Wasserstein space over the Wiener space

Shizan Fang

Université de Bourgogne, Dijon,France

IPAM, April 3

This talk is based on a joint work with **J. Shao** and **K.Th. Sturm**, preprint: December 2007, see

http://math.u-bourgogne.fr/IMB/IMB2-publication.html

Our work is directly motivated by

- a book by L. Ambrosio, N. Gigli and G. Savaré: *Gradient flows in metric spaces and in the space of probability measures*;

or the Lect. Notes by L. Ambrosio and G. Savaré: *Gradient flow of probability measures.*

- a paper by Feyel and Üstünel: *Monge-Kantorovitch measure transportation on the Wiener space.*

Overview

1) The classical Bochner-Weitzenböck formula on a Riemannian manifold

$$dd^* + d^*d = \Delta + \mathrm{Ric}^\#$$

gives a deep link between the Geometry, Analysis and Probability.

2) Bakry-Emery's Γ_2 theory gives an abstraction of the above classical case and is applied to other "regular cases". To explain it, we consider the linear operator $L = \Delta + \nabla\varphi$ on a Riemannian manifold, where $\varphi \in C^2$. Assume that L admits an invariant probability measure $\mu = e^\varphi \, dx$. Consider

$$\Gamma(f,g) = \frac{1}{2}\big(fLg + gLf - L(fg)\big)$$

$$\Gamma_2(f,g) = \frac{1}{2}\big(\Gamma(f, Lg) + \Gamma(Lf, g) - L\Gamma(f,g)\big)$$
$$= (\mathrm{Ric}\nabla f, \nabla g) - \mathrm{Hess}(\varphi)(\nabla f, \nabla g).$$

Bakry-Emery's condition

$$\Gamma_2(f,f) \geq c\,\Gamma(f,f) \text{ or } \mathrm{Ric} - \mathrm{Hess}(\varphi) \geq c > 0$$

implies that

$$||P_t f - \mu(f)||_{L^2} \leq e^{-ct/2}\,||f - \mu(f)||_{L^2}, \quad \mu(f) = \int f d\mu,$$

where $P_t f$ solves the heat equation

$$\frac{dP_t f}{dt} = LP_t f, \quad P_0 f = f.$$

In fact, the stronger result

$$c \int f^2 \log \frac{|f|}{||f||_{L^2}} \, d\mu \leq \int \Gamma(f,f) \, d\mu.$$

3) Consider the non linear equation on $\mathbb{R}^d$

$$\frac{d\rho_t}{dt} = \Delta(\rho_t^m), \quad m > 1,$$

with a positive initial function $\rho_0 \in L^{m+1}(\mathbb{R}^d)$ satisfying

$$\int \rho_0 \, dx = 1 \quad \int |x|^2 \rho_0 \, dx < +\infty.$$

Let $\mathcal{P}_2^a(\mathbb{R}^d)$ be the space of probability measures on $\mathbb{R}^d$ admitting density and finite second moment. By scaling the time and the space

$$\hat{\rho}(t,x) = e^{d\alpha t}\rho(e^t, e^{\alpha t}x),$$

where $\alpha = \frac{1}{d(m-1)+2}$, Otto showed that $\hat{\rho}(t,\cdot)$ is a "gradient flow" associated to a "α-convex functional" on $\mathcal{P}_2^a(\mathbb{R}^d)$:

$$\frac{d\hat{\rho}_t}{dt} = -\nabla F(\hat{\rho}_t),$$

where $F : \mathcal{P}_2^a(\mathbb{R}^d) \to \mathbb{R}$ is defined by: $F(\rho) = \frac{1}{m-1}\int \rho^m dx + \alpha \int |x|^2 \rho \, dx$.

4) In the book by L. Ambrosio, N. Gigli and G. Savaré: *Gradient flows in metric spaces and in the space of probability measures*, Lect. in Math. ETH Zürich, Birkhäuser Verlag,

Basel, 2005, the differential structure of $\mathcal{P}_2^a(\mathbb{R}^d)$ was rigourously introduced and a general theory for gradient flows of convex functionals was established.

