Optimal transport: introduction, applications and derivation
Lecture 3: Existence and uniqueness for the Fokker-Planck equation with pairwise interactions

Contents

In this lecture, we will be interested in the following PDE:

$$\begin{cases} \partial_t \rho = \operatorname{div}(\rho(\nabla V + \nabla W * \rho)) + \Delta \rho, \\ \rho|_{t=0} = \rho_0 \in \mathcal{P}(\mathbb{T}^d), \end{cases} \tag{1}$$

where V and W are nonnegative functions in $C^1(\mathbb{T}^d)$, and W is even. This PDE is to be understood in a distributional sense, that is ρ is called a solution if it satisfies the following property.

Definition 1. Let $\rho_0 \in \mathcal{P}(\mathbb{T}^d)$. A curve $\rho \in C^0(\mathbb{R}_+; \mathcal{P}(\mathbb{T}^d))$ is called a solution of (1) provided for all $\varphi \in C^1_c(\mathbb{R}_+ \times \mathbb{T}^d)$,

$$\int \varphi(0,x)\,\mathrm{d}\rho_0(x) + \int_0^{+\infty}\int \{\partial_t\varphi(t,x) + v(t,x)\cdot\nabla\varphi(t,x) + \Delta\varphi(t,x)\}\,\mathrm{d}\rho(t,x)\,\mathrm{d}t = 0,$$

where v is the continuous vector field defined for all $(t,x) \in \mathbb{R}_+ \times \mathbb{T}^d$ by

$$v(t,x) = -\nabla V(x) - \int \nabla W(x-y)\,\mathrm{d}\rho(t,y).$$

As we saw in the last lecture, equation (1) is the gradient flow in the Wasserstein space of the functional

$$\mathcal{E} : \rho \in \mathcal{P}(\mathbb{T}^d) \longmapsto \begin{cases} \displaystyle\int \left\{ V(x) + \frac{1}{2} W * \rho(x) + \log\rho(x) \right\} \mathrm{d}\rho(x) & \text{if } \rho \ll \text{Leb}, \\ +\infty & \text{else.} \end{cases}$$

The goal of this lecture is to show the following theorem, only using optimal transport tools.

Theorem 2. *Let $\rho_0 \in \mathcal{P}(\mathbb{T}^d)$ be such that $\mathcal{E}(\rho_0) < +\infty$. Then equation (1) admits a solution in the sense of Definition 1. If in addition, V and W are of class C^2, then this solution is unique.*

1 Existence: the JKO scheme

1.1 In the Euclidean space

Let us imagine that we are aiming to solve the ODE

$$\begin{cases} \dot{X}_t = -\nabla F(X_t), \\ X_0 = x, \end{cases} \tag{2}$$

where F is a nonnegative C^1 function on the Euclidean space $\mathbb{R}^n$ and $x \in \mathbb{R}^n$. A possible approach is to introduce the following scheme, called minimizing movement scheme, labeled by a positive number $\tau > 0$ interpreted as a time step:

$$\begin{cases} X_0^\tau = x, \\ \forall n \in \mathbb{N}, \quad X_{n+1}^\tau \in \arg\min_X \dfrac{|X - X_n^\tau|^2}{2\tau} + F(X). \end{cases} \tag{3}$$

Of course, the second line simply means that X_{n+1}^τ is a minimizer of the function in the right hand side. Such a minimizer exists because as $\tau > 0$, the function to minimize is coercive on $\mathbb{R}^n$. At first order, the optimality condition for this minimization problem writes

$$X_{n+1}^\tau = X_n^\tau - \tau \nabla F(X_{n+1}^\tau). \tag{4}$$

Writing $\in$ instead of $=$ is just a way to insist on the fact that uniqueness of the minimizer is not necessary to give a meaning to the scheme. Actually, in the reasonable case when $F \in C^2$, it is clear that such a minimizer will be unique for sufficiently small values of τ, but we are not interested in uniqueness issues in this section so we will not develop that idea further.

The theorem we prove in this context is the following.

