FROM POINCARÉ TO LOGARITHMIC SOBOLEV INEQUALITIES: A GRADIENT FLOW APPROACH

JEAN DOLBEAULT*, BRUNO NAZARET†, AND GIUSEPPE SAVARÉ‡

Abstract. We use the distances introduced in a previous joint paper to exhibit the gradient flow structure of some drift-diffusion equations for a wide class of entropy functionals. Functional inequalities obtained by the comparison of the entropy with the entropy production functional reflect the contraction properties of the flow. Our approach provides a unified framework for the study of the Kolmogorov-Fokker-Planck (KFP) equation.

Key words. Optimal transport, Kantorovich-Rubinstein-Wasserstein distance, Generalized Poincaré inequality, Continuity equation, Action functional, Gradient flows, Kolmogorov-Fokker-Planck equation

AMS subject classifications. 26A51, 26D10, 53D25

1. Setting of the problem. Our starting point concerns nonnegative solutions with finite mass of the heat equation in $\mathbb{R}^d$

$$\partial_t u_t = \Delta u_t \ . \tag{1.1}$$

It is straightforward to check that for any smooth enough solution of (1.1) and any C^2 convex function ψ,

$$\frac{d}{dt}\int_{\mathbb{R}^d} \psi(u_t)\,dx = -\int_{\mathbb{R}^d} \psi''(u_t)\,|\boldsymbol{D}u_t|^2\,dx$$

so that $\int_{\mathbb{R}^d} \psi(u_t)\,dx$ plays the role of a Lyapunov functional. To extract some information out of such an identity, one needs to analyze the relation between $\int_{\mathbb{R}^d} \psi(u_t)\,dx$ and $\int_{\mathbb{R}^d} \psi''(u_t)\,|\boldsymbol{D}u_t|^2\,dx$. This can be done using Green's function or moment estimates, with the drawback that these quantities are explicitly t-dependent. It is simpler to rewrite the equation in self-similar variables and replace (1.1) by the Fokker-Planck (FP) equation

$$\partial_t v_t = \Delta v_t + \nabla\cdot(x\,v) \ . \tag{1.2}$$

This can be done without changing the initial data by the time-dependent change of variables

$$u_t(x) = \frac{1}{R(t)^d}\,v_{\log R(t)}\left(\frac{x}{R(t)}\right), \quad R(t) = \sqrt{1+2t}\ .$$

We shall restrict our approach to nonnegative initial data $u_0 = v_0$. By linearity, we can further assume that

$$\int_{\mathbb{R}^d} v_t\,dx = \int_{\mathbb{R}^d} u_t\,dx = \int_{\mathbb{R}^d} u_0\,dx = 1$$

*Ceremade (UMR CNRS no. 7534), Université Paris-Dauphine, place de Lattre de Tassigny, 75775 Paris Cédex 16, France. E-mail: dolbeaul@ceremade.dauphine.fr

†Ceremade (UMR CNRS no. 7534), Université Paris-Dauphine, place de Lattre de Tassigny, 75775 Paris Cédex 16, France. E-mail: nazaret@ceremade.dauphine.fr

‡Università degli studi di Pavia, Dipartimento di Matematica "F. Casorati", Via Ferrata 1, 27100, Pavia, Italy. E-mail: giuseppe.savare@unipv.it

without loss of generality. We shall also assume that ψ is defined on $\mathbb{R}^+$. Up to the change of ψ into $\tilde\psi$ such that $\tilde\psi(s) = \psi(s) - \psi(1) - \psi'(1)(s-1)$, we can also assume that ψ is nonnegative on $\mathbb{R}^+$ and achieves its minimum value, zero, at $s = 1$.

Eq. (1.2) has a unique nonnegative stationary solution $v = \gamma$ normalized such that $\int_{\mathbb{R}^d} \gamma\, dx = 1$, namely

$$\gamma(x) = \frac{e^{-|x|^2/2}}{(2\pi)^{d/2}} \quad \forall\, x \in \mathbb{R}^d\,.$$

If we introduce $\rho_t = v_t/\gamma$, then ρ_t is a solution of the Ornstein-Uhlenbeck, or Kolmogorov-Fokker-Planck (KFP), equation

$$\partial_t \rho_t = \Delta \rho_t - x \cdot \boldsymbol{D}\rho_t \tag{1.3}$$

with initial data $\rho_0 = v_0/\gamma$. After identifying γ with the measure $\gamma\, \mathscr{L}^d$, the relevant Lyapunov functional, or *entropy*, is $\int_{\mathbb{R}^d} \psi(\rho_t)\, d\gamma$ and

$$\frac{d}{dt} \int_{\mathbb{R}^d} \psi(\rho_t)\, d\gamma = - \int_{\mathbb{R}^d} \psi''(\rho_t)\, |\boldsymbol{D}\rho_t|^2\, d\gamma\,.$$

We shall restrict our study to a class of functions ψ for which the entropy and the entropy production functional are related by the inequality

$$2\lambda \int_{\mathbb{R}^d} \psi(\rho)\, d\gamma \le \int_{\mathbb{R}^d} \psi''(\rho)\, |\boldsymbol{D}\rho|^2\, d\gamma \tag{1.4}$$

for some $\lambda > 0$ (it turns out that in the case of the Gaussian measure we can choose $\lambda = 1$). This allows us to prove that the entropy is exponentially decaying, namely

$$\int_{\mathbb{R}^d} \psi(\rho_t)\, d\gamma \le \left(\int_{\mathbb{R}^d} \psi(\rho_0)\, d\gamma \right)\, e^{-2\lambda t} \quad \forall\, t \ge 0\,, \tag{1.5}$$

if ρ_t is a solution of (1.3) and if λ is positive. A sufficient condition for such an inequality is that

$$\textit{the function } h := 1/\psi'' \textit{ is concave} \tag{1.6}$$

(see for instance [11, p. 201], [8, (2.15)]). At first sight, this may look like a technical condition but it has some deep implications. We are indeed interested in exhibiting a gradient flow structure for (1.2) associated with the entropy or, to be more precise, to establish that, for some distance, the gradient flow of the entropy is actually (1.2). It turns out that (1.6) is the natural condition as we shall see in Section 3.2.

The entropy decays exponentially according to (1.5) not only when one considers the $L^2_\gamma(\mathbb{R}^d)$ norm (the norm of the square integrable functions with respect to the Gaussian measure γ), i.e. the case $\psi(\rho) = (\rho-1)^2/2$, or the classical entropy built on $\psi(\rho) = \rho \log \rho$, for which (1.3) is the gradient flow with respect to the usual Wasserstein distance (according to the seminal paper [27] of Jordan, Kinderlehrer and Otto). We also have an exponential decay result of any entropy generated by

$$\psi(\rho) = \frac{\rho^{2-\alpha} - 1 - (2-\alpha)(\rho-1)}{(2-\alpha)(1-\alpha)} =: \psi_\alpha(\rho)\,, \quad \alpha \in [0,1)\,,$$

and more generally any ψ satisfying (1.6). Notice by the way that $\psi(\rho) = \psi_\alpha(\rho)$ is compatible with (1.6) if and only if $\alpha \in [0, 1)$ and that $\psi(\rho) = \rho \log \rho$ appears as the limit case when $\alpha \to 1_-$.

The exponential decay is a striking property which raises the issue of the hidden mathematical structure, a question asked long ago by F. Poupaud. As already mentionned, the answer lies in the gradient flow interpretation and the construction of the appropriate distances. Such distances, based on an action functional related to ψ, have been studied in [24]. Our purpose is to exploit this action functional for the construction of gradient flows, not only in the case corresponding to (1.3) but also for KFP equations based on general λ-convex potentials V. For the convenience of the reader, the main steps of the strategy have been collected in Section 2, without technical details (for instance on the measure theoretic aspects of our approach).

Coming back to our basic example, namely the solution of (1.3), we may observe that a solution can easily be represented using the Green kernel of the heat equation and our time-dependent change of variables. If $\psi(\rho) = \psi_\alpha(\rho)$, $\alpha \in [0, 1)$, we may observe that the exponential decay of the entropy can be obtained using the known properties of the heat flow and the homogeneity of ψ_α, while the contraction properties of the heat flow measured in the framework of the weighted Wasserstein distances introduced in [24] can be translated into the exponential decay of the distance of the solution of (1.3) to the gaussian measure γ, if we assume that $\rho\,\gamma$ is a probability measure. We shall however not pursue in this direction as it is very specific of the potential $V(x) = \frac12\,|x|^2$ and of the heat flow (for which an explicit Green function is available).

Functionals based on entropy densities ψ satisfying (1.6) are sometimes called *φ-entropies*. In this paper, we shall however avoid this denomination to prevent from possible confusions with the function ϕ and the functional Φ used below to define the action and the weighted Wasserstein distances W_h.

We shall refer to [19, 28] for a probabilistic point of view. A proof of (1.5) under Assumption (1.6) and the *Bakry-Emery condition* can be found for instance in [8] or in the more recent paper [17]. This approach is based on the Bakry-Emery method [11, 21] and heavily relies on the flow of KFP or, equivalently, on the geometric properties of the Ornstein-Uhlenbeck operator (using the *carré du champ:* see [17]). Strict convexity of the potential is usually required, but can be removed afterwards by various methods: see [8, 12, 23]. For capacity-measure approaches of (1.4), we shall refer to [13, 14, 22]. The inequality (1.4) itself has been introduced in [15] with a proof based on the hypercontractivity of the heat flow and spectral estimates, and later refined and adapted to general potentials in [6].

Concerning gradient flows and distances of Wasserstein type, there has been a huge activity over the last years. We can refer to [27, 16] for fundamental ideas, and to two books, [2, 31], for a large overview of the field. Many other contributions in this area will be quoted whenever needed in the proofs.

The key of our approach is based on the definition of the *h-Wasserstein distance* W_h. For any entropy density ψ satisfying (1.6), we build an action functional on the space of probability measures in the spirit of [24] using $h = 1/\psi''$, which provides the distance W_h and generalizes the Wasserstein distance.

The core of the paper is a decay estimate of the action functional (Theorem 2.1). The analysis of the equality case in (2.15) allows to identify the KFP flow as the gradient flow of the entropy functional w.r.t. W_h in (2.16). Our two main results are

Theorems 7.1 and 7.2.

In this paper, our goal is to clarify the relationship between the flow of the Fokker-Planck equation and the family of inequalities (1.4). The main observation is that any estimate on the action functional can be converted into an estimate on distances, an *Eulerian* point of view (as in [30]), which is a major advantage compared to previous approaches (based on *Lagrangian* formulations). It turns out that the Fokker-Planck equation admits a nice metric formulation in terms of an Evolution Variational Inequality involving W_h, that has interesting regularizing and stability properties [5] and could be extended to more general, non-smooth settings, as the metric-measure spaces with Ricci curvature bounded from below [31, 3].

The *flat case*, corresponding to a constant potential V and to the reference Lebesgue measure $\gamma = \mathscr{L}^d$ (so that $\lambda = 0$), has already been considered in [24] and further generalized and applied to nonlinear diffusion equation in [18, 29]. In the present paper, we deal with general Fokker-Planck equations.

Section 2 is devoted to formal computations and heuristics. Considerations on the distance and on the entropy functional can be found in Sections 3 and 4 respectively. The KFP flow and its first variation are studied in Section 5. A stronger version of Theorem 2.1 is established in Corollary 6.2. Main results can be found in Section 7.

2. Formal point of view: definitions, strategy and main results. In Section 1, we have considered the case of the harmonic potential $V(x) = \frac{1}{2}|x|^2$. We generalize the setting to any smooth, convex *potential* $V : \mathbb{R}^d \to \mathbb{R}$ with

$$D^2 V \geq \lambda \, \mathsf{I} \,, \quad \lambda \in \mathbb{R} \,, \tag{2.1}$$

known as the *Bakry-Emery condition* or $CD(\lambda, \infty)$ condition, and consider the *reference measure* γ given by

$$\gamma := e^{-V} \, \mathscr{L}^d \tag{2.2}$$

where $\mathscr{L}^d$ denotes Lebesgue measure on $\mathbb{R}^d$. We assume that

$$\gamma(\mathbb{R}^d) = \int_{\mathbb{R}^d} e^{-V} \, dx =: Z < \infty \,. \tag{2.3}$$

Next we define the *action density* $\phi : (0, \infty) \times \mathbb{R}^d \to \mathbb{R}$ as

$$\phi(\rho, \boldsymbol{w}) := \frac{|\boldsymbol{w}|^2}{h(\rho)}$$

for some concave, positive, non decreasing function h with *sublinear* growth. Our main example is $h(\rho) := \rho^\alpha$ for some $\alpha \in (0, 1)$.

Sometimes it will be more convenient to express some properties of ϕ in terms of the function $g(\rho) := 1/h(\rho)$: notice that h is concave if and only if

$$g(\rho) = \frac{1}{h(\rho)} \quad \text{satisfies} \quad 2\,(g')^2 \leq g\, g'' \,, \tag{2.4}$$

in particular g is convex (compare with [8, (2.12c)]).

Based on the action density, we can define the *action functional* by

$$\Phi(\rho, \boldsymbol{w}) := \int_{\mathbb{R}^d} \phi(\rho, \boldsymbol{w}) \, d\gamma \,. \tag{2.5}$$

The Kolmogorov-Fokker-Planck (KFP) equation. With the notations $\Delta_\gamma := \Delta - \boldsymbol{D}V \cdot \boldsymbol{D}$, the equation

$$\partial_t \rho_t - \Delta_\gamma \rho_t = 0 \tag{2.6}$$

determines the *Kolmogorov-Fokker-Planck (KFP) flow* $S_t : \rho_0 \mapsto \rho_t$. Its first variation, $\boldsymbol{R}_t : \boldsymbol{w}_0 \mapsto \boldsymbol{w}_t$, can be obtained as the solution of the *modified Kolmogorov-Fokker-Planck equation*

$$\partial_t \boldsymbol{w}_t - \Delta_\gamma \boldsymbol{w}_t + D^2 V \, \boldsymbol{w}_t = 0 \, .$$

If $\boldsymbol{w}_0 = \boldsymbol{D}\rho_0$, then $\boldsymbol{w}_t = \boldsymbol{D}\rho_t$, which can be summarized by

$$\boldsymbol{D}(S_t \rho_0) = \boldsymbol{R}_t(\boldsymbol{D}\rho_0) \, .$$

By duality, using the notations $\nabla_\gamma \cdot \boldsymbol{w} := \nabla \cdot \boldsymbol{w} - \boldsymbol{D}V \cdot \boldsymbol{w}$ and $\nabla \cdot \boldsymbol{w} := \sum_{i=1}^d \partial \boldsymbol{w}_i / \partial x_i$, if $\nabla_\gamma \cdot \boldsymbol{w}_0 = \rho_0$, we also find that $\nabla_\gamma \cdot \boldsymbol{w}_t = \rho_t$, which amounts to

$$\nabla_\gamma \cdot (\boldsymbol{R}_t \boldsymbol{w}_0) = S_t(\nabla_\gamma \cdot \boldsymbol{w}_0) \tag{2.7}$$

(see Theorem 5.4 for details). If $\mu = \rho \, \gamma$, we define the semigroup $\mathcal{S}_t$ acting on measures by $\mathcal{S}_t \mu := (S_t \rho) \, \gamma$.

The main estimate on which our method is based goes as follows.

THEOREM 2.1. *Under Assumptions* (2.1)–(2.4), *if* $\Phi(\rho_0, \boldsymbol{w}_0) < \infty$, $\rho_t = S_t \rho_0$ *and* $\boldsymbol{w}_t = \boldsymbol{R}_t \boldsymbol{w}_0$, *then*

$$\frac{d}{dt} \Phi(\rho_t, \boldsymbol{w}_t) + 2\lambda \, \Phi(\rho_t, \boldsymbol{w}_t) \leq 0 \quad \forall \, t \geq 0 \, .$$

In particular

$$\Phi(\rho_t, \boldsymbol{w}_t) \leq e^{-2\lambda t} \, \Phi(\rho_0, \boldsymbol{w}_0) \quad \forall \, t \geq 0 \, , \tag{2.8}$$

and the action functional decays exponentially if λ is positive. At formal level, this follows by an easy convexity argument. The rigorous proof requires many regularizations. See Theorem 6.1 for a more detailed version of this result. Now let us review some of the consequences of Theorem 2.1.

Entropy, entropy production and generalized Poincaré inequalities. Consider an *entropy density* function ψ such that $\psi(1) = \psi'(1) = 0$. If we define the *entropy* functional by

$$\Psi(\rho) := \int_{\mathbb{R}^d} \psi(\rho) \, d\gamma$$

and the *entropy production*, or generalized Fisher information functional, as the action functional for the particular choice $\boldsymbol{w} = \boldsymbol{D}\rho$, i.e.

$$P(\rho) := \Phi(\rho, \boldsymbol{D}\rho) \, ,$$

then, along the KFP flow, we get

$$\frac{d}{dt} \Psi(\rho_t) = -P(\rho_t) = -\Phi(\rho_t, \boldsymbol{D}\rho_t) \tag{2.9}$$

for a solution ρ_t of (2.6) if

$$\psi'' = \frac{1}{h} \, .$$

Notice that (1.6) and (2.4) are equivalent, so the assumption (1.6) on the entropy corresponds to the convexity of the action functional. See Section 3.2 for more details.

We can now apply Theorem 2.1 to the KFP flow. With $\boldsymbol{w} = \boldsymbol{D}\rho$, we find that the *entropy production* functional satisfies:

$$\frac{d}{dt} P(\rho_t) + 2\lambda \, P(\rho_t) \le 0 \, , \quad P(\rho_t) \le e^{-2\lambda t} P(\rho_0) \quad \forall \, t \ge 0 \, . \tag{2.10}$$

If λ is positive, $P(\rho_t)$ decays exponentially; by integrating (2.10) along the KFP flow when t varies in $\mathbb{R}^+$, using (2.9) and $\Psi(1) = 0$, we recover for $\rho = \rho_0$ the *generalized Poincaré inequalities*

$$\Psi(\rho) \le \frac{1}{2\lambda} \, P(\rho) \quad \text{when } \lambda > 0, \tag{2.11}$$

found by Beckner in [15] in the case of the harmonic potential and for $h(\rho) := \rho^\alpha$, $\alpha \in (0, 1)$, and generalized for instance in [8]. Such inequalities interpolate between Poincaré and logarithmic Sobolev inequalities. Inequalities (2.10) and (2.11) are not new and we refer the interested reader to [10, 11], [8, Ineq. (2.51)], [19, 9]. See Theorem 6.3 for details.

