Wasserstein barycenters from a PDE perspective

Guillaume Carlier [a]

Based on joint works with M. Agueh, K. Eichinger and A. Kroshnin,
Erlangen, FAU, CAA online seminar, June 9th 2020.

[a]CEREMADE, Université Paris Dauphine and MOKAPLAN (Inria-Dauphine).

Outline

① Definition and characterization

② Integrability, convexity

③ Numerical approximation

④ Regularized barycenters

⑤ Asymptotics (LLN, CLT)

Definition and characterization

Quadratic Wasserstein distance, let $\mathcal{P}_2(\mathbf{R}^d)$ be the set of probability measures on $\mathbf{R}^d$ with a finite second moment, let ρ_0 and ρ_1 be in $\mathcal{P}_2(\mathbf{R}^d)$, 2-Wasserstein distance between ρ_0 and ρ_1, $W_2(\rho_0, \rho_1)$

$$W_2^2(\rho_0, \rho_1) = \inf_{\gamma \in \Pi(\rho_0, \rho_1)} \int_{\mathbf{R}^d \times \mathbf{R}^d} |x - y|^2 \mathrm{d}\gamma(x, y)$$

where $\Pi(\rho_0, \rho_1)$ is the set of transport plans between ρ_0 and ρ_1 i.e. prob. on $\mathbf{R}^d \times \mathbf{R}^d$ having ρ_0 and ρ_1 as marginals. W_2 is a distance on $\mathcal{P}_2(\mathbf{R}^d)$, $(\mathcal{P}_2(\mathbf{R}^d), W_2)$ Wasserstein space.

Let $T : \mathbf{R}^d \to \mathbf{R}^k$ Borel and $\nu \in \mathcal{P}(\mathbf{R}^d)$, the push-forward of ν through T (or image measure) is the measure $T_\# \nu \in \mathcal{P}(\mathbf{R}^k)$ defined by

$$T_\# \nu(A) = \nu(T^{-1}(A)), \ \forall A \text{ Borel subset of } \mathbf{R}^k$$

or, equivalently,

$$\int_{\mathbf{R}^k} \varphi \mathrm{d} T_\# \nu = \int_{\mathbf{R}^d} \varphi(T(x)) \mathrm{d}\nu(x)$$

for every $\varphi \in C_b(\mathbf{R}^k)$. One says that T transports ν to ρ if $T_\# \nu = \rho$.

Brenier, McCann: if ρ_0 does not charge Lipschitz hypersurfaces, there is a unique solution, characterized by $\gamma = (\mathrm{id}, \nabla\varphi)_{\#}\rho_0$ with φ convex. Brenier's map: $\nabla\varphi$, extension of the monotone rearrangement to several dimensions. Link with Monge-Ampère:

$$\det(D^2\varphi)\rho_1(\nabla\varphi) = \rho_0, \quad \varphi \text{ convex.}$$

Regularity theory (Caffarelli, Figalli, De Philippis).
Interpolation (McCann): curve of measures $t \in [0,1] \mapsto \rho_t = ((1-t)\,\mathrm{id} + t\nabla\varphi)_{\#}\rho_0$, geodesic between ρ_0 and ρ_1.

Given $\nu_1, \dots, \nu_N$ in $\mathcal{P}_2(\mathbf{R}^d)^N$ and $\lambda_i > 0$, $\sum_{i=1}^N \lambda_i = 1$, Wasserstein barycenter problem

$$\inf_{\rho \in \mathcal{P}_2(\mathbf{R}^d)} \sum_{i=1}^{N} \lambda_i W_2^2(\nu_i, \rho). \tag{1}$$

This is a convex problem, existence is easy, uniqueness holds as soon as one of the measures ν_i does not charge hypersurfaces (in this case $\rho \mapsto W_2^2(\rho, \nu_i)$ is strictly convex). Solution: Wasserstein barycenter. No regularizing effect. Special instance of Fréchet mean.

Variants:

- Riemannian manifold instead of $\mathbf{R}^d$, Kim and Pass.
- Barycenter of a general probability P over $\mathcal{P}_2(\mathbf{R}^d)$ with

$$\int_{\mathcal{P}_2(\mathbf{R}^d)} m_2(\nu) \mathrm{d}P(\nu) < +\infty, \; m_2(\nu) := \int_{\mathbf{R}^d} |x|^2 \mathrm{d}\nu(x)$$

$$\inf_{\rho \in \mathcal{P}_2(\mathbf{R}^d)} \int_{\mathcal{P}_2(\mathbf{R}^d)} W_2^2(\nu, \rho) \mathrm{d}P(\nu). \tag{2}$$

Bigot and Klein, Loubès and Le Gouic.

