

# Nonlocal cross-interaction systems on graphs: Nonquadratic Finslerian structure and nonlinear mobilities

Georg Heinze<sup>\*,†</sup>, Jan-Frederik Pietschmann<sup>†</sup>, Markus Schmidtchen<sup>‡</sup>

January 12, 2022

## Abstract

We study the evolution of a system of two species with nonlinear mobility and nonlocal interactions on a graph whose vertices are given by an arbitrary, positive measure. To this end, we extend a recently introduced 2-Wasserstein-type quasi-metric on generalized graphs, which is based on an upwind-interpolation, to the case of two-species systems, concave, nonlinear mobilities, and  $p \neq 2$ . We provide a rigorous interpretation of the interaction system as a gradient flow in the Finslerian setting, arising from the new quasi-metric.

---

2010 *Mathematics Subject Classification.* 49J40 (Variational inequalities), 45G10 (Other nonlinear integral equations), 49J45 (Methods involving semicontinuity and convergence; relaxation), 28A33 (Spaces of measures, convergence of measures); 35B38 (Critical points of functionals in context of PDEs (e.g., energy functionals);

---

## 1. Introduction

The goal of this paper is the study of a two-species nonlocal interaction system with nonlinear mobility on a graph. It is well known that, in a local and continuous setting, evolution equations for a single species  $\rho_t(x)$  of the form

$$\partial_t \rho_t = \nabla \cdot (\rho_t \nabla K * \rho_t), \quad (1)$$

can be cast into a Wasserstein gradient flow framework, cf. [48, 15, 32, 17, 42]. Here, the corresponding functional

$$\mathcal{E}(\rho) = \frac{1}{2} \iint_{\mathbb{R}^d \times \mathbb{R}^d} K(x-y) d\rho(x) d\rho(y),$$

denotes the interaction energy which encodes the nature of the interactions among members of the species. The so-called aggregation equation, (1), can be obtained as the mean-field limit of a particle system associated to it

---

\*corresponding author

<sup>†</sup>Fakultät für Mathematik, Technische Universität Chemnitz Reichenhainer Straße 41, Chemnitz, Germany. ({georg.heinze,jfpietschmann}@math.tu-chemnitz.de)

<sup>‡</sup>Institute of Scientific Computing, Technische Universität Dresden, Zellescher Weg 25, 01069 Dresden, Germany. (markus.schmidtchen@tu-dresden.de).

by letting the number of particles,  $N$ , tend to infinity [30, 44, 38, 12]. It is straightforward to introduce a second species to the dynamics such that the energy functional becomes

$$\mathcal{E}(\rho) = \frac{1}{2} \sum_{i,k=1}^2 \iint_{\mathbb{R}^d \times \mathbb{R}^d} K^{(ik)}(x, y) d\rho^{(i)}(x) d\rho^{(k)}(y), \quad (2)$$

where  $\rho = (\rho^{(1)}, \rho^{(2)})$  and  $\rho^{(i)} \in \mathcal{P}(\mathbb{R}^d)$ , with  $i = 1, 2$ , denote the two species. Throughout we refer to  $K^{(11)}, K^{(22)}$  as the *self-interaction* potentials and we call  $K^{(12)}, K^{(21)}$  the *cross-interaction* potentials, respectively. It is worthwhile to highlight that, under the condition that  $K^{(12)} = \beta K^{(21)}$ , for some  $\beta > 0$ , the two-species interaction energy gives rise to a Wasserstein gradient flow on the product space and the evolution of the two densities is governed by the equations

$$\partial_t \rho_t^{(i)} = \nabla \cdot \left( \rho_t^{(i)} \nabla \left( K^{(i1)} * \rho_t^{(1)} + K^{(i2)} * \rho_t^{(2)} \right) \right), \quad (3)$$

with  $i = 1, 2$ , cf. [37, 29, 28, 22, 26]. For a suitable product space metric, the system is the epitome of interaction models found in many applied contexts for instance in cell-cell adhesion models [4, 45, 20, 8], chromatophore interactions in the skin of zebrafish [50, 51, 49], and multi-species systems with volume exclusion effects that result in cross-diffusion interaction systems, [10, 18, 27, 13].

### 1.1. Graph setting, non-linear mobility, and $p \neq 2$

While the space-time continuous dynamics and the Wasserstein gradient flow structure are well-understood, the situation is much more delicate when considering the flow of two densities, one per species, on graphs. In our work, we extend the recent work by Esposito et al. [35], which has established a graph analog of the aggregation equation. Their work shows that an appropriate definition of the geometry of the underlying space allows to understand a class of interaction equations on graphs as gradient flows, albeit in a Finslerian framework rather than the usual Riemannian setting.

The starting point is the dynamic formulation of the 2-Wasserstein distance due to [9]. There it was shown that the 2-Wasserstein distance can be characterized equivalently by minimizing (twice) the kinetic energy over all connecting paths

$$W_2^2(\mathcal{Q}_0, \mathcal{Q}_1) = \inf_{(\rho_t, j_t)_t} \int_0^1 \int_{\mathbb{R}^d} \frac{j_t^2}{\rho_t} d\mu dt, \quad (4)$$

where  $\mu \in \mathcal{M}^+(\mathbb{R}^d)$  is a suitable reference measure and the infimum is taken over all pairs  $(\rho_t, j_t)_{t \in [0,1]}$ ,  $\mu$ -a.e. satisfying

$$\partial_t \rho_t + \nabla \cdot j_t = 0,$$

as well as  $\rho_0 = \mathcal{Q}_0$ , and  $\rho_1 = \mathcal{Q}_1$ . While developed with numerical applications in mind, (4) turned out to be a starting point for various adaptations and generalizations of the classical Wasserstein distance. Many of these Wasserstein-type distances modify the action density  $\mathcal{A}(\rho, j) := \int_{\mathbb{R}^d} |j|^2 / \rho d\mu$  that appears inside the time integral in (4), introducing nonlinear mobilities [31], reaction terms that allow for initial and final measures having different masses [24] or even different actions in the interior and on the boundary of a given domain [43]. Of particular interest here are the works which identify nonlocal equations such as the nonlocal heat equation [41], nonlocal adaptations of the Fokker-Planck equation [25] and a nonlocal porous medium equation [34] on finite graphs or finite Markov chains as gradient flows for suitable Wasserstein-type metrics. Also an extension to treat the nonlocal heat equation on  $(\mathbb{R}^d, \mu)$  for some Radon measure  $\mu$  was considered in [33]. An important challenge faced when transferring the notion of a gradient flow to the setting of graphs is the need to compare fluxes or velocities (edge-based quantities) with densities (vertex-based quantities). This difficulty can be remedied by introducing a weight function  $\theta : [0, \infty) \times [0, \infty) \rightarrow [0, \infty)$ , which acts as an interpolation of quantities defined on connected vertices. More precisely, if  $\rho(x)$  and  $\rho(y)$  denote the densities on the vertices  $x$  and  $y$ , respectively, which share a connecting edge, then we shall think of  $\theta(\rho(x), \rho(y))$ , as the edge density. This definition allows for an adaptation of the action density to the graph setting [35], i.e.,

$$\mathcal{A}(\rho, j) = \iint_G \frac{|j(x, y)|^2}{\theta(\rho(x), \rho(y))} d\mu(x) d\mu(y),$$

where  $G$  is the set of edges and  $\mu \in \mathcal{M}^+(G)$  is again a reference measure, which, is arbitrary if  $\theta$  is 1-homogeneous. There are different choices for  $\theta$  that all seem reasonable but have a strong impact on the metric structure derived from  $\mathcal{A}$ . In [41] and [25] the logarithmic mean  $\theta_l(r, s) = \frac{r-s}{\log r - \log s}$  is shown to be a suitable choice for equations involving diffusive terms as it allows for a discrete chain rule, a fact that has already been observed and used in the finite volume community, cf. [23, 46, 16]. However, as was pointed out in [35], this choice does not allow for an increase of the support of the solution in the absence of diffusion. Thus, a different choice must be made to obtain physically reasonable solutions. Similar problems occur with the geometric mean  $\theta_g(r, s) = \sqrt{rs}$ , while dynamics using the arithmetic mean  $\theta_a(r, s) = \frac{r+s}{2}$  as an interpolation function may lead to negative densities and is therefore also not a reasonable choice. However, it is known that for transport equations, upwind schemes yield stable and structure-preserving discretizations, which motivates the choice

$$\theta_j(r, s) = r\mathbb{1}_{\{j>0\}}(r, s) + s\mathbb{1}_{\{j<0\}}(r, s)$$

as an interpolation function. This comes at the price of losing the antisymmetry of the action density, obtaining a Finslerian structure instead of a Riemannian structure. Yet, this structure is sufficient to define a notion of gradient flows as curves of maximal slope on graphs and leads to stability of gradient flows under narrow convergence, obtaining the existence of gradient flows for a large set of base measures  $\mu$  via approximation with finite graphs, [35].

In the context of finite volume discretizations, this upwinding has plenty of precedent, cf. [36], and references therein. In particular, for equations exhibiting an entropy-dissipation structure, it has been observed that certain finite volume discretizations can be used to preserve the structure at the discrete level, cf. [11] for a general class of drift-diffusion equations, [21, 5, 7, 47, 19, 6] for extensions to nonlocal drift-diffusion equations. The strategy was then extended to systems of cross-interaction species in [18, 16, 6, 40].

Even though cross-interactions between opposing species introduce a coupling in the velocity fields, using this upwinding, we can show that the evolution on the graph is, indeed, a gradient flow in the set of probability measures with respect to an appropriately defined Finsler product metric for the interaction energy

$$\mathcal{E}(\rho) = \frac{1}{2} \sum_{i,k=1}^2 \iint_{\mathbb{R}^d \times \mathbb{R}^d} K^{(ik)}(x, y) d\rho^{(i)}(x) d\rho^{(k)}(y). \quad ((2) \text{ revisited})$$

The interpretation of the dynamics as a gradient flow on a graph is established by noting that the quasi metric in the two species case decomposes into the sum of two instances of the metric for the single species, introduced in [35] — in analogy with the strategy for the continuous dynamics in [37].

A formal Finslerian structure also naturally appears in the space-time continuous setting when studying the  $p$ -Wasserstein distance  $W_p$ , see [1]. The dynamic formulation of  $W_p$  is given as

$$W_p^p(\rho_0, \rho_1) = \inf_{(\rho_t, j_t)_t} \int_0^1 \int_{\mathbb{R}^d} \frac{|j_t|^p}{(\rho_t)^{p-1}} d\mu dt. \quad (5)$$

Agueh still provides a norm in this setting, yet not an inner product. This allows to define the differential and gradient of functions on the set of probability measures with this distance, giving rise to a notion of gradient flows in this space.

Moreover, (5) can be generalized to the case  $p \neq 2$ , [31] studied generalizations of (5) including unbounded, concave mobilities, see also [14]. These nonlinear mobilities give rise to evolution equations of the form

$$\partial_t \rho = \nabla \cdot (m(\rho) \nabla K * \rho).$$

Depending on the specific choice of the mobility, it becomes necessary to add recession terms to the action functional to ensure its lower semicontinuity. Finally, [39] extended this study to bounded mobilities, which naturally appear in models with volume filling.

## 1.2. Our contribution

In this paper, we show that the setting of [35] can be carried over to systems of two (or potentially multiple) interacting species with a non-linear mobility and remains valid in the case  $p \neq 2$ . This way, we derive a corresponding two-species interaction equation as a gradient flow. It is given by

$$\begin{aligned} \partial_t \rho_t^{(i)}(x) &= -(\bar{\nabla} \cdot j_t^{(i)})(x), \\ j_t^{(i)}(x, y) &= \left[ m(\rho_t^{(i)}(x), \rho_t^{(i)}(y)) (v_t^{(i)})_+(x, y) \right]^{\frac{1}{p-1}} - \left[ m(\rho_t^{(i)}(y), \rho_t^{(i)}(x)) ((v_t^{(i)})_-(x, y)) \right]^{\frac{1}{p-1}}, \\ v_t^{(i)}(x, y) &= -\bar{\nabla} \left[ \left( K^{(i1)} * \rho_t^{(1)} \right)(x, y) + \left( K^{(i2)} * \rho_t^{(2)} \right)(x, y) \right], \end{aligned} \quad (6)$$

where  $p \in (1, \infty)$ ,  $i = 1, 2$ ,  $x, y \in \mathbb{R}^d$  and  $t \geq 0$ . Here, the quantities  $\mu$  and  $\eta$  encode the structure of the graph,  $m$  is a concave mobility, and the operators  $\bar{\nabla}$  and  $\bar{\nabla} \cdot$  are discrete analogues of gradient and divergence. Precise definitions will be given in Section 2 below.

The core novelties of our work are:

- Introduction of an action functional, which incorporates an upwind structure and a concave (un)bounded mobility.
- Extension of the Finslerian structure from [35] to the case  $p \in (1, \infty)$ , weakening the notion of Finsler metric, in the spirit of [1], and employing the notion of the metric gradient.
- Derivation of all the core results from [35] in the generalized framework, in particular a chain rule, a stability result for gradient flows, and an existence result beyond finite graphs.
- Extension of our framework to multiple species and derivation of a gradient flow structure for energies with symmetric cross-interaction.

The remainder of the paper is organized as follows. In section 2, we define the graph setting and introduce the notion of action functional and continuity equation, as well as the induced quasimetric. In Section 3 we discuss the Finsler geometry, give the interpretation of our system as a gradient flow using a suitable variational characterization and provide existence and stability results.

## 2. Analytical setting of the two species graph structure

This section provides the necessary extension of the dynamic 2-Wasserstein distance (4) to the graph setting, including a nonlinear mobility and for  $p \neq 2$ . To this end, we give a precise definition of the corresponding action functional and study some of its properties. Then, after introducing a notion of generalized continuity equations, we can define and analyze the corresponding quasimetric.

We start by introducing the graph setting. Throughout,  $\mathcal{M}(X)$  ( $\mathcal{M}_+(X)$ ) denotes the space of (nonnegative) Radon measures on the space  $X$ . The vertices of our graph are defined by the base measure  $\mu \in \mathcal{M}^+(\mathbb{R}^d)$ . We define a nonnegative weight function  $\eta : \mathbb{R}^d \times \mathbb{R}^d \rightarrow [0, \infty)$  and thereby the edges of the undirected graph as

$$G := \{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d : x \neq y, \eta(x, y) > 0\}.$$

Setting  $1 < p = q/(q-1) < \infty$ , throughout we shall make use of the following set of technical assumptions on  $\mu$  and the  $\eta$ :

$$\text{(continuous symmetric weight)} \quad \eta|_G \in C(G, [0, \infty)), \quad \forall x, y \in \mathbb{R}^d \text{ it holds } \eta(x, y) = \eta(y, x), \quad (\text{W})$$

$$\text{(individual moment bound)} \quad \int_{\mathbb{R}^d} (1 + |x|^p) d\mu(x) \leq C_\mu, \quad (\text{MB1})$$

$$\text{(joint moment bound)} \quad \sup_{x \in \mathbb{R}^d} \int_{\mathbb{R}^d} |x - y|^q \vee |x - y|^{pq} \eta(x, y) d\mu(y) \leq C_\eta, \quad (\text{MB2})$$

$$\text{(local blow-up control)} \quad \lim_{\varepsilon \rightarrow 0} \sup_{x \in \mathbb{R}^d} \int_{B_\varepsilon(x) \setminus \{x\}} |x - y|^q \eta(x, y) d\mu(y) = 0, \quad (\text{BC})$$

for some constants  $C_\eta, C_\mu > 0$  and where  $B_\varepsilon(x) := \{y \in \mathbb{R}^d : |x - y| < \varepsilon\}$ .

**Remark 2.1.** While the assumptions (W), (BC) and (MB2) are similar to the assumptions in [35], (MB1) is a new assumption required due to the added nonlinear mobilities. It is needed, whenever we apply Lemma 2.14 in the sequel. However, for a large class of mobilities Lemma 2.14 can be refined to allow dropping assumption (MB1). More details on this will be given in Remark 2.15.

Next, we establish nonlocal analogues for the gradient and the divergence as foreshadowed in the introduction.

**Definition 2.2** (Nonlocal gradient and nonlocal divergence). Given  $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ , we define the nonlocal gradient  $\bar{\nabla}\varphi$  by

$$\bar{\nabla}\varphi(x, y) := \varphi(y) - \varphi(x).$$

Given  $j \in \mathcal{M}(\mathbb{R}^d)$  we define the nonlocal divergence  $\bar{\nabla} \cdot j$  by

$$\int_{\mathbb{R}^d} \varphi(x) d\bar{\nabla} \cdot j(x) := -\frac{1}{2} \iint_G \bar{\nabla}\varphi(x, y) \eta(x, y) dj(x, y) \quad \forall \varphi : \mathbb{R}^d \rightarrow \mathbb{R}.$$

## 2.1. Definition of the action functional

**Definition 2.3** (Mobility and density functions). Given two thresholds  $R, S \in (0, \infty]$ , we define a mobility function  $m \in C([0, R) \times [0, S))$ , which is concave and strictly positive in  $(0, R) \times (0, S)$ . For  $(r, s) \in [0, R) \times [0, S)$  we denote  $m(r, S) = \lim_{s \rightarrow S} m(r, s)$  and  $m(R, s) = \lim_{r \rightarrow R} m(r, s)$ . We call such a mobility  $m$  upwind-admissible if for every  $s \geq 0$  we have  $m(0, s) = 0$ . Furthermore, if  $R = S = \infty$ , which implies that  $m$  is nondecreasing, we set

$$m_\infty(r, s) := \lim_{\lambda \rightarrow \infty} \frac{1}{\lambda} m(\lambda r, \lambda s). \quad (7)$$

We say that the growth of  $m$  is uniformly sublinear, if  $m_\infty \equiv 0$ .

Given a mobility  $m$  and an exponent  $p \in (1, \infty)$ , we define the convex, lsc. density function  $\alpha : \mathbb{R}^3 \rightarrow [0, \infty]$  by

$$\alpha_m(j, r, s) := \begin{cases} \frac{(j_+)^p}{m^{p-1}(r, s)}, & (r, s) \in [0, R] \times [0, S], \\ \infty, & \text{otherwise,} \end{cases} \quad (8)$$

where we use the conventions

$$a/b = \begin{cases} 0, & \text{if } a = b = 0, \\ \infty, & \text{if } a \neq b = 0. \end{cases}$$

In the case  $R = S = \infty$ , we define the recession function

$$\alpha_{m_\infty}(j, r, s) := \lim_{\lambda \rightarrow \infty} \frac{1}{\lambda} \alpha_m(\lambda j, \lambda r, \lambda s) = \frac{(j_+)^p}{m_\infty^{p-1}(r, s)}. \quad (9)$$

**Remark 2.4.** (i) Observe that (8) encodes an upwind structure as only the positive part of the flux enters the definition.

(ii) By definition,  $m_\infty$ , and thus  $\alpha_{m_\infty}$ , are positively 1-homogeneous. In particular, for  $r > 0$  we have  $m_\infty(r, s) = r m_\infty(1, s/r)$  and for  $s > 0$  we have  $s m_\infty(r/s, 1)$ .

(iii) The assumption  $m(0, s) = 0$  is crucial as it ensures the non-negativity of  $\rho$  when the action is finite. However, note that this assumption excludes, among others, the choice  $m \equiv 1$ .

(iv) If  $m$  is an upwind-admissible mobility, the homogeneity implies  $m_\infty(0, s) = 0$ .

(v) If  $m_\infty(r, s) = 0$ , then  $\alpha_{m_\infty}(j, r, s) = \infty$  for every  $j \neq 0$ . In particular,  $\alpha_{m_\infty}(j, r, s)$  is the convex indicator of the set  $\{j = 0\}$  (it is zero if  $j = 0$  and infinite otherwise), if the growth of  $m$  is uniformly sublinear.

To define the action density functional, we introduce the following notation.

**Definition 2.5.** The transpose of a Borel set  $B \in \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d)$  is denoted by  $B^\top := \{(y, x) \in \mathbb{R}^d \times \mathbb{R}^d : (x, y) \in B\}$ . The transpose of a measure  $\nu \in \mathcal{M}(\mathbb{R}^d \times \mathbb{R}^d)$  is then defined as  $(\nu)^\top(B) := \nu(B^\top)$ .

Given a pair of probability measures  $\rho = (\rho^{(1)}, \rho^{(2)}) \in (\mathcal{P}(\mathbb{R}^d))^2$ , abusing notation, we denote the Lebesgue decomposition  $d\rho^{(i)} = d\rho^{(i)\mu} + d\rho^{(i)\perp} = \rho^{(i)}d\mu + d\rho^{(i)\perp}$  for  $i = 1, 2$ . Similarly, given a pair of fluxes  $j \in (\mathcal{M}(G))^2$ , for  $i = 1, 2$  we denote  $dj^{(i)} = dj^{(i)\mu} + dj^{(i)\perp} = j^{(i)}d(\mu \otimes \mu) + dj^{(i)\perp}$ .

We emphasize that from now on, we always use non-italic, upright letters such as  $j, \rho, \beta$  etc. to indicate pairs of quantities indexed by superscript  $(i)$ , e.g.  $j = (j^{(1)}, j^{(2)})$ .

**Definition 2.6** (Single species action density functional). Let  $\mu \in \mathcal{M}^+(\mathbb{R}^d)$ ,  $\eta$  satisfy **(W)**. Now, given  $p \in (1, \infty)$  and an upwind-admissible  $m$ , for any  $(\rho, j) \in \mathcal{P}(\mathbb{R}^d) \times \mathcal{M}(G)$ , we define the single-species action density functional  $\bar{\mathcal{A}}_m$  as follows:

1. If  $R = S = \infty$ , for any  $\sigma \in \mathcal{M}^+(\mathbb{R}^d \times \mathbb{R}^d)$  such that  $\sigma = \sigma^\top$ ,  $d(\rho^\perp \otimes \mu) = \rho^\perp \otimes \tilde{\mu}d\sigma$ ,  $d(\mu \otimes \rho^\perp) = \tilde{\mu} \otimes \rho^\perp d\sigma$  and  $dj^\perp = j^\perp d\sigma$ , we define

$$\begin{aligned} \bar{\mathcal{A}}_m(\mu; \rho, j) := & \frac{1}{2} \iint_G \left[ \alpha_m(j, \rho \otimes \mu, \mu \otimes \rho) + \alpha_m(-j, \mu \otimes \rho, \rho \otimes \mu) \right] \eta d(\mu \otimes \mu) \\ & + \frac{1}{2} \iint_G \left[ \alpha_{m_\infty}(j^\perp, \rho^\perp \otimes \tilde{\mu}, \tilde{\mu} \otimes \rho^\perp) + \alpha_{m_\infty}(-j^\perp, \tilde{\mu} \otimes \rho^\perp, \rho^\perp \otimes \tilde{\mu}) \right] \eta d\sigma. \end{aligned} \quad (10)$$

2. If  $R \wedge S < \infty$ , similar to [39], we define  $\bar{\mathcal{A}}_m$  as follows:

$$\bar{\mathcal{A}}_m(\mu; \rho, j) := \begin{cases} \frac{1}{2} \iint_G \left[ \alpha_m(j, \rho \otimes \mu, \mu \otimes \rho) + \alpha_m(-j, \mu \otimes \rho, \rho \otimes \mu) \right] \eta d(\mu \otimes \mu), & \text{if } \rho^\perp = 0, j^\perp = 0, \\ \infty, & \text{otherwise.} \end{cases} \quad (11)$$

Let  $\beta = (\beta^{(1)}, \beta^{(2)}) \in (0, \infty) \times (0, \infty)$  be a pair of positive constants. Then, for any  $\rho \in (\mathcal{P}(\mathbb{R}^d))^2$ ,  $j \in (\mathcal{M}(\mathbb{R}^d))^2$ , the two-species action density functional is defined by

$$\mathcal{A}_{m,\beta}(\mu; \rho, j) := \frac{1}{\beta^{(1)}} \bar{\mathcal{A}}_m(\mu; \rho^{(1)}, j^{(1)}) + \frac{1}{\beta^{(2)}} \bar{\mathcal{A}}_m(\mu; \rho^{(2)}, j^{(2)}). \quad (12)$$

**Remark 2.7.** (i)  $\mathcal{A}_{m,\beta}$  is well-defined which is clear if  $R \wedge S < \infty$ . If  $R = S = \infty$ , the definition  $\bar{\mathcal{A}}_m$  is independent of the particular choice of  $\sigma$  due to the positive 1-homogeneity of  $\alpha_{m_\infty}$ . An admissible  $\sigma$  can always be constructed, e.g. by setting

$$\sigma = (\rho^\perp \otimes \mu + \rho^\perp \otimes \mu + |j^\perp| + |(j^\perp)^\top|),$$

for  $i = 1, 2$ . The definition may be adapted such that the symmetry assumption  $\sigma = \sigma^\top$  is not required. However, since this assumption is nonrestrictive and simplifies notation, we apply it throughout.

- (ii) If  $R = S = \infty$  and  $m_\infty \equiv 0$ , the definition of the single-species action density functional (10) simplifies to

$$\bar{\mathcal{A}}_m(\mu; \rho, j) := \begin{cases} \frac{1}{2} \iint_G \left[ \alpha_m(j, \rho \otimes \mu, \mu \otimes \rho) + \alpha_m(-j, \mu \otimes \rho, \rho \otimes \mu) \right] \eta d(\mu \otimes \mu), & \text{if } j^\perp = 0, \\ \infty, & \text{otherwise.} \end{cases}$$

- (iii) If  $R \wedge S < \infty$ , finiteness of the action density implies both  $\rho \ll \mu$  as well as  $j \ll \mu \otimes \mu$ .

- (iv) The two-species action density functional is fully decoupled with respect to the different species.

Next, we want to show an antisymmetry property of the action density. To do this, we introduce the following notation.

**Definition 2.8** (Antisymmetric velocities and fluxes). *We define the set of antisymmetric velocities by*

$$\mathcal{V}^{\text{as}}(G) := \{(v, v^\perp) : G \rightarrow \mathbb{R}^2, \text{ s.t. } v = -v^\top \text{ and } (v^\perp)^\top = -(v^\perp)^\top\}.$$

For  $\mu \in \mathcal{M}^+(\mathbb{R}^d)$  and  $\rho \in \mathcal{P}(\mathbb{R}^d)$  we define  $\zeta \in \mathcal{M}^+(\mathbb{R}^d)$  as

$$\zeta = \rho^\perp \otimes \mu + \mu \otimes \rho^\perp, \quad (13)$$

and denote the corresponding densities by  $d(\rho^\perp \otimes \mu) =: \rho^\perp \otimes \tilde{\mu} d\zeta$  and  $d(\mu \otimes \rho^\perp) =: \tilde{\mu} \otimes \rho^\perp d\zeta$ . With these densities, we further shorten notation by introducing

$$\begin{aligned} \mathbf{m}(x, y) &:= m(\rho(x), \rho(y)), \\ \mathbf{m}_\infty(x, y) &:= m_\infty(\tilde{\mu}(x)\rho^\perp(y), \rho^\perp(x)\tilde{\mu}(y)). \end{aligned} \quad (14)$$

With this, recalling  $q = p/(p-1)$ , for  $k = 1, 2$ , we define

$$\begin{aligned} d\gamma_1(x, y) &:= (\mathbf{m}(x, y))^{q-1} d\mu(x) d\mu(y), & d\gamma_2(x, y) &:= (\mathbf{m}(y, x))^{q-1} d\mu(x) d\mu(y), \\ d\gamma_1^\perp(x, y) &:= (\mathbf{m}_\infty(x, y))^{q-1} d\zeta(x, y), & d\gamma_2^\perp(x, y) &:= (\mathbf{m}_\infty(y, x))^{q-1} d\zeta(x, y). \end{aligned}$$

These measures satisfy  $\gamma_1^\top = \gamma_2$  and vice versa as well as  $(\gamma_1^\perp)^\top = \gamma_2^\perp$  and vice versa. By **(W)**, this does not change, when multiplied by  $\eta$ . Hence, it makes sense to define the following set of antisymmetric fluxes:

$$\mathcal{M}_{\eta\gamma_1}^{\text{as}}(G) := \{j \in \mathcal{M}(G) : j_+ \ll \eta\gamma_1, j_- = (j_+)^\top, j_+^\perp \ll \eta\gamma_1^\perp, j_-^\perp = (j_+^\perp)^\top\}.$$

**Remark 2.9.** Any  $j \in \mathcal{M}_{\eta\gamma_1}^{\text{as}}(G)$  satisfies  $j_- \ll \eta\gamma_2$  and  $j_-^\perp \ll \eta\gamma_2^\perp$ .

