Strong $L^2 H^2$ convergence of the JKO scheme for the Fokker-Planck equation

Filippo Santambrogio[*], Gayrat Toshpulatov[†]

May 14, 2024

Abstract

Following a celebrated paper by Jordan, Kinderleherer and Otto it is possible to discretize in time the Fokker-Planck equation $\partial_t \varrho = \Delta\varrho + \nabla\cdot(\varrho\nabla V)$ by solving a sequence of iterated variational problems in the Wasserstein space, and the sequence of piecewise constant curves obtained from the scheme is known to converge to the solution of the continuous PDE. This convergence is uniform in time valued in the Wasserstein space and also strong in L^1 in space-time. We prove in this paper, under some assumptions on the domain (a bounded and smooth convex domain) and on the initial datum (which is supposed to be bounded away from zero and infinity and belong to $W^{1,p}$ for an exponent p larger than the dimension), that the convergence is actually strong in $L^2_t H^2_x$, hence strongly improving the previously known results in terms of the order of derivation in space. The technique is based on some inequalities, obtained with optimal transport techniques, that can be proven on the discrete sequence of approximate solutions, and that mimic the corresponding continuous computations.

Contents

---

[*] Universite Claude Bernard Lyon 1, CNRS, Ecole Centrale de Lyon, INSA Lyon, Université Jean Monnet, Institut Camille Jordan, UMR5208, 43 bd du 11 novembre 1918, 69622 Villeurbanne Cedex, France. `santambrogio@math.univ-lyon1.fr`

[†] Institut für Analysis & Scientific Computing,Technische Universität Wien, Wiedner Hauptstr. 8, 1040 Wien, Austria, `gayrat.toshpulatov@tuwien.ac.at`

1 Introduction

More than 20 years ago Jordan, Kinderleherer and Otto wrote their seminal paper [15], where they identified a variational structure in the Fokker-Plank equation

$$\partial_t \varrho = \Delta \varrho + \nabla \cdot (\rho \nabla V)$$

as a gradient flow of the functional $J(\varrho) = \int \varrho(V + \log \varrho)$ in the Wasserstein space W_2. This is remarkable since the same equation has no gradient-flow structure for Hilbertian distances such as the L^2 or H^{-1} norms, differently from the heat equation (obtained for $V = 0$). The gradient flow interpretation also gives a natural discretization in time, where a time step $\tau > 0$ is fixed and a sequence $(\varrho^\tau_k)_k$ is iteratively defined using

$$\varrho^\tau_{k+1} \in \operatorname{argmin}_\varrho \left\{ J(\varrho) + \frac{W_2^2(\varrho, \varrho^\tau_k)}{2\tau} \right\}.$$

This sequence is then used to define a curve $t \mapsto \varrho^\tau(t)$ in the space of probability measures via $\varrho^\tau(0) = \varrho_0$ and

$$\varrho^\tau(t) = \varrho^\tau_{k+1} \text{ for } t \in (k\tau, (k+1)\tau].$$

In [15] the convergence of ϱ^τ to the solution of the Fokker-Planck equation was proven when the domain on which the equation is set is the whole space (of course under suitable decay assumptions on the initial datum, namely that it has finite second moment $\int |x|^2 d\varrho_0(x) < +\infty$), and the convergence is weak in L^1 for every t and strong in $L^1([0,T] \times \mathbb{R}^d)$. The analysis of the convergence can be adapted to the case of a bounded domain Ω (and in this case there is no moment condition) and the results are essentially the same. This is, for instance, the object of Chapter 8 in [24], where the Fokker-Plank equation is chosen as an example to present the JKO scheme and the gradient-flow approach to some diffusion equations. As it is a linear equation, it is also the simplest case where this analysis can be performed.

The goal of the present paper is to improve the nature of the above convergence, under some possible assumptions on the initial datum. We do not mean obtaining a better rate of convergence in terms of τ (note that [1] proves a convergence of order $O(\tau)$ in the Wasserstein distance W_2 whenever V is semi-convex), but obtaining strong convergences in the best possible spaces.

In particular, we obain in this paper strong convergence results for the space derivatives up to order two of the solutions (more precisely, we obtain strong $L^2_t H^2_x$ convergence under some assumptions). Besides its *per se* mathematical interest, the main application of this fact is numerical, since this means that one can "trust" the approximation found via the JKO scheme (which can be attacked by different numerical methods, see, for instance, [23, 22] or [25] for a general presentation) not only in order to predict the behavior of the true solution, but also of its derivatives. Of course, the interest of these estimates in the very case of the Fokker-Plack equation is limited, since this equation is a simple linear variant of the heat equation, for which efficient numerical solvers exist independently of its gradient-flow interpretation, and it would be interesting to generalize this approach to other PDEs which have a gradient flow structure in the Wasserstein space but feature non-linear diffusion, such as the porous medium or the fast diffusion equation (see, for instance, [21]). Yet, this extesion is not within reach so far and will be a matter of further study.

The paper will undergo these proofs of convergence by analyzing different steps. The first one consists in proving a very classical $L^2_t H^1_x$ bound on the discrete solutions ϱ^τ, which is obtained by a discrete analogue of a very standard computation, which is itself based on a simple integration-by-parts: the time derivative of $\int \frac{1}{2}\varrho^2$ equals $-\int |\nabla\varrho|^2 + \varrho\nabla\varrho\cdot\nabla V$ whenever ϱ solves the Fokker-Planck equation. A similar computation can be done for the sequence obtained via the JKO scheme, but strongly relies on the geodesic convexity of the functional $\varrho \mapsto \int \frac{1}{2}\varrho^2$ (see [20] or [24, Chapter 7]). This estimate provides strong L^2 compactness in space, and allows to obtain convergence in $L^2_t L^2_x$ via the Aubin-Lions lemma when coupled with bounds in time on a suitable interpolation. This $L^2_t L^2_x$ is just a small refinement of the original L^1 convergence already proven by Jordan-Kinderlehrer-Otto, but is a necessary step to go on. The next step consists in the strong convergence in $L^2_t H^1_x$. This is obtained by refining the same computations. Once we have a bound on $\int_0^T \int_\Omega |\nabla\varrho^\tau|^2$, this provides weak convergence in $L^2_t H^1_x$ and the limit can only be the solution ϱ to the limit Fokker-Planck equation. We do have strong convergence if we are able to prove $\limsup_{\tau\to 0} \int_0^T \int_\Omega |\nabla\varrho^\tau|^2 \le \int_0^T \int_\Omega |\nabla\varrho|^2$, which can be obtained by the very same estimates (using the strong $L^2_t L^2_x$ convergence to handle the extra term involving ∇V). This proof is presented in Section 4, after two preliminary sections, one on the properties of the solution in continuous-time (Section 2) and one on the properties of the JKO scheme (Section 3).

Then, a similar argument is proposed for the convergence of the second derivatives in space. The strategy consists in finding a first-order quantity which decreases along iterations of the JKO scheme such that its dissipation is a second-order quantity identical to the one which could be obtained along the continuous-in-time flow of the PDE up to terms which tend to 0 when $\tau \to 0$. This is done by looking at the evolution in time of the quantities

$$F_p(\varrho) := \frac{1}{p}\int_\Omega \left|\frac{\nabla\varrho}{\varrho} + \nabla V\right|^p d\varrho.$$

The particular case $p = 2$ is the most important one, as the functional F_2, sometimes called *Fisher information* (in particular in the case $V = 0$) naturally appears in the Fokker-Planck equation as the dissipation of the entropy J along the solution of the equation (more precisely, we do have $\partial_t J(\varrho_t) = -2F_2(\varrho_t)$). This fact is widely used in functional inequalities as for instance in the Bakry-Emery theory (see [2], for instance). Our analysis will be based on the evolution in time of the functionals F_p along the JKO scheme, and on the evolution of F_2 on both the JKO and the continuous-in-time equation. It is possible to differentiate F_2 in time and obtain several terms including the main one $-\int_\Omega \varrho|D^2(\log\varrho + V)|^2$. The same computation may be done on the JKO scheme using the so-called five-gradients inequality introduced by the second author and collaborators in [11] and applied to the Fokker-Planck equation in [12]. This requires a finer analysis than what is done in [12] since the remainders of the inequality will be crucial. Moreover, the dissipation along the steps of the JKO scheme does not provide exactly the desired term $-\int_\Omega \varrho^\tau|D^2(\log\varrho^\tau + V)|^2$ but includes an error term of the order of $||D^2\varphi^\tau_k||_{L^\infty}$, where φ^τ_k is the Kantorovich potential in the optimal transport from ϱ^τ_k to ϱ^τ_{k-1}.

In order to get rid of this error term we need uniform upper and lower bounds together with uniform $C^{0,\alpha}$ estimates on ϱ^τ, which provide (using Caffarelli's theory for the Monge-Ampère equation, [5, 6, 7, 8, 9, 10]) a uniform bound $\varphi^\tau_k \in C^{2,\alpha}$; this, combined with simpler bounds on $\nabla\varphi^\tau_k$ allows to obain $||D^2\varphi^\tau_k||_{L^\infty} \to 0$. Uniform $C^{0,\alpha}$ bounds on ϱ^τ

are obtained in a non-optimal way: we indeed suppose $F_p(\varrho_0) < +\infty$ for $p > d$ and prove that this quantity stays bounded in time, which implies the Hölder behavior because of standard Sobolev embeddings.

Once we get rid of the error terms, the fact that the estimates in discrete and in continuous time are essentially the same allows to obtain $\int_0^T \int_\Omega \varrho^\tau |D^2(\log \varrho^\tau + V)|^2 \to \int_0^T \int_\Omega \varrho |D^2(\log \varrho + V)|^2$ which provides strong $L^2_{t,x}$ convergence of $\sqrt{\varrho^\tau} D^2 \log \varrho^\tau$ and, after carefully using again the upper and lower bounds on ϱ^τ, we obtain $D^2\varrho^\tau \to D^2\varrho$ in $L^2_{t,x}$.

The experienced reader can see that some of the assumptions on th initial data are crucial, and some or essentially technical: first of all, we assume ϱ_0 to be bounded from above and from below which simplifies many statements and proofs in order to obtain, with our technique, dirct bounds on ϱ^τ and not on functions of it; once these bounds are accepted, the condition $F_2(\varrho_0) < +\infty$ becomes absolutely necessary since the solution of the PDE is $L^2_t H^2_x$ if and only if $\varrho_0 \in H^1$; on the other hand, $F_p(\varrho_0) < +\infty$ for $p > d$ is not a natural assumption, and is chosen for technical reasons. Getting rid of these extra assumptions, as well as extending this kind of results to other diffusion equations, would of course be interesting, but since this is, to the best of our knowledge, the first higher-order strong convergence result for a JKO scheme we believe that it has its own interest.

Acknowledgement. The first author acknowledges the support of the ANR project MAGA (ANR-16-CE40-0014), of the Lagrange Mathematics and Computation Research Center project on Optimal Transportation, and then of the European Union via the ERC AdG 101054420 EYAWKAJKOS project. The second author acknowledges the partial support of the Austrian Science Fund (FWF) through the project F65 “Taming Complexity in Partial Differential Systems”, as well as the support of École normale supérieure de Lyon and its scholarship program Ampère for the year spent in Lyon when the work leading to this paper started as a part of his master thesis.

2 Basics on the Fokker-Planck equation

In the present paper we will consider the Fokker-Planck equation in a finite interval $[0, T]$ and a convex bounded domain $\Omega \subset \mathbb{R}^d$ whose boundary $\partial\Omega$ is smooth enough. The drift in the equation will be of gradient type and autonomous. We denote by $\vec{n}$ the exterior unit normal vector of the boundary $\partial\Omega$. We consider the Cauchy problem for the Fokker-Planck equation with no-flux boundary condition, i.e.,

$$\begin{cases} \partial_t \varrho(t,x) = \Delta \varrho(t,x) + \operatorname{div}(\varrho(t,x)\nabla V(x)), & (t,x) \in (0,T] \times \Omega, \\ \nabla \varrho(t,x) \cdot \vec{n}(x) + \varrho(t,x) \nabla V(x) \cdot \vec{n}(x) = 0, & (t,x) \in [0,T] \times \partial\Omega, \\ \varrho(0,x) = \varrho_0(x), & x \in \Omega, \end{cases} \tag{1}$$

where $\varrho_0 \in \mathcal{P}(\Omega) \cap L^1_+(\Omega)$.
The initial total mass is $\int_\Omega \varrho_0(x)dx = 1$ and it can be formally seen that it is preserved.