Characterization of the barycenter ρ. Kantorovich duality formula:

$$\frac{1}{2}W_2^2(\rho,\nu_i) = \sup\{\int_{\mathbf{R}^d} u_i \mathrm{d}\rho + \int_{\mathbf{R}^d} v_i \mathrm{d}\nu_i \ : \ u_i(x)+v_i(y) \leq \frac{1}{2}|x-y|^2\}$$

achieved by a pair of potentials (u_i, v_i) which are semiconcave and related through

$$\varphi_i := \frac{1}{2}|.|^2 - u_i \text{ convex with } \varphi_i^* := \frac{1}{2}|.|^2 - v_i$$

optimal transport plan γ_i between ρ and ν_i: on the support of γ_i the constraint is an equality $u_i(x) + v_i(y) = \frac{1}{2}|x-y|^2$, i.e. $\varphi_i(x) + \varphi_i^*(y) = x \cdot y$ i.e.

$$y \in \partial\varphi_i(x).$$

Optimality condition for the barycenter ρ:

$$\sum_{i=1}^{N} \lambda_i u_i \geq 0, \quad \text{with equality on } \operatorname{spt}(\rho)$$

i.e.

$$\frac{1}{2}|x|^2 \geq \sum_{i=1}^{N} \lambda_i \varphi_i(x), \ \forall x \in \mathbf{R}^d \ \ \text{with equality on } \operatorname{spt}(\rho)$$

since the φ_i's are convex, they all are differentiable on the contact set and in particular everywhere on $\operatorname{spt}(\rho)$, and

$$x = \sum_{i=1}^{N} \lambda_i \nabla \varphi_i(x), \ \forall x \in \operatorname{spt}(\rho). \tag{3}$$

and $\varphi_{i\#}\rho = \nu_i$, there exists a (Lipschitz) optimal transport from the barycenter to the fixed measures ν_i (no assumption on ν_i).

The barycenter (actually its support) is characterized by a free-boundary problem for a system of Monge-Ampère equations: φ_i convex,

$$\frac{1}{2}|x|^2 \geq \sum_{i=1}^{N} \lambda_i \varphi_i(x),\ \forall x \in \mathbf{R}^d \ \text{ with equality on spt}(\rho)$$

and

$$\det(D^2\varphi_i)\nu_i(\nabla\varphi_i) = \rho,\ i = 1, \ldots, N.$$

Explicit examples:

- $N = 2$, two measures: their barycenters coincide with McCann's interpolation,
- $d = 1$, ν_1 atomless T_i monotone $T_{i\#}\nu_1 = \nu_i$, $\rho = (\sum_{i=1}^{N} \lambda_i T_i)_{\#}\nu_1$, (in particular barycenters are associative: false in higher dimensions),
- barycenters of Gaussians are Gaussians.

Integrability, convexity

McCann's displacement convexity, ν_1, ν_2, optimal transport $\nabla\varphi$, $\rho_\lambda = ((1-\lambda)\,\mathrm{id} + \lambda\nabla\varphi)_\# \nu_1$, $\lambda \in [0,1]$. A functional E: $\mathcal{P}_2(\mathbf{R}^d) \to \mathbf{R} \cup \{+\infty\}$ is displacement convex if

$$E(\rho_\lambda) \leq (1-\lambda)E(\nu_1) + \lambda E(\nu_2).$$

Convexity along barycenters if for ρ a barycenter of the ν_i's with weights λ_i one has

$$E(\rho) \leq \sum_{i=1}^{N} \lambda_i E(\nu_i).$$

Moments: Let $V\colon \mathbf{R}^d \to \mathbf{R}$ convex and such that $\int_{\mathbf{R}^d} V \mathrm{d}\nu_i < +\infty$ for every i. Recall (3) and $\nabla\varphi_{i\#}\rho = \nu_i$:

$$\begin{aligned}\int_{\mathbf{R}^d} V \mathrm{d}\rho &= \int_{\mathbf{R}^d} V(\sum_{i=1}^N \lambda_i \nabla\varphi_i(x))\mathrm{d}\rho(x) \\ &\leq \sum_{i=1}^N \lambda_i \int_{\mathbf{R}^d} V(\nabla\varphi_i(x))\mathrm{d}\rho(x) \\ &= \sum_{i=1}^N \lambda_i \int_{\mathbf{R}^d} V \mathrm{d}\nu_i.\end{aligned}$$

