


Nonlinear kinetic Fokker–Planck equations as gradient flows of the free energy


Giovanni Brigati, Guillaume Carlier, Jean Dolbeault, Filippo Quattrocchi


Abstract We consider nonhomogeneous kinetic equations that involve a free transport operator and a diffusion of porous medium type acting on velocities. The main novelty is a gradient flow interpretation of dynamics driven by an interplay of conservative and dissipative effects. We rely on a notion of discrepancy adapted to a phase space of positions and velocities, built upon second-order characteristics obeying Newton’s laws. The equation appears as the steepest descent of the free energy functional. We also prove that approximate solutions constructed with an implicit Euler scheme converge to a solution of the kinetic equation. Thus, we generalise to a family of nonlinear kinetic equations the celebrated JKO scheme in mass transport theory. Most of our results are new even in the case of the linear Vlasov–Fokker–Planck equation.


2020 MSC Primary: [49Q22](#). Secondary: [82C40](#); [35Q49](#); [35K55](#).

Keywords Kinetic optimal transport · JKO scheme · Gradient flow

Giovanni Brigati 
ISTA, Institute of Science and Technology Austria, Am Campus 1, Klosterneuburg, 3400, Austria,
e-mail: giovanni.brigati@ist.ac.at

Guillaume Carlier 
CEREMADE, Centre de Recherche en Mathématiques de la Décision (CNRS UMR n° 7534),
PSL University, Université Paris-Dauphine, Place de Lattre de Tassigny, 75775, Paris 16, France, and
Inria-Paris, Mokaplan,
e-mail: carlier@ceremade.dauphine.fr

Jean Dolbeault 
CEREMADE, Centre de Recherche en Mathématiques de la Décision (CNRS UMR n° 7534),
PSL University, Université Paris-Dauphine, Place de Lattre de Tassigny, 75775, Paris 16, France,
e-mail: dolbeaul@ceremade.dauphine.fr

Filippo Quattrocchi 
LMO, Laboratoire de Mathématiques d’Orsay, Faculté des Sciences d’Orsay, Université Paris-Saclay,
307 rue Michel Magat, 91405, Orsay, France,
e-mail: filippo.quattrocchi@universite-paris-saclay.fr

1 Introduction and main results

For any $m > 1$, we consider the nonlinear kinetic Fokker–Planck equation

$$\partial_t f + v \cdot \nabla_x f = \nabla_v \cdot (f (\nabla_v P_m(f) + v)) \quad (1)$$

for $t \geq 0$, over the phase space $\Gamma := \mathbb{R}^d \times \mathbb{R}^d \ni (x, v)$ of positions and velocities, where f is a nonnegative distribution function and

$$P_m(f) := \frac{m}{m-1} f^{m-1}$$

is the pressure variable. With $P_1(f) := \log f$, the case $m = 1$ is the linear Vlasov–Fokker–Planck equation corresponding to the second-order Langevin process. Equation (1) admits a Lyapunov functional, namely the *free energy* \mathcal{F}_m given by

$$\mathcal{F}_m(f) := \iint_{\Gamma} \left(\frac{1}{m} P_m(f) + \frac{1}{2} |v|^2 \right) f dx dv.$$

Indeed, for a smooth solution of (1) with sufficient decay properties, we have

$$\frac{d}{dt} \mathcal{F}_m(f(t, \cdot, \cdot)) = -\mathcal{I}_m(f) := -\iint_{\Gamma} |\nabla_v P_m(f) + v|^2 f dx dv \leq 0,$$

where $\mathcal{I}_m(f)$ denotes a generalized *Fisher information* functional. Characterising the nonhomogeneous kinetic equation (1) as a gradient flow of the energy \mathcal{F}_m requires a notion of *discrepancy*, tailored on the second-order nature of Newton’s equations

$$x'(t) = v(t), \quad v'(t) = F_t(x(t), v(t))$$

associated to a force field $\mathbb{R}^+ \times \Gamma \ni (t, x, v) \mapsto F_t(x, v)$ with values in \mathbb{R}^d . On a finite time interval $[0, T] \ni t$, a curve of measures $(\mu_t)_t$ subject to a measurable force field $(F_t)_t$ with values in $L^2(\mu_t; \mathbb{R}^d)$ gets transported according to the Vlasov equation

$$\partial_t \mu_t + v \cdot \nabla_x \mu_t + \nabla_v \cdot (F_t \mu_t) = 0 \quad \text{in } \mathcal{D}'((0, T) \times \Gamma). \quad (2)$$

Let $\mathcal{P}_2(\Gamma)$ be the set of probability measures μ on Γ with $\iint_{\Gamma} (|x|^2 + |v|^2) d\mu < \infty$. As in [6], given $\mu, \nu \in \mathcal{P}_2(\Gamma)$ and $T > 0$, we define

$$d_T^2(\mu, \nu) := \inf_{(\mu_t, F_t)_{t \in [0, T]}} T \int_0^T \|F_t\|_{L^2(\mu_t)}^2 dt, \quad \mu_0 = \mu, \quad \mu_T = \nu, \quad (3)$$

by minimising over all admissible $(\mu_t, F_t)_t$ to (2). In [6, Theorems 1.4–1.7], d_T -regular curves of measures $(\mu_t)_t$ are identified with solutions to (2) and

$$|\mu'_t|_d := \lim_{h \rightarrow 0_+} \frac{d_h(\mu_t, \mu_{t+h})}{h}$$

is well defined for almost every t . We shall write $\mathcal{F}_m(\mu_t) = +\infty$ and $\mathcal{I}_m(\mu_t) = +\infty$ if μ_t is not absolutely continuous with respect to Lebesgue's measure. For a smooth function f , we have $\nabla_v \cdot (f \nabla_v P_m(f)) = \Delta_v f^m$, $\sqrt{f} \nabla_v P_m(f) = \frac{m}{m-1/2} \nabla_v (f^{m-1/2})$ and

$$\mathcal{I}_m(f) = \iint_{\Gamma} \left(\left(\frac{2m}{2m-1} \right)^2 |\nabla_v (f^{m-1/2})|^2 + |v|^2 f - 2d f^m \right) dx dv \quad (4)$$

which are relevant quantities for nonsmooth solutions to (2). Nonnegative weak solutions $f \in L^1_{\text{loc}}(\mathbb{R}^+; L^1(1 + |v|^2, \Gamma)) \cap L^m_{\text{loc}}(\mathbb{R}^+; L^m(\Gamma))$ have to be understood in the distributional sense and written in terms of f and f^m only. See Lemma 1 in Section 2 for details and consequences. With these conventions, we can state our first main result.

Theorem 1 (Chain rule for the free energy and gradient flow). *Let $m \in [1, 3/2]$ and $T > 0$. If $\mu_t = f(t, \cdot, \cdot)$ is in $\mathcal{P}_2(\Gamma)$ for a.e. t and solves (2), with $\int_0^T (\mathcal{F}_m(\mu_t) + \mathcal{I}_m(\mu_t)) dt + \int_0^T \|F_t\|_{L^2(\mu_t)}^2 dt < \infty$, then, for a negligible set $A \subset [0, T]$ and a.e. $(a, b) \in ([0, T] \setminus A)^2$ with $b > a$,*

$$\begin{aligned} \mathcal{F}_m(\mu_b) - \mathcal{F}_m(\mu_a) &= \int_a^b \iint_{\Gamma} F_t \cdot (\nabla_v P_m(\mu_t) + v) \mu_t dx dv dt \\ &\geq -\frac{1}{2} \int_a^b \left(|\mu'_t|_{\text{d}}^2 + \mathcal{I}_m(\mu_t) \right) dt. \end{aligned} \quad (5)$$

Equality in the inequality of (5) is attained if and only if f is a solution to the nonlinear kinetic Fokker–Planck equation (1) on the time interval $[0, T]$. Moreover, in this case,

$$|\mu'_t|_{\text{d}} = \mathcal{I}_m(\mu_t). \quad (6)$$

In other words, among solutions of Vlasov's equations, all curves of maximal energy dissipation (Fisher information) solve (1). Our proof crucially relies on convexity properties of the Fisher information, which are only ensured under the condition $m \in [1, 3/2]$. Next, let us consider a second-order version of the JKO scheme.

Theorem 2 (Minimising-movement scheme and maximal dissipation).

Let $m \geq 1$. The scheme starting at $\mu_0^{(h)} = \mu_0 \in \mathcal{P}_2(\Gamma)$ with $\mathcal{F}_m(\mu_0) < \infty$, and

$$\mu_{k+1}^{(h)} \in \arg \min_{\nu \in \mathcal{P}_2(\Gamma)} \left(\mathcal{F}_m(\nu) + \frac{1}{2h} d_h^2(\mu_k^{(h)}, \nu) \right) \quad (7)$$

is well-posed for any $h \in (0, 1)$ and $k \in \mathbb{N}$. The piecewise constant interpolations $(\mu_t^{(h)})_t$ are relatively compact in $L^1_{\text{loc}}(\mathbb{R}^+; L^1(\Gamma))$. Up to the extraction of a sequence, any limit $\mu = \lim_{h \rightarrow 0_+} \mu^{(h)}$ solves (1) in $\mathcal{D}'(\mathbb{R}^+ \times \Gamma)$ with initial datum μ_0 . Moreover, if $m \in [1, 3/2]$, μ dissipates the energy maximally, in the sense that (5) is an equality.

One crucial point of our scheme (7) is *strong compactness*, which we establish by extending the L^1 -contraction argument of [19] (also see [22]) beyond the case

of bounded domains. Contraction entails uniform equi-continuity with respect to x -translations, while the Fisher information \mathcal{I}_m bounds the oscillations in the v -variable. Thus, we control the relevant nonlinear quantities along the scheme and identify limit curves μ with solutions to (1). The threshold $m = 3/2$ does not play a role here, as \mathcal{I}_m is lower semicontinuous in the strong topology for any $m \geq 1$.

Moreover, we prove that limit curves produced by the scheme satisfy some reverse form of Inequality (5) for any $m \geq 1$, see (55). When $m \in [1, 3/2]$, the maximal dissipation property of Theorem 1 follows, *i.e.*, we have equality in (5). This gradient flow interpretation of (1) is new, even in the case $m = 1$, and the strong compactness for (7) may be of independent interest.

