > 0$.
- If $A_\beta(u, v) \in \{\pm\infty\}$ then $D_\beta^-(u, v) = +\infty$ and the right-hand side of (4.54) is infinite.
- It remains to consider the case when $A_\beta(u, v) \in \mathbb{R}$, $\alpha(u, v) > 0$ and $w \neq 0$. In this situation

$$\big|B(u, v, w)\big| = \big|A_\beta(u, v)w\big| = \Big|A_\beta(u, v)\frac{w}{\alpha(u, v)}\Big|\alpha(u, v)$$

$$\leq \Psi\Big(\frac{w}{\alpha(u, v)}\Big)\alpha(u, v) + \Psi^*\big(A_\beta(u, v)\big)\alpha(u, v)$$

$$= \Upsilon(u, v, w) + D_\beta^-(u, v). \tag{4.57}$$

This proves (4.54).

It is now easy to study the case of equality in (4.55), when the right-hand side of (4.54) and (4.55) is finite. This in particular implies that $\alpha(u, v) > 0$ and $A_\beta(u, v) \in \mathbb{R}$ or $\alpha(u, v) = 0$ and $w = 0$. In the former case, calculations similar to (4.57) show that $-w = (\Psi^*)'\big(A_\beta(u, v)\big)\alpha(u, v)$. In the latter case, $\alpha(u, v) = w = 0$ yields that $B_\beta(u, v, w) = 0$, $\Upsilon(u, v, w) = \hat{\Psi}(w, \alpha(u, v)) = \hat{\Psi}(0, 0) = 0$, and $D_\beta(u, v) = \Psi^*(A_\beta(u, v))\alpha(u, v) = 0$. $\qquad\square$

As a consequence of Lemma 4.19, we conclude a chain-rule inequality involving the smallest functional D_β^- and thus, a fortiori, the functional D_β which, for $\beta = \phi$, shall enter into the definition of the Fisher information $\mathscr{D}$.

Corollary 4.20 (Chain rule II) *Let $(\rho, \boldsymbol{j}) \in \mathcal{A}(a, b)$ with $\rho_t = u_t \pi \ll \pi$ and $2\boldsymbol{j}_\lambda = w(\lambda \otimes \boldsymbol{\vartheta})$ satisfy*

$$\int_V \beta(u_a)\,\mathrm{d}\pi < +\infty, \quad \int_a^b \iint_E D_\beta^-(u_t(x), u_t(y))\,\boldsymbol{\vartheta}(\mathrm{d}x, \mathrm{d}y)\mathrm{d}t < +\infty. \quad (4.58)$$

Then the map $t \mapsto \int_V \beta(u_t)\,\mathrm{d}\pi$ is absolutely continuous in $[a, b]$ and

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\int_V \beta(u_t)\,\mathrm{d}\pi\right| \le \mathscr{R}(\rho_t, \boldsymbol{j}_t) + \frac{1}{2}\iint_E D_\beta^-(u_t(x), u_t(y))\,\boldsymbol{\vartheta}(\mathrm{d}x, \mathrm{d}y) \quad \text{for a.e. } t \in (a, b). \quad (4.59)$$

If moreover

$$-\frac{\mathrm{d}}{\mathrm{d}t}\int_V \beta(u_t)\,\mathrm{d}\pi = \mathscr{R}(\rho_t, \boldsymbol{j}_t) + \frac{1}{2}\iint_E D_\beta^-(u_t(x), u_t(y))\,\boldsymbol{\vartheta}(\mathrm{d}x, \mathrm{d}y)$$

then $2\boldsymbol{j} = \boldsymbol{j}^\flat$ and

$$-w_t(x, y) = (\Psi^*)'\big(A_\beta(u_t(x), u_t(y))\big)\alpha(u_t(x), u_t(y)) \quad \text{for } \boldsymbol{\vartheta}\text{-a.e. } (x, y) \in E. \quad (4.60)$$

In particular, $w_t = 0$ $\boldsymbol{\vartheta}$-a.e. in $\big\{(x, y) \in E : \alpha(u_t(x), u_t(y)) = 0\big\}$.

Proof We recall that for λ-a.e. $t \in (a, b)$

$$\mathscr{R}(\rho_t, \boldsymbol{j}_t) = \frac{1}{2}\iint_E \Upsilon(u_t(x), u_t(y), w_t(x, y))\,\boldsymbol{\vartheta}(\mathrm{d}x, \mathrm{d}y).$$

We can then apply Lemma 4.19 and Theorem 4.16, observing that

$$\iint_E \Upsilon(u_t(x), u_t(y), w_t^\flat(x, y))\,\boldsymbol{\vartheta}(\mathrm{d}x, \mathrm{d}y) \le \iint_E \Upsilon(u_t(x), u_t(y), w(x, y))\,\boldsymbol{\vartheta}(\mathrm{d}x, \mathrm{d}y) \quad (4.61)$$

since

$$\Upsilon(u_t(x), u_t(y), w_t^\flat(x, y)) = \Upsilon(u_t(x), u_t(y), \frac{1}{2}(w_t(x, y) - w_t(y, x)))$$

$$\le \frac{1}{2}\Upsilon(u_t(x), u_t(y), w_t(x, y)) + \frac{1}{2}\Upsilon(u_t(x), u_t(y), w_t(y, x))$$

and the integral of the last term coincides with the right-hand side of (4.61) thanks to the symmetry of $\boldsymbol{\vartheta}$. $\qquad\square$

4.5 Compactness properties of curves with uniformly bounded $\mathscr{R}$-action

The next result shows an important compactness property for collections of curves in $\mathcal{A}(a, b)$ with bounded action. Recalling the discussion and the notation of Sect. 2.4, we will systematically associate with a given $(\rho, j) \in \mathcal{A}(I), I = [a, b]$, a couple of measures $\rho_\lambda \in \mathcal{M}^+(I \times V)$, $j_\lambda \in \mathcal{M}(I \times E)$ by integrating with respect to the Lebesgue measure λ in I:

$$\rho_\lambda(\mathrm{d}t, \mathrm{d}x) = \lambda(\mathrm{d}t)\rho_t(\mathrm{d}x), \quad j_\lambda(\mathrm{d}t, \mathrm{d}x, \mathrm{d}y) = \lambda(\mathrm{d}t)j_t(\mathrm{d}x, \mathrm{d}y). \tag{4.62}$$

Similarly, we define

$$\begin{aligned}
\boldsymbol{\vartheta}^\pm_{\rho,\lambda}(\mathrm{d}t, \mathrm{d}x, \mathrm{d}y) :=& (\boldsymbol{\vartheta}^\pm_\rho)_\lambda(\mathrm{d}t, \mathrm{d}x, \mathrm{d}y) = \lambda(\mathrm{d}t)\boldsymbol{\vartheta}^\pm_{\rho_t}(\mathrm{d}x, \mathrm{d}y) \\
=& \lambda(\mathrm{d}t)\rho_t(\mathrm{d}x)\kappa(x, \mathrm{d}y) = \boldsymbol{\vartheta}^\pm_{\rho_\lambda}(\mathrm{d}t, \mathrm{d}x, \mathrm{d}y).
\end{aligned} \tag{4.63}$$

It is not difficult to check that

$$\int_I \mathscr{R}(\rho_t, j_t)\, \mathrm{d}t = \frac{1}{2}\mathscr{F}_\Upsilon(\boldsymbol{\vartheta}^-_{\rho,\lambda}, \boldsymbol{\vartheta}^+_{\rho,\lambda}, 2j_\lambda | \lambda \otimes \boldsymbol{\vartheta}). \tag{4.64}$$

Proposition 4.21 (Bounded $\int \mathscr{R}$ implies compactness and lower semicontinuity) *Let* $(\rho^n, j^n)_n \subset \mathcal{A}(a, b)$ *be a sequence such that the initial states* ρ^n_a *are* π*-absolutely-continuous and relatively compact with respect to setwise convergence. Assume that*

$$M := \sup_{n \in \mathbb{N}} \int_a^b \mathscr{R}(\rho^n_t, j^n_t)\, \mathrm{d}t < +\infty. \tag{4.65}$$

Then, there exist a subsequence (not relabelled) and a pair $(\rho, j) \in \mathcal{A}(a, b)$ *such that, for the measures* $j^n_\lambda \in \mathcal{M}([a, b] \times E)$ *defined as in* (4.62) *there holds*

$$\rho^n_t \to \rho_t \quad \text{setwise in } \mathcal{M}^+(V) \text{ for all } t \in [a, b]\,, \tag{4.66a}$$

$$j^n_\lambda \rightharpoonup j_\lambda \quad \text{setwise in } \mathcal{M}([a, b] \times E)\,, \tag{4.66b}$$

where j_λ *is induced (in the sense of* (4.62)*) by a* λ*-integrable family* $(j_t)_{t \in [a,b]} \subset \mathcal{M}(E)$*. In addition, for any sequence* (ρ^n, j^n) *converging to* (ρ, j) *in the sense of* (4.66)*, we have*

$$\int_a^b \mathscr{R}(\rho_t, j_t)\, \mathrm{d}t \le \liminf_{n \to \infty} \int_a^b \mathscr{R}(\rho^n_t, j^n_t)\, \mathrm{d}t. \tag{4.67}$$

Proof Let us first remark that the mass conservation property of the continuity equation yields

$$\rho^n_t(V) = \rho^n_a(V) \le M_1 \quad \text{for every } t \in [a, b],\ n \in \mathbb{N} \tag{4.68}$$

for a suitable finite constant M_1 independent of n. We deduce that for every $t \in [a, b]$ the measures $\boldsymbol{\vartheta}^\pm_{\rho^n_t}$ have total mass bounded by $M_1\|\kappa_V\|_\infty$, so that estimate (2.29) for $y = (c, c) \in D(\alpha_*)$ yields

$$\mathbf{v}_{\rho^n_t}(E) = \alpha[\boldsymbol{\vartheta}^+_{\rho^n_t}, \boldsymbol{\vartheta}^-_{\rho^n_t}|\boldsymbol{\vartheta}](E) \le M_2 \quad \text{for every } t \in [a, b],\ n \in \mathbb{N}, \tag{4.69}$$

where $M_2 := 2c\, M_1\|\kappa_V\|_\infty - \alpha_*(c, c)\boldsymbol{\vartheta}(E)$. Jensen's inequality (2.18) and the monotonicity property (2.19) yield

$$\mathscr{R}(\rho^n_t, j^n_t) \ge \frac{1}{2}\hat{\Psi}\Big(2j^n_t(E), \mathbf{v}_{\rho^n_t}(E)\Big) \ge \frac{1}{2}\hat{\Psi}\Big(2j^n_t(E), M_2\Big) = \frac{1}{2}\Psi\Big(\frac{2j^n_t(E)}{M_2}\Big)M_2, \tag{4.70}$$

with $\hat{\Psi}$ the perspective function associated with Ψ, cf. (2.14). Since Ψ has superlinear growth, we deduce that the sequence of functions $t \mapsto |j^n_t|(E)$ is equi-integrable.

Since the sequence $(\rho_a^n)_n$, with $\rho_a^n = u_a^n \pi \ll \pi$, is relatively compact with respect to setwise convergence, by Theorems 2.1(6) and 2.2(3) there exist a convex superlinear function $\beta : \mathbb{R}_+ \to \mathbb{R}_+$ and a constant $M_3 < +\infty$ such that

$$\mathscr{F}_\beta(\rho_a^n|\pi) = \int_V \beta(u_a^n) \, \mathrm{d}\pi \le M_3 \quad \text{for every } n \in \mathbb{N}. \tag{4.71}$$

Possibly adding M_1 to M_3, it is not restrictive to assume that $\beta'(r) \ge 1$. We can then apply Lemma C.3 and we can find a smooth convex superlinear function $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ such that (C.10) holds. In particular

$$\int_V \omega(u_a^n) \, \mathrm{d}\pi \le M_1, \tag{4.72}$$

$$\int_a^b \iint_E \mathrm{D}_\omega^-(u_r^n(x), u_r^n(y)) \, \vartheta(\mathrm{d}x, \mathrm{d}y) \, \mathrm{d}r \le \int_a^b \iint_E (u_r^n(x) + u_r^n(y)) \, \vartheta(\mathrm{d}x, \mathrm{d}y) \, \mathrm{d}r$$
$$\le 2(b-a)M_1 \|\kappa_V\|_\infty. \tag{4.73}$$

By Corollary 4.20 we obtain

$$\int_V \omega(u_t^n) \, \mathrm{d}\pi \le M + (b-a)M_1 \|\kappa_V\|_\infty + M_1 \quad \text{for every } t \in [a, b]. \tag{4.74}$$

By (4.7) we deduce that

$$\|u_t^n - u_s^n\|_{L^1(V,\pi)} \le \zeta(s, t) \quad \text{where} \quad \zeta(s, t) := 2 \sup_{n \in \mathbb{N}} \int_s^t |\boldsymbol{j}_r^n|(E) \, \mathrm{d}r. \tag{4.75}$$

Since $t \mapsto |\boldsymbol{j}_t^n|(E)$ is equi-integrable we have

$$\lim_{(s,t) \to (r,r)} \zeta(s, t) = 0 \quad \text{for all } r \in [a, b],$$

We conclude that the sequence of maps $(u_t^n)_{t \in [a,b]}$ satisfies the conditions of the compactness result [4, Prop. 3.3.1], which yields the existence of a (not relabelled) subsequence and of a $L^1(V, \pi)$-continuous (thus also weakly-continuous) function $[a, b] \ni t \mapsto u_t \in L^1(V, \pi)$ such that $u_t^n \rightharpoonup u_t$ weakly in $L^1(V, \pi)$ for every $t \in [a, b]$. By (2.5) we also deduce that (4.66a) holds, i.e.

$$\rho_t^n \to \rho_t = u_t \pi \quad \text{setwise in } \mathcal{M}(V) \text{ for all } t \in [a, b].$$

It is also clear that for every $t \in [a, b]$ we have $\vartheta_{\rho_t^n}^\pm \to \vartheta_{\rho_t}^\pm$ setwise. The Dominated Convergence Theorem and (2.4), (2.36) imply that the corresponding measures $\vartheta_{\rho^n,\lambda}^\pm$ converge setwise to $\vartheta_{\rho,\lambda}^\pm$, and are therefore equi-absolutely continuous with respect to $\vartheta_\lambda = \lambda \otimes \vartheta$ (recall (2.7)).

Let us now show that also the sequence $(\boldsymbol{j}_\lambda^n)_n$ is equi-absolutely continuous with respect to ϑ_λ, so that (4.66b) holds up to extracting a further subsequence.

Selecting a constant $c > 0$ sufficiently large so that $\alpha(u_1, u_2) \le c(1 + u_1 + u_2)$, the trivial estimate $\boldsymbol{v}_\rho \le c(\vartheta + \vartheta_\rho^- + \vartheta_\rho^+)$ and the monotonicity property (2.19) yield

$$M \ge \int_a^b \mathscr{R}(\rho_t^n, \boldsymbol{j}_t^n) \, \mathrm{d}t = \frac{1}{2} \mathscr{F}_\Psi(2\boldsymbol{j}_\lambda^n|\boldsymbol{v}_{\rho_\lambda^n}) \ge \mathscr{F}_\Psi(\boldsymbol{j}_\lambda^n|\boldsymbol{\varsigma}^n), \quad \boldsymbol{\varsigma}^n := c(\vartheta_\lambda + \vartheta_{\rho^n,\lambda}^+ + \vartheta_{\rho^n,\lambda}^-). \tag{4.76}$$

For every $B \in \mathfrak{A} \otimes \mathfrak{B}$, $\mathfrak{A}$ being the Borel σ-algebra of $[a, b]$, with $\boldsymbol{\vartheta}_\lambda(B) > 0$, Jensen's inequality (2.18) yields

$$\Psi\left(\frac{\boldsymbol{j}_\lambda^n(B)}{\varsigma^n(B)}\right)\varsigma^n(B) \le \mathscr{F}_\Psi(\boldsymbol{j}_\lambda^n \llcorner B | \varsigma^n \llcorner B) \le M. \tag{4.77}$$

Denoting by $U : \mathbb{R}_+ \to \mathbb{R}_+$ the inverse function of Ψ, we thus find

$$\boldsymbol{j}_\lambda^n(B) \le \varsigma^n(B)\, U\left(\frac{M}{\varsigma^n(B)}\right). \tag{4.78}$$

Since Ψ is superlinear, U is sublinear so that

$$\lim_{\delta \downarrow 0} \delta U(M/\delta) = 0. \tag{4.79}$$

For every $\varepsilon > 0$ there exists $\delta_0 > 0$ such that $\delta U(M/\delta) \le \varepsilon$ for every $\delta \in (0, \delta_0)$. Since ς^n is equi absolutely continuous with respect to $\boldsymbol{\vartheta}_\lambda$ we can also find $\delta_1 > 0$ such that $\boldsymbol{\vartheta}_\lambda(B) < \delta_1$ yields $\varsigma^n(B) \le \delta_0$. By (4.78) we eventually conclude that $\boldsymbol{j}_\lambda^n(B) \le \varepsilon$.

It is then easy to pass to the limit in the integral formulation (4.4) of the continuity equation. Finally, concerning (4.67), it is sufficient to use the equivalent representation given by (4.64).

$\square$

4.6 Definition and properties of the cost

We now define the Dynamical-Variational Transport cost $\mathcal{W} : (0, +\infty) \times \mathcal{M}^+(V) \times \mathcal{M}^+(V) \to [0, +\infty)$ by

$$\mathscr{W}(\tau, \rho_0, \rho_1) := \inf\left\{\int_0^\tau \mathscr{R}(\rho_t, \boldsymbol{j}_t)\, \mathrm{d}t \,:\, (\rho, \boldsymbol{j}) \in \mathcal{CE}(0, \tau; \rho_0, \rho_1)\right\}. \tag{4.80}$$

In studying the properties of $\mathcal{W}$, we will also often use the notation

$$\mathscr{A}(0, \tau; \rho_0, \rho_1) := \left\{(\rho, \boldsymbol{j}) \in \mathcal{A}(0, \tau) \,:\, \rho(0) = \rho_0,\ \rho(\tau) = \rho_1\right\}, \tag{4.81}$$

with $\mathcal{A}(0, \tau)$ the class from (4.33).

For given $\tau > 0$ and $\rho_0, \rho_1 \in \mathcal{M}^+(V)$, if the set $\mathscr{A}(0, \tau; \rho_0, \rho_1)$ is non-empty, then it contains an exact minimizer for $\mathscr{W}(\tau, \rho_0, \rho_1)$. This is stated by the following result that is a direct consequence of Proposition 4.21.

Corollary 4.22 (Existence of minimizers) *If $\rho_0, \rho_1 \in \mathcal{M}^+(V)$ and $\mathscr{A}(0, \tau; \rho_0, \rho_1)$ is not empty, then the infimum in (4.80) is achieved.*

Remark 4.23 (**Scaling invariance**) Let us consider the perspective function $\hat{\Psi}(r, s)$ associated wih Ψ as in (2.14), $\hat{\Psi}(r, s) = s\Psi(r/s)$ if $s > 0$. We call $\mathscr{R}_s(\rho, \boldsymbol{j})$ the dissipation functional induced by $\hat{\Psi}(\cdot, s)$, with induced Dynamic-Transport cost $\mathscr{W}_s$. For every $\tau > 0$, $\rho_0, \rho_1 \in \mathcal{M}^+(V)$ a rescaling argument yields

$$\mathscr{W}(\tau, \rho_0, \rho_1) = \mathscr{W}_{\tau/\sigma}(\sigma, \rho_0, \rho_1) = \inf\left\{\int_0^\sigma \mathscr{R}_{\tau/\sigma}(\rho_t, \boldsymbol{j}_t)\, \mathrm{d}t \,:\, (\rho, \boldsymbol{j}) \in \mathcal{CE}(0, \sigma; \rho_0, \rho_1)\right\}. \tag{4.82}$$

In particular, choosing $\sigma = 1$ we find

$$\mathscr{W}(\tau, \rho_0, \rho_1) = \mathscr{W}_\tau(1, \rho_0, \rho_1). \tag{4.83}$$

Since $\hat{\Psi}(\cdot, \tau)$ is convex, lower semicontinuous, and decreasing with respect to τ, we find that $\tau \mapsto \mathscr{W}(\tau, \rho_0, \rho_1)$ is decreasing and convex as well.

Currently, proving that *any* pair of measures can be connected by a curve with finite action $\int \mathscr{R}$ under general conditions on V, Ψ and α is an open problem: in other words, in the general case we cannot exclude that $\mathscr{A}(0, \tau; \rho_0, \rho_1) = \emptyset$, which would make $\mathscr{W}(\tau, \rho_0, \rho_1) = +\infty$. Nonetheless, in a more specific situation, Proposition 4.25 below provides sufficient conditions for this connectivity property, between two measures $\rho_0, \rho_1 \in \mathcal{M}^+(V)$ with the same mass and such that $\rho_i \ll \pi$ for $i \in \{0, 1\}$. Preliminarily, we give the following

Definition 4.24 Let $q \in (1, +\infty)$. We say that the measures (π, ϑ) satisfy a *q-Poincaré inequality* if there exists a constant $C_P > 0$ such that for every $\xi \in L^q(V; \pi)$ with $\int_V \xi(x)\pi(\mathrm{d}x) = 0$ there holds

$$\int_V |\xi(x)|^q \pi(\mathrm{d}x) \le C_P \int_E |\overline{\nabla}\xi(x, y)|^q \vartheta(\mathrm{d}x, \mathrm{d}y). \tag{4.84}$$

We are now in a position to state the connectivity result, where we specialize the discussion to dissipation densities with *p*-growth for some $p \in (1, +\infty)$.

Proposition 4.25 *Suppose that*

$$\exists\, p \in (1, +\infty),\ \overline{C}_p > 0\ \forall\, r \in \mathbb{R}: \qquad \Psi(r) \le \overline{C}_p(1+|r|^p), \tag{4.85}$$

and that the measures (π, ϑ) satisfy a q-Poincaré inequality for $q = \frac{p}{p-1}$. Let $\rho_0, \rho_1 \in \mathcal{M}^+(V)$ with the same mass be given by $\rho_i = u_i\pi$, with positive $u_i \in L^1(V; \pi) \cap L^\infty(V; \pi)$, for $i \in \{0, 1\}$. Then, for every $\tau > 0$ the set $\mathscr{A}(0, \tau; \rho_0, \rho_1)$ is non-empty and thus $\mathscr{W}(\tau, \rho_0, \rho_1) < \infty$.

We postpone the proof of Proposition 4.25 to Appendix D, where some preliminary results, also motivating the role of the *q*-Poincaré inequality, will be provided.

4.7 Abstract-level properties of $\mathscr{W}$

The main result of this section collects a series of properties of the cost that will play a key role in the study of the *Minimizing Movement* scheme (1.26). Indeed, as already hinted in the Introduction, the analysis that we will carry out in Sect. 7 ahead might well be extended to a scheme set up in a general topological space, endowed with a cost functional enjoying properties (4.86) below. We will now check them for the cost $\mathscr{W}$ associated with generalized gradient structure $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ fulfilling **Assumptions ($V\pi\kappa$) and ($\mathscr{R}^*\Psi\alpha$)**. In this section *all convergences will be with respect to the setwise topology.*

Theorem 4.26 *The cost $\mathscr{W}$ enjoys the following properties:*

(1) For all $\tau > 0$, $\rho_0, \rho_1 \in \mathcal{M}^+(V)$,

$$\mathscr{W}(\tau, \rho_0, \rho_1) = 0 \Leftrightarrow \rho_0 = \rho_1. \tag{4.86a}$$

(2) For all $\rho_1, \rho_2, \rho_3 \in \mathcal{M}^+(V)$ and $\tau_1, \tau_2 \in (0, +\infty)$ with $\tau = \tau_1 + \tau_2$,

$$\mathscr{W}(\tau, \rho_1, \rho_3) \le \mathscr{W}(\tau_1, \rho_1, \rho_2) + \mathscr{W}(\tau_2, \rho_2, \rho_3). \tag{4.86b}$$

(3) For $\tau_n \to \tau > 0$, $\rho_0^n \to \rho$, $\rho_1^n \to \rho_1$ in $\mathcal{M}^+(V)$,

$$\liminf_{n \to +\infty} \mathscr{W}(\tau_n, \rho_0^n, \rho_1^n) \geq \mathscr{W}(\tau, \rho_0, \rho_1). \tag{4.86c}$$

(4) For all $\tau_n \downarrow 0$ and for all $(\rho_n)_n$, $\rho \in \mathcal{M}^+(V)$,

$$\sup_{n \in \mathbb{N}} \mathscr{W}(\tau_n, \rho_n, \rho) < +\infty \quad \Rightarrow \quad \rho_n \to \rho. \tag{4.86d}$$

(5) For all $\tau_n \downarrow 0$ and all $(\rho_n)_n$, $(v_n)_n \subset \mathcal{M}^+(V)$ with $\rho_n \to \rho$, $v_n \to v$,

$$\limsup_{n \to \infty} \mathscr{W}(\tau_n, \rho_n, v_n) < +\infty \quad \Rightarrow \quad \rho = v. \tag{4.86e}$$

Proof *(1)* Since $\Psi(s)$ is strictly positive for $s \neq 0$ it is immediate to check that $\mathscr{R}(\rho, \boldsymbol{j}) = 0 \Rightarrow \boldsymbol{j} = 0$. For an optimal pair $(\rho, \boldsymbol{j})$ satisfying $\int_0^\tau \mathscr{R}(\rho_t, \boldsymbol{j}_t) \, dt = 0$ we deduce that $\boldsymbol{j}_t = 0$ for a.e. $t \in (0, \tau)$. The continuity equation then implies $\rho_0 = \rho_1$.

(2) This can easily be checked by using the existence of minimizers for $\mathscr{W}(\tau, \rho_0, \rho_1)$.