If $\lambda \ge 0$ and we combine (2.11) with (2.10), we find that the *entropy* decays according to

$$\frac{d}{dt} \Psi(\rho_t) + 2\lambda \Psi(\rho_t) \le 0 \, , \quad \Psi(\rho_t) \le e^{-2\lambda t} \, \Psi(\rho_0) \quad \forall \, t \ge 0 \, .$$

Setting $e_{2\lambda}(t) := \int_0^t e^{2\lambda r} \, dr$ and integrating from 0 to t the inequality

$$\begin{aligned} \frac{d}{dt} \Big(e_{2\lambda}(t) \, P(\rho_t) \Big) &= e^{2\lambda t} P(\rho_t) + e_{2\lambda}(t) \, \frac{d}{dt} P(\rho_t) \le e^{2\lambda t} P(\rho_t) - 2\lambda e_{2\lambda}(t) \, P(\rho_t) \\ &\le P(\rho_t) = -\frac{d}{dt} \Psi(\rho_t) \, , \end{aligned} \tag{2.12}$$

which itself follows from (2.9) and (2.10), we observe a *first regularization effect* along the KFP flow, namely

$$e_{2\lambda}(t) \, P(\rho_t) \le \Psi(\rho_0) \quad \forall \, t \ge 0 \, . \tag{2.13}$$

The h-Wasserstein distance. If μ is a measure with absolutely continuous part ρ with respect to γ, and singular part $\mu^\perp$, if $\boldsymbol{\nu}$ is a vector valued measure which is absolutely continuous with respect to γ and has a modulus of continuity $\boldsymbol{w}$, i.e. if

$$\mu = \rho \, \gamma + \mu^\perp \quad \text{and} \quad \boldsymbol{\nu} = \boldsymbol{w} \, \gamma \, , \tag{2.14}$$

we can extend the action functional Φ to the measures μ and $\boldsymbol{\nu}$ by setting

$$\Phi(\mu, \boldsymbol{\nu}) = \Phi(\rho, \boldsymbol{w}) = \int_{\mathbb{R}^d} \phi(\rho, \boldsymbol{w}) \, d\gamma \, .$$

We shall say that there is an *admissible path* connecting μ_0 to μ_1 if there is a solution $(\mu_s, \boldsymbol{\nu}_s)_{s\in[0,1]}$ to the *continuity equation*

$$\partial_s \mu_s + \nabla \cdot \boldsymbol{\nu}_s = 0 \,, \quad s \in [0,1] \,,$$

and will denote by $\Gamma(\mu_0, \mu_1)$ the set of all admissible paths. In terms of densities, that is if $\mu_s = \rho_s \gamma$ and $\boldsymbol{\nu}_s = \boldsymbol{w}_s \gamma$, the continuity equations reads

$$\partial_s \rho_s + \nabla_\gamma \cdot \boldsymbol{w}_s = 0 \,.$$

See [24] for more details. With these tools in hands, we can define the *h-Wasserstein distance* between μ_0 and μ_1 by

$$W_h^2(\mu_0, \mu_1) := \inf \Big\{ \int_0^1 \Phi(\mu_s, \boldsymbol{\nu}_s)\, ds \,:\, (\mu, \boldsymbol{\nu}) \in \Gamma(\mu_0, \mu_1) \Big\} \,.$$

Notice that h in "h-Wasserstein distance" refers to the dependence of Φ in h through the action density ϕ, the usual Wasserstein distance corresponding to $h(\rho) = \rho$. Notice that the above formula defines the *square* of the distance $W_h(\mu_0, \mu_1)$. If $(\mu_t)_{t\in(0,T)}$ is a curve of measures, its *h-Wasserstein velocity* $|\dot\mu_t|$ is determined by

$$|\dot\mu_t|^2 = \inf_{\boldsymbol{\nu}} \Big\{ \Phi(\mu, \boldsymbol{\nu}) \,:\, \nabla \cdot \boldsymbol{\nu} = -\,\partial_t \mu_t \Big\} \,.$$

Using the decomposition (2.14), we compute the *derivative of the entropy along an arbitrary curve* $(\mu_t)_{t\in(0,T)}$ as

$$\frac{d}{dt}\Psi(\rho_t) = \int_{\mathbb{R}^d} \psi'(\rho_t)\, \partial_t \rho_t \, d\gamma = \int_{\mathbb{R}^d} \psi''(\rho_t)\, \boldsymbol{D}\rho_t \cdot \boldsymbol{w}_t \, d\gamma \,,$$

for any $(\boldsymbol{w}_t)_{t\in(0,T)}$ such that $(\rho_t\,\gamma, \boldsymbol{w}_t\,\gamma)_{t\in(0,T)}$ is an admissible curve, and find that

$$-\frac{d}{dt}\Psi(\rho_t) = -\int_{\mathbb{R}^d} \sqrt{\psi''(\rho_t)}\, \boldsymbol{D}\rho_t \cdot \sqrt{\psi''(\rho_t)}\, \boldsymbol{w}_t \, d\gamma \le \sqrt{P(\rho_t)} \sqrt{\textstyle\int_{\mathbb{R}^d} \psi''(\rho_t)\, |\boldsymbol{w}_t|^2 \, d\gamma}$$

by the Cauchy-Schwarz inequality. By taking the optimal case in the infimum w.r.t. $(\boldsymbol{w}_t)_{t\in(0,T)}$, we find

$$-\frac{d}{dt}\Psi(\rho_t) \le \sqrt{P(\rho_t)}\, |\dot\mu_t| \,. \tag{2.15}$$

Along the KFP flow, we know that $\boldsymbol{w}_t = \nabla\rho_t$ is such that $(\rho_t\,\gamma, \boldsymbol{w}_t\,\gamma)_{t\in(0,T)}$ is an admissible curve, thus showing that

$$-\frac{d}{dt}\Psi(\rho_t) = P(\rho_t) \ge |\dot\mu_t|^2 \,.$$

By (2.15), we also know that

$$-\frac{d}{dt}\Psi(\rho_t) = P(\rho_t) \le \sqrt{P(\rho_t)}\, |\dot\mu_t| \,,$$

which proves that $P(\rho_t) \le |\dot\mu_t|^2$. Altogether, we have found that

$$\frac{d}{dt}\Psi(\rho_t) = -P(\rho_t) = -|\dot\mu_t|^2 = -\sqrt{P(\rho_t)}\, |\dot\mu_t| \,, \tag{2.16}$$

which is the equality case in (2.15). This characterizes the KFP flow as the steepest descent flow of the entropy Ψ, i.e. this is a first charaterization of KFP as *the gradient flow of* Ψ with respect to the h-Wasserstein distance.

The KFP flow connects $\mu = \rho\,\gamma$ with $\mu_\infty = \gamma$ and it has been established in [24] that one can estimate the length of the path by

$$W_h(\mu,\gamma) = \int_0^\infty \sqrt{P(\rho_t)}\,dt = \int_0^\infty |\dot\mu_t|\,dt \tag{2.17}$$

(see Section 3.5 for details). According to (2.10), when $\lambda > 0$ we get

$$W_h(\mu,\gamma) \le \sqrt{P(\rho)} \int_0^\infty e^{-\lambda t}\,dt = \frac{1}{\lambda}\,\sqrt{P(\rho)}\,.$$

This establishes the *entropy production – distance estimate*

$$W_h(\mu,\gamma) \le \frac{1}{\lambda}\,\sqrt{P(\rho)}\,, \quad \text{if} \quad \mu = \rho\,\gamma\,.$$

Along the KFP flow, we also find that

$$-\frac{d}{dt}\sqrt{\Psi(\rho_t)} = \frac{P(\rho_t)}{2\sqrt{\Psi(\rho_t)}} \ge \sqrt{\frac{\lambda}{2}\,P(\rho_t)}$$

using (2.11). By applying (2.17), this establishes the *(Talagrand) entropy – distance estimate*

$$W_h^2(\mu,\gamma) \le \frac{2}{\lambda}\,\Psi(\rho)\,.$$

Contraction properties and gradient flow structure. Here as in [24], we use the technique introduced in [30] and extended in [20, § 2]: we consider a geodesic (or an approximation of a geodesic), and evaluate the derivative of the action functional along a family of curves obtained by evolving the geodesic with the KFP flow. For simplicity, we will use the word *contraction* even in the case $\lambda \le 0$ and specify when the assumption $\lambda > 0$ is required (also in Section 7).

Consider an ε-geodesic $(\rho^s, \boldsymbol{w}^s)$ connecting $\mu^0 = \rho^0\,\gamma$ to $\mu^1 = \rho^1\,\gamma$, i.e. an admissible path in $\Gamma(\mu_0,\mu_1)$ such that $\Phi(\rho_0^s, \boldsymbol{w}_0^s) \le W_h^2(\rho_0^0,\rho_0^1) + \varepsilon$ for any $s \in (0,1)$ and observe that by (2.7), we know that $(\rho_t^s = S_t\rho^s, \boldsymbol{w}_t^s = \boldsymbol{R}_t\boldsymbol{w}^s)$ is still an admissible curve connecting $S_t\rho^0$ to $S_t\rho^1$. Therefore (2.8) yields

$$W_h^2(\rho_t^0,\rho_t^1) \le \int_0^1 \Phi(\rho_t^s,\boldsymbol{w}_t^s)\,ds \le e^{-2\lambda t}\int_0^1 \Phi(\rho_0^s,\boldsymbol{w}_0^s)\,ds \le e^{-2\lambda t}\left(W_h^2(\rho_0^0,\rho_0^1)+\varepsilon\right)\,,$$

which, by letting $\varepsilon \to 0$, proves that the KFP flow *contracts the distance:*

$$W_h(\mathcal{S}_t\mu^0, \mathcal{S}_t\mu^1) \le e^{-\lambda t}\,W_h(\mu^0,\mu^1) \quad \forall\, t \ge 0\,. \tag{2.18}$$

See Theorem 7.1 for more details. This is our first main result. In case of the Wasserstein distance, (2.18) is known to be equivalent to the *Bakry-Emery condition* (2.1); see [32, 33].

Next, we should again consider an ε-geodesic, but for simplicity we assume that there is a geodesic $(\rho^s,\boldsymbol{w}^s)$ connecting $\sigma = \mu^0 = \rho^0\,\gamma$ to $\mu = \mu^1 = \rho^1\,\gamma$, i.e. such that $\Phi(\rho^s,\boldsymbol{w}^s) = W_h^2(\sigma,\mu)$, and consider the path

$$(\rho_t^s, \boldsymbol{w}_t^s) := (S_{st}\rho^s, \boldsymbol{R}_{st}\boldsymbol{w}^s + t\,\boldsymbol{D}\rho_t^s)$$

connecting σ to $\mu_t := \mathcal{S}_t\mu$. Notice that our notations mean that $\rho^s = \rho_0^s$. Since

$$\partial_s \rho_t^s = \rho_t^s + t\,\Delta_\gamma \rho_t^s = \nabla_\gamma \cdot (\boldsymbol{w}_t^s + t\,\boldsymbol{D}\rho_t^s)\ ,$$

the path is admissible and, as a consequence,

$$W_h^2(\mu_t, \sigma) \le \int_0^1 \Phi(\rho_t^s, \boldsymbol{w}_t^s)\, ds\ .$$

We can therefore differentiate the right hand side in the above inequality instead of the distance and furthermore notice that it is sufficient to do it at $t = 0$; see Theorem 7.2 and its proof for details. Along the KFP flow we find that

$$\frac{1}{2}\,\frac{d}{dt} W_h^2(\mu_t, \sigma) + \frac{\lambda}{2}\, W_h^2(\mu_t, \sigma) \le \Psi(\sigma\,|\,\gamma) - \Psi(\mu_t\,|\,\gamma)\ , \tag{2.19}$$

which is our second main result. (2.19) provides the strongest formulation of a λ-*contracting gradient flow* in a metric setting as a solution of an *Evolution Variational Inequality* [2, Ch. 4, 11] and it is related to the *above-tangent lemma* for displacement convexity in the framework of the usual Wasserstein distance. Here we have defined the relative entropy as $\Psi(\mu\,|\,\gamma) := \psi(\rho)$ if $\mu \ll \gamma$ and $\mu = \rho\,\gamma$, and $\Psi(\sigma\,|\,\gamma) := +\infty$ otherwise. Hence we recover a second characterization of the fact that KFP *is the gradient flow of* Ψ with respect to W_h.

As another consequence, the entropy Ψ *is geodesically* λ*-convex*. This follows from (2.19). Fix a geodesic μ^s between μ^0 and μ^1, follow the evolution of μ^s by KFP taking first μ^0 and then μ^1 fixed, and apply (2.19) with $\mu_t := \mathcal{S}_t\mu^s$ and $\mu = \mu^0$ or $\mu = \mu^1$. Because of the minimality of the energy along the geodesic at time $t = 0$, by summing the two resulting inequalities we prove the convexity inequality of Ψ. See [20, Theorem 3.2] for more details.

Notice that the results of Theorems 2.1, 7.1 and 7.2 are still valid for negative values of λ, as well as the characterization (2.19) of KFP as a gradient flow.

As a final observation, let us notice that, directly from the metric formulation (2.19), it follows that the KFP flow also has the following *regularizing properties* (that we state in the case $\lambda \ge 0$ for simplicity)

$$\Psi(\rho_t) \le \frac{1}{2t}\, W_h^2(\rho_0, \gamma) \quad \text{and} \quad P(\rho_t) \le \frac{1}{t^2}\, W_h^2(\rho_0, \gamma) \quad \forall\, t \ge 0\ .$$

The first estimate can indeed be obtained by integrating (2.19) (with $\lambda = 0$ and $\sigma = \gamma$) from 0 to t and recalling that $t \mapsto \Psi(\rho_t)$ is decreasing. As for the second one, we observe that also $t \mapsto P(\rho_t)$ is decreasing by (2.10), so that (2.13) and (2.19) yield

$$\frac{d}{dt}\left(\frac{t^2}{2}\, P(\rho_t)\right) \le t\, P(\rho_t) \le \Psi(\rho_t) \le -\frac{1}{2}\,\frac{d}{dt} W_h^2(\mu_t, \gamma)\ .$$

A further integration in time from 0 to t completes the proof. Notice that it is crucial to start from a measure $\mu = \rho_0\,\gamma$ at *finite* distance from γ.

3. Definition and properties of the weighted Wasserstein distance. In this section we first recall some definitions and results taken from [24]. The measure γ and the functions ϕ and ψ are as in Section 2, and we assume that Conditions (2.1)–(2.4) are satisfied.

3.1. Properties of the potential. Let $V : \mathbb{R}^d \to \mathbb{R}$ be a λ-convex and continuous potential. λ-convexity means that the map $x \mapsto V(x) - \frac{\lambda}{2}|x|^2$ is convex. When V is smooth in $\mathbb{R}^d$, this condition is equivalent to (2.1). We are assuming that e^{-V} is integrable in $\mathbb{R}^d$, so that we can introduce the finite and positive measure γ defined by (2.2). For simplicity, we shall assume that γ is a probability measure, i.e. $Z = 1$, which can always be enforced by replacing V by $V + \log Z$.

When $\lambda \geq 0$ the potential V is convex and γ is log-concave; the integrability of e^{-V} is equivalent to the property that $V(x) \uparrow \infty$ at least linearly as $|x| \uparrow \infty$; see e.g. [5, Appendix]. As a consequence, there exist two constants $A > 0$, $B \geq 0$ such that

$$V(x) \geq A\,|x| - B \quad \forall\, x \in \mathbb{R}^d\,. \tag{3.1}$$

We recall that non smooth, λ-convex potentials V can be approximated from below by an increasing sequence of λ-convex potentials V_n:

$$V_n(x) := \frac{\lambda}{2}\,|x|^2 + \inf_{y \in \mathbb{R}^d} \Big(\frac{n}{2}\,|x-y|^2 + V(y) - \frac{\lambda}{2}\,|y|^2 \Big)\,.$$

Moreover, the potentials V_n are λ-convex and in the case $\lambda \geq 0$, they satisfy conditions (3.1) with respect to constants A and B which are *independent of* n. In particular, the log-concave measures $\gamma_n := e^{-V_n} \mathscr{L}^d$ weakly* and monotonically converge in $C_b^0(\mathbb{R}^d)'$ to γ.

By this regularization technique, many results could be extended to the case when V is just lower semicontinuous and can take the value $+\infty$.

3.2. Convexity of the action density. As in Section 2, consider h on $(0,\infty)$ such that $\phi(\rho, \boldsymbol{w}) = |\boldsymbol{w}|^2/h(\rho)$. The following result has already been observed in [24] but we reproduce it here for completeness.

LEMMA 3.1 (Convexity of the action density). *With the notations of Section 2, the action density ϕ is* convex *if and only if h is concave on $(0,\infty)$ or, equivalently, if $g := \frac{1}{h}$ satisfies Condition* (2.4).