Equation (1) is probably an oversimplified nonlinear kinetic equation for significant applications in physics. However, the gradient flow property not only makes sense in a functional framework of mathematical interest, but also admits an interpretation in mechanics that may apply in more realistic models.

Let us conclude this introduction with a brief review of the literature. The *JKO scheme* [20] (also see [27, Section 4.4]) characterises the linear Fokker–Planck equation as the maximal dissipation flow of the free energy according to the Wasserstein distance. This property was formally extended to nonlinear diffusions by *Otto’s calculus* in [23] and later justified as a gradient flow in [3]. In linear diffusion models, adapted free energy functionals and distances can be chosen simultaneously [10]. *Minimising-movement schemes for kinetic equations* appeared in [16, 17], and later in [4, 12, 11]. For the classical alternative of splitting schemes, we refer to [24] and references therein. Various notions of optimal transport distances adapted to Newton’s laws have been considered for instance in [17, 16, 15, 4, 6, 21], which also play a role in *Wasserstein stability* of solutions to Vlasov’s equation, including mean-field forces, Poisson kernels and Fokker–Planck diffusions in velocities (see, *e.g.*, [18] and references therein). Such an approach has interesting consequences for large time asymptotics, uniqueness, particle approximation and propagation of molecular chaos, which are out of our scope. However, we can refer to [21] for a recent numerical implementation with $m = 1$, with or without self-consistent mean field Poisson coupling. In this paper, we rely on the *kinetic optimal transport* discrepancy d_T of [6] in order to characterise (1) as the flow of maximal dissipation of the free energy for $m \geq 1$. The restriction $m \leq 3/2$ appears, for instance, also in [14]. Apart from [1], which is devoted to a model of granular media with a different approach, our gradient flow interpretation is new in kinetic theory, even in the linear case $m = 1$, while the (weak) convergence of (7) is already known from [12] if $m = 1$. The first key ingredient is the *chain rule*, introduced in [3, Chapter 10, pp. 228 & 233 (E)] in the context of Wasserstein gradient flows. The proof via regularisation by convolution originates from [13, 8] using the *Galilean group* as in [28]. The second key ingredient is strong compactness for (7), as discussed above. Similar considerations, adapted to a splitting scheme, are also crucial in [5].

This paper is organised as follows. Theorem 1 is proved in Section 2. The proof of Theorem 2 is split into two sections: convergence of the JKO scheme to a solution to (1) in Section 3; maximal dissipation in the range $m \in [1, 3/2]$ in Section 4.

2 Chain rule for the free energy and gradient flow

This section is devoted to the proof of Theorem 1. Without further notice, we shall assume that all corresponding assumptions are satisfied.

Step 1: Regularisation via Galilean convolution. Let $(\mu, F) = (\mu_t, F_t)_t$ be as in Theorem 1 and denote by f_t the density of μ_t . We extend (in time) the force field F by 0 outside $[0, T]$ and then extend f by solving (2) for all times. Let ψ be a smooth probability density supported on $[-1, 0]$ with $\psi > 0$ on $(-1, 0)$ and for $\epsilon \in (0, 1)$, let us consider the mollifier on $\mathbb{R} \times \Gamma$ defined by

$$\eta_\epsilon(t, x, v) := \frac{1}{\pi^d \epsilon^{1+2d}} \psi\left(\frac{t}{\epsilon}\right) e^{-\frac{1}{\epsilon^2}(|x|^2 + |v|^2)}.$$

Inspired by [28], we regularise f and F by

$$f^\epsilon := \eta_\epsilon \star f \quad \text{and} \quad F^\epsilon := (\eta_\epsilon \star (F f)) / f^\epsilon \quad (8)$$

where the *Galilean convolution* \star is defined by

$$(\eta_\epsilon \star f)(t, x, v) := \int_{\mathbb{R}} \iint_{\Gamma} \eta_\epsilon(t-s, x-y-s(v-w), v-w) f(s, y, w) dy dw ds.$$

According to [6, Lemma 4.9], $(f_t^\epsilon, F_t^\epsilon)_t$ then is a smooth classical solution to (2) on $\mathbb{R} \times \Gamma$. Moreover, it is a strong approximation in the sense that f^ϵ and $f^\epsilon F^\epsilon$ converge strongly in $L^1_{\text{loc}}(\mathbb{R}; L^1(\Gamma))$ to f and fF respectively. Passing to a vanishing sequence of ϵ 's, we may also assume that the previous convergence holds a.e. in (t, x, v) and, for a.e. t , in $L^1(\Gamma)$. Again following [6, Lemma 4.9], we have

$$\int_0^T \|F_t^\epsilon\|_{L^2(f_t^\epsilon)}^2 dt \leq \int_0^T \|F_t\|_{L^2(\mu_t)}^2 dt \quad (9)$$

for any $\epsilon > 0$ and

$$\begin{aligned} \lim_{\epsilon \rightarrow 0_+} \int_0^T \iint_{\Gamma} |x|^2 f_t^\epsilon dt &= \int_0^T \iint_{\Gamma} |x|^2 d\mu_t dt, \\ \lim_{\epsilon \rightarrow 0_+} \int_0^T \iint_{\Gamma} |v|^2 f_t^\epsilon dt &= \int_0^T \iint_{\Gamma} |v|^2 d\mu_t dt. \end{aligned} \quad (10)$$

By (9), $\sqrt{f^\epsilon} F^\epsilon$ is bounded in $L^2((0, T) \times \Gamma)$ and, up to the extraction of sequences, converges weakly to some limit G . By strong L^2 convergence of $\sqrt{f^\epsilon}$ to \sqrt{f} and strong L^1 convergence of $f^\epsilon F^\epsilon$ to fF we readily have $fF = \sqrt{f}G$, $G = \sqrt{f}F$. Weak L^2 convergence of $\sqrt{f^\epsilon} F^\epsilon$ to $\sqrt{f}F$ is in fact strong because of the convergence of the L^2 norm deduced from (9). Hence, we have

$$\lim_{\epsilon \rightarrow 0_+} \left\| \sqrt{f^\epsilon} F^\epsilon - \sqrt{f} F \right\|_{L^2((0, T) \times \Gamma)} = 0. \quad (11)$$

By the same reasoning, from the second condition in (10), we deduce, after passing to a subsequence if necessary,

$$\lim_{\epsilon \rightarrow 0_+} \|v(\sqrt{f^\epsilon} - \sqrt{f})\|_{L^2((0,T) \times \Gamma)} = 0 \quad \text{and} \quad \lim_{\epsilon \rightarrow 0_+} \iint_{\Gamma} |v|^2 f_t^\epsilon = \iint_{\Gamma} |v|^2 d\mu_t \quad (12)$$

where the second equality holds for a.e. t .

Step 2: Bounds in free energy and Fisher information. We rely on Jensen's inequality and on the *data processing inequality*: integrated convex functions decrease under convolution. See for instance [9, Chapter 2]. If $m > 1$, using the data processing inequality and letting $\epsilon \rightarrow 0_+$ gives

$$\|f_t\|_{L^m(\Gamma)}^m \leq \liminf_{\epsilon \rightarrow 0_+} \|f_t^\epsilon\|_{L^m(\Gamma)}^m \leq \limsup_{\epsilon \rightarrow 0_+} \frac{1}{\epsilon} \int_{\mathbb{R}} \psi\left(\frac{s}{\epsilon}\right) \|f_{t-s}\|_{L^m(\Gamma)}^m ds = \|f_t\|_{L^m(\Gamma)}^m$$

for a.e. t , where the first inequality holds by Fatou's Lemma. For $m = 1$, we can argue similarly, using second moments (see for instance [20]) to bound the negative part of $f^\epsilon \log(f^\epsilon)$, and Fatou's Lemma. Using again (12), for any $m \geq 1$, we obtain

$$\lim_{\epsilon \rightarrow 0_+} \mathcal{F}_m(\mu_t^\epsilon) = \mathcal{F}_m(\mu_t) \quad \text{for a.e. } t \in [0, T]. \quad (13)$$

Next, let us turn our attention to the Fisher information $\mathcal{I}_m(f_t)$ which is more delicate. First of all, to define $\mathcal{I}_m(f_t)$ rigorously via (4), we shall use the following lemma (whose standard proof by approximation is left to the reader).

Lemma 1 *Let $\alpha \in (0, +\infty)$ and let L be a homogeneous first-order differential operator. If $g \in L^1_{\text{loc}}(\Gamma)$ is nonnegative, $Lg \in L^1_{\text{loc}}(\Gamma)$, and $g^{\alpha-1} Lg \in L^1_{\text{loc}}(\Gamma)$, then*

$$L(g^\alpha) = \alpha g^{\alpha-1} Lg.$$

Applied with $L = \nabla_v$, $f = g^\alpha$, $\alpha = 1/(m - 1/2)$ and $m \in [1, 3/2]$, if $f \in L^m_{\text{loc}}(\Gamma)$, and $\nabla_v(f^{m-1/2}) \in L^2_{\text{loc}}(\Gamma)$, then

$$\nabla_v f = \frac{1}{m-1/2} f^{3/2-m} \nabla_v(f^{m-1/2}) \in L^1_{\text{loc}}(\Gamma).$$

The key observation now is that the function

$$\mathbb{R} \times \mathbb{R}^d \ni (s, X) \mapsto \Theta(s, X) := \begin{cases} s^{2m-3} |X|^2 & \text{if } s > 0, \\ 0 & \text{if } s = 0 \text{ and } X = 0, \\ +\infty & \text{otherwise,} \end{cases} \quad (14)$$

is jointly convex and lower semicontinuous for $1 \leq m \leq 3/2$ (and actually only in this range for m). Our assumptions on the curve f then ensure that $\nabla_v f^{m-1/2}$ is L^2 , and thanks to Lemma 1, the identity

$$\frac{m^2}{(m-1/2)^2} \iint_{\Gamma} |\nabla_v f^{m-1/2}|^2 = m^2 \iint_{\Gamma} \Theta(f, \nabla_v f)$$

holds for $f = f_t$ and a.e. t . Recalling that we interpret $\sqrt{f} \nabla_v P_m(f)$ as $\frac{m}{m-1/2} \nabla_v f^{m-1/2}$, we write

$$m^2 \iint_{\Gamma} \Theta(f, \nabla_v f) = \|\sqrt{f} \nabla_v P_m(f)\|_{L^2(\Gamma)}^2.$$

These considerations lead to the following.

