Application of Optimal Transport to Evolutionary PDEs

3 - Gradient flows of the potential, interaction, and internal energy functionals

Giuseppe Savaré
http://www.imati.cnr.it/∼savare
Department of Mathematics, University of Pavia, Italy

2010 CNA Summer School
New Vistas in Image Processing and PDEs

Carnegie Mellon Center for Nonlinear Analysis, Pittsburgh, *June 7–12, 2010*

Outline

An (incomplete...) list of basic questions

Prove

- Existence
- Stability
- Asymptotic behaviour

of Wasserstein gradient flows.

Some advantages of this approach:

- It is natural to deal with "transportation" mechanisms
- Non-negativity is for free (interesting for 4-th order problems)
- Covers both diffuse and discrete models, measure-valued solutions, concentration effects
- A general approximation scheme is available
- It is quite robust with respect to perturbations: second order evolution equation admits a derivative-free formulation.
- Interesting geometric aspects of the underlying space are involved.
- Gradient flows are often associated to useful functional inequalities.

The structure of the equations

In the **diffuse** case, interpret $t \mapsto u_t$ as the Lebesgue densities of the probability measures $\mu_t = u_t \mathscr{L}^d$ which evolve according to the continuity equation with velocity v

$$\begin{cases} \partial_t u + \mathrm{div}(u\, v) = 0 & (\textit{Continuity equation}) \\ v = -\nabla \dfrac{\delta \Phi}{\delta u} & (\textit{Nonlinear variational condition}) \\ u(0,\cdot) = u_0 & u_0 \in L^1(\mathbb{R}^d),\ u_0 \geq 0. \end{cases}$$

Φ is an integral functional and $\dfrac{\delta \Phi}{\delta u}$ is its Euler-Lagrange first variation

$$\Phi(u) := \int_{\mathbb{R}^d} \varphi(x, u, \mathrm{D}u)\, dx, \quad \frac{\delta \Phi}{\delta u} = \varphi_u(x, u, \mathrm{D}u) - \mathrm{div}\, \varphi_{\mathrm{D}u}(x, u, \mathrm{D}u)$$

We will look for solutions (μ, v) with

$$\int_0^T \Big(\int_{\mathbb{R}^d} |v_t(x)|^2 \, \mathrm{d}\mu_t(x) \Big)\, \mathrm{d}t < +\infty$$

and satisfying the continuity equation in the sense of distribution:

$$\int_0^T \int_{\mathbb{R}^d} \Big(\partial_t \zeta + \langle \nabla \zeta, v_t \rangle \Big)\, \mathrm{d}\mu_t\, \mathrm{d}t = 0 \quad \text{for every } \zeta \in \mathrm{C}^\infty_\mathrm{c}(\mathbb{R}^d \times (0,T)).$$

The simplest example: the potential energy and the linear transport equation

The linear transport equation asociated to a potential $V : \mathbb{R}^d \to \mathbb{R}$.

$$\boxed{\partial_t u - \mathsf{div}(u \nabla V) = 0} \quad u(x,0) = u_0(x). \tag{LTE}$$

is the gradient flow of the **potential energy**

$$\Phi = \mathcal{V} := \int_{\mathbb{R}^d} V(x)\, u \,\mathrm{d}x = \int_{\mathbb{R}^d} V(x)\,\mathrm{d}\mu(x), \quad \frac{\delta \mathcal{V}}{\delta u} = V$$

When $V \in C^2(\mathbb{R}^d)$ with $\mathrm{D}^2 V \geq \boldsymbol{\lambda} I$, the solution can be easily obtained by the characteristic method: we solve the gradient flow in $\mathbb{R}^d$ generated by V

$$\frac{\mathrm{d}}{\mathrm{d}t} X_t(x) = -\nabla V(X_t), \quad X_0(x) = x \tag{GF}$$

and then we represent the solution $\mu_t := u_t \mathscr{L}^d$ of (LTE) by the push-forward formula

$$\mu_t = (X_t)_\# \mu_0.$$

When $\boldsymbol{\lambda} = 0$, μ_t solves the Evolution Variational Inequality

$$\frac{\mathrm{d}}{\mathrm{d}t} \frac{1}{2} W_2^2(\mu_t, \sigma) \leq \mathcal{V}(\sigma) - \mathcal{V}(\mu_t) \quad \text{for every } \sigma \in \mathscr{P}_2(\mathbb{R}^d). \tag{EVI}$$

A direct proof of EVI

It is interesting to give a direct proof of (EVI) for the potential energy.
We start from the formula for the derivative of the Wasserstein distance

$$\frac{\mathrm{d}}{\mathrm{d}t} W_2^2(\mu_t, \sigma) = 2 \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle \boldsymbol{v}_t(x), x - y \rangle \, \mathrm{d}\boldsymbol{\mu}_t(x, y)$$

along a solution of $\partial_t \mu_t + \mathrm{div}(\mu_t \boldsymbol{v}_t) = 0$, where $\boldsymbol{\mu}_t$ is an optimal coupling between μ_t and σ.
Since $\boldsymbol{v}_t \equiv -\nabla V$ we obtain

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} \frac{1}{2} W_2^2(\mu_t, \sigma) &= \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle \nabla V(x), y - x \rangle \, \mathrm{d}\boldsymbol{\mu}_t(x, y) \\ &\le \int_{\mathbb{R}^d \times \mathbb{R}^d} \Big(V(y) - V(x) \Big) \, \mathrm{d}\boldsymbol{\mu}_t(x, y) = \mathcal{V}(\sigma) - \mathcal{V}(\mu_t) \end{aligned}$$

where we applied the subgradient inequality for V

$$\langle \nabla V(x), y - x \rangle \le V(y) - V(x)$$

and the fact that μ_t and σ are the marginal of $\boldsymbol{\mu}_t$.

The variational approximation: general strategy

1. **The starting point:**

   construct the gradient flow by the JKO/Minimizing Movement scheme.

   Existence of a discrete solution can be proved by the direct method of the Calculus of Variation, combining **lower-semicontinuity and compactness arguments**.

2. **Euler-Lagrange variation in the Wasserstein setting**: try to extract information by **taking suitable first variations** of the functional involved, which should **take care of the particular structure of the Wasserstein distance**: this leads to a crucial

   "discrete" formulation of the evolution PDE.