Proof. By standard approximations, it is not restrictive to assume that g, $h \in C^2(0,\infty)$. First of all observe that

$$g^3\,h'' = 2\,(g')^2 - g\,g''\,,$$

so that h'' is nonpositive if and only if $2\,(g')^2 \leq g\,g''$. Next we evaluate the second derivative of ϕ along the direction of the vector $\boldsymbol{z} = (x, \boldsymbol{y}) \in \mathbb{R} \times \mathbb{R}^d$ as

$$\big\langle \mathrm{D}^2\phi(\rho, \boldsymbol{w})\,\boldsymbol{z}, \boldsymbol{z} \big\rangle = g''(\rho)\,|\boldsymbol{w}|^2\,x^2 + 4\,g'(\rho)\,\boldsymbol{w}\cdot x\,\boldsymbol{y} + 2\,g(\rho)\,|\boldsymbol{y}|^2\,.$$

By minimizing with respect to $x \in \mathbb{R}$, we get

$$g''(\rho)\,|\boldsymbol{w}|^2\,\big\langle \mathrm{D}^2\phi(\rho, \boldsymbol{w})\,\boldsymbol{z}, \boldsymbol{z} \big\rangle \geq 2\,\big[g''(\rho)\,|\boldsymbol{w}|^2\,g(\rho)\,|\boldsymbol{y}|^2 - 2\,(g'(\rho)\,\boldsymbol{w}\cdot\boldsymbol{y})^2\big] \tag{3.2}$$

if $g''(\rho) > 0$, with equality for the appropriate choice of x. The convexity of ϕ is thus equivalent to

$$g''(\rho)\,|\boldsymbol{w}|^2\,g(\rho)\,|\boldsymbol{y}|^2 \geq 2\,(g'(\rho)\,\boldsymbol{w}\cdot\boldsymbol{y})^2 \quad \forall\,\rho > 0\,, \quad \forall\,\boldsymbol{y}\,,\ \boldsymbol{w} \in \mathbb{R}^d\,.$$

If ϕ is convex, by choosing $\boldsymbol{y} := h(\rho)\,g'(\rho)\,\boldsymbol{w}$ and using $h(\rho)\,g(\rho) = 1$, we get

$$g''(\rho)\,|\boldsymbol{w}|^2\,h(\rho)\,(g'(\rho))^2\,|\boldsymbol{w}|^2 \geq 2\,\big[h(\rho)\,(g'(\rho))^2\,|\boldsymbol{w}|^2\big]^2 \quad \forall\,\rho > 0\,, \quad \forall\,\boldsymbol{w} \in \mathbb{R}^d\,,$$

which yields (2.4). Conversely, the convexity of ϕ follows from $(\boldsymbol{w}\cdot\boldsymbol{y})^2 \le |\boldsymbol{w}|^2\,|\boldsymbol{y}|^2$. □

REMARK 3.2 (Main example). *Our main example is provided by the function*

$$h(\rho) := \rho^\alpha\,, \quad 0 \le \alpha \le 1\,, \quad \phi(\rho,\boldsymbol{w}) = \frac{|\boldsymbol{w}|^2}{\rho^\alpha}\,,$$

which satisfies (6.4). *When* $\alpha = 0$ *we simply get*

$$\phi(\rho,\boldsymbol{w}) := |\boldsymbol{w}|^2\,,$$

and for $\alpha = 1$ *we have the* 1*-homogeneous functional*

$$\phi(\rho,\boldsymbol{w}) := \frac{|\boldsymbol{w}|^2}{\rho}\,.$$

Notice that the above considerations can be generalized to matrix-valued functions g and h: see [24, Example 3.4].

3.3. The action functional on densities. The *action functional* Φ induced by ϕ has been defined by (2.5), with domain

$$\mathcal{D}(\Phi) := \Big\{(\rho,\boldsymbol{w}) \in L^1_\gamma(\mathbb{R}^d)\times L^1_\gamma(\mathbb{R}^d;\mathbb{R}^d)\,:\,\rho\ge 0\,,\ \Phi(\rho,\boldsymbol{w})<\infty\Big\}\,.$$

Assuming as in Section 3.2 that ϕ convex, it is well known that if $(\rho_k)_{k\in\mathbb{N}}$ and $(\boldsymbol{w}_k)_{k\in\mathbb{N}}$ are such that $(\rho_k,\boldsymbol{w}_k)\in\mathcal{D}(\Phi)$ for any $k\in\mathbb{N}$ and if $\rho_k \rightharpoonup \rho$ in $L^1_\gamma(\mathbb{R}^d)$, and $\boldsymbol{w}_k \rightharpoonup^* \boldsymbol{w} \in L^1_\gamma(\mathbb{R}^d;\mathbb{R}^d)$ as $n\uparrow\infty$, then by lower semi-continuity of Φ, we have

$$\liminf_{n\uparrow\infty} \Phi(\rho_k,\boldsymbol{w}_k) \ge \Phi(\rho,\boldsymbol{w})\,.$$

LEMMA 3.3 (Approximation by smooth bounded densities). *Consider two functions* $\rho\in L^1_\gamma(\mathbb{R}^d)$ *and* $\boldsymbol{w}\in L^1_\gamma(\mathbb{R}^d;\mathbb{R}^d)$ *such that* $\rho\ge 0$ *and* $\Phi(\rho,\boldsymbol{w})<\infty$. *Then there exist two sequences* $(\rho_k)_{k\in\mathbb{N}}$ *and* $(\boldsymbol{w}_k)_{k\in\mathbb{N}}$ *of bounded smooth functions (with bounded derivatives of arbitrary orders) such that* $\inf_{\mathbb{R}^d}\rho_k > 0$ *and*

$$\lim_{k\uparrow\infty}\rho_k = \rho \quad \textit{in } L^1_\gamma(\mathbb{R}^d)\,, \quad \lim_{k\uparrow\infty}\boldsymbol{w}_k = \boldsymbol{w} \quad \textit{in } L^1_\gamma(\mathbb{R}^d;\mathbb{R}^d)\,,$$
$$\int_{\mathbb{R}^d}\rho_k\,d\gamma = \int_{\mathbb{R}^d}\rho\,d\gamma \quad \forall\,k\in\mathbb{N} \quad \textit{and} \quad \lim_{k\uparrow\infty}\int_{\mathbb{R}^d}\phi(\rho_k,\boldsymbol{w}_k)\,d\gamma = \int_{\mathbb{R}^d}\phi(\rho,\boldsymbol{w})\,d\gamma\,.$$

Proof. We first truncate ρ and $\boldsymbol{w}$ from above as follows. Let $m := \int_{\mathbb{R}^d}\rho\,d\gamma$ and, for any $k\in\mathbb{N}$, $m_k := \int_{\mathbb{R}^d}(\rho\wedge k)\,d\gamma$, $R_k := \{x\in\mathbb{R}^d\,:\,\rho(x)\le k\}$. We set $\rho_k := m_k^{-1}\,m\,(\rho\wedge k)$ and

$$w_k(x) := \begin{cases} w(x) & \text{if } |w(x)|\le k \text{ and } x\in R_k\,,\\ 0 & \text{otherwise}\,. \end{cases}$$

Clearly $\rho_k\to\rho$, $\boldsymbol{w}_k\to\boldsymbol{w}$ pointwise γ a.e. in $\mathbb{R}^d$, so that Fatou's Lemma yields

$$\liminf_{k\uparrow\infty}\Phi(\rho_k,\boldsymbol{w}_k) \ge \Phi(\rho,\boldsymbol{w})\,. \tag{3.3}$$

Since $\rho \wedge k \to \rho$ in $L^1_\gamma(\mathbb{R}^d)$ as $k \uparrow \infty$, we have $m_k \to m$ and $\rho_k \to \rho$ in $L^1_\gamma(\mathbb{R}^d)$. The dominated convergence theorem also yields $\boldsymbol{w}_k \to \boldsymbol{w}$ in $L^1_\gamma(\mathbb{R}^d;\mathbb{R}^d)$. Finally, since $\rho_k \ge \rho$ and $|\boldsymbol{w}_k| \le |\boldsymbol{w}|$ on R_k, and since g is non increasing,

$$\Phi(\rho_k,\boldsymbol{w}_k) = \int_{\mathbb{R}^d} \phi(\rho_k,\boldsymbol{w}_k)\,d\gamma = \int_{R_k} \phi(\rho_k,\boldsymbol{w}_k)\,d\gamma \\ \le \int_{R_k} \phi(\rho,\boldsymbol{w})\,d\gamma \le \int_{\mathbb{R}^d} \phi(\rho,\boldsymbol{w})\,d\gamma = \Phi(\rho,\boldsymbol{w})\;,$$

so that the "lim inf" in (3.3) is in fact a limit.

Next we perform a lower truncation on ρ. By a diagonal argument, it is sufficient to approximate the functions ρ_k and $\boldsymbol{w}_k$ we have just introduced, so we can assume that ρ is essentially bounded by a constant k and we omit the dependence on k. For $\delta > 0$ we now set $\rho_\delta := (\rho+\delta)\,m/(m+\delta)$. Observe that

$$\rho_\delta - m = \frac{m}{m+\delta}\,(\rho-m) \quad \text{and} \quad \rho - \rho_\delta = \frac{\delta}{m+\delta}\,(\rho-m)$$

so that $m \le \rho_\delta \le \rho$ on the set R_m^c and, by convexity of g, we get

$$g(\rho_\delta) \le C_\delta\, g(\rho) \quad \text{where} \quad C_\delta = 1 + \delta\,\frac{|g'(m)|\,(k-m)}{g(k)\,(\delta+m)}\;.$$

On the other hand, on the set R_m, we have $\rho \le \rho_\delta$, and then $g(\rho_\delta) \le g(\rho)$. As a consequence,

$$\int_{\mathbb{R}^d} \phi(\rho_\delta,\boldsymbol{w})\,d\gamma \le C_\delta \int_{\mathbb{R}^d} \phi(\rho,\boldsymbol{w})\,d\gamma\;.$$

We can then pass to the limit as $\delta \downarrow 0$, since $\rho_\delta \to \rho$ pointwise.

The last step is to approximate the functions ρ and $\boldsymbol{w}$, with $\delta \le \rho \le k$, $|\boldsymbol{w}| \le k$, by smooth functions. We consider a family of smooth approximations $\tilde\rho_\varepsilon$ and $\boldsymbol{w}_\varepsilon$ obtained by convolution with a smooth kernel. We finally set $m_\varepsilon := \int_{\mathbb{R}^d} \tilde\rho_\varepsilon\,d\gamma$ and, in this framework, redefine $\rho_\varepsilon := m\,\tilde\rho_\varepsilon/m_\varepsilon$. Since $(\rho_\varepsilon,\boldsymbol{w}_\varepsilon)$ converges to $(\rho,\boldsymbol{w})$ pointwise a.e. in $\mathbb{R}^d$ and is uniformly bounded, we can pass to the limit as above when $\varepsilon \downarrow 0$. □

3.4. The action functional on measures. Since we assumed that h is concave and strictly positive for $\rho > 0$, h is an increasing map, so that g is decreasing. We extend h and g to $[0,\infty)$ by continuity and we still denote by ϕ the lower semicontinuous envelope of ϕ in the closure $[0,\infty)\times\mathbb{R}^d$. If $h(0) > 0$ then $g(0) < \infty$ and $\phi(0,\boldsymbol{w}) = g(0)\,|\boldsymbol{w}|^2$. When $h(0) = 0$ we have $g(0) = \infty$ and

$$\phi(0,\boldsymbol{w}) = \begin{cases} \infty & \text{if } \boldsymbol{w} \ne 0\;,\\ 0 & \text{if } \boldsymbol{w} = 0\;. \end{cases}$$

We also introduce the *recession functional*

$$\phi^\infty(\rho,\boldsymbol{w}) := \sup_{\lambda>0} \frac{1}{\lambda}\,\phi(\lambda\,\rho,\lambda\,\boldsymbol{w}) = \lim_{\lambda\uparrow\infty} \frac{1}{\lambda}\,\phi(\lambda\,\rho,\lambda\,\boldsymbol{w})\;,$$

which is still a convex and lower semicontinuous function with values in $[0,\infty]$, and 1-homogeneous. It is determined by the behaviour of $h(\rho)$ as $\rho \uparrow \infty$. If we set

$$h^\infty := \lim_{\rho\uparrow\infty} \frac{h(\rho)}{\rho}\;,$$

we have

$$\phi^\infty(\rho,\boldsymbol{w}) = \begin{cases} \infty & \text{if } \boldsymbol{w} \neq 0 \\ 0 & \text{if } \boldsymbol{w} = 0 \end{cases} \quad \text{when} \quad h^\infty = 0 \,,$$

and

$$\phi^\infty(\rho,\boldsymbol{w}) = \begin{cases} \frac{|\boldsymbol{w}|^2}{h^\infty \rho} & \text{if } \rho \neq 0 \\ \infty & \text{if } \rho = 0 \text{ and } \boldsymbol{w} \neq 0 \end{cases} \quad \text{when} \quad h^\infty > 0 \,.$$

Let $\mu \in \mathcal{M}^+(\mathbb{R}^d)$ be a nonnegative Radon measure and let $\boldsymbol{\nu} \in \mathcal{M}(\mathbb{R}^d;\mathbb{R}^d)$ be a vector Radon measure on $\mathbb{R}^d$. We write their Lebesgue decomposition with respect to the reference measure γ as

$$\mu := \rho\,\gamma + \mu^\perp \,, \quad \boldsymbol{\nu} := \boldsymbol{w}\,\gamma + \boldsymbol{\nu}^\perp \,.$$

We can always introduce a nonnegative Radon measure $\sigma \in \mathcal{M}^+(\mathbb{R}^d)$ such that $\mu^\perp = \rho^\perp\sigma \ll \sigma$, $\boldsymbol{\nu}^\perp = \boldsymbol{w}^\perp\sigma \ll \sigma$, e.g. $\sigma := \mu^\perp + |\boldsymbol{\nu}^\perp|$ and define the *action functional*

$$\Phi(\mu,\boldsymbol{\nu}\,|\,\gamma) := \int_{\mathbb{R}^d} \phi(\rho,\boldsymbol{w})\,d\gamma + \int_{\mathbb{R}^d} \phi^\infty(\rho^\perp,\boldsymbol{w}^\perp)\,d\sigma \,.$$

Since ϕ^∞ is 1-homogeneous, this definition is independent of σ. As we have done up to now, we shall simply write $\Phi(\mu,\boldsymbol{\nu}) = \Phi(\mu,\boldsymbol{\nu}\,|\,\gamma)$ when there is no ambiguity on the reference measure γ.

Remark 3.4. *If h has a* sublinear growth*, then $h^\infty = 0$ and, as a consequence, if $\Phi(\mu,\boldsymbol{\nu}) < +\infty$, then we have*

$$\boldsymbol{\nu} = \boldsymbol{w}\cdot\gamma \ll \gamma \quad \textit{and} \quad \Phi(\mu,\boldsymbol{\nu}) = \int_{\mathbb{R}^d} \phi(\rho,\boldsymbol{w})\,d\gamma \,,$$

so $\Phi(\mu,\boldsymbol{\nu})$ is independent of the singular part $\mu^\perp$. When h has a linear growth*, i.e. $h^\infty > 0$, if $\Phi(\mu,\boldsymbol{\nu}) < +\infty$, then we have*

$$\boldsymbol{\nu}^\perp = \boldsymbol{w}^\perp\cdot\mu^\perp \ll \mu^\perp \,.$$

In both cases, one can choose $\sigma = \mu^\perp$, so that

$$\boldsymbol{\nu} = \boldsymbol{w}\cdot\gamma \ll \gamma + \boldsymbol{w}^\perp\cdot\mu^\perp$$

and if $1/h^\infty$ is finite, then we have

$$\Phi(\mu,\boldsymbol{\nu}\,|\,\gamma) = \int_{\mathbb{R}^d} \phi(\rho,\boldsymbol{w})\,d\gamma + \frac{1}{h^\infty}\int_{\mathbb{R}^d} |\boldsymbol{w}^\perp|^2\,d\mu^\perp \,,$$

while the last term simply drops if $h^\infty = 0$.

Lemma 3.5 (Lower semicontinuity, regular approximation of the action functional). *The action functional is lower semicontinuous with respect to the weak convergence of measures, i.e. if (γ_n), (μ_n) and (ν_n) are sequences such that $\gamma_n \rightharpoonup \gamma$ weakly in $\mathcal{M}^+(\mathbb{R}^d)$, $\mu_n \rightharpoonup \mu$ weakly in $\mathcal{M}^+(\mathbb{R}^d)$ and $\boldsymbol{\nu}_n \overset{*}{\rightharpoonup} \boldsymbol{\nu}$ in $\mathcal{M}(\mathbb{R}^d;\mathbb{R}^d)$ as $n \uparrow \infty$, then*

$$\liminf_{n\uparrow\infty} \Phi(\mu_n,\boldsymbol{\nu}_n\,|\,\gamma_n) \geq \Phi(\mu,\boldsymbol{\nu}\,|\,\gamma) \,.$$

Moreover, for every $\mu \in \mathcal{M}^+(\mathbb{R}^d)$ and $\boldsymbol{\nu} \in \mathcal{M}(\mathbb{R}^d;\mathbb{R}^d)$ such that $\Phi(\mu,\boldsymbol{\nu}) < \infty$, there exist sequences (μ_n) and (ν_n) for which

$$\mu_n := \rho_n\,\gamma \text{ with } \rho_n \in C_b^0(\mathbb{R}^d) \text{ and } \inf \rho_n > 0\,, \quad \boldsymbol{\nu}_n := \boldsymbol{w}_n\,\gamma \text{ with } \boldsymbol{w}_n \in C_b^0(\mathbb{R}^d;\mathbb{R}^d)$$

such that

$$\mu_n \rightharpoonup \mu \text{ and } \boldsymbol{\nu}_n \rightharpoonup \boldsymbol{\nu}\,, \quad \lim_{n\uparrow\infty} \int_{\mathbb{R}^d} \phi(\rho_n,\boldsymbol{w}_n)\,d\gamma = \Phi(\mu,\boldsymbol{\nu}\,|\,\gamma)\,. \tag{3.4}$$

Proof. The first statement is a well known fact about lower semicontinuity of convex integrals (see e.g. [1]). Concerning the approximation property (3.4), general relaxation results provide a family of approximations in $L^1_\gamma(\mathbb{R}^d)$. We can then apply Lemma 3.3 and a standard diagonal argument. □

3.5. The weighted Wasserstein distance. Denote by $\mathcal{B}(\mathbb{R}^d)$ the collection of all Borel subsets of $\mathbb{R}^d$, by $\mathcal{M}^+(\mathbb{R}^d)$ the collection of all finite positive Borel measures defined on $\mathbb{R}^d$ and by $\mathcal{P}(\mathbb{R}^d)$ the convex subset of all probability measures i.e. all $\mu \in \mathcal{M}^+(\mathbb{R}^d)$ such that $\mu(\mathbb{R}^d) = 1$. If $\mathcal{M}(\mathbb{R}^d;\mathbb{R}^d)$ is the set of the vector valued Borel measures $\boldsymbol{\nu} : \mathcal{B}(\mathbb{R}^d) \to \mathbb{R}^d$ with finite variation, i.e. such that

$$|\boldsymbol{\nu}|(B) := \sup\Big\{ \sum_{j\le n} |\boldsymbol{\nu}(B_j)| \;:\; B = \bigcup_{j\le n} B_j\,,\ B_j \in \mathcal{B}(\mathbb{R}^d) \text{ pairwise disjoint}\,,\ n < \infty \Big\} < \infty$$

for any $B \in \mathcal{B}(\mathbb{R}^d)$, then $|\boldsymbol{\nu}|$ is in fact a finite positive measure in $\mathcal{M}^+(\mathbb{R}^d)$ and $\boldsymbol{\nu}$ admits the polar decomposition $\boldsymbol{\nu} = \boldsymbol{w}\,|\boldsymbol{\nu}|$ where the Borel vector field $\boldsymbol{w}$ belongs to $L^1_{|\boldsymbol{\nu}|}(\mathbb{R}^d;\mathbb{R}^d)$. We can also consider $\boldsymbol{\nu}$ as a vector $(\boldsymbol{\nu}^1,\boldsymbol{\nu}^2,\cdots,\boldsymbol{\nu}^d)$ of d measures in $\mathcal{M}(\mathbb{R}^d;\mathbb{R})$.

