

# Transport costs for PDEs: the coupling method

Nicolas Fournier\*

Benoît Perthame\*\*

June 18, 2020

## Abstract

We informally review a few PDEs for which some transport cost between pairs of solutions, possibly with some judicious cost function, decays: heat equation, Fokker-Planck equation, heat equation with varying coefficients, fractional heat equation with varying coefficients, homogeneous Boltzmann equation for Maxwell molecules, and some nonlinear integro-differential equations arising in neurosciences. We always use the same method, that consists in building a coupling between two solutions. This means that we double the variables and solve, globally in time, a well-chosen PDE posed on the Euclidian square of the physical space. Finally, although the above method fails, we recall a simple idea to treat the case of the porous media equation. We also introduce another method based on the dual Monge-Kantorovich problem.

---

2010 *Mathematics Subject Classification*: 35A05; 35K55; 60J99; 28A33

*Keywords and phrases*: Transport cost; Monge-Kantorovich distance; Coupling; Fokker-Planck equation; Fractional Laplacian; Homogeneous Boltzmann equation; Integro-differential equations; Porous media equation.

*Acknowledgments*: We warmly thank the referees for their numerous fruitful comments.

## Introduction

It is usual to study the well-posedness, stability and large-time behavior of stochastic processes (e.g. solutions to Stochastic Differential Equations) by using coupling methods: we consider two such processes, with different initial conditions, driven by *suitably correlated* randomness, and we measure the  $\varrho$ -transport cost  $\mathcal{T}_\varrho$  between their distributions.

We work in  $\mathbb{R}^d$  and we always assume that the cost function  $\varrho : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}$  satisfies  $\varrho(x, x) = 0$  and  $\varrho(x, y) = \varrho(y, x) > 0$  for  $x \neq y$ . We recall that for two probability densities  $u_1, u_2$  on  $\mathbb{R}^d$ ,

$$\begin{cases} \mathcal{T}_\varrho(u_1, u_2) = \inf_{v \in K(u_1, u_2)} \iint \varrho(x, y) v(x, y) dx dy, \\ K(u_1, u_2) = \{v : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}_+ \text{ such that } \int v(x, y) dy = u_1(x), \int v(x, y) dx = u_2(y)\}. \end{cases} \quad (1)$$

---

\*Sorbonne Université, CNRS, Laboratoire de Probabilité, Statistique et Modélisation, F-75005 Paris, France. Email: Nicolas.Fournier@sorbonne-universite.fr

\*\*Sorbonne Université, CNRS, Université de Paris, Inria, Laboratoire Jacques-Louis Lions, F-75005 Paris, France. Email: Benoit.Perthame@sorbonne-universite.fr. B.P. has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 740623).

When for some  $p \geq 1$ ,

$$\varrho_p(x, y) = \frac{|x - y|^p}{p},$$

and we put  $\mathcal{T}_p = \mathcal{T}_{\varrho_p}$ . The distances  $d_p = \mathcal{T}_p^{1/p}$  are also called Wasserstein distances, and one refers to the Monge-Kantorovich distance when  $p = 1$ .

The probabilistic coupling method consists in finding, for two solutions  $u_1$  and  $u_2$  of a given Partial Differential Equation (PDE in short), two coupled stochastic processes  $X_1$  and  $X_2$ , with time-marginals  $u_1$  and  $u_2$ , so that  $X_1$  and  $X_2$  remain as close as possible. One then controls  $\mathcal{T}_\varrho(u_1(t), u_2(t))$  by  $\mathbb{E}[\varrho(X_1(t), X_2(t))]$ . The goal of the present survey paper is to describe, in an informal way, this method, using only arguments based PDEs. This is what we call the *coupling method*: for each problem, we introduce a PDE, with doubled variables, describing the time-evolution of density  $v(x, y, t)$  of the law at time  $t \geq 0$  of two underlying coupled processes, in such a way that its two marginals solve the original equation, and such that  $v(x, y, t)$  is as concentrated as possible near the diagonal  $x = y$ . This is possible only when the underlying pair of stochastic coupled processes is Markov. We then study the quantity  $\iint \varrho(x, y)v(x, y, t)$ , which controls the transport cost  $\mathcal{T}_\varrho$  between two solutions. This can be seen as a translation, in terms of PDEs, of the probabilistic coupling method. For example, our proof of Tanaka's theorem [38] for the Boltzmann equation really relies on the same main arguments as Tanaka (who is using some Poisson-driven stochastic differential equations), but we avoid introducing any stochastic process.

The difficulties and novelties rely on the choice of the cost function and on the choice of coupling between two solutions by solving a well-chosen PDE posed on the Euclidian square of the physical space,  $\mathbb{R}^{2d}$  in general. Each time, we try to emphasize the main technical difficulties that would allow one to justify the computations.

Of course, any way to produce some non-expansion estimates along solutions of some PDEs for some transport cost relies on some coupling. However, for example, the deterministic methods in Carrillo [12] for the heat equation or Villani [41, Section 7.5.6], see also [5], for Tanaka's theorem, are really different in spirit. In [12], the main tool is that solutions to the heat equation can be represented by a convolution formula. In [41, Section 7.5.6] and [5], everything relies on the contractive property of the gain operator and on the Duhamel principle.

For example, considering the Brownian motion leads to the heat equation. We first give a simple proof that the heat equation is non-expansive (weak contraction) for any smooth cost function of the form  $\varrho(x, y) = r(|x - y|)$ . This is standard but the PDE literature seems to ignore this simple approach. The method can be extended to various cases. The Fokker-Planck equation is the simplest extension. The case of the heat equation with variable coefficients, of the form  $\partial_t u - \Delta(a(x)u) = 0$ , is more involved: in one dimension, the Monge-Kantorovich distance  $\mathcal{T}_1$  plays a central role and is always non-expansive (under technical conditions); we illustrate the general structure in higher dimension and show that if the cost function  $\varrho$  satisfies some elliptic PDE, which does seem to enter a class with generic existence results, then  $\mathcal{T}_\varrho$  is non-expansive along solutions. The method also applies to some jump processes: fractional heat equation with variable coefficients in dimension one, scattering equations, kinetic scattering equations, Boltzmann equation for Maxwell molecules.

For the porous media equation, the situation is more intricate and the above method does not seem to apply. However, we recall from [6] another, somehow related and rather simple, path to treat this equation.

Concerning piecewise deterministic jump processes and (inhomogeneous) kinetic scattering equations, we present a new result showing that some transport costs are non-expansive.

Finally, concerning jump processes, in particular those related to the discretized heat equation, we present another approach, based on the dual formulation of the transport costs.

Maybe the first example of use of the coupling method can be found in Dobrushin [20], where the Vlasov equation is derived as mean-field limit of a deterministic system of interacting particles, making use of some transport cost. No PDE is written for the coupling in [20], because everything may be written in terms of characteristics. See [24, Section 3] for a PDE analogue of Dobrushin's argument. In the same spirit, the Euler equation is derived from a deterministic system of interacting vortices in Marchioro-Pulvirenti [31, Section 5.3], using also a coupling argument. See [26] for a result with the strong transport distance  $d_\infty$  but for a problem slightly less singular than the Euler equation.

Recently, the topic of transport costs has developed quickly for PDEs and integro-differential equations (IDEs) after new understanding of optimal transportation and of the Brenier-Kantorovich map by [10, 2]. There are several approaches to use the transport costs in PDEs. A geometrical approach based on gradient flow structures has been introduced in [34] and extended in [14, 8], in particular for the porous media equation (with  $\mathcal{T}_2$ ), for interacting particle systems and for granular flows. Also, many results on PDEs have been derived from the discretization algorithm named JKO after [27]. See the book [42] for a complete presentation of these results. See also [39] who showed that contractivity for some other transport cost may fail for the porous media equation. Let us also mention that the special structure associated with dimension 1 has been used to prove strict contraction for the porous media equation [13] for the cost  $\mathcal{T}_2$ , and to treat other equations as scalar conservation laws [4], the Keller-Segel system [11] or the granular media equation [30]. See also the general reference [15] for nonlinear diffusion equations and [3] for the multi-dimensional Keller-Segel system. Methods based on optimal transportation have also been recently used to treat singular *congestion* (incompressible) equations arising in crowd modeling, see for instance [9, 32, 17].

Most of the recent papers using some transport cost for PDEs have been using the gradient flow structure which is closely related to a variational formulation of the fluxes. Here, with several examples of conservative equations, which do not necessarily have a gradient flow structure, we control the transport cost using the coupling method. We often borrow our examples from the stochastic processes which represent the PDEs thanks to their Kolmogorov equation. The cases of variable coefficients are particularly interesting because they often require some special choice of the cost function.

We organize our examples as follows. We begin with three simple examples: heat equation, Fokker-Planck equation, and a class of nonlinear transport equations. We show directly that the transport costs are non-expansive along these equations. Then we turn, in Section 2, to the heat equation with variable coefficients. In Section 3 we consider some IDEs: scattering equations, including kinetic scattering and inhomogeneous fractional heat equation. Zero-th order terms, describing absorption and re-emission, as they appear in models of neural networks, can also be treated by adapting the method; this is explained in Section 4. The famous Tanaka theorem for the homogeneous Boltzmann equation can be included in our framework and this is done in Section 5. We treat the porous media equation in Section 6. Finally, we exemplify in Section 7 how the same results can be proved using the dual formulation of the transport costs.

# 1 Heat, Fokker-Planck and transport equations

In order to explain the coupling method in a very simple, but still relevant, framework, we begin with the heat equation. Then we turn to drift and transport terms.

## 1.1 Heat equation

Here is the well-known result, see e.g. [42], we want to quickly recall.

**Theorem 1** Consider any increasing function  $r : [0, \infty) \mapsto [0, \infty)$  such that the cost function  $\varrho : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}_+$  defined by  $\varrho(x, y) = r(|x - y|)$  is of class  $C^2$ . Consider two probability densities  $u_1^0, u_2^0$  on  $\mathbb{R}^d$ , and the corresponding solutions  $u_1, u_2$  to the heat equation

$$\partial_t u - \Delta u = 0, \quad x \in \mathbb{R}^d, t \geq 0. \quad (2)$$

For any  $t \geq 0$ , one has

$$\mathcal{T}_\varrho(u_1(t), u_2(t)) \leq \mathcal{T}_\varrho(u_1^0, u_2^0).$$

**Proof.** We consider an initial density  $v^0 : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}_+$  with marginals  $u_1^0$  and  $u_2^0$ , that is such that  $v^0 \in K(u_1^0, u_2^0)$ . We next consider the solution  $v(x, y, t)$  to the degenerate heat equation

$$\frac{\partial v}{\partial t} - \Delta_x v - \Delta_y v - 2\nabla_x \cdot \nabla_y v = 0, \quad x, y \in \mathbb{R}^d, t \geq 0 \quad (3)$$

starting from  $v^0$ . Observe that (3) may be rewritten as

$$\frac{\partial v}{\partial t} - (\nabla_x + \nabla_y) \cdot (\nabla_x + \nabla_y)v = 0, \quad x, y \in \mathbb{R}^d, t \geq 0. \quad (4)$$

Clearly, it holds that  $v(x, y, t) \geq 0$ , because of the non-negativity of the operator in (3), which can be written in the variables  $(x + y, x - y)$  as  $-\Delta_{x+y}$ .

We then define the marginals

$$v_1(x, t) = \int v(x, y, t) dy, \quad v_2(y, t) = \int v(x, y, t) dx$$

and show that  $v_1 = u_1$  and  $v_2 = u_2$ : for instance, integrating (3) with respect to  $y$ , one finds

$$\begin{cases} \frac{\partial v_1(x, t)}{\partial t} - \Delta_x v_1(x, t) = 0, & x \in \mathbb{R}^d, t \geq 0, \\ v_1(x) = u_1^0(x), & x \in \mathbb{R}^d \end{cases}$$

and uniqueness of the solution of the heat equation gives us  $v_1 = u_1$ .