In particular

$$m_2(\rho) \leq \sum_{i=1}^N \lambda_i m_2(\nu_i), \ \int_{\mathbf{R}^d} x\mathrm{d}\rho(x) = \sum_{i=1}^N \lambda_i \int_{\mathbf{R}^d} y\mathrm{d}\nu_i(y)$$

Integral estimates: Internal energy $E(\rho) = \int_{\mathbf{R}^d} U(\rho(x))\mathrm{d}x$, $U\colon \mathbf{R}_+ \to \mathbf{R}$, $U(0) = 0$, satisfies McCann's condition:

$$t \mapsto t^d U\Big(\frac{1}{t^d}\Big) \text{ convex nonincreasing}$$

e.g. $U(t) = t^p$, $p > 1$, $U(t) = t\log(t)$. McCann showed displacement convexity of such energies. Generalizes to barycenters: assume ν_i absolutely continuous, and with finite energy

$$\sum_{i=1}^{N} \lambda_i \int_{\mathbf{R}^d} U(\nu_i(x))\mathrm{d}x = \sum_{i=1}^{N} \lambda_i \int_{\mathbf{R}^d} U\Big(\frac{\rho}{\det D^2\varphi_i}\Big) \det D^2\varphi_i \mathrm{d}x$$

Set $\Phi_\rho : (t) = t^d U(\frac{\rho}{t^d})$, then by convexity of Φ_ρ

$$\sum_{i=1}^N \lambda_i \int_{\mathbf{R}^d} U(\nu_i(x))\mathrm{d}x = \sum_{i=1}^N \lambda_i \int_{\mathbf{R}^d} \Phi_\rho((\det D^2\varphi_i)^{1/d})$$
$$\geq \int_{\mathbf{R}^d} \Phi_\rho(\sum_{i=1}^N \lambda_i(\det D^2\varphi_i)^{1/d})$$

Since $D^2\varphi_i \geq 0$, Minkowski's concavity inequality together with the optimallity condition $\sum \lambda_i D^2\varphi_i = \mathrm{id}$ on $\mathrm{spt}(\rho)$, gives

$$\sum_{i=1}^N \lambda_i(\det D^2\varphi_i)^{1/d} \leq (\sum_{i=1}^N \lambda_i \det D^2\varphi_i)^{1/d} = 1$$

so that, since Φ_ρ is nondecreasing

$$\Phi_\rho\Big(\sum_{i=1}^N \lambda_i(\det D^2\varphi_i)^{1/d}\Big) \geq \Phi_\rho(1) = U(\rho)$$

The desired convexity inequality follows

$$\int_{\mathbf{R}^d} U(\rho(x))\mathrm{d}x \leq \sum_{i=1}^{N} \lambda_i \int_{\mathbf{R}^d} U(\nu_i(x))\mathrm{d}x.$$

In particular:

- $p \in (1, +\infty)$, $\nu_i \in L^p \Rightarrow \rho \in L^p$ with $\|\rho\|_{L^p}^p \leq \sum_i \lambda_i \|\nu_i\|_{L^p}^p$,
- ν_i has finite entropy so does ρ, $\mathrm{Ent}(\rho) \leq \sum_{i=1}^N \lambda_i \mathrm{Ent}(\nu_i)$,
- if $d = 1$, one can use negative powers as well.

Limit cases:

- $\nu_i \in L^1 \Rightarrow \rho \in L^1$,
- $\nu_1 \in L^\infty \Rightarrow \rho \in L^\infty$ with $\|\rho\|_{L^\infty} \leq \lambda_1^{-d} \|\nu_1\|_{L^\infty}$,
- nothing known about regularity: $\nu_i \in C^{k,\alpha} \Rightarrow \rho \in C^{k,\alpha}$?

