Available online at www.sciencedirect.com
ScienceDirect
J. Math. Pures Appl. 95 (2011) 18–35
JOURNAL DE MATHÉMATIQUES PURES ET APPLIQUÉES
www.elsevier.com/locate/matpur

Contraction of general transportation costs along solutions to Fokker–Planck equations with monotone drifts

Luca Natile [a], Mark A. Peletier [b], Giuseppe Savaré [a,*]

[a] *Dipartimento di Matematica, Università di Pavia, Italy*
[b] *Department of Mathematics and Computer Science and Institute for Complex Molecular Systems, Technische Universiteit Eindhoven, The Netherlands*

Received 10 February 2010
Available online 3 August 2010

Abstract

We shall prove new contraction properties of general transportation costs along nonnegative measure-valued solutions to Fokker–Planck equations in $\mathbb{R}^d$, when the drift is a monotone (or λ-monotone) operator. A new duality approach to contraction estimates has been developed: it relies on the Kantorovich dual formulation of optimal transportation problems and on a variable-doubling technique. The latter is used to derive a new comparison property of solutions of the backward Kolmogorov (or dual) equation. The advantage of this technique is twofold: it directly applies to distributional solutions without requiring stronger regularity, and it extends the Wasserstein theory of Fokker–Planck equations with gradient drift terms, started by Jordan, Kinderlehrer and Otto (1998) [14], to more general costs and monotone drifts, without requiring the drift to be a gradient and without assuming any growth conditions.

Résumé

On démontre de nouvelles propriétés de contraction des coûts du transport global en suivant des solutions de type mesures positives de l'équation de Fokker–Planck où la déviation est un opératateur monotone ou λ-monotone. Une nouvelle approche duale a été développée pour les estimations de contraction : elle s'appuie sur la formulation duale de Kantorovitch des problèmes de transport optimal et sur une technique de doublement des variables. Cette dernière est utilisée pour obtenir une nouvelle propriété de comparaison de solutions de l'équation de Kolmogorov rétrograde (ou de son équation duale). Les avantages de cette technique sont de deux types : d'une part, elle s'applique directement aux solutions au sens des distributions sans demander plus de réguralité, d'autre part, elle généralise la théorie de Wasserstein des équations de Fokker–Planck avec une déviation de type gradient, introduite par Jordan, Kinderlehrer et Otto (1998) [14], à des coûts plus généraux et à des déviations monotones, sans demander que les déviations soient des gradients et sans condition de croissance.

MSC: 35Q84; 35K10; 60J60; 82C31

Keywords: Fokker–Planck equation; Wasserstein distance; Optimal transport; Monotone operators

* Corresponding author.
E-mail addresses: luca.natile@unipv.it (L. Natile), m.a.peletier@tue.nl (M.A. Peletier), giuseppe.savare@unipv.it (G. Savaré).

0021-7824/$ – see front matter
doi:10.1016/j.matpur.2010.07.003

1. Introduction

The aim of this paper is to obtain new uniqueness and contractivity results for nonnegative measure-valued solutions to the Fokker–Planck equation:

$$\partial_t \rho - \Delta\rho - \nabla\cdot(\rho B) = 0, \quad \rho|_{t=0} = \rho_0, \tag{1}$$

where $B:\mathbb{R}^d \to \mathbb{R}^d$ is a Borel λ-monotone operator, $\lambda \in \mathbb{R}$, i.e.

$$\big\langle B(x) - B(y), x - y\big\rangle \geqslant \lambda|x-y|^2 \quad \text{for every } x, y \in \mathbb{R}^d. \tag{2}$$

Here we consider a weakly continuous family of probability measures $(\rho_t)_{t\geqslant 0} \subset \mathscr{P}(\mathbb{R}^d)$ satisfying Eq. (1) in the sense of distributions:

$$\int_0^{+\infty}\int_{\mathbb{R}^d} (\partial_t\zeta + \Delta\zeta - B\cdot\nabla\zeta)\,\mathrm{d}\rho_t\,\mathrm{d}t = 0 \quad \text{for every } \zeta \in C_c^\infty\big(\mathbb{R}^d \times (0,+\infty)\big), \tag{3}$$

with the initial datum ρ_0.

Equations of this type are the subject of several papers by Bogachev, Da Prato, Krylov, Röckner, and Stannat, who consider a very general situation where the Laplacian is replaced by a second order elliptic operator with variable coefficients and B is locally bounded. Existence of solutions has been proved by Bogachev et al. [6, Cor. 3.3], uniqueness has been considered in [5] under general growth-coercivity conditions on B, and regularity has been investigated by Bogachev et al. [7]: in particular, it has been shown that ρ_t is absolutely continuous with respect to the Lebesgue measure for $\mathscr{L}^1$-a.e. t.

When B is Lipschitz continuous, uniqueness can be obtained by standard duality arguments, see e.g. [3, Sec. 3]. Here we want to obtain a more precise stability estimate on the solutions of (1), only assuming monotonicity of B without any growth condition. To achieve this aim, we adopt the point of view of optimal transportation.

The Wasserstein approach to the Fokker–Planck equation in the gradient case. When B is the gradient of a λ-convex function $V:\mathbb{R}^d \to \mathbb{R}$ then (1) can be considered as the *gradient flow* of the perturbed entropy functional

$$\mathscr{H}(\rho) := \int_{\mathbb{R}^d} u(x)\log u(x)\,\mathrm{d}x + \int_{\mathbb{R}^d} V(x)\,\mathrm{d}\rho(x), \quad \rho = u\mathscr{L}^d, \tag{4}$$

in the space $\mathscr{P}_2(\mathbb{R}^d)$ of probability measures with finite quadratic moments endowed with the so-called L^2-Kantorovich–Rubinstein–Wasserstein distance $W_2(\cdot,\cdot)$. This distance can be defined by

$$W_2^2\big(\rho^1,\rho^2\big) := \min\left\{ \int_{\mathbb{R}^d\times\mathbb{R}^d} |x_1-x_2|^2\,\mathrm{d}\boldsymbol{\rho}(x_1,x_2)\colon\ \boldsymbol{\rho}\in\mathscr{P}\big(\mathbb{R}^d\times\mathbb{R}^d\big),\ \boldsymbol{\rho} \text{ is a coupling between } \rho^1 \text{ and } \rho^2\right\}, \tag{5}$$

in terms of couplings, i.e. measures $\boldsymbol{\rho}$ on the product space $\mathbb{R}^d\times\mathbb{R}^d$ whose marginals are ρ^1 and ρ^2 respectively, so that $\boldsymbol{\rho}(E\times\mathbb{R}^d) = \rho^1(E)$ and $\boldsymbol{\rho}(\mathbb{R}^d\times E) = \rho^2(E)$ for every Borel subset $E\subset\mathbb{R}^d$. It is possible to prove that an *optimal coupling* realizing the minimum in (5) always exists.

This remarkable interpretation found in [14] gave rise to a series of studies on the relationships between certain classes of diffusion equations and distances between probability measures induced by optimal transport problems (see e.g. the general overviews of [21,2,22]). One of the strengths of this approach is a new geometric insight (developed in [16]) in the evolution process: in the case of (1) the λ-convexity of the potential V reflects a λ-convexity property (also called *displacement convexity*) of the functional $\mathscr{H}$ along the geodesics of $\mathscr{P}_2(\mathbb{R}^d)$. This nice feature, discovered by McCann [15], suggests that one can adapt some typical basic existence, approximation, and regularity results for gradient flows of convex functionals in Euclidean spaces or Riemannian manifolds to the measure-theoretic setting of $\mathscr{P}_2(\mathbb{R}^d)$. This program has been carried out (see e.g. [2]) and, among the most interesting estimates, it provides the λ-contraction property,

$$W_2\big(\rho_t^1,\rho_t^2\big) \leqslant \mathrm{e}^{-\lambda t}W_2\big(\rho_0^1,\rho_0^2\big) \quad \text{for every } t\geqslant 0, \tag{6}$$

where ρ_t^i, $i=1,2$, are the solutions to (1) starting from the initial data $\rho_0^i \in \mathscr{P}_2(\mathbb{R}^d)$.

Two strategies for the derivation of the contraction estimate (6) in the gradient case. In order to prove (6) in the gradient case $B = \nabla V$, essentially two basic strategies have been proposed:

1. A first approach, developed by [10] for smooth evolutions and by [2] in a measure-theoretic setting, starts from Eq. (1) written in the form:
$$\partial_t \rho + \nabla \cdot (\rho \boldsymbol{v}) = 0, \qquad \boldsymbol{v} = -\left(\frac{\nabla u}{u} + \nabla V\right), \qquad \rho = u\mathcal{L}^d, \tag{7}$$
and it is based on two ingredients: the first one is the formula which evaluates the derivative of the squared Wasserstein distance from a fixed measure σ along the (absolutely continuous) curve ρ in $\mathcal{P}_2(\mathbb{R}^d)$,
$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}W_2^2(\rho_t, \sigma) = \int\limits_{\mathbb{R}^d \times \mathbb{R}^d} \langle \boldsymbol{v}_t(x), y - x\rangle \,\mathrm{d}\boldsymbol{\rho}_t(x, y) \quad \text{for } \mathcal{L}^1\text{-a.e. } t > 0, \tag{8}$$
where $\boldsymbol{\rho}_t$ is an optimal coupling between ρ_t and σ.
The second ingredient is the "subgradient" property of the vector field $\boldsymbol{v}_t$ given by (7), related to the displacement convexity of $\mathcal{H}$: in the case $\lambda = 0$ it reads as
$$\int\limits_{\mathbb{R}^d \times \mathbb{R}^d} \langle \boldsymbol{v}_t(x), y - x\rangle \,\mathrm{d}\boldsymbol{\rho}_t(x, y) \leqslant \mathcal{H}(\sigma) - \mathcal{H}(\rho_t) \quad \text{if } \boldsymbol{v}_t = -\left(\frac{\nabla u_t}{u_t} + \nabla V\right). \tag{9}$$
Combination of (8) and (9) yields the so-called Evolution Variational Inequality,
$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}W_2^2(\rho_t, \sigma) \leqslant \mathcal{H}(\sigma) - \mathcal{H}(\rho_t) \quad \text{for every } \sigma \in \mathcal{P}_2(\mathbb{R}^d), \tag{10}$$
which easily yields (6) for $\lambda = 0$ by a variable-doubling argument (see [2, Theorem 11.1.4]).
The main technical point here is that (9) requires $\boldsymbol{v}_t \in L^2(\rho_t)$ and (8) holds if for every $0 < t_0 < t_1 < +\infty$,
$$\int\limits_{t_0}^{t_1}\int\limits_{\mathbb{R}^d} |\boldsymbol{v}_t|^2 \,\mathrm{d}\rho_t \,\mathrm{d}t = \int\limits_{t_0}^{t_1}\int\limits_{\mathbb{R}^d} \left|\frac{\nabla u_t}{u_t} + \nabla V\right|^2 \mathrm{d}\rho_t \,\mathrm{d}t < +\infty, \tag{11}$$
which should be imposed (in a suitable distributional sense) as an *a priori* regularity assumption on the solution of (1). We do not know if solutions to (3) exhibit a similar regularization effect. A second, even more difficult point prevents a simple extension of (10) to the general non-gradient case: it is the lack of a potential V and therefore of an entropy-like functional $\mathcal{H}$ satisfying an inequality similar to (9).
2. A second approach has been proposed by Otto and Westdickenberg [17] and further developed in [12,9]: it is based on the Benamou–Brenier [4] representation formula for the Wasserstein distance:
$$W_2^2(\rho_0, \rho_1) = \inf\left\{\int\limits_0^1\int\limits_{\mathbb{R}^d} |\boldsymbol{v}_t|^2 \,\mathrm{d}\rho_t \,\mathrm{d}t \colon \partial_t\rho_t + \nabla \cdot (\rho_t v_t) = 0 \text{ in } \mathbb{R}^d \times (0, 1),\ \rho_0 = \rho|_{t=0},\ \rho_1 = \rho|_{t=1}\right\}, \tag{12}$$
and on a careful analysis of the effect of the evolution semigroup generated by the equation on curves in $\mathcal{P}_2(\mathbb{R}^d)$ and its Riemannian tensor $\int_{\mathbb{R}^d} |\boldsymbol{v}|^2 \,\mathrm{d}\rho$. This technique involves various repeated differentiations and works quite well if a nice semigroup preserving smoothness and strict positivity of the densities has already been defined. Once contraction has been proved on smooth initial data, the evolution can be extended to more general ones but it seems hard to extend the uniqueness result to cover a general distributional solution to the equation.