Theorem 3. *Given $\tau > 0$, let (X_n^τ) be a sequence satisfying* (3). *Let us define the curve $X^\tau : \mathbb{R}_+ \to \mathbb{R}^n$ by*

$$\forall n \in \mathbb{N}, \forall s \in [0,1], \quad X^\tau((n+s)\tau) := (1-s)X_n^\tau + sX_{n+1}^\tau.$$

Then the family $(X^\tau)_{\tau>0}$ is precompact in $C^0_{\mathrm{loc}}(\mathbb{R}_+;\mathbb{R}^n)$ and any of its limiting curve satisfies (2) *in the classical sense.*

Remark 4. When F is C^2, then the Cauchy-Lipschitz theorem applies and the solution of (2) is unique. In that case, the whole family converges towards this unique solution, but once again we do not want to address here uniqueness issues.

Proof. Step 1: Compactness.

For all $n \in \mathbb{N}$, by optimality of X_{n+1}^τ, we have

$$\frac{|X_{n+1}^\tau - X_n^\tau|^2}{2\tau} + F(X_{n+1}^\tau) \leq F(X_n^\tau).$$

Therefore, summing this inequality over n and using nonnegativity of F, we find that

$$\sum_{n=0}^{+\infty} \frac{|X_{n+1}^\tau - X_n^\tau|^2}{2\tau} \leq F(x).$$

But the sum in the right hand side is nothing but

$$\frac{1}{2}\int_0^{+\infty} |\dot{X}^\tau(t)|^2 \,\mathrm{d}t.$$

Therefore, for all $0 \leq s < t$, we find that

$$|X^\tau(t) - X^\tau(s)| = \left| \int_s^t \dot{X}^\tau(u) \, \mathrm{d}u \right| \leq \sqrt{\int_s^t |\dot{X}^\tau(u)|^2 \, \mathrm{d}u} \sqrt{t-s} \leq \sqrt{2F(x)} \sqrt{t-s}.$$

Therefore, $(X^\tau)_{\tau>0}$ has a modulus of continuity which is uniform in τ, and compactness in $C^0_{\mathrm{loc}}(\mathbb{R}_+; \mathbb{R}^n)$ follows from the Ascoli-Arzela theorem. Now, up to extraction, we assume that the whole family (X^τ) converges locally uniformly towards a limiting curve X. Note that the uniform modulus of continuity also implies that

$$\begin{aligned} \left| X^\tau \left(\left\lceil \frac{t}{\tau} \right\rceil \tau \right) - X(t) \right| &\leq \left| X^\tau \left(\left\lceil \frac{t}{\tau} \right\rceil \tau \right) - X^\tau(t) \right| + |X^\tau(t) - X(t)| \\ &\leq \sqrt{2F(x)} \sqrt{\tau} + |X^\tau(t) - X(t)| \underset{\tau \to 0}{\longrightarrow} 0, \end{aligned} \tag{5}$$

locally uniformly in $t \in \mathbb{R}_+$.

Step 2: Identification of the limit.

Now, to identify the equation solved by X, we need to pass to the limit in (4). To do so, we observe that for all $0 \leq s < t$, we have

$$X^\tau(t) - X^\tau(s) = - \int_s^t \nabla F \left(X^\tau \left(\left\lceil \frac{u}{\tau} \right\rceil \tau \right) \right) \mathrm{d}u.$$

We can pass to the limit in this identity because of (5) and the continuity of ∇F. For instance, a way to do it is to observe that

$$\{X^\tau(u),\, u \in [s, t+\tau],\, \tau > 0\} \subset \mathbb{R}^n$$

is relatively compact. Therefore, ∇F is bounded on this set and the dominated convergence theorem applies. □

1.2 In the Wasserstein space

Now, our goal is to mimic the proof of the previous subsection, but for curves valued in the Wasserstein space instead of $\mathbb{R}^n$. To do it, we replace (3) by

$$\begin{cases} \rho_0^\tau = \rho_0, \\ \forall n \in \mathbb{N}, \quad \rho_{n+1}^\tau \in \arg\min\limits_{\rho} \dfrac{W_2^2(\rho_n^\tau, \rho)}{2\tau} + \mathcal{E}(\rho), \end{cases} \tag{6}$$

where $\rho_0 \in \mathcal{P}(\mathbb{T}^d)$. To justify the existence of such a sequence $(\rho_n^\tau)_{n \in \mathbb{N}}$ for a given $\tau > 0$, we have to use the following lemma.

