Ann. Inst. H. Poincaré
Anal. Non Linéaire 39 (2022), 1485–1517
DOI 10.4171/AIHPC/36

JKO estimates in linear and non-linear Fokker–Planck equations, and Keller–Segel: L^p and Sobolev bounds

Simone Di Marino and Filippo Santambrogio

Abstract. We analyze some parabolic PDEs with different drift terms which are gradient flows in the Wasserstein space and consider the corresponding discrete-in-time JKO scheme. We prove with optimal transport techniques how to control the L^p and L^∞ norms of the iterated solutions in terms of the previous norms, essentially recovering well-known results obtained on the continuous-in-time equations. Then we pass to higher-order results, and in particular to some specific BV and Sobolev estimates, where the JKO scheme together with the so-called "five gradients inequality" allows us to recover some estimates that can be deduced from the Bakry–Émery theory for diffusion operators, but also to obtain some novel ones, in particular for the Keller–Segel chemotaxis model.

1. Short introduction

The goal of this paper is to present some estimates on evolution PDEs in the space of probability densities which share two important features: they include a linear diffusion term, and they are gradient flows in the Wasserstein space W_2. These PDEs will be of the form

$$\partial_t \rho - \Delta\rho - \nabla\cdot(\rho\nabla u[\rho]) = 0,$$

complemented with no-flux boundary conditions and an initial condition on ρ_0.

We will in particular concentrate on the Fokker–Planck case, where $u[\rho] = V$ and V is a fixed function (with possible regularity assumptions) independent of ρ, on the case where $u[\rho] = W * \rho$ is obtained by convolution and models interaction between particles, and on the parabolic–elliptic Keller–Segel case where $u[\rho]$ is related to ρ via an elliptic equation. This last case models the evolution of a biological population ρ subject to diffusion but attracted by the concentration of a chemo-attractant produced by the population itself, so that its distribution is ruled by a PDE where the density ρ appears as a source term. In a certain regime (when the production rate is supposed to be much faster than the motion of the cells), its distribution is ruled by a static PDE, and gives rise to a parabolic–elliptic system which is a gradient flow in the variable ρ. The parabolic–parabolic case (see [11]), where the timescale for the cells and for the nutrient are comparable, is also

2020 Mathematics Subject Classification. Primary 35-XX; Secondary 58J35, 49Q22.
Keywords. Wasserstein space, parabolic PDEs, Sobolev bounds.

a gradient flow, in the product space $W_2 \times L^2$, but we will not consider this case. Since we concentrate on the case of bounded domains (supposed to be convex for technical reasons), in the Keller–Segel case the term $u[\rho]$ cannot be expressed as a convolution and requires ad hoc computations.

Throughout the paper, the estimates will be studied on a time-discretized version of these PDEs, consisting of the so-called JKO (Jordan–Kinderlehrer–Otto) scheme, based on iterated optimization problems involving the Wasserstein distance W_2. We will first present zero-order estimates, on the L^p and L^∞ norms of the solution. This is just a translation into the JKO language of well-known properties of these equations. The main goal of this part is hence to popularize the techniques which allow us to handle these estimates at a discrete level. Then we will turn to first-order estimates, i.e. on the gradient of the solutions. This includes in particular estimates on the BV norm of the solution ρ and $W^{1,p}$-like estimates (in particular, the quantity that we will consider is related to $\|\rho^{1/p}\|_{W^{1,p}}$). We point out that a first result in this direction (estimates on the gradient for the JKO scheme) can be found in Lee [25], where the Lipschitz constant of the solution is bounded for the JKO scheme corresponding to a Fokker–Planck equation. The same result, presented in [25] in the periodic case, has recently been extended to the case of convex domains in [16]. However, the technique and the result in this paper are quite different than those in [16, 25].

The estimates we present are non-trivial and seem novel at least in the Keller–Segel case. In the Fokker–Planck case they correspond to a suitable integral version of the well-known Bakry–Émery estimate $|\nabla(P_t f)| \le P_t(|\nabla f|)$ for drift-diffusion operators P_t (see [3]). The interest in this case is to obtain them at a discrete level, on the JKO scheme. Note that, as the Bakry–Émery analogy suggests, these estimates should certainly be obtainable at a continuous level as well, but the computations are not at all easy (and most likely there is some term in the estimates which cannot easily be seen to have a sign, while the discrete-in-time approach allows us to handle it without difficulties). This is an extra reason to also study the zero-order estimate at a discrete level, since some of these first-order estimates require us to use the corresponding zero-order ones.

We insist on the interest of studying these estimates at a discrete level, which can be both theoretical and numerical. From the theoretical point of view, these estimates can pave the way to similar ones for less classical equations, and lead to new existence and regularity results. Moreover, the discrete setting can be useful for some classes of results such as functional inequalities obtained by studying the asymptotic behavior of a flow, as in the Bakry–Émery theory: discrete versions of the relevant inequalities could be used when the existence of a continuous-in-time flow is non-trivial, while the JKO scheme always admits solutions. Regarding numerical schemes, many of them have now been developed using the JKO approach: these estimates can justify their convergence regarding the discretization both in time and in space, since it is classical in numerical analysis to obtain a quantified order of convergence according to a priori knowledge of the smoothness of the true solution. Finally, certain estimates that we present could also be turned into bounds on the optimal displacement in the transport problem appearing in each JKO step, and could help in reducing the complexity of some linear programming algorithms required to find the optimizers.

The JKO scheme provides, for a fixed time step $\tau > 0$, a sequence $(\rho_n^\tau)_n$, where each ρ_{n+1}^τ optimizes a functional depending on ρ_n^τ. All the estimates that we provide are of the following form: a norm, or a quantity comparable to a norm, computed at ρ_{n+1}^τ can be bounded in terms of the same expression computed at ρ_n^τ. Of course, we only want estimates which can be iterated (i.e. the possible increase passing from ρ_n^τ to ρ_{n+1}^τ should be of the order of τ) and which do not explode when $\tau \to 0$. When the same quantity is really decreasing along iterations – in particular if exponential decreasing behavior is obtained – this can be used to study the asymptotic behavior of the solution ρ_t of the PDE as $t \to \infty$. When there is no decreasing behavior, but the increase is controlled, this can be used to justify local-in-time bounds which can provide compactness (to be used either for the convergence of numerical schemes or for other stability results, when data are varying, for instance).

The paper and the results are organized as follows. After this introduction, in Section 2 we present the background that we need for the JKO scheme for gradient flows in the Wasserstein space, including some useful tools such as displacement convexity and the five-gradients inequality, together with general facts on optimal transportation and some details on the functionals that we will use. Section 3 presents the main estimates on the L^p and L^∞ norms of the solution of one step of the JKO scheme in the case where the functional is either a potential energy $\rho \mapsto \int V\,d\rho$ or an interaction energy $\rho \mapsto \frac{1}{2}\int W(x-y)\,d\rho(x)\,d\rho(y)$. In particular, we prove iterable bounds on the L^p norm, for $p < \infty$, when V or W are Lipschitz, as well as better bounds (which include an L^∞ estimate and an L^p estimate which can be used in the limit $p \to \infty$ and also provide an exponential L^∞ bound) in the case of the potential energy under a second-order condition on V. As far as the L^∞ norm is concerned, we also provide a uniform bound stating that the maximal value of ρe^V is decreasing in time under essentially no assumption on V, together with an adaptation for the interaction case, when W is Lipschitz continuous. These results are summarized in Proposition 3.8. Section 4 concentrates, then, on the Keller–Segel case, and reproduces, in this discretized JKO setting, a well-known two-dimensional result (based on [15] and [20]) which states that the L^p norm and the L^∞ norm do not grow too much in time as soon as we are in the subcritical regime, an assumption which allows us to control the entropy with the total energy itself. This very technical result is contained in Theorem 4.5. Finally, Section 5 is devoted to higher-order estimates, which are the core of the paper. The results are expressed in terms of the following quantity: given a convex function $H\colon \mathbb{R}^d \to \mathbb{R}$, we consider $\int H(Z_\rho)\,d\rho$, where

$$Z_\rho := \frac{\nabla\rho}{\rho} + \nabla u[\rho].$$

When $H(z) = |z|^p$ and $\nabla u[\rho]$ is bounded, this quantity (usually denoted by $J_{(p)}(\rho)$) is comparable to $\int |\nabla\rho|^p \rho^{1-p}\,dx$, which can be related by simple algebraic computations to the $W^{1,p}$ norm of $\rho^{1/p}$. On these quantities we prove iterable bounds in the case of the potential energy when V is semiconvex (Proposition 5.4); it is also useful to consider convex functions H other than only powers, which can provide Lipschitz bounds and $W^{1,1}$ regularity. A variant of this result exists for interaction energies, possibly combined with potential energies (Proposition 5.7, where we assume semiconvexity of V and $C^{1,1}$ regularity for W). The results for the potential and interaction cases are contained in the Section 5.1, while the Section 5.2 is devoted to the Keller–Segel case. In this case, the lack of semiconcavity for $u[\rho]$, which is only defined as a solution of an elliptic PDE involving ρ, prevents us from having easy estimates on the error terms, and a different technique is required to bound them: finally, we obtain an iterable estimate on $J_{(p)}(\rho)$ only for $p < 2$, and under the extra assumption that ρ is bounded in an L^r space, with $r = (4-p)/(2-p)$ depending on p and exploding as $p \to 2$. This explains the interest of the zero-order estimates at the JKO level for Keller–Segel, which can indeed guarantee such an L^r assumption.

2. Preliminaries on the JKO scheme

We refer to [2, 30, 31] for the whole theory about gradient flows in the Wasserstein space which justifies the few facts that we list below.

Throughout the sequel, Ω will be a compact convex subset of $\mathbb{R}^d$ with non-empty interior. The compactness of the domain allows us to obtain weak compactness for the set of probability measures on it, and the convexity allows us to handle all boundary terms appearing in the estimates that we will present. We note here that some of the results hold true also in a more general setting, but for the sake of simplicity we will always keep such an assumption implicitly on Ω. Moreover, with a slight abuse of notation, we will use the same letters for the probability measure and its density with respect to the Lebesgue measure.

Whenever a functional $\mathcal{F}: \mathcal{P}(\Omega) \to \mathbb{R} \cup \{+\infty\}$ is given, we fix a time step $\tau > 0$ and a measure $\eta \in \mathcal{P}(\Omega)$, and consider the following minimization problem:

$$\min_{\rho} \Big\{ \mathcal{F}(\rho) + \tfrac{W_2^2(\rho,\eta)}{2\tau} : \rho \in \mathcal{P}(\Omega) \Big\}, \tag{2.1}$$

where W_2 is the 2-Wasserstein distance (see [2, 30, 32]). Before going on with the discussion, let us recall a few important facts about optimal transport and the W_2 distance.

If two probabilities $\mu, \nu \in \mathcal{P}(\Omega)$ are given on a compact domain, the Monge–Kantorovich problem reads

$$\inf \left\{ \int |x - T(x)|^2 \, d\mu : T : \Omega \to \Omega, \; T_{\#}\mu = \nu \right\}.$$

This problem, introduced by Monge [28], was reformulated by Kantorovich [22] in the following convex form:

$$\inf \left\{ \int |x - y|^2 d\gamma : \gamma \in \mathcal{P}(\Omega \times \Omega), \; (\pi_x)_{\#}\gamma = \mu, \; (\pi_y)_{\#}\gamma = \nu \right\}.$$

The square root of the optimal value above defines a distance on the set of probability measures on a given compact space (in the case of non-compactness a condition on the second moments has to be added) which, by the way, metrizes the weak convergence of probabilities (again, under the assumption that Ω is compact; note that we call "weak convergence" the convergence of probability measures in duality with continuous and bounded functions; on compact spaces, this is the weak-* convergence in the dual of C^0). Kantorovich also provided a dual formulation for the above minimization problem, that we can state, for simplicity, using the cost function $|x-y|^2/2$:

$$\frac{1}{2}W_2^2(\mu,\nu) = \sup\left\{\int \varphi\, d\mu + \int \psi\, d\nu : \varphi(x)+\psi(y) \le \frac{1}{2}|x-y|^2\right\}.$$

It is possible to prove the existence of an optimal γ and of an optimal pair (φ,ψ), and, as soon as μ is absolutely continuous, there also exists an optimal transport map T (and the optimal γ will be a measure on $\Omega\times\Omega$ concentrated on the graph of such a map T). Moreover, the optimal φ, called the *Kantorovich potential,* is Lipschitz continuous, and is connected to the optimal T via $T(x) = x - \nabla\varphi(x)$ (we can also write $T = \nabla u$ with $u(x) = |x|^2/2 - \varphi(x)$, and u is a convex function, which is the result of the celebrated Brenier theorem [8, 9]).

Using these tools from optimal transport theory, if Ω is compact and $\mathcal{F}$ is l.s.c. (lower semicontinuous) for the weak convergence of probability measures, then problem (2.1) admits a solution. We will denote the set of solutions as $\mathrm{Prox}^{\tau}_{\mathcal{F}}(\eta)$, mimicking the notation for the proximal operator which is used in hilbertian settings. In some cases (in particular if $\mathcal{F}$ is strictly convex) this proximal operator is single valued (i.e. the minimizer is unique), but this will not be crucial in our analysis.

The JKO scheme (introduced in [21]) consists in iterating the above minimization problem, i.e. starting from ρ_0 and, for fixed $\tau > 0$, defining a sequence $(\rho^\tau_n)_n$ satisfying

$$\rho^\tau_0 = \rho_0, \quad \rho^\tau_{n+1} \in \mathrm{Prox}^{\tau}_{\mathcal{F}}(\rho^\tau_n). \tag{2.2}$$

The above sequence can be used to define a curve of measures $\rho^\tau(t)$ for $t\in[0,T]$, with $\rho^\tau(n\tau) = \rho^\tau_n$ (for instance by piecewise constant interpolation). Under suitable conditions on $\mathcal{F}$ it can be proven that the curves ρ^τ uniformly converge (as curves valued in the Wasserstein space) to a continuous curve ρ which is a solution of the PDE

$$\partial_t\rho - \nabla\cdot\left(\rho\nabla\frac{\delta\mathcal{F}}{\delta\rho}\right) = 0$$

(complemented with no-flux boundary conditions and the initial condition ρ_0), where $\frac{\delta\mathcal{F}}{\delta\rho}$ is the first variation of the functional $\mathcal{F}$ (see [30, Chapter 7]).