Main result of the paper: contraction estimates for distributional solutions. Our purpose is twofold:

- First of all we want to find a new approach working directly on measure-valued solutions to (1) just satisfying the usual distributional formulation (3).
We note that in general (1) does not exhibit the same regularization effect of the heat equation. Even in the gradient case $B = \nabla V$, there exist solutions ρ_t to (3) which are not of class $C^1(\mathbb{R}^d)$ for every $t \geqslant 0$: take, e.g., the invariant

measure $\rho_t \equiv Z^{-1}\mathrm{e}^{-V}$ for a suitable convex function $V \notin C^1(\mathbb{R}^d)$ with $\mathrm{e}^{-V} \in L^1(\mathbb{R}^d)$. Moreover, distributional solutions are easily obtained by approximation arguments, as regularization or splitting methods, and they should be better suited to deal with the infinite-dimensional case, as in [3]: a stability result for such a weak class of solutions should be useful in these cases.

- Second, we want to cover the case of an arbitrary monotone field B, without any growth restriction, and to extend contraction estimates to more general transportation costs.

To this aim, let us first introduce the general cost functional:

$$\mathcal{C}_h\big(\rho^1,\rho^2\big) := \inf\left\{ \int_{\mathbb{R}^d\times\mathbb{R}^d} h\big(|x_1-x_2|\big)\,\mathrm{d}\boldsymbol{\rho}(x_1,x_2)\colon\ \boldsymbol{\rho} \in \mathscr{P}\big(\mathbb{R}^d\times\mathbb{R}^d\big),\ \boldsymbol{\rho} \text{ is a coupling between } \rho^1 \text{ and } \rho^2 \right\}. \tag{13}$$

Throughout this paper we assume that

$$h : [0,+\infty) \to [0,+\infty) \text{ is a continuous and } \textit{non-decreasing} \text{ function with } h(0) = 0.$$

Among the possible interesting choices of h, the case $h(r) := r^p$ is associated with the family of L^p Wasserstein distances (whose L^2-version has been introduced in (5)) on the space $\mathscr{P}_p(\mathbb{R}^d)$ of all the probability measures with moment of order p. When h is a bounded concave function satisfying $h(r) > 0$ if $r > 0$, $d(x,y) := h(|x-y|)$ is a bounded and complete distance function on $\mathbb{R}^d$ inducing the usual Euclidean topology so that $\mathcal{C}_h(\cdot,\cdot)$ is a complete metric on the space $\mathscr{P}(\mathbb{R}^d)$ whose topology coincides with the usual weak one (see e.g. [2, Proposition 7.1.5]).

Since we are not assuming any homogeneity on the general cost function h, its rescaled versions,

$$h_s(r) := h\big(r\mathrm{e}^s\big), \quad s \in \mathbb{R},\ r \geqslant 0, \tag{14}$$

will be useful. Let us now state our main result:

Theorem 1.1. *If ρ^1, ρ^2 are two distributional solutions to* (3) *satisfying the summability condition,*

$$\int_{t_0}^{t_1}\int_{\mathbb{R}^d} \big|B(x) - \lambda x\big|\,\mathrm{d}\rho_t(x)\,\mathrm{d}t < +\infty \quad \textit{for every } 0 < t_0 < t_1 < +\infty, \tag{15}$$

then they satisfy:

$$\mathcal{C}_{h_{\lambda t}}\big(\rho_t^1,\rho_t^2\big) \leqslant \mathcal{C}_h\big(\rho_0^1,\rho_0^2\big) \quad \textit{for every } t \geqslant 0. \tag{16}$$

In particular, if $\rho_0^1 = \rho_0^2$ then ρ^1 and ρ^2 coincide for every time $t \geqslant 0$.

Let us make explicit some consequences of (16) according to the different signs of λ and the behaviour of h near 0 and $+\infty$:

Corollary 1.2. *Let ρ^1, ρ^2 be two distributional solutions to* (3) *satisfying* (15)*.*

a) *If B is monotone, i.e. $\lambda \geqslant 0$, then*

$$\mathcal{C}_h\big(\rho_t^1,\rho_t^2\big) \leqslant \mathcal{C}_h\big(\rho_0^1,\rho_0^2\big).$$

b) *If B is λ-monotone with $\lambda > 0$ and h satisfies for some exponent $p > 0$,*

$$h(\alpha r) \geqslant \alpha^p h(r) \quad \textit{for every } \alpha \geqslant 1 \textit{ and } r \geqslant 0, \tag{17}$$

then

$$\mathcal{C}_h\big(\rho_t^1,\rho_t^2\big) \leqslant \mathrm{e}^{-p\lambda t}\mathcal{C}_h\big(\rho_0^1,\rho_0^2\big).$$

c) *If B is λ-monotone with $\lambda < 0$ and h satisfies, for some exponent $p > 0$,*

$$h(\alpha r) \geqslant \alpha^p h(r) \quad \text{for every } \alpha \leqslant 1 \text{ and } r \geqslant 0,$$

then

$$\mathcal{C}_h\big(\rho_t^1, \rho_t^2\big) \leqslant \mathrm{e}^{-p\lambda t}\mathcal{C}_h\big(\rho_0^1, \rho_0^2\big).$$

d) *In the particular case of the Wasserstein distance W_p, $p \geqslant 1$, and for every $\lambda \in \mathbb{R}$ we have*:

$$W_p\big(\rho_t^1, \rho_t^2\big) \leqslant \mathrm{e}^{-\lambda t} W_p\big(\rho_0^1, \rho_0^2\big). \tag{18}$$

Theorem 1.1 has a simple application to invariant measures $\rho_\infty \in \mathscr{P}(\mathbb{R}^d)$, which are stationary solutions of (3) and therefore satisfy:

$$\int_{\mathbb{R}^d} (\Delta\zeta - B \cdot \nabla\zeta)\, \mathrm{d}\rho_\infty = 0 \quad \text{for every } \zeta \in C_{\mathrm{c}}^\infty\big(\mathbb{R}^d\big). \tag{19}$$

Corollary 1.3 *(Strongly monotone operators and invariant measures). Let us suppose that B is strongly monotone, i.e. $\lambda > 0$. Then Eq.* (19) *has at most one solution $\rho_\infty \in \mathscr{P}(\mathbb{R}^d)$ satisfying the integrability condition*:

$$\int_{\mathbb{R}^d} |Bx - \lambda x|\, \mathrm{d}\rho_\infty(x) < \infty. \tag{20}$$

For each solution ρ_t to (3)–(15) *and each cost h satisfying* (17) *we have*:

$$\mathcal{C}_h(\rho_t, \rho_\infty) \leqslant \mathrm{e}^{-p\lambda(t-t_0)}\mathcal{C}_h(\rho_{t_0}, \rho_\infty). \tag{21}$$

Note that in the case $\lambda > 0$ condition (20) is weaker than $B \in L^1(\rho_\infty; \mathbb{R}^d)$.

Remark 1.4 *(An equivalent formulation of the contraction estimate).* We can give an equivalent version of (16) by keeping fixed the cost but rescaling the measures. In fact, we can associate to the solutions ρ^1, ρ^2 of (3) their rescaled versions $\tilde\rho^1, \tilde\rho^2$ defined by:

$$\tilde\rho^j(E) := \rho^j\big(\mathrm{e}^{-\lambda t}E\big) \quad \text{for every Borel set } E \subset \mathbb{R}^d,\ j = 1, 2. \tag{22}$$

Then $\tilde\rho^j$ is the push-forward of ρ^j through the map $x \mapsto \mathrm{e}^{\lambda t}x$ and satisfies the change-of-variables formula:

$$\int_{\mathbb{R}^d} \zeta(y)\, \mathrm{d}\tilde\rho^j(y) = \int_{\mathbb{R}^d} \zeta\big(\mathrm{e}^{\lambda t}x\big)\, \mathrm{d}\rho^j(x) \quad \text{for every } \zeta \in C_{\mathrm{b}}\big(\mathbb{R}^d\big). \tag{23}$$

Inequality (16) is then equivalent to

$$\mathcal{C}_h\big(\tilde\rho_t^1, \tilde\rho_t^2\big) \leqslant \mathcal{C}_h\big(\rho_0^1, \rho_0^2\big) \quad \text{for every } t > 0. \tag{24}$$

Strategy of the proof: the Kantorovich duality and a variable-doubling technique. In order to prove Theorem 1.1 we develop a new strategy, generalizing [18]. It relies on the well-known dual Kantorovich formulation [21] of the transportation cost (13):

$$\mathcal{C}_h\big(\rho^1, \rho^2\big) = \sup\left\{\int_{\mathbb{R}^d} \phi^1\, \mathrm{d}\rho^1 + \int_{\mathbb{R}^d} \phi^2\, \mathrm{d}\rho^2\colon \phi^1, \phi^2 \in C_{\mathrm{b}}\big(\mathbb{R}^d\big),\ \phi^1(x_1) + \phi^2(x_2) \leqslant h\big(|x_1 - x_2|\big)\right\}. \tag{25}$$

This formula reduces the estimate of the cost $\mathcal{C}_h(\rho_T^1, \rho_T^2)$ of two solutions of (1) at a certain final time T to the estimate of

$$\Sigma\big(\phi^1, \phi^2; T\big) := \int_{\mathbb{R}^d} \phi^1\, \mathrm{d}\rho_T^1 + \int_{\mathbb{R}^d} \phi^2\, \mathrm{d}\rho_T^2, \tag{26}$$

for an arbitrary pair of functions ϕ^1, ϕ^2 satisfying the constraint

$$\phi^1(x_1) + \phi^2(x_2) \leqslant h\big(|x_1 - x_2|\big) \quad \text{for every } x_1, x_2 \in \mathbb{R}^d. \tag{27}$$

Assuming for the sake of simplicity that B is monotone, bounded and smooth, we can obtain an estimate of $\Sigma(\phi^1, \phi^2; T)$ by solving the final-value problem for the adjoint equation

$$\partial_t \phi^i + \Delta \phi^i - B \cdot \nabla \phi^i = 0 \quad \text{in } \mathbb{R}^d \times (0, T), \qquad \phi^i(\cdot, T) := \phi^i, \tag{28}$$

since the distributional formulation (3) yields

$$\Sigma\big(\phi_T^1, \phi_T^2; T\big) = \Sigma\big(\phi_0^1, \phi_0^2; 0\big). \tag{29}$$

The following crucial result, based on a "variable-doubling technique", provides the final step, showing that ϕ_0^1, ϕ_0^2 still satisfy the constraint (27) so that $\Sigma(\phi_0^1, \phi_0^2; 0) \leqslant \mathcal{C}_h(\rho_0^1, \rho_0^2)$.