5) For a measured metric space (X, d, m), the notion of **Ricci** has been introduced by Lott, Sturm and Villani through the entropy functional:

$$\mathrm{Ent}_m(\gamma_t) \leq (1-t)\,\mathrm{Ent}_m(\gamma_0) + t\,\mathrm{Ent}_m(\gamma_1) - c\frac{t(1-t)}{2} W_d(\gamma_0, \gamma_1).$$

The framework of our talk is the Wiener space (X, H, μ). For example:

$$X = C_0([0,1], \mathbb{R}),\ H = \left\{h;\ \int_0^1 |\dot{h}(s)|^2\, ds < +\infty\right\} \text{ and } \mu = \text{the Wiener measure.}$$

Elements on the Wiener space

Let (X, H, μ) be a Wiener space, that is, X is a separable Banach space, H a separable Hilbert space densely and continuously embedded in X such that

$$\int_X e^{\sqrt{-1}\ell(x)}\, d\mu(x) = e^{-|i^*(\ell)|_H^2/2}, \quad \ell \in X^*(\text{dual space}),$$

where $i^* : X^* \to H$. This means that $x \to \ell(x)$ is a Gaussian Random variable of variance $|\ell|_H^2$ defined on the probability space (X, μ). For $h \in X$, we denote by $\tau_h(x) = x + h$ the translation by h. The Wiener measure μ is quasi-invariant under τ_h if and only if $h \in H$. Here is a brief explanation. The space X can be seen as $\mathbb{R}^\infty$ via Wiener representation theorem

$$x(t) = G_0\, t + \sum_{k=1}^{\infty} G_k\, \sqrt{2} \frac{\sin k\pi t}{k\pi},$$

here $\{G_k;\ k \geq 0\}$ is a sequence of independent of normal Random variables. On the space $\mathbb{R}^n$, we consider $\gamma_n(dx) = e^{-|x|^2/2} \frac{dx}{(\sqrt{2\pi})^n}$ and we have

$$\int_{\mathbb{R}^n} f(x+a)\, d\gamma_n(x) = \int_{\mathbb{R}^n} f(x) e^{<x,a>_{\mathbb{R}^n} - (1/2)|a|^2_{\mathbb{R}^n}}\, d\gamma_n(x).$$

The problem is what happens as $n \to +\infty$? For $a = (a_0, a_1, \cdots) \in \ell_2$, the term $|a|^2_{\mathbb{R}^n} \to \sum_{k=0}^{\infty} a_k^2$, the term $< x, a >_{\mathbb{R}^n} = \sum_{k=0}^{n-1} a_k x_k$ does not converge for each fixed $x \in \mathbb{R}^\infty$; however under the probability measure γ_∞, it converges almost surely to a Gaussian

Random variable. So the Cameron-Martin theorem reads as:

$$\int_X f(x+h)\,d\mu(x) = \int_X f(x)e^{Y_h(x)-(1/2)|h|_H^2}\,d\mu(x).$$

Replacing h by εh, and taking the derivative at $\varepsilon = 0$, we get

$$\int_X D_h f\,d\mu = \int_X f\,Y_h\,d\mu.$$

The space H is called Cameron-Martin space, which can be seen as the tangent space of X. If we consider $Z : X \to H$ defined by $Z(x) = h$. The term $\mathrm{div}_\mu(Z) := Y_h$ is the divergence of h with respect to μ. Let $U_t(x) = x + th$ and

$$K(x) = \exp\Big(\int_0^1 \mathrm{div}_\mu(Z)(U_{-s})\,ds\Big).$$

We see that K is justly the density in Cameron-Martin theorem. A natural geometric distance on X is induced by H:

$$d_H(x,y) = |x-y|_H \text{ if } x-y \in H; \quad d_H(x,y) = +\infty \text{ otherwise.}$$

Two basic differences comparing to $\mathbb{R}^d$:

i) $\{||x||_X \leq R\}$ is not compact, $\{|x|_H \leq R\}$ is compact in X, but of measure μ zero.
ii) The notion of moments is not suitable.