3. **Convergence of dicrete solutions:** Two basic situations:
   - **The functional Φ is displacement $\boldsymbol{\lambda}$-convex**: one can apply a general theory, based on the EVI formulation (at continuous and discrete level)

     $$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}\mathsf{d}^2(\boldsymbol{u}_t,\boldsymbol{w}) \le \Phi(\boldsymbol{w}) - \Phi(\boldsymbol{u}_t) \qquad \text{(EVI, } \boldsymbol{\lambda}\text{=0)}$$

   - **The functional Φ is not displacement $\boldsymbol{\lambda}$-convex**: to pass to the limit in the discrete formulation as $\tau \downarrow 0$, one needs suitable **space–time compactness estimates**.
     Compactness in "space" (a priori bounds on the functional and its first variation/Wasserstein slope) and compactness in "time" can be deduced by a discrete version of the basic **Maximal slope inequality**

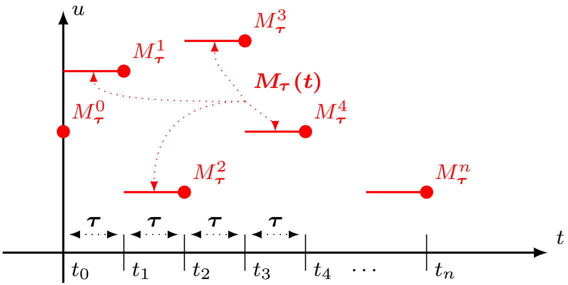
     $$\frac{1}{2}|\dot{\boldsymbol{u}}_t|^2 + \frac{1}{2}|\partial\Phi|^2(\boldsymbol{u}_t) \le -\frac{\mathrm{d}}{\mathrm{d}t}\Phi(\boldsymbol{u}_t). \qquad \text{(MSI)}$$

     but **often further estimates are needed.**

JKO/Minimizing Movement scheme

- Choose a partition of $(0, +\infty)$ of ***step size*** $\boldsymbol{\tau} > 0$ and look for

measures $M_\tau^n \in \mathscr{P}_2(\mathbb{R}^d)$ which approximate μ_t at the time $t = n\tau$.



Algorithm: starting from $M_\tau^0 := \mu_0$ **find recursively** $M_\tau^n, \; n = 1, 2, \ldots,$ such that

$$M_\tau^n \quad \text{minimizes} \quad M \mapsto \frac{W_2^2(M, M_\tau^{n-1})}{2\tau} + \Phi(M) \quad \text{in } \mathscr{P}_2(\mathbb{R}^d).$$

- $\boldsymbol{M_\tau}$ is the **piecewise constant** interpolant of $\{M_\tau^n\}_n$.

The discrete equation associated to a single minimization step

Problem: Let $\boldsymbol{M}$ be a minimizer of

$$M \mapsto \frac{W_2^2(M, M_{\boldsymbol{\tau}}^{n-1})}{2\tau} + \Phi(M) \quad \text{in } \mathscr{P}_2(\mathbb{R}^d) \tag{MM}$$

Which kind of "Euler" equation does $\boldsymbol{M}$ satisfy?
Analogy: If $\boldsymbol{U}$ minimizes in some $\mathbb{R}^m$

$$U \mapsto \frac{1}{2\tau}|U - U^{n-1}\tau|^2 + \phi(U)$$

we get

$$\frac{\boldsymbol{U} - U_\tau^{n-1}}{\tau} + \nabla\phi(\boldsymbol{U}) = 0. \tag{EE}$$

(EE) can be obtained by perturbing $\boldsymbol{U}$ along a direction Z by taking

$$\boldsymbol{U_\varepsilon} := \boldsymbol{U} + \varepsilon Z$$

and observing that the function

$$f(\varepsilon) := \frac{1}{2\tau}|\boldsymbol{U_\varepsilon} - U_\tau^{n-1}|^2 + \phi(\boldsymbol{U_\varepsilon}) \qquad \varepsilon > 0,$$

has a minimum at $\varepsilon = 0$, so that

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon} f(\varepsilon)\Big|_{\varepsilon=0^+} \geq 0 \quad \Leftrightarrow \quad \left\langle \frac{\boldsymbol{U} - U_\tau^{n-1}}{\tau}, Z \right\rangle + \langle \nabla\phi(\boldsymbol{U_\tau^n}), Z\rangle \geq 0 \quad \text{for every } Z \in \mathbb{R}^m.$$

Natural perturbations in the Wasserstein framework

Let M be a minimizer of

$$M \mapsto \frac{W_2^2(M, M_\tau^{n-1})}{2\tau} + \Phi(M) \quad \text{in } \mathscr{P}_2(\mathbb{R}^d) \tag{MM}$$

Main idea [JKO]:

replace **"linear"** perturbations $M + \varepsilon Z$ with **"transport"** perturbations.

Choose a **smooth vector field** $\boldsymbol{\xi} \in C_c^\infty(\mathbb{R}^d; \mathbb{R}^d)$ and consider the flow $X_t : \mathbb{R}^d \to \mathbb{R}^d$ associated to the ODE system

$$\frac{\mathrm{d}}{\mathrm{d}t} X_t(x) = \boldsymbol{\xi}(X_t(x)), \qquad X_0(x) = x,$$

We perturb the minimizer M by

$$\boxed{M_\varepsilon := (X_\varepsilon)_\# M}$$

Notice that M_ε is still a probability measure and solves the transport equation

$$\partial_\varepsilon M_\varepsilon + \mathsf{div}(M_\varepsilon \boldsymbol{\xi}) = 0, \qquad M_0 = M$$

Again, setting

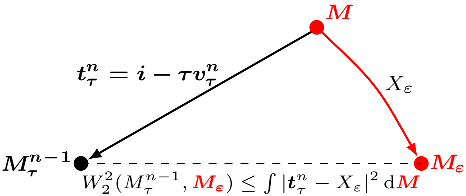
$$f(\varepsilon) := \frac{W_2^2(M_\varepsilon, M_\tau^{n-1})}{2\tau} + \Phi(M_\varepsilon) \quad \text{we have} \quad \frac{\mathrm{d}}{\mathrm{d}\varepsilon} f(\varepsilon)\big|_{\varepsilon=0^+} \geq 0.$$

