Lecture 3: Functionals over $\mathcal{P}(\Omega)$ and hidden convexity

Luca Nenna

March 2, 2022

1 Problem

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

$$\min_{\mu \in \mathcal{P}(\Omega)} \mathcal{E}_V(\mu) + \mathcal{E}_W(\mu) + \mathcal{E}_F(\mu), \tag{1.1}$$

$$\min_{\mu \in \mathcal{P}(\Omega)} \mathcal{T}_c(\mu, \nu) + \mathcal{E}_V(\mu) + \mathcal{E}_W(\mu) + \mathcal{E}_F(\mu), \tag{1.2}$$

where in the second case the probability measure ν is given and $\mathcal{T}_c(\mu, \nu)$ denotes the minimal transport cost between μ and ν with a cost function c. The functionals $\mathcal{E}_V$, $\mathcal{E}_W$ and $\mathcal{E}_F$ are called potential, interaction and internal energy and are defined as follows:

- The potential energy $\mathcal{E}_V$ is associated to a 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 towards areas where V is minimal.
- The interaction energy $\mathcal{E}_W$ is a sort of potential energy associated to pairs of particles, associated to a potential $W : \Omega \to \mathbb{R} \cup \{+\infty\}$ and defined as
$$\mathcal{E}_V(\mu) := \int_\Omega \int_\Omega W(x - y) \mathrm{d}\mu(x) \mathrm{d}\mu(y).$$
This term can both be attractive ($W(z) = \|z\|^2$) or repulsive ($W(z) = -\log(\|z\|)$).
- The internal energy is a generalization of the mathematical entropy[1] $\rho \in \mathcal{P}^{\mathrm{ac}} \mapsto \int_\Omega \rho \log \rho$, and is repulsive as it favors mass distributions that are evenly spread in the domain. To define it, one needs a function $F : \mathbb{R}^+ \to \mathbb{R} \cup \{+\infty\}$,
$$\mathcal{E}_F(\mu) = \begin{cases} \int_\Omega F(\rho(x)) \mathrm{d}x \text{ if } \mu \ll \lambda \text{ and } \rho := \frac{\mathrm{d}\mu}{\mathrm{d}\lambda} \\ +\infty \text{ if not,} \end{cases} \tag{1.3}$$
where λ is the Lebesgue measure.

Minimization problems of the type (4.7) and (1.2) occur very frequently in mathematical physics, chemistry, machine learning, economics, biology. Before treating existence, uniqueness of minimisers and optimality conditions, we need some definitions and important properties.

[1] Note that the mathematical entropy is equal to minus the physical entropy. In particular, it decreases in time when evaluated e.g. on solutions of the heat equation.

Definition 1.1 (Wasserstein (or Monge-Kantorovich) distance)**.** Let $\mu, \nu \in \mathcal{P}(\Omega)$ and $p \in [1, +\infty)$, we set

$$\mathcal{W}_p(\mu, \nu) = \mathcal{T}_c(\mu, \nu)^{\frac{1}{p}},$$

for $c(x, y) = d(x, y)^p$ and d is a distance on Ω (we will always consider $d(x, y)^p = |x - y|^p$. We refer to $\mathcal{W}_p$ as the $p-$Wasserstein (or Monge-Kantorovich) distance.

Proposition 1.2. *Let Ω a compact domain. $\mathcal{W}_p$ is a distance over $\mathcal{P}(\Omega)$*

Theorem 1.3. *If $\Omega \subset \mathbb{R}^d$ is compact and $p \in [1, +\infty)$, in the space $(\mathcal{P}(\Omega), \mathcal{W}_p)$, we have $\mu_n \rightharpoonup \mu$ if and only if $\mathcal{W}_p(\mu_n, \mu) \to 0$.*

Theorem 1.4 (Existence and uniqueness of optimal transport map)**.** *Given $\mu, \nu \in \mathcal{P}(\Omega)$, where Ω is a compact domain, there exists an optimal transport plan for the cost $c(x, y) = h(x - y)$, with h strictly convex. It is unique and of the form $(Id, T)_\sharp \mu$, provided μ is absolutely continuous and $\partial\Omega$ is negligible. Moreover, there exists a Kantorovich potential φ and T and the potential φ are linked by*

$$T(x) = x - (\nabla h)^{-1}(\nabla \varphi(x)).$$

2 Existence of minimizers to (4.7)

Since Ω is bounded, probability measures in $\mathcal{P}(\Omega)$ automatically have bounded second moment. Therefore, $\mathcal{W}_2$ metrizes the topology induced by $\mathcal{C}_b(\Omega) = \mathcal{C}^0(\Omega)$, and $(\mathcal{P}(\Omega), \mathcal{W}_2)$ is compact.