With respect to d_H, the Wasserstein distance between two probability measures ν_1 and ν_2 on X is defined by

$$W_2^2(\nu_1,\nu_2) = \inf\Big\{\int_{X\times X} |x-y|_H^2 \pi(dx,dy);\ \pi \in \mathcal{C}(\nu_1,\nu_2)\Big\}$$

W_2 could take $+\infty$, but Talagrand inequality holds:

$$W_2^2(\rho\mu,\mu) \le 2\,\mathrm{Ent}(\rho). \tag{1}$$

The results by Brenier, MaCann on optimal transport maps were extended on the Wiener space by Feyel and Üstünel (PTRF 2005):

Theorem 0.1 *Let $\nu_1 = \rho_1\,\mu$ and $\nu_2 = \rho_2\,\mu$ be such that $W_2(\nu_1,\nu_2) < +\infty$. Then there is a unique $\pi_0 \in \mathcal{C}(\nu_1,\nu_2)$ which realizes the distance; moreover $\pi_0 = (I, I+\xi)_*\nu_1$, where $\xi : X \to H$ and the map $T := I + \xi$ is invertible.*

The suitable class of probability measures is

$$\mathcal{P}^*(X) = \Big\{\nu = \rho\,\mu;\ \mathrm{Ent}(\rho) = \int_X \rho\log\rho\,d\mu < +\infty\Big\}.$$

For $\nu_1, \nu_2 \in \mathcal{P}^*(X)$, by Talagrand inequality, $W_2(\nu_1, \nu_2) < +\infty$ and the theorem yields

$$W_2^2(\nu_1, \nu_2) = \int_X |\xi(x)|_H^2 \, d\nu_1(x). \tag{2}$$

A compactness result

Theorem 0.2 *Let $R > 0$. The set $\mathcal{K}_R = \big\{\nu \in \mathcal{P}^*;\ \mathrm{Ent}(\nu) \le R\big\}$ is compact in $\mathcal{P}^*(x)$ for the narrow topology.*

Proof. Pick a compact $K \subset X$ such that $\mu(K^c) \le \varepsilon$. Let $B_H(r) = \big\{|x|_H \le R\big\}$, which is compact in X. For $\nu \in \mathcal{K}_R$, Using the theorem 0.1,

$$\nu((K + B_H(r))^c) = \int_X \mathbf{1}_{(K+B_H(r))^c}(x + \xi(x)) \, d\mu(x).$$

Splitting the integral into two parts, the above quantity is dominated by

$$\mu(K^c) + \int_K \mathbf{1}_{(K+B_H(r))^c}(x+\xi(x))\, d\mu(x).$$

The second term is majorized by

$$\mu(|\xi(x)|_H > r) \leq \frac{1}{r^2}\int_X |\xi|_H^2\, d\mu = \frac{1}{r^2}W_2^2(\nu,\mu) \leq \frac{2}{r^2}\mathrm{Ent}(\nu) \leq \frac{2R}{r^2}.$$

A result of convexity

Let $\nu_0, \nu_1 \in \mathcal{P}^*(X)$. Consider $\nu_t = (I + t\xi)_*\nu_0$. Then

$$\mathrm{Ent}(\nu_t) \leq (1-t)\mathrm{Ent}(\nu_0) + t\mathrm{Ent}(\nu_1) - \frac{t(1-t)}{2}W_2(\nu_0,\nu_1). \tag{3}$$

In particular, $\nu_t \in \mathcal{P}^*(X)$ and the subset $\mathcal{K}_R$ is convex. The space (X, d_H, μ) admits the Ricci bound 1.

