C. R. Acad. Sci. Paris, Ser. I 350 (2012) 35–40

Contents lists available at SciVerse ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Partial Differential Equations/Differential Geometry

Finsler structure in the p-Wasserstein space and gradient flows

Structure de Finsler dans l'espace de Wasserstein L^p et flux de gradient

Martial Agueh [1]

Department of Mathematics and Statistics, University of Victoria, P.O. Box 3060 STN CSC, Victoria, BC, V8W 3R4, Canada

A R T I C L E   I N F O

Article history:
Received 17 February 2011
Accepted after revision 24 November 2011
Available online 8 December 2011

Presented by Gilles Pisier

A B S T R A C T

It is known from the work of F. Otto (2001) [9], that the space of probability measures equipped with the quadratic Wasserstein distance, i.e., the 2-Wasserstein space, can be viewed as a Riemannian manifold. Here we show that when the quadratic cost is replaced by a general homogeneous cost of degree $p > 1$, the corresponding space of probability measures, i.e., the p-Wasserstein space, can be endowed with a Finsler metric whose induced distance function is the p-Wasserstein distance. Using this Finsler structure of the p-Wasserstein space, we give definitions of the differential and gradient of functionals defined on this space, and then of gradient flows in this space. In particular we show in this framework that the parabolic q-Laplacian equation is a gradient flow in the p-Wasserstein space, where $p = q/(q-1)$. When $p = 2$, we recover the Riemannian structure introduced by F. Otto, which confirms that the 2-Wasserstein space is a Riemann–Finsler manifold. Our approach is confined to a smooth situation where probability measures are absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^n$, and they have smooth and strictly positive densities.

R É S U M É

Il est connu que l'espace des mesures de probabilités muni de la distance de Wasserstein L^2 (l'espace de Wasserstein L^2) est une variété Riemanienne (voir F. Otto (2001) [9]). Ici, nous montrons que lorsqu'on change le coût quadratique en un coût plus general, homogène de degré $p > 1$, l'espace correspondant (l'espace de Wasserstein L^p) admet une structure de Finsler dont la distance induite est la distance de Wasserstein L^p. Grâce à cette structure de Finsler, nous donnons une définition de la différentiel et du gradient des fonctionelles définies sur cet espace, et aussi des flux de gradient sur cet espace. En particulier nous montrons que l'équation parabolique q-Laplacien est un flux de gradient dans l'espace de Wasserstein L^p pour $p = q/(q-1)$. Quand $p = 2$, nous retrouvons la structure Remannienne de F. Otto, ce qui confirme que l'espace de Wasserstein L^2 est une variété Riemanienne de Finsler. Notre méthode s'applique à des mesures de probabilité absolument continues par rapport à la mesure de Lebesgue dans $\mathbb{R}^n$, et dont les densités sont strictement positives.

E-mail address: agueh@math.uvic.ca.

[1] The author is supported by a grant from the Natural Science and Engineering Research Council of Canada.

1631-073X/$ – see front matter © 2011 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.crma.2011.11.014

Version française abrégée

Ce papier traite de la structure géometrique de l'espace de Wasserstein L^p, $(\mathbf{P}_p(\mathbb{R}^n), d_p)$, et de ses applications à des equations aux dérivées partielles. Il est connu (voir Jordan–Kinderlehrer–Otto [7], Otto [9], Carrillo–McCann–Villani [4]) que l'espace de Wasserstein L^2 admet une structure Riemanienne. Récemment, Agueh [1] a montré l'existence des solutions de l'equation $\frac{\partial\rho}{\partial t} = \operatorname{div}[\rho\nabla c^* \circ \nabla(F'(\rho) + U)]$ en généralisant à $(\mathbf{P}_p(\mathbb{R}^n), d_p)$ la méthode variationelle de [7]; ici $c^*(x) = |x|^q/q$ et $q = p/(p-1)$. Ce travail suggère que l'espace $(\mathbf{P}_p(\mathbb{R}^n), d_p)$ admet une structure pour laquelle cette équation peut être vue comme un flux de gradient. Ici nous montrons que $(\mathbf{P}_p(\mathbb{R}^n), d_p)$ admet une structure de Finsler, F_p, et nous donnons une définition de la différentiel et du gradient des fonctionelles définies sur cet espace, et aussi des flux de gradient sur cet espace. Finalement, nous prouvons que cette équation est éffectivement un flux de gradient de l'energie $E(\rho) = \int_{\mathbb{R}^n}(F(\rho) + \rho U)\,dx$ dans la variété de Finsler $(\mathbf{P}_p(\mathbb{R}^n), F_p)$.