Lemma 2 *For any a and b such that $0 < a < b < T$, we have*

$$\lim_{\epsilon \rightarrow 0^+} \left\| (\nabla_v P_m(f^\epsilon) + v) \sqrt{f^\epsilon} - (\nabla_v P_m(f) + v) \sqrt{f} \right\|_{L^2((a,b) \times \Gamma)} = 0. \quad (15)$$

In addition, for any fixed $\epsilon > 0$, there is a constant $C(T, \epsilon) > 0$ such that

$$\int_a^b \iint_{\Gamma} |\nabla_x P_m(f^\epsilon)|^2 f^\epsilon dx dv dt \leq C(T, \epsilon). \quad (16)$$

Proof. Thanks to (12), to establish (15), we just need to prove that $\sqrt{f^\epsilon} \nabla_v P_m(f^\epsilon)$ converges to $\sqrt{f} \nabla_v P_m(f)$ in $L^2((a, b) \times \Gamma)$. By (8) and Lemma 1, we have

$$\nabla_v f^\epsilon(t, x, v) = \int_t^{t+\epsilon} \iint_{\Gamma} \eta^\epsilon(t-s, x-y-s(v-w), v-w) \nabla_w f(s, y, w) dy dw ds.$$

We now apply Jensen's inequality to the function Θ defined in (14) and deduce from the assumption $\int_0^T \mathcal{I}_m(f) dt < \infty$ that

$$\begin{aligned} \int_a^b \iint_{\Gamma} |\nabla_v P_m(f^\epsilon)|^2 f^\epsilon dx dv dt &= m^2 \int_a^b \iint_{\Gamma} \Theta(f^\epsilon, \nabla_v f^\epsilon) dx dv dt \\ &\leq m^2 \int_a^{b+\epsilon} \iint_{\Gamma} \Theta(f, \nabla_v f) dx dv dt = \int_a^b \iint_{\Gamma} |\nabla_v P_m(f)|^2 f dx dv dt + o(1). \end{aligned} \quad (17)$$

This implies that $\sqrt{f^\epsilon} \nabla_v P_m(f^\epsilon)$ is bounded in L^2 and since $(f^\epsilon)^{m-1/2}$ converges to $f^{m-1/2}$ in $L^{1/(m-1/2)}$, the weak limit of $\sqrt{f^\epsilon} \nabla_v P_m(f^\epsilon)$ necessarily is $\sqrt{f} \nabla_v P_m(f) = \frac{m}{m-1/2} \nabla_v f^{m-1/2}$. But thanks to (17), we see that the convergence is in fact strong. This completes the proof of (15).

To show (16), we proceed in a similar way, except that the derivative with respect to x acts on the mollifier. More precisely, according to [28, p. 6], we have

$$\nabla_x f^\epsilon(t, x, v) = \int_t^{t+\epsilon} \iint_{\Gamma} (\nabla_y \eta^\epsilon)(s, y, w) f(t-s, x-y-s(v-w), v-w) dy dw ds.$$

Using the homogeneity of Θ and Jensen's inequality again yields

$$\int_a^b \iint_{\Gamma} \Theta(f^\epsilon, \nabla_x f^\epsilon) dx dv dt \leq (T+\epsilon)^{2m-1} \int_{-\epsilon}^0 \iint_{\Gamma} \Theta(\eta^\epsilon, \nabla_x \eta^\epsilon) dx dv dt =: C(T, \epsilon)$$

which is finite for fixed $\epsilon > 0$. This proves (16). \square

Step 3: Chain rule for regular solutions. Thanks to (17)-(16), the Wasserstein slope (see [3, Theorem 10.4.13]) of \mathcal{F}_m satisfies

$$\iint_{\Gamma} |\nabla_{x,v} P_m(f^\epsilon) - (0, v)|^2 f^\epsilon dx dv = |\partial \mathcal{F}_m(f^\epsilon)|^2 \in L^1(0, T)$$

and \mathcal{F}_m is convex along Wasserstein geodesics (see [3, Section 9.3]). At the same time, $(\mu^\epsilon)_t$ is a 2-absolutely continuous curve in the Wasserstein distance. According to [3, p. 233], the *chain rule* applies for a.e. $t \in (0, T)$, that is,

$$\frac{d}{dt} \mathcal{F}_m(\mu_t^\epsilon) = \iint_{\Gamma} (\nabla_x P_m(f^\epsilon) \cdot v f^\epsilon + (\nabla_v P_m(f^\epsilon) + v) \cdot F^\epsilon f^\epsilon) dx dv. \quad (18)$$

The first term in the right-hand side, however, vanishes thanks to Stoke's theorem. Indeed, the function $\nabla_x P_m(f^\epsilon) \cdot v f^\epsilon$ is integrable on Γ for a.e. $t \in (a, b)$. Then, if $(\varphi_n)_{n \in \mathbb{N}}$ is a sequence of smooth and compactly supported cut-off functions such that $\varphi_n = 1$ on centered balls of radius n , we obtain

$$\begin{aligned} \iint_{\Gamma} \nabla_x P_m(f^\epsilon) \cdot v f^\epsilon dx dv &= \lim_{n \rightarrow \infty} \iint_{\Gamma} \varphi_n \nabla_x P_m(f^\epsilon) \cdot v f^\epsilon dx dv \\ &= - \lim_{n \rightarrow \infty} \iint_{\Gamma} \nabla_x \varphi_n \cdot v (f^\epsilon)^m dx dv = 0 \end{aligned}$$

because the field $v (f^\epsilon)^m$ belongs to $L^1(\Gamma)$ for a.e. $t \in (a, b)$. Taking this observation into account, we integrate (18) over the interval (a, b) to get

$$\mathcal{F}_m(\mu_b^\epsilon) - \mathcal{F}_m(\mu_a^\epsilon) = \int_a^b \iint_{\Gamma} (\nabla_v P_m(f^\epsilon) + v) \cdot F^\epsilon f^\epsilon dx dv dt, \quad (19)$$

which is (5) for the regularised pair (f^ϵ, F^ϵ) .

Step 4: Conclusion. The equality in (5) is proved by passing to the limit $\epsilon \rightarrow 0_+$ in (19), thanks to (11)-(13)-(15). Since $(\nabla_v P_m(\mu_t) + v)$ is a v -gradient (precisely, in the $L^2(\mu)$ -closure of gradients of smooth functions), we may write

$$\mathcal{F}_m(\mu_b) - \mathcal{F}_m(\mu_a) = \int_a^b \iint_{\Gamma} (\nabla_v P_m(\mu_t) + v) \cdot (F_t + G_t) \mu_t dx dv dt,$$

for any $G_t \in L^2(\mu_t; \mathbb{R}^d)$ with $\nabla_v \cdot (G_t \mu_t) = 0$ for a.e. $t \in (a, b)$. We apply Young's inequality

$$\begin{aligned} \int_a^b \iint_{\Gamma} (\nabla_v P_m(\mu_t) + v) \cdot (F_t + G_t) \mu_t dx dv dt \\ \geq -\frac{1}{2} \int_a^b \left(\|F_t + G_t\|_{L^2(\mu_t)}^2 + \mathcal{I}_m(\mu_t) \right) dt, \end{aligned}$$

take the infimum in the right-hand side over all $(G_t)_t$, and obtain the norm of the projection of F_t over v -gradients, that is, $|\mu'_t|_d$, by [6, Proposition 5.23]. This proves (5).

Equality is achieved if and only if the two fields F_t and $\nabla_v P_m(f_t) + v$ are opposite, up to adding a force field $(G_t)_t$ such that $\nabla_v \cdot (G_t \mu_t) = 0$. In this case, (5) holds with

$$|\mu'_t|_d = \|\nabla_v P_m(f_t) + v\|_{L^2(\mu_t)} = \mathcal{I}_m(\mu_t).$$

3 Strong convergence of the minimising-movement scheme

All computations are performed for $m > 1$, but, with minor changes, can be extended to the linear case $m = 1$.

Step 1: Basic estimates for a JKO type scheme. Let $h \in (0, 1)$. On the phase space Γ , the transformation

$$\mathcal{G}_h(x, v) := (x + h v, v), \quad (x, v) \in \Gamma$$

preserves Lebesgue's measure as well as the velocity marginal. As a consequence we have $\mathcal{F}_m(\mu) = \mathcal{F}_m(\mathcal{G}_h \# \mu)$ for every $\mu \in \mathcal{P}_2(\Gamma)$, where $\mathcal{G}_h \# \mu$ denotes the push forward of μ through \mathcal{G}_h . Following [6, Section 3.1], one can rewrite d_h^2 defined by (3) as the value of the optimal transport problem

$$d_h^2(\mu, \nu) = \inf_{\gamma \in \Pi(\mu, \nu)} \iint_{\Gamma \times \Gamma} c_h((x, v), (y, w)) d\gamma((x, v), (y, w))$$

where $\Pi(\mu, \nu)$ is the set of probability measures on $\Gamma \times \Gamma$ with marginals μ and ν , and

$$c_h((x, v), (y, w)) := 12 \left| \frac{y-x}{h} - \frac{v+w}{2} \right|^2 + |w-v|^2 = |(y, w) - \mathcal{G}_h(x, v)|_h^2. \quad (20)$$

Here, the squared norm $|\cdot|_h^2$ on the phase space is defined by

$$|(a, b)|_h^2 := 12 \left| \frac{a}{h} - \frac{b}{2} \right|^2 + |b|^2 = |A_h(a, b)|^2, \quad \text{with } A_h(a, b) := \left(\frac{2\sqrt{3}}{h} a - \sqrt{3} b, b \right).$$