In this paper we will always consider the case where $\mathcal{F} = \mathcal{E} + \mathcal{G}$, where

$$\mathcal{E}(\rho) = \begin{cases} \int \rho\log\rho\, dx & \text{if } \rho \ll \mathcal{L}^d, \\ +\infty & \text{otherwise,} \end{cases}$$

is the entropy functional. For the functional $\mathcal{G}$, we will often write $u[\rho] = \delta\mathcal{G}/\delta\rho$ and we will consider three cases:

- The *Fokker–Planck* case $\mathcal{G}(\rho) = \int V \, d\rho$, for a fixed function $V:\Omega \to \mathbb{R}$ acting as a potential, in which case we have $u[\rho] = V$. This case will be called the Fokker–Planck case.
- The *interaction case* $\mathcal{G}(\rho) = \frac{1}{2}\int W(x-y)\, d\rho(x)\, d\rho(y)$, for an even function $W:\mathbb{R}^d \to \mathbb{R}$, in which case we have $u[\rho] = W * \rho$. This case will be called the interaction case; it can be mixed with the previous one by considering $u[\rho] = V + W * \rho$, if explicitly indicated.
- Finally, the *Keller–Segel* case is a particular case arising from mathematical biology: for $\chi > 0$ a given constant take

$$\mathcal{G}(\rho) := -\frac{\chi}{2}\int |\nabla h[\rho]|^2 \, dx = -\frac{\chi}{2}\int h[\rho]\, d\rho, \tag{2.3}$$

where $h[\rho]$ is the only solution of

$$\begin{cases} -\Delta h = \rho & \text{in } \Omega, \\ h = 0 & \text{on } \partial\Omega. \end{cases}$$

Note the negative sign before the integral in the definition of $\mathcal{G}$. It is not difficult to check that we have

$$\frac{\delta\mathcal{G}}{\delta\rho} = -\chi h[\rho].$$

Indeed, $h[\rho + \varepsilon\delta\rho] = h[\rho] + \varepsilon h[\delta\rho]$ and

$$\begin{aligned} \mathcal{G}(\rho + \varepsilon\delta\rho) &= \mathcal{G}(\rho) - \varepsilon\chi \int \nabla h[\rho] \cdot \nabla h[\delta\rho]\, dx + O(\varepsilon^2) \\ &= \mathcal{G}(\rho) + \varepsilon\chi \int h[\rho]\Delta h[\delta\rho]\, dx + O(\varepsilon^2), \end{aligned}$$

which allows us to conclude using $\Delta h[\delta\rho] = -\delta\rho$. This case will be called the Keller–Segel case and is motivated by chemotaxis modeling (see [18, 23] for a description of the model). In dimension $d = 2$, it is well known that this model is well posed and that there is existence (both for the minimization problems in the JKO scheme and for the continuous-in-time PDE, with global-in-time existence) as soon as $\chi < 8\pi$. This is due to a crucial inequality which states that we can bound $\mathcal{E}(\rho)$ in terms of $\mathcal{E}(\rho) + \mathcal{G}(\rho)$ (the problem being that $\mathcal{G}$ is in general not bounded from below, but $\mathcal{E} + \mathcal{G}$ is bounded from below on probability measures as soon as $\chi \le 8\pi$: this implies

$$\mathcal{E}(\rho) \le A + B(\mathcal{E}(\rho) + \mathcal{G}(\rho))$$

with $B = 8\pi/(8\pi - \chi)$). For a mathematical analysis of the Keller–Segel PDE and of the corresponding JKO scheme, we refer to [4–6, 10, 12] and to [29, Chapter 5].

The reader may need to be convinced of the bound from below of $\mathcal{E}(\rho) + \mathcal{G}(\rho)$ when $\chi \leq 8\pi$ in dimension 2, if we are on a bounded domain and $h[\rho]$ is defined with Dirichlet boundary conditions on $\partial\Omega$. This can be seen by observing the following facts. The logarithmic Hardy–Littlewood–Sobolev inequality provides a uniform bound from below on $\int \rho \log \rho \, dx - 4\pi \int \tilde{h}[\rho] \, d\rho$, where $\tilde{h}[\rho](x) := -(2\pi)^{-1} \int_{\mathbb{R}^2} \log(|x-y|) \, d\rho(y)$. Noting that we have $-\Delta \tilde{h}[\rho] = \rho$ and $\tilde{h}[\rho] + \log(R)/(2\pi) \geq 0$ on Ω (where R is the diameter of Ω), we deduce $h[\rho] \leq \tilde{h}[\rho] + \log(R)/(2\pi)$ (since $\tilde{h}[\rho] + \log(R)/(2\pi) - h[\rho]$ is harmonic and non-negative on the boundary). Hence, for $\chi \leq 8\pi$ we have $\mathcal{E}(\rho) + \mathcal{G}(\rho) = \int \rho \log \rho \, dx - 4\pi \int h[\rho] \, d\rho \geq \int \rho \log \rho \, dx - 4\pi \int \tilde{h}[\rho] \, d\rho - 2 \log R$ and this provides the desired bound from below.

A useful tool, introduced in [14] and already used in the framework of the JKO scheme in the same paper in order to obtain BV estimates is the so-called *five-gradients inequality* (note that this name is not present in [14], but the inequality was popularized under this name later on). This inequality states the following:

Lemma 2.1. *Let $\Omega \subset \mathbb{R}^d$ be bounded and convex, $\rho, \eta \in W^{1,1}(\Omega)$ be two probability densities, and $H \in C^1(\mathbb{R}^d)$ be a radially symmetric convex function. Then the following inequality holds:*

$$\int_\Omega \big(\nabla\rho \cdot \nabla H(\nabla\varphi) + \nabla\eta \cdot \nabla H(\nabla\psi)\big) \, dx \geq 0, \tag{2.4}$$

where φ and ψ are the corresponding Kantorovich potentials.

Note that the above result is first proven for $H \in C^2$ (second derivatives are used in the proof) and then, by approximation, it stays true for $H \in C^1$; the same approximation can also be applied to the quite common case $H \in C^1(\mathbb{R}^d \setminus \{0\})$, setting $\nabla H(0) := 0$ (which is coherent with the fact that H is radial), and the result stays true. In particular, we will sometimes apply this to $H(z) = |z|$. For the sake of completeness we sketch a proof in the smooth case: after an integration by parts, noting that the boundary terms have the correct sign thanks to the convexity of Ω, we end up having

$$\begin{aligned}
-\int_\Omega &\big(\nabla\rho \cdot \nabla H(\nabla\varphi) + \nabla\eta \cdot \nabla H(\nabla\psi)\big) \, dx \\
&\leq \int_\Omega D^2 H(\nabla\varphi) \cdot D^2\varphi \, d\rho + \int_\Omega D^2 H(\nabla\psi) \cdot D^2\psi \, d\eta \\
&= \int_{\Omega^2} \big(D^2 H(\nabla\varphi) \cdot D^2\varphi + D^2 H(\nabla\psi) \cdot D^2\psi\big) \, d\gamma \\
&= \int_{\Omega^2} D^2 H(\nabla\varphi(x)) \cdot (D^2\varphi(x) + D^2\psi(x)) \, d\gamma \leq 0,
\end{aligned}$$

where we used γ, the optimal plan between ρ and η. In the last line we used that on the support of γ we have $\nabla\varphi(x) = x - y = -\nabla\psi(y)$ (first-order condition) and $D^2\varphi(x) + D^2\psi(y) \leq 0$ (second-order condition).

Another useful notion in the study of gradient flow is that of displacement convexity, introduced by McCann in [27]. It corresponds to the convexity of a functional along the geodesics of the metric space $(\mathcal{P}(\Omega), W_2)$.

Definition 2.2. Let $\mathcal{H} : \mathcal{P}(\Omega) \to \mathbb{R} \cup \{+\infty\}$ be a functional defined on probability measures on a compact convex domain Ω. We say that $\mathcal{H}$ is displacement convex if for every pair of measures $\rho, \eta \in \mathcal{P}(\Omega)$ there exists a curve ρ_t which is geodesic for the W_2 distance, which connects ρ and ν (i.e. $\rho_0 = \rho$, $\rho_1 = \nu$) and such that $\mathcal{H}(\rho_t) \leq (1-t)\mathcal{H}(\rho) + t\mathcal{H}(\nu)$.

We recall that, whenever ρ is absolutely continuous, the geodesic curve between ρ and ν is unique and is given by

$$\rho_t = (\mathrm{id} - t\nabla\varphi)_{\#}\rho,$$

where φ is the Kantorovich potential between ρ and ν for the cost $c(x, y) = \frac{1}{2}|x-y|^2$. Indeed, $\mathrm{id} - t\nabla\varphi$ is the convex interpolation between the identity map and the optimal transport map $T = \mathrm{id} - \nabla\varphi$.

In [27], McCann provided the condition for the displacement convexity of functionals of the form $\mathcal{H}(\rho) := \int F(\rho(x))\, dx$.

Definition 2.3. Let F be a convex increasing function on $[0, +\infty)$ such that $F(0) = 0$. Then we say that F satisfies the d-McCann condition if $s \mapsto F(\frac{1}{s^d})s^d$ is convex and decreasing.

Note that $s \mapsto F(\frac{1}{s^d})s^d$ being convex and decreasing is enough to guarantee that F itself is convex.

The main result of [27] is indeed the fact that, if F satisfies the d-McCann condition, then the functional $\mathcal{H}$, defined via $\mathcal{H}(\rho) := \int F(\rho(x))\, dx$, is displacement convex in dimension d. In particular, this applies to $F(s) = s^q$, $q > 1$, and to $F(s) = s \log s$ (hence to $\mathcal{H} = \mathcal{E}$).

In [2] the general theory for gradient flows in a metric space is presented, and the assumption of geodesic convexity is crucial, in particular for uniqueness and stability. Here we do not insist on this aspect (by the way, the functional $\mathcal{G}$ in the Keller–Segel case is in general not displacement convex), but we are interested in another property related to displacement convexity. As was first observed in [26], estimates can be provided on $\mathcal{H}(\rho^\tau_{n+1})$ in terms of $\mathcal{H}(\rho^\tau_n)$ when $\mathcal{H}$ is displacement convex, even when the gradient flow that we are considering is the gradient flow of another functional $\mathcal{F}$ (in the case $\mathcal{F} = \mathcal{H}$, the inequality $\mathcal{H}(\rho^\tau_{n+1}) \leq \mathcal{H}(\rho^\tau_n)$ is trivial). The key point is to use the following general estimate.

Lemma 2.4. *Let us consider two absolutely continuous measures $\rho, \eta \in \mathcal{P}(\Omega)$, and a convex function F, such that $F(0) = 0$, satisfying the d-McCann condition. Suppose that the density of ρ is Lipschitz continuous, and that Ω is convex. Then, denoting by φ the Kantorovich potential in the transport from ρ to η, we have*

$$\int_\Omega F(\eta)\, dx \geq \int_\Omega F(\rho)\, dx - \int_\Omega \rho\nabla(F'(\rho)) \cdot \nabla\varphi\, dx. \tag{2.5}$$

Proof. Using the displacement convexity of $\mathcal{H}$, we have

$$\mathcal{H}(\eta) - \mathcal{H}(\rho) = \mathcal{H}(\rho_1) - \mathcal{H}(\rho_0) \geq \frac{d}{dt}\mathcal{H}(\rho_t)|_{t=0},$$

where $(\rho_t)_t$ is the geodesic interpolation between $\eta = \rho_1$ and $\rho = \rho_0$. We just need to prove that we have

$$\frac{d}{dt}\mathcal{H}(\rho_t)|_{t=0} \geq -\int_\Omega \rho\nabla(F'(\rho)) \cdot \nabla\varphi \, dx. \tag{2.6}$$

A formal computation gives

$$\frac{d}{dt}\mathcal{H}(\rho_t) = \int F'(\rho_t)\partial_t\rho_t = \int \nabla(F'(\rho_t)) \cdot v_t \, d\rho_t, \tag{2.7}$$

where v_t is the velocity field of the geodesic curve $(\rho_t)_t$, solving $\partial_t\rho_t + \nabla \cdot (\rho_t v_t) = 0$. The equality $v_0 = -\nabla\varphi$ provides the result.

The reader should be aware that this argument is only formal because of a lack of regularity. Yet everything could be justified by a precise computation of the density of the measure ρ_t. This is classical but technically delicate, and it is done, for instance, in [7, Appendix A2]. In particular, the possible presence of singular parts in the second derivatives of φ justifies the inequality in (2.6), instead of the equality we found using (2.7). ∎

This can then provide estimates on $\mathcal{F}(\rho)$ once we suppose $\rho \in \mathrm{Prox}^\tau_{\mathcal{F}}(\eta)$ and use the optimality condition in the optimization problem solved by ρ, which is of the form

$$\frac{\varphi}{\tau} + \frac{\delta F}{\delta\rho} = \text{const on } \{\rho > 0\},$$

(see [30, Chapter 7] for precise statements and justifications on these optimality conditions). The consequence in the case $\mathcal{F} = \mathcal{E} + \mathcal{G}$ is presented in the next section.

3. Warm-up: L^p and L^∞ estimates for Fokker–Planck and interaction equations

In this section we present various computations leading to L^p estimates (including $p = \infty$) for the simplest case that we consider, i.e. the linear Fokker–Planck case with $\mathcal{G}(\rho) = \int V \, d\rho$. We will then adapt them to the case where the first variation $u[\rho]$ depends on ρ (while for the Fokker–Planck case we do have $u[\rho] = V$ for every ρ) but in a very simple way, by convolution.

We start from the following result.