Theorem 1.5. *If $\phi^1, \phi^2 \in C_{\mathrm{b}}^{2,1}(\mathbb{R}^d \times [0, T])$ are solutions of* (28) *in the case when B is monotone, bounded and smooth, such that*

$$\phi^1(x_1, T) + \phi^2(x_2, T) \leqslant h\big(|x_1 - x_2|\big) \quad \textit{for every } x_1, x_2 \in \mathbb{R}^d,$$

then

$$\phi^1(x_1, 0) + \phi^2(x_2, 0) \leqslant h\big(|x_1 - x_2|\big) \quad \textit{for every } x_1, x_2 \in \mathbb{R}^d.$$

Remark 1.6. While we prove Theorem 1.5 for bounded and smooth drifts B, and solutions $\phi^{1,2} \in C_{\mathrm{b}}^{2,1}(\mathbb{R}^d \times [0, T])$, the property clearly carries over to any pointwise limit of such solutions. We therefore expect it to hold for a much larger class of monotone drifts B and solutions.

Plan of the paper. In Section 2, we collect some tools useful to our arguments: we present a slightly refined version of the Kantorovich duality, an approximation technique of the cost functional, the construction of a smooth and bounded approximation of the operator B, and a rescaling trick which allows to consider $\lambda = 0$ in the following arguments. Section 3 is devoted to the proof of Theorem 1.5, the last section contains the proof of Theorem 1.1.

2. Preliminaries

In this section we collect some preliminary and technical regularization results which will turn to be useful in the sequel.

2.1. $C_{\mathrm{c}}^\infty(\mathbb{R}^d)$ functions in Kantorovich duality

Let us first show that we can assume ϕ^1, ϕ^2 are smooth and compactly supported in the duality formula (25).

Proposition 2.1. *If the cost function h is Lipschitz continuous and satisfies $\lim_{r \uparrow +\infty} h(r) = +\infty$, then*

$$\mathcal{C}_h\big(\rho^1, \rho^2\big) = \sup\left\{ \int_{\mathbb{R}^d} \phi^1 \,\mathrm{d}\rho^1 + \int_{\mathbb{R}^d} \phi^2 \,\mathrm{d}\rho^2 \colon\ \phi^1, \phi^2 \in C_{\mathrm{c}}^\infty\big(\mathbb{R}^d\big),\ \phi^1(x_1) + \phi^2(x_2) \leqslant h\big(|x_1 - x_2|\big) \right\}. \tag{30}$$

Proof. Let us recall that the h-transform of a given bounded function $\zeta : \mathbb{R}^d \to \mathbb{R}$ is defined as

$$\zeta^h(y) := \inf_{x \in \mathbb{R}^d} h\big(|x - y|\big) - \zeta(x), \tag{31}$$

and it is a bounded and Lipschitz continuous function satisfying $\zeta(x) + \zeta^h(y) \leqslant h(|x - y|)$.

Let us fix $c < \mathcal{C}_h(\rho^1, \rho^2)$ and admissible $\phi^1, \phi^2 \in C_{\mathrm{b}}(\mathbb{R}^d)$ such that

$$\int_{\mathbb{R}^d} \phi^1 \,\mathrm{d}\rho^1 + \int_{\mathbb{R}^d} \phi^2 \,\mathrm{d}\rho^2 > c. \tag{32}$$

By possibly replacing ϕ^2 with $(\phi^1)^h \geqslant \phi^2$ and ϕ^1 with $(\phi^1)^{hh} \geqslant \phi^1$, it is not restrictive to assume that ϕ^1, ϕ^2 are also Lipschitz continuous. Adding to ϕ^1 and subtracting from ϕ^2 a suitable constant, we can also assume that $\phi^1 \geqslant 0$ and $\phi^2 \leqslant 0$.

Let us now consider a family of mollifiers κ_η and of cutoff functions χ_R defined by:

$$\kappa_\eta(x) := \eta^{-d}\kappa(x/\eta), \qquad \chi_R(x) := \chi(x/R), \quad x \in \mathbb{R}^d, \ \eta, R > 0, \tag{33a}$$

where $\kappa, \chi \in C_c^\infty(\mathbb{R}^d)$ satisfy,

$$\kappa \geqslant 0, \quad \int_{\mathbb{R}^d} \kappa(x)\,\mathrm{d}x = 1, \quad 0 \leqslant \chi \leqslant 1, \qquad \chi(x) = 0 \quad \text{if } |x| \geqslant 1, \qquad \chi(x) = 1 \quad \text{if } |x| \leqslant 1/2. \tag{33b}$$

We set $\phi^1_\eta := \phi^1 * \kappa_\eta$ and $\phi^2_\eta := \phi^2 * \kappa_\eta - \delta_\eta$, where

$$\delta_\eta := \sup\bigl(\phi^1 * \kappa_\eta - \phi^1\bigr)^+ + \sup\bigl(\phi^2 * \kappa_\eta - \phi^2\bigr)^+.$$

The definition of δ_η yields:

$$\phi^1_\eta(x_1) + \phi^2_\eta(x_2) \leqslant \phi^1 * \kappa_\eta(x_1) - \phi^1(x_1) + \phi^2 * \kappa_\eta(x_2) - \phi^2(x_2) - \delta_\eta + h\bigl(|x_1 - x_2|\bigr) \leqslant h\bigl(|x_1 - x_2|\bigr).$$

Moreover, since ϕ^1, ϕ^2 are Lipschitz, ϕ^1_η and ϕ^2_η converge to ϕ^1, ϕ^2 uniformly as $\eta \downarrow 0$, so that ϕ^1_η and ϕ^2_η are a smooth admissible pair still satisfying the sign condition $\phi^1_\eta \geqslant 0$, $\phi^2_\eta \leqslant 0$ and (32) for a sufficiently small $\eta > 0$.

Let us now choose $R_0 > 0$ such that

$$h(r) \geqslant \sup \phi^1_\eta \quad \text{for every } r \geqslant R_0. \tag{34}$$

Setting $\phi^1_{\eta,R} := \phi^1_\eta \chi_R \leqslant \phi^1_\eta$ we easily have for $R \geqslant R_0$

$$\inf_{x_1 \in \mathbb{R}^d} h\bigl(|x_1 - x_2|\bigr) - \phi^1_{\eta,R}(x_1) \geqslant 0 \quad \text{if } |x_2| \geqslant 2R \geqslant R + R_0.$$

Since $\phi^2_{\eta,4R} := \phi^2_\eta \chi_{4R}$ satisfies $\phi^2_{\eta,4R}(x_2) = \phi^2_\eta(x_2)$ if $|x_2| \leqslant 2R$ and $\phi^2_{\eta,4R}(x_2) \leqslant 0$ for every $x_2 \in \mathbb{R}^d$, it follows that $\phi^1_{\eta,R}, \phi^2_{\eta,4R}$ is an admissible couple in $C_c^\infty(\mathbb{R}^d)$, and, for R sufficiently large, it still satisfies (32). □

2.2. Regularization of the cost function

In this section we shall show that it is sufficient to consider nonnegative, Lipschitz, and unbounded costs (as those considered in Proposition 2.1) in the proof of Theorem 1.1.

Lemma 2.2. *If* (16) *holds for every nonnegative Lipschitz and nondecreasing cost function h with* $\lim_{r\uparrow+\infty} h(r) = +\infty$, *then it holds for every continuous and nondecreasing cost h.*

Proof. We first prove that it is sufficient to consider nonnegative Lipschitz costs; in a second step, we deal with the asymptotic requirement.

Step 1: h Lipschitz. Adding a suitable constant we can assume that $h(r) \geqslant h(0) = 0$. We can then approximate h from below by the increasing sequence of nonnegative Lipschitz functions,

$$h^n(r) := \inf_{s \geqslant 0} h(s) + n|r - s|,$$

which satisfies

$$0 = h^n(0) \leqslant h^n(r) \leqslant h(r), \qquad \lim_{n\uparrow+\infty} h^n(r) = h(r) \quad \text{for every } r \geqslant 0,$$

the convergence being uniform on each compact interval of $[0,+\infty)$. Applying Lemma 2.3 below we find:

$$\mathcal{C}_{h_{\lambda t}}\big(\rho_t^1,\rho_t^2\big) \overset{(37)}{=} \lim_{n\uparrow+\infty} \mathcal{C}_{h^n_{\lambda t}}\big(\rho_t^1,\rho_t^2\big) \overset{(16)}{\leqslant} \liminf_{n\uparrow+\infty} \mathcal{C}_{h^n}\big(\rho_0^1,\rho_0^2\big) \overset{(37)}{=} \mathcal{C}_h\big(\rho_0^1,\rho_0^2\big).$$

Step 2: $\lim_{r\uparrow+\infty} h(r)=+\infty$. Let us set $\rho_0:=\rho_0^1+\rho_0^2$, let us introduce the function

$$m(r):=\rho_0\big(\mathbb{R}^d\setminus rU\big),\quad U:=\big\{x\in\mathbb{R}^d\colon |x|<1\big\},$$

and let us consider a sequence r_n in $[0,+\infty)$ such that

$$r_0:=0,\qquad r_1:=1,\qquad r_{n+1}-r_n\geqslant r_n-r_{n-1},\quad \text{and}\quad m(r_{n+1})\leqslant 2^{-n}.$$

It is easy to check that r_n is a diverging increasing sequence; if g is the piecewise linear function satisfying $g(r_n)=n$, i.e.