The cost function c_h appears for instance in [15, (4.1)]. With this notation, we have

$$d_h^2(\mu, \nu) = W_h^2(\mathcal{G}_h \# \mu, \nu) = W^2((A_h \circ \mathcal{G}_h) \# \mu, A_h \# \nu) \quad (21)$$

where W_h is the Wasserstein distance associated with the norm $|\cdot|_h$ and W is the usual 2-Wasserstein distance. See [15] for details. Notice that the scheme (7) may also be written as

$$\mu_{k+1}^{(h)} \in \arg \min_{\nu \in \mathcal{P}_2(\Gamma)} \left\{ \mathcal{F}_m(\nu) + \frac{1}{2h} W_h^2(\mathcal{G}_h \# \mu_k^{(h)}, \nu) \right\}, \quad k \in \mathbb{N}, \quad \mu_0^{(h)} = \mu_0. \quad (22)$$

Standard lower semicontinuity and convexity arguments ensure that this scheme is well-posed, and the explicit form of the cost c_h guarantees the uniqueness of an optimal transport plan $\gamma_k^{(h)} \in \Pi(\mu_k^{(h)}, \mu_{k+1}^{(h)})$ for the cost c_h . We define the *piecewise constant in time map* $t \mapsto \mu_t^{(h)}$ by

$$\mu_t^{(h)} := \mu_{k+1}^{(h)}, \quad t \in (k h, (k+1) h], \quad k \in \mathbb{N}, \quad \mu^{(h)}(0) = \mu_0.$$

Taking $\mathcal{G}_{h\#}\mu_k^{(h)}$ as a competitor in the minimisation (22), we have

$$\mathcal{F}_m(\mu_{k+1}^{(h)}) + \frac{1}{2h} W_h^2(\mathcal{G}_{h\#}\mu_k^{(h)}, \mu_{k+1}^{(h)}) \leq \mathcal{F}_m(\mathcal{G}_{h\#}\mu_k^{(h)}) = \mathcal{F}_m(\mu_k^{(h)}), \quad (23)$$

and, in particular,

$$\max_{k \in \mathbb{N}} \mathcal{F}_m(\mu_k^{(h)}) \leq \mathcal{F}_m(\mu_0). \quad (24)$$

Let us fix a time horizon $T > 0$ and define $N = \lceil T/h \rceil$. Summing (23) over k and using the fact that \mathcal{F}_m is nonnegative yield

$$\sum_{k=0}^{N-1} \frac{1}{2h} W_h^2(\mathcal{G}_{h\#}\mu_k^{(h)}, \mu_{k+1}^{(h)}) = \frac{1}{2h} \sum_{k=0}^{N-1} d_h^2(\mu_k^{(h)}, \mu_{k+1}^{(h)}) \leq \mathcal{F}_m(\mu_0). \quad (25)$$

Since moments in velocity are bounded by (24), we have

$$W^2(\mu_k^{(h)}, \mathcal{G}_{h\#}\mu_k^{(h)}) \leq \mathcal{C} h^2, \quad k \in \mathbb{N}.$$

Together with the crude inequality $W^2 \leq 2 W_h^2$ and (25), we obtain

$$\sum_{k=0}^{N-1} \frac{1}{2h} W^2(\mu_k^{(h)}, \mu_{k+1}^{(h)}) \leq \mathcal{C} \frac{1}{h} \sum_{k=0}^{N-1} \left(h^2 + W_h^2(\mathcal{G}_{h\#}\mu_k^{(h)}, \mu_{k+1}^{(h)}) \right) \leq \mathcal{C}. \quad (26)$$

for some constant \mathcal{C} depending on T . In what follows, the constant \mathcal{C} may change from formula to formula but stays independent of h . As a consequence,

$$W(\mu_t^{(h)}, \mu_s^{(h)}) \leq \mathcal{C} \sqrt{|t-s| + h}, \quad 0 \leq s, t \leq T. \quad (27)$$

Finally, using the triangle inequality, Cauchy–Schwarz, and (26) we get

$$\begin{aligned} \max_{k \in \{0, \dots, N\}} \iint_{\Gamma} |(x, v)|^2 d\mu_k^{(h)} &\leq 2 \iint_{\Gamma} |(x, v)|^2 d\mu_0 + 2 \max_{k \in \{0, \dots, N\}} W^2(\mu_0, \mu_k^{(h)}) \\ &\leq \mathcal{C} + 2 \left(\sum_{k=0}^{N-1} W(\mu_k^{(h)}, \mu_{k+1}^{(h)}) \right)^2 \leq \mathcal{C} + 2N \sum_{k=0}^{N-1} W^2(\mu_k^{(h)}, \mu_{k+1}^{(h)}) \leq \mathcal{C}. \end{aligned} \quad (28)$$

Step 2: Euler-Lagrange equations and approximate nonlinear kinetic Fokker-Planck equation. Let $S_k^{(h)} = \left(\left[S_k^{(h)} \right]_1, \left[S_k^{(h)} \right]_2 \right)$ be the d_h -optimal transport map from $\mu_{k+1}^{(h)}$ to $\mu_k^{(h)}$. For every smooth vector field $V: \Gamma \rightarrow \Gamma$, the first variation

$$\frac{d^+}{d\epsilon} \left[\mathcal{F}_m \left((\text{id} + \epsilon V) \# \mu_{k+1}^{(h)} \right) + \frac{1}{2h} \iint_{\Gamma \times \Gamma} c_h d \left(S_k^{(h)}, \text{id} + \epsilon V \right) \# \mu_{k+1}^{(h)} \right]$$

(with c_h given by (20)) at $\epsilon = 0$ is nonnegative. Therefore, tedious but standard computations (see, e.g., [20, 7]) yield $\mu_{k+1}^{(h)} = f dx dv$ with $f^m \in W^{1,1}(\Gamma)$ and

$$\begin{aligned} \left[S_k^{(h)}(y, w) \right]_1 &= y - hw - \frac{h^2}{2} \left(\frac{\nabla_v(f^m)}{f} + w \right) - \frac{h^3}{6} \frac{\nabla_x(f^m)}{f}, \\ \left[S_k^{(h)}(y, w) \right]_2 &= w + h \left(\frac{D_h(f^m)}{f} + w \right), \end{aligned} \quad (29)$$

with $D_h := \frac{h}{2} \nabla_x + \nabla_v$. By plugging this in the cost, we obtain

$$\begin{aligned} d_h^2 \left(\mu_k^{(h)}, \mu_{k+1}^{(h)} \right) &= \iint_{\Gamma \times \Gamma} c_h d \left(S_k^{(h)}, \text{id} \right) \# \mu_{k+1}^{(h)} \\ &= \frac{h^4}{12} \iint_{\Gamma} \frac{|\nabla_x(f^m)|^2}{f} dy dw + h^2 \iint_{\Gamma} \left| \frac{D_h(f^m)}{f} + w \right|^2 d\mu_{k+1}^{(h)} \\ &\geq \frac{h^2}{4} \iint_{\Gamma} \left| \frac{\nabla_v(f^m)}{f} + w \right|^2 d\mu_{k+1}^{(h)}, \end{aligned} \quad (30)$$

where the inequality is obtained from $\frac{a^2}{4} \leq \frac{b^2}{3} + (a+b)^2$. By Lemma 1 applied with $\alpha = 1 - 1/(2m)$ and $g = f^m$, we find $f^{m-1/2} \in W_{\text{loc}}^{1,1}(\Gamma)$ and we can write

$$\frac{1}{f} L(f^m) = \frac{1}{\sqrt{f}} \frac{m}{m-1/2} L \left(f^{m-1/2} \right), \quad \text{for both } L = \nabla_v \text{ and } L = D_h. \quad (31)$$

In particular,

$$\begin{aligned} d_h^2 \left(\mu_k^{(h)}, \mu_{k+1}^{(h)} \right) &\geq \frac{h^2}{4} \iint_{\Gamma} \left(\left(\frac{2m}{2m-1} \right)^2 |\nabla_v(f^{m-1/2})|^2 + |v|^2 f - 2v \cdot \nabla_v(f^m) \right) dx dv \\ &= \frac{h^2}{4} \iint_{\Gamma} \left(\left(\frac{2m}{2m-1} \right)^2 |\nabla_v(f^{m-1/2})|^2 + |v|^2 f - 2d f^m \right) dx dv = \mathcal{I}_m \left(\mu_{k+1}^{(h)} \right), \end{aligned} \quad (32)$$

where integration by parts can be proven using $f \in L^m(\Gamma)$. Setting

$$F_t^{(h)} := - \left(\frac{D_h(f^m)}{f} + v \right) = - \left(\frac{m}{m-1/2} \frac{D_h(f^{m-1/2})}{\sqrt{f}} + v \right), \quad d\mu_{k+1}^{(h)} = f dx dv \quad (33)$$

for any $t \in (kh, (k+1)h)$, Equation (30) gives

$$\frac{1}{2h} \mathfrak{d}_h^2(\mu_k^{(h)}, \mu_{k+1}^{(h)}) \geq \frac{1}{2} \int_k^{(k+1)h} \|F_t^{(h)}\|_{L^2(\mu_t^{(h)})}^2 dt. \quad (34)$$

Combining (32), (34), and (25), we deduce

$$\int_0^T \|F_t^{(h)}\|_{L^2(\mu_t^{(h)})}^2 dt + \int_0^T \mathcal{I}_m(\mu_t^{(h)}) dt \leq \mathcal{C} \frac{1}{h} \sum_{n=0}^{N-1} \mathfrak{d}_h^2(\mu_k^{(h)}, \mu_{k+1}^{(h)}) \leq \mathcal{C}, \quad (35)$$

Let $\eta \in \mathcal{D}((0, T) \times \Gamma)$. Using the fact that $\mu_k^{(h)} = (S_k^{(h)})_{\#} \mu_{k+1}^{(h)}$, we have

$$\begin{aligned} \int_0^T \iint_{\Gamma} \partial_t \eta d\mu_t^{(h)} dt &= \sum_{k=1}^{N-1} \iint_{\Gamma} \eta(kh, \cdot) d(\mu_k^{(h)} - \mu_{k+1}^{(h)}) \\ &= \sum_{k=1}^{N-1} \iint_{\Gamma} (\eta(kh, S_k^{(h)}) - \eta(kh, \cdot)) d\mu_{k+1}^{(h)} \\ &= \sum_{k=1}^{N-1} \iint_{\Gamma} (\nabla_{x,v} \eta(kh, \cdot) \cdot (S_k^{(h)} - \text{id}) + R_k^{(h)}) d\mu_{k+1}^{(h)}, \end{aligned} \quad (36)$$

where the error term $R_k^{(h)}$ in the second-order Taylor expansion satisfies

$$|R_k^{(h)}| \leq \frac{1}{2} \|D^2 \eta\|_{L^\infty} |S_k^{(h)} - \text{id}|^2.$$