For any $T > 0$, let $\mathcal{CE}(0,T;\mathbb{R}^d)$ be the set of time dependent measures $(\mu_t)_{t\in[0,T]}$, $(\boldsymbol{\nu}_t)_{t\in(0,T)}$ such that

1. $t \mapsto \mu_t$ is weakly * continuous in $\mathcal{M}^+_{\rm loc}(\mathbb{R}^d)$,
2. $(\boldsymbol{\nu}_t)_{t\in(0,T)}$ is a Borel family with $\int_0^T |\boldsymbol{\nu}_t|(B_R)\,dt < \infty$ for any $R > 0$,
3. $(\mu,\boldsymbol{\nu})$ is a distributional solution of

$$\partial_t \mu_t + \nabla\cdot\boldsymbol{\nu}_t = 0 \quad \text{in } \mathbb{R}^d \times (0,T)\,.$$

As in [24], we define the weighted Wasserstein distance as follows.

Definition 3.6. *The (h,γ)-Wasserstein distance between μ_0 and $\mu_1 \in \mathcal{M}^+_{\rm loc}(\mathbb{R}^d)$ is defined by*

$$W_{h,\gamma}(\mu_0,\mu_1) := \inf\Big\{ \Big[\int_0^1 \Phi(\mu_t,\boldsymbol{\nu}_t\,|\,\gamma)\,dt\Big]^{1/2} \;:\; (\mu,\boldsymbol{\nu}) \in \mathcal{CE}(0,1;\mathbb{R}^d)\,,\ \mu_{t=0} = \mu_0\,,\ \mu_{t=1} = \mu_1 \Big\} \tag{3.5}$$

with $\Phi(\mu,\boldsymbol{\nu}\,|\,\gamma) := \Phi(\rho,\boldsymbol{w}) + \Phi^\infty(\boldsymbol{w}^\perp)$ if $\mu = \rho\,\gamma + \mu^\perp$ and $\boldsymbol{\nu} = \boldsymbol{w}\,\gamma + \boldsymbol{w}^\perp\,\mu^\perp$, $\Phi(\mu,\boldsymbol{\nu}\,|\,\gamma) := \infty$ otherwise, and $\Phi^\infty(\boldsymbol{w}) := \lim_{\lambda\uparrow\infty} \lambda\,\phi(\lambda,\boldsymbol{w})$.

We denote by $\mathcal{M}_{h,\gamma}[\sigma]$ the set of all measures $\mu \in \mathcal{M}^+_{\rm loc}(\mathbb{R}^d)$ which are at finite $W_{h,\gamma}$-distance from σ.

Notice that in [24] we were using the notation $W_{\phi,\gamma}$ instead of $W_{h,\gamma}$. Whenever there is no ambiguity on the choice of the measure γ, we shall simply write W_h. The next result is taken from [24, Theorem 5.6 and Proposition 5.14]

THEOREM 3.7 (Lower semicontinuity). *If ϕ satisfies (2.4) and (2.5), the map $(\mu_0,\mu_1) \mapsto W_{h,\gamma}(\mu_0,\mu_1)$ is lower semicontinuous with respect to the weak * convergence in $\mathcal{M}^+_{\rm loc}(\mathbb{R}^d)$. More generally, suppose that $\gamma_n \rightharpoonup^* \gamma$ in $\mathcal{M}^+_{\rm loc}(\mathbb{R}^d)$, h_n is monotonically decreasing w.r.t. n and pointwise converging to h, and $\mu_0^n \rightharpoonup^* \mu_0$, $\mu_1^n \rightharpoonup^* \mu_1$ in $\mathcal{M}^+_{\rm loc}(\mathbb{R}^d)$ as $n \uparrow \infty$. Then*

$$\liminf_{n\uparrow\infty} W_{h_n,\gamma_n}(\mu_0^n,\mu_1^n) \geq W_{h,\gamma}(\mu_0,\mu_1)\ .$$

If moreover $\gamma_n \geq \gamma$ we have

$$\lim_{n\to+\infty} W_{h_n,\gamma_n}(\mu_0^n,\mu_1^n) = W_{h,\gamma}(\mu_0,\mu_1)\ .$$

It is possible to reparametrize the path connecting μ_0 to μ_1 in the definition of $W_{h,\gamma}$ and establish that, for any $T>0$,

$$W_{h,\gamma}(\sigma,\eta) = \inf\Big\{\sqrt{T}\,\Big[\int_0^T \Phi(\mu_t,\boldsymbol{\nu}_t\,|\,\gamma)\,dt\Big]^{1/2} : (\mu,\boldsymbol{\nu}) \in \mathcal{C}E(0,T;\sigma\to\eta)\Big\}$$

where $\mathcal{C}E(0,T;\sigma\to\eta)$ denotes the set of the paths $(\mu,\boldsymbol{\nu}) \in \mathcal{C}\mathcal{E}(0,T;\mathbb{R}^d)$ such that $\mu_{t=0}=\sigma$ and $\mu_{t=T}=\eta$. By [24, Theorem 5.4 and Corollary 5.18], we have the

THEOREM 3.8 (Existence of geodesics). *Whenever the infimum in (3.5) has a finite value, it is attained by a curve $(\mu,\boldsymbol{\nu}) \in \mathcal{C}\mathcal{E}(0,1;\mathbb{R}^d)$ such that*

$$\Phi(\mu_t,\boldsymbol{\nu}_t\,|\,\gamma) = W^2_{h,\gamma}(\mu_0,\mu_1) \quad \forall\, t\in(0,1)\ \mathscr{L}^1\ a.e.$$

In this case we have the equivalent characterization

$$W_{h,\gamma}(\sigma,\eta) = \min\Big\{\int_0^T \big[\Phi(\mu_t,\boldsymbol{\nu}_t\,|\,\gamma)\big]^{1/2} dt\ :\ (\mu,\boldsymbol{\nu}) \in \mathcal{C}E(0,T;\sigma\to\eta)\Big\}\ .$$

The curve $(\mu_t)_{t\in[0,1]}$ associated to a minimum for (3.5) is a constant speed mimimal geodesic:

$$W_{h,\gamma}(\mu_s,\mu_t) = |t-s|\,W_{h,\gamma}(\mu_0,\mu_1) \quad \forall\, s\,,\ t\in[0,1]\ .$$

We may notice that the characterization of $W_{h,\gamma}(\sigma,\eta)$ in terms of $\int_0^T \sqrt{\Phi(\mu_t,\boldsymbol{\nu}_t\,|\,\gamma)}\,dt$ allows to consider the case $T=+\infty$. By [2, Chap. 1] (also see [24, p. 222]), one knows that

$$W_{h,\gamma}(\mu_0,\mu_T) \leq \int_0^T |\dot\mu_t|\,dt \quad \text{with } |\dot\mu_t| := \lim_{h\to0} \frac{W_{h,\gamma}(\mu_{t+h},\mu_t)}{h}$$

for any absolutely continuous curve $t\mapsto\mu_t$ such that $\mu_{t=0}=\mu_0$ and $\mu_{t=T}=\mu_T$.

Now let us come back to the formal point of view of Section 2 and establish (2.17) in this framework. Assume that ρ_t is given by KFP and $\boldsymbol{w}_t = \boldsymbol{D}\rho_t$. The curve $\mu_t = \rho_t\,\gamma$ connects $\mu_0=\rho_0\,\gamma$ with $\mu_\infty=\gamma$ and, using

$$\sqrt{P(\rho_t)} = \sqrt{\Phi(\rho_t,\boldsymbol{w}_t\,|\,\gamma)} = |\dot\mu_t|\ ,$$

it follows that

$$W_{h,\gamma}(\rho_0,\gamma) \le \int_0^\infty \sqrt{P(\rho_t)}\,dt = \int_0^\infty |\dot\mu_t|\,dt$$

as already noted in Section 2 (equality case in (2.15)). On the other hand, for any $(\mu,\boldsymbol{\nu}) \in \mathcal{C}E(0,T;\mu_0 \to \mu_T)$, $T \in (0,\infty)$, we have $|\dot\mu_t| \le \sqrt{\Phi(\mu_t,\boldsymbol{\nu}_t)}$ and so

$$\int_0^T |\dot\mu_t|\,dt \le \int_0^T \sqrt{\Phi(\mu_t,\boldsymbol{\nu}_t)}\,dt\ .$$

By taking first the infimum $(\mu,\boldsymbol{\nu}) \in \mathcal{C}E(0,T;\mu_0 \to \mu_T)$ and then the limit $T \to \infty$, we also find

$$\int_0^\infty |\dot\mu_t|\,dt \le W_{h,\gamma}(\rho_0,\gamma)\ ,$$

thus proving the equality in the above inequality. This completes the proof of (2.17).

4. Entropy and entropy production. Let us consider now a function ψ such that $\psi''(x) = g(x)$ for any $x > 0$. Among all possible choices of ψ, we consider in particular the convex functions $\psi_a : [0,\infty) \to [0,\infty)$ depending on $a > 0$ and characterized by the conditions

$$\psi_a''(x) = g(x)\ , \quad \psi_a(a) = \psi_a'(a) = 0\ , \quad \text{i.e.} \quad \psi_a(x) = \int_a^x (x-r)\,g(r)\,dr\ .$$

Observe that $\psi_a \in C^2(0,\infty)$ has a strict minimum at $a > 0$ and it satisfies the transformation rule

$$\psi_a(x) = \psi(x) - \psi(a) - \psi'(a)\,(x-a) \quad \forall\ a > 0\ ,$$

independently of the choice of ψ (for a given function g). When $g(x) = 1/x$ we obtain the logarithmic entropy density $\mathsf{E}(x) := x\log x$ and the family

$$\mathsf{E}_a(x) := \int_a^x (y-r)\,\frac{1}{r}\,dr = x\log x - a\log a - (1+\log a)\,(x-a)\ ,$$

which provides useful lower/upper bounds for ψ. In fact, h being concave, if $h(0) = 0$, then $h(x) \ge \frac{h(a)}{a}\,x$ if $0 < x < a$, so that

$$\psi(x) \le \frac{a}{h(a)}\,\mathsf{E}_a(x) \quad \forall\ x \in (0,a]\ .$$

On the other hand, when $x \ge a$, we have $h(x) \le \frac{h(a)}{a}\,x$, so that

$$\psi(x) \ge \frac{a}{h(a)}\,\mathsf{E}_a(x) \quad \forall\ x \in [a,+\infty)\ , \tag{4.1}$$

thus showing that $\psi(x)$ has a superlinear growth as $x \uparrow \infty$.

We can therefore introduce the *relative entropy functional*

$$\Psi(\rho) := \int_{\mathbb{R}^d} \psi_a(\rho(x))\,d\gamma(x) = \int_{\mathbb{R}^d} \Big(\psi(\rho(x)) - \psi(a)\Big)\,d\gamma \quad \text{with} \quad a = \int_{\mathbb{R}^d} \rho\,d\gamma\ .$$

In the particular case $\psi = \mathsf{E}$, we set

$$\mathcal{H}(\rho) := \int_{\mathbb{R}^d} \rho \log \rho \, d\gamma - a \log a \qquad \text{with} \quad a = \int_{\mathbb{R}^d} \rho \, d\gamma \, .$$

Since ψ is convex and superlinearly increasing, if $\sup_n \Psi(\rho_n) < \infty$, then there exists a subsequence weakly converging to ρ in $L^1_\gamma(\mathbb{R}^d)$ and

$$\liminf_{n \uparrow \infty} \Psi(\rho_n) \geq \Psi(\rho) \, .$$

REMARK 4.1. *If the function ψ satisfies $\psi'' = g$, $\psi(0) = 0$ and if (2.4) holds, then ψ also satisfies McCann's conditions, i.e. the map $x \mapsto e^x \, \psi(e^{-x})$ is convex and non increasing on $(0, \infty)$ or, equivalently,*

$$x \, \psi' - \, \psi \geq 0 \quad \textit{and} \quad x^2 \, \psi'' - x \, \psi' + \psi \geq 0 \quad \forall \, x > 0 \, .$$

The convexity of ψ indeed yields $x \, \psi'(x) - \, \psi(x) \geq - \psi(0) = 0$. Consider the function $\vartheta(x) := x^2 \, \psi''(x) - x \, \psi'(x) + \psi(x)$ and observe that $\lim_{x \downarrow 0} \vartheta(x) = 0$, since $\psi'' = 1/h$ and h is concave so that, in particular, $h(x) \geq c \, x$ near $x = 0$, for some positive constant c. On the other hand, setting as usual $g = \psi'' = 1/h$, we have

$$\vartheta'(x) = x^2 \, g'(x) + x \, g(x) = x \frac{d}{dx}\big(x g(x)\big) = x \, \frac{d}{dx} \left(\frac{x}{h(x)} \right)$$

and the function $x \mapsto h(x)/x$ being positive, non increasing, we deduce that $\vartheta'(x) \geq 0$, so that $\vartheta \geq 0$.

Let us introduce the Sobolev spaces

$$W^{1,p}_\gamma(\mathbb{R}^d) := \left\{ \rho \in W^{1,p}_{\rm loc}(\mathbb{R}^d) \, : \, \int_{\mathbb{R}^d} \Big(|\rho|^p + |\boldsymbol{D} \rho|^p \Big) \, d\gamma < \infty \right\} .$$

For $\rho \in W^{1,1}_\gamma(\mathbb{R}^d)$, $\rho \geq 0$, we define the *entropy production functional* as

$$P(\rho) := \Phi(\rho, \boldsymbol{D} \rho) \text{ with domain } \mathcal{D}(P) := \left\{ \rho \in W^{1,1}_\gamma(\mathbb{R}^d) \, : \, \rho \geq 0 \, , \quad P(\rho) < \infty \right\} .$$

We also introduce the absolutely continuous functions

$$f(r) := \int_0^r \frac{1}{\sqrt{h(\xi)}} \, d\xi \, , \quad L_\psi(r) := r \, \psi'(r) - \, \psi(r) \, ,$$

and observe that

$$\frac{d}{dr} L_\psi(r) = r \, \psi''(r) = \frac{r}{h(r)}$$

is bounded if and only if $h(r)$ has a linear growth as $r \uparrow \infty$. In the case $h(r) = r$, $\psi = \mathsf{E}$, to the entropy functional $\mathcal{H}$ corresponds the *entropy production functional*

$$\mathcal{I}(\rho) := \int_{\mathbb{R}^d} \frac{|\boldsymbol{D} \rho|^2}{\rho} \, d\gamma \, .$$

PROPOSITION 4.2. *Let ρ be nonnegative function in $L^1_\gamma(\mathbb{R}^d)$. Then $\rho \in W^{1,1}_\gamma(\mathbb{R}^d)$ and $P(\rho) < \infty$ if and only if $\boldsymbol{D} f(\rho) \in L^2_\gamma(\mathbb{R}^d;\mathbb{R}^d)$ and in this case we have*

$$P(\rho) = \int_{\mathbb{R}^d} |\boldsymbol{D} f(\rho)|^2 \, d\gamma \, .$$

If $\rho \in \mathcal{D}(P)$ and $h(r) \geq \mathsf{h}\, r$ for some constant $\mathsf{h} > 0$, then $L_\psi(\rho) \in W^{1,1}_\gamma(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d} \frac{|\boldsymbol{D} L_\psi(\rho)|^2}{\rho} \, d\gamma \leq \mathsf{h}^{-1} P(\rho) \quad \textit{and} \quad P(\rho) \leq \mathsf{h}^{-1}\, \mathcal{I}(\rho) \, . \tag{4.2}$$

Moreover, the functional $\rho \mapsto P(\rho)$ is lower semicontinuous with respect to the weak convergence in $L^1_\gamma(\mathbb{R}^d)$, i.e. if a sequence $(\rho_n)_{n\in\mathbb{N}}$ weakly converges to some ρ in $L^1_\gamma(\mathbb{R}^d)$ and $\sup_{n\in\mathbb{N}} P(\rho_n) < \infty$, then $\rho \in W^{1,1}_\gamma(\mathbb{R}^d)$ and

$$\liminf_{n\uparrow\infty} P(\rho_n) \geq P(\rho) \, . \tag{4.3}$$

Proof. Identity (4.3) and $\boldsymbol{D}\, L_\psi(\rho) = \frac{\rho}{h(\rho)}\, \boldsymbol{D}\rho$ are straightforward if ρ takes its values in a compact interval of $(0,\infty)$. The general case follows as in Lemma 3.3 by a standard truncation argument, while the lower semicontinuity is a consequence of convexity. □

5. The KFP flow and its first variation.

5.1. Variational solutions to the KFP flow. As in Section 2, let us introduce the differential operators

$$\begin{aligned} \nabla_\gamma \cdot \boldsymbol{v} &:= e^V \, \nabla \cdot (e^{-V} \boldsymbol{v}) = \nabla \cdot \boldsymbol{v} - \boldsymbol{v} \cdot \boldsymbol{D} V \, , \\ \Delta_\gamma \rho &:= \nabla_\gamma \cdot (\boldsymbol{D}\rho) = \Delta \rho - \boldsymbol{D}\rho \cdot \boldsymbol{D} V \, , \end{aligned}$$

which, with respect to the measure γ, satisfy the following "integration by parts formulae" against test functions $\zeta \in C^\infty_c(\mathbb{R}^d)$:

$$\int_{\mathbb{R}^d} \boldsymbol{v} \cdot \boldsymbol{D}\zeta \, d\gamma = -\int_{\mathbb{R}^d} \nabla_\gamma \cdot \boldsymbol{v}\, \zeta \, d\gamma \quad \text{and} \quad \int_{\mathbb{R}^d} \boldsymbol{D} v \cdot \boldsymbol{D}\zeta \, d\gamma = -\int_{\mathbb{R}^d} \Delta_\gamma v\, \zeta \, d\gamma \, .$$

We consider the Kolmogorov-Fokker-Planck equation

$$\partial_t \rho_t - \Delta_\gamma \rho_t = 0 \quad \text{in } (0,\infty) \times \mathbb{R}^d \, . \tag{5.1}$$

For simplicity, we will consider equations in the whole $\mathbb{R}^d$ (corresponding to the finiteness assumption on the potential V); necessary adaptations when this is not the case are straightforward and left to the reader. We will also assume that

$$\textit{the potential } V \textit{ is smooth with bounded second derivatives} \, . \tag{5.2}$$

Based on the integration by parts formula, the variational formulation of (5.1) in the Hilbert space $L^2_\gamma(\mathbb{R}^d)$ relies on the symmetric, closed Dirichlet form

$$a_\gamma(\rho,\eta) := \int_{\mathbb{R}^d} \langle \boldsymbol{D}\rho, \boldsymbol{D}\eta \rangle \, d\gamma \quad \forall\, \rho\,,\ \eta \in W^{1,2}_\gamma(\mathbb{R}^d) \, ,$$

where $W^{1,2}_\gamma(\mathbb{R}^d)$ is endowed with its natural norm $\|\rho\|^2_{W^{1,2}_\gamma(\mathbb{R}^d)} := \|\rho\|^2_{L^2_\gamma(\mathbb{R}^d)} + a_\gamma(\rho,\rho)$. Using smooth approximations, it is not difficult to prove that $W^{1,2}_\gamma(\mathbb{R}^d)$ is dense in $L^2_\gamma(\mathbb{R}^d)$. The abstract theory of variational evolution equation and the log-concavity of the measure γ yield the following result (see e.g. [4, Thm. 6.7]).