Recalling (1), we conclude that

$$\mathcal{T}_\varrho(u_1(t), u_2(t)) \leq \iint \varrho(x, y) v(x, y, t) dx dy = \iint r(|x - y|) v(x, y, t) dx dy.$$

Finally, we may also compute, using (4) and integrating by parts,

$$\frac{d}{dt} \iint r(|x - y|) v(x, y, t) dx dy = \iint v(x, y, t) \left( (\nabla_x + \nabla_y) \cdot (\nabla_x + \nabla_y) [r(|x - y|)] \right) dx dy = 0.$$

Therefore, for any initial data  $v^0 \in K(u_1^0, u_2^0)$ ,

$$\mathcal{T}_\varrho(u_1(t), u_2(t)) \leq \iint r(|x-y|)v^0(x, y)dxdy = \iint \varrho(x, y)v^0(x, y)dxdy$$

and minimizing among such  $v^0$  completes the proof.  $\square$

The only technical question is to justify the integration by parts, which is immediate if we assume enough moments initially (otherwise there is nothing to prove), at least when we restrict ourselves to power cost functions  $\varrho_p(x, y) = |x-y|^p/p$  with  $p \geq 2$ . Notice that the well-posedness for (3) follows from the observation that we actually deal with  $-\Delta_{x+y}$ . It is also possible, under some conditions, to treat the case of some non smooth cost functions, e.g.  $\varrho_p$  for some  $p \in [1, 2)$ : this issue is discussed in Section 2.

Remark that the solution to (3) can be represented as  $v(x, y, t) = \mathbb{E}[v^0(x + \sqrt{2}B_t, y + \sqrt{2}B_t)]$ , for  $(B_t)_{t \geq 0}$  a Brownian motion. Note also that in (3), diffusion occurs only in the  $x+y$  variable, while  $x-y$  remains invariant under the dynamics.

Another way to understand (3) is to see that  $v(t)$  is the law of the couple  $(X_0 + \sqrt{2}B_t, Y_0 + \sqrt{2}B_t)$ , where  $(X_0, Y_0)$  is  $v^0$ -distributed and where we use the *same* Brownian motion  $(B_t)_{t \geq 0}$  for both coordinates. If using e.g. independent Brownian motions  $(B_t)_{t \geq 0}$  and  $(C_t)_{t \geq 0}$ , i.e. if considering the coupling  $(X_0 + \sqrt{2}B_t, Y_0 + \sqrt{2}C_t)$ , one would find  $v$  with the good marginals but solving  $\partial_t v - (\Delta_x + \Delta_y)v = 0$ , leading for example to the far from optimal estimate  $\mathcal{T}_2(u_1(t), u_2(t)) \leq \mathcal{T}_2(u_1^0, u_2^0) + 2dt$ .

## 1.2 Fokker-Planck equation

The coupling method can be extended to the Fokker-Planck equation, see [7] for some more elaborate consequences. The result can be stated as follows

**Theorem 2** *Consider some function  $V : \mathbb{R}^d \times \mathbb{R}_+ \mapsto \mathbb{R}^d$  such that, for some  $\alpha \in \mathbb{R}$ ,*

$$(V(x, t) - V(y, t)) \cdot (x - y) \leq \alpha|x-y|^2, \quad x, y \in \mathbb{R}^d, t \geq 0. \quad (5)$$

*Consider two probability densities  $u_1^0, u_2^0$  on  $\mathbb{R}^d$  and the corresponding solutions  $u_1, u_2$  to the Fokker-Planck equation*

$$\partial_t u - \Delta u + \operatorname{div}(V(x, t)u) = 0, \quad x \in \mathbb{R}^d, t \geq 0. \quad (6)$$

*For any  $t \geq 0$ , any  $p \geq 1$ , one has*

$$\mathcal{T}_p(u_1(t), u_2(t)) \leq \mathcal{T}_p(u_1^0, u_2^0) \exp(\alpha pt).$$

This inequality is well-known, see for example [42] §9.1.5 and the references therein. One can also find relations to several deep and recent functional analysis tools. This goes far beyond our present purpose.

**Proof.** This is the same proof as for the heat equation, with longer expressions. We consider any  $v^0$ , with marginals  $u_1^0$  and  $u_2^0$ , and the solution  $v$  to the equation

$$\partial_t v - \Delta_x v - \Delta_y v - 2\nabla_x \cdot \nabla_y v + \operatorname{div}_x(V(x, t)v) + \operatorname{div}_y(V(y, t)v) = 0, \quad x, y \in \mathbb{R}^d, t \geq 0 \quad (7)$$

starting from  $v^0$ . One easily checks that  $v(x, y, t) \geq 0$  and, integrating (7) with respect to  $y$ , that  $v_1(x, t) := \int v(x, y, t) dy$  solves (6) and starts from  $u_1^0$ , whence  $v_1 = u_1$ . The second marginal is treated similarly, and we conclude that  $\mathcal{T}_p(u_1(t), u_2(t)) \leq p^{-1} \iint |x - y|^p v(x, y, t) dx dy$ . Finally, using the same computation as for the heat equation, with some additional terms, we see that

$$\begin{aligned} \frac{d}{dt} \iint \frac{|x - y|^p}{p} v(x, y, t) dx dy &= 0 + \iint v(x, y, t) |x - y|^{p-2} (x - y) \cdot (V(x, t) - V(y, t)) dx dy \\ &\leq \alpha \iint |x - y|^p v(x, y, t) dx dy \end{aligned}$$

by assumption (5). The result follows using the Gronwall lemma

$$\mathcal{T}_p(u_1(t), u_2(t)) \leq \iint \frac{|x - y|^p}{p} v(x, y, t) dx dy \leq \left( \iint \frac{|x - y|^p}{p} v^0(x, y) dx dy \right) e^{\alpha p t}$$

and minimizing in  $v^0$ . □

Observe that one can treat in a similar way the case with variable diffusion coefficients, of the form  $\partial_t u - \sum_{i,j=1}^d \partial_{ij}(a_{ij}(x, t)u) + \operatorname{div}(V(x, t)u) = 0$ , under suitable conditions on the nonnegative symmetric matrix  $a$  and on  $V$ . For example, we will find that  $\mathcal{T}_2(u_1(t), u_2(t)) \leq \mathcal{T}_2(u_1^0, u_2^0) \exp(\alpha t)$  as soon as

$$\operatorname{Tr}\left([\sigma(x, s) - \sigma(y, s)][\sigma(x, s) - \sigma(y, s)]^*\right) + (x - y) \cdot (V(x, t) - V(y, t)) \leq \alpha |x - y|^2,$$

where  $\sigma(x, t)$  is, for each  $(x, t) \in \mathbb{R}^d \times [0, \infty)$ , a matrix such that  $\sigma(x, t)[\sigma(x, t)]^* = a(x, t)$ .

To our knowledge, a divergence form equation  $\partial_t u - \sum_{i,j=1}^d \partial_i(a_{ij}(x, t)\partial_j u) = 0$  does not enjoy particular properties from this point of view.

### 1.3 A nonlinear transport equation

We next consider a fully deterministic problem which arises in several types of modeling, such as polymers, cell division, neuron networks, etc :

$$\partial_t u + \operatorname{div}[V(x, I(t))u] = 0, \quad x \in \mathbb{R}^d, t \geq 0, \quad (8)$$

where the nonlinearity stems from the quantity  $I(t)$  defined, with a given weight  $\psi : \mathbb{R}^d \mapsto \mathbb{R}$ ,

$$I(t) = \int_{\mathbb{R}^d} \psi(x) u(x, t) dx. \quad (9)$$

We again complement this equation with an initial condition  $u^0 \geq 0$  with mass  $\int u^0 = 1$ .

**Theorem 3** *Assume that  $V : \mathbb{R}^d \times \mathbb{R} \mapsto \mathbb{R}^d$  and  $\psi : \mathbb{R}^d \mapsto \mathbb{R}$  are of class  $C^1$ , and that for some  $\alpha > 0$*

$$(x - y) \cdot (V(x, I) - V(y, I)) \leq -\alpha |x - y|^2, \quad \forall x, y \in \mathbb{R}^d, I \geq 0. \quad (10)$$

*Setting  $\langle x(t) \rangle = \int x u(x, t) dx$ , we have*

$$\mathcal{T}_2(u(t), \delta_{\langle x(t) \rangle}) = \int \frac{|x - \langle x(t) \rangle|^2}{2} u(x, t) dx \leq e^{-2\alpha t} \int \frac{|x - \langle x(0) \rangle|^2}{2} u^0(x, t) dx = e^{-2\alpha t} \mathcal{T}_2(u(0), \delta_{\langle x(0) \rangle}).$$

Assume additionally that

$$\beta = \|D_I V\|_\infty \|D\psi\|_\infty < \alpha$$

and fix any initial point  $X^0 \in \mathbb{R}^d$ . Consider the solution  $X$  to  $X'(t) = -V(X(t), \psi(X(t)))$  starting from  $X^0$ . For all  $t \geq 0$ , one has

$$\mathcal{T}_2(u(t), \delta_{X(t)}) = \int \frac{|x - X(t)|^2}{2} u(x, t) dx \leq e^{2(\beta - \alpha)t} \int \frac{|x - X^0|^2}{2} u^0(x) dx = e^{2(\beta - \alpha)t} \mathcal{T}_2(u^0, \delta_{X^0}).$$

It holds that  $(\delta_{X(t)})_{t \geq 0}$  solves (8) in a weak sense. A more general result, involving any pair of solutions, can be found in [42]. One could prove Theorem 3 without using a PDE for the coupling, using only the characteristics, in the spirit of Dobrushin [20].

**Proof.** We consider two solutions  $u_1$  and  $u_2$  to (8), and denote by  $I_1(t)$  and  $I_2(t)$  the corresponding functions, see (9). As we are interested in the case where one of the two solutions is a Dirac mass (for each  $t \geq 0$ ,  $u_2(t) = \delta_{X(t)}$ ), we can consider the trivial coupling  $v(x, y, t) = u_1(x, t)u_2(y, t)$ , which of course has the correct marginals, and satisfies

$$\partial_t v + \operatorname{div}_x [V(x, I_1(t))v] + \operatorname{div}_y [V(y, I_2(t))v] = 0.$$

Therefore, we may compute

$$\begin{aligned} \frac{d}{dt} \iint \frac{|x - y|^2}{2} u_1(x, t)u_2(y, t) dx dy &= \iint (x - y) \cdot (V(x, I_1(t)) - V(y, I_2(t))) u_1(x, t)u_2(y, t) dx dy \\ &= \iint (x - y) \cdot (V(x, I_1(t)) - V(y, I_1(t))) u_1(x, t)u_2(y, t) dx dy \\ &\quad + \iint (x - y) \cdot (V(y, I_1(t)) - V(y, I_2(t))) u_1(x, t)u_2(y, t) dx dy \\ &\leq -\alpha \iint |x - y|^2 u_1(x, t)u_2(y, t) dx dy \\ &\quad + \|D_I V\|_\infty |I_1(t) - I_2(t)| \left( \iint |x - y|^2 u_1(x, t)u_2(y, t) dx dy \right)^{1/2}. \end{aligned}$$

We first apply this in the case of single solution  $u := u_1 = u_2$ , whence  $I_1 = I_2$ , and we directly conclude by Gronwall's lemma that

$$\iint \frac{|x - y|^2}{2} u(x, t)u(y, t) dx dy \leq e^{-2\alpha t} \iint \frac{|x - y|^2}{2} u^0(x, t)u^0(y, t) dx dy.$$

This classically rewrites as

$$\int |x - \langle x(t) \rangle|^2 u(x, t) dx \leq e^{-2\alpha t} \int |x - \langle x(0) \rangle|^2 u^0(x, t) dx.$$

as desired. Next, when considering two solutions, we notice that

$$I_1(t) - I_2(t) = \iint [\psi(x) - \psi(y)] u_1(x, t)u_2(y, t) dx dy,$$

whence

$$|I_1(t) - I_2(t)| \leq \|D\psi\|_\infty \iint |x - y| u_1(x, t)u_2(y, t) dx dy \leq \|D\psi\|_\infty \left( \iint |x - y|^2 u_1(x, t)u_2(y, t) dx dy \right)^{1/2}.$$

Therefore,

$$\frac{d}{dt} \iint \frac{|x-y|^2}{2} u_1(x,t) u_2(y,t) dx dy \leq (\beta - \alpha) \iint |x-y|^2 u_1(x,t) u_2(y,t) dx dy.$$

Applying this to the case where  $u_2(t) = \delta_{X(t)}$  concludes the proof.  $\square$

For consistency with the other presentations in this section, we have written this result for an  $L^1$  density  $v$  with a finite second moment, but the extension to a probability measure is immediate.

