Bull. London Math. Soc. 53 (2021) 1436–1452 doi:10.1112/blms.12509

On the extremal points of the ball of the Benamou–Brenier energy

Kristian Bredies, Marcello Carioni, Silvio Fanzon and Francisco Romero

Abstract

In this paper, we characterize the extremal points of the unit ball of the Benamou–Brenier energy and of a coercive generalization of it, both subjected to the homogeneous continuity equation constraint. We prove that extremal points consist of pairs of measures concentrated on absolutely continuous curves which are characteristics of the continuity equation. Then, we apply this result to provide a representation formula for sparse solutions of dynamic inverse problems with finite-dimensional data and optimal-transport based regularization.

1. *Introduction*

The classical theory of Optimal Transport deals with the problem of efficiently transporting mass from a probability distribution into a target one. In the last 30 years, great advances in the understanding of the underlying theory have been achieved [**3, 45, 48**]. However, only recently these techniques are starting to be applied in order to solve computational problems in a great variety of fields, with logistic problems [**8, 17–19**], crowd dynamics [**36, 37**], image processing [**28, 34, 38, 40, 43, 46, 47**], inverse problems [**15, 31**] and machine learning [**5, 26, 27, 39, 44, 51**] being a few examples.

In this paper, we focus on the so-called Benamou–Brenier formula, which provides an equivalent dynamic formulation of the classical Monge–Kantorovich transport problem [**30**]. Introduced by Benamou and Brenier in [**6**], such formula allows to compute an optimal transport between two probability measures ρ_0 and ρ_1 on a closed bounded domain $\overline{\Omega} \subset \mathbb{R}^d$ through the minimization of the kinetic energy

$$\frac{1}{2}\int_0^1 \int_{\overline{\Omega}} |v_t(x)|^2 d\rho_t(x)\,, \tag{1}$$

among all the pairs (ρ_t, v_t), where $t \mapsto \rho_t$ is a curve of probability measures on $\overline{\Omega}$, $v_t \colon \overline{\Omega} \to \mathbb{R}^d$ is a time-dependent vector field and the pair (ρ_t, v_t) satisfies distributionally the continuity equation

$$\partial_t \rho_t + \operatorname{div}(\rho_t v_t) = 0 \quad \text{subjected to} \quad \rho_{t=0} = \rho_0,\ \rho_{t=1} = \rho_1\,. \tag{2}$$

The interest around the Benamou–Brenier formulation is motivated by its remarkable properties. First, it allows to compute an optimal transport in an efficient way [**6**] by means of a convex reformulation of (1), by introducing the momentum $m_t = \rho_t v_t$. More precisely, denoting by $X := (0,1) \times \overline{\Omega}$ the time-space cylinder, the Benamou–Brenier energy (1) can

Received 30 July 2019; revised 16 October 2020; published online 26 June 2021.

2020 *Mathematics Subject Classification* 52A05, 49N45, 49J45, 35F05 (primary).

KB and SF gratefully acknowledge support by the Christian Doppler Research Association (CDG) and Austrian Science Fund (FWF) through the Partnership in Research project PIR-27 'Mathematical methods for motion-aware medical imaging' and project P 29192 'Regularization graphs for variational imaging'. MC is supported by the Royal Society (Newton International Fellowship NIF\R1\192048 Minimal partitions as a robustness boost for neural network classifiers).

be equivalently defined as a convex functional on the space of bounded Borel measures $\mathcal{M} := \mathcal{M}(X) \times \mathcal{M}(X; \mathbb{R}^d)$ by setting

$$B(\rho, m) := \frac{1}{2} \int_X \left| \frac{dm}{d\rho}(t,x) \right|^2 d\rho(t,x)\,, \tag{3}$$

whenever $(\rho, m) \in \mathcal{M}$ are such that $\rho \geqslant 0$, $m \ll \rho$, and $B := +\infty$ otherwise. With this change of variables, the continuity equation at (2) assumes the form

$$\partial_t \rho + \operatorname{div} m = 0\,. \tag{4}$$

In addition, the dynamic nature of the Benamou–Brenier reformulation of optimal transport is at the core of many recent developments in the fields of PDEs, optimal transport and inverse problems. Indeed, the dynamic formulation allows to endow the space of probability measures with a differentiable structure [**3, 48**], making possible the characterization of differential equations as gradient flows in spaces of measures [**4, 29, 41**] or the derivation of sharp inequalities [**22, 35, 42**]. Moreover, it motivated recent developments in unbalanced optimal transport theory [**20, 21, 32, 33**], that is, when the marginals are arbitrary positive measures. Finally, as the Benamou–Brenier energy provides a description of the optimal flow of the transported mass at each time t, which is a valuable information in applications, it was recently employed as a regularizer for variational inverse problems [**13, 15, 28, 34, 49**] (see also a forthcoming paper by Bredies, Carioni, Fanzon and Walter).

The goal of this paper is to characterize the extremal points of the unit ball of the Benamou–Brenier energy B at (3), and of a coercive version of it, which is obtained by adding the total variation of ρ to B. Both functionals are constrained via the continuity equation (2). Precisely, we introduce the functional

$$J_{\alpha,\beta}(\rho, m) := \beta B(\rho, m) + \alpha \|\rho\|_{\mathcal{M}(X)} \quad \text{subjected to} \quad \partial_t \rho + \operatorname{div} m = 0\,, \tag{5}$$

defined for all $(\rho, m) \in \mathcal{M}$ and $\alpha \geqslant 0$, $\beta > 0$. We then characterize the extremal points of the subset of $\mathcal{M}$ defined by

$$C_{\alpha,\beta} := \{(\rho, m) \in \mathcal{M} : \; J_{\alpha,\beta}(\rho, m) \leqslant 1\}\,.$$

We emphasize that we do not enforce initial conditions to the continuity equation in (5). To be more specific, we prove the following result (see Theorem 6).

THEOREM. *Let $\alpha \geqslant 0$, $\beta > 0$. The extremal points of the set $C_{\alpha,\beta}$ are exactly given by the zero measure* (0,0) *and the pairs of measures* (ρ, m) *such that*

$$\rho = a_\gamma \, dt \otimes \delta_{\gamma(t)}, \quad m = \dot\gamma(t) a_\gamma \, dt \otimes \delta_{\gamma(t)}, \quad a_\gamma = \left(\frac{\beta}{2} \int_0^1 |\dot\gamma(t)|^2 \, dt + \alpha \right)^{-1},$$

where $\gamma : [0,1] \to \overline{\Omega}$ is an absolutely continuous curve with weak derivative $\dot\gamma \in L^2$, and such that $a_\gamma < +\infty$. If $\alpha = 0$, the condition $a_\gamma < +\infty$ is satisfied if and only if γ is not constant.

We therefore show that the extremal points of the set $C_{\alpha,\beta}$ are pairs of measures (ρ, m), with ρ concentrated on some absolutely continuous curve γ in $\overline{\Omega}$, and the density of m with respect to ρ is given by $\dot\gamma$. Note that such conditions are equivalent to the existence of a measurable field $v \colon X \to \mathbb{R}^d$ such that

$$\dot\gamma(t) = v(t, \gamma(t)) \quad \text{for a.e.} \quad t \in (0,1)\,, \tag{6}$$

thus showing that γ is a characteristic associated to the continuity equation at (2) with respect to the field v. We prove the above theorem in Section 3, with the aid of a probabilistic version of the superposition principle for positive measure solutions to the continuity equation (2) on the domain $(0,1) \times \overline{\Omega}$ (see Theorem 7). We mention that the ideas behind such superposition principle are not new, and they were originally introduced in [**1**] for positive measures on $(0,1) \times \mathbb{R}^d$ (see also [**3, 7, 50**]). The result of Theorem 7 allows to decompose any measure solution (ρ, m) of the continuity equation (4) with bounded Benamou–Brenier energy, as superposition of measures concentrated on absolutely continuous characteristics of (4), that is, curves solving (6) with $v = dm/d\rho$. As a consequence, we show that any pair of measures that is not of such a form can be written as a proper convex combination of elements of $C_{\alpha,\beta}$ and thus it is not an extremal point. The opposite inclusion follows from the convexity of the energy and the properties of the continuity equation.

The interest in characterizing extremal points of the Benamou–Brenier energy is not only theoretical. It has been recently shown in [**12**] and [**11**] that in the context of variational inverse problems with finite-dimensional data, the structure of sparse solutions is linked to the extremal points of the unit ball of the regularizer. In the classical theory of variational inverse problems, one aims to solve

$$\min_{u \in \mathcal{U}} R(u) \quad \text{subjected to} \quad Au = y\,, \tag{7}$$

where $\mathcal{U}$ is the target space, R is a convex regularizer, A is a linear observation operator mapping to a finite-dimensional space and y is the observation. It has been empirically observed that the presence of the regularizer R is promoting the existence of sparse solutions, namely minimizers that can be represented as a finite linear combination of simpler atoms. While this effect has been well-understood in the case when $\mathcal{U}$ is finite dimensional, the infinite-dimensional case has been only recently addressed [**11, 12, 23, 24, 52–54**]. In particular, in [**11, 12**], it has been shown that, under suitable assumptions on R and A, there exists a minimizer of (7) that can be represented as a finite linear combination of extremal points of the unit ball of R; namely the atoms forming a sparse solution are extremal points of the ball of the regularizer.

In view of the above discussion, in Section 4 we apply our characterization of the extremal points of the energy $J_{\alpha,\beta}$ at (5) to understand the structure of sparse solutions for inverse problems with such transport energy acting as regularizer. We mention that the analysis is carried out for the case $\alpha > 0$, as the functional $J_{0,\beta}$, corresponding to the rescaled Benamou–Brenier energy, lacks of compactness properties (see Remark 1). We verify that the assumptions needed to apply the representation theorems in [**12**] and [**11**] are satisfied by $J_{\alpha,\beta}$, and consequently we deduce the existence of a minimizer that is given by a finite linear combination of measures concentrated on absolutely continuous curves in $\overline{\Omega}$ (see Theorem 10). As a specific application of Theorem 10, we consider the setting introduced in [**15**], where the regularizer $J_{\alpha,\beta}$ is coupled with a fidelity term that penalizes the distance between the unknown measure ρ_t computed at $t_1, \ldots, t_N \in [0,1]$, and the observation at such times (see Section 4.2). This setting is relevant for applications, such as variational reconstruction in undersampled dynamic MRI. Employing the previous results, we are able to prove the existence of a sparse solution represented with a finite linear combination of measures concentrated on absolutely continuous curves in $\overline{\Omega}$ (see Corollary 12).