Proposition 2.1. *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 increading 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 $(\mu_k \otimes \mu_k)$ converges narrowly to $\mu \otimes \mu$. □

Lemma 2.2. *Let $(\mu_k)_k$ and $(\nu_k)_k$ 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 $\varphi, \psi \in \mathcal{C}^0(\Omega)$. Then, by hypothesis,

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

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

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

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

Proposition 2.3. *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. Let $F^* : t \mapsto \sup_{t \geqslant 0} st - F(t)$, so that $F^*(s) + F(t) \geqslant st$. By superlinearity, one can see that $F^* : t \mapsto \sup_{t \geqslant 0} st - F(t)$ is finite on $\mathbb{R}^+$, and therefore continous on $\mathbb{R}^+$. If $\mu \in \mathcal{P}(\Omega)$ has density ρ with respect to the Lebesgue measure, then for any bounded measurable function f,

$$\mathcal{E}_F(\mu) = \int F(\rho)\mathrm{d}\lambda \geqslant \int \rho f - F^*(f)\mathrm{d}\lambda.$$

Moreover, by Fenchel-Moreau theorem ($F = F^{**}$ for F convex l.s.c.), one has $F(s) = F^{**}(s) = \sup_{t \in \mathbb{R}} st - F^*(t)$. We therefore get

$$\forall \mu \in \mathcal{P}^{\mathrm{ac}}(\Omega),\ \mathcal{E}_F(\mu) = \sup_{f \text{ measurable bounded}} \int f\mathrm{d}\mu - \int F^*(f)\mathrm{d}\lambda.$$

We now define

$$\overline{\mathcal{E}_F}(\mu) = \sup_{f \in \mathcal{C}^0(\Omega)} \int f\mathrm{d}\mu - \int F^*(f)\mathrm{d}\lambda,$$

and show that

$$\forall \mu \in \mathcal{P}(\Omega),\ \overline{\mathcal{E}_F}(\mu) = \sup_{f \text{ measurable bounded}} \int f\mathrm{d}\mu - \int F^*(f)\mathrm{d}\lambda.$$

Since the space of continuous functions is included in the space of measurable bounded function, we automatically have one inequality. To show the other inequality, we need to approximate measurable bounded functions by continuous ones. Using Lusin's theorem, for any $f : \Omega \to \mathbb{R}$ measurable, we have the existence of $K \subseteq \Omega$ compact and $g \in \mathcal{C}^0(\Omega)$ such that $f|\,K = g|\,K$ and $(\lambda + \mu)(\Omega \setminus K) \leqslant \varepsilon$. Moreover, one can impose that $\|g\|_\infty \leqslant \|f\|_\infty + \mathrm{diam}(\Omega)$. Then,

$$\left| \int f\mathrm{d}\mu - \int g\mathrm{d}\mu \right| = \left| \int_{\Omega \setminus K} (f - g)\mathrm{d}\mu \right| \leqslant \varepsilon(2\,\|f\|_\infty + \mathrm{diam}(\Omega)).$$

$$\left| \int F^*(f)\mathrm{d}\lambda - \int F^*(g)\mathrm{d}\lambda \right| = \left| \int_{\Omega \setminus K} (F^*(f) - F^*(g))\mathrm{d}\lambda \right| \leqslant 2\varepsilon \max_{[0, \|f\|_\infty + \mathrm{diam}(\Omega)]} |F^*|\,.$$

Since this can be done for any $\varepsilon > 0$, the second inequality is established.

This shows that $\mathcal{E}_F = \overline{\mathcal{E}_F}$ on $\mathcal{P}^{\mathrm{ac}}(\Omega)$. Now, let $\mu \in \mathcal{P}(\Omega) \setminus \mathcal{P}^{\mathrm{ac}}(\Omega)$. This implies the existence of a set $S \subseteq \Omega$ such that $\lambda(S) = 0$ and $\mu(S) > 0$. Defining $f = N\mathbf{1}_S$ for $N \in \mathbb{N}$ we get

$$\overline{\mathcal{E}_F}(\mu) \geqslant N\mu(S) - \lambda(\Omega)F^*(0) \xrightarrow{N \to +\infty} +\infty.$$

Therefore $\mathcal{E}_F$ coincides with convex lsc function $\mathcal{E}_F$. □

Proposition 2.4. *Given any $\sigma \in \mathcal{P}(\Omega)$ and $c(x, y) = h(x - y)$ with h strictly convex, the function $\rho \in \mathcal{P}(\Omega) \mapsto \mathcal{T}_c(\sigma, \rho)$ is convex along curves of the form $\rho_t = (1 - t)\rho_0 + t\rho_1$, and it is even strictly convex if $\sigma \in \mathcal{P}^{\mathrm{ac}}(\Omega)$. Moreover, assume there exists a unique pair (ψ_ρ^c, ψ_ρ) of Kantorovich potentials between ρ and σ, then $\frac{\delta \mathcal{T}_c(\sigma, \cdot)}{\delta \mu}(\rho) = \psi_\rho$.*

Proof. Let $\rho_0, \rho_1 \in \mathcal{P}(\Omega)$ and $\gamma_i \in \Pi(\sigma, \rho_i)$ 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

$$\mathcal{T}_c(\sigma, \rho_t) \leqslant \int h(x - y)\mathrm{d}\gamma_t(x, y) \leqslant (1 - t)\mathcal{T}_c(\sigma, \rho_0) + t\mathcal{T}_c(\sigma, \rho_1).$$

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 by contradiction that $\rho_0 \neq \rho_1$ and $t \in (0,1)$ are such that

$$\mathcal{T}_c(\sigma, \rho_t) = (1-t)\mathcal{T}_c(\sigma, \rho_0) + t\mathcal{T}_c(\sigma, \rho_1) = h(x-y)\mathrm{d}\gamma_t(x,y).$$