Numerical approximation

Reformulate the barycenter problem (1) as a linear pproblem in terms of transport plans γ_i between the fixed measures ν_i and the unkown barycenter ρ, minimize with respect to $(\gamma_1, \ldots, \gamma_N) \in \mathcal{P}(\mathbf{R}^d \times \mathbf{R}^d)$ the cost

$$\sum_{i=1}^{N} \lambda_i \int_{\mathbf{R}^d \times \mathbf{R}^d} \frac{1}{2} |x_i - x|^2 \mathrm{d}\gamma_i(x_i, x)$$

subject to the fixed marginal constraints $((\pi_1, \pi_2)(a, b) = (a, b)$ denote the canonical projections):

$$\pi_{1\#}\gamma_i = \nu_i, \; i = 1, \ldots, N, \qquad (4)$$

as well as the constraint all the γ_i's share the same second marginal (which is the barycenter):

$$\pi_{2\#}\gamma_1 = \pi_{2\#}\gamma_2 = \ldots = \pi_{2\#}\gamma_N (= \rho). \tag{5}$$

Entropic regularization: $\varepsilon > 0$ minimize under the same constraints

$$\sum_{i=1}^{N} \lambda_i \int_{\mathbf{R}^d \times \mathbf{R}^d} \Big(\frac{1}{2}|x_i - x|^2 + \varepsilon \log(\gamma_i(x_i, x)\Big) \mathrm{d}\gamma_i(x_i, x)$$

which is an entropic projection problem:

$$\inf \sum_{i=1}^{N} \lambda_i H(\gamma_i|\theta_i) \, : \, \theta_i(x_i, x) = \exp(-\frac{|x_i - x|^2}{2\varepsilon})$$

where H denotes relative (aka Kullback-Leibler divergence). In other words, we have to Kullback-Leibler project Gaussians onto the linear constraints (4)-(5).

This is a popular, simple and efficient approximation method for computational OT: Sinkhorn scaling algorithm (Cuturi, Cuturi and Peyré's recent book). Alternate projection algorithm: perform one projection at a time (same as coordinate descent, aka nonlinear Gauss-Seidel, on the dual). The key-point is that these projections are totally explicit. KL Projection onto a fixed marginal constraint:

$$\inf H(\gamma_1|\theta_1) = \int \Big(\log\Big(\frac{\gamma_1}{\theta_1}\Big) - 1\Big)\gamma_1 \ : \ \pi_{1\#}\gamma_1 = \nu_1$$

optimality condition (with u_1 a Lagrange multiplier for the fixed marginal constraint):

$$\log\Big(\frac{\gamma_1(x_1, x)}{\theta_1(x_1, x)}\Big) = u_1(x_1) \text{ i.e. } \gamma_1(x_1, x) = a_1(x_1)\theta_1(x_1, x).$$

One finds a_1 thanks to the marginal constraint:

$$a_1(x_1) = \frac{\nu_1(x_1)}{\int_{\mathbf{R}^d} \theta_1(x_1, x) \mathrm{d}x}$$

this is just a scaling as in Sinkhorn. Consider now the KL projection onto the common marginal constraint (5):

$$\inf \sum_{i=1}^{N} \lambda_i H(\gamma_i | \theta_i) \text{ subject to } (5)$$

trick: use Lagrange multipliers by observing that (5) is the same as:

$$\sum_{i=1}^{N} u_i = 0 \Rightarrow \sum_{i=1}^{N} \int_{\mathbf{R}^d \times \mathbf{R}^d} u_i(x) \mathrm{d}\gamma_i(x_i, x) = 0.$$

Optimality conditions:

$$\lambda_i \log\left(\frac{\gamma_i(x_i,x)}{\theta_i(x_i,x)}\right) = u_i(x),\ \sum_{i=1}^{N} u_i = 0, \text{that is}$$

$$\gamma_i(x_i,x) = \theta_i(x_i,x)a_i(x)^{\frac{1}{\lambda_i}},\ \prod_{i=1}^{N} a_i = 1$$

now we solve using the constraint that

$$\rho(x) = \int_{\mathbf{R}^d} \gamma_i(x_i,x)\mathrm{d}x_i = a_i(x)^{\frac{1}{\lambda_i}} \int_{\mathbf{R}^d} \theta_i(x_i,x)\mathrm{d}x_i,\ i = 1,\ldots,N$$

and find an explict geometric mean formula for ρ:

$$\rho(x) = \prod_{i=1}^{N} \rho(x)^{\lambda_i} = \prod_{i=1}^{N} a_i(x) \Big(\int_{\mathbf{R}^d} \theta_i(x_i, x) \mathrm{d}x_i \Big)^{\lambda_i}$$

$$= \prod_{i=1}^{N} \Big(\int_{\mathbf{R}^d} \theta_i(x_i, x) \mathrm{d}x_i \Big)^{\lambda_i}.$$

All steps of Sinkhorn for barycenters are totally explicit. Easy to implement (few lines of codes), converges pretty fast (for ε not too small). Cost of computing the integrals (can be parallelized, well suited for GPU).