PROPOSITION 5.1. *Assume that* (2.1)–(2.3) *hold. For every* $\rho_0 \in L^2_\gamma(\mathbb{R}^d)$*, the solution of* (5.1) *has the following properties:*

1. *There exists a unique* $\rho_t = S_t\rho_0 \in W^{1,2}_{\rm loc}\left(0,\infty; L^2_\gamma(\mathbb{R}^d)\right)$, $t>0$, *such that*

$$\frac{d}{dt}\langle \rho_t, \eta\rangle_{L^2_\gamma(\mathbb{R}^d)} + a_\gamma(\rho_t,\eta) = 0 \quad \forall\, \eta \in W^{1,2}_\gamma(\mathbb{R}^d)\,, \quad \lim_{t\downarrow 0} \rho_t = \rho_0 \text{ in } L^2_\gamma(\mathbb{R}^d)\,. \tag{5.3}$$

   If $\rho_{\min} \le \rho_0 \le \rho_{\max}$*, then* ρ_t *satisfies the same uniform bounds. The semigroup* $(S_t)_{t\ge 0}$ *is an analytic Markov semigroup in* $L^2_\gamma(\mathbb{R}^d)$ *which can be extended by continuity to a contraction semigroup in* $L^p_\gamma(\mathbb{R}^d)$ *for every* $p\in[1,\infty)$ *and to a weakly* * *continuous semigroup in* $L^\infty_\gamma(\mathbb{R}^d)$*.*

2. *For every* ρ, $\sigma \in L^2_\gamma(\mathbb{R}^d)$*, we have*

$$\int_{\mathbb{R}^d} \left(S_t\rho\right)\sigma\, d\gamma = \int_{\mathbb{R}^d} \rho\left(S_t\sigma\right) d\gamma \quad \forall\, t\ge 0\,.$$

3. *For every* $t>0$*,* S_t *maps* $L^\infty_\gamma(\mathbb{R}^d)$ *into* $C_b(\mathbb{R}^d)$ *and* $\mathrm{Lip}_b(\mathbb{R}^d)$ *into itself, with the uniform bound*

$$[S_t\rho]_{\mathrm{Lip}(\mathbb{R}^d)} \le [\rho]_{\mathrm{Lip}(\mathbb{R}^d)} \quad \forall\, t\ge 0\,, \quad \forall\, \rho\in \mathrm{Lip}_b(\mathbb{R}^d)\,.$$

4. *If* $\rho_0 \ge 0$*,* $\int_{\mathbb{R}^d} |x|^2\,\rho_0\, d\gamma < \infty$ *and* $\mathcal{H}(\rho_0) < \infty$*, then the map* $t\mapsto \mathcal{H}(\rho_t)$ *is convex,* $\rho_t \in W^{1,1}_\gamma(\mathbb{R}^d)$ *for every time* $t>0$*, and*

$$\sup_{t\in[0,T]} \int_{\mathbb{R}^d} |x|^2 \rho_t\, d\gamma < \infty\,, \quad \frac{d}{dt}\,\mathcal{H}(\rho_t) = -\mathcal{I}(\rho_t)\,, \quad \frac{d}{dt}\Big(e^{2\lambda t}\,\mathcal{I}(\rho_t)\Big) \le 0\,.$$

Notice that the Assumption $\rho_0 \in L^2_\gamma(\mathbb{R}^d)$ is not needed in Property 4, according to [4, Thm. 6.7].

5.2. Measure valued solutions to the FP flow. We first recall some basic results on measure-valued solutions of the Fokker-Planck (FP) equation

$$\partial_t \mu_t = \Delta\mu_t + \nabla\cdot(\boldsymbol{D}V\mu_t) \quad (t,x)\in(0,+\infty)\times\mathbb{R}^d\,. \tag{5.4}$$

Solutions of (5.4) are understood in the sense of distributions, i.e. for any $T>0$ and $\varphi \in C^\infty_c([0,T]\times\mathbb{R}^d)$, we have

$$\int_{\mathbb{R}^d} \varphi_T\, d\mu_T = \int_{\mathbb{R}^d} \varphi\, d\mu_0 + \int_0^T \int_{\mathbb{R}^d} \Big(\partial_t\varphi_t + \Delta\varphi_t - \boldsymbol{D}V\cdot\boldsymbol{D}\varphi_t\Big)\, d\mu_t\, dt\,. \tag{5.5}$$

For any $\mu \in \mathcal{M}^+(\mathbb{R}^d)$, we denote by $\mathsf{m}_p(\mu)$, $p\in[1,\infty)$, the p-moment of μ, i.e. $\mathsf{m}_p(\mu) := \int_{\mathbb{R}^d} |x|^p\, d\mu(x)$. By $\mathcal{P}_2(\mathbb{R}^d)$ we denote the space of probability measures on $\mathbb{R}^d$ with finite second moment m_2. The *relative entropy* of μ with respect to γ is defined as

$$\mathcal{H}(\mu\,|\,\gamma) := \int_{\mathbb{R}^d} \rho\log\rho\, d\gamma \text{ if } \mu \ll \gamma \text{ and } \mu = \rho\,\gamma\,, \quad \mathcal{H}(\mu\,|\,\gamma) := +\infty \text{ otherwise}\,.$$

Given two probability measures μ and ν in $\mathcal{P}(\mathbb{R}^d)$, the classical *Wasserstein distance* W_2 is defined as $W_2(\mu,\nu) := \inf\{[\int_{\mathbb{R}^d\times\mathbb{R}^d} |y-x|^2\,d\Sigma]^{1/2} \,:\, \Sigma\in\Gamma(\mu,\nu)\}$. Here $\Gamma(\mu,\nu)$ is the set of all *couplings* between μ and ν: it consists of all probability measures Σ on $\mathbb{R}^d\times\mathbb{R}^d$ whose first and second marginals are respectively μ and ν, i.e. $\Sigma(B\times\mathbb{R}^d) = \mu(B)$ and $\Sigma(\mathbb{R}^d\times B) = \nu(B)$ for any $B\in\mathcal{B}(\mathbb{R}^d)$. Notice that the notation W_2 is not consistent with the one for weighted distances W_h; we shall however use it as it is classical.

The next lemma is a particular case (for the Relative Entropy functional) of the general result [2, Theorem 11.2.1]; see also [4, Sect. 6.3] and [5, Sect. 3].

PROPOSITION 5.2 (Uniqueness and stability of the solutions of FP). *Let* $\mu_0\in\mathcal{P}_2(\mathbb{R}^d)$.

1. *The FP equation* (5.5) *has a unique solution* $\mu_t = S_t\mu_0$ *in the class of weakly continuous maps* $t\mapsto\mu_t\in\mathcal{M}^+(\mathbb{R}^d)$ *with* $\sup_{t\in(0,T)}\mathsf{m}_2(\mu_t)<+\infty$.
2. *The unique solution* μ_t *is continuous with respect to the Wasserstein distance* W_2 *and Lipschitz continuous in all compact intervals* $[t_0,t_1]\subset(0,+\infty)$.
3. *It is characterized by the family of evolution variational inequalities*
$$\frac12\,\frac{d}{dt}W_2^2(\mu_t,\nu)+\frac{\lambda}{2}\,W_2^2(\mu_t,\nu)+\mathcal{H}(\mu_t\,|\,\gamma)\le\mathcal{H}(\nu\,|\,\gamma)\quad\forall\,\nu\in\mathcal{P}_2(\mathbb{R}^d)\,,$$
which hold for a.e. $t>0$.
4. *In addition, it is stable:* $\mu_0^n\to\mu_0$ *in* $\mathcal{P}_2(\mathbb{R}^d)$ *implies that* $\mu_t^n\to\mu_t$ *in* $\mathcal{P}_2(\mathbb{R}^d)$ *for all* $t\ge0$.

Notice that the measure γ provides a *stationary* solution of (5.4). All solutions μ_t weakly converge to γ as $t\to+\infty$. Finally, μ_t is absolutely continuous with respect to γ for any $t>0$, with density ρ_t, and ρ_t is a solution of the KFP flow.

5.3. Variational solutions to the modified KFP equation. We consider the first variation of the KFP flow, i.e. the *modified Kolmogorov-Fokker-Planck equation*

$$\partial_t\boldsymbol{w}_t-\Delta_\gamma\boldsymbol{w}_t+D^2V\,\boldsymbol{w}_t=0\quad\text{in }(0,\infty)\times\mathbb{R}^d\,,\quad\lim_{t\downarrow0}\boldsymbol{w}_t=\boldsymbol{w}_0\quad\text{in }L^2_\gamma(\mathbb{R}^d;\mathbb{R}^d)\quad(5.6)$$

for the vector field $\boldsymbol{w}:(0,\infty)\times\mathbb{R}^d\to\mathbb{R}^d$. In the Hilbert space $\boldsymbol{W}:=W^{1,2}_\gamma(\mathbb{R}^d;\mathbb{R}^d)$, we consider the continuous (recall (5.2)) bilinear form

$$\boldsymbol{a}_\gamma(\boldsymbol{v},\boldsymbol{w}):=\int_{\mathbb{R}^d}\Big(\boldsymbol{D}\boldsymbol{v}:\boldsymbol{D}\boldsymbol{w}+\boldsymbol{D}^2V\,\boldsymbol{v}\cdot\boldsymbol{w}\Big)\,d\gamma\,,$$

$\boldsymbol{D}\boldsymbol{v}:\boldsymbol{D}\boldsymbol{w}:=\sum_{i,j}\partial_j\boldsymbol{v}_i\,\partial_j\boldsymbol{w}_i$. We look for solutions $\boldsymbol{w}\in W^{1,2}_{\rm loc}((0,\infty);L^2_\gamma(\mathbb{R}^d;\mathbb{R}^d))\cap L^2_{\rm loc}([0,\infty);\boldsymbol{W})$ solving the variational formulation

$$\frac{d}{dt}\int_{\mathbb{R}^d}\boldsymbol{w}_t\cdot\boldsymbol{\zeta}\,d\gamma+\boldsymbol{a}_\gamma(\boldsymbol{w}_t,\boldsymbol{\zeta})=0\quad\forall\,\boldsymbol{\zeta}\in\boldsymbol{W}.\qquad(5.7)$$

Observe that vector fields in $C^1_c(\mathbb{R}^d;\mathbb{R}^d)$ belong to $\boldsymbol{W}$. Actually the space of smooth compactly supported functions $C^\infty_c(\mathbb{R}^d;\mathbb{R}^d)$ is dense in $\boldsymbol{W}$, and $\boldsymbol{W}$ itself is dense in $L^2_\gamma(\mathbb{R}^d;\mathbb{R}^d)$. Notice moreover that if $\zeta:\mathbb{R}\to[0,\infty)$ is a smooth convex function with bounded second order derivatives and $\zeta(0)=0$, and $\boldsymbol{z}(\boldsymbol{w}):=\frac{\zeta'(|\boldsymbol{w}|)}{|\boldsymbol{w}|}\boldsymbol{w}$ (with $\boldsymbol{z}(0)=0$), an easy calculation shows that solutions of (5.7) satisfy

$$-\frac{d}{dt}\int_{\mathbb{R}^d}\zeta(|\boldsymbol{w}_t|)\,\mathrm{d}\gamma=\boldsymbol{a}_\gamma(\boldsymbol{w}_t,\boldsymbol{z}(\boldsymbol{w}_t))\ge0\quad\text{a.e. in }(0,\infty).$$

With these observations in hand, we can apply the variational theory of evolution equations and a simple regularization argument to prove the next result.

PROPOSITION 5.3. *For every $\boldsymbol{w}_0 \in L^2_\gamma(\mathbb{R}^d;\mathbb{R}^d)$, there exists a unique solution $\boldsymbol{w} = \boldsymbol{R}\boldsymbol{w}_0$ of (5.7) in $W^{1,2}_{\rm loc}((0,\infty);L^2_\gamma(\mathbb{R}^d;\mathbb{R}^d)) \cap L^2_{\rm loc}([0,\infty);\boldsymbol{W})$ with $\lim_{t\downarrow 0}\boldsymbol{w}_t = \boldsymbol{w}_0$ in $L^2_\gamma(\mathbb{R}^d;\mathbb{R}^d)$. The semigroup $\boldsymbol{R}$ is symmetric*

$$\int_{\mathbb{R}^d} \boldsymbol{R}_t\boldsymbol{w}\cdot\boldsymbol{z}\,d\gamma = \int_{\mathbb{R}^d} \boldsymbol{w}\cdot\boldsymbol{R}_t\boldsymbol{z}\,d\gamma \quad \forall\,\boldsymbol{w}\,,\ \boldsymbol{z}\in L^2_\gamma(\mathbb{R}^d;\mathbb{R}^d)\,,\quad \forall\,t>0\,,$$

and satisfies

$$\int_{\mathbb{R}^d} \zeta\big(|\boldsymbol{R}_t\boldsymbol{w}_0|\big)\,\mathrm{d}\gamma \le \int_{\mathbb{R}^d} \zeta\big(|\boldsymbol{w}_0|\big) \quad \textit{for every } \boldsymbol{w}_0\in L^2_\gamma(\mathbb{R}^d;\mathbb{R}^d)$$

and for every convex function $\zeta:\mathbb{R}\to[0,\infty)$ with $\zeta(0)=0$ (which is therefore non decreasing). In particular $\boldsymbol{R}$ can be extended by density to a contraction semigroup in $L^p_\gamma(\mathbb{R}^d;\mathbb{R}^d)$, $p\in[1,\infty]$.

The link between (5.1) and (5.6) is enlightened by the next result.

THEOREM 5.4. *If ρ_t is a variational solution of the KFP equation (5.1) with initial datum $\rho_0\in W^{1,2}_\gamma(\mathbb{R}^d)$, then $\boldsymbol{w}_t := \boldsymbol{D}\rho_t$ belongs to $C^0([0,\infty);L^2_\gamma(\mathbb{R}^d))$ and it is the solution of the modified KFP equation (5.6) with initial datum $\boldsymbol{w}_0 := \boldsymbol{D}\rho_0$. In particular we have*

$$\int_{\mathbb{R}^d} \boldsymbol{D}S_t\rho\cdot\boldsymbol{w}\,d\gamma = \int_{\mathbb{R}^d} \boldsymbol{D}\rho\cdot\boldsymbol{R}_t\boldsymbol{w}\,d\gamma \quad \forall\,\rho\in W^{1,2}_\gamma(\mathbb{R}^d)\,,\quad \forall\,\boldsymbol{w}\in L^2_\gamma(\mathbb{R}^d;\mathbb{R}^d)\,.$$

The same result holds if ρ_0 belongs to $W^{1,1}_\gamma(\mathbb{R}^d)$.