Tangent spaces to $\mathcal{P}^*(X)$

For a smooth curve $t \to c(t)$ on a Riemannian manifold M, the derivative $c'(t) \in T_{c(t)}M$. The convenient substitute of smooth curves on $\mathcal{P}^*(X)$ is the class of absolutely continuous curves. We say that a curve $t \to \nu_t \in \mathcal{P}^*(X)$ is absolutely continuous if

$$W_2(\nu_{t_1}, \nu_{t_2}) \leq \int_{t_1}^{t_2} m(s)\, ds, \quad m \in L^2([0,1]).$$

Before going further, let's introduce some notions. We say that $F \in \mathrm{Cylin}(X)$ if

$$F(x) = f(e_1(x), \cdots, e_m(x)), \quad e_1, \cdots, e_m \in X^*, \ f \in C_c^\infty(\mathbb{R}^m).$$

For $F \in \mathrm{Cylin}(X)$, we define its gradient by

$$\nabla F(x) = \sum_{i=1}^{m} \frac{\partial f}{\partial \xi_i}(e_1(x), \cdots, e_m(x))\, e_i. \tag{4}$$

A function $Z : X \to H$ is said cylindrical vector field on X if

$$Z = \sum_{j=1}^{N} F_j\, h_j, \quad F_j \in \mathrm{Cylin}(X),\ h_j \in X^*. \tag{5}$$

For such a vector field Z, there is a flow of continuous map $U_t : X \to X$ such that

$$U_t(x) = x + \int_0^t Z(U_s(x))\, ds. \tag{6}$$

The flow U_t leaves the Wiener measure quasi-invariant: $(U_t)_*\mu = K_t\, \mu$.

Theorem 0.3 *Let $(\nu_t)_{t\in[0,1]}$ be an absolutely continuous curve on $\mathcal{P}^*(X)$. Then there exists a unique*

$$Z_t \in T_{\nu_t}\mathcal{P}^* = \overline{\Big\{ \sum_{i,finite} \nabla F_i;\ F_i \in \mathrm{Cylin}(X) \Big\}}^{L^2(\nu_t)},$$

such that

$$\frac{d\nu_t}{dt} + \nabla \cdot (Z_t \nu_t) = 0 \tag{7}$$

in the sense that

$$\int_{[0,1]}\int_X \Big(\alpha'(t)F(x) + \big\langle Z_t(x), \nabla F(x)\big\rangle_H \alpha(t)\Big)\, d\nu_t(x)dt = 0, \quad \alpha \in C_c^\infty(]0,1[, F \in \mathrm{Cylin}(X).$$

Definition 0.4 *We say that* $Z_t := \frac{d^o \nu_t}{dt}$ *is the derivative process of* $t \to \nu_t$*, in Otto-Ambrosio-Savaré's sense.*

Using the notation $\frac{d^o\nu_t}{dt}$, we get the following interpretation for Benamou-Brenier's formula:

$$W_2^2(\nu_0,\nu_1) = \inf\Big\{\int_0^1 \Big\|\frac{d^o\nu_t}{dt}\Big\|^2_{T_{\nu_t}\mathcal{P}^*} dt;\ \nu_t \text{ A.C. connecting } \nu_0,\nu_1\Big\}. \tag{8}$$

Let $\xi : X \to H$ be given by Theorem 0.1. Define $T_t = I + t\xi$ and $\nu_t = (T_t)_* \nu_0$ and

$V_t = \xi(T_t^{-1})$. Then for a.e. $t \in]0,1[$,

$$\frac{d^o \nu_t}{dt} = V_t, \quad W_2^2(\nu_0, \nu_1) = \int_0^1 \left\| \frac{d^o \nu_t}{dt} \right\|^2_{T_{\nu_t}\mathcal{P}^*} dt.$$