1. Introduction

Let $n \geqslant 1$ be an integer and $p > 1$ be a real number. Denote by $\mathbf{P}_p(\mathbb{R}^n)$ the space of smooth and strictly positive probability densities on $\mathbb{R}^n$. The p-Wasserstein distance between two densities ρ_0 and ρ_1 in $\mathbf{P}_p(\mathbb{R}^n)$ is defined as

$$d_p^p(\rho_0, \rho_1) = \inf_T \left\{ \int_{\mathbb{R}^n} \left|T(x) - x\right|^p \rho_0(x)\,dx;\ T : \mathbb{R}^n \to \mathbb{R}^n,\ T_\#\rho_0 = \rho_1 \right\}, \tag{1}$$

where $T_\#\rho_0 = \rho_1$ means that $\rho_1(B) = \rho_0(T^{-1}(B))$ for all Borel sets $B \subset \mathbb{R}^n$. When $\mathbf{P}_p(\mathbb{R}^n)$ is equipped with the distance d_p, it will be called the p-Wasserstein space, and denoted by $(\mathbf{P}_p(\mathbb{R}^n), d_p)$. This paper deals with the geometric structure of the p-Wasserstein space and its applications to partial differential equations (pde). The starting point is the pioneering work of Jordan–Kinderlehrer–Otto [7] where existence of the solution to the linear Fokker–Planck equation, $\partial_t\rho = \Delta\rho + \operatorname{div}(\rho\nabla U)$, is proved via a time-discrete iterative variational scheme in $(\mathbf{P}_2(\mathbb{R}^n), d_2)$. Then, it is shown that $(\mathbf{P}_2(\mathbb{R}^n), d_2)$ can be endowed with a "Riemannian" structure (see Otto [9], Carrillo–McCann–Villani [4]). Using a similar variational scheme in $(\mathbf{P}_p(\mathbb{R}^n), d_p)$, the author [1] proved existence of solutions to a larger class of pde's which includes the q-Laplacian equation ($q = p/(p-1)$), namely, $\frac{\partial\rho}{\partial t} = \operatorname{div}[\rho\nabla c^* \circ \nabla(F'(\rho) + U)]$, where $c^*(x) = |x|^q/q$ is the Legendre transform of $c(x) = |x|^p/p$, and the functions F and U satisfy some regularity assumptions. This result suggests that there is a deeper geometric structure in the p-Wasserstein space through which this pde is a gradient flow. This is precisely the aim of this work. Here, we showed that the p-Wasserstein space can be endowed with a Finsler structure, F_p. Using F_p, we give definitions of the differential and gradient of functionals on this space. Precisely, we show that the gradient of a smooth functional E on $\mathbf{P}_p(\mathbb{R}^n)$ w.r.t. F_p is $\nabla_{F_p}E(\rho) = -\operatorname{div}[\rho\nabla c^* \circ \nabla(\frac{\delta E}{\delta\rho})]$, where $\delta E/\delta\rho$ denotes the gradient of E w.r.t. the standard L^2-Euclidean structure. When $E(\rho) = \int_{\mathbb{R}^n}(F(\rho) + \rho U)\,dx$, we deduce that its gradient flow, $\partial_t\rho = -\nabla_{F_p}E(\rho)$, in the p-Wasserstein Finsler space $(\mathbf{P}_p(\mathbb{R}^n), F_p)$ is the pde studied in [1]. In particular when $p = 2$, we recover Otto's interpretation [9]. So the main result of this work can be summarized as follows: *for all $p > 1$, the p-Wasserstein space $(\mathbf{P}_p(\mathbb{R}^n), d_p)$ can be viewed as a Finsler manifold $(\mathbf{P}_p(\mathbb{R}^n), F_p)$. When $p = 2$, this Finsler structure is Riemannian. Therefore the 2-Wasserstein space $(\mathbf{P}_2(\mathbb{R}^n), d_2)$ is a Riemann–Finsler manifold $(\mathbf{P}_2(\mathbb{R}^n), F_2)$*. We end this introduction by mentioning that definitions of subdifferential and gradient flows in the p-Wasserstein space were previously given for more general functionals in [2], using a method based on metric arguments. Though our approach is different from theirs, we will show later in Remark 2 that both approaches match via a certain isomorphism. Throughout the paper, $c^*(x) = |x|^q/q$ denotes the Legendre transform of a cost function $c(x) = |x|^p/p$, where $p > 1$ and $1/p + 1/q = 1$, and for $V : \mathbb{R}^n \to \mathbb{R}$ and $\rho \in \mathbf{P}(\mathbb{R}^n)$, $\|V\|^p_{L^p_\rho(\mathbb{R}^n)} := \int_{\mathbb{R}^n} |V(x)|^p \rho(x)\,dx$.