Proof. Since $\mathcal{D}(\Delta_\gamma)$ is dense in $W^{1,2}_\gamma(\mathbb{R}^d)$, we can assume that $\rho_0\in\mathcal{D}(\Delta_\gamma)$. Then the regularity result of Proposition 5.1 shows that $\rho_t\in\mathcal{D}(\Delta_\gamma)$ for every $t\ge 0$. Setting $\boldsymbol{w}_t := \boldsymbol{D}\rho_t$, we know (see e.g. the argument in the proof of [25, Lemma 5.2]) that $\boldsymbol{a}_\gamma(\boldsymbol{w}_t,\boldsymbol{w}_t)\le\|\Delta_\gamma\rho_0\|_{L^2_\gamma(\mathbb{R}^d)}<+\infty$. For a fixed $\boldsymbol{\zeta}\in C^\infty_c(\mathbb{R}^d;\mathbb{R}^d)$, we can then evaluate

$$\frac{d}{dt}\int_{\mathbb{R}^d}\boldsymbol{w}_t\cdot\boldsymbol{\zeta}\,d\gamma = \frac{d}{dt}\int_{\mathbb{R}^d}\boldsymbol{D}\rho_t\cdot\boldsymbol{\zeta}\,d\gamma = -\frac{d}{dt}\int_{\mathbb{R}^d}\rho_t\,\nabla_\gamma\cdot\boldsymbol{\zeta}\,d\gamma = \int_{\mathbb{R}^d}\boldsymbol{D}\rho_t\cdot\boldsymbol{D}(\nabla_\gamma\cdot\boldsymbol{\zeta})\,d\gamma\,. \tag{5.8}$$

With the notations $\partial_i = \partial/\partial x_i$ and $\partial_{ij} = \partial^2/\partial x_i\partial x_j$ for $i,\,j = 1,\,2\ldots d$, let us observe that

$$\big(\boldsymbol{D}\,(\nabla_\gamma\cdot\boldsymbol{\zeta})\big)_j = \sum_i \partial_j\big(\partial_i\zeta_i - \zeta_i\,\partial_i V\big) = \sum_i \partial^2_{ij}\zeta_i - \partial_j\zeta_i\,\partial_i V - \zeta_i\,\partial^2_{ij}V$$

and

$$\boldsymbol{D}\rho_t\cdot\boldsymbol{D}(\nabla_\gamma\cdot\boldsymbol{\zeta}) = \sum_{i,j}\partial_j\rho_t\,\partial^2_{ij}\zeta_i - \partial_j\rho_t\,\partial_j\zeta_i\,\partial_i V - \partial_j\rho_t\,\zeta_i\,\partial^2_{ij}V\,.$$

Inserting this expression in (5.8) and integrating by parts the first term we get

$$\begin{aligned}
&\int_{\mathbb{R}^d} \boldsymbol{D}\rho_t \cdot \boldsymbol{D}(\nabla_\gamma \cdot \boldsymbol{\zeta})\, d\gamma \\
&\qquad = \sum_{i,j} \int_{\mathbb{R}^d} \Big(\partial_j \rho_t\, \partial^2_{ij} \zeta_i\, d\gamma - \sum_{i,j} \int_{\mathbb{R}^d} \Big(\partial_j \rho_t\, \partial_j \zeta_i\, \partial_i V + \partial_j \rho_t\, \zeta_i\, \partial^2_{ij} V \Big)\, d\gamma \\
&\qquad = \sum_{i,j} \int_{\mathbb{R}^d} \Big(-\partial^2_{ij} \rho_t\, \partial_j \zeta_i + \partial_i V\, \partial_j \rho_t\, \partial_j \zeta_i \Big)\, d\gamma \\
&\qquad\qquad\qquad\qquad - \sum_{i,j} \int_{\mathbb{R}^d} \Big(\partial_j \rho_t\, \partial_j \zeta_i\, \partial_i V + \partial_j \rho_t\, \zeta_i\, \partial^2_{ij} V \Big)\, d\gamma \\
&\qquad = - \sum_{i,j} \int_{\mathbb{R}^d} \Big(\partial^2_{ij} \rho_t\, \partial_j \zeta_i \; + \partial_j \rho_t\, \zeta_i\, \partial^2_{ij} V \Big)\, d\gamma \\
&\qquad = - \int_{\mathbb{R}^d} \Big(\boldsymbol{D}\boldsymbol{w}_t : \boldsymbol{D}\boldsymbol{\zeta} + D^2 V \boldsymbol{w}_t \cdot \boldsymbol{\zeta} \Big)\, d\gamma = -\, \boldsymbol{a}_\gamma(\boldsymbol{w}_t, \boldsymbol{\zeta})\ .
\end{aligned}$$

Combined with (5.8), this shows that $\boldsymbol{w}_t := \boldsymbol{D}\rho_t$ satisfies the variational formulation of (5.6). The case of $\rho_0 \in W^{1,1}_\gamma(\mathbb{R}^d)$ follows by a standard approximation procedure, the fact that $\boldsymbol{D} S_t \rho_0 = \boldsymbol{R}_t \boldsymbol{D}\rho_0$, and the L^1_γ-contraction property of $\boldsymbol{R}$. □

5.4. Measure valued solutions to the modified KFP equation. Exactly like the (K)FP equation, the modified system can be extended to vector-valued measures initial data. To $\boldsymbol{w}_t$, we associate the vector valued measures $\boldsymbol{\nu}_t := \boldsymbol{w}_t\, \gamma \in \mathcal{M}(\mathbb{R}^d; \mathbb{R}^d)$; starting from (5.7), a further integration by parts yields the weak formulation

$$\frac{d}{dt} \int_{\mathbb{R}^d} \boldsymbol{\zeta} \cdot d\boldsymbol{\nu}_t = \int_{\mathbb{R}^d} \Big(\Delta \boldsymbol{\zeta} - \boldsymbol{D}\boldsymbol{\zeta}\, \boldsymbol{D}V - D^2 V\, \boldsymbol{\zeta} \Big) \cdot\, d\boldsymbol{\nu}_t \quad \forall\, \boldsymbol{\zeta} \in C^2_c(\mathbb{R}^d)\ , \tag{5.9}$$

of the system

$$\partial_t \boldsymbol{\nu}_t = \Delta \boldsymbol{\nu}_t + \nabla \cdot (\boldsymbol{D}V \otimes \boldsymbol{\nu}_t) - D^2 V\, \boldsymbol{\nu}_t\ ,$$

where $(\boldsymbol{D}\boldsymbol{\zeta}\, \boldsymbol{D}V)_j = \sum_i \partial_i \boldsymbol{\zeta}_j\, \partial_i V$ and $\big(\nabla \cdot (\boldsymbol{D}V \otimes \boldsymbol{\nu})\big)_j = \sum_i \partial_i \left(\partial_i V\, \boldsymbol{\nu}_j\right)$.

The semigroup can be extended to initial data which are vector valued measures with finite total variation using equi-integrability and moment estimates taken from [24].

PROPOSITION 5.5 (Equi-integrability and moment estimates). *Let ζ be a non-negative Borel function such that $\mu(\zeta^2) = \int_{\mathbb{R}^d} \zeta^2\, d\mu$ and $\gamma(\zeta^2) = \int_{\mathbb{R}^d} \zeta^2\, d\gamma$ are finite. If $\Phi(\mu, \boldsymbol{\nu}) < \infty$, we have*

$$\Big(\int_{\mathbb{R}^d} \zeta\, d\,|\boldsymbol{\nu}| \Big)^2 \le \Phi(\mu, \boldsymbol{\nu})\, \gamma(\zeta^2)\, h\left(\mu(\zeta^2)/\gamma(\zeta^2)\right)\ .$$

In particular, for every Borel set $A \in \mathcal{B}(\mathbb{R}^d)$ we have

$$\Big(|\boldsymbol{\nu}|(A) \Big)^2 \le \Phi(\mu, \boldsymbol{\nu})\, \gamma(A)\, h\big(\mu(A)/\gamma(A)\big) \tag{5.10}$$

which in particular yields ($\gamma(\mathbb{R}^d) = 1$)

$$\Big(|\boldsymbol{\nu}|(\mathbb{R}^d) \Big)^2 \le \Phi(\mu, \boldsymbol{\nu})\, h\big(\mu(\mathbb{R}^d)\big)\ .$$

If moreover $\mathsf{m}_2(\mu) < \infty$, we can bound the first moment of $|\boldsymbol{\nu}|$ by

$$\mathsf{m}_1(|\boldsymbol{\nu}|) = \int_{\mathbb{R}^d} |x|\, d\,|\boldsymbol{\nu}| \le \Big(\Phi(\mu,\boldsymbol{\nu})\, \mathsf{m}_2(\gamma)\, h\big(\mathsf{m}_2(\mu)/\mathsf{m}_2(\gamma)\big)\Big)^{1/2}. \tag{5.11}$$

THEOREM 5.6. *For every $\boldsymbol{\nu}_0 \in \mathcal{M}(\mathbb{R}^d;\mathbb{R}^d)$ with $\mathsf{m}_1(|\boldsymbol{\nu}_0|) < +\infty$, there exists a unique solution $\boldsymbol{\nu}_t = \boldsymbol{\mathcal{R}}_t\boldsymbol{\nu}_0$ in the class of weakly continuous maps $t \mapsto \boldsymbol{\nu}_t \in \mathcal{M}(\mathbb{R}^d;\mathbb{R}^d)$ with $\sup_{t\in[0,T]} \mathsf{m}_1(|\boldsymbol{\nu}_t|) < +\infty$, for every final time $T > 0$. When $\boldsymbol{\nu}_0 = \boldsymbol{w}_0\,\gamma$ then $\boldsymbol{\mathcal{R}}_t\boldsymbol{\nu}_0 = \boldsymbol{R}_t\,\boldsymbol{w}_0\,\gamma$. The map $\boldsymbol{\nu}_0 \mapsto \boldsymbol{\mathcal{R}}\boldsymbol{\nu}_0$ is stable in the following sense: if*

$$\boldsymbol{\nu}_0^n \rightharpoonup^* \boldsymbol{\nu}_0 \quad \textit{weakly}^* \textit{ in } \mathcal{M}(\mathbb{R}^d;\mathbb{R}^d) \textit{ with } \sup_n \mathsf{m}_1(|\boldsymbol{\nu}_0^n|) < +\infty$$

then $\boldsymbol{\mathcal{R}}_t\boldsymbol{\nu}_0^n \rightharpoonup^* \boldsymbol{\mathcal{R}}_t\boldsymbol{\nu}_0$ in $\mathcal{M}(\mathbb{R}^d;\mathbb{R}^d)$.

Proof. We divide the proof in three steps.

Step 1. Let us first associate to $\boldsymbol{\nu} = \boldsymbol{w}\,\gamma \in \mathcal{M}(\mathbb{R}^d;\mathbb{R}^d)$ the probability measure

$$\upsilon := \frac{1}{M}\,\frac{1}{\sqrt{1+|x|^2}}\,|\boldsymbol{\nu}| = \frac{1}{M}\,\frac{|\boldsymbol{w}|}{\sqrt{1+|x|^2}}\,\gamma \in \mathcal{P}(\mathbb{R}^d)\,, \tag{5.12}$$

where the constant M is a renormalization factor such that $\upsilon(\mathbb{R}^d) = 1$. Observe that if $\mathsf{m}_1(|\boldsymbol{\nu}|) = \int_{\mathbb{R}^d} |x|\,|\boldsymbol{w}(x)|\,d\gamma$ is finite, then $\upsilon \in \mathcal{P}_2(\mathbb{R}^d)$ and

$$\mathsf{m}_2(\upsilon) \le \frac{1}{M}\,\mathsf{m}_1(|\boldsymbol{\nu}|)\,. \tag{5.13}$$

We also choose the action density to be $\phi_2(\rho,\boldsymbol{w}) := |\boldsymbol{w}|^2/\rho$ corresponding to $h(\rho) = \rho$, and observe that the corresponding functional writes

$$\Phi_2(\upsilon,\boldsymbol{w}) = M\int_{\mathbb{R}^d} \sqrt{1+|x|^2}\,|\boldsymbol{w}(x)|\,d\gamma \le M\Big(|\boldsymbol{\nu}|(\mathbb{R}^d) + \mathsf{m}_1(|\boldsymbol{\nu}|)\Big)\,. \tag{5.14}$$

PROPOSITION 5.7. *Let us suppose that $\boldsymbol{w} \in L^1_\gamma(\mathbb{R}^d;\mathbb{R}^d)$ with $\mathsf{m}_1(|\boldsymbol{\nu}|) < +\infty$ and set $\boldsymbol{\nu} := \boldsymbol{w}\,\gamma$, υ as in (5.12), $\boldsymbol{w}_t = \boldsymbol{R}_t\boldsymbol{w}$, $\boldsymbol{\nu}_t = \boldsymbol{w}_t\,\gamma$, $\upsilon_t = S_t\upsilon$. Then*

$$|\boldsymbol{\nu}_t|(\mathbb{R}^d) \le |\boldsymbol{\nu}|(\mathbb{R}^d)\,, \quad \mathsf{m}_1(|\boldsymbol{\nu}_t|) \le |\boldsymbol{\nu}|(\mathbb{R}^d) + 2\,\mathsf{m}_1(|\boldsymbol{\nu}|) + 4\,M\,\mathsf{m}_2(\gamma), \tag{5.15}$$

and for any $t > 0$, we have

$$\Big(|\boldsymbol{\nu}_t|(A)\Big)^2 \le M\Big(|\boldsymbol{\nu}|(\mathbb{R}^d) + \mathsf{m}_1(|\boldsymbol{\nu}|)\Big)\,\upsilon_t(A) \quad \forall\, A \in \mathcal{B}(\mathbb{R}^d)\,. \tag{5.16}$$

Proof. The first inequality of (5.15) follows by the L^1_γ-contraction property of $\boldsymbol{R}$. Since the FP flow contracts the Wasserstein distance by Proposition 5.2 and since γ is a stationary solution, the triangle inequality for the Wasserstein distance and the fact that $\sqrt{\mathsf{m}_2(\mu)} = W_2(\mu,\delta_0)$ yield

$$\sqrt{\mathsf{m}_2(\upsilon_t)} \le W_2(\upsilon_t,\gamma) + \sqrt{\mathsf{m}_2(\gamma)} \le W_2(\upsilon,\gamma) + \sqrt{\mathsf{m}_2(\gamma)} \le \sqrt{\mathsf{m}_2(\upsilon)} + 2\sqrt{\mathsf{m}_2(\gamma)}\,.$$

On the other hand, (5.11) yields

$$\mathsf{m}_1(|\boldsymbol{\nu}_t|) \le \sqrt{\mathsf{m}_2(\upsilon_t)}\,\sqrt{\Phi_2(\upsilon_t,\boldsymbol{\nu}_t)} \le \Big(\sqrt{\mathsf{m}_2(\upsilon)} + 2\sqrt{\mathsf{m}_2(\gamma)}\Big)\,\sqrt{\Phi_2(\upsilon,\boldsymbol{\nu})}\,.$$

Here we used the fact that $\Phi_2(v_t, \boldsymbol{\nu}_t) \le \Phi_2(v, \boldsymbol{\nu})$. This will appear later as a consequence of Theorem 6.1, and is independent of the present result. Combined with (5.13) and (5.14), this proves the estimate on $\mathsf{m}_1(|\boldsymbol{\nu}_t|)$.

Applying (5.10) and (5.14), we get (5.16). □

Step 2: existence. Let us approximate a given $\boldsymbol{\nu}_0 \in \mathcal{M}(\mathbb{R}^d;\mathbb{R}^d)$ with $\mathsf{m}_1(|\boldsymbol{\nu}_0|) < +\infty$ by a sequence $\boldsymbol{\nu}^k = \boldsymbol{w}^k\,\gamma \rightharpoonup^* \boldsymbol{\nu}_0$ as $k \to \infty$ in $\mathcal{M}(\mathbb{R}^d;\mathbb{R}^d)$ with $\boldsymbol{w}_k \in L^2_\gamma(\mathbb{R}^d;\mathbb{R}^d)$ and $\mathsf{m}_1(|\boldsymbol{\nu}_k|) \to \mathsf{m}_1(|\boldsymbol{\nu}_0|)$. We set $\boldsymbol{\nu}^k_t := \boldsymbol{w}^k_t\,\gamma$ with $\boldsymbol{w}^k_t = \boldsymbol{R}_t \boldsymbol{w}^k$ so that $\boldsymbol{\nu}^k_t$ solves (5.9). Thanks to Proposition 5.7, we know that the first order moment of $\boldsymbol{\nu}^k_t$ are uniformly bounded. This is sufficient to pass to the limit (up to extraction of a suitable subsequence) in (5.9) and to find a solution $\boldsymbol{\nu}_t$ which is weakly* continuous in $\mathcal{M}(\mathbb{R}^d;\mathbb{R}^d)$ and satisfies the initial condition in the sense that $\nu_t \rightharpoonup^* \nu_0$ in $\mathcal{M}(\mathbb{R}^d;\mathbb{R}^d)$ as $t \downarrow 0$.

Step 3: uniqueness and stability. It follows by a standard duality argument, like in the case of Equation (5.6). If $\boldsymbol{\nu}^1_t$ and $\boldsymbol{\nu}^2_t$ are two weakly continuous solutions of (5.9), their difference $\boldsymbol{\sigma}_t := \boldsymbol{\nu}^1_t - \boldsymbol{\nu}^2_t$ solves

$$\int_{\mathbb{R}^d} \boldsymbol{\zeta}_T \cdot d\boldsymbol{\sigma}_T = \int_0^T \int_{\mathbb{R}^d} \Big(\partial_t \boldsymbol{\zeta}_t + \Delta \boldsymbol{\zeta}_t - \boldsymbol{D}\boldsymbol{\zeta}_t \otimes \boldsymbol{D}V - D^2V\,\boldsymbol{\zeta}_t\Big) \cdot d\boldsymbol{\sigma}_t\,dt \tag{5.17}$$

for every $T > 0$ and $\boldsymbol{\zeta} \in C^\infty_c([0,T]\times\mathbb{R}^d;\mathbb{R}^d)$. By a mollification technique, it is not difficult to check that (5.17) also holds for every function $\boldsymbol{\zeta} \in C([0,T]\times\mathbb{R}^d;\mathbb{R}^d)$ with $\partial_t\varphi$, $\boldsymbol{D}\boldsymbol{\zeta}$ and $D^2\boldsymbol{\zeta}$ continuous and bounded in $[0,T]\times\mathbb{R}^d$.