#### 1.4 Agregation type equation

Among the nonlinear transport equations let us mention the agregation equation

$$\partial_t u - \operatorname{div}[u(K * u)] = 0, \quad x \in \mathbb{R}^d, t \geq 0, \quad (11)$$

where the convolution term is defined by  $K * u(x, t) = \int_{\mathbb{R}^d} K(x-y)u(y, t)dy$ . It is well-known, under various assumptions on  $K$  and for several variants of the equation, that it is contractive for transport costs and we refer to the papers [34, 14, 15, 30, 24].

Here we assume that the smooth kernel satisfies, for some  $\alpha \geq 0$ ,

$$K \in C^1(\mathbb{R}^d; \mathbb{R}^d), \quad (x-y) \cdot (K(x) - K(y)) \geq -\alpha|x-y|^2 \quad K(x) = -K(-x). \quad (12)$$

We also assume that the  $L^1(\mathbb{R}^d)$ -initial conditions  $u_i^0 \geq 0$  with unit mass  $\int u_i^0 = 1$ , for  $i = 1, 2$  satisfy

$$\int_{\mathbb{R}^d} |x|^2 u_i^0 dx < \infty, \quad \int_{\mathbb{R}^d} x u_i^0(x) dx = 0. \quad (13)$$

These imply that  $\int_{\mathbb{R}^d} x u_i(x, t) dx = 0$  for all  $t > 0$ .

**Theorem 4** *Assume (12). Consider two initial data  $u_1^0$  and  $u_2^0$  satisfying (13) and  $u_1, u_2$  the corresponding solutions to (11). One has for all  $t \geq 0$ ,*

$$\mathcal{T}_2(u_1(t), u_2(t)) \leq e^{2\alpha t} \mathcal{T}_2(u_1^0, u_2^0).$$

Notice that, combining the coupling below with that in Section 1.1, it is immediate to also treat the case with diffusion

$$\partial_t u - \Delta u - \operatorname{div}[u K * u] = 0,$$

**Proof.** Again the simple coupling

$$\partial_t v(x, y, t) - \operatorname{div}_x[v K * u_1] - \operatorname{div}_y[v K * u_2] = 0, \quad x, y \in \mathbb{R}^d, t \geq 0 \quad (14)$$

is enough. Indeed, we can verify immediately the properties that it preserves nonnegativity and the

marginales satisfy the correct equations. Finally, one can compute

$$\begin{aligned}
& \frac{d}{dt} \iint \frac{|x-y|^2}{2} v(x,y,t) dx dy \\
&= - \iint (x-y) \cdot \left[ \int K(x-x') u_1(x') dx' - \int K(y-y') u_2(y') dy' \right] v(x,y,t) dx dy \\
&= - \iiint (x-y) \cdot [K(x-x') - K(y-y')] v(x',y',t) v(x,y,t) dx' dy' dx dy \\
&= -\frac{1}{2} \iiint (x-x'-y+y') \cdot [K(x-x') - K(y-y')] v(x',y',t) v(x,y,t) dx' dy' dx dy \\
&\leq \frac{\alpha}{2} \iiint |x-x'-y+y'|^2 v(x',y',t) v(x,y,t) dx' dy' dx dy \\
&= \alpha \iint |x-y|^2 v(x,y,t) dx dy.
\end{aligned}$$

We have used a symmetry argument and that  $\iint (x-y)v(x,y,t) dx dy = 0$ . We complete the proof as in Section 1.1.  $\square$

## 2 Heat equation with variable coefficients

We consider the heat equation with variable coefficient. This is much more intricate than the previous examples. In 1 dimension, we use the  $\mathcal{T}_1$  distance and recover a result implicitly included in [29]. In higher dimension, we indicate a general way to construct cost functions. This leads to a poorly explored degenerate elliptic PDE, see however [36] and the references therein.

### 2.1 One-dimensional case

We consider some  $a : \mathbb{R} \mapsto \mathbb{R}_+$  and the following heat equation.

$$\frac{\partial u}{\partial t} - \frac{\partial^2}{\partial x^2} [a(x)u] = 0, \quad x \in \mathbb{R}, t \geq 0. \tag{15}$$

**Theorem 5** *Assume that  $d = 1$  and that  $a = \sigma^2$  for some  $\sigma \in C^{1/2}(\mathbb{R})$ . Consider two probability densities  $u_1^0, u_2^0$  on  $\mathbb{R}$  and the corresponding solutions  $u_1, u_2$  to (15). For all  $t \geq 0$ , one has*

$$\mathcal{T}_1(u_1(t), u_2(t)) \leq \mathcal{T}_1(u_1^0, u_2^0).$$

**Proof.** We give a proof for  $\sigma \in C^\alpha(\mathbb{R})$ , with  $\alpha > 1/2$ , the remark below explains how to treat  $\alpha = 1/2$ . We consider any probability density  $v^0(x,y)$  with marginals  $u_1^0$  and  $u_2^0$  and consider the coupling equation

$$\partial_t v - \partial_{xx}(\sigma^2(x)v) - \partial_{yy}(\sigma^2(y)v) - 2\partial_x \partial_y [\sigma(x)\sigma(y)v] = 0, \quad x, y \in \mathbb{R}, t \geq 0 \tag{16}$$

starting from  $v^0$ . This equation preserves non-negativity. A simple way to see this is the following computation: multiplying (16) by  $-v_-$ , integrating on  $\mathbb{R}^2$  and using some integrations by parts, one

can check that

$$\begin{aligned}
& \frac{1}{2} \frac{d}{dt} \iint v_-^2(x, y, t) dx dy \\
&= - \iint \left[ |\sigma(x) \partial_x v_-(x, y, t)|^2 - 2\sigma(x)\sigma(y) \partial_x v_-(x, y, t) \partial_y v_-(x, y, t) + |\sigma(y) \partial_y v_-(x, y, t)|^2 \right] dx dy \\
& \quad + \frac{1}{2} \iint v_-^2(x, y, t) [\partial_{xx}(\sigma^2(x)) + \partial_{yy}(\sigma^2(y)) - 2\partial_x \sigma(x) \partial_y \sigma(y)] dx dy \\
&\leq \frac{1}{2} \iint v_-^2(x, y, t) [\partial_{xx}(\sigma^2(x)) + \partial_{yy}(\sigma^2(y)) - 2\partial_x \sigma(x) \partial_y \sigma(y)] dx dy.
\end{aligned}$$

Since  $\iint v_-^2(x, y, 0) dx dy = 0$ , the result follows from the Gronwall lemma if  $\sigma$  is smooth. Otherwise, one can work by approximation.

Integrating (16) with respect to  $y$ , we see that  $v_1(x, t) := \int v(x, y, t) dy$  solves (15) and starts from  $u_1^0$ , whence  $v_1 = u_1$ . The second marginal is treated similarly, and we conclude that  $\mathcal{T}_1(u_1(t), u_2(t)) \leq \iint |x - y| v(x, y, t) dx dy$ . Because of its singularity, we need to regularize the absolute value as a  $W^{2, \infty}$  function and define

$$\omega_\varepsilon(r) = \begin{cases} \frac{r^2}{2\varepsilon} & \text{for } r \leq \varepsilon, \\ r - \frac{\varepsilon}{2} & \text{for } r \geq \varepsilon. \end{cases}$$

Using the Hölder constant  $C_\sigma$  of  $\sigma(\cdot)$ , we see that

$$\begin{aligned}
\frac{d}{dt} \iint \omega_\varepsilon(|x - y|) v(x, y, t) dx dy &= \iint v(x, y, t) \omega_\varepsilon''(|x - y|) [\sigma(x) - \sigma(y)]^2 dx dy \\
&\leq C_\sigma^2 \iint v(x, y, t) \frac{\mathbf{1}_{\{|x - y| \leq \varepsilon\}}}{\varepsilon} |x - y|^{2\alpha} dx dy \\
&\leq C_\sigma^2 \varepsilon^{2\alpha - 1},
\end{aligned}$$

because  $v(t)$  is a probability measure. Since now  $2\alpha - 1 > 0$ , we may let  $\varepsilon \rightarrow 0$  and we find that

$$\mathcal{T}_1(u_1(t), u_2(t)) \leq \iint |x - y| v(x, y, t) dx dy \leq \iint |x - y| v^0(x, y) dx dy.$$

We conclude, as usual, by minimizing in  $v^0$ . □

**Remark 6** *The condition  $\sigma \in C^{1/2}(\mathbb{R})$  is enough. To treat this exponent, a better construction of the regularization is required, using the so-called Yamada function:*

$$\omega_\varepsilon(r) = 0 \text{ for } r \leq \varepsilon^{3/2}, \quad \omega_\varepsilon''(r) = \frac{2}{r |\ln(\varepsilon)|} \text{ for } \varepsilon^{3/2} \leq r \leq \varepsilon, \quad \omega_\varepsilon'(r) = 1 \text{ for } r \geq \varepsilon.$$

There are other technical issues here. For example, the well-posedness of (16), which is necessary to identify the marginals of the solution  $v$  to the coupling equation, is not so easy. A possible direction is to use results established in [21], in the spirit of [19].

## 2.2 A general construction of the weight

In order to unravel the algebraic structure behind the choice of the weight  $\varrho$ , we now consider the general case of dimension  $d$ . We assume that  $a : \mathbb{R}^d \mapsto \mathcal{M}_{d \times d}(\mathbb{R})$  is everywhere symmetric and nonnegative, of the form

$$a_{ij}(x) = \sum_{k=1}^K \sigma_{ik}(x) \sigma_{jk}(x), \quad (17)$$

for some  $\sigma : \mathbb{R}^d \mapsto \mathcal{M}_{d \times K}(\mathbb{R})$ , and we consider the heat equation

$$\frac{\partial u}{\partial t} - \sum_{i,j=1}^d \frac{\partial^2}{\partial x_i \partial x_j} [a_{ij}(x)u] = 0, \quad x \in \mathbb{R}^d, t \geq 0, \quad (18)$$

completed with an initial probability density  $u^0$  on  $\mathbb{R}^d$ .