2. Generalities on Finsler manifolds

Let M be a manifold, and denote by T_xM the tangent space at $x \in M$ and by $TM := \bigcup_{x\in M} T_xM$ the tangent bundle of M, that is the set of all pairs $(x, v) \in M \times T_xM$. A *Finsler metric* on M is a function $F : TM \to [0, \infty)$ such that:

(i) *Positivity*: $F(x, v) > 0$ for all $x \in M$ and $0 \neq v \in T_xM$;
(ii) *Positive homogeneity*: $F(x, \lambda v) = \lambda F(x, v)$ for all $\lambda > 0$, $x \in M$ and $v \in T_xM$;
(iii) *Strong convexity*: $F(x, v + v') \leqslant F(x, v) + F(x, v')$ for all $x \in M$ and $v, v' \in T_xM$, with equality (when $v, v' \neq 0$) if and only if $v = \lambda v'$ for some $\lambda > 0$.

Condition (iii) is slightly weaker than the one formulated in many differential geometry text-books (e.g. [10]) where the definition of a Finsler metric is given for a finite dimensional smooth manifold, namely:

(iii)′ *The Hessian matrix $[F^2]_{v_iv_j}(x, v)$ is positive definite for any non-zero vector $v \in T_xM$.*

But since our applications will involve only infinite dimensional spaces (space of probability densities), it is most suitable to use (iii). Of course (iii)′ implies (iii), but they are not equivalent (see [10]). If F satisfies $F(x, -v) = F(x, v)$, the Finsler metric is said to be *reversible*. In that case we have the *absolute homogeneity*: $F(x,\lambda v)=|\lambda|F(x,v)$ *for all* $\lambda\in\mathbb{R}$, $x\in M$ *and* $v\in T_xM$. When M is equipped with a Finsler metric F, we call (M,F) a *Finsler manifold*. On every tangent space T_xM at $x\in M$, F defines a norm (without the reversibility condition), $\|\ \|_{T_xM}:=F(x,.)$, called a *Minkowski norm* on T_xM. When F is reversible, then the Minkowski norm is a genuine norm. Because of (iii), the Minkowski norm satisfies the *strong triangle inequality*: $\|v+v'\|_{T_xM}\leqslant\|v\|_{T_xM}+\|v'\|_{T_xM}$ *for all* $v,v'\in T_xM$ *with equality* (*when* $v,v'\neq 0$), *if and only if* $v=\lambda v'$ *for some* $\lambda>0$. When the Minkowski norm is Euclidean at every $x\in M$ (i.e., admits an inner product $\|v\|_{T_xM}=\sqrt{\langle v,v\rangle_{T_xM}}$), then the Finsler metric is called a *Riemann–Finsler metric*, and the manifold (M,F) is then called a *Riemann–Finsler manifold*.

Finsler metrics are used to measure the length of smooth curves in a manifold. Indeed, if $c=c(t):[0,1]\to M$ is a C^1-curve and $p>1$, we define the p-length of c in (M,F) as

$$\mathbf{L}_F(c):=\big\|F\big(c(t),\dot c(t)\big)\big\|_{L^p(0,1)}=\left(\int_0^1\big[F\big(c(t),\dot c(t)\big)\big]^p\,\mathrm{d}t\right)^{1/p},\tag{2}$$

