A note on the differentiability of the Hellinger-Kantorovich distances

Florentine Catharina Fleißner *

Abstract

This paper will deal with differentiability properties of the class of Hellinger-Kantorovich distances $H_{\Lambda,\Sigma} (\Lambda, \Sigma > 0)$ which was recently introduced on the space $\mathcal{M}(\mathbb{R}^d)$ of finite nonnegative Radon measures. The $L^1$-a.e.-differentiability of

$$t \mapsto H_{\Lambda,\Sigma}(\mu_t, \nu)^2,$$

for $\nu \in \mathcal{M}(\mathbb{R}^d)$ and absolutely continuous curves $(\mu_t)_t$ in $(\mathcal{M}(\mathbb{R}^d), H_{\Lambda,\Sigma})$, will be examined and the corresponding derivatives will be computed. The characterization of absolutely continuous curves in $(\mathcal{M}(\mathbb{R}^d), H_{\Lambda,\Sigma})$ will be refined.

1 Introduction

Recently, a new class of distances on the space $\mathcal{M}(\mathbb{R}^d)$ of finite nonnegative Radon measures was established by three independent teams [8, 9, 7, 3, 4]. We will follow the presentation of these distances by Liero, Mielke and Savaré [8, 9] who named it Hellinger-Kantorovich distances. The class of Hellinger-Kantorovich distances $H_{\Lambda,\Sigma} (\Lambda, \Sigma > 0)$ is based on the conversion of one measure into another one (possibly having different total mass) by means of transport and creation / annihilation of mass. The parameters $\Lambda$ and $\Sigma$ serve as weightings of the transport part and the mass creation/annihilation part respectively. To be more precise, the square $H_{\Lambda,\Sigma}(\mu_1, \mu_2)^2$ of the Hellinger-Kantorovich distance $H_{\Lambda,\Sigma}$ between two measures $\mu_1, \mu_2 \in \mathcal{M}(\mathbb{R}^d)$ on $\mathbb{R}^d$ corresponds to

$$\min \left\{ \sum_{i=1}^{2} \frac{4}{\Sigma} \int_{\mathbb{R}^d} (\sigma_i \log \sigma_i - \sigma_i + 1) \, d\mu_i + \int_{\mathbb{R}^d \times \mathbb{R}^d} c_{\Lambda,\Sigma}(|x_1 - x_2|) \, d\gamma : \gamma \in \mathcal{M}(\mathbb{R}^d \times \mathbb{R}^d), \gamma_i \ll \mu_i \right\},$$

(1.1)

with entropy cost functions $\frac{4}{\Sigma} (\sigma_i \log \sigma_i - \sigma_i + 1)$,

$$\sigma_i := \frac{d\gamma_i}{d\mu_i} \quad (\gamma_i \text{ i-th marginal of } \gamma),$$

(1.2)
and transportation cost function
\[ c_{\Lambda, \Sigma}(d) := \begin{cases} 
-\frac{8}{\Sigma} \log(\cos(\sqrt{\Sigma/(4\Lambda)}d)) & \text{if } d < \pi \sqrt{\Lambda/\Sigma}, \\
+\infty & \text{if } d \geq \pi \sqrt{\Lambda/\Sigma}. 
\end{cases} \] (1.3)

There exists an optimal plan \( \gamma \) for the Logarithmic Entropy-Transport problem (1.1) (cf. Thm. 3.3 in [9]), and if \( \mu_1 \) is absolutely continuous with respect to the Lebesgue measure and \( \gamma \) is such optimal plan, then there exists a Borel optimal transport mapping \( t : \mathbb{R}^d \to \mathbb{R}^d \) so that \( \gamma \) takes the form
\[ \gamma = (I \times t)_{\#} \gamma_1 = (I \times t)_{\#} \sigma_1 \mu_1 \] (cf. Thm. 4.5 in [6] and Thm. 6.6 in [9]). We refer the reader to ([9], Cor. 7.14, Thms. 7.17 and 7.20) for the proofs that \( \mathcal{H}_k(\Lambda, \Sigma) \) defined via the Logarithmic Entropy-Transport problem (1.1) indeed represents a distance on the space of finite nonnegative Radon measures and that \((\mathcal{M}(\mathbb{R}^d), \mathcal{H}_k(\Lambda, \Sigma))\) is a complete metric space. Furthermore, the Hellinger-Kantorovich distance \( \mathcal{H}_k(\Lambda, \Sigma) \) metrizes the weak topology on \( \mathcal{M}(\mathbb{R}^d) \) in duality with continuous and bounded functions (cf. Thm. 7.15 in [9]) and can be interpreted as weighted infimal convolution of the Kantorovich-Wasserstein distance and the Hellinger-Kakutani distance. A representation formula à la Benamou-Brenier which can be proved for \( \mathcal{H}_k(\Lambda, \Sigma) \) (cf. ([9], Thm. 8.18; [8], Thm. 3.6(v))) justifies this interpretation:
\[ \mathcal{H}_k(\Lambda, \Sigma)(\mu_1, \mu_2)^2 = \min \left\{ \int_0^1 \int_{\mathbb{R}^d} (\Lambda |v_t|^2 + \Sigma |w_t|^2) \, d\mu_t \, dt : \mu_1^{(\mu,v,w)} \sim \mu_2 \right\} \tag{1.4} \]
where \( \mu_1^{(\mu,v,w)} \sim \mu_2 \) means that \( \mu : [0,1] \to \mathcal{M}(\mathbb{R}^d) \) is a continuous curve connecting \( \mu(0) = \mu_1 \) and \( \mu(1) = \mu_2 \) and satisfying the continuity equation with reaction
\[ \partial_t \mu_t = -\Lambda \text{div}(v_t \mu_t) + \Sigma \nu_t \mu_t, \tag{1.5} \]
governed by Borel functions \( v : (0,1) \times \mathbb{R}^d \to \mathbb{R}^d \) and \( w : (0,1) \times \mathbb{R}^d \to \mathbb{R} \) with
\[ \int_0^1 \int_{\mathbb{R}^d} (\Lambda |v_t|^2 + \Sigma |w_t|^2) \, d\mu_t \, dt < +\infty, \tag{1.6} \]
in duality with \( C^\infty \)-functions with compact support in \((0.1) \times \mathbb{R}^d\), i.e.
\[ \int_0^1 \int_{\mathbb{R}^d} (\partial_t \psi(t,x) + \Lambda \langle \nabla \psi(t,x), v(t,x) \rangle + \Sigma \psi(t,x)w(t,x)) \, d\mu_t(x) \, dt = 0 \tag{1.7} \]
for all \( \psi \in C^\infty_c((0,1) \times \mathbb{R}^d) \).

The class of such continuous curves \( \mu \) satisfying (1.5), (1.6) for some Borel vector field \((v, w)\) coincides with the class of absolutely continuous curves \((\mu_t)_{t \in [0,1]}\) in \((\mathcal{M}(\mathbb{R}^d), \mathcal{H}_k(\Lambda, \Sigma))\) with square-integrable metric derivatives (cf. Thms. 8.16 and 8.17 in [9], see Sect. 3 in this paper).
In order to deepen our understanding of a distance, it is always worth studying its differentiability along absolutely continuous curves (e.g. see Chap. 8 in [1] for the corresponding analysis of the Kantorovich-Wasserstein distance on the space of Borel probability measures with finite second order moments). The present paper addresses this issue for the class of Hellinger-Kantorovich distances on the space of finite nonnegative Radon measures. Clearly, if $(\mu_t)_{t \in [0,1]}$ is an absolutely continuous curve in $(\mathcal{M}(\mathbb{R}^d), \mathcal{H}_{\Lambda, \Sigma})$ and $\nu \in \mathcal{M}(\mathbb{R}^d)$, then the mapping

\[ t \mapsto \mathcal{H}_{\Lambda, \Sigma}(\mu_t, \nu)^2 \]  

is $\mathcal{L}^1$-a.e. differentiable. A natural question that arises is the one of the concrete form of the corresponding derivatives. We will answer this question for absolutely continuous curves with square-integrable metric derivatives (for which such characterization (1.5) is available), refine that characterization by providing more information on $(v, w)$ (see Prop. 3.1) and determine

\[ \frac{d}{dt} \mathcal{H}_{\Lambda, \Sigma}(\mu_t, \nu)^2 \]  

at $\mathcal{L}^1$-a.e. $t \in [0,1]$ (see Thm. 3.4). This piece of work can be viewed as continuation of Sect. 2 in the author’s paper [5] constituting a starting point for the study of differentiability properties of the Hellinger-Kantorovich distances. Therein, we identified elements of the Fréchet subdifferential of mappings

\[ t \mapsto -\mathcal{H}_{\Lambda, \Sigma}((I + tv)\#(1 + tR)^2 \mu_0, \nu)^2 \]

at $t = 0$, for $\mu_0, \nu \in \mathcal{M}(\mathbb{R}^d)$ and bounded Borel functions $v: \mathbb{R}^d \to \mathbb{R}^d$ and $R: \mathbb{R}^d \to \mathbb{R}$. That subdifferential calculus was an essential ingredient for our Minimizing Movement approach to a class of scalar reaction-diffusion equations [5] substantiating their gradient-flow-like structure in the space of finite nonnegative Radon measures endowed with the Hellinger-Kantorovich distance $\mathcal{H}_{\Lambda, \Sigma}$.