Proposition 3.1. *Let us consider an absolutely continuous measure $\eta \in \mathcal{P}(\Omega)$ and $F \in C([0,\infty)) \cap C^2((0,\infty))$ a convex function satisfying the d-McCann condition. Let $\mathcal{G}\colon \mathcal{P}(\Omega) \to \overline{\mathbb{R}}$ be a given functional, $\rho \in \mathrm{Prox}^{\tau}_{\mathcal{E}+\mathcal{G}}(\eta)$, and $u[\rho] := \delta\mathcal{G}/\delta\rho$. Suppose that $u[\rho]$ is Lipschitz continuous. Then ρ is also Lipschitz continuous, and bounded from below by a positive constant, and, if Ω is convex, we have*

$$\int_\Omega F(\eta)\,dx \geq \int_\Omega F(\rho)\,dx + \tau \int_\Omega \big(F''(\rho)|\nabla\rho|^2 + \rho F''(\rho)\nabla\rho\cdot\nabla u[\rho]\big). \tag{3.1}$$

Proof. This estimate is a combination of the one in Lemma 2.4 with the optimality conditions characterizing ρ. Indeed, we have (see [30, Chapter 8] and adapt the computations which are presented there just in the case $u[\rho] = V$)

$$\log\rho + u[\rho] + \frac{\varphi}{\tau} = \text{const}, \quad \text{hence } \frac{\nabla\rho}{\rho} + \nabla u[\rho] + \frac{\nabla\varphi}{\tau} = 0,$$

these equalities being true a.e. on Ω since we have $\rho > 0$ a.e. (for this, see the proof in [30, Chapter 8]). As a consequence, $\log\rho$ is Lipschitz continuous, and we can apply the result of Lemma 2.4, replacing $\nabla\varphi$ with $-\tau(\nabla u[\rho] - \nabla\rho/\rho)$. ∎

Let us analyze first the purely linear Fokker–Planck case, i.e. the case where $u[\rho] = V$ does not depend on ρ, and let us concentrate on L^p estimates.

Proposition 3.2. *Let $\eta \in L^p(\Omega)$, with $p < \infty$, be a probability measure and $\mathcal{G}\colon \mathcal{P}(\Omega) \to \mathbb{R}$ be defined via $\mathcal{G}(\rho) := \int V\,d\rho$ for a given Lipschitz function $V\colon \Omega \to \mathbb{R}$. Take $\rho \in \mathrm{Prox}^{\tau}_{\mathcal{E}+\mathcal{G}}(\eta)$: then ρ is Lipschitz continuous and bounded from below by a positive constant, and*

- *denoting by* $\mathrm{Lip}(V)$ *the Lipschitz constant of V, we have*

$$\int \eta^p\,dx \geq \Big(1 - \tau\frac{p(p-1)}{4}\mathrm{Lip}(V)^2\Big)\int \rho^p\,dx;$$

- *if $\Delta V \leq A$ in Ω and $\nabla V \cdot n \geq 0$ on $\partial\Omega$, then*

$$\int \eta^p\,dx \geq (1 - \tau(p-1)A)\int \rho^p\,dx.$$

If moreover $\eta \in L^\infty(\Omega)$ we have

$$\|\rho\|_\infty \leq \|\eta\|_\infty\Big(1 + \tau\frac{A}{d}\Big)^d.$$

Proof. The estimates on the L^p norm are consequences of the estimate in Proposition 3.1, applied to $F(s) = s^p$. In this case we obtain

$$\int_\Omega \eta^p\,dx \geq \int_\Omega \rho^p\,dx + \tau p(p-1)\Big(\int_\Omega \rho^{p-2}|\nabla\rho|^2 + \rho^{p-1}\nabla\rho\cdot\nabla V\,dx\Big).$$

For the first estimate, we apply Young's inequality which gives

$$\int \rho^{p-2}\nabla\rho\cdot(\rho\nabla V)\,dx \geq -\int \rho^{p-2}|\nabla\rho|^2 - \frac{1}{4}\int \rho^p|\nabla V|^2\,dx,$$

which proves the claim.

For the second, we ignore the positive term $\int_\Omega \rho^{p-2}|\nabla\rho|^2$ and we rewrite the remaining terms as

$$\tau p(p-1)\int_\Omega \rho^{p-1}\nabla\rho\cdot\nabla V\,dx = \tau(p-1)\int_\Omega \nabla(\rho^p)\cdot\nabla V\,dx$$

and we integrate by parts, using our assumptions on V.

The last statement, about the L^∞ norm, can be proven differently, following the same strategy as in [30, Proposition 7.32]. The adaptations to be performed in the proof are the following: instead of a minimum point of φ we take a minimum point for $\varphi+\tau V$; we first use $\det(I-D^2\varphi)\leq(1-\Delta\varphi/d)^d$, which is a consequence of the geometric–arithmetic mean inequality, and then $-\Delta\varphi\leq\tau\Delta V\leq\tau A$. The assumption on $\partial V/\partial n$ is needed to handle a possible maximizer on the boundary. This statement can first be proven for $V\in C^2$ and then by approximation, for a less regular V, where the assumptions on the Laplacian and on the sign of the normal derivative should be interpreted as a condition on the distributional divergence of the vector field ∇V, extended as 0 outside Ω. ■

Remark 3.3. We observe that, in the continuous-time limit of the JKO scheme, the second estimate would give $\int\rho_t^p\,dx\leq e^{(p-1)At}\int\rho_0^p\,dx$. It is then possible to raise to the power $1/p$, take the limit $p\to\infty$, and obtain $\|\rho_t\|_\infty\leq e^{At}\|\rho_0\|_\infty$. Unfortunately, this cannot be done in discrete time since if we first send $p\to\infty$ for fixed $\tau>0$, the coefficient on the right-hand side in front of $\int\rho^p$ can become negative. This is why we presented a different technique for the L^∞ estimate, but we can notice that, asymptotically as $\tau\to 0$, the two results coincide.

Also, we observe that the first estimate is not suitable for a limit $p\to\infty$, as the coefficient on the right-hand side is quadratic in p, so that its continuous-in-time version is $\int\rho_t^p\,dx\leq e^{p(p-1)\,\mathrm{Lip}(V)^2t/4}\int\rho_0^p\,dx$.

Proposition 3.4. *Let us consider a probability measure $\eta\in L^\infty(\Omega)$, and let $\mathcal{G}:\mathcal{P}(\Omega)\to\mathbb{R}$ be defined via $\mathcal{G}(\rho):=\int V\,d\rho$ for a given bounded function $V:\Omega\to\mathbb{R}$. Take $\rho\in\mathrm{Prox}^\tau_{\mathcal{E}+\mathcal{G}}(\eta)$; then ρ is bounded, and satisfies*

$$\|\rho e^V\|_\infty\leq\|\eta e^V\|_\infty.$$

Proof. The proof, consisting in looking at the maximum point of ρe^V, i.e. of $\log\rho+V$, is exactly the same as in [19, Lemma 2.4]. Indeed, [19] concerns the case of a functional of the form

$$\rho\mapsto\int U\Big(\frac{\rho(x)}{m(x)}\Big)m(x)\,dx,$$

but here we can write

$$\mathcal{E}(\rho) + \int V\, d\rho = \int (\rho \log \rho + \rho V)\, dx = \int F\Big(\frac{\rho}{m}\Big) dm,$$

for $F(s) = s \log s$ and $m = e^{-V}$, thus falling into the setting of [19, Lemma 2.4]. ■

Remark 3.5. Actually, as is proven in [19], the same bound also holds from below. Moreover, morally this upper bound weighted by e^V should also work for the L^p estimates, but it would require the geodesic convexity of the L^p norm with respect to e^{-V}, which requires V to be convex (and if V is only λ-convex, it does not work). It is interesting to see that the assumption is the opposite of the one in Proposition 3.2, where we needed upper bounds on $D^2 V$.

The case where $u[\rho]$ depends on ρ, but in a very good way, is easy to handle. Take an even function $W\colon \mathbb{R}^d \to \mathbb{R}$ and consider

$$u[\rho] = W * \rho, \quad \text{i.e. } \mathcal{G}(\rho) = \frac{1}{2} \int W\, d\rho(x)\, d\rho(y).$$

The first of the two estimates in Proposition 3.2 is easy to adapt, while unfortunately the other one, based on second-order assumptions but also on the boundary behavior, cannot be easily translated in terms of W. The same problem occurs for the L^∞ estimate of Proposition 3.2. We can therefore state the following:

Proposition 3.6. *Let $p < \infty$ and $\eta \in L^p(\Omega)$ be a probability measure, and let $\mathcal{G}\colon \mathcal{P}(\Omega) \to \mathbb{R}$ be defined via $\mathcal{G}(\rho) := \frac{1}{2} \int W(x-y)\, d\rho(x)\, d\rho(y)$ for an even Lipschitz function $W\colon \mathbb{R}^d \to \mathbb{R}$. Take $\rho \in \mathrm{Prox}^\tau_{\mathcal{E}+\mathcal{G}}(\eta)$: then ρ is Lipschitz continuous and bounded from below by a positive constant, and*

$$\int \eta^p\, dx \geq \Big(1 - \tau \frac{p(p-1)}{4} \mathrm{Lip}(W)^2\Big) \int \rho^p\, dx.$$

Proof. The proof is identical to that in Proposition 3.2, which does not depend on the fact that $u[\rho]$ depends or not upon ρ, but only on its Lipschitz bounds. Hence, we just have to observe that we have $\mathrm{Lip}(W * \rho) \leq \mathrm{Lip}(W)$. ■

On the other hand, it is easier to extend the estimate in Proposition 3.4, but this requires an adaptation if one wants to iterate it.

Proposition 3.7. *Let $\eta \in L^\infty(\Omega)$ be a probability measure and $\mathcal{G}\colon \mathcal{P}(\Omega) \to \mathbb{R}$ be defined via $\mathcal{G}(\rho) := \frac{1}{2} \int W(x-y)\, d\rho(x)\, d\rho(y)$ for an even Lipschitz function $W\colon \mathbb{R}^d \to \mathbb{R}$. Take $\rho \in \mathrm{Prox}^\tau_{\mathcal{E}+\mathcal{G}}(\eta)$: then ρ is Lipschitz continuous and bounded from below by a positive constant, and*

$$\|\rho e^{u[\rho]}\|_\infty \leq \|\eta e^{u[\rho]}\|_\infty.$$

This implies in particular the (more useful) estimate

$$\|\rho e^{u[\rho]}\|_\infty e^{\mathcal{E}(\rho)+\mathcal{G}(\rho)} \leq \|\eta e^{u[\eta]}\|_\infty e^{\mathcal{E}(\eta)+\mathcal{G}(\eta)} e^{\tau\, \mathrm{Lip}(W)^2/2}.$$

Proof. The estimate $\|\rho e^{u[\rho]}\|_\infty \le \|\eta e^{u[\rho]}\|_\infty$ can be trivially obtained in the same way as in the case where $u[\rho] = V$ does not depend on ρ. Then we observe that we have

$$u[\rho] \le u[\eta] + \|W * (\rho - \eta)\|_\infty \le u[\eta] + \mathrm{Lip}(W) W_1(\rho, \eta).$$

We then use

$$\begin{aligned}\mathrm{Lip}(W) W_1(\rho, \eta) \le \mathrm{Lip}(W) W_2(\rho, \eta) &\le \frac{\tau}{2} \mathrm{Lip}(W)^2 + \frac{W_2^2(\rho, \eta)}{2\tau} \\ &\le \frac{\tau}{2} \mathrm{Lip}(W)^2 + \mathcal{F}(\eta) - \mathcal{F}(\rho),\end{aligned}$$

where $\mathcal{F} = \mathcal{E} + \mathcal{G}$. This provides the claimed result. ∎

We can now deduce, from the various estimates of this section, the following bounds, whose proofs are just combinations of the arguments above.

Proposition 3.8. *Suppose $\rho_0 \in L^p(\Omega)$ is a probability measure. Let us consider the sequence ρ_n^τ defined via the JKO scheme as in (2.2), where $\mathcal{F} = \mathcal{E} + \mathcal{G}$; we will denote by ρ_t the piecewise constant interpolation of $\rho_{n\tau} = \rho_n^\tau$.*

In the Fokker–Planck case $\mathcal{G}(\rho) = \int V\, d\rho$, we have the following:

- *if $p < +\infty$ and V is Lipschitz continuous, then the norm $\|\rho_t\|_p$ grows at most exponentially in time;*
- *if $p \in [1, +\infty]$ is arbitrary and V is such that ΔV is bounded from above and $\nabla V \cdot n \ge 0$ on $\partial\Omega$, then $\|\rho_t\|_p$ grows at most exponentially in time.*
- *if $p = \infty$ and V is bounded, then the norm $\|\rho_t e^V\|_\infty$ is non-increasing in time.*

In the case of the interaction functional $\mathcal{G}(\rho) = \frac{1}{2} \int W(x - y)\, d\rho(x)\, d\rho(y)$, we have the following:

- *if $p < +\infty$ and W is Lipschitz continuous, then the norm $\|\rho_t\|_p$ grows at most exponentially in time;*
- *if $p = \infty$ and W is Lipschitz continuous, then the quantity*

$$\|\rho_t e^{W * \rho_t}\|_\infty e^{\mathcal{F}(\rho_t)}$$

grows at most exponentially in time, and in particular the same is true for the norm $\|\rho_t\|_\infty$.

4. L^p and L^∞ estimates for the Keller–Segel case

Proposition 4.1. *For every $p \ge 1$ and $K > 0$ the function $F_{p,K}(s) = (s^p - K s^{\frac{d-1}{d}})_+$ is convex and satisfies the d-McCann condition.*

Proof. It is clear that $F_{p,K}$ is convex and increasing. It is then sufficient to compute

$$F_{p,K}\Big(\frac{1}{s^d}\Big)s^d = \Big(\frac{1}{s^{d(p-1)}} - Ks\Big)_+,$$

which is clearly convex as it is the positive part of a convex function. ∎

Remark 4.2. The goal of Proposition 4.1 is to find a function, satisfying the d-McCann condition, growing as a power s^p, but vanishing before a certain threshold. We could have considered, instead, the simpler function $\widetilde{F}_{p,K}(s) = (s-K)_+^p$. However, the computations are messier and the d-McCann condition seems to be true only in the case $p \geq \frac{4d}{3d+1}$, which explains why we preferred to use $F_{p,K}$, for which the condition is satisfied for any p.

For the study of the Keller–Segel case we will use the following functional $\mathcal{G}$:

$$\mathcal{G}(\rho) := -\frac{\chi}{2}\int |\nabla h[\rho]|^2\,dx = -\frac{\chi}{2}\int h[\rho]\,d\rho,$$

where $\chi > 0$ is a given constant and $h[\rho]$ is the only solution of

$$\begin{cases} -\Delta h = \rho & \text{in } \Omega, \\ h = 0 & \text{on } \partial\Omega. \end{cases}$$

Note the negative sign before the integral in the definition of $\mathcal{G}$. As we previously did in Section 2, it is not difficult to check that we have

$$\frac{\delta \mathcal{G}}{\delta \rho} = -\chi h[\rho].$$

Proposition 4.3. *Let us consider an absolutely continuous $\eta \in \mathcal{P}(\Omega)$ and F a convex (but not necessarily smooth) function satisfying the d-McCann condition, and let $\mathcal{H}(\rho) = \frac{1}{q}\int_\Omega \rho^q\,dx$ with $q > d/2$. Then, let $\delta > 0$ and consider $\rho \in \mathrm{Prox}^\tau_{\mathcal{E}+\delta\mathcal{H}+\mathcal{G}}(\eta)$; we have*

$$\begin{aligned}\int_\Omega F(\eta)\,dx \geq \int_\Omega F(\rho)\,dx - \tau\chi\int_\Omega [\rho F'^{,r}(\rho) - F(\rho)]\rho\,dx \\ + \tau\int_\Omega |\nabla\rho|^2 F''_{\mathrm{ac}}(\rho)\,dx, \end{aligned} \tag{4.1}$$

where F''_{ac} is the absolutely continuous part of the derivative of F' and $F'^{,r}$ is the right derivative of F.