**Proposition 7** *Assume that  $\sigma$  is regular enough and consider two probability densities  $u_1^0, u_2^0$  on  $\mathbb{R}^d$  and the corresponding solutions  $u_1, u_2$  to (18). For all  $t \geq 0$ , one has*

$$\mathcal{T}_\varrho(u_1(t), u_2(t)) \leq \mathcal{T}_\varrho(u_1^0, u_2^0),$$

for any smooth cost  $\varrho : \mathbb{R}^d \mapsto \mathbb{R}_+$  satisfying

$$\sum_{i,j=1}^d a_{ij}(x) \frac{\partial^2 \varrho(x, y)}{\partial x_i \partial x_j} + \sum_{i,j=1}^d a_{ij}(y) \frac{\partial^2 \varrho(x, y)}{\partial y_i \partial y_j} + 2 \sum_{i,j=1}^d \sum_{k=1}^K \sigma_{ik}(x) \sigma_{jk}(y) \frac{\partial^2 \varrho(x, y)}{\partial x_i \partial y_j} \leq 0, \quad x, y \in \mathbb{R}^d. \quad (19)$$

When  $a$  is constant, we recover that any  $C^2$  cost function of the form  $\varrho(x, y) = r(|x - y|)$  works. In dimension 1,  $\varrho(x, y) = |x - y|$  is indeed a (weak) solution to (19). We do not know of a theory to solve (19), in dimension  $d \geq 2$ , for a general coefficient  $a$ , so that we do not know if this result is useful. Notice that equation (19) should be completed by the boundary value  $\varrho(x, x) = 0$  with some growth condition to mimic  $|x - y|^p$ .

**Proof.** We consider any probability density  $v^0(x, y)$  with marginals  $u_1^0$  and  $u_2^0$  and consider the coupling equation

$$\frac{\partial v}{\partial t} - \sum_{i,j=1}^d \frac{\partial^2}{\partial x_i \partial x_j} [a_{ij}(x)v] - \sum_{i,j=1}^d \frac{\partial^2}{\partial y_i \partial y_j} [a_{ij}(y)v] - 2 \sum_{i,j=1}^d \sum_{k=1}^K \frac{\partial^2}{\partial x_i \partial y_j} [\sigma_{ik}(x) \sigma_{jk}(y)v] = 0$$

starting from  $v^0$ . We show as usual that  $\int v(x, y, t) dy = u_1(x, t)$  and that  $\int v(x, y, t) dx = u_2(y, t)$ . Moreover, we have  $v(x, y, t) \geq 0$ : we multiply the coupling equation by  $-v_-$  and integrate, finding

$$\frac{1}{2} \frac{d}{dt} \iint v_-^2(x, y, t) dx dy = - \sum_{k=1}^K I_k - J,$$

with

$$I_k = \iint \sum_{i,j=1}^d \left[ \sigma_{ik}(x) \frac{\partial v_-(x, y, t)}{\partial x_i} \sigma_{jk}(x) \frac{\partial v_-(x, y, t)}{\partial x_j} + \sigma_{ik}(y) \frac{\partial v_-(x, y, t)}{\partial y_i} \sigma_{jk}(y) \frac{\partial v_-(x, y, t)}{\partial y_j} + 2 \sigma_{ik}(x) \sigma_{jk}(y) \frac{\partial v_-(x, y, t)}{\partial x_i} \frac{\partial v_-(x, y, t)}{\partial y_j} \right] dx dy$$

which can also be written

$$I_k = \iint \left| \sum_{i=1}^d \sigma_{ik}(x) \frac{\partial v_-(x, y, t)}{\partial x_i} + \sum_{i=1}^d \sigma_{ik}(y) \frac{\partial v_-(x, y, t)}{\partial y_i} \right|^2 dx dy \geq 0.$$

The other term is

$$J = \iint \sum_{i,j=1}^d \left[ \frac{\partial v_-(x, y, t)}{\partial x_i} \frac{\partial a_{ij}(x)}{\partial x_j} v_-(x, y, t) + \frac{\partial v_-(x, y, t)}{\partial y_i} \frac{\partial a_{ij}(y)}{\partial y_j} v_-(x, y, t) + 2 \frac{\partial v_-(x, y, t)}{\partial x_i} v_-(x, y, t) \frac{\partial}{\partial y_j} \sum_{k=1}^K \sigma_{ik}(x) \sigma_{jk}(y) \right] dx dy,$$

which can also be written after integration by parts

$$J = -\frac{1}{2} \iint (v_-(x, y, t))^2 \sum_{i,j=1}^d \left[ \frac{\partial^2 a_{ij}(x)}{\partial x_i \partial x_j} + \frac{\partial^2 a_{ij}(y)}{\partial y_i \partial y_j} + 2 \frac{\partial^2}{\partial x_i \partial y_j} \sum_{k=1}^K \sigma_{ik}(x) \sigma_{jk}(y) \right] dx dy.$$

Assuming that the entries  $\sigma_{ik}$  are bounded with two bounded derivatives, we conclude by Gronwall's lemma that  $v_- \equiv 0$ , since we initially have  $\iint v_-^2(x, y, 0) dx dy = 0$ .

Recalling (1), we conclude that  $\mathcal{T}_\varrho(u_1(t), u_2(t)) \leq \iint \varrho(x, y) v(x, y, t) dx dy$ . Since finally

$$\begin{aligned} & \frac{d}{dt} \iint \varrho(x, y) v(x, y, t) dx dy \\ &= \iint v(x, y, t) \sum_{i,j=1}^d \left[ a_{ij}(x) \frac{\partial^2 \varrho(x, y)}{\partial x_i \partial x_j} + a_{ij}(y) \frac{\partial^2 \varrho(x, y)}{\partial y_i \partial y_j} + 2 \sum_{k=1}^K \sigma_{ik}(x) \sigma_{jk}(y) \frac{\partial^2 \varrho(x, y)}{\partial x_i \partial y_j} \right] dx dy \leq 0 \end{aligned}$$

by assumption, we conclude as usual.  $\square$

We leave open the question to formalize this approach rigorously, in particular for degenerate coefficients  $\sigma$ , and to build other examples where one can prove the existence of a weight  $\varrho$ .

### 3 Scattering and integral kernels

We now turn to equations that describe the probability law of various jump processes. These are well-known results except the case of kinetic scattering in Subsection 3.2 which seems to be new.

#### 3.1 Simple scattering

For  $x \in \mathbb{R}^d$ , we parameterize the pre-jump location  $X = \Phi(x, h)$  by  $h \in \mathbb{R}^d$ , distributed according to a bounded measure  $\mu$ . We assume that for all fixed  $h \in \mathbb{R}^d$ ,

$$x \mapsto X = \Phi(x, h) \text{ is invertible on } \mathbb{R}^d \text{ and } D_x \Phi(x, h) \text{ is an invertible matrix,} \quad (20)$$

and we use the notation  $X \mapsto x = \Phi^{-1}(X, h)$  for the inverse in  $x$  (with  $h$  fixed).

We consider the scattering problem

$$\partial_t u(x, t) = \int \left[ u(\Phi(x, h), t) \det(D_x \Phi(x, h)) - u(x, t) \right] d\mu(h), \quad (21)$$

with initial condition  $u^0$ , a probability density on  $\mathbb{R}^d$ . Actually, this equation is to be understood in the weak sense: integrating the right hand side against a test function  $\varphi(x)$ , we see that

$$\iint \varphi(x) \left[ u(\Phi(x, h), t) \det(D_x \Phi(x, h)) - u(x, t) \right] d\mu(h) = \iint u(X, t) [\varphi(\Phi^{-1}(X, h)) - \varphi(X)] dX d\mu(h),$$

which shows that the determinant  $\det(D_x \Phi(x, h))$  is only used informally. We briefly prove the following result, which is classical, see for instance [1].

**Theorem 8** *Assume (20), fix  $p \in [1, \infty)$  and suppose there is  $\delta \in \mathbb{R}$  such that for all  $X, Y \in \mathbb{R}^d$ ,*

$$\int |\Phi^{-1}(X, h) - \Phi^{-1}(Y, h)|^p d\mu(h) \leq KL|X - Y|^p, \quad \text{where } K = \mu(\mathbb{R}^d). \quad (22)$$

*Consider two probability densities  $u_1^0, u_2^0$  on  $\mathbb{R}^d$  and the corresponding solutions  $u_1, u_2$  to (21). For all  $t \geq 0$ , one has*

$$\mathcal{T}_p(u_1(t), u_2(t)) \leq e^{K(L-1)t} \mathcal{T}_p(u_1^0, u_2^0),$$

The homogeneous scattering corresponds to  $\Phi(x, h) = x + h$  and obviously fulfills the above assumptions.

**Proof.** For a probability density  $v^0$  on  $\mathbb{R}^d \times \mathbb{R}^d$  with marginals  $u_1^0$  and  $u_2^0$ , we consider the solution  $v$  of the coupled equation built in such a way that the jumps parameter  $h$  is common to the two variables. Namely, we choose

$$\partial_t v(x, y, t) = \int [v(\Phi(x, h), \Phi(y, h), t) \det(D_x \Phi(x, h)) \det(D_y \Phi(y, h)) - v(x, y, t)] d\mu(h), \quad (23)$$

starting from  $v^0$ . We clearly have  $v \geq 0$ , and integrating in  $y$  and using the change of variable  $y \mapsto \Phi(y, h)$ , we find that  $v_1(x, t) = \int v(x, y, t) dy$  satisfies (21). Since it starts from  $u_1^0$ , we conclude that  $v_1 = u_1$ . The second marginal is treated similarly, and we conclude as usual that  $\mathcal{T}_p(u_1(t), u_2(t)) \leq p^{-1} \iint |x - y|^p v(x, y, t) dx dy$ . Next, we compute, using (23):

$$\begin{aligned} \frac{d}{dt} \iint |x - y|^p v(x, y, t) dx dy &+ K \iint |x - y|^p v(x, y, t) dx dy \\ &= \iiint |x - y|^p v(\Phi(x, h), \Phi(y, h), t) \det(D_x \Phi(x, h)) \det(D_y \Phi(y, h)) d\mu(h) dx dy \\ &= \iiint |\Phi^{-1}(X, h) - \Phi^{-1}(Y, h)|^p v(X, Y, t) d\mu(h) dX dY. \end{aligned}$$

We used the changes of variables  $X = \Phi(x, h)$  and  $Y = \Phi(y, h)$  (with  $h$  fixed). Recalling (22), we conclude that

$$\frac{d}{dt} \iint |x - y|^p v(x, y, t) dx dy \leq K(L - 1) \iint |x - y|^p v(x, y, t) dx dy.$$

Using the Gronwall lemma, we thus find that

$$\mathcal{T}_p(u_1(t), u_2(t)) \leq p^{-1} \iint |x - y|^p v(x, y, t) dx dy \leq p^{-1} e^{K(L-1)t} \iint |x - y|^p v^0(x, y) dx dy$$

and we conclude as usual, minimizing in  $v^0$ . □

The most general scattering equation reads

$$\partial_t u(x, t) = \int [\pi(x, x_*)u(x_*) - \pi(x_*, x)u(x, t)] dx_*, \quad (24)$$

and equation (21) corresponds to the homogeneous cases when  $\int \pi(x_*, x) dx_* = 1$ , and the above method can easily be adapted. For the inhomogeneous case, see Section 4.

### 3.2 Kinetic scattering

We next consider some kinetic scattering models, that means we work in the phase space. We consider some finite measure  $\mu$  on  $\mathbb{R}^d$ , some application  $V : \mathbb{R}^d \mapsto \mathbb{R}^d$  such that, for all  $h \in \mathbb{R}^d$ ,

$$v \mapsto V = \Phi(v, h) \text{ is invertible and } D_v \Phi(v, h) \text{ is an invertible matrix,} \quad (25)$$

and the kinetic scattering equation

$$\partial_t f(x, v, t) + v \cdot \nabla_x f = \int [f(x, \Phi(v, h), t) \det(D_v \Phi(v, h)) - f(x, v, t)] d\mu(h) \quad (26)$$

completed with an initial data  $f^0(x, v) \geq 0$  with  $\int f^0 dx dv = 1$ .