The meaning of the no-flux boundary condition in (1) is the following: we only consider solutions which are at least $L^2([0,T];H^1(\Omega))\cap C([0,T];L^2(\Omega))$ and we require them to satisfy

$$\int_0^T\int_\Omega \varrho(t,x)\partial_t\varphi(t,x)dxdt - \int_0^T\int_\Omega(\nabla\varrho(t,x)+\varrho(t,x)\nabla V(x))\cdot\nabla\varphi(t,x)dxdt$$
$$= \int_\Omega \varphi(T,x)\varrho(T,x)dx - \int_\Omega \varphi(0,x)\varrho(0,x))dx$$

for every test function $\varphi\in C^1([0,T]\times\Omega)$, without requiring φ to vanish on $\partial\Omega$.

Well-known results on parabolic differential equations (see [16], [19]) let us to state the following properties of (1):

Proposition 2.1. *Let $[0,T]$ be a finite interval and Ω be bounded domain whose boundary $\partial\Omega$ is Lipschitz continuous. Then the following hold:*

- *if V is Lipschitz continuous and the initial data $\varrho_0\in L^2(\Omega)$, then there exists a unique solution ϱ of (1) in $L^2([0,T];H^1(\Omega))\cap C([0,T];L^2(\Omega))$. If $\varrho_0\in H^1(\Omega)$, then we have $\varrho\in L^2([0,T];H^2(\Omega))\cap C([0,T];H^1(\Omega))$.*
- *if $\varrho_0\in C(\bar\Omega)$ and $V\in C^2(\bar\Omega)$, then $\varrho\in C([0,T]\times\Omega)$ is differentiable with respect to t in $(0,T]\times\bar\Omega$; $\varrho(t,\cdot)$ belongs to $W^{2,p}(\Omega)$ for every $p\geq 1$.*
- *If ϱ_0 is bounded from below and above by two positive constants, then the same (for possibly different constants) holds for ϱ.*
- *if $\varrho_0\in C(\bar\Omega)$, $\partial\Omega$ has $C^{2+\alpha}$ regularity and $V\in C^{2+\alpha}(\bar\Omega)$ for $\alpha\in(0,1)$, then $\varrho\in C^{1+\frac{\alpha}{2},2+\alpha}((0,T]\times\bar\Omega)$. If moreover $\partial\Omega$ and V are $C^{3+\alpha}$, then $\varrho(t,\cdot)\in C^{3+\alpha}(\bar\Omega)$ for every $t>0$.*

Once we know that the solution of (1) exists, is unique, and is smooth, we are interested in evaluating, and in particular differentiating in time, some quantities involving the solution. First we consider the following classical statement

Proposition 2.2. *Let $\varrho_0\in L^2(\Omega)$ and ϱ_t be the unique solution of (1). Then we have*

$$\int_\Omega \varrho_T^2(x)\,dx - \int_\Omega \varrho_0^2(x)\,dx = -2\int_0^T\int_\Omega |\nabla\varrho_t(x)|^2\,dxdt - 2\int_0^T\int_\Omega \varrho_t(x)\,\nabla\varrho_t(x)\cdot\nabla V(x)\,dxdt.$$

Proof. In order to obtain this result it is enough to differentiate in time the function $t\mapsto\int_\Omega\varrho_t^2$. To do this, we use a Lemma 2.3, which is a general lemma from functional analysis. Since $\varrho_t\in L^2([0,T];H^1(\Omega))$ we have $\partial_t\varrho_t=\Delta\varrho_t+\operatorname{div}(\varrho_t\nabla V)\in L^2([0,T];H^{-1}(\Omega))$. Thus, ϱ_t satisfies Lemma 2.3 in the case of $V=H^1(\Omega)$ and $H=L^2(\Omega)$. Consequently, we have that $\varrho_t\in C([0,T];L^2(\Omega))$ and

$$\int_\Omega \varrho_T^2(x)\,dx - \int_\Omega \varrho_0^2(x)\,dx = 2\int_0^T\langle\partial_t\varrho_t,\varrho_t\rangle_{H^{-1},H^1}\,dt$$

and it is enough to use the expression for $\partial_t\varrho$ and integrate in time to obtain the result. □

In the above proof we mentioned a general functional analysis fact, which is recalled here below. To introduce it, let us consider a Hilbert space H endowed with the norm $||\cdot||_H$, a Banach space V, and we assume that V is reflexive, $V \subset H$ with dense and bounded embedding. The following Lemma is proven, for instance in [27, Lemma 1.2, page 260].

Lemma 2.3. *The following inclusion*

$$L^2([0,T];V)\cap H^1([0,T];V') \subset C([0,T];H)$$

holds true. Moreover, for any $g \in L^2([0,T];V)\cap H^1([0,T];V')$ *there holds*

$$t \to ||g(t)||_H^2 \in W^{1,1}(0,T)$$

and

$$\frac{d}{dt}||g(t)||_H^2 = 2\langle g'(t), g(t)\rangle_{V',V} \ \textit{a.e. on } (0,T). \tag{2}$$

In the above computation, we saw that a zero-order quantity (here $\int \varrho^2$) is the integral in time of a first-order quantity (which is in our case given by $\int |\nabla\varrho^2| + \varrho\nabla\varrho\cdot\nabla V$). We now need to look at higher-order quantities. In particular, we will consider the functional

$$F_p(\varrho) := \frac{1}{p}\int_\Omega \left|\frac{\nabla\varrho}{\varrho} + \nabla V\right|^p d\varrho \tag{3}$$

defined for $\varrho \in W^{1,1}(\Omega)\cap\mathcal{P}(\Omega)$. We will mainly look at F_2.

Lemma 2.4. *Suppose* $0 < T < +\infty$*,* Ω *is a bounded domain with* $C^{3+\alpha}$ *boundary,* $V \in C^{3+\alpha}(\bar\Omega)$ *for some* $\alpha \in (0,1)$ *and* $\varrho_0 \in C(\bar\Omega)$ *is positive. If* ϱ *is the solution of (1), then for* $t > 0$ *we have*

$$\partial_t F_2(\varrho_t) = -\int_\Omega |D^2(\log\varrho + V)|^2 \varrho\, dx - \int_\Omega (\nabla(\log\varrho+V))^T\cdot D^2V\cdot\nabla(\log\varrho+V)\varrho\, dx +$$
$$\int_{\partial\Omega} (\nabla(\log\varrho+V))^T\cdot D^2(\log\varrho+V)\cdot\vec{n}\varrho\, d\mathcal{H}^{d-1}. \tag{4}$$

Proof. The assumptions provide that $\varrho(t,\cdot)$ is also positive and belongs to $C^{3+\alpha}(\bar\Omega)$ for every $t \in (0,T]$. We have

$$2\partial_t F_2(\varrho_t) = \partial_t \int_\Omega \sum_{i=1}^d \left(\frac{\partial_i\varrho}{\varrho} + \partial_i V\right)^2 \varrho\, dx =$$
$$\sum_{i=1}^d \int_\Omega \partial_t\varrho\left(\frac{\partial_i\varrho}{\varrho} + \partial_i V\right)^2 dx + 2\sum_{i=1}^d\int_\Omega\left(\frac{\partial_i\varrho}{\varrho} + \partial_i V\right)\left(\partial_t\partial_i\varrho - \frac{\partial_i\varrho\,\partial_t\varrho}{\varrho}\right)dx =$$
$$\int_\Omega\sum_{i=1}^d \partial_t\varrho\left[(\partial_i V)^2 - \left(\frac{\partial_i\varrho}{\varrho}\right)^2\right]dx + 2\sum_{i=1}^d\int_\Omega\left(\frac{\partial_i\varrho}{\varrho}+\partial_i V\right)\partial_t\partial_i\varrho\, dx.$$

We look at the different parts of the last integral. First we use the equation $\partial_t\varrho = \Delta\varrho + \mathrm{div}(\varrho\nabla V)$ and integrate by parts.

$$\int_\Omega\sum_{i=1}^d\partial_t\varrho\left[(\partial_i V)^2 - \left(\frac{\partial_i\varrho}{\varrho}\right)^2\right]dx = \int_\Omega\sum_{i,j=1}^d\left[(\partial_i V)^2 - \left(\frac{\partial_i\varrho}{\varrho}\right)^2\right]\partial_j\left(\partial_j\varrho + \varrho\,\partial_j V\right)dx =$$

$$-2\sum_{i,j=1}^{d}\int_{\Omega}\left[\varrho\,\partial_i V\partial_{ij}V-\frac{\partial_i\varrho\,\partial_{ij}\varrho}{\varrho}+\frac{(\partial_i\varrho)^2\,\partial_j\varrho}{\varrho^2}\right]\left(\frac{\partial_j\varrho}{\varrho}+\partial_j V\right)dx+$$

$$\sum_{i,j=1}^{d}\int_{\partial\Omega}\left[(\partial_i V)^2-\left(\frac{\partial_i\varrho}{\varrho}\right)^2\right]\left(\frac{\partial_j\varrho}{\varrho}+\partial_j V\right)\varrho n_j\,d\mathcal{H}^{d-1}.$$

We now compute the second part

$$2\sum_{i=1}^{d}\int_{\Omega}\left(\frac{\partial_i\varrho}{\varrho}+\partial_i V\right)\partial_t\partial_i\varrho\,dx=2\sum_{i,j=1}^{d}\int_{\Omega}\left(\frac{\partial_i\varrho}{\varrho}+\partial_i V\right)\left(\partial_{ijj}\varrho+\partial_{ij}(\varrho\partial_j V)\right)dx=$$

$$-2\sum_{i,j=1}^{d}\int_{\Omega}\left(\frac{\partial_{ij}\varrho}{\varrho}+\partial_{ij}V-\frac{\partial_i\varrho\,\partial_j\varrho}{\varrho^2}\right)\left(\partial_{ij}\varrho+\varrho\partial_{ij}V+\partial_i\varrho\,\partial_j V\right)dx+$$

$$2\sum_{i,j=1}^{d}\int_{\partial\Omega}\left(\frac{\partial_i\varrho}{\varrho}+\partial_i V\right)(\partial_{ij}\varrho+\varrho\partial_{ij}V+\partial_i\varrho\,\partial_j V)n_j\,d\mathcal{H}^{d-1}=$$

$$-2\sum_{i,j=1}^{d}\int_{\Omega}\left(\frac{\partial_{ij}\varrho}{\varrho}+\partial_{ij}V-\frac{\partial_i\varrho\,\partial_j\varrho}{\varrho^2}\right)^2\varrho\,dx$$

$$-2\sum_{i,j=1}^{d}\int_{\Omega}\left(\frac{\partial_{ij}\varrho}{\varrho}+\partial_{ij}V-\frac{\partial_i\varrho\,\partial_j\varrho}{\varrho^2}\right)\partial_i\varrho\left(\frac{\partial_j\varrho}{\varrho}+\partial_j V\right)dx$$

$$+2\sum_{i,j=1}^{d}\int_{\partial\Omega}\left(\frac{\partial_i\varrho}{\varrho}+\partial_i V\right)(\partial_{ij}\varrho+\varrho\partial_{ij}V+\partial_i\varrho\,\partial_j V)n_j\,d\mathcal{H}^{d-1}.$$

We consider the integrals over the boundary of Ω. Because of the no-flux boundary condition the first boundary integral vanishes:

$$\sum_{i,j=1}^{d}\int_{\partial\Omega}\left[(\partial_i V)^2-\left(\frac{\partial_i\varrho}{\varrho}\right)^2\right]\left(\frac{\partial_j\varrho}{\varrho}+\partial_j V\right)\varrho n_j\,d\mathcal{H}^{d-1}=0.$$