Problem: compute $\quad \frac{\mathrm{d}}{\mathrm{d}\varepsilon} W_2^2(M_\varepsilon, M_\tau^{n-1}), \qquad \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Phi(M_\varepsilon).$

Perturbation of the Wasserstein distance

$\boldsymbol{t}_\tau^n :=$ optimal transport map pushing $\boldsymbol{M} = \boldsymbol{M}_\tau^n$ on M_τ^{n-1}

$\boldsymbol{v}_\tau^n := \frac{\boldsymbol{i}-\boldsymbol{t}_\tau^n}{\tau} :=$ the **discrete velocity**



$$\boxed{\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Big(\frac{1}{2\tau}W_2^2(\boldsymbol{M}_\varepsilon, M_\tau^{n-1})\Big)\Big|_{\varepsilon=0} \le \int \langle \boldsymbol{\xi}, \boldsymbol{v}_\tau^n\rangle \,\mathrm{d}\boldsymbol{M}}$$

$$W_2^2(\boldsymbol{M}, M_\tau^{n-1}) = \int_{\mathbb{R}^d} |x - \boldsymbol{t}_\tau^n(x)|^2 \,\mathrm{d}\boldsymbol{M}, \quad W_2^2(\boldsymbol{M}_\varepsilon, M_\tau^{n-1}) \boxed{\le} \int_{\mathbb{R}^d} |\boldsymbol{t}_\tau^n - X_\varepsilon|^2 \,\mathrm{d}\boldsymbol{M}$$

$$\begin{aligned}
&\lim_{\varepsilon\downarrow 0} \varepsilon^{-1}\Big(W_2^2(\boldsymbol{M}_\varepsilon, M_\tau^{n-1}) - W_2^2(\boldsymbol{M}, M_\tau^{n-1})\Big) \\
\le &\lim_{\varepsilon\downarrow 0} \int_{\mathbb{R}^d} \varepsilon^{-1}\Big(|\boldsymbol{t}_\tau^n - X_\varepsilon|^2 - |\boldsymbol{t}_\tau^n(x) - x|^2\Big) \,\mathrm{d}\boldsymbol{M} \\
= &2\int_{\mathbb{R}^d} \langle \boldsymbol{\xi}, \boldsymbol{t}_\tau^n(x) - x\rangle \,\mathrm{d}\boldsymbol{M} = 2\tau \int \langle \boldsymbol{\xi}, \boldsymbol{v}_\tau^n\rangle \,\mathrm{d}\boldsymbol{M}
\end{aligned}$$

Perturbation of the functional

$$\frac{\mathrm{d}}{\mathrm{d}t}X_t(x) = \boldsymbol{\xi}(X_t(x)),\ X_0(x) = x; \qquad \boldsymbol{M_\varepsilon} := (X_\varepsilon)_\# \boldsymbol{M}; \quad \rightsquigarrow \quad \boxed{\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Phi(\boldsymbol{M_\varepsilon})|_{\varepsilon=0}.}$$

Definition (Weak Wasserstein gradient)

Is a vector field $\boldsymbol{g} := \partial\Phi(\boldsymbol{M}) \in L^2(\boldsymbol{M};\mathbb{R}^d)$ satisfying

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Phi(\boldsymbol{M_\varepsilon})|_{\varepsilon=0} = \int_{\mathbb{R}^d} \langle \boldsymbol{g}, \boldsymbol{\xi}\rangle \, d\boldsymbol{M} \quad \text{for every smooth vector field } \boldsymbol{\xi} \in C^\infty_c(\mathbb{R}^d;\mathbb{R}^d).$$

Corollary (Euler equation for the minimizing movement)

If $\boldsymbol{M^n_\tau}$ *is a minimizer of* $\quad M \mapsto \dfrac{W_2^2(M, M^{n-1}_{\boldsymbol{\tau}})}{2\tau} + \Phi(M) \quad$ *in* $\mathscr{P}_2(\mathbb{R}^d)$

and if for every $\boldsymbol{\xi} \in C^\infty_c(\mathbb{R}^d;\mathbb{R}^d)$ *the function* $\varepsilon \mapsto \Phi(\boldsymbol{M_\varepsilon})$ *is differentiable at* $\varepsilon = 0$ *then*

$$\boxed{\boldsymbol{v}^n_\tau = -\partial\Phi(\boldsymbol{M^n_\tau})} \quad \text{in } L^2(\boldsymbol{M^n_\tau};\mathbb{R}^d).$$

Proof: take $\boldsymbol{\xi}$ and $-\boldsymbol{\xi}$ in the following inequality.

$$0 \le \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Big(\frac{1}{2\tau}W_2^2(\boldsymbol{M_\varepsilon}, M^{n-1}_\tau) + \Phi(\boldsymbol{M_\varepsilon})\Big)|_{\varepsilon=0} \le \int_{\mathbb{R}^d} \langle \boldsymbol{v}^n_\tau, \boldsymbol{\xi}\rangle \, d\boldsymbol{M} + \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Phi(\boldsymbol{M_\varepsilon})|_{\varepsilon=0}.$$

The potential energy

$$\frac{\mathrm{d}}{\mathrm{d}t}X_t(x) = \boldsymbol{\xi}(X_t(x)), \quad \boldsymbol{M_\varepsilon} := (X_\varepsilon)_\# \boldsymbol{M} \rightsquigarrow \boxed{\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathcal{V}(\boldsymbol{M_\varepsilon})|_{\varepsilon=0} = \int \langle \partial\mathcal{V}(\boldsymbol{M}), \boldsymbol{\xi}\rangle \, d\boldsymbol{M}.}$$

The functional is linear

$$\mathcal{V}(\boldsymbol{M}) := \int_{\mathbb{R}^d} V(x)\, d\boldsymbol{M}, \qquad \boxed{\partial\mathcal{V}(\boldsymbol{M}) = \nabla V}$$

$$\mathcal{V}(\boldsymbol{M_\varepsilon}) = \int_{\mathbb{R}^d} V(x)\, \mathrm{d}\boldsymbol{M_\varepsilon} = \int V(X_\varepsilon(x))\, d\boldsymbol{M},$$