(3) Assume without loss of generality that $\liminf_{n \to +\infty} \mathscr{W}(\tau_n, \rho_0^n, \rho_1^n) < \infty$. By (4.83) we use that, for every $n \in \mathbb{N}$ and setting $\overline{\tau} = \sup_n \tau_n$,

$$\mathscr{W}(\tau_n, \rho_n^0, \rho_n^1) = \mathscr{W}_{\tau_n}(1, \rho_n^0, \rho_n^1) \leq \mathscr{W}_{\overline{\tau}}(1, \rho_n^0, \rho_n^1) \overset{(*)}{=} \int_0^1 \mathscr{R}_{\overline{\tau}}(\rho_t^n, \boldsymbol{j}_t^n) \, dt,$$

where the identity $(*)$ holds for an optimal pair $(\rho^n, \boldsymbol{j}^n) \in \mathcal{CE}(0, 1; \rho_0^n, \rho_1^n)$. Applying Proposition 4.21, we obtain the existence of $(\rho, \boldsymbol{j}) \in \mathcal{CE}(0, 1; \rho_0, \rho_1)$ such that, up to a subsequence,

$$\begin{aligned} &\rho_s^n \to \rho_s \text{ setwise in } \mathcal{M}^+(V) \quad \text{for all } s \in [0, 1], \\ &\boldsymbol{j}^n \to \boldsymbol{j} \text{ setwise in } \mathcal{M}([0, 1] \times E), \end{aligned} \tag{4.87}$$

Arguing as in Proposition 4.21 and using the joint lower semicontinuity of $\hat{\Psi}$, we find that

$$\liminf_{n \to \infty} \int_0^1 \mathscr{R}_{\tau_n}(\rho_s^n, \boldsymbol{j}_s^n) \, ds \geq \int_0^1 \mathscr{R}_\tau(\rho_s, \boldsymbol{j}_s) \, ds \geq \mathscr{W}_\tau(1, \rho_0, \rho_1) = \mathscr{W}(\tau, \rho_0, \rho_1).$$

(4) If we denote by $\mathscr{R}_0$ the dissipation associated with $\hat{\Psi}(\cdot, 0)$, given by $\hat{\Psi}(w, 0) = +\infty$ for $w \neq 0$ and $\hat{\Psi}(0, 0) = 0$, we find

$$\mathscr{R}_0(\rho, \boldsymbol{j}) < +\infty \quad \Rightarrow \quad \boldsymbol{j} = 0. \tag{4.88}$$

By the same argument as for part (3), every subsequence of ρ_n has a converging subsequence in the setwise topology; the lower semicontinuity result of the proof of part (3) shows that any limit point must coincide with ρ.

(5) The argument combines (4.88) and part (3). □

4.8 The action functional $\mathbb{W}$ and its properties

The construction of $\mathscr{R}$ and $\mathscr{W}$ above proceeded in the order $\mathscr{R} \rightsquigarrow \mathscr{W}$: we first constructed $\mathscr{R}$, and then $\mathscr{W}$ was defined in terms of $\mathscr{R}$. It is a natural question whether one can invert this construction: given $\mathscr{W}$, can one reconstruct $\mathscr{R}$, or at least integrals of the form $\int_a^b \mathscr{R} \, dt$? The answer is positive, as we show in this section.

Given a functional $\mathscr{W}$ satisfying the properties (4.86), we define the '$\mathscr{W}$-action' of a curve $\rho : [a, b] \to \mathcal{M}^+(V)$ as

$$\mathbb{W}(\rho; [a, b]) := \sup \left\{ \sum_{j=1}^{M} \mathscr{W}(t^j - t^{j-1}, \rho(t^{j-1}), \rho(t^j)) \; : \; (t^j)_{j=0}^{M} \in \mathfrak{P}_f([a, b]) \right\},$$
(4.89)

for all $[a, b] \subset [0, T]$ where $\mathfrak{P}_f([a, b])$ denotes the set of all partitions of a given interval $[a, b]$.

If $\mathscr{W}$ is defined by (4.80), then each term in the sum above is defined as an optimal version of $\int_{t^{j-1}}^{t^j} \mathscr{R}(\rho_t, \cdot) \, dt$, and we might expect that $\mathbb{W}(\rho; [a, b])$ is an optimal version of $\int_a^b \mathscr{R}(\rho_t, \cdot) \, dt$. This is indeed the case, as is illustrated by the following analogue of [20, Th. 5.17]:

Proposition 4.27 *Let $\mathscr{W}$ be given by (4.80), and let $\rho : [0, T] \to \mathcal{M}^+(V)$. Then $\mathbb{W}(\rho; [0, T]) < +\infty$ if and only if there exists a measurable map $\boldsymbol{j} : [0, T] \to \mathcal{M}(E)$ such that $(\rho, \boldsymbol{j}) \in \mathcal{CE}(0, T)$ with $\int_0^T \mathscr{R}(\rho_t, \boldsymbol{j}_t) dt < +\infty$. In that case,*

$$\mathbb{W}(\rho; [0, T]) \le \int_0^T \mathscr{R}(\rho_t, \boldsymbol{j}_t) \, dt,$$
(4.90)

and there exists a unique $\boldsymbol{j}_{\mathrm{opt}}$ such that equality is achieved. The optimal $\boldsymbol{j}_{\mathrm{opt}}$ is skew-symmetric, i.e. $\boldsymbol{j}_{\mathrm{opt}} = \boldsymbol{j}_{\mathrm{opt}}^{\flat}$ (cf. Remark 4.12).

Prior to proving Proposition 4.27, we establish the following approximation result.

Lemma 4.28 *Let $\rho : [0, T] \to \mathcal{M}^+(V)$ satisfy $\mathbb{W}(\rho; [0, T]) < +\infty$. For a sequence of partitions $P_n = (t_n^j)_{j=0}^{M_n} \in \mathfrak{P}_f([0, T])$ with fineness $\tau_n = \max_{j=1,\dots,M_n}(t_n^j - t_n^{j-1})$ converging to zero, let $\rho^n : [0, T] \to \mathcal{M}^+(V)$ satisfy*

$$\rho^n(t_n^j) = \rho(t_n^j) \;\; \text{for all } j = 1, \dots, M_n \quad \text{and} \quad \sup_{n \in \mathbb{N}} \mathbb{W}(\rho^n; [0, T]) < +\infty.$$

Then $\rho^n(t) \to \rho(t)$ setwise for all $t \in [0, T]$ as $n \to \infty$.

Proof First of all, observe that by the symmetry of Ψ, also the time-reversed curve $\check{\rho}(t) := \rho(T - t)$ satisfies $\mathbb{W}(\check{\rho}; [0, T]) < +\infty$. Let $\bar{\mathfrak{t}}_n$ and $\underline{\mathfrak{t}}_n$ be the piecewise constant interpolants associated with the partitions P_n, cf. (7.5). Fix $t \in [0, T]$; we estimate

$$\mathscr{W}\big(2(\bar{\mathfrak{t}}_n - t), \rho^n(t), \rho(t)\big) \overset{(1)}{\le} \mathscr{W}\big(\bar{\mathfrak{t}}_n - t, \rho^n(t), \rho^n(\bar{\mathfrak{t}}_n(t))\big) + \mathscr{W}\big(\bar{\mathfrak{t}}_n - t, \rho(\bar{\mathfrak{t}}_n(t)), \rho(t)\big)$$
$$= \mathscr{W}\big(\bar{\mathfrak{t}}_n - t, \rho^n(t), \rho^n(\bar{\mathfrak{t}}_n(t))\big) + \mathscr{W}\big(\bar{\mathfrak{t}}_n - t, \check{\rho}(T - \bar{\mathfrak{t}}_n(t)), \check{\rho}(T - t)\big)$$
$$\le \mathbb{W}(\rho^n; [t, \bar{\mathfrak{t}}_n(t)]) + \mathbb{W}(\check{\rho}; [T - \bar{\mathfrak{t}}_n(t), T - t])$$
$$\le \sup_{n \in \mathbb{N}} \mathbb{W}(\rho^n; [0, T]) + \mathbb{W}(\check{\rho}; [0, T]) =: C < +\infty,$$

where (1) follows from property (4.86b) of $\mathscr{W}$. Consequently, by property (4.86d) it follows that $\rho^n(t) \to \rho(t)$ setwise in $\mathcal{M}^+(V)$ for all $t \in [0, T]$. $\qquad\square$

We are now in a position to prove Proposition 4.27:

Proof of Proposition 4.27 One implication is straightforward: if a pair (ρ, j) exists, then

$$\mathscr{W}(t-s, \rho_s, \rho_t) \overset{(4.80)}{\leq} \int_s^t \mathscr{R}(\rho_r, j_r) \, dr, \qquad \text{for all } 0 \leq s < t \leq T,$$

and therefore $\mathbb{W}(\rho; [0, T]) < +\infty$ and (4.90) holds.

To prove the other implication, assume that $\mathbb{W}(\rho; [0, T]) < +\infty$. Choose a sequence of partitions $P_n = (t_n^j)_{j=0}^{M_n} \in \mathfrak{P}_f([0, T])$ that becomes dense in the limit $n \to \infty$. For each $n \in \mathbb{N}$, construct a pair $(\rho^n, j^n) \in \mathcal{CE}(0, T)$ as follows: On each time interval $[t_n^{j-1}, t_n^j]$, let (ρ^n, j^n) be given by Corollary 4.22 as the minimizer under the constraint $\rho^n(t_n^{j-1}) = \rho(t_n^{j-1})$ and $\rho^n(t_n^j) = \rho(t_n^j)$, namely

$$\mathscr{W}(t_n^j - t_n^{j-1}, \rho(t_n^{j-1}), \rho(t_n^j)) = \int_{t_n^{j-1}}^{t_n^j} \mathscr{R}(\rho_r^n, j_r^n) \, dr. \tag{4.91}$$

By concatenating the minimizers on each of the intervals a pair $(\rho^n, j^n) \in \mathcal{CE}(0, T)$ is obtained, thanks to Lemma 4.6. By construction we have the property

$$\mathbb{W}(\rho^n; [0, T]) = \int_0^T \mathscr{R}(\rho_t^n, j_t^n) \, dt. \tag{4.92}$$

Also by optimality we have

$$\mathbb{W}(\rho^n; [t_n^{j-1}, t_n^j]) = \mathscr{W}(t_n^j - t_n^{j-1}, \rho(t_n^{j-1}), \rho(t_n^j)) \leq \mathbb{W}(\rho; [t_n^{j-1}, t_n^j]),$$

which implies by summing that

$$\mathbb{W}(\rho^n; [0, T]) \leq \mathbb{W}(\rho; [0, T]). \tag{4.93}$$

By Lemma 4.28 we then find that $\rho^n(t) \to \rho(t)$ setwise as $n \to \infty$ for each $t \in [0, T]$.

Applying Proposition 4.21, we find that $j^n(dt \, dx \, dy) := j_t^n(dx \, dy) \, dt$ setwise converges along a subsequence to a limit j. The limit j can be disintegrated as $j(dt \, dx \, dy) = \lambda(dt) \, j_t(dx \, dy)$ for a measurable family $(j_t)_{t \in [0,T]}$, and the pair (ρ, j) is an element of $\mathcal{CE}(0, T)$. In addition we have the lower-semicontinuity property

$$\liminf_{n \to \infty} \int_0^T \mathscr{R}(\rho_t^n, j_t^n) \, dt \geq \int_0^T \mathscr{R}(\rho_t, j_t) \, dt. \tag{4.94}$$

We then have the series of inequalities

$$\mathbb{W}(\rho; [0, T]) \overset{(4.93)}{\geq} \limsup_{n \to \infty} \mathbb{W}(\rho^n; [0, T]) \overset{(4.92)}{=} \limsup_{n \to \infty} \int_0^T \mathscr{R}(\rho_t^n, j_t^n) \, dt$$
$$\overset{(4.94)}{\geq} \int_0^T \mathscr{R}(\rho_t, j_t) \, dt \overset{(4.90)}{\geq} \mathbb{W}(\rho; [0, T]),$$

which implies that $\int_0^T \mathscr{R}(\rho_t, j_t) \, dt = \mathbb{W}(\rho; [0, T])$.

Finally, the uniqueness of j is a consequence of the strict convexity of $\Upsilon(u_1, u_2, \cdot)$, cf. Lemma 4.7. Similarly, the skew-symmetry of j follows from the strict convexity of $\Upsilon(u_1, u_2, \cdot)$, the symmetry of $\Upsilon(\cdot, \cdot, w)$, and the invariance of the continuity equation (4.3) under the 'skew-symmetrization' $j \mapsto j^\flat$, cf. Remark 4.12. $\qquad\square$

5 The fisher information $\mathscr{D}$ and the definition of solutions

With the definitions and the properties that we established in the previous section we have given a rigorous meaning to the first term in the functional $\mathscr{L}$ in (1.18). In this section we continue with the second term in the integral, often called *Fisher information*, after the canonical version in diffusion problems [68]. Section 5.2 is devoted to

(a) A rigorous definition of the Fisher information $\mathscr{D}(\rho)$ (Definition 5.1).

In several practical settings, such as the proof of existence that we give in Sect. 7, it is important to have lower semicontinuity of $\mathscr{D}$: this is proved in Proposition 5.3.

We are then in a position to give

(b) a rigorous definition of solutions to the $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ system (Definition 5.4).

In Sect. 1.2.1 we explained that the Energy-Dissipation balance approach to defining solutions is based on the fact that $\mathscr{L}(\rho, \boldsymbol{j}) \geq 0$ for all $(\rho, \boldsymbol{j})$ by the validity of a suitable chain-rule inequality.

(c) A rigorous proof of this chain-rule inequality, involving $\mathscr{R}$ and $\mathscr{D}$, is given in Corollary 5.6, which is based on Theorem 4.16).

This establishes the inequality $\mathscr{L}(\rho, \boldsymbol{j}) \geq 0$. Hence, we can rigorously deduce that the opposite inequality $\mathscr{L}(\rho, \boldsymbol{j}) \leq 0$ characterizes the property that $(\rho, \boldsymbol{j})$ is a solution to the $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ system. Theorem 5.7 provides an additional characterization of this solution concept.

Finally, in Sects. 5.3 and 5.4,

(d) we prove existence, uniqueness and stability of solutions under suitable convexity/l.s.c. conditions on $\mathscr{D}$ (Theorems 5.10 and 5.9). We also discuss their asymptotic behaviour and the role of the invariant measures π.

Throughout this section we adopt **Assumptions $(V\pi\kappa)$, $(\mathscr{R}^*\Psi\alpha)$, and $(\mathscr{E}\phi)$**.

5.1 The fisher information $\mathscr{D}$

Formally, the Fisher information is the second term in (1.18), namely

$$\mathscr{D}(\rho) = \mathscr{R}^*\Big(\rho, -\overline{\nabla}\phi(u)\Big) = \frac{1}{2}\iint_E \Psi^*\big(-(\phi'(u(y)) - \phi'(u(x)))\big)\boldsymbol{v}_\rho(\mathrm{d}x\,\mathrm{d}y), \qquad \rho = u\pi.$$

In order to give a precise meaning to this formulation when ϕ is not differentiable at 0 (as, for instance, in the case of the Boltzmann entropy function (3.10)), we use the function D_ϕ defined in (4.53c).

Definition 5.1 (**The Fisher-information functional** $\mathscr{D}$) The Fisher information $\mathscr{D} : \mathrm{D}(\mathscr{E}) \to [0, +\infty]$ is defined as

$$\mathscr{D}(\rho) := \frac{1}{2}\iint_E \mathrm{D}_\phi\big(u(x), u(y)\big)\,\boldsymbol{\vartheta}(\mathrm{d}x\,\mathrm{d}y) \qquad \text{for } \rho = u\pi. \tag{5.1}$$

Example 5.2 (**The Fisher information in the quadratic and in the** cosh **case**) For illustration we recall the two expressions for D_ϕ from Example 4.18 for the linear equation (1.2) with quadratic and cosh-type potentials Ψ^*:

(1) If $\Psi^*(s) = s^2/2$, then

$$\mathrm{D}_\phi(u, v) = \begin{cases} \frac{1}{2}(\log(u) - \log(v))(u - v) & \text{if } u, \ v > 0, \\ 0 & \text{if } u = v = 0, \\ +\infty & \text{if } u = 0 \text{ and } v \neq 0, \text{ or vice versa.} \end{cases}$$

(2) If $\Psi^*(s) = 4\big(\cosh(s/2) - 1\big)$, then

$$\mathrm{D}_\phi(u, v) = 2\Big(\sqrt{u} - \sqrt{v}\Big)^2 \qquad \forall \, (u, v) \in [0, +\infty) \times [0, +\infty).$$

These two examples of D_ϕ are convex. The convexity of D_ϕ in the case of potentials Ψ^* generated by the power means (1.33) is discussed in Appendix E.

Let us discuss the lower-semicontinuity properties of $\mathscr{D}$. In accordance with the Minimizing-Movement approach carried out in Sect. 7.1, we will just be interested in lower semicontinuity of $\mathscr{D}$ along sequences with bounded energy $\mathscr{E}$. Now, since sublevels of the energy $\mathscr{E}$ are relatively compact with respect to setwise convergence (by part 2 of Theorem 2.2), there is no difference between narrow and setwise lower semicontinuity of $\mathscr{D}$.

Proposition 5.3 *(Lower semicontinuity of $\mathscr{D}$) Assume either that π is purely atomic or that the function D_ϕ is convex on $\mathbb{R}_+^2$. Then $\mathscr{D}$ is (sequentially) lower semicontinuous with respect to setwise convergence, i.e., for all $(\rho^n)_n, \ \rho \in \mathrm{D}(\mathscr{E})$*

$$\rho^n \to \rho \text{ setwise in } \mathcal{M}^+(V) \implies \mathscr{D}(\rho) \leq \liminf_{n \to \infty} \mathscr{D}(\rho^n). \tag{5.2}$$

Proof When π is purely atomic, setwise convergence implies pointwise convergence π-a.e. for the sequence of the densities, so that (5.2) follows by Fatou's Lemma.

A standard argument, still based on Fatou's Lemma, shows that the functional

$$u \mapsto \iint_E \mathrm{D}_\phi(u(x), u(y)) \, \boldsymbol{\vartheta}(\mathrm{d}x, \mathrm{d}y) \tag{5.3}$$

is lower semicontinuous with respect to the strong topology in $L^1(V, \pi)$: it is sufficient to check that $u_n \to u$ in $L^1(V, \pi)$ implies $(u_n^-, u_n^+) \to (u^-, u^+)$ in $L^1(E, \boldsymbol{\vartheta})$. If D_ϕ is convex on $\mathbb{R}_+^2$, then the functional (5.3) is also lower semicontinuous with respect to the weak topology in $L^1(V, \pi)$. On the other hand, since ρ_n and ρ are absolutely continuous with respect to π, $\rho_n \to \rho$ setwise if and only if $\mathrm{d}\rho_n/\mathrm{d}\pi \rightharpoonup \mathrm{d}\rho/\mathrm{d}\pi$ weakly in $L^1(V, \pi)$ (see Theorem 2.1). $\qquad\square$

5.2 The definition of solutions: $\mathscr{R}/\mathscr{R}^*$ energy-dissipation balance

We are now in a position to formalize the concept of solution.

Definition 5.4 *($(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ **Energy-Dissipation balance**) We say that a curve $\rho : [0, T] \to \mathcal{M}^+(V)$ is a solution of the $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ evolution system, if it satisfies the $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ Energy-Dissipation balance*:

(1) $\mathscr{E}(\rho_0) < +\infty$;
(2) There exists a measurable family $(\boldsymbol{j}_t)_{t \in [0, T]} \subset \mathcal{M}(E)$ such that $(\rho, \boldsymbol{j}) \in \mathcal{CE}(0, T)$ with

$$\int_s^t \big(\mathscr{R}(\rho_r, \boldsymbol{j}_r) + \mathscr{D}(\rho_r)\big) \, \mathrm{d}r + \mathscr{E}(\rho_t) = \mathscr{E}(\rho_s) \qquad \text{for all } 0 \leq s \leq t \leq T. \tag{5.4}$$

Remark 5.5

(1) Since $(\rho, \boldsymbol{j}) \in \mathcal{CE}(0, T)$, the curve ρ is absolutely continuous with respect to the total variation distance.

(2) The Energy-Dissipation balance (5.4) written for $s = 0$ and $t = T$ implies that $(\rho, \boldsymbol{j}) \in \mathcal{A}(0, T)$ as well. Moreover, $t \mapsto \mathcal{E}(\rho_t)$ takes finite values and it is absolutely continuous in the interval $[0, T]$.

(3) The chain-rule estimate (4.59) implies the following important corollary:

Corollary 5.6 (Chain-rule estimate III) *For any curve $(\rho, \boldsymbol{j}) \in \mathcal{CE}(0, T)$,*

$$\mathcal{L}_T(\rho, \boldsymbol{j}) := \int_0^T \big(\mathcal{R}(\rho_r, \boldsymbol{j}_r) + \mathcal{D}(\rho_r)\big)\, dr + \mathcal{E}(\rho_T) - \mathcal{E}(\rho_0) \geq 0. \tag{5.5}$$

It follows that the Energy-Dissipation balance (5.4) is equivalent to the Energy-Dissipation Inequality

$$\mathcal{L}_T(\rho, \boldsymbol{j}) \leq 0. \tag{5.6}$$

Let us give an equivalent characterization of solutions to the $(\mathcal{E}, \mathcal{R}, \mathcal{R}^*)$ evolution system. Recalling the definition (1.11) of the map F in the interior of $\mathbb{R}_+^2$ and the definition (4.40) of A_ϕ, we first note that F can be extended to a function defined in $\mathbb{R}_+^2$ with values in the extended real line $[-\infty, +\infty]$ by

$$F_0(u, v) := \begin{cases} (\Psi^*)'\big(A_\phi(u, v)\big)\alpha(u, v) & \text{if } \alpha(u, v) > 0, \\ 0 & \text{if } \alpha(u, v) = 0. \end{cases} \tag{5.7}$$

where we set $(\Psi^*)'(\pm\infty) := \pm\infty$. The function F_0 is skew-symmetric.

Theorem 5.7 *A curve $(\rho_t)_{t \in [0,T]}$ in $\mathcal{M}^+(V)$ is a solution of the $(\mathcal{E}, \mathcal{R}, \mathcal{R}^*)$ system iff*

(1) *$\rho_t = u_t \pi \ll \pi$ for every $t \in [0, T]$ and $t \mapsto u_t$ is an absolutely continuous a.e. differentiable map with values in $L^1(V, \pi)$;*

(2) *$\mathcal{E}(\rho_0) < +\infty$;*

(3) *We have*

$$\int_0^T \iint_E |F_0(u_t(x), u_t(y))|\, \boldsymbol{\vartheta}(dx, dy)\, dt < +\infty; \tag{5.8}$$

and

$$D_\phi(u_t(x), u_t(y)) = D_\phi^-(u_t(x), u_t(y)) \quad \textit{for } \lambda \otimes \boldsymbol{\vartheta}\textit{-a.e. } (t, x, y) \in [0, T] \times E. \tag{5.9}$$

In particular the complement U' of the set

$$U := \{(t, x, y) \in [0, T] \times E : F_0(u_t(x), u_t(y)) \in \mathbb{R}\} \tag{5.10}$$

is $(\lambda \otimes \boldsymbol{\vartheta})$-negligible and F_0 takes finite values $(\lambda \otimes \boldsymbol{\vartheta})$-a.e. in $[0, T] \times E$;

(4) *Setting*

$$2\boldsymbol{j}_t(dx, dy) = -F_0(u_t(x), u_t(y))\, \boldsymbol{\vartheta}(dx, dy), \tag{5.11}$$

we have $(\rho, \boldsymbol{j}) \in \mathcal{CE}(0, T)$. In particular,

$$\dot{u}_t(x) = \int_V F_0(u_t(x), u_t(y))\, \kappa(x, dy) \quad \textit{for } (\lambda \otimes \pi)\textit{-a.e. } (t, x, y) \in [0, T] \times E. \tag{5.12}$$

Proof Let $\rho_t = u_t \pi$ be a solution of the $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ system with the corresponding flux $\boldsymbol{j}_t$. By Corollary 4.14 we can find a skew-symmetric measurable map $\xi : (0, T) \times E \to \mathbb{R}$ such that $\boldsymbol{j}_\lambda = \xi \alpha(u^-, u^+) \lambda \otimes \boldsymbol{\vartheta}$ and (4.36), (4.37) hold. Taking into account that $D_\phi^- \le D_\phi$ and applying the equality case of Corollary 4.20, we complete the proof of one implication.

Suppose now that ρ_t satisfies all the above conditions (1)–(4); we want to apply formula (4.45) of Theorem 4.16 for $\beta = \phi$. For this we write the shorthand u^-, u^+ for $u_t(x), u_t(y)$ and set $w = -F_0(u^-, u^+)$. We verify the equality conditions (4.56) of Lemma 4.19:

- At (t, x, y) where $\alpha(u^-, u^+) = 0$, we have by definition $w = -F_0(u^-, u^+) = 0$;
- At $(\lambda \otimes \boldsymbol{\vartheta})$–a.e. (t, x, y) where $\alpha(u^-, u^+) > 0$, $F_0(u^-, u^+)$ is finite by condition (3), and by (5.7) it follows that $(\Psi^*)'\big(A_\phi(u^-, u^+)\big)$ is finite and therefore $A_\phi(u^-, u^+)$ is finite. The final condition $-w = (\Psi^*)'\big(A_\beta(u, v)\big)\alpha(u, v)$ then follows by the definition of w.