Next, we introduce a family of smooth convex potentials V_n with bounded derivatives of arbitrary orders, which satisfies a uniform Lipschitz condition

$$|\boldsymbol{D}V_n(x) - \boldsymbol{D}V_n(y)| \le L\,|x-y| \quad \forall\, x,\, y \in \mathbb{R}^d\,,$$

for some positive constant L which is independent of n and such that

$$V_n \to V\,, \quad \boldsymbol{D}V_n \to \boldsymbol{D}V\,, \quad D^2V_n \to D^2V \quad \text{pointwise as } n \to \infty\,.$$

For a given $\boldsymbol{\eta} \in C^\infty_c(\mathbb{R}^d;\mathbb{R}^d)$, we consider the solution $\boldsymbol{\zeta}_t$ of the time reversed (adjoint) parabolic equation

$$\partial_t \boldsymbol{\zeta}_t + \Delta\boldsymbol{\zeta}_t - \boldsymbol{D}\boldsymbol{\zeta}_t \cdot \boldsymbol{D}V - D^2V\,\boldsymbol{D}\boldsymbol{\zeta}_t = 0 \quad \text{in } (0,T)\times\mathbb{R}^d\,, \quad \boldsymbol{\zeta}_T = \boldsymbol{\eta}\,.$$

We denote by $\boldsymbol{\zeta}_n$ the solution corresponding to V_n. Using a maximum principle that can be found in [26] and the fact that the first and second order spatial derivatives of $\boldsymbol{\zeta}$ solve an analogous equation, standard parabolic regularity theory shows that $\boldsymbol{\zeta}$ is sufficiently regular to be used as a test function in (5.17) and satisfies the uniform bound (observe that the second and third derivatives of V_n are still uniformly bounded)

$$\sup_{t,x} |\boldsymbol{\zeta}_n| + |\boldsymbol{D}\boldsymbol{\zeta}_n| \le C < +\infty\,.$$

This leads to

$$\Big|\int_{\mathbb{R}^d} \boldsymbol{\eta}\cdot\,d\boldsymbol{\sigma}_T\Big| \le C\int_0^T\int_{\mathbb{R}^d} \Big(|\boldsymbol{D}V - \boldsymbol{D}V_n| + |D^2V - D^2V_n|\Big)\,d\,|\boldsymbol{\sigma}_t|\,dt\,.$$

Since the first order moment of $|\boldsymbol{\sigma}_t|$ is uniformly bounded, we can pass to the limit as $n \to \infty$ obtaining $\int_{\mathbb{R}^d} \boldsymbol{\eta}\cdot d\sigma_T = 0$. As $\boldsymbol{\eta}$ is arbitrary, we conclude that $\boldsymbol{\nu}^1_T = \boldsymbol{\nu}^2_T$. The stability is then a simple consequence of uniqueness. □

6. Action decay along the KFP flow and consequences. We can prove now our main estimate, which is a refined version of Theorem 2.1, under the assumption that

$$\text{for some } \alpha \in (0,1] \text{ the function } \quad h^{1/\alpha} \text{ is concave in } (0,\infty)\,. \tag{6.1}$$

(6.1) is equivalent to

$$\begin{aligned} &\textit{the function } h \textit{ is concave and, for } \beta = \tfrac{1-\alpha}{1+\alpha} \in [0,1),\\ &(1-\beta)\, h\, h'' + 2\,\beta\,(h')^2 \le 0 \textit{ holds in the sense of distributions.} \end{aligned} \tag{6.2}$$

In terms of $g = 1/h$ (6.1) can also be formulated as

$$g(\rho)\, g''(\rho) \ge (1+\alpha^{-1})\,(g'(\rho))^2 \quad \forall\, \rho > 0\,, \tag{6.3}$$

and by (3.2) it yields the refined action estimate

$$\left\langle \mathrm{D}^2\phi(\rho,\boldsymbol{w})\,\boldsymbol{z},\boldsymbol{z}\right\rangle \ge 2\,\beta\,\phi(\rho,\boldsymbol{y}) \quad \forall\, \boldsymbol{z}=(x,\boldsymbol{y}) \in \mathbb{R}^{d+1}\,, \quad \text{with } \beta := \frac{1-\alpha}{1+\alpha}\,. \tag{6.4}$$

Such a refinement has interesting consequences, which have been investigated in [6, 7, 23].

THEOREM 6.1. *Assume that* (2.1) *and* (2.3) *hold, and let* $(\rho,\boldsymbol{w}) \in L^1_\gamma(\mathbb{R}^d,\mathbb{R}^+)\times L^1_\gamma(\mathbb{R}^d,\mathbb{R}^d)$ *be such that* $\Phi(\rho,\boldsymbol{w}) < \infty$. *If* (6.1) *is satisfied, then*

$$\Phi(S_t\rho,\boldsymbol{R}_t\boldsymbol{w}) + 2\,\beta\sum_{i=1}^d \int_0^t \Phi(S_s\rho\,,\partial_i \boldsymbol{R}_s\boldsymbol{w})\, e^{2\lambda(s-t)}\,ds \le e^{-2\lambda t}\,\Phi(\rho,\boldsymbol{w}) \quad \forall\, t \ge 0\,,$$

with $\beta = \frac{1-\alpha}{1+\alpha} \in [0,1)$.

Proof. We first prove the result with the additional assumptions that $0 < \rho_{\min} \le \rho \le \rho_{\max}$ and $|\boldsymbol{w}| \le w_{\max}\ \gamma$ a.e. in $\mathbb{R}^d$. Assume that h is of class $C^2(0,\infty)$. It follows that $\rho_t = S_t\rho$ and $\boldsymbol{w}_t = \boldsymbol{R}_t\boldsymbol{w}$ satisfy the same bounds and, for all $t > 0$, ρ_t, $\partial_t\rho_t \in W^{1,2}_\gamma(\mathbb{R}^d)$ and $\boldsymbol{w}_t$, $\partial_t\boldsymbol{w}_t \in W^{1,2}_\gamma(\mathbb{R}^d;\mathbb{R}^d)$. The function ϕ is of class C^2 in the strip

$$Q := [\rho_{\min},\rho_{\max}] \times \{\boldsymbol{w}\in\mathbb{R}^d\,:\,|\boldsymbol{w}| \le w_{\max}\}$$

and its differential $D\phi(\rho,\boldsymbol{w})$ can be decomposed in terms of $g = 1/h$ as

$$D_\rho\phi(\rho,\boldsymbol{w}) = g'(\rho)\,|\boldsymbol{w}|^2\,, \quad D_{\boldsymbol{w}}\phi(\rho,\boldsymbol{w}) = 2\,g(\rho)\,\boldsymbol{w}\,.$$

Since $g(\rho)$ and $g'(\rho)$ are bounded, the differential is also in $L^2_\gamma(\mathbb{R}^d;\mathbb{R}^{d+1})$. As a consequence, the time derivative of $t \mapsto \Phi(\rho_t,\boldsymbol{w}_t)$ exists and

$$\frac{d}{dt}\,\Phi(\rho_t,\boldsymbol{w}_t) = \int_{\mathbb{R}^d}\Big(g'(\rho_t)\,|\boldsymbol{w}_t|^2\partial_t\rho_t + 2\,g(\rho_t)\,\boldsymbol{w}_t\cdot\partial_t\boldsymbol{w}_t\Big)\,d\gamma\,.$$

In order to apply (5.3) and (5.7) we have to verify that all components of $D\phi(\rho_t,\boldsymbol{w}_t)$ are in $W^{1,2}_\gamma(\mathbb{R}^d)$. We have already seen that they are in $L^2(\mathbb{R}^d)$. Let us compute their x-derivative:

$$\begin{aligned} &\boldsymbol{D}\left(D_\rho\phi(\rho_t,\boldsymbol{w}_t)\right) = g''(\rho_t)|\,\boldsymbol{w}_t|^2\,\boldsymbol{D}\rho_t + 2\,g'(\rho_t)\,\boldsymbol{w}_t\cdot\boldsymbol{D}\boldsymbol{w}_t\,,\\ &\boldsymbol{D}\left(D_{\boldsymbol{w}^i}\phi(\rho_t,\boldsymbol{w}_t)\right) = 2\,g'(\rho_t)\,\boldsymbol{w}_t^i\,\boldsymbol{D}\rho_t + 2\,g(\rho_t)\,\boldsymbol{D}\boldsymbol{w}_t^i \quad \text{for any } i = 1\,,\ 2\,,\dots\ d\,. \end{aligned}$$

The above functions are in $L^2_\gamma(\mathbb{R}^d)$, since $g(\rho_t)$, $g'(\rho_t)$, $g''(\rho_t)$ and $\boldsymbol{w}_t$ are bounded, so we get

$$\frac{d}{dt}\,\Phi(\rho_t,\boldsymbol{w}_t) = -\sum_{i=1}^d \int_{\mathbb{R}^d} \left\langle D^2\phi(\rho_t,\boldsymbol{w}_t)(\partial_i\rho_t,\partial_i\boldsymbol{w}_t),(\partial_i\rho_t,\partial_i\boldsymbol{w}_t)\right\rangle\, d\gamma \\ - 2\int_{\mathbb{R}^d} g(\rho_t)\, D^2V\,\boldsymbol{w}_t\cdot\boldsymbol{w}_t\, d\gamma\,.$$

Recalling (6.4) and the convexity assumption on V, we find

$$\frac{d}{dt}\,\Phi(\rho_t,\boldsymbol{w}_t) \le -2\,\beta\sum_{i=1}^d \Phi(\rho_t,\partial_i\boldsymbol{w}_t) - 2\lambda\,\Phi(\rho_t,\boldsymbol{w}_t)\,.$$

It follows from Gronwall's lemma that for all $s\in(0,t)$,

$$e^{2\lambda t}\,\Phi(\rho_t,\boldsymbol{w}_t) + 2\,\beta\sum_{i=1}^d\int_s^t \Phi(\rho_r,\partial_i\boldsymbol{w}_r)\,e^{2\lambda r}\,dr \le e^{2\lambda s}\,\Phi(\rho_s,\boldsymbol{w}_s)\,.$$

The result follows by passing to the limit as $s\downarrow 0$ and recalling that ϕ is continuous and bounded on Q and ρ_s, $\boldsymbol{w}_s$ converge to ρ, $\boldsymbol{w}$ as $s\downarrow 0$ in $L^2_\gamma(\mathbb{R}^d)$ and $L^2_\gamma(\mathbb{R}^d;\mathbb{R}^d)$ respectively. The general result for an arbitrary concave function h easily follows by approximating h by a decreasing family of smooth concave functions in the interval $[\rho_{\min},\rho_{\max}]$. Finally, the general case $\rho\in L^1_\gamma(\mathbb{R}^d)$, $\boldsymbol{w}\in L^1_\gamma(\mathbb{R}^d;\mathbb{R}^d)$, without upper and lower bounds, follows by approximation, using Lemmas 3.3 and 3.5. □

We can extend the results of Theorem 6.1 to measure valued initial data.

COROLLARY 6.2. *Assume that* (2.1)–(2.3) *and* (6.1) *hold. Let* $\mu\in\mathcal{P}_2(\mathbb{R}^d)$ *and* $\boldsymbol{\nu}\in\mathcal{M}(\mathbb{R}^d;\mathbb{R}^d)$ *with* $\mathsf{m}_1(|\boldsymbol{\nu}|)<+\infty$. *Then for every* $t>0$ *we have* $\mu_t=\mathcal{S}_t\mu=\rho_t\,\gamma$, $\boldsymbol{\nu}_t=\boldsymbol{\mathcal{R}}_t\boldsymbol{\nu}=\boldsymbol{w}_t\,\gamma$ *with* $\rho_t\in W^{1,1}_\gamma(\mathbb{R}^d)$, $\boldsymbol{w}_t\in W^{1,1}_\gamma(\mathbb{R}^d;\mathbb{R}^d)$ *if* $\alpha<1$, *and*

$$\Phi(\rho_t,\boldsymbol{w}_t) + 2\beta\sum_{i=1}^n\int_0^t \Phi(\rho_s,\partial_i\boldsymbol{w}_s)\,e^{2\lambda(s-t)}\,ds \le e^{-2\lambda t}\,\Phi(\mu,\boldsymbol{\nu}\,|\,\gamma)\quad\forall\,t>0\,,$$

where, as usual, $\beta=\frac{1-\alpha}{1+\alpha}$.

Proof. This follows directly from the measure formulation of the KFP flow (Proposition 5.2 and Theorem 5.6). □

Let us now consider the entropy functional $\Psi(\rho):=\int_{\mathbb{R}^d}\psi(\rho)\,d\gamma$, for a function ψ as in Section 4. We can easily recover the result of [11, p. 201] in our setting and deduce some consequences which have already been mentioned in Section 2. For completeness, we state them with a short proof.

THEOREM 6.3. *Let* $\rho\in\mathcal{D}(\Psi)$ *with* $\int_{\mathbb{R}^d}|x|^2\,\rho\,d\gamma<\infty$ *and let* $\rho_t:=S_t\rho$. *Assume that* (2.1)–(2.3) *holds and* h *is concave. Then* $\Psi(\rho_t)<\infty$ *and* $P(\rho_t)<\infty$ *for every* $t>0$, *and we have*

$$\frac{d}{dt}\Psi(\rho_t) = -P(\rho_t)\quad\textit{and}\quad \frac{d}{dt}P(\rho_t) + 2\lambda\,P(\rho_t)\le 0\,.$$

As a consequence, setting $e_\lambda(t)=\int_0^t e^{\lambda r}\,dr$, *we have*

$$e_{2\lambda}(t)P(\rho_t)\le\,\Psi(\rho)\,,\quad P(\rho_t)\le e^{-2\lambda t}\,P(\rho)$$

for any $t \geq 0$. If $\lambda > 0$ we also get

$$\Psi(\rho_t) \leq e^{-2\lambda t}\, \Psi(\rho)$$

and the following entropy – entropy production inequality*, or* generalized Poincaré inequality*, holds*

$$\Psi(\rho) \leq \frac{1}{2\,\lambda}\, P(\rho)\;, \quad \forall\, \rho \in \mathcal{D}(\Psi) \quad \textit{such that} \quad \int_{\mathbb{R}^d} |x|^2\, \rho\, d\gamma < \infty\;.$$

Proof. It is not restrictive to assume that $\int_{\mathbb{R}^d} \rho\, d\gamma = 1$. We first prove Theorem 6.3 for a function h which grows at least linearly at ∞, and therefore satisfies $h(r) \geq \mathsf{h}\, r$ for some constant $\mathsf{h} > 0$. The general result follows by writing h as the limit of a decreasing sequence of such concave functions h_n, observing that the corresponding actions ϕ_n and entropies ψ_n converge increasingly to ϕ and ψ respectively.

By (4.1) we know that $\mathcal{H}(\rho)$ is finite and therefore we have

$$\int_0^\infty \mathcal{I}(\rho_t)\, dt \leq \mathcal{H}(\rho) < \infty \quad \text{and} \quad \int_0^\infty \frac{|\boldsymbol{D}\, L_\psi(\rho_t)|^2}{\rho_t}\, d\gamma < \infty\;,$$

where the second estimate follows from (4.2).

Applying the chain rule for convex functionals in Wasserstein spaces (see for instance [2, p. 233]), we obtain that the map $t \mapsto \Psi(\rho_t)$ is absolutely continuous and

$$-\frac{d}{dt}\Psi(\rho_t) = \int_{\mathbb{R}^d} \frac{\boldsymbol{D}\, L_\psi(\rho_t)}{\rho_t} \cdot \frac{\boldsymbol{D}\rho_t}{\rho_t}\, \rho_t\, d\gamma = P(\rho_t)\;.$$

By combining Theorems 5.4 and 6.1 applied with $\boldsymbol{w}_t := \boldsymbol{D}\rho_t$ and differentiating with respect to t, we get that $-\frac{d}{dt}P(\rho_t) \geq 2\lambda\, P(\rho_t)$. All other estimate are easy consequences that have already been established in Section 2. □

7. Contraction of the h-Wasserstein distance and KFP as a gradient flow. Consider the space $\mathscr{P}_{h,\gamma}(\mathbb{R}^d)$ of probability measures at finite $W_{h,\gamma}$ distance from γ. From (3.1), we know that γ has finite quadratic moments and, as a consequence of [23, Theorem 5.9], any measure in $\mathscr{P}_{h,\gamma}(\mathbb{R}^d)$ also has finite quadratic moments. The same result holds for moments of higher order.

THEOREM 7.1. *For every $\sigma, \eta \in \mathscr{P}_{h,\gamma}(\mathbb{R}^d)$, we have*

$$W_{h,\gamma}(\mathcal{S}_t\sigma\;, \mathcal{S}_t\eta) \leq e^{-\lambda t}\, W_{h,\gamma}(\sigma, \eta) \quad \forall\, t \geq 0\;.$$

Proof. It is a straightforward consequence of Corollary 6.2 and Theorem 3.8. □

THEOREM 7.2. *For every $\mu \in \mathscr{P}_{h,\gamma}(\mathbb{R}^d)$, we have*

$$\frac{1}{2}\frac{d}{dt} W^2_{h,\gamma}(\mathcal{S}_t\mu, \sigma) + \frac{\lambda}{2}\, W^2_{h,\gamma}(\mathcal{S}_t\mu, \sigma) + \Psi(\mathcal{S}_t\mu\,|\,\gamma) \leq \Psi(\sigma\,|\,\gamma) \quad \forall\, \sigma \in \mathcal{D}(\Psi)\;. \tag{7.1}$$

Proof. Let us first notice that since $\mathscr{P}_{h,\gamma}(\mathbb{R}^d)$ is stable under the action of the semigroup $(\mathcal{S}_t)$, it is sufficient to prove (7.1) only at $t = 0$, under the assumption that μ writes as $\mathcal{S}_\tau\tilde{\mu}$, for some $\tau > 0$. We make the additional assumption on the function h that there exists some $\mathsf{h} > 0$ for which

$$h(r) \geq \mathsf{h}\, r \quad \forall\, r > 0\;. \tag{7.2}$$

This assumption will be removed later in the proof. Let $\varepsilon > 0$ fixed and $(\rho^s, \boldsymbol{w}^s) \in L^1(\mathbb{R}^d) \times L^1(\mathbb{R}^d, \mathbb{R}^d)$, $s \in [0,1]$, be an admissible curve connecting σ to μ such that

$$W^2_{h,\gamma}(\mu, \sigma) \le \mathscr{E}_\Phi(\rho^s, \boldsymbol{w}^s) \le W^2_{h,\gamma}(\mu, \sigma) + \varepsilon \,,$$

where $\mathscr{E}_\Phi(\rho^s, \boldsymbol{w}^s) := \int_0^1 \Phi(\rho^s, \boldsymbol{w}^s)\, ds$. For any $\kappa > 0$, we take $\rho^s_\kappa = \rho^s + \kappa \ge \kappa$. Since h is non decreasing, we still have

$$\mathscr{E}_\Phi(\rho^s_\kappa, \boldsymbol{w}^s) \le W^2_{h,\gamma}(\mu, \sigma) + \varepsilon \,. \tag{7.3}$$

Notice that, thanks to [24, Theorem 5.17] (also see Theorem 3.8), it is possible to assume that

$$\mathscr{E}_\Phi(\rho^s_\kappa, \boldsymbol{w}^s) = \Phi(\rho^s_\kappa, \boldsymbol{w}^s) \tag{7.4}$$

is constant with respect to $s \in [0,1]$. For $t > 0$, we set

$$\begin{cases} \rho^{s,t}_\kappa = S_{st} \rho^s_\kappa \,, \\ \boldsymbol{w}^{s,t}_\kappa = \boldsymbol{R}_{st} \boldsymbol{w}^s - t\, \boldsymbol{D} \rho^{s,t}_\kappa \,. \end{cases}$$