Because of the same reason the second boundary integral can be written as follows

$$2\sum_{i,j=1}^{d}\int_{\partial\Omega}\left(\frac{\partial_i\varrho}{\varrho}+\partial_i V\right)(\partial_{ij}\varrho+\varrho\partial_{ij}V+\partial_i\varrho\cdot\partial_j V)n_j\,d\mathcal{H}^{d-1}=$$

$$2\sum_{i,j=1}^{d}\int_{\partial\Omega}\varrho\left(\frac{\partial_i\varrho}{\varrho}+\partial_i V\right)\left(\frac{\partial_{ij}\varrho}{\varrho}+\partial_{ij}V-\frac{\partial_i\varrho\cdot\partial_j\varrho}{\varrho^2}\right)n_j\,d\mathcal{H}^{d-1}.$$

Consequently, we have

$$\partial_t F_2(\varrho)=-\sum_{i,j=1}^{d}\int_{\Omega}\left(\frac{\partial_{ij}\varrho}{\varrho}+\partial_{ij}V-\frac{\partial_i\varrho\,\partial_j\varrho}{\varrho^2}\right)^2\varrho\,dx-$$

$$\sum_{i,j=1} \int_\Omega \left(\frac{\partial_i \varrho}{\varrho} + \partial_i V\right) \partial_{ij} V \left(\frac{\partial_j \varrho}{\varrho} + \partial_j V\right) \varrho \, dx$$

$$+ \sum_{i,j=1}^{d} \int_{\partial\Omega} \varrho \left(\frac{\partial_i \varrho}{\varrho} + \partial_i V\right) \left(\frac{\partial_{ij} \varrho}{\varrho} + \partial_{ij} V - \frac{\partial_i \varrho \, \partial_j \varrho}{\varrho^2}\right) n_j \, d\mathcal{H}^{d-1}.$$

If we take into account that we have

$$\partial_i(\log \varrho + V) = \left(\frac{\partial_i \varrho}{\varrho} + \partial_i V\right) \text{ and } \partial_{i,j}(\log \varrho + V) = \frac{\partial_{ij} \varrho}{\varrho} + \partial_{ij} V - \frac{\partial_i \varrho \, \partial_j \varrho}{\varrho^2},$$

we get the desired equality. □

The last term in formula (4) can be re-written using the following lemma.

Lemma 2.5. *Suppose* $\Omega = \{h < 0\}$ *for a smooth function* $h : \mathbb{R}^d \to \mathbb{R}$ *with* $\nabla h \neq 0$ *on* $\{h = 0\}$*, so that the exterior normal vector at* $x \in \partial\Omega$ *is given by* $\vec{n}(x) = \nabla h(x)/|\nabla h(x)|$*. Let* $v : \Omega \to \mathbb{R}^d$ *be a smooth vector field such that* $v \cdot \vec{n} = 0$ *on* $\partial\Omega$*. Then we have the following equality for every* $x \in \partial\Omega$

$$v(x)^T \cdot Dv(x) \cdot n(x) = -\frac{v(x)^T \cdot D^2 h(x) \cdot v(x)}{|\nabla h(x)|}.$$

Proof. Given $x \in \partial\Omega$, we consider a smooth curve $\gamma : (-t_0, t_0) \to \partial\Omega$ with $\gamma(0) = x$ and write the equality $v(\gamma(t)) \cdot \nabla h(\gamma(t)) = 0$ for every t. Differentiating w.r.t. t we obtain

$$\gamma'(t)^T \cdot Dv(\gamma(t)) \cdot \nabla h(\gamma(t)) + v(\gamma(t))^T \cdot D^2 h(\gamma(t)) \cdot \gamma'(t) = 0.$$

We can take $t = 0$ and choose a curve with $\gamma'(0) = v(x)$ since v is tangent to the surface $\partial\Omega$, thus obtaining

$$v(x)^T \cdot Dv(x) \cdot \nabla h(x) = -v(x)^T \cdot D^2 h(x) \cdot v(x).$$

It is then enough to divide by $|\nabla h(x)|$ in order to get the claim. □

We then obtain the following formula.

Corollary 2.6. *Suppose* $0 < T < +\infty$*, take* $\Omega = \{h < 0\}$ *a bounded domain defined as the negativity set of a function* $h \in C^2$ *with* $\nabla h \neq 0$ *on* $\{h = 0\}$*, and* $V \in C^2(\bar{\Omega})$*. Given a strictly positive* $\varrho_0 \in H^1(\Omega)$ *initial datum, let* ϱ *be the solution of (1). We then have*

$$\begin{aligned} F_2(\varrho_T) - F_2(\varrho_0) \quad = \quad & -\int_0^T dt \int_\Omega |D^2(\log \varrho + V)|^2 \varrho \, dx \\ & -\int_0^T dt \int_\Omega (\nabla(\log \varrho + V))^T \cdot D^2 V \cdot \nabla(\log \varrho + V) \varrho \, dx \\ & -\int_0^T dt \int_{\partial\Omega} (\nabla(\log \varrho + V))^T \cdot D^2 h \cdot (\nabla(\log \varrho + V)) \varrho \, d\mathcal{H}^{d-1}. \end{aligned}$$

Proof. The result is obtained by first generalizing formula (4) to the case of C^2 regularity by approximation, then integrating in time, using the continuity in H^1 of the solution at $t = 0$, and finally re-writing the boundary term using Lemma 2.5 □

3 The JKO scheme for the Fokker-Planck equation

We define the functional $J\colon \mathcal{P}(\Omega) \to \mathbb{R}$ as follows

$$J(\mu) = \begin{cases} \displaystyle\int_\Omega \frac{d\mu}{dx} \log\left(\frac{d\mu}{dx}\right) dx + \int_\Omega V\, d\mu & \text{if } \mu \ll dx \\ +\infty & \text{otherwise.} \end{cases} \tag{5}$$

The main achievement of [15] is to view the Fokker-Planck equation (1) as a gradient flow of the functional J in the metric space $(\mathcal{P}(\Omega), W_2)$ (see also [1, 24, 28]) and to define a discrete iterated scheme converging to the solution. This scheme is nowadays called Jordan-Kinderlehrer-Otto scheme.

Given $N \in \mathbb{N}$, we set $\tau := \frac{T}{N}$ and we assume that the initial data ϱ_0 satisfies $J(\varrho_0) < \infty$. We define recursively a sequence of probability measures $\{\varrho_k^\tau\}_{k=0}^N$ such that $\varrho_0^\tau := \varrho_0$ and

$$\varrho_{k+1}^\tau \in \operatorname{argmin}_\varrho \left\{ J(\varrho) + \frac{W_2^2(\varrho, \varrho_k^\tau)}{2\tau} \right\} \tag{6}$$

We use this sequence to build a curve ϱ^τ in the space of probability measures defined via $\varrho^\tau(0) = \varrho_0$ and

$$\varrho^\tau(t) = \varrho_{k+1}^\tau \text{ for } t \in (k\tau, (k+1)\tau] \text{ and } k \in \mathbb{N} \cup \{0\}.$$

In the original work by R. Jordan, D. Kinderlehrer and F. Otto ([15]) the above scheme was considered for $\Omega = \mathbb{R}^d$, and the following important theorem was proven.

Theorem 3.1. *Let $V \in C^\infty(\mathbb{R}^d)$, $\varrho_0 \in \mathcal{P}_2(\mathbb{R}^d)$ satisfies $J(\varrho_0) < \infty$, and for given $\tau > 0$, let $\{\varrho_k^\tau\}_{k\in\mathbb{N}}$ be defined recursively by (6). Define the interpolation*

$$\varrho^\tau(t) = \varrho_{k+1}^\tau \textit{ for } t \in (k\tau, (k+1)\tau] \textit{ and } k \in \mathbb{N} \cup \{0\}.$$

Then as $\tau \to 0$, $\varrho^\tau(t) \to \varrho(t)$ weakly in $L^1(\mathbb{R}^d)$ for all $t \in (0, \infty)$, where $\varrho \in C^\infty((0,\infty) \times \mathbb{R}^d)$ is the unique solution of

$$\partial_t \varrho = \Delta \varrho + \operatorname{div}(\varrho \nabla V)$$

with initial condition $\varrho(t) \to \varrho_0$ strongly in $L^1(\mathbb{R}^d)$ for $t \to 0$. Moreover, $\varrho^\tau \to \varrho$ strongly in $L^1((0,T) \times \mathbb{R}^d)$ for all $T < \infty$.

In this paper we are instead interested in the case where Ω is an open bonded subset of $\mathbb{R}^d$. The details of the JKO scheme for bounded domains are, for instance, given in [24, Chapter 8]. In this case it is important to emphasize that the limit curve of the JKO scheme not only solves the Fokker-Planck equation but also satisfies the *no-flux* boundary condition. More precisely, we summarize here the properties we need about the JKO scheme for bounded Ω :

Theorem 3.2. *Let $[0,T]$ be a finite interval and Ω be a bounded domain of $\mathbb{R}^d$ whose boundary $\partial\Omega$ is Lipschitz continuous. We assume V is Lipschitz continuous and the initial data $\varrho_0 \in \mathcal{P}(\Omega) \cap L^1_+(\Omega)$ satisfies $J(\varrho_0) < \infty$. Then the following hold:*

1. *The functional J has a unique minimum over $\mathcal{P}(\Omega)$. In particular J is bounded from below. Moreover, for each $\tau > 0$, the sequence $\{\varrho_k^\tau\}_{k=0}^N$ defined by the formula (6) is well-defined (there is a unique minimizer at every step).*

2. *For any $k \in \{1, ..., N\}$, the optimizer ϱ_k^τ is strictly positive, Lipschitz continuous, and satisfies*
$$\log \varrho_{k+1}^\tau + V + \frac{\varphi_k}{\tau} = constant \tag{7}$$
where φ_k is the Kantorovich potential from ϱ_{k+1}^τ to ϱ_k^τ.

3. *For every $\tau > 0$, the sequence $\{\varrho_k^\tau\}_{k=0}^N$ satisfies*
$$\sum_{k=0}^{N-1} \frac{W_2^2(\varrho_k^\tau, \varrho_{k+1}^\tau)}{\tau} \leq 2(J(\varrho_0) - \inf J). \tag{8}$$

4. *There exists a $\frac{1}{2}$-Hölder and absolutely continuous curve ϱ in $(\mathcal{P}(\Omega), W_2)$ such that $W_2(\varrho_t^\tau, \varrho_t) \to 0$ uniformly as $\tau \to 0$. Moreover, ϱ satisfies the Fokker-Planck equation (1) in the distributional sense.*

Some estimates in L^p have been established in [12] when the domain Ω is convex.

Theorem 3.3. *Let $[0, T]$ be a finite interval and Ω be a convex bounded domain of $\mathbb{R}^d$ whose boundary $\partial\Omega$ is Lipschitz continuous. We assume V is Lipschitz continuous and the initial data $\varrho_0 \in \mathcal{P}(\Omega) \cap L_+^1(\Omega)$ satisfies $J(\varrho_0) < \infty$.*

(i) Given $f : \mathbb{R}_+ \to \mathbb{R}$ satisfying the d-McCann condition (i.e. $[0, \infty) \ni s \mapsto f(s^{-d})s^d$ is convex and decreasing), then we have
$$\int_\Omega f(\varrho_k^\tau)\, dx \geq \int_\Omega f(\varrho_{k+1}^\tau)\, dx + \tau \int_\Omega \left(f''(\varrho_{k+1}^\tau)|\nabla \varrho_{k+1}^\tau|^2 + \varrho_{k+1}^\tau f''(\varrho_{k+1}^\tau)\nabla\varrho_{k+1}^\tau \cdot \nabla V\right) dx. \tag{9}$$

(ii) Suppose $\varrho_0 \in L^p(\Omega)$, with $p < \infty$. Then, for any $k \in \{1, ..., N\}$, we have
$$\int_\Omega (\varrho_k^\tau)^p dx \geq \left(1 - \tau\frac{p(p-1)}{4} Lip(V)^2\right) \int_\Omega (\varrho_{k+1}^\tau)^p dx. \tag{10}$$

(iii) In particular, if τ is small enough (depending on V and p) and under the assumptions of (i) and (ii), the norm $||\varrho_t^\tau||_{L^p}$ grows at most exponentially in time for $p \in [1, \infty]$

4 Weak and strong convergence of the JKO scheme

In this section we use the bounds provided by Theorem 3.3 on the solutions of the JKO scheme to improve its convergence to the solution of the Fokker-Planck equation up to strong convergence in $L^2([0, T]; H^1(\Omega))$. As a starting point, we first consider weak convergence in the same space.