Lemma 5. *Let $\mu \in \mathcal{P}(\mathbb{T}^d)$. The functional*

$$\rho \in \mathcal{P}(\mathbb{T}^d) \longmapsto \frac{W_2^2(\mu, \rho)}{2\tau} + \mathcal{E}(\rho)$$

admits at least one minimizer.

Proof. The set $\mathcal{P}(\mathbb{T}^d)$ is compact for the topology of narrow convergence. Therefore, we just have to prove that the functional is lower semi-continuous. The distance part is continuous, because the Wasserstein distance metrizes the narrow topology. So we just need to show that $\mathcal{E}$ is l.s.c.. But it is clear that

$$\rho \in \mathcal{P}(\mathbb{T}^d) \longmapsto \int V \,\mathrm{d}\rho \qquad \text{and} \qquad \rho \in \mathcal{P}(\mathbb{T}^d) \longmapsto \frac{1}{2}\int W * \rho \,\mathrm{d}\rho$$

are narrowly continuous. So it remains to prove that the entropy, namely

$$H : \rho \in \mathcal{P}(\mathbb{T}^d) \longmapsto \begin{cases} \displaystyle\int \log \rho(x) \,\mathrm{d}\rho(x) & \text{if } \rho \ll \text{Leb}, \\ +\infty & \text{else}, \end{cases}$$

is l.s.c. for the narrow topology. But this is a consequence of the following formula, valid for all $\rho \in \mathcal{P}(\mathbb{T}^d)$,

$$H(\rho) = \sup_{\varphi \in C(\mathbb{T}^d)} \int \varphi(x) \,\mathrm{d}\rho(x) - \int \exp(\varphi(x) - 1) \,\mathrm{d}x.$$

□

Therefore, the goal is now to prove the following analogue of Theorem 3.

Theorem 6. *Given $\tau > 0$ and $\rho_0 \in \mathcal{P}(\mathbb{T}^d)$ with $\mathcal{E}(\rho_0) < +\infty$, let $(\rho_n^\tau)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{P}(\mathbb{T}^d)$ satisfying* (6). *For a given $n \in \mathbb{N}$ and $s \in [0,1]$, call $\rho^\tau((n+s)\tau)$ the position at time s of the Benamou-Brenier interpolation between ρ_n^τ and ρ_{n+1}^τ. The family (ρ^τ) is precompact in $C^0_{\mathrm{loc}}(\mathbb{R}_+; \mathcal{P}(\mathbb{T}^d))$, where $\mathcal{P}(\mathbb{T}^d)$ is endowed with the narrow topology, and any of it limit points is a distributional solution of* (1)*.*

We first prove a series of lemmas before attacking the proof of this theorem.

Lemma 7 (Time rescaling of Benamou-Brenier)**.** *Let (ρ, v) be a solution of the continuity equation in the sense of the previous chapter, on the time interval $[s,t]$, where $s < t$. Then*

$$\frac{W_2^2(\rho(s), \rho(t))}{t - s} \le \int_s^t \int |v|^2 \,\mathrm{d}\rho(u) \,\mathrm{d}u.$$

Proof. Let us define for $\theta \in [0,1]$ and $x \in \mathbb{T}^d$:

$$\bar{\rho}(\theta) := \rho((1-\theta)s + \theta t) \qquad \text{and} \qquad \bar{v}(\theta, x) := (t-s)v((1-\theta)s + \theta t, x).$$