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathcal{V}(\boldsymbol{M_\varepsilon})|_{\varepsilon=0} = \int_{\mathbb{R}^d} \langle \nabla V(x), \boldsymbol{\xi}\rangle \, d\boldsymbol{M}.$$

Wasserstein gradient of the interaction energy

$$\frac{\mathrm{d}}{\mathrm{d}t}X_t(x) = \boldsymbol{\xi}(X_t(x)), \quad \boldsymbol{M_\varepsilon} := (X_\varepsilon)_\# \boldsymbol{M} \rightsquigarrow \boxed{\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Phi(\boldsymbol{M_\varepsilon})|_{\varepsilon=0} = \int \langle \partial\Phi(\boldsymbol{M}), \boldsymbol{\xi}\rangle \, d\boldsymbol{M}.}$$

The interaction potential: $W \in \mathrm{C}^1(\mathbb{R}^d)$, even, with bounded derivatives.

$$\Phi = \mathcal{W}(\boldsymbol{M}) := \frac{1}{2}\iint W(x-y)\,\mathrm{d}\boldsymbol{M}(x)\,\mathrm{d}\boldsymbol{M}(y), \qquad \boxed{\partial\mathcal{W}(\boldsymbol{M}) = \int_{\mathbb{R}^d} \nabla W(x-y)\,\mathrm{d}\boldsymbol{M}(y)}$$

$$\begin{aligned}
\mathcal{W}(\boldsymbol{M_\varepsilon}) &= \frac{1}{2}\iint W(x'-y')\,\mathrm{d}\boldsymbol{M_\varepsilon}(x')\,\mathrm{d}\boldsymbol{M}_\varepsilon(y') \\
&= \iint W(X_\varepsilon(x) - X_\varepsilon(y))\,\mathrm{d}\boldsymbol{M}(x)\,\mathrm{d}\boldsymbol{M}(y), \\
\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathcal{W}(\boldsymbol{M_\varepsilon})|_{\varepsilon=0} &= \frac{1}{2}\iint \langle \nabla W(x-y), \boldsymbol{\xi}(x) - \boldsymbol{\xi}(y)\rangle\,\mathrm{d}\boldsymbol{M}(x)\,\mathrm{d}\boldsymbol{M}(y) \\
&= \frac{1}{2}\int_{\mathbb{R}^d}\Big\langle \boxed{\int_{\mathbb{R}^d} \nabla W(x-y)\,\mathrm{d}\boldsymbol{M}(y)}, \boldsymbol{\xi}(x)\Big\rangle\,\mathrm{d}\boldsymbol{M}(x) \\
&+ \frac{1}{2}\int_{\mathbb{R}^d}\Big\langle\Big(\int_{\mathbb{R}^d} -\nabla W(x-y)\,\mathrm{d}\boldsymbol{M}(x)\Big), \boldsymbol{\xi}(y)\Big\rangle\,\mathrm{d}\boldsymbol{M}(y)
\end{aligned}$$

Switch the variable x and y in the last integral and use the fact that $-\nabla W(x-y) = \nabla W(y-x)$.

Wasserstein Gradient of the logarithmic entropy (I)

$$\frac{\mathrm{d}}{\mathrm{d}t}X_t(x) = \boldsymbol{\xi}(X_t(x)), \quad \boldsymbol{M_\varepsilon} := (X_\varepsilon)_\# \boldsymbol{M} \rightsquigarrow \boxed{\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Phi(\boldsymbol{M_\varepsilon})|_{\varepsilon=0} = \int \langle \partial\Phi(\boldsymbol{M}), \boldsymbol{\xi}\rangle \, d\boldsymbol{M}.}$$

The Logarithmic entropy

$$\Phi = \mathcal{H}(\boldsymbol{M}) := \int_{\mathbb{R}^d} U \log U \,\mathrm{d}x = \int \log U \,\mathrm{d}\boldsymbol{M}, \quad \boldsymbol{M} = U\mathscr{L}^d$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathcal{H}(\boldsymbol{M_\varepsilon}) = -\int_{\mathbb{R}^d} \mathsf{div}\,\boldsymbol{\xi}\, d\boldsymbol{M}.}$$

$$\mathcal{H}(\boldsymbol{M_\varepsilon}) = \int_{\mathbb{R}^d} \log\Big(U_\varepsilon(y)\Big)\,\mathrm{d}\boldsymbol{M_\varepsilon}(y) = \int_{\mathbb{R}^d} \log\Big(U_\varepsilon(X_\varepsilon(x))\Big)\,\mathrm{d}\boldsymbol{M}$$
$$= \int_{\mathbb{R}^d} \log U \,\mathrm{d}\boldsymbol{M} - \int_{\mathbb{R}^d} \log\Big(\det \mathrm{D}X_\varepsilon(x)\Big)\,\mathrm{d}\boldsymbol{M}$$

thanks to the change of variable formula: $U_\varepsilon(X_\varepsilon(x)) \det \mathrm{D}X_\varepsilon(x) = U(x)$.

In order to calculate the derivative of $\log\Big(\det \mathrm{D}X_\varepsilon(x)\Big)$ w.r.t. ε, we differentiate the ODE with respect to x:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{D}X_t(x) = \mathrm{D}\boldsymbol{\xi}(X_t)\cdot \mathrm{D}X_t(x), \quad \frac{\mathrm{d}}{\mathrm{d}t}\det \mathrm{D}X_t(x) = \mathrm{trace}\Big(\mathrm{D}\boldsymbol{\xi}(X_t)\Big)\det \mathrm{D}X_t(x)$$

so that

$$\frac{\mathrm{d}}{\mathrm{d}t}\log\Big(\det \mathrm{D}X_t(x)\Big) = \mathsf{div}\,\boldsymbol{\xi}(X_t(x))$$

Wasserstein gradient of the logarithmic entropy (II)

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathcal{H}(\boldsymbol{M}_{\boldsymbol{\varepsilon}}) = -\int_{\mathbb{R}^d} \operatorname{div}\boldsymbol{\xi}\, d\boldsymbol{M}.$$