**Theorem 9** *Assume (25). Set  $K = \mu(\mathbb{R}^d)$  and suppose that for some  $L \in \mathbb{R}_+$ , for all  $v, w \in \mathbb{R}^d$ ,*

$$\int |\Phi^{-1}(V, h) - \Phi^{-1}(W, h)| d\mu(h) \leq KL|V - W|. \quad (27)$$

*Suppose that  $K \geq KL + 1$ . Consider two probability densities  $f_1^0, f_2^0$  on  $\mathbb{R}^d \times \mathbb{R}^d$  and the corresponding solutions  $f_1, f_2$  to (21). It holds that for all  $t \geq 0$ , (here  $\mathcal{T}_1$  is associated to the cost function  $\varrho((x, v), (y, w)) = |x - y| + |v - w|$ )*

$$\mathcal{T}_1(f_1(t), f_2(t)) \leq \mathcal{T}_1(f_1^0, f_2^0).$$

**Proof.** As usual, we consider any probability density  $F^0((x, v), (y, w))$  on  $(\mathbb{R}^d \times \mathbb{R}^d)^2$  with marginals  $f_1^0$  and  $f_2^0$ , and we consider  $F((x, v), (y, w), t)$  starting from  $F^0$  and solving

$$\begin{aligned} & \partial_t F + v \cdot \nabla_x F + w \cdot \nabla_y F \\ &= \int \left[ F((x, \Phi(v, h)), (y, \Phi(w, h)), t) \det(D_v \Phi(v, h)) \det(D_w \Phi(w, h)) - F((x, v), (y, w), t) \right] d\mu(h). \end{aligned}$$

This function is clearly nonnegative and has the correct marginals. For example, with  $F_1(x, v, t) = \int F(x, y, v, w, t) dy dw$ , we see that

$$\partial_t F_1 + v \cdot \nabla_x F_1 = \int [F_1(x, \Phi(v, h), t) \det(D_v \Phi(v, h)) - F_1(x, v, t)] d\mu(h)$$

because  $\int F((x, \Phi(v, h)), (y, \Phi(w, h)), t) \det(D_w \Phi(w, h)) dy dw = F_1(x, \Phi(v, h), t)$ : use the substitution  $V = \Phi(w, h)$  (with  $h$  fixed). Since  $F_1(0) = f_1(0)$ , we conclude that  $F_1(t) = f_1(t)$ . Hence we conclude that  $\mathcal{T}_1(f_1(t), f_2(t)) \leq \iint (|x - y| + |v - w|) F((x, v), (y, w), t) dx dy dv dw$ .

Next, using the equation for  $F$ , we find with  $V = \Phi(v, h)$  and  $W = \Phi(w, h)$ ,

$$\begin{aligned}
& \frac{d}{dt} \iint (|x - y| + |v - w|) F(x, y, v, w, t) dx dy dv dw \\
= & \iint \frac{x - y}{|x - y|} \cdot (v - w) F(x, y, v, w, t) dx dy dv dw \\
& - K \iint (|x - y| + |v - w|) F(x, y, v, w, t) dx dy dv dw \\
& + \iiint (|x - y| + |\Phi^{-1}(V, h) - \Phi^{-1}(W, h)|) F((x, V), (y, W), t) dx dy dV dW d\mu(h) \\
\leq & (1 - K + KL) \iint |v - w| F(x, y, v, w, t) dx dy dv dw.
\end{aligned}$$

by (27). Since now  $K \geq 1 + KL$  by assumption, we deduce that

$$\mathcal{T}_1(f_1(t), f_2(t)) \leq \iint (|x - y| + |v - w|) F^0((x, v), (y, w)) dx dy dv dw$$

and complete the proof as usual, minimizing in  $F^0$ .  $\square$

**Remark 10** Fix  $a > 0$ . If using the transport cost with weight  $\varrho = a|x - y| + |v - w|$ , the condition  $K \geq 1 + KL$  is replaced by the condition  $K > a + KL$ .

### 3.3 Fractional heat equation with variable coefficients

Informally, the fractional Laplacian is a variant of the integral equation treated in Subsection 3.1. However there is a particular interest when the coefficients depend on space, an example we borrow from [28, 22]. Consider the parabolic equation with derivatives of order  $\alpha \in (0, 2)$

$$\begin{cases} \partial_t u(x, t) = \mathcal{L}_\alpha[u], & x \in \mathbb{R}, t \geq 0, \\ \mathcal{L}_\alpha^*[\varphi](x) := \int [\varphi(x + \sigma(x)h) - \varphi(x) - h\sigma(x)\varphi'(x)] \frac{dh}{|h|^{1+\alpha}}. \end{cases} \quad (28)$$

**Theorem 11** Assume that  $\alpha \in (1, 2)$  and that  $\sigma \in C^{1/\alpha}$  and consider two initial probability densities  $u_1^0$  and  $u_2^0$  on  $\mathbb{R}$  and the corresponding solutions  $u_1$  and  $u_2$  to (28). For all  $t \geq 0$ ,

$$\mathcal{T}_{\alpha-1}(u_1(t), u_2(t)) \leq \mathcal{T}_{\alpha-1}(u_1^0, u_2^0).$$

**Proof.** We consider an initial probability density  $v^0$  on  $\mathbb{R}^2$  with marginals  $u_1^0$  and  $u_2^0$  and the solution  $v$  to the problem (written in weak form): for all smooth  $\varphi : \mathbb{R}^2 \mapsto \mathbb{R}$ ,

$$\begin{aligned}
& \frac{d}{dt} \iint \varphi(x, y) v(x, y, t) dx dy dt \\
= & \iint v(x, y, t) \int \left( \varphi(x + \sigma(x)h, y + \sigma(y)h) - \varphi(x, y) - h[\sigma(x)\partial_x \varphi(x, y) + \sigma(y)\partial_y \varphi_y(x, y)] \right) \frac{dh}{|h|^{1+\alpha}} dx dy
\end{aligned} \quad (29)$$

starting from  $v^0$ . The solution is clearly nonnegative and one checks as usual that for each  $t \geq 0$ , the marginals of  $v(t)$  are  $u_1(t)$  and  $u_2(t)$ : for example, we apply the above formula with  $\varphi$  depending

only on  $x$  and deduce that  $v_1(x, t) = \int v(x, y, t) dy$  solves the weak form of (28), whence  $v_1 = u_1$  since  $v_1(0) = u_1(0)$ . Consequently, we have  $\mathcal{T}_{\alpha-1}(u_1(t), u_2(t)) \leq (\alpha - 1)^{-1} \iint |x - y|^{\alpha-1} v(x, y, t) dx dy$ , and, using the same arguments as usual, it suffices to show that

$$\frac{d}{dt} \iint |x - y|^{\alpha-1} v(x, y, t) dx dy dt \leq 0.$$

This follows from the fact that for all  $x, y \in \mathbb{R}$ , setting  $u = \frac{\sigma(x) - \sigma(y)}{x - y}$ ,

$$\begin{aligned} & \int_{\mathbb{R}} \left[ |x + \sigma(x)h - y - \sigma(y)h|^{\alpha-1} - |x - y|^{\alpha-1} - (\alpha - 1)h[\sigma(x) - \sigma(y)]|x - y|^{\alpha-3}(x - y) \right] \frac{dh}{|h|^{1+\alpha}} \\ &= |x - y|^{\alpha-1} \int_{\mathbb{R}} \left[ |1 + hu|^{\alpha-1} - 1 - (\alpha - 1)hu \right] \frac{dh}{|h|^{1+\alpha}} \\ &= |x - y|^{\alpha-1} |u|^\alpha \int_{\mathbb{R}} \left[ |1 + h|^{\alpha-1} - 1 - (\alpha - 1)h \right] \frac{dh}{|h|^{1+\alpha}} = 0. \end{aligned}$$

The proof of this last equality can be found in [22, Lemma 9-(ii)], case  $a_+ = a_-$  and  $\beta = \alpha - 1$ . Observe that

$$|x - y|^{\alpha-1} |u|^\alpha = \frac{|\sigma(x) - \sigma(y)|^\alpha}{|x - y|} \leq C_\sigma$$

so that (29) makes sense with  $\varphi(x, y) = |x - y|^\alpha$ , thanks to our regularity assumption on  $\sigma$ .  $\square$

Here again, as in Section 2.1, the main technical difficulty is to prove the well-posedness of (28), in particular when  $\sigma$  may degenerate. This is useful to check that the solution  $v$  to the coupled equation has the correct marginals.

## 4 Inhomogeneous integral equations

Our next purpose is to give an example on the way to take into account  $x$ -dependency in IPDE models, for instance when considering a measure  $\mu(x, h)$  in the scattering equation (21). We exemplify this issue with a simple equation we borrow from [23]. Consider an interval  $I$  of  $\mathbb{R}$ , a rate function  $d \geq 0$  defined on  $I$  and some probability density  $b$  on  $I$ . We consider the conservative equation

$$\partial_t u(x, t) + d(x)u = b(x)A(t), \quad A(t) = \int_I d(x)u(x, t) dx \quad (30)$$

starting from an initial probability density  $u^0$  on  $I$ . We notice at once that this equation makes sense for probability measures  $u(dx, t)$  (for each  $t \geq 0$ ,  $u(dx, t)$  is a probability measure on  $I$ ) in the following weak sense: for all smooth  $\varphi : I \mapsto \mathbb{R}$ ,

$$\frac{d}{dt} \int \varphi(x)u(dx, t) = \iint [\varphi(z) - \varphi(x)]b(z)d(x)u(dx, t)dz. \quad (31)$$

**Theorem 12** *Consider two probability densities  $u_1^0, u_2^0$  on  $\mathbb{R}^d$  and the corresponding solutions  $u_1, u_2$  to (31). Under one of the two conditions (a) or (b) below, for all  $t \geq 0$ ,*

$$\mathcal{T}_\varrho(u_1(t), u_2(t)) \leq \mathcal{T}_\varrho(u_1^0, u_2^0).$$

- (a)  $I = \mathbb{R}_+$ ,  $d(0) = 0$ ,  $d$  is increasing,  $b = \delta_0$ , and  $\varrho(x, y) = |d^p(x) - d^p(y)|$  for some  $p \geq 1$ .
- (b)  $I = \mathbb{R}_+$ ,  $d(x) = \alpha x^p + \beta$  for some  $\alpha, \beta \geq 0$  and  $p \geq 1$ , with  $\varrho(x, y) = |x^p - y^p|$ , under the condition that  $\beta \geq \alpha \int_0^\infty z^p b(z) dz$ .

Other assumptions on  $I$ ,  $b$ ,  $d$  are possible: it suffices that  $\varrho$ ,  $b$  and  $d$  satisfy the dual inequality (33) below, which corresponds to (19) for the heat equation with variable coefficients.