Proof. Since $\rho \in L^q$ with $q > d/2$ (this is the main reason for adding the term $\delta\mathcal{H}$ in this statement, since the condition $\rho \in L^q$ allows us to start the regularity argument on ρ) we have that $h[\rho] \in W^{2,q}$ is bounded and Hölder continuous. Looking at the optimality condition

$$\log\rho + \delta\rho^{q-1} = c - \varphi/\tau + \chi h[\rho],$$

we deduce that ρ is also Hölder continuous, and bounded from above and below (in particular we deduce $\rho \in L^\infty$, but we would not have been able to do so without starting from $\rho \in L^q$). As a consequence, by elliptic regularity, $h[\rho]$ is a C^2 function. Then ρ has the same regularity as the worst between φ and $h[\rho]$, and in particular ρ is Lipschitz continuous. Now let us assume for a while that F is convex and C^2: we start from (3.1) and replace $u[\rho]$ with $\delta\rho^{q-1} - \chi h[\rho]$. This provides

$$\begin{aligned}\int_\Omega F(\eta)\,dx \geq \int_\Omega F(\rho)\,dx + \tau \int_\Omega F''(\rho)|\nabla\rho|^2\,dx \\ + \delta\tau \int_\Omega \rho F''(\rho)\nabla\rho\cdot\nabla(\rho^{p-1})\,dx - \chi\tau\int_\Omega \rho F''(\rho)\nabla\rho\cdot\nabla h[\rho]\,dx.\end{aligned}$$

Noting that $s \mapsto sF''(s)$ is the derivative of $s \mapsto sF'(s) - F(s)$, we can integrate the last term by parts, thus obtaining

$$\begin{aligned}\int_\Omega \rho F''(\rho)\nabla\rho\cdot\nabla h[\rho]\,dx &= -\int_\Omega (\rho F'(\rho) - F(\rho))\Delta h[\rho]\,dx \\ &\quad + \int_{\partial\Omega} (\rho F'(\rho) - F(\rho))\nabla h[\rho]\cdot\nu\,d\sigma \\ &\leq \int_\Omega (\rho F'(\rho) - F(\rho))\rho\,dx.\end{aligned}$$

In the above inequality, $d\sigma$ denotes the uniform $(d-1)$-measure on the boundary $\partial\Omega$ (and the terms we integrate on the boundary make sense, because of regularity). Moreover, we used $\nabla h[\rho]\cdot\nu \leq 0$ (a consequence of the positivity of $h[\rho]$ together with its Dirichlet boundary condition on $\partial\Omega$) and $\rho F'(\rho) - F(\rho) \geq 0$ (a consequence of the convexity of F together with $F(0) = 0$). Using this information, and the positivity of $\nabla\rho\cdot\nabla(\rho^{p-1})$, we get

$$\int_\Omega F(\eta)\,dx \geq \int_\Omega F(\rho)\,dx + \tau\int_\Omega F''(\rho)|\nabla\rho|^2 - \chi\tau\int_\Omega (\rho F'(\rho) - F(\rho))\rho\,dx. \quad (4.2)$$

Now let us consider any convex function F (also not smooth) and let us approximate it by smooth convex functions satisfying McCann conditions. In order to do this, instead of directly approximating F by convolution, we approximate by convolution the function M given by $M(s) = s^d F(s^{-d})$. We want to define M_ε to be the convolution of M with a standard convolution kernel supported on $[-\varepsilon, \varepsilon]$, so that M_ε is also convex and decreasing. Yet M could be impossible to extend with finite values on $s < 0$, so that we will have $M_\varepsilon = +\infty$ close to 0. Therefore, before convolving, we also modify M close to 0, replacing M with its tangent at $s = \varepsilon$. We then define $F_\varepsilon(s) = sM_\varepsilon(s^{-1/d})$ so that F_ε satisfies McCann conditions. Note that any function satisfying McCann conditions is automatically convex, and so is F_ε. Note that the modification that we performed on M only modifies the values of $F_\varepsilon(s)$ for $s > \varepsilon^{-1/d}$, which is irrelevant since ρ only takes bounded values. If we denote by I an interval, bounded away from 0 and $+\infty$, where ρ takes its values, we also have $F_\varepsilon \to F$ uniformly on I, as a consequence of the local uniform convergence of M_ε to M.

For convex functions this implies convergence of the derivatives: we have $F'_\varepsilon \to F'$ at any differentiability point of F (and $\limsup F'_\varepsilon \le F'^{,r}$ at every non-differentiability point, where $F'^{,r}$ stands for the right-derivative of F), and $F''_\varepsilon \to F''_{\mathrm{ac}}$ almost everywhere, where F''_{ac} is the absolutely continuous part of F''. This implies $|\nabla\rho|^2 F''_\varepsilon(\rho) \to |\nabla\rho|^2 F''_{\mathrm{ac}}(\rho)$ since the convergence holds on a.e. level set of ρ, and we have $|\nabla\rho| = 0$ a.e. on $\{x \in \Omega : \rho(x) \in A\}$, where A is the set of values on which we do not have the convergence $F''_\varepsilon \to F''_{\mathrm{ac}}$. Finally, we note that $F'_\varepsilon(\rho)$, $F'^{,r}(\rho)$ are bounded since $\rho \in I$ is bounded from above and from below.

In particular, we can pass to the limit in (4.2) using Fatou's lemma, thus obtaining

$$\int_\Omega F(\eta)\,dx \ge \int_\Omega F(\rho)\,dx + \tau \int_\Omega F''_{\mathrm{ac}}(\rho)|\nabla\rho|^2\,dx - \chi\tau \int_\Omega (\rho F'^{,r}(\rho) - F(\rho))\rho\,dx,$$

which proves the claim. ■

The following estimates require $d = 2$ and χ sufficiently small.

Indeed, we first use the following fact, which we mentioned in Section 2: for $\chi \le 8\pi$, if $d = 2$, the functional $\mathcal{F} := \mathcal{E} + \mathcal{G}$ is bounded from below; hence, if $\chi < 8\pi$, then there exist constants $A, B > 0$ such that $\mathcal{E}(\rho) \le A + B(\mathcal{E}(\rho) + \mathcal{G}(\rho))$ is true for every $\rho \in \mathcal{P}(\Omega)$.

Moreover, another point of the proof where we use $d = 2$ is an inequality where we use the BV norm to estimate the L^2 norm of a given function, which is a two-dimensional fact. The precise statement that we need is the following lemma:

Lemma 4.4. *Let $\Omega \subset \mathbb{R}^d$ be a bounded connected domain with Lipschitz boundary. Then, for every number $a \in (0, |\Omega|)$ there exists a constant $C = C(a,\Omega)$ such that*

$$\|u\|_{L^{\frac{d}{d-1}}} \le C \int_\Omega |\nabla u| \quad \textit{for every } u \in W^{1,1}(\Omega) \textit{ with } |\{u = 0\}| \ge a.$$

In particular, this applies to convex domains, and in dimension $d = 2$ provides a bound on the L^2 norm in terms of the L^1 norm of the gradient.

Note that the statement would also be true for BV functions instead of $W^{1,1}$, but we will only apply it to functions which actually belong to a Sobolev space.

Proof. The continuous embedding of $W^{1,1}$ and BV into L^{1^*} with $1^* = d/(d-1)$ is a well-known fact for which we refer, for instance, to [1, Corollary 3.49]. This would give the desired bound, with a constant only depending on Ω but not on a, in terms of the full BV norm of u, i.e. $\int |\nabla u| + |u|$. Hence, we just need to prove a sort of Poincaré inequality of the form

$$\|u\|_{L^1} \le C(a,\Omega) \int_\Omega |\nabla u| \quad \text{for every } u \in W^{1,1}(\Omega) \text{ with } |\{u = 0\}| \ge a.$$

This can be proven by contradiction. If it is false, for fixed a, we have a sequence of functions u_n satisfying

$$1 = \|u_n\|_{L^1} \ge n\|\nabla u_n\|_{L^1},$$

together with $|\{u_n = 0\}| \geq a$, which means that we can choose a sequence of functions $v_n \in L^\infty$ with $0 \leq v_n \leq 1$, $\int v_n \geq a$, and $\int u_n v_n = 0$ (take $v_n = I_{A_n}$ for some sets $A_n \subset \{u_n = 0\}$). Then the compact embedding of BV into L^1 allows us to extract a converging subsequence such that $u_n \to u$ in L^1 and $u \in \mathrm{BV}$ with $\|Du\| = 0$, i.e. u is a constant. We write $u = \alpha$ and we have $\|u\|_{L^1} = 1$ and hence $\alpha \neq 0$. If we also extract a weakly-* converging subsequence $v_n \rightharpoonup v$ in L^∞, then we have the following contradiction: $\alpha \int v = \lim \int u_n v_n = 0$ but $\alpha \neq 0$ and $\int v \geq a > 0$. ■

Theorem 4.5. *Let $\eta \in L^p(\Omega)$ be a probability measure and $\mathcal{G}$ be as in (2.3). For every $c_0 > 0$ there exists $D_1 = D_1(c_0, p) > 0$ such that whenever $\mathcal{F}(\eta) := \mathcal{E}(\eta) + \mathcal{G}(\eta) \leq c_0$, for $K \geq K(c_0, p)$ there exists $\rho \in \mathrm{Prox}^\tau_{\mathcal{E}+\mathcal{G}}(\eta)$ with*

$$\int_\Omega F_{p,K}(\rho)\,dx \leq \int_\Omega F_{p,K}(\eta)\,dx + \tau D_1. \tag{4.3}$$

In particular, $\rho \in L^p(\Omega)$.

Moreover, let ρ^τ_n as in (2.2) with $\rho_0 := \eta$, paying attention to selecting the measures so that the above estimate applies, we have $\|\rho^\tau_n\|^p_p \leq \|\rho^\tau_0\|^p_p + (1 + n\tau)D_2$ with D_2 again depending only on c_0 and p. A similar estimate holds for $p = \infty$: $\|\rho_n\|_\infty \leq (1 + n\tau)C(c_0, \|\rho_0\|_\infty)$.

Proof. Let us consider $\mathcal{H}$ and $\rho_\delta = \mathrm{Prox}^\tau_{\mathcal{E}+\delta\mathcal{H}+\mathcal{G}}(\eta)$ as in Proposition 4.3. Then we use $F = F_{p,K}$ with $K = k^{p-\frac{d-1}{d}}$ in (4.1). The conditions on K (or k), which has to be chosen large enough, will be made precise later.

First, note that we have $F_{p,K}(s) > 0$ if and only if $s > k$, as well as

$$0 \leq \rho F'^{,r}_{p,K}(\rho) - F_{p,K}(\rho) \leq p\rho^p \mathbb{1}_{\rho \geq k}, \quad (F''_{p,K})_{\mathrm{ac}}(\rho) \geq p(p-1)\rho^{p-2}\mathbb{1}_{\rho \geq k}.$$

This allows us to write the inequality

$$\begin{aligned}\int_\Omega F_{p,K}(\eta)\,dx - \int_\Omega F_{p,K}(\rho_\delta)\,dx \geq -\tau\chi p \int_{\rho_\delta \geq k} \rho_\delta^{p+1}\,dx \\ + \tau p(p-1)\int_{\rho_\delta \geq k} |\nabla\rho_\delta|^2 \rho_\delta^{p-2}\,dx,\end{aligned} \tag{4.4}$$

Consider a constant $c_1 = c_1(\Omega)$ such that we have $|Du|(\Omega) \geq c_1\|u\|_2$ for every function u satisfying $2|\{|u| > 0\}| \leq |\Omega|$ (a constant which exists because of Lemma 4.4). Then, whenever $2|\{\rho \geq k\}| \leq \Omega$ (note that $k \geq 2|\Omega|^{-1}$ is enough) and $p > -1$, we have

$$\begin{aligned}\int_{\rho \geq k} \rho\,dx \int_{\rho \geq k} |\nabla\rho|^2\rho^{p-2}\,dx &\geq \left(\int_{\rho \geq k} |\nabla\rho|\rho^{\frac{p-1}{2}}\,dx\right)^2 \\ &= \frac{4}{(p+1)^2}\left(\int_\Omega |\nabla(\rho^{\frac{p+1}{2}} - k^{\frac{p+1}{2}})_+|\,dx\right)^2 \\ &\geq \frac{4c_1^2}{(p+1)^2}\int_\Omega (\rho^{\frac{p+1}{2}} - k^{\frac{p+1}{2}})_+^2\,dx.\end{aligned}$$

Now we can use the inequality (valid for $p > 0$)

$$\rho^{p+1} \leq \begin{cases} \dfrac{(\rho^{\frac{p+1}{2}} - k^{\frac{p+1}{2}})_+^2}{(1 - 2^{-(p+1)/2p})^2} & \text{if } \rho \geq 2^{1/p}k, \\ 2k^p \rho & \text{if } 0 \leq \rho \leq 2^{1/p}k. \end{cases}$$

Using $\frac{p+1}{2p} > \frac{1}{2}$, we have $1/(1 - 2^{-\frac{p+1}{2p}})^2 \leq (1 - 1/\sqrt{2})^{-2} = 6 + 4\sqrt{2} \leq 12$, and we find

$$\int_\Omega \rho^{p+1}\, dx \leq 2k^p \int_\Omega \rho\, dx + 12 \int_\Omega (\rho^{\frac{p+1}{2}} - k^{\frac{p+1}{2}})_+^2\, dx,$$

$$12 \int_\Omega (\rho^{\frac{p+1}{2}} - k^{\frac{p+1}{2}})_+^2\, dx \geq \int_\Omega \rho^{p+1}\, dx - 2k^p.$$

In particular,

$$\int_{\rho_\delta \geq k} |\nabla \rho_\delta|^2 \rho_\delta^{p-2}\, dx \geq \frac{c_1^2}{3(p+1)^2 \int_{\rho_\delta \geq k} \rho_\delta\, dx} \left(\int_\Omega \rho_\delta^{p+1}\, dx - 2k^p \right).$$

Now, it is sufficient to find k such that

$$\int_{\rho_\delta \geq k} \rho_\delta\, dx \leq \alpha(p) := \frac{(p-1)c_1^2}{3(p+1)^2 \chi}$$

in order to obtain

$$\int_\Omega F_{p,K}(\eta)\, dx - \int_\Omega F_{p,K}(\rho_\delta)\, dx \geq -2\tau\chi p k^p,$$

which would give the first part of the claim.