$$g(r):=n+\frac{r-r_n}{r_{n+1}-r_n}\quad \text{if } r\in[r_n,r_{n+1}],$$

then g is Lipschitz continuous, increasing, unbounded, concave, it satisfies $g(0)=0$, and

$$\begin{aligned} G:=\int_{\mathbb{R}^d} g\big(|x|\big)\,\mathrm{d}\rho_0(x) &= \int_{\mathbb{R}^d}\Bigg(\int_0^{|x|} g'(r)\,\mathrm{d}r\Bigg)\mathrm{d}\rho_0(x)=\int_{\mathbb{R}^d}\Bigg(\int_0^{+\infty} g'(r)\mathbb{1}_{r\leqslant|x|}\,\mathrm{d}r\Bigg)\mathrm{d}\rho_0(x)\\ &=\int_0^{\infty} g'(r)m(r)\,\mathrm{d}r=\sum_{n=1}^{+\infty}\frac{1}{r_n-r_{n-1}}\int_{r_{n-1}}^{r_n} m(r)\,\mathrm{d}r\leqslant\sum_{n=0}^{+\infty} m(r_n)<+\infty. \end{aligned}$$

We can thus consider the perturbed cost,

$$h^\varepsilon(r):=h(r)+\varepsilon g(r),$$

which is Lipschitz, increasing, unbounded. Since g is concave, increasing, and $g(0)=0$, we have:

$$g\big(|x_1-x_2|\big)\leqslant g\big(|x_1|+|x_2|\big)\leqslant g\big(|x_1|\big)+g\big(|x_2|\big)\quad \text{for every } x_1,x_2\in\mathbb{R}^d, \tag{35}$$

so that if $\boldsymbol{\rho}_0$ is an optimal coupling between ρ_0^1 and ρ_0^2 for the cost h (we can assume that the initial cost is finite), then

$$\begin{aligned} \mathcal{C}_h\big(\rho_0^1,\rho_0^2\big) &\leqslant \mathcal{C}_{h^\varepsilon}\big(\rho_0^1,\rho_0^2\big)\leqslant \mathcal{C}_h\big(\rho_0^1,\rho_0^2\big)+\varepsilon\int_{\mathbb{R}^d\times\mathbb{R}^d} g\big(|x_1-x_2|\big)\,\mathrm{d}\boldsymbol{\rho}_0(x_1,x_2)\\ &\overset{(35)}{\leqslant} \mathcal{C}_h\big(\rho_0^1,\rho_0^2\big)+\varepsilon\int_{\mathbb{R}^d\times\mathbb{R}^d}\big(g\big(|x_1|\big)+g\big(|x_2|\big)\big)\,\mathrm{d}\boldsymbol{\rho}_0(x_1,x_2)=\mathcal{C}_h\big(\rho_0^1,\rho_0^2\big)+\varepsilon G. \end{aligned}$$

Therefore, if Theorem 1.1 holds for h^ε we have:

$$\mathcal{C}_h\big(\rho_t^1,\rho_t^2\big)\leqslant \mathcal{C}_{h^\varepsilon}\big(\rho_t^1,\rho_t^2\big)\leqslant \mathcal{C}_{h^\varepsilon}\big(\rho_0^1,\rho_0^2\big)\leqslant \mathcal{C}_h\big(\rho_0^1,\rho_0^2\big)+\varepsilon G.$$

Passing to the limit as $\varepsilon\downarrow 0$ we conclude. □

The following result provides a variant of the well-known stability properties of transportation costs (see [19, Theorem 3], [22, Theorem 5.20]) and holds in the much more general setting of optimal transportation in Radon metric spaces [2, Chapter 6].

Lemma 2.3 *(Lower semicontinuity of the cost functional w.r.t. local uniform convergence of h). Let* $h:[0,+\infty)\to[0,+\infty)$ *be a continuous cost function and let* $h^n:[0,+\infty)\to[0,+\infty)$ *be a sequence of lower semicontinuous functions converging to h locally uniformly in* $[0,+\infty)$. *For every couple* $\rho^1,\rho^2\in\mathcal{P}(\mathbb{R}^d)$ *we have*:

$$\liminf_{n\uparrow+\infty}\mathcal{C}_{h^n}\big(\rho^1,\rho^2\big)\geqslant\mathcal{C}_h\big(\rho^1,\rho^2\big). \tag{36}$$

In particular, if $h^n \leqslant h$ for every $n \in \mathbb{N}$, then

$$\lim_{n\to+\infty} \mathcal{C}_{h^n}\big(\rho^1,\rho^2\big) = \mathcal{C}_h\big(\rho^1,\rho^2\big). \tag{37}$$

Proof. Let us set $H^n(x_1,x_2) := h^n(|x_1-x_2|)$ and observe that H^n converges to $H(x_1,x_2) := h(|x_1-x_2|)$ uniformly on compact sets of $\mathbb{R}^d\times\mathbb{R}^d$. If $\boldsymbol{\rho}_n \in \mathscr{P}(\mathbb{R}^d\times\mathbb{R}^d)$ is an optimal coupling between ρ^1,ρ^2 with respect to the cost h^n, then

$$\mathcal{C}_{h^n}\big(\rho^1,\rho^2\big) = \int_{[0,+\infty)} z\,\mathrm{d}\rho_n(z), \quad \text{where } \rho_n = \big(H^n\big)_\#\boldsymbol{\rho}_n \in \mathscr{P}\big([0,+\infty)\big).$$

Since the marginals of $\boldsymbol{\rho}_n$ are fixed, the sequence $(\boldsymbol{\rho}_n)_{n\in\mathbb{N}}$ is tight and up to the extraction of a suitable subsequence (still denoted by $\boldsymbol{\rho}_n$) we can suppose that $\boldsymbol{\rho}_n$ converge to some limit coupling $\boldsymbol{\rho}$ between ρ^1,ρ^2 in $\mathscr{P}(\mathbb{R}^d\times\mathbb{R}^d)$. Since ρ_n weakly converge to $\rho = H_\#\boldsymbol{\rho}$ in $\mathscr{P}([0,+\infty))$ by [2, Lemma 5.2.1], standard lower semicontinuity of integrals with nonnegative continuous integrands [2, Lemma 5.1.7] yields:

$$\liminf_{n\to+\infty} \int_{[0,+\infty)} z\,\mathrm{d}\rho_n(z) \geqslant \int_{[0,+\infty)} z\,\mathrm{d}\rho(z) = \int_{\mathbb{R}^d\times\mathbb{R}^d} H(x_1,x_2)\,\mathrm{d}\boldsymbol{\rho}(x_1,x_2) \geqslant \mathcal{C}_h\big(\rho^1,\rho^2\big). \qquad \square$$

2.3. Bounded, smooth approximations of a monotone operator

If $A:\mathbb{R}^d\to\mathbb{R}^d$ is a monotone operator then there exists [8, Corollary 2.1] a maximal monotone multivalued extension $\mathsf{A}:\mathbb{R}^d \rightrightarrows \mathbb{R}^d$ (thus taking values in $2^{\mathbb{R}^d}$) such that $A(x)\in\mathsf{A}(x)$ for every $x\in\mathbb{R}^d$. We denote by $\mathsf{A}^\circ(x)$ the element of minimal norm in (the closed convex set) $\mathsf{A}(x)$. [1, Corollary 1.4] shows that the set $\mathsf{A}(x)\subset\mathbb{R}^d$ reduces to the singleton $\{A(x)\}$ $\mathscr{L}^d$-almost everywhere: in fact it satisfies

$$\mathsf{A}(x) = \big\{\mathsf{A}^\circ(x)\big\} = \big\{A(x)\big\} \quad \text{for } \mathscr{L}^d\text{-a.e. } x\in\mathbb{R}^d, \qquad \mathsf{A}(x) = \operatorname{conv}\Big\{\lim_{n\to\infty} A(x_n) \text{ for some } x_n\to x\Big\}. \tag{38}$$

We recall the following important approximation result [13, Theorem 4.1]: we denote by U the open unit ball in $\mathbb{R}^d$.

Theorem *(Fitzpatrick–Phelps). For every maximal monotone operator $\mathsf{A}:\mathbb{R}^d\rightrightarrows\mathbb{R}^d$, there exists a sequence of maximal monotone operators $\mathsf{A}_n:\mathbb{R}^d\rightrightarrows\mathbb{R}^d$ such that, for each $x\in\mathbb{R}^d$ and all n,*

$$\mathsf{A}(x)\cap nU \subset \mathsf{A}_n(x) \subset n\overline{U}, \qquad \mathsf{A}_n(x)\setminus\mathsf{A}(x)\subset n\partial U \quad \textit{for every } x\in\mathbb{R}^d. \tag{39}$$

Notice that (39) yields in particular

$$\big|\mathsf{A}_n^\circ(x)\big| = \min\big(\big|\mathsf{A}^\circ(x)\big|, n\big) \quad \text{for every } x\in\mathbb{R}^d. \tag{40}$$

Theorem 2.4. *Let $\mathsf{A}:\mathbb{R}^d\rightrightarrows\mathbb{R}^d$ be a maximal monotone operator and $(\beta_n)_{n\in\mathbb{N}}$ a vanishing sequence of positive real numbers. There exists a sequence of smooth, globally Lipschitz, and bounded monotone operators $A_n:\mathbb{R}^d\to\mathbb{R}^d$ such that*

$$\operatorname{Lip}(A_n)\leqslant n, \qquad \big|A_n(x)\big| \leqslant \min\big(\big|\mathsf{A}^\circ(x)\big|, n\big) + \beta_n, \qquad \lim_{n\to+\infty} A_n(x) = \mathsf{A}^\circ(x) \quad \textit{for every } x\in\mathbb{R}^d. \tag{41}$$

Proof. Let A_n be a sequence of maximal monotone operators satisfying (39) and let $Y_n:\mathbb{R}^d\to\mathbb{R}^d$ be the Moreau–Yosida approximation of A_n of parameter n^{-1} [8, Proposition 2.6],

$$Y_n(x) := n\big(x - \big(I + n^{-1}\mathsf{A}_n\big)^{-1}x\big).$$

Note that Y_n is an n-Lipschitz monotone map satisfying:

$$\big|Y_n(x)\big| \leqslant \big|\mathsf{A}_n^\circ(x)\big| \overset{(40)}{=} \min\big(\big|\mathsf{A}^\circ(x)\big|, n\big) \quad \text{for every } x\in\mathbb{R}^d. \tag{42}$$

Let us fix $x \in \mathbb{R}^d$ and let $x_n \in \mathbb{R}^d$ be the unique solution of

$$x_n + n^{-1}\mathsf{A}_n(x_n) \ni x \quad \text{so that} \quad Y_n(x) = n(x - x_n) \in \mathsf{A}_n(x_n). \tag{43}$$

If $n > |\mathsf{A}^\circ(x)|$ then (42) yields $Y_n(x) \notin n\partial U$; applying (39) and (42) again we get:

$$Y_n(x) \in \mathsf{A}(x_n), \quad \big|Y_n(x)\big| \leqslant \big|\mathsf{A}^\circ(x)\big|, \qquad |x - x_n| \leqslant n^{-1}\big|\mathsf{A}^\circ(x)\big| \quad \text{for every } n > \big|\mathsf{A}^\circ(x)\big|. \tag{44}$$

Since the graph of A is closed, any accumulation point y of the bounded sequence $Y_n(x)$ satisfies:

$$y \in \mathsf{A}(x), \quad |y| \leqslant \big|\mathsf{A}^\circ(x)\big|. \tag{45}$$

We thus conclude that $\lim_{n\uparrow+\infty} Y_n(x) = \mathsf{A}^\circ(x)$ for every $x \in \mathbb{R}^d$.