Proposition 4.1. *Under the same assumptions of Theorem 3.3 and $\varrho_0 \in L^2(\Omega)$, the curve ϱ^τ is uniformly bounded with respect to τ in $L^\infty([0, T]; L^2(\Omega))$ and $L^2([0, T]; H^1(\Omega))$. Moreover, if ϱ is the solution of (1), then $\varrho^\tau \rightharpoonup \varrho$ in $L^2([0, T]; H^1(\Omega))$.*

Proof. We use in the inequality (9) in Theorem 3.3 for the function $f(s) = s^2$. Then, for each $k \in \{0, , ..., N - 1\}$, $\varrho_{k+1}^\tau \in L^2(\Omega)$ and we have
$$\int_\Omega (\varrho_k^\tau)^2\, dx \geq \int_\Omega (\varrho_{k+1}^\tau)^2\, dx + 2\tau \int_\Omega \left(|\nabla \varrho_{k+1}^\tau|^2 + \varrho_{k+1}^\tau \nabla \varrho_{k+1}^\tau \cdot \nabla V\right) dx. \tag{11}$$

By the Young's inequality we have

$$\int_\Omega \varrho^\tau_{k+1} \nabla \varrho^\tau_{k+1} \cdot \nabla V \, dx \geq -\frac{1}{2}\int_\Omega |\nabla \varrho^\tau_{k+1}|^2 - \frac{1}{2}\int_\Omega (\varrho^\tau_{k+1})^2 |\nabla V|^2 \, dx$$

The estimate above implies

$$\int_\Omega (\varrho^\tau_k)^2 dx - \int_\Omega (\varrho^\tau_{k+1})^2 dx + \tau \operatorname{Lip}(V)^2 \int_\Omega (\varrho^\tau_{k+1})^2 dx \geq \tau \int_\Omega |\nabla \varrho^\tau_{k+1}|^2 dx.$$

We sum the inequalities above with respect to k, then

$$\int_\Omega (\varrho^\tau_0)^2 dx - \int_\Omega (\varrho^\tau_N)^2 dx + \operatorname{Lip}(V)^2 \sum_{k=0}^{N-1} \tau \int_\Omega (\varrho^\tau_{k+1})^2 dx \geq \sum_{k=0}^{N-1} \tau \int_\Omega |\nabla \varrho^\tau_{k+1}|^2 dx. \tag{12}$$

By the definition of the curve ϱ^τ we have

$$||\varrho^\tau||^2_{L^2([0,T];L^2(\Omega))} = \sum_{k=0}^{N-1} \tau \int_\Omega (\varrho^\tau_{k+1})^2 dx$$

and

$$||\varrho^\tau||^2_{L^2([0,T];H^1(\Omega))} = \sum_{k=0}^{N-1} \left(\tau \int_\Omega (\varrho^\tau_{k+1})^2 \, dx + \tau \int_\Omega |\nabla \varrho^\tau_{k+1}|^2 dx \right).$$

The inequality (10) in Theorem 3.3 (which is actually proven exactly as in the computations above, choosing better coefficients in the Young inequality so that the H^1 part disappears), provides uniform bounds on $||\varrho^\tau_t||_{L^2(\Omega)}$, which guarantees the bound in $L^\infty([0,T];L^2(\Omega))$ and, using (12), also in $L^2([0,T];H^1(\Omega))$.

Since $L^2([0,T];H^1(\Omega))$ is reflexive, when $\tau \to 0$ it is easy to find a weak limit $\rho \in L^2([0,T];H^1(\Omega))$ and this limit necessarily coincides with the limit in the Wasserstein sense, i.e. the unique solution of (1), which proves the last part of the statement. □

We now want to start proving strong convergences of ϱ^τ to ϱ. A first step will make use of the well-known Aubin-Lions lemma for time-dependent functions valued into function spaces, but we need to handle the time derivative. As the functions $t \mapsto \varrho^\tau$ are discontinuous, for simplicity (instead of evoking modified versions of the Aubin-Lions compactness criterion using functions which are BV in time), we define a new family of interpolations which help in obtaining the desired result.

Given $\varepsilon \in (0,1)$, we consider another curve $\varrho^{\tau,\varepsilon}$ such that $\varrho^{\tau,\varepsilon}_0 = \varrho_0$ and for $t \in (0,T]$

$$\varrho^{\tau,\varepsilon}_t = \begin{cases} \varrho^\tau_{k+1} & \text{if } t \in (k\tau, k\tau + (1-\varepsilon)\tau] \text{ and } k \in \{0,...,N-2\} \\ \varrho^\tau_{k+1} \frac{(k+1)\tau - t}{\varepsilon\tau} + \varrho^\tau_{k+2} \frac{t-(k+1)\tau+\varepsilon\tau}{\varepsilon\tau} & \text{if } t \in (k\tau + (1-\varepsilon)\tau, (k+1)\tau] \text{ and } k \in \{0,...,N-2\} \\ \varrho^\tau_N & \text{if } t \in ((N-1)\tau, N\tau] \end{cases}.$$

This curve is Lipschitz continuous in space and time and it time-derivative equals

$$\partial_t \varrho^{\tau,\varepsilon}_t = \sum_{k=0}^{N-2} \frac{\varrho^\tau_{k+2} - \varrho^\tau_{k+1}}{\varepsilon\tau} \cdot \mathbf{1}_{(k\tau+(1-\varepsilon)\tau,(k+1)\tau)} \text{ a.e. } t \in [0,T]. \tag{13}$$

Let $\mathcal{X} := \{\psi \in \mathrm{Lip}(\Omega) : \int_\Omega \psi = 0\}$ and $||\cdot||_{\mathrm{Lip}}$ be the Lipschitz semi-norm, which is actually a norm on this space where the average is prescribed, then it is easy to check that $(\mathcal{X}, ||\cdot||_{\mathrm{Lip}})$ forms a Banach space. We denote by $\mathcal{X}'$ the dual space of $\mathcal{X}$. This space includes all measures on Ω and it is well-known that we have $W_1(\mu,\nu) = ||\mu - \nu||_{\mathcal{X}'}$ for every $\mu, \nu \in \mathcal{P}(\Omega)$.

Proposition 4.2. *Under the same assumptions of Theorem 3.3 and $\varrho_0 \in L^2(\Omega)$, the following facts hold:*

1. *the curve $\varrho^{\tau,\varepsilon}$ converges to the solution of the Fokker-Planck equation uniformly in W_2 distance when $\tau \to 0$ (i.e. for every sequence $\varepsilon_j \in (0,1)$ and $\tau_j \to 0$ we do have this convergence).*
2. *the curve $\varrho^{\tau,\varepsilon}$ is uniformly bounded in $L^\infty([0,T]; L^2(\Omega))$ and $L^2([0,T]; H^1(\Omega))$ with respect to ε and τ.*
3. *we have $||\varrho^\tau - \varrho^{\tau,\varepsilon}||_{L^2([0,T];H^1(\Omega))} \leq C\sqrt{\varepsilon}$ for a constant C independent of τ and ε.*
4. *$\partial_t \varrho_t^{\tau,\varepsilon}$ is uniformly bounded in $L^1([0,T]; \mathcal{X}')$ with respect to ε and τ.*

Proof. Using the convexity of $\mu \mapsto W_2^2(\mu,\nu)$ it is easy to check that we have

$$W_2(\varrho_t^\tau, \varrho_t^{\tau,\varepsilon}) \leq \sup_k W_2(\varrho_k^\tau, \varrho_{k+1}^\tau) \leq \sqrt{\tau} \cdot \sqrt{2(J(\varrho_0) - \inf J)}. \tag{14}$$

This shows that, as $\tau \to 0$, the curves ϱ^τ and $\varrho^{\tau,\varepsilon}$ tend to the same limit (which is the solution of the Fokker-Planck equation), independently of ε.
Since the curve ϱ^τ is uniformly bounded with respect to τ in $L^\infty([0,T]; L^2(\Omega))$ by $C(T)||\varrho_0||_{L^2(\Omega)}$, and $\varrho^{\tau,\varepsilon}$ is obtained by convex combinations of values of ϱ_t^τ, the same bound is also true for $\varrho^{\tau,\varepsilon}$. Moreover, we have

$$||\varrho^\tau - \varrho^{\tau,\varepsilon}||_{L^2([0,T];L^2(\Omega))} = \sqrt{\frac{\varepsilon}{3}\sum_{k=0}^{N-2} \tau \int_\Omega (\varrho_{k+2}^\tau - \varrho_{k+1}^\tau)^2 dx} \leq C(T)\sqrt{\varepsilon} \cdot ||\varrho_0||_{L^2(\Omega)}. \tag{15}$$

Similar computations show

$$\begin{aligned} ||\varrho^\tau - \varrho^{\tau,\varepsilon}||_{L^2([0,T];H^1(\Omega))} &= \sqrt{\frac{\varepsilon}{3}\sum_{k=0}^{N-2} \tau ||\varrho_{k+2}^\tau - \varrho_{k+1}^\tau||_{H^1}^2} \\ &\leq C\sqrt{\varepsilon}\sqrt{\sum_{k=0}^{N-2} \tau||\varrho_{k+2}^\tau||_{H^1}^2 + \tau||\varrho_{k+1}^\tau||_{H^1}^2} \leq C\sqrt{\varepsilon}||\varrho^\tau||_{L^2([0,T];H^1(\Omega))}. \end{aligned}$$

By Proposition 4.1 and the last two estimates we have $||\varrho^\tau - \varrho^{\tau,\varepsilon}||_{L^2([0,T];H^1(\Omega))} \to 0$ as $\varepsilon \to 0$, uniformly in τ. We now look at the time derivative and we use the inequality $W_1 \leq W_2$ together with the identification of the distance W_1 as the dual norm of $\mathcal{X}$. We then have

$$\sum_{k=0}^{N-1} \frac{1}{\tau} ||\varrho_k^\tau - \varrho_{k+1}^\tau||_{\mathcal{X}'}^2 \leq 2(J(\varrho_0) - \inf J). \tag{16}$$

By the Cauchy–Schwarz inequality

$$\sum_{k=0}^{N-1} ||\varrho_k^\tau - \varrho_{k+1}^\tau||_{\mathcal{X}'} \leq \sqrt{N \sum_{k=0}^{N-1} ||\varrho_k^\tau - \varrho_{k+1}^\tau||^2_{\mathcal{X}'}} = \sqrt{\frac{T}{\tau} \sum_{k=0}^{N-1} ||\varrho_k^\tau - \varrho_{k+1}^\tau||^2_{\mathcal{X}'}} \leq \sqrt{2T(J(\varrho_0) - \inf J)}. \tag{17}$$

The estimate (17) and the expression (13) we obtain

$$||\partial_t \varrho^{\tau,\varepsilon}||_{L^1([0,T];\mathcal{X}')} = \sum_{k=0}^{N-2} \varepsilon\tau ||(\varrho_{k+2}^\tau - \varrho_{k+1}^\tau)/\varepsilon\tau||_{\mathcal{X}'} \leq \sum_{k=0}^{N-1} ||\varrho_k^\tau - \varrho_{k+1}^\tau||_{\mathcal{X}'} \leq \sqrt{2T(J(\varrho_0) - \inf J)}. \tag{18}$$

The estimate above shows that $\partial_t \varrho_t^{\tau,\varepsilon}$ is uniformly bounded in $L^1([0,T];\mathcal{X}')$. ☐

We state here a version of the Aubin-Lions-Simon Compactness Theorem which gives a compactness criterion in $L^p([0,T];B)$ for a Banach space B. Its proof can be found, for instance, in [26, Corollary 6, page 87].

