Evolution of the Wasserstein distance between the marginals of two Markov processes

Benjamin Jourdain (ENPC CERMICS and INRIA Mathrisk)

Joint work with: Aurélien Alfonsi and Jacopo Corbetta

PDE and Probability Methods for Interactions
March 31, 2017

Wasserstein distance

Let $\varrho \geq 1$, P and $\widetilde{P}$ be probability meas. on $\mathbb{R}^d$ s.t. $\int_{\mathbb{R}^d} |x|^\varrho (P + \widetilde{P})(\mathrm{d}x) < \infty$.

Definition of the ϱ–Wasserstein distance

$$W_\varrho(P, \widetilde{P}) = \left(\inf_{\pi \in \Pi(P, \widetilde{P})} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^\varrho \, \pi(\mathrm{d}x, \mathrm{d}y) \right)^{\frac{1}{\varrho}}$$

where $\Pi(P, \widetilde{P})$ is the set of "coupling" measures on $\mathbb{R}^d \times \mathbb{R}^d$ with respective marginals P and $\widetilde{P}$.

Dual Representation

$W_\varrho^\varrho(P, \widetilde{P}) = \sup \left\{ -\int_{\mathbb{R}^d} \phi(x) P(\mathrm{d}x) - \int_{\mathbb{R}^d} \widetilde{\phi}(y) \widetilde{P}(\mathrm{d}y) \right\}$ where the supremum runs over all pairs $(\phi, \widetilde{\phi}) \in L^1(P) \times L^1(\widetilde{P})$ such that $\forall (x, y) \in \mathbb{R}^d \times \mathbb{R}^d$, $-\phi(x) - \widetilde{\phi}(y) \leq |x - y|^\varrho$.

Kantorovich potentials

- There exists a couple of Kantorovich potentials $(\psi, \widetilde{\psi}) \in L^1(P) \times L^1(\widetilde{P})$ satisfying $-\psi(x) - \widetilde{\psi}(y) \leq |x-y|^\varrho$ and such that

$$W_\varrho^\varrho(P, \widetilde{P}) = -\int_{\mathbb{R}^d} \psi(x) P(\mathrm{d}x) - \int_{\mathbb{R}^d} \widetilde{\psi}(y) \widetilde{P}(\mathrm{d}y).$$

- One is the ϱ-transform of the other :

$$\psi(x) = -\inf_{y \in \mathbb{R}^d} \{|x-y|^\varrho + \widetilde{\psi}(y)\}, \widetilde{\psi}(y) = -\inf_{x \in \mathbb{R}^d} \{|x-y|^\varrho + \psi(x)\}.$$

- For an (the when $\varrho > 1$) optimal coupling π, since

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} -(\psi(x) + \widetilde{\psi}(y)) \pi(\mathrm{d}x, \mathrm{d}y) = W_\varrho^\varrho(P, \widetilde{P}) = \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^\varrho \pi(\mathrm{d}x, \mathrm{d}y)$$

$$\pi(\mathrm{d}x, \mathrm{d}y) \text{ a.e.}, -\psi(x) - \widetilde{\psi}(y) = |x-y|^\varrho.$$

Objective

- Obtain a formula for the evolution of $W_\varrho^\varrho(P_t, \widetilde{P}_t)$ where $(P_t)_{t\geq 0}$ and $(\widetilde{P}_t)_{t\geq 0}$ are the marginals of two Markov processes with respective generators L and $\widetilde{L}$.
- Possible applications :
  - stability estimates when $L = \widetilde{L}$ and $P_0 \neq \widetilde{P}_0$,
  - Convergence of $\widetilde{P}_t$ to P_t when $(\widetilde{P}_0, \widetilde{L}) \to (P_0, L)$.