With this choice, it is straightforward to check that $(\bar{\rho}, \bar{v})$ is a solution of the continuity equation between times 0 and 1, with $\bar{\rho}(0) = \rho(s)$ and $\bar{\rho}(1) = \rho(t)$. Therefore, by the Benamou-Brenier formula,

$$W_2^2(\rho(s), \rho(t)) \le \int_0^1 \int |\bar{v}|^2 \,\mathrm{d}\bar{\rho}(\theta) \,\mathrm{d}\theta.$$

The result follows from changing the variable according to $u = (1-\theta)s + \theta t$. □

Lemma 8 (The optimizer in Benamou-Brenier is defined for all time)**.** *Let $\mu, \nu \in \mathcal{P}(\mathbb{T}^d)$, (ρ, v) be an optimal solution of the continuity equation for the Benamou-Brenier formulation of optimal transport between μ and ν and P be a corresponding generalized flow. Then the formula*

$$\forall \xi \in C(\mathbb{T}^d; \mathbb{R}^d), \qquad \int \xi \cdot v(t) \, \mathrm{d}\rho(t) = \int \xi(\omega(t)) \cdot \dot{\omega}(t) \, \mathrm{d}P(\omega) \tag{7}$$

defines a version of v which is defined for $\underline{\text{all}}$ $t \in [0,1]$ in $L^2(\rho(t))$. Moreover, the corresponding moment $m = \rho v$ is continuous with respect to time with values in vector valued Radon measures on $\mathbb{T}^d$, endowed with the topology of narrow convergence. If in addition, a test function ξ is Lipschitz, for all $0 \leq s \leq t \leq 1$,

$$\left| \int \xi \cdot \mathrm{d}m(t) - \int \xi \cdot \mathrm{d}m(s) \right| \leq (t-s) \mathrm{Lip}(\xi) W_2^2(\mu, \nu). \tag{8}$$

Proof. What we know from the last chapter is that P only charge geodesic curves, hence curves with constant speed, and such that for almost all t,

$$\dot{\omega}(t) = v(t, \omega(t)).$$

Take t_0 be a point such that this identity holds for P-almost all ω, and call

$$\mathcal{V}(\omega) := v(t_0, \omega(t_0)).$$

The map $\mathcal{V}$ belongs to $L^2(P)$, because

$$\frac{1}{2} \int |\mathcal{V}|^2 \, \mathrm{d}P = \mathcal{A}(P).$$

Then, for any $t \in [0,1]$, the right-hand side in (7) can be estimated as

$$\left| \int \xi(\omega(t)) \cdot \mathcal{V}(\omega) \, \mathrm{d}P(\omega) \right| \leq \sqrt{2\mathcal{A}(P)} \|\xi\|_{L^2(\rho(t))},$$

hence defines a continuous linear form on $L^2(\rho(t))$, so it indeed defines a velocity $v(t)$, which has to coincide with the v we started from by definition of P.

To check the continuity with respect to time of m, take $0 \leq s \leq t \leq 1$ and $\xi \in C(\mathbb{T}^d; \mathbb{R}^d)$. We have

$$\left| \int \xi \cdot \mathrm{d}m(t) - \int \xi \cdot \mathrm{d}m(s) \right| \leq \int |\xi(\omega(t)) - \xi(\omega(s))| |\mathcal{V}(\omega)| \, \mathrm{d}P(\omega),$$

which converges to 0 by dominated convergence. If in addition, ξ is Lipschitz, then we can pursue this estimate by using $\xi(\omega(t)) - \xi(\omega(s)) \leq (t-s)\mathrm{Lip}(\xi)|\mathcal{V}(\omega)|$, and deduce (8). $\square$

Now, we are ready to prove Theorem 6.

Proof of Theorem 6. Step 1: Notations, solutions of the continuity equation.