Theorem 4.3. *Let X, B and Y be Banach spaces such that $X \subset B \subset Y$ with compact embedding $X \hookrightarrow B$. Let E be a bounded set in $L^q([0,T];B) \cap L^1_{loc}([0,T];X)$ for $1 < q \leq \infty$. If $\frac{\partial E}{\partial t} := \left\{ \frac{\partial f}{\partial t} : f \in E \right\}$ is bounded in $L^1_{loc}([0,T];Y)$, then E is relatively compact in $L^p([0,T];B)$ for all $p < q$.*

Corollary 4.4. *Under the same assumptions of Theorem 3.3 and supposing $\varrho_0 \in L^2(\Omega)$, then the curves ϱ^τ and $\varrho^{\tau,\varepsilon}$ converge strongly to the solution of the Fokker-Planck equation in $L^2([0,T];L^2(\Omega))$ as $\tau \to 0$.*

Proof. Let us consider Theorem 4.3 for the Banach spaces $X = H^1(\Omega), B = L^2(\Omega), Y = \mathcal{X}'$ and the indexes $p = 2, q = +\infty$. By the result of Proposition 4.2, for each $\varepsilon \in (0,1)$, the family of curves $E := (\varrho^{\tau,\varepsilon})_{\tau>0}$ satisfies Theorem 4.3, thus it is relatively compact in $L^2([0,T];L^2(\Omega))$. By the first part of Proposition 4.2, the curve $\varrho^{\tau,\varepsilon}$ converges to the unique solution of (1) in W_2 distance. Both the strong convergence in $L^2([0,T];L^2(\Omega))$ and the convergence in W_2 imply the weak convergence as measures in space-time, so we can conclude that the whole family of curves $E = (\varrho^{\tau,\varepsilon})_{\tau>0}$ converges strongly to this solution in $L^2([0,T];L^2(\Omega))$.
Let ϱ be the unique solution of (1). By $\lim_{\tau\to 0} ||\varrho - \varrho^{\tau,\varepsilon}||_{L^2([0,T];L^2(\Omega))} = 0$ and (15), we have

$$\begin{aligned} \lim_{\tau\to 0} ||\varrho - \varrho^\tau||_{L^2([0,T];L^2(\Omega))} &\leq \lim_{\tau\to 0} ||\varrho - \varrho^{\tau,\varepsilon}||_{L^2([0,T];L^2(\Omega))} + \lim_{\tau\to 0} ||\varrho^{\tau,\varepsilon} - \varrho^\tau||_{L^2([0,T];L^2(\Omega))} \\ &\leq \sqrt{\varepsilon} \cdot C(T) ||\varrho_0||_{L^2(\Omega)}. \end{aligned}$$

Letting $\varepsilon \to 0$, we get $\lim_{\tau\to 0} ||\varrho - \varrho^\tau||_{L^2([0,T];L^2(\Omega))} = 0$. ☐

The strong convergence in L^2 that we just obtained is not surprising, and is a necessary preliminary to pass on to the main goal of this section, which is a higher-order strong convergence result.

Theorem 4.5 (**Main Theorem I**)**.** *Let $[0,T]$ be a finite interval and Ω be a convex domain of $\mathbb{R}^d$ whose boundary $\partial\Omega$ is Lipschitz continuous. We assume V is Lipschitz continuous and the initial data ϱ_0 is in $\mathcal{P}(\Omega) \cap L^2(\Omega)$. As $\tau \to 0$, the curve ϱ^τ strongly converges to the solution of the Fokker-Planck equation (1) in $L^2([0,T];H^1(\Omega))$.*

Proof. Denote by ϱ_t the unique solution of (1). We saw in Proposition 2.2 that we have

$$\int_\Omega \varrho_T^2(x)\,dx - \int_\Omega \varrho_0^2(x)\,dx = -2\int_0^T\int_\Omega |\nabla\varrho_t(x)|^2\,dxdt - 2\int_0^T\int_\Omega \varrho_t(x)\,\nabla\varrho_t(x)\cdot\nabla V(x)\,dxdt. \tag{19}$$

By the estimate (11), we have

$$\begin{aligned}\int_\Omega (\varrho_T^\tau)^2\,dx - \int_\Omega \varrho_0^2\,dx = \int_\Omega (\varrho_N^\tau)^2 dx - \int_\Omega \varrho_0^2\,dx \leq \\ -2\sum_{k=0}^{N-1}\tau\int_\Omega \left(|\nabla\varrho_{k+1}^\tau|^2 + \varrho_{k+1}^\tau\,\nabla\varrho_{k+1}^\tau\cdot\nabla V\right)\,dx = \\ -2\int_0^T\int_\Omega \left(|\nabla\varrho_t^\tau|^2 + \varrho_t^\tau\,\nabla\varrho_t^\tau\cdot\nabla V\right)\,dxdt.\end{aligned} \tag{20}$$

The estimate (20) implies that

$$\limsup_{\tau\to 0}\int_\Omega (\varrho_T^\tau)^2\,dx - \int_\Omega \varrho_0^2\,dx \leq -2\limsup_{\tau\to 0}\int_0^T\int_\Omega \left(|\nabla\varrho_T^\tau|^2 + \varrho_t^\tau\,\nabla\varrho_t^\tau\cdot\nabla V\right)\,dxdt.$$

By Theorem 3.2, we know that for any $t\in[0,T]$,

$$W_2(\varrho_t^\tau,\varrho_t)\to 0$$

uniformly as $\tau\to 0$. The convergence in Wasserstein distance for every t implies

$$\liminf_{\tau\to 0}||\varrho_t^\tau||_{L^2(\Omega)} \geq ||\varrho_t||_{L^2(\Omega)}. \tag{21}$$

As a consequence of (21), we have

$$\limsup_{\tau\to 0}\int_\Omega (\varrho_T^\tau)^2\,dx \geq \int_\Omega (\varrho_T)^2\,dx. \tag{22}$$

Since ϱ_t^τ tends to ϱ_t strongly in $L^2([0,T];L^2(\Omega))$ and weakly in $L^2([0,T];H^1(\Omega))$, we have

$$\lim_{\tau\to 0}\int_0^T\int_\Omega \varrho_t^\tau\,\nabla\varrho_t^\tau\cdot\nabla V\,dxdt = \int_0^T\int_\Omega \varrho_t\,\nabla\varrho_t\cdot\nabla V\,dxdt \tag{23}$$

and

$$\liminf_{\tau\to 0}\int_0^T\int_\Omega |\nabla\varrho_t^\tau|^2\,dxdt \geq \int_0^T\int_\Omega |\nabla\varrho_t|^2\,dxdt. \tag{24}$$

Equations (22) and (23) show that

$$2\limsup_{\tau\to 0}\int_0^T\int_\Omega |\nabla\varrho_t^\tau|^2 dxdt \leq \int_\Omega \varrho_0^2 dx - \int_\Omega \varrho_T^2 dx - 2\int_0^T\int_\Omega \varrho_t\cdot\nabla\varrho_t\cdot\nabla V dxdt.$$

Using (19) and (24), we obtain

$$\lim_{\tau\to 0}\int_0^T\int_\Omega |\nabla\varrho_t^\tau|^2\,dxdt = \int_0^T\int_\Omega |\nabla\varrho_t|^2\,dxdt.$$

Together with the weak L^2 convergence of $\nabla\varrho^\tau$ to $\nabla\varrho$ this implies that ϱ^τ converges strongly to ϱ in $L^2([0,T];H^1(\Omega))$. ☐

5 Sobolev estimates on the JKO scheme

In this section we want to present important estimates in Sobolev spaces which were established in [12]. These type estimates require to consider some important inequalities in optimal transportation.

Lemma 5.1. *[**Five-gradients inequality**] Let Ω be a bounded, uniformly convex domain with C^2 boundary and $\varrho, g \in W^{1,1}(\Omega) \cap C^\alpha(\bar{\Omega})$, $0 < \alpha < 1$, be two strictly positive probability densities. Let $H \in C^2(\mathbb{R}^d)$ be a radially symmetric convex function. Then the following hold*

$$\int_\Omega (\nabla \varrho \cdot \nabla H(\nabla \varphi) + \nabla g \cdot \nabla H(\nabla \psi))\, dx$$
$$= \int_\Omega \varrho \mathrm{Tr}\{D^2 H(\nabla \varphi) \cdot (D^2\varphi)^2 \cdot (I - D^2 \varphi)^{-1}\} dx + \int_{\partial\Omega} (\varrho \nabla H(\nabla\varphi) \cdot \vec{n} + g \nabla H(\nabla \psi) \cdot \vec{n})\, d\mathcal{H}^{d-1} \geq 0, \tag{25}$$

where (φ, ψ) is a choice of Kantorovich potentials in the transport from ϱ to g.

Proof. Because of our assumptions on the densities, Caffarelli's regularity theory (see [5]-[10]) implies that the Kantorovich potentials, φ and ψ, are in $C^{2+\alpha}(\bar{\Omega})$, so we can use integration by part to write

$$\int_\Omega (\nabla \varrho \cdot H(\nabla \varphi) + \nabla g \cdot H(\nabla \psi))\, dx =$$
$$- \int_\Omega (\varrho \operatorname{div}[\nabla H(\nabla \varphi)] + g \operatorname{div}[\nabla H(\nabla \psi)])\ dx + \int_{\partial\Omega} (\varrho \nabla H(\nabla \varphi) \cdot \vec{n} + g \nabla H(\nabla \psi) \cdot \vec{n})\ d\mathcal{H}^{d-1}$$

Due to the radial symmetry of H we have

$$\nabla H(\nabla \psi) = -\nabla H(-\nabla \psi), \tag{26}$$

and $\nabla H(z)$ is a positive scalar multiple of z for every vector z. Since the gradients of the Kantorovich potentials $\nabla \varphi$ and $\nabla \psi$ calculated at boundary points are pointing outward Ω, (as a consequence of the optimal transport map $T(x) = x - \nabla \varphi(x) \in \Omega$ and its inverse $S(x) = x - \nabla \psi \in \Omega$) we see that the boundary term is non-negative:

$$\nabla H(\nabla \varphi(x)) \cdot \vec{n}(x) \geq 0 \text{ and } \nabla H(\nabla \psi(x)) \cdot \vec{n}(x) \geq 0 \tag{27}$$

We now consider the term with the divergence, and we have

$$\operatorname{div}[\nabla H(\nabla \varphi)] = \mathrm{Tr}(D^2 H(\nabla \varphi) \cdot D^2 \varphi), \qquad \operatorname{div}[\nabla H(\nabla \psi)] = \mathrm{Tr}(D^2 H(\nabla \psi) \cdot D^2 \psi).$$

We then use $g = T_\# \varrho$ to write

$$\int_\Omega g \mathrm{Tr}(D^2 H(\nabla \psi) \cdot D^2 \psi) = \int_\Omega \varrho \mathrm{Tr}(D^2 H(\nabla \psi \circ T) \cdot D^2 \psi \circ T).$$

Since H is radially symmetric and hence even, D^2H is also even, and using $\nabla\psi \circ T = -\nabla\varphi$ we can write

$$\int_\Omega \left(\varrho \operatorname{div}[\nabla H(\nabla\varphi)] + g \operatorname{div}[\nabla H(\nabla\psi)]\right) dx = \int_\Omega \varrho \mathrm{Tr}(D^2H(\varphi) \cdot (D^2\varphi + D^2\psi \circ T)).$$

We now use the relation $(id - \nabla\psi) \circ T = id$ with $T = id - \nabla\varphi$ to see that we have $D^2\psi \circ T = I - (I - D^2\varphi)^{-1}$ and then $D^2\varphi + D^2\psi \circ T = -(D^2\varphi)^2(I - D^2\varphi)^{-1}$, which proves the claimed formula. We are only left to see the positivity of the integral on Ω on the r.h.s. but the integrand is pointwise positive as a consequence of $D^2H \geq 0$, $(D^2\varphi)^2 \geq 0$ and $I - D^2\varphi \geq 0$ since $\frac{|x|^2}{2} - \varphi$ is convex. □

The result of previous lemma, that we proved in details, was first proven in [11] , but both in [11] and in [12] only the positivity was used, without considering the precise remainder terms. Indeed, even ignoring these terms and just using the positivity of $\int \nabla\varrho \cdot \nabla H(\nabla\varphi) + \nabla g \cdot \nabla H(\nabla\psi)$, it is possible to obtain estimates on the optimizers of the JKO scheme for the Fokker-Planck equation, as it is explained in [12].