Formal derivation

For each $t \geq 0$, let $(\psi_t, \widetilde{\psi}_t)$ be Kantorovich potentials for $(P_t, \widetilde{P}_t)$

$$W_\varrho^\varrho(P_t, \widetilde{P}_t) = -\int_{\mathbb{R}^d} \psi_t(x)\, P_t(\mathrm{d}x) - \int_{\mathbb{R}^d} \widetilde{\psi}_t(x)\, \widetilde{P}_t(\mathrm{d}x).$$

For every $s \geq 0$

$$W_\varrho^\varrho(P_s, \widetilde{P}_s) \geq -\int_{\mathbb{R}^d} \psi_t(x)\, P_s(\mathrm{d}x) - \int_{\mathbb{R}^d} \widetilde{\psi}_t(x)\, \widetilde{P}_s(\mathrm{d}x)\,.$$

As a consequence

$$\begin{aligned} &W_\varrho^\varrho(P_s, \widetilde{P}_s) - W_\varrho^\varrho(P_t, \widetilde{P}_t) \\ &\qquad \geq \int_{\mathbb{R}^d} \psi_t(x)(P_t(\mathrm{d}x) - P_s(\mathrm{d}x)) + \int_{\mathbb{R}^d} \widetilde{\psi}_t(x)(\widetilde{P}_t(\mathrm{d}x) - \widetilde{P}_s(\mathrm{d}x)) \end{aligned}$$

Formal derivation (2)

$$\int_{\mathbb{R}^d} \psi_t(x)(P_t(\mathrm{d}x) - P_s(\mathrm{d}x)) = \int_s^t \int_{\mathbb{R}^d} L\psi_t(x) P_r(\mathrm{d}x)\mathrm{d}r$$

For the choice $s = t + h$ with $h > 0$, we deduce

$$\frac{1}{h}\Big(W_\varrho^\varrho(P_{t+h}, \widetilde{P}_{t+h}) - W_\varrho^\varrho(P_t, \widetilde{P}_t)\Big)$$
$$\geq \frac{1}{h}\left(-\int_t^{t+h}\int_{\mathbb{R}^d} L\psi_t(x)P_r(\mathrm{d}x)\mathrm{d}r - \int_t^{t+h}\int_{\mathbb{R}^d} \widetilde{L}\widetilde{\psi}_t(x)P_r(\mathrm{d}x)\mathrm{d}r\right)$$

Taking the limit $h \to 0^+$ yields

$$\frac{\mathrm{d}}{\mathrm{d}t^+} W_\varrho^\varrho(P_t, \widetilde{P}_t) \geq -\int_{\mathbb{R}^d} L\psi_t(x)P_t(\mathrm{d}x) - \int_{\mathbb{R}^d} \widetilde{L}\,\widetilde{\psi}_t(x)\widetilde{P}_t(\mathrm{d}x)$$

In a symmetric way, the choice $s = t - h$ leads to

$$\frac{\mathrm{d}}{\mathrm{d}t^-} W_\varrho^\varrho(P_t, \widetilde{P}_t) \leq -\int_{\mathbb{R}^d} L\psi_t(x)P_t(\mathrm{d}x) - \int_{\mathbb{R}^d} \widetilde{L}\,\widetilde{\psi}_t(x)\widetilde{P}_t(\mathrm{d}x)\,.$$

A generic heuristic formula

$$W_\varrho^\varrho(P_t,\widetilde{P}_t) = -\int_{\mathbb{R}^d}\psi_t(x)P_t(\mathrm{d}x) - \int_{\mathbb{R}^d}\widetilde{\psi}_t(y)\widetilde{P}_t(\mathrm{d}y)$$

For all $t \geq 0$

$$\frac{\mathrm{d}}{\mathrm{d}t}W_\varrho^\varrho(P_t,\widetilde{P}_t) = -\int_{\mathbb{R}^d}L\psi_t(x)P_t(\mathrm{d}x) - \int_{\mathbb{R}^d}\widetilde{L}\,\widetilde{\psi}_t(x)\widetilde{P}_t(\mathrm{d}x)\,.$$

Integral formulation: for all $0 \leq s \leq t$

$$W_\varrho^\varrho(P_t,\widetilde{P}_t) - W_\varrho^\varrho(P_s,\widetilde{P}_s) = -\int_s^t\left[\int_{\mathbb{R}^d}L\psi_r(x)P_r(\mathrm{d}x) + \int_{\mathbb{R}^d}\widetilde{L}\,\widetilde{\psi}_r(x)\widetilde{P}_r(\mathrm{d}x)\right]\mathrm{d}r\,.$$