In order to estimate $\int_{\rho_\delta \geq k} \rho_\delta$ we just use

$$\begin{aligned} \int \rho_\delta |\log \rho_\delta| &\leq \mathcal{E}(\rho_\delta) + 2|\Omega|e^{-1} \leq A + 2|\Omega|e^{-1} + B\mathcal{F}(\rho_\delta) \\ &\leq A + 2|\Omega|e^{-1} + B \min_\rho \left\{ \mathcal{F}(\rho) + \delta\mathcal{H}(\rho) + \frac{W_2^2(\rho, \eta)}{2\tau} \right\}. \end{aligned}$$

It is easy to prove the Γ-convergence of $\mathcal{F} + \delta\mathcal{H}$ to $\mathcal{F}$ (the Γ-liminf inequality is trivial, and the Γ-limsup is straightforward for $\rho \in L^q$, which is a subset of $\mathcal{P}(\Omega)$ which is dense in energy for the functional $\mathcal{F}$), and this implies (using the continuity of the Wasserstein distance for the weak convergence)

$$\lim_{\delta \to 0} \min_\rho \left\{ \mathcal{F}(\rho) + \delta\mathcal{H}(\rho) + \frac{W_2^2(\rho, \eta)}{2\tau} \right\} = \min_\rho \left\{ \mathcal{F}(\rho) + \frac{W_2^2(\rho, \eta)}{2\tau} \right\}.$$

Hence, we have, for small δ,

$$\begin{aligned} \min_\rho \left\{ \mathcal{F}(\rho) + \delta\mathcal{H}(\rho) + \frac{W_2^2(\rho, \eta)}{2\tau} \right\} &\leq \min_\rho \left\{ \mathcal{F}(\rho) + \frac{W_2^2(\rho, \eta)}{2\tau} \right\} + 1 \\ &\leq \mathcal{F}(\eta) + 1 \leq c_0 + 1. \end{aligned}$$

Then we can choose k looking at

$$\int_{\rho_\delta \geq k} \rho_\delta \leq \frac{1}{\log k} \int_{\rho_\delta \geq k} \rho_\delta |\log \rho_\delta| \leq \frac{A + 2|\Omega| e^{-1} + B(c_0 + 1)}{\log k}.$$

It is then enough to take k large enough depending on c_0 and p, and in particular we impose $k \geq 2|\Omega|^{-1}$ and

$$k \geq k(c_0, p) \quad \text{with } \frac{A + 2|\Omega| e^{-1} + B(c_0 + 1)}{\log k(c_0, p)} = \frac{(p-1)c_1^2}{3(p+1)^2 \chi}.$$

The value of $K(c_0, p)$ is defined accordingly, and then we find, for $K \geq K(c_0, p)$,

$$\int_\Omega F_{p,K}(\rho_\delta)\, dx \leq \int_\Omega F_{p,K}(\eta)\, dx + \tau D_1.$$

Now we let $\delta \to 0$: up to a subsequence we have $\rho_\delta \rightharpoonup \rho$ and the limit ρ satisfies the same inequality; moreover we also have that $\rho \in \mathrm{Prox}^\tau_{\mathcal{E}+\mathcal{F}}(\eta)$ thanks to the Γ-convergence of the functionals. For the global-in-time estimate we can iterate the previous result, thanks to the fact that $\mathcal{F}(\rho_n)$ is decreasing in n, in order to get

$$\int_\Omega F_{p,K}(\rho_n)\, dx \leq \int_\Omega F_{p,K}(\rho_0)\, dx + n\tau D_1.$$

Then we can use the inequalities

$$-k^{p-1}\rho + \rho^p \leq (\rho^p - k^{p-1}\rho)_+ \leq F_{p,K}(\rho) \leq \rho^p$$

(remember $K = k^{p-(d-1)/d}$) to conclude

$$\begin{aligned}\int_\Omega |\rho_n|^p\, dx &\leq \int_\Omega F_{p,K}(\rho_n)\, dx + k(c_0, p)^{p-1} \leq \int_\Omega |\rho_0|^p\, dx + n\tau D_1 + k(c_0, p)^{p-1} \\ &\leq \int_\Omega |\rho_0|^p\, dx + (1 + \chi n\tau) 2pk(c_0, p)^p,\end{aligned}$$

where in the last inequality we use the dependence of D_1 in terms of k and we suppose $k \geq 1$.

For the L^∞ estimate we cannot simply pass to the limit in the L^p inequality we just found since $k(c_0, p) \to \infty$ as $p \to \infty$. However, we can iterate the procedure, finding better estimates for $k(c_0, p)$ using the fact that we have some explicit bounds on $\|\rho_m\|_p$: let us fix n and let us consider $T = n\tau$. We will consider iteratively $p_i = 2^i + 1$. Always choosing $k(c_0, p) \geq \|\rho_0\|_\infty$ we find that

$$D(p_i) = \sup_{m \leq n} \|\rho_m\|_{p_i} \leq ((2 + \chi T) 2p_i)^{1/p_i} k(c_0, p_i); \tag{4.5}$$

For the iterative step, we have

$$\int_{\rho \geq k} \rho_\delta \leq \frac{1}{k^{p_i - 1}} \int_{\rho \geq k} \rho_\delta^{p_i}\, dx \leq \frac{D(p_i)^{p_i}}{k^{p_i - 1}};$$

in particular it is sufficient to choose

$$k(c_0, p_{i+1}) = \Big(\frac{D(p_i)^{p_i}}{\alpha(p_i)}\Big)^{\frac{1}{p_i-1}}.$$

Now we can use $\alpha(p_i) \geq C/p_i \geq C2^{-i}$ (where C also depends on χ) and (4.5) to obtain

$$k(c_0, p_{i+1}) \leq (C2^{1+2i}(2+\chi T))^{\frac{1}{2^i}} k(c_0, p_i)^{1+\frac{1}{2^i}}.$$

From here we can derive a uniform bound for $k(c_0, p_i)$: indeed, defining $c_{i+1} = \prod_{j=2}^{i}(1+\frac{1}{2^j})$ and writing $k_i := k(c_0, p_i)$, we have

$$\begin{aligned} k_{i+1}^{\frac{1}{c_{i+1}}} &\leq (C2^{1+2i}(2+\chi T))^{\frac{1}{2^i c_{i+1}}} k_i^{\frac{1}{c_i}} \\ &\leq k_2 \prod_{j=2}^{i+1} (C2^{1+2j}(2+\chi T))^{\frac{1}{2^j}} \\ &\leq Dk_2\sqrt{2+\chi T}. \end{aligned}$$

In particular, using (4.5) together with the last estimate and the fact that $c_i \leq 2$, we can say that we have

$$\|\rho_n\|_\infty \leq \lim_{i\to\infty} D(p_i) \leq \lim_{i\to\infty} k(c_0, p_i) \leq D(2+\chi T)\max\{k(c_0, 5), \|\rho_0\|_\infty\}^2. \qquad \blacksquare$$

The above estimate is the discrete counterpart of a well-known result studied in continuous time, which can be found for instance in [29], and is proven in [20] and [15] (more precisely, [20] showed that equi-integrability of ρ is enough to propagate in time the estimates on the L^p norms, and [15] found the sharp condition to bound the entropy of the solution, and hence provide equi-integrability).

We note that in the above L^p estimate we obtain linear growth in time of the L^p norm raised to the power p, i.e. on $\int \rho^p\, dx$, so that the norm itself has much slower growth. In this regard, the estimate on the L^∞ norm is most likely not sharp, as it is the norm itself which grows linearly.

Regarding the L^∞ estimates, we recall that other L^∞ bounds have been found on a (perturbed) JKO scheme in [13], but those bounds always explode in finite time (at time $T = 1/\|\rho_0\|_\infty$). On the other hand, they have the advantage that they are true for any form of diffusion, and that they require no condition on χ, nor on the dimension.

Remark 4.6. With similar but more tedious calculation it is also possible to get hypercontractivity estimates (improvement in time of the summability exponent) in the JKO setting: this could potentially weaken the integrability requirement on ρ_0 in Theorem 5.11, but then we would need a different analysis for the first steps, where the integrability assumption $\rho \in L^r$ is still not satisfied. For this reason we do not want to pursue this direction here, but it would certainly be interesting.

In a similar but different spirit we also mention that it is possible to obtain L^q estimates in time and space starting from L^p assumptions on ρ_0, as is done in [24, Lemma 2.11], but we do not investigate this question here since we decided to concentrate on bounds which are not integrated in time but derived from the decreasing behavior from one step of the JKO scheme to the next.

5. Sobolev estimates

In this section we pass to the core of the paper, i.e. the higher-order estimates. The goal will be to obtain results comparable to those of Proposition 3.8, but for norms involving the gradient of ρ.

Lemma 5.1. *Let us consider an absolutely continuous $\eta \in \mathcal{P}(\Omega)$ and a functional $\mathcal{G}$: $\mathcal{P}(\Omega) \to \mathbb{R} \cup \{\infty\}$. Take $\rho \in \mathrm{Prox}^{\tau}_{\mathcal{E}+\mathcal{G}}(\eta)$ and set $u[\rho] := \delta\mathcal{G}/\delta\rho$. Set*

$$Z_\rho := \frac{\nabla \rho}{\rho} + \nabla u[\rho], \quad Z_\eta := \frac{\nabla \eta}{\eta} + \nabla u[\eta]$$

and consider φ and ψ the optimal Kantorovich potentials for the dual formulation of $W_2(\rho, \eta)$, with $T = \mathrm{id} - \nabla\varphi$ the optimal transport map from ρ to η. Then for every radial convex function $H: \mathbb{R}^d \to \mathbb{R}$ we have

$$\int H(Z_\eta)\, d\eta \geq \int H(Z_\rho)\, d\rho + \int \nabla H\Big(\frac{\nabla\varphi}{\tau}\Big) \cdot (\nabla u[\rho] - \nabla u[\eta] \circ T)\, d\rho.$$

Proof. We first notice that the optimality condition on ρ gives $Z_\rho + \frac{\nabla\varphi}{\tau} = 0$. As the statement involves both Z_ρ and Z_η, it is important to underline that we do not have, instead, any relation between Z_η and $\frac{\nabla\psi}{\tau} = 0$; since φ and ψ are the Kantorovich potentials in the transport between ρ and η, we have $\rho \in \mathrm{Prox}^{\tau}_{\mathcal{E}+\mathcal{G}}(\eta)$ but we do not have $\eta \in \mathrm{Prox}^{\tau}_{\mathcal{E}+\mathcal{G}}(\rho)$. However, we can heuristically expect Z_η to look like $\frac{\nabla\psi}{\tau}$ (note the sign, which is different from that in the equality $Z_\rho = -\frac{\nabla\varphi}{\tau}$). Therefore, we first estimate $H(Z_\eta)$ from below in terms of $H(-\frac{\nabla\psi}{\tau})$, using the convexity of H and the fact that it is radial and hence even:

$$\begin{aligned}\int H(Z_\eta)\, d\eta &= \int H(-Z_\eta)\, d\eta \\ &\geq \int H\Big(\frac{-\nabla\psi}{\tau}\Big)\, d\eta + \int \nabla H\Big(\frac{-\nabla\psi}{\tau}\Big) \cdot \Big(-Z_\eta + \frac{\nabla\psi}{\tau}\Big)\, d\eta.\end{aligned}$$

We look at the different parts of the right-hand side. First we use $\eta = T_{\#}\rho$ and $-\nabla\psi \circ T = \nabla\varphi$, together with the optimality condition $Z_\rho + \frac{\nabla\varphi}{\tau} = 0$ and again the fact that H is even, in order to get

$$\int H\Big(\frac{-\nabla\psi}{\tau}\Big)\, d\eta = \int H\Big(\frac{\nabla\varphi}{\tau}\Big)\, d\rho = \int H(Z_\rho)\, d\rho.$$

Using $\eta = T_\# \rho$ and $-\nabla\psi \circ T = \nabla\varphi$ again, we obtain

$$\int \nabla H\Big(\frac{-\nabla\psi}{\tau}\Big)\cdot\frac{\nabla\psi}{\tau}\,d\eta = -\int \nabla H\Big(\frac{\nabla\varphi}{\tau}\Big)\cdot\frac{\nabla\varphi}{\tau}\,d\rho = \int \nabla H\Big(\frac{\nabla\varphi}{\tau}\Big)\cdot Z_\rho\,d\rho$$
$$= \int \nabla H\Big(\frac{\nabla\varphi}{\tau}\Big)\cdot\nabla\rho\,dx + \int \nabla H\Big(\frac{\nabla\varphi}{\tau}\Big)\cdot\nabla u[\rho]\,d\rho.$$

We now pass to the part involving Z_η, and write

$$\int \nabla H\Big(\frac{-\nabla\psi}{\tau}\Big)\cdot(-Z_\eta)\,d\eta = \int \nabla H\Big(\frac{\nabla\psi}{\tau}\Big)\cdot(Z_\eta)\,d\eta$$
$$= \int \nabla H\Big(\frac{\nabla\psi}{\tau}\Big)\cdot\nabla\eta\,dx + \int \nabla H\Big(\frac{\nabla\psi}{\tau}\Big)\cdot\nabla u[\eta]\,d\eta$$
$$= \int \nabla H\Big(\frac{\nabla\psi}{\tau}\Big)\cdot\nabla\eta\,dx - \int \nabla H\Big(\frac{\nabla\varphi}{\tau}\Big)\cdot\nabla u[\eta]\circ T\,d\rho.$$

Summing all the terms, and using the five-gradients inequality, Lemma 2.1 (which requires H to be radial in order to handle the boundary terms),

$$\int \nabla H\Big(\frac{\nabla\psi}{\tau}\Big)\cdot\nabla\eta\,dx + \int \nabla H\Big(\frac{\nabla\varphi}{\tau}\Big)\cdot\nabla\rho\,dx \geq 0,$$

we obtain the desired result. ∎

The quantities of the form $\int H(Z_\rho)\,d\rho$ will be crucial for the Sobolev regularity of the solutions of the JKO scheme. We will then often note $J_{(p)}(\rho) := \int H(Z_\rho)\,d\rho$ when $H(z) = |z|^p$, without explicit reference to the term $u[\rho]$, which will be clear from the context.