To conclude, we mention that characterizing the extremal points for a given regularizer has important consequences in devising algorithms able to compute a sparse solution. Notable examples have been proposed for the total variation regularizer in the space of measures [**10, 16**] using so-called *generalized conditional gradient methods* (or Frank–Wolfe-type algorithms [**25**]). Inspired by the previous methods, and building on the theoretical results obtained in the present paper, we plan to develop numerical algorithms to compute sparse solutions of dynamic inverse problems with the optimal transport energy $J_{\alpha,\beta}$ as a regularizer [**13**] (see also a forthcoming paper by Bredies, Carioni, Fanzon and Walter), effectively providing a numerical counterpart to the theoretical framework established in [**15**]. Finally, we remark that similar results to the ones presented in this paper can be obtained for unbalanced optimal transport energies. This has been recently achieved in [14], by introducing a novel superposition principle for measure solutions to the inhomogeneous continuity equation.

2. *Mathematical setting and preliminaries*

In this section, we give the basic notions about the continuity equation, the Benamou–Brenier energy, and its coercive version $J_{\alpha,\beta}$ anticipated in the introduction. We refer to [3, 6, 15] for a more detailed overview. For measure theoretical notions, we refer to the definitions in [2].

Given a metric space Y, we will denote by $\mathcal{M}(Y)$ (respectively, $\mathcal{M}(Y;\mathbb{R}^d)$) the space of bounded Borel measures (respectively, bounded vector Borel measures) on Y. Similarly, $\mathcal{M}^+(Y)$ and $\mathcal{P}(Y)$ denote the set of bounded positive Borel measures and Borel probability measures on Y, respectively. Let $\Omega \subset \mathbb{R}^d$ be an open, bounded domain with $d \in \mathbb{N}, d \geqslant 1$. Set $X := (0,1) \times \overline{\Omega}$,

$$\mathcal{M} := \mathcal{M}(X) \times \mathcal{M}(X;\mathbb{R}^d)\,,$$

and

$$\mathcal{D} := \{(\rho, m) \in \mathcal{M} \ : \ \partial_t \rho + \operatorname{div} m = 0 \ \text{ in } \ X\}\,,$$

where the solutions of the continuity equation are intended in a distributional sense, that is,

$$\int_X \partial_t \varphi \, d\rho + \int_X \nabla \varphi \cdot dm = 0 \quad \text{for all} \quad \varphi \in C_c^\infty(X)\,. \tag{8}$$

We remark that the above weak formulation includes no-flux boundary conditions for the momentum m on $\partial\Omega$. Also, no initial and final data are prescribed in (8). Moreover, by standard approximation arguments, we can consider in (8) test functions in $C_c^1(X)$ (see [3, Remark 8.1.1]).

We now introduce the Benamou–Brenier energy. For this purpose, define the convex, lower semicontinuous and one-homogeneous map $\Psi\colon \mathbb{R} \times \mathbb{R}^d \to [0,\infty]$ by setting

$$\Psi(t,x) := \begin{cases} \frac{|x|^2}{2t} & \text{if } t > 0\,, \\ 0 & \text{if } t = |x| = 0\,, \\ +\infty & \text{otherwise}\,. \end{cases}$$

The Benamou–Brenier energy $B : \mathcal{M} \to [0,\infty]$ is defined for every pair $(\rho, m) \in \mathcal{M}$ as

$$B(\rho, m) := \int_X \Psi\left(\frac{d\rho}{d\lambda}, \frac{dm}{d\lambda}\right) d\lambda\,, \tag{9}$$

$\lambda \in \mathcal{M}^+(X)$ is such that $\rho, m \ll \lambda$. Since Ψ is one-homogeneous, the above representation of B does not depend on λ. For some fixed $\alpha \geqslant 0$, $\beta > 0$, we consider the following functional

$$J_{\alpha,\beta}(\rho, m) := \begin{cases} \beta B(\rho, m) + \alpha \|\rho\|_{\mathcal{M}(X)} & \text{if } (\rho, m) \in \mathcal{D}, \\ +\infty & \text{otherwise,} \end{cases} \tag{10}$$

where $\|\cdot\|_{\mathcal{M}(X)}$ denotes the total variation norm in $\mathcal{M}(X)$.

Remark 1. Note that in the definition of $J_{\alpha,\beta}$ we add the total variation of ρ to the Benamou–Brenier energy. If $\alpha > 0$, this choice enforces the balls of the energy $J_{\alpha,\beta}$ to be compact in the weak* topology of $\mathcal{M}$ (see Lemma 4). As a consequence, the functional $J_{\alpha,\beta}$ is a natural regularizer for dynamic inverse problems when the initial and final data are not prescribed [15]. We remark that, although in the case $\alpha = 0$ the unit ball of the energy $J_{0,\beta}$

is not compact, we can still characterize its extremal points. However, in this case, due to the lack of coercivity, $J_{0,\beta}$ has limited use as a regularizer for dynamic inverse problems.

For a measure $\rho \in \mathcal{M}(X)$, we say that ρ disintegrates with respect to time if there exists a Borel family of measures $\{\rho_t\}_{t\in[0,1]}$ in $\mathcal{M}(\overline{\Omega})$ such that

$$\int_X \varphi(t,x)\, d\rho(t,x) = \int_0^1 \int_{\overline{\Omega}} \varphi(t,x)\, d\rho_t(x)\, dt \quad \text{ for all } \quad \varphi \in L^1_\rho(X)\,.$$

We denote such disintegration with the symbol $\rho = dt \otimes \rho_t$. Further, we say that a curve of measures $t \in [0,1] \mapsto \rho_t \in \mathcal{M}(\overline{\Omega})$ is narrowly continuous if the map

$$t \mapsto \int_{\overline{\Omega}} \varphi(x)\, d\rho_t(x)$$

is continuous for each fixed $\varphi \in C(\overline{\Omega})$. The family of narrowly continuous curves will be denoted by $C_{\mathrm{w}}([0,1];\mathcal{M}(\overline{\Omega}))$. We also introduce $C_{\mathrm{w}}([0,1];\mathcal{M}^+(\overline{\Omega}))$, as the family of narrowly continuous curves with values into the positive measures on $\overline{\Omega}$.

We now recall several results about B, $J_{\alpha,\beta}$ and measure solutions of the continuity equation (8), which will be useful in the following analysis. For proofs of such results, we refer the interested reader to [**15**, Propositions 2.2, 2.4, 2.6 and Lemmas 4.5, 4.6].

LEMMA 2 (Properties of B). *The functional B defined in* (9) *is convex, positively one-homogeneous and sequentially lower semicontinuous with respect to the weak∗ topology on $\mathcal{M}$. Moreover, it satisfies the following properties.*

(i) $B(\rho,m) \geqslant 0$ *for all* $(\rho,m) \in \mathcal{M}$.
(ii) *If* $B(\rho,m) < +\infty$, *then* $\rho \geqslant 0$ *and* $m \ll \rho$, *that is, there exists a measurable map* $v\colon X \to \mathbb{R}^d$ *such that* $m = v\rho$.
(iii) *If* $\rho \geqslant 0$ *and* $m = v\rho$ *for some* $v\colon X \to \mathbb{R}^d$ *measurable, then*

$$B(\rho,m) = \int_X \Psi(1,v)\, d\rho = \frac{1}{2}\int_X |v|^2\, d\rho\,. \tag{11}$$

LEMMA 3 (Properties of the continuity equation). *Assume that $(\rho,m) \in \mathcal{M}$ satisfies* (8) *and that $\rho \in \mathcal{M}^+(X)$. Then ρ disintegrates with respect to time into $\rho = dt \otimes \rho_t$, where $\rho_t \in \mathcal{M}^+(\overline{\Omega})$ for almost every (a.e.) t. Moreover, $t \mapsto \rho_t(\overline{\Omega})$ is constant, with $\rho_t(\overline{\Omega}) = \rho(X)$ for a.e. $t \in (0,1)$. If in addition $B(\rho,m) < +\infty$, that is,*

$$\int_0^1 \int_{\overline{\Omega}} |v|^2\, d\rho_t(x)\, dt < +\infty\,,$$

where $m = v\rho$ for some $v\colon X \to \mathbb{R}^d$ measurable, then $t \mapsto \rho_t$ belongs to $C_{\mathrm{w}}([0,1];\mathcal{M}^+(\overline{\Omega}))$.