where $\dot c(t)\in T_{c(t)}M$ is the tangent vector at $c(t)\in M$ along the curve c. Actually, in many differential geometry text-books (e.g. [10]), the L^1-norm is customarily used to define the length of curves, i.e., $\mathbf{L}_F(c):=\int_0^1F(c(t),\dot c(t))\,\mathrm{d}t$. Here, we use the L^p-norm of $(0,1)\ni t\mapsto F(c(t),\dot c(t))$ as it more convenient for our applications. Equipped with this length structure, we can define the distance between two points x,y in (M,F) in the standard way, as

$$d_F(x,y):=\inf_c\big\{\mathbf{L}_F(c):\ c:[0,1]\to M\text{ is }C^1,\ c(0)=x,\ c(1)=y\big\}.\tag{3}$$

d_F is called the *distance function* of F, or the *distance induced* by F on M. A *geodesic* between two points x,y in (M,F) is defined as a length-minimizing curve with constant-speed connecting x and y, i.e., a minimizing curve $\bar c(t)$ in $d_F(x,y)$ s.t. $d_F(\bar c(s),\bar c(t))=(t-s)\,d_F(x,y)$ for all $0\leqslant s\leqslant t\leqslant 1$. Hence if $\bar c$ is such a geodesic, we have $d_F(x,y)=\mathbf{L}_F(\bar c)$. Now, consider a functional $L:M\to\mathbb{R}$ and a point $x\in M$. We define the *differential* of L at x, as the bounded linear functional, $D_FL(x)$, on the tangent space $(T_xM,\|\ \|_{TxM}:=F(x,.))$, i.e., the element on the cotangent space T_x^*M, defined by

$$\big\langle D_FL(x);v\big\rangle=\big[D_FL(x)\big](v):=\frac{\mathrm{d}}{\mathrm{d}t}L\big(c(t)\big)\bigg|_{t=0}\quad\forall v\in T_xM,\tag{4}$$

where $c:[0,1]\to M$ is an arbitrary C^1-curve emanating from $c(0)=x$ with tangent vector $\dot c(0)=v$.

The norm of the differential $D_FL(x)$ in T_x^*M is defined in the standard way, as the dual norm,

$$\big\|D_FL(x)\big\|_*=\big\|D_FL(x)\big\|_{T_x^*M}:=\sup_v\big\{\big|\big\langle D_FL(x);v\big\rangle\big|:\ v\in T_xM,\ \|v\|_{TxM}\leqslant 1\big\}.\tag{5}$$

Since the tangent space T_xM is not in general Euclidean (except when (M,F) is a Riemann–Finsler manifold), then to define the gradient of $L:M\to\mathbb{R}$, we are inspired by the definition of the *metric gradient* in normed linear spaces by Golomb and Tapia [6].

Definition 2.1. Let $p>1$ and set $q=p/(p-1)$. The p-gradient of $L:M\to\mathbb{R}$ at $x\in M$ with respect to the Finsler structure F is the unique element (if it exists), $\nabla_{F_p}L(x)$, of T_xM that satisfies

$$\big\langle D_FL(x);\nabla_{F_p}L(x)\big\rangle=\big\|\nabla_{F_p}L(x)\big\|_{T_xM}^p=\big\|D_FL(x)\big\|_*^q.\tag{6}$$

In particular when $p=2$, we recover the standard definition of the metric gradient in a normed space formulated in [6], which clearly extends the usual definition of gradient in a Hilbert space. The proof of the uniqueness of the p-gradient is a consequence of the strong convexity of the Minkowski norm. Its existence follows from the Hahn–Banach theorem *provided* T_xM *at every* $x\in M$ *is reflexive* (see [6]).

3. Application to the *p*-Wasserstein space

For simplicity, we restrict our discussion to bounded domains of $\mathbb{R}^n$. So, let Ω be an open, bounded, convex and smooth subset of $\mathbb{R}^n$, and let $p>1$. Denote by $P(\Omega)$ the set of strictly positive C^1-probability densities on Ω, and by $P_p(\Omega)$ the p-Wasserstein space $(P(\Omega),d_p)$. It is known [5], that the Monge–Kantorovich problem (1) has a unique minimizer $T(x)=x-\nabla c^*(\nabla\phi(x))$ where $\phi:\Omega\to\mathbb{R}$ is a c-concave function, i.e., $\phi(x)=\inf_{y\in\Omega}\{c(x-y)-\varphi(y)\}$ for some function $\varphi:\Omega\to\mathbb{R}$. Moreover, T is one-to-one, and its inverse $T^{-1}(y)=y-\nabla c^*(\nabla\varphi(y))$ transports ρ_1 to ρ_0. Hence,

$$d_p(\rho_0,\rho_1)^p=\int_\Omega\big|T(x)-x\big|^p\rho_0(x)\,\mathrm{d}x=\int_\Omega\big|y-T^{-1}(y)\big|^p\rho_1(y)\,\mathrm{d}y.$$