In order to correctly interpret this fomula, we use the fact that we know *a priori* by the minimization scheme that the **discrete velocity** $\boldsymbol{v}_\tau^n$ satisfies

$$\int_{\mathbb{R}^d} \langle \boldsymbol{v}_\tau^n, \boldsymbol{\xi}\rangle\, \mathrm{d}\boldsymbol{M} = -\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathcal{H}(\boldsymbol{M}_{\boldsymbol{\varepsilon}})$$

for every smooth vector field $\boldsymbol{\xi}$. It follows that

$$\int_{\mathbb{R}^d} \langle \boldsymbol{v}_\tau^n, \boldsymbol{\xi}\rangle\, U\, \mathrm{d}x = \int_{\mathbb{R}^d} \operatorname{div}\boldsymbol{\xi}\, U\, \mathrm{d}x$$

This has two consequences:

- $U \in W^{1,1}(\mathbb{R}^d)$, since

$$\int_{\mathbb{R}^d} \operatorname{div}\boldsymbol{\xi}\, U\, \mathrm{d}x = \int_{\mathbb{R}^d} \langle \boldsymbol{h}, \boldsymbol{\xi}\rangle\, \mathrm{d}x \quad \text{for every } \boldsymbol{\xi} \in \mathrm{C}_\mathrm{c}^\infty(\mathbb{R}^d;\mathbb{R}^d) \quad \text{and } \boldsymbol{h} := U\boldsymbol{v}_\tau^n \in L^1(\mathbb{R}^d).$$

- The (distributional) gradient of U is $-U\boldsymbol{v}_\tau^n$. This is the appropriate formulation of

$$\boldsymbol{v}_\tau^n = -\frac{\nabla U}{U}.$$

In this sense, we can also say that

$$\partial\mathcal{H}(\boldsymbol{M}) = \frac{\nabla U}{U} = \text{“}\nabla(\log U)\text{''}$$

Wasserstein gradient of the internal energy functional

$$\frac{\mathrm{d}}{\mathrm{d}t}X_t(x) = \boldsymbol{\xi}(X_t(x)), \quad \boldsymbol{M_\varepsilon} := (X_\varepsilon)_\# \boldsymbol{M} \rightsquigarrow \boxed{\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Phi(\boldsymbol{M_\varepsilon})_{|\varepsilon=0} = \int \langle \partial\Phi(\boldsymbol{M}), \boldsymbol{\xi}\rangle \, d\boldsymbol{M}.}$$

A convex internal energy functional: choose a smooth convex function $F : [0,+\infty) \to [0,+\infty)$ and the associated integral

$$\mathcal{F}(\boldsymbol{M}) := \int_{\mathbb{R}^d} F(U)\,\mathrm{d}x, \qquad \boldsymbol{M} = U\mathscr{L}^d$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathcal{F}(\boldsymbol{M_\varepsilon}) = -\int \mathrm{div}\,\boldsymbol{\xi}\,\boxed{L(U)}\,\mathrm{d}x, \quad L(r) := rF'(r) - F(r).}$$

E.g. $F(r) = r\log r \rightsquigarrow L(r) = r, \qquad F(r) = \frac{1}{\beta-1}r^\beta \rightsquigarrow L(r) = r^\beta.$

Notice that $\boxed{L'(r) = rF''(r)}$.

In this case, the existence of the Wasserstein gradient means

$$L(\boldsymbol{U}) \in W^{1,1}(\mathbb{R}^d), \quad -\nabla L(\boldsymbol{U}) = \boldsymbol{U}\boldsymbol{v}_\tau^n$$

which formally yields

$$\boldsymbol{v}_\tau^n = -\frac{\nabla L(\boldsymbol{U})}{\boldsymbol{U}} = -\frac{L'(U)\nabla U}{U} = -F''(U)\nabla U = \boxed{-\nabla F'(U)}.$$

i.e.

$$\textbf{discrete velocity} = -\nabla\frac{\delta\mathcal{F}}{\delta u}$$

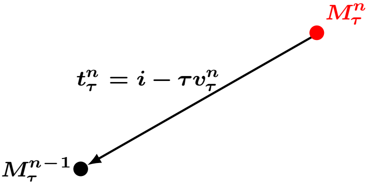
Summary

$\boldsymbol{M_\tau^n}$ is a minimizer in $\mathscr{P}_2(\mathbb{R}^d)$ of

$$\frac{W_2^2(M, M_\tau^{n-1})}{2\tau} + \Phi(M)$$

$\boldsymbol{t}_\tau^n$:= optimal transport map pushing $\boldsymbol{M_\tau^n}$ on M_τ^{n-1}

$\boldsymbol{v}_\tau^n := \frac{\boldsymbol{i}-\boldsymbol{t}_\tau^n}{\tau}$:= the **discrete velocity**



For every $\boldsymbol{\xi} \in \mathrm{C}_\mathrm{c}^\infty(\mathbb{R}^d;\mathbb{R}^d)$

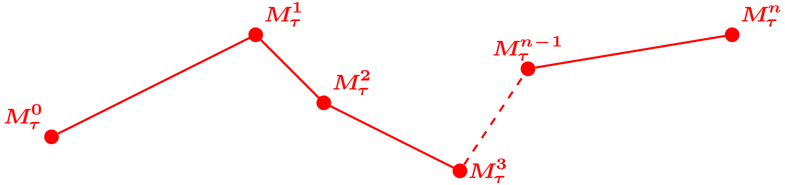
$$\int_{\mathbb{R}^d} \langle \boldsymbol{v}_\tau^n, \boldsymbol{\xi}\rangle \,\mathrm{d}\boldsymbol{M_\tau^n} + \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Phi\big((X_\varepsilon)_\# \boldsymbol{M_\tau^n}\big) = 0, \quad \dot{X} = \boldsymbol{\xi}(X) \quad \Rightarrow \quad \boxed{\boldsymbol{v}_\tau^n = -\partial\Phi(\boldsymbol{M_\tau^n})}$$

$$\Phi = \mathcal{V} = \int_{\mathbb{R}^d} V(x)\,\mathrm{d}\boldsymbol{M} \quad \rightsquigarrow \quad \partial\mathcal{V}(\boldsymbol{M}) = \nabla V$$

$$\Phi = \mathcal{W} = \iint W(x-y)\,\mathrm{d}\boldsymbol{M}(x)\,\mathrm{d}\boldsymbol{M}(y) \quad \rightsquigarrow \quad \partial\mathcal{W}(\boldsymbol{M}) = \nabla W * \boldsymbol{M}$$

$$\Phi = \mathcal{F} = \int_{\mathbb{R}^d} F(U)\,\mathrm{d}x \quad \rightsquigarrow \quad \partial\mathcal{F}(\boldsymbol{M}) = \frac{\nabla L(U)}{U}, \; L(r) = rF'(r) - F(r)$$

Passing to the limit: the continuity equation



Equation for the discrete velocity: at each step $\boxed{v_\tau^n = -\partial\Phi(M_\tau^n)}$.