To conclude the proof we need to regularize Y_n: to this aim we consider the family of mollifiers κ_η as in (33a) and we set:

$$A_n := Y_n * \kappa_\eta \quad \text{with } \eta := (nk)^{-1}\beta_n \text{ where } k := \int_{\mathbb{R}^d} |x|\kappa(x)\,\mathrm{d}x, \tag{46}$$

so that

$$\big|A_n(x) - Y_n(x)\big| \leqslant \eta k \operatorname{Lip}(Y_n) \leqslant n\eta k \leqslant \beta_n. \qquad \square$$

We consider now a radial smoothing:

Proposition 2.5. *Let $A_n : \mathbb{R}^d \to \mathbb{R}^d$ be smooth, Lipschitz, and bounded monotone operators satisfying* (41). *For every $m \in \mathbb{N}$ there exist bounded, smooth, Lipschitz, and monotone operators $A_{n,m}$ such that*

$$\operatorname{Lip}(A_{n,m}) \leqslant n, \qquad \sup_{x\in\mathbb{R}^d}\big|A_{n,m}(x)\big| \leqslant n + \beta_n, \qquad \sup_{x\in\mathbb{R}^d}\big|\mathrm{D}A_{n,m}(x)\cdot x\big| \leqslant 2m(n+\beta_n), \tag{47}$$

$$\lim_{m\uparrow+\infty} A_{n,m}(x) = A_n(x) \quad \textit{for every } x \in \mathbb{R}^d. \tag{48}$$

Proof. We consider a family of mollifiers $\kappa_\eta = \eta^{-1}\kappa(\cdot/\eta) \in C^\infty_{\mathrm{c}}(\mathbb{R})$, where κ satisfies:

$$\operatorname{supp}(\kappa) \subset [0,2], \quad 0 \leqslant \kappa \leqslant \kappa(1) = 1, \qquad (1-x)\kappa'(x) \geqslant 0, \qquad \int_{\mathbb{R}} \kappa(x)\,\mathrm{d}x = 1, \tag{49}$$

and the function $\vartheta \in C^\infty_{\mathrm{c}}(0,+\infty)$ defined by $\vartheta(r) := \kappa(-\log r)$, $r > 0$. We set:

$$A_{n,m}(x) := m\int_0^{+\infty} A(rx)\vartheta\big(r^m\big)\frac{\mathrm{d}r}{r}. \tag{50}$$

The change of variable $r = \mathrm{e}^{-z}$ shows that

$$A_{n,m}(x) = m\int_{\mathbb{R}} A_n\big(x\mathrm{e}^{-z}\big)\kappa(mz)\,\mathrm{d}z = A_n^x * \kappa_{1/m}(0), \quad \text{where } A_n^x(z) := A_n\big(x\mathrm{e}^{z}\big) \text{ for } z \in \mathbb{R}.$$

It is then easy to check that $|\mathrm{D}A_{n,m}| \leqslant n$, since

$$\big|\mathrm{D}A_{n,m}(x)\big| \leqslant m\int_{\mathbb{R}} \big|\mathrm{D}A_n\big(x\mathrm{e}^{-z}\big)\big|\mathrm{e}^{-z}\kappa(mz)\,\mathrm{d}z \overset{(41)}{\leqslant} n\int_{\mathbb{R}} \mathrm{e}^{-y/m}\kappa(y)\,\mathrm{d}y \overset{(49)}{\leqslant} n,$$

and $A_{n,m}$ converges pointwise to A_n as $m \uparrow +\infty$.

Concerning the third bound of (47) we easily have:

$$\mathrm{D}A_{n,m}(x)\cdot x = m\int_0^{+\infty}\mathrm{D}A_n(rx)\cdot x\vartheta\big(r^m\big)\,\mathrm{d}r = m\int_0^{+\infty}\frac{\mathrm{d}}{\mathrm{d}r}\big(A_n(rx)\big)\vartheta\big(r^m\big)\,\mathrm{d}r$$
$$= -m^2\int_0^{+\infty}A_n(rx)\tilde\vartheta\big(r^m\big)\frac{\mathrm{d}r}{r}\quad \text{where } \tilde\vartheta(r) := r\vartheta'(r);$$

the inequality follows since by (49) the total variation of κ and thus of ϑ is 2, so that

$$m^2\int_0^{+\infty}\big|\tilde\vartheta\big(r^m\big)\big|\frac{\mathrm{d}r}{r} = m\int_0^{+\infty}\big|\tilde\vartheta(r)\big|\frac{\mathrm{d}r}{r} = m\int_0^{+\infty}\big|\vartheta'(r)\big|\,\mathrm{d}r = 2m. \qquad \square$$

2.4. λ-Monotonicity and rescaling

We show here a simple rescaling argument (inspired by [11], where the rescaling technique has been applied to a wide class of diffusion equations), which is useful to deduce the estimates in the general λ-monotone case to the simpler case of a monotone operator.

We therefore assume that $\lambda \neq 0$, and we introduce the time rescaling functions:

$$\mathsf{s}(t) := \int_0^t \mathrm{e}^{2\lambda r}\,\mathrm{d}r = \frac{1}{2\lambda}\big(\mathrm{e}^{2\lambda t}-1\big), \qquad \mathsf{t}(s) := \frac{1}{2\lambda}\log(1+2\lambda s), \quad s\in[0,S_\infty), \tag{51}$$

where

$$S_\infty := \begin{cases} +\infty & \text{if } \lambda > 0,\\ -1/(2\lambda) & \text{if } \lambda < 0; \end{cases} \tag{52}$$

notice that $\mathsf{t}(\mathsf{s}(t)) = t$.

We associate to a family of probability measures ρ_t, $t\in[0,T]$, their rescaled versions σ_s, $s\in[0,S_\infty)$, defined by:

$$\sigma_s(E) := \rho_{\mathsf{t}(s)}\big(\mathrm{e}^{-\lambda\mathsf{t}(s)}E\big)\quad \text{for every Borel set } E\subset\mathbb{R}^d. \tag{53}$$

If $B:\mathbb{R}^d\to\mathbb{R}^d$ is a λ-monotone Borel map we set $A := B-\lambda I$, and

$$\tilde B(y,s) := \mathrm{e}^{-\lambda\mathsf{t}(s)}B\big(\mathrm{e}^{-\lambda\mathsf{t}(s)}y\big), \qquad \tilde A(y,s) = \mathrm{e}^{-\lambda\mathsf{t}(s)}A\big(\mathrm{e}^{-\lambda\mathsf{t}(s)}y\big) \quad \text{for } y\in\mathbb{R}^d,\ s\in\mathbb{R}. \tag{54}$$

Notice that if B is λ-monotone, then A and $\tilde A(\cdot,s)$, $s\in[0,S_\infty)$, are monotone.

Proposition 2.6. *A continuous family $\rho_t\in\mathscr{P}(\mathbb{R}^d)$ is a distributional solution of* (3) *if and only if the rescaled measures σ_s defined by* (53) *and* (51) *satisfy*:

$$\int_0^{S_\infty}\int_{\mathbb{R}^d}\big(\partial_s\varphi+\Delta\varphi-\tilde A(\cdot,s)\cdot\nabla\varphi\big)\,\mathrm{d}\sigma_s\,\mathrm{d}s = 0 \quad \textit{for every } \varphi\in C_c^\infty\big(\mathbb{R}^d\times(0,S_\infty)\big). \tag{55}$$

If ρ satisfies (15)*, then*

$$\int_{s_0}^{s_1}\int_{\mathbb{R}^d}\big|\tilde A(x,s)\big|\,\mathrm{d}\sigma_s\,\mathrm{d}s < +\infty \quad \textit{for every } 0<s_0<s_1<S_\infty, \tag{56}$$

and in this case σ satisfies

$$\int_{\mathbb{R}^d}\varphi(\cdot,s_1)\,\mathrm{d}\sigma_{s_1} - \int_{\mathbb{R}^d}\varphi(\cdot,s_0)\,\mathrm{d}\sigma_{s_0} = \int_{s_0}^{s_1}\int_{\mathbb{R}^d}\big(\partial_s\varphi+\Delta\varphi-\tilde A(y,s)\cdot\nabla\varphi\big)\,\mathrm{d}\rho_s\,\mathrm{d}s, \tag{57}$$

for every test function $\varphi\in C_b^{2,1}(\mathbb{R}^d\times[s_0,s_1])$ with bounded first and second derivatives.

Proof. We introduce the change of variable map $\mathsf{X}(x,t) := (e^{\lambda t}x, \mathsf{s}(t))$ and for a given smooth function $\varphi \in C_c^\infty(\mathbb{R}^d \times (0, S_\infty))$ we set $\zeta(x,t) := \varphi(e^{\lambda t}x, \mathsf{s}(t)) = \varphi \circ \mathsf{X}$. Denoting by $(y,s) \in \mathbb{R}^d \times [0, S_\infty)$ the new variables, easy calculations show that in $\mathbb{R}^d \times (0, +\infty)$ we have

$$\partial_t \zeta = \mathsf{s}'\big(\partial_s\varphi + \lambda e^{-2\lambda t}\nabla_y\varphi \cdot y\big) \circ \mathsf{X}, \qquad \nabla_x\zeta = e^{\lambda t}(\nabla_y\varphi \circ \mathsf{X}),$$
$$\Delta_x\zeta = e^{2\lambda t}(\Delta_y\varphi \circ \mathsf{X}), \qquad B \cdot \nabla_x\zeta = e^{2\lambda t}\big(\big(\tilde{B}(y,s) \cdot \nabla_y\varphi\big) \circ \mathsf{X}\big),$$

where we used the fact that $B = e^{\lambda t}(\tilde{B} \circ \mathsf{X})$; in particular we have:

$$\partial_t\zeta - B \cdot \nabla_x\zeta = \mathsf{s}'\big(\partial_s\varphi - \tilde{A}(y,s) \cdot \nabla_y\varphi\big) \circ \mathsf{X}.$$