Now we shall compute the "gradient" of the entropy functional $\mathrm{Ent} : \mathcal{P}^*(X) \to \mathbb{R}$. Let U_t be the flow associated to ∇F. For $\nu_0 \in \mathcal{P}^*(X)$, we define $\nu_t = (U_t)_* \nu_0$. Then

$$\frac{d}{dt}|_{t=0} \mathrm{Ent}(\nu_t) = \int_X LF \, d\nu_0, \tag{9}$$

where $LF = \mathrm{div}_\mu(\nabla F)$ admits the expression:

$$LF = -\sum_{i,j=1}^m (\partial_j \partial_i f) \left\langle e_i, e_j \right\rangle_H + \sum_{i=1}^m (\partial_i f)\, e_i(x).$$

Proof. We have $(U_t)_*\mu = K_t\,\mu$ with $K_t(x) = \exp\Big(\int_0^t \operatorname{div}_\mu(\nabla F)(U_{-s}(x))\,ds\Big)$, here div_μ is the divergence with respect to μ:

$$\int_X F \operatorname{div}_\mu(Z)\,d\mu = \int_X \big\langle \nabla F, Z\big\rangle_H\,d\mu.$$

For the cylindrical vector field given in (5), we have

$$\operatorname{div}_\mu(Z) = \sum_{j=1}^N \Big(F_j\,h_j(x) - \big\langle \nabla F_j, h_j\big\rangle_H\Big).$$

If we denote by $\nu_t = (U_t)_*\nu_0 = \rho_t\mu$, we have $\rho_t = \rho_0(U_{-t})K_t$ and

$$\operatorname{Ent}(\rho_t) = \operatorname{Ent}(\rho_0) + \int_X \log K_t(U_t)\ \rho_0 d\mu.$$

Taking the derivative under the integral is guaranteed by

$$||K_t||_{L^p}^p \leq \int_X \exp\Big(\frac{p^2}{p-1}|\mathrm{div}_\mu(\nabla F)|\Big)\, d\mu.$$

If $\nu_0 = \rho_0\,\mu$ with good $\rho_0 \in \mathrm{Cylin}(X)$, then

$$(\partial_{\nabla F}\mathrm{Ent})(\nu_0) := \frac{d}{dt}|_{t=0}\mathrm{Ent}(\nu_t) = \int_X \big\langle \nabla F, \nabla \log \rho_0 \big\rangle_H \, d\nu_0.$$

We say that the gradient $\nabla\mathrm{Ent}$ exists at ν_0; more general, we say that the gradient $\nabla\mathrm{Ent}$ exists at $\nu \in \mathcal{P}^*(X)$ if there exists $v \in T_\nu\mathcal{P}^*$ such that

$$\big\langle v, \nabla F \big\rangle_{T_\nu \mathcal{P}^*} = (\partial_{\nabla F}\mathrm{Ent})(\nu) \quad \text{for all } F \in \mathrm{Cylin}(X). \tag{10}$$

Theorem 0.5 *Fix $\nu_0 \in \mathcal{P}^*(X)$. Then for any $\eta > 0$, there exists a unique $\hat{\nu} \in \mathcal{P}^*$ such that*

$$\frac{1}{2}W_2^2(\nu_0, \hat{\nu}) + \eta\,\mathrm{Ent}(\hat{\nu}) = \inf\Big\{\frac{1}{2}W_2^2(\nu_0, \nu) + \eta\,\mathrm{Ent}(\nu);\ \nu \in \mathcal{P}^*\Big\}.$$

Moreover the gradient $\nabla\mathrm{Ent}$ exists at $\hat{\nu}$.

Proof. Using the compactness result and the semi-lower continuity of

$$\nu \longrightarrow \frac{1}{2}W_2^2(\nu_0, \nu) + \eta\,\mathrm{Ent}(\nu),$$

such a $\hat{\nu}$ exists. Let $(U_t)_{t\in\mathbb{R}}$ be the flow associated to $Z = \nabla F$. Let $\pi \in \mathcal{C}(\nu_0, \hat{\nu})$ be the optimal plan. We define $\pi_t \in \mathcal{C}(\nu_0, (U_t)_*\hat{\nu})$ by $\pi_t = (I \times U_t)_*\pi$:

$$\int_{X\times X} \psi(x, y)\pi_t(dx, dy) = \int_{X\times X} \psi(x, U_t(y))\pi(dx, dy).$$