Regularized barycenters

Given Ω an open convex subset of $\mathbf{R}^d$ (can be the whole of $\mathbf{R}^d$) and $\mu > 0$, entropic barycenter

$$\inf_{\rho \in \mathcal{P}_2(\mathbf{R}^d),\ \int_\Omega \rho = 1} \sum_{i=1}^{N} \lambda_i W_2^2(\nu_i, \rho) + \mu \int_\Omega \rho \log(\rho). \tag{6}$$

Proposed by Bigot, Cazelles and Papadakis for numerical (avoids discretization artifacts) and statistical puposes. Forces full support in Ω. Different from the entropic regularization using plans in Sinkhorn. Unique solution, the entropic regularized barycenter. As we shall see, this is a purely PDE problem.

More generally let P be a Borel measure over Wasserstein $\mathcal{P}_2(\mathbf{R}^d)$ with

$$\int_{\mathcal{P}_2(\mathbf{R}^d)} m_2(\nu)\mathrm{d}P(\nu) < +\infty$$

the entropic regularized barycenter of P, $\rho = \mathrm{Bar}_{\mu,\Omega}(P)$ (just $\mathrm{Bar}_{\mu}(P)$ if $\Omega = \mathbf{R}^d$) is the solution of

$$\inf_{\rho \in \mathcal{P}_2(\mathbf{R}^d),\ \int_\Omega \rho = 1} \int_{\mathcal{P}_2(\mathbf{R}^d)} W_2^2(\nu, \rho)\mathrm{d}P(\nu) + \mu \int_\Omega \rho \log(\rho). \tag{7}$$

Euler-Lagrange equation is a (possibly infinite) system of Monge-Ampère equations: $\nabla\varphi_\rho^\nu$ the OT from ρ to ν. Then

$$\mu\log(\rho(x)) + \frac{|x|^2}{2} = \int_{\mathcal{P}_2(\mathbf{R}^d)} \varphi_\rho^\nu(x)\mathrm{d}P(\nu)(\text{convex})$$

i.e.

$$\rho(x) = \exp\Big(-\frac{|x|^2}{2\mu} + \frac{1}{\mu}\int_{\mathcal{P}_2(\mathbf{R}^d)} \varphi_\rho^\nu(x)\mathrm{d}P(\nu)\Big)$$

which couples the family of Monge-Ampère equations:

$$\rho = \det(D^2\varphi_\rho^\nu)\nu(\nabla\varphi_\rho^\nu).$$

In particular ρ is less log concave than a Gaussian, $\log(\rho)$ is locally Lipschitz, has a locally BV gradient and

$$\mu\nabla\log(\rho) + \mathrm{id} = \int_{\mathcal{P}_2(\mathbf{R}^d)} \nabla\varphi_\rho^\nu(x)\mathrm{d}P(\nu),\ {\nabla\varphi_\rho^\nu}_{\#}\rho = \nu. \qquad (8)$$

We expect regularizing effects and global bounds, which ones?

Moment bounds Let $V : \mathbf{R}^d \to \mathbf{R}_+$ convex, assume that

$$\int_{\mathcal{P}_2(\mathbf{R}^d)} \int_{\mathbf{R}^d} V \mathrm{d}\nu \mathrm{d}P(\nu) < +\infty$$

on the one hand, convexity and (8) give

$$\int_{\mathbf{R}^d} V(x + \mu \nabla \log \rho)\rho = \int_{\mathbf{R}^d} V\Big(\int_{\mathcal{P}_2(\mathbf{R}^d)} \nabla \varphi_\rho^\nu(x) \mathrm{d}P(\nu)\Big)\rho(x)\mathrm{d}x$$

$$\leq \int_{\mathbf{R}^d} \int_{\mathcal{P}_2(\mathbf{R}^d)} V(\nabla \varphi_\rho^\nu(x))\mathrm{d}P(\nu) d\rho(x) = \int_{\mathcal{P}_2(\mathbf{R}^d)} \int_{\mathbf{R}^d} V \mathrm{d}\nu \mathrm{d}P(\nu)$$