Let $\tau > 0$ and $n \in \mathbb{N}$. Let (ρ, v) be the solution of the Benamou-Brenier problem between ρ_n^τ and ρ_{n+1}^τ. We defined in the statement of the theorem, for all $s \in [0,1]$,

$$\rho^\tau((n+s)\tau) := \rho(s).$$

Also, we define

$$v^\tau((n+s)\tau) := \frac{1}{\tau} v((n+s)\tau) \qquad \text{and} \qquad m^\tau((n+s)\tau) := \rho^\tau((n+s)\tau) v^\tau((n+s)\tau).$$

Doing so, we observe that (ρ^τ, v^τ) is a distributional solution of the continuity equation on $\mathbb{R}_+$.

Step 2: Compactness in $C_{\mathrm{loc}}(\mathbb{R}_+; \mathcal{P}(\mathbb{T}^d))$.

The exact same argument as in the Euclidean case provides

$$\sum_{n=0}^{+\infty} \frac{W_2^2(\rho_n^\tau, \rho_{n+1}^\tau)}{2\tau} \leq \mathcal{E}(\rho_0.)$$

But from the previous step and Lemma 7, we see that for all $0 \leq s < t$,

$$\frac{W_2^2(\rho^\tau(s), \rho^\tau(t))}{2(t-s)} \leq \frac{1}{2} \int_s^t \!\! \int |v^\tau(u)|^2 \, \mathrm{d}\rho(u) \, \mathrm{d}u.$$

On the other hand, by definition of (ρ^τ, v^τ)

$$\frac{W_2^2(\rho_n^\tau, \rho_{n+1}^\tau)}{2\tau} = \frac{1}{2} \int_{n\tau}^{(n+1)\tau} |v^\tau(u)|^2 \, \mathrm{d}\rho^\tau(u) \, \mathrm{d}u.$$

Therefore, we conclude that as in the Euclidean case,

$$W_2(\rho^\tau(s), \rho^\tau(t)) \leq \sqrt{2\mathcal{E}(\rho_0)} \sqrt{t-s}.$$

So the Ascoli-Arzela theorem applies and compactness follows.

Step 3: First variation in one step of the minimization movement scheme: strategy.

Now, we consider $\mu \in \mathcal{P}(\mathbb{T}^d)$ with $\mathcal{E}(\mu) < +\infty$, and ν a minimizer of

$$J^\tau : \rho \longmapsto \frac{W_2^2(\mu, \rho)}{2\tau} + \mathcal{E}(\rho).$$

Our goal is to derive some optimality condition for ν. Of course, these optimality conditions will then be used for each measure of the sequence (ρ_n^τ). The strategy to compute these optimality conditions is to compare the value $J^\tau(\nu)$ with $J^\tau(\nu^\varepsilon)$, for ν^ε a small variation of ν. The question is hence what kind of variations do we consider?

To perturb ν, we consider a vector field $\xi \in C^1(\mathbb{T}^d; \mathbb{R}^d)$, we call $\Phi = \Phi(t,x)$, $t \in \mathbb{R}$, $x \in \mathbb{T}^d$, the associated Cauchy-Lipschitz flow, and for $\varepsilon > 0$, $\nu^\varepsilon := \Phi(\varepsilon)_{\#}\nu$. Somehow, we use the "horizontal" structure on the space of measures. Another proof relying on the "vertical" structure can be find in [1]. Now, the goal of the next steps will be to upper-bound $J^\tau(\nu^\varepsilon)$ at order one in ε.

Step 4: First variation of the potential terms.

As a warm up, let us give an upper bound for

$$\int V \,\mathrm{d}\nu^\varepsilon \qquad \text{and} \qquad \frac{1}{2}\int W * \nu^\varepsilon \,\mathrm{d}\nu^\varepsilon.$$

We have

$$\int V \,\mathrm{d}\nu^\varepsilon = \int V(\Phi(\varepsilon, x))\,\mathrm{d}\nu(x) = \int V\,\mathrm{d}\nu(x) + \varepsilon \int \nabla V(x) \cdot \xi(x)\,\mathrm{d}\nu(x) + o(\varepsilon).$$