In the sequel, any indices attached to  $\rho$  will be passed to  $\mathbf{m}$ ,  $\mathbf{m}_\infty$ ,  $\gamma_k$ ,  $\gamma_k^\perp$  and  $\zeta$ , i.e. when  $\rho = \rho_t^{n, (i)\perp}$  in (14), we write

$$\mathbf{m}_{\infty, t}^{n, (i)}(x, y) := m_\infty(\tilde{\mu}(x)\rho_t^{n, (i)\perp}(y), \rho_t^{n, (i)\perp}(x)\tilde{\mu}(y)),$$

and similarly for the other expressions.

We are now in the position to establish a connection between fluxes and velocities.

**Lemma 2.10** (Dual representation). *Let  $\mu \in \mathcal{M}^+(\mathbb{R}^d)$ ,  $\rho$  and  $j$  be such that  $\mathcal{A}_{m, \beta}(\mu; \rho, j) < \infty$ . Then, there exists a pair of measurable functions  $(v^{(1)}, v^{(2)}) : G \rightarrow \mathbb{R}^2$  such that for  $i = 1, 2$  we have*

$$dj^{(i)\mu} = (v_+^{(i)})^{q-1} d\gamma_1^{(i)} - (v_-^{(i)})^{q-1} d\gamma_2^{(i)}. \quad (15)$$

Further,  $\zeta$  from Definition 2.8 is an admissible choice for  $\sigma$  in Definition 2.6, i.e.  $j^{(i)\top} \ll \zeta^{(i)}$ , and there exists another pair of measurable functions  $(v^{(1)\perp}, v^{(2)\perp}) : G \rightarrow \mathbb{R}^2$  (which is zero if  $R \wedge S < \infty$ ), such that for  $i = 1, 2$  we have

$$dj^{(i)\perp} = (v_+^{(i)\perp})^{q-1} d\gamma_1^{(i)\perp} - (v_-^{(i)\perp})^{q-1} d\gamma_2^{(i)\perp}. \quad (16)$$

We can rewrite the action density in terms of  $\mathbf{v} = (v^{(1)}, v^{(1)\perp}, v^{(2)}, v^{(2)\perp})$  as

$$\begin{aligned} \mathcal{A}_{m, \beta}(\mu; \rho, j) &= \frac{1}{2} \sum_{i=1}^2 \frac{1}{\beta^{(i)}} \left[ \iint_G \mathbf{m}^{(i)} \left( (v_+^{(i)})^q + ((v_-^{(i)})^\top)^q \right) \eta d\mu \otimes \mu \right. \\ &\quad \left. + \iint_G \mathbf{m}_\infty^{(i)} \left( (v_+^{(i)\perp})^q + (((v_-^{(i)\perp})^\top)^q \right) \eta d\zeta^{(i)} \right] \\ &=: \tilde{\mathcal{A}}_{m, \beta}(\mu; \rho, \mathbf{v}), \end{aligned} \quad (17)$$

while for  $R \wedge S < \infty$  we have  $\rho^\perp = 0$  and  $j^\perp = 0$ . In particular, if  $\mathbf{v} \in (\mathcal{V}^{\text{as}}(G))^2$ , we have

$$\tilde{\mathcal{A}}_{m, \beta}(\mu; \rho, \mathbf{v}) = \sum_{i=1}^2 \frac{1}{\beta^{(i)}} \left[ \iint_G \mathbf{m}^{(i)} (v_+^{(i)})^q \eta d\mu \otimes \mu + \iint_G \mathbf{m}_\infty^{(i)} (v_+^{(i)\perp})^q \eta d\zeta^{(i)} \right]. \quad (18)$$

*Proof.* Assume  $R = S = \infty$ . The case  $R \wedge S < \infty$  then works similarly. First, we show that the finiteness of the action implies that  $j_+^{(i)\perp} \ll \gamma_1^{(i)\perp}$  and  $j_-^{(i)\perp} \ll \gamma_2^{(i)\perp}$ . Indeed, assume there exists a set  $B \in \mathcal{B}(G)$  such that  $j_+^{(i)\perp}(B) > 0 = \gamma_1^{(i)\perp}(B)$ . By Remark 2.7  $\sigma^{(i)} = \gamma_1^{(i)\perp} + \gamma_2^{(i)\perp} + |j^{(i)\perp}|$  is admissible in (10) and  $\sigma^{(i)}(B) > 0$ . Since we have

$$m_\infty^{q-1} \left( \frac{d(\rho^{(i)\perp} \otimes \mu)}{d\sigma^{(i)}}, \frac{d(\mu \otimes \rho^{(i)\perp})}{d\sigma^{(i)}} \right) = m_\infty^{q-1} \left( \frac{d(\mu \otimes \rho^{(i)\perp})}{d\sigma^{(i)}}, \frac{d(\rho^{(i)\perp} \otimes \mu)}{d\sigma^{(i)}} \right) = m_\infty^{q-1}(0, 0) = 0$$

$\sigma^{(i)}$ -a.e. in  $B$ , we obtain a contradiction to  $\mathcal{A}_{m,\beta}(\mu; \rho, j) < \infty$ . Similar arguments hold true for  $j_-^{(i)\perp}$  as well as  $j_+^{(i)}$  and  $j_-^{(i)}$ . Therefore, the nonnegative functions  $v_+^{(i)}$  and  $v_-^{(i)}$  satisfying (15) are well-defined  $\gamma_1^{(i)}$ -a.e. and  $\gamma_2^{(i)}$ -a.e., respectively, which gives us  $v^{(i)} = v_+^{(i)} - v_-^{(i)}$ . Similarly, we obtain  $v^{(i)\perp}$  satisfying (16). Finally, (17), follows, when we insert

$$dj^{(i)\mu}(x, y) = \left[ \left( m^{(i)}(x, y) v_+^{(i)}(x, y) \right)^{q-1} - \left( m^{(i)}(y, x) v_-^{(i)}(x, y) \right)^{q-1} \right] d\mu(x) d\mu(y)$$

and

$$dj^{(i)\perp}(x, y) = \left[ \left( m_\infty^{(i)}(x, y) v_+^{(i)\perp}(x, y) \right)^{q-1} - \left( m_\infty^{(i)}(y, x) v_-^{(i)\perp}(x, y) \right)^{q-1} \right] d\zeta^{(i)}(x, y)$$

into (10). □

**Definition 2.11.** *In light of (17), we define*

$$\tilde{\alpha}_m(v, r, s) := \begin{cases} m(r, s)(v_+)^q, & (r, s) \in [0, R] \times [0, S], \\ \infty, & \text{otherwise,} \end{cases} \quad (19)$$

which gives us a representation of  $\tilde{\mathcal{A}}_{m,\beta}$  similar to the one in Definition 2.6, only replacing  $j$  with  $v$  and  $\alpha$  with  $\tilde{\alpha}$ .

We observe that antisymmetric fluxes admit lower action densities while preserving their nonlocal divergence.

**Corollary 2.12** (Antisymmetric vector fields have lower action density). *Let  $\mu \in \mathcal{M}^+(\mathbb{R}^d)$ ,  $\rho \in (\mathcal{P}(\mathbb{R}^d))^2$  and  $j \in (\mathcal{M}(G))^2$  s.t.  $\mathcal{A}_{m,\beta}(\mu; \rho, j) < \infty$ . Then, there exists  $\bar{j} \in (\mathcal{M}_{\eta\gamma_1}^{\text{as}}(G))^2$  such that*

$$\bar{\nabla} \cdot j^{(i)} = \bar{\nabla} \cdot \bar{j}^{(i)}, \quad i = 1, 2,$$

and

$$\mathcal{A}_{m,\beta}(\mu; \rho, \bar{j}) \leq \mathcal{A}_{m,\beta}(\mu; \rho, j).$$

*Proof.* Analogous to [35, Corollary 2.8], one defines  $d\bar{j}^{(i)} := d(j^{(i)} - (j^{(i)})^\top)/2$ , compares (17) with (18), and applies Jensen's inequality. □

Next, we establish properties of the action density as well as important bounds for the subsequent analysis.

**Lemma 2.13** (Lower semicontinuity of the action density). *The action is lower semicontinuous with respect to weak-\* convergence in  $\mathcal{M}^+(\mathbb{R}^d) \times (\mathcal{M}^+(\mathbb{R}^d))^2 \times (\mathcal{M}(G))^2$ . That is, for  $\mu^n \rightharpoonup^* \mu$  in  $\mathcal{M}^+(\mathbb{R}^d)$ ,  $\rho_n \rightharpoonup^* \rho$  in  $(\mathcal{M}^+(\mathbb{R}^d))^2$ , and  $j^n \rightharpoonup^* j$  in  $(\mathcal{M}(G))^2$ , we have*

$$\liminf_{n \rightarrow \infty} \mathcal{A}_{m,\beta}(\mu^n; \rho^n, j^n) \geq \mathcal{A}_{m,\beta}(\mu; \rho, j).$$

*Proof.* See [2, Theorem 2.34], while keeping in mind that  $\mu^n \otimes \rho^{n,(i)} \rightharpoonup^* \mu \otimes \rho^{(i)}$  in  $\mathcal{M}(\mathbb{R}^d \times \mathbb{R}^d)$  if and only if both  $\mu^n \rightharpoonup^* \mu$  in  $\mathcal{M}(\mathbb{R}^d)$  and  $\rho^{n,(i)} \rightharpoonup^* \rho^{(i)}$  in  $\mathcal{M}(\mathbb{R}^d)$ . □

**Lemma 2.14.** Let  $\mu \in \mathcal{M}^+(\mathbb{R}^d)$ ,  $\rho \in (\mathcal{P}(\mathbb{R}^d))^2$  and  $j \in (\mathcal{M}(G))^2$ , such that  $\mathcal{A}_{m,\beta}(\mu; \rho, j) < \infty$ . Then, there exists a constant  $M = M(m, p, \beta) > 0$ , such that for any measurable  $\Phi : G \rightarrow \mathbb{R}_+$ , it holds

$$\iint_G \Phi \eta d|j| \leq M \mathcal{A}_{m,\beta}^{1/p}(\mu; \rho, j) \sum_{i=1}^2 \left( \iint_G \Phi^q \eta d(\mu \otimes \mu + \rho^{(i)} \otimes \mu + \mu \otimes \rho^{(i)}) \right)^{1/q}. \quad (20)$$

*Proof.* First, let  $R = S = \infty$  and  $\sigma^{(i)} = \zeta^{(i)} \in \mathcal{M}^+(G)$ ,  $i = 1, 2$  be as in Definition 2.8. Since  $\mathcal{A}_{m,\beta}(\mu; \rho, j) < \infty$ , we have that

$$B := \left\{ (x, y) \in G : \sum_{i=1}^2 \left[ \alpha_m(j_+^{(i)}(x, y), \rho^{(i)}(x), \rho^{(i)}(y)) + \alpha_m(j_-^{(i)}(x, y), \rho^{(i)}(y), \rho^{(i)}(x)) \right. \right. \\ \left. \left. + \alpha_{m_\infty}(j_+^{(i)\perp}(x, y), \rho^{(i)\perp}(x)\tilde{\mu}(y), \tilde{\mu}(x)\rho^{(i)\perp}(y)) + \alpha_{m_\infty}(j_-^{(i)\perp}(x, y), \tilde{\mu}(x)\rho^{(i)\perp}(y), \rho^{(i)\perp}(x)\tilde{\mu}(y)) \right] = \infty \right\}$$

is a  $\zeta^{(i)}$ -nullset for  $i = 1, 2$ . By definition of  $\alpha_m$ , we have,  $\zeta^{(i)}$ -a.e. in  $B^c$ , the inequality

$$\begin{aligned} \left( j_+^{(i)}(x, y) \right)^p + \left( j_-^{(i)}(x, y) \right)^p &\leq \max \left\{ m^{p-1}(\rho^{(i)}(y), \rho^{(i)}(x)), m^{p-1}(\rho^{(i)}(x), \rho^{(i)}(y)) \right\} \\ &\quad \cdot \left( \alpha_m(j_+^{(i)}, \rho^{(i)}(x), \rho^{(i)}(y)) + \alpha_m(j_-^{(i)}, \rho^{(i)}(y), \rho^{(i)}(x)) \right) \\ &\leq \bar{M} [1 + \rho^{(i)}(x) + \rho^{(i)}(y)]^{p-1} \\ &\quad \cdot \left( \alpha_m(j_+^{(i)+}, \rho^{(i)}(x), \rho^{(i)}(y)) + \alpha_m(j_-^{(i)-}, \rho^{(i)}(y), \rho^{(i)}(x)) \right), \end{aligned} \quad (21)$$

where  $\bar{M}$  only depends on  $m$ . Indeed, such an  $\bar{M}$  exists, since  $m$  is concave and  $m(0, s) = 0$  by definition. Similarly, we have the bound

$$\begin{aligned} \left( j_+^{(i)\perp}(x, y) \right)^p + \left( j_-^{(i)\perp}(x, y) \right)^p &\leq \bar{M} [1 + \rho^{(i)\perp}(x)\tilde{\mu}(y) + \tilde{\mu}(x)\rho^{(i)\perp}(y)]^{p-1} \\ &\quad \cdot \left( \alpha_{m_\infty}(j_+^{(i)\perp}, \rho^{(i)\perp}(x)\tilde{\mu}(y), \tilde{\mu}(x)\rho^{(i)\perp}(y)) \right. \\ &\quad \left. + \alpha_{m_\infty}(j_-^{(i)\perp}, \tilde{\mu}(x)\rho^{(i)\perp}(y), \rho^{(i)\perp}(x)\tilde{\mu}(y)) \right). \end{aligned}$$

By the complementarity of the positive and negative parts, this gives

$$\begin{aligned} |j^{(i)}|(x, y) &\leq \tilde{M} \left( 1 + \rho^{(i)}(x) + \rho^{(i)}(y) \right)^{1/q} \left( \frac{1}{\beta^{(i)}} \alpha_m(j_+^{(i)}, \rho^{(i)}(x), \rho^{(i)}(y)) \right. \\ &\quad \left. + \frac{1}{\beta^{(i)}} \alpha_m(j_-^{(i)}, \rho^{(i)}(y), \rho^{(i)}(x)) \right)^{1/p}, \\ |j^{(i)\perp}|(x, y) &\leq \tilde{M} \left( 1 + \rho^{(i)\perp}(x)\tilde{\mu}(y) + \rho^{(i)\perp}(y)\tilde{\mu}(x) \right)^{1/q} \left( \frac{1}{\beta^{(i)}} \alpha_{m_\infty}(j_+^{(i)\perp}, \rho^{(i)\perp}(x)\tilde{\mu}(y), \tilde{\mu}(x)\rho^{(i)\perp}(y)) \right. \\ &\quad \left. + \frac{1}{\beta^{(i)}} \alpha_{m_\infty}(j_-^{(i)\perp}, \tilde{\mu}(x)\rho^{(i)\perp}(y), \rho^{(i)\perp}(x)\tilde{\mu}(y)) \right)^{1/p}, \end{aligned}$$

where  $\tilde{M}$  depends only on  $m$ ,  $p$  and  $\beta$ . Since  $|j| + |j^\perp| = \sum_{i=1}^2 (|j^{(i)}| + |j^{(i)\perp}|)$ , these estimates together with Hölder's inequality yield

$$\begin{aligned} \iint_G \Phi \eta d|j| &= \sum_{i=1}^2 \left( \iint_{B^c} \Phi \eta |j^{(i)}| d(\mu \otimes \mu) + \iint_{B^c} \Phi \eta |j^{(i)\perp}| d\zeta^{(i)} \right) \\ &\leq 4\tilde{M} (2\mathcal{A}_{m,\beta}(\mu; \rho, j))^{1/p} \sum_{i=1}^2 \left( \left( \iint_G \Phi^q \eta d(\mu \otimes \mu + \rho^{(i)\mu} \otimes \mu + \mu \otimes \rho^{(i)\mu}) \right)^{1/q} \right. \\ &\quad \left. + \left( \iint_G \Phi^q \eta d(\zeta^{(i)} + \rho^{(i)\perp} \otimes \mu + \mu \otimes \rho^{(i)\perp}) \right)^{1/q} \right). \end{aligned}$$

Thus, recalling that  $\varsigma^{(i)} = \rho^{(i)\perp} \otimes \mu + \mu \otimes \rho^{(i)\perp}$ , we obtain (20) with  $M = 16 \cdot 2^{1/p} \tilde{M}$ . If  $R \wedge S < \infty$ , we argue similarly, only replacing in (21)  $m$  by  $m_{\uparrow}(r, s) := \sup_{(\tilde{r}, \tilde{s}) \in [0, r] \times [0, s]} m(\tilde{r}, \tilde{s})$ , which is still concave and satisfies  $m_{\uparrow}(0, s) = 0$ .  $\square$

**Remark 2.15.** *If there exists  $C > 0$  such that  $m$  satisfies*

$$m(r, s) \leq C(r + s) \quad \forall r, s \in [0, R] \times [0, S], \quad (\text{M})$$

then (20) can be replaced by the refined bound

$$\iint_G \Phi \eta d|j| \leq M \mathcal{A}_{m, \beta}^{1/p}(\mu; \rho, j) \sum_{i=1}^2 \left( \iint_G \Phi^q \eta d(\rho^{(i)} \otimes \mu + \mu \otimes \rho^{(i)}) \right)^{1/q}. \quad (22)$$

Indeed, observe that in this case the summand 1 on the right-hand side (21) can be omitted, which leads to dropping the integral with respect to  $\mu \otimes \mu$ . It is straightforward to check that replacing (20) by (22), whenever it is employed in the sequel, allows to drop the assumption (MB1) altogether.

**Corollary 2.16.** *Let  $\mu \in \mathcal{M}^+(\mathbb{R}^d)$ ,  $\rho \in (\mathcal{P}(\mathbb{R}^d))^2$  and  $j \in (\mathcal{M}(G))^2$ , such that  $\mathcal{A}_{m, \beta}(\mu; \rho, j) < \infty$ . Then, for  $\Phi_1(x, y) = 2 \wedge |x - y|$  and  $\Phi_2 = |x - y| \vee |x - y|^p$ , we have*

$$\iint_G \Phi_k \eta d|j| \leq M C_{\eta}^{1/q} \mathcal{A}_{m, \beta}^{1/p}(\mu; \rho, j), \quad k = 1, 2,$$

where  $M = M(m, p, \beta)$  is different from that in Lemma 2.14.

*Proof.* Note that  $\Phi_1(x, y) \leq |x - y| \leq \Phi_2(x, y)$ . Therefore, Lemma 2.14 yields for  $k = 1, 2$

$$\begin{aligned} & \iint_G \Phi_k(x, y) \eta(x, y) d|j|(x, y) \\ & \leq \bar{M} \mathcal{A}_{m, \beta}^{1/p}(\mu; \rho, j) \sum_{i=1}^2 \left( \iint_G \Phi_k^q(x, y) \eta(x, y) d(\mu \otimes \mu + \rho^{(i)} \otimes \mu + \mu \otimes \rho^{(i)})(x, y) \right)^{1/q} \\ & = \bar{M} \mathcal{A}_{m, \beta}^{1/p}(\mu; \rho, j) \sum_{i=1}^2 \left( \iint_G |x - y|^q \vee |x - y|^{pq} \eta(x, y) d(\mu \otimes \mu + 2\rho^{(i)} \otimes \mu)(x, y) \right)^{1/q} \\ & \leq \bar{M} \mathcal{A}_{m, \beta}^{1/p}(\mu; \rho, j) 2(C_{\mu} + 2)^{1/q} C_{\eta}^{1/q}, \end{aligned}$$

where we used (MB2) together with (MB1) and the fact that  $\rho^{(i)} \in \mathcal{P}(\mathbb{R}^d)$  for  $i = 1, 2$ .  $\square$

**Lemma 2.17** (Convexity of the action). *Let  $\mu_0, \mu_1 \in \mathcal{M}^+(\mathbb{R}^d)$ ,  $\rho_0, \rho_1 \in (\mathcal{P}(\mathbb{R}^d))^2$  and  $j_0, j_1 \in (\mathcal{M}(G))^2$ . For  $\tau \in (0, 1)$  define  $\mu_{\tau} = (1 - \tau)\mu_0 + \tau\mu_1$ ,  $\rho_{\tau} = (1 - \tau)\rho_0 + \tau\rho_1$  and  $j_{\tau} = (1 - \tau)j_0 + \tau j_1$ . Then, we have*

$$\mathcal{A}_{m, \beta}(\mu_{\tau}; \rho_{\tau}, j_{\tau}) \leq (1 - \tau) \mathcal{A}_{m, \beta}(\mu^0; \rho_0, j_0) + \tau \mathcal{A}_{m, \beta}(\mu_1; \rho_1, j_1),$$

*Proof.* This immediately follows from the convexity of  $\alpha_m$  and  $\alpha_{m_{\infty}}$ . A detailed argument for one species and  $m(r, s) = r$ , which upon small adjustments is also applicable here, can be found in [35, Lemma 2.12].  $\square$

## 2.2. Generalized continuity equation and properties

In this subsection we study the nonlocal continuity equation and properties of its solutions with finite action.

**Definition 2.18** (Continuity equation). *We say that the pair  $(\boldsymbol{\rho}, \mathbf{j}) = ((\rho_t)_{t \in [0, T]}, (\mathbf{j}_t)_{t \in [0, T]})$  with  $\rho_t \in (\mathcal{P}(\mathbb{R}^d))^2$  and  $\mathbf{j}_t \in (\mathcal{M}(G))^2$ , is a weak solution of the continuity equation*

$$\partial_t \rho + \bar{\nabla} \cdot \mathbf{j}_t = 0 \text{ on } (0, T) \times \mathbb{R}^d,$$

if we have

- (i)  $\boldsymbol{\rho}$  is a weakly continuous curve in  $(\mathcal{P}(\mathbb{R}^d))^2$
- (ii)  $\mathbf{j}$  is a Borel-measurable curve in  $(\mathcal{M}(G))^2$
- (iii) For any  $\varphi \in C_c^\infty(\mathbb{R}^d \times (0, T))$  and  $i = 1, 2$ , we have

$$\int_0^T \int_{\mathbb{R}^d} \partial_t \varphi_t(x) d\rho_t^{(i)}(x) dt + \frac{1}{2} \int_0^T \iint_G \bar{\nabla} \varphi_t(x, y) \eta(x, y) d\mathbf{j}_t^{(i)}(x, y) dt = 0. \quad (23)$$

We denote the set of all weak solutions on the time interval  $[0, T]$  by  $\text{CE}_T$ . For  $\mathbf{q}_0, \mathbf{q}_1 \in (\mathcal{P}(\mathbb{R}^d))^2$ , we write  $(\boldsymbol{\rho}, \mathbf{j}) \in \text{CE}_T(\mathbf{q}_0, \mathbf{q}_1)$  if  $(\boldsymbol{\rho}, \mathbf{j}) \in \text{CE}_T$  and, additionally,  $\rho_0 = \mathbf{q}_0, \rho_T = \mathbf{q}_1$ . We will often shorten notation and write  $\text{CE} := \text{CE}_1$ .

We make the following observations:

**Remark 2.19.** (i) Since  $|\bar{\nabla} \varphi(x, y)| \leq \|\varphi\|_{C^1(\mathbb{R}^d)} 2 \wedge |x - y|$ , the continuity equation is well-defined under the integrability condition

$$\int_0^T \iint_G 2 \wedge |x - y| \eta(x, y) d|\mathbf{j}_t^{(i)}|(x, y) dt < \infty, \quad \text{for } i = 1, 2. \quad (24)$$

By Corollary 2.16, this condition is satisfied for any pair  $(\boldsymbol{\rho}, \mathbf{j})$  with  $\int_0^T \mathcal{A}_{m, \beta}(\mu; \rho_t, \mathbf{j}_t) dt < \infty$ .

- (ii) The continuity equation holds for more general test functions. Indeed, regularizing via convolution, we immediately see that (23) also holds for  $\varphi \in C_c^1(\mathbb{R}^d \times (0, T))$ . Under the integrability condition (24) we can also consider bounded test functions  $\varphi \in C_b^1(\mathbb{R}^d \times (0, T))$ , whose support has a compact projection in  $(0, T)$ . To see this, we approximate  $\varphi$  by  $\varphi \chi_R$ , where  $\chi_R \in C_c^\infty(\mathbb{R}^d)$ ,  $0 \leq \chi_R \leq 1$  and  $\chi_R \equiv 1$  on  $B_R(0)$ .
- (iii) The continuity equation is decoupled with respect to the different components  $i \in \{1, 2\}$  of  $\boldsymbol{\rho}$  and  $\mathbf{j}$ .

Since both the action density functional  $\mathcal{A}_{m, \beta}$  as well as the continuity equations are fully decoupled with respect to the different species, previous remarks yield analogues of [35, Lemma 2.15, Lemma 2.16 and Proposition 2.17] for the two-species case.