Furthermore, if $t\in[0,1]$ and $T_t(x):=(1-t)x+tT(x)$ is McCann's interpolation [8], then the curve $\bar\rho(t)=(T_t)_\#\rho_0:[0,1]\to P(\Omega)$ is the unique (constant-speed) geodesic joining ρ_0 and ρ_1 in the p-Wasserstein space $(P(\Omega),d_p)$ (see [2]).

To realize the p-Wasserstein space $P_p(\Omega)$ as a Finsler manifold $(P(\Omega), F_p)$, we must identify the tangent space $T_\rho P(\Omega)$ at every point $\rho \in P(\Omega)$, and define a Finsler metric, F_p, so that the induced distance, d_{F_p}, coincides with the p-Wasserstein distance, d_p. So matching the geodesic in a Finsler manifold with that in $(P(\Omega), d_p)$, we must have $d_p(\rho_0, \rho_1) = d_{F_p}(\rho_0, \rho_1) = \mathbf{L}_{F_p}(\bar\rho)$ which suggests via (2),

$$F_p\big(\bar\rho(t), \dot{\bar\rho}(t)\big) := \bigg(\int_\Omega \big|T(x) - x\big|^p \rho_0(x)\,\mathrm{d}x\bigg)^{1/p} = \bigg(\int_\Omega \bigg|\frac{\partial}{\partial t} T_t(x)\bigg|^p \rho_0(x)\,\mathrm{d}x\bigg)^{1/p}.$$

To get an explicit formula for F_p, we rewrite the rhs of the subsequent equation in terms of $\bar\rho(t)$ and $\dot{\bar\rho}(t) = \frac{\partial \bar\rho(t)}{\partial t}$. For that, consider the velocity field $\bar V(t,x)$ associated with the trajectory $[0,1] \times \Omega \ni (t,x) \mapsto T_t(x) \in \Omega$, i.e., $\bar V(t, T_t(x)) = \frac{\partial}{\partial t} T_t(x)$. It is easy to check that $\bar\rho(t) = (T_t)_\# \rho_0$ satisfies the transport equation

$$\dot{\bar\rho}(t,x) = -\mathrm{div}\big(\bar\rho(t,x)\bar V(t,x)\big) \quad \text{in } \Omega, \quad \text{with } \bar V(t,x)\cdot \nu = 0 \quad \text{on } \partial\Omega \quad \text{and} \quad \bar V(t,x) = \nabla c^*\big(\nabla \bar\varphi_t(x)\big), \tag{7}$$

where we use that $\bar V(t,x) = \frac{x-(T_t)^{-1}(x)}{t}$, $(T_t)^{-1}(x) = x - \nabla c^*(\nabla\varphi_t(x))$, meaning that (see [5]) $(T_t)^{-1}$ is the optimal map in $d_p(\rho_t, \rho_0)$, and $\bar\varphi_t(x) := \frac{1}{t^{1/(q-1)}}\varphi_t(x)$. The boundary condition in (7) can be seen by integrating the transport equation in (7) against a test function $\varphi \in C^1(\Omega)$, use an integration by parts, the definitions of $\bar\rho(t)$ and $\bar V(t, T_t(x))$ and then $\bar\rho(t) > 0$. Hence the formula of $F_p(\bar\rho(t), \dot{\bar\rho}(t))$ becomes:

$$\big[F_p\big(\bar\rho(t), \dot{\bar\rho}(t)\big)\big]^p = \int_\Omega \big|\bar V\big(t, T_t(x)\big)\big|^p \rho_0(x)\,\mathrm{d}x = \int_\Omega \big|\bar V(t,x)\big|^p \bar\rho(t,x)\,\mathrm{d}x := \big\|\bar V(t,.)\big\|^p_{L^p_{\bar\rho(t,.)}(\Omega)}. \tag{8}$$