5.1. Fokker–Planck and aggregation

We will see some consequences of Lemma 5.1, starting from the easiest case, i.e. the purely linear Fokker–Planck case: $u[\rho] = V$ and $\mathcal{G}(\rho) := \int V\,d\rho$.

Proposition 5.2. *Let $\eta \in \mathcal{P}(\Omega)$ and let us consider $\mathcal{G}(\rho) := \int V\,d\rho$ and $\rho \in \mathrm{Prox}^\tau_{\mathcal{E}+\mathcal{G}}(\eta)$. Suppose that V is λ-convex, i.e. $D^2V \geq \lambda I$.*

Then, if $\lambda = 0$ and $H : \mathbb{R}^d \to \mathbb{R}$ is a radial convex function, we have

$$\int H(Z_\rho)\,d\rho \leq \int H(Z_\eta)\,d\eta.$$

If $\lambda > 0$ and $H(0) = 0$, we also have

$$(1+\lambda\tau)\int H(Z_\rho)\,d\rho \leq \int H(Z_\eta)\,d\eta.$$

For $\lambda < 0$, if H satisfies $\nabla H(z)\cdot z \leq C(H(z)+1)$ then

$$(1-|\lambda|C\tau)\int H(Z_\rho)\,d\rho \leq \int H(Z_\eta)\,d\eta + |\lambda|C\tau.$$

Proof. All these results are just consequences of Lemma 5.1. They can be obtained if one estimates the term $\int \nabla H(\frac{\nabla\varphi}{\tau}) \cdot (\nabla u[\rho] - \nabla u[\eta] \circ T)\, d\rho$. First, note that since H is radial, the vectors $\nabla H(\nabla\varphi/\tau)$ and $\nabla\varphi$ are parallel and oriented in the same direction. We also use $u[\rho] = u[\eta] = V$ and the assumptions on V. Indeed, we have

$$\big(\nabla V(x) - \nabla V(T(x))\big) \cdot \nabla\varphi(x) = \big(\nabla V(x) - \nabla V(x - \nabla\varphi(x))\big) \cdot \nabla\varphi(x) \geq \lambda |\nabla\varphi(x)|^2,$$

thanks to the λ-convexity of V. In the case $\lambda = 0$ this is enough to obtain the claim.

For $\lambda > 0$, we write

$$\begin{aligned}\int \nabla H\Big(\frac{\nabla\varphi}{\tau}\Big) \cdot (\nabla V - \nabla V \circ T)\, d\rho &= \int \frac{|\nabla H(\frac{\nabla\varphi}{\tau})|}{|\nabla\varphi|} (\nabla V - \nabla V \circ T) \cdot \nabla\varphi\, d\rho \\ &\geq \lambda\tau \int H\Big(\frac{\nabla\varphi}{\tau}\Big)\, d\rho,\end{aligned}$$

where we used the inequality $|\nabla H(z)| \geq H(z)/|z|$ which is valid for radial convex functions with $H(0) = 0$ and the same estimate due to the λ-convexity of V as above. This allows us to prove the second part of the claim.

In the case $\lambda < 0$, the estimate is similar, but since we bound the scalar product $(\nabla V(x) - \nabla V(T(x))) \cdot \nabla\varphi(x)$ from below with $\lambda|\nabla\varphi(x)|^2$, which is negative, we need to estimate $|\nabla H(z)|$ from above, and for this we use our assumption on H (which is essentially an assumption of polynomial growth for H; note that, H being radial, we have $\nabla H(z) \cdot z = |\nabla H(z)|\,|z|$). ∎

Remark 5.3. As we underlined in the introduction, the above result is a time-discrete translation of a suitable integral version of a well-known estimate in Bakry–Émery theory (again, we refer for instance to [3]). Indeed, the time-continuous equation satisfied by ρ when taking the gradient flow of $\mathcal{E} + \mathcal{G}$ is $\partial_t\rho - \Delta\rho - \nabla \cdot (\rho\nabla V) = 0$. If one defines $u = \rho e^V$ then u satisfies the drift–diffusion PDE $\partial_t u = \Delta u - \nabla V \cdot \nabla u$. If we call P_t the semigroup associated with this PDE, the celebrated Bakry–Émery estimates provide $|\nabla(P_t f)| \leq e^{-\lambda t} P_t(|\nabla f|)$ when $D^2 V \geq \lambda I$. Taking H convex and radially increasing (and writing, by abuse of notation, $H(z) = H(|z|)$), and using this inequality, for instance for $\lambda = 0$, together with the convexity of the function $(s, y) \mapsto H(y/s)s$, one can prove

$$\int H\Big(\frac{\nabla(P_t f)}{P_t f}\Big) P_t f\, de^{-V} \leq \int H\Big(\frac{P_t(|\nabla f|)}{P_t f}\Big) P_t f\, de^{-V} \leq \int H\Big(\frac{|\nabla f|}{f}\Big) f\, de^{-V},$$

which can be seen to be equivalent to the result of Proposition 5.2 since

$$\frac{\nabla(P_t f)}{P_t f} = \frac{\nabla u}{u} = \frac{(\nabla\rho + \rho\nabla V)e^V}{\rho e^V} = Z_\rho.$$

As a consequence, we obtain the following information on the JKO scheme:

Proposition 5.4. *Suppose that V is λ-convex and Lipschitz, and consider $\mathcal{G}(\rho) = \int_\Omega V\, d\rho$; let $\rho_0 \in \mathcal{P}(\Omega)$ and let ρ_n^τ be provided by JKO scheme as in (2.2). Fix a time $T > 0$ and only consider (n, τ) such that $\tau n \leq T$. Then we have the following bounds:*

- *if* $\lambda \geq 0$*, if* $\log \rho_0$ *is Lipschitz continuous, then* $\log \rho_n^\tau$ *is also Lipschitz continuous, with bounded Lipschitz constant, and* $\mathrm{Lip}(\log \rho_n^\tau + V)$ *decreases in time (in* n*);*
- *if* $\rho_0^{1/p} \in W^{1,p}(\Omega)$*, then* $(\rho_n^\tau)^{1/p}$ *is bounded in* $W^{1,p}(\Omega)$ *independently of* τ *and* n*;*
- *if* $\rho_0 \in \mathrm{BV}(\Omega)$*, then* ρ_n^τ *is bounded in* $\mathrm{BV}(\Omega)$ *independently of* τ *and* n*;*
- *if* $\rho_0 \in W^{1,1}(\Omega)$*, then all the densities* ρ_n^τ *belong to a weakly compact subset of* $W^{1,1}(\Omega)$.

The above bounds but the first are also valid for $\lambda < 0$*, but if* $\lambda \geq 0$ *then they are independent of* T*. Moreover, if* $\lambda > 0$ *then the gradients of the functions* $(\rho_n^\tau e^V)^{1/p}$ *converge exponentially fast to* 0 *in* $L^p(e^{-V})$*, uniformly with respect to* τ*, and hence the functions* $(\rho_n^\tau e^V)^{1/p}$ *converge in* $W^{1,p}(e^{-V})$ *to a constant.*

Proof. In the case $\lambda \geq 0$, for the first part of the statement, take a measure η and $\rho \in \mathrm{Prox}^\tau_{\mathcal{E}+\mathcal{G}}(\eta)$. Let us suppose $\mathrm{Lip}(\log \eta + V) \leq L$ and use as a function H the convex indicator function of $\overline{B(0,L)}$. From $\int H(Z_\eta)\,d\eta = 0$ we deduce that we also have $\int H(Z_\rho)\,d\rho = 0$. This means $|\nabla(\log \rho + V)| \leq L$ a.e. on $\{\rho > 0\}$. Yet we know from the optimality conditions that ρ is a continuous density which is bounded away from 0 since $\log \rho = C - V - \varphi/\tau$, hence we get $\mathrm{Lip}(\log \rho + V) \leq L$. This can be iterated along the JKO scheme thus obtaining the first part of the statement.

For the second part of the statement, we use $H(z) = |z|^p$ and

$$\|\rho^{1/p}\|^p_{W^{1,p}} = c(p)\int (\rho^{1/p-1}|\nabla\rho|)^p + \int \rho \leq c(p)\int H(Z_\rho)\,d\rho + C,$$

where we used the boundedness of ∇V. Since $\int H(Z_{\rho_n^\tau})\,d\rho_n^\tau$ decreases with n (if $\lambda \geq 0$) or at least its growth is exponentially controlled (if $\lambda < 0$), then $(\rho_n^\tau)^{1/p}$ is bounded in $W^{1,p}$.

The third part of the statement is proven in a similar way, using $p = 1$. Indeed, given $\eta \in \mathrm{BV}(\Omega)$ and $\rho \in \mathrm{Prox}^\tau_{\mathcal{E}+\mathcal{G}}(\eta)$, we can approximate η with smoother densities η_j with $\|\nabla\eta_j + \eta_j\nabla V\| \to \|\nabla\eta + \eta\nabla V\|$ (the norm being taken in the space of vector measures). For each j we have a measure $\rho_j \in \mathrm{Prox}^\tau_{\mathcal{E}+\mathcal{G}}(\eta_j)$, which is Lipschitz continuous and satisfies

$$\|\nabla\rho_j + \rho_j\nabla V\| = \int \left|\frac{\nabla\rho_j}{\rho_j} + \nabla V\right| d\rho_j \leq \|\nabla\eta_j + \eta_j\nabla V\|. \tag{5.1}$$

It is moreover clear (from a triangular inequality on the Wasserstein term, which is bounded because Ω is bounded) that the functionals $\mathcal{F}_{\eta_j} = \mathcal{E}(\cdot) + \mathcal{G}(\cdot) + \frac{1}{2\tau}W_2^2(\cdot,\eta_j)$ are Γ-converging, with respect to the Wasserstein distance, to $\mathcal{F}_\eta$: we then have $\rho_j \rightharpoonup \rho$. In particular, passing to the limit in j in the inequality (5.1), we get

$$\|\nabla\rho + \rho\nabla V\| \leq \|\nabla\eta + \eta\nabla V\|.$$

This proves that $\|\nabla\rho_n^\tau + \rho_n^\tau\nabla V\|$ is decreasing, and hence $\|\nabla\rho_n^\tau\|$ stays bounded.

Regarding the $W^{1,1}$ estimate, we use a convex and superlinear function H such that $\int H(Z_{\rho_0})\,d\rho_0 < \infty$ (which exists since $\nabla\rho_0 \in L^1$ implies $\nabla\rho_0/\rho_0 + \nabla V \in L^1(\rho_0)$ and we know that L^1 functions are also integrable when composed with a suitable superlinear function, which can be taken convex). The results of Proposition 5.7 allow us then to keep the same integrability of the gradient along the iterations of the JKO scheme: this is easy if $\lambda \geq 0$, while for $\lambda < 0$ we just need to note that H can be taken superlinear but with polynomial growth (actually, its growth can be taken as close to linear as we want), and hence we can apply the last claim in Proposition 5.7. This guarantees equi-integrability for $\nabla \rho_n^\tau$ and hence the claim.

We are now left to consider the behavior for $n \to \infty$ in the case $\lambda > 0$. In this case we have exponential decay of the quantity $J_{(p)}(\rho_n^\tau) := \int H(Z_{\rho_n^\tau})\, d\rho_n^\tau$ for $H(z) = |z|^p$. We can then observe that we have

$$\int \left| \frac{\nabla \rho}{\rho} + \nabla V \right|^p d\rho = \int |\nabla \log(\rho e^V)|^p e^{-V} d(\rho e^V)$$
$$= c \int \left|\nabla((\rho e^V)^{1/p})\right|^p d(e^{-V}).$$

This last result provides a sort of rate of convergence to the steady state of the Fokker–Planck equation $\rho = e^{-V}$. For the $W^{1,p}$ convergence we only have to use an appropriate local Sobolev inequality and exploit uniform integrability. ■

Remark 5.5. The bound on the Lipschitz constant could have been obtained as a limit on the L^p norms for $p \to \infty$. Indeed, for $\lambda \geq 0$ we can also easily obtain

$$\int |Z_\rho|^p\, d\rho \leq \int |Z_\eta|^p d\eta$$

which, raising to the power $1/p$ and sending $p \to \infty$ also gives a bound on the L^∞ norm of Z_ρ, and hence on the Lipschitz constant of $\log \rho + V$. On the other hand, the approach with $H(z) = |z|^p$ is interesting for $\lambda < 0$, as it provides

$$\int |Z_\rho|^p\, d\rho \leq (1 - |\lambda| p \tau)^{-1} \int |Z_\eta|^p d\eta.$$

This estimate provides exponential bounds on $\int |Z_{\rho_t}|^p\, d\rho_t$, i.e.

$$\int |Z_{\rho_t}|^p\, d\rho_t \leq e^{|\lambda| p t} \int |Z_{\rho_0}|^p\, d\rho_0.$$

By taking the power $1/p$ and the limit as $p \to \infty$, one gets $\|Z_{\rho_t}\|_\infty \leq e^{|\lambda| t} \|Z_{\rho_0}\|_\infty$. Yet, this last computation can only be performed in continuous time. More precisely, we first need to send $\tau \to 0$ and then $p \to \infty$. Indeed, if we first send $p \to \infty$ while $\tau > 0$ is fixed, we would get $1 - |\lambda| p < 0$ which prevents any interesting estimate being obtained.

In the next remark we use the following notation: when a vector z and an exponent $\alpha > 0$ are given, by z^α we mean $|z|^{\alpha-1} z$ (if $z \neq 0$, and 0 if $z = 0$), i.e. a vector whose norm is $|v|^\alpha$ and the direction is the same as that of v.