Main Issues

We have to check the following facts in order to make the previous heuristic rigorous:

- $\psi_t \in Dom(L)$ and $\widetilde{\psi}_t \in Dom(\widetilde{L})$?
- $L\psi_t \in L^{1,\mathrm{loc}}_t(L^1(P_t))$ and $\widetilde{L}\widetilde{\psi}_t \in L^{1,\mathrm{loc}}_t L^1(\widetilde{P}_t)$?
- Differentiability of $t \mapsto W^{\varrho}_{\varrho}(P_t, \widetilde{P}_t)$?

Pure Jumps Markov Processes

$$Lf(x) = \lambda(x) \int_{\mathbb{R}^d} (f(y) - f(x))\, k(x, \mathrm{d}y)$$

$$\widetilde{L}f(x) = \widetilde{\lambda}(x) \int_{\mathbb{R}^d} (f(y) - f(x))\, \widetilde{k}(x, \mathrm{d}y).$$

- λ, $\widetilde{\lambda}$ jump rates,
- k, $\widetilde{k}$ probability kernels

Assumptions

- $\sup_{x \in \mathbb{R}^d} \max(\lambda(x), \widetilde{\lambda}(x)) < \infty$
- $t \mapsto \int_{\mathbb{R}^d} |x|^{\varrho(1+\varepsilon)}(P_t(\mathrm{d}x) + \widetilde{P}_t(\mathrm{d}x))$ is locally bounded.

Lemma

Let $\alpha \geq 1$. If $\int_{\mathbb{R}^d} |x|^\alpha P_0(\mathrm{d}x) < \infty$ and $\sup_{x \in \mathbb{R}^d} \int_{\mathbb{R}^d} k(x, \mathrm{d}y)|y - x|^\alpha < +\infty$, then $t \mapsto \int_{\mathbb{R}^d} |x|^\alpha P_t(\mathrm{d}x)$ is locally bounded.

Main Result

Theorem

- $t \mapsto \int_{\mathbb{R}^d} |L\psi_t(x)| P_t(\mathrm{d}x) + \int_{\mathbb{R}^d} |\widetilde{L}\,\widetilde{\psi}_t(x)| \widetilde{P}_t(\mathrm{d}x)$ is locally bounded on $(0,+\infty)$.
- $t \mapsto W_\varrho^\varrho(P_t, \widetilde{P}_t)$ is locally Lipschitz on $(0,+\infty)$ and for almost every $t \in (0,\infty)$

$$\frac{\mathrm{d}}{\mathrm{d}t} W_\varrho^\varrho(P_t, \widetilde{P}_t) = -\int_{\mathbb{R}^d} L\psi_t(x) P_t(\mathrm{d}x) - \int_{\mathbb{R}^d} \widetilde{L}\,\widetilde{\psi}_t(x) \widetilde{P}_t(\mathrm{d}x)\,.$$

- for every $t \geq 0$

$$W_\varrho^\varrho(P_t, \widetilde{P}_t) - W_\varrho^\varrho(P_0, \widetilde{P}_0) = \\ -\int_0^t \left[\int_{\mathbb{R}^d} L\psi_r(x) P_r(\mathrm{d}x) + \int_{\mathbb{R}^d} \widetilde{L}\,\widetilde{\psi}_r(x) \widetilde{P}_r(\mathrm{d}x)\right] \mathrm{d}r$$

How to deal with the issues?

Proposition

Let $P, \widetilde{P}$ be two probability measures on $\mathbb{R}^d$ such that $\int_{\mathbb{R}^d} |x|^{\varrho(1+\varepsilon)} P(\mathrm{d}x) + \int_{\mathbb{R}^d} |y|^{\varrho(1+\varepsilon)} \widetilde{P}(\mathrm{d}y) < \infty$ for some $\varepsilon \geq 0$. Then $(\psi, \widetilde{\psi}) \in L^{1+\varepsilon}(P) \times L^{1+\varepsilon}(\widetilde{P})$.