On the other hand if V is $C^{1,1}$ (say), since $V(x + \mu\nabla \log \rho) \geq V(x) + \nabla V(x)\nabla \log(\rho)$ we also have

$$\int_{\mathbf{R}^d} V(x + \mu \nabla \log \rho)\rho \geq \int_{\mathbf{R}^d} (V - \mu \Delta V)\rho$$

So

$$\int_{\mathbf{R}^d} (V - \mu \Delta V)\rho \leq \int_{\mathcal{P}_2(\mathbf{R}^d)} \int_{\mathbf{R}^d} V \mathrm{d}\nu \mathrm{d}P(\nu)$$

for instance for V quadratic this gives

$$m_2(\rho) \leq 2\mu d + \int_{\mathcal{P}_2(\mathbf{R}^d)} m_2(\nu)\mathrm{d}P(\nu).$$

More interestingly, if P has higher moments, $p > 2$, $m_p(\nu) := \int |x|^p \mathrm{d}\nu$ and

$$\int_{\mathcal{P}_2(\mathbf{R}^d)} m_p(\nu)\mathrm{d}P(\nu) < +\infty$$

then $m_p(\rho) < +\infty$ (with an explicit bound).

Sobolev regularity Fisher Information bound: square (8), $\mu^2|\nabla\log(\rho)|^2$ multiply by ρ, integrate and use Jensen and Fubini:

$$\begin{aligned}\mu^2\int_{\mathbf{R}^d}|\nabla\log(\rho)|^2\rho &= \int_{\mathbf{R}^d}\left|\int_{\mathcal{P}_2(\mathbf{R}^d)}(\nabla\varphi_\rho^\nu(x)-x)\mathrm{d}P(\nu)\right|^2\rho(x)\mathrm{d}x\\ &\leq \int_{\mathcal{P}_2(\mathbf{R}^d)}\int_{\mathbf{R}^d}|\nabla\varphi_\rho^\nu(x)-x|^2\rho(x)\mathrm{d}x\mathrm{d}P(\nu)\\ &= \int_{\mathcal{P}_2(\mathbf{R}^d)}W_2^2(\mu,\rho)\mathrm{d}P(\nu)\end{aligned}$$

so that $\sqrt{\rho}\in H^1(\mathbf{R}^d)$. If for $p>2$, $\int_{\mathcal{P}_2(\mathbf{R}^d)}m_p(\nu)\mathrm{d}P(\nu)<+\infty$ then by a similar argument and the bound on $m_p(\rho)$ one gets $\rho^{1/p}\in W^{1,p}(\mathbf{R}^d)$. In particular if $p>d$, $\rho\in C^{0,1-d/p}(\mathbf{R}^d)$.

Strong stability Wasserstein metric between measures on $\mathcal{P}_2(\mathbf{R}^d)$ (with finite second moments):

$$\mathcal{W}_2^2(P,Q) = \inf_{\Gamma \in \Pi(P,Q)} \int_{\mathcal{P}_2(\mathbf{R}^d)\times\mathcal{P}_2(\mathbf{R}^d)} W_2^2(\mu,\nu)\mathrm{d}\Gamma(\mu,\nu). \tag{9}$$

Wasserstein over Wasserstein. Assume $\mathcal{W}_2^2(P_n, P) \to 0$, set $\rho_n = \mathrm{Bar}_\mu(P_n)$, $\rho = \mathrm{Bar}_\mu(P)$, then not only $W_2(\rho_n, \rho) \to 0$ but also

$$\sqrt{\rho_n} \to \sqrt{\rho} \text{ strongly in } H^1(\mathbf{R}^d).$$

Useful for the law of large numbers (see later).

Maximum principle Again by the optimality condition (8), looking at a maximum point of $\log(\rho)$ (suitably regularizing the problem if necessary) one gets that if

$$P(\{\nu \in L^\infty(\mathbf{R}^d),\ \nu \leq M\}) = \alpha > 0$$

then $\rho \in L^\infty(\mathbf{R}^d)$ with the bound

$$\rho \leq \frac{M}{\alpha^d}.$$

Higher regularity I: the bounded case If the data are smooth and compactly supported, $\Omega = B = B(0,r)$:

$$P(\{\nu \ : \ \mathrm{spt}(\nu) \subset B, \ \|\nu\|_{C^{k,\alpha}(B)} + \|\log(\nu)\|_{L^\infty(B)} \leq M\}) = 1$$