The proof in [9] that absolutely continuous curves in $(\mathcal{M}(\mathbb{H}), \mathcal{H}_{\Lambda, \Sigma})$ with square-integrable metric derivatives are characterized via ((1.5), (1.6)) was carried out only for $\mathbb{H} = \mathbb{R}^d$, endowed with usual scalar product $\langle \cdot, \cdot \rangle$ and norm $| \cdot | := \sqrt{\langle \cdot, \cdot \rangle}$, but according to a comment at the beginning of Sect. 8.5 in [9], it should be possible to prove such characterization result in a more general setting. We would like to remark that also our computation of the derivatives (1.9) may be adapted for general separable Hilbert spaces $\mathbb{H}$.

Our plan for the paper is to give an equivalent characterization of the Hellinger-Kantorovich distances in Sect. 2 and to perform the computation of the derivatives (1.9) in Sect. 3.

2 Optimal transportation on the cone

According to ([8], Sect. 3) and ([9], Sect. 7), the Logarithmic Entropy-Transport problem (1.1) translates into a problem of optimal transportation on the geometric cone $\mathcal{C}$ on $\mathbb{R}^d$, see (2.7),
The geometric cone is defined as the quotient space
\[ \mathcal{C} := \mathbb{R}^d \times [0, +\infty) / \sim \] (2.1)
with
\[ (x_1, r_1) \sim (x_2, r_2) \iff r_1 = r_2 = 0 \text{ or } r_1 = r_2, \ x_1 = x_2 \] (2.2)
and is endowed with a class of distances \( d_{\varepsilon,\Lambda,\Sigma} \) \((\Lambda, \Sigma > 0)\). The vertex \( \phi \) \((\text{for } r = 0)\) and \([x, r]\) \((\text{for } x \in \mathbb{R}^d \text{ and } r > 0)\) denote the corresponding equivalence classes and
\[ d_{\varepsilon,\Lambda,\Sigma}([x_1, r_1], [x_2, r_2])^2 := \frac{4}{\Sigma} \left( r_1^2 + r_2^2 - 2r_1r_2 \cos \left( \left( \sqrt{\Sigma / 4\Lambda} |x_1 - x_2| \right) \right) \right)^2 \] (2.3)
(where \( \phi \) is identified with \([\bar{x}, 0]\) for some \( \bar{x} \in \mathbb{R}^d \)). The distance \( d_{\varepsilon,\Lambda,\Sigma} \) gives rise to an optimal transport problem on the cone and therewith to an extended quadratic Kantorovich-Wasserstein distance \( W_{\varepsilon,\Lambda,\Sigma} \) on the space \( M_2(\mathcal{C}) \) of finite nonnegative Radon measures on \( \mathcal{C} \) with finite second order moments, i.e. \( \int_{\mathcal{C}} d_{\varepsilon,\Lambda,\Sigma}([x, r], \alpha)^2 \, d\alpha([x, r]) < +\infty \). The extended Kantorovich-Wasserstein distance \( W_{\varepsilon,\Lambda,\Sigma}(\alpha_1, \alpha_2) \) between two measures \( \alpha_1, \alpha_2 \in M_2(\mathcal{C}) \) is equal to \(+\infty\) if \( \alpha_1(\mathcal{C}) \neq \alpha_2(\mathcal{C}) \) and is given by
\[ W_{\varepsilon,\Lambda,\Sigma}(\alpha_1, \alpha_2)^2 := \min \left\{ \int_{\mathcal{C} \times \mathcal{C}} d_{\varepsilon,\Lambda,\Sigma}([x_1, r_1], [x_2, r_2])^2 \, d\beta \mid \beta \in M(\alpha_1, \alpha_2) \right\} \] (2.4)
if \( \alpha_1(\mathcal{C}) = \alpha_2(\mathcal{C}) \), with \( M(\alpha_1, \alpha_2) \) being the set of finite nonnegative Radon measures on \( \mathcal{C} \times \mathcal{C} \) whose first and second marginals coincide with \( \alpha_1 \) and \( \alpha_2 \). Every measure \( \alpha \in M_2(\mathcal{C}) \) on the cone is assigned a measure \( \mathfrak{h}\alpha \in M(\mathbb{R}^d) \) on \( \mathbb{R}^d \),
\[ \mathfrak{h}\alpha := x_\#(r^2\alpha), \] (2.5)
with \((x, r) : \mathcal{C} \to \mathbb{R}^d \times [0, +\infty) \) defined as
\[ (x, r)([x, r]) := (x, r) \text{ for } [x, r] \in \mathcal{C}, \ r > 0, \ (x, r)(\phi) := (\bar{x}, 0), \] (2.6)
which means \( \int_{\mathbb{R}^d} \phi(x) \, d(\mathfrak{h}\alpha) = \int_{\mathbb{R}^d} r^2\phi(x) \, d\alpha \) for all continuous and bounded functions \( \phi : \mathbb{R}^d \to \mathbb{R} \) \((\text{short } \phi \in C^0_b(\mathbb{R}^d))\). Please note that the mapping \( \mathfrak{h} : M_2(\mathcal{C}) \to M(\mathbb{R}^d) \) is not injective.

Now, an equivalent characterization of the Hellinger-Kantorovich distance \( HK_{\Lambda,\Sigma} \) is given by the transportation problems
\[ HK_{\Lambda,\Sigma}(\mu_1, \mu_2)^2 = \min \left\{ \left. W_{\varepsilon,\Lambda,\Sigma}(\alpha_1, \alpha_2)^2 \right| \alpha_i \in M_2(\mathcal{C}), \ \mathfrak{h}\alpha_i = \mu_i \right\} \] (2.7)
\[ = \min \left\{ \left. W_{\varepsilon,\Lambda,\Sigma}(\alpha_1, \alpha_2)^2 + \frac{4}{\Sigma} \sum_{i=1}^2 (\mu_i - \mathfrak{h}\alpha_i)(\mathbb{R}^d)^2 \right| \alpha_i \in M_2(\mathcal{C}), \ \mathfrak{h}\alpha_i \leq \mu_i \right\}, \] (2.8)
cf. Probl. 7.4, Thm. 7.6, Lem. 7.9, Thm. 7.20 in [9]. Every solution $\gamma \in \mathcal{M}(\mathbb{R}^d \times \mathbb{R}^d)$ to the Logarithmic Entropy-Transport problem (1.1) induces a solution $\beta \in \mathcal{M}(\mathcal{C} \times \mathcal{C})$ to ((2.8), (2.4)): if $\gamma$ is an optimal plan for (1.1) with Lebesgue decompositions 

$$\mu_i = \rho_i \gamma_i + \mu_i^\perp,$$  

then

$$\beta := ([x_1, \sqrt{\rho_1(x_1)}], [x_2, \sqrt{\rho_2(x_2)}]) \# \gamma \in \mathcal{M}(\mathcal{C} \times \mathcal{C})$$

(2.10)
is an optimal plan for the transport problem ((2.8), (2.4)) (cf. (9, Thm. 7.20(iii))). Furthermore, if $\beta \in \mathcal{M}(\mathcal{C} \times \mathcal{C})$ is a solution to ((2.8), (2.4)) or a solution to ((2.7), (2.4)) (which exists by ([9], Thm. 7.6)), then

$$\beta\left(\{(x_1, r_1), [x_2, r_2]\} \in \mathcal{C} \times \mathcal{C}: \ r_1, r_2 > 0, \ |x_1 - x_2| > \pi \sqrt{\Lambda/\Sigma}\right) = 0,$$  

(2.11)

cf. ([9], Lem. 7.19)).

Finally, we show how to construct geodesics in $(\mathcal{C}, d_{\mathcal{C},\Lambda,\Sigma})$ (cf. Sect. 8.1 in [9]) as they will play an important role in our analysis of (1.9). We suppose that $|x_1 - x_2| \leq \pi \sqrt{\Lambda/\Sigma}, \ r_1, r_2 > 0$, and search for functions $\mathcal{R} : [0, 1] \to [0, +\infty)$ and $\theta : [0, 1] \to [0, 1]$ so that the curve $\eta : [0, 1] \to \mathcal{C}$ defined as $\eta(s) := [x_1 + \theta(s)(x_2 - x_1)], \mathcal{R}(s))$ is a (constant speed) geodesic connecting $[x_1, r_1]$ and $[x_2, r_2]$, which means $d_{\mathcal{C},\Lambda,\Sigma}(\eta(s), \eta(t)) = |s - t|d_{\mathcal{C},\Lambda,\Sigma}([x_1, r_1], [x_2, r_2])$ for all $s, t \in [0, 1]$. If $x_1 = x_2$, we set $\theta \equiv 0$. We note that

$$d_{\mathcal{C},\Lambda,\Sigma}(\eta(s), \eta(t))^2 = |z(s) - z(t)|^2_{\mathcal{C}},$$

(2.12)

where $z : [0, 1] \to \mathbb{C}$ is the curve in the complex plane $\mathbb{C}$ defined as

$$z(s) := \frac{2}{\sqrt{\Sigma}}\mathcal{R}(s) \exp \left(\frac{i\theta(s)}{\sqrt{\Sigma/4\Lambda}} |x_1 - x_2| \right),$$