**Lemma 2.20.** *Let  $\boldsymbol{\rho}$  and  $\mathbf{j}$  be Borel families of measures in  $(\mathcal{P}(\mathbb{R}^d))^2$  and  $(\mathcal{M}(\mathbb{R}^d))^2$  satisfying (23) and (24). Then, there exist weakly continuous curves  $\bar{\boldsymbol{\rho}} \subset (\mathcal{P}(\mathbb{R}^d))^2$  such that  $\bar{\rho}_t^{(i)} = \rho_t^{(i)}$  for a.e.  $t \in [0, T]$  and  $i = 1, 2$ . Moreover, for any  $\varphi \in C_b^1([0, T] \times \mathbb{R}^d)$  and any  $0 \leq s \leq t \leq T$  and  $i = 1, 2$  it holds*

$$\begin{aligned} \int_{\mathbb{R}^d} \varphi_t(x) d\bar{\rho}_t^{(i)}(x) - \int_{\mathbb{R}^d} \varphi_s(x) d\bar{\rho}_s^{(i)}(x) &= \int_s^t \int_{\mathbb{R}^d} \partial_t \varphi_t(x) d\rho_t^{(i)}(x) dt \\ &+ \frac{1}{2} \int_s^t \iint_G \bar{\nabla} \varphi_t(x, y) \eta(x, y) d\mathbf{j}_t^{(i)}(x, y) dt. \end{aligned} \quad (25)$$

*Proof.* This is an adaptation of [3, Lemma 8.1.2]. The required estimate on the time derivatives  $\partial_t \rho_t^{(i)}$  is provided by Corollary 2.16 as described in Remark 2.19 (i). Finally, similar to Remark 2.19 (ii), we can lower the regularity and compactness assumptions on the test functions  $\varphi$ .  $\square$

**Lemma 2.21** (Time-uniformly bounded  $p$ -th moments). *Let  $(\mu^n)_{n \in \mathbb{N}} \subset \mathcal{M}^+(\mathbb{R}^d)$  satisfy (MB1) and (MB2) uniformly in  $n$ . Let  $(\rho_0^{n,(i)})_{n \in \mathbb{N}} \subset \mathcal{P}(\mathbb{R}^d)$  be such that  $\sup_{n \in \mathbb{N}} M_p(\rho_0^{n,(i)}) < \infty$  and let  $(\rho^n, \mathbf{j}^n)_{n \in \mathbb{N}} \subset \text{CE}_T$  be such that  $\sup_{n \in \mathbb{N}} \int_0^T \mathcal{A}_{m,\beta}(\mu^n; \rho_t^n, \mathbf{j}_t^n) dt < \infty$ . Then, we have*

$$\sup_{n \in \mathbb{N}} \sup_{t \in [0, T]} M_p(\rho_t^{n,(i)}) < \infty.$$

*Proof.* We argue similarly to [35]. Since  $|x|^p$  is not admissible in (23), we introduce a smooth cut-off  $\varphi_R \in C_c^\infty(\mathbb{R}^d; [0, 1])$  satisfying  $\varphi_R|_{B_R(0)} \equiv 1$ ,  $\text{supp } \varphi_R \subset B_{2R}(0)$  and  $|\nabla \varphi_R| \leq 2/R$ . Then, we define  $\psi_R(x) := \varphi_R^p(x)(1 + |x|)^p$ , which is admissible in (23), giving us

$$\begin{aligned} \frac{d}{dt} \sum_{i=1}^2 \int_{\mathbb{R}^d} \psi_R d\rho_t^{n,(i)} &= -\frac{1}{2} \sum_{i=1}^2 \iint_G \bar{\nabla} \psi_R \eta d\mathbf{j}_t^{n,(i)} \\ &\leq \frac{1}{2} M \mathcal{A}_{m,\beta}^{1/p}(\mu; \rho_t, \mathbf{j}_t) \sum_{i=1}^2 \left( \iint_G |\bar{\nabla} \psi_R|^q \eta d(\mu \otimes \mu + \rho_t^{n,(i)} \otimes \mu + \mu \otimes \rho_t^{n,(i)}) \right)^{1/q}, \end{aligned}$$

where we used Lemma 2.14 in the last step. To bound the right-hand side by the moment, we shorten notation by introducing  $\zeta := \varphi_R(x)|x|$  and  $\xi := \varphi_R(y)|y|$ , and calculate

$$|\bar{\nabla} \psi_R(x, y)|^q = |\varphi_R^p(x)(1 + |x|)^p - \varphi_R^p(y)(1 + |y|)^p|^q \leq C_1 (|\varphi_R^p(x) - \varphi_R^p(y)|^q + |\zeta^p - \xi^p|^q).$$

Since  $\varphi_R \leq 1$  and  $|\nabla \varphi_R| \leq 2/R$ , we have

$$|\varphi_R^p(x) - \varphi_R^p(y)|^q \leq C_2 |\varphi_R(x) - \varphi_R(y)|^q \leq 2^q C_2 / R^q |x - y|^q \leq C_2 |x - y|^q,$$

for every  $R \geq 2$  and where we can always choose  $C_2 \leq [p]$ . To bound the second term we employ the mean value theorem for the function  $z \mapsto z^p$  and obtain

$$|\zeta^p - \xi^p|^q \leq (p|\zeta - \xi|(\zeta + \xi)^{p-1})^q = p^q |\zeta - \xi|^q (2r - \zeta + \xi)^p \leq C_3 (|\zeta - \xi|^{pq} + |\zeta - \xi|^q \zeta^p).$$

Since  $x \mapsto \varphi(x)|x|$  is globally Lipschitz, there exists a constant  $C_4 > 0$  independent of  $R$ , such that

$$|\bar{\nabla} \psi_R(x, y)|^q \leq C_4 (1 + |x|^p) (|x - y|^q \vee |x - y|^{pq}).$$

Thus, sending  $R \rightarrow \infty$  and using (MB1) as well as (MB2), we find

$$\begin{aligned} \frac{d}{dt} \sum_{i=1}^2 \int_{\mathbb{R}^d} (1 + |x|^p) d\rho_t^{n,(i)}(x) &\leq M C_4 C_\eta^{1/q} \mathcal{A}_{m,\beta}^{1/p}(\mu; \rho_t, \mathbf{j}_t) \sum_{i=1}^2 \left( \frac{1}{2} C_\mu + \int_{\mathbb{R}^d} (1 + |x|^p) d\rho_t^{n,(i)}(x) \right)^{1/q} \\ &\leq \frac{1}{2} (C_\mu + 2) M C_4 C_\eta^{1/q} \mathcal{A}_{m,\beta}^{1/p}(\mu; \rho_t, \mathbf{j}_t) \sum_{i=1}^2 \left( \int_{\mathbb{R}^d} (1 + |x|^p) d\rho_t^{n,(i)}(x) \right)^{1/q}. \end{aligned}$$

Integrating this inequality in time, estimating the  $p$ -th power of the sum by the sum of  $p$ -th powers and using Hölder's inequality, we arrive at

$$\sum_{i=1}^2 \int_{\mathbb{R}^d} (1 + |x|^p) d\rho_t^{n,(i)}(x) \leq C \sum_{i=1}^2 \int_{\mathbb{R}^d} (1 + |x|^p) d\rho_0^{n,(i)}(x) + C T^{p-1} \int_0^T \mathcal{A}_{m,\beta}(\mu; \rho, \mathbf{j}) dt,$$

for some constant  $C > 0$ . Taking the supremum over  $n \in \mathbb{N}$  and  $t \in [0, T]$  finishes the proof.  $\square$

**Proposition 2.22** (Compactness of solutions to the nonlocal continuity equation). *Let  $(\mu^n)_{n \in \mathbb{N}} \subset \mathcal{M}^+(\mathbb{R}^d)$  and suppose that  $(\mu^n)_{n \in \mathbb{N}}$  weakly-\* converges to  $\mu \in \mathcal{M}^+(\mathbb{R}^d)$ . Moreover, assume that the base measures  $\mu^n$  and  $\mu$  satisfy (MB1), (MB2) and (BC) uniformly in  $n$ . Let  $((\rho^n, \mathbf{j}^n))_{n \in \mathbb{N}} \subset \text{CE}_T$  be such that  $\sup_{n \in \mathbb{N}} M_p(\rho_0^{n,(i)}) < \infty$*

and  $\sup_{n \in \mathbb{N}} \int_0^T \mathcal{A}_{m,\beta}(\mu_t^n; \rho_t^n, j_t^n) dt < \infty$ . Then, there exists  $(\rho, \mathbf{j}) \in \text{CE}_T$  such that, up to a subsequence (still indexed by  $n$ ), as  $n \rightarrow \infty$ , it holds

$$\begin{aligned} \rho_t^n &\rightharpoonup \rho_t && \text{narrowly for all } t \in [0, T], \\ \mathbf{j}^n &\rightharpoonup^* \mathbf{j} && \text{in } (\mathcal{M}(G \times [0, T]))^2. \end{aligned}$$

Moreover, the action is lower semicontinuous along the above subsequences i.e., we have

$$\liminf_{n \rightarrow \infty} \int_0^T \mathcal{A}_{m,\beta}(\mu_t^n; \rho_t^n, j_t^n) dt \geq \int_0^T \mathcal{A}_{m,\beta}(\mu_t; \rho_t, j_t) dt.$$

*Proof.* We argue similarly to [35, Proposition 2.17] and present only the main steps. At first, we employ Lemma 2.21, Corollary 2.16, Hölder's inequality, Assumption (W) and the disintegration theorem (see e.g. [3, Theorem 5.3.1]) to obtain a subsequence still denoted by  $(j_t^n)_n$  which weakly-\* converges to a Borel family  $(j_t)_{t \in [0, T]}$  such that for  $i = 1, 2$  we have  $j^{(i)}(K \times I) = \int_I j_t^{(i)}(K) dt$  as well as (24) for any compact sets  $I \subset [0, T]$ ,  $K \subset G$ . Then, for  $0 \leq s \leq t \leq T$  and  $\varphi \in C_c^\infty(\mathbb{R}^d)$ , we obtain the equality

$$\lim_{n \rightarrow \infty} \int_s^t \iint_G \bar{\nabla} \varphi(x, y) \eta(x, y) d j_t^{n, (i)}(x, y) dt = \int_s^t \iint_G \bar{\nabla} \varphi(x, y) \eta(x, y) d j_t^{(i)}(x, y) dt$$

by employing a truncation argument and using the Assumptions (MB1), (MB2), (BC) and (W) as well as Hölder's inequality and Lemma 2.14. Since  $(\rho_0^{n, (i)})_n$  has uniformly bounded  $p$ -th moments, it is uniformly tight. Hence, Prokhorov's theorem (see e.g. [3, Theorem 5.1.3]), the above convergence result and (25), where we choose  $\varphi_t = \xi$ , yield local narrow convergence of  $(\rho_t^{n, (i)})_n$  to some  $\rho_t^{(i)} \in \mathcal{M}^+(\mathbb{R}^d)$  for  $i = 1, 2$ . In the last step we use (24) and Corollary 2.16 to find that  $\rho_t^{(i)} \in \mathcal{P}(\mathbb{R}^d)$  and employ Lemma 2.21 to obtain that the narrow convergence of  $(\rho_t^{n, (i)})_n$  towards  $\rho_t^{(i)}$  is in fact global. Therefore, we have  $(\rho, \mathbf{j}) \in \text{CE}_T$ . Finally, since narrow convergence implies weak-\* convergence, Lemma 2.13 shows the claim of lower semicontinuity of the action.  $\square$

**Remark 2.23.** Note that in Proposition 2.22 we have compactness of the action density not only in  $\rho_t$  and  $j_t$ , but also in the base measure  $\mu_t$ . This will play a crucial role later in the proof of existence.

### 2.3. Definition of a quasimetric

Having defined an action density and a continuity equation, we are now ready to define the induced quasimetric:

**Definition 2.24** (Nonlocal upwind transportation cost for two species). For  $\mu \in \mathcal{M}^+(\mathbb{R}^d)$ ,  $\eta$  satisfying (MB1), (MB2), (BC), and  $\mathcal{Q}_0, \mathcal{Q}_1 \in (\mathcal{P}(\mathbb{R}^d))^2$ , the nonlocal upwind transportation cost between  $\mathcal{Q}_0$  and  $\mathcal{Q}_1$  is defined as

$$\mathcal{T}_{m,\beta,\mu}(\mathcal{Q}_0, \mathcal{Q}_1) = \left( \inf \left\{ \int_0^1 \mathcal{A}_{m,\beta}(\mu; \rho_t, j_t) dt : (\rho, \mathbf{j}) \in \text{CE}(\mathcal{Q}_0, \mathcal{Q}_1) \right\} \right)^{1/p}. \quad (26)$$

**Remark 2.25** (Decoupling of the transportation cost). Let us denote the nonlocal upwind transportation cost for one species by  $\bar{\mathcal{T}}_{m,\mu}$ . Then, since both the action and the continuity equation are decoupled with respect to the components of  $\rho$  and  $\mathbf{j}$ , the infima are also independent of each other, which implies that for any  $\mathcal{Q}_0, \mathcal{Q}_1 \in (\mathcal{P}(\mathbb{R}^d))^2$  we have

$$\mathcal{T}_{m,\beta,\mu}^p(\mathcal{Q}_0, \mathcal{Q}_1) = \frac{1}{\beta^{(1)}} \bar{\mathcal{T}}_{m,\mu}^p(\varrho_0^{(1)}, \varrho_1^{(1)}) + \frac{1}{\beta^{(2)}} \bar{\mathcal{T}}_{m,\mu}^p(\varrho_0^{(2)}, \varrho_1^{(2)}). \quad (27)$$

**Theorem 2.26** (Optimal curves exist and are constant speed geodesics). For any  $\mu \in \mathcal{M}^+(\mathbb{R}^d)$  satisfying (MB1), (MB2) and (BC), any  $T \geq 0$  and any  $\mathcal{Q}_0, \mathcal{Q}_1 \in (\mathcal{P}(\mathbb{R}^d))^2$  with  $\mathcal{T}_{m,\beta,\mu}(\mathcal{Q}_0, \mathcal{Q}_1) < \infty$ , the infimum in (26) is attained by a curve  $(\rho, \mathbf{j}) \in \text{CE}(\mathcal{Q}_0, \mathcal{Q}_1)$  with  $\mathcal{A}_{m,\beta}(\mu; \rho_t, j_t) = \mathcal{T}_{\beta,\mu}^2(\mathcal{Q}_0, \mathcal{Q}_1)$  for a.e.  $t \in [0, 1]$ . This curve is a constant-speed geodesic, i.e., it satisfies

$$\mathcal{T}_{m,\beta,\mu}(\rho_s, \rho_t) = |t - s| \mathcal{T}_{m,\beta,\mu}(\mathcal{Q}_0, \mathcal{Q}_1), \text{ for every } s, t \in [0, 1].$$

*Proof.* This can be proved by using [35, Theorem 2.20] and the decoupling of  $\mathcal{A}_{m,\beta}$ . Alternatively, one could infer this from Lemma 2.27 similar to the proof of [33, Theorem 4.3].  $\square$

**Lemma 2.27** (Reparametrization). *For any  $\mu \in \mathcal{M}^+(\mathbb{R}^d)$  satisfying (MB1), (MB2) and (BC), any  $T \geq 0$  and any  $\varrho_0, \varrho_1 \in (\mathcal{P}(\mathbb{R}^d))^2$  it holds*

$$\mathcal{T}_{m,\beta,\mu}(\varrho_0, \varrho_1) = \inf \left\{ \int_0^T \mathcal{A}_{m,\beta}^{1/p}(\mu; \rho_t, \mathbf{j}_t) dt : (\rho, \mathbf{j}) \in \text{CE}_T(\varrho_0, \varrho_1) \right\}.$$

*Proof.* This immediately follows from Theorem 2.26. Alternatively, one can argue via a reparametrization argument similar to the one in the proof of [31, Theorem 5.4].  $\square$

**Proposition 2.28.** *Let  $\mu \in \mathcal{M}^+(\mathbb{R}^d)$  satisfy (MB1) and (MB2). Then, for any  $\varrho_0, \varrho_1 \in (\mathcal{P}(\mathbb{R}^d))^2$ , again denoting by  $\overline{\mathcal{T}}_{m,\mu}$  the transportation cost for one species, there exists  $C > 0$  such that*

$$W_1(\varrho_0, \varrho_1) := \sum_{i=1}^2 W_1(\varrho_0^{(i)}, \varrho_1^{(i)}) \leq C \overline{\mathcal{T}}_{m,\beta,\mu}^{1/p}(\varrho_0, \varrho_1).$$

*Proof.* By Remark 2.19, in (25) we can choose a test function  $\psi$ , which is constant in time, 1-Lipschitz in space and satisfies  $0 \leq \psi \leq 1$ . Then, a quick calculation yields a uniform bound, which allows us to take the supremum over all  $\psi$  and employ the Kantorovich-Rubinstein formula to obtain the result. A detailed proof for  $p = 2$  and  $m(r, s) = r$  (which does not change the argument) can be found in [35, Proposition 2.21].  $\square$

Observe that, by the previous Proposition, Young's inequality and (27), we have

$$W_1(\varrho_0, \varrho_1) \leq \bar{C} \sum_{i=1}^2 \overline{\mathcal{T}}_{m,\mu}^{1/p}(\varrho_0^{(i)}, \varrho_1^{(i)}) \leq \bar{C} \left( \sum_{i=1}^2 \overline{\mathcal{T}}_{m,\mu}^p(\varrho_0^{(i)}, \varrho_1^{(i)}) \right)^{1/p} \leq C \overline{\mathcal{T}}_{m,\beta,\mu}^{1/p}(\varrho_0, \varrho_1),$$

where  $C$  depends only on the  $m, p, \beta, \eta$  and  $\mu$ . Hence,  $\mathcal{T}_{m,\beta,\mu}$  defines a quasimetric on  $(\mathcal{P}(\mathbb{R}^d))^2$  and induces a topology stronger than the  $W_1$ -topology:

**Theorem 2.29.** *Let  $\mu \in \mathcal{M}^+(\mathbb{R}^d)$  satisfy (MB1), (MB2) and (BC). Then, the nonlocal upwind transportation cost for two species  $\mathcal{T}_{m,\beta,\mu}$  defines a quasimetric on  $(\mathcal{P}_p(\mathbb{R}^d))^2$  and the map  $(\varrho_0, \varrho_1) \mapsto \mathcal{T}_{m,\beta,\mu}(\varrho_0, \varrho_1)$  is lower semicontinuous with respect to the narrow convergence. The topology induced by  $\mathcal{T}_{m,\beta,\mu}$  is stronger than the  $W_1$ -topology and the narrow topology. In particular, bounded sets in  $((\mathcal{P}_p(\mathbb{R}^d))^2, \mathcal{T}_{m,\beta,\mu})$  are narrowly relatively compact.*

*Proof.* Similar to [35, Theorem 2.22] we have that if  $\mathcal{T}_{m,\beta,\mu}(\varrho_0, \varrho_1) = 0$ , then the minimizing pair  $(\rho, \mathbf{j}) \in \text{CE}(\varrho_0, \varrho_1)$  satisfies  $\mathcal{A}_{m,\beta}(\mu; \rho_t, \mathbf{j}_t) = 0$  for a.e.  $t \in [0, T]$ . Thus, for  $i = 1, 2$  we have  $\mathbf{j}_t^{(i)} \equiv 0$ ,  $(\mu \otimes \mu + \mathcal{S}_t^{(i)})$ -a.e. and hence  $\rho_0^{(i)} \equiv \rho_1^{(i)}$ . The triangle inequality follows from Lemma 2.27 by concatenating the solutions of the nonlocal continuity equation. The compactness and lower semicontinuity of  $\mathcal{A}_{m,\beta}$  shown in Proposition 2.22 are inherited by  $\mathcal{T}_{m,\beta,\mu}$ . Lastly, the claims about the topology immediately follow from Proposition 2.28.  $\square$

Now, we adapt the definition of absolutely continuous curves to our setting.

**Definition 2.30** (Absolutely continuous curves). *Let  $\mu \in \mathcal{M}^+(\mathbb{R}^d)$  satisfy (MB1), (MB2) and (BC). A curve  $\rho \subset (\mathcal{P}(\mathbb{R}^d))^2$  belongs to  $AC^p([0, T]; ((\mathcal{P}_p(\mathbb{R}^d))^2, \mathcal{T}_{m,\beta,\mu}))$  if there exists  $f \in L^p(0, T)$  such that for any  $0 < s \leq t < T$  we have*

$$\mathcal{T}_{m,\beta,\mu}(\rho_s, \rho_t) \leq \int_s^t f(t) dt. \quad (28)$$

Such a curve is called  $(p)$ -absolutely continuous. For any  $\rho \in \text{AC}^p([0, T]; ((\mathcal{P}_p(\mathbb{R}^d))^2, \mathcal{T}_{m, \beta, \mu}))$  and a.e.  $t \in [0, T]$  the limit

$$|\rho'_t| := \lim_{h \rightarrow 0} \frac{\mathcal{T}_{m, \beta, \mu}(\rho_t, \rho_{t+h})}{|h|}$$

is well-defined.<sup>1</sup> It is called the metric derivative of  $\rho$  at  $t$ . The map  $t \mapsto |\rho'_t|$  belongs to  $L^p(0, T)$  and satisfies  $|\rho'_t| \leq f(t)$  for any  $m$  satisfying (28), making it the minimal integrand in (28).

**Proposition 2.31** (Metric velocity). *Let  $\mu \in \mathcal{M}^+(\mathbb{R}^d)$  satisfy (MB1), (MB2) and (BC). A curve  $\rho \subset (\mathcal{P}_p(\mathbb{R}^d))^2$  belongs to  $\text{AC}^p([0, T]; ((\mathcal{P}_p(\mathbb{R}^d))^2, \mathcal{T}_{m, \beta, \mu}))$  if and only if there exists a family  $\mathbf{j}$  such that  $(\rho, \mathbf{j}) \in \text{CE}_T$  and*

$$\int_0^T \mathcal{A}_{m, \beta}^{1/p}(\mu; \rho_t, \mathbf{j}_t) dt < \infty.$$

In this case, the metric derivative satisfies  $|\rho'|^p(t) \leq \mathcal{A}_{m, \beta}(\mu; \rho_t, \mathbf{j}_t)$  for a.e.  $t$ . Additionally, there exists a unique family  $\mathbf{j}$  such that  $(\rho, \mathbf{j}) \in \text{CE}_T$  and

$$|\rho'_t|^p = \mathcal{A}_{m, \beta}(\mu; \rho_t, \mathbf{j}_t) \text{ for a.e. } t \in [0, T].$$

This identity holds if and only if  $\mathbf{j}_t \in T_\rho(\mathcal{P}_p(\mathbb{R}^d))^2$  for a.e.  $t$ , where we define the tangent space at  $\rho$  as

$$T_\rho(\mathcal{P}_p(\mathbb{R}^d))^2 := \{\mathbf{j} \in (\mathcal{M}_{\eta\gamma_1}^{\text{as}}(G))^2 : \mathcal{A}_{m, \beta}(\mu; \rho, \mathbf{j}) \leq \mathcal{A}_{m, \beta}(\mu; \rho, \mathbf{j} + d), \forall d \in (\mathcal{M}_{\text{div}}(G))^2\} \quad (29)$$

and the space of divergence free-fluxes as

$$\mathcal{M}_{\text{div}}(G) := \left\{ d \in \mathcal{M}(G) : \iint_G \bar{\nabla} \varphi \eta dd = 0 \text{ for any } \varphi \in C_c^\infty(\mathbb{R}^d) \right\}.$$

*Proof.* The first statement about the characterization of absolutely continuous curves follows from [31, Theorem 5.17], due to Theorem 2.26, Lemma 2.27 and Proposition 2.22. Since, by Corollary 2.12, we have that antisymmetric fluxes have lower action, it is not restrictive to require the minimizing flux to lie in  $(\mathcal{M}_{\eta\gamma_1}^{\text{as}}(G))^2$ . For the converse statement, we argue as in the proof of [35, Proposition 2.25]. Here we use that the map  $j \mapsto \mathcal{A}_{m, \beta}(\mu; \rho, j)$  is strictly convex for  $j \in (\mathcal{M}_{\eta\gamma_1}^{\text{as}}(G))^2$  with  $\mathcal{A}_{m, \beta}(\mu; \rho, j) < \infty$  and that the set  $\{j \in (\mathcal{M}_{\eta\gamma_1}^{\text{as}}(G))^2 : \bar{\nabla} j = \bar{\nabla} j_t\}$  is closed with respect to weak-\* convergence. Additionally, we employ Corollary 2.16 and obtain that  $j \mapsto \mathcal{A}_{m, \beta}(\mu; \rho, j)$  has locally relatively compact sublevel sets with respect to narrow convergence, by arguing as in the proof of Proposition 2.22. Finally, applying the direct method of calculus of variations, we see that  $\mathbf{j}$  is well-defined.  $\square$

**Definition 2.32.** Recall that for any  $\rho \in (\mathcal{P}_p(\mathbb{R}^d))^2$ , by Lemma 2.10, we can identify any  $j \in (\mathcal{M}(G))^2$  such that  $\mathcal{A}_{m, \beta}(\mu; \rho, j) < \infty$  with a velocity  $\mathbf{v} = (v^{(1)}, v^{(1)\perp}, v^{(2)}, v^{(2)\perp})$  given as

$$\begin{aligned} dj^{(i)\mu} &= (v_+^{(i)})^{q-1} d\gamma_1^{(i)} - (v_-^{(i)})^{q-1} d\gamma_2^{(i)}, \\ dj^{(i)\perp} &= (v_+^{(i)\perp})^{q-1} d\gamma_1^{(i)\perp} - (v_-^{(i)\perp})^{q-1} d\gamma_2^{(i)\perp}. \end{aligned}$$

We define as  $\tilde{T}_\rho(\mathcal{P}_p(\mathbb{R}^d))^2$  the set of velocities  $\mathbf{v}$  associated this way to  $j \in T_\rho(\mathcal{P}_p(\mathbb{R}^d))^2$ .

In the following proposition we present a characterization of tangent velocities in cases, when  $R \wedge S < \infty$  or  $m_\infty \equiv 0$ . In these cases they lie in the closure of the set of gradients of smooth functions.

**Proposition 2.33** (Tangent velocities are almost gradient). *Assume that either  $R \wedge S < \infty$  or  $m_\infty \equiv 0$ . Let  $\mu \in \mathcal{M}^+(\mathbb{R}^d)$  satisfy (MB1), (MB2) and (BC). Let  $\rho \in (\mathcal{P}(\mathbb{R}^d))^2$  and  $\mathbf{v} = (v^{(1)}, 0, v^{(2)}, 0) : G \rightarrow \mathbb{R}^4$  be associated to  $j \in (\mathcal{M}(G))^2$  satisfying  $\mathcal{A}_{m, \beta}(\mu; \rho, j) < \infty$  as before. Then, we have  $\mathbf{v} \in \tilde{T}_\rho(\mathcal{P}_p(\mathbb{R}^d))^2$  if and only if*

$$v^{(i)} \in \overline{\{\bar{\nabla} \varphi : \varphi \in C_c^\infty(\mathbb{R}^d)\}}^{L^q(\eta \hat{\gamma}_v^{(i)})}, \quad \text{where } d\hat{\gamma}_v^{(i)} = \mathbb{1}_{\{v^{(i)} > 0\}} d\gamma_1^{(i)} + \mathbb{1}_{\{v^{(i)} < 0\}} d\gamma_2^{(i)}.$$

<sup>1</sup>For details see e.g. [3, Theorem 1.1.2].

*Proof.* The boundedness of the action implies that the singular part vanishes. Thus, we can argue similarly to the proof of [35, Proposition 2.26]. Let  $\mathbf{j} \in (\mathcal{M}(G))^2$  be the flux associated to  $\nu$ . We define  $J_+^{(i)} := \text{supp } j_+^{(i)}$  and  $\gamma_+^{(i)} := \gamma_1^{(i)}|_{J_+^{(i)}}$ . Observe that by the antisymmetry of  $j^{(i)}$  we have  $(J_+^{(i)})^\top = \text{supp } j_-^{(i)}$ . Therefore, if  $\mathcal{A}_{m,\beta}(\mu; \rho, \mathbf{j}) < \infty$ , by Lemma 2.10 and the assumptions on  $m$ , for  $i = 1, 2$  there exist antisymmetric functions  $f^{(i)}$  such that  $\text{d}j^{(i)} = f^{(i)} \text{d}(\gamma_+^{(i)} + (\gamma_+^{(i)})^\top)$ . These functions satisfy

$$\bar{\mathcal{A}}_m(\mu; \rho^{(i)}, j^{(i)}) = \|f_+^{(i)}\|_{L^p(\eta\gamma_1^{(i)})}^p = \|f^{(i)}\|_{L^p(\eta\gamma_+^{(i)})}^p.$$

By symmetry, we can rewrite the divergence as

$$\frac{1}{2} \iint_G \bar{\nabla} \phi \eta \text{d}j^{(i)} = \iint_G \bar{\nabla} \phi \eta \text{d}j_+^{(i)} = \iint_G \bar{\nabla} \phi f^{(i)} \eta \text{d}\gamma_+^{(i)}.$$

Now, we observe that (29) is equivalent to

$$\iint_G |f^{(i)}|^p - |f^{(i)} + g^{(i)}|^p \eta \text{d}\gamma_+^{(i)} \leq 0, \quad (30)$$

for all antisymmetric  $g^{(i)} \in L^p(\eta\gamma_+^{(i)})$ , which satisfy  $\iint_G \bar{\nabla} \cdot \psi g^{(i)} \eta \text{d}\gamma_+^{(i)} = 0$  for every  $\psi \in C_c^\infty(\mathbb{R}^d)$ . Since the sign of  $g^{(i)}$  may be negative, (30) is equivalent to

$$(f^{(i)})^{p-1} g^{(i)} = 0, \quad \eta\gamma_+^{(i)}\text{-a.e.},$$

which is equivalent to

$$(f_+^{(i)})^{p-1} g^{(i)} = 0, \quad \eta\gamma^{(i)}\text{-a.e.}$$

Now, note that we have  $\nu_+^{(i)} = (f_+^{(i)})^{p-1}$ . Hence,  $\nu_+^{(i)}$  belongs to the closure of  $\{\bar{\nabla} \varphi : \varphi \in C_c^\infty(\mathbb{R}^d)\}$  in  $L^q(\eta\gamma_1^{(i)})$ . Finally, recalling that  $\nu^{(i)}$  are antisymmetric and that  $(\gamma_1^{(i)})^\top = \gamma_2^{(i)}$ , the claim follows.  $\square$

**Remark 2.34.** Proposition 2.33 shows that if  $R \wedge S < \infty$  or  $m_\infty \equiv 0$  and for  $\mu$  and  $\rho$  as in the statement, for  $\mathbf{j}$  chosen from a dense subset of  $T_\rho(\mathcal{P}_p(\mathbb{R}^d))^2$ , there exists a measurable function  $\varphi = (\varphi^{(1)}, 0, \varphi^{(2)}, 0) : \mathbb{R}^d \rightarrow \mathbb{R}^4$ , such that we have

$$\mathcal{A}_{m,\beta}(\mu; \rho, \mathbf{j}) = \tilde{\mathcal{A}}_{m,\beta}(\mu; \rho, \bar{\nabla} \varphi),$$

and we can then write

$$\text{d}j^{(i)} = ((\bar{\nabla} \varphi^{(i)})_+)^{q-1} \text{d}\gamma_1^{(i)} - ((\bar{\nabla} \varphi^{(i)})_-)^{q-1} \text{d}\gamma_2^{(i)}. \quad (31)$$

**Proposition 2.35** (Absolutely continuous curves stay supported in  $\text{supp } \mu$ ). *Let  $\mu \in \mathcal{M}^+(\mathbb{R}^d)$  satisfy (MB1), (MB2) and (BC) and let  $\rho \in \text{AC}^p([0, T]; ((\mathcal{P}_p(\mathbb{R}^d))^2, \mathcal{T}_{m,\beta,\mu}))$  be such that  $\text{supp } \rho_0^{(i)} \subseteq \text{supp } \mu$ ,  $i = 1, 2$ . Additionally, assume that  $m$  satisfies the following condition:*

$$R \wedge S < \infty \quad \text{or} \quad m_\infty \equiv 0 \quad \text{or} \quad m(r, s) = 0 \iff r = 0. \quad (\text{A})$$

Then, we have  $\text{supp } \rho_t^{(i)} \subseteq \text{supp } \mu$  for all  $t \in [0, T]$ ,  $i = 1, 2$ .