- For $t > 0$, P_t is equivalent to $Q = \sum_{n \in \mathbb{N}} \frac{Q_n}{n!}$ where
$$Q_n(\mathrm{d}x_n) = \int_{(\mathbb{R}^d)^n} P_0(\mathrm{d}x_0) \prod_{j=0}^{n-1} \lambda(x_j) k(x_j, \mathrm{d}x_{j+1}).$$
Permits to transfer integrability from one marginal to another.
- Moreover $\int \lambda(x) k(x, \mathrm{d}y) P_t(\mathrm{d}x)$ is absolutely continuous with respect to Q and for f measurable and $0 \leq s \leq t$, if $\int_s^t \int_{\mathbb{R}^d \times \mathbb{R}^d} \lambda(x) |f(y) - f(x)| k(x, \mathrm{d}y) P_r(\mathrm{d}x) \mathrm{d}r < +\infty$, then $\int_{\mathbb{R}^d} f(x)(P_t(\mathrm{d}x) - P_s(\mathrm{d}x)) = \int_s^t \int_{\mathbb{R}^d} Lf(x) P_r(\mathrm{d}x) \mathrm{d}r$.

Application to Birth and Death process

$$\begin{cases} Lf(x) = \eta(x)(f(x+1) - f(x)) + \nu(x)(f(x-1) - f(x)),\ x \in \mathbb{N} \\ \nu(0) = 0 \end{cases}.$$

Theorem (Joulin 2007)

$$\forall t \geq 0,\ W_1(P_t, \widetilde{P}_t) \leq e^{-\kappa t} W_1(P_0, \widetilde{P}_0)$$

where $\kappa := \inf_{x \in \mathbb{N}}(\eta(x) + \nu(x+1) - \eta(x+1) - \nu(x))$ is the Wasserstein curvature

Estimation of $W_\varrho(P_t, \widetilde{P}_t)$ with $\varrho > 1$?

Extension to general ϱ

Proposition

Assume affine growth of the birth rate η. Let $\varrho \geq 1$, P_0 and $\widetilde{P}_0$ such that $\int_{\mathbb{N}} x(P_0(\mathrm{d}x) + \widetilde{P}_0(\mathrm{d}x)) < \infty$. Then, there exists a constant $C_\varrho \in [0, +\infty)$ such that for any $t \geq 0$,

$$W_\varrho^\varrho(P_t, \widetilde{P}_t) \leq W_\varrho^\varrho(P_0, \widetilde{P}_0)e^{-\kappa\varrho t} \\ + C_\varrho(\mathrm{Lip}(\eta) + \mathrm{Lip}(\nu)) \int_0^t e^{\kappa\varrho(r-t)}\big(W_1(P_r, \widetilde{P}_r) \\ + 1_{\{\varrho>2\}} W_{\varrho-1}^{\varrho-1}(P_r, \widetilde{P}_r)\big)\mathrm{d}r.$$

When $\varrho \in (1, 2]$, $C_\varrho = 1$ and the second term is bounded from above by $(\mathrm{Lip}(\eta) + \mathrm{Lip}(\nu))W_1(P_0, \widetilde{P}_0)\frac{e^{-\kappa t} - e^{-\kappa\varrho t}}{\kappa(\varrho-1)}$.

Sketch of the proof

Let π_r be an optimal coupling at time r

$$W_\varrho^\varrho(P_t, \widetilde{P}_t) = W_\varrho^\varrho(P_0, \widetilde{P}_0)$$
$$+ \int_0^t \sum_{x,y\in\mathbb{N}} \pi_r(x,y)\Big(\eta(x)(\psi_r(x) - \psi_r(x+1)) + \nu(x)(\psi_r(x) - \psi_r(x-1)$$
$$+ \eta(y)(\widetilde{\psi}_r(y) - \widetilde{\psi}_r(y+1)) + \nu(y)(\widetilde{\psi}_r(y) - \widetilde{\psi}_r(y-1))\Big)\mathrm{d}r$$

