Optimal transport

V Functionals on Wasserstein space

Seongho, Joo

Seoul National University

Problem

- Let Ω be a compact domain, and will be interested in minimization problem involving the sum of three of four terms, namely

$$\min_{\mu \in \mathcal{P}(\Omega} \mathcal{E}_V(\nu) + \mathcal{E}_W(\mu) + \mathcal{E}_F(\mu), \tag{0.1}$$

$$\min_{\mu \in \mathcal{P}(\Omega)} W_2^2(\mu, \nu) + \mathcal{E}_V(\mu) + \mathcal{E}_W(\mu) + \mathcal{E}_F(\mu), \tag{0.2}$$

where in the second case the probability measure ν is given. The functionals $\mathcal{E}_V, \mathcal{E}_W$ and $\mathcal{E}_f$ are called potentials, interaction and internal energy and are defined as follows:

- The potential energy $\mathcal{E}_v$ is associated to potential $V : \Omega \to \mathbb{R} \cup \{+\infty\}$ and defined as

$$\mathcal{E}_V(\mu) := \int_\Omega V \,\mathrm{d}\mu$$

  It tends to attract the mass of μ toward areas where V is minimal.

- The interaction energy $\mathcal{E}_W$ is a sort of potential energy associated with pairs of particles, associated to a potential $W : \Omega \to \mathbb{R} \cup \{+\infty\}$ and defined as

$$\mathcal{E}_W(\mu) := \int_\Omega \int_\Omega W(x - y) \,\mathrm{d}\mu(x) \,\mathrm{d}\mu(y).$$

  The term can both be attractive $(W(z) = \|z\|^2)$ or repulsive $(W(z) = -\log(\|z\|))$.

Cont.

- The internal energy is a generalization of the mathematical entropy $\rho \in \mathcal{P}^{ac} \to \int_\Omega \rho \log \rho$, and is repulsive as it favors mass distributions that are evenly spread in the domain. To define it, one needs a function F : $\mathsf{r}^+ \to \mathsf{r} \cup \{+\infty\}$,

$$\mathcal{E}_F(\mu) = \begin{cases} \int_\Omega F(\rho(x))\,\mathrm{d}x & \text{if } \mu << \lambda \text{ and } \rho := \frac{\mathrm{d}\mu}{\mathrm{d}\lambda} \\ +\infty & \text{if not} \end{cases} \tag{0.3}$$

Minimization problems of type 0.1 and 0.2 occur very frequently in mathematical physics, economics, and biology.

Existence of minimizers to 0.1

Since Ω is bounded, probability measures in $\mathcal{P}$ automatically bounded second moment. Therefore W_2 metrizes the topology induced by $C_b(\Omega) = C^0(\Omega)$, and $(\mathcal{P}(\Omega), W_2)$ is compact.

Proposition 1

If V and W are lower semi-continuous, then the energies $\mathcal{E}_V$ (resp. $\mathcal{E}_W$) are lower semi-continuous on $\mathcal{P}(\Omega)$ with respect to narrow convergence. Moreover, $\mathcal{E}_V$ is convex.

Proof.

Existence of minimizers to 0.1

Since Ω is bounded, probability measures in $\mathcal{P}$ automatically bounded second moment. Therefore W_2 metrizes the topology induced by $C_b(\Omega) = C^0(\Omega)$, and $(\mathcal{P}(\Omega), W_2)$ is compact.

Proposition 1

If V and W are lower semi-continuous, then the energies $\mathcal{E}_V$ (resp. $\mathcal{E}_W$) are lower semi-continuous on $\mathcal{P}(\Omega)$ with respect to narrow convergence. Moreover, $\mathcal{E}_V$ is convex.

Proof. For $\mathcal{E}_V$, the proof is the same as for the lower semi-continuity of the optimal transport problem (i.e. write $V = \sup_k V_k$ where V_k is k-Lipschitz and bounded and pointwise increasing in k). The same strategy works for $\mathcal{E}_W$, but in addition one has to prove that if (μ_k) converges narrowly to μ, then $(\mu_k \otimes \nu_k)$ converges narrowly to $\mu \otimes \mu$.

Lemma 1

Let $\{\mu_k\}_{k=1}^{\infty}$ and $\{\nu_k\}_{k=1}^{\infty}$ be sequences in $\mathcal{P}(\Omega)$ converging narrowly to μ, ν. Then, $\mu_k \otimes \nu_k$ converges narrowly to $\mu \otimes \nu$.

Cont.

Lemma 2

Let $\{\mu_k\}_{k=1}^{\infty}$ and $\{\nu_k\}_{k=1}^{\infty}$ be sequences in $\mathcal{P}(\Omega)$ converging narrowly to μ, ν. Then, $\mu_k \otimes \nu_k$ converges narrowly to $\mu \otimes \nu$.