(2.13)

and $|\cdot|_{\mathbb{C}}$ denotes the absolute value for complex numbers. Thus, if $z$ is a geodesic in the complex plane between $z_1 := \frac{2}{\sqrt{\Sigma}}r_1$ and $z_2 := \frac{2}{\sqrt{\Sigma}}r_2 \exp \left(\frac{i\sqrt{\Sigma/4\Lambda}}{\sqrt{\Sigma/4\Lambda}} |x_1 - x_2| \right)$, i.e.

$$z(s) = z_1 + s(z_2 - z_1) \quad \text{for all } s \in [0, 1],$$

(2.14)

then $\eta$ is a geodesic in $(\mathcal{C}, d_{\mathcal{C},\Lambda,\Sigma})$ between $[x_1, r_1]$ and $[x_2, r_2]$. This condition yields an appropriate choice for $\mathcal{R} : [0, 1] \to [0, +\infty)$ and $\theta : [0, 1] \to [0, 1]$, and it is not difficult to see that they are both smooth functions, their first derivatives satisfy

$$\frac{4}{\Sigma} (\mathcal{R}'(s))^2 + \frac{1}{\Lambda} \mathcal{R}(s)^2(\theta'(s))^2 |x_1 - x_2|^2 = d_{\mathcal{C},\Lambda,\Sigma}([x_1, r_1], [x_2, r_2])^2 \quad \text{for all } s \in (0, 1),$$

(2.15)

and they are right differentiable at $s = 0$. We obtain a geodesic from $[x_1, r_1]$ to the vertex $\mathfrak{o}$ by setting $\theta \equiv 0$ and $\mathcal{R}(s) := (1 - s)r_1$ and identifying $\mathfrak{o}$ with $[x_1, 0]$. Also in this case, (2.15) holds good.

---

1 according to Lem. 2.3 in [9], there exist Borel functions $\rho_i : \mathbb{R}^d \to [0, +\infty)$ and nonnegative finite Radon measures $\mu_i^\perp \in \mathcal{M}(\mathbb{R}^d)$, $\mu_i^\perp \perp \gamma_i$, so that (2.9) holds good.
3 Differentiability results

We fix $\Lambda, \Sigma > 0$ and examine the behaviour of the Hellinger-Kantorovich distance $H_{\Lambda, \Sigma}$ along absolutely continuous curves.

Let $(\mu_t)_{t \in [0,1]}$ be an absolutely continuous curve in $\mathcal{M}(\mathbb{R}^d, H_{\Lambda, \Sigma})$ with square-integrable metric derivative, i.e. the limit

$$|\mu'_t| := \lim_{h \to 0} \frac{H_{\Lambda, \Sigma}(\mu_{t+h}, \mu_t)}{|h|}$$

exists for $\mathcal{L}^1$-a.e. $t \in (0,1)$, the function $t \mapsto |\mu'_t|$ which is called metric derivative of $(\mu_t)_t$ belongs to $L^2((0,1))$ and

$$H_{\Lambda, \Sigma}(\mu_s, \mu_t) \leq \int_s^t |\mu'_r| \, dr \quad \text{for all } 0 \leq s \leq t \leq 1$$

(cf. Def. 1.1.1 and Thm. 1.1.2 in [1]). According to Thms. 8.16 and 8.17 in [9], there exists a Borel vector field $(v, w) : (0, 1) \times \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}$ so that the continuity equation with reaction

$$\partial_t \mu_t = -\text{div}(v_t \mu_t) + \Sigma w_t \mu_t$$

(3.3)

($v_t := v(t, \cdot), w_t := w(t, \cdot)$) holds good, in duality with $C^\infty$-functions with compact support in $(0, 1) \times \mathbb{R}^d$ (see (1.7)), and

$$\int_{\mathbb{R}^d} (\Lambda |v_t|^2 + \Sigma |w_t|^2) \, d\mu_t = |\mu'_t|^2 \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (0,1).$$

(3.4)

For every $t \in (0,1)$ and $h \in (-t, 1-t)$, there exists a plan $\beta_{t,t+h} \in \mathcal{M}(\mathcal{C} \times \mathcal{C})$ which is optimal in the definition of $H_{\Lambda, \Sigma}(\mu_{t+h}, \mu_{t})^2$ according to ((2.7), (2.4)) and whose first marginal $\pi^1_{\#} \beta_{t,t+h}$ satisfies

$$\int_{\mathbb{R}^d} \phi([x, r]) \, d(\pi^1_{\#} \beta_{t,t+h}) = \int_{\mathbb{R}^d} \phi([x, 1]) \, d\mu_t + h^2 \phi(o)$$

(3.5)

for all $\phi \in C^0_b(\mathcal{C})$ (cf. Thm. 7.6 and Lem. 7.10 in [9]).

We fix $\nu \in \mathcal{M}(\mathbb{R}^d)$. It follows from (3.2) that

$$t \mapsto H_{\Lambda, \Sigma}(\mu_t, \nu)$$

is an absolutely continuous mapping from $[0,1]$ to $[0, +\infty)$ and thus $\mathcal{L}^1$-a.e. differentiable.

The plan of this section is as follows. First, Prop. 3.1 will identify $(v_t, w_t)$ as belonging to a particular class of functions. Second, the push-forwards of $\beta_{t,t+h}$ through mappings

$$(y_1, y_2) \mapsto \left( (x(y_1), r(y_1)), \left( \frac{1}{h\Lambda} \mathcal{R}_{y_1,y_2}(s)\theta'_{y_1,y_2}(s)(x(y_2) - x(y_1)), \frac{2}{h\Sigma} \mathcal{R}_{y_1,y_2}(s) \right) \right)$$

(3.7)
from \((\mathcal{C} \times \mathcal{C}) \setminus \{(x_1, r_1), [x_2, r_2) \in \mathcal{C} \times \mathcal{C} : r_1, r_2 > 0, |x_1 - x_2| > \pi \sqrt{\Lambda/\Sigma}\) to \((\mathbb{R} \times \mathbb{R}) \times (\mathbb{R} \times \mathbb{R})\) will be considered, for \(s \in (0, 1)\), with \(y_i \equiv [x_i, r_i], \mathbf{x}, \mathbf{r}\) as in (2.6), and
\[
[0, 1] \ni s \mapsto (\theta_{[x_1, r_1], [x_2, r_2]}(s), \mathcal{R}_{[x_1, r_1], [x_2, r_2]}(s)) \in [0, 1] \times [0, +\infty)
\]
being constructed according to Sect. 2 (cf. (2.12)-(2.15)) so that
\[
s \mapsto [x_1 + \theta_{[x_1, r_1], [x_2, r_2]}(s)(x_2 - x_1), \mathcal{R}_{[x_1, r_1], [x_2, r_2]}(s)) is a geodesic from \([x_1, r_1]\) to \([x_2, r_2]\). (3.8)
\]
Please recall (2.11) in this context and note that, by (2.15), the mappings (3.7) are Borel measurable. Their second components may be interpreted as blow-ups of tangent vectors to geodesics in \((\mathcal{C}, d_{\mathcal{C}, \Lambda, \Sigma})\) and Prop. 3.3 will provide information on the limits of the corresponding push-forwards of \(\beta_{l, t, h}\) as \(h \to 0\), linking them to \((v_t, w_t)\). That result will be helpful in studying the \(\mathcal{L}^1\)-a.e.-differentiability of the mapping (3.6) and finally, in Thm. 3.4 we will determine the derivatives by computing
\[
\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{H}_{\Lambda, \Sigma}(\mu_t, \nu)^2
\]
at \(\mathcal{L}^1\)-a.e. \(t \in (0, 1)\).

The above notation holds good throughout this section.

**Proposition 3.1.** For \(\mathcal{L}^1\)-a.e. \(t \in (0, 1)\), the Borel function \((v_t, w_t)\) belongs to the closure in \(L^2(\mu_t, \mathbb{R} \times \mathbb{R})\) of the subspace \(\{(\nabla \zeta, \zeta) : \zeta \in \mathcal{C}_c^\infty(\mathbb{R}^d)\}\).

Here \((L^2(\mu_t, \mathbb{R} \times \mathbb{R}), ||\cdot||_{L^2(\mu_t, \mathbb{R} \times \mathbb{R})})\) denotes the normed space of all \(\mu_t\)-measurable functions \((\bar{v}, \bar{w})\) from \(\mathbb{R}^d\) to \(\mathbb{R} \times \mathbb{R}\) satisfying
\[
|||(\bar{v}, \bar{w})|||_{L^2(\mu_t, \mathbb{R} \times \mathbb{R})} := \left( \int_{\mathbb{R}^d} (\Lambda|\bar{v}|^2 + \Sigma|\bar{w}|^2) \, \mathrm{d}\mu_t \right)^{1/2} < +\infty. \quad (3.10)
\]

**Proof.** We construct a Borel vector field \((\bar{v}, \bar{w}) : (0, 1) \times \mathbb{R} \to \mathbb{R} \times \mathbb{R}\) satisfying (3.3) so that, for \(\mathcal{L}^1\)-a.e. \(t \in (0, 1)\), the function \((\bar{v}_t, \bar{w}_t)\) belongs to the closure in \(L^2(\mu_t, \mathbb{R} \times \mathbb{R})\) of the subspace \(\{(\nabla \zeta, \zeta) : \zeta \in \mathcal{C}_c^\infty(\mathbb{R}^d)\}\) and
\[
|||((\bar{v}_t, \bar{w}_t)|||_{L^2(\mu_t, \mathbb{R} \times \mathbb{R})} = \int_{\mathbb{R}^d} (\Lambda|\bar{v}_t|^2 + \Sigma|\bar{w}_t|^2) \, \mathrm{d}\mu_t \leq |\mu_t|^2. \quad (3.11)
\]