Then we have

$$W_2^2(\nu_0, (U_t)_*\hat{\nu}) - W_2^2(\nu_0, \hat{\nu}) \leq \int_{X\times X} \Big\{ |x - U_t(y)|_H^2 - |x-y|_H^2 \Big\} \, \pi(dx, dy).$$

Since $|x - U_t(y)|_H^2 = |x - y + y - U_t(y)|_H^2 = |x-y|_H^2 + |y - U_t(y)| + 2\big\langle x-y, y-U_t(y)\big\rangle_H$, it follows that

$$\overline{\lim_{t\downarrow 0}} \frac{1}{2t}\Big[W_2^2(\nu_0, (U_t)_*\hat{\nu}) - W_2^2(\nu_0, \hat{\nu}) \Big] \leq -\int_{X\times X} \big\langle Z(y), x-y \big\rangle_H \, \pi(dx, dy).$$

By construction of $\hat{\nu}$, for $t > 0$,

$$\frac{\eta}{t}\Big[\mathrm{Ent}((U_t)_*\hat{\nu}) - \mathrm{Ent}(\hat{\nu}) \Big] + \frac{1}{2t}\Big[W_2^2(\nu_0, (U_t)_*\hat{\nu}) - W_2^2(\nu_0, \hat{\nu}) \Big] \geq 0.$$

Letting $t \downarrow 0$ gives

$$\eta\,(\partial_{\nabla F}\mathrm{Ent})(\hat{\nu}) - \int_{X\times X} \big\langle Z(y), x-y \big\rangle_H \,\pi(dx,dy) \geq 0.$$

Changing F into $-F$, we get the equality:

$$(\partial_{\nabla F}\mathrm{Ent})(\hat{\nu}) = \frac{1}{\eta}\int_{X\times X} \big\langle Z(y), x-y \big\rangle_H \,\pi(dx,dy).$$

Let ξ be given in Theorem 0.1; then $T_1 = I + \xi$ pushes ν_0 forward to $\hat{\nu}$. We have

$$(\partial_{\nabla F}\mathrm{Ent})(\hat{\nu}) = \frac{1}{\eta}\int_X \big\langle Z(T_1), -\xi \big\rangle_H \,d\nu_0 = -\int_X \big\langle \nabla F, \xi(T_1^{-1})/\eta \big\rangle_H \,d\hat{\nu}.$$

Note that $\int_X |\xi(T^{-1})|_H^2\, d\hat{\nu} = \int_X |\xi|_H^2\, d\nu_0 = W_2^2(\nu_0, \hat{\nu}) < +\infty$. So the gradient

$$(\nabla \mathrm{Ent})(\hat{\nu}) \in T_{\hat{\nu}}\mathcal{P}^*$$

exists. □

We denote $\hat{\nu}$ by $\nu^{(1)}$. Using this result step by step, we get a sequence of $\nu^{(n)} \in \mathcal{P}^*$. According to Jordan, Kinderlehrer and Otto's approach, we consider

$$\nu_\eta(t, dx) = \sum_{k=1}^{N+1} \nu^{(k)}(dx)\, \mathbf{1}_{](k-1)\eta, k\eta]}(t), \quad \text{where } N\eta \leq 1.$$

We see that $\nu_\eta(t, \cdot) \in \mathrm{Dom}(\nabla \mathrm{Ent})$ for $t > 0$. Again by compactness result,

$$\nu_\eta(t, dx)dt \text{ converges weakly to } \nu_t(dx)dt = \rho(x, t)d\mu(x)dt.$$

Theorem 0.6 *The curve $t \to \nu_t \in \mathcal{P}^*(X)$ is absolutely continuous and $\nu_t \in Dom(\nabla \mathrm{Ent})$ such that*

$$\frac{d^o \nu_t}{dt} = -(\nabla \mathrm{Ent})(\nu_t).$$

THANK YOU!