**Proof.** We consider some probability density  $v^0$  on  $I^2$  with marginals  $u_1^0$  and  $u_2^0$  and define the probability measure  $v(dx, dy, t)$  as solving, for all smooth  $\varphi : I^2 \mapsto \mathbb{R}$ ,

$$\begin{aligned} \frac{d}{dt} \iint \varphi(x, y) v(dx, dy, t) &= \iiint [\varphi(z, z) - \varphi(x, y)] b(z) \min(d(x), d(y)) v(dx, dy, t) dz \\ &+ \iiint [\varphi(z, y) - \varphi(x, y)] b(z) (d(x) - d(y))_+ v(dx, dy, t) dz \\ &+ \iiint [\varphi(x, z) - \varphi(x, y)] b(z) (d(y) - d(x))_+ v(dx, dy, t) dz. \end{aligned} \quad (32)$$

It holds true that  $v(t)$  is a probability measure on  $I^2$  for each  $t \geq 0$ , and that its marginals are  $u_1(t)$  and  $u_2(t)$ . For example, applying the coupling equation with  $\varphi$  depending only on  $x$  and using that

$$\min(d(x), d(y)) + (d(x) - d(y))_+ = d(x),$$

one verifies that  $\int_{y \in I} v(dx, dy, t)$  solves (31), whence  $\int_{y \in I} v(dx, dy, t) = u_1(dx, t)$  by uniqueness. Hence for any cost function  $\varrho : I^2 \mapsto \mathbb{R}_+$ , we have  $\mathcal{T}_\varrho(u_1(t), u_2(t)) \leq \iint \varrho(x, y) v(dx, dy, t)$ . Furthermore, we easily compute, using that  $\varrho(z, z) = 0$  for all  $z \in I$ , that  $b$  is a probability density, and that  $\min(r, s) + (r - s)_+ + (s - r)_+ = \max(r, s)$ ,

$$\begin{aligned} &\frac{d}{dt} \iint \varrho(x, y) v(dx, dy, t) + \iint \varrho(x, y) \max(d(x), d(y)) v(dx, dy, t) \\ &= \iiint \varrho(z, y) b(z) (d(x) - d(y))_+ v(dx, dy, t) dz + \iiint \varrho(x, z) b(z) (d(y) - d(x))_+ v(dx, dy, t) dz. \end{aligned}$$

Therefore, using the same arguments as usual, the result will follow from the fact that for all  $x, y \in I$ ,

$$\varrho(x, y) \max(d(x), d(y)) \geq \int [\varrho(z, y) b(z) (d(x) - d(y))_+ + \varrho(x, z) b(z) (d(y) - d(x))_+ dz. \quad (33)$$

(a) Assuming that  $I = \mathbb{R}_+$ , that  $d(0) = 0$ , that  $d$  is increasing, that  $\varrho(x, y) = |d(x) - d(y)|^p$  for some  $p \geq 1$  and that  $b = \delta_0$ , we check that, when e.g.  $x \geq y \geq 0$ ,

$$(d^p(x) - d^p(y))d(x) \geq d(y)^p(d(x) - d(y)),$$

which holds true since indeed, for any  $s \geq t \geq 0$ ,  $(s^p - t^p)s \geq t^p(s - t)$  because  $p \geq 1$ .

(b) Assume next that  $I = \mathbb{R}_+$ ,  $d(x) = \alpha x^p + \beta$  for some  $\alpha, \beta \geq 0$  and  $p \geq 1$  and that  $\varrho(x, y) = |x^p - y^p|$ . We have to verify that, for all  $x \geq y \geq 0$ ,

$$(x^p - y^p)(\alpha x^p + \beta) \geq (\alpha x^p - \alpha y^p) \int_0^\infty |z^p - y^p| b(z) dz.$$

Setting  $m = \int_0^\infty z^p b(z) dz$ , it suffices to check that  $\alpha x^p + \beta \geq \alpha(m + y^p)$ . This of course holds true if  $\beta \geq \alpha m$ .  $\square$

Observe that the strong equation corresponding to the weak form (32) is nothing but

$$\begin{aligned} \partial_t v + \max(d(x), d(y)) v &= b(x) \delta(x - y) \int \min(d(x'), d(y')) v(dx', dy', t) \\ &+ b(x) \int (d(x') - d(y))_+ v(dx', y, t) + b(y) \int (d(y') - d(x))_+ v(x, dy', t). \end{aligned}$$

## 5 Homogeneous Boltzmann equation

In his seminal paper [38], Tanaka observed, using a probabilistic approach based on nonlinear Poisson-driven stochastic differential equations, that the homogeneous Boltzmann equation for Maxwell molecules is non-expansive for the 2-transport cost. A deterministic proof was provided by Villani in [41, Section 7.5.6] and was extended to inelastic collisions in [5]. A survey of results concerning homogeneous kinetic equations can be found in [16]. The approach in [41, 5, 16] is to first derive a contractivity result for the gain operator by coupling, and then to conclude *via* Duhamel's principle. Here we follow Tanaka's original approach, or rather we show how to write down his main arguments without introducing stochastic processes.

Let us also mention that [37] managed to study the corresponding dissipation in order to quantify the convergence to equilibrium of the solutions and, even more interesting, to prove the convergence to equilibrium of the Kac particle system, with a rate of convergence not depending on the number of particles.

The homogeneous Boltzmann writes

$$\begin{cases} \partial_t f(v, t) = Q(f) := \int_{\mathbb{R}^3} \int_{\mathbb{S}^2} [f(v', t)f(v'_*, t) - f(v, t)f(v_*, t)]B(\theta)dv_*d\sigma, \\ v' = \frac{1}{2}(v + v_*) + \frac{1}{2}|v - v_*|\sigma, \quad v'_* = \frac{1}{2}(v + v_*) - \frac{1}{2}|v - v_*|\sigma, \\ \cos(\theta) = \frac{v - v_*}{|v - v_*|} \cdot \frac{v' - v'_*}{|v' - v'_*|}. \end{cases} \quad (34)$$

The collision kernel  $B : (0, \pi) \mapsto \mathbb{R}_+$  is assumed to satisfy  $\int_0^\pi B(\theta)d\theta = 1$ . As is well-known, this equation writes, in weak form, for all mapping  $\phi : \mathbb{R}^3 \mapsto \mathbb{R}$ ,

$$\frac{d}{dt} \int_{\mathbb{R}^3} \phi(v)f(v, t)dv = \int_{\mathbb{R}^3 \times \mathbb{R}^3} \int_{\mathbb{S}^2} [\phi(v') + \phi(v'_*) - \phi(v) - \phi(v_*)]B(\theta)f(v, t)f(v_*, t)dv dv_* d\sigma. \quad (35)$$

**Theorem 13** *Consider two initial probability densities  $f_1^0, f_2^0$  on  $\mathbb{R}^3$  with a finite moment of order 2 and the corresponding solutions  $f_1, f_2$  to (34). Then, for all  $t \geq 0$ , one has*

$$\mathcal{T}_2(f_1(t), f_2(t)) \leq \mathcal{T}_2(f_1^0, f_2^0).$$

**Proof.** We fix a probability density  $F^0$  on  $(\mathbb{R}^3)^2$  with marginals  $f_1^0$  and  $f_2^0$  and build a coupled equation with the same principle as for scattering, that is the jump parameters are taken in common to the two variables, in such a way that the post-collisional velocities are as close as possible. We consider the solution  $F(v, w, t)$ , starting from  $F^0$ , to the following coupling equation written in weak form: for all mapping  $\Psi : \mathbb{R}^3 \times \mathbb{R}^3 \mapsto \mathbb{R}$ ,

$$\begin{aligned} \frac{d}{dt} \iint_{(\mathbb{R}^3)^2} \Psi(v, w)F(v, w, t)dv dw &= \iint_{(\mathbb{R}^3 \times \mathbb{R}^3)^2} \int_0^\pi \int_0^{2\pi} [\Psi(v', w') + \Psi(v'_*, w'_*) - \Psi(v_*, w_*) - \Psi(v, w)] \\ &\quad B(\theta)F(v, w, t)F(v_*, w_*, t)dv dw dv_* dw_* d\theta d\varphi, \end{aligned} \quad (36)$$

where, for  $v, w, v_*, w_* \in \mathbb{R}^3$ ,  $\theta \in (0, \pi)$  and  $\varphi \in (0, 2\pi)$ , we have set

$$\sigma = \cos(\theta) \frac{v - v_*}{|v - v_*|} + \sin(\theta)[I \cos(\varphi) + I_1 \sin(\varphi)] \quad \text{and} \quad \omega = \cos(\theta) \frac{w - w_*}{|w - w_*|} + \sin(\theta)[I \cos(\varphi) + I_2 \sin(\varphi)],$$

where  $I = \frac{(v-v_*) \wedge (w-w_*)}{|(v-v_*) \wedge (w-w_*)|}$ , and  $I_1, I_2$  are chosen so that  $(\frac{v-v_*}{|v-v_*|}, I, I_1)$  and  $(\frac{w-w_*}{|w-w_*|}, I, I_2)$  are two direct orthonormal bases, and where

$$\begin{aligned} v' &= \frac{1}{2}(v + v_*) + \frac{1}{2}|v - v_*|\sigma, & v'_* &= \frac{1}{2}(v + v_*) - \frac{1}{2}|v - v_*|\sigma, \\ w' &= \frac{1}{2}(w + w_*) + \frac{1}{2}|w - w_*|\omega, & w'_* &= \frac{1}{2}(w + w_*) - \frac{1}{2}|w - w_*|\omega. \end{aligned}$$

It is not difficult to show that (36) is well-posed in  $L_{loc}^\infty([0, \infty), L^1(\mathbb{R}^3 \times \mathbb{R}^3))$  and that the solution remains nonnegative for all times. Indeed, applying (36) with the test function  $\Psi(v, w) = -\mathbb{1}_{\{F(v, w) < 0\}}$ , using that

$$\begin{aligned} & \left[ -\mathbb{1}_{\{F(v', w') < 0\}} - \mathbb{1}_{\{F(v'_*, w'_*) < 0\}} + \mathbb{1}_{\{F(v, w) < 0\}} + \mathbb{1}_{\{F(v_*, w_*) < 0\}} \right] F(v, w) F(v_*, w_*) \\ & \leq 4 \left[ F_-(v, w) |F(v_*, w_*)| + |F(v, w)| F_-(v_*, w_*) \right] \end{aligned}$$

and that  $\int_0^\pi B(\theta) d\theta = 1$ , we conclude that

$$\frac{d}{dt} \iint_{(\mathbb{R}^3)^2} F_-(v, w, t) dv dw \leq 8 \iint_{(\mathbb{R}^3)^2} F_-(v, w, t) dv dw \times \iint_{(\mathbb{R}^3)^2} |F(v, w, t)| dv dw,$$

whence the conclusion by the Gronwall lemma.

Also, it holds that  $\int_{\mathbb{R}^3} F(v, w, t) dw = f_1(v, t)$  and  $\int_{\mathbb{R}^3} F(v, w, t) dv = f_2(w, t)$ . For example concerning  $f_1$ , we apply the weak coupling equation to some  $\Psi$  depending only on  $v$  and we show that  $\int_{\mathbb{R}^3} F(v, w, t) dw$  solves (35). This follows from the fact that, when fixing  $(v, w)$  and  $(v_*, w_*)$ , the expression between brackets in (36) only depends on  $\sigma$ , so that for any function  $H : \mathbb{S}^2 \mapsto \mathbb{R}$ , we may write

$$\int_{\mathbb{S}^2} H(\sigma) B(\theta) d\sigma = \int_0^\pi \int_0^{2\pi} H\left(\cos(\theta) \frac{v - v_*}{|v - v_*|} + \sin(\theta)[I \cos(\varphi) + I_1 \sin(\varphi)]\right) B(\theta) d\varphi d\theta.$$

We conclude that

$$\begin{aligned} & \iint_{(\mathbb{R}^3 \times \mathbb{R}^3)^2} \int_0^\pi \int_0^{2\pi} [\Psi(v') + \Psi(v'_*)] B(\theta) F(v, w, t) F(v_*, w_*, t) dv dw dv_* dw_* d\theta d\varphi \\ & = \iint_{(\mathbb{R}^3 \times \mathbb{R}^3)^2} \int_{\mathbb{S}^2} [\Psi(v') + \Psi(v'_*)] B(\theta) F(v, w, t) F(v_*, w_*, t) dv dw dv_* dw_* d\sigma \\ & = \iint_{(\mathbb{R}^3 \times \mathbb{R}^3)^2} \int_{\mathbb{S}^2} [\Psi(v') + \Psi(v'_*)] B(\theta) f_1(v, t) f_1(v_*, t) dv dv_* d\sigma. \end{aligned}$$

Consequently, it holds that  $\mathcal{T}_2(f_1(t), f_2(t)) \leq \iint |v - w|^2 F(v, w, t) dv dw =: h(t)$ , and it suffices, as usual, to show that  $h'(t) \leq 0$ . For this, it suffices to verify that for all fixed  $v, w, v_*, w_* \in \mathbb{R}^3$ , all  $\theta \in (0, \pi)$ ,

$$\Delta = \int_0^{2\pi} [|v' - w'|^2 + |v'_* - w'_*|^2 - |v - w|^2 - |v_* - w_*|^2] d\varphi \leq 0.$$