By Lemma 4.19 therefore we have at $(\lambda \otimes \boldsymbol{\vartheta})$–a.e. (t, x, y)

$$-B_\phi(u^-, u^+, w) = \Upsilon(u^-, u^+, -w) + D_\phi^-(u^-, u^+) \overset{(5.9)}{=} \Upsilon(u^-, u^+, -w) + D_\phi(u^-, u^+).$$

In particular B_ϕ is nonpositive, and the integrability condition (4.44) is trivially satisfied. Integrating (4.45) in time we find (5.4). $\qquad\square$

Remark 5.8 By Theorem 5.7(3), along a solution $\rho_t = u_t \pi$ of the $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ system, the functions D_ϕ and D_ϕ^- coincide. Recall that, in general, we only have $D_\phi^- \le D_\phi$, and the inequality can be strict, as in the examples of the linear equation (1.2) with the Boltzmann entropy and the quadratic and cosh-dissipation potentials discussed in Ex. 4.18. There, D_ϕ and D_ϕ^- differ on the boundary of $\mathbb{R}^2$. Therefore, (5.9) encompasses the information that the pair $(u_t(x), u_t(y))$ stays in the interior of $\mathbb{R}^2$ $(\lambda \otimes \boldsymbol{\vartheta})$-a.e. in $[0, T] \times E$.

5.3 Existence and uniqueness of solutions of the $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ system

Let us now collect a few basic structural properties of solutions of the $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ Energy-Dissipation balance. Recall that we will always adopt **Assumptions $(V\pi\kappa)$, $(\mathscr{R}^*\Psi\alpha)$, and $(\mathscr{E}\phi)$**.

Following an argument by Gigli [34] we first use convexity of $\mathscr{D}$ to deduce uniqueness.

Theorem 5.9 (Uniqueness) *Suppose that $\mathscr{D}$ is convex and the energy density ϕ is strictly convex. Suppose that ρ^1, ρ^2 satisfy the $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ Energy-Dissipation balance (5.4) and are identical at time zero. Then $\rho_t^1 = \rho_t^2$ for every $t \in [0, T]$.*

Proof Let $\boldsymbol{j}^i \in \mathcal{M}((0, T) \times E)$ satisfy $\mathscr{L}_t(\rho^i, \boldsymbol{j}^i) = 0$ and let us set

$$\rho_t := \frac{1}{2}(\rho_t^1 + \rho_t^2), \quad \boldsymbol{j} := \frac{1}{2}(\boldsymbol{j}^1 + \boldsymbol{j}^2).$$

By the linearity of the continuity equation we have that $(\rho, \boldsymbol{j}) \in \mathcal{CE}(0, T)$ with $\rho_0 = \rho_0^1 = \rho_0^2$, so that by convexity

$$\begin{aligned}
\mathscr{E}(\rho_t) &\ge \mathscr{E}(\rho_0) - \int_0^t \big(\mathscr{R}(\rho_r, \boldsymbol{j}_r) + \mathscr{D}(\rho_r)\big)\,\mathrm{d}r \\
&\ge \mathscr{E}(\rho_0) - \frac{1}{2}\int_0^t \big(\mathscr{R}(\rho_r^1, \boldsymbol{j}_r^1) + \mathscr{D}(\rho_r^1)\big)\,\mathrm{d}r - \frac{1}{2}\int_0^t \big(\mathscr{R}(\rho_r^2, \boldsymbol{j}_r^2) + \mathscr{D}(\rho_r^2)\big)\,\mathrm{d}r \\
&= \frac{1}{2}\mathscr{E}(\rho_t^1) + \frac{1}{2}\mathscr{E}(\rho_t^2).
\end{aligned}$$

Since $\mathscr{E}$ is strictly convex we deduce $\rho_t^1 = \rho_t^2$.   □

Theorem 5.10 (Existence and stability) *Let us suppose that the Fisher information functional $\mathscr{D}$ is lower semicontinuous with respect to setwise convergence (e.g. if π is purely atomic, or D_ϕ is convex, see Proposition 5.3).*

(1) *For every $\rho_0 \in \mathcal{M}^+(V)$ with $\mathscr{E}(\rho_0) < +\infty$ there exists a solution $\rho : [0, T] \to \mathcal{M}^+(V)$ of the $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ evolution system starting from ρ_0.*
(2) *Every sequence $(\rho_t^n)_{t \in [0,T]}$ of solutions to the $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ evolution system such that*

$$\sup_{n \in \mathbb{N}} \mathscr{E}(\rho_0^n) < +\infty \tag{5.13}$$

has a subsequence setwise converging to a limit $(\rho_t)_{t \in [0,T]}$ for every $t \in [0, T]$.
(3) *Let $(\rho_t^n)_{t \in [0,T]}$ is a sequence of solutions, with corresponding fluxes $(j_t^n)_{t \in [0,T]}$. Let ρ_t^n converge setwise to ρ_t for every $t \in [0, T]$, and assume that*

$$\lim_{n \to \infty} \mathscr{E}(\rho_0^n) = \mathscr{E}(\rho_0). \tag{5.14}$$

Then ρ is a solution as well, with flux j, and the following additional convergence properties hold:

$$\lim_{n \to \infty} \int_0^T \mathscr{R}(\rho_t^n, j_t^n)\, dt = \lim_{n \to \infty} \int_0^T \mathscr{R}(\rho_t, j_t)\, dt, \tag{5.15a}$$

$$\lim_{n \to \infty} \int_0^T \mathscr{D}(\rho_t^n)\, dt = \lim_{n \to \infty} \int_0^T \mathscr{D}(\rho_t, j_t)\, dt, \tag{5.15b}$$

$$\lim_{n \to \infty} \mathscr{E}(\rho_t^n) = \mathscr{E}(\rho_t) \quad \textit{for every } t \in [0, T]. \tag{5.15c}$$

If moreover $\mathscr{E}$ is strictly convex then ρ^n converges uniformly in $[0, T]$ with respect to the total variation distance.

Proof Part *(2)* follows immediately from Proposition 4.21.

For part *(3)*, the three statements of (5.15) *as inequalities* $\leq$ follow from earlier results: for (5.15a) this follows again from Proposition 4.21, for (5.15b) from Proposition 5.3, and for (5.15c) from Lemma 2.3. Using these inequalities to pass to the limit in the equation $\mathscr{L}_T(\rho^n, j^n) = 0$ we obtain that $\mathscr{L}_T(\rho, j) \leq 0$. On the other hand, since $\mathscr{L}_T(\rho, j) \geq 0$ by the chain-rule estimate (5.5), standard arguments yield the equalities in (5.15).

When $\mathscr{E}$ is strictly convex, we obtain the convergence in $L^1(V, \pi)$ of the densities $u_t^n = d\rho_t^n/d\pi$ for every $t \in [0, T]$. We then use the equicontinuity estimate (4.75) of Proposition 4.21 to conclude uniform convergence of the sequence $(\rho_n)_n$ with respect to the total variation distance.

For part *(1)*, when the density u_0 of ρ_0 takes value in a compact interval $[a, b]$ with $0 < a < b < \infty$, the existence of a solution follows by Theorem 6.6 below. The general case follows by a standard approximation of u_0 by truncation and applying the stability properties of parts *(2)* and *(3)*.   □

5.4 Stationary states and attraction

Let us finally make a few comments on stationary measures and on the asymptotic behaviour of solutions of the $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ system. The definition of invariant measures was already given in Sect. 2.4, and we recall it for convenience.

Definition 5.11 (**Invariant and stationary measures**) Let $\rho = u\pi \in D(\mathscr{E})$ be given.

(1) We say that ρ is *invariant* if $\kappa_\rho(\mathrm{d}x\mathrm{d}y) = \rho(\mathrm{d}x)\kappa(x, \mathrm{d}y)$ has equal marginals, i.e. $\mathsf{x}_\#\kappa_\rho = \mathsf{y}_\#\kappa_\rho$.

(2) We say that ρ is *stationary* if the constant curve $\rho_t \equiv \rho$ is a solution of the $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ system.

Note that we always assume that π is invariant (see Assumption $(V\pi\kappa)$). It is immediate to check that

$$\rho \text{ is stationary} \iff \mathscr{D}(\rho) = 0 \iff \mathrm{D}_\phi(u(x), u(y)) = 0 \quad \vartheta\text{-a.e.} \tag{5.16}$$

If a measure ρ is invariant, then $u = \mathrm{d}\rho/\mathrm{d}\pi$ satisfies

$$u(x) = u(y) \quad \text{for } \vartheta\text{-a.e. } (x, y) \in E, \tag{5.17}$$

which implies (5.16); therefore invariant measures are stationary. Depending on the system, the set of stationary measures might also contain non-invariant measures, as the next example shows.

Example 5.12 Consider the example of the cosh-type dissipation (1.17a),

$$\alpha(u, v):=\sqrt{uv}, \quad \Psi^*(\xi):=4\big(\cosh(\xi/2) - 1\big),$$

but combine this with a Boltzmann entropy with an additional multiplicative constant $0 < \gamma \le 1$:

$$\phi(s):=\gamma(s \log s - s + 1).$$

The case $\gamma = 1$ corresponds to the example of (1.17a), and for general $0 < \gamma \le 1$ we find that

$$\mathrm{F}(u, v) = u^{\frac{1-\gamma}{2}} v^{\frac{1+\gamma}{2}} - u^{\frac{1+\gamma}{2}} v^{\frac{1-\gamma}{2}},$$

resulting in the evolution equation (see (1.12))

$$\partial_t u(x) = \int_{y\in V} \left[u(x)^{\frac{1-\gamma}{2}} u(y)^{\frac{1+\gamma}{2}} - u(x)^{\frac{1+\gamma}{2}} u(y)^{\frac{1-\gamma}{2}} \right] \kappa(x, \mathrm{d}y).$$

When $0 < \gamma < 1$, any function of the form $u(x) = \mathbb{1}\{x \in A\}$ for $A \subset V$ is a stationary point of this equation, and equivalently any measure $\pi \llcorner A$ is a stationary solution of the $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ system. For $0 < \gamma < 1$ therefore the set of stationary measures is much larger than just invariant measures.

As in the case of linear evolutions, $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ systems behave well with respect to decomposition of π into mutually singular invariant measures.

Theorem 5.13 (Decomposition) *Let us suppose that* $\pi = \pi^1 + \pi^2$ *with* $\pi^1, \pi^2 \in \mathcal{M}^+(V)$ *mutually singular and invariant. Let* $\rho : [0, T] \to \mathcal{M}^+(V)$ *be a curve with* $\rho_t = u_t\pi \ll \pi$ *and let* $\rho_t^i:=u_t\pi^i$ *be the decomposition of* ρ_t *with respect to* π^1 *and* π^2*. Then* ρ *is a solution of the* $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ *system if and only if each curve* ρ_t^i, $i = 1, 2$, *is a solution of the* $(\mathscr{E}^i, \mathscr{R}^i, (\mathscr{R}^i)^*)$ *system, where* $\mathscr{E}^i(\mu):=\mathscr{F}_\phi(\mu|\pi^i)$ *is the relative entropy with respect to the measures* π^i *and and* $\mathscr{R}^i, (\mathscr{R}^i)^*$ *are induced by* π^i.

Remark 5.14 It is worth noting that when α is 1-homogeneous then $\mathscr{R}^i = \mathscr{R}$ and $(\mathscr{R}^i)^* = \mathscr{R}^*$ do not depend on π^i, cf. Corollary 4.11. The decomposition is thus driven just by the splitting of the entropy $\mathscr{E}$.

Proof of Theorem 5.13 Note that the assumptions of invariance and mutual singularity of π^1 and π^2 imply that ϑ has a singular decomposition $\vartheta = \vartheta^1 + \vartheta^2 := \kappa_{\pi^1} + \kappa_{\pi^2}$, where the κ_{π^i} are symmetric. It then follows that $\mathscr{E}(\rho_t) = \mathscr{E}^1(\rho_t^1) + \mathscr{E}^2(\rho_t^2)$ and $\mathscr{D}(\rho_t) = \mathscr{D}^1(\rho_t^1) + \mathscr{D}^2(\rho_t^2)$, where

$$\mathscr{D}^i(\rho^i) = \frac{1}{2} \iint_E D_\phi(u(x), u(y)) \, \vartheta^i(\mathrm{d}x, \mathrm{d}y).$$

Finally, Corollary 4.11 shows that decomposing j as the sum $j^1 + j^2$ where $j^i \ll \vartheta^i$, the pairs (ρ^i, j^i) belong to $\mathcal{CE}(0, T)$ and $\mathscr{R}(\rho_t, j_t) = \mathscr{R}^1(\rho_t^1, j_t^1) + \mathscr{R}^2(\rho_t^2, j_t^2)$.   □

Theorem 5.15 (Asymptotic behaviour) *Let us suppose that the only stationary measures are multiples of π, and that $\mathscr{D}$ is lower semicontinuous with respect to setwise convergence. Then every solution $\rho : [0, \infty) \to \mathcal{M}^+(V)$ of the $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ evolution system converges setwise to $c\pi$, where $c := \rho_0(V)/\pi(V)$.*

Proof Let us fix a vanishing sequence $\tau_n \downarrow 0$ such that $\sum_n \tau_n = +\infty$. Let ρ_∞ be any limit point with respect to setwise convergence of the curve ρ_t along a diverging sequence of times $t_n \uparrow +\infty$. Such a point exists since the curve ρ is contained in a sublevel set of $\mathscr{E}$. Up to extracting a further subsequence, it is not restrictive to assume that $t_{n+1} \geq t_n + \tau_n$.

Since

$$\sum_{n \in \mathbb{N}} \int_{t_n}^{t_n + \tau_n} \left(\mathscr{R}(\rho_t, j_t) + \mathscr{D}(\rho_t) \right) \mathrm{d}t$$

$$\leq \int_0^{+\infty} \left(\mathscr{R}(\rho_t, j_t) + \mathscr{D}(\rho_t) \right) \mathrm{d}t \leq \mathscr{E}(\rho_0) < \infty$$

and the series of τ_n diverges, we find

$$\liminf_{n \to +\infty} \frac{1}{\tau_n} \int_{t_n}^{t_n + \tau_n} \mathscr{D}(\rho_t) \, \mathrm{d}t = 0, \quad \lim_{n \to \infty} \int_{t_n}^{t_n + \tau_n} \mathscr{R}(\rho_t, j_t) \, \mathrm{d}t = 0.$$

Up to extracting a further subsequence, we can suppose that the above lim inf is a limit and we can select $t_n' \in [t_n, t_n + \tau_n]$ such that

$$\lim_{n \to \infty} \mathscr{D}(\rho_{t_n'}) = 0, \quad \lim_{n \to \infty} \int_{t_n}^{t_n'} \mathscr{R}(\rho_t, j_t) \, \mathrm{d}t = 0.$$

Recalling the definition (4.80) of the Dynamical-Variational Transport cost and the monotonicity with respect to τ, we also get $\lim_{n \to \infty} \mathscr{W}(\tau_n, \rho_{t_n}, \rho_{t_n'}) = 0$, so that Theorem 4.26(5) and the relative compactness of the sequence $(\rho_{t_n'})_n$ yield $\rho_{t_n'} \to \rho_\infty$ setwise.

The lower semicontinuity of $\mathscr{D}$ yields $\mathscr{D}(\rho_\infty) = 0$ so that $\rho_\infty = c\pi$ thanks to the uniqueness assumption and to the conservation of the total mass. Since we have uniquely identified the limit point, we conclude that the *whole* curve ρ_t converges setwise to ρ_∞ as $t \to +\infty$.   □

6 Dissipative evolutions in $L^1(V, \pi)$

In this section we construct solutions of the $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ formulation by studying their equivalent characterization as abstract evolution equations in $L^1(V, \pi)$. Throughout this section we adopt Assumption $(V\pi\kappa)$.

6.1 Integro-differential equations in L^1

Let $J \subset \mathbb{R}$ be a closed interval (not necessarily bounded) and let us first consider a map $G : E \times J^2 \to \mathbb{R}$ with the following properties:

(1) measurability with respect to $(x, y) \in E$:

$$\text{for every } u, v \in J \text{ the map } (x, y) \mapsto G(x, y; u, v) \text{ is measurable;} \tag{6.1a}$$

(2) continuity with respect to u, v and linear growth: there exists a constant $M > 0$ such that

$$\text{for every } (x, y) \in E \quad (u, v) \mapsto G(x, y; u, v) \text{ is continuous and}$$
$$|G(x, y; u, v)| \leq M(1 + |u| + |v|) \quad \text{for every } u, v \in J, \tag{6.1b}$$

(3) skew-symmetry:

$$G(x, y; u, v) = -G(y, x; v, u), \quad \text{for every } (x, y) \in E, \ u, v \in J, \tag{6.1c}$$

(4) ℓ-dissipativity: there exists a constant $\ell \geq 0$ such that for every $(x, y) \in E, u, u', v \in J$:

$$u \leq u' \quad \Rightarrow \quad G(x, y; u', v) - G(x, y; u, v) \leq \ell(u' - u). \tag{6.1d}$$

Remark 6.1 Note that (6.1d) is surely satisfied if G is ℓ-Lipschitz in (u, v), uniformly with respect to (x, y). The 'one-sided Lipschitz condition' (6.1d) however is weaker than the standard Lipschitz condition; this type of condition is common in the study of ordinary differential equations, since it is still strong enough to guarantee uniqueness and non-blowup of the solutions (see e.g. [41, Ch. IV.12]).

Let us also remark that (6.1c) and (6.1d) imply the reverse monotonicity property of G with respect to v,

$$v \geq v' \quad \Rightarrow \quad G(x, y; u, v') - G(x, y; u, v) \leq \ell(v - v'), \tag{6.2}$$

and the joint estimate

$$u \leq u', \ v \geq v' \quad \Rightarrow \quad G(x, y; u', v') - G(x, y; u, v) \leq \ell\big[(u' - u) + (v - v')\big]. \tag{6.3}$$

Let us set $L^1(V, \pi; J) := \{u \in L^1(V, \pi) : u(x) \in J \text{ for } \pi\text{-a.e. } x \in V\}$.

Lemma 6.2 *Let $u : V \to J$ be a measurable π-integrable function.*

(1) We have

$$\int_V \big|G(x, y; u(x), u(y))\big| \kappa(x, \mathrm{d}y) < +\infty \ \text{ for } \pi\text{-a.e. } x \in V, \tag{6.4}$$

and the formula

$$\boldsymbol{G}[u](x) := \int_V G(x, y; u(x), u(y)) \kappa(x, \mathrm{d}y) \tag{6.5}$$

defines a function $\boldsymbol{G}[u]$ in $L^1(V, \pi)$ that only depends on the Lebesgue equivalence class of u in $L^1(V, \pi)$.

(2) The map $\boldsymbol{G} : L^1(V, \pi; J) \to L^1(V, \pi)$ is continuous.

(3) The map $\boldsymbol{G}$ is $(\ell \|\kappa_V\|_\infty)$-dissipative, in the sense that for all $h > 0$,

$$\big\|(u_1 - u_2) - h(\boldsymbol{G}[u_1] - \boldsymbol{G}[u_2])\big\|_{L^1(V, \pi)} \geq (1 - 2\ell\|\kappa_V\|_\infty h)\|u_1 - u_2\|_{L^1(V, \pi)} \tag{6.6}$$

for every $u_1, u_2 \in L^1(V, \pi; J)$.

(4) If $a \in J$ satisfies

$$0 = G(x, y; a, a) \leq G(x, y; a, v) \quad \text{for every } (x, y) \in E, \ v \geq a, \tag{6.7}$$

then for every function $u \in L^1(V, \pi; J)$ we have

$$u \geq a\pi \text{ -a.e.} \quad \Rightarrow \quad \lim_{h \downarrow 0} \frac{1}{h} \int_V \Big(a - (u + h\mathbf{G}[u]) \Big)_+ \mathrm{d}\pi = 0. \tag{6.8}$$

If $b \in J$ satisfies

$$0 = G(x, y; b, b) \geq G(x, y; b, v) \quad \text{for every } (x, y) \in E, \ v \leq b, \tag{6.9}$$

then for every function $u \in L^1(V, \pi; J)$ we have

$$u \leq b\pi \text{ -a.e.} \quad \Rightarrow \quad \lim_{h \downarrow 0} \frac{1}{h} \int_V \Big(u + h\mathbf{G}[u] - b \Big)_+ \mathrm{d}\pi = 0. \tag{6.10}$$

Proof *(1)* Since G is a Carathéodory function, for every measurable u and every $(x, y) \in E$ the map $(x, y) \mapsto G(x, y; u(x), u(y))$ is measurable. Since

$$\iint_E |G(x, y; u(x), u(y))| \, \kappa(x, \mathrm{d}y)\pi(\mathrm{d}x) = \iint_E |G(x, y; u(x), u(y))| \, \vartheta(\mathrm{d}x, \mathrm{d}y)$$
$$\leq M \|\kappa_V\|_\infty \Big(1 + 2 \int_V |u| \, \mathrm{d}\pi \Big), \tag{6.11}$$

the first claim follows by Fubini's Theorem [18, II, 14].

(2) Let $(u_n)_{n \in \mathbb{N}}$ be a sequence of functions strongly converging to u in $L^1(V, \pi; J)$. Up to extracting a further subsequence, it is not restrictive to assume that u_n also converges to u pointwise π-a.e. We have

$$\big\| \mathbf{G}[u_n] - \mathbf{G}[u] \big\|_{L^1(V,\pi)} = \iint_E \Big| G(x, y; u_n(x), u_n(y)) - G(x, y; u(x), u(y)) \Big| \, \vartheta(\mathrm{d}x, \mathrm{d}y). \tag{6.12}$$

Since the integrand g_n in (6.12) vanishes ϑ-a.e. in E as $n \to \infty$, by the generalized Dominated Convergence Theorem (see for instance [27, Thm. 4, page 21] it is sufficient to show that there exist positive functions h_n pointwise converging to h such that

$$g_n \leq h_n \ \vartheta\text{-a.e. in } E, \quad \lim_{n \to \infty} \iint_E h_n \, \mathrm{d}\vartheta = \iint_E h \, \mathrm{d}\vartheta.$$

We select $h_n(x, y) := M(2 + |u_n(x)| + |u_n(y)| + |u(x)| + |u(y)|)$ and $h(x, y) := 2M(1 + |u(x)| + |u(y)|)$. This proves the result.

(3) Let us set

$$\mathfrak{s}(r) := \begin{cases} 1 & \text{if } r > 0, \\ -1 & \text{if } r \leq 0, \end{cases}$$

and observe that the left-hand side of (6.6) may be estimated from below by

$$\big\| (u_1 - u_2) - h(\mathbf{G}[u_1] - \mathbf{G}[u_2]) \big\|_{L^1(V,\pi)} \geq \|u_1 - u_2\|_{L^1(V,\pi)}$$
$$- h \int_V \mathfrak{s}(u_1 - u_2)\big(\mathbf{G}[u_1] - \mathbf{G}[u_2]\big) \, \mathrm{d}\pi$$

for all $h > 0$. Therefore, estimate (6.6) follows if we prove that

$$\delta := \int_V \mathfrak{s}(u_1 - u_2)\big(\mathbf{G}[u_1] - \mathbf{G}[u_2]\big) \, \mathrm{d}\pi \leq 2\ell \|\kappa_V\|_\infty \|u_1 - u_2\|_{L^1(V,\pi)}. \tag{6.13}$$

Let us set

$$\Delta_G(x, y) := G(x, y; u_1(x), u_1(y)) - G(x, y; u_2(x), u_2(y)),$$

and

$$\Delta_{\mathfrak{s}}(x, y) := \mathfrak{s}(u_1(x) - u_2(x)) - \mathfrak{s}(u_1(y) - u_2(y)). \tag{6.14}$$

Since $\Delta_G(x, y) = -\Delta_G(y, x)$, using (6.1c) we have

$$\delta = \int_V \mathfrak{s}(u_1 - u_2)\left(G[u_1] - G[u_2]\right) d\pi = \iint_E \mathfrak{s}(u_1(x) - u_2(x))\Delta_G(x, y)\,\vartheta(dx, dy)$$

$$= \frac{1}{2}\iint_E \Delta_{\mathfrak{s}}(x, y)\Delta_G(x, y)\,\vartheta(dx, dy).$$

Setting $\Delta(x) := u_1(x) - u_2(x)$ we observe that by (6.3)

$$\Delta(x) > 0, \ \Delta(y) > 0 \ \Rightarrow \ \Delta_{\mathfrak{s}}(x, y) = 0,$$
$$\Delta(x) \le 0, \ \Delta(y) \le 0 \ \Rightarrow \ \Delta_{\mathfrak{s}}(x, y) = 0,$$
$$\Delta(x) \le 0, \ \Delta(y) > 0 \ \Rightarrow \ \Delta_{\mathfrak{s}}(x, y) = -2, \ \Delta_G(x, y) \ge -\ell\left(\Delta(y) - \Delta(x)\right)$$
$$\Delta(x) > 0, \ \Delta(y) \le 0 \ \Rightarrow \ \Delta_{\mathfrak{s}}(x, y) = 2, \ \Delta_G(x, y) \le \ell\left(\Delta(x) - \Delta(y)\right).$$

We deduce that

$$\delta \le \ell \iint_E \left[|u_1(x) - u_2(x)| + |u_1(y) - u_2(y)|\right]\vartheta(dx, dy) \le 2\ell\|\kappa_V\|_\infty \|u_1 - u_2\|_{L^1(V, \pi)}.$$

(4) We will only address the proof of property (6.8), as the argument for (6.10) is completely analogous. Suppose that $u \ge a$ π-a.e. Let us first observe that if $u(x) = a$, then from (6.7),

$$G[u](x) = \int_V G(x, y; a, u(y))\,\kappa(x, dy) \ge 0.$$

We set $f_h(x) := h^{-1}(a - u(x)) - G[u](x)$, observing that $f_h(x)$ is monotonically decreasing to $-\infty$ if $u(x) > a$ and $f_h(x) = -G[u](x) \le 0$ if $u(x) = a$, so that $\lim_{h\downarrow 0}\left(f_h(x)\right)_+ = 0$. Since $\left(f_h\right)_+ \le \left(-G[u]\right)_+$ we can apply the Dominated Convergence Theorem to obtain

$$\lim_{h\downarrow 0}\int_V \left(f_h(x)\right)_+ \pi(dx) = 0,$$

thereby concluding the proof.   □

In what follows, we shall address the Cauchy problem

$$\dot{u}_t = G[u_t] \ \text{ in } L^1(V, \pi) \text{ for every } t \ge 0, \tag{6.15a}$$

$$u|_{t=0} = u_0. \tag{6.15b}$$

Lemma 6.3 (Comparison principles) *Let us suppose that the map* G *satisfies* (6.1a,b,c) *with* $J = \mathbb{R}$.