Based on (7) and (8), we can formulate the following definitions:

Definition 3.1. The tangent space, $T_\rho P(\Omega)$, at $\rho \in P(\Omega)$ in the p-Wasserstein space, is the subset of $-\mathrm{div}(\rho L^p_\rho(\Omega))$ whose elements $v := -\mathrm{div}(\rho V)$ satisfy,

$$\|V\|_{L^p_\rho(\Omega)} < \infty, \quad V = \nabla c^* \circ \nabla\phi \quad \text{in } \Omega, \quad V\cdot\nu = 0 \quad \text{on } \partial\Omega, \tag{9}$$

for some $W^{1,q}_\rho(\Omega)$-function $\phi : \Omega \to \mathbb{R}$, where $q = p/(p-1)$ and $\|V\|^p_{L^p_\rho(\Omega)} := \int_\Omega |V(x)|^p \rho(x)\,\mathrm{d}x$.

Definition 3.2. The Finsler metric in the p-Wasserstein space is the nonnegative function, F_p, defined on the tangent bundle $TP(\Omega) := \bigcup_{\rho\in P(\Omega)} T_\rho P(\Omega)$ by

$$F_p(\rho, v) := \|V\|_{L^p_\rho(\Omega)} = \big\|\nabla c^* \circ \nabla\phi\big\|_{L^p_\rho(\Omega)}, \tag{10}$$

where $\rho \in P(\Omega)$ and $v \in T_\rho P(\Omega)$ with $v := -\mathrm{div}(\rho V)$, $V = \nabla c^* \circ \nabla\phi$.

The following propositions further justify these definitions:

Proposition 3.1. *If $[0,1] \ni t \mapsto \rho(t) \in P(\Omega)$ is any C^1-curve, then the variational problem*

$$\inf_{V(t)\in L^p_{\rho(t)}(\Omega)} \bigg\{ \int_\Omega \big|V(t,x)\big|^p \rho(t,x)\,\mathrm{d}x\text{:}\ \dot\rho(t) + \mathrm{div}\big(\rho(t)V(t)\big) = 0 \textit{ in } \Omega,\ V(t)\cdot\nu = 0 \textit{ on } \partial\Omega \bigg\} \tag{11}$$

has at most one minimizer V, which is characterized by

$$V(t,x) = \nabla c^* \circ \nabla\phi_t(x) \quad \textit{in } \Omega, \quad \textit{and} \quad V(t,x)\cdot\nu = 0 \quad \textit{on } \partial\Omega \tag{12}$$

for some $W^{1,q}_{\rho(t)}(\Omega)$ function $\phi_t : \Omega \to \mathbb{R}$, where $q = p/(p-1)$.

In fact, among all the velocity fields leading to the same flow $\rho(t)$, we select this minimal velocity field as the tangent vector $\dot\rho(t)$ in Definition 3.1 of the tangent space in $P_p(\Omega)$.

Proof. If V is a minimizer in (11) and $V_\epsilon(t) := V(t) + \epsilon W/\rho(t)$ a variation of V, with $\epsilon \neq 0$ and $W \in C^1_0(\Omega; \Omega)$ s.t. $\mathrm{div}\, W = 0$, we have

$$\bigg[\frac{\mathrm{d}}{\mathrm{d}\epsilon} \int_\Omega \big|V_\epsilon(t,x)\big|^p \rho(t,x)\,\mathrm{d}x\bigg]_{\epsilon=0} = p\int_\Omega \big(\nabla c \circ V(t,x)\big)\cdot W(x)\,\mathrm{d}x = 0$$

which shows that $\nabla c \circ V(t,x) = \nabla\phi_t(x)$ or $V(t,x) = \nabla c^* \circ \nabla\phi_t(x)$ for some function $\phi_t \in W^{1,q}_{\rho(t)}(\Omega)$. □

Proposition 3.2. *For any $\rho_0, \rho_1 \in P(\Omega)$, defining $\mathbf{L}_{F_p}(\rho)$ by* (2)*, we have*

$$d_p(\rho_0,\rho_1) = \inf_{\rho(t)}\big\{\mathbf{L}_{F_p}(\rho);\ \rho:[0,1]\to P(\Omega),\ \rho(0)=\rho_0,\ \rho(1)=\rho_1\big\} := d_{F_p}(\rho_0,\rho_1). \tag{13}$$