In the same way,

$$\begin{aligned}
\frac{1}{2}\int W * \nu^\varepsilon \,\mathrm{d}\nu^\varepsilon &= \frac{1}{2}\iint W(\Phi(\varepsilon, y) - \Phi(\varepsilon, x))\,\mathrm{d}\nu(y)\,\mathrm{d}\nu(x) \\
&= \frac{1}{2}\iint W * \nu \,\mathrm{d}\nu + \frac{\varepsilon}{2}\iint \nabla W(y - x)\cdot(\xi(y) - \xi(x))\,\mathrm{d}\nu(y)\,\mathrm{d}\nu(x) + o(\varepsilon) \\
&= \frac{1}{2}\iint W * \nu\,\mathrm{d}\nu + \varepsilon \int \nabla W * \nu\,\mathrm{d}\nu + o(\varepsilon).
\end{aligned}$$

Step 5: First variation of the entropic term.

A quick computation using changes of variables provides the following formula, when identifying ν and ν^ε with their densities with respect to Lebesgue:

$$\nu^\varepsilon(x) = \nu(\Phi(-\varepsilon, y))\det \mathrm{D}\Phi(-\varepsilon, y).$$

Therefore, we have

$$\begin{aligned}
\int \log \nu^\varepsilon \,\mathrm{d}\nu^\varepsilon &= \int \log \nu^\varepsilon(\Phi(\varepsilon, x))\,\mathrm{d}\nu(x) \\
&= \int \Big\{ \log \nu(x) + \log\det \mathrm{D}\Phi(-\varepsilon, \Phi(\varepsilon, x))\Big\}\,\mathrm{d}\nu(x) \\
&= H(\nu) + \int \log\det \mathrm{D}\Phi(-\varepsilon, \Phi(\varepsilon, x))\,\mathrm{d}\nu(x) \\
&= H(\nu) + \int \log\det \mathrm{D}\Phi(-\varepsilon, y)\,\mathrm{d}\nu^\varepsilon(y)
\end{aligned}$$

Now, observe that

$$\det \mathrm{D}\Phi(-\varepsilon, y) = 1 - \varepsilon \operatorname{div}\xi(y) + o(\varepsilon),$$

where o is uniform in y. Hence,

$$\int \det \mathrm{D}\Phi(-\varepsilon, y)\, \mathrm{d}\nu^\varepsilon(y) = 1 - \varepsilon \int \operatorname{div} \xi \, \mathrm{d}\nu^\varepsilon + o(\varepsilon),$$
$$= 1 - \varepsilon \int \operatorname{div} \xi \, \mathrm{d}\nu + o(\varepsilon).$$

Step 6: First variation of the distance term.

The derivation of the distance term goes as follows. Let (ρ, v) be an optimal solution of the continuity equation given by the Benamou-Brenier formula, and P be a corresponding generalized flow. Remember that because of Lemma 8, v is well defined for all time $t \in [0,1]$. Let Ψ^ε be the map

$$\Psi^\varepsilon : \omega \in C([0,1];\mathbb{T}^d) \longmapsto \Big(t \mapsto \Phi(\varepsilon t, \omega(t))\Big) \in C([0,1];\mathbb{T}^d). \tag{9}$$

Finally, let $P^\varepsilon := {\Psi^\varepsilon}_\# P$. We have $X_{0\#}P^\varepsilon = \mu$ and $X_{1\#}P^\varepsilon = \nu^\varepsilon$. Therefore

$$\frac{1}{2} W_2^2(\mu, \nu^\varepsilon) \leq \mathcal{A}(P^\varepsilon).$$

But we have

$$\mathcal{A}(P^\varepsilon) = \iint_0^1 \left| \frac{\mathrm{d}}{\mathrm{d}t} \Phi(\varepsilon t, \omega(t)) \right|^2 \mathrm{d}t \, \mathrm{d}P(\omega) = \iint_0^1 \left| \dot\omega(t) + \frac{\mathrm{d}}{\mathrm{d}t} \Phi(\varepsilon t, \omega(t)) - \dot\omega(t) \right|^2 \mathrm{d}t \, \mathrm{d}P(\omega).$$