*Proof.* If  $R \wedge S < \infty$  or  $m_\infty \equiv 0$ , this is immediate from the finiteness of the action. Thus, let  $R = S = \infty$  and assume  $m(r, s) > 0$  for every  $r > 0$ . This allows us to argue similarly to [35, Proposition 2.28] by employing Proposition 2.31, Lemma 2.10 and Lemma 2.20 to obtain a pair  $(\rho, \mathbf{j}) \in (\mathcal{P}_p(\mathbb{R}^d))^2 \times (\mathcal{M}_{\eta\gamma_1}^{\text{as}}(G))^2$  satisfying (25) and we have for  $i = 1, 2$

$$\begin{aligned} \text{d}j_t^{(i)\mu} &= (f_t^{(i)})_+ \text{d}(\mu \otimes \mu) - (f_t^{(i)})_- \text{d}(\mu \otimes \mu), \\ \text{d}j_t^{(i)\perp} &= (f_t^{(i)\perp})_+ \text{d}(\rho_t^{(i)\perp} \otimes \mu) - (f_t^{(i)\perp})_- \text{d}(\mu \otimes \rho_t^{(i)\perp}), \end{aligned}$$

with suitable antisymmetric  $f_t^{(i)}$  and  $f_t^{(i)\perp}$ . Here we used that  $\gamma_{1,t}^{(i)\perp} \ll \rho_t^{(i)\perp} \otimes \mu$  and  $\gamma_{2,t}^{(i)\perp} \ll \mu \otimes \rho_t^{(i)\perp}$  since we assumed  $m(r, s) = 0$  if and only if  $r = 0$ . Inserting  $\varphi \in C_c^\infty(\mathbb{R}^d)$  with  $\text{supp } \varphi \subset \mathbb{R}^d \setminus \text{supp } \mu$  and  $\varphi \geq 0$  into (25), by the antisymmetry of  $v_t^{(i)}$  and  $v_t^{(i)\perp}$ , we obtain for both  $i = 1, 2$

$$\begin{aligned} \int_{\mathbb{R}^d} \varphi(x) d\rho_t^{(i)}(x) &= \int_{\mathbb{R}^d} \varphi(x) d\rho_0^{(i)}(x) + \int_0^t \iint_G \bar{\nabla} \varphi(x, y) (f_\tau^{(i)})_+(x, y) \eta(x, y) d\mu(x) d\mu(y) d\tau \\ &\quad + \int_0^t \iint_G (\varphi(y) - \varphi(x)) (f_\tau^{(i)\perp})_+(x, y) \eta(x, y) d\rho^{(i)\perp}(x) d\mu(y) d\tau \\ &\leq - \int_0^t \iint_G \varphi(x) (f_\tau^{(i)\perp})_+(x, y) \eta(x, y) d\rho^{(i)\perp}(x) d\mu(y) d\tau \leq 0. \end{aligned}$$

Since  $\rho_t^{(i)\perp}$  and  $\mu$  are nonnegative measures this finishes the proof.  $\square$

**Remark 2.36.** (i) Assumption (A) plays an important role in the proof of existence of gradient flows. By Proposition 2.35 it guarantees that when  $\mu$  is a counting measure and  $\text{supp } \rho_0^{(i)} \subseteq \text{supp } \mu$ , the same is true for all times. This will reduce the continuity equation to a finite system of ordinary differential equations.

(ii) On a finite graph  $m(r, s) = 0 \iff r = 0$  means that the mobility vanishes if and only if there is no mass on the node from which the mass is flowing away. This is reasonable from a model point of view.

### 3. Two nonlocally interacting species as Finsler gradient flows

In this section we define a Minkowski norm on  $T_\rho(\mathcal{P}_\rho(\mathbb{R}^d))^2$ , thereby inducing a Finslerian structure. Moreover, the inner product gives rise to a notion of gradient and divergence. Subsequently, we show that this gradient of the nonlocal cross-interaction energy,

$$\mathcal{E}(\rho) = \frac{1}{2} \sum_{i,k=1}^2 \iint_{\mathbb{R}^d \times \mathbb{R}^d} K^{(ik)}(x, y) d\rho^{(i)}(x) d\rho^{(k)}(y), \quad (2) \text{ revisited}$$

exists and is unique, whenever  $K^{(12)}$  and  $K^{(21)}$  are positive multiples of one another. Then, system (6) reads

$$\begin{aligned} \partial_t \rho_t^{(i)}(x) + \beta^{(i)} \int_{\mathbb{R}^d} (\mathbf{m}_t^{(i)}(x, y) \bar{\nabla} (K^{(i1)} * \rho_t^{(1)} + K^{(i2)} * \rho_t^{(2)})(x, y)_-)^{q-1} \eta(x, y) d\mu(y) \\ - \beta^{(i)} \int_{\mathbb{R}^d} (\mathbf{m}_t^{(i)}(y, x) \bar{\nabla} (K^{(i1)} * \rho_t^{(1)} + K^{(i2)} * \rho_t^{(2)})(x, y)_+)^{q-1} \eta(x, y) d\mu(y) \\ + \beta^{(i)} \int_{\mathbb{R}^d} (\mathbf{m}_{\infty,t}^{(i)}(x, y) \bar{\nabla} (K^{(i1)} * \rho_t^{(1)} + K^{(i2)} * \rho_t^{(2)})(x, y)_-)^{q-1} \eta(x, y) \mathcal{S}_t^{(i)}(x, dy) \\ - \beta^{(i)} \int_{\mathbb{R}^d} (\mathbf{m}_{\infty,t}^{(i)}(y, x) \bar{\nabla} (K^{(i1)} * \rho_t^{(1)} + K^{(i2)} * \rho_t^{(2)})(x, y)_+)^{q-1} \eta(x, y) \mathcal{S}_t^{(i)}(x, dy) = 0, \end{aligned} \quad (32)$$

is a gradient flow of  $\mathcal{E}$  with respect to the Finslerian structure. Here  $\beta^{(1)}, \beta^{(2)} > 0$  and, after a rescaling, we shall assume that  $\beta^{(1)} = 1$  and  $K^{(12)} = K^{(21)}$ .

This will finally allow us to deduce that weak solutions of (32) exist for a large family of base measures  $\mu$  via approximation with finite graphs. These considerations are based on known results for one species [35].

Before we construct the Finslerian structure on the product space, let us introduce our notion of weak solutions.

**Definition 3.1.** A curve  $\rho : [0, T] \rightarrow (\mathcal{P}_\rho(\mathbb{R}^d))^2$  is called a weak solution to (32) if the pair  $(\rho, \mathbf{j})$  is a weak solution of the continuity equation

$$\partial_t \rho_t + \bar{\nabla} \cdot \mathbf{j}_t = 0 \text{ on } [0, T] \times \mathbb{R}^d,$$

in the sense of Definition 2.18, where the flux  $\mathbf{j} : [0, T] \rightarrow (\mathcal{M}(G))^2$  for  $i = 1, 2$  is given by

$$\begin{aligned} d\mathbf{j}_t^{(i)\mu} &= (\beta^{(i)} (\bar{\nabla} \delta_{\rho^{(i)}} \mathcal{E}(\rho_t))_-)^{q-1} d\gamma_{1,t}^{(i)} - (\beta^{(i)} (\bar{\nabla} \delta_{\rho^{(i)}} \mathcal{E}(\rho_t))_+)^{q-1} d\gamma_{2,t}^{(i)}, \\ d\mathbf{j}_t^{(i)\perp} &= (\beta^{(i)} (\bar{\nabla} \delta_{\rho^{(i)}} \mathcal{E}(\rho_t))_-)^{q-1} d\gamma_{1,t}^{(i)\perp} - (\beta^{(i)} (\bar{\nabla} \delta_{\rho^{(i)}} \mathcal{E}(\rho_t))_+)^{q-1} d\gamma_{2,t}^{(i)\perp}. \end{aligned} \quad (33)$$

**Remark 3.2.** *There was no coupling between the two species until this point, neither in the definition of the action densities nor in the continuity equations. Only now, via the dependence of  $\frac{\delta \mathcal{E}}{\delta \rho^{(i)}}$  on both  $\rho^{(1)}$  and  $\rho^{(2)}$ , cross-interaction occurs.*

Throughout the rest of this paper, we make the following assumptions on the kernels  $K^{(ik)} : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ ,  $i, k = 1, 2$ :

$$\text{(Symmetry)} \quad \forall x, y \in \mathbb{R}^d \text{ it holds } K^{(ik)}(x, y) = K^{(ik)}(y, x), \quad (\text{K1})$$

$$\text{(Growth)} \quad \exists L_K \in (0, \infty) \text{ such that } \forall (x, y), (x', y') \in \mathbb{R}^d \times \mathbb{R}^d, \text{ and } i, k = 1, 2, \text{ we have} \quad (\text{K2})$$

$$\left| K^{(ik)}(x, y) - K^{(ik)}(x', y') \right| \leq L_K (|(x, y) - (x', y')| \vee |(x, y) - (x', y')|^p)$$

In particular, (K2) implies continuity and guarantees that the proper domain of  $\mathcal{E}$  contains  $(\mathcal{P}_p(\mathbb{R}^d))^2$ . Indeed, by (K2) there exists  $C > 0$  s.t. for all  $x, y \in \mathbb{R}^d$  we have  $|K^{(ik)}(x, y)| \leq C(1 + |x|^p + |y|^p)$  (see [35, Remark 3.2] for details).

**Proposition 3.3** (Continuity of the energy). *Let the potentials  $K^{(ik)}$ ,  $i, k = 1, 2$  satisfy (K1), (K2). Then, for any sequence  $(\rho^n)_{n \in \mathbb{N}} \subset (\mathcal{P}_p(\mathbb{R}^d))^2$  narrowly converging to some  $\rho \in (\mathcal{P}_p(\mathbb{R}^d))^2$ , we have*

$$\lim_{n \rightarrow \infty} \mathcal{E}(\rho^n) = \mathcal{E}(\rho).$$

*Proof.* Keeping in mind that  $\rho^n \rightarrow \rho$  if and only if  $\rho^{n,(i)} \rightarrow \rho^{(i)}$  for both  $i = 1, 2$  and using the assumptions on  $K^{(ik)}$ ,  $i, k = 1, 2$ , we can argue as in [35, Proposition 3.3]; we truncate the kernels to obtain bounded continuous test functions. Then, we employ Lebesgue's dominated convergence theorem and a diagonal argument.  $\square$

### 3.1. Finslerian geometry

**Definition 3.4.** (Finsler metric). *Given  $\rho \in (\mathcal{P}_p(\mathbb{R}^d))^2$  we define the function  $l_\rho : T_\rho(\mathcal{P}_p(\mathbb{R}^d))^2 \rightarrow (T_\rho(\mathcal{P}_p(\mathbb{R}^d))^2)^*$  as follows: for any  $\mathbf{j}, \bar{\mathbf{j}} \in T_\rho(\mathcal{P}_p(\mathbb{R}^d))^2$ , in the case  $R = S = \infty$ , we set*

$$l_\rho(\mathbf{j})[\bar{\mathbf{j}}] := \frac{1}{2} \sum_{i=1}^2 \frac{1}{\beta^{(i)}} \left[ \iint_G \bar{j}^{(i)}(x, y) \left( \frac{j_+^{(i)}(x, y)}{\mathbf{m}^{(i)}(x, y)} - \frac{j_-^{(i)}(x, y)}{\mathbf{m}^{(i)}(y, x)} \right)^{p-1} \eta(x, y) d\mu(x) d\mu(y) \right. \\ \left. + \iint_G \bar{j}^{(i)\perp}(x, y) \left( \frac{j_+^{(i)\perp}(x, y)}{\mathbf{m}_\infty^{(i)}(x, y)} - \frac{j_-^{(i)\perp}(x, y)}{\mathbf{m}_\infty^{(i)}(y, x)} \right)^{p-1} \eta(x, y) d\zeta^{(i)}(x, y) \right],$$

where  $\zeta^{(i)}$  are as in Lemma 2.10. In the case  $R \wedge S < \infty$ , we analogously define

$$l_\rho(\mathbf{j})[\bar{\mathbf{j}}] := \frac{1}{2} \sum_{i=1}^2 \frac{1}{\beta^{(i)}} \iint_G \bar{j}^{(i)}(x, y) \left( \frac{j_+^{(i)}(x, y)}{\mathbf{m}^{(i)}(x, y)} - \frac{j_-^{(i)}(x, y)}{\mathbf{m}^{(i)}(y, x)} \right)^{p-1} \eta(x, y) d\mu(x) d\mu(y).$$

Next, we define the Finsler metric  $F_\rho : T_\rho(\mathcal{P}_p(\mathbb{R}^d))^2 \rightarrow \mathbb{R}$  as

$$F_\rho(\mathbf{j}) := (l_\rho(\mathbf{j})[\mathbf{j}])^{1/p} = \mathcal{A}_{m, \beta}^{1/p}(\mu; \rho, \mathbf{j}).$$

Our first goal is to show that  $F$  is a Minkowski norm. To this end, we establish a Hölder-type inequality:

**Lemma 3.5** (Hölder-type inequality). *For  $\mathbf{j}, \bar{\mathbf{j}} \in T_\rho(\mathcal{P}_p(\mathbb{R}^d))^2$  it holds*

$$l_\rho(\mathbf{j})[\bar{\mathbf{j}}] \leq (l_\rho(\bar{\mathbf{j}})[\bar{\mathbf{j}}])^{1/p} (l_\rho(\mathbf{j})[\mathbf{j}])^{1/q}. \quad (34)$$

We have equality in (34) if and only if there exists  $\lambda \geq 0$  such that for  $i = 1, 2$  we have  $\bar{j}^{(i)} = \lambda j^{(i)}$ ,  $\eta(\mu \otimes \mu)$ -a.e. and, if  $R = S = \infty$ ,  $\bar{j}^{(i)\perp} = \lambda j^{(i)\perp}$ ,  $\eta \zeta^{(i)}$ -a.e.

*Proof.* We only consider the case  $R = S = \infty$ . A proof in the case  $R \wedge S < \infty$  can then be obtained by setting the recession terms to zero. First note that we have the simple estimate

$$\frac{\bar{j}(j_+)^{p-1}}{m^{p-1}(r, s)} - \frac{\bar{j}(j_-)^{p-1}}{m^{p-1}(s, r)} \leq \frac{\bar{j}_+(j_+)^{p-1}}{m^{p-1}(r, s)} + \frac{\bar{j}_-(j_-)^{p-1}}{m^{p-1}(s, r)}, \quad (35)$$

with equality if and only if  $\bar{j}$  is nonnegative, where  $j$  is positive and nonpositive, where  $j$  is negative. Recalling that  $p - 1 = p/q$ , we shorten the notation by introducing

$$\begin{aligned} a_1^{(i)}(x, y) &:= \frac{\bar{j}_+^{(i)}(x, y)}{(\mathbf{m}(x, y))^{1/q}}, & b_1^{(i)}(x, y) &:= \frac{(j_+^{(i)}(x, y))^{p/q}}{(\mathbf{m}(x, y))^{p/q^2}}, \\ a_2^{(i)}(x, y) &:= \frac{\bar{j}_-^{(i)}(x, y)}{(\mathbf{m}(y, x))^{1/q}}, & b_2^{(i)}(x, y) &:= \frac{(j_-^{(i)}(x, y))^{p/q}}{(\mathbf{m}(y, x))^{p/q^2}}, \end{aligned}$$

and similarly the recession terms  $a_{k,\infty}^{(i)}$  and  $b_{k,\infty}^{(i)}$  for  $i, k = 1, 2$ . Note that  $a_1^{(i)} a_2^{(i)} = 0$  and similar for all the other terms. We use this fact, (35), and apply Hölder's inequality for sums and integrals, to obtain

$$\begin{aligned} l_\rho(j)[\bar{j}] &\leq \frac{1}{2} \sum_{i=1}^2 \frac{1}{\beta^{(i)}} \left[ \iint_G \sum_{k=1}^2 a_k^{(i)} b_k^{(i)} \eta d(\mu \otimes \mu) + \iint_G \sum_{k=1}^2 a_{k,\infty}^{(i)} b_{k,\infty}^{(i)} \eta d\mathcal{S}^{(i)} \right] \\ &\leq \frac{1}{2} \sum_{i=1}^2 \frac{1}{\beta^{(i)}} \left[ \iint_G \left( \sum_{k=1}^2 (a_k^{(i)})^p \right)^{1/p} \left( \sum_{k=1}^2 (b_k^{(i)})^q \right)^{1/q} \eta d(\mu \otimes \mu) \right. \\ &\quad \left. + \iint_G \left( \sum_{k=1}^2 (a_{k,\infty}^{(i)})^p \right)^{1/p} \left( \sum_{k=1}^2 (b_{k,\infty}^{(i)})^q \right)^{1/q} \eta d\mathcal{S}^{(i)} \right] \\ &\leq \frac{1}{2} \sum_{i=1}^2 \frac{1}{\beta^{(i)}} \left[ \left( \iint_G \sum_{k=1}^2 (a_k^{(i)})^p \eta d(\mu \otimes \mu) \right)^{1/p} \left( \iint_G \sum_{k=1}^2 (b_k^{(i)})^q \eta d(\mu \otimes \mu) \right)^{1/q} \right. \\ &\quad \left. + \left( \iint_G \sum_{k=1}^2 (a_{k,\infty}^{(i)})^p \eta d\mathcal{S}^{(i)} \right)^{1/p} \left( \iint_G \sum_{k=1}^2 (b_{k,\infty}^{(i)})^q \eta d\mathcal{S}^{(i)} \right)^{1/q} \right] \\ &\leq \frac{1}{2} \sum_{i=1}^2 \frac{1}{\beta^{(i)}} \left[ \left( \iint_G \sum_{k=1}^2 (a_k^{(i)})^p \eta d(\mu \otimes \mu) + \iint_G \sum_{k=1}^2 (a_{k,\infty}^{(i)})^p \eta d\mathcal{S}^{(i)} \right)^{1/p} \right. \\ &\quad \left. \left( \iint_G \sum_{k=1}^2 (b_k^{(i)})^q \eta d(\mu \otimes \mu) + \iint_G \sum_{k=1}^2 (b_{k,\infty}^{(i)})^q \eta d\mathcal{S}^{(i)} \right)^{1/q} \right] \\ &\leq (l_\rho(j)[\bar{j}])^{1/p} (l_\rho(j)[j])^{1/q}. \end{aligned} \quad (36)$$

The third inequality in (36) is satisfied with equality if and only if there exist  $\lambda^{(i)}, \lambda_\infty^{(i)} > 0$ ,  $i = 1, 2$ , s.t.  $\sum_{k=1}^2 (a_k^{(i)})^p = \lambda^{(i)} \sum_{k=1}^2 (b_k^{(i)})^q$ ,  $\eta(\mu \otimes \mu)$ -a.e. and  $\sum_{k=1}^2 (a_{k,\infty}^{(i)})^p = \lambda^{(i)} \sum_{k=1}^2 (b_{k,\infty}^{(i)})^q$ ,  $\eta\mathcal{S}^{(i)}$ -a.e. Assuming equality in the third inequality, equality in the first and second inequality is achieved if and only if for  $i, k = 1, 2$  we have  $(a_k^{(i)})^p = \lambda^{(i)} (b_k^{(i)})^q$ ,  $\eta(\mu \otimes \mu)$ -a.e. and  $(a_{k,\infty}^{(i)})^p = \lambda^{(i)} (b_{k,\infty}^{(i)})^q$ ,  $\eta\mathcal{S}^{(i)}$ -a.e. Assuming equality in the first three inequalities, equality in the fourth inequality is obtained if and only if for  $i = 1, 2$  we have  $\lambda^{(i)} = \lambda_\infty^{(i)}$ , while equality in the last inequality is given if and only if  $\lambda^{(1)} = \lambda^{(2)}$  and  $\lambda_\infty^{(1)} = \lambda_\infty^{(2)}$ . Combining these considerations finishes the proof.  $\square$

**Theorem 3.6** (Minkowski norm). *For any  $\rho \in (\mathcal{P}_p(\mathbb{R}^d))^2$  we have that  $F_\rho$  is a Minkowski-norm on  $T_\rho(\mathcal{P}_p(\mathbb{R}^d))^2$ , i.e. it is smooth away from zero and satisfies*

- (i) *Positivity:*  $F_\rho(j) > 0$  for all  $\rho \in (\mathcal{P}_p(\mathbb{R}^d))^2$  and all  $0 \neq j \in T_\rho(\mathcal{P}_p(\mathbb{R}^d))^2$ .
- (ii) *Positive 1-homogeneity:*  $F_\rho(\lambda j) = \lambda F_\rho(j)$  for all  $\lambda > 0$ ,  $\rho \in (\mathcal{P}_p(\mathbb{R}^d))^2$  and  $j \in T_\rho(\mathcal{P}_p(\mathbb{R}^d))^2$ .
- (iii) *Strong convexity:*  $F_\rho(j + \bar{j}) \leq F_\rho(j) + F_\rho(\bar{j})$  for  $\rho \in (\mathcal{P}_p(\mathbb{R}^d))^2$  and  $j, \bar{j} \in T_\rho(\mathcal{P}_p(\mathbb{R}^d))^2$ , with equality if and only if there exists  $C \geq 0$  such that  $j = C\bar{j}$ .

*Proof.* Positivity, positive 1-homogeneity and smoothness, when  $j^{(i)} \neq 0$ ,  $\eta(\mu \otimes \mu)$ -a.e. (and, for  $R = S = \infty$ ,  $j^{(i)\perp} \neq 0$ ,  $\eta\zeta^{(i)}$ -a.e.) are immediate from the definition. Strong convexity is obtained from Lemma 3.5 as follows:

$$\begin{aligned} (F_\rho(j + \bar{j}))^p &= l_\rho(j + \bar{j})[j + \bar{j}] = l_\rho(j + \bar{j})[j] + l_\rho(j + \bar{j})[\bar{j}] \\ &\leq ((l_\rho(j)[j])^{1/p} + (l_\rho(\bar{j})[\bar{j}])^{1/p})(l_\rho(j + \bar{j})[j + \bar{j}])^{1/q} \\ &= (F_\rho(j) + F_\rho(\bar{j}))(F_\rho(j + \bar{j}))^{p-1}. \end{aligned}$$

Dividing by  $F_\rho^{p-1}$  yields the statement.  $\square$

**Remark 3.7.** We use a different notion of Minkowski norm, compared to [35]. There, (i), (ii) and (iii) are replaced by the stronger assumption that the second variation of  $F_\rho(j)$  is a symmetric positive definite bilinear form if  $j$  is nonzero ( $\rho \otimes \mu + \mu \otimes \rho$ )-a.e. However, the notion used here is better suited to the case  $p \neq 2$  and also commonly used, e.g. in [1].

**Proposition 3.8.** Let  $\rho \in (\mathcal{P}_p(\mathbb{R}^d))^2$  and  $j \in T_\rho(\mathcal{P}_p(\mathbb{R}^d))^2$  such that for  $i = 1, 2$  we have density of  $j^{(i)} \neq 0$ ,  $\eta(\mu \otimes \mu)$ -a.e. and  $j^{(i)\perp} \neq 0$ ,  $\eta\zeta^{(i)}$ -a.e., if  $R = S = \infty$ . Then, for any  $\bar{j} \in T_\rho(\mathcal{P}_p(\mathbb{R}^d))^2$ , we have

$$\frac{1}{p} \frac{d}{d\tau} (F_\rho[j + \tau\bar{j}])^p \Big|_{\tau=0} = l_\rho(j)[\bar{j}].$$

*Proof.* Recalling that  $(a + \tau b)^p = \sum_{k=0}^{\infty} \binom{p}{k} a^{p-k} \tau^k b^k = a^p + p a^{p-1} \tau b + \mathcal{O}(\tau^2)$  and that the different species as well as the absolutely continuous and the singular parts can be treated individually, this follows as in [35, Appendix A].  $\square$

**Definition 3.9** (Differential and metric gradient). Given a functional  $\mathcal{F} : (\mathcal{P}(\mathbb{R}^d))^2 \rightarrow \mathbb{R} \cup \{+\infty\}$ , we define its differential at  $\rho \in (\mathcal{P}_p(\mathbb{R}^d))^2$  in direction  $\bar{j} \in T_\rho(\mathcal{P}_p(\mathbb{R}^d))^2$  by

$$\text{diff } \mathcal{F}(\rho)[\bar{j}] := \frac{d}{dt} \mathcal{F}(\tilde{\rho}_t) \Big|_{t=0},$$

where  $\tilde{\rho}_t$  solves  $\frac{d}{dt} \tilde{\rho}_t = -\bar{\nabla} \cdot \bar{j}$  on a small interval according to Definition 2.18 and satisfies  $\tilde{\rho}_0 = \rho$ .

We further define the metric gradient (if it exists) via the equation

$$\text{diff } \mathcal{F}(\rho)[\bar{j}] = l_\rho(\text{grad } \mathcal{F}(\rho))[\bar{j}], \quad \text{for any } \bar{j} \in T_\rho(\mathcal{P}_p(\mathbb{R}^d))^2.$$

**Theorem 3.10** (Uniqueness of the gradient). Given a functional  $\mathcal{F} : (\mathcal{P}(\mathbb{R}^d))^2 \rightarrow \mathbb{R} \cup \{+\infty\}$ , if for  $\rho \in (\mathcal{P}_p(\mathbb{R}^d))^2$  the differential  $\text{diff } \mathcal{F}(\rho)$  exists, then it is unique.