(1) If $\bar{u} \in \mathbb{R}$ *satisfies*

$$0 = G(x, y; \bar{u}, \bar{u}) \le G(x, y; \bar{u}, v) \ \text{ for every } (x, y) \in E, \ v \ge \bar{u}, \tag{6.16}$$

then for every initial datum $u_0 \ge \bar{u}$ *the solution* u *of* (6.15) *satisfies* $u_t \ge \bar{u}$ π-a.e. *for every* $t \ge 0$.

(2) If $\bar{u} \in \mathbb{R}$ satisfies

$$0 = G(x, y; \bar{u}, \bar{u}) \geq G(x, y; \bar{u}, v) \quad \text{for every } (x, y) \in E, \ v \leq \bar{u}, \tag{6.17}$$

then for every initial datum $u_0 \leq \bar{u}$ the solution u of (6.15) satisfies $u_t \leq \bar{u}$ π-a.e. for every $t \geq 0$.

Proof *(1)* Let us first consider the case $\bar{u} = 0$. We define a new map $\overline{G}$ by symmetry:

$$\overline{G}(x, y; u, v) := G(x, y; u, |v|) \tag{6.18}$$

which satisfies the same structural properties (6.1a,b,c), and moreover

$$0 = \overline{G}(x, y; 0, 0) \leq \overline{G}(x, y; 0, v) \quad \text{for every } x, y \in V, \ v \in \mathbb{R}. \tag{6.19}$$

We call $\overline{\boldsymbol{G}}$ the operator induced by $\overline{G}$, and $\bar{u}$ the solution curve of the corresponding Cauchy problem starting from the same (nonnegative) initial datum u_0. If we prove that $\bar{u}_t \geq 0$ for every $t \geq 0$, then $\bar{u}_t$ is also the unique solution of the original Cauchy problem (6.15) induced by G, so that we obtain the positivity of u_t.

Note that (6.19) and property (6.1d) yield

$$\overline{G}(x, y; u, v) \geq \overline{G}(x, y; u, v) - \overline{G}(x, y; 0, v) \geq \ell u \quad \text{for } u \leq 0. \tag{6.20}$$

We set $\beta(r) := r_- = \max(0, -r)$ and $P_t := \{x \in V : \bar{u}_t(x) < 0\}$ for each $t \geq 0$. Due to the Lipschitz continuity of β, the map $t \mapsto b(t) := \int_V \beta(\bar{u}_t) \, d\pi$ is absolutely continuous. Hence, the chain-rule formula applies, which, together with (6.20) gives

$$\frac{d}{dt} b(t) = - \int_{P_t} \overline{\boldsymbol{G}}[\bar{u}_t](x) \, \pi(dx) = - \iint_{P_t \times V} \overline{G}(x, y; \bar{u}_t(x), \bar{u}_t(y)) \, \boldsymbol{\vartheta}(dx, dy)$$

$$\leq \ell \iint_{P_t \times V} (-\bar{u}_t(x)) \, \boldsymbol{\vartheta}(dx, dy) = \ell \iint_E \beta(\bar{u}_t(x)) \, \boldsymbol{\vartheta}(dx, dy) \leq \ell \|\kappa_V\|_\infty b(t).$$

Since b is nonnegative and $b(0) = 0$, we conclude, by Gronwall's inequality, that $b(t) = 0$ for every $t \geq 0$ and therefore $\bar{u}_t \geq 0$. In order to prove the the statement for a general $\bar{u} \in \mathbb{R}$ it is sufficient to consider the new operator $\widetilde{G}(x, y; u, v) := G(x, y; u + \bar{u}, v + \bar{u})$, and to consider the curve $\widetilde{u}_t := u_t - \bar{u}$ starting from the nonnegative initial datum $\widetilde{u}_0 := u_0 - \bar{u}$.
(2) It suffices to apply the transformation $\widetilde{G}(x, y; u, v) := -G(x, y; -u, -v)$ and set $\widetilde{u}_t := -u_t$. We then apply the previous claim, yielding the lower bound $-\bar{u}$. $\qquad\square$

We can now state our main result concerning the well-posedness of the Cauchy problem (6.15).

Theorem 6.4 *Let $J \subset \mathbb{R}$ be a closed interval of $\mathbb{R}$ and let $G : E \times J^2 \to \mathbb{R}$ be a map satisfying conditions (6.1). Let us also suppose that, if $a = \inf J > -\infty$ then (6.7) holds, and that, if $b = \sup J < +\infty$ then (6.9) holds.*

(1) For every $u_0 \in L^1(V, \pi; J)$ there exists a unique curve $u \in C^1([0, \infty); L^1(V, \pi; J))$ solving the Cauchy problem (6.15).
(2) $\int_V u_t \, d\pi = \int_V u_0 \, d\pi$ for every $t \geq 0$.
(3) If u, v are two solutions with initial data $u_0, v_0 \in L^1(V, \pi; J)$ respectively, then

$$\|u_t - v_t\|_{L^1(V, \pi)} \leq e^{2\|\kappa_V\|_\infty \ell t} \|u_0 - v_0\|_{L^1(V, \pi)} \quad \text{for every } t \geq 0. \tag{6.21}$$

(4) If $\bar{a} \in J$ satisfies condition (6.7) and $u_0 \geq \bar{a}$, then $u_t \geq \bar{a}$ for every $t \geq 0$. Similarly, if $\bar{b} \in J$ satisfies condition (6.9) and $u_0 \leq \bar{b}$, then $u_t \leq \bar{b}$ for every $t \geq 0$.

(5) If $\ell = 0$, then the evolution is order preserving: if u, v are two solutions with initial data u_0, v_0 then

$$u_0 \leq v_0 \quad \Rightarrow \quad u_t \leq v_t \quad \text{for every } t \geq 0. \tag{6.22}$$

Proof Claims *(1), (3), (4)* follow by the abstract generation result of [50, §6.6, Theorem 6.1] applied to the operator G defined in the closed convex subset $D := L^1(V, \pi; J)$ of the Banach space $L^1(V, \pi)$. For the theorem to apply, one has to check the continuity of $G : D \to L^1(V, \pi)$ (Lemma 6.2(2)), its dissipativity (6.6), and the property

$$\liminf_{h \downarrow 0} h^{-1} \inf_{v \in D} \|u + hG[u] - v\|_{L^1(V, \pi)} = 0 \quad \text{for every } u \in D \,.$$

When $J = \mathbb{R}$, the inner infimum always is zero; if J is a bounded interval $[a, b]$ then the property above follows from the estimates of Lemma 6.2(4), since for any $u \in D$,

$$\inf_{v \in D} \int_V |u + hG[u] - v| \, \mathrm{d}\pi \leq \int_V \Big(a - (u + hG[u])\Big)_+ \mathrm{d}\pi + \int_V \Big(u + hG[u] - b\Big)_+ \mathrm{d}\pi \,.$$

When $J = [a, \infty)$ or $J = (-\infty, b]$ a similar reasoning applies.

Claim *(2)* is an immediate consequence of (6.1c). Finally, when $\ell = 0$, claim *(5)* follows from the Crandall-Tartar Theorem [14], stating that a non-expansive map in L^1 (cf. (6.21)) that satisfies claim *(2)* is also order preserving. □

6.2 Applications to dissipative evolutions

Let us now consider the map $\mathrm{F} : (0, +\infty)^2 \to \mathbb{R}$ induced by the system (Ψ^*, ϕ, α), first introduced in (1.11),

$$\mathrm{F}(u, v) := (\Psi^*)'\big(\phi'(v) - \phi'(u)\big)\,\alpha(u, v) \quad \text{for every } u, v > 0, \tag{6.23}$$

with the corresponding integral operator:

$$\boldsymbol{F}[u](x) := \int_V \mathrm{F}(u(x), u(y))\,\kappa(x, \mathrm{d}y)\,. \tag{6.24}$$

Since Ψ^*, ϕ are C^1 convex functions on $(0, +\infty)$ and α is locally Lipschitz in $(0, +\infty)^2$ it is easy to check that F satisfies properties (6.1a,b,c,d) in every compact subset $J \subset (0, +\infty)$ and conditions (6.7), (6.9) at every point $a, b \in J$. In order to focus on the structural properties of the associated evolution problem, cf. (6.28) below, we will mostly confine our analysis to the regular case, according to the following:

Assumption (F). The map F defined by (6.23) satisfies the following properties:

$$F \text{ admits a continuous extension to } [0, \infty), \quad (6.25)$$

and for every $R > 0$ there exists $\ell_R \geq 0$ such that

$$v \leq v' \quad \Rightarrow \quad F(u, v) - F(u, v') \leq \ell_R (v' - v) \quad \text{for every } u, v, v' \in [0, R]. \quad (6.26)$$

If moreover (6.26) is satisfied in $[0, +\infty)$ for some constant $\ell_\infty \geq 0$ and there exists a constant M such that

$$|F(u, v)| \leq M(1 + u + v) \quad \text{for every } u, v \geq 0, \quad (6.27)$$

we say that (F_∞) holds.

Note that (6.25) is always satisfied if ϕ is differentiable at 0. Estimate (6.26) is also true if in addition α is Lipschitz. However, as we have shown in Sect. 1.3, there are important examples in which $\phi'(0) = -\infty$, but (6.25) and (6.26) hold nonetheless. All of the examples given in Sect. 1.3 indeed provide families of maps F that satisfy Assumption (F).

Theorem 6.4 yields the following general result:

Theorem 6.5 *Consider the Cauchy problem*

$$\dot{u}_t = F[u_t] \quad t \geq 0, \quad u_{|t=0} = u_0. \quad (6.28)$$

for a given nonnegative $u_0 \in L^1(V, \pi)$.

(1) For every $u_0 \in L^1(V, \pi; J)$ with J a compact subinterval of $(0, +\infty)$ there exists a unique bounded and nonnegative solution $u \in C^1([0, \infty); L^1(V, \pi; J))$ of (6.28). We will denote by $(S_t)_{t \geq 0}$ the corresponding C^1-semigroup of nonlinear operators, mapping u_0 to the value $u_t = S_t[u_0]$ at time t of the solution u.
(2) $\int_V u_t \, d\pi = \int_V u_0 \, d\pi$ for every $t \geq 0$.
(3) If $a \leq u_0 \leq b$ π-a.e. in V, then $a \leq u_t \leq b$ π-a.e. for every $t \geq 0$.
(4) The solution satisfies the Lipschitz estimate (6.21) (with $\ell = \ell_R$) and the order preserving property if $\ell_R = 0$.
(5) If Assumption (F) holds, then $(S_t)_{t \geq 0}$ can be extended to a semigroup defined on every essentially bounded nonnegative $u_0 \in L^1(V, \pi)$ and satisfying the same properties (1)–(4) above.
(6) If additionally (F_∞) holds, then $(S_t)_{t \geq 0}$ can be extended to a semigroup defined on every nonnegative $u_0 \in L^1(V, \pi)$ and satisfying the same properties (1)–(4) above.

We now show that the solution u given by Theorem 6.5 is also a solution in the sense of the $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ Energy-Dissipation balance.

Theorem 6.6 *Assume $(V\pi\kappa)$, $(\mathscr{R}^*\Psi\alpha)$, $(\mathscr{E}\phi)$. Let $u_0 \in L^1(V; \pi)$ be nonnegative and π-essentially valued in a compact interval J of $(0, \infty)$ and let $u = S[u_0] \in C^1([0, +\infty);$ $L^1(V, \pi; J))$ be the solution to (6.28) given by Theorem 6.5. Then the pair (ρ, j) given by*

$$\rho_t(dx) := u_t(x)\pi(dx),$$
$$2j_t(dx \, dy) := w_t(x, y) \, \vartheta(dx \, dy), \qquad w_t(x, y) := -F(u_t(x), u_t(y)),$$

is an element of $\mathcal{CE}(0, +\infty)$ and satisfies the $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ Energy-Dissipation balance (5.4).

If F *satisfies the stronger assumption* (F)*, then the same result holds for every essentially bounded and nonnegative initial datum. Finally, if also* (F$_\infty$) *holds, the above result is valid for every nonnegative* $u_0 \in L^1(V, \pi)$ *with* $\rho_0 = u_0\pi \in D(\mathscr{E})$.

Proof Let us first consider the case when u_0 satisfies $0 < a \le u_0 \le b < +\infty$ π-a.e.. Then, the solution $u = \mathsf{S}[u_0]$ satisfies the same bounds, the map w_t is uniformly bounded and $\alpha(u_t(x), u_t(y)) \ge \alpha(a, a) > 0$, so that $(\rho, \boldsymbol{j}) \in \mathcal{A}(0, T)$. We can thus apply Theorem 5.7, obtaining the Energy-Dissipation balance

$$\mathscr{E}(\rho_0) - \mathscr{E}(\rho_T) = \int_0^T \mathscr{R}(\rho_t, \boldsymbol{j}_t)\, \mathrm{d}t + \int_0^T \mathscr{D}(\rho_t)\, \mathrm{d}t, \quad \text{or equivalently} \quad \mathscr{L}(\rho, \boldsymbol{j}) = 0. \tag{6.29}$$

In the case $0 \le u_0 \le b$ we can argue by approximation, setting $u_0^a := \max\{u_0, a\}$, $a > 0$, and considering the solution $u_t^a := \mathsf{S}_t[u_0^a]$ with divergence field $2\boldsymbol{j}_t^a(\mathrm{d}x, \mathrm{d}y) = -\mathrm{F}(u_t^a(x), u_t^a(y))\boldsymbol{\vartheta}(\mathrm{d}x, \mathrm{d}y)$. Theorem 6.5(4) shows that $u_t^a \to u_t$ strongly in $L^1(V, \pi)$ as $a \downarrow 0$, and consequently also $\boldsymbol{j}_\lambda^a \to \boldsymbol{j}_\lambda$ setwise. Hence, we can pass to the limit in (6.29) (written for $(\rho^a, \boldsymbol{j}^a)$ thanks to Proposition 4.21 and Proposition 5.3), obtaining $\mathscr{L}(\rho, \boldsymbol{j}) \le 0$, which is still sufficient to conclude that $(\rho, \boldsymbol{j})$ is a solution thanks to Remark 5.5(3).

Finally, if (F$_\infty$) holds, we obtain the general result by a completely analogous argument, approximating u_0 by the sequence $u_0^b := \min\{u_0, b\}$ and letting $b \uparrow +\infty$. $\qquad\square$

7 Existence via minimizing movements

In this section we construct solutions to the $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ formulation via the *Minimizing Movement* approach. The method uses only fairly general properties of $\mathscr{W}$, $\mathscr{E}$, and the underlying space, and it may well have broader applicability than the measure-space setting that we consider here (see Remark 7.8). Therefore we formulate the results in a slightly more general setup.

We consider a topological space

$$(X, \sigma) = \mathcal{M}^+(V) \text{ endowed with the setwise topology.} \tag{7.1}$$

For consistency with the above definition, in this section we will use use the abstract notation $\overset{\sigma}{\rightharpoonup}$ to denote setwise convergence in $X = \mathcal{M}^+(V)$. Although throughout this paper we adopt the Assumptions $(V\pi\kappa)$, $(\mathscr{R}^*\Psi\alpha)$, and $(\mathscr{E}\phi)$, in this chapter we will base the discussion only on the following properties:

Assumption (Abs).

(1) the Dynamical-Variational Transport (DVT) cost $\mathscr{W}$ enjoys properties (4.86);
(2) the driving functional $\mathscr{E}$ enjoys the typical lower-semicontinuity and coercivity properties underlying the variational approach to gradient flows:

$$\mathscr{E} \ge 0 \quad \text{and} \quad \mathscr{E} \text{ is } \sigma\text{-sequentially lower semicontinuous;} \tag{7.2a}$$

$$\exists \rho^* \in X \quad \text{such that} \quad \forall \tau > 0, \tag{7.2b}$$

the map $\rho \mapsto \mathscr{W}(\tau, \rho^*, \rho) + \mathscr{E}(\rho)$ has σ-sequentially compact sublevels.

Assumption **(Abs)** is implied by Assumptions **(V$\pi\kappa$)**, **($\mathscr{R}^*\Psi\alpha$)**, and **($\mathscr{E}\phi$)**. The properties (4.86) are the content of Theorem 4.26; condition (7.2a) follows from Assumption **($\mathscr{E}\phi$)** and Lemma 5.3; condition (7.2b) follows from the superlinearity of ϕ at infinity and Prokhorov's characterization of compactness in the space of finite measures [9, Th. 8.6.2].

7.1 The minimizing movement scheme and the convergence result

The classical 'Minimizing Movement' scheme for metric-space gradient flows [4,17] starts by defining approximate solutions through incremental minimization,

$$\rho^n \in \operatorname*{argmin}_{\rho} \left(\frac{1}{2\tau} d(\rho^{n-1}, \rho)^2 + \mathscr{E}(\rho) \right).$$

In the context of this paper the natural generalization of the expression to be minimized is $\mathscr{W}(\tau, \rho^{n-1}, \rho) + \mathscr{E}(\rho)$. This can be understood by remarking that if $\mathscr{R}(\rho, \cdot)$ is quadratic, then it formally generates a metric

$$\frac{1}{2} d(\mu, \nu)^2 = \inf \left\{ \int_0^1 \mathscr{R}(\rho_t, \boldsymbol{j}_t) \, dt \; : \; \partial_t \rho_t + \overline{\operatorname{div}} \boldsymbol{j}_t = 0, \; \rho_0 = \mu, \text{ and } \rho_1 = \nu \right\}$$

$$= \tau \inf \left\{ \int_0^\tau \mathscr{R}(\rho_t, \boldsymbol{j}_t) \, dt \; : \; \partial_t \rho_t + \overline{\operatorname{div}} \boldsymbol{j}_t = 0, \; \rho_0 = \mu, \text{ and } \rho_\tau = \nu \right\}$$

$$= \tau \mathscr{W}(\tau, \mu, \nu).$$

In this section we set up the approximation scheme featuring the cost $\mathscr{W}$.

We consider a partition $\{t_\tau^0 = 0 < t_\tau^1 < \ldots < t_\tau^n < \ldots < t_\tau^{N_\tau - 1} < T \leq t_\tau^{N_\tau}\}$, with fineness $\tau := \max_{i=n,\ldots,N_\tau} (t_\tau^n - t_\tau^{n-1})$, of the time interval $[0, T]$. The sequence of approximations $(\rho_\tau^n)_n$ is defined by the following recursive minimization scheme. Fix $\rho^\circ \in X$.

Problem 7.1 *Given $\rho_\tau^0 := \rho^\circ$, find $\rho_\tau^1, \ldots, \rho_\tau^{N_\tau} \in X$ fulfilling*

$$\rho_\tau^n \in \operatorname*{argmin}_{v \in X} \left\{ \mathscr{W}(t_\tau^n - t_\tau^{n-1}, \rho_\tau^{n-1}, v) + \mathscr{E}(v) \right\} \quad \text{for } n = 1, \ldots, N_\tau. \tag{7.3}$$

Lemma 7.2 *Under assumption **(Abs)**, for any $\tau > 0$ Problem 7.1 admits a solution $\{\rho_\tau^n\}_{n=1}^{N_\tau} \subset X$.*

We denote by $\overline{\rho}_\tau$ and $\underline{\rho}_\tau$ the left-continuous and right-continuous piecewise constant interpolants of the values $\{\rho_\tau^n\}_{n=1}^{N_\tau}$ on the nodes of the partition, fulfilling $\overline{\rho}_\tau(t_\tau^n) = \underline{\rho}_\tau(t_\tau^n) = \rho_\tau^n$ for all $n = 1, \ldots, N_\tau$, i.e.,

$$\overline{\rho}_\tau(t) = \rho_\tau^n \quad \forall t \in (t_\tau^{n-1}, t_\tau^n], \qquad \underline{\rho}_\tau(t) = \rho_\tau^{n-1} \quad \forall t \in [t_\tau^{n-1}, t_\tau^n), \quad n = 1, \ldots, N_\tau. \tag{7.4}$$

Likewise, we denote by $\overline{t}_\tau$ and $\underline{t}_\tau$ the piecewise constant interpolants $\overline{t}_\tau(0) := \underline{t}_\tau(0) := 0$, $\overline{t}_\tau(T) := \underline{t}_\tau(T) := T$, and

$$\overline{t}_\tau(t) = t_\tau^n \quad \forall t \in (t_\tau^{n-1}, t_\tau^n], \qquad \underline{t}_\tau(t) = t_\tau^{n-1} \quad \forall t \in [t_\tau^{n-1}, t_\tau^n). \tag{7.5}$$

We also introduce another notion of interpolant of the discrete values $\{\rho_\tau^n\}_{n=0}^{N_\tau}$ introduced by De Giorgi, namely the *variational interpolant* $\widetilde{\rho}_\tau : [0, T] \to X$, which is defined in the

following way: the map $t \mapsto \widetilde{\rho}_\tau(t)$ is Lebesgue measurable in $(0, T)$ and satisfies

$$
\begin{cases}
\widetilde{\rho}_\tau(0) = \rho^\circ, \quad \text{and, for } t = t_\tau^{n-1} + r \in (t_\tau^{n-1}, t_\tau^n], \\
\widetilde{\rho}_\tau(t) \in \underset{\mu \in X}{\operatorname{argmin}} \left\{ \mathscr{W}(r, \rho_\tau^{n-1}, \mu) + \mathscr{E}(\mu) \right\}
\end{cases}
\tag{7.6}
$$

The existence of a measurable selection is guaranteed by [15, Cor. III.3, Thm. III.6].

It is natural to introduce the following extension of the notion of *(Generalized) Minimizing Movement*, which is typically given in a metric setting [4,5]. For simplicity, we will continue to use the classical terminology.

Definition 7.3 We say that a curve $\rho : [0, T] \to X$ is a *Generalized Minimizing Movement* for the energy functional $\mathscr{E}$ starting from the initial datum $\rho^\circ \in D(\mathscr{E})$, if there exist a sequence of partitions with fineness $(\tau_k)_k$, $\tau_k \downarrow 0$ as $k \to \infty$, and, correspondingly, a sequence of discrete solutions $(\overline{\rho}_{\tau_k})_k$ such that, as $k \to \infty$,

$$
\overline{\rho}_{\tau_k}(t) \overset{\sigma}{\rightharpoonup} \rho(t) \qquad \text{for all } t \in [0, T].
\tag{7.7}
$$

We shall denote by $\mathrm{GMM}(\mathscr{E}, \mathscr{W}; \rho^\circ)$ the collection of all Generalized Minimizing Movements for $\mathscr{E}$ starting from ρ°.

We can now state the main result of this section.

Theorem 7.4 *Under **Assumptions** ($V\pi\kappa$), ($\mathscr{R}^*\Psi\alpha$), and ($\mathscr{E}\phi$), let the lower-semicontinuity Property* (5.2) *be satisfied.*

Then $\mathrm{GMM}(\mathscr{E}, \mathscr{W}; (0, T), \rho^\circ) \neq \emptyset$ *and every* $\rho \in \mathrm{GMM}(\mathscr{E}, \mathscr{W}; (0, T), \rho^\circ)$ *satisfies the* $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ *Energy-Dissipation balance (Definition 5.4).*

Throughout Sects. 7.2–7.4 we will first prove an abstract version of this theorem as Theorem 7.7 below, under **Assumption (Abs)**. Indeed, therein we could 'move away' from the context of the 'concrete' gradient structure for the Markov processes, and carry out our analysis in a general topological setup (cf. Remark 7.8 ahead). In Sect. 7.5 we will 'return' to the problem under consideration and deduce the proof of Theorem 7.4 from Theorem 7.7.

7.2 Moreau–Yosida approximation and generalized slope

Preliminarily, let us observe some straightforward consequences of the properties of the transport cost:

(1) the 'generalized triangle inequality' from (4.86b) entails that for all $m \in \mathbb{N}$, for all $(m + 1)$-ples $(t, t_1, \ldots, t_m) \in (0, +\infty)^{m+1}$, and all $(\rho_0, \rho_1, \ldots, \rho_m) \in X^{m+1}$, we have

$$
\mathscr{W}(t, \rho_0, \rho_m) \leq \sum_{k=1}^m \mathscr{W}(t_k, \rho_{k-1}, \rho_k) \qquad \text{if } t = \sum_{k=1}^m t_k.
\tag{7.8}
$$

(2) Combining (4.86a) and (4.86b) we deduce that

$$
\mathscr{W}(t, \rho, \mu) \leq \mathscr{W}(s, \rho, \mu) \quad \text{for all } 0 < s < t \text{ and for all } \rho, \mu \in X.
\tag{7.9}
$$

In the context of metric gradient-flow theory, the 'Moreau-Yosida approximation' (see e.g. [11, Ch. 7] or [4, Def. 3.1.1]) provides an approximation of the driving functional that is finite and sub-differentiable everywhere, and can be used to define a generalized slope. We now construct the analogous objects in the situation at hand.