Writing $\eta_k := \eta(kh, \cdot)$ and, once again, $\mu_{k+1}^{(h)} = f dx dv$, we observe that, by (29),

$$\begin{aligned} - \iint_{\Gamma} \nabla_x \eta_k \cdot ([S_k^{(h)}]_1 - y) d\mu_{k+1}^{(h)} &= \left(h + \frac{h^2}{2}\right) \iint_{\Gamma} \nabla_x \eta_k \cdot w d\mu_{k+1}^{(h)} \\ &\quad - \frac{h^2}{2} \iint_{\Gamma} \text{div}_v(\nabla_x \eta_k) f^m dx dv - \iint_{\Gamma} \frac{h^3}{6} \Delta_x \eta_k f^m dx dv, \end{aligned}$$

whence, recalling (24), and allowing \mathcal{C} to depend on η ,

$$\left| \iint_{\Gamma} \nabla_x \eta_k \cdot ([S_k^{(h)}]_1 - y + hw) d\mu_{k+1}^{(h)} \right| \leq \mathcal{C} h^2 \mathcal{F}_m(\mu_{k+1}^{(h)}) \leq \mathcal{C} h^2. \quad (37)$$

Furthermore,

$$\begin{aligned} \frac{1}{2} \|S_k^{(h)} - \text{id}\|_{L^2(\mu_{k+1}^{(h)})}^2 &\leq \|S_k^{(h)} - \mathcal{G}_h \circ S_k^{(h)}\|_{L^2(\mu_{k+1}^{(h)})}^2 + \|\mathcal{G}_h \circ S_k^{(h)} - \text{id}\|_{L^2(\mu_{k+1}^{(h)})}^2 \\ &\leq \|\text{id} - \mathcal{G}_h\|_{L^2(\mu_k^{(h)})}^2 + 2 \iint_{\Gamma} \|\mathcal{G}_h \circ S_k^{(h)} - \text{id}\|_h^2 d\mu_{k+1}^{(h)} \\ &\leq h^2 \iint_{\Gamma} \|v\|^2 d\mu_k^{(h)} + 2 \mathfrak{d}_h^2(\mu_k^{(h)}, \mu_{k+1}^{(h)}) \leq \mathcal{C} \left(h^2 + \mathfrak{d}_h^2(\mu_k^{(h)}, \mu_{k+1}^{(h)})\right) \end{aligned}$$

where we used $\mu_k^{(h)} = S_k^{(h)} \# \mu_{k+1}^{(h)}$, and $|\cdot|^2 \leq 2|\cdot|_h^2$ in the second line, and the optimality of $S_k^{(h)}$ and (28) in the third line. By (25), we obtain that $\sum_{k=1}^{N-1} \iint_{\Gamma} |R_k^{(h)}| d\mu_{k+1}^{(h)} \leq \mathcal{C} h$. Hence, using (36), (37), and (29), we obtain

$$\begin{aligned} \int_0^T \iint_{\Gamma} \partial_t \eta d\mu_t^{(h)} dt &= - \int_0^T \iint_{\Gamma} \nabla_x \eta \cdot v d\mu_t^{(h)} dt \\ &\quad - \int_0^T \iint_{\Gamma} \nabla_v \eta \cdot F^{(h)} d\mu_t^{(h)} dt + O(h). \end{aligned}$$

In other words, we have shown that, in the distributional sense,

$$\partial_t \mu^{(h)} + v \cdot \nabla_x \mu^{(h)} + \nabla_v \cdot (F^{(h)} \mu^{(h)}) \rightarrow 0 \quad \text{as } h \rightarrow 0_+. \quad (38)$$

In conclusion of this step, observe that the vector measure $Q^{(h)} := F^{(h)} \mu^{(h)}$ is bounded in total variation uniformly in h , because, by the Cauchy–Schwarz inequality and (35),

$$\|Q^{(h)}\|_{\text{TV}} = \int_0^T \|F_t^{(h)}\|_{L^1(\mu_t^{(h)})} dt \leq \int_0^T \|F_t^{(h)}\|_{L^2(\mu_t^{(h)})} dt \leq \mathcal{C}. \quad (39)$$

Step 3: L^1 -contraction and L^1 -equicontinuity in x . In [19], an L^1 -contraction principle is established for JKO steps which, in particular, implies the following. Given $R > 0$ and absolutely continuous probabilities μ_1 and μ_2 supported on B_R , and denoting by

$$v_i := \arg \min_{\substack{v \in \mathcal{P}_2(\Gamma) \\ \text{supp}(v) \subset B_R}} \left\{ \mathcal{F}_m(v) + \frac{1}{2h} W_h^2(\mu_i, v) \right\}, \quad i = 1, 2,$$

we have

$$\|(v_1 - v_2)_+\|_{L^1(B_R)} \leq \|(\mu_1 - \mu_2)_+\|_{L^1(B_R)}.$$

Let us briefly explain how to extend this principle to the full space by approximation and Γ -convergence arguments. Let μ_i be in $\mathcal{P}_2(\Gamma) \cap L^1(\Gamma)$, let $(R_n)_{n \in \mathbb{N}}$ be a sequence of radii such that $R_n \uparrow \infty$ and $\mu_i(B_{R_n}) \geq 1/2$, and let μ_i^n be the normalised restriction of μ_i to B_{R_n} , i.e., $\mu_i^n(A) = \mu_i(A \cap B_{R_n}) / \mu_i(B_{R_n})$. For $v \in \mathcal{P}_2(\Gamma)$, define the functionals

$$J_i^n(v) := \begin{cases} \mathcal{F}_m(v) + \frac{1}{2h} W_h^2(\mu_i^n, v) & \text{if } \text{supp}(v) \subset B_{R_n}, \\ +\infty & \text{otherwise,} \end{cases}$$

and denote by v_i^n the minimiser of J_i^n . It is easy to check that the sequence $(J_i^n)_{n \in \mathbb{N}}$ Γ -converges to $\mathcal{F}_m + \frac{1}{2h} W_h^2(\mu_i, \cdot)$ as $n \rightarrow \infty$ for the narrow topology (to construct recovery sequences, one can use the same approximation by restriction and normalisation as in the definition of μ_i^n). This entails narrow (in fact, weak L^m) convergence of v_i^n to a minimiser v_i of $\mathcal{F}_m + \frac{1}{2h} W_h^2(\mu_i, \cdot)$. By the L^1 -contraction principle of [19],

$$\|(v_1^n - v_2^n)_+\|_{L^1(B_{R_n})} \leq \|(\mu_1^n - \mu_2^n)_+\|_{L^1(B_{R_n})},$$

and, since μ_i^n converges to μ_i in L^1 and using the weak lower semicontinuity properties of the left hand side, by passing to the liminf, we readily obtain the desired inequality

$$\|(v_1 - v_2)_+\|_{L^1(\Gamma)} \leq \|(\mu_1 - \mu_2)_+\|_{L^1(\Gamma)}.$$

In our JKO scheme, recall that $\mu_{k+1}^{(h)}$ is obtained from $\mu_k^{(h)}$ through (22). For $\delta \in \mathbb{R}^d$, set $\tau_\delta(x, v) := (x - \delta, v)$. The invariance of \mathcal{F}_m by horizontal translations gives

$$\tau_{\delta\#} \mu_{k+1}^{(h)} = \arg \min_{v \in \mathcal{P}_2(\Gamma)} \left\{ \mathcal{F}_m(v) + \frac{1}{2h} W_h^2 \left((\tau_\delta \circ \mathcal{G}_h)\# \mu_k^{(h)}, v \right) \right\},$$

which, combined with the above L^1 -contraction principle and the commutation of \mathcal{G}_h with horizontal translations, yields

$$\|\tau_{\delta\#} \mu_{k+1}^{(h)} - \mu_{k+1}^{(h)}\|_{L^1(\Gamma)} \leq \|(\tau_\delta \circ \mathcal{G}_h)\# \mu_k^{(h)} - \mathcal{G}_h\# \mu_k^{(h)}\|_{L^1(\Gamma)} = \|\tau_{\delta\#} \mu_k^{(h)} - \mu_k^{(h)}\|_{L^1(\Gamma)},$$

so that for all k and all $\delta \in \mathbb{R}^d$, we have

$$\|\tau_{\delta\#} \mu_k^{(h)} - \mu_k^{(h)}\|_{L^1(\Gamma)} \leq \Delta_0(\delta) := \|\tau_{\delta\#} \mu_0 - \mu_0\|_{L^1(\Gamma)}.$$

We have thus shown that

$$\sup_{t \geq 0} \|\tau_{\delta\#} \mu_t^{(h)} - \mu_t^{(h)}\|_{L^1(\Gamma)} \leq \Delta_0(\delta). \quad (40)$$

Since μ_0 lies in $L^1(\Gamma)$, we have $\lim_{\delta \rightarrow 0} \Delta_0(\delta) = 0$.