*Proof.* Similar to [35, Subsection 3.1], the uniqueness of the gradient is an immediate consequence of the injectivity of the map  $j \mapsto l_\rho(j)[\bar{j}]$  for given  $\bar{j} \in T_\rho(\mathcal{P}_p(\mathbb{R}^d))^2$ . To show this injectivity, let  $j, \tilde{j} \in T_\rho(\mathcal{P}_p(\mathbb{R}^d))^2$  with  $l_\rho(j) = l_\rho(\tilde{j})$ . If either  $j^{(i)} = 0$  or  $\tilde{j}^{(i)} = 0$ ,  $\eta(\mu \otimes \mu)$ -a.e., then  $l_\rho(j) = l_\rho(\tilde{j})$  implies  $j^{(i)} = \tilde{j}^{(i)} = 0$ ,  $\eta(\mu \otimes \mu)$ -a.e. Similarly, if either  $j^{(i)\perp} = 0$  or  $\tilde{j}^{(i)\perp} = 0$ ,  $\eta\zeta^{(i)}$ -a.e., then  $l_\rho(j) = l_\rho(\tilde{j})$  implies  $j^{(i)\perp} = \tilde{j}^{(i)\perp} = 0$ ,  $\eta\zeta^{(i)}$ -a.e. Now, for at least one  $i \in \{1, 2\}$  let  $j^{(i)}(A) \neq 0$  for some  $A \subset G$  with  $\mu \otimes \mu(A) > 0$  or  $j^{(i)\perp}(B) \neq 0$  for some  $B \subset G$  with  $\zeta^{(i)}(B) > 0$ , and let  $\tilde{j}^{(i)}(\tilde{A}) \neq 0$  for some  $\tilde{A} \subset G$  with  $\mu \otimes \mu(\tilde{A}) > 0$  or  $\tilde{j}^{(i)\perp}(\tilde{B}) \neq 0$  for some  $\tilde{B} \subset G$  with  $\zeta^{(i)}(\tilde{B}) > 0$ . Then, by (34) we obtain

$$0 < l_\rho(j)[\bar{j}] = l_\rho(\tilde{j})[\bar{j}] \leq (l_\rho(\tilde{j})[\bar{j}])^{1/p} (l_\rho(j)[\bar{j}])^{1/q},$$

which gives us  $l_\rho(j)[\bar{j}] \leq l_\rho(\tilde{j})[\bar{j}]$ . Inverting the roles of  $j$  and  $\tilde{j}$ , we obtain  $l_\rho(\tilde{j})[\bar{j}] \leq l_\rho(j)[\bar{j}]$ , i.e. we have  $l_\rho(j)[\bar{j}] = l_\rho(\tilde{j})[\bar{j}]$ . This implies equality in the Hölder-type inequality (34), thus yielding  $j^{(i)} = C\tilde{j}^{(i)}$ ,  $\eta(\mu \otimes \mu)$ -a.e. and  $j^{(i)\perp} = C\tilde{j}^{(i)\perp}$ ,  $\eta\zeta^{(i)}$ -a.e. for some  $C \geq 0$  and both  $i = 1, 2$ . Using the positive 1-homogeneity of  $l_\rho$  we obtain  $l_\rho(j) = l_\rho(C\tilde{j}) = Cl_\rho(\tilde{j}) = Cl_\rho(j)$ , which yields  $C = 1$  since  $l_\rho(j)[\bar{j}] \neq 0$ . This proves the claimed injectivity.  $\square$

Since the map  $j \mapsto l_\rho(j)[\bar{j}]$  is not antisymmetric, i.e.  $l_\rho(j)[\bar{j}] \neq -l_\rho(-j)[\bar{j}]$ , we separately need to define the negative gradient:

**Definition 3.11** (Negative metric gradient). *Given  $\mathcal{F} : (\mathcal{P}(\mathbb{R}^d))^2 \rightarrow \mathbb{R}$  and  $\rho \in (\mathcal{P}_p(\mathbb{R}^d))^2$ , we define the negative metric gradient of  $\mathcal{F}$  at  $\rho$  by*

$$l_\rho(\text{grad}^- \mathcal{F}(\rho))[\bar{j}] := -\text{diff } \mathcal{F}(\rho)[\bar{j}], \quad \text{for all } \bar{j} \in T_\rho(\mathcal{P}_p(\mathbb{R}^d))^2.$$

Since  $l_\rho(\cdot)$  is not antisymmetric, in general  $\text{grad}^- \mathcal{F}(\rho) \neq -\text{grad } \mathcal{F}(\rho)$ . To define the (unique) direction of steepest descent at  $\rho$  we use the following criterion, as in the Riemannian case:

**Definition 3.12.** *Given  $\rho$  and  $\mathcal{F}$  such that  $\text{diff } \mathcal{F}(\rho) \neq 0$ , we define the direction of steepest descent as*

$$j^* := \arg \min \{ \text{diff } \mathcal{F}(\rho)[\bar{j}] \mid \bar{j} \in T_\rho(\mathcal{P}_p(\mathbb{R}^d))^2, \text{ s.t. } l_\rho(\bar{j})[\bar{j}] = 1 \}, \quad (37)$$

*if it exists.*

Note that  $\text{diff } \mathcal{F}(\rho) = 0$  implies  $\text{grad}^- \mathcal{F}(\rho) = 0$ . Otherwise, the negative metric gradient determines the direction of steepest descent, as the next lemma shows.

**Lemma 3.13.** *Let  $j^*$  be as in (37). Then, there exists  $C > 0$  such that  $j^* = C \text{grad}^- \mathcal{F}(\rho)$  holds.*

*Proof.* We argue similar to [35, Subsection 3.1] and start by adding the constraint of the optimization problem with Lagrange multiplier  $C \in \mathbb{R}$ . For  $j \in T_\rho(\mathcal{P}_p(\mathbb{R}^d))^2$  we define the functional

$$\mathcal{H}(C, j) := \text{diff } \mathcal{F}(\rho)[j] + \frac{C}{p}(l_\rho(j)[j] - 1).$$

We employ that, by Proposition 3.8, for any  $j, \bar{j} \in T_\rho(\mathcal{P}_p(\mathbb{R}^d))^2$ , we have the equality

$$\left. \frac{d}{d\tau} l_\rho(j + \tau \bar{j})[j + \tau \bar{j}] \right|_{\tau=0} = p l_\rho(j)[\bar{j}].$$

Then, using the linearity of the differential  $\text{diff } \mathcal{F}(\rho)$ , any minimizer  $(C^*, j^*)$  of  $\mathcal{H}$  must satisfy the condition

$$\text{diff } \mathcal{F}(\rho)[\cdot] = -C^* l_\rho(j^*)[\cdot].$$

By the linearity of the map  $j \mapsto -C^* l_\rho(j^*)[j]$ , and the symmetry of the constraint, we find  $0 > \text{diff } \mathcal{F}(\rho)[j^*] = -C^* l_\rho(j^*)[j^*]$ , which implies  $C^* > 0$ . Thus, the previously proven injectivity and positive 1-homogeneity of  $l_\rho$  yield

$$j^* = l_\rho^{-1} \left( -\frac{1}{C^*} \text{diff } \mathcal{F}(\rho) \right) = \frac{1}{C^*} l_\rho^{-1} (-\text{diff } \mathcal{F}(\rho)) = \frac{1}{C^*} \text{grad}^- \mathcal{F}(\rho).$$

□

In light of Lemma 3.13, it makes sense to write metric gradient flows with respect to  $\mathcal{F}$  in the Finsler space  $((\mathcal{P}_p(\mathbb{R}^d))^2, \mathcal{T}_\beta)$  as

$$\partial_t \rho_t = \bar{\nabla} \cdot \text{grad}^- \mathcal{F}(\rho_t).$$

Since the previous considerations did not use any specific structure of  $\mathcal{F}$ , they stay valid for general functionals  $\mathcal{F} : (\mathcal{P}_p(\mathbb{R}^d))^2 \rightarrow \mathbb{R} \cup \{+\infty\}$ . However, even though by Theorem 3.10 we know that the (negative) metric gradient of a functional is unique, we have yet to show its existence. For the case where  $\mathcal{F}$  is the nonlocal cross-interaction energy (2), the following theorem ensures existence.

**Theorem 3.14** (Existence of the negative metric gradient for the nonlocal cross-interaction energy). *Let  $\mathcal{E}$  be the nonlocal cross-interaction energy. Then, for any  $\rho \in (\mathcal{P}_p(\mathbb{R}^d))^2$  and  $\eta(\mu \otimes \mu)$ -a.e. the negative metric gradient  $d \text{grad}^- \mathcal{E}(\rho) = \sum_{i=1}^2 (\text{grad}^- \mathcal{E}(\rho))^{(i)} d(\mu \otimes \mu) + (\text{grad}^- \mathcal{E}(\rho))^{(i)\perp} d\zeta^{(i)}$  is given for  $i = 1, 2$  by*

$$\begin{aligned} (\text{grad}^- \mathcal{E}(\rho))^{(i)} &= (\mathbf{m}^{(i)})^\top ((-\beta^{(i)} \bar{\nabla} \delta_{\rho^{(i)}} \mathcal{E}(\rho))_-)^{q-1} - \mathbf{m}^{(i)} ((-\beta^{(i)} \bar{\nabla} \delta_{\rho^{(i)}} \mathcal{E}(\rho))_+)^{q-1}, \\ (\text{grad}^- \mathcal{E}(\rho))^{(i)\perp} &= (\mathbf{m}_\infty^{(i)})^\top ((-\beta^{(i)} \bar{\nabla} \delta_{\rho^{(i)}} \mathcal{E}(\rho))_-)^{q-1} - \mathbf{m}_\infty^{(i)} ((-\beta^{(i)} \bar{\nabla} \delta_{\rho^{(i)}} \mathcal{E}(\rho))_+)^{q-1}, \end{aligned} \quad (38)$$

if  $R = S = \infty$ . If  $R \wedge S < \infty$ , then we have  $(\text{grad}^- \mathcal{E}(\rho))^{(i)\perp} = 0$  for  $i = 1, 2$ .

*Proof.* We calculate the differential according to Definition 3.9. For  $\rho \in (\mathcal{P}_p(\mathbb{R}^d))^2$  and  $j \in T_\rho(\mathcal{P}_p(\mathbb{R}^d))^2$  take any curve  $\tilde{\rho} \in \text{AC}^p([0, T]; ((\mathcal{P}_p(\mathbb{R}^d))^2, \mathcal{T}_{m, \beta, \mu}))$ , s.t.  $\tilde{\rho}_0 = \rho$  and  $\frac{d}{dt} \tilde{\rho}_t = -\bar{\nabla} \cdot \bar{j}_t$  according to Definition 2.18 and  $\bar{j}_0 = j$ . Then, using the equality  $K^{(21)} = K^{(12)}$  and Lemma 2.20, we find

$$\begin{aligned} -\text{diff } \mathcal{E}(\rho)[j] &= -\frac{d}{dt} \mathcal{E}(\tilde{\rho}_t) \Big|_{t=0} = -\lim_{\tau \rightarrow 0} \frac{\mathcal{E}(\tilde{\rho}_\tau) - \mathcal{E}(\tilde{\rho}_0)}{\tau} \\ &= -\lim_{\tau \rightarrow 0} \frac{1}{2\tau} \sum_{i,k=1}^2 \left[ \int_{\mathbb{R}^d} (K^{(ik)} * \tilde{\rho}_\tau^{(k)})(x) d\tilde{\rho}_\tau^{(i)}(x) - \int_{\mathbb{R}^d} (K^{(ik)} * \tilde{\rho}_0^{(k)})(x) d\tilde{\rho}_0^{(i)}(x) \right] \\ &= -\lim_{\tau \rightarrow 0} \frac{1}{2\tau} \sum_{i,k=1}^2 \left[ \int_0^\tau \frac{d}{dt} \iint_{\mathbb{R}^d \times \mathbb{R}^d} K^{(ik)}(x, y) d\tilde{\rho}_t^{(k)}(y) d\tilde{\rho}_t^{(i)}(x) dt \right] \\ &= -\frac{1}{2} \sum_{i,k=1}^2 \iint_G \bar{\nabla} (K^{(ik)} * \rho^{(k)})(x, y) \eta(x, y) dj^{(i)}(x, y) \\ &= \frac{1}{2} \sum_{i,k=1}^2 \iint_G -\bar{\nabla} (K^{(ik)} * \rho^{(k)})(x, y) \eta(x, y) dj^{(i)}(x, y). \end{aligned}$$

Since  $\delta_{\rho^{(1)}} \mathcal{E}(\rho) = K^{(11)} * \rho^{(1)} + K^{(12)} * \rho^{(2)}$  and  $\delta_{\rho^{(2)}} \mathcal{E}(\rho) = K^{(22)} * \rho^{(2)} + K^{(21)} * \rho^{(1)}$ , and  $\beta^{(i)} > 0$  for  $i = 1, 2$ , we rewrite this in terms of the negative metric gradient of the energy functional:

$$\begin{aligned} -\text{diff } \mathcal{E}(\rho)[j] &= \frac{1}{2} \sum_{i=1}^2 \left[ \iint_G -\bar{\nabla} \delta_{\rho^{(i)}} \mathcal{E}(\rho)(x, y) \eta(x, y) j^{(i)}(x, y) d\mu(x) d\mu(y) \right. \\ &\quad \left. + \iint_G -\bar{\nabla} \delta_{\rho^{(i)}} \mathcal{E}(\rho)(x, y) \eta(x, y) j^{(i)\perp}(x, y) d\zeta^{(i)}(x, y) \right] \\ &= \frac{1}{2} \sum_{i=1}^2 \frac{1}{\beta^{(i)}} \iint_G j^{(i)}(x, y) \left( \frac{\mathbf{m}^{(i)}(x, y) (\beta^{(i)} (-\bar{\nabla} \delta_{\rho^{(i)}} \mathcal{E}(\rho))_+(x, y))^{q-1}}{\mathbf{m}^{(i)}(x, y)} \right. \\ &\quad \left. - \frac{\mathbf{m}^{(i)}(y, x) (\beta^{(i)} (-\bar{\nabla} \delta_{\rho^{(i)}} \mathcal{E}(\rho))_-(x, y))^{q-1}}{\mathbf{m}^{(i)}(y, x)} \right)^{p-1} \eta(x, y) d\mu(x) d\mu(y) \\ &\quad + \iint_G j^{(i)\perp}(x, y) \left( \frac{\mathbf{m}_\infty^{(i)}(x, y) (\beta^{(i)} (-\bar{\nabla} \delta_{\rho^{(i)}} \mathcal{E}(\rho))_+(x, y))^{q-1}}{\mathbf{m}_\infty^{(i)}(x, y)} \right. \\ &\quad \left. - \frac{\mathbf{m}_\infty^{(i)}(y, x) (\beta^{(i)} (-\bar{\nabla} \delta_{\rho^{(i)}} \mathcal{E}(\rho))_-(x, y))^{q-1}}{\mathbf{m}_\infty^{(i)}(y, x)} \right)^{p-1} \eta(x, y) d\zeta^{(i)}(x, y). \end{aligned}$$

Comparing this expression with the definition of the gradient, (38) follows.  $\square$

**Remark 3.15.** For  $R \wedge S < \infty$  or  $m_\infty \equiv 0$  the structure of the negative gradient closely resembles the structure in (31) with  $\varphi^{(i)} = -\beta^{(i)} \delta_{\rho^{(i)}} \mathcal{E}(\rho)$  since  $-\beta^{(i)} \bar{\nabla} \delta_{\rho^{(i)}} \mathcal{E}(\rho) = -\bar{\nabla} \beta^{(i)} \delta_{\rho^{(i)}} \mathcal{E}(\rho)$ .

### 3.2. Variational characterization for the nonlocal nonlocal cross-interaction equation

We now want to characterize (32) as a gradient flow in the sense of curves of maximal slope and start by defining the one-sided strong upper gradient.

**Definition 3.16.** (*One-sided strong upper gradient*). A function  $h : (\mathcal{P}_p(\mathbb{R}^d))^2 \rightarrow [0, \infty]$  is called a one-sided strong upper gradient for  $\mathcal{E}$  if for every  $\rho \in \text{AC}^p([0, T]; ((\mathcal{P}_p(\mathbb{R}^d))^2, \mathcal{T}_{m,\beta,\mu}))$  the function  $h \circ \rho : [0, T] \rightarrow [0, \infty]$  is measurable and we have

$$\mathcal{E}(\rho_t) - \mathcal{E}(\rho_s) \geq - \int_s^t h(\rho_\tau) |\rho'_\tau| d\tau, \quad \text{for all } 0 \leq s \leq t \leq T.$$

As before  $|\rho'_\tau|$  denotes the metric derivative of  $\rho_\tau$  with respect to  $\mathcal{T}_{m,\beta,\mu}$ .

The one-sided strong upper gradient is sufficient to characterize curves of maximal slope:

**Definition 3.17.** (*Curve of maximal slope*). Given a strong one-sided upper gradient  $h$  for  $\mathcal{E}$ , a curve  $\rho \in \text{AC}^p([0, T]; ((\mathcal{P}_p(\mathbb{R}^d))^2, \mathcal{T}_{m,\beta,\mu}))$  is called a curve of maximal slope for  $\mathcal{E}$  with respect to  $h$  if and only if

$$\mathcal{E}(\rho_t) - \mathcal{E}(\rho_s) + \int_s^t \frac{1}{q} (h(\rho_\tau))^q + \frac{1}{p} |\rho'_\tau|^p d\tau \leq 0, \quad \text{for all } 0 \leq s \leq t \leq T. \quad (39)$$

**Remark 3.18.** Note that inequality (39) implies that  $t \mapsto \mathcal{E}(\rho_t)$  is nonincreasing. Further, observe that by Young's inequality we immediately see that any strong one-sided upper gradient for  $\mathcal{E}$  satisfies

$$\mathcal{E}(\rho_t) - \mathcal{E}(\rho_s) + \int_s^t \frac{1}{q} (h(\rho_\tau))^q + \frac{1}{p} |\rho'_\tau|^p d\tau \geq 0, \quad \text{for all } 0 \leq s \leq t \leq T,$$

i.e., if  $\rho$  is a curve of maximal slope for  $\mathcal{E}$  with respect to its strong one-sided upper gradient  $h$ , then we have equality in (39).

Our next goal is to derive a chain rule. However, we have seen in Theorem 3.14, the relation between  $(\text{grad}^- \mathcal{E}(\rho))^{(i)}$  and  $-\beta^{(i)} \bar{\nabla} \delta_{\rho^{(i)}} \mathcal{E}(\rho)$  is not linear, but contains a power  $q - 1$ . To account for this, we make the following definition:

**Definition 3.19.** Given two maps  $\mathbf{v} = (v^{(1)}, v^{(1)\perp}, v^{(2)}, v^{(2)\perp}), \bar{\mathbf{v}} = (\bar{v}^{(1)}, \bar{v}^{(1)\perp}, \bar{v}^{(2)}, \bar{v}^{(2)\perp}) : G \rightarrow \mathbb{R}^4$ , we define

$$\begin{aligned} \tilde{l}_\rho(\mathbf{v})[\bar{\mathbf{v}}] = & \frac{1}{2} \sum_{i=1}^2 \frac{1}{\beta^{(i)}} \left[ \iint_G \bar{v}^{(i)} \left( \mathbf{m}^{(i)}(v_+^{(i)}(\rho))^{q-1} - (\mathbf{m}^{(i)})^\top (v_-^{(i)})^{q-1} \right) \eta d(\mu \otimes \mu) \right. \\ & \left. + \iint_G \bar{v}^{(i)\perp} \left( \mathbf{m}_\infty^{(i)}(v_+^{(i)\perp})^{q-1} - (\mathbf{m}_\infty^{(i)})^\top (v_-^{(i)\perp})^{q-1} \right) \eta d\mathcal{S}^{(i)} \right], \end{aligned}$$

if  $R = S = \infty$ . For  $R \wedge S < \infty$ , we define  $\tilde{l}$  by

$$\tilde{l}_\rho(\mathbf{v})[\bar{\mathbf{v}}] = \frac{1}{2} \sum_{i=1}^2 \frac{1}{\beta^{(i)}} \iint_G \bar{v}^{(i)} \left( \mathbf{m}^{(i)}(v_+^{(i)})^{q-1} - (\mathbf{m}^{(i)})^\top (v_-^{(i)})^{q-1} \right) \eta d(\mu \otimes \mu).$$

**Remark 3.20.** Let  $\mathbf{v}, \bar{\mathbf{v}}$  be associated to  $\mathbf{j}, \bar{\mathbf{j}}$  as in (15) and (16). Then, in general  $l_\rho(\mathbf{j})[\bar{\mathbf{j}}] \neq \tilde{l}_\rho(\mathbf{v})[\bar{\mathbf{v}}]$ . However,  $\tilde{l}_\rho(\mathbf{v})[\bar{\mathbf{v}}]$  is linear in  $\bar{\mathbf{v}}$  and it still holds that

$$\tilde{l}_\rho(\mathbf{v})[\bar{\mathbf{v}}] = \tilde{\mathcal{A}}_{m,\beta}(\mu; \rho, \mathbf{v}) = \mathcal{A}_{m,\beta}(\mu; \rho, \mathbf{j}) = l_\rho(\mathbf{j})[\bar{\mathbf{j}}]. \quad (40)$$

With this machinery in place, we can now adapt (25) as follows:

**Lemma 3.21** (Chain rule for test functions). *For  $\boldsymbol{\rho} \in \text{AC}^p([0, T]; ((\mathcal{P}_p(\mathbb{R}^d))^2, \mathcal{T}_{m, \beta, \mu}))$  let  $\mathbf{j} \subset T_\rho(\mathcal{P}_p(\mathbb{R}^d))^2$  such that  $(\boldsymbol{\rho}, \mathbf{j}) \in \text{CE}_T$  and  $|\rho'_t|^p = \mathcal{A}_{m, \beta}(\mu; \rho_t; \mathbf{j}_t)$  for a.e.  $t \in [0, T]$ , as given in Proposition 2.31. For this  $\mathbf{j}$ , let  $\mathbf{v} = (v_t)_t, v_t = (v_t^{(1)}, v_t^{(1)\perp}, v_t^{(2)}, v_t^{(2)\perp}) : G \rightarrow \mathbb{R}^4$  be as in (15) and (16). Then, for any  $\varphi = (\varphi^{(1)}, \varphi^{(1)\perp}, \varphi^{(2)}, \varphi^{(2)\perp}) \in (C_c^\infty(\mathbb{R}^d))^4$ ,  $i = 1, 2$ ,  $0 \leq s \leq t \leq T$  and  $i = 1, 2$  it holds*

$$\sum_{i=1}^2 \int_{\mathbb{R}^d} \varphi^{(i)}(x) d\rho_t^{(i)}(x) - \int_{\mathbb{R}^d} \varphi^{(i)}(x) d\rho_s^{(i)}(x) = \int_s^t \tilde{l}_{\rho_\tau}(v_\tau) [\beta \bar{\nabla} \varphi] d\tau.$$

*Proof.* Let  $i \in \{1, 2\}$ . Starting from the continuity equation (23) we calculate

$$\begin{aligned} & \int_{\mathbb{R}^d} \varphi^{(i)}(x) d\rho_t^{(i)}(x) - \int_{\mathbb{R}^d} \varphi^{(i)}(x) d\rho_s^{(i)}(x) = \frac{1}{2} \int_s^t \iint_G \bar{\nabla} \varphi(x, y) \eta(x, y) dj_\tau^{(i)}(x, y) d\tau \\ & = \frac{1}{2\beta^{(i)}} \int_s^t \iint_G \beta^{(i)} \bar{\nabla} \varphi^{(i)}(x, y) \eta(x, y) \left( ((v_\tau^{(i)})_+(x, y))^{q-1} d\gamma_{1, \tau}^{(i)}(x, y) - ((v_\tau^{(i)})_-(x, y))^{q-1} d\gamma_{2, \tau}^{(i)}(x, y) \right) d\tau. \end{aligned}$$

From this, we conclude by summing over both species.  $\square$

As for  $l$ , for  $\tilde{l}$  we too have a Hölder-type inequality:

**Lemma 3.22** (Hölder-type inequality). *For all  $v, \bar{v} \in \tilde{T}_\rho(\mathcal{P}_p(\mathbb{R}^d))^2$  we have*

$$\tilde{l}_\rho(v) [\bar{v}] \leq (\tilde{l}_\rho(v) [v])^{1/p} (\tilde{l}_\rho(\bar{v}) [\bar{v}])^{1/q}, \quad (41)$$

*with equality if and only if, for some  $\lambda > 0$ , for  $i = 1, 2$  we have  $v_+^{(i)} = \lambda \bar{v}_+^{(i)}$ ,  $\eta \gamma_1^{(i)}$ -a.e. as well as  $v_+^{(i)\perp} = \lambda \bar{v}_+^{(i)\perp}$ ,  $\eta \gamma_1^{(i)\perp}$ -a.e. and hence, by antisymmetry, also  $v_-^{(i)} = \lambda \bar{v}_-^{(i)}$ ,  $\eta \gamma_2^{(i)}$ -a.e. as well as  $v_-^{(i)\perp} = \lambda \bar{v}_-^{(i)\perp}$ ,  $\eta \gamma_2^{(i)\perp}$ -a.e.*

*Proof.* The argument is analogous to that for  $l_\rho(\mathbf{j})[\mathbf{j}]$  in the proof of Lemma 3.5.  $\square$

**Definition 3.23.** (Dissipation and De Giorgi functional). *For  $\rho \in (\mathcal{P}_p(\mathbb{R}^d))^2$ , we define the dissipation at  $\rho$  by*

$$\mathcal{D}(\rho) := \tilde{\mathcal{A}}_{m, \beta}(\mu; \rho, -\beta \bar{\nabla} \delta_\rho \mathcal{E}(\rho)) = \tilde{l}_\rho(-\beta \bar{\nabla} \delta_\rho \mathcal{E}(\rho) [-\beta \bar{\nabla} \delta_\rho \mathcal{E}(\rho)]),$$

*where  $\delta_\rho \mathcal{E}(\rho) := (\delta_{\rho^{(1)}} \mathcal{E}(\rho), \delta_{\rho^{(2)}} \mathcal{E}(\rho), \delta_{\rho^{(1)}} \mathcal{E}(\rho), \delta_{\rho^{(2)}} \mathcal{E}(\rho))$ . For any  $\boldsymbol{\rho} \in \text{AC}^p([0, T]; ((\mathcal{P}_p(\mathbb{R}^d))^2, \mathcal{T}_{m, \beta, \mu}))$ , we define the De Giorgi functional at  $\boldsymbol{\rho}$  by*

$$\mathcal{G}_T(\boldsymbol{\rho}) := \mathcal{E}(\rho_T) - \mathcal{E}(\rho_0) + \int_0^T \frac{1}{q} \mathcal{D}(\rho_t) + \frac{1}{p} |\rho'_t|^p dt.$$

*When the dependence on the base measure needs to be made explicit, we write  $\mathcal{D}(\mu; \boldsymbol{\rho})$  and  $\mathcal{G}_T(\mu; \boldsymbol{\rho})$ .*

### 3.3. Characterization of weak solutions

In this subsection we show that weak solutions of (32) can be characterized as minimizers for the De Giorgi functional  $\mathcal{G}_T$  introduced in Definition 3.23. To achieve this, we need the chain rule for the gradient velocity of  $\mathcal{E}$ . Its proof is based on a mollification and truncation argument (and can be found in Appendix A).