Given $r > 0$ and $\rho \in X$, we define the subset $J_r(\rho) \subset X$ by

$$J_r(\rho) := \operatorname*{argmin}_{\mu \in X} \left\{ \mathscr{W}(r, \rho, \mu) + \mathscr{E}(\mu) \right\}$$

(by Lemma 7.2, this set is non-empty) and define

$$\mathscr{E}_r(\rho) := \inf_{\mu \in X} \left\{ \mathscr{W}(r, \rho, \mu) + \mathscr{E}(\mu) \right\} = \mathscr{W}(r, \rho, \rho_r) + \mathscr{E}(\rho_r) \quad \forall \, \rho_r \in J_r(\rho). \tag{7.10}$$

In addition, for all $\rho \in D(\mathscr{E})$, we define the *generalized slope*

$$\mathscr{S}(\rho) := \limsup_{r \downarrow 0} \frac{\mathscr{E}(\rho) - \mathscr{E}_r(\rho)}{r} = \limsup_{r \downarrow 0} \frac{\sup_{\mu \in X} \left\{ \mathscr{E}(\rho) - \mathscr{W}(r, \rho, \mu) - \mathscr{E}(\mu) \right\}}{r}. \tag{7.11}$$

Recalling the *duality formula* for the local slope (cf. [4, Lemma 3.15]) and the fact that $\mathscr{W}(\tau, \cdot, \cdot)$ is a proxy for $\frac{1}{2\tau} d^2(\cdot, \cdot)$, it is immediate to recognize that the generalized slope is a surrogate of the local slope. Furthermore, as we will see that its definition is somehow tailored to the validity of Lemma 7.5 ahead. Heuristically, the generalized slope $\mathscr{S}(\rho)$ coincides with the Fisher information $\mathscr{D}(\rho) = \mathscr{R}^*(\rho, -D\mathscr{E}(\rho))$. This can be recognized, again heuristically, by fixing a point ρ_0 and considering curves $\rho_t := \rho_0 - t \overline{\operatorname{div}} \, \boldsymbol{j}$, for a class of fluxes $\boldsymbol{j}$. We then calculate

$$\mathscr{R}^*(\rho_0, -D\mathscr{E}(\rho_0)) = \sup_{\boldsymbol{j}} \left\{ -D\mathscr{E}(\rho_0) \cdot \boldsymbol{j} - \mathscr{R}(\rho_0, \boldsymbol{j}) \right\}$$

$$= \sup_{\boldsymbol{j}} \lim_{r \to 0} \frac{1}{r} \left\{ \mathscr{E}(\rho_0) - \mathscr{E}(\rho_r) - \int_0^r \mathscr{R}(\rho_t, \boldsymbol{j}) \, dt \right\}.$$

In Theorem 7.9 below we rigorously prove that $\mathscr{S} \geq \mathscr{D}$ using this approach.

The following result collects some properties of $\mathscr{E}_r$ and $\mathscr{S}$.

Lemma 7.5 *For all $\rho \in D(\mathscr{E})$ and for every selection $\rho_r \in J_r(\rho)$*

$$\mathscr{E}_{r_2}(\rho) \leq \mathscr{E}_{r_1}(\rho) \leq \mathscr{E}(\rho) \quad \text{for all } 0 < r_1 < r_2; \tag{7.12}$$

$$\rho_r \overset{\sigma}{\to} \rho \text{ as } r \downarrow 0, \quad \mathscr{E}(\rho) = \lim_{r \downarrow 0} \mathscr{E}_r(\rho); \tag{7.13}$$

$$\frac{d}{dr} \mathscr{E}_r(\rho) \leq -\mathscr{S}(\rho_r) \quad \text{for a.e. } r > 0. \tag{7.14}$$

In particular, for all $\rho \in D(\mathscr{E})$

$$\mathscr{S}(\rho) \geq 0 \quad \text{and} \tag{7.15}$$

$$\mathscr{W}(r_0, \rho, \rho_{r_0}) + \int_0^{r_0} \mathscr{S}(\rho_r) \, dr \leq \mathscr{E}(\rho) - \mathscr{E}(\rho_{r_0}) \tag{7.16}$$

for every $r_0 > 0$ and $\rho_{r_0} \in J_{r_0}(\rho)$.

Proof Let $r > 0$, $\rho \in D(\mathscr{E})$, and $\rho_r \in J_r(\rho)$. It follows from (7.10) and (4.86a) that

$$\mathscr{E}_r(\rho) = \mathscr{W}(r, \rho, \rho_r) + \mathscr{E}(\rho_r) \leq \mathscr{W}(r, \rho, \rho) + \mathscr{E}(\rho) = \mathscr{E}(\rho) \quad \forall r > 0, \rho \in X; \tag{7.17}$$

in the same way, one checks that for all $\rho \in X$ and $0 < r_1 < r_2$,

$$\mathscr{E}_{r_2}(\rho) - \mathscr{E}_{r_1}(\rho) \leq \mathscr{W}(r_2, \rho_{r_1}, \rho) + \mathscr{E}(\rho_{r_1}) - \mathscr{W}(r_1, \rho_{r_1}, \rho) - \mathscr{E}(\rho_{r_1}) \overset{(7.9)}{\leq} 0,$$

which implies (7.12). Thus, the map $r \mapsto \mathscr{E}_r(\rho)$ is non-increasing on $(0, +\infty)$, and hence almost everywhere differentiable. Let us fix a point of differentiability $r > 0$. For $h > 0$ and $\rho_r \in J_r(\rho)$ we then have

$$\frac{\mathscr{E}_{r+h}(\rho) - \mathscr{E}_r(\rho)}{h} = \frac{1}{h} \inf_{v \in X} \left\{ \mathscr{W}(r + h, \rho, v) + \mathscr{E}(v) - \mathscr{W}(r, \rho, \rho_r) - \mathscr{E}(\rho_r) \right\}$$

$$\leq \frac{1}{h} \inf_{v \in X} \left\{ \mathscr{W}(h, \rho_r, v) + \mathscr{E}(v) - \mathscr{E}(\rho_r) \right\},$$

the latter inequality due to (4.86b), so that

$$\frac{\mathrm{d}}{\mathrm{d}r} \mathscr{E}_r(\rho) \leq \liminf_{h \downarrow 0} \frac{1}{h} \inf_{v \in X} \left\{ \mathscr{W}(h, \rho_r, v) + \mathscr{E}(v) - \mathscr{E}(\rho_r) \right\}$$

$$= -\limsup_{h \downarrow 0} \frac{1}{h} \sup_{v \in X} \left\{ -\mathscr{W}(h, \rho_r, v) - \mathscr{E}(v) + \mathscr{E}(\rho_r) \right\},$$

whence (7.14). Finally, (7.17) yields that, for any $\rho \in \mathrm{D}(\mathscr{E})$ and any selection $\rho_r \in J_r(\rho)$, one has $\sup_{r>0} \mathscr{W}(r, \rho, \rho_r) < +\infty$. Therefore, (4.86d) entails the first convergence in (7.13). Furthermore, we have

$$\mathscr{E}(\rho) \geq \limsup_{r \downarrow 0} \mathscr{E}_r(\rho) \geq \liminf_{r \downarrow 0} (\mathscr{W}(r, \rho, \rho_r) + \mathscr{E}(\rho_r)) \geq \liminf_{r \downarrow 0} \mathscr{E}(\rho_r) \geq \mathscr{E}(\rho),$$

where the first inequality again follows from (7.17), and the last one from the σ-lower semicontinuity of $\mathscr{E}$. This implies the second statement of (7.13). □

7.3 *A priori* estimates

Our next result collects the basic estimates on the discrete solutions. In order to properly state it, we need to introduce the 'density of dissipated energy' associated with the interpolant $\overline{\rho}_\tau$, namely the piecewise constant function $\overline{W}_\tau : [0, T] \to [0, +\infty)$ defined by

$$\overline{W}_\tau(t) := \frac{\mathscr{W}(t_\tau^n - t_\tau^{n-1}, \rho_\tau^{n-1}, \rho_\tau^n)}{t_\tau^n - t_\tau^{n-1}} \quad t \in (t_\tau^{n-1}, t_\tau^n], \quad n = 1, \ldots, N_\tau,$$

so that $\displaystyle \int_{t_\tau^{j-1}}^{t_\tau^n} \overline{W}_\tau(t) \, \mathrm{d}t = \sum_{k=j}^n \mathscr{W}(t_\tau^k - t_\tau^{k-1}, \rho_\tau^{k-1}, \rho_\tau^k)$ for all $1 \leq j < n \leq N_\tau$. (7.18)

Proposition 7.6 (Discrete energy-dissipation inequality and a priori estimates) *We have*

$$\mathscr{W}(t - \underline{t}_\tau(t), \underline{\rho}_\tau(t), \widetilde{\rho}_\tau(t))$$

$$+ \int_{\underline{t}_\tau(t)}^t \mathscr{S}(\widetilde{\rho}_\tau(r)) \, \mathrm{d}r + \mathscr{E}(\widetilde{\rho}_\tau(t)) \leq \mathscr{E}(\underline{\rho}_\tau(t)) \quad \text{for all } 0 \leq t \leq T, \tag{7.19}$$

$$\int_{\underline{t}_\tau(s)}^{\bar{t}_\tau(t)} \overline{W}_\tau(r) \, \mathrm{d}r$$

$$+ \int_{\underline{t}_\tau(s)}^{\bar{t}_\tau(t)} \mathscr{S}(\widetilde{\rho}_\tau(r)) \, \mathrm{d}r + \mathscr{E}(\overline{\rho}_\tau(t)) \leq \mathscr{E}(\underline{\rho}_\tau(s)) \quad \text{for all } 0 \leq s \leq t \leq T, \tag{7.20}$$

and there exists a constant $C > 0$ such that for all $\tau > 0$

$$\int_0^T \overline{W}_\tau(t) \, \mathrm{d}t \leq C, \quad \int_0^T \mathscr{S}(\widetilde{\rho}_\tau(t)) \, \mathrm{d}t \leq C. \tag{7.21}$$

Finally, there exists a σ-sequentially compact subset $K \subset X$ such that

$$\overline{\rho}_\tau(t), \, \underline{\rho}_\tau(t), \, \widetilde{\rho}_\tau(t) \in K \quad \forall \, t \in [0, T] \, and \, \tau > 0. \tag{7.22}$$

Proof From (7.16) we directly deduce, for $t \in (t_\tau^{j-1}, t_\tau^j]$,

$$\mathscr{W}(t - t_\tau^{j-1}, \rho_\tau^{j-1}, \widetilde{\rho}_\tau(t)) + \int_{t_\tau^{j-1}}^t \mathscr{S}(\widetilde{\rho}_\tau(r)) \, dr + \mathscr{E}(\widetilde{\rho}_\tau(t)) \le \mathscr{E}(\rho_\tau^{j-1}), \tag{7.23}$$

which implies (7.19); in particular, for $t = t_\tau^j$ one has

$$\int_{t_\tau^{j-1}}^{t_\tau^j} \overline{W}_\tau(t) \, dt + \int_{t_\tau^{j-1}}^{t_\tau^j} \mathscr{S}(\widetilde{\rho}_\tau(t)) \, dt + \mathscr{E}(\rho_\tau^j) \le \mathscr{E}(\rho_\tau^{j-1}). \tag{7.24}$$

The estimate (7.20) follows upon summing (7.24) over the index j. Furthermore, applying (7.8)–(7.9) one deduces for all $1 \le n \le N_\tau$ that

$$\mathscr{W}(n\tau, \rho_0, \rho_\tau^n) + \mathscr{E}(\rho_\tau^n) \le \int_0^{t_\tau^n} \overline{W}_\tau(r) \, dr + \int_0^{t_\tau^n} \mathscr{S}(\widetilde{\rho}_\tau(r)) \, dr + \mathscr{E}(\rho_\tau^n) \le \mathscr{E}(\rho_0). \tag{7.25}$$

In particular, (7.21) follows, as well as $\sup_{n=0,\dots,N_\tau} \mathscr{E}(\rho_\tau^n) \le C$. Then, (7.23) also yields $\sup_{t \in [0,T]} \mathscr{E}(\widetilde{\rho}_\tau(t)) \le C$.

Next we show the two estimates

$$\mathscr{W}(2T, \rho^*, \overline{\rho}_\tau(t)) + \mathscr{E}(\overline{\rho}_\tau(t)) \le C, \tag{7.26}$$

$$\mathscr{W}(2T, \rho^*, \widetilde{\rho}_\tau(t)) + \mathscr{E}(\widetilde{\rho}_\tau(t)) \le C. \tag{7.27}$$

Recall that ρ^* is introduced in Assumption **(Abs)**.

To deduce (7.26), we use the triangle inequality for $\mathscr{W}$. Preliminarily, we observe that $\mathscr{W}(t, \rho^*, \rho_0) < +\infty$ for all $t > 0$. In particular, let us fix an arbitrary $m \in \{1, \dots, N_\tau\}$ and let $C^* := \mathscr{W}(t_\tau^m, \rho^*, \rho_0)$. We have for any n,

$$\mathscr{W}(2T, \rho^*, \rho_\tau^n) \le \mathscr{W}(2T - t_\tau^n, \rho^*, \rho_0) + \mathscr{W}(t_\tau^n, \rho_0, \rho_\tau^n) \overset{(1)}{\le} \mathscr{W}(t_\tau^m, \rho^*, \rho_0) + \mathscr{W}(t_\tau^n, \rho_0, \rho_\tau^n)$$
$$\le C^* + \mathscr{W}(t_\tau^n, \rho_0, \rho_\tau^n) \quad \text{for all } n \in \{1, \dots, N_\tau\},$$

where for (1) we have used that $\mathscr{W}(2T - t_\tau^n, \rho^*, \rho_0) \le \mathscr{W}(t_\tau^m, \rho^*, \rho_0)$ since $2T - t_\tau^n \ge t_\tau^m$. Thus, in view of (7.25) we deduce

$$\mathscr{W}(2T, \rho^*, \overline{\rho}_\tau(t)) + \mathscr{E}(\overline{\rho}_\tau(t)) \le C^* + \mathscr{W}(\bar{t}_\tau(t), \rho_0, \overline{\rho}_\tau(t)) + \mathscr{E}(\overline{\rho}_\tau(t))$$
$$\le C^* + \mathscr{E}(\rho_0) \le C \quad \text{for all } t \in [0, T], \tag{7.28}$$

i.e. the desired (7.26).

Likewise, adding (7.23) and (7.24) one has $\mathscr{W}(t, \rho_0, \widetilde{\rho}_\tau(t)) + \mathscr{E}(\widetilde{\rho}_\tau(t)) \le \mathscr{E}(\rho_0)$, whence (7.27) with arguments similar to those in the previous lines. □

7.4 Compactness result

The main result of this section, Theorem 7.7 below, states that GMM$(\mathscr{E}, \mathscr{W}; (0, T), \rho^\circ)$ is non-empty, and that any curve $\rho \in$ GMM$(\mathscr{E}, \mathscr{W}; (0, T), \rho^\circ)$ fulfills an 'abstract' version (7.31) of the $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ Energy-Dissipation estimate (5.6), obtained by passing to the limit in the discrete inequality (7.20).

We recall the $\mathscr{W}$-action of a curve $\rho : [0, T] \to X$, defined in (4.89) as

$$\mathbb{W}(\rho; [a, b]) := \sup \left\{ \sum_{j=1}^{M} \mathscr{W}(t^j - t^{j-1}, \rho(t^{j-1}), \rho(t^j)) \; : \; (t^j)_{j=0}^{M} \in \mathfrak{P}_f([a, b]) \right\}$$

for all $[a, b] \subset [0, T]$, where $\mathfrak{P}_f([a, b])$ is the set of all finite partitions of the interval $[a, b]$. We also introduce the *relaxed generalized slope* $\mathscr{S}^- : \mathrm{D}(\mathscr{E}) \to [0, +\infty]$ of the driving energy functional $\mathscr{E}$, namely the relaxation of the generalized slope $\mathscr{S}$ along sequences with bounded energy:

$$\mathscr{S}^-(\rho) := \inf \left\{ \liminf_{n \to \infty} \mathscr{S}(\rho_n) \; : \; \rho_n \overset{\sigma}{\to} \rho, \; \sup_{n \in \mathbb{N}} \mathscr{E}(\rho_n) < +\infty \right\}. \tag{7.29}$$

We are now in a position to state and prove the 'abstract version' of Theorem 7.4.

Theorem 7.7 *Under Assumption (Abs), let $\rho^\circ \in \mathrm{D}(\mathscr{E})$. Then, for every vanishing sequence $(\tau_k)_k$ there exist a (not relabeled) subsequence and a σ-continuous curve $\rho : [0, T] \to X$ such that $\rho(0) = \rho^\circ$, and*

$$\overline{\rho}_{\tau_k}(t), \; \underline{\rho}_{\tau_k}(t), \; \widetilde{\rho}_{\tau_k}(t) \overset{\sigma}{\to} \rho(t) \quad \text{for all } t \in [0, T], \tag{7.30}$$

and ρ satisfies the Energy-Dissipation estimate

$$\mathbb{W}(\rho; [0, t]) + \int_0^t \mathscr{S}^-(\rho(r)) \mathrm{d}r + \mathscr{E}(\rho(t)) \leq \mathscr{E}(\rho_0) \quad \text{for all } t \in [0, T]. \tag{7.31}$$

Remark 7.8 Theorem 7.7 could be extended to a topological space where the cost $\mathscr{W}$ and the energy functional $\mathscr{E}$ satisfy the properties listed at the beginning of the section.

Proof Consider a sequence $\tau_k \downarrow 0$ as $k \to \infty$.

Step 1: Construct the limit curve $\overline{\rho}$. We first define the limit curve $\overline{\rho}$ on the set $A := \{0\} \cup N$, with N a countable dense subset of $(0, T)$. Indeed, in view of (7.22), with a diagonalization procedure we find a function $\overline{\rho} : A \to X$ and a (not relabeled) subsequence such that

$$\overline{\rho}_{\tau_k}(t) \overset{\sigma}{\to} \overline{\rho}(t) \quad \text{for all } t \in A \quad \text{and} \quad \overline{\rho}(t) \in K \text{ for all } t \in A. \tag{7.32}$$

In particular, $\overline{\rho}(0) = \rho^\circ$.

We next show that $\overline{\rho}$ can be uniquely extended to a σ-continuous curve $\overline{\rho} : [0, T] \to X$. Let $s, t \in A$ with $s < t$. By the lower-semicontinuity property (4.86c) we have

$$\mathscr{W}(t - s, \overline{\rho}(s), \overline{\rho}(t)) \leq \liminf_{k \to \infty} \mathscr{W}(t - s, \overline{\rho}_{\tau_k}(s), \overline{\rho}_{\tau_k}(t)) \overset{(7.18)}{\leq} \liminf_{k \to \infty} \int_{\underline{t}_{\tau_k}(s)}^{\overline{t}_{\tau_k}(t)} \overline{W}_{\tau_k}(r) \, \mathrm{d}r$$

$$\overset{(1)}{\leq} \liminf_{k \to \infty} \mathscr{E}(\overline{\rho}_{\tau_k}(t_1)) \overset{(2)}{\leq} \mathscr{E}(\rho_0),$$

where (1) follows from (7.20) (using the lower bound on $\mathscr{E}$), and (2) is due to the fact that $t \mapsto \mathscr{E}(\overline{\rho}_{\tau_k}(t))$ is nonincreasing.

By the property (4.86e) of $\mathscr{W}$, this estimate is a form of uniform continuity of $\overline{\rho}$, and we now use this to extend $\overline{\rho}$. Fix $t \in [0, T] \setminus A$, and choose a sequence $t_m \in A$, $t_m \to t$, with the property that $\overline{\rho}(t_m)$ σ-converges to some $\tilde{\rho}$. For any sequence $s_m \in A$, $s_m \to t$, we then have

$$\sup_m \mathscr{W}(|t_m - s_m|, \overline{\rho}(s_m), \overline{\rho}(t_m)) < +\infty,$$

and since $|t_m - s_m| \to 0$, property (4.86e) implies that $\overline{\rho}(s_m) \overset{\sigma}{\to} \tilde{\rho}$. This implies that along any converging sequence $t_m \in A$, $t_m \to t$ the sequence $\overline{\rho}(t_m)$ has the same limit; therefore there is a unique extension of $\overline{\rho}$ to $[0, T]$, that we again indicate by $\overline{\rho}$. By again applying the lower-semicontinuity property (4.86c) we find that

$$\mathscr{W}(|t - s|, \overline{\rho}(s), \overline{\rho}(t)) \leq \mathscr{E}(\rho_0) \qquad \text{for all } t, s \in [0, T], \ s \neq t,$$

and therefore the curve $[0, T] \ni t \mapsto \overline{\rho}(t)$ is σ-continuous.

Step 2: Show convergence at all $t \in [0, T]$. Now fix $t \in [0, T]$; we show that $\overline{\rho}_{\tau_k}(t)$, $\underline{\rho}_{\tau_k}(t)$, and $\widetilde{\rho}_{\tau_k}(t)$ each converge to $\overline{\rho}(t)$. Since $\overline{\rho}_{\tau_k}(t) \in K$, there exists a convergent subsequence $\overline{\rho}_{\tau_{k_j}}(t) \overset{\sigma}{\to} \tilde{\rho}$. Take any $s \in A$ with $s \neq t$. Then

$$\mathscr{W}(|t - s|, \tilde{\rho}, \overline{\rho}(s)) \leq \liminf_{j \to \infty} \mathscr{W}(|t - s|, \overline{\rho}_{\tau_{k_j}}(t), \overline{\rho}_{\tau_{k_j}}(s)) \leq \mathscr{E}(\rho_0) \leq C,$$

by the same argument as above. Taking the limit $s \to t$, property (4.86e) and the continuity of $\overline{\rho}$ imply $\tilde{\rho} = \overline{\rho}(t)$. Therefore $\overline{\rho}_{\tau_{k_j}}(t) \overset{\sigma}{\to} \overline{\rho}(t)$ along each subsequence τ_{k_j}, and consequently also along the whole sequence τ_k.

Estimates (7.19) & (7.20) also give at each $t \in (0, T]$

$$\limsup_{k \to \infty} \mathscr{W}(t - \underline{t}_{\tau_k}(t), \underline{\rho}_{\tau_k}(t), \overline{\rho}_{\tau_k}(t)) \leq \mathscr{E}(\rho_0),$$
$$\limsup_{k \to \infty} \mathscr{W}(t - \underline{t}_{\tau_k}(t), \underline{\rho}_{\tau_k}(t), \widetilde{\rho}_{\tau_k}(t)) \leq \mathscr{E}(\rho_0),$$

so that, again using the compactness information provided by (7.22) and property (4.86e) of the cost $\mathscr{W}$, it is immediate to conclude (7.30).

Step 3: Derive the energy-dissipation estimate. Finally, let us observe that

$$\liminf_{k \to \infty} \int_0^{\overline{t}_{\tau_k}(t)} \overline{W}_{\tau_k}(r) \mathrm{d}r \geq \mathbb{W}(\rho; [0, t]) \quad \text{for all } t \in [0, T]. \tag{7.33}$$

Indeed, for any partition $\{0 = t^0 < \ldots < t^j < \ldots < t^M = t\}$ of $[0, t]$ we find that

$$\sum_{j=1}^{M} \mathscr{W}(t^j - t^{j-1}, \rho(t^{j-1}), \rho(t^j)) \overset{(1)}{\leq} \liminf_{k \to \infty} \sum_{j=1}^{M} \mathscr{W}(\overline{t}_{\tau_k}(t^j) - \overline{t}_{\tau_k}(t^{j-1}), \overline{\rho}_{\tau_k}(t^{j-1}), \overline{\rho}_{\tau_k}(t^j))$$
$$= \liminf_{k \to \infty} \int_0^{\overline{t}_{\tau_k}(t)} \overline{W}_{\tau_k}(r) \, \mathrm{d}r,$$

with (1) due to (4.86c). Then (7.33) follows by taking the supremum over all partitions. On the other hand, by Fatou's Lemma we find that

$$\liminf_{k \to \infty} \int_0^{\overline{t}_{\tau_k}(t)} \mathscr{S}(\widetilde{\rho}_{\tau_k}(r)) \, \mathrm{d}r \geq \int_0^t \mathscr{S}^-(\rho(r)) \mathrm{d}r,$$

while the lower semicontinuity of $\mathscr{E}$ gives

$$\liminf_{k \to \infty} \mathscr{E}(\overline{\rho}_{\tau_k}(t)) \geq \mathscr{E}(\rho(t))$$

so that (7.31) follows from taking the $\liminf_{k \to \infty}$ in (7.20) for $s = 0$. $\qquad\square$

7.5 Proof of Theorem 7.4

Having established the abstract compactness result of Theorem 7.7, we now apply this to the proof of Theorem 7.4. As described above, under **Assumptions ($V\pi\kappa$), ($\mathscr{R}^*\Psi\alpha$), and ($\mathscr{E}\phi$)** the conditions of Theorem 7.7 are fulfilled, and Theorem 7.7 provides us with a curve $\rho :$ $[0, T] \to \mathcal{M}^+(V)$ that is continuous with respect to setwise convergence such that

$$\mathbb{W}(\rho; [0, t]) + \int_0^t \mathscr{S}^-(\rho(r))\mathrm{d}r + \mathscr{E}(\rho(t)) \leq \mathscr{E}(\rho_0) \qquad \text{for all } t \in [0, T]. \qquad (7.34)$$

To conclude the proof of Theorem 7.4, we now show that the Energy-Dissipation inequality (5.6) can be derived from (7.34).

We first note that Corollary 4.22 implies the existence of a flux j such that $(\rho, j) \in \mathcal{CE}(0, T)$ and $\mathbb{W}(\rho; [0, T]) = \int_0^T \mathscr{R}(\rho_t, j_t)\, \mathrm{d}t$. Then from Corollary 7.11 below, we find that $\mathscr{S}^-(\rho(r)) \geq \mathscr{D}(\rho(r))$ for all $r \in [0, T]$. Combining these results with (7.34) we find the required estimate (5.6).

It remains to prove the inequality $\mathscr{S}^- \geq \mathscr{D}$, which follows from the corresponding inequality $\mathscr{S} \geq \mathscr{D}$ for the non-relaxed slope (Theorem 7.9) with the lower semicontinuity of $\mathscr{D}$ that is assumed in Theorem 7.4. This is the topic of the next section.

7.6 The generalized slope bounds the fisher information

We recall the definition of the generalized slope $\mathscr{S}$ from (7.11):

$$\mathscr{S}(\rho):=\limsup_{r\downarrow 0} \sup_{\mu\in X} \frac{1}{r}\left\{\mathscr{E}(\rho) - \mathscr{E}(\mu) - \mathscr{W}(r, \rho, \mu)\right\}.$$

Given the structure of this definition, the proof of the inequality $\mathscr{S} \geq \mathscr{D}$ naturally proceeds by constructing an admissible curve $(\rho, j) \in \mathcal{CE}(0, T)$ such that $\rho_{|t=0} = \rho$ and such that the expression in braces can be related to $\mathscr{D}(\rho)$.