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 $0 < t < 1$, since γ_t must be induced by an optimal transport map, we get that $T_0 = T_1 = T_t$ σ-almost everywhere. But this is a contradiction with $\rho_0 \neq \rho_1$ and proves strict convexity. We now have to prove that $\frac{\delta \mathcal{T}_c(\sigma,\cdot)}{\delta \mu}(\rho) = \psi_\rho$. Take $\rho_\varepsilon = \rho + \varepsilon \chi$ with $\chi = \tilde{\rho} - \rho$ and estimate the ration $(\mathcal{T}_c(\sigma, \rho_\varepsilon) - \mathcal{T}_c(\sigma, \rho))/\varepsilon$. By using that (ψ_ρ^c, ψ_ρ) is optimal for ρ but not necessarily for ρ_ε we get

$$\frac{\mathcal{T}_c(\sigma, \rho_\varepsilon) - \mathcal{T}_c(\sigma, \rho)}{\varepsilon} \geqslant \frac{\int \psi_\rho \mathrm{d}\rho_\varepsilon + \int \psi_\rho^c \mathrm{d}\sigma - \int \psi_\rho \mathrm{d}\rho - \int \psi_\rho^c \mathrm{d}\sigma}{\varepsilon} = \int \psi_\rho \mathrm{d}\chi,$$

which give the lower bound $\liminf_{\varepsilon \to 0} (\mathcal{T}_c(\sigma, \rho_\varepsilon) - \mathcal{T}_c(\sigma, \rho))/\varepsilon \geqslant \int \psi_\rho \mathrm{d}\chi$. Consider now a sequence of values of ε_k realising the limp, then we can estimate the same ratio using the optimality of the pair $(\psi_{\varepsilon_k}^c, \psi_{\varepsilon_k})$ between ρ_{ε_k} and σ and get

$$\frac{\mathcal{T}_c(\sigma, \rho_{\varepsilon_k}) - \mathcal{T}_c(\sigma, \rho)}{\varepsilon_k} \leqslant \frac{\int \psi_{\varepsilon_k} \mathrm{d}\rho_{\varepsilon_k} + \int \psi_{\varepsilon_k}^c \mathrm{d}\sigma - \int \psi_{\varepsilon_k} \mathrm{d}\rho - \int \psi_{\varepsilon_k}^c \mathrm{d}\sigma}{\varepsilon_k} = \int \psi_{\varepsilon_k} \mathrm{d}\chi,$$

and we now have to pass to the limit in k. As in the proof of existence for the dual problem we have uniform convergence (up to a subsequence) $(\psi_{\varepsilon_k}^c, \psi_{\varepsilon_k}) \to (\psi^c, \psi)$ and (ψ^c, ψ) must be optimal for the transport between ρ and σ. By uniqueness we have that $\psi_\rho = \psi$. We finally obtain that $\limsup_{\varepsilon \to 0} (\mathcal{T}_c(\sigma, \rho_\varepsilon) - \mathcal{T}_c(\sigma, \rho))/\varepsilon \leqslant \int \psi_\rho \mathrm{d}\chi$. □

Remark 2.5. A direct consequence of the theorem above is the (strict) convexity of $\mathcal{W}_p^p$ with $p > 1$.

As a consequence, (1.2) admits a uniques solution if $\mathcal{E}_W = 0$.

3 Optimality conditions

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

$$\mathcal{J}(\rho) = \frac{1}{2\tau} \mathcal{W}_2^2(\sigma, \rho) + \int V \mathrm{d}\rho + \int \rho \log \rho, \tag{3.4}$$

where we assume that V is a Lipschitz on the compact domain Ω.

Proposition 3.1. *$\mathcal{J}$ admits a unique minimiser on Ω, denoted ρ. Moreover:*

- $\rho > 0$ *a.e.*
- $\log(\rho) \in \mathrm{L}^1(\Omega)$
- *if $(\varphi, \psi) \in \mathrm{Lip}(\Omega)^2$ are Kantorovich potentials associated to the optimal transport problem between ρ and σ, then*

$$\frac{\varphi}{2\tau} + V + \log \rho = C \ a.e.$$

- $\log\rho \in \mathrm{Lip}(\Omega)$, *and if* $T = \mathrm{id} - \frac{\nabla\varphi}{2}$ *is the optimal transport map between* ρ *and* σ,

$$\frac{\mathrm{id} - T}{2\tau} + \nabla V + \nabla \log \rho = 0 \ \textit{a.e.}$$

Proof. **Step 1.** Let χ be the probability measure with constant positive density $\kappa = \frac{1}{|\Omega|}$ on Ω. We let $\rho_\varepsilon = (1-\varepsilon)\rho + \varepsilon\chi$. Then, by convexity of $\varepsilon \mapsto \mathcal{W}_2^2(\sigma, \rho_\varepsilon)$,

$$\mathcal{W}_2^2(\sigma, \rho_\varepsilon) \leqslant \mathcal{W}_2^2(\sigma, \rho) + \varepsilon(\mathcal{W}_2^2(\sigma, \chi) - \mathcal{W}_2^2(\sigma, \rho))$$

and by convexity of $\varepsilon \mapsto \mathcal{E}_V(\rho_\varepsilon)$,

$$\mathcal{E}_V(\rho_\varepsilon) \leqslant \mathcal{E}_V(\rho) + \varepsilon(\mathcal{E}_V(\chi) - \mathcal{E}_V(\rho)).$$