**Proposition 3.24** (Chain rule for  $\mathcal{E}$ ). *Let  $K^{(ik)}$ ,  $i, k = 1, 2$  satisfy (K1), (K2) and  $K^{(21)} = K^{(12)}$ , let  $\boldsymbol{\rho} \subset \text{AC}^p([0, T]; ((\mathcal{P}_p(\mathbb{R}^d))^2, \mathcal{T}_{m, \beta, \mu}))$  and let  $0 \leq s \leq t \leq T$ . Denote by  $\mathbf{v}$  the unique velocity in  $\tilde{T}_\rho(\mathcal{P}_p(\mathbb{R}^d))^2$ , which is associated to  $\boldsymbol{\rho}$  by Proposition 2.31 and Lemma 2.10. Then, we have the chain rule identity*

$$\mathcal{E}(\rho_t) - \mathcal{E}(\rho_s) = \int_s^t \tilde{l}_{\rho_\tau}(v_\tau) [\beta \bar{\nabla} \delta_\rho \mathcal{E}(\rho_\tau)] d\tau. \quad (42)$$

Using the chain rule, we infer that  $\mathcal{D}^{1/q}$  is a one-sided strong upper gradient for  $\mathcal{E}$ .

**Corollary 3.25.** *For any curve  $\rho \in AC^p([0, T]; ((\mathcal{P}_p(\mathbb{R}^d))^2, \mathcal{T}_{m,\beta,\mu}))$  and any  $0 \leq s \leq t \leq T$  we have*

$$\mathcal{E}(\rho_t) - \mathcal{E}(\rho_s) \geq - \int_s^t (\mathcal{D}(\rho_\tau))^{1/q} |\rho'_\tau| d\tau,$$

i.e.,  $\mathcal{D}^{1/q}$  is a one-sided strong upper gradient for  $\mathcal{E}$  in the sense of Definition 3.16.

*Proof.* Without loss of generality, assume that  $\int_s^t (\mathcal{D}(\rho_\tau))^{1/q} |\rho'_\tau| d\tau < \infty$  as otherwise there is nothing to show. We employ (42) from Proposition 3.24 and apply the Hölder-type inequality from Lemma 3.22. For  $0 \leq s \leq t \leq T$ , we have

$$\begin{aligned} \mathcal{E}(\rho_t) - \mathcal{E}(\rho_s) &= \int_s^t \tilde{l}_{\rho_\tau}(\mathbf{v}_\tau) [\beta \bar{\nabla} \delta_\rho \mathcal{E}(\rho_\tau)] d\tau = - \int_s^t \tilde{l}_{\rho_\tau}(\mathbf{v}_\tau) [-\beta \bar{\nabla} \delta_\rho \mathcal{E}(\rho_\tau)] d\tau \\ &\geq - \int_s^t (\tilde{l}_{\rho_\tau}(\mathbf{v}_\tau) [\mathbf{v}_\tau])^{1/p} (\tilde{l}_{\rho_\tau}(-\beta \bar{\nabla} \delta_\rho \mathcal{E}(\rho_\tau)) [-\beta \bar{\nabla} \delta_\rho \mathcal{E}(\rho_\tau)])^{1/q} d\tau \\ &= - \int_s^t (\tilde{\mathcal{A}}_{m,\beta}(\rho_\tau, \mathbf{v}_\tau))^{1/p} (\mathcal{D}(\rho_\tau))^{1/q} d\tau = - \int_s^t |\rho'_\tau| (\mathcal{D}(\rho_\tau))^{1/q} d\tau. \end{aligned}$$

Here, the last two inequalities are provided by (40) and Proposition 2.31.  $\square$

Now we are ready to identify weak solutions to (32) as minimizers of  $\mathcal{G}_T$ .

**Theorem 3.26** (Characterization of weak solutions to the nonlocal nonlocal cross-interaction system). *Suppose  $\mu$  satisfies (MB1), (MB2) and (BC), and the kernels  $K^{(ik)}$  satisfy (K1), (K2) for  $i, k = 1, 2$  as well as  $K^{(21)} = K^{(12)}$ . A curve  $\rho : [0, T] \rightarrow (\mathcal{P}_p(\mathbb{R}^d))^2$  is a weak solution of (32) according to Definition 3.1 if and only if  $\rho \in AC^p([0, T]; ((\mathcal{P}_p(\mathbb{R}^d))^2, \mathcal{T}_{m,\beta,\mu}))$  is a curve of maximal slope for  $\mathcal{E}$  with respect to  $(\mathcal{D}(\rho))^{1/q}$  in the sense of Definition 3.17, i.e., it satisfies*

$$\mathcal{G}_T(\rho) = 0, \tag{43}$$

where  $\mathcal{G}_T$  is the De Giorgi functional given in Definition 3.23.

*Proof.* Assume that  $\rho$  is a weak solution to (32) according to Definition 3.1. To construct a weak solution for the continuity equation (23), we define the flux  $\mathbf{j}$  by

$$\begin{aligned} dj_t^{(i)\mu} &= (\beta^{(i)} (\bar{\nabla} \delta_{\rho^{(i)}} \mathcal{E}(\rho))_-)^{q-1} d\gamma_{1,t}^{(i)} - (\beta^{(i)} (\bar{\nabla} \delta_{\rho^{(i)}} \mathcal{E}(\rho))_+)^{q-1} d\gamma_{2,t}^{(i)}, \\ dj_t^{(i)\perp} &= (\beta^{(i)} (\bar{\nabla} \delta_{\rho^{(i)}} \mathcal{E}(\rho))_-)^{q-1} d\gamma_{1,t}^{(i)\perp} - (\beta^{(i)} (\bar{\nabla} \delta_{\rho^{(i)}} \mathcal{E}(\rho))_+)^{q-1} d\gamma_{2,t}^{(i)\perp}, \end{aligned}$$

for  $i = 1, 2$ . Using the abbreviation  $v_t^{\mathcal{E},(i)} := \beta^{(i)} \bar{\nabla} \delta_{\rho^{(i)}} \mathcal{E}(\rho)$ , we immediately obtain

$$\int_0^T \mathcal{A}_{m,\beta}(\mu; \rho_t, \mathbf{j}_t) dt = \int_0^T \tilde{\mathcal{A}}_{m,\beta}(\mu; \rho_t, v_t^{\mathcal{E}}) dt = \int_0^T \mathcal{D}(\rho_t) dt < \infty.$$

The first two equalities are clear from the definitions. For the finiteness, recall that due to the concavity and finiteness of the mobility  $m$ , for any  $B \in \mathcal{B}(G)$  we have the bound

$$(\gamma_{1,\tau}^{(i)} + \gamma_{1,\tau}^{(i)\perp})(B) \leq M(\mu \otimes \mu + \rho^{(i)} \otimes \mu + \mu \otimes \rho^{(i)})(B),$$

where  $M$  only depends on  $m$  and  $G$ . With this, Jensen's inequality, (K2), (MB1) and (MB2), we obtain

$$\begin{aligned}
\mathcal{D}(\rho_t) &= \sum_{i=1}^2 (\beta^{(i)})^{q-1} \left[ \iint_G \left( \left( \sum_{k=1}^2 \bar{\nabla}(K^{(ik)} * \rho_t^{(k)}) \right) \right)_-^q \eta d\gamma_{1,t}^{(i)} + \iint_G \left( \left( \sum_{k=1}^2 \bar{\nabla}(K^{(ik)} * \rho_t^{(k)}) \right) \right)_-^q \eta d\gamma_{1,t}^{(i)\perp} \right] \\
&\leq ML_K^q \sum_{i,k=1}^2 (\beta^{(i)})^{q-1} \int_{\mathbb{R}^d} \iint_G (|x-y|^q \vee |x-y|^{pq}) \eta(x,y) d\mu(y) d\rho_t^{(k)}(x) d(\mu + 2\rho_t^{(i)})(z) \\
&\leq ML_K^q C_\eta \sum_{i,k=1}^2 (\beta^{(i)})^{q-1} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} d\rho_t^{(k)}(x) d(\mu + 2\rho_t^{(i)})(z) \\
&= 2(C_\mu + 2)((\beta^{(1)})^{q-1} + (\beta^{(2)})^{q-1}) ML_K^q C_\eta < \infty.
\end{aligned}$$

By Proposition 2.31, this also proves that  $\boldsymbol{\rho} \in \text{AC}^p([0, T]; ((\mathcal{P}_p(\mathbb{R}^d))^2, \mathcal{T}_\beta))$  and that  $|\rho_t'|^p \leq \mathcal{D}(\rho_t)$  for a.e.  $t \in [0, T]$ . The latter together with Proposition 3.24 yields

$$\begin{aligned}
\mathcal{E}(\rho_t) - \mathcal{E}(\rho_s) &= \int_s^t \tilde{l}_{\rho_\tau}(v_\tau^\mathcal{E}) [\beta \bar{\nabla} \delta_\rho \mathcal{E}(\rho_\tau)] d\tau = - \int_s^t \tilde{l}_{\rho_\tau}(v_\tau^\mathcal{E}) [-\beta \bar{\nabla} \delta_\rho \mathcal{E}(\rho_\tau)] d\tau \\
&= - \int_s^t \tilde{l}_{\rho_\tau}(-\beta \bar{\nabla} \delta_\rho \mathcal{E}(\rho_\tau)) [-\beta \bar{\nabla} \delta_\rho \mathcal{E}(\rho_\tau)] d\tau \\
&= - \int_s^t \mathcal{D}(\rho_\tau) d\tau \leq - \int_s^t \frac{1}{q} \mathcal{D}(\rho_\tau) + \frac{1}{p} |\rho_\tau'|^p d\tau.
\end{aligned}$$

Hence, Corollary 3.25 in conjunction with Remark 3.18 yields

$$\mathcal{E}(\rho_t) - \mathcal{E}(\rho_s) + \int_s^t \frac{1}{q} \mathcal{D}(\rho_\tau) + \frac{1}{p} |\rho_\tau'|^p d\tau = 0.$$

Therefore, the first implication of the theorem follows for the choices  $s = 0$  and  $t = T$  implying  $\mathcal{G}_T(\boldsymbol{\rho}) = 0$ .

To prove the converse implication, now consider  $\boldsymbol{\rho} \in \text{AC}^p([0, T]; ((\mathcal{P}_p(\mathbb{R}^d))^2, \mathcal{T}_\beta))$  satisfying (43). We verify that  $\boldsymbol{\rho}$  is a weak solution of (32) according to Definition 3.1. By Proposition 2.31, there exists a unique family  $\mathbf{j} \subset T_\rho((\mathcal{P}(\mathbb{R}^d))^2)$ , such that  $(\boldsymbol{\rho}, \mathbf{j}) \in \text{CE}_T$ ,  $\int_0^T \mathcal{A}_{m,\beta}^{1/p}(\rho_t, \mathbf{j}_t) dt < \infty$  and  $|\rho_t'|^p = \mathcal{A}_{m,\beta}(\rho_t, \mathbf{j}_t)$ , for a.e.  $t \in [0, T]$ . Moreover, by Lemma 2.10 we find a family of antisymmetric measurable vector fields  $\mathbf{v} = (\mathbf{v}^{(1)}, \mathbf{v}^{(1)\perp}, \mathbf{v}^{(2)}, \mathbf{v}^{(2)\perp}) : [0, T] \times G \rightarrow \mathbb{R}^4$  such that for every  $t \in [0, T]$  and  $i = 1, 2$  we have

$$\begin{aligned}
d\mathbf{j}^{(i)\mu} &= (v_+^{(i)})^{q-1} d\gamma_1^{(i)} - (v_-^{(i)})^{q-1} d\gamma_2^{(i)}, \\
d\mathbf{j}^{(i)\perp} &= (v_+^{(i)\perp})^{q-1} d\gamma_1^{(i)\perp} - (v_-^{(i)\perp})^{q-1} d\gamma_2^{(i)\perp}.
\end{aligned}$$

Employing Proposition 3.24, the Hölder-type inequality (41), the identity (40), Definition 3.23, and Young's inequality, we obtain

$$\begin{aligned}
\mathcal{E}(\rho_T) - \mathcal{E}(\rho_0) &= \int_0^T \tilde{l}_{\rho_\tau}(v_\tau) [\beta \bar{\nabla} \delta_\rho \mathcal{E}(\rho_\tau)] d\tau = - \int_0^T \tilde{l}_{\rho_\tau}(v_\tau) [-\beta \bar{\nabla} \delta_\rho \mathcal{E}(\rho_\tau)] d\tau \\
&\geq - \int_0^T (\mathcal{A}_{m,\beta}(\rho_\tau, \mathbf{j}_\tau))^{1/p} (\mathcal{D}(\rho_\tau))^{1/q} d\tau = - \int_0^T |\rho_\tau'| (\mathcal{D}(\rho_\tau))^{1/q} d\tau \\
&\geq - \int_0^T \frac{1}{q} \mathcal{D}(\rho_\tau) + \frac{1}{p} |\rho_\tau'|^p d\tau.
\end{aligned}$$

Equation (43) implies that the inequalities are actually equalities. By Lemma 3.5, equality holds if and only if for  $i = 1, 2$  and a.e.  $t \in [0, T]$  we have

$$\begin{aligned}
(v_t^{(i)})_+ &= -\beta^{(i)} \bar{\nabla} \delta_{\rho^{(i)}} \mathcal{E}(\rho_t)_+, & \gamma_{1,t}^{(i)} \text{-a.e. on } G, & & (v_t^{(i)})_- &= -\beta^{(i)} \bar{\nabla} \delta_{\rho^{(i)}} \mathcal{E}(\rho_t)_-, & \gamma_{2,t}^{(i)} \text{-a.e. on } G, \\
(v_t^{(i)\perp})_+ &= -\beta^{(i)} \bar{\nabla} \delta_{\rho^{(i)}} \mathcal{E}(\rho_t)_+, & \gamma_{1,t}^{(i)\perp} \text{-a.e. on } G, & & (v_t^{(i)\perp})_- &= -\beta^{(i)} \bar{\nabla} \delta_{\rho^{(i)}} \mathcal{E}(\rho_t)_-, & \gamma_{2,t}^{(i)\perp} \text{-a.e. on } G
\end{aligned}$$

Hence,  $(\boldsymbol{\rho}, \mathbf{j}) \in \text{CE}_T$  is a weak solution to (32).  $\square$

### 3.4. Stability and existence of weak solutions

In this section, we utilize the characterization of weak solutions to (32) as minimizers of  $\mathcal{G}_T$ , attaining  $\mathcal{G}_T = 0$ . To show the existence of minimizers, we employ the direct method of calculus of variations. This way, we will prove the compactness and stability of gradient flows, which we will then utilize to approximate the desired problem by discrete problems. The existence of solutions is easy to show.

**Lemma 3.27.** *Let  $(\mu^n)_{n \in \mathbb{N}} \subset \mathcal{M}^+(\mathbb{R}^d)$  and suppose  $\mu^n \rightharpoonup^* \mu$  for some  $\mu \in \mathcal{M}^+(\mathbb{R}^d)$  as  $n \rightarrow \infty$ . Assume that  $\mu^n$  and  $\mu$  satisfy (MB1), (MB2) and (BC) uniformly in  $n$ . For  $i, k = 1, 2$ , let  $K^{(ik)}$  satisfy (K1), (K2) and  $K^{(21)} = K^{(12)}$ . Moreover, let  $(\rho^n)_{n \in \mathbb{N}}$  be a sequence in  $(\mathcal{P}_p(\mathbb{R}^d))^2$ , which satisfies  $\sup_{n \in \mathbb{N}} M_p(\rho^{n,(i)}) < \infty$  and is such that  $\rho^n \rightarrow \rho$  for some  $\rho \in (\mathcal{P}_p(\mathbb{R}^d))^2$ , as  $n \rightarrow \infty$ . Then, we have*

$$\liminf_{n \rightarrow \infty} \mathcal{D}(\mu^n; \rho^n) \geq \mathcal{D}(\mu; \rho).$$

*Proof.* For every  $n \in \mathbb{N}$  and  $i = 1, 2$ , we define  $u^{n,(i)} := \beta^{(i)} \sum_{k=1}^2 \bar{\nabla}(K^{(ik)} * \rho^{n,(k)})$  and  $u^{(i)} := \beta^{(i)} \sum_{k=1}^2 \bar{\nabla}(K^{(ik)} * \rho^{(k)})$ . Further, we define the convex and continuous map  $f : \mathbb{R} \rightarrow \mathbb{R}, r \mapsto (r_-)^q$  and note that we have

$$\begin{aligned} \mathcal{D}(\mu^n; \rho^n) &= \sum_{i=1}^2 \frac{1}{\beta^{(i)}} \iint_G f(u^{n,(i)}) \eta d(\gamma_1^{n,(i)} + \gamma_1^{n,(i)\perp}), \\ \mathcal{D}(\mu; \rho) &= \sum_{i=1}^2 \frac{1}{\beta^{(i)}} \iint_G f(u^{(i)}) \eta d(\gamma_1^{(i)} + \gamma_1^{(i)\perp}), \end{aligned}$$

where  $\gamma_k^{(i)}$  and  $\gamma_k^{(i)\perp}$  are as in Lemma 2.8 and Remark 2.9. We want to employ [3, Theorem 5.4.4 (ii)] to prove the desired inequality. To this end, we observe that  $u^{(i)} \in L^q(\eta \gamma_1^{(i)})$  and  $u^{n,(i)} \in L^q(\eta \gamma_1^{n,(i)})$ . Indeed, (K2), (MB1) and (MB2) and the bound

$$(\gamma_1^{n,(i)} + \gamma_1^{n,(i)\perp})(B) \leq M(\mu \otimes \mu + \rho^{n,(i)} \otimes \mu + \mu \otimes \rho^{n,(i)})(B) \quad \forall B \in \mathcal{B}(G),$$

from Lemma 2.14 imply

$$\begin{aligned} & \iint_G |u^{n,(i)}(x, y)|^q \eta(x, y) d\gamma_1^{n,(i)}(x, y) \\ &= \left(\beta^{(i)}\right)^q \iint_G \left| \sum_{k=1}^2 K^{(ik)} * \rho^{n,(k)}(y) - K^{(ik)} * \rho^{n,(k)}(x) \right|^q \eta(x, y) d(\gamma_1^{n,(i)} + \gamma_1^{n,(i)\perp})(x, y) \\ &\leq 2M(C_\mu + 2) \left(\beta^{(i)}\right)^q L_K^q C_\eta. \end{aligned}$$

Now, let  $\varphi \in C_c^\infty(G)$ . We find for  $i = 1, 2$ :

$$\begin{aligned} & \left(\beta^{(i)}\right)^{-q} \iint_G u^{n,(i)}(x, y) \varphi(x, y) \eta(x, y) d(\gamma_1^{n,(i)} + \gamma_1^{n,(i)\perp})(x, y) \\ &= \sum_{k=1}^2 \iint_G \int_{\mathbb{R}^d} \left( K^{(ik)}(y, z) - K^{(ik)}(x, z) \right) d\rho^{n,(k)}(z) \varphi(x, y) \eta(x, y) d(\gamma_1^{n,(i)} + \gamma_1^{n,(i)\perp})(x, y) \\ &= \sum_{k=1}^2 \iint_{\text{supp } \varphi} \int_{\mathbb{R}^d \cap B_R} \left( K^{(ik)}(y, z) - K^{(ik)}(x, z) \right) d\rho^{n,(k)}(z) \varphi(x, y) \eta(x, y) d(\gamma_1^{n,(i)} + \gamma_1^{n,(i)\perp})(x, y) \\ &+ \sum_{k=1}^2 \iint_{\text{supp } \varphi} \int_{\mathbb{R}^d \setminus B_R} \left( K^{(ik)}(y, z) - K^{(ik)}(x, z) \right) d\rho^{n,(k)}(z) \varphi(x, y) \eta(x, y) d(\gamma_1^{n,(i)} + \gamma_1^{n,(i)\perp})(x, y). \end{aligned}$$

The terms which are integrated over  $\mathbb{R}^d \setminus B_R$  vanish as  $R \rightarrow \infty$  since  $\rho^{n,(k)}(\mathbb{R}^d \setminus B_R) \xrightarrow{R \rightarrow \infty} 0$  by Prokhorov's Theorem. (K2) together with (MB1) and (MB2) yields

$$\begin{aligned} & \left| \iint_{\text{supp } \varphi} \int_{\mathbb{R}^d \setminus B_R} \left( K^{(ik)}(y, z) - K^{(ik)}(x, z) \right) \varphi(x, y) \eta(x, y) d(\rho^{n,(k)} \otimes (\gamma_1^{n,(i)} + \gamma_1^{n,(i)\perp}))(z, x, y) \right| \\ & \leq ML_K \|\varphi\|_{\infty} \rho^{n,(k)}(\mathbb{R}^d \setminus B_R) \iint_{\text{supp } \varphi} (|x - y| \vee |x - y|^p) \eta(x, y) d(\mu + 2\rho^{n,(i)})(x) d\mu^n(y) \\ & \leq \frac{(C_\mu + 2)ML_K C_\eta \|\varphi\|_{\infty} \rho^{n,(k)}(\mathbb{R}^d \setminus B_R)}{\inf_{\text{supp } \varphi} (|x - y|^{q/p} \vee |x - y|^q)}. \end{aligned}$$

By (W) and (K2), the function  $(z, y, x) \mapsto (K(y, z) - K(x, z))\varphi(x, y)\eta(x, y)$  is continuous and bounded on  $(\mathbb{R}^d \cap B_R) \times G$ . On the other hand we have  $\rho^{n,(k)} \otimes (\gamma_1^{n,(i)} + \gamma_1^{n,(i)\perp}) \rightarrow \rho^{(k)} \otimes (\gamma_1^{(i)} + \gamma_1^{(i)\perp})$  in  $\mathcal{P}(\mathbb{R}^d) \times \mathcal{M}^+(G)$  for  $i, k = 1, 2$ . Therefore, for any  $R > 0$  and  $i, k = 1, 2$ , we obtain

$$\begin{aligned} & \lim_{n \rightarrow \infty} \iint_{\text{supp } \varphi} \int_{\mathbb{R}^d \cap B_R} \left( K^{(ik)}(y, z) - K^{(ik)}(x, z) \right) \varphi(x, y) \eta(x, y) d(\rho^{n,(k)} \otimes (\gamma_1^{n,(i)} + \gamma_1^{n,(i)\perp}))(z, x, y) \\ & = \iint_{\text{supp } \varphi} \int_{\mathbb{R}^d \cap B_R} \left( K^{(ik)}(y, z) - K^{(ik)}(x, z) \right) \varphi(x, y) \eta(x, y) d(\rho^{(k)} \otimes (\gamma_1^{(i)} + \gamma_1^{(i)\perp}))(z, x, y). \end{aligned}$$

Letting  $R \rightarrow \infty$ , we obtain

$$\lim_{n \rightarrow \infty} \iint_G u^{n,(i)} \varphi \eta d(\gamma_1^{n,(i)} + \gamma_1^{n,(i)\perp}) = \iint_G u^{(i)} \varphi \eta d(\gamma_1^{(i)} + \gamma_1^{(i)\perp}).$$

Therefore,  $u^{n,(i)}$  converges weakly to  $u^{(i)}$  in the sense of [3, Definition 5.4.3]. This allows the application of [3, Theorem 5.4.4 (ii)] to conclude

$$\begin{aligned} \liminf_{n \rightarrow \infty} \mathcal{D}(\mu^n; \rho^n) & = \liminf_{n \rightarrow \infty} \sum_{i=1}^2 \frac{1}{\beta^{(i)}} \iint_G f(u^{n,(i)}) \eta d(\gamma_1^{n,(i)} + \gamma_1^{n,(i)\perp}) \\ & \geq \sum_{i=1}^2 \frac{1}{\beta^{(i)}} \iint_G f(u^{(i)}) \eta d(\gamma_1^{(i)} + \gamma_1^{(i)\perp}) = \mathcal{D}(\mu; \rho), \end{aligned}$$

which finishes the proof.  $\square$

**Lemma 3.28** (Compactness and lower semicontinuity of the De Giorgi functional). *Let  $(\mu^n)_{n \in \mathbb{N}} \subset \mathcal{M}^+(\mathbb{R}^d)$  and suppose  $\mu^n \rightharpoonup^* \mu$  for some  $\mu \in \mathcal{M}^+(\mathbb{R}^d)$  as  $n \rightarrow \infty$ . Assume that  $\mu^n$  and  $\mu$  satisfy (MB1), (MB2) and (BC) uniformly in  $n$ . For  $i, k = 1, 2$ , let  $K^{(ik)}$  satisfy (K1), (K2) and  $K^{(21)} = K^{(12)}$ . Moreover, let  $(\rho^n)_{n \in \mathbb{N}}$  be such that  $\rho^n \in AC^p([0, T]; ((\mathcal{P}_p(\mathbb{R}^d))^2, \mathcal{T}_{m,\beta,\mu^n}))$ , for all  $n \in \mathbb{N}$  with  $\sup_{n \in \mathbb{N}} M_p(\rho_0^{n,(i)}) < \infty$  for  $i = 1, 2$  and  $\sup_{n \in \mathbb{N}} \mathcal{G}_T(\mu^n; \rho^n) < \infty$ . Then, there exists  $\rho \in AC^p([0, T]; ((\mathcal{P}_p(\mathbb{R}^d))^2, \mathcal{T}_{m,\beta,\mu}))$  such that up to a subsequence we have  $\rho_t^n \rightarrow \rho_t$  as  $n \rightarrow \infty$  for all  $t \in [0, T]$  and it holds*

$$\liminf_{n \rightarrow \infty} \mathcal{G}_T(\mu^n; \rho^n) \geq \mathcal{G}_T(\mu; \rho).$$

*Proof.* Let  $n \in \mathbb{N}$ . Recall

$$\mathcal{G}_T(\mu^n; \rho^n) = \mathcal{E}(\rho_T^n) - \mathcal{E}(\rho_0^n) + \int_0^T \frac{1}{q} \mathcal{D}(\mu^n; \rho_t^n) + \frac{1}{p} |(\rho_t^n)'|_{\mathcal{T}_{m,\beta,\mu^n}}^p dt,$$

where the metric derivative of  $\rho_t^n$  is taken with respect to  $\mathcal{T}_{m,\beta,\mu^n}$ . Since the domain of  $\mathcal{E}$  is all of  $(\mathcal{P}_p(\mathbb{R}^d))^2$  and  $\mathcal{D}$  is nonnegative, the bound  $\sup_{n \in \mathbb{N}} \mathcal{G}_T(\mu^n; \rho^n) < \infty$  ensures that

$$\sup_{n \in \mathbb{N}} \int_0^T |(\rho_t^n)'|_{\mathcal{T}_{m,\beta,\mu^n}}^p dt < \infty.$$

Since for any  $n$  we have  $\boldsymbol{\rho}^n \in \text{AC}^p([0, T]; ((\mathcal{P}_p(\mathbb{R}^d))^2, \mathcal{T}_{m, \beta, \mu^n}))$ , Proposition 2.31 yields the existence of a unique flux  $\mathbf{j}^n$  such that  $(\boldsymbol{\rho}^n, \mathbf{j}^n) \in \text{CE}_T$  and  $|(\rho_t^n)'|_{\mathcal{T}_{m, \beta, \mu^n}}^p = \mathcal{A}_{m, \beta}(\mu^n; \rho_t^n, \mathbf{j}_t^n)$  for a.e.  $t \in [0, T]$ . We therefore obtain

$$\sup_{n \in \mathbb{N}} \int_0^t \mathcal{A}_{m, \beta}(\mu^n; \rho_t^n, \mathbf{j}_t^n) dt = \sup_{n \in \mathbb{N}} \int_0^T |(\rho_t^n)'|_{\mathcal{T}_{m, \beta, \mu^n}}^p dt < \infty.$$

Thus, by Proposition 2.22, there exists  $(\boldsymbol{\rho}, \mathbf{j}) \in \text{CE}_T$  such that, up to subsequences,  $\rho_t^n \rightharpoonup \rho_t$  and  $\mathbf{j}_t^n \rightharpoonup^* \mathbf{j}_t$  as  $n \rightarrow \infty$  for a.e.  $t \in [0, T]$ , and we have

$$\int_0^t \mathcal{A}_{m, \beta}(\mu; \rho_t, \mathbf{j}_t) dt \leq \liminf_{n \rightarrow \infty} \int_0^t \mathcal{A}_{m, \beta}(\mu^n; \rho_t^n, \mathbf{j}_t^n) dt < \infty.$$