A simple computation, using that  $\int_0^{2\pi} \sigma d\varphi = \int_0^{2\pi} \omega d\varphi = 0$ , shows that

$$\begin{aligned} \Delta &= \int_0^{2\pi} [(v - w) \cdot (v_* - w_*) - |v - v_*| |w - w_*| \sigma \cdot \omega] d\varphi \\ &= \int_0^{2\pi} [(1 - \cos^2 \theta - \sin^2 \theta \sin^2 \varphi)(v - w) \cdot (v_* - w_*) - \sin^2 \theta \cos^2 \varphi |v - v_*| |w - w_*|] d\varphi. \end{aligned}$$

We used that  $|v - v_*||w - w_*|I_1 \cdot I_2 = (v - w) \cdot (v_* - w_*)$ . All in all, we arrive at

$$\Delta = [(v - w) \cdot (v_* - w_*) - |v - v_*||w - w_*|] \sin^2 \theta \int_0^{2\pi} \cos^2 \varphi d\varphi,$$

and the proof is complete.  $\square$

## 6 Porous media equation

We now consider the generalized porous media equation written, with  $A : \mathbb{R}_+ \mapsto \mathbb{R}$  of class  $C^2$ , as

$$\partial_t u - \operatorname{div}(u \nabla[A'(u)]) = 0, \quad x \in \mathbb{R}^d, t \geq 0. \quad (37)$$

Using a gradient flow approach taking advantage of displacement convexity, as introduced in [33], it was discovered by Otto [34], see also [14, 35, 18], that this equation is non-expansive for  $\mathcal{T}_2$ , under a few conditions on  $A$ , including convexity. The method was improved by Bolley and Carrillo [6] who make the contraction property an equivalence to displacement convexity for the energy functional. The coupling method, as defined in the introduction and used in the whole present paper, does not seem to apply directly. However, using Brenier's map, the argument in [6] can somehow be presented in relation with a coupling. We present this argument, staying at an informal level.

**Theorem 14** ([34, 6]) *Consider some  $C^2$  function  $A : \mathbb{R}_+ \mapsto \mathbb{R}$  such that  $B(r) = \int_0^r w A''(w) dw \geq 0$  for all  $r \geq 0$  and such that  $r \mapsto r^{1/d-1} B(r)$  is non-decreasing. Consider two probability densities  $u_1^0, u_2^0$  on  $\mathbb{R}^d$  and the corresponding solutions  $u_1, u_2$  to (37). Then for all  $t \geq 0$ , one has*

$$\mathcal{T}_2(u_1(t), u_2(t)) \leq \mathcal{T}_2(u_1^0, u_2^0).$$

This applies to the porous media equation, i.e. with  $A(u) = m^{-1}u^m$ , as soon as  $m \geq 1$ . The justification of the computation requires at least that  $\int_{\mathbb{R}^d} B(u_1(x, t)) < \infty$ . See [34, 6] for the rigorous proof, which uses that general weak solutions of the porous media equation can be approximated by smooth and positive solutions of suitably regularized nonlinear diffusion equations.

**Proof.** We consider Brenier's map [10] for  $u_1^0$  and  $u_2^0$ , i.e. a convex function  $\Phi : \mathbb{R}^d \mapsto \mathbb{R}$  such that  $\mathcal{T}_2(u_1^0, u_2^0) = \frac{1}{2} \int_{\mathbb{R}^d} |x - \nabla \Phi(x)|^2 u_1^0(x) dx$  and  $\nabla \Phi \# u_1^0 = u_2^0$ . We next consider the probability measure  $v(dx, dy, t)$  (for each  $t \geq 0$ ,  $v(t) \in \mathcal{P}((\mathbb{R}^d)^2)$ ) solving the coupling equation

$$\frac{\partial v}{\partial t} = \operatorname{div}_x(v \nabla_x A'(u_1(x, t))) + \operatorname{div}_y(v \nabla_y A'(u_2(y, t)))$$

and starting from  $v^0 \in K(u_1^0, u_2^0)$  defined by the formula  $v^0(\mathcal{A}) = \int_{\mathbb{R}^d} \mathbf{1}_{\{(x, \nabla \Phi(x)) \in \mathcal{A}\}} u_1^0(x) dx$  for all Borel set  $\mathcal{A} \subset \mathbb{R}^d$ , in short notation  $v^0(x, y) = u_1^0(x) \delta_{\nabla \Phi(x)}(y)$ . Because of this specific initial data, the equation on  $v$  will not provide a global coupling as in the equations treated so far. Again, we only use the weak form, for all smooth  $\varphi : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}$ ,

$$\frac{d}{dt} \iint \varphi(x, y) v(dx, dy, t) = - \iint \left[ \nabla_x \varphi(x, y) \cdot \nabla_x [A'(u_1(x, t))] + \nabla_y \varphi(x, y) \cdot \nabla_y [A'(u_2(y, t))] \right] v(dx, dy, t).$$

One easily verifies, as usual, that for each  $t \geq 0$ ,  $v(t)$  has  $u_1(t)$  and  $u_2(t)$  for marginals: for example, applying the weak equation of  $v$  to some  $\varphi$  depending only on  $x$  shows that  $\int_{y \in \mathbb{R}^d} v(dx, dy, t)$  is a

(weak) solution to (37) and since it starts from  $u_1^0$ , we conclude by uniqueness. As a conclusion, for all  $t \geq 0$ ,  $\mathcal{T}_2(u_1(t), u_2(t)) \leq I(t)$ , where  $I(t) = \frac{1}{2} \iint |x - y|^2 v(dx, dy, t)$ . Next, following [6], we observe that

$$I'(t) = - \iint (\nabla_x[A'(u_1(x, t))] - \nabla_y[A'(u_2(y, t))]) \cdot (x - y)v(dx, dy, t) = D_1(t) + D_2(t),$$

where

$$D_1(t) = - \iint \nabla_x[A'(u_1(x, t))] \cdot (x - y)v(dx, dy, t) \quad \text{and} \quad D_2(t) = - \iint \nabla_y[A'(u_2(y, t))] \cdot (y - x)v(dx, dy, t).$$

In particular, by definition of  $v^0$ ,

$$D_1(0) = - \int \nabla_x[A'(u_1^0(x))] \cdot (x - \nabla\Phi(x))u_1^0(x)dx = - \int \nabla_x[B(u_1^0(x))] \cdot (x - \nabla\Phi(x))dx,$$

where we recall that  $B(r) = \int_0^r wA''(w)dw$ . Integrating by parts, we thus find

$$D_1(0) = \int B(u_1^0(x))[d - \Delta\Phi(x)]dx \leq d \int B(u_1^0(x))[1 - (\det(D^2\Phi(x)))^{1/d}]dx.$$

This uses that for any convex function  $\Phi : \mathbb{R}^d \mapsto \mathbb{R}$ , we have  $d^{-1}\Delta\Phi(x) \geq [\det(D^2\Phi(x))]^{1/d}$ . But since  $\nabla\Phi \# u_1^0 = u_2^0$ , we have, for any  $\varphi : \mathbb{R}^d \mapsto \mathbb{R}$ ,

$$\int \varphi(x)u_1^0(x)dx = \int \varphi((\nabla\Phi)^{-1}(y))u_2^0(y)dy = \int \varphi(x)u_2(\nabla\Phi(x)) \det D^2\Phi(x)dx,$$

so that  $\det D^2\Phi(x) = u_1^0(x)/u_2(\nabla\Phi(x))$  (see [25] for an account on this Monge-Ampère equation). All in all, we have checked that

$$D_1(0) \leq d \int B(u_1^0(x)) \left[ 1 - \left( \frac{u_1^0(x)}{u_2(\nabla\Phi(x))} \right)^{1/d} \right] dx.$$

Proceeding similarly, we see that

$$D_2(0) \leq d \int B(u_2^0(y)) \left[ 1 - \left( \frac{u_2^0(y)}{u_1^0((\nabla\Phi)^{-1}(y))} \right)^{1/d} \right] dy.$$

Performing the substitution  $x = (\nabla\Phi)^{-1}(y)$ , we end with

$$D_2(0) \leq d \int B(u_2^0(\nabla\Phi(x))) \left[ 1 - \left( \frac{u_2^0(\nabla\Phi(x))}{u_1^0(x)} \right)^{1/d} \right] \frac{u_1^0(x)}{u_2(\nabla\Phi(x))} dx.$$

We thus find, with the notation  $y = \nabla\Phi(x)$

$$\begin{aligned} \frac{I'(0)}{d} &\leq \int \left[ B(u_1^0(x)) \left( 1 - \left( \frac{u_1^0(x)}{u_2(y)} \right)^{1/d} \right) + B(u_2^0(y)) \left( 1 - \left( \frac{u_2^0(y)}{u_1^0(x)} \right)^{1/d} \right) \frac{u_1^0(x)}{u_2(y)} \right] dx \\ &= \int u_1^0(x) \left[ \frac{B(u_1^0(x))}{u_1^0(x)} \left( 1 - \left( \frac{u_1^0(x)}{u_2^0(y)} \right)^{1/d} \right) + \frac{B(u_2^0(y))}{u_2^0(y)} \left( 1 - \left( \frac{u_2^0(y)}{u_1^0(x)} \right)^{1/d} \right) \right] dx \\ &= \int u_1^0(x) \left[ \frac{B(u_1^0(x))}{u_1^0(x)} [u_1^0(x)]^{1/d} - \frac{B(u_2^0(y))}{u_2^0(y)} [u_2^0(y)]^{1/d} \right] \left( [u_1^0(x)]^{-1/d} - [u_2^0(y)]^{-1/d} \right) dx. \end{aligned}$$

Since  $r \mapsto r^{1/d-1}B(r)$  is non-decreasing by assumption, we conclude that  $I'(0) \leq 0$ .

The above considerations hold true at any time, and not only at  $t = 0$ . In other words, for all  $t \geq 0$ , we can find a function  $I_t : [t, \infty) \mapsto \mathbb{R}$  such that  $I_t(t) = \mathcal{T}_2(u_1(t), u_2(t))$ ,  $I_t'(t) \leq 0$  and  $\mathcal{T}_2(u_1(s), u_2(s)) \leq I_t(s)$  for all  $0 \leq t \leq s$ . One immediately concludes that for all  $t \geq 0$ ,

$$\limsup_{h \downarrow 0} \frac{\mathcal{T}_2(u_1(t+h), u_2(t+h)) - \mathcal{T}_2(u_1(t), u_2(t))}{h} \leq I_t'(t) \leq 0,$$

so that  $t \mapsto \mathcal{T}_2(u_1(t), u_2(t))$  is non-increasing.  $\square$

## 7 An approach by duality

In order to complete the presentation, we quickly mention another possible and original approach, based on duality. We consider the simplest model, i.e. the heat equation in dimension 1, but all the models treated in the present paper, except the porous media equation, may be treated similarly, with more complicated discretization procedures and more involved computations. See also Villani [40, pages 41-43] for a similar approach concerning the Vlasov equation, in the case of the Monge-Kantorovich distance  $\mathcal{T}_1$ , of which the dual expression is particularly simple.