It is clear that $(\rho^{s,t}_\kappa)$ connects $\sigma + \kappa\gamma$ to $\mathcal{S}_t(\mu + \kappa\gamma) = \mathcal{S}_t\mu + \kappa\gamma$. Note that, thanks to the maximum principle, we have $\rho^{s,t}_\kappa \ge \kappa$. We claim that $(\rho^{s,t}_\kappa, \boldsymbol{w}^{s,t}_\kappa)$ is admissible. Indeed,

$$\begin{aligned} \partial_s \rho^{s,t}_\kappa &= S_{st}(\partial_s \rho^s_\kappa) + t\, \partial_\tau (S_\tau \rho^s_\kappa)_{|\tau = st} \\ &= -S_{st}(\nabla_\gamma \cdot \boldsymbol{w}^s) + \nabla_\gamma \cdot (t\, \boldsymbol{D} \rho^{s,t}_\kappa) \,, \end{aligned}$$

since $(\rho^s_\kappa, \boldsymbol{w}^s)$ is admissible. Hence,

$$\partial_s \rho^{s,t}_\kappa = \nabla_\gamma \cdot \left(-\boldsymbol{R}_{st} \boldsymbol{w}^s + t\, \boldsymbol{D} \rho^{s,t}_\kappa \right) = -\nabla_\gamma \cdot (\boldsymbol{w}^{s,t}_\kappa) \,.$$

It follows from the definition of $W^2_{h,\gamma}$ that

$$W^2_{h,\gamma}(\mathcal{S}_t\mu + \kappa\gamma, \sigma + \kappa\gamma) \le \mathscr{E}_\Phi(\rho^{s,t}_\kappa, \boldsymbol{w}^{s,t}_\kappa) \,,$$

hence, with (7.3) and (7.4), we obtain

$$\frac{1}{2} \left[W^2_{h,\gamma}(\mathcal{S}_t\mu + \kappa\gamma, \sigma + \kappa\gamma) - W^2_{h,\gamma}(\mu, \sigma) \right] \le \frac{1}{2} \left[\mathscr{E}_\Phi(\rho^{s,t}_\kappa, \boldsymbol{w}^{s,t}_\kappa) - \mathscr{E}_\Phi(\rho^s_\kappa, \boldsymbol{w}^s) \right] + \frac{\varepsilon}{2} \,. \tag{7.5}$$

By definition of $\mathscr{E}_\Phi$, we have

$$\mathscr{E}_\Phi(\rho^{s,t}_\kappa, \boldsymbol{w}^{s,t}_\kappa) = \int_0^1 \Phi\left(S_{st}\rho^s_\kappa, \boldsymbol{R}_{st}\boldsymbol{w}^s - t\, \boldsymbol{D}\rho^{s,t}_\kappa \right) ds \,,$$

where

$$\begin{aligned} \Phi\left(S_{st}\rho^s_\kappa, \boldsymbol{R}_{st}\boldsymbol{w}^s - t\, \boldsymbol{D}\rho^{s,t}_\kappa \right) &= \int_{\mathbb{R}^d} \frac{\left| \boldsymbol{R}_{st}\boldsymbol{w}^s - t\, \boldsymbol{D}\rho^{s,t}_\kappa \right|^2}{h(\rho^{s,t}_\kappa)}\, d\gamma \\ &= \int_{\mathbb{R}^d} \frac{\left| \boldsymbol{R}_{st}\boldsymbol{w}^s \right|^2}{h(\rho^{s,t}_\kappa)}\, d\gamma - 2t \int_{\mathbb{R}^d} \frac{\boldsymbol{D}\rho^{s,t}_\kappa \cdot \boldsymbol{R}_{st}\boldsymbol{w}^s}{h(\rho^{s,t}_\kappa)}\, d\gamma + t^2 \int_{\mathbb{R}^d} \frac{\left| \boldsymbol{D}\rho^{s,t}_\kappa \right|^2}{h(\rho^{s,t}_\kappa)}\, d\gamma \\ &\le \int_{\mathbb{R}^d} \frac{\left| \boldsymbol{R}_{st}\boldsymbol{w}^s \right|^2}{h(\rho^{s,t}_\kappa)}\, d\gamma - 2t \int_{\mathbb{R}^d} \frac{\boldsymbol{D}\rho^{s,t}_\kappa \cdot \boldsymbol{w}^{s,t}_\kappa}{h(\rho^{s,t}_\kappa)}\, d\gamma \,, \end{aligned}$$

and hence,

$$\mathscr{E}_\Phi(\rho_\kappa^{s,t}, \boldsymbol{w}_\kappa^{s,t}) \le \mathscr{E}_\Phi(S_{st}\rho_\kappa^s, \boldsymbol{R}_{st}\boldsymbol{w}^s) - 2t \int_0^1 \int_{\mathbb{R}^d} \frac{\boldsymbol{D}\rho_\kappa^{s,t} \cdot \boldsymbol{w}_\kappa^{s,t}}{h(\rho_\kappa^{s,t})}\, d\gamma\, ds\ . \tag{7.6}$$

LEMMA 7.3. *If* (7.2) *holds, then we have*

$$\int_0^1 \int_{\mathbb{R}^d} \frac{\boldsymbol{D}\rho_\kappa^{s,t} \cdot \boldsymbol{w}_\kappa^{s,t}}{h(\rho_\kappa^{s,t})}\, d\gamma\, ds = \Psi(\mathcal{S}_t\mu + \kappa\gamma \,|\, \gamma) - \Psi(\sigma + \kappa\gamma \,|\, \gamma)\ . \tag{7.7}$$

Proof. Recall that $\rho_\kappa^{s,t} = S_{st}\rho_\kappa^s$, with $\rho_\kappa^s \in L^1_\gamma(\mathbb{R}^d)$. Then, acting as in the proof of Proposition 4.2, we get that

$$\int_\tau^1 \int_{\mathbb{R}^d} \frac{\left|\boldsymbol{D}\rho_\kappa^{s,t}\right|^2}{\rho_\kappa^{s,t}}\, d\gamma\, ds \,<\, \infty \quad \forall\, \tau > 0\ . \tag{7.8}$$

The assumption (7.2) on h then leads to

$$\int_\tau^1 \int_{\mathbb{R}^d} \frac{\left|\boldsymbol{D}L_\psi(\rho_\kappa^{s,t})\right|^2}{\rho_\kappa^{s,t}}\, d\gamma\, ds \le \frac{1}{\mathsf{h}^2} \int_\tau^1 \int_{\mathbb{R}^d} \frac{\left|\boldsymbol{D}\rho_\kappa^{s,t}\right|^2}{\rho_\kappa^{s,t}}\, d\gamma\, ds \,<\, \infty\ .$$

The next step consists in proving that

$$\int_\tau^1 \int_{\mathbb{R}^d} \frac{\left|\boldsymbol{w}_\kappa^{s,t}\right|^2}{\rho_\kappa^{s,t}}\, d\gamma\, ds \,<\, \infty\ . \tag{7.9}$$

Note that $\rho_\kappa^{s,t} \ge \kappa$ and the concavity of h implies that

$$h(\rho_\kappa^{s,t}) \le \frac{h(\kappa)}{\kappa}\, \rho_\kappa^{s,t}\ ,$$

hence

$$\begin{aligned}\int_\tau^1 \int_{\mathbb{R}^d} \frac{\left|\boldsymbol{R}_{st}\boldsymbol{w}^s\right|^2}{\rho_\kappa^{s,t}}\, d\gamma\, ds &\le \frac{h(\kappa)}{\kappa} \int_\tau^1 \int_{\mathbb{R}^d} \frac{\left|\boldsymbol{R}_{st}\boldsymbol{w}^s\right|^2}{h(\rho_\kappa^{s,t})}\, d\gamma\, ds \\ &\le \frac{h(\kappa)}{\kappa} \int_\tau^1 \int_{\mathbb{R}^d} \frac{\left|\boldsymbol{w}^s\right|^2}{h(\rho_\kappa^s)}\, d\gamma\, ds < \infty\end{aligned}$$

since the KFP flow decreases the action. The bound (7.9) immediately follows from the previous one and (7.8). As a consequence, we can apply the chain rule in Wasserstein space, which implies that the function $s \mapsto \Psi(\rho_\kappa^{s,t})$ is absolutely continuous on $[\tau, 1]$ and, for all $s \in [\tau, 1]$,

$$\frac{d}{ds} \int_{\mathbb{R}^d} \psi(\rho_\kappa^{s,t})\, d\gamma = \int_{\mathbb{R}^d} \frac{\boldsymbol{D}L_\psi(\rho_\kappa^{s,t})}{\rho_\kappa^{s,t}} \cdot \frac{\boldsymbol{w}_\kappa^{s,t}}{\rho_\kappa^{s,t}}\, \rho_\kappa^{s,t}\, d\gamma = \int_{\mathbb{R}^d} \frac{\boldsymbol{D}\rho_\kappa^{s,t} \cdot \boldsymbol{w}_\kappa^{s,t}}{h(\rho_\kappa^{s,t})}\, d\gamma\ . \tag{7.10}$$

Integrating (7.10) on $[\tau, 1]$ and letting τ go to 0 finally leads to (7.7). □

Let us go back to the proof of Theorem 7.2. We put (7.5) and (7.6) together and obtain

$$\begin{aligned}&\frac{1}{2}\left[W^2_{h,\gamma}(\mathcal{S}_t\mu + \kappa\gamma, \sigma + \kappa\gamma) - W^2_{h,\gamma}(\mu, \sigma)\right) \\ &\le \frac{1}{2}\left[\mathscr{E}_\Phi(S_{st}\rho_\kappa^s, \boldsymbol{R}_{st}\boldsymbol{w}^s) - \mathscr{E}_\Phi(\rho_\kappa^s, \boldsymbol{w}^s)\right] + t\left[\Psi(\sigma + \kappa\gamma \,|\, \gamma) - \Psi(\mathcal{S}_t\mu + \kappa\gamma \,|\, \gamma)\right] + \frac{\varepsilon}{2}\ .\end{aligned}$$

We then use the main estimate in Theorem 6.1 with $\beta = 0$ and (7.4) to write

$$\frac{1}{2}\left[\mathscr{E}_\Phi(S_{st}\rho^s_\kappa, \boldsymbol{R}_{st}\boldsymbol{w}^s) - \mathscr{E}_\Phi(\rho^s_\kappa, \boldsymbol{w}^s)\right] \le -\frac{1}{2}\, I_\lambda(t)\, \mathscr{E}_\Phi(\rho^s_\kappa, \boldsymbol{w}^s)$$
$$\le -\frac{1}{2}\, I_\lambda(t)\, W^2_{h,\gamma}(\mathcal{S}_t\mu + \kappa\gamma, \sigma + \kappa\gamma)\ ,$$

where $I_\lambda(t) := \int_0^1 \left(1 - e^{-2\lambda t s}\right) ds$. It follows that

$$\frac{1}{2}\left[W^2_{h,\gamma}(\mathcal{S}_t\mu + \kappa\gamma, \sigma + \kappa\gamma) - W^2_{h,\gamma}(\mu, \sigma)\right] + \frac{1}{2}\, I_\lambda(t)\, W^2_{h,\gamma}(\mathcal{S}_t\mu + \kappa\gamma, \sigma + \kappa\gamma)$$
$$\le t\left[\Psi(\sigma + \kappa\gamma\,|\,\gamma) - \Psi(\mathcal{S}_t\mu + \kappa\gamma\,|\,\gamma)\right] + \frac{\varepsilon}{2}\ .$$

If we first let ε and then κ go to 0 in the above estimate, we get that

$$\frac{1}{2}\left[W^2_{h,\gamma}(\mathcal{S}_t\mu, \sigma) - W^2_{h,\gamma}(\mu, \sigma)\right] + \frac{1}{2}\, I_\lambda(t)\, W^2_{h,\gamma}(\mathcal{S}_t\mu, \sigma) \le t\left[\Psi(\sigma\,|\,\gamma) - \Psi(\mathcal{S}_t\mu\,|\,\gamma)\right] \tag{7.11}$$

as soon as h satisfies the assumption (7.2). Now, any concave and non decreasing function h can be decreasingly approched by a sequence (h_n) satisfying (7.2), and the corresponding entropies converge increasingly. Then, with Theorem 3.7, Inequality (7.11) turns out to be valid for any general h. To complete the proof of Theorem 7.2, it just remains to divide (7.11) by t and let t go to 0. □

Acknowledgements. This work has been partially supported by the projects CBDif, EVOL and OTARIE of the French National Research Agency (ANR) and by a PRIN08-grant from MIUR for the project *Optimal transport theory, geometric and functional inequalities, and applications.*

REFERENCES

[1] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, Oxford Mathematical Monographs, Clarendon Press, Oxford, 2000.

[2] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, second ed., 2008.

[3] L. Ambrosio, N. Gigli, and G. Savaré, *Metric measure spaces with Riemannian Ricci curvature bounded from below*, Arxiv 1109.0222, (2011), pp. 1–60.

[4] L. Ambrosio and G. Savaré, *Gradient flows of probability measures*, in Handbook of Evolution Equations (III), Elsevier, 2006.

[5] L. Ambrosio, G. Savaré, and L. Zambotti, *Existence and stability for Fokker-Planck equations with log-concave reference measure*, Probab. Theory Related Fields, 145 (2009), pp. 517–564.

[6] A. Arnold, J.-P. Bartier, and J. Dolbeault, *Interpolation between logarithmic Sobolev and Poincaré inequalities*, Commun. Math. Sci., 5 (2007), pp. 971–979.

[7] A. Arnold and J. Dolbeault, *Refined convex Sobolev inequalities*, J. Funct. Anal., 225 (2005), pp. 337–351.

[8] A. Arnold, P. Markowich, G. Toscani, and A. Unterreiter, *On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker-Planck type equations*, Comm. Partial Differential Equations, 26 (2001), pp. 43–100.

[9] D. Bakry, *Functional inequalities for Markov semigroups*, in Probability measures on groups: recent directions and trends, Tata Inst. Fund. Res., Mumbai, 2006, pp. 91–147.

[10] D. Bakry and M. Émery, *Hypercontractivité de semi-groupes de diffusion*, C. R. Acad. Sci. Paris Sér. I Math., 299 (1984), pp. 775–778.

[11] D. Bakry and M. Émery, *Diffusions hypercontractives*, in Séminaire de probabilités, XIX, 1983/84, vol. 1123 of Lecture Notes in Math., Springer, Berlin, 1985, pp. 177–206.
[12] D. Bakry, M. Ledoux, and F.-Y. Wang, *Perturbations of functional inequalities using growth conditions*, J. Math. Pures Appl. (9), 87 (2007), pp. 394–407.
[13] F. Barthe, P. Cattiaux, and C. Roberto, *Interpolated inequalities between exponential and Gaussian, Orlicz hypercontractivity and isoperimetry*, Rev. Mat. Iberoam., 22 (2006), pp. 993–1067.
[14] F. Barthe, P. Cattiaux, and C. Roberto, *Isoperimetry between exponential and Gaussian*, Electron. J. Probab., 12 (2007), pp. no. 44, 1212–1237.
[15] W. Beckner, *A generalized Poincaré inequality for Gaussian measures*, Proc. Amer. Math. Soc., 105 (1989), pp. 397–400.
[16] 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.
[17] F. Bolley and I. Gentil, *Phi-entropy inequalities for diffusion semigroups*, J. Math. Pures Appl. (9), 93 (2010), pp. 449–473.
[18] J. A. Carrillo, S. Lisini, G. Savaré, and D. Slepčev, *Nonlinear mobility continuity equations and generalized displacement convexity*, J. Funct. Anal., 258 (2010), pp. 1273–1309.
[19] D. Chafaï, *Entropies, convexity, and functional inequalities: on Φ-entropies and Φ-Sobolev inequalities*, J. Math. Kyoto Univ., 44 (2004), pp. 325–363.
[20] S. Daneri and G. Savaré, *Eulerian calculus for the displacement convexity in the Wasserstein distance*, SIAM J. Math. Anal., 40 (2008), pp. 1104–1122.
[21] P. Deng and F. Wang, *Beckner inequality on finite- and infinite-dimensional manifolds*, Chinese Ann. Math. Ser. B, 27 (2006), pp. 581–594.
[22] J. Dolbeault, I. Gentil, A. Guillin, and F.-Y. Wang, *L^q-functional inequalities and weighted porous media equations*, Potential Anal., 28 (2008), pp. 35–59.
[23] J. Dolbeault, B. Nazaret, and G. Savaré, *On the Bakry-Emery criterion for linear diffusions and weighted porous media equations*, Commun. Math. Sci., 6 (2008), pp. 477–494.
[24] J. Dolbeault, B. Nazaret, and G. Savaré, *A new class of transport distances between measures*, Calc. Var. Partial Differential Equations, 34 (2009), pp. 193–231.
[25] U. Gianazza, G. Savaré, and G. Toscani, *The Wasserstein gradient flow of the Fisher information and the quantum drift-diffusion equation*, Arch. Ration. Mech. Anal., 194 (2009), pp. 133–220.
[26] F. John, *Hyperbolic and parabolic equations*, in Partial Differential Equations (Proc. Summer Seminar, Boulder, Col., 1957), Interscience, New York, 1964, pp. 1–129.
[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] R. Latała and K. Oleszkiewicz, *Between Sobolev and Poincaré*, in Geometric aspects of functional analysis, vol. 1745 of Lecture Notes in Math., Springer, Berlin, 2000, pp. 147–168.
[29] S. Lisini, D. Matthes, and G. Savaré, *Cahn-Hilliard and Thin Film equations with nonlinear mobility as gradient flows in weighted-Wasserstein metrics*, Arxiv 1201.2367, (2012), pp. 1–28.
[30] F. Otto and M. Westdickenberg, *Eulerian calculus for the contraction in the Wasserstein distance*, SIAM J. Math. Anal., 37 (2005), pp. 1227–1255.
[31] C. Villani, *Optimal transport. Old and new*, vol. 338 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, 2009.
[32] M.-K. von Renesse and K.-T. Sturm, *Transport inequalities, gradient estimates, entropy, and Ricci curvature*, Comm. Pure Appl. Math., 58 (2005), pp. 923–940.
[33] F. Wang, *Functional inequalities, markov processes and spectral theory*, Sci. Press, Beijing, (2004).