By optimality, $\pi_r(\mathrm{d}x, \mathrm{d}y)$ a.e., $-\psi_r(x) - \widetilde{\psi}_r(y) = |x-y|^\varrho$.
Moreover, for all $(z,w) \in \mathbb{N}^2$, $-\psi_r(z) - \widetilde{\psi}_r(w) \leq |z-w|^\varrho$ so that

$$\psi_r(x) + \widetilde{\psi}_r(y) - (\psi_r(x+1) + \widetilde{\psi}_r(y)) \leq |x+1-y|^\varrho - |x-y|^\varrho,$$
$$\psi_r(x) + \widetilde{\psi}_r(y) - (\psi_r(x) + \widetilde{\psi}_r(y+1)) \leq |y+1-x|^\varrho - |x-y|^\varrho,$$
$$\psi_r(x) + \widetilde{\psi}_r(y) - (\psi_r(x+1) + \widetilde{\psi}_r(y+1)) \leq |x-y|^\varrho - |x-y|^\varrho = 0.$$

One-dimensional Piecewise Deterministic Markov Processes

$$Lf(x) = V(x)f'(x) + \lambda(x)\int_{\mathbb{R}} (f(y) - f(x))\, k(x, \mathrm{d}y)$$

(i) The vector field V is locally Lipschitz and bounded.

(ii) $\lambda(x)$ is continuous and $sup_x \lambda(x) < \infty$.

(iii) $\exists M < \infty,\ \sup_{x\in\mathbb{R}} \int_{\mathbb{R}} 1_{|x-y|>M} k(x, \mathrm{d}y) = 0$.

(iv)
- either $F_0(x) = P_0((-\infty, x])$ is continuous and $\forall x,\ P(\{x\}) = 0 \Rightarrow \forall x, \int_{\mathbb{R}} k(y, \mathrm{d}x)P(\mathrm{d}y) = 0$
- or $P_0(\mathrm{d}x)$ and $\int_{\mathbb{R}} e^{-\frac{|y|^2}{2}} k(y, \mathrm{d}x)\mathrm{d}y$ have densities w.r.t. $\mathrm{d}x$, .

(v) $x \mapsto k(x, \mathrm{d}y)$ is continuous for the weak topology.

(vi) $\int_{\mathbb{R}} |x|^{\varrho(1+\varepsilon)} P_0(\mathrm{d}x) < \infty$.

(vii) F_0 is increasing and

$$\forall y \geq 0,\ \sup_{x\in\mathbb{R}} \frac{F_0(x+y)}{F_0(x)} \vee \frac{1 - F_0(x-y)}{1 - F_0(x)} \leq ce^{Cy}.$$

The hypotheses on the initial marginal are satisfied by $P_0^\varepsilon = \frac{1}{2\varepsilon} e^{-\frac{|x|}{\varepsilon}} \star P_0$ such that $W_\varrho^\varrho(P_0^\varepsilon, P_0) = \varepsilon^{1+\varrho}\Gamma(1+\varrho)$.

Evolution of the Wassertein distance

Theorem

Let $(P_t)_{t\geq 0}$ and $(\widetilde{P}_t)_{t\geq 0}$ be the time-marginals of two real valued PDMP satisfying the previous Assumptions. Then for every $t \geq 0$

$$W_\varrho^\varrho(P_t, \widetilde{P}_t) - W_\varrho^\varrho(P_0, \widetilde{P}_0) = -\int_0^t \left(\int_{\mathbb{R}} L\,\psi_r\, P_r + \int_{\mathbb{R}} \widetilde{L}\,\widetilde{\psi}_r\, \widetilde{P}_r \right) \mathrm{d}r\,.$$

- Approximation of L by $\frac{e^{\varepsilon L}-1}{\varepsilon}$ which is a pure jump generator with intensity $\leq \frac{1}{\varepsilon}$,
- In general, no equivalent measure permitting to transfer integrability from one marginal to another,
- Use of the explicit optimal coupling in dimension $d = 1$ given by the comonotonic inverse transform sampling :

$$\psi_t'(x) = \varrho(\widetilde{F}_t^{-1}(F_t(x)) - x)|x - \widetilde{F}_t^{-1}(F_t(x))|^{\varrho-2}.$$

Thank you for your attention.