Hence, Proposition 2.31 implies that  $\boldsymbol{\rho} \in \text{AC}^p([0, T]; ((\mathcal{P}_p(\mathbb{R}^d))^2, \mathcal{T}_{m, \beta, \mu}))$  and  $|\rho_t'|_{\mathcal{T}_{m, \beta, \mu}}^p \leq \mathcal{A}_{m, \beta}(\mu; \rho_t, \mathbf{j}_t)$  for a.e.  $t \in [0, T]$ , which now yields

$$\int_0^T |\rho_t'|_{\mathcal{T}_{m, \beta, \mu}}^p dt \leq \liminf_{n \rightarrow \infty} \int_0^T |(\rho_t^n)'|_{\mathcal{T}_{m, \beta, \mu^n}}^p dt. \quad (44)$$

By Proposition 3.3, we have that the  $\mathcal{E}$  is narrowly continuous, i.e.

$$\lim_{n \rightarrow \infty} \mathcal{E}(\rho_0^n) = \mathcal{E}(\rho_0) \text{ and } \lim_{n \rightarrow \infty} \mathcal{E}(\rho_T^n) = \mathcal{E}(\rho_T). \quad (45)$$

Lastly, Fatou's lemma and the narrow lower semicontinuity of  $\mathcal{D}$ , shown in Lemma 3.27, give us

$$\int_0^T \mathcal{D}(\mu; \rho_t) dt \leq \int_0^T \liminf_{n \rightarrow \infty} \mathcal{D}(\mu^n; \rho_t^n) dt \leq \liminf_{n \rightarrow \infty} \int_0^T \mathcal{D}(\mu^n; \rho_t^n) dt. \quad (46)$$

Combining (44), (45) and (46), we finally obtain

$$\mathcal{G}_T(\mu; \boldsymbol{\rho}) = \mathcal{E}(\rho_T) - \mathcal{E}(\rho_0) + \int_0^T \frac{1}{q} \mathcal{D}(\mu; \rho_t) + \frac{1}{p} |\rho_t'|_{\mathcal{T}_{m, \beta, \mu}}^p dt \leq \liminf_{n \rightarrow \infty} \mathcal{G}_T(\mu^n; \boldsymbol{\rho}^n),$$

which finishes the proof.  $\square$

**Theorem 3.29** (Closedness of the Null Space of the DeGiorgi Functional). *Let  $(\mu^n)_{n \in \mathbb{N}} \subset \mathcal{M}^+(\mathbb{R}^d)$  and suppose  $\mu^n \rightharpoonup^* \mu$  for some  $\mu \in \mathcal{M}^+(\mathbb{R}^d)$  as  $n \rightarrow \infty$ . Assume that  $\mu^n$  and  $\mu$  satisfy (MB1), (MB2) and (BC) uniformly in  $n$ . For  $i, k = 1, 2$  let  $K^{(ik)}$  satisfy (K1), (K2) and  $K^{(21)} = K^{(12)}$ . Let  $\boldsymbol{\rho}^n$  be a gradient flow of  $\mathcal{E}$  with respect to  $\mu^n$  for all  $n \in \mathbb{N}$ , i.e.*

$$\mathcal{G}_T(\mu^n; \boldsymbol{\rho}^n) = 0, \quad \text{for all } n \in \mathbb{N}.$$

*Additionally, assume  $\sup_{n \in \mathbb{N}} M_p(\rho_0^{n, (i)}) < \infty$  for  $i = 1, 2$  and  $\rho_t^n \rightharpoonup \rho_t$  for all  $t \in [0, T]$  for some  $\boldsymbol{\rho} \subset (\mathcal{P}_p(\mathbb{R}^d))^2$  as  $n \rightarrow \infty$ . Then,  $\boldsymbol{\rho} \in \text{AC}^p([0, T]; ((\mathcal{P}_p(\mathbb{R}^d))^2, \mathcal{T}_{m, \beta, \mu}))$  is a gradient flow of  $\mathcal{E}$  with respect to  $\mu$ , i.e.*

$$\mathcal{G}_T(\mu; \boldsymbol{\rho}) = 0.$$

*Proof.* By Lemma 3.28, we immediately obtain that  $\boldsymbol{\rho} \in \text{AC}^p([0, T]; ((\mathcal{P}_p(\mathbb{R}^d))^2, \mathcal{T}_{m, \beta, \mu}))$  and that, up to a subsequence, we have

$$\liminf_{n \rightarrow \infty} \mathcal{G}_T(\mu^n; \boldsymbol{\rho}^n) \geq \mathcal{G}_T(\mu; \boldsymbol{\rho}).$$

Finally, since  $\mathcal{G}_T(\mu; \boldsymbol{\rho}) \geq 0$ , Young's inequality and Corollary 3.25 yield  $\mathcal{G}_T(\mu; \boldsymbol{\rho}) = 0$ .  $\square$

**Theorem 3.30** (Existence of weak solutions). *Let  $m$  satisfy assumption (A) from Proposition 2.35 and for  $i, k = 1, 2$ , let  $K^{(ik)}$  satisfy (K1), (K2) and  $K^{(21)} = K^{(12)}$ . Suppose that  $\mu \in \mathcal{M}^+(\mathbb{R}^d)$  satisfies (MB1) and (BC). Assume further that there exists  $C'_\eta > 0$  such that we have*

$$\sup_{(x, y) \in G \cap \text{supp } \mu \otimes \mu} (|x - y|^q \vee |x - y|^{pq}) \eta(x, y) \leq C'_\eta. \quad (\text{MB2}')$$

*Let  $\boldsymbol{\rho}_0 \in (\mathcal{P}_p(\mathbb{R}^d))^2$  be  $\mu$ -absolutely continuous. Then, there exists a weakly continuous curve  $\boldsymbol{\rho} : [0, T] \rightarrow (\mathcal{P}_p(\mathbb{R}^d))^2$  s.t.  $\text{supp } \rho_t \subseteq \text{supp } \mu$  for all  $t \in [0, T]$ , which is a weak solution of (32) and satisfies the initial condition  $\rho_0 = \boldsymbol{\rho}_0$ .*

*Proof.* Let  $(\mu^n)_{n \in \mathbb{N}} \subset \mathcal{M}^+(\mathbb{R}^d)$  be a sequence of atomic measures with finitely many atoms that narrowly converges to  $\mu$ . This means every  $\mu^n$  is of the form

$$\mu^n = \sum_{l=1}^{N_n} \mu_l^n \delta_{x_l^n}, \quad (47)$$

for some  $N_n \in \mathbb{N}$ ,  $\mu_l^n \in \mathbb{R} \setminus \{0\}$  and  $x_l^n \in \mathbb{R}^d$ . We further assume, without loss of generality, for any  $n \in \mathbb{N}$  that  $\mu^n(\mathbb{R}^d) \leq \mu(\mathbb{R}^d)$  and  $\text{supp } \mu^n \subset \text{supp } \mu$ .

Since every  $\mu^n$  consists of finitely many atoms and their limit  $\mu$  satisfies **(BC)**, the family  $(\mu^n)_{n \in \mathbb{N}}$  satisfies **(BC)** uniformly in  $n$ . Indeed, as  $\mu^n \rightarrow \mu$ , for any  $\varepsilon > 0$  and  $N \in \mathbb{N}$ , there exists  $\tilde{\varepsilon} = \tilde{\varepsilon}(\varepsilon, N) > 0$  s.t.  $\tilde{\varepsilon} \rightarrow 0$  when  $\varepsilon \rightarrow 0$  and  $N \rightarrow \infty$  and s.t. we have

$$\begin{aligned} & \sup_{n \geq N} \sup_{x \in \mathbb{R}^d} \int_{B_\varepsilon(x) \setminus \{x\}} |x - y|^q \eta(x, y) d\mu^n(y) \\ & \leq \sup_{x \in \mathbb{R}^d} \left( \sup_{n \geq N} \int_{B_\varepsilon(x) \setminus \{x\}} |x - y|^q \eta(x, y) |d\mu^n - d\mu|(y) + \int_{B_\varepsilon(x) \setminus \{x\}} |x - y|^q \eta(x, y) d\mu(y) \right) \\ & \leq \tilde{\varepsilon} + \sup_{x \in \mathbb{R}^d} \int_{B_\varepsilon(x) \setminus \{x\}} |x - y|^q \eta(x, y) d\mu(y). \end{aligned}$$

On the other hand, since all the  $\mu^n$  consist of only finitely many atoms, for any  $\varepsilon > 0$  there exists  $N = N(\varepsilon) \in \mathbb{N}$  such that  $N \rightarrow \infty$  when  $\varepsilon \rightarrow 0$  and such that we have

$$\sup_{n < N} \sup_{x \in \mathbb{R}^d} \int_{B_\varepsilon(x) \setminus \{x\}} |x - y|^q \eta(x, y) d\mu^n(y) = 0.$$

Thus, choosing  $N(\varepsilon)$  and  $\tilde{\varepsilon}(\varepsilon, N(\varepsilon))$  as above, letting  $\varepsilon \rightarrow 0$  and using the fact that  $\mu$  satisfies **(BC)**, we obtain

$$\lim_{\varepsilon \rightarrow 0} \sup_{n \in \mathbb{N}} \sup_{x \in \mathbb{R}^d} \int_{B_\varepsilon(x) \setminus \{x\}} |x - y|^q \eta(x, y) d\mu^n(y) = 0.$$

Next, denote by  $\tilde{\mu}^n$  the normalization of  $\mu^n$ , i.e.

$$\tilde{\mu}^n = \frac{\mu(\mathbb{R}^d)}{\mu^n(\mathbb{R}^d)} \mu^n,$$

and let  $\pi^n$  be an optimal transportation plan between  $\mu$  and  $\tilde{\mu}^n$  for the quadratic cost. Since we have  $\tilde{\mu}^n \rightarrow \mu$  narrowly, we have  $\pi^n \rightarrow (\text{id} \times \text{id})_{\#} \mu$  narrowly. For  $i = 1, 2$  let  $\tilde{\varrho}_0^{(i)}$  be the density of  $\varrho_0^{(i)}$  with respect to  $\mu$  and let  $\varrho_0^{n,(i)}$  be the second marginal of  $(\tilde{\varrho}_0^{(i)} \times \mathbb{1}) d\pi^n$ , i.e., for any  $B \in \mathcal{B}(\mathbb{R}^d)$ , we have  $\varrho_0^{n,(i)}(B) = \int_{\mathbb{R}^d \times B} \tilde{\varrho}_0^{(i)}(x) d\pi^n(x, y)$ . Then, by construction, for any  $n \in \mathbb{N}$  and  $i = 1, 2$ , we find  $\varrho_0^{n,(i)}(\mathbb{R}^d) = \varrho_0^{(i)}(\mathbb{R}^d)$  and  $\varrho_0^{n,(i)} \ll \mu^n$ . Also, by the convergence of  $\pi^n$  and the fact that  $(\tilde{\varrho}_0^{(i)} \times \mathbb{1}) \pi^n$  is a transport plan between  $\varrho_0^{(i)}$  and  $\varrho_0^{n,(i)}$ , we find that  $\varrho_0^{n,(i)} \rightarrow \varrho_0^{(i)}$  for  $i = 1, 2$  as  $n \rightarrow \infty$ . By **(MB2')**, for all  $n \in \mathbb{N}$ , we obtain the bound

$$\mu - \text{ess sup}_{x \in \mathbb{R}^d} \int_{\mathbb{R}^d} (|x - y|^q \vee |x - y|^{pq}) \eta(x, y) d\mu^n(y) \leq C'_\eta \mu^n(\mathbb{R}^d) \leq C'_\eta \mu(\mathbb{R}^d).$$

Since, by construction,  $\varrho_0^{n,(i)} \ll \mu^n$ , we have  $\text{supp } \varrho_0^{n,(i)} \subset \text{supp } \mu^n \subset \text{supp } \mu$ . By Proposition 2.35, the nested support is preserved in time for any  $\rho^n \in \text{AC}^p([0, T]; ((\mathcal{P}_p(\mathbb{R}^d))^2, \mathcal{T}_{m,\beta,\mu}))$  with  $\rho_0^{n,(i)} = \varrho_0^{n,(i)}$ , i.e., we have  $\text{supp } \rho_t^{n,(i)} \subset \text{supp } \mu^n \subset \text{supp } \mu$  for any  $t \in [0, T]$  and any  $n \in \mathbb{N}$ . Therefore, **(MB2')** can be used to replace **(MB2)** uniformly in  $n$ , when employing Lemma 3.28 and Theorem 3.29 later in this proof. Further note that  $\{\mu^n\}_n$  satisfy **(MB1)** uniformly in  $n$ , since  $\mu$  satisfies **(MB1)**. These considerations now allow us to construct curves  $\rho^n \in \text{AC}^p([0, T]; ((\mathcal{P}_p(\mathbb{R}^d))^2, \mathcal{T}_{m,\beta,\mu}))$ , which are gradient flows and converge to a gradient flow  $\rho$ . Indeed, since any  $\mu^n$  is a counting measure, we have  $\rho_t^{n,(i)} \ll \mu^n$  for any  $i = 1, 2$ ,  $t \in [0, T]$ , and  $n \in \mathbb{N}$ . Thus, we can write

$$\rho_t^{n,(i)} = \sum_{v=1}^{N_n} \rho_v^{n,(i)}(t) \mu_v^n \delta_{x_v^n}, \quad (48)$$

for suitable functions  $\rho_\nu^{n,(i)} : [0, T] \rightarrow \mathbb{R}$  and the points  $x_\nu^n \in \mathbb{R}^d$  from (47). Now, let  $\varphi_\nu^n \in C_c^\infty(\mathbb{R}^d)$ ,  $\nu \in \{1, \dots, N_n\}$  satisfy  $\varphi_\nu^n(x_\nu^n) \neq 0$  and  $\varphi_\nu^n(x_\kappa^n) = 0$  for  $x_\kappa \neq \nu$ . Then, inserting (47) and (48) into Equation (33), we find for  $i = 1, 2$  and  $\nu \in \{1, \dots, N_n\}$

$$\begin{aligned} \partial_t \rho_l^{n,(i)} = & \sum_{m=1}^{N_n} \left[ \left( \beta^{(i)} m \left( \rho_l^{n,(i)}, \rho_m^{n,(i)} \right) \left( \sum_{k=1}^2 \sum_{h=1}^{N_n} \left( K^{(ik)}(x_m^n, x_h^n) - K^{(ik)}(x_l^n, x_h^n) \right) \rho_h^{n,(k)} \mu_h^n \right)_+ \right)^{q-1} \right. \\ & \left. - \left( \beta^{(i)} m \left( \rho_m^{n,(i)}, \rho_l^{n,(i)} \right) \left( \sum_{k=1}^2 \sum_{h=1}^{N_n} \left( K^{(ik)}(x_m^n, x_h^n) - K^{(ik)}(x_l^n, x_h^n) \right) \rho_h^{n,(k)} \mu_h^n \right)_- \right)^{q-1} \right] \eta(x_l^n, x_m^n) \mu_m^n. \end{aligned} \quad (49)$$

Since  $m(0, s) = 0$ , we see that the simplex defined by

$$\rho_\nu^{n,(i)} \in \left[ 0, \left( \min_{1 \leq m \leq N_n} \mu_m^n \right)^{-1} \right], \quad \sum_{\nu=1}^{N_n} \mu_\nu^n \rho_\nu^{n,(i)} = 1. \quad (50)$$

is an invariant region of the dynamics. Due to the continuity of  $m$ , the right-hand side of (49) is continuous with respect to  $\rho^{n,(i)}$  for any  $n \in \mathbb{N}$  and  $i = 1, 2$ . With this, the Peano existence theorem provides us with a strong solution  $\rho^n$  of (49) on an interval  $[0, \tau_n]$  for some  $\tau_n > 0$ . Due to (50),  $\tau_n$  only depends on  $n$  and  $\mu^n$ . Thus, by a standard continuation argument, a piecewise  $C^1$  solution exists on the whole interval  $[0, T]$ . By construction, this solution is a weak solution for (32) in the sense of Definition 3.1 with respect to  $\mu^n$  starting from  $\mathfrak{q}_0^n$ . Therefore, by Theorem 3.26,  $\rho^n$  is a gradient flow of  $\mathcal{E}$  with respect to  $\mu^n$  with initial datum  $\mathfrak{q}_0^n$ , for any  $n \in \mathbb{N}$ . This allows us to apply the compactness from Lemma 3.28 and the stability from Theorem 3.29 to find that, up to a subsequence,  $\rho_t^n \rightarrow \rho_t$  as  $n \rightarrow \infty$  for all  $t \in [0, T]$ , where  $\rho \in \text{AC}^p([0, T]; ((\mathcal{P}_p(\mathbb{R}^d))^2, \mathcal{T}_{m,\beta,\mu}))$  is a gradient flow of  $\mathcal{E}$  with respect to  $\mu$  starting from  $\mathfrak{q}_0$ .  $\square$

**Remark 3.31.** Assumption (MB2') is needed to obtain an atomic approximating sequence  $(\mu^n)_n$  for  $\mu$ , which satisfies (MB1), (MB2) and (BC) uniformly in  $n$ . There might be cases, where it is possible drop this assumption if one is able to explicitly construct a sequence  $(\mu^n)_n$  satisfying these bounds uniformly in  $n$ .

## Appendix

### A. Chain rule

**Remark A.1.** (Approximate energies). Let  $K^{(ik)}$ ,  $i, k = 1, 2$  satisfy (K1), (K2) and  $K^{(21)} = K^{(12)}$ . Let  $m \in C_c^\infty(\mathbb{R}^d \times \mathbb{R}^d)$  be a standard mollifier. For  $\varepsilon > 0$  and  $z \in \mathbb{R}^{2d}$ , we set  $m_\varepsilon(z) := \frac{1}{\varepsilon^{2d}} m(z/\varepsilon)$ . Moreover, for  $R > 0$  let  $\varphi_R \in C_c^\infty(\mathbb{R}^{2d})$  be a cut-off function with  $\text{supp } \varphi_R \subset B_{2R}(0)$ ,  $\varphi_R|_{B_R(0)} \equiv 1$  and  $|\nabla \varphi_R| \leq \frac{2}{R}$ . Using the mollifier and the cut-off we define for  $i, k = 1, 2$

$$K_R^{\varepsilon,(ik)}(x, y) := \varphi_R(x, y) (K^{(ik)} * m_\varepsilon)(x, y).$$

Note that the functions  $K_R^{\varepsilon,(ik)}$  still satisfy (K1), (K2) and, additionally, lie in  $C_c^\infty(\mathbb{R}^{2d})$ . For  $\rho \in (\mathcal{P}_p(\mathbb{R}^d))^2$ , we define the approximate energies

$$\mathcal{E}_R^\varepsilon(\rho) := \frac{1}{2} \sum_{i,k=1}^2 \iint_{\mathbb{R}^d \times \mathbb{R}^d} K_R^{\varepsilon,(ik)}(x, y) d\rho^{(i)}(x) d\rho^{(k)}(y).$$

**Lemma A.2** (Mollification). Let  $(\rho, \mathbf{j}) \in \text{CE}_T$  with  $\int_0^T \mathcal{A}_{m,\beta}(\mu; \rho_t, \mathbf{j}_t) dt < \infty$ . Let  $n \in C_c^\infty(\mathbb{R})$  be a standard mollifier with  $\text{supp } n \subseteq [-1, 1]$ . For  $\bar{\varepsilon} > 0$  set  $n_{\bar{\varepsilon}}(t) := \frac{1}{\bar{\varepsilon}} n\left(\frac{t}{\bar{\varepsilon}}\right)$ . Extending  $\rho$  and  $\mathbf{j}$  periodically to  $[-T, 2T]$ , i.e.,

$\rho_{-t} = \rho_{T-t}$  and  $\rho_{T+t} = \rho_t$  for any  $t \in (0, T]$  and likewise for  $\mathbf{j}$ , we define the regularizations  $\rho_t^{\bar{\varepsilon}} = (\rho_t^{\bar{\varepsilon},(1)}, \rho_t^{\bar{\varepsilon},(2)})^\top$  and  $\mathbf{j}_t^{\bar{\varepsilon}} = (\mathbf{j}_t^{\bar{\varepsilon},(1)}, \mathbf{j}_t^{\bar{\varepsilon},(2)})^\top$  by

$$\begin{aligned}\rho_t^{\bar{\varepsilon},(i)}(A) &:= (n_{\bar{\varepsilon}} * \rho_t^{(i)})(A) = \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} n_{\bar{\varepsilon}}(s) \rho_{t-s}(A) ds, \quad \forall A \subseteq \mathbb{R}^d, \\ \mathbf{j}_t^{\bar{\varepsilon},(i)}(B) &:= (n_{\bar{\varepsilon}} * \mathbf{j}_t^{(i)})(B) = \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} n_{\bar{\varepsilon}}(s) \mathbf{j}_{t-s}(B) ds, \quad \forall B \subseteq G,\end{aligned}$$

for  $i = 1, 2$  and any  $\bar{\varepsilon} \in (0, T)$ . Then, we obtain that the integral  $\int_0^T \mathcal{A}_{m,\beta}(\mu; \rho_t^{\bar{\varepsilon}}, \mathbf{j}_t^{\bar{\varepsilon}}) dt$  is uniformly bounded with respect to  $\bar{\varepsilon}$  and that the pair  $(\rho^{\bar{\varepsilon}}, \mathbf{j}^{\bar{\varepsilon}})$  lies in  $\text{CE}_T$ . Furthermore, if  $\rho_t \in (\mathcal{P}_p(\mathbb{R}^d))^2$ , then  $(\rho_t^{\bar{\varepsilon}})_{\bar{\varepsilon}} \subset (\mathcal{P}_p(\mathbb{R}^d))^2$  with uniformly bounded  $p$ -th moments.

*Proof.* If  $\rho_t \in (\mathcal{P}_p(\mathbb{R}^d))^2$  for all  $t \in [0, T]$ , it is immediate that  $\rho_t^{\bar{\varepsilon}} \in (\mathcal{P}_p(\mathbb{R}^d))^2$  for all  $t \in [0, T]$ . Indeed, for  $i \in \{1, 2\}$  and  $f: \mathbb{R}^d \rightarrow \mathbb{R}$  integrable with respect to  $\rho_t$  for  $t \in [0, T]$ , Fubini's theorem gives us

$$\int_{\mathbb{R}^d} f(x) d\rho_t^{\bar{\varepsilon},(i)}(x) = \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} n_{\bar{\varepsilon}}(s) \int_{\mathbb{R}^d} f(x) d\rho_{t-s}^{(i)}(x) ds.$$

In particular, for  $f \equiv 1$  the inner integral is equal to 1 and hence the whole expression, while for  $f(x) = |x|^p$  the inner integral and consequently the whole expression are both finite. In particular, the family  $(\rho_t^{\bar{\varepsilon}})_{\bar{\varepsilon}}$  has uniformly bounded  $p$ -th moments.

Next, we prove the uniform bound. To shorten notation, we only consider the case  $R \wedge S < \infty$ . The recession terms in the case  $R = S = \infty$  can then be treated analogously by employing  $\sigma_t^{(i)} = \zeta_t^{(i)} = \rho_t^{(i)\perp} \otimes \mu + \mu \otimes \rho_t^{(i)\perp}$  for  $i = 1, 2$ . By the joint convexity of the density function  $\alpha$ , Jensen's inequality and Fubini's theorem, we obtain

$$\begin{aligned}\int_0^T \mathcal{A}_{m,\beta}(\mu; \rho_t^{\bar{\varepsilon}}, \mathbf{j}_t^{\bar{\varepsilon}}) dt &= \frac{1}{2} \sum_{i=1}^2 \frac{1}{\beta^{(i)}} \int_0^T \iint_G \left[ \alpha \left( \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \frac{d\mathbf{j}_{t-s}^{(i)}}{d(\mu \otimes \mu)} n_{\bar{\varepsilon}}(s) ds, \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \frac{d\gamma_{1,t-s}^{(i)}}{d(\mu \otimes \mu)} n_{\bar{\varepsilon}}(s) ds \right) \right. \\ &\quad \left. + \alpha \left( - \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \frac{d\mathbf{j}_{t-s}^{(i)}}{d(\mu \otimes \mu)} n_{\bar{\varepsilon}}(s) ds, \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \frac{d\gamma_{2,t-s}^{(i)}}{d(\mu \otimes \mu)} n_{\bar{\varepsilon}}(s) ds \right) \right] \eta d(\mu \otimes \mu) dt \\ &\leq \frac{1}{2} \sum_{i=1}^2 \frac{1}{\beta^{(i)}} \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \int_0^T \iint_G \left[ \alpha \left( \frac{d\mathbf{j}_{t-s}^{(i)}}{d(\mu \otimes \mu)} n_{\bar{\varepsilon}}(s), \frac{d\gamma_{1,t-s}^{(i)}}{d(\mu \otimes \mu)} n_{\bar{\varepsilon}}(s) \right) \right. \\ &\quad \left. + \alpha \left( - \frac{d\mathbf{j}_{t-s}^{(i)}}{d(\mu \otimes \mu)} n_{\bar{\varepsilon}}(s), \frac{d\gamma_{2,t-s}^{(i)}}{d(\mu \otimes \mu)} n_{\bar{\varepsilon}}(s) \right) \right] n_{\bar{\varepsilon}}(s) ds \eta d\mu \otimes \mu \\ &= \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} \int_0^T \mathcal{A}_{m,\beta}(\mu; \rho_{t-s}, \mathbf{j}_{t-s}) dt ds \leq \int_{-T}^{2T} \mathcal{A}_{m,\beta}(\mu; \rho_t, \mathbf{j}_t) dt \\ &= 3 \int_0^T \mathcal{A}_{m,\beta}(\mu; \rho_t, \mathbf{j}_t) dt < \infty.\end{aligned}$$

Let us now check that  $(\rho^{\bar{\varepsilon}}, \mathbf{j}^{\bar{\varepsilon}}) \in \text{CE}_T$ . The first two requirements are immediate. Hence, it only remains to check that the continuity equation (23) holds. To this end, let  $i \in \{1, 2\}$ ,  $\varphi \in C_c^\infty([0, T] \times \mathbb{R}^d)$ , and periodically extend  $\varphi$  to  $[-T, 2T]$ . Then, for  $i = 1, 2$  and denoting by  $\bar{\rho}^{(i)}$  the continuous representative of  $\rho^{(i)}$  defined in

Lemma 2.20, we obtain

$$\begin{aligned}
& \int_0^T \int_{\mathbb{R}^d} \partial_t \varphi_t(x) d\rho_t^{\bar{\varepsilon},(i)}(x) dt + \frac{1}{2} \int_0^T \iint_G (\bar{\nabla} \varphi_t)(x, y) \eta(x, y) dj_t^{\bar{\varepsilon},(i)}(x, y) dt \\
&= \int_0^T \int_{\mathbb{R}^d} \partial_t \varphi_t(x) \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} n_{\bar{\varepsilon}}(s) d\rho_{t-s}^{(i)}(x) ds dt + \frac{1}{2} \int_0^T \iint_G (\bar{\nabla} \varphi_t)(x, y) \eta(x, y) \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} n_{\bar{\varepsilon}}(s) dj_{t-s}^{(i)}(x, y) ds dt \\
&= \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} n_{\bar{\varepsilon}}(s) \left[ \int_0^T \int_{\mathbb{R}^d} \partial_t \varphi_t(x) d\rho_{t-s}^{(i)}(x) dt + \frac{1}{2} \int_0^T \iint_G (\bar{\nabla} \varphi_t)(x, y) \eta(x, y) dj_{t-s}^{(i)}(x, y) dt \right] ds \\
&= \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} n_{\bar{\varepsilon}}(s) \left[ \int_{-s}^{T-s} \int_{\mathbb{R}^d} \partial_t \varphi_{t+s}(x) d\rho_t^{(i)}(x) dt + \frac{1}{2} \int_{-s}^{T-s} \iint_G (\bar{\nabla} \varphi_{t+s})(x, y) \eta(x, y) dj_t^{(i)}(x, y) dt \right] ds \\
&= \int_{-\bar{\varepsilon}}^{\bar{\varepsilon}} n_{\bar{\varepsilon}}(s) \left[ \int_{\mathbb{R}^d} \varphi_T(x) d\bar{\rho}_{T-s}^{(i)}(x) - \int_{\mathbb{R}^d} \varphi_0(x) d\bar{\rho}_{-s}^{(i)}(x) \right] ds = \int_{\mathbb{R}^d} \varphi_T(x) d\bar{\rho}_T^{\bar{\varepsilon},(i)}(x) - \int_{\mathbb{R}^d} \varphi_0(x) d\bar{\rho}_0^{\bar{\varepsilon},(i)}(x).
\end{aligned}$$

□

**Remark A.3** (Chain rule in the mollified case). For  $(\rho^{\bar{\varepsilon}}, \mathbf{j}^{\bar{\varepsilon}}) \in \text{CE}_T$  as in Lemma A.2, by the regularity of  $\rho^{\bar{\varepsilon}}$  as well as the continuity equation (23) in conjunction with Remark 2.19, we have

$$\begin{aligned}
\frac{d}{dt} \mathcal{E}_R^{\varepsilon}(\rho_t^{\bar{\varepsilon}}) &= \sum_{i,k=1}^2 \int_{\mathbb{R}^d} (K_R^{\varepsilon,(ik)} * \rho_t^{\bar{\varepsilon},(i)})(x) \partial_t \rho_t^{\bar{\varepsilon},(k)}(x) d\mu(x) \\
&= \frac{1}{2} \sum_{i,k=1}^2 \iint_G \bar{\nabla} (K_R^{\varepsilon,(ik)} * \rho_t^{\bar{\varepsilon},(i)})(x, y) \eta(x, y) dj_t^{\bar{\varepsilon},(k)}(x, y) \\
&= \frac{1}{2} \sum_{i=1}^2 \iint_G \bar{\nabla} \delta_{\rho^{(i)}} \mathcal{E}_R^{\varepsilon}(\rho_t^{\bar{\varepsilon}})(x, y) \eta(x, y) dj_t^{\bar{\varepsilon},(i)}(x, y).
\end{aligned}$$

Now, we are in the position to prove Proposition 3.24.