More precisely, we have the following estimates.

Proposition 5.2. *Suppose Ω is a bounded and uniformly convex domain with C^2 boundary, and $V\colon \bar\Omega \to \mathbb{R}$ is Lipschitz continuous. Let $\varrho_0 \in W^{1,1}(\Omega) \cap C^\alpha(\bar\Omega)$, $0 < \alpha < 1$, (ϱ_k^τ) be the sequence obtained in the JKO scheme (6) and H be a convex radially symmetric function. Let (φ_k, ψ_k) denote the pair of Kantorovich potentials in the transport from ϱ_{k+1}^τ to ϱ_k^τ and T_k the corresponding optimal map, i.e. $T_k(x) = x - \nabla\varphi_k(x)$. Then we have*

$$\begin{aligned}\int_\Omega H(\nabla(\log \varrho_k^\tau + V)) d\varrho_k^\tau \;\; &\geq \;\; \int_\Omega H(\nabla(\log \varrho_{k+1}^\tau + V)) d\varrho_{k+1}^\tau \\ &+ \;\; \int_\Omega \nabla H\left(\frac{\nabla\varphi_k}{\tau}\right) \cdot (\nabla V - \nabla V \circ T_k) d\varrho_{k+1}^\tau + R, \end{aligned} \tag{28}$$

where $R \geq 0$ is the remainder term in the statement of Lemma 5.1, i.e.

$$\begin{aligned} R := \frac{1}{\tau}\int_\Omega \varrho \mathrm{Tr}\{D^2H\left(\frac{\nabla\varphi_k}{\tau}\right) \cdot (D^2\varphi_k)^2 \cdot (I - D^2\varphi_k)^{-1}\} dx \; + \\ \int_{\partial\Omega} (\varrho_{k+1}^\tau \nabla H\left(\frac{\nabla\varphi_k}{\tau}\right) \cdot \vec{n} + \varrho_k^\tau \nabla H\left(\frac{\nabla\psi_k}{\tau}\right) \cdot \vec{n}) d\mathcal{H}^{d-1}. \end{aligned}$$

In particular, when $H(z) = \frac{|z|^p}{p}$, $V \in V^2(\bar\Omega)$ and $\lambda \in \mathbb{R}$ such that $D^2V \geq \lambda I$, ignoring the positive terms and using $\nabla H(z) \cdot z = pH(z)$, we obtain

$$F_p(\varrho_k^\tau) \geq (1 + p\lambda\tau) F_p(\varrho_{k+1}^\tau).$$

Proof. The proof is exactly the same as in Lemma 5.1 of [12], with the only difference that the above statement requires to keep track of the positive remainder terms. We observe that at every step of the JKO scheme the obtained densities are smooth enough to justify the application of Lemma 5.1 of our paper (it is only by chance that the lemmas are both numbered 5.1 in the two different papers). □

We now proceed to some uniform estimates on the minimizers of the JKO scheme and on the corresponding potentials.

Proposition 5.3. *Suppose Ω is a bounded and uniformly convex domain with C^2 boundary, and $V \in C^2(\bar{\Omega})$. Let (ϱ_k^τ) be the sequence obtained in the JKO scheme (6), (φ_k, ψ_k) denote the pair of Kantorovich potentials in the transport from ϱ_{k+1}^τ to ϱ_k^τ and T_k the corresponding optimal map, i.e. $T_k(x) = x - \nabla\varphi_k(x)$. Suppose that ϱ_0 is bounded from below and above by positive constants and denote by a, b two constants such that $a \le \log \varrho_0 + V \le b$. Suppose moreover that we have $\varrho_0 \in W^{1,p}(\Omega)$ for some exponent $p > d$. Then we have:*

1. *For each k we have $a \le \log(\varrho_k^\tau) + V \le b$. In particular, all ϱ_k^τ are bounded from below and above by some uniform positive constants.*
2. *All the potentials φ_k satisfy $||id - T_k||_{L^\infty} = ||\nabla\varphi_k||_{L^\infty} \le C\tau^{1/(d+2)}$.*
3. *If τ is small enough (depending on V and p), then the values of $F_p(\varrho_k^\tau)$ are uniformly bounded by a constant depending on ϱ_0 and on T. In particular, the densities ϱ_k^τ are bounded in $C^{0,\alpha}$ for $\alpha = 1 - \frac{d}{p} > 0$.*
4. *All the potentials φ_k belong to $C^{2+\alpha}(\bar{\Omega})$ and $||\varphi_k||_{C^{2+\alpha}(\bar{\Omega})}$ is bounded by a uniform constant.*
5. *The potentials φ_k also satisfy $||D^2\varphi_k|| \le C\tau^\beta$ for a certain exponent $\beta > 0$.*

Proof.

1. The uniform estimates on ϱ^τ, already cited in [12], is the same as in Lemma 2.4 of [14].
2. Whenever μ, ν are two measures in a convex domain $\Omega \subset \mathbb{R}^d$ and T is the corresponding optimal map sending μ onto ν, if the density of μ is bounded from below by a constant $c_0 > 0$, then the following remarkable estimate is proven in [4] :

   $$||T - id||_{L^\infty} \le C(d, c_0) W_2(\mu, \nu)^{2/(d+2)}.$$

   If we combine this with

   $$W_2^2(\varrho_{k+1}^\tau, \varrho_k^\tau) \le 2\tau \left(J(\varrho_k^\tau) - J(\varrho_{k+1}^\tau)\right) \le C\tau.$$

   and the fact that the density ϱ_{k+1}^τ is bounded from below by a universal constant, we then have

   $$||\nabla\varphi_k||_{L^\infty(\Omega)}^{2+d} = ||id - T_k||_{L^\infty(\Omega)}^{2+d} \le C W_2^2(\varrho_{k+1}^\tau, \varrho_k^\tau) \le C\tau. \qquad (29)$$
3. First we note that, since ϱ_0 is bounded and $W^{1,p}$, we have $F_p(\varrho_0) < +\infty$. Then, Proposition 5.2 guarantees that $F_p(\varrho_k^\tau)$ grows at most exponentially (here we use the smallness assumption on τ, as we need $1 + p\lambda\tau > 0$, where λ is the lower bound for the second derivatives of V, which could be negative). We then obtain a uniform bound on $F_p(\varrho_k^\tau)$ and on $||\varrho_k^\tau||_{W^{1,p}}$, as a consequence of the uniform lower bound on ϱ_k^τ. The well-known injection of Sobolev spaces into Hölder spaces gives the rest of the claim.

4. The bound on $||\varphi_k||_{C^{2+\alpha}(\bar{\Omega})}$ is a consequence of Caffarelli's regularity theory for the Monge-Ampère equation (see [5]-[10]), once we have proven that the densities are uniformly bounded from above, from below, and in $C^{0,\alpha}$, when the domain is uniformly convex and C^2.

5. If we apply the interpolation inequality (30) to $u = \nabla\varphi$ in the case $\gamma = 1+\alpha$ and $\theta = 1$, we get
$$||\nabla\varphi_k||_{C^1(\bar{\Omega})} \leq c||\nabla\varphi_k||^{\frac{1}{1+\alpha}}_{C^{1+\alpha}(\bar{\Omega})} ||\nabla\varphi_k||^{\frac{\alpha}{1+\alpha}}_{C(\bar{\Omega})}.$$
Using (29) and the uniform bounds on $||\varphi_k||_{C^{2+\alpha}(\bar{\Omega})}$ we obtain the claim with $\beta = \frac{\alpha}{(1+\alpha)(2+d)}$.

□

We mentioned an interpolation inequality involving higher-order Hölder norms: here below is a precise statement, whose proofs can be found in [19, Proposition 1.1.3] .

Theorem 5.4 (**Interpolation inequality for Hölder continuous functions**). *Let $0 < \theta < \gamma$ and Ω be a open set in $\mathbb{R}^d$ with uniformly C^γ boundary. Then there exists a positive constant c depending on $\bar{\Omega}, \theta$ and γ such that*
$$||u||_{C^\theta(\bar{\Omega})} \leq c||u||^{\frac{\theta}{\gamma}}_{C^\gamma(\bar{\Omega})} ||u||^{1-\frac{\theta}{\gamma}}_{C(\bar{\Omega})} \tag{30}$$
for all $u \in C^\gamma(\bar{\Omega})$.

6 Strong convergence for second derivatives

In this section we are going to prove that the approximate solution which comes from the JKO scheme (6) convergences strongly to the solution of Fokker-Planck equation in higher order Sobolev spaces.

We assume Ω is a bounded and uniformly convex domain given by
$$\Omega = \{x \in \mathbb{R}^d : h(x) < 0\} \text{ and } \partial\Omega = \{x \in \mathbb{R}^d : h(x) = 0\}, \tag{31}$$
where h is a uniformly convex function of $C^2(\mathbb{R}^d)$ such that $D^2h \geq cId$ for some positive constant c. We also assume that $0 \in h(\mathbb{R}^d)$ is a regular value of h, i.e. $|\nabla h(x)| \neq 0$ for all $x \in \{x \in \mathbb{R}^d : h(x) = 0\}$. Then, it is easy to show that the boundary $\partial\Omega$ is a regular surface of class C^2 and its exterior unit normal vector $\vec{n}$ at $x \in \partial\Omega$ is defined by
$$\vec{n}(x) = \frac{\nabla h(x)}{|\nabla h(x)|}.$$

Let $\{\varrho^\tau_k\}^N_{k=0}$ be the sequence defined in the JKO scheme (6) and ϱ^τ_t be obtained from it by piecewise constant interpolation.

We consider the integral
$$F_2(\varrho^\tau_t) = \frac{1}{2}\int_\Omega \left|\frac{\nabla\varrho^\tau_t(x)}{\varrho^\tau_t(x)} + \nabla V(x)\right|^2 \varrho^\tau_t(x)\,dx. \tag{32}$$

and we will prove that the approximate curve constructed in the JKO scheme satisfies a discrete analogue of what is shown in Lemma 2.4. To get our purpose we mention here the trace theorem in the Sobolev spaces and some properties of Hausdorff measures.

Theorem 6.1 (Trace Theorem). *Assume $p \in [1, +\infty)$, Ω is a bounded open set of $\mathbb{R}^d$ and $\partial\Omega$ is C^1. Then there exists a bounded operator*

$$\mathrm{Tr} : W^{1,p}(\Omega) \to L^p(\partial\Omega)$$

such that

(i) $\mathrm{Tr}\, u = u\,|_{\partial\Omega}$ *for all* $u \in W^{1,p}(\Omega) \cap C(\bar{\Omega})$.

(ii)

$$||\mathrm{Tr}\, u||_{L^p(\partial\Omega)} \leq C||u||_{W^{1,p}(\Omega)}$$

for each $u \in W^{1,p}(\Omega)$, *with constant* C *depending on only* p *and* Ω.

Proof. See [17, section 18]. □

Proposition 6.2. *Let $\mathcal{H}^k$ be the $k-$dimensional Hausdorff measure on a metric space (X, d). If $Y \subset X$ is any set and $f, g \colon Y \to X$ satisfy $d(f(y), f(z)) \leq Cd(g(y), g(z))$ for all $y, z \in Y$, then $\mathcal{H}^k(f(A)) \leq C^k\mathcal{H}^k(g(A))$ for all $A \subset Y$.*

Proof. See [13, Proposition 11.18]. □

In the following computations, $\varepsilon(\tau)$ denotes any quantity which depends only on $\bar{\Omega}$, ϱ_0, V and τ such that $\varepsilon(\tau) \to 0$ as $\tau \to 0$.