A quick computation gives

$$\frac{\mathrm{d}}{\mathrm{d}t} \Phi(\varepsilon t, \omega(t)) - \dot\omega(t) = \varepsilon \xi(\omega(t)) + \{\mathrm{D}\Phi(\varepsilon t, \omega(t)) - \mathrm{Id}\} \cdot \dot\omega(t) = O(\varepsilon),$$

where the "O" is uniform in ω as soon as it is chosen in the set of geodesics of $\mathbb{T}^d$. Therefore

$$\mathcal{A}(P^\varepsilon) \leq \mathcal{A}(P) + \iint_0^1 \dot\omega(t) \cdot \frac{\mathrm{d}}{\mathrm{d}t} \{\Phi(\varepsilon t, \omega(t)) - \omega(t)\} \, \mathrm{d}t \, \mathrm{d}P(\omega) + o(\varepsilon).$$

But for P-almost all curve, $\dot\omega(t) = \mathcal{V}(\omega)$ does not depend on t. So we find

$$\begin{aligned}
\mathcal{A}(P^\varepsilon) &\leq \mathcal{A}(P) + \int \mathcal{V}(\omega) \cdot \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} \{\Phi(\varepsilon t, \omega(t)) - \omega(t)\} \, \mathrm{d}t \, \mathrm{d}P(\omega) + o(\varepsilon) \\
&= \mathcal{A}(P) + \int \mathcal{V}(\omega) \cdot \{\Phi(\varepsilon, \omega(1)) - \omega(1)\} \, \mathrm{d}P(\omega) + o(\varepsilon) \\
&= \mathcal{A}(P) + \varepsilon \int \mathcal{V}(\omega) \cdot \xi(\omega(1)) \, \mathrm{d}P(\omega) + o(\varepsilon) \\
&= \mathcal{A}(P) + \varepsilon \int \dot\omega(1) \cdot \xi(\omega(1)) \, \mathrm{d}P(\omega) + o(\varepsilon) = \mathcal{A}(P) + \varepsilon \int v(1) \cdot \xi \, \mathrm{d}\nu + o(\varepsilon).
\end{aligned}$$

Step 7: Conclusion of the first variation analysis.

Gathering everything and using the optimality of ν, we find for all $\varepsilon > 0$:

$$J^\tau(\nu) \leq J^\tau(\nu^\varepsilon) \leq J^\tau(\nu) + \varepsilon \int \left\{ \frac{v(1)}{\tau} + \nabla V + \nabla W * \nu \right\} \cdot \xi \, \mathrm{d}\nu - \varepsilon \int \operatorname{div} \xi \, \mathrm{d}\nu + o(\varepsilon).$$

Dividing by ε and letting ε go to 0, we conclude that for all $\xi \in C^1(\mathbb{T}^d; \mathbb{R}^d)$,

$$\int \left\{ \frac{v(1)}{\tau} + \nabla V + \nabla W * \nu \right\} \cdot \xi \, \mathrm{d}\nu - \int \operatorname{div} \xi \, \mathrm{d}\nu = 0,$$

which means that in the sense of distributions,

$$\frac{m(1)}{\tau} = -\Big\{ \nabla V + \nabla W * \nu \Big\} \nu - \nabla \nu.$$

In terms of the objects defined at Step 1, we conclude that for all $n \in \mathbb{N}^*$,

$$m^\tau(n\tau) = -\Big\{ \nabla V + \nabla W * \rho^\tau(n\tau) \Big\} \rho^\tau(n\tau) - \nabla \rho^\tau(n\tau).$$

Step 8: Quantitative continuity of the moments and conclusion.