Proof. Let $\varphi, \psi \in C^0(\Omega)$. Then, by the assumption,

$$\int \varphi \otimes \psi \, \mathrm{d}\mu_k \otimes \nu_k = \left(\int \varphi_k \, \mathrm{d}\mu_k \right) \left(\int \psi \, \mathrm{d}\nu_k \right) \overset{k \to +\infty}{\longrightarrow} \int \varphi \otimes \psi \, \mathrm{d}\mu \otimes \nu,$$

so that $\mathcal{A}$ is the algebra generated by the set $\{\varphi \otimes \psi \,|\, \varphi \in C^0(\Omega\}$, then

$$\forall f \in \mathcal{A}, \quad \int f \, \mathrm{d}\mu_k \otimes \nu_k \overset{k \to +\infty}{\longrightarrow} \int f \, \mathrm{d}\mu \otimes \nu$$

By Stone-Weierstrass, this algebra is dense in $C^0(\Omega \times \Omega)$, showing that $\mu_k \otimes \mu_k$ converges narrowly to $\mu \otimes \nu$.

Existence of minimizers to 0.1

Proposition 2

Let $\Omega \subseteq \mathbb{R}^d$ compact and let $F : \mathbb{R}^+ \to \mathbb{R} \cup \{+\infty\}$ be convex, lower semicontinuous, and superlinear (i.e. $\lim_{r\to\infty} F(r)/r = +\infty$), then $\mathcal{E}_F$ is lower semi-continuous on $\mathcal{P}(\Omega)$ and convex along curves of the form $t \mapsto (1-t)\rho_0 + t\rho_1$.

Proof.

Existence of minimizers to 0.1

Proposition 3 (Convexity of the functional)

Given any $\sigma \in \mathcal{P}(\Omega)$, the function $\rho \in \mathcal{P}(\Omega) \to W_2^2(\sigma, \rho)$ is convex along the curves of the form $\rho_t = (1-t)\rho_0 + t\rho_1$, and it is even strictly convex if $\sigma \in \mathcal{P}^{ac}$.

Proof.

Existence of minimizers to 0.1

Proposition 3 (Convexity of the functional)

Given any $\sigma \in \mathcal{P}(\Omega)$, the function $\rho \in \mathcal{P}(\Omega) \to W_2^2(\sigma, \rho)$ is convex along the curves of the form $\rho_t = (1-t)\rho_0 + t\rho_1$, and it is even strictly convex if $\sigma \in \mathcal{P}^{ac}$.

Proof. Let $\rho_0, \rho_1 \in \mathcal{P}(\Omega)$ and $\gamma_i \in \Gamma(\sigma, \rho_i),\ i \in \{0,1\}$ be optimal transport plans. Then $\gamma_t = (1-t)\gamma_0 + t\gamma_1$ is a transport plan between σ and $\rho_t = (1-t)\rho_0 + t\rho_1$ so that

$$W_2^2(\sigma, \rho_t) \leq \int \|x-y\|^2 \, \mathrm{d}\gamma_t(x,y) \leq (1-t)W_2^2(\sigma, \rho_0) + tW_2^2(\sigma, \rho_1).$$

If σ is absolutely continuous, $\gamma_i = (\mathrm{id}, T_i)_{\#}\rho_i$ where T_i is an optimal transport map between σ and ρ_i. Assume that

$$W_2^2(\sigma, \rho_t) = (1-t)W_2^2(\sigma, \rho_0) + tW_2^2(\sigma, \rho_1) = \int \|x-y\|^2 \, \mathrm{d}\gamma_t(x,y)$$

Thus, γ_t is the unique optimal transport plan between σ and ρ_t, i.e. $\gamma_t = (\mathrm{id}, T_t)_{\#}\sigma$ where T_t is the optimal transport map between σ and ρ_t. Thus,

$$(\mathrm{id}, T_t)_{\#}\sigma = (1-t)(\mathrm{id}, T_0)_{\#}\sigma + t(\mathrm{id}, T_1)_{\#}\sigma.$$

If $0 < t < 1$, this implies that $T_0 = T_1 = T_t$ σ-almost everywhere.

Optimality conditions

- Here we will deal in more detail with the following example, where $\sigma \in \mathcal{P}(\Omega)$:

$$\mathcal{J}(\rho) = \frac{1}{2\tau} W_2^2(\sigma, \rho) + \int V \, \mathrm{d}\rho + \int \rho \log \rho \qquad (0.4)$$

where we assume that V is a Lipschitz vector field.

Proposition 4

$\mathcal{J}$ admits a unique minimizer on Ω, denoted ρ. Moreover:

- $\rho > 0$ a.e.
- $\log(\rho) \in L^1(\Omega)$
- if $(\varphi, \psi) \in Lip(\Omega)^2$ are Kantorovich potentials associated to the optimal transport problem between ρ and σ, then
$$\frac{\varphi}{2\tau} + V + \log \rho = C \text{ a.e.}$$
- $\log \rho \in Lip(\Omega)$, and if $T = \mathrm{id} - \frac{\nabla \varphi}{2}$ is the optimal transport map between ρ and σ,
$$\frac{\mathrm{id} - T}{2\tau} + \nabla V + \nabla \log \rho = 0 \text{ a.e.}$$