LEMMA 4 (Properties of $J_{\alpha,\beta}$). *Let $\alpha \geqslant 0$, $\beta > 0$. The functional $J_{\alpha,\beta}$ is non-negative, convex, positively one-homogeneous and sequentially lower semicontinuous with respect to weak∗ convergence on $\mathcal{M}$. Assume now $\alpha > 0$. For $(\rho,m) \in \mathcal{M}$ such that $J_{\alpha,\beta}(\rho,m) < +\infty$, we have*

$$\alpha\|\rho\|_{\mathcal{M}(X)} \leqslant J_{\alpha,\beta}(\rho,m)\,, \quad \min(2\alpha,\beta)\|m\|_{\mathcal{M}(X;\mathbb{R}^d)} \leqslant J_{\alpha,\beta}(\rho,m)\,. \tag{12}$$

Moreover, if $\{(\rho^n,m^n)\}_n$ is a sequence in $\mathcal{M}$ such that

$$\sup_n\, J_{\alpha,\beta}(\rho^n,m^n) < +\infty\,,$$

then $\rho^n = dt \otimes \rho^n_t$ *for some* $(t \mapsto \rho^n_t) \in C_{\mathrm{w}}([0,1];\mathcal{M}^+(\overline{\Omega}))$ *and there exists some* $(\rho, m) \in \mathcal{D}$ *with* $\rho = dt \otimes \rho_t$, $\rho_t \in C_{\mathrm{w}}([0,1];\mathcal{M}^+(\overline{\Omega}))$ *such that, up to subsequences,*

$$\begin{cases} (\rho^n, m^n) \overset{*}{\rightharpoonup} (\rho, m) \ \text{ weakly* in } \ \mathcal{M}\,, \\ \rho^n_t \overset{*}{\rightharpoonup} \rho_t \ \text{ weakly* in } \ \mathcal{M}(\overline{\Omega})\,, \ \text{ for every } \ t \in [0,1]\,. \end{cases} \tag{13}$$

3. *Characterization of extremal points*

The aim of this section is to characterize the extremal points of the unit ball of the functional $J_{\alpha,\beta}$ at (10) for all $\alpha \geqslant 0, \beta > 0$, namely, of the convex set

$$C_{\alpha,\beta} := \{(\rho, m) \in \mathcal{M} : \ J_{\alpha,\beta}(\rho, m) \leqslant 1\}\,.$$

To this end, let us first introduce the following set.

Definition 5 (Characteristics). For $\alpha \geqslant 0$, $\beta > 0$, define the set $\mathcal{C}_{\alpha,\beta}$ of all pairs $(\rho, m) \in \mathcal{M}$ such that

$$\rho = a_\gamma \, dt \otimes \delta_{\gamma(t)}\,, \quad m = \dot\gamma(t)\, a_\gamma \, dt \otimes \delta_{\gamma(t)}\,, \quad a_\gamma := \left(\frac{\beta}{2} \int_0^1 |\dot\gamma(t)|^2 \, dt + \alpha \right)^{-1}\,,$$

where $\gamma \in \mathrm{AC}^2([0,1];\mathbb{R}^d)$ satisfies $\gamma(t) \in \overline{\Omega}$ for each $t \in [0,1]$ and $a_\gamma < +\infty$.

We remind that $\mathrm{AC}^2([0,1];\mathbb{R}^d)$ denotes the space of absolutely continuous curves having a weak derivative in L^2. We point out that by definition $a_\gamma > 0$ for all choices of $\alpha \geqslant 0$, $\beta > 0$. Moreover the condition $a_\gamma < +\infty$ is always satisfied if $\alpha > 0$. When $\alpha = 0$, we instead have $a_\gamma < +\infty$ if and only if $\int_0^1 |\dot\gamma(t)|^2 \, dt > 0$, that is, the set $\mathcal{C}_{0,\beta}$ does not contain constant curves.

For the extremal points of $C_{\alpha,\beta}$, we have the following characterization.

Theorem 6. *Let* $\alpha \geqslant 0$, $\beta > 0$ *be fixed. Then*

$$\mathrm{Ext}(C_{\alpha,\beta}) = \{(0,0)\} \cup \mathcal{C}_{\alpha,\beta}\,.$$

The proof of Theorem 6 is postponed to Section 3.2. In order to show the inclusion $\mathrm{Ext}(C_{\alpha,\beta}) \subset \{(0,0)\} \cup \mathcal{C}_{\alpha,\beta}$, we will make use of a superposition principle for measure solutions of the continuity equation (8). This result is not new, and it is proved in [**3**, Chapter 8.2] for the case $\Omega = \mathbb{R}^d$. In Section 3.1, we show that it also holds for bounded closed domains.

3.1. *The superposition principle*

Before stating the superposition principle in $\overline{\Omega}$, we introduce the following notation. Let

$$\Gamma := \left\{\gamma \colon [0,1] \to \mathbb{R}^d : \ \gamma \text{ continuous}\right\}$$

be equipped with the supremum norm, that is, $\|\gamma\|_\infty := \max_{t\in[0,1]} |\gamma(t)|$. For every $t \in [0,1]$ let $e_t \colon \Gamma \to \mathbb{R}^d$ be the evaluation at t, that is, $e_t(\gamma) := \gamma(t)$. Note that e_t is continuous. For a measurable vector field $v \colon (0,1) \times \mathbb{R}^d \to \mathbb{R}^d$, we define the following subset of Γ consisting of AC^2 curves solving the ODE (6) in the sense of Carathéodory:

$$\Gamma_v(\mathbb{R}^d) := \left\{\gamma \in \Gamma : \ \gamma \in \mathrm{AC}^2([0,1];\mathbb{R}^d), \ \dot\gamma(t) = v(t, \gamma(t)) \text{ for a.e. } t \in (0,1)\right\}.$$

Moreover define the set of solutions to the ODE which live inside $\overline{\Omega}$ for all times:

$$\Gamma_v(\overline{\Omega}) := \left\{\gamma \in \Gamma_v(\mathbb{R}^d) : \ \gamma(t) \in \overline{\Omega} \ \text{ for all } \ t \in [0,1]\right\}.$$

The superposition principle for probability solutions to (8) states as follows.

THEOREM 7. *Let $t \in [0,1] \mapsto \rho_t \in \mathcal{P}(\overline{\Omega})$ be a narrowly continuous solution of the continuity equation in the sense of (8), for some measurable $v \colon (0,1) \times \overline{\Omega} \to \mathbb{R}^d$ such that*

$$\int_0^1 \int_{\overline{\Omega}} |v(t,x)|^2 \, d\rho_t(x) \, dt < +\infty \, . \tag{14}$$

Then there exists a probability measure $\sigma \in \mathcal{P}(\Gamma)$ concentrated on $\Gamma_v(\overline{\Omega})$ and such that $\rho_t = (e_t)_{\#}\sigma$ for every $t \in [0,1]$, that is,

$$\int_{\overline{\Omega}} \varphi(x) \, d\rho_t(x) = \int_\Gamma \varphi(\gamma(t)) \, d\sigma(\gamma) \quad \textit{for every} \ \ \varphi \in C(\overline{\Omega}), \ t \in [0,1] \, . \tag{15}$$

Proof. Let $\bar{v} : (0,1) \times \mathbb{R}^d \to \mathbb{R}^d$ be the extension to zero of v to the whole $\mathbb{R}^d$. Similarly, for each $t \in [0,1]$, let $\bar{\rho}_t \in \mathcal{P}(\mathbb{R}^d)$ be the extension to zero of ρ_t in $\mathbb{R}^d$. Note that the pair $(\bar{\rho}, \bar{v}\,\bar{\rho})$ is a solution of the continuity equation in $(0,1) \times \mathbb{R}^d$ in the sense of (8). Moreover $\bar{\rho}$ and $\bar{v}$ satisfy (14) in $(0,1) \times \mathbb{R}^d$. Therefore we can apply [**3**, Theorem 8.2.1] and obtain a probability measure $\sigma \in \mathcal{P}(\Gamma)$ concentrated on $\Gamma_{\bar{v}}(\mathbb{R}^d)$ and such that $\bar{\rho}_t = (e_t)_{\#}\sigma$ for all $t \in [0,1]$, that is,

$$\int_{\mathbb{R}^d} \varphi(x) \, d\bar{\rho}_t(x) = \int_\Gamma \varphi(\gamma(t)) \, d\sigma(\gamma) \quad \text{for every} \ \ \varphi \in C_b(\mathbb{R}^d), \ t \in [0,1] \, . \tag{16}$$

We claim that σ is concentrated on $\Gamma_v(\overline{\Omega})$. In order to show that, partition $\Gamma_{\bar{v}}(\mathbb{R}^d)$ into

$$\Gamma_{\bar{v}}(\mathbb{R}^d) = \Gamma_{\bar{v}}(\overline{\Omega}) \cup A \, ,$$

where

$$A := \left\{ \gamma \in \Gamma_{\bar{v}}(\mathbb{R}^d) : \text{ there exists } \ \hat{t} \in [0,1] \text{ such that } \gamma(\hat{t}) \in \overline{\Omega}^c \right\} .$$

Note that, since $\overline{\Omega}^c$ is open and $v \equiv 0$ in $\overline{\Omega}^c$, the curves in A are constant, so that we can write

$$A = \left\{ \gamma \in \Gamma_{\bar{v}}(\mathbb{R}^d) : \ \gamma(0) \in \overline{\Omega}^c \right\} .$$

From this, it follows that $A \subset e_0^{-1}(\overline{\Omega}^c)$. Moreover, (16) implies $\bar{\rho}_0(\overline{\Omega}^c) = \sigma(e_0^{-1}(\overline{\Omega}^c))$. Therefore, using that $\bar{\rho}_t$ is concentrated on $\overline{\Omega}$, we conclude that $\sigma(A) = 0$, showing that σ is concentrated on $\Gamma_{\bar{v}}(\overline{\Omega})$. Finally, (16) implies (15) since $\bar{\rho}_t$ is supported in $\overline{\Omega}$ and it coincides with ρ_t in $\overline{\Omega}$. Also $\Gamma_{\bar{v}}(\overline{\Omega}) = \Gamma_v(\overline{\Omega})$ by definition of $\bar{v}$, thus concluding the proof. □

3.2. *Proof of Theorem 6*

Let $\alpha \geqslant 0$, $\beta > 0$. We divide the proof into two parts.

Part 1: $\{(0,0)\} \cup \mathcal{C}_{\alpha,\beta} \subset \operatorname{Ext}(C_{\alpha,\beta})$.