Proof. (13) is an analogue of Benamou–Brenier [3] characterization of d_2, for d_p. □

Remark 1. If $\rho \in P(\Omega)$, the Minkowski norm, $-\mathrm{div}(\rho L^p_\rho(\Omega)) \ni -\mathrm{div}(\rho V) := \nu \mapsto F_p(\rho,\nu) := \|V\|_{L^p_\rho(\Omega)}$, can be identified with the L^p_ρ-norm. Then $P_p(\Omega) = (P(\Omega), F_p)$ is a reversible Finsler manifold. Moreover, $T_\rho P(\Omega)$ is reflexive at every $\rho \in P(\Omega)$. In particular when $p = 2$, the Minkowski norm $F_2(\rho, .)$ is identified with the L^2_ρ-norm which comes from an inner product. Therefore, $P_2(\Omega) = (P(\Omega), F_2)$ is a Riemann–Finsler manifold as shown by Otto [9].

Remark 2. In [2] (in the context of bounded domains), the tangent space, $\mathrm{Tan}_\rho P(\Omega)$, of $P_p(\Omega)$ is a subset of $L^p_\rho(\Omega)$, while here, $T_\rho P(\Omega)$ is a subset of the space of distributions, $-\mathrm{div}(\rho L^p_\rho(\Omega))$, which is the image of $\mathrm{Tan}_\rho P(\Omega)$ under the isomorphism $L^p_\rho(\Omega) \ni V \mapsto -\mathrm{div}(\rho V) \in -\mathrm{div}(\rho L^p_\rho(\Omega))$.

Next we derive the gradient of functionals in the p-Wasserstein Finsler manifold $(P(\Omega), F_p)$.

Proposition 3.3. *Let $E : P(\Omega) \to \mathbb{R}$ be a functional, and ξ a $C^2(\Omega)$-vector field related to E by the rule*

$$\frac{\mathrm{d}}{\mathrm{d}t}E\big(\rho(t)\big) = \int_\Omega \xi(t,x)\dot\rho(t,x)\,\mathrm{d}x.$$

Then the gradient of E with respect to the Finsler structure F_p is

$$\nabla_{F_p}E(\rho) = -\mathrm{div}\big[\rho\nabla c^*\circ\nabla\xi\big] \quad in\ \Omega, \qquad \big[\nabla c^*\circ\nabla\xi\big]\cdot\nu = 0 \quad on\ \partial\Omega. \tag{14}$$

Therefore the gradient flow of E in the p-Wasserstein Finsler manifold $(P(\Omega), F_p)$ is the pde

$$\frac{\partial\rho}{\partial t} := -\nabla_{F_p}E(\rho) = \mathrm{div}\big[\rho\nabla c^*\circ\nabla\xi\big] \quad in\ \Omega, \qquad \big[\nabla c^*\circ\nabla\xi\big]\cdot\nu = 0 \quad on\ \partial\Omega. \tag{15}$$

Proof. First we compute the dual norm $\|D_{F_p}E(\rho)\|_*$ via (4) and (5). By definition,

$$\big\langle D_{F_p}E(\rho);\dot\rho(t)\big\rangle = \frac{\mathrm{d}}{\mathrm{d}t}E\big(\rho(t)\big) \quad\Longrightarrow\quad \big\langle D_{F_p}E(\rho);\nu\big\rangle = \int_\Omega \xi(x)\nu(x)\,\mathrm{d}x, \quad \forall\nu\in T_\rho P(\Omega). \tag{16}$$

Then using $\nu = -\mathrm{div}(\rho V)$ with $V\cdot\nu = 0$ on $\partial\Omega$, an integration by parts and Hölder inequality, we have:

$$\big|\big\langle D_{F_p}E(\rho);\nu\big\rangle\big| = \left|\int_\Omega \rho V\cdot\nabla\xi\,\mathrm{d}x\right| \leqslant \left(\int_\Omega \rho|\nabla\xi|^q\,\mathrm{d}x\right)^{\frac{1}{q}}\left(\int_\Omega \rho|V|^p\,\mathrm{d}x\right)^{\frac{1}{p}} = \|\nabla\xi\|_{L^q_\rho(\Omega)}\|\nu\|_{T_\rho P(\Omega)},$$