For the systems of this paper, the construction of such a curve faces three technical difficulties: the first is that ρ needs to remain nonnegative, the second is that ϕ' may be unbounded at zero, and the third is that the function $\mathrm{D}_\phi(u, v)$ in (4.53c) that defines $\mathscr{D}$ may be infinite when u or v is zero (see Example 5.2).

We first prove a lower bound for the generalized slope $\mathscr{S}$ involving D_ϕ^-, under the basic conditions on the $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ system presented in Sect. 3.

Theorem 7.9 *Assume ($V\pi\kappa$), ($\mathscr{R}^*\Psi\alpha$), and ($\mathscr{E}\phi$). Then*

$$\mathscr{S}(\rho) \geq \frac{1}{2}\iint_E \mathrm{D}_\phi^-(u(x), u(y))\, \vartheta(\mathrm{d}x, \mathrm{d}y) \quad \text{for all } \rho = u\pi \in \mathrm{D}(\mathscr{E}). \qquad (7.35)$$

Proof Let us fix $\rho_0 = u_0\pi \in \mathrm{D}(\mathscr{E})$, a bounded measurable skew-symmetric map

$$\xi : E \to \mathbb{R} \quad \text{with } \xi(y, x) = -\xi(x, y), \quad |\xi(x, y)| \leq \Xi < \infty \quad \text{for every } (x, y) \in E,$$

the Lipschitz functions $q(r):=\min(r, 2(r - 1/2)_+)$ (approximating the identity far from 0) and $h(r):=\max(0, \min(2 - r, 1))$ (cutoff for $r \geq 2$), and the Lipschitz regularization of α

$$\alpha_\varepsilon(u, v):=\varepsilon q(\alpha(u, v)/\varepsilon).$$

We introduce the field $\mathrm{G}_\varepsilon : E \times \mathbb{R}_+^2 \to \mathbb{R}$

$$\mathrm{G}_\varepsilon(x, y; u, v):=\xi(x, y)g_\varepsilon(u, v), \qquad (7.36)$$

where

$$g_\varepsilon(u, v) := \alpha_\varepsilon(u, v)\, h(\varepsilon \max(u, v))\, q(\min(1, \min(u, v)/\epsilon))\,,$$

which vanishes if $\alpha(u, v) < \varepsilon/2$ or $\min(u, v) < \varepsilon/2$ or $\max(u, v) \geq 2/\varepsilon$, and coincides with α if $\alpha \geq \varepsilon$, $\min(u, v) \geq \varepsilon$, and $\max(u, v) \leq 1/\varepsilon$. Since g_ε is Lipschitz, it is easy to check that G_ε satisfies all the assumptions (6.1a,b,c,d) and also (6.7) for $a = 0$, since $0 = g_\varepsilon(0, 0) \leq g_\varepsilon(0, v)$ for every $v \geq 0$ and every $(x, y) \in E$.

It follows that for every nonnegative $u_0 \in L^1(X, \pi)$ there exists a unique nonnegative solution $u^\varepsilon \in C^1([0, \infty); L^1(V, \pi))$ of the Cauchy problem (6.15) induced by G_ε with initial datum u_0 and the same total mass. Henceforth, we set $\rho_t^\varepsilon = u_t^\varepsilon \pi$ for all $t \geq 0$.

Setting $2j_t^\varepsilon(\mathrm{d}x, \mathrm{d}y) := w_t^\varepsilon(x, y)\boldsymbol{\vartheta}(\mathrm{d}x, \mathrm{d}y)$, where $w_t^\varepsilon(x, y) := G_\varepsilon(x, y; u_t(x), u_t(y))$, it is also easy to check that $(\rho^\varepsilon, j^\varepsilon) \in \mathcal{A}(0, T)$, since $g_\varepsilon(u, v) \leq \alpha(u, v)$ and

$$|w_t^\varepsilon(x, y)| \leq |\xi|\alpha(u_t^\varepsilon(x), u_t^\varepsilon(y))\mathcal{X}_{U_\varepsilon(t)}(x, y) \qquad \text{for } (x, y) \in E\,,$$

where $U_\varepsilon(t) := \{(x, y) \in E : g_\epsilon(u_t^\varepsilon(x), u_t^\varepsilon(y)) > 0\}$, thereby yielding

$$\Upsilon(u_t^\varepsilon(x), u_t^\varepsilon(y), w_t^\varepsilon(x, y)) \leq \Psi(\Xi)\alpha(2/\varepsilon, 2/\varepsilon)\,.$$

Finally, recalling (4.40) and (4.42), we get

$$|B_\phi(u_t^\varepsilon(x), u_t^\varepsilon(y), w_t^\varepsilon(x, y))| \leq \Xi\big(\phi'(2/\varepsilon) - \phi'(\varepsilon/2)\big)\alpha(2/\varepsilon, 2/\varepsilon).$$

Thus, we can apply Theorem 4.16 obtaining

$$\mathscr{E}(\rho_0) - \mathscr{E}(\rho_\tau^\varepsilon) = -\frac{1}{2}\int_0^\tau \iint_E B_\phi(u_t^\varepsilon(x), u_t^\varepsilon(y), w_t^\varepsilon(x, y))\,\boldsymbol{\vartheta}(\mathrm{d}x, \mathrm{d}y)\,\mathrm{d}t, \qquad (7.37)$$

and consequently

$$\begin{aligned}
\mathscr{S}(\rho_0) &\geq \limsup_{\tau\downarrow 0}\tau^{-1}\Big(\mathscr{E}(\rho_0) - \mathscr{E}(\rho_\tau^\varepsilon) - \int_0^\tau \mathscr{R}(\rho_t^\varepsilon, j_t^\varepsilon)\,\mathrm{d}t\Big)\\
&= \frac{1}{2}\iint_E \Big(B_\phi(u_0(x), u_0(y), w_0^\varepsilon(x, y)) - \Upsilon(u_0(x), u_0(y), w_0^\varepsilon(x, y))\Big)\boldsymbol{\vartheta}(\mathrm{d}x, \mathrm{d}y).
\end{aligned} \qquad (7.38)$$

Let us now set Δ_k to be the truncation of $\phi'(u_0(x)) - \phi'(u_0(y))$ to $[-k, k]$, i.e.

$$\Delta_k(x, y) := \max\Big\{-k, \min\big[k, \phi'(u_0(x)) - \phi'(u_0(y))\big]\Big\}\,,$$

and $\xi_k(x, y) := (\Psi^*)'(\Delta_k(x, y))$ for each $k \in \mathbb{N}$. Notice that ξ_k is a bounded measurable skew-symmetric map satisfying $|\xi_k(x, y)| \leq k$ for every $(x, y) \in E$ and $k \in \mathbb{N}$. Therefore, inequality (7.38) holds for $w_0^\varepsilon(x, y) = \xi_k(x, y)\, g_\varepsilon(u_0(x), u_0(y))$, $(x, y) \in E$. We then observe from Lemma 4.19(3) that

$$\begin{aligned}
(\phi'(u_0(x)) - \phi'(u_0(y))) \cdot \xi_k(x, y) &\geq \Delta_k(x, y)\xi_k(x, y)\\
&= \Psi(\xi_k(x, y)) + \Psi^*(\Delta_k(x, y))\,,
\end{aligned} \qquad (7.39)$$

and from $g_\epsilon(u, v) \leq \alpha(u, v)$ that

$$\begin{aligned}
\Upsilon(u_0(x), u_0(y), w_0^\varepsilon(x, y)) &= \Psi\left(\frac{\xi_k(x, y)g_\varepsilon(u_0(x), u_0(y))}{\alpha(u_0(x), u_0(y))}\right)\alpha(u_0(x), u_0(y))\\
&\leq \Psi(\xi_k(x, y))\alpha(u_0(x), u_0(y))\,.
\end{aligned} \qquad (7.40)$$

Substituting these bounds in (7.38) and passing to the limit as $\varepsilon \downarrow 0$ we obtain

$$\mathscr{S}(\rho) \geq \frac{1}{2} \iint_E \Psi^*(\Delta_k(x, y)) \alpha(u_0(x), u_0(y)) \, \vartheta(\mathrm{d}x, \mathrm{d}y). \tag{7.41}$$

We eventually let $k \uparrow \infty$ and obtain (7.35). $\qquad\qquad \square$

In the next proposition we finally bound $\mathscr{S}$ from below by the Fisher information, by relying on the existence of a solution to the $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ system, as shown in Sect. 6.

Proposition 7.10 *Let us suppose that for $\rho \in D(\mathscr{E})$ there exists a solution to the $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ system. Then the generalized slope bounds the Fisher information from above:*

$$\mathscr{S}(\rho) \geq \mathscr{D}(\rho) \quad \text{for all } \rho \in D(\mathscr{E}). \tag{7.42}$$

Proof. Let $\rho_t = u_t \pi$ be a solution to the $(\mathscr{E}, \mathscr{R}, \mathscr{R}^*)$ system with initial datum $\rho_0 \in D(\mathscr{E})$. Then, we can find a family $(j_t)_{t \geq 0} \in \mathcal{M}(E)$ such that $(\rho, j) \in \mathcal{CE}(0, +\infty)$ and

$$\mathscr{E}(\rho_t) + \int_0^t \left[\mathscr{R}(\rho_r, j_r) + \mathscr{D}(\rho_r)\right] \mathrm{d}r = \mathscr{E}(\rho_0) \quad \text{for all } t \geq 0.$$

Therefore

$$\mathscr{S}(\rho_0) \geq \liminf_{t \downarrow 0} \frac{1}{t}\left[\mathscr{E}(\rho_0) - \mathscr{E}(\rho_t) - \mathscr{W}(t, \rho_0, \rho_t)\right]$$

$$\geq \liminf_{t \downarrow 0} \frac{1}{t}\left[\mathscr{E}(\rho_0) - \mathscr{E}(\rho_t) - \int_0^t \mathscr{R}(\rho_r, j_r) \, \mathrm{d}r\right] = \liminf_{t \downarrow 0} \frac{1}{t} \int_0^t \mathscr{D}(\rho_r) \, \mathrm{d}r.$$

Since $u_t \to u_0$ in $L^1(V; \pi)$ as $t \to 0$ and since $\mathscr{D}$ is lower semicontinuous with respect to $L^1(V, \pi)$-convergence (see the proof of Proposition 5.3), with a change of variables we find

$$\mathscr{S}(\rho_0) \geq \liminf_{t \downarrow 0} \int_0^1 \mathscr{D}(\rho_{ts}) \, \mathrm{d}s \geq \mathscr{D}(\rho_0). \qquad\qquad \square$$

We then easily get the desired lower bound for $\mathscr{S}^-$ in terms of $\mathscr{D}$, under the condition that the latter functional is lower semicontinuous (recall that Proposition 5.3 provides sufficient conditions for the lower semicontinuity of $\mathscr{D}$):

Corollary 7.11 *Let us suppose that **Assumptions** ($V\pi\kappa$), ($\mathscr{R}^*\Psi\alpha$), ($\mathscr{E}\phi$) hold and that $\mathscr{D}$ is lower semicontinuous with respect to setwise convergence. Then*

$$\mathscr{S}^-(\rho) \geq \mathscr{D}(\rho) \quad \text{for all } \rho \in D(\mathscr{E}). \tag{7.43}$$

As discussed in Example 4.18, the cosh and the quadratic case provide examples in which the Fisher information functional $\mathscr{D}$ is lower semicontinuous. When π is purely atomic, then $\mathscr{D}$ is lower semicontinuous for all the examples mentioned in Sect. 1.3. In the case when π is not purely atomic, the lower semicontinuity of $\mathscr{D}$ is related to the convexity of the function D_ϕ (4.53). We show in Appendix E that all the power means in (1.34) for $p \in [-\infty, -1] \cup [0, 1)$, together with Ψ^* in (1.35), do lead to D_ϕ's that are convex and lower semicontinuous, and ultimately to the lower semicontinuity of $\mathscr{D}$.

Remark 7.12 The combination of Theorem 7.9, Proposition 7.10, and Corollary 7.11 illustrates why we introduced both D_ϕ and D_ϕ^-. For the duration of this remark, consider both

the functional $\mathscr{D}$ that is defined in (5.1) in terms of D_ϕ, and a corresponding functional $\mathscr{D}^-$ defined in terms of the function D_ϕ^-:

$$\mathscr{D}^-(\rho) := \frac{1}{2} \iint_E D_\phi^-\big(u(x), u(y)\big)\, \boldsymbol{\vartheta}(\mathrm{d}x\,\mathrm{d}y) \qquad \text{for } \rho = u\pi\,.$$

In the two guiding cases of Example 4.18, D_ϕ is convex and lower semicontinuous, but D_ϕ^- is only lower semicontinuous. As a result, $\mathscr{D}$ is lower semicontinuous with respect to setwise convergence, but $\mathscr{D}^-$ is not (indeed, consider e.g. a sequence ρ_n converging setwise to ρ, with $\mathrm{d}\rho_n/\mathrm{d}\pi$ given by characteristic functions of some sets A_n, where the sets A_n are chosen such that for the limit the density $\mathrm{d}\rho/\mathrm{d}\pi$ is strictly positive and non-constant; then $\mathscr{D}^-(\rho_n) = 0$ for all n while $\mathscr{D}^-(\rho) > 0$). Setwise lower semicontinuity of $\mathscr{D}$ is important for two reasons: first, this is required for stability of solutions of the Energy-Dissipation balance under convergence in some parameter (evolutionary Γ-convergence), which is a hallmark of a good variational formulation; and secondly, the proof of existence using the Minimizing-Movement approach requires the bound (7.43), for which $\mathscr{D}$ also needs to be lower semicontinuous. This explains the importance of D_ϕ, and it also explains why we defined the Fisher information $\mathscr{D}$ in terms of D_ϕ and not in terms of D_ϕ^-.

On the other hand, D_ϕ^- is straightforward to determine, and in addition the weaker control of D_ϕ^- is still sufficient for the chain rule: it is D_ϕ^- that appears on the right-hand side of (4.59). Note that if D_ϕ^- itself is convex, then it coincides with D_ϕ.

Appendix A: Derivation of the cosh-structure from large deviations

We mentioned in the introduction that the choices

$$\phi(s) = s \log s - s + 1, \qquad \Psi^*(\xi) = 4\big(\cosh(\xi/2) - 1\big), \quad \text{and} \quad \alpha(u, v) = \sqrt{uv} \quad \text{(A.1)}$$

arise in the context of large deviations. In this section we describe this context. Throughout this section we work under Assumptions $(V\pi\kappa)$, $(\mathscr{E}\phi)$, and $(\mathscr{R}^*\Psi\alpha)$, and since we are interested in the choices above, we will also assume (A.1), implying that

$$\boldsymbol{\nu}_\rho(\mathrm{d}x\,\mathrm{d}y) = \sqrt{u(x)u(y)}\,\pi(\mathrm{d}x)\kappa(x, \mathrm{d}y), \qquad \text{if } \rho = u\pi \ll \pi.$$

Consider a sequence of independent and identically distributed stochastic processes X^i, $i = 1, 2, \ldots$ on V, each described by the jump kernel κ, or equivalently by the generator Q in (1.3). With probability one, a realization of each process has a countable number of jumps in the time interval $[0, +\infty)$, and we write t_k^i for the kth jump time of X^i. We can assume that X^i is a càdlàg function of time.

We next define the empirical measure ρ^n and the empirical flux j^n by

$$\rho^n : [0, T] \to \mathcal{M}^+(V), \qquad\qquad \rho_t^n := \frac{1}{n} \sum_{i=1}^n \delta_{X_t^i},$$

$$j^n \in \mathcal{M}^+((0, T) \times E), \qquad j^n(\mathrm{d}t\, \mathrm{d}x\, \mathrm{d}y) := \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^\infty \delta_{t_k^i}(\mathrm{d}t) \delta_{(X_{t_-}^i, X_t^i)}(\mathrm{d}x\, \mathrm{d}y),$$

where t_k^i is the kth jump time of X^i, and $X_{t_-}^i$ is the left limit (pre-jump state) of X^i at time t. Equivalently, j^n is defined by

$$\langle j^n, \varphi \rangle := \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^\infty \varphi\big(t_k^i, X_{t_k^i-}^i, X_{t_k^i}^i\big), \qquad \text{for } \varphi \in C_b([0, T] \times E).$$

A standard application of Sanov's theorem yields a large-deviation characterization of the pair (ρ^n, j^n) in terms of two rate functions I_0 and I,

$$\text{Prob}\big((\rho^n, j^n) \approx (\rho, j)\big) \sim \exp\Big[-n\big(I_0(\rho_0) + I(\rho, j)\big)\Big], \qquad \text{as } n \to \infty.$$

The rate function I_0 describes the large deviations of the initial datum ρ_0^n; this functional is determined by the choices of the initial data of X_0^i and is independent of the stochastic process itself, and we therefore disregard it here.

The functional I characterizes the large-deviation properties of the dynamics of the pair (ρ^n, j^n) conditional on the initial state, and has the expression

$$I(\rho, j) = \int_0^T \mathscr{F}_\eta(j_t | \vartheta_{\rho_t}^-) \, \mathrm{d}t. \tag{A.2}$$

In this expression we write $\vartheta_{\rho_t}^-$ for the measure $\rho_t(\mathrm{d}x)\kappa(x, \mathrm{d}y) \in \mathcal{M}(E)$ (see also (4.18) ahead). The function η is the Boltzmann entropy function that we have seen above,

$$\eta(s) := s \log s - s + 1, \qquad \text{for } s \geq 0,$$

and the functional $\mathscr{F}_\eta : \mathcal{M}^+(E) \times \mathcal{M}^+(E) \to [0, \infty]$ is given by (2.11). Even though the function η coincides in this section with ϕ, we choose a different notation to emphasize that the roles of ϕ and η are different: the function ϕ defines the entropy of the system, which is related to the large deviations of the empirical measures ρ^n in equilibrium (see [59]); the function η characterizes the large deviations of the time courses of ρ^n and j^n.

Remark A.1 Sanov's theorem can be found in many references on large deviations (e.g. [24, Sec. 6.2]); the derivation of the expression (A.2) is fairly well known and can be found in e.g. [56, Eq. (8)] or [43, App. A]. Instead of proving (A.2) we give an interpretation of the expression (A.2) and the function η in terms of exponential clocks. An exponential clock with rate parameter r has large-deviation behaviour given by $r\eta(\cdot/r)$ (see [24, Exercise 5.2.12] or [57, Th. 1.5]) in the following sense: for each $t > 0$,

$$\text{Prob}\big(\approx \beta nt \text{ firings in time } nt\big) \sim \exp\Big[-ntr\, \eta(\beta/r)\Big] \qquad \text{as } n \to \infty.$$

The expression (A.2) generalizes this to a field of exponential clocks, one for each edge (x, y). In this case, the rescaled rate parameter r for the clock at edge (x, y) is equal to $\rho_t(dx)\kappa(x, dy)$, since it is proportional to the number of particles $n\rho_t(dx)$ at x and to the rate of jump $\kappa(x, dy)$ from x to y. The flux $n\boldsymbol{j}_t(dx\,dy)$ is the observed number of jumps from x to y, corresponding to firings of the clock associated with the edge (x, y). In this way, the functional I in (A.2) can be interpreted as characterizing the large-deviation fluctuations in the clock-firings for each edge $(x, y) \in E$.

The expression (A.2) leads to the functional $\mathscr{L}$ in (1.18) after a symmetry reduction, which we now describe (see also [43, App. A]). Assuming that we are more interested in the fluctuation properties of ρ than those of $\boldsymbol{j}$, we might decide to minimize $I(\rho, \boldsymbol{j})$ over a class of fluxes $\boldsymbol{j}$ for a fixed choice of ρ. Here we choose to minimize over the class of fluxes with the same skew-symmetric part,

$$A_j := \left\{ \boldsymbol{j}' \in \mathcal{M}([0, T] \times E) : \boldsymbol{j}' - s_\# \boldsymbol{j}' = \boldsymbol{j} - s_\# \boldsymbol{j} \right\}.$$

By the form (1.23) of the continuity equation and the definition (1.6b) of the divergence we have $\overline{\mathrm{div}}\,\boldsymbol{j}' = \overline{\mathrm{div}}\,\boldsymbol{j}$ for all $\boldsymbol{j}' \in A_j$, so that replacing $\boldsymbol{j}$ by $\boldsymbol{j}'$ preserves the continuity equation.

Formal Lemma A.2 *The minimum of $I(\rho, \boldsymbol{j}')$ over all $\boldsymbol{j}' \in A_j$ is achieved for the 'skew-symmetrization' $\boldsymbol{j}^\flat = \frac{1}{2}(\boldsymbol{j} - s_\# \boldsymbol{j})$, and for $\boldsymbol{j}^\flat$ the result equals $\frac{1}{2}\mathscr{L}$:*

$$\inf_{\boldsymbol{j}' \in A_j} I(\rho, \boldsymbol{j}') = \inf_{\boldsymbol{j}' \in A_j} \tfrac{1}{2}\mathscr{L}(\rho, \boldsymbol{j}') = \tfrac{1}{2}\mathscr{L}(\rho, \boldsymbol{j}^\flat). \tag{A.3}$$

Consequently, for a given curve $\rho : [0, T] \to \mathcal{M}^+(V)$,

$$\inf_{\boldsymbol{j}} \left\{ I(\rho, \boldsymbol{j}) : \partial_t \rho + \overline{\mathrm{div}}\,\boldsymbol{j} = 0 \right\} = \inf_{\boldsymbol{j}} \left\{ \tfrac{1}{2}\mathscr{L}(\rho, \boldsymbol{j}) : \partial_t \rho + \overline{\mathrm{div}}\,\boldsymbol{j} = 0 \right\},$$

and in this final expression the flux can be assumed to be skew-symmetric:

$$= \inf_{\boldsymbol{j}} \left\{ \tfrac{1}{2}\mathscr{L}(\rho, \boldsymbol{j}) : \partial_t \rho + \overline{\mathrm{div}}\,\boldsymbol{j} = 0 \text{ and } s_\# \boldsymbol{j} = -\boldsymbol{j} \right\}.$$

This implies that the two functionals I and $\mathscr{L}$ can be considered to be the same, if one is only interested in ρ, not in $\boldsymbol{j}$. By the Contraction Principle (e.g. [24, Sec. 4.2.1]) the functional $\rho \mapsto \inf_{\boldsymbol{j}} I(\rho, \boldsymbol{j}) = \inf_{\boldsymbol{j}} \tfrac{1}{2}\mathscr{L}(\rho, \boldsymbol{j})$ also can be viewed as the large-deviation rate function of the sequence of empirical measures ρ^n.

The above lemma is only formal because we have not given a rigorous definition of the functional $\mathscr{L}$. While it would be possible to do so, using the construction of Lemma 2.3 and the arguments of the proof below, actually the rest of this paper deals with this question in a more detailed manner. In addition, in the context of this paper, this lemma only serves to explain why we consider this specific class of functionals $\mathscr{L}$. Therefore here we only give heuristic arguments.

Proof We assume throughout this (formal) proof that all measures are absolutely continuous, strictly positive, and finite where necessary. Note that writing $\rho_t = u_t \pi$ we have $\vartheta^-_{\rho_t}(dx\,dy) = u_t(x)\vartheta(dx\,dy)$, and using (A.1) we therefore have

$$\sqrt{\vartheta^-_{\rho_t} s_\# \vartheta^-_{\rho_t}}\,(dx\,dy) = \sqrt{u_t(x)u_t(y)}\,\vartheta(dx\,dy) = \boldsymbol{v}_{\rho_t}(dx\,dy), \qquad \text{and}$$

$$\log \frac{ds_\# \vartheta^-_{\rho_t}}{d\vartheta^-_{\rho_t}}(x, y) = \log \frac{u_t(y)}{u_t(x)} = \overline{\nabla}\phi'(u_t)(x, y).$$

For the length of this proof we write $\hat{\eta}$ for the perspective function corresponding to η (see (2.14) in Lemma 2.3)

$$\hat{\eta}(a,b) := \begin{cases} a\log\dfrac{a}{b} - a + b & \text{if } a, b > 0, \\ 0 & \text{if } a = 0, \\ +\infty & \text{if } a > 0,\ b = 0. \end{cases}$$

We now rewrite $\inf_{j' \in A_j} I(\rho, j')$ as

$$\inf_{j' \in A_j} \int_0^T \iint_E \eta\left(\frac{dj'_t}{d\vartheta^-_{\rho_t}}\right) d\vartheta^-_{\rho_t} dt = \inf_{j' \in A_j} \int_0^T \iint_E u_t\, \eta\left(\frac{1}{u_t}\frac{dj'_t}{d\vartheta}\right) d\vartheta\, dt$$

$$= \inf_{j' = \zeta\vartheta \in A_j} \int_0^T \iint_E \hat{\eta}\big(\zeta_t(x,y), u_t(x)\big)\vartheta(dx, dy)\, dt$$

$$= \frac{1}{2}\inf_{j' = \zeta\vartheta \in A_j} \int_0^T \iint_E \left\{\hat{\eta}\big(\zeta_t(x,y), u_t(x)\big) + \hat{\eta}\big(\zeta_t(y,x), u_t(y)\big)\right\}\vartheta(dx, dy)\, dt.$$

Since $\zeta(x,y) - \zeta(y,x) = d(j' - \mathsf{s}_\# j')/d\vartheta$ is constrained in A_j, we follow the expression inside the second integral and set

$$\psi : \mathbb{R} \times [0,+\infty)^2 \to [0,+\infty], \qquad \psi(s\,;c,d) := \inf_{a,b \geq 0}\left\{[\hat{\eta}(a,c) + \hat{\eta}(b,d)] : a - b = 2s\right\},$$

for which a calculation gives the explicit formula (for $c, d > 0$)

$$\psi(s\,;c,d) = \frac{\sqrt{cd}}{2}\left\{\Psi\left(\frac{2s}{\sqrt{cd}}\right) + \Psi^*\left(-\log\frac{d}{c}\right)\right\} + s\,\log\frac{d}{c},$$

in terms of the function $\Psi^*(\xi) = 4\big(\cosh\xi/2 - 1\big)$ and its Legendre dual Ψ. This minimization corresponds to minimizing over all fluxes for which the 'net flux' $j - \mathsf{s}_\# j = 2j^\flat$ is the same; see e.g. [43,75] for discussions.