then by Caffarelli's regularity theory one gets

$$\rho \in C^{k+2,\alpha}(B).$$

Higher regularity II: the log concave case Observe that $\log(\rho)$ is less concave than $-\frac{1}{2\mu}|.|^2$. Caffarelli's contraction principle: the OT between a standard gaussian γ and $e^{-W}\gamma$ with W convex is 1-Lipschitz. We can use this principle here to deduce that, if there is some $A > 0$ such that P-almost every ν writes as $\mathrm{d}\nu = e^{-V(y)}\mathrm{d}y$ with $D^2V \geq A\,\mathrm{id}$ (in the sense of distributions), then $\log\rho \in C^{1,1}(\mathbf{R}^d)$ and more precisely there holds

$$-\,\mathrm{id} \leq \mu D^2 \log\rho \leq \Big(\frac{1}{\sqrt{\mu A}} - 1\Big)\,\mathrm{id}\,. \tag{10}$$

Asymptotics (LLN, CLT)

A stochastic perspective on the problem, Bigot and Klein, Le Gouic and Loubès, Kroshnin. P is a probability measure on $\mathcal{P}_2(\mathbf{R}^d)$ (with a finite second moment), N large, $\hat{\nu}_1, \ldots, \hat{\nu}_N$ drawn in an i.i.d. way according to P. Empirical barycenter (this is a random measure), $\mu \geq 0$ (regularization or not)

$$\hat{\rho}_N = \mathrm{Bar}_\mu \Big(\frac{1}{N} \sum_{i=1}^{N} \delta_{\hat{\nu}_i} \Big)$$

and the true barycenter

$$\rho = \mathrm{Bar}_\mu(P).$$

Law of large numbers, a.s. convergence of $\hat{\rho}_N$ to ρ,

- Bigot and Klein, Le Gouic and Loubès, weak convergence, $W_2(\hat{\rho}_N, \rho) \to 0$ a.s.,
- $\mu > 0$, strong convergence thanks to the control of the Fisher information (and its convergence), C., Eichinger, Kroshnin.

Can we go one step futher: speed of convergence, asymptotic normality of the error (CLT)? In the un-regularized case, serious nonsmoothness issues (free boundary problem for Monge-Ampère), it is not even clear that the barycenter behaves in a (Wasserstein) Lipschitz way with respect to P.

What we mean by CLT has to be made precise, two ways to look at the Wasserstein space:

- as a (formal) Riemannian manifold (Otto) where the tangent space at ρ consists of $L^2(\rho)$ vector fields (and see variations of ρ as the image of ρ by the corresponding flow over a short time), Benamou-Brenier dynamical formulation...
- as a convex subset of the vector space of measures (or of a space of more regular densities, L^2, H^1...),

With the geometric viewpoint, LLN suggests to work in the tangent $L^2(\rho)$ and to consider $\hat{T}_N$ as the OT from ρ to $\hat{\rho}_N$ ($L^2(\rho, \mathbf{R}^d)$ random variable, natural framework for CLT). Roughly, LLN expresses that $\|\hat{T}_N - \mathrm{id}\,\|_{L^2(\rho)} \to 0$. CLT holds if

$$\hat{h}_N := \sqrt{N}(\hat{T}_N - \mathrm{id}) \text{ converges in law to a Gaussian } \mathcal{N}(0, \Sigma)$$

where Σ is a self-adjoint, nonnegative operator trace class operator. No general result of this kind, except in particular cases we considered with Agueh ($\mu = 0$):

- $d = 1$,
- P is supported by Gaussians.

For $\mu > 0$, we saw that there are regularizing effects, it makes sense to expect that $\sqrt{N}(\hat{\rho}_N - \rho)$ is asymptotically gaussian). Requires control on the linearization of Monge-Ampère.

Theorem 1 (C., Eichinger, Kroshnin) *Under suitable Assumptions on P, $\sqrt{N}(\hat{\rho}_N - \rho)$ converges in law in L^2 to the gaussian with covariance $\Sigma = G^{-1}\mathrm{Var}(\varphi_\rho^\nu)G^{-1}$ where*

$$G(u) = \mu\Big(\frac{u}{\rho} - \frac{1}{|B|}\int_B \frac{u}{\rho}\Big) - \mathbf{E}(\mathbf{\Phi}^\nu)'(\rho).$$