We start by showing that $\{(0,0)\} \cup \mathcal{C}_{\alpha,\beta} \subset C_{\alpha,\beta}$. The fact that $(0,0) \in C_{\alpha,\beta}$ follows immediately, since (0,0) solves the continuity equation and $J_{\alpha,\beta}(0,0) = 0$ (by Lemma 2). Consider now $(\rho, m) \in \mathcal{C}_{\alpha,\beta}$. Note that $(\rho, m) \in \mathcal{C}_{\alpha,\beta}$ satisfies the continuity equation in the sense of (8): indeed for every $\varphi \in C_c^1((0,1) \times \overline{\Omega})$, we have

$$\begin{aligned} \int_{(0,1)\times\overline{\Omega}} \partial_t \varphi \, d\rho + \nabla\varphi \cdot dm &= a_\gamma \int_0^1 \partial_t \varphi(t,\gamma(t)) + \nabla\varphi(t,\gamma(t)) \cdot \dot{\gamma}(t) \, dt \\ &= a_\gamma \int_0^1 \frac{d}{dt} \varphi(t,\gamma(t)) \, dt = a_\gamma(\varphi(1,\gamma(1)) - \varphi(0,\gamma(0))) = 0 \end{aligned} \tag{17}$$

since φ is compactly supported in $(0,1)\times\overline{\Omega}$. Moreover, due to the fact that $\rho \geqslant 0$ and $m=\dot\gamma\rho$, we can invoke (11) to obtain

$$J_{\alpha,\beta}(\rho,m) = a_\gamma\left(\frac{\beta}{2}\int_0^1 |\dot\gamma(t)|^2\,dt + \alpha\right) = 1\,, \tag{18}$$

proving that $(\rho,m)\in C_{\alpha,\beta}$.

We now want to show that any $(\rho,m)\in\{(0,0)\}\cup\mathcal{C}_{\alpha,\beta}$ is an extremal point for $C_{\alpha,\beta}$. Hence assume that $(\rho^1,m^1),(\rho^2,m^2)\in C_{\alpha,\beta}$ are such that

$$(\rho,m) = \lambda(\rho^1,m^1) + (1-\lambda)(\rho^2,m^2) \tag{19}$$

for some $\lambda\in(0,1)$. We need to show that $(\rho,m)=(\rho^1,m^1)=(\rho^2,m^2)$. Let $j\in\{1,2\}$. Since (ρ^j,m^j) is such that $J_{\alpha,\beta}(\rho^j,m^j)\leqslant 1$, from (ii) in Lemma 2 we have that $\rho^j\geqslant 0$ and $m^j=v^j\rho^j$ for some Borel field $v^j\colon X\to\mathbb{R}^d$. In particular, if $(\rho,m)=(0,0)$, (19) forces $(\rho^j,m^j)=0$, hence showing that (0,0) is an extremal point of $C_{\alpha,\beta}$.

Let us now consider the case $(\rho,m)\in\mathcal{C}_{\alpha,\beta}$. By (18), we have $J_{\alpha,\beta}(\rho,m)=1$. From (19), convexity of $J_{\alpha,\beta}$, and the fact that $J_{\alpha,\beta}(\rho^j,m^j)\leqslant 1$, $\lambda\in(0,1)$, we conclude

$$J_{\alpha,\beta}(\rho^j,m^j) = 1\,. \tag{20}$$

Since (ρ^j,m^j) solves the continuity equation, $\rho^j\geqslant 0$ and $J_{\alpha,\beta}(\rho^j,m^j)=1$, from Lemma 3 we deduce that $\rho^j = dt\otimes\rho^j_t$ for some narrowly continuous curve $t\mapsto\rho^j_t\in\mathcal{M}^+(\overline{\Omega})$, with $\rho^j_t(\overline{\Omega})$ constant in time. We define $a_j:=\rho^j_0(\overline{\Omega})$ and note that $a_j>0$: Indeed, $a_j=0$ would imply $\rho^j=0$, yielding $J_{\alpha,\beta}(\rho^j,m^j)=J_{\alpha,\beta}(0,0)=0$. This would contradict (20). Now, from condition (19) and uniqueness of the disintegration, we deduce

$$a_\gamma\,\delta_{\gamma(t)} = \lambda\rho^1_t + (1-\lambda)\rho^2_t \qquad \text{for every } t\in[0,1]\,. \tag{21}$$

Since $a_j>0$ (and hence $\rho^j_t\neq 0$), the above equality implies that $\operatorname{supp}\rho^j_t=\{\gamma(t)\}$, that is,

$$\rho^j_t = a_j\,\delta_{\gamma(t)} \qquad \text{for every } t\in[0,1]\,. \tag{22}$$

We now show that $v^j=\dot\gamma$ on $\operatorname{supp}\rho=\operatorname{graph}(\gamma):=\{(t,\gamma(t))\colon t\in(0,1)\}$, that is

$$v^j(t,\gamma(t)) = \dot\gamma(t) \qquad \text{for a.e. } \ t\in(0,1)\,. \tag{23}$$

By assumption, $\partial_t\rho^j+\operatorname{div}m^j=0$ in the sense of (8). Therefore, recalling (22) and the fact that $a_j>0$, we get that for each $\varphi\in C^1_c((0,1)\times\overline{\Omega})$,

$$\begin{aligned} 0 &= \int_0^1 \partial_t\varphi(t,\gamma(t)) + \nabla\varphi(t,\gamma(t))\cdot v^j(t,\gamma(t))\,dt \\ &= \int_0^1 \partial_t\varphi(t,\gamma(t)) + \nabla\varphi(t,\gamma(t))\cdot\dot\gamma(t)\,dt + \int_0^1 \nabla\varphi(t,\gamma(t))\cdot(v^j(t,\gamma(t))-\dot\gamma(t))\,dt \\ &= \int_0^1 \nabla\varphi(t,\gamma(t))\cdot(v^j(t,\gamma(t))-\dot\gamma(t))\,dt\,, \end{aligned} \tag{24}$$

where the last equality follows from (17), since $a_\gamma>0$. Let $\psi\in C^1_c((0,1))$ and define $\varphi(t,x):=x_i\psi(t)$, where $x=(x_1,\dots,x_d)$, so that φ is a test function for (24). By plugging φ into (24), we obtain

$$\int_0^1 \psi(t)(v^j_i(t,\gamma(t))-\dot\gamma_i(t))\,dt = 0 \qquad \text{for all } \ i\in\{1,\dots,d\},\ \ j\in\{1,2\}$$

where v^j_i and $\dot\gamma_i$ are the ith component of v^j and $\dot\gamma$, respectively. This implies that $v^j(t,\gamma(t))=\dot\gamma(t)$ for a.e. $t\in(0,1)$, that is, $v^j=\dot\gamma$ a.e. on $\operatorname{graph}(\gamma)$. With this at hand, by means of (11) we can see that $J_{\alpha,\beta}(\rho^j, m^j) = a_j/a_\gamma$. Since (20) holds, we obtain $a_j = a_\gamma$, thus proving $(\rho, m) = (\rho^j, m^j)$ and hence extremality for (ρ, m) in $C_{\alpha,\beta}$.

Part 2: $\mathrm{Ext}(C_{\alpha,\beta}) \subset \{(0,0)\} \cup \mathcal{C}_{\alpha,\beta}$.

Let $(\rho, m) \in C_{\alpha,\beta}$ be an extremal point. In particular, $J_{\alpha,\beta}(\rho, m) \leqslant 1$ so that by Lemma 2(*ii*), we obtain $\rho \geqslant 0$ and $m = v\rho$ for some Borel field $v\colon X \to \mathbb{R}^d$. Note that by extremality of (ρ, m) and one-homogeneity of $J_{\alpha,\beta}$, we immediately infer that either $J_{\alpha,\beta}(\rho, m) = 0$ or $J_{\alpha,\beta}(\rho, m) = 1$. If $J_{\alpha,\beta}(\rho, m) = 0$, by decomposing (ρ, m) as

$$(\rho, m) = \frac{1}{2}(2\rho, 2m) + \frac{1}{2}(0, 0)$$

and using the extremality of (ρ, m) together with the one-homogeneity of $J_{\alpha,\beta}$, we deduce that $(\rho, m) = (0, 0)$. Thus, we consider the case

$$J_{\alpha,\beta}(\rho, m) = 1\,. \tag{25}$$

Since by definition (ρ, m) solves the continuity equation in the sense of (8) and $J_{\alpha,\beta}(\rho, m) = 1$, we can apply Lemma 3 to obtain that $\rho = a\, dt \otimes \rho_t$ for some narrowly continuous curve $t \mapsto \rho_t \in \mathcal{P}(\overline{\Omega})$, where $a := \rho(X) > 0$.

Claim. $\operatorname{supp} \rho_t$ *is a singleton for each* $t \in [0, 1]$.

Proof of Claim. The hypotheses of Theorem 7 are satisfied, therefore there exists a measure $\sigma \in \mathcal{P}(\Gamma)$ concentrated on $\Gamma_v(\overline{\Omega})$ and such that $\rho_t = (e_t)_{\#}\sigma$ for every $t \in [0, 1]$. Assume by contradiction that there exists a time $\hat{t} \in [0, 1]$ such that $\operatorname{supp} \rho_{\hat{t}}$ is not a singleton. Therefore, we can find a Borel set $E \subset \overline{\Omega}$ such that

$$0 < \rho_{\hat{t}}(E) < 1, \quad 0 < \rho_{\hat{t}}(\overline{\Omega} \setminus E) < 1\,. \tag{26}$$