i.e., $\|D_{F_p}E(\rho)\|_* \leqslant \|\nabla\xi\|_{L^q_\rho(\Omega)}$. Now setting $\bar\nu = -\mathrm{div}(\rho\bar V)$ with $\bar V = \frac{1}{\lambda}\nabla c^*\circ\nabla\xi$ and $\lambda = \|\nabla\xi\|^{q/p}_{L^q_\rho(\Omega)}$, we have $\|\bar\nu\|_{T_\rho P(\Omega)} = \|\bar V\|_{L^p_\rho(\Omega)} = 1$ and $\langle D_{F_p}E(\rho);\bar\nu\rangle = \|\nabla\xi\|_{L^q_\rho(\Omega)}$. Hence, $\|D_{F_p}E(\rho)\|_* = \|\nabla\xi\|_{L^q_\rho(\Omega)}$. Next we compute $\nabla_{F_p}E(\rho)$ via (6). Since $\nabla_{F_p}E(\rho) \in T_\rho P(\Omega)$, then $\nabla_{F_p}E(\rho) = -\mathrm{div}(\rho V)$ for some $V = \nabla c^*\circ\nabla\phi \in L^p_\rho(\Omega)$ with $V\cdot\nu = 0$ on $\partial\Omega$. Then (6) reads as $\int_\Omega \rho V\cdot\nabla\xi\,\mathrm{d}x = \int_\Omega \rho|V|^p\,\mathrm{d}x = \int_\Omega \rho|\nabla\xi|^q\,\mathrm{d}x$. It is easy to check that $V = |\nabla\xi|^{q-2}\nabla\xi = \nabla c^*\circ\nabla\xi$ solves this equation. Then by uniqueness (see Definition 2.1), we deduce that $\nabla_{F_p}E(\rho)$ is given by (14). We conclude (15) by the definition of the gradient flow. □

Example 1. If E is the sum of the internal energy, potential energy and interaction energy, $E(\rho) = \int_\Omega (F(\rho) + U\rho + \frac{1}{2}(W\star\rho)\rho)\,\mathrm{d}x$, where $F : [0,\infty) \to \mathbb{R}$, $U : \Omega \to \mathbb{R}$ and $W : \mathbb{R}^n \to \mathbb{R}$ are sufficiently regular and W is even, then (15) gives that the gradient flow of E w.r.t. the Finsler structure F_p is:

$$\frac{\partial\rho}{\partial t} = \mathrm{div}\big[\rho\nabla c^*\circ\nabla\big(F'(\rho)+U+W\star\rho\big)\big] \quad \text{in } \Omega, \qquad \big[\nabla c^*\circ\nabla\big(F'(\rho)+U+W\star\rho\big)\big]\cdot\nu = 0 \quad \text{on } \partial\Omega.$$

References

[1] M. Agueh, Existence of solutions to degenerate parabolic equations via the Monge–Kantorovich theory, Adv. Differential Equations 10 (2005) 309–360.
[2] L. Ambrosio, N. Gigli, G. Savaré, Gradient Flows in Metric Spaces and in the Space of Probability Measures, Lectures in Mathematics, Birkhäuser, 2005.
[3] J.-D. Benamou, Y. Brenier, A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem, Numer. Math. 84 (2000) 375–393.
[4] J.A. Carrillo, R.J. McCann, C. Villani, Contractions in the 2-Wasserstein length space and thermalization of granular media, Arch. Ration. Mech. Anal. 179 (2006) 217–263.
[5] W. Gangbo, R.J. McCann, Optimal maps in the Monge's mass transport problem, C. R. Acad. Sci. Paris, Ser. I 321 (1995) 1653–1658.
[6] M. Golomb, R.A. Tapia, The metric gradient in normed linear spaces, Numer. Math. 20 (1972) 115–124.
[7] R. Jordan, D. Kinderlehrer, F. Otto, The variational formulation of the Fokker–Planck equation, SIAM J. Math. Anal. 29 (1998) 1–17.
[8] R.J. McCann, A convexity principle for interacting gases, Adv. Math. 128 (1997) 153–179.
[9] F. Otto, The geometry of dissipative evolution equations: the porous medium equation, Comm. Partial Differential Equations 26 (2001) 101–174.
[10] H. Rund, The Differential Geometry of Finsler Spaces, Springer-Verlag, 1959.