We will now upper bound the internal energy $\mathcal{E}_F(\rho_\varepsilon)$. Let $D \subseteq \Omega$ be a measurable set on which ρ vanishes. First,

$$\int_D \rho_\varepsilon \log \rho_\varepsilon = \varepsilon\kappa \log(\varepsilon\kappa)|D|$$

Second, by convexity of $F(r) = r\log r$, and using $F'(r) = \log r + 1$ we have

$$\begin{aligned} F(\rho) &\geqslant F(\rho_\varepsilon) + (\rho - \rho_\varepsilon)F'(\rho_\varepsilon) \\ &= F(\rho_\varepsilon) + \varepsilon(\rho - \kappa)(\log(\rho_\varepsilon) + 1) \\ &\geqslant F(\rho_\varepsilon) + \varepsilon(\rho - \kappa)(\log(\kappa) + 1) \end{aligned}$$

so that

$$\mathcal{E}_F(\rho_\varepsilon) \leqslant \mathcal{E}_F(\rho) + \varepsilon\kappa\log(\varepsilon\kappa)|D| - \varepsilon \int_{\Omega\setminus D} (\rho - \kappa)(\log(\kappa) + 1).$$

Finally we have

$$\mathcal{J}(\rho) \leqslant \mathcal{J}(\rho_\varepsilon) \leqslant \mathcal{J}(\rho) + \varepsilon\kappa\log(\varepsilon\kappa)|D| + C\varepsilon,$$

implying that $-\kappa\log(\varepsilon\kappa)|D| \leqslant C$. Letting $\varepsilon \to 0$ we get a contradiction unless $\lambda_\Omega(D) = 0$.

Step 2. Let us now show that $\log(\rho) \in \mathrm{L}^1(\Omega)$. We already know that

$$(\rho - \kappa)(\log(\rho) + 1) \geqslant (\rho - \kappa)(\log(\kappa) + 1),$$

and this lower bound is integrable. In addition, using the same arguments as above and Fatou's lemma we get

$$\int_\Omega (\rho - \kappa)(\log(\rho_\varepsilon) + 1) \leqslant C \Longrightarrow \int_\Omega (\rho - \kappa)(\log(\rho) + 1) \leqslant C.$$

We therefore get that $(\rho - \kappa)(\log(\rho) + 1) \in \mathrm{L}^1(\Omega)$. Since in addition ρ and $\rho\log\rho \in \mathrm{L}^1(\Omega)$ we get $\log\rho \in \Omega$.

Step 3. Let $\chi \in \mathcal{P}(\Omega) \cap \mathrm{L}^\infty(\Omega)$ and $\rho_\varepsilon = (1-\varepsilon)\rho + \varepsilon\chi$. Then, one can show that

$$\left.\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0} \mathcal{W}_2^2(\sigma, \rho_\varepsilon) = \int \varphi \mathrm{d}(\chi - \rho).$$

Easy computations also show

$$\left.\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0} \mathcal{E}_F(\rho_\varepsilon) + \mathcal{E}_V(\rho_\varepsilon) = \int (\log\rho + V)\chi,$$

thus implying by optimality of ρ that

$$\forall \chi \in \mathcal{P}(\Omega) \cap \mathrm{L}^\infty(\Omega), \quad \int \varphi + \log \rho + V \mathrm{d}(\chi - \rho) \geqslant 0.$$

Set $g = \varphi + \log \rho + V$. The previous inequality can be reformulated as

$$\forall \chi \in \mathcal{P}(\Omega) \cap \mathrm{L}^\infty(\Omega), \quad \int g \mathrm{d}\chi \geqslant \int g \mathrm{d}\rho.$$

This implies that ρ is supported on the set $\{x \mid g(x) = \ell\}$ where ℓ is the essential infimum of g. Since $\mathrm{spt}(\rho) = \Omega$, this shows that g is constant. □

Remark 3.2 (Gradient flow in the Wasserstein space)**.** One can look at (3.4) of an iterate of the following gradient flow scheme

$$\rho_\tau^{(k+1)} \in \operatorname{argmin} \frac{1}{2\tau} \mathcal{W}_2^2(\rho, \rho_\tau^{(k)}) + \int V \mathrm{d}\rho + \int \rho \log \rho.$$

It can be shown that at the limit $\tau \to 0$ one can find a solution to the equation

$$\partial_t \rho - \Delta \rho - \operatorname{div}(\rho \nabla V) = 0,$$

with no-flux boundary condition.

4 Convexity along geodesics and generalized geodesics

For simplifying the exposition, we will study geodesic convexity only on the set of absolutely continuous measures, and for the exponent $p = 2$ only. Given two measures μ_0, μ_1 in $\mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ (:=the space of ac probability measures having finite second moment), we recall that the unique minimizing geodesic between μ_0 and μ_1 is given by

$$\mu_t := [(1-t)\mathrm{id} + tT]_\# \mu_0,$$

where T is the optimal transport plan between μ_0 and μ_1 for $c = \|\cdot\|^2$.

Definition 4.1 (Geodesic convexity for sets)**.** A set $S \subseteq \mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ is called geodesically convex if for any $\mu_0, \mu_1 \in S$, the $\mathcal{W}_2$–geodesic μ_t remains in S.