Define the Borel set

$$A := \{\gamma \in \Gamma :\ \gamma(\hat{t}) \in E\} = e_{\hat{t}}^{-1}(E)\,.$$

By the relation $\rho_t = (e_t)_{\#}\sigma$ and definition of A, we obtain $\rho_{\hat{t}}(E) = \sigma(A)$. Therefore, from (26)

$$0 < \sigma(A) < 1, \quad 0 < \sigma(\Gamma \setminus A) < 1\,. \tag{27}$$

Define

$$\lambda_1 := a\left(\frac{\beta}{2}\int_0^1 \int_A |\dot{\gamma}(t)|^2 \, d\sigma(\gamma)\, dt + \alpha\, \sigma(A)\right),$$

$$\lambda_2 := a\left(\frac{\beta}{2}\int_0^1 \int_{A^c} |\dot{\gamma}(t)|^2 \, d\sigma(\gamma)\, dt + \alpha\, \sigma(A^c)\right),$$

where $A^c := \Gamma \setminus A$. Note that λ_1, λ_2 are well defined (possibly being equal to $+\infty$) as the map

$$L(\gamma) := \int_0^1 |\dot{\gamma}(t)|^2 \, dt \quad \text{if } \ \gamma \in \mathrm{AC}^2([0, 1]; \mathbb{R}^d)\,, \quad L(\gamma) := +\infty \quad \text{otherwise}, \tag{28}$$

is lower semicontinuous on Γ, and hence measurable. Note that

$$\lambda_1 + \lambda_2 = a\left(\frac{\beta}{2}\int_0^1 \int_\Gamma |\dot{\gamma}(t)|^2 \, d\sigma(\gamma)\, dt + \alpha\right) = a\left(\frac{\beta}{2}\int_0^1 \int_\Gamma |v(t, \gamma(t))|^2 \, d\sigma(\gamma)\, dt + \alpha\right), \tag{29}$$

because σ is concentrated on $\Gamma_v(\overline{\Omega})$. Since $v(t,\cdot)$ belongs to $L^2_{\rho_t}(\overline{\Omega};\mathbb{R}^d)$ for a.e. $t \in (0,1)$, by [9, Theorem 3.6.1], we obtain that the representation formula (15) holds for $\varphi(x) := v(t,x)$ and a.e. $t \in (0,1)$, that is,

$$\int_\Gamma |v(t,\gamma(t))|^2 \, d\sigma(\gamma) = \int_{\overline{\Omega}} |v(t,x)|^2 \, d\rho_t(x) \quad \text{for a.e. } t \in (0,1)\,. \tag{30}$$

Therefore, from (11), (25), (29) and (30), we deduce $\lambda_1 + \lambda_2 = J_{\alpha,\beta}(\rho,m) = 1$.

We now proceed with the proof of the claim separately for the cases $\alpha > 0$ and $\alpha = 0$. Suppose first $\alpha > 0$. Note that $\lambda_1, \lambda_2 > 0$ due to (27) and the fact that $a > 0$. Decompose (ρ,m) as

$$(\rho,m) = \lambda_1(\rho^1,m^1) + \lambda_2(\rho^2,m^2)\,, \tag{31}$$

where we defined

$$\rho^j := \frac{a}{\lambda_j}\, dt \otimes (e_t)_\#\sigma_j\,, \quad m^j := \rho^j v\,, \tag{32}$$

for $j = 1,2$, with $\sigma_1 := \sigma \llcorner A$ and $\sigma_2 := \sigma \llcorner A^c$. Note that $\rho^j \in \mathcal{M}^+(X)$, since σ is a positive measure concentrated on $\Gamma_v(\overline{\Omega})$, and $a, \lambda_j > 0$. We now claim that $(\rho^j,m^j) \in C_{\alpha,\beta}$. First, we prove that $\partial_t \rho^j + \operatorname{div} m^j = 0$ in the sense of (8). Let $j = 1$ and fix $\varphi \in C^1_c((0,1)\times\overline{\Omega})$. Since $v(t,\cdot)$ belongs to $L^2_{\rho_t}(\overline{\Omega};\mathbb{R}^d)$ for a.e. $t \in (0,1)$, by [9, Theorem 3.6.1], (15) and the definition of σ_1, we get

$$\begin{aligned}\int_X \partial_t\varphi \, d\rho^1 + \nabla\varphi \cdot dm^1 =& \frac{a}{\lambda_1}\int_0^1 \int_{\overline{\Omega}} \partial_t\varphi(t,x) + \nabla\varphi(t,x)\cdot v(t,x)\, d((e_t)_\#\sigma_1)(x)\, dt \\ =& \frac{a}{\lambda_1}\int_0^1 \int_A \partial_t\varphi(t,\gamma(t)) + \nabla\varphi(t,\gamma(t))\cdot v(t,\gamma(t))\, d\sigma(\gamma)\, dt\,.\end{aligned} \tag{33}$$

Now recall that σ is concentrated on $\Gamma_v(\overline{\Omega})$ and that φ is compactly supported in time, so that

$$\begin{aligned}\int_X \partial_t\varphi \, d\rho^1 + \nabla\varphi \cdot dm^1 =& \frac{a}{\lambda_1}\int_0^1 \int_A \partial_t\varphi(t,\gamma(t)) + \nabla\varphi(t,\gamma(t))\cdot \dot\gamma(t)\, d\sigma(\gamma)\, dt \\ =& \frac{a}{\lambda_1}\int_A \left(\int_0^1 \frac{d}{dt}\varphi(t,\gamma(t))\, dt\right) d\sigma(\gamma) = 0\,.\end{aligned} \tag{34}$$

The calculation for $j = 2$ is similar. Also, by definition of (ρ^j,m^j) and of λ_j, one can perform similar calculations to the ones in (29), (30), and prove that $J_{\alpha,\beta}(\rho^j,m^j) = 1$. Hence $(\rho^j,m^j) \in C_{\alpha,\beta}$. We now claim that $(\rho^1,m^1) \neq (\rho^2,m^2)$. Suppose by contradiction that $(\rho^1,m^1) = (\rho^2,m^2)$. Then in particular $\rho^1 = \rho^2$, so that by (32) we get

$$\frac{(e_t)_\#\sigma_1}{\lambda_1} = \frac{(e_t)_\#\sigma_2}{\lambda_2} \quad \text{for a.e. } t \in (0,1)\,. \tag{35}$$

As (ρ^j,m^j) are solutions of the continuity equation and $J_{\alpha,\beta}(\rho^j,m^j) = 1$, from Lemma 3 it follows that the maps $t \mapsto (e_t)_\#\sigma_j$ are narrowly continuous. In particular, (35) holds for each $t \in [0,1]$. However, by (27) and by definition of A, σ_1, σ_2, we have

$$[(e_{\hat t})_\#\sigma_1](E) = \sigma(A) > 0\,, \qquad [(e_{\hat t})_\#\sigma_2](E) = \sigma(\emptyset) = 0\,,$$

which contradicts (35). Therefore $(\rho^1,m^1) \neq (\rho^2,m^2)$, which shows that the decomposition (31) is non-trivial. This is a contradiction, since we are assuming that (ρ,m) is an extremal point for $C_{\alpha,\beta}$. Thus the claim follows.

Suppose now that $\alpha = 0$ and define the set

$$Z := \left\{ \gamma \in \Gamma : \int_0^1 |\dot\gamma(t)|^2 \, dt = 0 \right\}.$$

Note that Z is measurable, due to the measurability of the map L at (28). We claim that $\sigma(Z) = 0$. In order to prove that, let $Z^c := \Gamma \setminus Z$ and define the measures $\sigma_Z := \sigma \llcorner Z$, $\sigma_{Z^c} := \sigma \llcorner Z^c$, so that $\sigma = \sigma_Z + \sigma_{Z^c}$. Recalling that $\rho_t = (e_t)_\# \sigma$ for all $t \in [0,1]$, we can decompose

$$(\rho, m) = \frac{1}{2}(\rho^1, m^1) + \frac{1}{2}(\rho^2, m^2), \tag{36}$$

where

$$\rho^1 := a \, dt \otimes (e_t)_\# \sigma_{Z^c}, \qquad m^1 := v\rho^1,$$

$$\rho^2 := a \, dt \otimes (e_t)_\# \sigma_{Z^c} + 2a \, dt \otimes (e_t)_\# \sigma_Z, \qquad m^2 := v\rho^2.$$

Let $j = 1, 2$. Note that $\rho^j \in \mathcal{M}^+(X)$ since σ is a positive measure concentrated on $\Gamma_v(\overline{\Omega})$ and $a > 0$. Following similar computation as (33) and (34), we infer that (ρ^j, m^j) solves the continuity equation in the sense of (8). Moreover, by definition of Z and the fact that σ is concentrated on $\Gamma_v(\overline{\Omega})$, we obtain

$$\begin{aligned} \int_0^1 \int_{Z^c} |v(t,\gamma(t))|^2 \, d\sigma(\gamma) \, dt &= \int_0^1 \int_{Z^c} |\dot\gamma(t)|^2 \, d\sigma(\gamma) \, dt \\ &= \int_{Z^c} \int_0^1 |\dot\gamma(t)|^2 \, dt \, d\sigma(\gamma) \\ &= \int_\Gamma \int_0^1 |\dot\gamma(t)|^2 \, dt \, d\sigma(\gamma) = \int_0^1 \int_\Gamma |v(t,\gamma(t))|^2 \, d\sigma(\gamma) \, dt, \end{aligned} \tag{37}$$

where we employed Fubini's Theorem, which holds due to the measurability of the map L and the identity (30), the latter implying boundedness of the last term in (37). By (37) and arguing as in (29) and (30), it is immediate to check that $J_{0,\beta}(\rho^j, m^j) = J_{0,\beta}(\rho, m)$. Recalling (25), we then obtain $(\rho^j, m^j) \in C_{0,\beta}$. As (ρ, m) is an extremal point of $C_{0,\beta}$, from (36), we deduce that $(\rho^1, m^1) = (\rho^2, m^2)$ and thus $dt \otimes (e_t)_\# \sigma_Z = 0$. In particular, there exists $\hat t \in [0,1]$ such that $(e_{\hat t})_\# \sigma_Z = 0$. Hence for every $E \subset \Gamma$ measurable, by the positivity of σ, we have $\sigma_Z(E) \leqslant (e_{\hat t})_\# \sigma_Z(e_{\hat t}(E)) = 0$, implying that $\sigma_Z = 0$. By (27) and the definition of λ_1 and λ_2, we conclude that $\lambda_1, \lambda_2 > 0$. With this property established, the claim that $\operatorname{supp} \rho_t$ is a singleton for each $t \in [0,1]$ follows by repeating the same arguments of the case $\alpha > 0$, employing the decomposition of (ρ, m) as in (31). □