We begin the proof with some estimations. Let \(\phi \in \mathcal{C}_c^\infty(\mathbb{R}^d)\). It follows from the construction of \(\mathcal{R}_{[x_1, r_1], [x_2, r_2]}\) and \(\theta_{[x_1, r_1], [x_2, r_2]}\) according to (2.12)-(2.15) that
\[
\frac{2}{\sum \mathrm{d}^2 s} \mathcal{R}_{[x_1, r_1], [x_2, r_2]}(s)^2 = d_{\mathcal{C}, \Lambda, \Sigma}(x_1, r_1, [x_2, r_2])^2,
\]
\[
\theta_{[x_1, r_1], [x_2, r_2]}(s) \mathcal{R}_{[x_1, r_1], [x_2, r_2]}(s)^2 (x_2 - x_1) \leq C_{\Lambda, \Sigma} d_{\mathcal{C}, \Lambda, \Sigma}(x_1, r_1, [x_2, r_2]),
\]
\[
2\theta_{[x_1, r_1], [x_2, r_2]}(s) \mathcal{R}_{[x_1, r_1], [x_2, r_2]}(s)^2 (x_2 - x_1) \leq C_{\Lambda, \Sigma} d_{\mathcal{C}, \Lambda, \Sigma}(x_1, r_1, [x_2, r_2]),
\]
\[
\frac{\mathrm{d}}{\mathrm{d}^2 s} \left[ \phi(x_1 + \theta_{[x_1, r_1], [x_2, r_2]}(s)(x_2 - x_1)) \mathcal{R}_{[x_1, r_1], [x_2, r_2]}(s)^2 \right] \leq C_{\phi} C_{\Lambda, \Sigma} d_{\mathcal{C}, \Lambda, \Sigma}(x_1, r_1, [x_2, r_2]),
\]
for $s \in (0, 1)$, with $C_{\phi} > 0$ only depending on $\phi$ and $C_{\Sigma, \Lambda} := 2\Sigma + 4\Lambda$; we refer the reader to the proof of Prop. 2.5 in [3] for details. With (2.15), and these estimations on hand, it is straightforward to prove that there exists a constant $C_{\phi, \Lambda, \Sigma} > 0$ only depending on $\phi$, $\Lambda$ and $\Sigma$ so that

$$
|\varphi'_{y_1,y_2}(s) - \varphi'_{y_1,y_2}(s)| \leq C_{\phi,\Lambda,\Sigma} d_{\varepsilon,\Lambda,\Sigma}(y_1, y_2)^2,
$$

(3.12)

$$
|\varphi'_{y_1,y_2}(s) - (\nabla \phi(x_1), \theta'_{y_1,y_2}(s)(x_2-x_1)) R_{y_1,y_2}(s)^2 + 2\phi(x_1) R'_{y_1,y_2}(s) R_{y_1,y_2}(s)| \leq C_{\phi,\Lambda,\Sigma} d_{\varepsilon,\Lambda,\Sigma}(y_1, y_2)^2
$$

(3.13)

and

$$
|\left( (\nabla \phi(x_1), \theta'_{y_1,y_2}(s)(x_2-x_1)) R_{y_1,y_2}(s) + 2\phi(x_1) R'_{y_1,y_2}(s) \right) (R_{y_1,y_2}(s) - r_1)| \leq C_{\phi,\Lambda,\Sigma} d_{\varepsilon,\Lambda,\Sigma}(y_1, y_2)^2
$$

(3.14)

for all $s, \bar{s} \in (0, 1)$, with $y_i := [x_i, r_i]$, $\varphi_{y_1,y_2}(s) := \phi(x_1 + \theta_{x_1[,],x_2[,]}(s)(x_2-x_1)) R_{x_1[,],x_2[,]}(s)^2$.

Now, let $t \in (0, 1)$ so that the limit (3.11) exists and $\mathcal{E}_\emptyset := \mathcal{E} \setminus \{\emptyset\}$. By applying (2.11), (3.13), (3.14), (3.5), Hölder’s inequality and (2.15), we obtain

$$
\left| \int_{\mathbb{R}^d} \phi \, d\mu_{t+h} - \int_{\mathbb{R}^d} \phi \, d\mu_t \right| = \left| \int_{\mathbb{R}^d} (\varphi(x_2) - \varphi(x_1)) \, d\beta_{t,t+h} \right| \leq \int_{\mathbb{R}^d} \int_0^1 \left| \varphi'_{y_1,y_2}(s) \right| \, d\beta_{t,t+h} \leq \int_{\mathbb{R}^d} \int_0^1 \left| \nabla \phi(x_1), \theta'_{x_1[,],x_2[,]}(s)(x_2-x_1) \right| R_{x_1[,],x_2[,]}(s)^2 + 2\phi(x_1) R'_{x_1[,],x_2[,]}(s) \, d\beta_{t,t+h} + 2C_{\phi,\Lambda,\Sigma} |H|_{\Lambda,\Sigma}(\mu_t, \mu_{t+h})^2 \leq \int_{\mathbb{R}^d} \left| (\nabla \phi(x_1) + \Sigma \phi^2) \right| d(\varphi_1 + \Sigma \phi^2)^2 \left( \int_{\mathbb{R}^d} \left( \frac{1}{\Lambda} R^2(\theta')^2 |x_2 - x_1|^2 + \frac{4}{\Sigma} (R')^2 \right) \, d\beta_{t,t+h} \right)^{1/2} + 2C_{\phi,\Lambda,\Sigma} |H|_{\Lambda,\Sigma}(\mu_t, \mu_{t+h})^2 \leq \int_{\mathbb{R}^d} \left| (\nabla \phi, \phi) \right|_{L^2(\mu, \mathbb{R}^d \times \mathbb{R})} |H|_{\Lambda,\Sigma}(\mu_t, \mu_{t+h}) + 2C_{\phi,\Lambda,\Sigma} |H|_{\Lambda,\Sigma}(\mu_t, \mu_{t+h})^2
$$

and thus,

$$
\lim_{h \to 0} \frac{1}{|h|} \left| \int_{\mathbb{R}^d} \phi \, d\mu_{t+h} - \int_{\mathbb{R}^d} \phi \, d\mu_t \right| \leq \left| (\nabla \phi, \phi) \right|_{L^2(\mu, \mathbb{R}^d \times \mathbb{R})} |H|_{\Lambda,\Sigma}(\mu_t, \mu_{t+h}) \quad \text{for all } \phi \in \mathcal{C}_h(0, 1) \times \mathbb{R}^d,
$$

(3.15)

At this point, we may follow the proof of Thm. 8.3.1 in [1]. Therein, a similar characterization of absolutely continuous curves in the space of Borel probability measures with finite second order moments, endowed with the Kantorovich-Wasserstein distance, was given by solving a suitable minimum problem. We adapt that approach. Let $\mu \in \mathcal{M}((0, 1) \times \mathbb{R}^d)$ be defined by

$$
\int_{(0,1) \times \mathbb{R}^d} \psi(t, x) \, d\mu(t, x) = \int_0^1 \int_{\mathbb{R}^d} \psi(t, x) \, d\mu_t(x) \, dt
$$

for all $\psi \in C^0_\emptyset((0, 1) \times \mathbb{R}^d)$, and let $(L^2(\mu, \mathbb{R}^d \times \mathbb{R}), |.|_{L^2(\mu, \mathbb{R}^d \times \mathbb{R})})$ denote the normed space of all $\mu$-measurable vector fields $(\hat{v}, \hat{w})$ from $(0, 1) \times \mathbb{R}^d$ to $\mathbb{R}^d \times \mathbb{R}$ satisfying

$$
|(|\hat{v}|, |\hat{w}|)|_{L^2(\mu, \mathbb{R}^d \times \mathbb{R})} := \left( \int_0^1 \int_{\mathbb{R}^d} (|\Lambda \hat{v}|^2 + \Sigma |\hat{w}|^2) \, d\mu_t \, dt \right)^{1/2} < +\infty.
$$

(3.16)
An application of (3.15), Fatou’s Lemma, Hölder’s inequality and Hahn-Banach Theorem shows that there exists a unique bounded linear functional $L$ defined on the closure $\mathcal{V}$ in $L^2(\mu, \mathbb{R}^d \times \mathbb{R})$ of the subspace $\{(\nabla \zeta, \zeta) : \zeta \in C^\infty_c((0,1) \times \mathbb{R}^d)\}$, satisfying

$$L((\nabla \zeta, \zeta)) := - \int_0^1 \int_{\mathbb{R}^d} \partial_t \zeta(t,x) \, d\mu(t) \, dt \quad \text{for all } \zeta \in C^\infty_c((0,1) \times \mathbb{R}^d). \tag{3.17}$$

We consider the minimum problem

$$\min \left\{ \frac{1}{2} ||(\hat{v}, \hat{w})||^2_{L^2(\mu, \mathbb{R}^d \times \mathbb{R})} - L((\hat{v}, \hat{w})) : (\hat{v}, \hat{w}) \in \mathcal{V} \right\}. \tag{3.18}$$