Limit equation: $\partial_t \mu + \mathrm{div}(\mu v) = 0$, $v = -\partial\Phi(\mu)$.

Is there a "discrete" version of the continuity equation?

Distributional formulation of the continuity equation: for every $\zeta \in \mathrm{C}_c^\infty(\mathbb{R}^d)$

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}^d}\zeta\,\mathrm{d}\mu_t = \int_{\mathbb{R}^d}\langle\nabla\zeta, v_t\rangle\,\mathrm{d}\mu_t \quad\Leftrightarrow\quad \int_{\mathbb{R}^d}\zeta\,\mathrm{d}\mu_t - \int_{\mathbb{R}^d}\zeta\,\mathrm{d}\mu_s = \int_s^t\Big(\int\langle\nabla\zeta, v_r\rangle\,\mathrm{d}\mu_r\Big)\,\mathrm{d}r$$

$$\int_{\mathbb{R}^d}\zeta\,\mathrm{d}\mu_t \quad\rightsquigarrow\quad \int\zeta\,\mathrm{d}M_\tau^n \qquad\qquad \int_{\mathbb{R}^d}\zeta\,\mathrm{d}\mu_t \quad\rightsquigarrow\quad \int\zeta\,\mathrm{d}M_\tau^m$$

$$\int_s^t\cdots\,\mathrm{d}r \quad\rightsquigarrow\quad \tau\sum_{k=m+1}^{n}\cdots \qquad\qquad \int_{\mathbb{R}^d}\langle\nabla\zeta, v_r\rangle\,\mathrm{d}\mu_r \quad\rightsquigarrow\quad \int_{\mathbb{R}^d}\langle\nabla\zeta, v_\tau^k\rangle\,\mathrm{d}M_\tau^k$$

Try to estimate the difference

$$\int\zeta\,\mathrm{d}M_\tau^n - \int\zeta\,\mathrm{d}M_\tau^m - \tau\sum_{k=m+1}^{n}\int_{\mathbb{R}^d}\langle\nabla\zeta, v_\tau^k\rangle\,\mathrm{d}M_\tau^k \qquad 0 \le m < n$$

Discrete continuity equation

Theorem

If $\|D^2\zeta\|_\infty \le H$ *then for every* $0 \le m < n$

$$\Big|\int \zeta \,\mathrm{d}\boldsymbol{M}_{\boldsymbol{\tau}}^{\boldsymbol{n}} - \int \zeta \,\mathrm{d}\boldsymbol{M}_{\boldsymbol{\tau}}^{\boldsymbol{m}} - \tau \sum_{k=m+1}^{n} \int_{\mathbb{R}^d} \langle \nabla\zeta, \boldsymbol{v}_{\boldsymbol{\tau}}^{\boldsymbol{k}}\rangle \,\mathrm{d}\boldsymbol{M}_\tau^k\Big| \le H \sum_{k=m+1}^{k} W_2^2(\boldsymbol{M}_{\boldsymbol{\tau}}^{\boldsymbol{k}}, \boldsymbol{M}_{\boldsymbol{\tau}}^{\boldsymbol{k-1}})$$

Proof. Start from two consecutive steps and use the fact that $(\boldsymbol{i} - \tau\boldsymbol{v}_{\boldsymbol{\tau}}^{\boldsymbol{k}})_\# \boldsymbol{M}_{\boldsymbol{\tau}}^{\boldsymbol{k}} = \boldsymbol{M}_{\boldsymbol{\tau}}^{\boldsymbol{k-1}}$:

$$\begin{aligned}\int \zeta(x)\,\mathrm{d}\boldsymbol{M}_{\boldsymbol{\tau}}^{\boldsymbol{k}}(x) - \int \zeta(y)\,\mathrm{d}\boldsymbol{M}_{\boldsymbol{\tau}}^{\boldsymbol{k-1}}(y) &= \int \zeta(x)\,\mathrm{d}\boldsymbol{M}_{\boldsymbol{\tau}}^{\boldsymbol{k}}(x) - \int \zeta(x - \tau\boldsymbol{v}_{\boldsymbol{\tau}}^{\boldsymbol{k}}(x))\,\mathrm{d}\boldsymbol{M}_{\boldsymbol{\tau}}^{\boldsymbol{k}}(x)\\ &= \int \Big(\zeta(x) - \zeta(x - \tau\boldsymbol{v}_{\boldsymbol{\tau}}^{\boldsymbol{k}}(x))\Big)\,\mathrm{d}\boldsymbol{M}_{\boldsymbol{\tau}}^{\boldsymbol{k}}(x)\\ &= \tau\int \langle \nabla\zeta(x), \boldsymbol{v}_{\boldsymbol{\tau}}^{\boldsymbol{k}}(x)\rangle\,\mathrm{d}\boldsymbol{M}_{\boldsymbol{\tau}}^{\boldsymbol{k}}(x) + E_\tau^k\end{aligned}$$

where

$$\begin{aligned}|E_\tau^k| &\le \int \Big|\zeta(x) - \zeta(x - \tau\boldsymbol{v}_{\boldsymbol{\tau}}^{\boldsymbol{k}}(x)) - \tau\langle\nabla\zeta(x), \boldsymbol{v}_{\boldsymbol{\tau}}^{\boldsymbol{k}}(x)\rangle\Big|\,\mathrm{d}\boldsymbol{M}_{\boldsymbol{\tau}}^{\boldsymbol{k}}(x) \le H\tau^2 \int_{\mathbb{R}^d} |\boldsymbol{v}_{\boldsymbol{\tau}}^{\boldsymbol{k}}|^2\, d\boldsymbol{M}_{\boldsymbol{\tau}}^{\boldsymbol{k}}\\ &= H\int_{\mathbb{R}^d} |x - \boldsymbol{t}_\tau^k(x)|^2\, d\boldsymbol{M}_{\boldsymbol{\tau}}^{\boldsymbol{k}} = H\, W_2^2(\boldsymbol{M}_{\boldsymbol{\tau}}^{\boldsymbol{k}}, \boldsymbol{M}_\tau^{k-1})\end{aligned}$$

since $\boldsymbol{t}_\tau^k = \boldsymbol{i} - \tau\boldsymbol{v}_{\boldsymbol{\tau}}^{\boldsymbol{k}}$ is an optimal map.