We have shown that for each $t \in [0,1]$, $\operatorname{supp} \rho_t$ is a singleton. We now conclude the proof of Theorem 6. Since $\rho_t \in \mathcal{P}(\overline{\Omega})$, the latter implies the existence of a curve $\gamma\colon [0,1] \to \overline{\Omega}$ such that $\rho_t = \delta_{\gamma(t)}$ for each $t \in [0,1]$. We will now prove that $\gamma \in \mathrm{AC}^2([0,1]; \mathbb{R}^d)$. By narrow continuity of $t \mapsto \rho_t$, we have that the map $t \mapsto \varphi(\gamma(t))$ is continuous for all $\varphi \in C(\overline{\Omega})$. By testing against the coordinate functions $\varphi(x) := x_i$, we obtain continuity for γ. Consider now $\varphi(t,x) := \xi(t)\eta(x)$ with $\xi \in C_c^\infty((0,1))$, $\eta \in C^1(\overline{\Omega})$. Note that the scalar map $t \mapsto \eta(\gamma(t))$ is continuous. Moreover, by testing the continuity equation $\partial_t \rho + \operatorname{div}(v\rho) = 0$ against φ, we get

$$\int_0^1 \xi'(t) \, \eta(\gamma(t)) \, dt = -\int_0^1 \xi(t) \, \nabla\eta(\gamma(t)) \cdot v(t,\gamma(t)) \, dt,$$

which implies that the distributional derivative of the map $t \mapsto \eta(\gamma(t))$ is given by

$$t \mapsto \nabla\eta(\gamma(t)) \cdot v(t,\gamma(t)).$$

We now remark that the above map belongs to $L^2((0,1))$, since

$$\int_0^1 |\nabla\eta(\gamma(t))\cdot v(t,\gamma(t))|^2\,dt \leqslant \|\nabla\eta\|_\infty \int_0^1 |v(t,\gamma(t))|^2\,dt \leqslant C\,\|\nabla\eta\|_\infty J_{\alpha,\beta}(\rho,m) < +\infty\,.$$

Therefore, $t\mapsto\eta(\gamma(t))$ belongs to $\mathrm{AC}^2([0,1])$ for every fixed $\eta\in C^1(\overline{\Omega})$. By choosing $\eta(x):=x_i$, $i=1,\dots,d$, we conclude that $\gamma\in\mathrm{AC}^2([0,1];\mathbb{R}^d)$. Since $\partial_t\rho+\operatorname{div}(v\rho)=0$, we can repeat the same argument employed to prove (23), and infer

$$v(t,\gamma(t))=\dot\gamma(t)\quad\text{a.e. in}\quad(0,1)\,. \tag{38}$$

From (38) and the fact that $\rho=a\,dt\otimes\delta_{\gamma(t)}$, $m=v\rho$, we then conclude $m=\dot\gamma\rho$. As $a>0$, we can apply (11) to compute

$$J_{\alpha,\beta}(\rho,m)=a\left(\frac{\beta}{2}\int_0^1|\dot\gamma(t)|^2\,dt+\alpha\right). \tag{39}$$

Recalling that $J_{\alpha,\beta}(\rho,m)=1$ (see (25)), from (39) we conclude that $a=a_\gamma$ with $a_\gamma<+\infty$. Therefore (ρ,m) belongs to $\mathcal{C}_{\alpha,\beta}$ according to Definition 5, and the proof of Theorem 6 is concluded.

4. *Application to sparse representation for inverse problems with optimal transport regularization*

In this section, we deal with the problem of reconstructing a family of time-dependent Radon measures given a finite number of observations. To be more specific, let H be a finite-dimensional Hilbert space and $A: C_{\mathrm{w}}([0,1];\mathcal{M}(\overline{\Omega}))\to H$ be a linear continuous operator, where continuity is understood in the following sense: given a sequence $(t\mapsto\rho_t^n)$ in $C_{\mathrm{w}}([0,1];\mathcal{M}(\overline{\Omega}))$, we require that

$$\rho_t^n \overset{*}{\rightharpoonup} \rho_t \quad\text{weakly* in}\quad \mathcal{M}(\overline{\Omega})\quad\text{for all}\quad t\in[0,1]\quad\text{implies}\quad A\rho^n\to A\rho\quad\text{in } H\,, \tag{40}$$

where, with a little abuse of notation, we will denote by ρ^n both the curve $t\mapsto\rho_t^n$, as well as the measure $\rho^n:=dt\otimes\rho_t^n$.

For some given data $y\in H$, we aim to reconstruct a solution $\rho\in C_{\mathrm{w}}([0,1];\mathcal{M}(\overline{\Omega}))$ to the dynamic inverse problem

$$A\rho=y\,. \tag{41}$$

For $\alpha>0$ and $\beta>0$, we regularize the above inverse problem by means of the energy $J_{\alpha,\beta}$ defined in (10), following the approach in [**15**]. In practice, upon introducing the space

$$\widetilde{\mathcal{M}}:=C_{\mathrm{w}}([0,1];\mathcal{M}(\overline{\Omega}))\times\mathcal{M}(X;\mathbb{R}^d)\,,$$

we consider the Tikhonov functional $G:\widetilde{\mathcal{M}}\to\mathbb{R}\cup\{+\infty\}$ defined as

$$G(\rho,m)=J_{\alpha,\beta}(\rho,m)+F(A\rho)\,, \tag{42}$$

where $F: H\to\mathbb{R}\cup\{+\infty\}$ is a given fidelity functional for the data y, which is assumed to be convex, lower semicontinuous and bounded from below. Additionally, we assume that G is proper. We then replace (41) by

$$\min_{(\rho,m)\in\widetilde{\mathcal{M}}} G(\rho,m)\,. \tag{43}$$

Remark 8. Two common choices for the fidelity term F in the case $H=\mathbb{R}^k$ are, for example:

(i) $F(x)=I_{\{y\}}(x)$ for a given $y\in\mathbb{R}^k$ that forces the constraint $A\rho=y$;
ii) $F(x)=\frac{1}{2}\|x-y\|_2^2$ that recovers a classical l^2 penalization.

REMARK 9. Under the above assumptions on A and F, problem (43) admits a solution. Indeed, since G is proper, any minimizing sequence $\{(\rho^n, m^n)\}_n$ is such that $\{G(\rho^n, m^n)\}_n$ is bounded. As F is bounded from below and $J_{\alpha,\beta} \geqslant 0$, we deduce that $\{J_{\alpha,\beta}(\rho^n, m^n)\}_n$ is bounded. Therefore, Lemma 4 implies that (ρ^n, m^n) converges (up to subsequences) to some $(\rho, m) \in \widetilde{\mathcal{M}}$, in the sense of (13). By weak$*$ lower semicontinuity of $J_{\alpha,\beta}$ in $\mathcal{M}$ (see Lemma 4) and by (40) together with the lower-semicontinuity of F, we infer that (ρ, m) solves (43).

It is well known that the presence of a finite-dimensional constraint in an inverse problem, such as (41), promotes sparsity in the reconstruction. This observation has been recently made rigorous in [**12**] and [**11**], where it has been shown that the atoms of a sparse minimizer are the extremal points of the ball of the regularizers. In Theorem 6, we provided a characterization for the extremal points of the ball of $J_{\alpha,\beta}$. Therefore, specializing the above-mentioned results to our setting yields the following characterization theorem for sparse minimizers to (43).

THEOREM 10. *Let $\alpha, \beta > 0$. There exists a minimizer $(\hat\rho, \hat m) \in \widetilde{\mathcal{M}}$ of (43) that can be represented as*

$$(\hat\rho, \hat m) = \sum_{i=1}^{p} c_i \,(\rho^i, m^i)\,, \tag{44}$$

where $p \leqslant \dim(H)$, $c_i > 0$, $\sum_{i=1}^p c_i = J_{\alpha,\beta}(\hat\rho, \hat m)$, and

$$\rho^i = a_{\gamma_i}\, dt \otimes \delta_{\gamma_i(t)}\,, \quad m^i = \dot\gamma_i\, \rho^i\,,$$

where $\gamma_i \in \mathrm{AC}^2([0,1];\mathbb{R}^d)$ with $\gamma(t) \in \overline{\Omega}$ for each $t \in [0,1]$, and $a_{\gamma_i}^{-1} := \frac{\beta}{2}\int_0^1 |\dot\gamma_i|^2\, dt + \alpha$.

In other words, the above theorem ensures the existence of a minimizer of (43) which is a finite linear combination of measures concentrated on the graphs of AC^2-trajectories contained in $\overline{\Omega}$. In Section 4.1, we give a proof of Theorem 10, and we conclude the paper with Section 4.2, where we apply the sparsity result of Theorem 10 to dynamic inverse problems with optimal transport regularization, following the approach of [**15**].

4.1. *Proof of Theorem* 10

As already mentioned, the proof is an immediate consequence of Theorem 6 and a particular case of [**11**, Corollary 2] (see also [**11**, Theorem 1]). Before proceeding with the proof, for the reader's convenience, we recall [**11**, Corollary 2]. The definitions appearing in the statement below will be briefly recalled in the proof of Theorem 10.

THEOREM 11 [**11**]. *Let $\mathcal{U}$ be a locally convex space, H be a finite-dimensional Hilbert space, $R : \mathcal{U} \to [-\infty, +\infty]$, $F : H \to [-\infty, +\infty]$ be convex, and $A : \mathcal{U} \to H$ be linear. Consider the variational problem*

$$\inf_{u \in \mathcal{U}} R(u) + F(Au)\,. \tag{45}$$

Suppose that the set of minimizers of (45), denoted by S, is non-empty. Additionally, assume that there exists $\hat u \in \mathrm{Ext}(S)$ such that the set

$$\overline{C} = \{u \in \mathcal{U} : R(u) \leqslant R(\hat u)\} \tag{46}$$

is linearly closed, *the* lineality space *of $\overline{C}$ is $\{(0,0)\}$ and $\inf_{u\in\mathcal{U}} R(u) < R(\hat u)$. Then, exactly one of the following conditions holds.*

(i) *$\hat u$ is a convex combination of at most $\dim(H)$ extremal points of $\overline{C}$.*

(ii) *$\hat{u}$ is as a convex combination of at most* $\dim(H)-1$ *points, which are either extremal points of* $\overline{C}$*, or belong to an extreme ray of* $\overline{C}$*.*

Proof of Theorem 10. We just need to verify that we can apply Theorem 11 to the variational problem (43). So, we choose $\mathcal{U}=\widetilde{M}$, $R=J_{\alpha,\beta}$ and F and A satisfying the assumptions stated above. Let S be the set of solutions to (43).