The same argument as in the proof of Thm. 8.3.1 in [1] proves that the unique solution $(\tilde{v}, \tilde{w})$ to (3.18) (which clearly exists) satisfies (3.3) and, for $L^1$-a.e. $t \in (0,1)$, the function $(\tilde{v}_t, \tilde{w}_t)$ belongs to the closure in $L^2(\mu_t, \mathbb{R}^d \times \mathbb{R})$ of the subspace $\{(\nabla \zeta, \zeta) : \zeta \in C^\infty_c(\mathbb{R}^d)\}$ and (3.11) holds good. By Thm. 8.17 in [9], for every Borel vector field $(\hat{v}, \hat{w}) \in L^2(\mu, \mathbb{R}^d \times \mathbb{R})$ satisfying the continuity equation with reaction (3.3) the opposite inequality holds good, i.e.

$$\int_{\mathbb{R}^d} (\Lambda |\hat{v}_t|^2 + \Sigma |\hat{w}_t|^2) \, d\mu(t) \geq |\mu'_t|^2 \quad \text{for } L^1\text{-a.e. } t \in (0,1).$$

It follows from this and from the strict convexity of $|| \cdot ||^2_{L^2(\mu_t, \mathbb{R}^d \times \mathbb{R})}$ that the Borel vector field $(\tilde{v}, \tilde{w})$ solves (3.3), (3.4) and that it coincides $L^1$-a.e. with any other vector field solving (3.3), (3.4)). This completes the proof of Prop. 3.1. \qed

**Definition 3.2.** Let $D(\mathbb{R}^d)$ be a countable subset of $C^\infty_c(\mathbb{R}^d)$ so that every function in $C^\infty_c(\mathbb{R}^d)$ can be approximated in the $C^1$-norm by a sequence of functions in $D(\mathbb{R}^d)$.

We define $\mathcal{N}$ as the set of points $t \in (0,1)$ at which the following holds good:

(i) the limit (3.1) exists,

(ii) $(v_t, w_t)$ belongs to the closure in $L^2(\mu_t, \mathbb{R}^d \times \mathbb{R})$ of the subspace $\{(\nabla \zeta, \zeta) : \zeta \in C^\infty_c(\mathbb{R}^d)\}$ and satisfies (3.4),

(iii) the mapping $t \mapsto \frac{1}{2} \text{H}K_{\Lambda, \Sigma}(\mu_t, \nu)^2$ is differentiable at $t$,

(iv) and, for all $\psi \in D(\mathbb{R}^d),$

$$\lim_{h \to 0} \frac{1}{h} \left( \int_{\mathbb{R}^d} \psi \, d\mu_{t+h} - \int_{\mathbb{R}^d} \psi \, d\mu_t \right) = \int_{\mathbb{R}^d} (\Lambda(\nabla \psi, v_t) + \Sigma \psi w_t) \, d\mu_t. \tag{3.20}$$

9
Please note that \( (0, 1) \setminus \mathbb{N} \) is an \( \mathcal{L}^1 \)-negligible set; it follows from (1.17) that, for fixed \( \psi \in C^\infty_c(\mathbb{R}^d) \), the mapping \( t \mapsto \int_{\mathbb{R}^d} \psi \, d\mu_t \) is absolutely continuous from \([0, 1] \) to \( \mathbb{R} \) and \( (3.20) \) holds good at \( \mathcal{L}^1 \)-a.e. \( t \in (0, 1) \).

We turn to the push-forward \( \Delta_{t,h,s} \in \mathcal{M}(\mathbb{R}^d \times \mathbb{R}) \times (\mathbb{R}^d \times \mathbb{R}) \) of \( \beta_{t,t+h} \) through (3.7), defined by

\[
\int_{(\mathbb{R}^d \times \mathbb{R}) \times (\mathbb{R}^d \times \mathbb{R})} \Phi(y) \, d\Delta_{t,h,s} = \int_{\mathbb{R}^d} \Phi((x, 1), (v_t(x), w_t(x))) \, d\mu_t
\]

for all \( \Phi \in C^0_b((\mathbb{R}^d \times \mathbb{R}) \times (\mathbb{R}^d \times \mathbb{R})) \).

**Proposition 3.3.** The following holds good for all \( t \in \mathbb{N} \).

(i) Let \( s \in (0, 1) \). Then

\[
\lim_{h \to 0} \int_{(\mathbb{R}^d \times \mathbb{R}) \times (\mathbb{R}^d \times \mathbb{R})} \Phi(y) \, d\Delta_{t,h,s} = \int_{\mathbb{R}^d} \Phi((x, 1), (v_t(x), w_t(x))) \, d\mu_t \tag{3.21}
\]

for all continuous functions \( \Phi : (\mathbb{R}^d \times \mathbb{R}) \times (\mathbb{R}^d \times \mathbb{R}) \to \mathbb{R} \) satisfying the growth condition

\[
|\Phi((x, 1), (x_2, r_2))| \leq C \left( 1 + |x_2|^2 + |r_2|^2 \right) \tag{3.22}
\]

for some \( C > 0 \).

(ii) Define \( \mathcal{E}_{t,h} := \{ [x, r] \in \mathcal{E} \setminus \{0\} : |v_t(x)| < \frac{1}{\sqrt{|h|}} \) and \( |w_t(x)| < \frac{2}{\sqrt{|h|}} \} \) and \( \Xi_{t,h} : \mathcal{E} \to \mathcal{E} \),

\[
\Xi_{t,h}([x, r]) := \begin{cases} [x + \Lambda hv_t(x), r(1 + \Sigma \frac{h}{2} hv_t(x))] & \text{if } [x, r] \in \mathcal{E}_{t,h}, \\ [x, r] & \text{else}. \end{cases} \tag{3.23}
\]

Let \( \chi_{t,h} := (\Xi_{t,h})\#(\pi^1_{t,h}\beta_{t,t+h}) \) be the push-forward of the first marginal of \( \beta_{t,t+h} \) through \( \Xi_{t,h} \), i.e.

\[
\int_{\mathcal{E}} \phi([x, r]) \, d\chi_{t,h} = \int_{\mathcal{E}} \phi(\Xi_{t,h}([x, r])) \, d(\pi^1_{t,h}\beta_{t,t+h})
\]

for all \( \phi \in C^0_b(\mathcal{E}) \). Then

\[
\lim_{h \to 0} \frac{H_{\Lambda, \Sigma}(\mu_{t+h}, h\chi_{t,h})^2}{h^2} = 0. \tag{3.24}
\]

**Proof.** We set \( Y := \mathbb{R}^d \times \mathbb{R} \).

(i) Let \( t \in \mathbb{N} \) and \( s \in (0, 1) \). We note that, by (2.15) and Def. 3.2(i),

\[
\int_{Y \times Y} (\Lambda |x_2|^2 + \Sigma |r_2|^2) \, d\Delta_{t,h,s}((x_1, r_1), (x_2, r_2)) = \frac{H_{\Lambda, \Sigma}(\mu_{t+h}, \mu_{t+h})^2}{h^2} \to |\mu_t|^2 \quad \text{as } h \to 0. \tag{3.25}
\]
We may apply Prokhorov’s Theorem to any sequence \((\Delta_{t,h,k})_{k \in \mathbb{N}}\), \(h_k \to 0\), of measures from the family \((\Delta_{t,h,k})_{h \in (-t,1-t)} \subset \mathcal{M}(Y \times Y)\), since such sequence is bounded and equally tight by (3.5) and (3.25), and we obtain a subsequence \(h_{k_l} \to 0\) and a measure \(\Delta \in \mathcal{M}(Y \times Y)\) so that \((\Delta_{t,h_{k_l}})_{l \in \mathbb{N}}\) converges to \(\Delta\) in the weak topology on \(\mathcal{M}(Y \times Y)\), in duality with continuous and bounded functions. So let \((\Delta_{t,h_{k_l}})_{l \in \mathbb{N}}\) be a convergent sequence with limit measure \(\Delta \in \mathcal{M}(Y \times Y)\), i.e.

\[
\lim_{l \to \infty} \int_{Y \times Y} \Phi(y) \, d\Delta_{t,h_{k_l}} = \int_{Y \times Y} \Phi(y) \, d\Delta \tag{3.26}
\]

for all \(\Phi \in C_b^0(Y \times Y)\). We want to identify \(\Delta\) as \((\langle x, 1 \rangle, (v_1(x), w_1(x)))_{#_\mu t}\). It is not difficult to infer from (3.5) that the first marginal \(\pi_{#_\mu} \Delta\) of \(\Delta\) coincides with \((x, 1)_{#_\mu t}\), i.e.