In this last step, we assume that the whole family $(\rho^\tau)_{\tau>0}$ converges in $C_{\mathrm{loc}}(\mathbb{R}_+; \mathcal{P}(\mathbb{T}^d))$ towards a limiting curve ρ. Let us call

$$\overline{\rho}^\tau(t) := \rho^\tau \left(\left\lceil \frac{t}{\tau} \right\rceil \tau \right) \quad \text{and} \quad \overline{m}^\tau(t) := m^\tau \left(\left\lceil \frac{t}{\tau} \right\rceil \tau \right) = -\Big\{ \nabla V + \nabla W * \overline{\rho}^\tau(t) \Big\} \overline{\rho}^\tau(t) - \nabla \overline{\rho}^\tau(t).$$

We observe that $\overline{\rho}^\tau \to \rho$ in $C_{\mathrm{loc}}(\mathbb{R}_+; \mathcal{P}(\mathbb{T}^d))$ thanks to the estimate in Step 2, and that

$$\overline{m}^\tau \underset{\tau \to 0}{\longrightarrow} -\Big\{ \nabla V + \nabla W * \rho \Big\} \rho - \nabla \rho \qquad \text{in } \mathcal{D}'(\mathbb{R}_+^* \times \mathbb{T}^d).$$

The last thing to prove is that

$$\overline{m}^\tau - m^\tau \underset{\tau \to 0}{\longrightarrow} 0 \qquad \text{in } \mathcal{D}'(\mathbb{R}_+^* \times \mathbb{T}^d).$$

Indeed, if we do so, as we know that

$$\partial_t \rho^\tau + \operatorname{div} m^\tau = 0,$$

we would conclude that

$$\partial_t \rho^\tau + \operatorname{div} \overline{m}^\tau \underset{\tau \to 0}{\longrightarrow} 0 \qquad \text{in } \mathcal{D}'(\mathbb{R}_+^* \times \mathbb{T}^d).$$

But as

$$\partial_t \rho^\tau + \operatorname{div} \overline{m}^\tau \underset{\tau \to 0}{\longrightarrow} \partial_t \rho - \operatorname{div} \left(\Big\{ \nabla V + \nabla W * \rho \Big\} \rho + \nabla \rho \right) \qquad \text{in } \mathcal{D}'(\mathbb{R}_+^* \times \mathbb{T}^d),$$

the result would follow.

We need to show that for all $\xi \in C_c^1(\mathbb{R}_+^* \times \mathbb{T}^d)$,

$$\lim_{\tau \to 0} \int_0^{+\infty} \int \xi(t,x) \cdot \mathrm{d}m^\tau \left(\left\lceil \frac{t}{\tau} \right\rceil \tau, x \right) \mathrm{d}t - \int_0^{+\infty} \int \xi(t,x) \cdot \mathrm{d}m^\tau(t,x) \, \mathrm{d}t = 0.$$

But this is direct using (8) and calling $T > 0$ an upper-bound for the temporal support of ξ:

$$\begin{aligned}
&\left| \int_0^{+\infty} \int \xi(t,x) \cdot \mathrm{d}m^\tau \left(\left\lceil \frac{t}{\tau} \right\rceil \tau, x \right) \mathrm{d}t - \int_0^{+\infty} \int \xi(t,x) \cdot \mathrm{d}m^\tau(t,x) \, \mathrm{d}t \right| \\
&\quad \leq \mathrm{Lip}(\xi) \int_0^T \left(\left\lceil \frac{t}{\tau} \right\rceil \tau - t \right) W_2^2 \left(\rho^\tau(t), \rho^\tau \left(\left\lceil \frac{t}{\tau} \right\rceil \tau \right) \right) \mathrm{d}t \\
&\quad \leq 2\mathrm{Lip}(\xi) \mathcal{E}(\rho_0) \int_0^T \left(\left\lceil \frac{t}{\tau} \right\rceil \tau - t \right)^2 \mathrm{d}t \leq 2\mathrm{Lip}(\xi) \mathcal{E}(\rho_0) T \tau^2,
\end{aligned}$$

which converges to 0 as $\tau \to 0$. □

2 Uniqueness: geodesic convexity

2.1 Geodesic convexity of the energy functional

2.2 Uniqueness of the PDE

References

[1] F. Santambrogio. Optimal transport for applied mathematicians. *Birkäuser, NY*, 2015.