Main steps of the convergence proof

1. Show that the **piecewise constant** or the **geodesic interpolant** $M_\tau(t)$ converge as $\tau \downarrow 0$ to some curve μ_t, at least along a subsequence: **time-compactness estimate.**
2. Show that the discrete velocities converge *in a suitable sense* to some limit vector field v and that (μ, v) solve the continuity equation

$$\partial_t \mu + \mathsf{div}(\mu v) = 0$$

3. Passing to the limit in the (possibly nonlinear) relation for the discrete velcity $v_\tau^n = -\partial\Phi(M_\tau^n)$: **here the structure of the functional plays a crucial role.**

A basic (and very simple...) discrete energy inequality

Start from the minimum problem

$$\frac{W_2^2(M_\tau^n, M_\tau^{n-1})}{2\tau} + \Phi(M_\tau^n) \le \frac{W_2^2(V, M_\tau^{n-1})}{2\tau} + \Phi(V)$$

and choose $V := M_\tau^{n-1}$:

$$\Phi(M_\tau^n) + \frac{\tau}{2}\frac{W_2^2(M_\tau^n, M_\tau^{n-1})}{\tau^2} \le \Phi(M_\tau^{n-1})$$

so that

- $\Phi(M_\tau^n)$ is non-increasing and bounded by $\Phi(u_0)$
- $\tau \sum_{n=1}^{N} \Big(\frac{W_2(M_\tau^n, M_\tau^{n-1})}{\tau}\Big)^2 \le 2\Big(\Phi(u_0) - \Phi(M_\tau^N)\Big) \le C$
  if Φ is bounded from below.

The coefficient 2 is non-optimal, since in the continuous case one has

$$\int_0^T |\dot{\boldsymbol{u}}_t|^2\,\mathrm{d}t \rightsquigarrow \tau \sum_{n=1}^{N} \Big(\frac{W_2(M_\tau^n, M_\tau^{n-1})}{\tau}\Big)^2$$

and

$$\int_0^T |\dot{\boldsymbol{u}}_t|^2 = \Phi(\boldsymbol{u}_0) - \Phi(\boldsymbol{u}_T)$$

Applications of the energy estimate

$$\tau \sum_{n=1}^{N} \Big(\frac{W_2(M_\tau^n, M_\tau^{n-1})}{\tau} \Big)^2 \leq C \text{ if } \Phi \text{ is bounded from below.}$$

Recall that $\boldsymbol{M_\tau}$ is the piecewise constant interpolation so that $\boldsymbol{M_\tau(t)} = \boldsymbol{M_\tau^n}$ if $t \in ((n-1)\tau, n\tau]$.
We also set $\boldsymbol{V_\tau(t)} := \boldsymbol{v_\tau^n}$ if $t \in ((n-1)\tau, n\tau]$

- **Equi-continuity:**

$$W_2(\boldsymbol{M_\tau^n}, \boldsymbol{M_\tau^m}) \leq C\sqrt{\tau(n-m)} \quad \text{which in terms of } \boldsymbol{M_\tau} \text{ yields}$$
$$W_2(\boldsymbol{M_\tau(t)}, \boldsymbol{M_\tau(s)}) \leq C\sqrt{t-s+\tau} \quad \text{if } 0 \leq s < t$$

- **Tightness:** On every fixed finite time interval $[0,T]$ the measures $\boldsymbol{M_\tau(t)}$ belong to a fixed bounded set of $\mathscr{P}_2(\mathbb{R}^d)$, so that they are tight.

Corollary

Up to extracting a suitable subsequence, we have $\boldsymbol{M_\tau(t)} \rightharpoonup \boldsymbol{\mu_t}$ *in* $\mathscr{P}(\mathbb{R}^d)$ *as* $\tau \downarrow 0$ *and the limit is an absolutely continuous curve in* $\mathscr{P}_2(\mathbb{R}^d)$.

Convergence of the velocities and the continuity equation

$$\tau \sum_{n=1}^{N} \Big(\frac{W_2(M_\tau^n, M_\tau^{n-1})}{\tau}\Big)^2 \le C \text{ if } \Phi \text{ is bounded from below.}$$

- **L^2-bound for the velocities:** since

$$W_2^2(M_\tau^n, M_\tau^{n-1}) = \int_{\mathbb{R}^d} |\boldsymbol{v}_{\boldsymbol{\tau}}^{\boldsymbol{n}}|^2 \, \mathrm{d}\boldsymbol{M}_{\boldsymbol{\tau}}^{\boldsymbol{n}} \quad \text{we have} \quad \int_0^T \int_{\mathbb{R}^d} |\boldsymbol{V}_{\boldsymbol{\tau}}(\boldsymbol{t})|^2 \, \mathrm{d}\boldsymbol{M}_{\boldsymbol{\tau}}(\boldsymbol{t}) \, \mathrm{d}t \le C.$$

Corollary

Up to extracting a subsequence, the vector measure $\boldsymbol{\nu}_\tau = \boldsymbol{V}_{\boldsymbol{\tau}} \boldsymbol{M}_{\boldsymbol{\tau}}$ converges weakly in $\mathscr{P}(\mathbb{R}^d \times (0,T))$ (after a renormalization by T^{-1}) to a limit measure $\boldsymbol{\nu} = \boldsymbol{\mu}\boldsymbol{v}$. In particular, for every smooth function ζ

$$\lim_{\tau \downarrow 0} \int_s^t \int_{\mathbb{R}^d} \langle \nabla\zeta, \boldsymbol{V}_{\boldsymbol{\tau}} \rangle \, \mathrm{d}\boldsymbol{M}_\tau \, \mathrm{d}r = \int_s^t \int_{\mathbb{R}^d} \langle \nabla\zeta, \boldsymbol{v}_{\boldsymbol{r}} \rangle \, \mathrm{d}\boldsymbol{\mu}_{\boldsymbol{r}} \, \mathrm{d}r$$

The limit $(\boldsymbol{\mu}, \boldsymbol{v})$ satisfies the continuity equation.