Definition 4.2 (Geodesic convexity for functions)**.** A function $\mathcal{E}$ from $\mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ to $\mathbb{R} \cup \{+\infty\}$ is *geodesically convex* if and only if for any $\mu_0, \mu_1 \in \mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$,

$$\mathcal{E}(\mu_t) \leqslant (1-t)\mathcal{E}(\mu_0) + t\mathcal{E}(\mu_1) \tag{4.5}$$

where (μ_t) is the $\mathcal{W}_2$–geodesic.

Following McCann, a geodesically convex function is often called displacement convex. A function $\mathcal{E}$ is *strictly geodesically convex* (or strictly displacement convex) if for any $t \in (0,1)$, the inequality (4.5) is strict unless $\mu_0 = \mu_1$.

Proposition 4.3. *The set $\mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ is geodesically convex. More precisely, given $\mu_0 \in \mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ and $\mu_1 \in \mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$, one has $\mu_t \in \mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ for any $t \in [0,1)$.*

Proof. Let $\mu_0 \in \mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d), \mu_1 \in \mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ and $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be a convex Kantorovich potential so that $\mu_t = ((1-t)\mathrm{id} + t\nabla\varphi)_{\#}\mu_0$ is the unique Wasserstein geodesic between μ_0 and μ_1. Define $T_t = (1-t)\mathrm{id} + t\nabla\varphi$. Then, for any $x, y \in \mathrm{spt}(\mu_0)$,

$$\begin{aligned} \langle T_t(x) - T_t(y) | x - y \rangle &= (1-t) \|x-y\|^2 + t\langle \nabla\varphi(x) - \nabla\varphi(y) | x - y \rangle \\ &\geqslant (1-t) \|x-y\|^2 , \end{aligned}$$

where we used the monotonicity of the gradient of convex functions to get the inequality. In particular, if $x \neq y$ and $t < 1$, then $T_t(x) \neq T_t(y)$ and the inverse map T_t^{-1} is well-defined. Moreover, the same inequality shows that T_t^{-1} is Lipschitz with constant $L = 1/(1-t)$. In addition, T_t^{-1} transports μ_t to μ_0, i.e. $\mu_t(B) = \mu_0(T_t^{-1}(B))$ for any Borel set B. Thus, if N is Lebesgue-negligible, $T_t^{-1}(N)$ is also negligible (by the next lemma), so that $\mu_t(N) = \mu_0(T_t^{-1}(N)) = 0$. This implies that $\mu_t \ll \lambda$. □

Lemma 4.4. *If N is Lebesgue-negligible, and if S is Lipschitz, then $S(N)$ is Lebesgue-negligible.*

Proof. By definition, for any $\varepsilon > 0$, there exists $(x_k, r_k)_{1 \leqslant k \leqslant +\infty}$ such that $N \subseteq \bigcup_k \mathrm{B}(x_k, r_k)$ and $\sum_k \lambda(B(x_k, r_k)) \leqslant \varepsilon$. Then, by the Lipschitz property,

$$T_t^{-1}(N) \subseteq \bigcup_k B(T_t^{-1}(x_k), Lr_k),$$

so that $\lambda(T_t^{-1}(N)) \leqslant L^d \sum_k \lambda(B(x_k, r_k)) \leqslant L^d\varepsilon$. □

4.1 Displacement convexity of $\mathcal{E}_V, \mathcal{E}_W$ and $\mathcal{E}_F$

Theorem 4.5 (McCann). *If $V, W : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ are convex, then $\mathcal{E}_V$ and $\mathcal{E}_W$ are displacement convex on $\mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$. Moreover,*

- *If V is strictly convex, then so is $\mathcal{E}_V$, i.e. for $t \in (0,1)$*

$$\mathcal{E}_V(\mu_t) \leqslant (1-t)\mathcal{E}_V(\mu_0) + t\mathcal{E}_V(\mu_1),$$

  with equality if and only if $\mu_0 = \mu_1$.

- *If W is strictly convex, then $\mathcal{E}_W$ is "strictly convex up to translations". More precisely, for any $t \in (0,1)$,*

$$\mathcal{E}_W(\mu_t) \leqslant (1-t)\mathcal{E}_W(\mu_0) + t\mathcal{E}_W(\mu_1),$$

  with equality if and only if μ_1 is a translation of μ_0.

Remark 4.6. Under the same assumption, $\mathcal{E}_V$ and $\mathcal{E}_W$ are also displacement convex on $\mathcal{P}_2(\mathbb{R}^d)$. To prove this, one needs to replace the optimal transport map in the definition of the Wasserstein geodesic by an optimal transport plan (i.e. $\mu_t = \pi_{t\#}\gamma$ where $\pi_t(x,y) = (1-t)x + ty$, see the previous lesson). Taking $\mu_0 = \delta_{x_0}$ and $\mu_1 = \delta_{x_1}$, one obtains that $\mathcal{E}_V$ is convex iff V is.