Step 4: Strong L^1 -compactness. We are going to show that $\mu^{(h)}$ is relatively compact in $L^1((0, T) \times \Gamma)$. To this end, we will make use of the refined version of the Aubin-Lions compactness Theorem due to Rossi and Savaré [25]. Notice that, in addition to the contractivity (40), we have a uniform bound C on the moments $\sup_{t \in [0, T]} \iint_{\Gamma} |x, v|^2 d\mu_t^{(h)}$ by (28), as well as on $\int_0^T \mathcal{I}_m(\mu_t^{(h)}) dt$ thanks to (35). Define K as the set of functions $f \in L^1 \cap \mathcal{P}_2(\Gamma)$ that satisfy

$$\iint_{\Gamma} |x, v|^2 f dx dv \leq C, \quad \|f \circ \tau_\delta^{-1} - f\|_{L^1(\Gamma)} \leq \Delta_0(\delta) \quad \forall \delta \in \mathbb{R}^d, \quad (41)$$

and set

$$\mathcal{J}_m(f) := \begin{cases} \mathcal{I}_m(f) & \text{if } f \in K, \\ +\infty & \text{if } f \in L^1(\Gamma) \setminus K. \end{cases}$$

We have

$$\sup_h \int_0^T \mathcal{J}_m(\mu^{(h)}(t)) dt \leq C$$

and $W(\mu^{(h)}(t), \mu^{(h)}(s)) \leq \mathcal{C} \sqrt{|t - s| + h}$ for any $(t, s) \in [0, T]^2$ by (27). If we prove that sublevel sets of \mathcal{J}_m are relatively compact in $L^1(\Gamma)$, [25, Theorem 2] will enable

us to conclude. To this end, we will invoke the Riesz–Fréchet–Kolmogorov Theorem on a sublevel set of \mathcal{J}_m . Indeed, it follows from the uniform moment bound that any such sublevel is equitight, and we already have equicontinuity for x -translations thanks to the second property in (41). There only remains to prove equicontinuity for v -translations. Fix a sublevel set S of \mathcal{J}_m . There exists a constant C_S such that $\|\nabla_v f^{m-1/2}\|_{L^2(\Gamma)} \leq C_S$ for all $f \in S$. Fix one such f . Denoting $g := f^{m-1/2}$, one can express $f = g^{2\theta}$ for the exponent $\theta = 1/(2m - 1)$, which is smaller than 1 since $m \geq 1$. For every $\delta \in \mathbb{R}^d$ and $R > 0$, denoting by B_R the ball of radius R centered at 0, it follows that

$$\begin{aligned}
& \iint_{B_R \times B_R} |f(x, v + \delta) - f(x, v)| \, dv \, dx \\
&= \iint_{B_R \times B_R} |g(x, v + \delta)^\theta - g(x, v)^\theta| (g(x, v + \delta)^\theta + g(x, v)^\theta) \, dv \, dx \\
&\leq \iint_{B_R \times B_R} |g(x, v + \delta) - g(x, v)|^\theta (g(x, v + \delta)^\theta + g(x, v)^\theta) \, dv \, dx \\
&\leq 2 |B_R|^{1-\theta} \left(\iint_{B_R \times B_R} |g(x, v + \delta) - g(x, v)|^2 \, dv \, dx \right)^{\theta/2} \left(\iint_{\Gamma} g^{2\theta} \, dv \, dx \right)^{1/2} \\
&\leq 2 |\delta|^\theta |B_R|^{1-\theta} \left(\iint_{\Gamma} |\nabla_v (f^{m-1/2})|^2 \, dv \, dx \right)^{\theta/2} \left(\iint_{\Gamma} f \, dv \, dx \right)^{1/2} \leq C_R (C_S |\delta|)^\theta,
\end{aligned}$$

where we used the fact that $\theta \leq 1$ in the third line, and Hölder's inequality with exponents $2/(1 - \theta)$, $2/\theta$, and 2 in the third line. Therefore,

$$\|f(\cdot, \delta + \cdot) - f\|_{L^1(\Gamma)} \leq C_R (C_S |\delta|)^\theta + C \left(R^{-2} + (R - |\delta|)^{-2} \right)$$

for $R > |\delta|$, which yields the desired v -equicontinuity.

Step 5: Limit points solve (1). Let μ be a $L^1_{\text{loc}}(\mathbb{R}^+; L^1(\Gamma))$ -limit point of $\mu^{(h)}$. We observe the following. First, thanks to (27) and the refined version of the Arzelà–Ascoli Theorem from [3, Proposition 3.3.1], we may assume that $\mu \in C^{0,1/2}([0, T], (\mathcal{P}_2(\Gamma), W))$ with

$$\lim_{h \rightarrow 0^+} \sup_{t \in [0, T]} W(\mu_t^{(h)}, \mu_t) = 0.$$

In particular, $\mu(0) = \mu_0$. Secondly, by indentifying $\mu^{(h)}$ and μ with their densities with respect to the Lebesgue measure, we have $\sqrt{\mu^{(h)}} \rightarrow \sqrt{\mu}$ in $L^2((0, T) \times \Gamma)$. Thirdly, by the uniform bound (24) on the L^m -norm and the Dunford–Pettis Theorem, we may assume that $(\mu^{(h)})^{m-1/2}$ converges weakly to $\mu^{m-1/2}$ in $L^1((0, T) \times \omega)$ for all bounded domain $\omega \subset \Gamma$. Lastly, by (35), we may also assume that $F^{(h)} \sqrt{\mu^{(h)}}$ converges weakly in $L^2((0, T) \times \Gamma; \mathbb{R}^d)$ to some function G .

From the previous observations and using (33), we obtain, for any test function $\eta \in \mathcal{D}((0, T) \times \Gamma)$,

$$\begin{aligned}\langle \eta, G \rangle &= \lim_{h \rightarrow 0_+} \left(\frac{m}{m-1/2} \langle D_h \eta, (\mu^{(h)})^{m-1/2} \rangle - \left\langle v \eta, \sqrt{\mu^{(h)}} \right\rangle \right) \\ &= \frac{m}{m-1/2} \langle \nabla_v \eta, \mu^{m-1/2} \rangle - \langle v \eta, \sqrt{\mu} \rangle,\end{aligned}$$

hence

$$G = - \left(\frac{m}{m-1/2} \nabla_v (\mu^{m-1/2}) + v \sqrt{\mu} \right).$$

In the light of (38), and again by the three convergences observed above, for any test function η , we find

$$\begin{aligned}0 &= \lim_{h \rightarrow 0_+} \left(\left\langle (\partial_t + v \cdot \nabla_x) \eta, \mu^{(h)} \right\rangle + \left\langle \nabla_v \eta \sqrt{\mu^{(h)}}, F^{(h)} \sqrt{\mu^{(h)}} \right\rangle \right) \\ &= \langle (\partial_t + v \cdot \nabla_x) \eta, \mu \rangle + \langle \nabla_v \eta \sqrt{\mu}, G \rangle \\ &= \langle (\partial_t + v \cdot \nabla_x) \eta, \mu \rangle - \frac{m}{m-1/2} \langle \nabla_v \eta, \nabla_v (\mu^{m-1/2}) \sqrt{\mu} \rangle - \langle \nabla_v \eta, v \mu \rangle,\end{aligned}$$

which is the desired conclusion.

4 Maximal dissipation for $m \in [1, 3/2]$

Now we assume that $m \in [1, 3/2]$ and fix an horizon $T > 0$. We know from Section 3 that we may assume that the family $\mu^{(h)}$ constructed by the JKO scheme converges in $L^1((0, T) \times \Gamma)$ and uniformly in time in Wasserstein distance to some μ . The goal of this section is to prove that μ dissipates the energy maximally.

Step 1: De Giorgi interpolation and discrete energy dissipation. We first need to improve (23) into a sharper discrete energy dissipation inequality. To do so, following [3, 26, 27], we consider a De Giorgi variational interpolation scheme. For every $k = 0, \dots, N-1$ and $\theta \in (0, 1]$, let

$$\hat{v}_k^{(h)}(\theta) := \arg \min_{v \in \mathcal{P}_2(\Gamma)} \left\{ \mathcal{F}_m(v) + \frac{1}{2\theta h} W_h^2 \left(\mathcal{G}_{h\#} \mu_k^{(h)}, v \right) \right\}$$

so that $\hat{v}_k^{(h)}(1) = \mu_{k+1}^{(h)}$. From

$$\mathcal{F}_m(\hat{v}_k^{(h)}(\theta)) + \frac{1}{2\theta h} W_h^2 \left(\mathcal{G}_{h\#} \mu_k^{(h)}, \hat{v}_k^{(h)}(\theta) \right) \leq \mathcal{F}_m(\mathcal{G}_{h\#} \mu_k^{(h)})$$

we infer that $\theta \mapsto \hat{v}_k^{(h)}(\theta)$ can be continuously extended (in the Wasserstein sense) at 0 with $\hat{v}_k^{(h)}(0) = \mathcal{G}_{h\#} \mu_k^{(h)}$, and also that

$$\sup_{\theta \in [0, 1]} W_h^2 \left(\mathcal{G}_{h\#} \mu_k^{(h)}, \hat{v}_k^{(h)}(\theta) \right) \leq \mathcal{C}. \quad (42)$$

It follows from classical arguments (see [27, paragraph 3.2] and [3, paragraphs 3.1-3.2]) that the following dissipation inequality holds

$$\mathcal{F}_m(\mu_{k+1}^{(h)}) - \mathcal{F}_m(\mu_k^{(h)}) + \frac{1}{2h} W_h^2(\mathcal{G}_{h\#}\mu_k^{(h)}, \mu_{k+1}^{(h)}) + \frac{h}{2} \int_0^1 \left| \partial_h \mathcal{F}_m(\hat{\nu}_k^{(h)}(\theta)) \right|^2 d\theta \leq 0, \quad (43)$$

where

$$|\partial_h \mathcal{F}_m(\mu)| := \limsup_{\nu \rightarrow \mu} \frac{(\mathcal{F}_m(\mu) - \mathcal{F}_m(\nu))_+}{W_h(\mu, \nu)}$$

denotes the descending metric slope. Here we use again $\mathcal{F}_m(\mu_k^{(h)}) = \mathcal{F}_m(\mathcal{G}_{h\#}\mu_k^{(h)})$, and the convergence $\nu \rightarrow \mu$ is intended in the W_h sense or, equivalently, in the W sense, as long as $h \in (0, 1)$ is fixed.