Remark 5.6. The estimate with $H(z) = |z|^p$ when V is λ-convex with $\lambda < 0$ can also be concluded in a different way when $p < 2$. Indeed, we can use

$$\begin{aligned}\int_\Omega (\nabla V - \nabla V \circ T) \cdot \Big(\frac{x - T(x)}{\tau}\Big)^{p-1} d\rho &\geq \frac{\lambda}{\tau^{p-1}} \int_\Omega |x - T(x)|^p \, d\rho \\ &= -\frac{|\lambda|}{\tau^{p-1}} \|\mathrm{id} - T\|^p_{L^p(\rho)}.\end{aligned}$$

We then use $p < 2$ to bound the L^p norm with the L^2 norm and write

$$\|\mathrm{id} - T\|_{L^p(\rho)} \leq \|\mathrm{id} - T\|_{L^2(\rho)} = W_2(\rho, \eta),$$

thus obtaining

$$\int_\Omega (\nabla V - \nabla V \circ T) \cdot \Big(\frac{x - T(x)}{\tau}\Big)^{p-1} d\rho \geq \frac{\lambda}{\tau^{p-1}} W_2(\rho, \eta)^p.$$

In particular, if ρ_n is the sequence generated by the JKO scheme we have, by induction,

$$J_{(p)}(\rho_{n+1}) \leq J_{(p)}(\rho_k) + p|\lambda| \sum_{i=k}^n \tau \Big(\frac{W_2(\rho_i, \rho_{i+1})}{\tau}\Big)^p.$$

Now we can use the Hölder inequality (again based on $p < 2$) and, together with $W_2^2(\rho_n, \rho_{n+1}) \leq 2\tau(\mathcal{F}(\rho_n) - \mathcal{F}(\rho_{n+1}))$, we obtain

$$\begin{aligned}J_{(p)}(\rho_{n+1}) &\leq J_{(p)}(\rho_k) + p|\lambda|((n-k)\tau)^{1-p/2} \Big(\sum_{i=k}^n \tau \Big(\frac{W_2(\rho_i, \rho_{i+1})}{\tau}\Big)^2\Big)^{p/2} \\ &\leq J_{(p)}(\rho_k) + p|\lambda|(t-s)^{1-p/2}(2\mathcal{F}(\rho_{n+1}) - 2\mathcal{F}(\rho_k))^{p/2}.\end{aligned}$$

Hence, we deduce that $J_{(p)}(\rho_t)$ is locally bounded in time, and grows sublinearly as $t \to \infty$.

We also want to consider the case where V depends on ρ via a smooth convolution kernel. This is typical in aggregation equations. We consider the case

$$\mathcal{G}(\rho) := \int V \, d\rho + \frac{1}{2} \int \int W(x-y) \, d\rho(x) \, d\rho(y) \tag{5.2}$$

for an interaction potential $W: \mathbb{R}^d \to \mathbb{R}$ which is supposed to be even ($W(z) = W(-z)$) and $C^{1,1}$. This last assumption is very demanding and non-optimal, but allows for a simple presentation of the estimates. The Keller–Segel case that we will see later is in some sense obtained from a singular interaction potential ($W(z) = \log|z|$ in dimension $d = 2$, when in the whole space), and will be treated in detail, but in a different way. We will set

$$J_{(p)}(\rho) := \int \Big|\frac{\nabla\rho}{\rho} + \nabla V + \nabla W * \rho\Big|^p \, d\rho,$$

i.e. $J_{(p)}(\rho) = \int H(Z_\rho) \, d\rho$, where $u[\rho] = V + W * \rho$. For simplicity, in the notation $J_{(p)}$, we are omitting the dependence on V and W.

Proposition 5.7. *Let us consider $\mathcal{G}$ as in (5.2), and $\rho \in \mathrm{Prox}^{\tau}_{\mathcal{E}+\mathcal{G}}(\eta)$. Suppose that Ω is convex, that V is λ-convex, i.e. $D^2V \geq \lambda I$, and that W is $C^{1,1}$, with $\mathrm{Lip}(\nabla W) = \mu$. Then we have*

$$(1 + p(\lambda - 2\mu)\tau) J_{(p)}(\rho) \leq J_{(p)}(\eta).$$

Proof. The starting point is, of course, the result of Lemma 5.1 applied to $H(z) = |z|^p$. This gives

$$\begin{aligned} J_{(p)}(\eta) \leq J_{(p)}(\rho) & \\ + p \int \Big(\frac{x - T(x)}{\tau}\Big)^{p-1} & \cdot \big(\nabla V(x) - \nabla V(T(x)) + (\nabla W * \rho)(x) \\ & - (\nabla W * \eta)(T(x))\big)\, d\rho(x), \end{aligned}$$

where by v^{p-1}, when v is a vector (here $v = (x - T(x))/\tau$), we mean $|v|^{p-2}v$.

Using the same argument as in Remark 5.6 we can obtain

$$\begin{aligned} \int \Big(\frac{x - T(x)}{\tau}\Big)^{p-1} \cdot \big(\nabla V(x) - \nabla V(T(x))\big)\, d\rho(x) &\geq \tau\lambda \int \Big|\frac{x - T(x)}{\tau}\Big|^p d\rho(x) \\ &= \tau\lambda J_{(p)}(\rho), \end{aligned}$$

as well as

$$\int \Big(\frac{x - T(x)}{\tau}\Big)^{p-1} \cdot \big((\nabla W * \eta)(x) - (\nabla W * \eta)(T(x))\big)\, d\rho(x) \geq -\tau\mu J_{(p)}(\rho),$$

since the function $W * \eta$ is $(-\mu)$-convex.

We are left to estimate the remaining term

$$\int \Big(\frac{x - T(x)}{\tau}\Big)^{p-1} \cdot \big((\nabla W * \eta)(x) - (\nabla W * \rho)(x)\big)\, d\rho(x).$$

This term will be bounded in absolute value, and we first note that we have

$$|(\nabla W * \rho)(x) - (\nabla W * \eta)(x)| = \Big|\int \nabla W(x - y)\, d(\rho - \eta)(y)\Big| \leq \mu W_1(\rho, \eta),$$

since $y \mapsto \nabla W(x - y)$ is μ-Lipschitz for every x. We also use

$$\begin{aligned} W_1(\rho, \eta) \leq \int |x - T(x)|\, d\rho = \tau \int \frac{|x - T(x)|}{\tau}\, d\rho &\leq \tau \Big(\int \Big|\frac{x - T(x)}{\tau}\Big|^p d\rho\Big)^{1/p} \\ &= \tau J_{(p)}(\rho)^{1/p}. \end{aligned}$$

Then we have

$$\begin{aligned} &\Big|\int \Big(\frac{x - T(x)}{\tau}\Big)^{p-1} \cdot \big((\nabla W * (\eta - \rho))(x)\big)\, d\rho(x)\Big| \\ &\quad \leq \mu\tau J_{(p)}(\rho)^{1/p} \int \Big|\frac{x - T(x)}{\tau}\Big|^{p-1} d\rho(x) \\ &\quad \leq \mu\tau J_{(p)}(\rho)^{1/p} J_{(p)}(\rho)^{(p-1)/p} = \mu\tau J_{(p)}(\rho). \end{aligned}$$

Putting all the results together provides the claim. ∎

Let us note that in the above result, in order to obtain an estimate which could be iterated, we needed to replace $\nabla W * \eta$ with $\nabla W * \rho$, and hence we used the Lipschitz behavior of ∇W: for this estimate, lower bounds on $D^2 W$ (as we required on V) were not enough. We also note that the same kind of exponential asymptotic behavior providing convergence to the steady state as in the last point of Proposition 5.4 could be obtained, provided $\lambda > 2\mu$, but these computations will not be detailed.

5.2. Keller–Segel case

We come back to the case $\mathcal{G}(\rho) = -\frac{\chi}{2}\int h[\rho]\,d\rho$, where $-\Delta h[\rho] = \rho$ in Ω with Dirichlet boundary conditions. In this case we have that the first variation of $\mathcal{G}$ is $u[\rho] = -\chi h[\rho]$. The main goal of this section will be to have an estimate of

$$J_{(p)}(\rho) = \int_\Omega \left|\frac{\nabla\rho}{\rho} - \chi\nabla h[\rho]\right|^p d\rho. \tag{5.3}$$

As we will see, in order to deal with some of the error terms, we will need to estimate $J_{(p)}$ with $W_2(\rho,\eta)$, and use $p < 2$. Moreover, we will also need an a priori bound on the L^r norm of ρ and η, for an exponent r depending on p. All these restrictions are mainly due to the fact that $-h[\rho]$ does not necessarily satisfy the semiconvexity assumptions that we usually used to handle remainder terms.

Thanks to Lemma 5.1 we have

$$\begin{aligned} J_{(p)}(\rho) &\le J_{(p)}(\eta) + p\chi\int_\Omega (\nabla h[\rho] - \nabla h[\eta]\circ T)\cdot\Big(\frac{x-T(x)}{\tau}\Big)^{p-1} d\rho \\ &= J_{(p)}(\eta) + p\chi\int_\Omega (\nabla h[\rho] - \nabla h[\eta])\cdot\Big(\frac{x-T(x)}{\tau}\Big)^{p-1} d\rho \\ &\quad - p\chi\int_\Omega (\nabla h[\eta]\circ T - \nabla h[\eta])\cdot\Big(\frac{x-T(x)}{\tau}\Big)^{p-1} d\rho. \end{aligned} \tag{5.4}$$

In order to treat the two remainder terms, we state a general comparison result between Sobolev dual norms and Wasserstein distances ([30, Exercise 38]).

Proposition 5.8. *Let $\rho,\eta \in \mathcal{P}(\Omega)$ be two absolutely continuous measures. Then, supposing that $\|\rho\|_r, \|\eta\|_r \le C$ with $\frac{1}{q} + \frac{1}{r} + \frac{1}{p} = 1 + \frac{1}{rq}$, for every $\varphi \in C^1(\Omega)$ we have*

$$\int \varphi\, d(\rho-\eta) \le \|\nabla\varphi\|_p \cdot C^{1/q'} \cdot W_q(\eta,\rho),$$

where $q' = q/(q-1)$ is the dual exponent of q. In particular, we have $\|\rho-\eta\|_{(W^{1,p})^*(\Omega)} \le \sqrt{\max\{\|\rho\|_r, \|\eta\|_r\}}\, W_2(\eta,\rho)$ for $r = p/(p-2)$.

The following lemma is very classical, and can be found, for instance, in [17, Theorem 10.15] (where the case $p \ge 2$ is treated; for $p < 2$ one can then argue by duality).

Lemma 5.9. *Let Ω be a bounded convex domain, and $f \in (W_0^{1,p})^*(\Omega)$ be given. Denoting by $h[f]$ the unique solution in $W_0^{1,p'}(\Omega)$ (with $p' = p/(p-1)$ the dual exponent* *to p) of $-\Delta h = f$, there exists a constant $C > 0$, depending only on the dimension, on p, and possibly on Ω, such that*

$$\|\nabla h[f]\|_{p'} \le C\|f\|_{(W^{1,p})^*(\Omega)}.$$

Lemma 5.10. *Given $p \in (1,2)$, set $r = \frac{4-p}{2-p}$. Let us assume $\rho, \eta \in \mathcal{P}(\Omega) \cap L^r(\Omega)$ with Ω convex and let us denote by T the optimal map between ρ and η. Then there exists a constant C, only depending on Ω, p, and d, such that*

$$\int_\Omega (\nabla h[\eta] - \nabla h[\eta]\circ T)\cdot\Big(\frac{x-T(x)}{\tau}\Big)^{p-1} d\rho \le C\tau \max\{\|\rho\|_r^r, \|\eta\|_r^r\} + \frac{W_2^2(\eta,\rho)}{\tau},$$
$$\int_\Omega (\nabla h[\rho] - \nabla h[\eta])\cdot\Big(\frac{x-T(x)}{\tau}\Big)^{p-1} d\rho \le C\tau \max\{\|\rho\|_r^r, \|\eta\|_r^r\} + \frac{W_2^2(\eta,\rho)}{\tau}.$$

Proof. We begin with the first inequality: if we set $T_t(x) = x + t(T(x)-x)$ then we know that $(T_t)_\sharp\rho := \rho_t \in \mathcal{P}(\Omega)$ is the Wasserstein geodesic from ρ to η, and the displacement convexity properties of the L^r norms imply that we have $\|\rho_t\|_r \le \max\{\|\eta\|_r, \|\rho\|_r\} =: M$. We have

$$\begin{aligned}
&\int_\Omega (\nabla h[\eta] - \nabla h[\eta]\circ T)\cdot\Big(\frac{x-T(x)}{\tau}\Big)^{p-1} d\rho \\
&\quad= \int_\Omega\int_0^1 (x-T(x))\cdot\big(D^2h[\eta](T_t(x))\big)\cdot\Big(\frac{x-T(x)}{\tau}\Big)^{p-1} dt\, d\rho \\
&\quad\le \int_{[0,1]\times\Omega} \tau^{1-p/2}|D^2h[\eta](T_t(x))|\cdot\frac{|x-T(x)|^p}{\tau^{p/2}}\, dt\, d\rho \\
&\quad\le \tau^{\frac{1}{q}}\Big(\int_0^1\int_\Omega |D^2h[\eta](T_t(x))|^q\, d\rho\, dt\Big)^{\frac{1}{q}}\cdot\Big(\int_\Omega \frac{|x-T(x)|^2}{\tau}\, d\rho\Big)^{\frac{p}{2}} \\
&\quad= \tau^{\frac{1}{q}}\Big(\int_0^1\int_\Omega |D^2h[\eta](x)|^q\rho_t(x)\, dx\, dt\Big)^{\frac{1}{q}}\cdot\Big(\frac{W_2(\eta,\rho)^2}{\tau}\Big)^{\frac{p}{2}},
\end{aligned}$$

where q denotes here the dual exponent of $2/p$, i.e. $q = \frac{2}{2-p} = r-1$. Using the Hölder inequality with exponents $\frac{q+1}{q}$ and $q+1$ and then the classical estimate $\|D^2h[\rho]\|_r \le C\|\Delta h[\rho]\|_r = C\|\rho\|_r$, we obtain

$$\int |D^2h[\rho]|^q\rho_t \le \|D^2h[\rho]\|_{q+1}^q\|\rho_t\|_{q+1} \le CM^{q+1}.$$