\[
\int_{Y} \phi((x, r)) \, d(\pi_{#_\mu} \Delta) = \int_{\mathbb{R}^d} \phi((x, 1)) \, d\mu_t \tag{3.27}
\]

for all \(\phi \in C_b^0(Y)\). Let \(\psi \in \mathcal{D}(\mathbb{R}^d)\). Then (3.26) also holds good for \(\Phi((x_1, r_1), (x_2, r_2)) := [\Lambda(\nabla \psi(x_1), x_2) + \Sigma \psi(x_1) r_2] r_1\); Indeed, we have

\[
\lim_{l \to \infty} \int_{Y \times Y} (\Phi_N) \, d\Delta_{t,h_{k_l}} = \int_{Y \times Y} (\Phi_N) \, d\Delta \tag{\text{for all } N > 0, with } \Phi_N := (\Phi \wedge N) \vee (-N). Setting Y := \{(x, r) \in Y : |x| + |r| > N\}, C_\psi := \sup_{x \in \mathbb{R}^d} \{|\nabla \psi(x)| + |\psi(x)|\}, and applying (3.25), (3.5) and (3.27), we conclude that for every \(\epsilon > 0\) there exists \(N_\epsilon > 0\) so that

\[
\int_{Y \times Y_N} (|x_2| + |r_2|) \, d\Delta_{t,h_{k_l}} + \int_{Y \times Y_N} (|x_2| + |r_2|) \, d\Delta \leq \epsilon \quad \text{for all } N \geq N_\epsilon, \ l \in \mathbb{N},
\]

and

\[
\lim_{l \to \infty} \sup \left| \int_{Y \times Y} \Phi \, d\Delta_{t,h_{k_l}} - \int_{Y \times Y} \Phi \, d\Delta \right| \\
\leq \lim_{l \to \infty} \sup \left| \int_{Y \times Y} (\Phi_{C_\psi(\Lambda + \Sigma) N_\epsilon}) \, d\Delta_{t,h_{k_l}} - \int_{Y \times Y} \Phi_{C_\psi(\Lambda + \Sigma) N_\epsilon} \, d\Delta \right| \\
\quad + C_\psi(\Lambda + \Sigma) \lim_{l \to \infty} \sup \int_{Y \times Y_N} (|x_2| + |r_2|) \, d(\Delta_{t,h_{k_l}} + \Delta) \\
\leq C_\psi(\Lambda + \Sigma) \epsilon.
\]

Hence, taking (3.27) into account, we obtain

\[
\lim_{l \to \infty} \int_{Y \times Y} \left[\Lambda(\nabla \psi(x_1), x_2) + \Sigma \psi(x_1) r_2\right] r_1 \, d\Delta_{t,h_{k_l}} = \int_{Y \times Y} \left[\Lambda(\nabla \psi(x_1), x_2) + \Sigma \psi(x_1) r_2\right] \, d\Delta. \tag{3.28}
\]
It holds that
\[
\int_{\mathbb{R}^d} \psi \, d\mu_{t+h} - \int_{\mathbb{R}^d} \psi \, d\mu_t = \int_{\mathbb{R}^d} (\psi(x_2)r_2^2 - \psi(x_1)r_1^2) \, d\beta_{t+h}
\]
\[
= \int_{\mathbb{R}^d} \int_0^1 \frac{d}{ds} \left[ \psi(x_1 + \theta_{[x_1,x_1],[x_2,x_2]}(s)(x_2 - x_1)) \mathcal{R}_{[x_1,x_1],[x_2,x_2]}(s)^2 \right] \, ds \, d\beta_{t+h}
\]
so that (3.20), (3.12), (3.13), (3.14), Def. 3.2(i) and (3.28) yield
\[
\int_{\mathbb{R}^d} (\Lambda(\nabla \psi, v_1) + \Sigma \psi w_1) \, d\mu_t = \lim_{t \to \infty} \frac{1}{h_t} \left( \int_{\mathbb{R}^d} \psi \, d\mu_{t+h} - \int_{\mathbb{R}^d} \psi \, d\mu_t \right)
\]
\[
= \lim_{t \to \infty} \int_{Y \times Y} \Lambda(\nabla \psi(1), x_2) + \Sigma \psi(x_2) \, r_1 \, d\Delta_{t,h_1,s} = \int_{Y \times Y} \Lambda(\nabla \psi(1), x_2) + \Sigma \psi(x_2) \, r_1 \, d\Delta
\]
According to the Disintegration Theorem (see e.g. Thm. 5.3.1 in [1]) and (3.27), there exists a Borel family of probability measures \((\Delta_{x_1})_{x_1 \in \mathbb{R}^d} \subset \mathcal{M}(Y), \Delta_{x_1}(Y) = 1\), so that
\[
\int_{Y \times Y} \Phi \, d\Delta = \int_{\mathbb{R}^d} \left( \int_Y \Phi((x_1, 1), (x_2, r_2)) \, d\Delta_{x_1}((x_2, r_2)) \right) \, d\mu_t(x_1)
\]
for all \(\Delta\)-integrable maps \(\Phi : Y \times Y \to \mathbb{R}\). We infer from (3.25) that, for \(\mu_t\)-a.e. \(x_1 \in \mathbb{R}^d\), the measure \(\Delta_{x_1}\) has finite second order moment and we define the function \((v_\Delta, w_\Delta) : \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}\) by
\[
v_\Delta(x_1) := \int_Y x_2 \, d\Delta_{x_1}((x_2, r_2)), \quad w_\Delta(x_1) := \int_Y r_2 \, d\Delta_{x_1}((x_2, r_2)) \quad \text{for } \mu_t\text{-a.e. } x_1 \in \mathbb{R}^d.
\]
The function \((v_\Delta, w_\Delta)\) is Borel measurable (cf. (5.3.1) and Def. 5.4.2 in [1]), and
\[
\int_{Y \times Y} \Lambda(\nabla \psi(x_1), x_2) + \Sigma \psi(x_2) \, r_1 \, d\Delta
\]
\[
= \int_{\mathbb{R}^d} \left( \int_Y \Lambda(\nabla \psi(x_1), x_2) + \Sigma \psi(x_2) \, r_1 \, d\Delta_{x_1}((x_2, r_2)) \right) \, d\mu_t(x_1)
\]
\[
= \int_{\mathbb{R}^d} (\Lambda(\nabla \psi, v_\Delta) + \Sigma \psi w_\Delta) \, d\mu_t.
\]
All in all, we have found that
\[
\int_{\mathbb{R}^d} (\Lambda(\nabla \psi, v_t) + \Sigma \psi w_t) \, d\mu_t = \int_{\mathbb{R}^d} (\Lambda(\nabla \psi, v_\Delta) + \Sigma \psi w_\Delta) \, d\mu_t
\]
for all \(\psi \in \mathcal{D}(\mathbb{R}^d)\). Since every function in \(C^\infty_c(\mathbb{R}^d)\) can be approximated in the \(C^1\)-norm by a sequence of functions in \(\mathcal{D}(\mathbb{R}^d)\) (cf. Def. 3.2 and, by (3.25) and Def. 3.2(ii), the functions \(v_\Delta, w_\Delta, v_t, w_t\) are square-integrable w.r.t. \(\mu_t\), (3.30) holds good for all \(\psi \in C^\infty_c(\mathbb{R}^d)\) and for all
pairs in the $L^2(\mu_t, \mathbb{R}^d \times \mathbb{R})$-closure of $\{(\nabla \zeta, \zeta) : \zeta \in C_c^\infty(\mathbb{R}^d)\}$. It follows from this and from Def. 3.2(ii) that

$$\|(v_t, w_t)\|_{L^2(\mu_t, \mathbb{R}^d \times \mathbb{R})} = \int_{\mathbb{R}^d} (\Lambda(v_t, v_\Delta) + \Sigma w_t w_\Delta) \, d\mu_t. \tag{3.31}$$

Applying Hölder’s inequality to (3.31), taking the definition (3.29) of $v_\Delta$, $w_\Delta$, Jensen’s inequality, (3.26), (3.25) and Def. 3.2(ii) into account, we obtain

$$\|(v_t, w_t)\|_{L^2(\mu_t, \mathbb{R}^d \times \mathbb{R})} \leq \|(v_\Delta, w_\Delta)\|_{L^2(\mu_t, \mathbb{R}^d \times \mathbb{R})} \leq \left( \int_{Y \times Y} (\Lambda|x_2|^2 + \Sigma|r_2|^2) \, d\Delta \right)^{1/2} \leq \lim_{t \to \infty} \left( \int_{Y \times Y} (\Lambda|x_2|^2 + \Sigma|r_2|^2) \, d\Delta_{t,h_t,s} \right)^{1/2} = \|(v_t, w_t)\|_{L^2(\mu_t, \mathbb{R}^d \times \mathbb{R})} \tag{3.32}$$

so that, in fact, equality holds good everywhere in (3.32) and (3.33). We infer from this and from (3.31) that

$$\|(v_t, w_t) - (v_\Delta, w_\Delta)\|_{L^2(\mu_t, \mathbb{R}^d \times \mathbb{R})} = 0 \tag{3.34}$$

which means

$$v_t(x) = v_\Delta(x) \quad \text{and} \quad w_t(x) = w_\Delta(x) \quad \text{for } \mu_t\text{-a.e. } x \in \mathbb{R}^d. \tag{3.35}$$

Moreover, the fact that the second inequality in (3.32), resulting from Jensen’s inequality, is in fact an equality and (3.34) yield $\Delta_{x_1} = \delta_{v_t(x_1)} \otimes \delta_{w_t(x_1)}$ for $\mu_t$-a.e. $x_1 \in \mathbb{R}^d$ (cf. a canonical proof of Jensen’s inequality), i.e.

$$\int_Y \phi((x, r)) \, d\Delta_{x_1} = \phi(v_t(x_1), w_t(x_1)) \tag{3.36}$$

for all $\phi \in C^0_c(Y)$, for $\mu_t$-a.e. $x_1 \in \mathbb{R}^d$.

Altogether, we may conclude that $\Delta = ((x, 1), (v_t(x), w_t(x))) \neq \mu_t$,

$$\int_{Y \times Y} (\Lambda|x_2|^2 + \Sigma|r_2|^2) \, d\Delta = |\mu_t'|^2 = \lim_{t \to \infty} \int_{Y \times Y} (\Lambda|x_2|^2 + \Sigma|r_2|^2) \, d\Delta_{t,h_t,s} \tag{3.37}$$

and that (3.21) holds good for all $\Phi \in C^0_c(Y \times Y)$.