We thus get:

$$\begin{aligned}\int_{\mathbb{R}^d} (\partial_t\zeta + \Delta_x\zeta - B \cdot \nabla_x\zeta)\,d\rho_t &= \mathsf{s}'(t)\int_{\mathbb{R}^d}\big(\partial_s\varphi + \Delta_y\varphi - \tilde{A}(y,s) \cdot \nabla_y\varphi\big) \circ \mathsf{X}\,d\rho_t \\ &= \mathsf{s}'(t)\int_{\mathbb{R}^d}\big(\partial_s\varphi + \Delta_y\varphi - \tilde{A}(y,s) \cdot \nabla_y\varphi\big)\,d\sigma_{\mathsf{s}(t)},\end{aligned}$$

since $\sigma_{\mathsf{s}(t)}(E) = \rho_t(e^{-\lambda t}E)$ for every Borel set $E \subset \mathbb{R}^d$. Eventually we obtain:

$$\int_0^{+\infty}\int_{\mathbb{R}^d}(\partial_t\zeta + \Delta_x\zeta - B \cdot \nabla_x\zeta)\,d\rho_t\,dt = \int_0^{S_\infty}\int_{\mathbb{R}^d}\big(\partial_s\varphi + \Delta_y\varphi - \tilde{A}(y,s) \cdot \nabla_y\varphi\big)\,d\sigma_s\,ds.$$

Eq. (56) follows by a simple application of the change-of-variable formula (23), since for every $t > 0$,

$$\begin{aligned}\int_{\mathbb{R}^d}\big|\tilde{A}(y,s)\big|\,d\sigma_s(y) &\overset{(54)}{=} e^{-\lambda \mathsf{t}(s)}\int_{\mathbb{R}^d}\big|A\big(e^{-\lambda \mathsf{t}(s)}y\big)\big|\,d\sigma_s(y) \\ &\overset{(53)}{=} e^{-\lambda \mathsf{t}(s)}\int_{\mathbb{R}^d}\big|A(x)\big|\,d\rho_{\mathsf{t}(s)}(x) = e^{-\lambda \mathsf{t}(s)}\int_{\mathbb{R}^d}\big|B(x) - \lambda x\big|\,d\rho_{\mathsf{t}(s)}(x).\end{aligned}$$

Since $\mathsf{t}'(s) = e^{-2\lambda \mathsf{t}(s)}$ we eventually get for $t_i = \mathsf{t}(s_i)$:

$$\begin{aligned}\int_{s_0}^{s_1}\int_{\mathbb{R}^d}\big|\tilde{A}(x,s)\big|\,d\sigma_s\,ds &= \int_{s_0}^{s_1}\bigg(\int_{\mathbb{R}^d}\big|B(x) - \lambda x\big|\,d\rho_{\mathsf{t}(s)}(x)\bigg)e^{\lambda \mathsf{t}(s)}\mathsf{t}'(s)\,ds \\ &= \int_{t_0}^{t_1}\bigg(\int_{\mathbb{R}^d}\big|B(x) - \lambda x\big|\,d\rho_t(x)\bigg)e^{\lambda t}\,dt \overset{(15)}{<} +\infty.\end{aligned}$$

(57) follows from (55) when φ belongs to $C_c^\infty(\mathbb{R}^d \times [s_0, s_1])$. If $\varphi \in C_b^{2,1}(\mathbb{R}^d \times [s_0, s_1])$ via a standard convolution and truncation argument we find an approximation sequence $\varphi_k \in C_c^\infty(\mathbb{R}^d \times [s_0, s_1])$ such that $\varphi_k, \partial_t\varphi_k, \nabla\varphi_k, \Delta\varphi_k$ remains uniformly bounded and converge pointwise to $\varphi, \partial_t\varphi, \nabla\varphi, \Delta\varphi$ respectively. By (56) we can apply the Lebesgue Dominated Convergence Theorem to pass to the limit in (57) written for φ_k, thus obtaining the same identity for φ. □

We conclude this section by a simple remark combining the regularization technique of Section 2.3 and the time rescaling (54).

Lemma 2.7. *Let $A := B - \lambda I$ be a monotone operator, let us consider a sequence $A_{n,m}$, $n, m \in \mathbb{N}$, of smooth monotone operators given by Theorem* 2.4 *and Proposition* 2.5*, and let us set*:

$$\tilde{A}_{n,m}(y,s) := e^{-\lambda \mathsf{t}(s)}A_{n,m}\big(e^{-\lambda \mathsf{t}(s)}y\big), \quad y \in \mathbb{R}^d,\ s \in [0, S_\infty), \tag{58}$$

defined as in (54), (51)*. Then $\tilde{A}_{n,m}$ are Lipschitz in $\mathbb{R}^d \times [0, S]$ for every $S \in [0, S_\infty)$.*

Proof. We just have to check that $|\partial_s \tilde{A}_{n,m}(\cdot, s)|$ is uniformly bounded in $\mathbb{R}^d \times [0, S]$: since $\mathrm{t}'(s) = \mathrm{e}^{-2\lambda \mathrm{t}(s)}$ a simple calculation yields

$$\begin{aligned}\partial_s \tilde{A}_{n,m}(y, s) &= -\lambda \mathrm{e}^{-3\lambda \mathrm{t}(s)} A_{n,m}\big(\mathrm{e}^{-\lambda \mathrm{t}(s)} y\big) - \lambda \mathrm{e}^{-3\lambda \mathrm{t}(s)} \mathrm{D}A_{n,m}\big(\mathrm{e}^{-\lambda \mathrm{t}(s)} y\big) \cdot y \\ &= -\lambda \mathrm{e}^{-2\lambda \mathrm{t}(s)} \tilde{A}_{n,m}(y, s) - \lambda \mathrm{e}^{-\lambda \mathrm{t}(s)} \tilde{Q}_{n,m}(y, s),\end{aligned}$$

where $Q_{n,m}(x) := \mathrm{D}A_{n,m}(x) \cdot x,\ x \in \mathbb{R}^d$.

Since $\mathrm{e}^{-\lambda \mathrm{t}(s)}$ is uniformly bounded with all its derivative in each compact interval $[0, S]$, $S < \infty$, (47) shows that $Q_{n,m}$ is bounded and therefore $\tilde{A}_{n,m}$ is Lipschitz with respect to s. □

3. A comparison result for the backward equation

In this section we give the proof of Theorem 1.5 in a slightly more general form, in order to be applied to (a suitably regularized version of) the rescaled formulation considered in Proposition 2.6.

Let us suppose that $\tilde{A} : (y, s) \in \mathbb{R}^d \times [0, S_\infty) \to \tilde{A}(y, s) \in \mathbb{R}^d$ is a smooth vector field satisfying:

$$\sup_{\mathbb{R}^d \times [0,S]} |\tilde{A}_s| + |\partial_s \tilde{A}| + |\mathrm{D}\tilde{A}| < +\infty \quad \text{for every } S \in [0, S_\infty), \tag{59}$$

$$\tilde{A}(\cdot, s) \text{ is monotone for every } s \in [0, S_\infty). \tag{60}$$

We denote by $\mathcal{L}[\cdot]$ the differential operator defined by:

$$\mathcal{L}[\varphi](y, s) := \Delta_y \varphi(y, s) - \tilde{A}(y, s) \cdot \nabla_y \varphi(y, s), \quad \varphi(\cdot, s) \in C^2\big(\mathbb{R}^d\big),\ (y, s) \in \mathbb{R}^d \times [0, S_\infty). \tag{61}$$

Thanks to (59) and (60), we can apply the existence result [20, Theorem 3.2.1] and for every $S \in [0, S_\infty)$ and $\phi \in C_c^\infty(\mathbb{R}^d)$ we can find a solution $\varphi \in C_b^{2,1}(\mathbb{R}^d \times [0, S))$ of the backward evolution equation

$$\partial_s \varphi + \mathcal{L}[\varphi] = 0 \quad \text{in } \mathbb{R}^d \times [0, S], \qquad \varphi(\cdot, S) = \phi(\cdot). \tag{62}$$

We have:

Theorem 3.1. *Let $h : [0, +\infty) \to \mathbb{R}$ be a continuous and non-decreasing function. Let $\varphi^1, \varphi^2 \in C_b^{2,1}(\mathbb{R}^d \times [0, S])$ be solutions of the "backward" inequality*

$$\partial_s \varphi + \mathcal{L}[\varphi] \geqslant 0 \quad \text{in } \mathbb{R}^d \times [0, S], \tag{63}$$

such that

$$\varphi^1(y_1, S) + \varphi^2(y_2, S) \leqslant h\big(|y_1 - y_2|\big) \quad \text{for every } y_1, y_2 \in \mathbb{R}^d.$$

Then

$$\varphi^1(y_1, 0) + \varphi^2(y_2, 0) \leqslant h\big(|y_1 - y_2|\big) \quad \text{for every } y_1, y_2 \in \mathbb{R}^d.$$

Proof. By approximating h from above, it is not restrictive to assume that $h \in C^1[0, +\infty)$ with $h'(0) = 0$; in particular the map $H(y_1, y_2) := h(|y_1 - y_2|)$ is of class C^1 in $\mathbb{R}^d \times \mathbb{R}^d$ and satisfies

$$\nabla_{y_1} H(y_1, y_2) = -\nabla_{y_2} H(y_1, y_2) = g(y_1, y_2)(y_1 - y_2), \tag{64}$$

where

$$0 \leqslant g(y_1, y_2) = g(y_2, y_1) := \begin{cases} \frac{h'(|y_1 - y_2|)}{|y_1 - y_2|} & \text{if } y_1 \neq y_2, \\ 0 & \text{if } y_1 = y_2. \end{cases} \tag{65}$$

The argument combines a variable-doubling technique and a classical variant of the maximum principle. Let us first show that if φ^1, φ^2 satisfy the *strict* inequality:

$$\partial_s \varphi^j + \mathcal{L}\big[\varphi^j\big] > 0 \quad \text{in } \mathbb{R}^d \times [0, S),\ j = 1, 2, \tag{66}$$

then the function

$$f(y_1, y_2, s) := \varphi^1(y_1, s) + \varphi^2(y_2, s) - H(y_1, y_2)$$

cannot attain a (local) maximum in a point $(\bar{y}_1, \bar{y}_2, \bar{s})$ with $\bar{s} < S$. We argue by contradiction and we suppose that $(\bar{y}_1, \bar{y}_2, \bar{s})$ is a local maximizer of f with $\bar{s} < S$; we thus have:

$$\partial_s f(\bar{y}_1, \bar{y}_2, \bar{s}) \leqslant 0, \qquad \nabla_{y_1} f(\bar{y}_1, \bar{y}_2, \bar{s}) = 0, \qquad \nabla_{y_2} f(\bar{y}_1, \bar{y}_2, \bar{s}) = 0,$$

so that

$$\begin{aligned} \partial_s\varphi^1(\bar{y}_1, \bar{s}) + \partial_s\varphi^2(\bar{y}_2, \bar{s}) &\leqslant 0, \\ \nabla_{y_1}\varphi^1(\bar{y}_1, \bar{s}) = \nabla_{y_1} H(\bar{y}_1, \bar{y}_2) &\overset{(64)}{=} g(\bar{y}_1, \bar{y}_2)(y_1 - y_2), \\ \nabla_{y_2}\varphi^2(\bar{y}_2, \bar{s}) = \nabla_{y_2} H(\bar{y}_1, \bar{y}_2) &\overset{(64)}{=} g(\bar{y}_1, \bar{y}_2)(y_2 - y_1). \end{aligned} \tag{67}$$