Remark 4.7. Note that the potential energy is always convex in the classical sense (i.e. $\mathcal{E}_V((1-t)\mu_0 + t\mu_1) \leqslant (1-t)\mathcal{E}_V(\mu_0) + t\mathcal{E}_V(\mu_1)$), but the interaction energy can be non-convex. For instance, when $W = \|\cdot\|^2$,

$$\begin{aligned}
\mathcal{E}_W(\mu) &= \int\int \|x-y\|^2 \,\mathrm{d}\mu(x)\mathrm{d}\mu(y) \\
&= 2\int \|x\|^2 \,\mathrm{d}\mu - 2\int\int \langle x|y\rangle \mathrm{d}\mu(x)\mathrm{d}\mu(y) \\
&= 2\left(\int \|x\|^2 \,\mathrm{d}\mu - \left(\int x\mathrm{d}\mu(x)\right)^2\right),
\end{aligned}$$

which is concave with respect to μ.

Proof. Let $\mu_0, \mu_1 \in \mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ and $\mu_t = ((1-t)\mathrm{id} + tT)_\#\mu_0$ with T the optimal transport map between μ_0 and μ_1. Then

$$\begin{aligned}
\mathcal{E}_V(\mu_t) &= \int_{\mathbb{R}^d} V(x)\mathrm{d}\mu_t(x) \\
&= \int_{\mathbb{R}^d} V((1-t)x + tT(x))\mathrm{d}\mu_0(x) \\
&\leqslant (1-t)\int_{\mathbb{R}^d} V(x)\mathrm{d}\mu_0(x) + t\int_{\mathbb{R}^d} V(T(x))\mathrm{d}\mu_0(x) \\
&= (1-t)\mathcal{E}_V(\mu_0) + t\mathcal{E}_V(\mu_1).
\end{aligned}$$

Equality holds if all inequalities are equalities. In particular, this implies that for μ_0-almost every x one has $V((1-t)x + tT(x)) = (1-t)V(x) + tV(T(x))$. If $t \in (0,1)$, this implies by strict convexity of V, this gives $T = \mathrm{id}$ μ_0–a.e., so that $\mu_1 = \mathrm{id}_\#\mu_0 = \mu_0$.

For $\mathcal{E}_W$ the proof is similar,

$$\begin{aligned}
\mathcal{E}_W(\mu_t) &= \int_{\mathbb{R}^d} W(x-y)\mathrm{d}\mu_t(x)\mathrm{d}\mu_t(y) \\
&= \int_{\mathbb{R}^d} W((1-t)x + tT(x) - (1-t)y + tT(y))\mathrm{d}\mu_0(x)\mathrm{d}\mu_0(y) \\
&\leqslant \int_{\mathbb{R}^d} (1-t)W(x-y) + tW(T(x) - T(y))\mathrm{d}\mu_0(x)\mathrm{d}\mu_0(y) \\
&= (1-t)\mathcal{E}_W(\mu_0) + t\mathcal{E}_W(\mu_1)
\end{aligned}$$

Note that equality holds if and only if all inequalities are equalities. For $t \in (0,1)$ and using the strict convexity of W, this gives that for $\mu_0 \otimes \mu_0$-almost every (x,y) one must have $x - y = T(x) - T(y)$. This implies that $x - T(x) = y - T(y)$ is constant. Hence, T is a translation. ☐

Theorem 4.8 (McCann)**.** *Let $F : [0, +\infty) \to \mathbb{R} \cup \{+\infty\}$ be such that*

(i) $F(0) = 0$ and

(ii) $r \mapsto F(r^{-d})r^d$ is convex non-increasing.

Then $\mathcal{E}_F$ is displacement convex on $\mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$.

This theorem is a corollary of the more general result below. Indeed, take $\mu_0 = \mu \in \mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$, $\varphi_0 = \frac{1}{2}\|\cdot\|^2$ and $\varphi_1 = \varphi$ a Kantorovich potential for the optimal transport problem between μ_0 and μ_1, i.e. $\nabla\varphi_{1\#}\mu_0 = \mu_1$. Then,

$$((1-t)\nabla\varphi_0 + t\nabla\varphi_1)_{\#}\mu = ((1-t)\mathrm{id} + t\nabla\varphi)_{\#}\mu_0 = \mu_t$$

is the Wasserstein geodesic between μ_0 and μ_1.

Theorem 4.9. *Let $\mu \in \mathcal{P}^{\mathrm{ac}}(\mathbb{R}^d)$ and let $\varphi_0, \varphi_1 : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be two convex functions such that $\mathrm{spt}(\mu) \subseteq \overline{\mathrm{dom}(\varphi_i)}$. If $F : [0, +\infty[\to [0, +\infty[$ is such that*

(i) $F(0) = 0$,

(ii) $r \mapsto F(r^{-d})r^d$ *is convex non-increasing,*

then

$$t \in [0, 1] \mapsto \mathcal{E}_F\left[((1-t)\nabla\varphi_0 + t\nabla\varphi_1)_{\#}\mu\right].$$

is convex

We only prove this theorem when the functions φ_0 and φ_1 are $\mathcal{C}^2$ and uniformly convex. The proof in the general case can be found in the article of McCann [1] or in Villani's first book [2].

