A Conjecture on the Structure of Support-Preserving Mappings between Probability Measures

Gabriel Peyré

July 30, 2024

In the following, we let X and Y be compact metric spaces and denote $\mathbb{P}(X)$ and $\mathbb{P}(Y)$ as the spaces of probability measures on these spaces, endowed with the weak* topology.

Definition 1 (Support-preserving mapping)**.** *A mapping $f : \mathbb{P}(X) \to \mathbb{P}(Y)$ is said to be "support-preserving" if it satisfies that for any n and any set of points $\{x_i\}_{i=1}^n \subset X^n$, there exists $\{y_j\}_{j=1}^n \subset Y^n$ (not necessarily distinct) such that*

$$f\Big(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}\Big) = \frac{1}{n}\sum_{j=1}^n \delta_{y_j}. \tag{S}$$

Remark 1 (Rationale for the name)**.** *A map f satisfying condition (S) is called a support-preserving mapping because it maps a uniform distribution on points to a uniform distribution on points, such that the cardinality of the support is preserved up to multiplicity (because the y_j are not necessarily distinct).*

Remark 2 (Push-forwards)**.** *A push-forward map $f(\mu) = T_\sharp \mu$ where $T : X \to Y$ is continuous satisfies (S). My conjecture is that these are the only ones, with the caveat that $T = T(\mu)$ might depend on μ.*

Conjecture 1. *f is weak* continuous and satisfies (S) if and only if there exists a map $T(\mu)$ (which might depend on μ)*

$$T(\mu) : x \in X \mapsto T(\mu)(x) \in Y$$

such that $(\mu, x) \mapsto T(\mu)(x)$ is continuous for the product topology (weak* on $\mathbb{P}(X)$) and

$$f(\mu) = T(\mu)_\sharp \mu. \tag{P}$$

Remark 3 (Forward direction)**.** *A map f of the form (P) ("parametric" push-forward) satisfies (S). The converse direction is not clear.*

Remark 4 (Intuitions)**.** *A first intuition of why this might be true is that all the maps f I am aware of (see Remark 1 below) satisfy this conjecture. A second intuition is that if one first works for a fixed n it is easy to construct a valid $T(\mu)$ "locally" around μ, for instance by projecting μ and $f(\mu)$ on lines such that there is no collision and then constructing T by assigning points in order along the line (a 1-D optimal transport). But this construction is not valid globally, and it is not clear how to glue them globally, avoiding issues when points collapse. Also, one needs to check that such gluing is consistent across all values of n.*

Remark 5 (Necessity of (S))**.** *Without condition (S), the conclusion (P) is false. For instance, for $X = \mathbb{R}^d$, a convolution $f(\mu) = \mu \star g$ against a smooth kernel g maps discrete measures to measures with density.*

Remark 6 (Necessity of uniform distribution)**.** *If one weakens (S) by only requiring that the cardinality of the support is preserved, but not the uniform distribution, the conjecture is false, because one can modify the mass by setting*

$$\frac{\mathrm{d}f(\mu)}{\mathrm{d}\mu}(x) = \frac{g(x)}{\int g \mathrm{d}\mu}$$

where $g(x) > 0$. Such a map is in general not a push-forward.

Remark 7 (Hypotheses on the ground spaces)**.** *It might be necessary to add constraints on X and Y, with the important case being finite-dimensional Euclidean spaces.*

Remark 8 (Smoothness hypothesis)**.** *It might be necessary to strengthen the smoothness conditions on both f and T, for instance, by requiring Lipschitz continuity.*

Example 1 (Wasserstein flows)**.** *Examples of maps that satisfy (S) on $X = Y = \mathbb{R}^d$ and are parameterized push-forwards of the form (P) are $f(\mu) = \rho_{t=1}$ where ρ_t is given by the Wasserstein gradient flow*

$$\frac{d\rho_t}{dt} = div\,(\rho_t v(\rho_t)) \quad \text{with} \quad \rho_{t=0} = \mu$$

where the vector field $v(\rho_t)$ is the Wasserstein gradient

$$v(\mu) = \nabla_{\mathbb{R}^d}\left[\delta E(\rho_t)\right] = 2\int \nabla_1 k(\cdot, y)\,\mu(y)\,dy$$

of interaction energies (assuming k symmetric)

$$E(\mu) := \int k(x,y)\,d\mu(x)\,d\mu(y).$$

If $k(x,y)$ does not depend on y, then $T(\mu) = T$ (the PDE is just an advection and T is the flow map) does not depend on μ, but otherwise it does. A more complex example is given by the action of transformer neural networks on a distribution of tokens (it is more complex because $v(\mu)$ is not linear in μ).