Lemma 6.3. *Under the assumptions in Proposition 5.3 and the above assumptions on Ω we have*

$$F_2(\varrho_0) - F_2(\varrho_T^\tau) \geq \int_0^T \int_\Omega |D^2(\log \varrho_t^\tau(x) + V(x))|^2 \varrho_t^\tau(x)\, dxdt +$$

$$\int_0^T \int_\Omega (\nabla \log \varrho_t^\tau(x) + \nabla V(x))^T \cdot D^2V(x) \cdot (\nabla \log \varrho_t^\tau(x) + \nabla V(x)) \varrho_t^\tau(x)\, dxdt +$$

$$\int_0^T \int_{\partial\Omega} \frac{\varrho_t^\tau(x)}{|\nabla h(x)|} (\nabla \log \varrho_t^\tau(x) + \nabla V(x))^T \cdot D^2h(x) \cdot (\nabla \log \varrho_t^\tau(x) + \nabla V(x))\, d\mathcal{H}^{d-1}dt + \varepsilon(\tau). \quad (33)$$

Proof. We apply Proposition 5.2 for ϱ_{k-1}^τ and ϱ_k^τ ($k \in \{1, ..., N\}$) in the case of $H(z) = \frac{1}{2}|z|^2$, and get

$$\begin{aligned} \frac{1}{2}\int_\Omega \left|\frac{\nabla \varrho_{k-1}^\tau}{\varrho_{k-1}^\tau} + \nabla V\right|^2 d\varrho_{k-1}^\tau \quad - \quad & \frac{1}{2}\int_\Omega \left|\frac{\nabla \varrho_k^\tau}{\varrho_k^\tau} + \nabla V\right|^2 d\varrho_k^\tau \qquad (34) \\ \geq \quad & \frac{1}{\tau}\int_\Omega \nabla\varphi_k \cdot (\nabla V - \nabla V \circ T_k)\, d\varrho_k^\tau \\ & + \frac{1}{\tau}\int_\Omega \mathrm{tr}[(D^2\varphi_k)^2 \cdot (Id - D^2\varphi_k)^{-1}]\, d\varrho_k^\tau \\ & + \frac{1}{\tau}\int_{\partial\Omega} \left(\varrho_k^\tau \nabla\varphi_k \cdot \vec{n} + \varrho_{k-1}^\tau \nabla\psi_k \cdot \vec{n}\right) d\mathcal{H}^{d-1}. \end{aligned}$$

If we sum the discrete inequality above with respect to k, we obtain

$$\begin{aligned} F_2(\varrho_0) - F_2(\varrho_T^\tau) \quad \geq \quad & \frac{1}{\tau}\sum_{k=1}^{N} \int_\Omega \nabla\varphi_k \cdot (\nabla V - \nabla V \circ T_k)\, d\varrho_k^\tau \\ & + \frac{1}{\tau}\sum_{k=1}^{N} \int_\Omega \mathrm{tr}[(D^2\varphi_k)^2 \cdot (Id - D^2\varphi_k)^{-1}]\, d\varrho_k^\tau \\ & + \frac{1}{\tau}\sum_{k=1}^{N} \int_{\partial\Omega} \left(\varrho_k^\tau \nabla\varphi_k \cdot \vec{n} + \varrho_{k-1}^\tau \nabla\psi_k \cdot \vec{n}\right)\, d\mathcal{H}^{d-1}. \end{aligned} \tag{35}$$

Each of the three terms in the right-hand side of this inequality corresponds to one of the terms in the right-hand side of the inequality of the claim. In particular,

- the term $\frac{1}{\tau}\int_\Omega \nabla\varphi_k \cdot (\nabla V - \nabla V \circ T_k)\, d\varrho_k^\tau$ corresponds to the term involving D^2V; the main difficulty here consists in replacing the difference of ∇V in two points with a Hessian in a intermediate point, and estimating the error due to changing the point where D^2V is computed;

- the term $\frac{1}{\tau}\int_\Omega \mathrm{tr}[(D^2\varphi_k)^2 \cdot (Id - D^2\varphi_k)^{-1}]\, d\varrho_k^\tau$ corresponds to the one involving $|D^2(\log\varrho_t^\tau(x) + V(x))|^2$; here the main difficulty consists in getting rid of the matrix factor $(Id - D^2\varphi_k)^{-1}$ which should be very close to the identity.

- finally, the boundary term will also correspond to the boundary term of the claim; here we will use the characterization of $\partial\Omega$ as $\{h = 0\}$ and a second-order Taylor expansion of h, which also requires to bound the error in the expansion.

We consider now each term of the inequalities (34) and (35) separately. First, the equality $T_k = id - \nabla\varphi_k$ and the optimality condition (7) imply

$$\frac{1}{\tau}\nabla\varphi_k \cdot (\nabla V - \nabla V \circ T_k) = \frac{1}{\tau}\nabla\varphi_k^T \cdot D^2V \circ \xi \cdot \nabla\varphi_k,$$

where $\xi(x)$ is, for every point x, a suitable point on the line segment connecting x and $T(x)$, obtained from the Taylor expansion of V. This can be re-written as

$$\begin{aligned} \frac{1}{\tau}\nabla\varphi_k \cdot (\nabla V - \nabla V \circ T) \quad = \quad & \tau\nabla(\log\varrho_k^\tau + V)^T \cdot D^2V \cdot \nabla(\log\varrho_k^\tau + V) \\ & + \tau\nabla(\log\varrho_k^\tau + V)^T (D^2V \circ \xi - D^2V) \cdot \nabla(\log\varrho_k^\tau + V). \end{aligned}$$

Since ξ satisfies

$$|\xi - x| \leq |T(x) - x| = |\nabla\varphi_k(x)| \leq ||\nabla\varphi_k||_{L^\infty(\Omega)} \leq C\,\tau^{\frac{1}{2+d}},$$

the uniform continuity of D^2V implies $|D^2V \circ \xi - D^2V| = \varepsilon(\tau)$, while the uniform bounds on $||\log\varrho_k^\tau + V||_{W^{1,p}}$ with respect to k and τ show that

$$\frac{1}{\tau}\sum_{k=1}^{N} \int_\Omega \nabla\varphi_k \cdot (\nabla V - \nabla V \circ T_k)\, d\varrho_k^\tau =$$

$$\tau\sum_{k=1}^{N}\int_{\Omega}(\nabla\log\varrho_k^\tau(x)+\nabla V(x))^T\cdot D^2V(x)\cdot(\nabla\log\varrho_k^\tau(x)+\nabla V(x))\varrho_k^\tau(x)\,dx+\varepsilon(\tau)=$$

$$\int_0^T\int_{\Omega}(\nabla\log\varrho_t^\tau(x)+\nabla V(x))^T\cdot D^2V(x)\cdot(\nabla\log\varrho_t^\tau(x)+\nabla V(x))\varrho_t^\tau(x)\,dxdt+\varepsilon(\tau). \quad (36)$$

Let us now consider the matrix $D^2\varphi_k$. Point 5 in Proposition 5.3 insures that $D^2\varphi_k$ tends uniformly to zero as $\tau\to 0$. This enables us to write

$$\begin{aligned}\frac{1}{\tau}\int_{\Omega}\mathrm{tr}[(D^2\varphi_k)^2\cdot(Id-D^2\varphi_k)^{-1}]\,d\varrho_k^\tau &= \frac{(1+\varepsilon(\tau))}{\tau}\int_{\Omega}|D^2\varphi_k|^2\,d\varrho_k^\tau\\ &= (1+\varepsilon(\tau))\tau\int_{\Omega}|D^2(\log\varrho_k^\tau+V)|^2\,d\varrho_k^\tau,\end{aligned}$$

where we used again the optimality condition (7). Subsequently, we have

$$\frac{1}{\tau}\sum_{k=1}^{N}\int_{\Omega}\mathrm{tr}[(D^2\varphi_k)^2\cdot(Id-D^2\varphi_k)^{-1}]\,d\varrho_k^\tau=(1+\varepsilon(\tau))\int_0^T\int_{\Omega}|D^2(\log\varrho_t^\tau(x)+V(x))|^2\varrho_t^\tau(x)\,dxdt. \quad (37)$$

We now concentrate on the boundary integrals. Since the optimal transport map T_k sends $\bar\Omega$ to itself, for any $x\in\partial\Omega$, we have $T_k(x)=x-\nabla\varphi_k(x)\in\bar\Omega$. Thus,

$$0\geq h(x-\nabla\varphi_k(x))=h(x)-\nabla h(x)\cdot\nabla\varphi_k(x)+\frac{1}{2}(\nabla\varphi_k(x))^T\cdot D^2h(\zeta(x))\cdot\nabla\varphi_k(x)$$

for some point $\zeta(x)$ lying in the line connecting x and $T_k(x)=x-\nabla\varphi_k(x)$. If we use $h(x)=0$ and $\nabla h(x)=|\nabla h(x)|\vec{n}(x)$

$$|\nabla h(x)|\vec{n}(x)\cdot\nabla\varphi_k(x)=\nabla h(x)\cdot\nabla\varphi_k(x)\geq\frac{1}{2}(\nabla\varphi_k(x))^T\cdot D^2h(\zeta(x))\cdot\nabla\varphi_k(x).$$

If we multiply the inequality above by $\frac{\varrho_k^\tau(x)}{\tau|\nabla h(x)|}$ and integrate over $\partial\Omega$, we get

$$\frac{1}{\tau}\int_{\partial\Omega}\varrho_k^\tau(x)\nabla\varphi_k(x)\cdot\vec{n}(x)\,d\mathcal{H}^{d-1}\geq\frac{1}{2\tau}\int_{\partial\Omega}\frac{\varrho_k^\tau(x)}{|\nabla h(x)|}(\nabla\varphi_k(x))^T\cdot D^2h(\zeta(x))\cdot\nabla\varphi_k(x)\,d\mathcal{H}^{d-1}=$$

$$\frac{\tau}{2}\int_{\partial\Omega}\frac{\varrho_k^\tau(x)}{|\nabla h(x)|}(\nabla\log\varrho_k^\tau(x)+\nabla V(x))^T\cdot D^2h(\zeta(x))\cdot(\nabla\log\varrho_k^\tau(x)+\nabla V(x))\,d\mathcal{H}^{d-1}. \quad (38)$$

The uniform continuity of D^2h in $\bar\Omega$ and the uniform estimates

$$|\zeta(x)-x|\leq|x-T_k(x)|\leq C\,\tau^{\frac{1}{2+d}}$$

show that we have

$$D^2h(\zeta(x))=D^2h(x)+\varepsilon(\tau). \quad (39)$$

As a result of (39) and the uniform bounds on ϱ_k^τ, (38) equals

$$\frac{\tau}{2}\int_{\partial\Omega}\frac{\varrho_k^\tau(x)}{|\nabla h(x)|}(\nabla\log\varrho_k^\tau(x)+\nabla V(x))^T\cdot D^2h(x)\cdot(\nabla\log\varrho_k^\tau(x)+\nabla V(x))\,d\mathcal{H}^{d-1}+$$

$$\tau\varepsilon(\tau)||\nabla \log \varrho_k^\tau + \nabla V||^2_{L^2(\partial\Omega)}. \tag{40}$$

Since the inverse of the optimal transport map $S_k := T_k^{-1}$ is defined by $S_k(x) = x - \nabla\psi_k(x) \in \bar{\Omega}$, similar arguments as we have done above show that we have

$$\frac{1}{\tau}\int_{\partial\Omega} \varrho_{k-1}^\tau(x)\nabla\psi_k(x)\cdot\vec{n}(x)d\mathcal{H}^{d-1} \geq \frac{1}{2\tau}\int_{\partial\Omega}\frac{\varrho_{k-1}^\tau(x)}{|\nabla h(x)|}(\nabla\psi_k(x))^T\cdot D^2h(\zeta'(x))\cdot\nabla\psi_k(x)\,d\mathcal{H}^{d-1} \tag{41}$$

for some point $\zeta'(x)$ lying in the line connecting x and $T_k(x) = x - \nabla\varphi_k(x)$ and satisfying

$$|\zeta'(x) - x| \leq |x - T_k(x)| \leq C\,\tau^{\frac{1}{2+d}}. \tag{42}$$

By the equalities

$$-\nabla\psi_k(x) = S_k(x) - x = S_k(x) - T_k(S_k(x)) = \nabla\varphi_k(S_k(x)),$$

$$\nabla\varphi_k(S_k(x)) = -\tau\nabla\log\varrho_k^\tau(S_k(x)) - \tau\nabla V(S_k(x))$$

and the Monge-Ampére equation

$$\det(DS_k(x))\varrho_k^\tau(S_k(x)) = \varrho_{k-1}^\tau(x),$$

the right hand side of (41) equals

$$\frac{\tau}{2}\int_{\partial\Omega}\frac{\det(DS_k(x))\varrho_k^\tau(S_k(x))}{|\nabla h(x)|}([\nabla\log\varrho_k^\tau+\nabla V]\circ S_k(x))^T\cdot D^2h(\zeta'(x))\cdot([\nabla\log\varrho_k^\tau+\nabla V]\circ S_k(x))\,d\mathcal{H}^{d-1}. \tag{43}$$