Let $w^\flat(x,y) := (w(x,y) - w(y,x)) = \frac{d(2j^\flat)}{d\vartheta}$ and $\alpha_t := \alpha_t(x,y) = \sqrt{u_t(x)u_t(y)}$. We find

$$\inf_{j' \in A_j} \int_0^T \mathscr{F}_\eta\big(j'_t | \vartheta^-_{\rho_t}\big) dt$$

$$= \frac{1}{2}\int_0^T \iint_E \psi\big(\tfrac{1}{2}w^\flat_t(x,y)\,;\, u_t(x), u_t(y)\big)\vartheta(dx, dy)dt$$

$$= \frac{1}{2}\int_0^T \iint_E \left\{\frac{\alpha_t}{2}\Psi\left(\frac{w^\flat_t}{\alpha_t}\right) + \frac{\alpha_t}{2}\Psi^*\left(-\overline{\nabla}\phi'(u_t)\right) + \frac{1}{2}w^\flat_t\overline{\nabla}\phi'(u_t)\right\} d\vartheta\, dt$$

$$= \frac{1}{2}\int_0^T \iint_E \frac{1}{2}\left\{\Psi\left(\frac{2dj^\flat_t}{dv_{\rho_t}}\right) + \Psi^*\left(-\overline{\nabla}\phi'(u_t)\right)\right\} dv_{\rho_t}\, dt + \frac{1}{2}\mathscr{E}(\rho_T) - \frac{1}{2}\mathscr{E}(\rho_0). \tag{A.4}$$

In the last identity we used the fact that since $\overline{\mathrm{div}}\, j^\flat_t = -\partial_t\rho_t$, formally we have

$$\int_0^T \iint_E \frac{1}{2}w^\flat_t\overline{\nabla}\phi'(u_t)\, d\vartheta dt = \int_0^T \iint_E \overline{\nabla}\phi'(u_t)\, dj^\flat_t\, dt = \int_0^T \langle\phi'(u_t), \partial_t\rho_t\rangle\, dt = \mathscr{E}(\rho_T) - \mathscr{E}(\rho_0).$$

The expression on the right-hand side of (A.4) is one half times the functional $\mathscr{L}$ defined in (1.18) (see also (1.21)). This proves that

$$\inf_{\boldsymbol{j}' \in A_j} I(\rho, \boldsymbol{j}') = \frac{1}{2}\mathscr{L}(\rho, \boldsymbol{j}^{\flat}).$$

From convexity of Ψ and symmetry of $\mathbf{v}_{\rho}$ we deduce that $\mathscr{L}(\rho, \boldsymbol{j}^{\flat}) \leq \mathscr{L}(\rho, \boldsymbol{j})$ for any $\boldsymbol{j}$; see Remark 4.12. The identity $\mathscr{L}(\rho, \boldsymbol{j}^{\flat}) = \inf_{\boldsymbol{j}' \in A_j} \mathscr{L}(\rho, \boldsymbol{j}')$ then follows immediately; this proves (A.3).

To prove the second part of the Lemma, we write

$$\inf_{\boldsymbol{j}} \left\{ I(\rho, \boldsymbol{j}) : \partial_t \rho + \overline{\mathrm{div}}\,\boldsymbol{j} = 0 \right\} = \inf_{\boldsymbol{j}} \left\{ \left[\inf_{\boldsymbol{j}' \in A_j} I(\rho, \boldsymbol{j}') \right] : \partial_t \rho + \overline{\mathrm{div}}\,\boldsymbol{j} = 0 \right\},$$

$$= \inf_{\boldsymbol{j}} \left\{ \left[\inf_{\boldsymbol{j}' \in A_j} \tfrac{1}{2}\mathscr{L}(\rho, \boldsymbol{j}') \right] : \partial_t \rho + \overline{\mathrm{div}}\,\boldsymbol{j} = 0 \right\},$$

$$= \inf_{\boldsymbol{j}} \left\{ \tfrac{1}{2}\mathscr{L}(\rho, \boldsymbol{j}^{\flat}) : \partial_t \rho + \overline{\mathrm{div}}\,\boldsymbol{j} = 0 \right\},$$

$$= \inf_{\boldsymbol{j}} \left\{ \tfrac{1}{2}\mathscr{L}(\rho, \boldsymbol{j}^{\flat}) : \partial_t \rho + \overline{\mathrm{div}}\,\boldsymbol{j}^{\flat} = 0 \right\}.$$

This concludes the proof.　　　　　　　　　　　　　　　　　　　　　　　　　□

Appendix B: Continuity equation

In this Section we complete the analysis of the continuity equation by carrying out the proofs of Lemma 4.4 and Corollary 4.3.

Proof of Lemma 4.4 The distributional identity (4.6) yields that for every $\zeta \in C_b(V, \tau)$ the map

$$t \mapsto \rho_t(\zeta) := \int_V \zeta(x)\rho_t(\mathrm{d}x) \quad \text{belongs to } W^{1,1}(a, b),$$

with distributional derivative

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho_t(\zeta) = \iint_E \overline{\nabla}\zeta \,\mathrm{d}\boldsymbol{j}_t = -\int_V \zeta \,\mathrm{d}\overline{\mathrm{div}}\,\boldsymbol{j}_t \quad \text{for almost all } t \in [a, b]. \qquad (\text{B.1})$$

Hence, setting $\eth_t := |\overline{\mathrm{div}}\,\boldsymbol{j}_t| \in \mathcal{M}^+(V)$, we have

$$\left| \frac{\mathrm{d}}{\mathrm{d}t}\rho_t(\zeta) \right| \leq \int_V |\zeta| \,\mathrm{d}\eth_t \leq \|\zeta\|_{C_b(V)} |\overline{\mathrm{div}}\,\boldsymbol{j}_t|(V) \leq 2\|\zeta\|_{C_b(V)} |\boldsymbol{j}_t|(E), \qquad (\text{B.2})$$

where we used the fact that

$$\eth_t = |\mathsf{x}_\sharp(\boldsymbol{j}_t - \mathsf{s}_\sharp \boldsymbol{j}_t)| = |\mathsf{x}_\sharp \boldsymbol{j}_t - \mathsf{y}_\sharp \boldsymbol{j}_t| \leq |\mathsf{x}_\sharp \boldsymbol{j}_t| + |\mathsf{y}_\sharp \boldsymbol{j}_t|$$

which implies

$$\eth_t(V) \leq 2|\boldsymbol{j}_t|(E).$$

Hence, the set L_ζ of the Lebesgue points of $t \mapsto \rho_t(\zeta)$ has full Lebesgue measure. Choosing $\zeta \equiv 1$ one immediately recognizes that $\rho_t(V)$ is (essentially) constant: it is not

restrictive to normalize it to 1 for convenience. Let us now consider a countable set $Z = \{\zeta_k\}_{k \in \mathbb{N}}$ of uniformly bounded functions in $C_b(V)$ such that

$$|\zeta_k| \leq 1, \quad d(\mu, \nu) := \sum_{k=1}^{\infty} 2^{-k} \left| \int_V \zeta_k \, d(\mu - \nu) \right|$$

is a distance inducing the weak topology in $\mathcal{M}^+(V)$ (see e.g. [4, § 5.1.1]). By introducing the set $L_Z := \bigcap_{\zeta \in Z} L_\zeta$, it follows from (B.2) that

$$d(\rho_s, \rho_t) \leq 2 \int_s^t |j_r|(E) \, dr \tag{B.3}$$

showing that the restriction of ρ to L_Z is continuous in $\mathcal{M}^+(V)$. Estimate (B.2) also shows that for all $s, t \in L_Z$ with $s \leq t$ we have

$$|\rho_t(\zeta) - \rho_s(\zeta)| \leq \int_s^t \int_V |\zeta| \, d\eth_r \, dr \leq 2\|\zeta\|_{C_b(V)} \int_s^t |j_r|(E) \, dr \quad \text{for all } \zeta \in C_b(V). \tag{B.4}$$

Taking the supremum with respect to ζ we obtain

$$\|\rho_t - \rho_s\|_{TV} \leq 2 \int_s^t |j_r|(E) \, dr \quad \text{and all } s, t \in L_Z, \ s \leq t, \tag{B.5}$$

which shows that the measures $(\rho_t)_{t \in L_Z}$ are uniformly continuous with respect to the total variation metric in $\mathcal{M}^+(V)$ and thus can be extended to an absolutely continuous curve $\tilde{\rho} \in AC(I; \mathcal{M}^+(V))$ satisfying (B.5) for every $s, t \in I$.

When $\varphi \in C_b(V)$, (4.4) immediately follows from (B.1). By a standard argument based on the functional monotone class Theorem [9, §2.12] we can extend the validity of (4.4) to every bounded Borel function.

If $\varphi \in C^1([a, b]; B_b(V))$, combining (B.1) and the fact that the map $t \mapsto \int_V \varphi(t, x) \, \tilde{\rho}_t(dx)$ is absolutely continuous we easily get (4.8). □

Proof of Corollary 4.3 Keeping the same notation of the previous proof, if we define

$$\gamma := \rho_0 + \int_0^T \eth_t \, dt$$

then the estimate (B.2) shows that

$$\rho_t(B) \leq \gamma(B) \quad \text{for every } B \in \mathfrak{B},$$

thus showing that $\rho_t = \tilde{u}_t \gamma$ for every $t \in [0, T]$ and

$$\|\rho_t - \rho_s\|_{TV} = \int_V |\tilde{u}_t - \tilde{u}_s| \, d\gamma \leq 2 \int_s^t |j_r|(E) \, dr \quad \text{for every } 0 \leq s < t \leq T. \tag{B.6}$$

□

We conclude with a result on the decomposition of the measure $j - s_\# j = 2j^\flat$ into its positive and negative part.

Lemma B.1 *If $j \in \mathcal{M}(E)$ and we set*

$$j^+ := (j - s_\# j)_+, \quad j^- := (j - s_\# j)_-, \tag{B.7}$$

then we have

$$j^- = s_\# j^+, \quad \overline{\mathrm{div}}\, j^+ = \overline{\mathrm{div}}\, j. \tag{B.8}$$

When j is skew-symmetric, we also have

$$j^+ = 2j_+, \quad j^- = -2j_-. \tag{B.9}$$

Proof By definition, we have $j^+ = 2j_+^\flat$, $j^- = 2j_-^\flat$. Furthermore, $j^\flat = -s_\# j^\flat = s_\# j_-^\flat - s_\# j_+^\flat$, where the first equality follows from the fact that $j^\flat$ is skew-symmetric. Since $s_\# j_-^\flat \perp s_\# j_+^\flat$ we deduce that $s_\# j_+^\flat = j_-^\flat$, $s_\# j_-^\flat = j_+^\flat$ and $j^\flat = j_+^\flat - s_\# j_+^\flat$, so that $\overline{\mathrm{div}}\, j = \overline{\mathrm{div}}\, j^\flat = 2\overline{\mathrm{div}}\, j_+^\flat = \overline{\mathrm{div}}\, j^+$. $\qquad \square$

Appendix C: Slowly increasing superlinear entropies

The main result of this Section is Lemma C.3 ahead, invoked in the proof of Proposition 4.21. It provides the construction of a *smooth* function estimating the entropy density ϕ from below and such that the function $(r, s) \mapsto \Psi^*(A_\omega(r, s))\alpha(r, s)$ fulfills a suitable bound, cf. (C.10) ahead. Prior to that, we prove the preliminary Lemmas C.1 and C.2 below.

Lemma C.1 *Let us suppose that α satisfies Assumptions ($\mathscr{R}^*\Psi\alpha$). Then for every $a \geq 0$*

$$\lim_{r \to +\infty} \frac{\alpha(r, a)}{r} = \lim_{r \to +\infty} \frac{\alpha(a, r)}{r} = 0. \tag{C.1}$$

Proof Since α is symmetric it is sufficient to prove the first limit. Let us first observe that the concavity of α yields the existence of the limit since the map $r \mapsto r^{-1}(\alpha(r, a) - \alpha(0, a))$ is decreasing, so that

$$\lim_{r \to +\infty} \frac{\alpha(r, a)}{r} = \lim_{r \to +\infty} \frac{\alpha(r, a) - \alpha(0, a)}{r} = \inf_{r > 0} \frac{\alpha(r, a) - \alpha(0, a)}{r}.$$

Let us call $L(a) \in \mathbb{R}_+$ the above quantity. The inequality (following by the concavity of α and the fact that $\alpha(0, 0) \geq 0$)

$$\alpha(r, a) \leq \lambda\alpha(r/\lambda, a/\lambda) \quad \text{for every } \lambda \geq 1 \tag{C.2}$$

yields

$$L(a) = \lim_{r \to +\infty} \frac{\alpha(r, a)}{r} \leq \lim_{r \to +\infty} \frac{\alpha(r/\lambda, a/\lambda)}{r/\lambda} = L(a/\lambda) \quad \text{for every } \lambda \geq 1. \tag{C.3}$$

For every $b \in (0, a)$ and $r > 0$, setting $\lambda := a/b > 1$, we thus obtain

$$L(a) \leq L(b) \leq \frac{\alpha(r, b) - \alpha(0, b)}{r}$$

Passing first to the limit as $b \downarrow 0$ and using the continuity of α we get

$$L(a) \leq \frac{\alpha(r, 0) - \alpha(0, 0)}{r} \quad \text{for every } r > 0.$$

Eventually, we pass to the limit as $r \uparrow +\infty$ and we get $L(a) \leq \alpha^\infty(1, 0) = 0$ thanks to (3.13). $\qquad \square$

Lemma C.2 *Let* $f : \mathbb{R}_+ \to \mathbb{R}_+$ *be an increasing continuous function and* $f_0 \geq 0$ *with*

$$\lim_{r \to +\infty} f(r) = \sup f = +\infty, \qquad \liminf_{r \downarrow 0} \frac{f(r) - f_0}{r} \in (0, +\infty]. \qquad (C.4)$$

Then for every $g_0 \in [0, f_0]$ *there exists a* C^∞ *concave function* $g : \mathbb{R}_+ \to \mathbb{R}_+$ *such that*

$$\forall r \in \mathbb{R}_+ : g(r) \leq f(r), \qquad g(0) = g_0, \qquad \lim_{r \to +\infty} g(r) = +\infty. \qquad (C.5)$$

Proof. By subtracting f_0 and g_0 from f and g, respectively, it is not restrictive to assume $f_0 = g_0 = 0$. We will use a recursive procedure to construct a concave piecewise-linear function g satisfying (C.5); a standard regularization yields a C^∞ map.

We set

$$a := \frac{1}{3} \liminf_{r \downarrow 0} \frac{f(r)}{r}, \quad x_1 := \sup \left\{ x \in (0, 1] : f(r) \geq 2ar \text{ for every } r \in (0, x] \right\}, \qquad (C.6)$$

and $\delta := ax_1$. We consider a strictly increasing sequence $(x_n)_{n \in \mathbb{N}}, n \in \mathbb{N}$, defined by induction starting from $x_0 = 0$ and x_1 as in (C.6), according to

$$x_{n+1} := \min \left\{ x \geq 2x_n - x_{n-1} : f(x) \geq f(x_n) + \delta \right\}, \quad n \geq 1. \qquad (C.7)$$

Since $\lim_{r \to +\infty} f(r) = +\infty$, the minimizing set in (C.7) is closed and not empty, so that the algorithm is well defined. It yields a sequence x_n satisfying

$$x_{n+1} - x_n \geq x_n - x_{n-1}, \quad x_{n+1} \geq x_n + \delta \quad \text{for every } n \geq 0, \qquad (C.8)$$

so that $(x_n)_{n \in \mathbb{N}}$ is strictly increasing and unbounded, and induces a partition $\{0 = x_0 < x_1 < x_1 < \cdots < x_n < \cdots\}$ of $\mathbb{R}_+$. We can thus consider the piecewise linear function $g : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$g(x_n) := n\delta, \quad g((1-t)x_n + tx_{n+1}) := (n+t)\delta \quad \text{for every } n \in \mathbb{N}, \ t \in [0, 1]. \qquad (C.9)$$

We observe that g is increasing, $\lim_{r \to +\infty} g(r) = +\infty$ and it is concave since

$$\frac{g(x_{n+1}) - g(x_n)}{x_{n+1} - x_n} = \frac{\delta}{x_{n+1} - x_n} \overset{(C.8)}{\leq} \frac{\delta}{x_n - x_{n-1}} = \frac{g(x_n) - g(x_{n-1})}{x_n - x_{n-1}}.$$

Furthermore, g is also dominated by f: in the interval $[x_0, x_1]$ this follows by (C.6). For $x \in [x_n, x_n + 1]$ and $n \geq 1$, we observe that (C.7) yields $f(x_{n+1}) \geq f(x_n) + \delta$ so that by induction $f(x_n) \geq (n+1)\delta$; on the other hand

$$\text{for every } x \in [x_n, x_{n+1}] : \quad g(x) \leq g(x_{n+1}) = (n+1)\delta \leq f(x_n) \leq f(x). \qquad \square$$

Lemma C.3 *Let* Ψ^*, α *be satisfying Assumptions* $(\mathscr{R}^*\Psi\alpha)$ *and let* $\beta : \mathbb{R}_+ \to \mathbb{R}_+$ *be a convex superlinear function with* $\beta'(r) \geq \beta'_0 > 0$ *for a.e.* $r \in \mathbb{R}_+$. *Then, there exists a* C^∞ *convex superlinear function* $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ *such that*

$$\omega(r) \leq \beta(r), \qquad \Psi^*(\omega'(s) - \omega'(r))\alpha(s, r) \leq r + s \quad \text{for every } r, s \in \mathbb{R}_+. \qquad (C.10)$$

Proof By a standard regularization, we can always approximate β by a smooth convex superlinear function $\tilde{\beta} \leq \beta$ whose derivative is strictly positive, so that it is not restrictive to assume that β is of class C^2. Let us set $r_0 := \inf\{r > 0 : \Psi^*(r) > 0\}$ and let $P : (0, +\infty) \to (r_0, +\infty)$ be the inverse map of Ψ^*: P is continuous, strictly increasing, and of class C^1.

Since α is concave, the function $x \mapsto \alpha(x, 1)/x$ is nonincreasing in $(0, +\infty)$; we can thus define the nondecreasing function $Q(x):=P(x/\alpha(x, 1))$ and the function

$$\gamma(x):=2g_0 + \int_1^x \min(\beta''(y), Q'(y))\,dy \quad \text{for every } x \geq 1, \quad g_0:=\frac{1}{2}\min(\beta'_0, Q(1)) > 0.$$

By construction $\gamma(1) = 2g_0 = \min(\beta'_0, Q(1)) \leq \beta'(1)$ so that $\gamma(x) \leq \min(\beta'(x), Q(x))$ for every $x \geq 1$. We eventually set

$$f(t):=\frac{e^t}{\gamma(e^t)} \quad t \geq 0.$$

Clearly, we have $f(0) = 2g_0$. Furthermore, we combine the estimate $\gamma(e^t) \leq Q(e^t) = P(e^t/\alpha(e^t, 1))$ with the facts that $e^t/\alpha(e^t, 1) \to +\infty$ as $t \to +\infty$, thanks to Lemma C.1, and that P has sublinear growth at infinity, being the inverse function of Ψ^*. All in all, we conclude that

$$\lim_{t \to +\infty} f(t) = +\infty.$$

Therefore, we are in a position to apply Lemma C.2, obtaining an increasing concave function $g : \mathbb{R}_+ \to \mathbb{R}_+$ such that $g_0 = g(0) \leq g(t) \leq f(t)$ and $\lim_{t \to +\infty} g(t) = +\infty$. Since $g(0) \geq 0$, the concaveness of g yields $g(t'') - g(t') \leq g(t'' - t')$ for every $0 \leq t' \leq t''$, so that the function $h(x):=g(\log(x \vee 1))$ satisfies $h(x) = g_0 \leq \beta'(x)$ for $x \in [0, 1]$, and

$$h(z) \leq \min(\beta'(z), Q(z)) \quad \text{for every } z \geq 1, \quad h(y) - h(x) \leq h(y/x) \quad \text{for every } 0 < x \leq y. \tag{C.11}$$

In fact, if $x \leq 1$ we get

$$h(y) - h(x) = h(y) - g_0 \leq h(y) \leq h(y/x)$$

and if $x \geq 1$ we get

$$h(y) - h(x) \leq g(\log y) - g(\log x) \leq g(\log y - \log x) = g(\log(y/x)) = h(y/x).$$

Let us now define the convex function $\omega(x):=\int_0^x h(y)\,dy$ with $\omega(0) = 0$ and $\omega' = h$. In particular $\omega(x) \leq \beta(x)$ for every $x \geq 0$.

It remains to check the second inequality of (C.10). The case $r, s \leq 1$ is trivial since $\omega'(s) - \omega'(r) = h(r) - h(s) = 0$. We can also consider the case $\omega'(r) \neq \omega'(s)$ and $\alpha(r, s) > 0$; since (C.10) is also symmetric, it is not restrictive to assume $r \leq s$; by continuity, we can assume $r > 0$.

Recalling that $\alpha(s, r) \leq r\alpha(s/r, 1)$ if $0 < r \leq s$, and $(r + s)/r > s/r$, (C.10) is surely satisfied if

$$\Psi^*(\omega'(s) - \omega'(r))\alpha(s/r, 1) \leq s/r \quad \text{for every } 0 < r < s. \tag{C.12}$$

Recalling that $\omega'(s) - \omega'(r) \leq \omega'(s/r)$ by (C.11) and Ψ^* is nondecreasing, (C.12) is satisfied if

$$\Psi^*(\omega'(s/r))\alpha(s/r, 1) \leq s/r \quad \text{for every } 0 < r < s. \tag{C.13}$$

After the substitution $t:=r/s$, (C.13) corresponds to

$$\omega'(t) \leq P(t/\alpha(t, 1)) = Q(t) \quad \text{for every } t \geq 1, \tag{C.14}$$

which is a consequence of the first inequality of (C.11). $\qquad\square$

Appendix D: Connectivity by curves of finite action

Preliminarily, with the reference measure $\pi \in \mathcal{M}_+(V)$ and with the 'jump equilibrium rate' $\boldsymbol{\vartheta}$ from (3.5) we associate the 'graph divergence' operator $\overline{\mathrm{div}}_{\pi,\boldsymbol{\vartheta}} : L^p(E; \boldsymbol{\vartheta}) \to L^p(V; \pi)$, $p \in [1, +\infty]$, defined as the transposed of the 'graph gradient' $\overline{\nabla} : L^q(V; \pi) \to L^q(E; \boldsymbol{\vartheta})$, with $q = p'$. Namely

$$\text{for } \zeta \in L^p(E; \boldsymbol{\vartheta}), \quad \xi = -\overline{\mathrm{div}}_{\pi,\boldsymbol{\vartheta}}(\zeta) \quad \text{if and only if}$$

$$\int_V \xi(x)\omega(x)\pi(\mathrm{d}x) = \int_E \zeta(x,y)\overline{\nabla}\omega(x,y)\boldsymbol{\vartheta}(\mathrm{d}x,\mathrm{d}y) \quad \text{for all } \omega \in L^q(V; \pi)$$

or, equivalently,

$$\xi\pi = -\overline{\mathrm{div}}(\zeta\boldsymbol{\vartheta}) \tag{D.1}$$

(with $\overline{\mathrm{div}}$ the divergence operator from (1.6)) in the sense of measures.

We can now first address the connectivity problem in the very specific setup

$$\alpha(u,v) \equiv 1 \quad \text{for all } (u,v) \in [0, +\infty) \times [0, +\infty). \tag{D.2}$$

Then, the action functional $\int \mathcal{R}$ is translation-invariant. Let us consider two measures $\rho_0, \rho_1 \in \mathcal{M}_+(V)$ such that for $i \in \{0, 1\}$ there holds $\rho_i = u_i\pi$ with $u_i \in L^p_+(V; \pi)$ for some $p \in (1, +\infty)$. Thus, we look for curves $\rho \in \mathscr{A}(0, \tau; \rho_0, \rho_1)$, with finite action, such that $\rho_t \ll \pi$, with density u_t, for almost all $t \in (0, \tau)$. Consequently, any flux $(\boldsymbol{j}_t)_{t\in(0,\tau)}$ shall satisfy $\boldsymbol{j}_t \ll \boldsymbol{\vartheta}$ for a.a. $t \in (0, \tau)$ (cf. Lemma 4.10). Taking into account (D.1), the continuity equation reduces to

$$\dot{u}_t = -\overline{\mathrm{div}}_{\pi,\boldsymbol{\vartheta}}(\zeta_t) \quad \text{for a.e. } t \in (0, \tau) \tag{D.3}$$

with $\zeta_t = \frac{\mathrm{d}\boldsymbol{j}_t}{\mathrm{d}\boldsymbol{\vartheta}}$. Furthermore, we look for a connecting curve $\rho_t = u_t\pi$ with $u_t = (1-t)u_0 + tu_1$, so that (D.3) becomes $-\overline{\mathrm{div}}_{\pi,\boldsymbol{\vartheta}}(\zeta_t) \equiv u_1 - u_0$. Hence, we can restrict to flux densities that are constant in time, i.e. $\zeta_t \equiv \zeta$ with $\zeta \in L^p(E; \boldsymbol{\vartheta})$. In this specific context, and if we further confine the discussion to the case $\Psi(r) = \frac{1}{p}|r|^p$ for $p \in (1, +\infty)$, the minimal action problem becomes

$$\inf \left\{ \frac{1}{p} \int_E |w|^p \, \boldsymbol{\vartheta}(\mathrm{d}x, \mathrm{d}y) \, : \, w = 2\zeta \in L^p(E; \boldsymbol{\vartheta}), \, -\overline{\mathrm{div}}_{\pi,\boldsymbol{\vartheta}}(\zeta) \equiv u_1 - u_0 \right\} \tag{D.4}$$

Now, by a general duality result on linear operators, the operator $-\overline{\mathrm{div}}_{\pi,\boldsymbol{\vartheta}} : L^p(E; \boldsymbol{\vartheta}) \to L^p(V; \pi)$ is surjective if and only if the graph gradient $\overline{\nabla} : L^q(V; \pi) \to L^q(E; \boldsymbol{\vartheta})$ fulfills the following property:

$$\exists C > 0 \; \forall \xi \in L^q(V; \pi) \text{ with } \int_V \xi\pi(\mathrm{d}x) = 0 \text{ there holds } \|\xi\|_{L^q(V;\pi)} \leq C\|\overline{\nabla}\xi\|_{L^q(E;\boldsymbol{\vartheta})},$$

namely the q-Poincaré inequality (4.84). We can thus conclude the following result.