A similar argument as in the proof of (3.28), making use of (3.36), will show (3.21) for all continuous functions $\Phi : Y \times Y \to \mathbb{R}$ satisfying the growth condition (3.22) (cf. Thm. 7.12 in [10] where the space of Borel probability measures with finite second order moments is considered and the equivalence between convergence in the Kantorovich-Wasserstein distance and convergence in duality with continuous functions satisfying a suitable growth condition is proved). This completes the proof of Prop. 3.3(i).

(ii) Let $t \in \mathbb{N}$. According to (2.7), (2.4), we have

$$\frac{\mathcal{H}K_{\Lambda, \Sigma}(\mu_{t+h}, \mathfrak{h}Y_{t+h})}{h^2} \leq \frac{1}{h^2} \int_{\mathbb{R} \times \mathbb{R}} d\mathfrak{e}_{\Lambda, \Sigma}(Y_{t+h}([x_1, r_1]), [x_2, r_2])^2 \, d\beta_{t+h}. \tag{3.37}$$
We will prove that the right-hand side of (3.37) converges to 0 as \( h \to 0 \).

First we note that, by Prokhorov’s Theorem, Def. (3.2(ii) and the proof of Prop. (3.3(i), every sequence \( \left( ((v_t(x_1), w_t(x_1)), (x_2, r_2))_{\#} \Delta_{t,h_1,s}\right)_{t \in \mathbb{N}}, h_t \to 0 \), is relatively compact w.r.t. the weak topology in \( \mathcal{M}(Y \times Y) \) and in duality with continuous functions \( \Phi : Y \times Y \to \mathbb{R} \) satisfying (3.22), and the second marginals of the corresponding limit measures coincide with \( (v_t(x), w_t(x))_{\#} \mu_t \). It follows from this and from an application of the Dominated Convergence Theorem that

\[
\lim_{N \to \infty} \limsup_{h \to 0} \frac{1}{h^2} \int_{(\mathcal{C}(\mathcal{C}_{t,1/N}) \times \mathcal{C})} d\mathcal{C}_{t,\Lambda,\Sigma}([x_1, r_1], [x_2, r_2])^2 d\beta_{t,t+h} = 0,
\]

which implies

\[
\lim_{h \to 0} \frac{1}{h^2} \int_{(\mathcal{C}(\mathcal{C}_{t,1}) \times \mathcal{C})} d\mathcal{C}_{t,\Lambda,\Sigma}([x_1, r_1], [x_2, r_2])^2 d\beta_{t,t+h} = 0. \tag{3.38}
\]

Next we consider \( \frac{1}{h} \int_{\mathcal{C}_{t,1} \times \mathcal{C}} d\mathcal{C}_{t,\Lambda,\Sigma}(\Xi_{t,h}([x_1, r_1]), [x_2, r_2])^2 d\beta_{t,t+h} \). According to (2), Sect. 3.6 and (9), Sect. 8.1, the geometric cone \((\mathcal{C}, d_{\mathcal{C},\Lambda,\Sigma})\) is a length space and it holds that any curve \( \eta := [x, r] : [0, 1] \to \mathcal{C} \) for \( C^1 \)-functions \( x : [0, 1] \to \mathbb{R}^d \) and \( r : [0, 1] \to [0, +\infty) \) is absolutely continuous in \((\mathcal{C}, d_{\mathcal{C},\Lambda,\Sigma})\) and

\[
d_{\mathcal{C},\Lambda,\Sigma}(\eta(1), \eta(0))^2 \leq \int_0^1 \left( \frac{4}{\Sigma}(r'(s))^2 + \frac{1}{\Lambda}r(s)^2 \right) ds
\]

(cf. (9), Lem. 8.1). We define, for \( y_1 := [x_1, r_1] \in \mathcal{C}_{t,h}, y_2 := [x_2, r_2] \in \mathcal{C}, \) with \(|x_1 - x_2| \leq \pi \sqrt{\Lambda/\Sigma} \) if \( r_2 > 0 \), an absolutely continuous curve \( \mathcal{C}_{h,\Xi(y_1),y_2} : [0, 1] \to \mathcal{C} \) connecting \( \Xi(y_1) = [x_1 + \Lambda h v_t(x_1), r_1(1 + \Sigma h w_t(x_1)/2)] \) and \( y_2 \) by setting \( \mathcal{C}_{h,\Xi(y_1),y_2} := [x_1, y_2] \),

\[
\mathcal{X}_{h,\Xi(y_1),y_2}(s) := x_1 + \theta_{y_1,y_2}(s)(x_2 - x_1) + \Lambda (1 - s) h v_t(x_1), \tag{3.39}
\]

\[
\mathcal{R}_{h,\Xi(y_1),y_2}(s) := \mathcal{R}_{y_1,y_2}(s) \left( 1 + \Sigma (1 - s) h w_t(x_1)/2 \right). \tag{3.40}
\]

(cf. (3.8), (2.11)). The functions \( \mathcal{X}_{h,\Xi(y_1),y_2} : [0, 1] \to \mathbb{R}^d \) and \( \mathcal{R}_{h,\Xi(y_1),y_2} : [0, 1] \to [0, +\infty) \) are continuously differentiable with

\[
(R'_{h,\Xi(y_1),y_2}(s))^2 = \left( \Sigma R'_{y_1,y_2}(s)(1 - s) h w_t(x_1)/2 + R'_{y_1,y_2}(s) - \Sigma R_{y_1,y_2}(s) h w_t(x_1)/2 \right)^2
\]

\[
\leq 2|h| \Sigma d_{\mathcal{C},\Lambda,\Sigma}(y_1, y_2)^2 + 2 \left( R'_{y_1,y_2}(s) - \Sigma r_1 h w_t(x_1)/2 \right)^2
\]

and

\[
\mathcal{R}_{h,\Xi(y_1),y_2}(s)^2 \mathcal{X}'_{h,\Xi(y_1),y_2}(s)^2 \leq 4 R_{y_1,y_2}(s)^2 |\theta'_{y_1,y_2}(s)(x_2 - x_1) - \Lambda h v_t(x_1)|^2
\]

\[
\leq 8 \left( |\mathcal{R}_{y_1,y_2}(s)\theta'_{y_1,y_2}(s)(x_2 - x_1) - \Lambda r_1 h v_t(x_1)|^2 + \Lambda^2 |h| |\mathcal{R}_{y_1,y_2}(s) - r_1|^2 \right)
\]

\[
\leq 8 \left( |\mathcal{R}_{y_1,y_2}(s)\theta'_{y_1,y_2}(s)(x_2 - x_1) - \Lambda r_1 h v_t(x_1)|^2 + \Lambda^2 |h|/4 d_{\mathcal{C},\Lambda,\Sigma}(y_1, y_2)^2 \right),
\]

14
where we have made use of \((3.8), (2.15)\) and the fact that \(y_1 = [x_1, r_1] \in \mathcal{C}_{t,h}\). It follows from the above estimations and an application of Fubini’s Theorem that

\[
\frac{1}{h^2} \int_{\mathcal{C}_{t,h}\times \varepsilon} d_{\varepsilon, \Lambda, \Sigma}(\Xi_{t,h}([x_1, r_1]), [x_2, r_2])^2 \, d\beta_{t,t+h}
\]

\[
\leq \frac{1}{h^2} \int_{\mathcal{C}_{t,h}\times \varepsilon} \int_0^1 \left( \frac{4}{\Lambda} R_{h, \Xi(y_1), y_2}(s) + \frac{1}{\Lambda} R_{h, \Xi(y_1), y_2}(s)^2 \right) ds \, d\beta_{t,t+h}
\]

\[
\leq \int_0^1 \int_{Y \times Y} \left( 2\Sigma(r_2 - r_1 w_t(x_1))^2 + 8\Lambda |x_2 - r_1 v_t(x_1)|^2 \right) d\Delta_{t,h,s}((x_1, r_1), (x_2, r_2)) ds
\]

\[
+ C_{\Lambda, \Sigma} \frac{H_{\Lambda, \Sigma}(\mu_t, \mu_{t+h})^2}{|h|}
\]

with \(C_{\Lambda, \Sigma}\) only depending on \(\Lambda\) and \(\Sigma\). According to Def. \((3.2)\)\(\text{(ii)}, \) there exists a sequence of functions \(\zeta_n \in C_c(\mathbb{R}^d)\) \((n \in \mathbb{N})\) so that \(((\nabla \zeta_n, \zeta_n))_{n \in \mathbb{N}}\) converges to \((v_t, w_t)\) in \(L^2(\mu_t, \mathbb{R}^d \times \mathbb{R})\), which means

\[
\lim_{n \to \infty} \int_{Y \times Y} \left( r_1^2 (\zeta_n(x_1) - w_t(x_1))^2 + r_1^2 |\nabla \zeta_n(x_1) - v_t(x_1)|^2 \right) d\Delta_{t,h,s} = 0 \quad (3.41)
\]

uniformly in \(h \in (-t, 1-t)\) and \(s \in (0, 1)\). Moreover, Prop. \((3.3)\)\(\text{(i)}, \) and \((3.5)\) yield

\[
\lim_{h \to 0} \int_{Y \times Y} \left( \Sigma(r_2 - r_1 \zeta_n(x_1))^2 + \Lambda |x_2 - r_1 \nabla \zeta_n(x_1)|^2 \right) d\Delta_{t,h,s} = ||(v_t, w_t) - (\nabla \zeta_n, \zeta_n)||^2_{L^2(\mu_t, \mathbb{R}^d \times \mathbb{R})} \quad (3.42)
\]

for all \(n \in \mathbb{N}\) and \(s \in (0, 1)\). Altogether, by applying Def. \((3.2)\)\(\text{(i)}, \) \((3.41), \) \((3.42)\) and Fatou’s Lemma to the above estimation of \(\frac{1}{h^2} \int_{\mathcal{C}_{t,h}\times \varepsilon} d_{\varepsilon, \Lambda, \Sigma}(\Xi_{t,h}([x_1, r_1]), [x_2, r_2])^2 \, d\beta_{t,t+h}\), we obtain

\[
\lim_{h \to 0} \frac{1}{h^2} \int_{\mathcal{C}_{t,h}\times \varepsilon} d_{\varepsilon, \Lambda, \Sigma}(\Xi_{t,h}([x_1, r_1]), [x_2, r_2])^2 \, d\beta_{t,t+h} = 0, \quad (3.43)
\]

which completes the proof of Prop. \((3.3)\)\(\text{(ii)}.\)