Using this estimate and eventually Young's inequality with exponents q and $2/p$ we obtain

$$\begin{aligned}
\Big(\int_0^1\int_\Omega \tau|D^2h[\rho](x)|^q\rho_t(x)\, dx\, dt\Big)^{\frac{1}{q}}\cdot\Big(\frac{W_2(\eta,\rho)^2}{\tau}\Big)^{\frac{p}{2}} &\le \big(\tau^{\frac{1}{q}}CM^{\frac{q+1}{q}}\big)\cdot\Big(\frac{W_2(\eta,\rho)^2}{\tau}\Big)^{\frac{p}{2}} \\
&\le \tau CM^r + \frac{W_2(\eta,\rho)^2}{\tau}.
\end{aligned}$$

Now we can pass to the second inequality. We perform directly the Hölder inequality with exponents $\frac{2}{3-p}$ and $\frac{2}{p-1}$, and then the Hölder inequality with exponents $\frac{4-p}{2}$ and $\frac{4-p}{2-p}$:

$$\begin{aligned}
&\int_\Omega (\nabla h[\rho] - \nabla h[\eta]) \cdot \Big(\frac{x - T(x)}{\tau}\Big)^{p-1} \, d\rho \\
&\quad \le \frac{1}{\tau^{p-1}} \Big(\int_\Omega |\nabla h[\rho - \eta]|^{\frac{2}{3-p}} \, d\rho \Big)^{\frac{3-p}{2}} \cdot \Big(\int_\Omega |T(x) - x|^2 \, d\rho \Big)^{\frac{p-1}{2}} \\
&\quad \le \frac{1}{\tau^{p-1}} \|\nabla h[\eta - \rho]\|_{\frac{4-p}{3-p}} \|\rho\|_r^{\frac{3-p}{2}} W_2(\rho, \eta)^{p-1}.
\end{aligned}$$

We then use Proposition 5.8 and Lemma 5.9 in order to write

$$\|\nabla h[\eta - \rho]\|_{\frac{4-p}{3-p}} \le C \|\eta - \rho\|_{(W^{1,4-p}(\Omega))^*} \le C\sqrt{M} W_2(\eta, \rho).$$

Using this and then Young's inequality, we obtain

$$\begin{aligned}
\frac{1}{\tau^{p-1}} \|\nabla h[\eta - \rho]\|_{\frac{4-p}{3-p}} \|\rho\|_r^{\frac{3-p}{2}} W_2(\rho, \eta)^{p-1} &\le \frac{C}{\tau^{p-1}} M^{\frac{4-p}{2}} W_2(\rho, \eta)^p \\
&\le \tau C M^r + \frac{W_2(\rho, \eta)^2}{\tau}.
\end{aligned}$$

■

We can then collect all the previous results to obtain the following estimate.

Theorem 5.11. *Let $\mathcal{G}$ be defined as in (2.3) and let $\rho_0 \in \mathcal{P}(\Omega)$; let us consider $(\rho_n^\tau)_n$, a sequence obtained from the JKO scheme (2.2) for the Keller–Segel functional $\mathcal{F} = \mathcal{E} + \mathcal{G}$, and let $J_{(p)}$ be defined as in (5.3), for $p < 2$. Then we have*

$$J_{(p)}(\rho_{n+1}^\tau) + \mathcal{F}(\rho_{n+1}^\tau) \le J_{(p)}(\rho_n^\tau) + \mathcal{F}(\rho_n^\tau) + C\tau \max\{\|\rho_n^\tau\|_r^r, \|\rho_{n+1}^\tau\|_r^r\}. \tag{5.5}$$

Hence, if $J_{(p)}(\rho_0) < \infty$, if $\mathcal{F}$ is bounded from below, and if $\|\rho_n^\tau\|_r^r$ stays bounded along iterations, then $(\rho_n^\tau)^{1/p}$ is bounded in $W^{1,p}$. In particular, for every $\alpha > 0$, in the two-dimensional Keller–Segel model with $\chi < 8\pi$, when $(\rho_0)^{1/p} \in L^\infty \cap W^{1,p}$ there exists a solution ρ_t which satisfies a bound of the form $J_{(p)}(\rho_t) \le C + Ct^{1+\alpha}$ (the constant C depending on α, on p, on $|\Omega|$, and on the initial datum). If we do not suppose $(\rho_0)^{1/p} \in W^{1,p}$, then the same bound will be true for $t \ge t_0 > 0$, with a constant depending on t_0.

Proof. The iterated estimate (5.5) is just a consequence of (5.4) and of Lemma 5.10, together with

$$\frac{W_2^2(\rho_n^\tau, \rho_{n+1}^\tau)}{\tau} \le \mathcal{F}(\rho_n^\tau) - \mathcal{F}(\rho_{n+1}^\tau).$$

If $\mathcal{F}$ is bounded from below and the L^r norm from above, we obtain a bound on $J_{(p)}(\rho_n^\tau)$ and then we use

$$\|(\rho)^{1/p}\|_{W^{1,p}} \le C + CJ_{(p)}(\rho) + C \int |\nabla h[\rho]|^p \, d\rho.$$

Using $\rho \in L^r$, we just need to bound $\nabla h[\rho]$ in $L^{pr'}$ in order to obtain the desired Sobolev bound. From elliptic regularity, we have $h[\rho] \in W^{2,r}$ and hence $\nabla h[\rho] \in L^{r^*}$, with $r^* = dr/(d-r)$, if $r < d$ (the case $r \geq d$ is easy, since in this case $\nabla h[\rho]$ belongs to all Lebesgue spaces). We just need to check $r^* \geq pr'$, which is true using $r \geq 3$ and $p < 2$.

In order to apply this estimate to the two-dimensional Keller–Segel model with $\chi < 8\pi$, we first note that in this case $\mathcal{F}$ is bounded from below (this would also be true for $\chi = 8\pi$). We just need to bound the L^r norm of the solution, and for this we use the estimates of Theorem 4.5. We obtain $\|\rho_t\|_q \leq C(1+t)^{1/q}$ for arbitrary $q > 1$, since $\rho_0 \in L^\infty$. Taking $q = r/\alpha$ one gets $\|\rho_t\|_r^r \leq C|\Omega|^{1-\alpha}(1+t)^\alpha$ and the conclusion follows iterating (5.5).

If we do not suppose $J_{(p)}(\rho_0) < \infty$, then we can use the dissipation of the energy $\mathcal{F}$ itself. Indeed, this dissipation provides $\mathcal{F}(\rho_0) = \int_0^{t_0} J_{(2)}(\rho_t)\,dt + \mathcal{F}(\rho_{t_0})$, which means that we can assume $J_{(2)}(\rho_{t_1}) \leq (\mathcal{F}(\rho_0) + C)/t_0$ for some $t_1 \in (0, t_0)$. This provides finiteness of $J_{(p)}(\rho_{t_1})$ for every $p < 2$. We then start a JKO scheme from ρ_{t_1}. ■

Funding. The work started when both the authors were at Laboratoire de Mathématiques d'Orsay, the first author being supported as a post-doc researcher by the ANR project ISOTACE (ANR-12-MONU-0013). This work is also supported by the Agence nationale de la recherche via the project ANR-16-CE40-0014 - MAGA - Monge-Ampère et Géométrie Algorithmique, and by INdAM, since the first author is member of GNAMPA.

References

[1] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*. Oxford Math. Monogr., Clarendon Press, Oxford University Press, New York, 2000 Zbl 0957.49001 MR 1857292

[2] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*. Lect. Math. ETH Zürich, Birkhäuser, Basel, 2005 Zbl 1090.35002 MR 2129498

[3] D. Bakry, I. Gentil, and M. Ledoux, *Analysis and geometry of Markov diffusion operators*. Grundlehren Math. Wiss. 348, Springer, Cham, 2014 Zbl 1376.60002 MR 3155209

[4] A. Blanchet, V. Calvez, and J. A. Carrillo, Convergence of the mass-transport steepest descent scheme for the subcritical Patlak-Keller-Segel model. *SIAM J. Numer. Anal.* **46** (2008), no. 2, 691–721 Zbl 1205.65332 MR 2383208

[5] A. Blanchet, J. A. Carrillo, D. Kinderlehrer, M. Kowalczyk, P. Laurençot, and S. Lisini, A hybrid variational principle for the Keller-Segel system in $\mathbb{R}^2$. *ESAIM Math. Model. Numer. Anal.* **49** (2015), no. 6, 1553–1576 Zbl 1334.35086 MR 3423264

[6] A. Blanchet, J. A. Carrillo, and N. Masmoudi, Infinite time aggregation for the critical Patlak-Keller-Segel model in $\mathbb{R}^2$. *Comm. Pure Appl. Math.* **61** (2008), no. 10, 1449–1481 Zbl 1155.35100 MR 2436186

[7] A. Blanchet, P. Mossay, and F. Santambrogio, Existence and uniqueness of equilibrium for a spatial model of social interactions. *Internat. Econom. Rev.* **57** (2016), no. 1, 31–59 Zbl 1404.91220 MR 3464513

[8] Y. Brenier, Décomposition polaire et réarrangement monotone des champs de vecteurs. *C. R. Acad. Sci. Paris Sér. I Math.* **305** (1987), no. 19, 805–808 Zbl 0652.26017 MR 923203

[9] Y. Brenier, Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.* **44** (1991), no. 4, 375–417 Zbl 0738.46011 MR 1100809

[10] V. Calvez and J. A. Carrillo, Volume effects in the Keller-Segel model: energy estimates preventing blow-up. *J. Math. Pures Appl. (9)* **86** (2006), no. 2, 155–175 Zbl 1116.35057 MR 2247456

[11] V. Calvez and L. Corrias, The parabolic-parabolic Keller-Segel model in $\mathbb{R}^2$. *Commun. Math. Sci.* **6** (2008), no. 2, 417–447 Zbl 1149.35360 MR 2433703

[12] J. A. Carrillo, S. Lisini, and E. Mainini, Uniqueness for Keller-Segel-type chemotaxis models. *Discrete Contin. Dyn. Syst.* **34** (2014), no. 4, 1319–1338 Zbl 1277.35009 MR 3117843

[13] J.-A. Carrillo and F. Santambrogio, L^∞ estimates for the JKO scheme in parabolic-elliptic Keller-Segel systems. *Quart. Appl. Math.* **76** (2018), no. 3, 515–530 Zbl 1391.35197 MR 3805040

[14] G. De Philippis, A. R. Mészáros, F. Santambrogio, and B. Velichkov, BV estimates in optimal transportation and applications. *Arch. Ration. Mech. Anal.* **219** (2016), no. 2, 829–860 Zbl 1333.49051 MR 3437864

[15] J. Dolbeault and B. Perthame, Optimal critical mass in the two-dimensional Keller-Segel model in $\mathbb{R}^2$. *C. R. Math. Acad. Sci. Paris* **339** (2004), no. 9, 611–616 Zbl 1056.35076 MR 2103197

[16] V. Ferrari and F. Santambrogio, Lipschitz estimates on the JKO scheme for the Fokker-Planck equation on bounded convex domains. *Appl. Math. Lett.* **112** (2021), Paper No. 106806 Zbl 1454.35385 MR 4162959

[17] E. Giusti, *Direct methods in the calculus of variations*. World Scientific, River Edge, NJ, 2003 Zbl 1028.49001 MR 1962933

[18] T. Hillen and K. J. Painter, A user's guide to PDE models for chemotaxis. *J. Math. Biol.* **58** (2009), no. 1-2, 183–217 Zbl 1161.92003 MR 2448428

[19] M. Iacobelli, F. S. Patacchini, and F. Santambrogio, Weighted ultrafast diffusion equations: from well-posedness to long-time behaviour. *Arch. Ration. Mech. Anal.* **232** (2019), no. 3, 1165–1206 Zbl 1480.35284 MR 3928748

[20] W. Jäger and S. Luckhaus, On explosions of solutions to a system of partial differential equations modelling chemotaxis. *Trans. Amer. Math. Soc.* **329** (1992), no. 2, 819–824 Zbl 0746.35002 MR 1046835

[21] R. Jordan, D. Kinderlehrer, and F. Otto, The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.* **29** (1998), no. 1, 1–17 Zbl 0915.35120 MR 1617171

[22] L. Kantorovich, On the translocation of masses. *C. R. (Doklady) Acad. Sci. URSS (N.S.)* **37** (1942), 199–201 Zbl 0061.09705 MR 0009619

[23] E. F. Keller and L. A. Segel, Initiation of slime mold aggregation viewed as an instability. *J. Theoret. Biol.* **26** (1970), no. 3, 399–415 Zbl 1170.92306 MR 3925816

[24] D. Kwon and A. R. Mészáros, Degenerate nonlinear parabolic equations with discontinuous diffusion coefficients. *J. Lond. Math. Soc. (2)* **104** (2021), no. 2, 688–746 Zbl 1478.35134 MR 4311108

[25] P. W. Y. Lee, On the Jordan-Kinderlehrer-Otto scheme. *J. Math. Anal. Appl.* **429** (2015), no. 1, 131–142 Zbl 1318.65064 MR 3339068

[26] D. Matthes, R. J. McCann, and G. Savaré, A family of nonlinear fourth order equations of gradient flow type. *Comm. Partial Differential Equations* **34** (2009), no. 10-12, 1352–1397 Zbl 1187.35131 MR 2581977

[27] R. J. McCann, A convexity principle for interacting gases. *Adv. Math.* **128** (1997), no. 1, 153–179 Zbl 0901.49012 MR 1451422

[28] G. Monge, Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences de Paris 1781*, pp. 666–704

[29] B. Perthame, *Transport equations in biology*. Front. Math., Birkhäuser, Basel, 2007 Zbl 1185.92006 MR 2270822

[30] F. Santambrogio, *Optimal transport for applied mathematicians*. Progr. Nonlinear Differential Equations Appl. 87, Birkhäuser/Springer, Cham, 2015 Zbl 1401.49002 MR 3409718

[31] F. Santambrogio, {Euclidean, metric, and Wasserstein} gradient flows: an overview. *Bull. Math. Sci.* **7** (2017), no. 1, 87–154 Zbl 1369.34084 MR 3625852

[32] C. Villani, *Optimal transport*. Grundlehren Math. Wiss. 338, Springer, Berlin, 2009 Zbl 1156.53003 MR 2459454

Received 6 October 2020; accepted 10 September 2021.

Simone Di Marino
Dipartimento di Matematica (DIMA), MaLGa, Università di Genova, Via Dodecaneso 35, 16146 Genova, Italy; simone.dimarino@unige.it

Filippo Santambrogio
Institut Camille Jordan, Université Claude Bernard Lyon 1, 43 boulevard du 11 novembre 1918, 69622 Villeurbanne Cedex, France; santambrogio@math.univ-lyon1.fr