Lemma D.1 *Suppose that $\alpha \equiv 1$, that Ψ has p-growth (cf. (4.85)), and that the measures $(\pi, \boldsymbol{\vartheta})$ satisfy a q-Poincaré inequality for $q = \frac{p}{p-1}$. Let $\rho_0, \rho_1 \in \mathcal{M}^+(V)$ be given by $\rho_i = u_i\pi$, with positive $u_i \in L^p(V; \pi)$, for $i \in \{0, 1\}$. Then, for every $\tau \in (0, 1)$ we have $\mathscr{W}(\tau, \rho_0, \rho_1) < +\infty$. If $\Psi(r) = \frac{1}{p}|r|^p$, q-Poincaré inequality is also necessary for having $\mathscr{W}(\tau, \rho_0, \rho_1) < +\infty$.*

We are now in a position to carry out the

Proof of Proposition 4.25 Assume that $\rho_0(V) = \int_V u_0(x)\pi(dx) = \pi(V)$. Hence, it is sufficient to provide a solution for the connectivity problem between u_0 and $u_1 \equiv 1$. We may also assume without loss of generality that $\alpha(u, v) \geq \alpha_0(u, v)$ with $\alpha_0(u, v) = c_0 \min(u, v, 1)$ for some $c_0 > 0$, so that

$$\Psi\left(\frac{w}{\alpha(u, v)}\right) \alpha(u, v) \leq \Psi\left(\frac{w}{\alpha_0(u, v)}\right) \alpha_0(u, v) \leq C_p \left(1 + \left|\frac{w}{\alpha_0(u, v)}\right|^p\right) \alpha_0(u, v)$$
$$\leq C_p c_0 + C_p |w|^p (\alpha_0(u, v))^{1-p},$$
$$(D.5)$$

where the first estimate follows from the convexity of Ψ and the fact that $\Psi(0) = 0$, yielding that $\lambda \mapsto \lambda\Psi(w/\lambda)$ is non-increasing. It is therefore sufficient to consider the case in which $c_0 = C_p = 1$, $\alpha_0(u, v) = \min(u, v, 1)$, and to solve the connectivity problem for $\tilde{\Psi}(r) = \frac{1}{p}|r|^p$. By Lemma D.1, we may first find $w \in L^p(E; \vartheta)$ solving the minimum problem (D.4) in the case $\alpha \equiv 1$, so that the flux density $\zeta_t \equiv \frac{1}{2}w$ is associated with the curve $u_t = (1-t)u_0 + tu_1$, $t \in [0, \tau]$. Then, we fix an exponent $\gamma > 0$ and we consider the rescaled curve $\tilde{u}_t := u_{t^\gamma}$, that fulfills $\partial_t \tilde{u}_t = -\overline{\text{div}}_{\pi,\vartheta}(\tilde{\zeta}_t)$ with $\tilde{\zeta}_t = \frac{1}{2}\tilde{w}_t = \frac{1}{2}\gamma t^{\gamma-1}w$. Moreover,

$$\alpha_0(\tilde{u}_t(x), \tilde{u}_t(y)) = \min\{(1-t^\gamma)u_0(x) + t^\gamma u_1(x), (1-t^\gamma)u_0(y) + t^\gamma u_1(y), 1\}$$
$$\geq \min(t^\gamma, 1) = t^\gamma$$

since $u_1(x) = u_1(y) = 1$. By (D.5) we thus get

$$\int_E \Psi\left(\frac{\tilde{w}_t(x, y)}{\alpha(\tilde{u}_t(x), \tilde{u}_t(y))}\right) \alpha(\tilde{u}_t(x), \tilde{u}_t(y))\vartheta(dx, dy)$$
$$\leq C_p c_0 \vartheta(E) + \int_E \gamma^p t^{p(\gamma-1)}|w(x, y)|^p t^{\gamma(1-p)}\vartheta(dx, dy) = C_p c_0 \vartheta(E) + \gamma^p t^{\gamma-p}\|w\|^p_{L^p(E;\vartheta)}.$$

Choosing $\gamma > p - 1$ we conclude that

$$\int_0^\tau \int_E \Psi\left(\frac{\tilde{w}_t(x, y)}{\alpha(\tilde{u}_t(x), \tilde{u}_t(y))}\right) \alpha(\tilde{u}_t(x), \tilde{u}_t(y))\vartheta(dx, dy) < +\infty$$

hence $\mathscr{A}(0, \tau; \rho_0, \rho_1) \neq \emptyset$. □

Appendix E: Convexity of D_ϕ induced by power means

Here, we prove the claim that D_ϕ is convex when ϕ is the Boltzmann entropy (1.5) and the pair (α, Ψ^*) is induced by the power means $\mathfrak{m}_p$ (see Example 1 in Sect. 1.3), for $p \in [-\infty, -1] \cup [0, 1)$. More precisely, we consider the case, $\alpha(u, v) := \mathfrak{m}_p(u, v) = v\mathfrak{f}_p(u/v)$, $p \in [-\infty, 1)$, where $\mathfrak{f}_p$ is the concave generating function

$$\mathfrak{f}_p(r) := \begin{cases} 2^{-1/p}(r^p + 1)^{1/p} & \text{if } p \in (-\infty, 1) \setminus \{0\}, \\ \sqrt{r} & \text{if } p = 0, \\ \min(r, 1) & \text{if } p = -\infty, \end{cases} \quad r > 0,$$

and

$$\Psi_p^*(\xi):=2^{1/p}\int_1^{\exp\xi}\frac{r-1}{\mathfrak{f}_p(r)}\frac{\mathrm{d}r}{r},\qquad \xi\in\mathbb{R}.$$

We consider here only the case $p<1$ since $\mathfrak{m}_p$ is not concave for $p>1$ and Ψ_p^* is not superlinear for $p=1$. We have for $u,v>0$

$$D_{p,\phi}(u,v)=vK_p(u/v)\quad\text{where}\quad K_p(\xi):=\mathfrak{f}_p(\xi)\Psi_p^*(\ln\xi)\quad\xi>0,$$

so that

$$D_{p,\phi}(u,v)=(u^p+v^p)^{1/p}\int_1^{u/v}\frac{r-1}{(r^p+1)^{1/p}}\frac{\mathrm{d}r}{r}\quad\text{if }p\in(-\infty,1)\setminus\{0\},$$

$$D_{0,\phi}(u,v)=2\big(\sqrt{u}-\sqrt{v}\big)^2,$$

$$D_{-\infty,\phi}(u,v)=|u-v|-\min(u,v)|\ln u-\ln v|.$$

$D_{p,\phi}$ takes a more explicit and remarkable form also in the case of the harmonic mean $p=-1$:

$$D_{-1,\phi}(u,v)=\frac{(u-v)^2}{u+v}.$$

It is also interesting to note that $D_{p,\phi}$ has a natural lower semicontinuous extension to the boundary $\{0\}\times(0,\infty)\cup(0,\infty)\times\{0\}=\partial(0,\infty)^2\setminus\{(0,0)\}$ given by

$$D_{p,\phi}(u,v)=\begin{cases}+\infty & \text{if }p>0,\\ 2(u+v) & \text{if }p=0,\\ u+v & \text{if }p<0\end{cases}\qquad (u,v)\in\partial(0,\infty)^2\setminus\{(0,0)\}.$$

The main result of this section is summarized in the following lemma (where we just consider the range $p\in[-\infty,1)$).

Lemma E.1 $D_{p,\phi}$ *is convex on* $(0,\infty)^2$ *if and only if* $p\in[-\infty,-1]\cup[0,1)$.

Proof The cases $p=0$ and $p=-\infty$ can be easily checked, so that we assume $p\in(-\infty,1)\setminus\{0\}$. It is not difficult to see that the convexity of $D_{p,\phi}$ on $(0,\infty)^2$ is determined by the convexity of K_p on $(0,\infty)$.

Hence, it suffices to check that $K_p''(\xi)\geq0$ for $\xi>0$.

Elementary computations yield

$$K_p''(\xi)=\xi^{p-2}(\xi^p+1)^{1/p-2}\underbrace{\left\{(\xi^p+1)^{1-1/p}\big(\xi^{-p}+\xi\big)+(p-1)\int_1^\xi\frac{r-1}{(r^p+1)^{1/p}}\frac{\mathrm{d}r}{r}\right\}}_{=:H_p(\xi)}.$$

Therefore the assertion holds if and only if $H_p(\xi)\geq0$ for $\xi>0$. Notice that $H_p(1)=2^{2-1/p}\geq0$. We now show that if $p\in(-\infty,-1]\cup(0,1)$ then H_p takes its minimum at $\xi=1$. Since

$$H_p'(\xi)=2\,p(\xi^p+1)^{-1/p}\xi^{-1/2}\underbrace{\left\{\frac{\xi^{1/2}-\xi^{-1/2}}{2}+\frac{\xi^{p+1/2}-\xi^{-(p+1/2)}}{2}\right\}}_{=:G_p(\xi)},$$

we can make a change of variables $\xi=\exp(\eta)$, $\eta\in\mathbb{R}$, to obtain

$$G_p(e^\eta) = \tilde{G}_p(\eta) = \sinh(\eta/2) + \sinh((p+1/2)\eta).$$

Since $\tilde{G}_p(\eta)$ is an odd function, H_p takes its minimum at $\xi = 1$ if $\tilde{G}_p(\eta) \geq 0$ for $\eta > 0$. This property clearly holds if $p \in (0, 1)$. On the other hand, if $p < 0$, $\tilde{G}_p(\eta) \geq 0$ if and only if $(1 + p)\eta \leq 0$ for all $\eta \geq 0$, i.e. the cases $p \leq -1$ apply. Altogether, we conclude that $H_p(\xi) \geq H_p(1) \geq 0$ and therefore $K_p''(\xi) \geq 0$ for every $\xi > 0$ if $p \in (-\infty, -1] \cup (0, 1)$.

If $p \in (-1, 0)$ then

$$(\xi^p + 1)^{1-1/p} \left(\xi^{-p} + \xi\right) = \xi + o(\xi) \quad \text{as } \xi \to +\infty$$

and

$$\int_1^\xi \frac{r-1}{(r^p + 1)^{1/p}} \frac{\mathrm{d}r}{r} = \xi + o(\xi) \quad \text{as } \xi \to +\infty$$

so that

$$H_p(\xi) = p\xi + o(\xi) \to -\infty \quad \text{as } \xi \to +\infty.$$

It follows that K_p'' takes strictly negative values in $(0, +\infty)$ and therefore it is not convex.
$\square$

References

1. Adams, S., Dirr, N., Peletier, M.A., Zimmer, J.: From a large-deviations principle to the Wasserstein gradient flow: a new micro-macro passage. Commun. Math. Phys. **307**, 791–815 (2011)
2. Adams, S., Dirr, N., Peletier, M.A., Zimmer, J.: Large deviations and gradient flows. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. **371**(2005), 20120341 (2013)
3. Ambrosio, L., Fusco, N., Pallara, D.: Functions of Bounded Variation and Free Discontinuity Problems. Oxford University Press, Oxford (2005)
4. Ambrosio, L., Gigli, N., Savaré, G.: Gradient Flows in Metric Spaces and in the Space of Probability Measures. Lectures in Mathematics ETH Zürich. Birkhäuser, Basel (2008)
5. Ambrosio, L.: Minimizing movements. Rend. Accad. Naz. Sci. XL Mem. Mat. Appl. **19**(5), 1773–1799 (1995)
6. Arnrich, S., Mielke, A., Peletier, M.A., Savaré, G., Veneroni, M.: Passing to the limit in a Wasserstein gradient flow: from diffusion to reaction. Calc. Var. Partial. Differ. Equ. **44**, 419–454 (2012)
7. Almgren, F., Taylor, J.E., Wang, L.: Curvature-driven flows: a variational approach. SIAM J. Control. Optim. **31**, 387–437 (1993)
8. Benamou, J.-D., Brenier, Y.: A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. Numer. Math. **84**(3), 375–393 (2000)
9. Bogachev, V.I.: Measure Theory, vol. I, II. Springer, Berlin (2007)
10. Bonaschi, G.A., Peletier, M.A.: Quadratic and rate-independent limits for a large-deviations functional. Contin. Mech. Thermodyn. **28**, 1191–1219 (2016)
11. Brezis, H.: Functional Analysis. Sobolev Spaces and Partial Differential Equations. Springer, New York (2011)
12. Bullen, P.S.: Handbook of Means and their Inequalities. Mathematics and its Applications. Springer, Netherlands (2003)
13. Chow, S.-N., Huang, W., Li, Y., Zhou, H.: Fokker–Planck equations for a free energy functional or Markov process on a graph. Arch. Ration. Mech. Anal. **203**(3), 969–1008 (2012)
14. Crandall, M.G., Tartar, L.: Some relations between nonexpansive and order preserving mappings. Proc. Am. Math. Soc. **78**(3), 385–390 (1980)
15. Castaing, C., Valadier, M.: Convex Analysis and Measurable Multifunctions. Lectures Notes in Mathematics, vol. 580. Springer, Berlin, New York (1977)
16. Dondl, P., Frenzel, T., Mielke, A.: A gradient system with a wiggly energy and relaxed EDP-convergence. ESAIM Control Optim. Calc. Var. **25**, 68 (2019)
17. De Giorgi, E., Marino, A., Tosques, M.: Problems of evolution in metric spaces and maximal decreasing curve. Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8) **68**(3), 180–187 (1980)

18. Dellacherie, C., Meyer, P.-A.: Probabilities and potential, volume 29 of North-Holland Mathematics Studies. North-Holland Publishing Co., Amsterdam-New York; North-Holland Publishing Co., Amsterdam-New York (1978)
19. Dal Maso, G., DeSimone, A., Mora, M.G.: Quasistatic evolution problems for linearly elastic-perfectly plastic materials. Arch. Ration. Mech. Anal. **180**(2), 237–291 (2006)
20. Dolbeault, J., Nazaret, B., Savaré, G.: A new class of transport distances between measures. Calc. Var. Part. Differ. Equ. **34**(2), 193–231 (2009)
21. Dolbeault, J., Nazaret, B., Savaré, G.: A new class of transport distances between measures. Calc. Var. Part. Differ. Equ. **34**(2), 193–231 (2009)
22. Duong, M.H., Peletier, M.A., Zimmer, J.: GENERIC formalism of a Vlasov–Fokker–Planck equation and connection to large-deviation principles. Nonlinearity **26**, 2951–2971 (2013)
23. Dirr, N., Stamatakis, M., Zimmer, J.: Entropic and gradient flow formulations for nonlinear diffusion. J. Math. Phys. **57**(8), 081505 (2016)
24. Dembo, A., Zeitouni, O.: Large Deviations Techniques and Applications. Springer, New York (1998)
25. Erbar, M., Fathi, M., Laschos, V., Schlichting, A.: Gradient flow structure for McKean–Vlasov equations on discrete spaces. Disc. Contin. Dyn. Syst. **36**(12), 6799–6833 (2016)
26. Erbar, M., Fathi, M., Schlichting, A.: Entropic curvature and convergence to equilibrium for mean-field dynamics on discrete spaces. arXiv preprint arXiv:1908.03397 (2019)
27. Evans, L.C., Gariepy, R.F.: Measure Theory and Fine Properties of Functions. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL (1992)
28. El Hajj, A., Ibrahim, H., Monneau, R.: Dislocation dynamics: from microscopic models to macroscopic crystal plasticity. Continuum Mech. Thermodyn. **21**(2), 109–123 (2009)
29. Erbar, M., Maas, J.: Gradient flow structures for discrete porous medium equations. Disc. Contin. Dyn. Syst. **34**(4), 1355–1374 (2014)
30. Erbar, M.: Gradient flows of the entropy for jump processes. Ann. Inst. Henri Poincaré Probab. Stat. **50**(3), 920–945 (2014)
31. Erbar, M.: A gradient flow approach to the Boltzmann equation. Arxiv preprint arXiv:01603.00540 (2016)
32. Feinberg, M.: On chemical kinetics of a certain class. Arch. Ration. Mech. Anal. **46**(1), 1–41 (1972)
33. Figalli, A., Gangbo, W., Yolcu, T.: A variational method for a class of parabolic PDEs. Annali della Scuola Normale Superiore di Pisa-Classe di Scienze **10**(1), 207–252 (2011)
34. Gigli, N.: On the heat flow on metric measure spaces: existence, uniqueness and stability. Calc. Var. Partial. Differ. Equ. **39**(1–2), 101–120 (2010)
35. Glitzky, A., Mielke, A.: A gradient structure for systems coupling reaction-diffusion effects in bulk and interfaces. Z. Angew. Math. Phys. **64**(1), 29–52 (2013)
36. Gavish, N., Nyquist, P., Peletier, M.: Large deviations and gradient flows for the Brownian one-dimensional hard-rod system. arXiv preprint arXiv:1909.02054 (2019). https://doi.org/10.1007/s11118-021-09933-0
37. Grmela, M.: Particle and bracket formulations of kinetic equations. In: Marsden, J.E. (ed) Proceedings of the AMS–IMS–SIAM Joint Summer Research Conference in the Mathematical Sciences on Fluids and Plasmas: Geometry and Dynamics, pp 125–132 (1984)
38. Grmela, M.: Multiscale equilibrium and nonequilibrium thermodynamics in chemical engineering. Adv. Chem. Eng. **39**, 75–129 (2010)
39. Girault, V., Wheeler, M.F.: Numerical discretization of a Darcy–Forchheimer model. Numer. Math. **110**(2), 161–198 (2008)
40. Hudson, T., van Meurs, P., Peletier, M.A.: Atomistic origins of continuum dislocation dynamics. Math. Models Methods Appl. Sci. **30**(13), 2557–2618 (2020)
41. Hairer, E., Wanner, G.: Solving ordinary differential equations. II, volume 14 of Springer Series in Computational Mathematics, 2nd edn. Springer, Berlin (1996). Stiff and differential-algebraic problems
42. Jordan, R., Kinderlehrer, D., Otto, F.: The variational formulation of the Fokker–Planck equation. SIAM J. Math. Anal. **29**(1), 1–17 (1998)
43. Kaiser, M., Jack, R.L., Zimmer, J.: Canonical structure and orthogonality of forces and currents in irreversible Markov chains. J. Stat. Phys. **170**(6), 1019–1050 (2018)
44. Knupp, P.M., Lage, J.L.: Generalization of the Forchheimer-extended Darcy flow model to the tensor permeability case via a variational principle. J. Fluid Mech. **299**, 97–104 (1995)
45. Kipnis, C., Olla, S., Varadhan, S.R.S.: Hydrodynamics and large deviation for simple exclusion processes. Commun. Pure Appl. Math. **42**(2), 115–137 (1989)
46. Liero, M., Mielke, A., Peletier, M.A., Renger, D.R.M.: On microscopic origins of generalized gradient structures. Disc. Cont. Dyn. Syst. Ser. S **10**(1), 1 (2017)
47. Liero, M., Mielke, A., Savaré, G.: Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures. Invent. Math. **211**(3), 969–1117 (2018)

48. Luckhaus, S., Sturzenhecker, T.: Implicit time discretization for the mean curvature flow equation. Calc. Var. Partial. Differ. Equ. **3**(2), 253–271 (1995)
49. Maas, J.: Gradient flows of the entropy for finite Markov chains. J. Funct. Anal. **261**(8), 2250–2292 (2011)
50. Martin, R.H.: Jr.: Nonlinear operators and differential equations in Banach spaces. Pure and Applied Mathematics. Wiley, New York (1976)
51. Mielke, A.: Evolution in rate-independent systems. In: Handbook of Differential Equations: Evolutionary Differential Equations, pp 461–559. North-Holland (2005)
52. Mielke, A.: Geodesic convexity of the relative entropy in reversible Markov chains. Calc. Var. Part. Differ. Equ. **48**(1–2), 1–31 (2013)
53. Mielke, A.: Deriving effective models for multiscale systems via evolutionary Γ-convergence. In: Control of Self-Organizing Nonlinear Systems, pp. 235–251. Springer, New York (2016)
54. Mielke, A.: On evolutionary Γ-convergence for gradient systems. In: Macroscopic and large scale phenomena: coarse graining, mean field limits and ergodicity, vol 3 of Lecture Notes Applied Mathematics Mechanical, pp 187–249. Springer, New York (2016)
55. Mielke, A., Montefusco, A., Peletier, M.A.: Exploring families of energy-dissipation landscapes via tilting—three types of EDP convergence. arXiv preprint arXiv:2001.01455 (2020). https://doi.org/10.1007/s00161-020-00932-x
56. Maes, C., Netočný, K.: Canonical structure of dynamical fluctuations in mesoscopic nonequilibrium steady states. Europhys. Lett. **82**(3), 30003 (2008)
57. Mörters, P.: Introduction to large deviations. Technical report, University of Bath (2010)
58. Mielke, A., Patterson, R.I.A., Peletier, M.A., Renger, D.R.M.: Non-equilibrium thermodynamical principles for chemical reactions with mass-action kinetics. SIAM J. Appl. Math. **77**(4), 1562–1585 (2017)
59. Mielke, A., Peletier, M.A., Renger, D.R.M.: On the relation between gradient flows and the large-deviation principle, with applications to Markov chains and diffusion. Potential Anal. **41**(4), 1293–1327 (2014)
60. Mielke, A., Peletier, M.A., Renger, D.R.M.: A generalization of Onsagers reciprocity relations to gradient flows with nonlinear mobility. J. Non-Equil. Thermodyn. **41**(2), 141–149 (2016)
61. Mielke, A., Roubíček, T.: Rate-independent systems. Theory and application, volume 193 of Applied Mathematical Sciences. Springer, New York (2015)
62. Mielke, A., Rossi, R., Savaré, G.: BV solutions and viscosity approximations of rate-independent systems. ESAIM Control Optim. Calc. Variat. **18**(01), 36–80 (2012)
63. Mielke, A., Rossi, R., Savaré, G.: Nonsmooth analysis of doubly nonlinear evolution equations. Calc. Var. Part. Differ. Equ. **46**(1–2), 253–310 (2013)
64. Mielke, A., Rossi, R., Savaré, G.: Balanced viscosity (BV) solutions to infinite-dimensional rate-independent systems. J. Eur. Math. Soc. (JEMS) **18**(9), 2107–2165 (2016)
65. Mirrahimi, S., Souganidis, P.E.: A homogenization approach for the motion of motor proteins. Nonlinear Differ. Equ. Appl. **20**(1), 129–147 (2013)
66. Mielke, A., Stephan, A.: Coarse-graining via EDP-convergence for linear fast-slow reaction systems. Math. Models Methods Appl. Sci. **30**(9), 1765–1807 (2020)
67. Mielke, A., Theil, F., Levitas, V.I.: A variational formulation of rate-independent phase transformations using an extremum principle. Arch. Ration. Mech. Anal. **162**(2), 137–177 (2002)
68. Otto, F.: The geometry of dissipative evolution equations: the porous medium equation. Commun. Part. Differ. Equ. **26**, 101–174 (2001)
69. Öttinger, H.C.: On the combined use of friction matrices and dissipation potentials in thermodynamic modeling. J. Non-Equilib. Thermodyn. **44**(3), 295–302 (2019)
70. Peletier, M.A.: Variational modelling: energies, gradient flows, and large deviations. Arxiv preprint arXiv:1402:1990 (2014)
71. Peletier, M.A., Redig, F., Vafayi, K.: Large deviations in stochastic heat-conduction processes provide a gradient-flow structure for heat conduction. J. Math. Phys. **55**(9), 093301 (2014)
72. Perthame, B., Souganidis, P.E.: Asymmetric potentials and motor effect: a large deviation approach. Arch. Ration. Mech. Anal. **193**(1), 153–169 (2009)
73. Perthame, B., Souganidis, P.E.: Asymmetric potentials and motor effect: a homogenization approach. Ann. de l'Institut Henri Poincare (C) Non-Linear Anal. **26**(6), 2055–2071 (2009)
74. Peletier, M.A., Schlottke, M.C.: Large-deviation principles of switching Markov processes via Hamilton–Jacobi equations. arXiv preprint arXiv:1901.08478 (2019)
75. Renger, D.R.M.: Flux large deviations of independent and reacting particle systems, with implications for macroscopic fluctuation theory. J. Stat. Phys. **172**(5), 1291–1326 (2018)
76. Rossi, R., Savaré, G., Segatti, A., Stefanelli, U.: Weighted energy-dissipation principle for gradient flows in metric spaces. J. Math. Pures Appl. **9**(127), 1–66 (2019)
77. Tyrrell Rockafellar, R., Roger, J.-B.: Wets: Variational Analysis. Springer, Berlin (1998)

78. Renger, M., Zimmer, J.: Orthogonality of fluxes in general nonlinear reaction networks. Disc. Contin. Dyn. Syst. Ser. S (2019)
79. Serfaty, S.: Gamma-convergence of gradient flows on Hilbert and metric spaces and applications. Disc. Contin. Dyn. Syst. **31**(4), 1427–1451 (2011)
80. Sandier, E., Serfaty, S.: Gamma-convergence of gradient flows with applications to Ginzburg–Landau. Commun. Pure Appl. Math. **57**(12), 1627–1672 (2004)