First, note that in Remark 9 we have already shown that the set of minimizers for (43) is non-empty, so that $S\neq\emptyset$. Moreover S is compact with respect to the weak$*$ topology. Indeed, given a sequence (ρ^n,m^n) in S we can use Lemma 4 to extract a subsequence (not relabelled) such that $(\rho^n,m^n)\overset{*}{\rightharpoonup}(\rho,m)$ in $\mathcal{M}$ and $\rho_t^n\overset{*}{\rightharpoonup}\rho_t$ in $\mathcal{M}(\overline{\Omega})$ for every $t\in[0,1]$. Using the sequential lower semicontinuity of $J_{\alpha,\beta}$ with respect to weak$*$ convergence combined with the continuity of A (according to (40)) and the lower semicontinuity of F, we obtain $(\rho,m)\in S$. We conclude that S is sequentially weakly$*$ compact and hence weakly$*$ compact, due to the metrizability of the weak$*$ convergence on bounded sets. Finally note that S is convex due to the convexity of F and $J_{\alpha,\beta}$ (Lemma 4). By Krein–Milman's Theorem, we then infer the existence of a $(\hat{\rho},\hat{m})\in\operatorname{Ext}(S)$.

The lineality space of $\overline{C}$ is defined as $\operatorname{lin}(\overline{C})=\operatorname{rec}(\overline{C})\cap(-\operatorname{rec}(\overline{C}))$, where $\operatorname{rec}(\overline{C})$ is the recession cone of $\overline{C}$ defined as the set of all $(\rho,m)\in\mathcal{U}$ such that $\overline{C}+\mathbb{R}_+(\rho,m)\subset\overline{C}$. Hence, from the coercivity of $J_{\alpha,\beta}$ in Lemma 4, it is immediate to conclude that $\operatorname{lin}(\overline{C})=\{(0,0)\}$. Moreover, $\overline{C}$ is linearly closed if the intersection of $\overline{C}$ with every line is closed. It is easy to verify that, as $\overline{C}$ is weakly$*$ closed (Remark 9), it is also linearly closed. Finally, the assumption $\inf_{(\rho,m)\in\widetilde{M}}J_{\alpha,\beta}(\rho,m)<J_{\alpha,\beta}(\hat{\rho},\hat{m})$ is satisfied whenever $(\hat{\rho},\hat{m})\neq 0$, as in this case $J_{\alpha,\beta}(\hat{\rho},\hat{m})>0$, while $\inf_{(\rho,m)\in\widetilde{M}}J_{\alpha,\beta}(\rho,m)=0$. Hence, the hypotheses of Theorem 11 for the functional (42) are verified. Note also that $\overline{C}$ does not contain extreme rays. In order to prove that, we first recall that a ray of $\overline{C}$ is any set of the form $r_{p,v}=\{p+tv:t>0\}$ for $p,v\in\overline{C}$, $v\neq 0$. An extreme ray of $\overline{C}$ is a ray $r_{p,v}$ such that for every segment intersecting $r_{p,v}$, the whole segment is contained in $r_{p,v}$. Due to the coercivity of $J_{\alpha,\beta}$ in Lemma 4, it is immediate to see that $\overline{C}$ contains no rays and thus no extreme rays. Hence, from either of the conclusions (i) and (ii) in Theorem 11, we deduce that there exists a minimizer $(\hat{\rho},\hat{m})\in\mathcal{M}$ of (43) that can be represented as

$$(\hat{\rho},\hat{m})=\sum_{i=1}^{p}c_i(\rho_i,m_i)\,, \tag{47}$$

where $(\rho_i,m_i)\in\operatorname{Ext}(C_{\alpha,\beta})$, $p\leqslant\dim(H)$, $c_i>0$ and $\sum_{i=1}^{p}c_i=J_{\alpha,\beta}(\hat{\rho},\hat{m})$. We remark that if $(\hat{\rho},\hat{m})=0$, the assumption $\inf_{(\rho,m)\in\widetilde{M}}J_{\alpha,\beta}(\rho,m)<J_{\alpha,\beta}(\hat{\rho},\hat{m})$ in Theorem 11 is not satisfied, but the representation (47) holds trivially. Using the characterization of extremal points in Theorem 6 and (47), we obtain an explicit sparse representation for solutions of (43) and the proof is achieved. □

4.2. *Dynamic inverse problems*

Theorem 10 provides a representation formula for sparse solutions of (43) that holds for every A and F satisfying the above-stated hypotheses. A relevant choice for A and F is proposed in [**15**] as a model for dynamic inverse problems: in particular, the authors apply their framework to variational reconstruction in undersampled dynamic MRI. In what follows we make an explicit choice of F and A in order to apply Theorem 10 to a special case of the framework in [**15**], namely the case of discrete time sampling, and finite-dimensionality of the data for each sampled time.

To be more specific, consider a discretization of the interval [0,1] in N points $t_1<t_2<\ldots<t_N$ and assume that we want to reconstruct an element of $C_{\mathrm{w}}([0,1];\mathcal{M}(\overline{\Omega}))$, by only making observations at the time instants $t_1,\ldots,t_N$. To this aim, let H_{t_i} be a family of

finite-dimensional Hilbert spaces and introduce the product space $\mathcal{H} := \times_{i=1}^N H_{t_i}$, normed by $\|y\|_{\mathcal{H}}^2 := \sum_{i=1}^N \|y_i\|_{H_{t_i}}^2$. Let $A_{t_i} : \mathcal{M}(\overline{\Omega}) \to H_{t_i}$ be linear operators, which are assumed to be weak$*$ continuous for each $i = 1, \dots, N$. For a given observation $(y_{t_1}, \dots, y_{t_N}) \in \mathcal{H}$, consider the problem of finding $\rho \in C_{\rm w}([0,1]; \mathcal{M}(\overline{\Omega}))$ such that

$$A_{t_i}\rho_{t_i} = y_{t_i} \quad \text{for each} \quad i = 1, \dots, N\,.$$

Following [**15**], we regularize the above problem by

$$\min_{(\rho,m)\in\widetilde{\mathcal{M}}} J_{\alpha,\beta}(\rho,m) + \frac{1}{2}\sum_{i=1}^N \|A_{t_i}\rho_{t_i} - y_{t_i}\|_{H_{t_i}}^2\,. \tag{48}$$

In order to recast the above problem into the form (43), let $A : C_{\rm w}([0,1]; \mathcal{M}(\overline{\Omega})) \to \mathcal{H}$ be the linear operator defined by

$$A\rho := (A_{t_1}\rho_{t_1}, \dots, A_{t_N}\rho_{t_N})\,.$$

Note that A is continuous in the sense of (40), due to the assumptions on A_{t_i}. We can then equivalently rewrite (48) as

$$\min_{(\rho,m)\in\widetilde{\mathcal{M}}} J_{\alpha,\beta}(\rho,m) + \frac{1}{2}\|A\rho - y\|_{\mathcal{H}}^2\,. \tag{49}$$

In this way, we recover a problem of the type of (43), where $F(x) := \frac{1}{2}\|x-y\|_{\mathcal{H}}^2$. Note that F is convex, lower semicontinuous and bounded from below. Moreover, the functional in (49) is proper, since $J_{\alpha,\beta}(0,0) = 0$. Hence, we can apply Theorem 10 to conclude the following result.

Corollary 12. *Let $\alpha, \beta > 0$. There exists a minimizer $(\hat\rho, \hat m) \in \widetilde{\mathcal{M}}$ of (48) that can be represented as*

$$(\hat\rho, \hat m) = \sum_{i=1}^p c_i\,(\rho^i, m^i)\,, \tag{50}$$

where $p \leqslant \dim(\mathcal{H}) = \sum_{i=1}^N \dim(H_i)$, $c_i > 0$, $\sum_{i=1}^p c_i = J_{\alpha,\beta}(\hat\rho,\hat m)$, and

$$\rho^i = a_{\gamma_i}\,dt \otimes \delta_{\gamma_i(t)}\,, \quad m^i = \dot\gamma_i\,\rho^i\,,$$

where $\gamma_i \in \mathrm{AC}^2([0,1];\mathbb{R}^d)$ with $\gamma(t) \in \overline{\Omega}$ for each $t \in [0,1]$, and $a_{\gamma_i}^{-1} := \frac{\beta}{2}\int_0^1 |\dot\gamma_i|^2\,dt + \alpha$.

Remark 13. The upper bound $p \leqslant \sum_{i=1}^N \dim(H_i)$ in the representation formula (50) might not be optimal. However, a careful analysis of the faces of the ball of the Benamou–Brenier energy, possibly under additional assumptions on the operator A and fidelity term F, could be needed to substantiate such conjecture. We leave this question open for future research.

Acknowledgements. We thank the referee for the useful suggestions provided, particularly for encouraging us to include the case $\alpha = 0$ in the characterization of Theorem 6, which was previously missing. The Institute of Mathematics and Scientific Computing, to which KB, SF and FR are affiliated, is a member of NAWI Graz (http://www.nawigraz.at/en/). The authors KB, SF and FR are further members of/associated with BioTechMed Graz (https://biotechmedgraz.at/en/).

References

1. L. Ambrosio, 'Transport equation and Cauchy problem for BV vector fields', *Invent. Math.* 158 (2004) 227–260.

2. L. Ambrosio, N. Fusco and D. Pallara, *Functions of bounded variation and free discontinuity problems* (Oxford Science Publications, Oxford, 2000).