**Proof of Theorem 1 when  $d = 1$  for  $\rho(x, y) = |x - y|^p$  with  $p \geq 1$ .** We consider two solutions  $u_1, u_2$  to (2), starting from probability measures  $u_1^0, u_2^0$  with finite  $p$ -moment. For  $h > 0$ , we consider the solutions  $u_{1,h}, u_{2,h}$ , starting from  $u_1^0, u_2^0$ , to the discrete heat equation

$$\partial_t u(x, t) - \frac{1}{h^2} [u(x+h, t) + u(x-h, t) - 2u(x, t)] = 0.$$

It can be written in weak form

$$\frac{d}{dt} \int_{\mathbb{R}^d} \varphi(x) u(x, t) dx = \int_{\mathbb{R}^d} \frac{\varphi(x+h) + \varphi(x-h) - 2\varphi(x)}{h^2} u(x, t) dx.$$

It is standard that  $u_{1,h} \rightarrow u_1$  and  $u_{2,h} \rightarrow u_2$  as  $h \rightarrow 0$  (in the weak topology of measures for instance). We will verify that for each  $h > 0$ , it holds that  $\mathcal{T}_p(u_{1,h}(t), u_{2,h}(t)) \leq \mathcal{T}_p(u_1^0, u_2^0)$ , for any  $p \geq 1$ , and this will complete the proof.

We fix  $p \geq 1$  and introduce the set  $Q_p$  of pairs  $(\varphi, \psi)$  of functions from  $\mathbb{R}^d$  to  $\mathbb{R}$  such that for all  $x, y \in \mathbb{R}^d$ ,  $\varphi(x) + \psi(y) \leq |x - y|^p$ . For any pair of probability densities  $f, g$  on  $\mathbb{R}^d$ , the transport cost can also be expressed by duality, see [42], as

$$\mathcal{T}_p(f, g) = \frac{1}{p} \sup_{(\varphi, \psi) \in Q_p} \left[ \int \varphi(x) f(x) dx + \int \psi(y) g(y) dy \right].$$

For  $(\varphi, \psi) \in Q_p$ , we set  $\Delta_{\varphi, \psi}(t) = \int \varphi(x) u_{1,h}(x, t) dx + \int \psi(y) u_{2,h}(y, t) dy$ . Using that  $(\varphi(\cdot+h), \psi(\cdot+h))$  and  $(\varphi(\cdot-h), \psi(\cdot-h))$  both belong to  $Q_p$ , we find

$$\frac{d}{dt} \Delta_{\varphi, \psi}(t) \leq -2h^{-2} \Delta_{\varphi, \psi}(t) + 2ph^{-2} \mathcal{T}_p(u_{1,h}(t), u_{2,h}(t)).$$

This implies that

$$\begin{aligned} e^{2h^{-2}t} \Delta_{\varphi, \psi}(t) &\leq \Delta_{\varphi, \psi}(0) + 2ph^{-2} \int_0^t e^{2h^{-2}s} \mathcal{T}_p(u_{1,h}(s), u_{2,h}(s)) ds \\ &\leq p \mathcal{T}_p(u_1^0, u_2^0) + 2ph^{-2} \int_0^t e^{2h^{-2}s} \mathcal{T}_p(u_{1,h}(s), u_{2,h}(s)) ds. \end{aligned}$$

Taking the supremum over all pairs  $(\varphi, \psi)$  in  $Q_p$  and dividing by  $p$ , we conclude that

$$e^{2h^{-2}t}\mathcal{T}_p(u_{1,h}(t), u_{2,h}(t)) \leq \mathcal{T}_p(u_1^0, u_2^0) + 2h^{-2} \int_0^t e^{2h^{-2}s}\mathcal{T}_p(u_{1,h}(s), u_{2,h}(s))ds.$$

By the Gronwall lemma, we conclude that  $e^{2h^{-2}t}\mathcal{T}_p(u_{1,h}(t), u_{2,h}(t)) \leq e^{2h^{-2}t}\mathcal{T}_p(u_1^0, u_2^0)$  as desired.  $\square$

Unfortunately, we are not able to use a similar procedure *directly* on the (non discretized) heat equation.

## References

- [1] G. BAL AND A. JOLLIVET, *Generalized stability estimates in inverse transport theory*, Inverse Probl. Imaging, 12 (2018), pp. 59–90.
- [2] J.-D. BENAMOU AND Y. BRENIER, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numer. Math., 84 (2000), pp. 375–393.
- [3] A. BLANCHET, V. CALVEZ, AND J. A. CARRILLO, *Convergence of the mass-transport steepest descent scheme for the subcritical Patlak-Keller-Segel model*, SIAM J. Numer. Anal., 46 (2008), pp. 691–721.
- [4] F. BOLLEY, Y. BRENIER, AND G. LOEPER, *Contractive metrics for scalar conservation laws*, J. Hyperbolic Differ. Equ., 2 (2005), pp. 91–107.
- [5] F. BOLLEY AND J. A. CARRILLO, *Tanaka theorem for inelastic Maxwell models*, Comm. Math. Phys., 276 (2007), pp. 287–314.
- [6] F. BOLLEY AND J. A. CARRILLO, *Nonlinear diffusion: geodesic convexity is equivalent to Wasserstein contraction*, Comm. Partial Differential Equations, 39 (2014), pp. 1860–1869.
- [7] F. BOLLEY, I. GENTIL, AND A. GUILLIN, *Convergence to equilibrium in Wasserstein distance for Fokker-Planck equations*, J. Funct. Anal., 263 (2012), pp. 2430–2457.
- [8] ———, *Uniform convergence to equilibrium for granular media*, Arch. Ration. Mech. Anal., 208 (2013), pp. 429–445.
- [9] L. BRASCO, G. CARLIER, AND F. SANTAMBROGIO, *Congested traffic dynamics, weak flows and very degenerate elliptic equations [corrected version of mr2584740]*, J. Math. Pures Appl. (9), 93 (2010), pp. 652–671.
- [10] Y. BRENIER, *Polar factorization and monotone rearrangement of vector-valued functions*, Comm. Pure Appl. Math., 44 (1991), pp. 375–417.
- [11] V. CALVEZ AND J. A. CARRILLO, *Refined asymptotics for the subcritical Keller-Segel system and related functional inequalities*, Proc. Amer. Math. Soc., 140 (2012), pp. 3515–3530.
- [12] J. A. CARRILLO, *EDPs de difusión y transporte óptimo de masa*, Bol. Soc. Mat. Apl., 28 (2004), pp. 129–154.

- [13] J. A. CARRILLO, M. DI FRANCESCO, AND G. TOSCANI, *Strict contractivity of the 2-Wasserstein distance for the porous medium equation by mass-centering*, Proc. Amer. Math. Soc., 135 (2007), pp. 353–363.
- [14] J. A. CARRILLO, R. J. MCCANN, AND C. VILLANI, *Contractions in the 2-Wasserstein length space and thermalization of granular media*, Arch. Ration. Mech. Anal., 179 (2006), pp. 217–263.
- [15] J. A. CARRILLO AND G. TOSCANI, *Wasserstein metric and large-time asymptotics of nonlinear diffusion equations*, in New trends in mathematical physics, World Sci. Publ., Hackensack, NJ, 2004, pp. 234–244.
- [16] ———, *Contractive probability metrics and asymptotic behavior of dissipative kinetic equations*, Riv. Mat. Univ. Parma (7), 6 (2007), pp. 75–198.
- [17] K. CRAIG, I. KIM, AND Y. YAO, *Congested aggregation via Newtonian interaction*, Arch. Ration. Mech. Anal., 227 (2018), pp. 1–67.
- [18] S. DANERI AND G. SAVARÉ, *Eulerian calculus for the displacement convexity in the Wasserstein distance*, SIAM J. Math. Anal., 40 (2008), pp. 1104–1122.
- [19] R. J. DIPERNA AND P.-L. LIONS, *Ordinary differential equations, transport theory and Sobolev spaces*, Invent. Math., 98 (1989), pp. 511–547.
- [20] R. L. DOBRUŠIN, *Vlasov equations*, Funktsional. Anal. i Prilozhen., 13 (1979), pp. 48–58, 96.
- [21] A. FIGALLI, *Existence and uniqueness of martingale solutions for SDEs with rough or degenerate coefficients*, J. Funct. Anal., 254 (2008), pp. 109–153.
- [22] N. FOURNIER, *On pathwise uniqueness for stochastic differential equations driven by stable Lévy processes*, Ann. Inst. Henri Poincaré Probab. Stat., 49 (2013), pp. 138–159.
- [23] N. FOURNIER AND E. LÖCHERBACH, *On a toy model of interacting neurons*, Ann. Inst. Henri Poincaré Probab. Stat., 52 (2016), pp. 1844–1876.
- [24] F. GOLSE, C. MOUHOT, AND T. PAUL, *On the mean field and classical limits of quantum mechanics*, Comm. Math. Phys., 343 (2016), pp. 165–205.
- [25] C. E. GUTIÉRREZ, *The Monge-Ampère equation*, vol. 89 of Progress in Nonlinear Differential Equations and their Applications, Birkhäuser/Springer, [Cham], 2016. Second edition [of MR1829162].
- [26] M. HAURAY, *Wasserstein distances for vortices approximation of Euler-type equations*, Math. Models Methods Appl. Sci., 19 (2009), pp. 1357–1384.
- [27] R. JORDAN, D. KINDERLEHRER, AND F. OTTO, *The variational formulation of the Fokker-Planck equation*, SIAM J. Math. Anal., 29 (1998), pp. 1–17.
- [28] T. KOMATSU, *On the pathwise uniqueness of solutions of one-dimensional stochastic differential equations of jump type*, Proc. Japan Acad. Ser. A Math. Sci., 58 (1982), pp. 353–356.

- [29] J.-F. LE GALL, *Applications du temps local aux équations différentielles stochastiques unidimensionnelles*, in Seminar on probability, XVII, vol. 986 of Lecture Notes in Math., Springer, Berlin, 1983, pp. 15–31.
- [30] H. LI AND G. TOSCANI, *Long-time asymptotics of kinetic models of granular flows*, Arch. Ration. Mech. Anal., 172 (2004), pp. 407–428.
- [31] C. MARCHIORO AND M. PULVIRENTI, *Mathematical theory of incompressible nonviscous fluids*, vol. 96 of Applied Mathematical Sciences, Springer-Verlag, New York, 1994.
- [32] B. MAURY, A. ROUDNEFF-CHUPIN, AND F. SANTAMBROGIO, *Congestion-driven dendritic growth*, Discrete Contin. Dyn. Syst., 34 (2014), pp. 1575–1604.
- [33] R. J. MCCANN, *A convexity principle for interacting gases*, Adv. Math., 128 (1997), pp. 153–179.
- [34] F. OTTO, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations, 26 (2001), pp. 101–174.
- [35] F. OTTO AND M. WESTDICKENBERG, *Eulerian calculus for the contraction in the Wasserstein distance*, SIAM J. Math. Anal., 37 (2005), pp. 1227–1255.
- [36] M. PIERRE, *Global existence in reaction-diffusion systems with control of mass: a survey*, Milan J. Math., 78 (2010), pp. 5417–455.
- [37] M. ROUSSET, *A  $N$ -uniform quantitative Tanaka’s theorem for the conservative Kac’s  $N$ -particle system with Maxwell molecules*. Preprint, 2014.
- [38] H. TANAKA, *Probabilistic treatment of the Boltzmann equation of Maxwellian molecules*, Z. Wahrsch. Verw. Gebiete, 46 (1978/79), pp. 67–105.
- [39] J. L. VÁZQUEZ, *The porous medium equation. New contractivity results*, in Elliptic and parabolic problems, vol. 63 of Progr. Nonlinear Differential Equations Appl., Birkhäuser, Basel, 2005, pp. 433–451.
- [40] C. VILLANI, *Limite de champ moyen (notes de cours de D.E.A.)*, unpublished, available on [https://cedricvillani.org/sites/dev/files/old\\_images/2013/03/champ-moyen.pdf](https://cedricvillani.org/sites/dev/files/old_images/2013/03/champ-moyen.pdf). 2001–2002.
- [41] —, *Topics in optimal transportation*, vol. 58 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2003.
- [42] —, *Optimal transport*, vol. 338 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, 2009. Old and new.