Lemma 4.10. *Let $\mu \in \mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ with density ρ, $\varphi \in \mathcal{C}^2(\mathbb{R}^d)$ be uniformly convex (that is $\exists\lambda > 0$ such that $\mathrm{D}^2\varphi \geqslant \lambda\mathrm{id}$), and $F(0) = 0$, then*

$$\mathcal{E}_F(\nabla\varphi_{\#}\mu) = \int_{\mathbb{R}^d} F\left(\frac{\rho(x)}{\det(\mathrm{D}^2\varphi(x))}\right)\det(\mathrm{D}^2\varphi(x))\mathrm{d}x.$$

Proof. Since $\mathrm{D}^2\varphi \geqslant \lambda$, setting $x_t = (1-t)y + tx$, one gets

$$\langle x - y | \nabla\varphi(x) - \nabla\varphi(y)\rangle = \langle x - y | \int_0^1 \mathrm{D}^2\varphi(x_t)\cdot(x-y)\rangle \geqslant \lambda\|x-y\|^2,$$

so that $T := \nabla\varphi$ is bijective and has Lipschitz inverse. As in Proposition 4.3, this implies that $T_{\#}\mu$ is absolutely continuous with respect to the Lebesgue measure. We denote σ the density of $T_{\#}\mu$. Then, by the change of variable formula $y = T(x)$ and using $\det(\mathrm{D}T(x)) = |\det \mathrm{D}T(x))|$,

$$\mathcal{E}_F(\nabla\varphi_{\#}\mu) = \int F(\sigma(y))\mathrm{d}y = \int F(\sigma(T(x)))\det(\mathrm{D}T(x))\mathrm{d}x. \tag{4.6}$$

Combining $T_{\#}\mu = \sigma$ and the change of variable formula one gets,

$$\begin{aligned}\forall\varphi \in \mathcal{C}_b(\mathbb{R}^d), \int \rho(x)\varphi(x)\mathrm{d}x &= \int \sigma(y)\varphi(T^{-1}(y))\mathrm{d}y \\ &= \int \sigma(T(x))\det(\mathrm{D}T(x))\varphi(x)\mathrm{d}x\end{aligned}$$

Then, the equality $\rho(x) = \sigma(T(x))\det(\mathrm{D}T(x))$ holds for a set with full measure in $\mathbb{R}^d$. Putting this equality into Eq. (4.6), gives the desired formula. □

Lemma 4.11. *The map $M \mapsto \det(M)^{1/d}$ is concave over the set of symmetric positive d-by-d matrices.*

Proof. Recall Hadamard's formula for a symmetric positive matrix M:

$$\det(M) = \min_{e_1,\dots,e_d} \langle e_1|Me_1\rangle \cdots \langle e_d|Me_d\rangle,$$

where the minimum is taken over orthonormal bases. Given a fixed orthonormal basis $e_1,\dots,e_d$ consider $f(M) = (\langle e_1|Me_1\rangle \cdots \langle e_d|Me_d\rangle)^{1/d}$. Then f is concave over the set of matrices M satisfying $\langle e_i|Me_i\rangle \geqslant 0$ as the composition of the geometric mean $(x \in (\mathbb{R}^+)^d \mapsto (x_1\cdots x_d)^{1/d})$ with linear functions. Then, $\det(\cdot)^{1/d}$ is concave over the set of symmetric positive matrices, as a minimum of concave functions. □

Proof of Theorem 4.9. We prove the theorem only when φ_i are $\mathcal{C}^2$ and uniformly convex. Then, $\varphi_t := (1-t)\varphi_0 + t\varphi_1$ is also $\mathcal{C}^2$ and uniformly convex. Hence, by Lemma 4.10, $\mathcal{E}_F(\nabla\varphi_{t\#}\mu) = \int_{\mathbb{R}^d} B(D(x,t))\rho(x)\mathrm{d}x$, where we have set $B(r) = F(r^{-d})r^d$ and $D(x,t) = (\det(\mathrm{D}^2\varphi_t(x))/\rho(x))^{1/d}$. By Lemma 4.11, for all $x \in \mathbb{R}^d$, $t \in [0,1] \mapsto D(x,t)$ is concave so that

$$D(x,t) \geqslant (1-t)D(x,0) + tD(x,1).$$

Hence, since B is non-decreasing and convex,

$$B(D(x,t)) \leqslant B((1-t)D(x,0) + tD(x,1)) \leqslant (1-t)B(D(x,0)) + tB(D(x,1)).$$

Integrating this inequality gives the desired convexity result. □

Corollary 4.12. *The functionals $\mathcal{E}_F$ generated by the following functions are displacement convex:*

- $F(r) = r^q$ *for* $q > 1$*;*
- $F(r) = r\log r$*;*
- $F(r) = -r^m$ *for* $m \in [1-1/d, 1)$*. (Note that in this case the function is not superlinear at infinity.)*

Proof. Let $B(r) = F(r^{-d})r^d$. In the three cases, the functions B are given respectively by $B(r) = r^{d(1-q)}$, $B(r) = -d\log r$ and $B(r) = -r^{m(1-d)}$, which are all three convex non-increasing under the given assumptions. □

Corollary 4.13. *Given $q \in (1,+\infty]$ and any constant $C \geqslant 0$, the set*

$$\left\{\mu \in \mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d) \mid \left\|\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}\right\|_{\mathrm{L}^q(\mathbb{R}^d)} \leqslant C\right\}$$

is geodesically convex.