3. L. Ambrosio, N. Gigli and G. Savaré. *Gradient flows: in metric spaces and in the space of probability measures* (Birkhäuser, Basel, 2005).
4. L. Ambrosio, N. Gigli and G. Savaré, 'Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below', *Invent. Math.* 195 (2014) 289–391.
5. M. Arjovsky, S. Chintala and L. Bottou, 'Wasserstein generative adversarial networks', *Proceedings of the 34th International Conference on Machine Learning* 70 (2017) 214–223.
6. J.-D. Benamou and Y. Brenier, 'A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem', *Numer. Math.* 84 (2000) 375–393.
7. P. Bernard, 'Young measures, superposition and transport', *Indiana Univ. Math. J.* 57 (2008) 247–275.
8. M. Bernot, V. Caselles and J.-M. Morel, *Optimal transportation networks*, Lecture Notes in Mathematics (Springer, Berlin, 2009).
9. V. I. Bogachev, *Measure theory* (Springer, Berlin Heidelberg, 2007).
10. N. Boyd, G. Schiebinger and B. Recht, 'The alternating descent conditional gradient method for sparse inverse problems', *SIAM J. Optim.* 27 (2017) 616–639.
11. C. Boyer, A. Chambolle, Y. De Castro, V. Duval, F. De Gournay and P. Weiss, 'On representer theorems and convex regularization', *SIAM J. Optim.* 29 (2019) 1260–1281.
12. K. Bredies and M. Carioni, 'Sparsity of solutions for variational inverse problems with finite-dimensional data', *Calc. Var. Partial Differential Equations* 59 (2020) 14.
13. K. Bredies, M. Carioni, S. Fanzon and F. Romero, 'A generalized conditional gradient method for dynamic inverse problems with optimal transport regularization'. Preprint, 2020, arXiv:2012.11706.
14. K. Bredies, M. Carioni and S. Fanzon, 'A superposition principle for the inhomogeneous continuity equation with Hellinger-Kantorovich-regular coefficients'. Preprint, 2020, arXiv:2007.06964.
15. K. Bredies and S. Fanzon, 'An optimal transport approach for solving dynamic inverse problems in spaces of measures', *ESAIM Math. Model. Numer. Anal.* 54 (2020) 2351–2382.
16. K. Bredies and H. K. Pikkarainen, 'Inverse problems in spaces of measures', *ESAIM Control Optim. Calc. Var.* 19 (2013) 190–218.
17. G. Buttazzo and F. Santambrogio, 'A mass transportation model for the optimal planning of an urban region', *SIAM Rev.* 51 (2009) 593–610.
18. G. Carlier and I. Ekeland, 'The structure of cities', *J. Global Optim.* 29 (2004) 371–376.
19. G. Carlier and I. Ekeland, 'Equilibrium structure of a bidimensional asymmetric city', *Nonlinear Anal. Real World Appl.* 8 (2007) 725–748.
20. L. Chizat, G. Peyré, B. Schmitzer and F.-X. Vialard, 'An interpolating distance between optimal transport and Fisher–Rao metrics', *Found. Comput. Math.* 18 (2018) 1–44.
21. L. Chizat, G. Peyré, B. Schmitzer and F.-X. Vialard, 'Unbalanced optimal transport: Dynamic and Kantorovich formulations', *J. Funct. Anal.* 274 (2018) 3090–3123.
22. D. Cordero-Erausquin, B. Nazaret and C. Villani, 'A mass-transportation approach to sharp Sobolev and Gagliardo–Nirenberg inequalities', *Adv. Math.* 182 (2004) 307–332.
23. I. Daubechies, M. Defrise and C. De Mol, 'An iterative thresholding algorithm for linear inverse problems with a sparsity constraint', *Comm. Pure Appl. Math.* 57 (2004) 1413–1457.
24. A. Flinth and P. Weiss, 'Exact solutions of infinite dimensional total-variation regularized problems', *Inf. Inference* 8 (2018) 407–443.
25. M. Frank and P. Wolfe, 'An algorithm for quadratic programming', *Naval Res. Logist.* 3 (1956) 95–110.
26. C. Frogner, C. Zhang, H. Mobahi, M. Araya-Polo and T. Poggio, 'Learning with a Wasserstein loss', *Proceedings of the 28th International Conference on Neural Information Processing Systems* 2 (2015) 2053–2061, 2015.
27. A. Genevay, G. Peyré and M. Cuturi, 'Learning generative models with Sinkhorn divergences', *Proc. Mach. learn. Res.* 84 (2018) 1608–1617.
28. R. Hug, E. Maitre and N. Papadakis, 'Multi-physics optimal transportation and image interpolation', *ESAIM Math. Model. Numer. Anal.* 49 (2015) 1671–1692.
29. R. Jordan, D. Kinderlehrer and F. Otto, 'The variational formulation of the Fokker–Planck equation', *SIAM J. Math. Anal.* 29 (1998) 1–17.
30. L. Kantorovitch, 'On the translocation of masses', *Comptes Rendus (Doklady) de l'Académie des Sciences de l'URSS* 37 (1942) 199–201.
31. J. Karlsson and A. Ringh, 'Generalized Sinkhorn iterations for regularizing inverse problems using optimal mass transport', *SIAM J. Imaging Sci. SIAM* 10 (2017) 1935–1962.
32. S. Kondratyev, L. Monsaingeon and D. Vorotnikov, 'A new optimal transport distance on the space of finite Radon measures', *Adv. Differential Equations* 21 (2016) 1117–1164.
33. M. Liero, A. Mielke and G. Savaré, 'Optimal Entropy-Transport problems and a new Hellinger–Kantorovich distance between positive measures', *Invent. Math.* 211 (2018) 969–1117.
34. J. Maas, M. Rumpf, C. Schönlieb and S. Simon, 'A generalized model for optimal transport of images including dissipation and density modulation', *ESAIM Math. Model. Numer. Anal.* 49 (2015) 1745–1769.
35. F. Maggi and C. Villani, 'Balls have the worst best Sobolev inequalities', *J. Geom. Anal.* 15 (2005) 83–121.
36. B. Maury, A. Roudneff-Chupin and F. Santambrogio, 'A macroscopic crowd motion model of gradient flow type', *Math. Models Methods Appl. Sci.* 20 (2010) 1787–1821.
37. B. Maury, A. Roudneff-Chupin, F. Santambrogio and J. Venel, 'Handling congestion in crowd motion modeling', *Netw. Heterog. Media* 6 (2011) 485–519.

38. L. Métivier, R. Brossier, Q. Mérigot, E. Oudet and J. Virieux, 'An optimal transport approach for seismic tomography: application to 3D full waveform inversion', *Inverse Problems* 32 (2016) 115008.
39. G. Montavon, K.-R. Müller and M. Cuturi, 'Wasserstein training of restricted Boltzmann machines', *Adv. Neural Inf. Process. Syst.* 29 (2016) 3718–3726.
40. K. Ni, X. Bresson, T. Chan and S. Esedoglu, 'Local histogram based segmentation using the Wasserstein distance', *Int. J. Comput. Vis.* 84 (2009) 97–111.
41. F. Otto, 'The geometry of dissipative evolution equations: the porus medium equation', *Comm. Partial Differential Equations* 26 (2001) 101–174.
42. F. Otto and C. Villani, 'Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality', *J. Funct. Anal.* 173 (2000) 361–400.
43. N. Papadakis, 'Optimal transport for image processing', Habilitation Thesis, 2015, https://hal.archives-ouvertes.fr/tel-01246096/file/hdr_hal2.pdf.
44. G. Patrini, R. van den Berg, P. Forré, M. Carioni, S. Bhargav, M. Welling, T. Genewein and F. Nielsen, 'Sinkhorn autoencoders', *Proc. Mach. Learn. Res.* 115 (2020) 733–743.
45. G. Peyré and M. Cuturi, 'Computational optimal transport: with applications to data science', *Found. Trends Mach. Learn.* 11 (2019) 355–607.
46. G. Peyré, J. M. Fadili and J. Rabin, 'Wasserstein active contours', *ICIP'*12 (2012) 2541–2544.
47. F. Pitie and A. Kokaram, 'The linear Monge-Kantorovitch linear colour mapping for example-based colour transfer', 4*th European Conference on Visual Media Production* (IET, London, 2007) 1–9.
48. F. Santambrogio, *Optimal transport for applied mathematicians* (Birkhäuser, Basel, 2015).
49. B. Schmitzer, K. P. Schafers and B. Wirth, 'Dynamic cell imaging in PET with optimal transport regularization', *IEEE Trans. Med. Imaging* 39 (2020) 1626–1635.
50. E. Stepanov and D. Trevisan, 'Three superposition principles: Currents, continuity equations and curves of measures', *J. Funct. Anal.* 272 (2017) 1044–1103.
51. I. Tolstikhin, O. Bousquet, S. Gelly and B. Schoelkopf, 'Wasserstein Auto-Encoders', 6*th International Conference on Learning Representations* (*ICLR*), May 2018.
52. M. Unser, 'A unifying representer theorem for inverse problems and machine learning', *Found. Comput. Math.* (2020), https://doi.org/10.1007/s10208-020-09472-x.
53. M. Unser and J. Fageot, 'Native Banach spaces for splines and variational inverse problems', Preprint, 2019, 1904.10818.
54. M. Unser, J. Fageot and J. P. Ward, 'Splines are universal solutions of linear inverse problems with generalized TV regularization', *SIAM Rev.* 59 (2017) 769–793.

Kristian Bredies, Silvio Fanzon and Francisco Romero
Institute of Mathematics and Scientific Computing
University of Graz
Heinrichstraße 36
Graz 8010
Austria

kristian.bredies@uni-graz.at
silvio.fanzon@uni-graz.at
francisco.romero-hinrichsen@uni-graz.at

Marcello Carioni
Department of Applied Mathematics and Theoretical Physics
University of Cambridge
Wilberforce Road
Cambridge CB3 0WA
United Kingdom

mc2250@maths.cam.ac.uk