Let us compute the metric slope. To do this, we define

$$\mathcal{F}_m^{(h)}(f) := \mathcal{F}_m(A_h^{-1}\#f) = \frac{\det(A_h)^{m-1}}{m-1} \iint_{\Gamma} f^m dx dv + \frac{1}{2} \iint_{\Gamma} |v|^2 f dx dv. \quad (44)$$

Thanks to (21), we observe that

$$|\partial \mathcal{F}_m^{(h)}(A_{h\#}\mu)| = \limsup_{\nu \rightarrow \mu} \frac{(\mathcal{F}_m^{(h)}(A_{h\#}\mu) - \mathcal{F}_m^{(h)}(A_{h\#}\nu))_+}{W(A_{h\#}\mu, A_{h\#}\nu)} = |\partial_h \mathcal{F}_m(\mu)|$$

where $|\partial \mathcal{F}_m^{(h)}|$ stands for the slope in the usual Wasserstein sense. Using the explicit form of (44), $\mathcal{F}_m^{(h)}$ is a displacement convex functional in the usual Wasserstein sense on Γ . Thus, when $|\partial \mathcal{F}_m^{(h)}(\mu)| < +\infty$, [3, Theorem 10.4.13] ensures that $d\mu = f dx dv$ with $f^m \in W_{\text{loc}}^{1,1}(\Gamma)$, and

$$|\partial \mathcal{F}_m^{(h)}(\mu)|^2 = \iint_{\Gamma} \left| \det(A_h)^{m-1} \frac{\nabla_{x,v}(f^m)}{f} + (0, v) \right|^2 d\mu.$$

From the change of variables $d(A_{h\#}\mu) = \det(A_h)^{-1} f \circ A_h^{-1} dx dv$, we deduce

$$\begin{aligned} |\partial_h \mathcal{F}_m(\mu)|^2 &= |\partial \mathcal{F}_m^{(h)}(A_{h\#}\mu)|^2 \\ &= \iint_{\Gamma} \left| \frac{[(\nabla_{x,v}(f^m)) \circ A_h^{-1}] \cdot A_h^{-1}}{f \circ A_h^{-1}} + (0, v) \right|^2 d(A_{h\#}\mu) \\ &= \iint_{\Gamma} \left| \frac{(\nabla_{x,v}(f^m)) \cdot A_h^{-1}}{f} + (0, v) \right|^2 d\mu \geq \iint_{\Gamma} \left| \frac{1}{f} \left(\frac{h}{2} \nabla_x + \nabla_v \right) (f^m) + v \right|^2 d\mu. \end{aligned}$$

Consider $D_h := \frac{h}{2} \nabla_x + \nabla_v$. When $|\partial \mathcal{F}_m^{(h)}(\mu)|^2$ is finite, by Lemma 1 applied with $\alpha = 1 - 1/(2m)$ and $g = f^m$, we may write $D_h(f^m)$ in terms of \sqrt{f} and $D_h(f^{m-1/2})$, see (31). We infer that

$$|\partial_h \mathcal{F}_m(\mu)|^2 \geq \mathcal{I}_m^{(h)}(f) := \iint_{\Gamma} \left| \frac{m}{m-1/2} D_h(f^{m-\frac{1}{2}}) + v \sqrt{f} \right|^2 dx dv.$$

Moreover, by a direct change of variable argument, one can easily obtain the alternative expression for this modified Fisher information as

$$\mathcal{I}_m^{(h)}(f) = \mathcal{I}_m\left(f \circ \mathcal{G}_{\frac{h}{2}}\right) = \mathcal{I}_m\left(\mathcal{G}_{-\frac{h}{2}} \# \mu\right). \quad (45)$$

Let us define the curve $[0, T] \ni t \mapsto \hat{\mu}_t^h$ by

$$\hat{\mu}_t^{(h)} := \hat{\nu}_k^{(h)}\left(\frac{t - kh}{h}\right), \quad t \in (kh, (k+1)h], \quad 0 \leq k \leq N-1.$$

Inequality (43), together with (21), (45), and (34) implies

$$\begin{aligned} 0 &\geq \mathcal{F}_m\left(\mu_{k+1}^{(h)}\right) - \mathcal{F}_m\left(\mu_k^{(h)}\right) + \frac{1}{2h} W_h^2\left(\mathcal{G}_{h\#}\mu_k^{(h)}, \mu_{k+1}^{(h)}\right) + \frac{1}{2} \int_{kh}^{(k+1)h} \mathcal{I}_m^{(h)}\left(\hat{\mu}_s^{(h)}\right) ds \\ &= \mathcal{F}_m\left(\mu_{k+1}^{(h)}\right) - \mathcal{F}_m\left(\mu_k^{(h)}\right) + \frac{1}{2h} d_h^2\left(\mu_k^{(h)}, \mu_{k+1}^{(h)}\right) + \frac{1}{2} \int_{kh}^{(k+1)h} \mathcal{I}_m\left(\mathcal{G}_{-\frac{h}{2}\#}\hat{\mu}_s^{(h)}\right) ds \\ &\geq \mathcal{F}_m\left(\mu_{k+1}^{(h)}\right) - \mathcal{F}_m\left(\mu_k^{(h)}\right) + \frac{1}{2} \int_{kh}^{(k+1)h} \left[\|F_s^{(h)}\|_{L^2(\mu^{(h)}(s))}^2 + \mathcal{I}_m\left(\mathcal{G}_{-\frac{h}{2}\#}\hat{\mu}_s^{(h)}\right) \right] ds. \end{aligned} \quad (46)$$

From $W^2 \leq 2W_h^2$, the triangle inequality, and (25), (42), we deduce

$$\sup_{t \in [0, T]} W\left(\mu_t^{(h)}, \hat{\mu}_t^{(h)}\right) \leq \mathcal{C} \sup_{t \in [0, T]} W_h\left(\mu_t^{(h)}, \hat{\mu}_t^{(h)}\right) \leq \mathcal{C} \sqrt{h}.$$

Moreover, by the very same L^1 -contraction and invariance by horizontal translations arguments as in Section 3, the interpolation $\hat{\mu}^{(h)}$ satisfies

$$\sup_{t \in [0, T]} \|\tau_{\delta\#}\hat{\mu}_t^{(h)} - \hat{\mu}_t^{(h)}\|_{L^1(\Gamma)} \leq \max_{0 \leq k \leq N} \|\tau_{\delta\#}\mu_k^{(h)} - \mu_k^{(h)}\|_{L^1(\Gamma)} \leq \Delta_0(\delta). \quad (47)$$

Step 2: Maximal dissipation. In view of (46), it is natural to define the family of curves of measures $\tilde{\mu}^{(h)} := \mathcal{G}_{-\frac{h}{2}\#}\hat{\mu}^{(h)}$. Then, we have

$$\begin{aligned} W^2\left(\hat{\mu}_t^{(h)}, \tilde{\mu}_t^{(h)}\right) &\leq \iint_{\Gamma} \left| \text{id} - \mathcal{G}_{-\frac{h}{2}} \right|^2 d\hat{\mu}_t^{(h)} = \frac{h^2}{4} \iint_{\Gamma} |v|^2 d\hat{\mu}_t^{(h)} \\ &\leq \frac{h^2}{2} \mathcal{F}_m\left(\hat{\mu}_t^{(h)}\right) \leq \mathcal{C} h^2. \end{aligned}$$

As a consequence, we obtain

$$\lim_{h \rightarrow 0^+} \sup_{t \in [0, T]} W\left(\tilde{\mu}_t^{(h)}, \mu_t\right) = 0$$

and we have uniform moment bounds on $\tilde{\mu}^{(h)}$. We deduce from (47):

$$\sup_{t \in [0, T]} \|\tau_{\delta\#} \tilde{\mu}_t^{(h)} - \tilde{\mu}_t^{(h)}\|_{L^1(\Gamma)} = \sup_{t \in [0, T]} \|\tau_{\delta\#} \hat{\mu}_t^{(h)} - \hat{\mu}_t^{(h)}\|_{L^1(\Gamma)} \leq \Delta_0(\delta). \quad (48)$$

Observe now that (24) and (46) implies

$$\int_0^T \mathcal{I}_m(\tilde{\mu}_s^{(h)}) ds \leq \mathcal{C}. \quad (49)$$

With uniform moment bounds, (48) and (49), we can argue exactly as in Section 3 to deduce that $\tilde{\mu}^{(h)}$ is relatively compact in $L^1((0, T) \times \Gamma)$ hence we may assume

$$\lim_{h \rightarrow 0_+} \|\tilde{\mu}^{(h)} - \mu\|_{L^1((0, T) \times \Gamma)} = 0.$$

If $t \in (k h, (k+1) h]$, by construction $\mathcal{F}_m(\mu_t^{(h)}) = \mathcal{F}_m(\mu_{k+1}^{(h)})$ hence summing the inequalities (46), from $k = 0$ to ℓ , we obtain

$$\begin{aligned} 0 &\geq \mathcal{F}_m(\mu_t^{(h)}) - \mathcal{F}_m(\mu_0) \\ &\quad + \frac{1}{2} \int_0^{(\ell+1)h} \|F_\tau^{(h)}\|_{L^2(\mu_\tau^{(h)})}^2 d\tau + \frac{1}{2} \int_0^{(\ell+1)h} \mathcal{I}_m(\mathcal{G}_{\frac{h}{2}\#} \hat{\mu}_\tau^{(h)}) d\tau \\ &\geq \mathcal{F}_m(\mu_t^{(h)}) - \mathcal{F}_m(\mu_0) + \frac{1}{2} \int_0^t \|F_\tau^{(h)}\|_{L^2(\mu_\tau^{(h)})}^2 d\tau + \frac{1}{2} \int_0^t \mathcal{I}_m(\tilde{\mu}^{(h)}) d\tau. \end{aligned} \quad (50)$$

Recalling (33), using (39) and a standard lower semicontinuity argument (see [2, Theorem 2.34]), we deduce that $Q^{(h)}$ converges in the sense of distributions to some vector measure Q with $Q_t = F_t \mu_t$, $F_t \in L^2(\mu_t)$, such that the pair (μ_t, F_t) solves (2) and

$$\int_0^t \iint_{\Gamma} |F_\tau| d\mu_\tau d\tau = \int_0^t \iint_{\Gamma} \left| \frac{dQ}{d\mu_\tau} \right|^2 d\mu_\tau d\tau \leq \liminf_{h \rightarrow 0_+} \int_0^t \iint_{\Gamma} |F_\tau^{(h)}|^2 d\mu_\tau^{(h)} d\tau \quad (51)$$

for any $t \in [0, T]$.