It follows that

$$\tilde{A}(\bar{y}_1, \bar{s}) \cdot \nabla_{y_1}\varphi^1(\bar{y}_1, \bar{s}) + \tilde{A}(\bar{y}_2, \bar{s}) \cdot \nabla_{y_2}\varphi^2(\bar{y}_2, \bar{s}) = g(\bar{y}_1, \bar{y}_2)\big(\tilde{A}(\bar{y}_1, \bar{s}) - \tilde{A}(\bar{y}_2, \bar{s})\big) \cdot (\bar{y}_1 - \bar{y}_2) \overset{(65)}{\geqslant} 0. \tag{68}$$

On the other hand, since $H(\bar{y}_1 + z, \bar{y}_2 + z) = H(\bar{y}_1, \bar{y}_2)$, the function,

$$\mathbb{R}^d \ni z \mapsto \varphi^1(\bar{y}_1 + z, \bar{s}) + \varphi^2(\bar{y}_2 + z, \bar{s}) - H(\bar{y}_1, \bar{y}_2) = f(\bar{y}_1 + z, \bar{y}_2 + z, \bar{s}),$$

has a local maximum at $z = 0$ so that

$$\Delta_{y_1}\varphi^1(\bar{y}_1, \bar{s}) + \Delta_{y_2}\varphi^2(\bar{y}_2, \bar{s}) \leqslant 0. \tag{69}$$

Combining (68), (68), and (69) we obtain:

$$\big(\partial_s\varphi^1 + \mathcal{L}[\varphi^1]\big)(\bar{y}_1, \bar{s}) + \big(\partial_s\varphi^2 + \mathcal{L}[\varphi^2]\big)(\bar{y}_2, \bar{s}) \leqslant 0,$$

which contradicts (66).

Suppose now that φ^1, φ^2 satisfy the inequality (63) and let us set for $\varepsilon, \delta > 0$,

$$\varphi^j_{\varepsilon,\delta}(y_j, s) := \varphi^j(y_j, s) - \delta(S - s) - \varepsilon e^{-s}|y_j|^2, \quad j = 1, 2.$$

We easily get:

$$\begin{aligned} \partial_s\varphi^j_{\varepsilon,\delta} &= \partial_s\varphi^j + \delta + \varepsilon e^{-s}|y_j|^2, \\ \mathcal{L}\big[\varphi^j_{\varepsilon,\delta}\big] &= \mathcal{L}\big[\varphi^j\big] - e^{-s}\big(2d\varepsilon + 2\varepsilon\tilde{A}(y_j, s) \cdot y_j\big), \\ \partial_s\varphi^j_{\varepsilon,\delta} + \mathcal{L}\big[\varphi^j_{\varepsilon,\delta}\big] &\geqslant \delta + \varepsilon e^{-s}\big(|y_j|^2 - 2d - C_n|y_j|\big), \end{aligned}$$

where $C_n = \sup_{y,s} |\tilde{A}_n(y, s)| < +\infty$.

It follows that for every $\delta > 0$ there exists a coefficient $\varepsilon > 0$ sufficiently small such that $\varphi^1_{\varepsilon,\delta}, \varphi^2_{\varepsilon,\delta}$ satisfy (66). On the other hand, the continuous function

$$(y_1, y_2, s) \mapsto f_{\varepsilon,\delta}(y_1, y_2, s) := \varphi^1_{\varepsilon,\delta}(y_1, s) + \varphi^2_{\varepsilon,\delta}(y_2, s) - h(|y_1 - y_2|), \quad y_1, y_2 \in \mathbb{R}^d,\ s \in [0, S],$$

attains its maximum at some point $(\bar{y}_1, \bar{y}_2, \bar{s}) \in \mathbb{R}^d \times \mathbb{R}^d \times [0, S]$; by the previous argument, we conclude that $\bar{s} = S$ and therefore for every $y_1, y_2 \in \mathbb{R}^d$,

$$\varphi^1_{\varepsilon,\delta}(y_1, 0) + \varphi^2_{\varepsilon,\delta}(y_2, 0) - h\big(|y_1 - y_2|\big) \leqslant f_{\varepsilon,\delta}(\bar{y}_1, \bar{y}_2, S) \leqslant \varphi^1(\bar{y}_1, S) + \varphi^2(\bar{y}_2, S) - h\big(|\bar{y}_1 - \bar{y}_2|\big) \leqslant 0.$$

Passing to the limit as $\varepsilon, \delta \downarrow 0$ we conclude. □

We conclude this section by recalling two well-known estimates:

Lemma 3.2 *(Uniform estimates). Let $\varphi \in C^{2,1}_{\mathrm{b}}(\mathbb{R}^d \times [0,S]) \cap C^\infty(\mathbb{R}^d \times (0,S))$ be the solution of* (62). *Then*

$$\sup_{\mathbb{R}^d \times [0,S]} |\varphi| \leqslant \sup_{\mathbb{R}^d} |\phi|, \qquad \sup_{\mathbb{R}^d \times [0,S]} |\nabla\varphi| \leqslant \sup_{\mathbb{R}^d} |\nabla\phi|. \tag{70}$$

Proof. The first inequality is a direct application of the maximum principle (see e.g. [20, Theorem 3.1.1]). By differentiating the equation with respect to y we obtain:

$$\partial_s \mathrm{D}\varphi + \mathcal{L}[\mathrm{D}\varphi] - \mathrm{D}\tilde{A}\mathrm{D}\varphi = 0,$$

and then

$$\frac{1}{2}\partial_s |\mathrm{D}\varphi|^2 + \frac{1}{2}\mathcal{L}\big[|\mathrm{D}\varphi|^2\big] - \mathrm{D}\tilde{A}\mathrm{D}\varphi \cdot \mathrm{D}\varphi - \big|\mathrm{D}^2\varphi\big|^2 = 0.$$

Since $\tilde{A}$ is monotone the quadratic form associated to $\mathrm{D}\tilde{A}$ is nonnegative and therefore,

$$\partial_s |\mathrm{D}\varphi|^2 + \mathcal{L}\big[|\mathrm{D}\varphi|^2\big] \geqslant 0.$$

A further application of the maximum principle yields (70). □

4. Proof of Theorem 1.1

We split the proof in various steps. Just to fix some notation, we consider a family $A_{n,m}$ of smooth, bounded, Lipschitz, and monotone operators approximating $A := B - \lambda I$ as in Proposition 2.5 and their rescaled version $\tilde{A}_{n,m}$ defined by (58). $\mathcal{L}_{n,m}[\cdot]$ are the associated differential operators:

$$\mathcal{L}_{n,m}[\varphi](y,s) := \Delta_y \varphi(y,s) - \tilde{A}_{n,m}(y,s) \cdot \nabla_y \varphi(y,s), \quad \varphi(\cdot,s) \in C^2\big(\mathbb{R}^d\big),\ (y,s) \in \mathbb{R}^d \times [0, S_\infty), \tag{71}$$

as in (61). Lemma 2.7 shows that $\tilde{A}$ satisfy (59).

Step 1: Reduction to the monotone case $\lambda = 0$. When $\lambda \neq 0$ we apply the rescaling argument of Section 2.4: we thus introduce the time rescaling $\mathrm{t}(s)$ defined by (51) and the corresponding measures $\sigma^i_s = \tilde{\rho}^i_{\mathrm{t}(s)}$ as in (53), which satisfy (56) and (57) for the rescaled operators $\tilde{A}$ of (54). Taking into account Remark 1.4 and the fact that $\sigma^i_s = \tilde{\rho}^i_{\mathrm{t}(s)}$, the thesis follows if we show that

$$\mathcal{C}_h\big(\sigma^1_s, \sigma^2_s\big) \leqslant \mathcal{C}_h\big(\sigma^1_0, \sigma^2_0\big) \quad \text{for every } s \in [0, S_\infty) \tag{72}$$

(see (52) for the definition of S_∞).

Step 2: If

$$\mathcal{C}_h\big(\sigma^1_{s_1}, \sigma^2_{s_1}\big) \leqslant \mathcal{C}_h\big(\sigma^1_{s_0}, \sigma^2_{s_0}\big) \quad \text{for every } 0 < s_0 < s_1 < S_\infty, \tag{73}$$

then (72) *holds*. When h is bounded, (73) implies (72) by taking a simple limit as $s_0 \downarrow 0$ and using the fact that the map $(\sigma^1, \sigma^2) \mapsto \mathcal{C}_h(\sigma^1, \sigma^2)$ is continuous with respect to weak convergence in $\mathscr{P}(\mathbb{R}^d) \times \mathscr{P}(\mathbb{R}^d)$. If (72) holds for every bounded Lipschitz cost, then it holds for every continuous and nondecreasing cost by Lemma 2.2.

Step 3: We claim the following:

Let $\phi^1, \phi^2 \in C^\infty_{\mathrm{c}}(\mathbb{R}^d)$ be satisfying the constraint $\phi^1(y_1) + \phi^2(y_2) \leqslant h(|y_1 - y_2|)$. Then

$$\int_{\mathbb{R}^d} \phi^1 \, \mathrm{d}\sigma^1_{s_1} + \int_{\mathbb{R}^d} \phi^2 \sigma^2_{s_1} \leqslant \mathcal{C}_h\big(\sigma^1_{s_0}, \sigma^2_{s_0}\big) + \ell K_{n,m}, \tag{74}$$

where $\ell := \sup_{\mathbb{R}^d} |\nabla\phi^1| + \sup_{\mathbb{R}^d} |\nabla\phi^2|$ *and*

$$K_{n,m} := \int_{s_0}^{s_1} \int_{\mathbb{R}^d} |\tilde{A}_{n,m} - \tilde{A}| \, \mathrm{d}\sigma^1_s \, \mathrm{d}s + \int_{s_0}^{s_1} \int_{\mathbb{R}^d} |\tilde{A}_{n,m} - \tilde{A}| \, \mathrm{d}\sigma^2_s \, \mathrm{d}s.$$

Indeed, applying [20, Theorem 3.2.1] we can introduce the solutions $\varphi^1_{n,m}, \varphi^2_{n,m} \in C^{2,1}_{\mathrm{b}}(\mathbb{R}^d \times [s_0, s_1])$ of the backward equations:

$$\partial_s \varphi^j_{n,m} + \mathcal{L}_{n,m}\big[\varphi^j\big] = 0 \quad \text{in } \mathbb{R}^d \times [s_0, s_1], \qquad \varphi^j_{n,m}(\cdot, s_1) = \phi^j(\cdot) \quad \text{in } \mathbb{R}^d.$$