We are now in a position to compute the derivative \((3.9)\) at every \(t \in \mathbb{N}.\)

**Theorem 3.4.** If \(t \in \mathbb{N}\) and \(\beta_{t,*} \in \mathcal{M}(\mathcal{C} \times \mathcal{C})\) is optimal in the definition of \(H_{\Lambda, \Sigma}(\mu_t, \nu)^2\) according to \((2.8), \) \((2.4), \) with first marginal \(\alpha_t \in \mathcal{M}_2(\mathcal{C})\), \(\hbar \alpha_t \leq \mu_t, \) and second marginal \(\alpha_* \in \mathcal{M}_2(\mathcal{C}), \) \(\hbar \alpha_* \leq \nu, \) then the derivative \(\frac{d}{dt} [H_{\Lambda, \Sigma}(\mu_t, \nu)^2] \) of \((3.19)\) at \(t\) coincides with

\[
\mathcal{F}_{t,*} + 2 \int_{\mathbb{R}^d} w_t(x) \, d(\mu_t - \hbar \alpha_t)
\]

where \(\mathcal{F}_{t,*}\) is defined as

\[
2 \int_{\varepsilon \times \varepsilon} \left[ r_1^2 w_t(x_1) - r_1 r_2 w_t(x_1) \cos(\sqrt{\Sigma/4\Lambda} |x_1 - x_2|) - r_1 r_2 \sqrt{\Lambda/\Sigma} \langle S_{\Lambda, \Sigma}(x_1, x_2), v_t(x_1) \rangle \right] \, d\beta_{t,*},
\]

\[
(3.44)
\]
the mapping $\chi$ with $\phi$ for all nonnegative bounded Borel functions $w$ defined as in Prop. 3.3(ii). Let $\bar{\chi}_{t,h}$ be the push-forward of $\alpha_t$ through the mapping $\Xi_{t,h}$ defined as in (3.23). We have

$$\int_{\mathbb{R}^d} \phi \, d(\bar{\chi}_{t,h}) = \int_{\mathbb{R}^d} r^2 \phi(x) \, d\bar{\chi}_{t,h} = \int_{\mathbb{R}^d} \phi(x) \, d\alpha_t + \int_{\mathbb{R}^d} r^2 \phi(x) \, d\alpha_t$$

$$= \int_{\mathbb{R}^d} (1 + \Sigma h w_t(x)/2)^2 (\phi(x) + \Lambda h \nu_t(x)) \, d\alpha_t + \int_{\mathbb{R}^d} \phi(x) \, d\alpha_t$$

$$\leq \int_{\mathbb{R}^d} (1 + \Sigma h w_t(x)/2)^2 (\phi(x) + \Lambda h \nu_t(x)) \, d\mu_t + \int_{\mathbb{R}^d} \phi(x) \, d\mu_t = \int_{\mathbb{R}^d} \phi(x) \, d(\bar{\chi}_{t,h})$$

for all nonnegative bounded Borel functions $\phi : \mathbb{R}^d \to \mathbb{R}$ (cf. (2.5), (2.6)), from which we infer that

$$(\bar{\chi}_{t,h} - \bar{\chi}_{t,h})(\mathbb{R}^d) = (\mu_t - \bar{\chi}_{t,h})(\mathbb{R}^d) + \int_{\mathbb{R}^d} (\Sigma h w_t(x) + \frac{\Sigma^2}{2} h^2 w_t(x)^2) \, d(\mu_t - \bar{\chi}_{t,h})$$

We obtain

$$\frac{1}{2} \left( \mathcal{H}_{\Lambda,\Sigma}(\bar{\chi}_{t,h}, \nu)^2 - \mathcal{H}_{\Lambda,\Sigma}(\mu_t, \nu)^2 \right) \leq \frac{1}{2} \left( \mathcal{W}_{\nu}(\bar{\chi}_{t,h}, \alpha_*)^2 - \mathcal{W}_{\nu}(\mu_t, \alpha_*)^2 \right)$$

$$+ 2 \int_{\mathbb{R}^d} \left( h w_t(x) + \frac{\Sigma^2}{4} h^2 w_t(x)^2 \right) \, d(\mu_t - \bar{\chi}_{t,h})$$

The same argument as in the proof of Lem. 2.2 in [5] yields

$$\limsup_{h \to 0} \frac{1}{2} \mathcal{W}_{\nu}(\bar{\chi}_{t,h}, \alpha_*)^2 - \frac{1}{2} \mathcal{W}_{\nu}(\mu_t, \alpha_*)^2 \leq$$

$$2 \int_{\mathbb{R}^d} \left[ r_1^2 w_t(x_1) - r_1 r_2 w_t(x_1) \cos(\sqrt{\Sigma/4 \Lambda} |x_1 - x_2|) - r_1 r_2 \sqrt{\Lambda/\Sigma} \langle S_{\Lambda,\Sigma}(x_1, x_2), v_t(x_1) \rangle \right] \, d\beta_\cdot$$

$$\leq \liminf_{h \to 0} \frac{1}{2} \mathcal{W}_{\nu}(\bar{\chi}_{t,h}, \alpha_*)^2 - \frac{1}{2} \mathcal{W}_{\nu}(\mu_t, \alpha_*)^2.$$
with $S_{\Lambda, \Sigma}$ defined as in (3.46). Since the limit (3.47) exists and
\[
\lim_{h \to 0} \int_{x(e_{t,h})} \left( w_t(x) + \frac{\Sigma}{4} hw_t(x) \right) d(\mu_t - h\alpha_t) = \int_{\mathbb{R}^d} w_t(x) d(\mu_t - h\alpha_t),
\]
it follows from the above computations that
\[
\lim_{h \to 0} \frac{1}{h} H_{\Lambda, \Sigma}(h\chi_{t,h}, \nu) - \frac{1}{2} H_{\Lambda, \Sigma}(\mu_t, \nu)^2 = \mathcal{F}_{t,*} + 2 \int_{\mathbb{R}^d} w_t(x) d(\mu_t - h\alpha_t).
\]

The proof of Thm. 3.4 is complete. \qed

We would like to remark that the derivatives of (3.19) at $t \in \mathbb{N}$ can be expressed equally in terms of the Logarithmic Entropy-Transport characterization (1.1) of the Hellinger-Kantorovich distance $H_{\Lambda, \Sigma}$, by applying (2.10) to the above representation (3.44), (3.45) of the derivatives.

Acknowledgement I gratefully acknowledge support from the Erwin Schrödinger International Institute for Mathematics and Physics (Vienna) during my participation in the programme “Optimal Transport”.

References

[1] L. Ambrosio, N. Gigli, and G. Savaré, Gradient Flows in Metric Spaces and in the Space of Probability Measures, Lectures in mathematics ETH Zürich, Birkhäuser, 2005.

[2] D. Burago, Y. Burago, and S. Ivanov, A course in metric geometry, vol. 33 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2001.

[3] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard, An interpolating distance between optimal transport and Fisher–Rao metrics, Foundations of Computational Mathematics, 18 (2018), pp. 1–44.

[4] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard, Unbalanced optimal transport: Dynamic and Kantorovich formulations, Journal of Functional Analysis, 274 (2018), pp. 3090–3123.

[5] F. Fleißner, A Minimizing Movement approach to a class of scalar reaction-diffusion equations, submitted, arXiv preprint arXiv 2002.04496, (2020).

[6] W. Gangbo and R. J. McCann, The geometry of optimal transportation, Acta Math., 177 (1996), pp. 113–161.

[7] S. Kondratyev, L. Monsaingeon, D. Vorotnikov, et al., A new optimal transport distance on the space of finite Radon measures, Advances in Differential Equations, 21 (2016), pp. 1117–1164.
[8] M. Liero, A. Mielke, and G. Savaré, *Optimal transport in competition with reaction: The Hellinger–Kantorovich distance and geodesic curves*, SIAM Journal on Mathematical Analysis, 48 (2016), pp. 2869–2911.

[9] ———, *Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures*, Inventiones mathematicae, 211 (2018), pp. 969–1117.

[10] C. Villani, *Topics in optimal transportation*, vol. 58 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2003.