Let us now write the Fisher information of $\tilde{\mu}_\tau^{(h)}$ as

$$\mathcal{I}_m(\tilde{\mu}_\tau^{(h)}) = \iint_{\Gamma} \left(\left(\frac{2m}{2m-1} \right)^2 |\nabla_v ((\tilde{\mu}_\tau^{(h)})^{m-1/2})|^2 + |v|^2 \tilde{\mu}_\tau^{(h)} \right) dx dv \quad (52)$$

$$+ \iint_{\Gamma} \frac{2m}{m-1/2} \nabla_v (\tilde{\mu}_\tau^{(h)})^{m-1/2} \cdot v \sqrt{\tilde{\mu}_\tau^{(h)}} dx dv \quad (53)$$

By Lemma 3, the first term in (52) is lower semicontinuous for the strong L^1 topology.

Lemma 3 *Let $\alpha \geq 0$, and $(f_n)_{n \in \mathbb{N}}$ be a sequence of probability densities on Γ . If $(f_n)_{n \in \mathbb{N}}$ converges strongly in $L^1(\Gamma)$ to f , then we have*

$$\liminf_{n \rightarrow \infty} \iint_{\Gamma} |\nabla_v f_n^\alpha|^2 dx dv \geq \iint_{\Gamma} |\nabla_v f^\alpha|^2 dx dv.$$

In (53), we have to handle the product of $v\sqrt{\tilde{\mu}^{(h)}}$ by $\nabla_v(\tilde{\mu}^{(h)})^{m-1/2}$. Note that $v\sqrt{\tilde{\mu}^{(h)}}$ converges weakly in L^2 to $v\sqrt{\mu}$, but, by Wasserstein convergence, we also have convergence of second moments, which implies that the convergence of these square roots is in fact strong in L^2 . As for the L^2 -bounded sequence $\nabla_v(\tilde{\mu}^{(h)})^{m-1/2}$, one easily deduces from the strong convergence of $\tilde{\mu}^{(h)}$ to μ in L^1 , that $\nabla_v(\tilde{\mu}^{(h)})^{m-1/2}$ converges to $\nabla_v(\mu)^{m-1/2}$ weakly in L^2 . Using these observations and Fatou's lemma for the second term in (52), we get

$$\liminf_{h \rightarrow 0^+} \int_0^t \mathcal{I}_m(\tilde{\mu}_\tau^{(h)}) d\tau \geq \int_0^t \mathcal{I}_m(\mu_\tau) d\tau. \quad (54)$$

To pass to the liminf in (50), we notice that \mathcal{F}_m is lower semicontinuous for the convergence in Wasserstein distance, so that with (51), (54) and Fatou's lemma, we conclude that, for a.e. $t \in [0, T]$, the limit curve satisfies

$$\begin{aligned} 0 &\geq \mathcal{F}_m(\mu_t) - \mathcal{F}_m(\mu_0) + \frac{1}{2} \int_0^t \|F_\tau\|_{L^2(\mu_\tau)}^2 d\tau + \frac{1}{2} \int_0^t \mathcal{I}_m(\mu_\tau) d\tau \\ &\geq \mathcal{F}_m(\mu_t) - \mathcal{F}_m(\mu_0) + \frac{1}{2} \int_0^t \left(|\mu'_\tau|_{\mathfrak{d}}^2 + \mathcal{I}_m(\mu_\tau) \right) d\tau \end{aligned} \quad (55)$$

where we used the fact that $|\mu'_t|_{\mathfrak{d}} \leq \|F_t\|_{L^2(\mu_t)}$ for a.e. t by [6, Proposition 5.5].

For $m \in [1, 3/2]$, we can now use the chain rule (5) in Theorem 1 (note that until now, we have not used any restriction on $m \geq 1$). Indeed, fix t and a sequence $t_n \rightarrow 0$ in the full-measure set for which (5) holds. Since $t \mapsto \mu_t$ is W -continuous and \mathcal{F}_m is lower semicontinuous, we obtain

$$\begin{aligned} \mathcal{F}_m(\mu_0) &\leq \liminf_{n \rightarrow \infty} \mathcal{F}_m(\mu_{t_n}) \leq \mathcal{F}_m(\mu_t) + \frac{1}{2} \liminf_{n \rightarrow \infty} \int_{t_n}^t \left(|\mu'_\tau|_{\mathfrak{d}}^2 + \mathcal{I}_m(\mu_\tau) \right) d\tau \\ &\leq \mathcal{F}_m(\mu_t) + \frac{1}{2} \int_0^t \left(|\mu'_\tau|_{\mathfrak{d}}^2 + \mathcal{I}_m(\mu_\tau) \right) d\tau. \end{aligned}$$

It follows that (55) is, in fact an equality. Writing this equality for two times $t = a$ and $t = b$, and by subtraction, we obtain equality in (5).

Acknowledgements This work has been supported by the Project *Conviviality* (JD, ANR-23-CE40-0003) and the PEPR PDE-AI project (FQ, ANR-23-PEIA-0004) of the French National Research Agency. GC and JD thank B. Nazaret for early discussions on kinetic equations.

© 2026 by the authors. This paper may be reproduced, in its entirety, for non-commercial purposes. [CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)

References

1. M. AGUEH AND G. CARLIER, *Generalized solutions of a kinetic granular media equation by a gradient flow approach*, Calc. Var. Partial Differential Equations, 55 (2016), pp. 1–26, Art. 37.
2. L. AMBROSIO, N. FUSCO, AND D. PALLARA, *Functions of bounded variation and free discontinuity*

- problems*, Oxford Mathematical Monographs, The Clarendon Press, Oxford University Press, New York, 2000.
3. 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.
 4. F. BOLLEY, A. GUILLIN, AND F. MALRIEU, *Trend to equilibrium and particle approximation for a weakly selfconsistent Vlasov-Fokker-Planck equation*, Math. Model. Numer. Anal. (M2AN), 44 (2010), pp. 867–884.
 5. G. BRIGATI, G. CARLIER, AND J. DOLBEAULT, *The fundamental solution of a nonlinear kinetic fokker-planck equation*, preprint arXiv 2603.26650, (2026).
 6. G. BRIGATI, J. MAAS, AND F. QUATTROCCHI, *Kinetic Optimal Transport (OTIKIN)—Part I: Second-Order Discrepancies Between Probability Measures*, arXiv 2502.15665v2, (2025).
 7. G. CARLIER AND M. LABORDE, *A splitting method for nonlinear diffusions with nonlocal, nonpotential drifts*, Nonlinear Anal., 150 (2017), pp. 1–18.
 8. J. A. CARRILLO, M. G. DELGADINO, L. DESVILLETES, AND J. S.-H. WU, *The Landau equation as a gradient flow*, Anal. PDE, 17 (2024), pp. 1331–1375.
 9. T. M. COVER AND J. A. THOMAS, *Elements of information theory*, Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second ed., 2006.
 10. 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.
 11. M. H. DUONG, M. A. PELETIER, AND J. ZIMMER, *GENERIC formalism of a Vlasov-Fokker-Planck equation and connection to large-deviation principles*, Nonlinearity, 26 (2013), pp. 2951–2971.
 12. ———, *Conservative-dissipative approximation schemes for a generalized Kramers equation*, Math. Methods Appl. Sci., 37 (2014), pp. 2517–2540.
 13. M. ERBAR, *A gradient flow approach to the Boltzmann equation*, J. Eur. Math. Soc. (JEMS), 26 (2024), pp. 4441–4490.
 14. M. ERBAR AND G. MEGLIOLI, *Gradient flow for a class of diffusion equations with Dirichlet boundary data*, arXiv 2408.05987v2, (2025).
 15. W. GANGBO AND M. WESTDICKENBERG, *Optimal transport for the system of isentropic Euler equations*, Comm. Partial Differential Equations, 34 (2009), pp. 1041–1073.
 16. C. HUANG, *A variational principle for the Kramers equation with unbounded external forces*, J. Math. Anal. Appl., 250 (2000), pp. 333–367.
 17. C. HUANG AND R. JORDAN, *Variational formulations for Vlasov-Poisson-Fokker-Planck systems*, Math. Methods Appl. Sci., 23 (2000), pp. 803–843.
 18. M. IACOBELLI, *A new perspective on Wasserstein distances for kinetic problems*, Arch. Ration. Mech. Anal., 244 (2022), pp. 27–50.
 19. M. JACOBS, I. KIM, AND J. TONG, *The L^1 -contraction principle in optimal transport*, Ann. Sc. Norm. Super. Pisa Cl. Sci. (5), 23 (2022), pp. 1871–1919.
 20. R. JORDAN, D. KINDERLEHRER, AND F. OTTO, *The variational formulation of the Fokker-Planck equation*, SIAM J. Math. Anal., 29 (1998), pp. 1–17.
 21. W. LEE, L. WANG, AND W. LI, *Deep kinetic JKO schemes for Vlasov-Fokker-Planck equations*, arXiv 2603.23901v1, (2026).
 22. F. LÉGER AND M. SYLVESTRE, *A synthetic approach to comparison principles for variational problems, with applications to optimal transport*, arXiv 2506.18884v2, (2025).
 23. F. OTTO, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations, 26 (2001), pp. 101–174.
 24. S. PARK, *A variational perspective on the dissipative Hamiltonian structure of the Vlasov-Fokker-Planck equation*, arXiv 2406.13682v2, (2024).
 25. R. ROSSI AND G. SAVARÉ, *Tightness, integral equicontinuity and compactness for evolution problems in Banach spaces*, Ann. Sc. Norm. Super. Pisa Cl. Sci. (5), 2 (2003), pp. 395–431.
 26. F. SANTAMBROGIO, *Optimal transport for applied mathematicians*, vol. 87 of Progress in Nonlinear Differential Equations and their Applications, Birkhäuser/Springer, Cham, 2015. Calculus of variations, PDEs, and modeling.
 27. ———, *Euclidean, metric, and Wasserstein gradient flows: an overview*, Bull. Math. Sci., 7 (2017), pp. 87–154.
 28. L. SYLVESTRE, *Hölder estimates for kinetic Fokker-Planck equations up to the boundary*, Ars Inven. Anal., (2022).