Identity (57) shows that, for $j = 1, 2$,

$$\begin{aligned}\int_{\mathbb{R}^d} \varphi^j_{n,m}(\cdot, s_1)\,\mathrm{d}\sigma^j_{s_1} - \int_{\mathbb{R}^d} \varphi^j_{n,m}(\cdot, s_0)\,\mathrm{d}\sigma^j_{s_0} &= \int_{s_0}^{s_1}\int_{\mathbb{R}^d} (\tilde{A}_{n,m} - \tilde{A}) \cdot \nabla\varphi^j_{n,m}\,\mathrm{d}\sigma^j_s\,\mathrm{d}s \\ &\overset{(70)}{\leqslant} \ell \int_{s_0}^{s_1}\int_{\mathbb{R}^d} |\tilde{A}_{n,m} - \tilde{A}|\,\mathrm{d}\sigma^j_s\,\mathrm{d}s.\end{aligned}$$

Summing up the these equation for $j = 1, 2$ we obtain:

$$\int_{\mathbb{R}^d} \varphi^1\,\mathrm{d}\sigma^1_{s_1} + \int_{\mathbb{R}^d} \varphi^2\,\mathrm{d}\sigma^2_{s_1} \leqslant \int_{\mathbb{R}^d} \varphi^1_{n,m}(\cdot, s_0)\,\mathrm{d}\sigma^1_{s_0} + \int_{\mathbb{R}^d} \varphi^2_{n,m}(\cdot, s_0)\,\mathrm{d}\sigma^2_{s_0} + \ell K_{n,m}. \tag{75}$$

Theorem 3.1 yields $\varphi^1_{n,m}(y_1, s_0) + \varphi^2_{n,m}(y_2, s_0) \leqslant h(|y_1 - y_2|)$ which implies (74).

Step 4:

$$\limsup_{n\uparrow+\infty}\Big(\limsup_{m\uparrow+\infty} K_{n,m}\Big) = 0. \tag{76}$$

Let us first notice that setting $t_i := \mathsf{t}(s_i)$ and recalling that $\mathsf{t}'(s) = \mathrm{e}^{-\lambda \mathsf{t}(s)}$ we have:

$$\int_{s_0}^{s_1}\int_{\mathbb{R}^d} |\tilde{A}_{n,m} - \tilde{A}|\,\mathrm{d}\sigma^1_s\,\mathrm{d}s = \int_{\mathsf{s}(t_0)}^{\mathsf{s}(t_1)} \mathsf{t}'(s) \int_{\mathbb{R}^d} |A_{n,m} - A|\,\mathrm{d}\rho^i_{\mathsf{t}(s)}\,\mathrm{d}s = \int_{t_0}^{t_1}\int_{\mathbb{R}^d} |A_{n,m} - A|\,\mathrm{d}\rho^i_t\,\mathrm{d}t,$$

so that

$$K_{n,m} = K^1_{n,m} + K^2_{n,m}, \qquad K^j_{n,m} := \int_{t_0}^{t_1}\int_{\mathbb{R}^d} |A_{n,m} - A|\,\mathrm{d}\rho^j_t\,\mathrm{d}t, \quad j = 1, 2.$$

We can estimate $K^j_{n,m}$ by

$$K^j_{n,m} \leqslant \int_{t_0}^{t_1}\int_{\mathbb{R}^d} |A_{n,m} - A_n|\,\mathrm{d}\rho^j_t\,\mathrm{d}t + \int_{t_0}^{t_1}\int_{\mathbb{R}^d} |A_n - A|\,\mathrm{d}\rho^j_t\,\mathrm{d}t,$$

observing that by (47), (48), and the Lebesgue Dominated Convergence Theorem we get:

$$\lim_{m\uparrow+\infty} K^j_{n,m} = \int_{t_0}^{t_1}\int_{\mathbb{R}^d} |A_n - A|\,\mathrm{d}\rho^j_t\,\mathrm{d}t.$$

Since $|A_n(x)| \leqslant |\mathsf{A}^\circ(x)| \leqslant |A(x)| = |B(x) - \lambda x|$ for every $x \in \mathbb{R}^d$, the integrability assumption (15), a further application of the Lebesgue Theorem, and (41) yield:

$$\lim_{n\uparrow+\infty}\Big(\lim_{m\uparrow+\infty} K_{n,m}\Big) = \int_{t_0}^{t_1}\int_{\mathbb{R}^d} \big|\mathsf{A}^\circ - A\big|\,\mathrm{d}\rho^j_t\,\mathrm{d}t. \tag{77}$$

This last integrand is 0 if A coincides with the minimal selection of A, in particular when A is continuous. In the general case, the regularity result of [7] shows that $\rho_t^j \ll \mathcal{L}^d$ for $\mathcal{L}^1$ a.e. $t \in (0, +\infty)$ and (38) says that $\mathsf{A}^\circ = A$ $\mathcal{L}^d$-a.e. in $\mathbb{R}^d$; therefore the last integral of (77) vanishes and we get (76).

Step 5: Conclusion.
Thanks to (76), passing to the limit in (74) we obtain

$$\int_{\mathbb{R}^d} \phi^1 \,\mathrm{d}\sigma_{s_1}^1 + \int_{\mathbb{R}^d} \phi^2 \,\mathrm{d}\sigma_{s_1}^2 \leqslant \mathcal{C}_h\big(\sigma_{s_0}^1, \sigma_{s_0}^2\big).$$

Taking the supremum with respect to $\phi^1, \phi^2 \in C_c^\infty(\mathbb{R}^d)$ and recalling Proposition 2.1 we obtain (73).

Remark 4.1. As it appears from the final argument of the previous step 4, in the case when $A = B - \lambda I$ is the minimal selection A° of A (in particular when B is continuous), we do not need to invoke the regularity result of [7] to conclude our proof.

Proof of Corollary 1.2. For (a), it is sufficient to observe that $\mathrm{e}^{\lambda t} \geqslant 1$; this implies $h(r) \leqslant h_{\lambda t}(r)$ and so,

$$\mathcal{C}_h\big(\rho_t^1, \rho_t^2\big) \leqslant \mathcal{C}_{h_{\lambda t}}\big(\rho_t^1, \rho_t^2\big) \overset{(16)}{\leqslant} \mathcal{C}_h\big(\rho_0^1, \rho_0^2\big).$$

Similarly, for (a) and (b),

$$\mathrm{e}^{p\lambda t}\mathcal{C}_h\big(\rho_t^1, \rho_t^2\big) \leqslant \mathcal{C}_{h_{\lambda t}}\big(\rho_t^1, \rho_t^2\big) \overset{(16)}{\leqslant} \mathcal{C}_h\big(\rho_0^1, \rho_0^2\big).$$

We conclude recalling that

$$W_p\big(\rho^1, \rho^2\big) = \mathcal{C}_h\big(\rho^1, \rho^2\big)^{1/p} \quad \text{with } h(r) = |r|^p,$$

and applying (a) and (b). □

References

[1] G. Alberti, L. Ambrosio, A geometrical approach to monotone functions in R^n, Math. Z. 230 (1999) 259–316.
[2] L. Ambrosio, N. Gigli, G. Savaré, Gradient Flows in Metric Spaces and in the Space of Probability Measures, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2005.
[3] L. Ambrosio, G. Savaré, L. Zambotti, Existence and stability for Fokker–Planck equations with log-concave reference measure, Probab. Theory Related Fields 145 (2009) 517–564.
[4] J.D. Benamou, Y. Brenier, A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem, Numer. Math. 84 (2000) 375–393.
[5] V.I. Bogachev, G. Da Prato, M. Röckner, W. Stannat, Uniqueness of solutions to weak parabolic equations for measures, Bull. Lond. Math. Soc. 39 (2007) 631–640.
[6] V.I. Bogachev, G. Da Prato, M. Röckner, On parabolic equations for measures, Comm. Partial Differential Equations 33 (2008) 397–418.
[7] V.I. Bogachev, N.V. Krylov, M. Röckner, On regularity of transition probabilities and invariant measures of singular diffusions under minimal conditions, Comm. Partial Differential Equations 26 (2001) 2037–2080.
[8] H. Brézis, Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert, North-Holland Math. Stud., vol. 5, North-Holland Publishing Co., Amsterdam, 1973. Notas de Matemática (50).
[9] J.A. Carrillo, S. Lisini, G. Savaré, D. Slepcev, Nonlinear mobility continuity equations and generalized displacement convexity, J. Funct. Anal. 258 (2010) 1273–1309.
[10] J.A. Carrillo, R.J. McCann, C. Villani, Contractions in the 2-Wasserstein length space and thermalization of granular media, Arch. Ration. Mech. Anal. 179 (2006) 217–263.
[11] J.A. Carrillo, G. Toscani, Asymptotic L^1-decay of solutions of the porous medium equation to self-similarity, Indiana Univ. Math. J. 49 (2000) 113–142.
[12] S. Daneri, G. Savaré, Eulerian calculus for the displacement convexity in the Wasserstein distance, SIAM J. Math. Anal. 40 (2008) 1104–1122.
[13] S. Fitzpatrick, R. Phelps, Bounded approximants to monotone operators on Banach spaces, Ann. Inst. H. Poincaré Anal. Non Linéaire 9 (1992) 573–595.
[14] R. Jordan, D. Kinderlehrer, F. Otto, The variational formulation of the Fokker–Planck equation, SIAM J. Math. Anal. 29 (1998) 1–17.
[15] R.J. McCann, A convexity principle for interacting gases, Adv. Math. 128 (1997) 153–179.

[16] F. Otto, The geometry of dissipative evolution equations: the porous medium equation, Comm. Partial Differential Equations 26 (2001) 101–174.
[17] F. Otto, M. Westdickenberg, Eulerian calculus for the contraction in the Wasserstein distance, SIAM J. Math. Anal. 37 (2005) 1227–1255.
[18] J.W. Portegies, M.A. Peletier, Well-posedness of a parabolic moving-boundary problem in the setting of Wasserstein gradient flows, arXiv:0812.1269, 2008.
[19] W. Schachermayer, J. Teichmann, Characterization of optimal transport plans for the Monge–Kantorovich problem, Proc. Amer. Math. Soc. 137 (2009) 519–529.
[20] D.W. Stroock, S.R.S. Varadhan, Multidimensional Diffusion Processes, Classics Math., Springer, Berlin, 2006, xii, p. 338 (reprint of the 2nd correted printing (1997)).
[21] C. Villani, Topics in Optimal Transportation, Grad. Stud. Math., vol. 58, American Mathematical Society, Providence, RI, 2003.
[22] C. Villani, Optimal Transport. Old and New, Grundlehren Math. Wiss., vol. 338, Springer-Verlag, Berlin, 2009.