Proof. Pass to the limit in the discrete continuity equation

$$\Big| \int \zeta \, \mathrm{d}\boldsymbol{M}_{\boldsymbol{\tau}}^{\boldsymbol{n}} - \int \zeta \, \mathrm{d}\boldsymbol{M}_{\boldsymbol{\tau}}^{\boldsymbol{m}} - \tau \sum_{k=m+1}^{n} \int_{\mathbb{R}^d} \langle \nabla\zeta, \boldsymbol{v}_{\boldsymbol{\tau}}^{\boldsymbol{k}} \rangle \, \mathrm{d}\boldsymbol{M}_\tau^k \Big| \le H \sum_{k=m+1}^{k} W_2^2(\boldsymbol{M}_{\boldsymbol{\tau}}^{\boldsymbol{k}}, \boldsymbol{M}_{\boldsymbol{\tau}}^{\boldsymbol{k-1}})$$

$$\text{recalling that} \qquad \sum_{k=0}^{+\infty} W_2^2(\boldsymbol{M}_{\boldsymbol{\tau}}^{\boldsymbol{k}}, \boldsymbol{M}_{\boldsymbol{\tau}}^{\boldsymbol{k-1}}) \le C\tau.$$

Equation for the limit velocity: the Fokker-Planck equation

Let us consider the case of the functional

$$\Phi(\boldsymbol{\mu}) := \mathcal{H}(\boldsymbol{\mu}) + \mathcal{V}(\boldsymbol{\mu}) = \int_{\mathbb{R}^d} u \log u \,\mathrm{d}x + \int_{\mathbb{R}^d} V \,\mathrm{d}\boldsymbol{\mu}, \quad \boldsymbol{\mu} = u\mathscr{L}^d$$

where V is a C^1 potential bounded from below such that $Z := \int_{\mathbb{R}^d} \mathrm{e}^{-V}\,\mathrm{d}x < +\infty$. In this case **$\Phi$ is bounded from below**, since $\Phi(\boldsymbol{\mu}) \geq -\log Z$ for every $\boldsymbol{\mu} \in \mathscr{P}(\mathbb{R}^d)$.

Theorem (JKO '98)

Let us suppose that $\Phi(\boldsymbol{\mu_0}) < +\infty$. The discrete solution $\boldsymbol{M_\tau}$ of the JKO/Minimizing Movement scheme converges to the solution $\boldsymbol{\mu}$ of the Fokker-Planck equation

$$\partial_t \boldsymbol{\mu} - \Delta\boldsymbol{\mu} - \mathsf{div}(\boldsymbol{\mu}\nabla V) = 0.$$

Similar results hold for the other functionals, even if the case of the logarithmic entropy $\mathcal{H}$ is simpler, since gives raise to a linear equation.

Proof

$\boldsymbol{M_\tau}$ converges pointwise to $\boldsymbol{\mu}$ in $\mathscr{P}(\mathbb{R}^d)$ and for every test function $\zeta \in \mathrm{C}^\infty_c(\mathbb{R}^d \times (0,T))$ and for every $0 \le s < r$

$$\lim_{\tau\downarrow 0} \int_0^T \int_{\mathbb{R}^d} \langle \nabla\zeta, \boldsymbol{V_\tau}\rangle \,\mathrm{d}\boldsymbol{M}_\tau \,\mathrm{d}r = \int_0^T \int_{\mathbb{R}^d} \langle \nabla\zeta, \boldsymbol{v_r}\rangle \,\mathrm{d}\boldsymbol{\mu_r} \,\mathrm{d}r = -\int_0^T \int_{\mathbb{R}^d} \partial_t\zeta \,\mathrm{d}\boldsymbol{\mu_r} \,\mathrm{d}r \tag{1}$$

On the other hand, we also know that

$$\int_{\mathbb{R}^d} \langle \boldsymbol{v^n_\tau}, \boldsymbol{\xi}\rangle \,\mathrm{d}\boldsymbol{M^n_\tau} = \int \Big(\operatorname{div}\boldsymbol{\xi} - \langle \nabla V, \boldsymbol{\xi}\rangle\Big) \,\mathrm{d}\boldsymbol{M^n_\tau} \quad \text{for every } \boldsymbol{\xi} \in \mathrm{C}^\infty_c(\mathbb{R}^d;\mathbb{R}^d). \tag{2}$$

In terms of the piecewise constant interpolant (2) is equivalent to

$$\int_{\mathbb{R}^d} \langle \boldsymbol{V_\tau}, \boldsymbol{\xi}\rangle \,\mathrm{d}\boldsymbol{M_\tau} = \int \Big(\operatorname{div}\boldsymbol{\xi} - \langle \nabla V, \boldsymbol{\xi}\rangle\Big) \,\mathrm{d}\boldsymbol{M_\tau} \tag{3}$$

Choosing $\boldsymbol{\xi} = \nabla\zeta$ (time dependent) in (3) and integrating with respect to time we have

$$\int_0^T \int_{\mathbb{R}^d} \langle \nabla\zeta, \boldsymbol{V_\tau}\rangle \,\mathrm{d}\boldsymbol{M_\tau(r)} \,\mathrm{d}r = \int_0^T \Big[\int_{\mathbb{R}^d} \Big(\Delta\zeta - \langle \nabla V, \nabla\zeta\rangle\Big) \,\mathrm{d}\boldsymbol{M_\tau(r)}\Big] \,\mathrm{d}r$$

Passing to the limit in the last identity and using (1) we eventually get

$$-\int_0^T \int_{\mathbb{R}^d} \partial_t\zeta \,\mathrm{d}\boldsymbol{\mu_r} \,\mathrm{d}r = \int_0^T \Big[\int_{\mathbb{R}^d} \Big(\Delta\zeta - \langle \nabla V, \nabla\zeta\rangle\Big) \,\mathrm{d}\boldsymbol{\mu_r}\Big] \,\mathrm{d}r$$