Corollary 4.14 (Brunn-Minkowski inequality)**.** *Let K_0, K_1 be two compact subsets of $\mathbb{R}^d$ and $K_t = (1-t)K_0 + tK_1$. Then,*

$$\log\lambda(K_t) \geqslant (1-t)\log\lambda(K_0) + t\log\lambda(K_1).$$

Proof. If K_0 or K_1 have zero volume, there is nothing to prove. If not, consider the probability measures $\mu_i = \frac{1}{\lambda(K_i)}\,\lambda|_{K_i}$ and take $F(r) = r\log r$. Then,

$$\mathcal{E}_F(\mu_i) = \int_{K_i} \frac{1}{\lambda(K_i)}\log\left(\frac{1}{\lambda(K_i)}\right)\mathrm{d}x = -\log(\lambda(K_i)),$$

and, setting $\rho_t = \mathrm{d}\mu_t/\mathrm{d}\lambda$,

$$\begin{aligned}\mathcal{E}_F(\mu_t) = \int_{\mathrm{spt}(\mu_t)} F(\rho_t(x))\mathrm{d}x &= \lambda(\mathrm{spt}(\mu_t)) \left(\frac{1}{\lambda(\mathrm{spt}(\mu_t))} \int_{\mathrm{spt}(\mu_t)} F(\rho_t(x))\mathrm{d}x \right) \\ &\geqslant \lambda(\mathrm{spt}(\mu_t)) F \left(\frac{1}{\lambda(\mathrm{spt}(\mu_t))} \int_{\mathrm{spt}(\mu_t)} \rho_t \right) \\ &= -\log \lambda(\mathrm{spt}(\mu_t))\end{aligned}$$

Since $T(K_0) \subseteq K_1$ we have $\mathrm{spt}(\mu_t) \subseteq ((1-t)\mathrm{id} + tT)(K_0) \subseteq K_t$. We conclude using the displacement convexity of $\mathcal{E}_F$:

$$\begin{aligned}-\log \lambda(K_t) &\leqslant \mathcal{E}_F(\mu_t) \leqslant (1-t)\mathcal{E}_F(\mu_0) + t\mathcal{E}_F(\mu_1) \\ &= -\left[(1-t)\log \lambda(K_0) + t \log \lambda(K_1)\right]\end{aligned} \qquad \square$$

Exercise 4.15. Prove the Brunn-Minkowski inequality in the case $\lambda(K_0) = \lambda(K_1) = 1$ using Corollary 4.13 with $q = +\infty$

4.2 On interacting gas and ground state

An important application (which actually was the initial motivation in [1]) of all the theory we have developed so far consists in establishing the existence of stationary configurations, particularly optimisers, and their properties of interacting gas models.
Consider a $d-$ dimensional gas of particles. The state of the gas is represented by it mass density $\rho \in \mathcal{P}^{ac}(\mathbb{R}^d)$. An attraction between the particles with increases with distance is represented by a strictly convex interaction potential W. Resistance of the gas to the compression is modelled by an equation of state in which the pressure depends on the local density only. Notice that the thermodynamical pressure is given by

$$P(\rho) = \rho F'(\rho) - F(\rho).$$

The question is then: is it possible for these two forces to balance each other and if they do, must the system be in a uniquely determined, stable equilibrium state?
This problem can formulated in a variational formulation which turns out to be

$$\min_{\mu \in \mathcal{P}(\mathbb{R}^d)} \mathcal{E}_V(\mu) + \mathcal{E}_W(\mu) + \mathcal{E}_F(\mu), \tag{4.7}$$

notice that one can also add the effect of an electrostatic potential V. In the original paper by McCann V is taken to be 0. When $F(r) = r^q$ and $q = 5/3$ in $d = 3$, the internal energy is the semi-classical approximation of the quantum kinetic energy of a gas of fermions. Then the following theorem holds

Theorem 4.16. *Consider the following functional*

$$\min_{\mu \in \mathcal{P}(\mathbb{R}^d)} \mathcal{E}_V(\mu) + \mathcal{E}_W(\mu) + \mathcal{E}_F(\mu).$$

Assume that V and W are l.s.c and strictly convex. Let F be l.s.c. and such that the hypothesis of 4.8 are satisfied. Then, there exists at most one minimiser on the set of absolutely continuous probability measures on $\mathbb{R}^d$.

Remark 4.17. We want to give some additional remarks in order to understand the physical meaning of hypothesis (ii) in 4.8.

- Consider a uniform cloud of $d-$dimensional gas with mass M in a volume V, so that the density is constant and equal to M/V. Assume that the gas expands: its dimensions are multiplied by a factor λ, so its volume is multiplied by λ^d and its density divided by λ^d. The internal energy, as a function of the dilation factor λ, is then $V\lambda^d F(\lambda^{-d}M/V)$, which is proportional to $r^d F(r^{-d})$. Condition (ii) means that the internal energy is a convex non-increasing function of this dilation factor. Note that physical realism requires at least that the internal energy be a non-increasing function.
- The first derivative of $r \mapsto r^d F(r^{-d})$ is $-dr^{d-1}P(r^{-d})$ so the non-increasing property is equivalent to the non-negativity of the pressure which makes physical sense. By computing the second derivative of $r \mapsto r^d F(r^{-d})$ and knowing that $P(0) = 0$, one easily sees that P should be non-decreasing and moreover since $P'(\rho) = \rho F''(\rho)$, it follows that F must be convex.

References

[1] Robert J McCann, *A convexity principle for interacting gases*, Advances in mathematics **128** (1997), no. 1, 153–179.

[2] Cédric Villani, *Topics in optimal transportation*, no. 58, American Mathematical Soc., 2003.