*Proof of Proposition 3.24.* Recall that choosing  $\rho, \mathbf{j}$  and  $\mathbf{v}$  as above, we have  $(\rho, \mathbf{j}) \in \text{CE}_T$ ,  $|\rho_t'|^p = \mathcal{A}_{m,\beta}(\mu; \rho_t, \mathbf{j}_t)$  for a.e.  $t \in [0, T]$ ,  $\int_0^T \mathcal{A}_{m,\beta}(\mu; \rho_t, \mathbf{j}_t) dt < \infty$  and  $dj_t^{(i)\mu} = ((v_t^{(i)})_+)^{q-1} d\gamma_{1,t}^{(i)} - ((v_t^{(i)})_-)^{q-1} d\gamma_{2,t}^{(i)}$ ,  $(\mu \otimes \mu)$ -a.e. in  $G$  as well as  $dj_t^{(i)\perp} = ((v_t^{(i)\perp})_+)^{q-1} d\gamma_{1,t}^{(i)\perp} - ((v_t^{(i)\perp})_-)^{q-1} d\gamma_{2,t}^{(i)\perp}$ ,  $\zeta_t^{(i)}$ -a.e. in  $G$ , for a.e.  $t \in [0, T]$  and  $i = 1, 2$ . Inserting the definition of  $\tilde{l}$  as well as the connection between  $\mathbf{v}$  and  $\mathbf{j}$ , we can rewrite (42) as

$$\begin{aligned}
\mathcal{E}(\rho_t) - \mathcal{E}(\rho_s) &= \int_s^t \tilde{l}_\rho(v_\tau) [\beta \bar{\nabla} \delta_\rho \mathcal{E}(\rho_\tau)] \\
&= \frac{1}{2} \sum_{i=1}^2 \frac{1}{\beta^{(i)}} \iint_G \beta^{(i)} \bar{\nabla} \delta_{\rho^{(i)}} \mathcal{E}(\rho_\tau) \eta \left[ ((v_\tau^{(i)})_+)^{q-1} d\gamma_{1,\tau}^{(i)} - ((v_\tau^{(i)})_-)^{q-1} d\gamma_{2,\tau}^{(i)} \right. \\
&\quad \left. + ((v_\tau^{(i)\perp})_+)^{q-1} d\gamma_{1,\tau}^{(i)\perp} - ((v_\tau^{(i)\perp})_-)^{q-1} d\gamma_{2,\tau}^{(i)\perp} \right] \\
&= \frac{1}{2} \sum_{i=1}^2 \int_s^t \iint_G \bar{\nabla} \delta_{\rho^{(i)}} \mathcal{E}(\rho_\tau)(x, y) \eta(x, y) dj_\tau^{(i)}(x, y) d\tau.
\end{aligned} \tag{51}$$

We regularize  $(\rho, \mathbf{j})$  to obtain  $(\rho^{\bar{\varepsilon}}, \mathbf{j}^{\bar{\varepsilon}})$  as in Lemma A.2, approximate the energies as in Remark A.1 and integrate in time to obtain

$$\mathcal{E}_R^{\varepsilon}(\rho_t^{\bar{\varepsilon}}) - \mathcal{E}_R^{\varepsilon}(\rho_s^{\bar{\varepsilon}}) = \frac{1}{2} \sum_{i=1}^2 \int_s^t \iint_G \bar{\nabla} \delta_{\rho^{(i)}} \mathcal{E}_R^{\varepsilon}(\rho_\tau^{\bar{\varepsilon}})(x, y) \eta(x, y) dj_\tau^{\bar{\varepsilon},(i)}(x, y) d\tau. \tag{52}$$

Our goal is to pass to the limit as  $\varepsilon \rightarrow 0$ ,  $\bar{\varepsilon} \rightarrow 0$  and  $R \rightarrow \infty$ , which will yield (51). Due to Proposition 3.3, we immediately obtain  $\mathcal{E}_R^{\varepsilon}(\rho_t^{\bar{\varepsilon}}) \rightarrow \mathcal{E}_R^{\varepsilon}(\rho_t)$  as  $\bar{\varepsilon} \rightarrow 0$ . By definition we also have that  $K_R^{\varepsilon,(ik)} \rightarrow \varphi_R K^{(ik)} =: K_R^{(ik)}$ ,

uniformly as  $\varepsilon \rightarrow 0$ . Then, letting  $R \rightarrow \infty$ , we obtain convergence of the left-hand side of (52) to the left-hand side of (51).

It remains to show convergence of the right-hand side. To this end, we use a truncation argument. Let  $\tilde{\varepsilon} > 0$  and set  $N_{\tilde{\varepsilon}} := \overline{B_{\tilde{\varepsilon}^{-1}}} \times \overline{B_{\tilde{\varepsilon}^{-1}}}$ , where  $B_{\tilde{\varepsilon}^{-1}} = \{x \in \mathbb{R}^d : |x| < \tilde{\varepsilon}^{-1}\}$ , and set  $G_{\tilde{\varepsilon}} := \{(x, y) \in G : \tilde{\varepsilon} \leq |x - y|\}$ . Let  $(\varphi_{\tilde{\varepsilon}})_{\tilde{\varepsilon} > 0} \subset C_c^\infty(\mathbb{R}^d \times G; [0, 1])$  be a family of truncation functions, which is s.t., for any  $\tilde{\varepsilon} > 0$ , we have  $\{\varphi_{\tilde{\varepsilon}} = 1\} \supset \overline{B_{\tilde{\varepsilon}^{-1}}} \times G_{\tilde{\varepsilon}} \cap N_{\tilde{\varepsilon}}$ . We add and subtract  $\varphi_{\tilde{\varepsilon}}$  on the right-hand side of (52). Since  $\rho_t^{\tilde{\varepsilon}} \otimes j_t^{\tilde{\varepsilon}} \rightarrow \rho_t \otimes j_t$  for any  $T \in [0, T]$  as  $\tilde{\varepsilon} \rightarrow 0$ , and  $K_R^{\varepsilon, (ik)} \rightarrow K_R^{(ik)}$  uniformly as  $\varepsilon \rightarrow 0$ , we can pass to the limit in  $\tilde{\varepsilon}$  and  $\varepsilon$  for any  $R, \tilde{\varepsilon} > 0$ :

$$\begin{aligned} & \lim_{\substack{\tilde{\varepsilon} \rightarrow 0 \\ \varepsilon \rightarrow 0}} \frac{1}{2} \int_s^t \iint_G \int_{\mathbb{R}^d} \varphi_{\tilde{\varepsilon}}(z, x, y) \left( K_R^{\varepsilon, (ik)}(y, z) - K_R^{\varepsilon, (ik)}(x, z) \right) \eta(x, y) d\rho_\tau^{\tilde{\varepsilon}, (i)}(z) dj_\tau^{\tilde{\varepsilon}, (k)}(x, y) d\tau \\ &= \frac{1}{2} \int_s^t \iint_G \int_{\mathbb{R}^d} \varphi_{\tilde{\varepsilon}}(z, x, y) \left( K_R^{(ik)}(y, z) - K_R^{(ik)}(x, z) \right) \eta(x, y) d\rho_\tau^{(i)}(z) dj_\tau^{(k)}(x, y) d\tau, \end{aligned}$$

for  $i, k = 1, 2$ . By using  $\varphi_{\tilde{\varepsilon}} \leq 1$ , (K2) and Corollary 2.16 in conjunction with (MB1) and (MB2), for any  $\tau \in [s, t]$ , we obtain the bound

$$\begin{aligned} & \left| \frac{1}{2} \iint_G \int_{\mathbb{R}^d} \varphi_{\tilde{\varepsilon}}(z, x, y) \left( K_R^{(ik)}(y, z) - K_R^{(ik)}(x, z) \right) \eta(x, y) d\rho_\tau^{(i)}(z) dj_\tau^{(k)}(x, y) \right| \\ & \leq \frac{1}{2} \iint_G \int_{\mathbb{R}^d} \frac{|K_R^{(ik)}(y, z) - K_R^{(ik)}(x, z)|}{|x - y| \vee |x - y|^p} |x - y| \vee |x - y|^p \eta(x, y) d\rho_\tau^{(i)}(z) dj_\tau^{(k)}(x, y) \\ & \leq L_K M C_\eta^{1/q} \mathcal{A}_{m, \beta}^{1/p}(\mu; \rho_\tau, j_\tau). \end{aligned}$$

Hence, the integral is bounded in time uniformly with respect to  $\tilde{\varepsilon}$  and  $R$ . Since, by definition of  $K_R^{(ik)}$ , we also have uniform boundedness in space, this allows us to apply Lebesgue's dominated convergence theorem to pass to the limit in  $\tilde{\varepsilon}$  and  $R$  to obtain:

$$\begin{aligned} & \lim_{\substack{\tilde{\varepsilon} \rightarrow 0 \\ R \rightarrow \infty}} \frac{1}{2} \int_s^t \iint_G \int_{\mathbb{R}^d} \varphi_{\tilde{\varepsilon}}(z, x, y) \left( K_R^{(ik)}(y, z) - K_R^{(ik)}(x, z) \right) \eta(x, y) d\rho_\tau^{(i)}(z) dj_\tau^{(k)}(x, y) d\tau \\ &= \frac{1}{2} \int_s^t \iint_G \int_{\mathbb{R}^d} \left( K^{(ik)}(y, z) - K^{(ik)}(x, z) \right) \eta(x, y) d\rho_\tau^{(i)}(z) dj_\tau^{(k)}(x, y) d\tau. \end{aligned}$$

The remaining step of the proof is to control the integral involving  $1 - \varphi_{\tilde{\varepsilon}}(z, x, y)$ . To do this, note that for  $\tilde{\varepsilon} > 0$ , we have

$$(\mathbb{R}^d \times G) \setminus \{\varphi_{\tilde{\varepsilon}} = 1\} \subseteq (\overline{B_{\tilde{\varepsilon}^{-1}}}^c \times G) \cup (G \setminus (G_{\tilde{\varepsilon}} \cap N_{\tilde{\varepsilon}})) =: M_{\tilde{\varepsilon}}$$

For  $i, k = 1, 2$ , as before, using (K2) and splitting the contributions, we obtain

$$\begin{aligned} & \left| \iint_G \int_{\mathbb{R}^d} (1 - \varphi_{\tilde{\varepsilon}}(z, x, y)) \left( K_R^{\varepsilon, (ik)}(y, z) - K_R^{\varepsilon, (ik)}(x, z) \right) \eta(x, y) d\rho_t^{\tilde{\varepsilon}, (i)}(z) dj_t^{\tilde{\varepsilon}, (k)}(x, y) \right| \\ & \leq L_K \iiint_{M_{\tilde{\varepsilon}}} |x - y| \vee |x - y|^p 2\eta(x, y) d\rho_t^{\tilde{\varepsilon}, (i)}(z) dj_t^{\tilde{\varepsilon}, (k)}(x, y) \\ & \leq L_K \int_{\overline{B_{\tilde{\varepsilon}^{-1}}}^c} d\rho_t^{\tilde{\varepsilon}, (i)}(z) \iint_G |x - y| \vee |x - y|^p \eta(x, y) dj_t^{\tilde{\varepsilon}, (k)}(x, y) \\ & \quad + L_K \int_{\mathbb{R}^d} d\rho_t^{\tilde{\varepsilon}, (i)}(z) \iint_{G_\delta^c} |x - y| \vee |x - y|^p \eta(x, y) dj_t^{\tilde{\varepsilon}, (k)}(x, y) \\ & \quad + L_K \int_{\mathbb{R}^d} d\rho_t^{\tilde{\varepsilon}, (i)}(z) \iint_{N_{\tilde{\varepsilon}}^c} |x - y| \vee |x - y|^p \eta(x, y) dj_t^{\tilde{\varepsilon}, (k)}(x, y). \end{aligned}$$

Thus, we rid ourselves of the dependence on  $R$ . In the first term we apply Corollary 2.16 together with (MB1) and (MB2) to obtain

$$L_K \int_{\overline{B_{\tilde{\varepsilon}^{-1}}}^c} d\rho_t^{\tilde{\varepsilon}, (i)}(z) \iint_G |x - y| \vee |x - y|^p \eta(x, y) dj_t^{\tilde{\varepsilon}, (k)}(x, y) \leq L_K M C_\eta^{1/q} \mathcal{A}_{m, \beta}^{1/p}(\mu; \rho_t^{\tilde{\varepsilon}}, j_t^{\tilde{\varepsilon}}) \rho_t^{\tilde{\varepsilon}, (i)}(\overline{B_{\tilde{\varepsilon}^{-1}}}^c),$$

hence this term vanishes as  $\tilde{\varepsilon} \rightarrow 0$ . To show that the second term also vanishes as  $\tilde{\varepsilon} \rightarrow 0$ , we assume, without loss of generality, that  $\tilde{\varepsilon} \leq 1$ , which implies that  $|x - y| \vee |x - y|^p = |x - y|$  on  $G_{\tilde{\varepsilon}}^c$ . Applying Lemma 2.14 with  $\Phi(x, y) = |x - y| \mathbb{1}_{G_{\tilde{\varepsilon}}^c}(x, y)$  yields

$$\begin{aligned} & L_K \int_{\mathbb{R}^d} d\rho_t^{\tilde{\varepsilon},(i)}(z) \iint_{G_{\tilde{\varepsilon}}^c} |x - y| \vee |x - y|^p \eta(x, y) d|j_t^{\tilde{\varepsilon},(k)}|(x, y) \\ & \leq L_K M \mathcal{A}_{m,\beta}^{1/p}(\mu; \rho_t^{\tilde{\varepsilon}}, j_t^{\tilde{\varepsilon}}) \sum_{l=1}^2 \left( \iint_{G_{\tilde{\varepsilon}}^c} |x - y|^q \eta(x, y) d(\mu \otimes \mu + \rho^{(l)} \otimes \mu + \mu \otimes \rho^{(l)})(x, y) \right)^{1/q}. \end{aligned}$$

By the local blow-up control Assumption (BC), this also vanishes as  $\tilde{\varepsilon} \rightarrow 0$ . Similarly, for the third term, we apply Lemma 2.14 with  $\Phi(x, y) = |x - y| \mathbb{1}_{N_{\tilde{\varepsilon}}^c}(x, y)$  and obtain

$$\begin{aligned} & L_K \int_{\mathbb{R}^d} d\rho_t^{\tilde{\varepsilon},(i)}(z) \iint_{N_{\tilde{\varepsilon}}^c} |x - y| \vee |x - y|^p \eta(x, y) d|j_t^{\tilde{\varepsilon},(k)}|(x, y) \\ & \leq L_K M C_{\eta} \mathcal{A}_{m,\beta}^{1/p}(\mu; \rho_t^{\tilde{\varepsilon}}, j_t^{\tilde{\varepsilon}}) \left( \mu \left( \overline{B}_{\tilde{\varepsilon}^{-1}}^c \right) + \sum_{l=1}^2 \rho_t^{\tilde{\varepsilon},(l)} \left( \overline{B}_{\tilde{\varepsilon}^{-1}}^c \right) \right). \end{aligned}$$

The uniform  $p$ -th moment bound of the family  $(\rho_t^{\tilde{\varepsilon},(l)})_{\tilde{\varepsilon}}$  implies tightness by the de la Vallée-Poussin theorem and the single measure  $\mu \in \mathcal{M}^+(\mathbb{R}^d)$  is tight as well. Thus, the third term also vanishes as  $\tilde{\varepsilon} \rightarrow 0$ , which concludes the proof.  $\square$

## References

- [1] Martial Agueh. Finsler structure in the  $p$ -wasserstein space and gradient flows. *Comptes Rendus Mathématique*, 350(1):35–40, 2012.
- [2] Luigi Ambrosio, Nicola Fusco, and Diego Pallara. *Functions of bounded variation and free discontinuity problems*. Oxford Mathematical Monographs. Oxford University Press, London, England, March 2000.
- [3] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.
- [4] Nicola J Armstrong, Kevin J Painter, and Jonathan A Sherratt. A continuum approach to modelling cell–cell adhesion. *Journal of theoretical biology*, 243(1):98–113, 2006.
- [5] Rafael Bailo, José A Carrillo, Hideki Murakawa, and Markus Schmidtchen. Convergence of a fully discrete and energy-dissipating finite-volume scheme for aggregation-diffusion equations. *Mathematical Models and Methods in Applied Sciences*, pages 1–36, 2020.
- [6] Rafael Bailo, José A. Carrillo, and Jingwei Hu. Bound-preserving finite-volume schemes for systems of continuity equations with saturation, 2021.
- [7] Rafael Bailo, José Antonio Carrillo, and Jingwei Hu. Fully discrete positivity-preserving and energy-dissipating schemes for aggregation-diffusion equations with a gradient flow structure. *Commun. Math. Sci.*, 18, 2020.
- [8] Julien Barre, Pierre Degond, Diane Peurichard, and Ewelina Zatorska. Modelling pattern formation through differential repulsion. *arXiv preprint arXiv:1906.00704*, 2019.
- [9] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, Jan 2000.
- [10] Judith Berendsen, Martin Burger, and Jan-Frederik Pietschmann. On a cross-diffusion model for multiple species with nonlocal interaction and size exclusion. *Nonlinear Analysis*, 159:10–39, 2017.

- [11] Marianne Bessemoulin-Chatard and Francis Filbet. A finite volume scheme for nonlinear degenerate parabolic equations. *SIAM Journal on Scientific Computing*, 34(5):B559–B583, 2012.
- [12] Giovanni A Bonaschi, José A Carrillo, Marco Di Francesco, and Mark A Peletier. Equivalence of gradient flows and entropy solutions for singular nonlocal interaction equations in 1d. *ESAIM: Control, Optimisation and Calculus of Variations*, 21(2):414–441, 2015.
- [13] Martin Burger, José A Carrillo, Jan-Frederik Pietschmann, and Markus Schmidtchen. Segregation effects and gap formation in cross-diffusion models. *Interfaces and Free Boundaries*, 22(2):175–203, 2020.
- [14] J.A. Carrillo, S. Lisini, G. Savaré, and D. Slepčev. Nonlinear mobility continuity equations and generalized displacement convexity. *Journal of Functional Analysis*, 258(4):1273–1309, 2010.
- [15] José A Carrillo, Marco DiFrancesco, Alessio Figalli, Thomas Laurent, and Dejan Slepčev. Global-in-time weak measure solutions and finite-time aggregation for nonlocal interaction equations. *Duke Mathematical Journal*, 156(2):229 – 271, 2011.
- [16] José A Carrillo, Francis Filbet, and Markus Schmidtchen. Convergence of a finite volume scheme for a system of interacting species with cross-diffusion. *Numerische Mathematik*, 145:473–511, 2020.
- [17] José A Carrillo, Massimo Fornasier, Giuseppe Toscani, and Francesco Vecil. Particle, kinetic, and hydrodynamic models of swarming. In *Mathematical modeling of collective behavior in socio-economic and life sciences*, pages 297–336. Springer, 2010.
- [18] José A Carrillo, Yanghong Huang, and Markus Schmidtchen. Zoology of a nonlocal cross-diffusion model for two species. *SIAM Journal on Applied Mathematics*, 78(2):1078–1104, 2018.
- [19] José A Carrillo, Francois James, Frédéric Lagoutière, and Nicolas Vauchelet. The Filippov characteristic flow for the aggregation equation with mildly singular potentials. *Journal of Differential Equations*, 260(1):304–338, 2016.
- [20] Jose A Carrillo, Hideki Murakawa, Makoto Sato, Hideru Togashi, and Olena Trush. A population dynamics model of cell-cell adhesion incorporating population pressure and density saturation. *Journal of theoretical biology*, 474:14–24, 2019.
- [21] José A. Carrillo, Alina Chertock, and Yanghong Huang. A finite-volume method for nonlinear nonlocal equations with a gradient flow structure. *Communications in Computational Physics*, 17(1):233–258, 2015.
- [22] José Antonio Carrillo, Marco Di Francesco, Antonio Esposito, Simone Fagioli, and Markus Schmidtchen. Measure solutions to a system of continuity equations driven by newtonian nonlocal interactions. *Discrete & Continuous Dynamical Systems*, 40(2):1191–1231, 2020.
- [23] Christopher J. Chang and Kenneth G. Cooper. A practical difference scheme for Fokker-Planck equations. *Journal of Computational Physics*, 6(1):1–16, 1970.
- [24] Lenaïc Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. An interpolating distance between optimal transport and fisher-rao metrics. *Foundations of Computational Mathematics*, 18(1):1–44, 2018.
- [25] Shui-Nee Chow, Wen Huang, Yao Li, and Haomin Zhou. Fokker-Planck equations for a free energy functional or Markov process on a graph. *Archive for Rational Mechanics and Analysis*, 203(3):969–1008, 2012.
- [26] Marco Di Francesco, Antonio Esposito, and Simone Fagioli. Nonlinear degenerate cross-diffusion systems with nonlocal interaction. *Nonlinear Analysis*, 169:94–117, 2018.
- [27] Marco Di Francesco, Antonio Esposito, and Simone Fagioli. Nonlinear degenerate cross-diffusion systems with nonlocal interaction. *Nonlinear Analysis*, 169:94–117, 2018.
- [28] Marco Di Francesco, Antonio Esposito, and Markus Schmidtchen. Many-particle limit for a system of interaction equations driven by newtonian potentials. *Calculus of Variations and Partial Differential Equations*, 60(2):1–44, 2021.

- [29] Marco Di Francesco and Simone Fagioli. A nonlocal swarm model for predators–prey interactions. *Mathematical Models and Methods in Applied Sciences*, 26(02):319–355, 2016.
- [30] Roland L’vovich Dobrushin. Vlasov equations. *Functional Analysis and Its Applications*, 13(2):115–123, 1979.
- [31] Jean Dolbeault, Bruno Nazaret, and Giuseppe Savaré. A new class of transport distances between measures. *Calculus of Variations and Partial Differential Equations*, 34(2):193–231, Jun 2008.
- [32] Maria R D’Orsogna, Yao-Li Chuang, Andrea L Bertozzi, and Lincoln S Chayes. Self-propelled particles with soft-core interactions: patterns, stability, and collapse. *Physical review letters*, 96(10):104302, 2006.
- [33] Matthias Erbar. Gradient flows of the entropy for jump processes. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 50(3):920 – 945, 2014.
- [34] Matthias Erbar and Jan Maas. Gradient flow structures for discrete porous medium equations. *Discrete Contin. Dyn. Syst.*, 34(4):1355–1374, 2014.
- [35] Antonio Esposito, Francesco S Patacchini, André Schlichting, and Dejan Slepčev. Nonlocal-interaction equation on graphs: gradient flow structure and continuum limit. *Archive for Rational Mechanics and Analysis*, pages 1–62, 2021.
- [36] Robert Eymard, Thierry Gallouët, and Raphaële Herbin. Finite volume methods. *Handbook of numerical analysis*, 7:713–1018, 2000.
- [37] Marco Di Francesco and Simone Fagioli. Measure solutions for non-local interaction PDEs with two species. *Nonlinearity*, 26(10):2777–2808, sep 2013.
- [38] François Golse. On the dynamics of large particle systems in the mean field limit. In *Macroscopic and large scale phenomena: coarse graining, mean field limits and ergodicity*, pages 1–144. Springer, 2016.
- [39] Stefano Lisini and Antonio Marigonda. On a class of modified wasserstein distances induced by concave mobility functions defined on bounded intervals. *manuscripta mathematica*, 133(1-2):197–224, June 2010.
- [40] Nadia Loy and Mattia Zanella. Structure preserving schemes for fokker–planck equations with nonconstant diffusion matrices. *Mathematics and Computers in Simulation*, 188:342–362, 2021.
- [41] Jan Maas. Gradient flows of the entropy for finite Markov chains. *J. Funct. Anal.*, 261(8):2250–2292, 2011.
- [42] Alexander Mogilner and Leah Edelstein-Keshet. A non-local model for a swarm. *Journal of mathematical biology*, 38(6):534–570, 1999.
- [43] Léonard Monsaingeon. A new transportation distance with bulk/interface interactions and flux penalization. *Calculus of Variations and Partial Differential Equations*, 60(3):101, Apr 2021.
- [44] Helmut Neunzert. An introduction to the nonlinear boltzmann-vlasov equation. In *Kinetic theories and the Boltzmann equation*, pages 60–110. Springer, 1984.
- [45] Kevin J Painter and Jonathan A Sherratt. Modelling the movement of interacting cell populations. *Journal of theoretical biology*, 225(3):327–339, 2003.
- [46] Lorenzo Pareschi and Mattia Zanella. Structure preserving schemes for nonlinear fokker–planck equations and applications. *Journal of Scientific Computing*, 74(3):1575–1600, 2018.
- [47] André Schlichting and Christian Seis. The Scharfetter–Gummel scheme for aggregation-diffusion equations. *arXiv preprint arXiv:2004.13981*, 2020.
- [48] Chad M Topaz, Andrea L Bertozzi, and Mark A Lewis. A nonlocal continuum model for biological aggregation. *Bulletin of mathematical biology*, 68(7):1601–1623, October 2006.
- [49] Alexandria Volkening, Madeline R Abbott, N Chandra, B Dubois, F Lim, D Sexton, and Bjorn Sandstede. Modeling stripe formation on growing zebrafish tailfins. *Bulletin of mathematical biology*, 82(5):1–33, 2020.

- [50] Alexandria Volkening and Björn Sandstede. Modelling stripe formation in zebrafish: an agent-based approach. *Journal of the Royal Society Interface*, 12(112):20150812, 2015.
- [51] Alexandria Volkening and Björn Sandstede. Iridophores as a source of robustness in zebrafish stripes and variability in danio patterns. *Nature communications*, 9(1):1–14, 2018.