Let $(S_k)_\sharp\mathcal{H}^{d-1}$ denotes the image measure of S_k on $\partial\Omega$ with respect to the Hausdorff measure defined by $(S_k)_\sharp\mathcal{H}^{d-1}(A) := \mathcal{H}^{d-1}(S_k^{-1}(A)) = \mathcal{H}^{d-1}(T_k(A))$ for all measurabele (w.r.t $\mathcal{H}^{d-1}$) $A \subset \partial\Omega$. Our assumptions provide that the maps $T_k, S_k\colon \partial\Omega \to \partial\Omega$ are homeomorphisms and hence $(S_k)_\sharp\mathcal{H}^{d-1}$ is well-defined. With the help of this image measure, (43) can be written in the following form

$$\frac{\tau}{2}\int_{\partial\Omega}\frac{\varrho_k^\tau(x)}{|\nabla h(T_k(x))|}(\nabla\log\varrho_k^\tau(x)+\nabla V(x))^T\cdot D^2h(\zeta'(T_k(x)))\cdot(\nabla\log\varrho_k^\tau(x)+\nabla V(x))\,d(S_k)_\sharp\mathcal{H}^{d-1}. \tag{44}$$

On the one hand, the estimate $|T_k(x)-T_k(y)| \leq \mathrm{Lip}(T_k)\,|x-y|$ for all $x, y \in \partial\Omega$ and Theorem 6.2 show that

$$(S_k)_\sharp\mathcal{H}^{d-1}(A) = \mathcal{H}^{d-1}(T_k(A)) \leq (\mathrm{Lip}(T_k))^{d-1}\mathcal{H}^{d-1}(A)$$

for all $A \subset \partial\Omega$. On the other hand, the estimate $|x - y| = |S_k(T_k(x)) - S_k(T_k(y))| \leq \mathrm{Lip}(S_k)\,|T_k(x) - T_k(y)|$ for all $x, y \in \partial\Omega$ and again Theorem 6.2 show that

$$\mathcal{H}^{d-1}(A) \leq (\mathrm{Lip}(S_k))^{d-1}\mathcal{H}^{d-1}(T_k(A)) = (\mathrm{Lip}(S_k))^{d-1}(S_k)_\sharp\mathcal{H}^{d-1}(A).$$

Therefore, we have

$$\frac{1}{(\mathrm{Lip}(S_k))^{d-1}}\mathcal{H}^{d-1}(A) \leq (S_k)_\sharp\mathcal{H}^{d-1}(A) \leq (\mathrm{Lip}(T_k))^{d-1}\mathcal{H}^{d-1}(A) \tag{45}$$

for all $A \subset \partial\Omega$. Using $DT_k(x) = I - D^2\varphi_k(x)$, $DS_k(x) = I - D^2\psi_k(x) = (DT_k)^{-1} \circ S_k$ together with $||D^2\varphi_k|| = \varepsilon(\tau)$ we have

$$\mathrm{Lip}(S_k) = 1 + \varepsilon(\tau), \quad \mathrm{Lip}(T_k) = 1 + \varepsilon(\tau). \tag{46}$$

The estimate (45) and (46) show that

$$(S_k)_\sharp \mathcal{H}^{d-1}(A) = (1 + \varepsilon(\tau))\, \mathcal{H}^{d-1}(A) \tag{47}$$

for all $A \subset \partial\Omega$. The regularity of h and point 2 in Proposition 5.3 provide

$$\frac{1}{|\nabla h(T_k(x))|} = \frac{1}{|\nabla h(x)|}\frac{|\nabla h(x)|}{|\nabla h(T_k(x))|} = (1 + \varepsilon(\tau))\frac{1}{|\nabla h(x)|} \tag{48}$$

and

$$D^2 h(\zeta'(T_k(x))) = D^2 h(x) + \varepsilon(\tau). \tag{49}$$

By considering (44), (47), (48) and (49), the right hand side of (41) equals

$$\frac{\tau}{2}\int_{\partial\Omega} \frac{\varrho_k^\tau(x)}{|\nabla h(x)|}(\nabla \log \varrho_k^\tau(x) + \nabla V(x))^T \cdot D^2 h(x) \cdot (\nabla \log \varrho_k^\tau(x) + \nabla V(x))\, d\mathcal{H}^{d-1} +$$

$$\tau\varepsilon(\tau)||\nabla(\log \varrho_k^\tau + V)||^2_{L^2(\partial\Omega)}. \tag{50}$$

Since h is convex function, the integrals (38) and (41) are positive. The positivity of these integrals and the estimates (36) and (37) provide the inequality

$$F_2(\varrho_0) - F_2(\varrho_T^\tau) \geq \int_0^T \int_\Omega |D^2(\log \varrho_t^\tau(x) + V(x))|^2 \varrho_t^\tau(x)\, dx dt +$$

$$\int_0^T \int_\Omega (\nabla \log \varrho_t^\tau(x) + \nabla V(x))^T \cdot D^2 V(x) \cdot (\nabla \log \varrho_t^\tau(x) + \nabla V(x)) \varrho_t^\tau(x)\, dx dt$$

We proved in the previous sections that ϱ^τ is bounded in $L^2([0,T]; H^1(\Omega))$, which provides a bound on the last integral term. This, together with the lower bounds on ρ^τ, implies that $\log \varrho_t^\tau + V$ is uniformly bounded in $L^2([0,T]; H^2(\Omega))$. Theorem 6.1 lets us conclude that $\nabla(\log \varrho^\tau + V)$, $i \in \{1, ..., d\}$, is uniformly bounded in $L^2([0,T]; L^2(\partial\Omega))$.

Considering (38), (40), (41), (50) and the fact that $\log \varrho_t^\tau + V$ and $\nabla \log \varrho_t^\tau + \nabla V$ are bounded in $L^2([0,T]; H^2(\Omega))$ and $L^2([0,T]; L^2(\partial\Omega))$ respectively, we conclude that

$$\frac{1}{\tau}\sum_{k=1}^N \int_{\partial\Omega} \left(\varrho_k^\tau \nabla\varphi_k \cdot \vec{n} + \varrho_{k-1}^\tau \nabla\psi_k \cdot \vec{n}\right) d\mathcal{H}^{d-1} \geq$$

$$\sum_{k=1}^N \tau \int_{\partial\Omega} \frac{\varrho_k^\tau(x)}{|\nabla h(x)|}(\nabla \log \varrho_k^\tau(x) + \nabla V(x))^T \cdot D^2 h(x) \cdot (\nabla \log \varrho_k^\tau(x) + \nabla V(x))\, d\mathcal{H}^{d-1} + \varepsilon(\tau) =$$

$$\int_0^T \int_{\partial\Omega} \frac{\varrho_t^\tau(x)}{|\nabla h(x)|}(\nabla \log \varrho_t^\tau(x) + \nabla V(x))^T \cdot D^2 h(x) \cdot (\nabla \log \varrho_t^\tau(x) + \nabla V(x))\, d\mathcal{H}^{d-1} dt + \varepsilon(\tau). \tag{51}$$

The estimates (36), (37) and (51) give the desired estimate. □

Lemma 6.4. *Let D be bounded Lipschitz domain of $\mathbb{R}^d$, $\{u_k\}_{k=1}^{\infty}$ be a sequence in $L^2([0,T];H^2(D))$ and bounded in $L^{\infty}([0,T]\times D)$. If there exists $u \in L^{\infty}([0,T];C^2(D))$ such that $u_k \to u$ strongly in $L^2([0,T];H^2(D))$, then, for any $f \in C^2(\mathbb{R})$, $\{f(u_k)\}_{k=1}^{\infty}$ converges to $f(u)$ in $L^2([0,T];H^2(D))$.*

Proof. Using our assumptions and the Gagliardo-Nirenberg inequality (see [3, section 9])

$$||\nabla v||^4_{L^4(D)} \le C||v||^2_{H^2(D)}||v||^2_{L^{\infty}(D)} \tag{52}$$

applied to $v = u_k - u$ we obtain $u_k \to u$ in $L^4([0,T];W^{1,4}(D))$. A simple computation shows

$$D^2(f(u_k)) = f'(u_k)D^2u_k + f''(u_k)\nabla u_k \otimes \nabla u_k$$

and the L^2 convergence of this matrix-valued function to $D^2(f(u)) = f'(u)D^2u + f''(u)\nabla u \otimes \nabla u$ is due to the following facts:

- $D^2u_k \to D^2u$ in $L^2([0,T]\times D)$;
- $\nabla u_k \otimes \nabla u_k \to \nabla u \otimes \nabla u$ in $L^2([0,T]\times D)$;
- both $f''(u_k)$ and $f'(u_k)$ converge a.e. (to $f''(u)$ and $f'(u)$, respectively) as a consequence of the convergence of u_k to u; moreover, these terms are bounded in L^{∞} as a consequence of the regularity of f and of the L^{∞} bound on u_k. □

We are now ready to prove our main theorem.

Theorem 6.5 (Main Theorem II). *Suppose $0 < T < +\infty$, Ω is a bounded and uniformly convex domain given by (31). Let $V \in C^2(\bar{\Omega})$, $\varrho_0 \in W^{1,p}(\Omega)$ for $p > d$, $\lambda \le \varrho_0 \le \Lambda$ for some strictly positive constants λ, Λ and ϱ be the solution of the Fokker-Planck equation (1). Then, $\varrho^{\tau} \to \varrho$ strongly in $L^2([0,T];H^2(\Omega))$ as $\tau \to 0$.*

Proof. In Lemma 6.3 and its proof, we get the estimate (33) and showed that $\log \varrho_t^{\tau} + V$ is uniformly bounded in $L^2([0,T];H^2(\Omega))$ with respect to τ. Since the space $L^2([0,T];H^2(\Omega))$ is reflexive and ϱ^{τ} converges strongly in $L^2([0,T];H^1(\Omega))$ to ϱ (see Theorem 4.5), we obtain that $\log \varrho^{\tau} + V$ converges weakly to $\log \varrho + V$ in $L^2([0,T];H^2(\Omega))$. This also implies that $\nabla \log \varrho^{\tau} + \nabla V$ converges weakly to $\nabla \log \varrho + \nabla V$ in $L^2([0,T];L^2(\partial\Omega))$. If τ tends to zero in (33), then by the lower semicontinuity of the norm for the weak convergence, we obtain

$$F_2(\varrho_0) - F_2(\varrho_T) \ge F_2(\varrho_0) - \liminf_{\tau\to 0} F_2(\varrho_T^{\tau})) \ge \limsup_{\tau\to 0} \int_0^T \int_{\Omega} |D^2(\log \varrho_t^{\tau}(x) + V(x))|^2 \varrho_t^{\tau}(x)\, dxdt +$$

$$\limsup_{\tau\to 0} \int_0^T \int_{\Omega} (\nabla \log \varrho_t^{\tau}(x) + \nabla V(x))^T \cdot D^2V(x) \cdot (\nabla \log \varrho_t^{\tau}(x) + \nabla V(x))\varrho_t^{\tau}(x)\, dxdt +$$

$$\limsup_{\tau\to 0} \int_0^T \int_{\partial\Omega} \frac{\varrho_t^{\tau}(x)}{|\nabla h(x)|} (\nabla \log \varrho_t^{\tau}(x) + \nabla V(x))^T \cdot D^2h(x) \cdot (\nabla \log \varrho_t^{\tau}(x) + \nabla V(x))\, d\mathcal{H}^{d-1}dt \ge$$

$$\limsup_{\tau\to 0} \int_0^T \int_{\Omega} |D^2(\log \varrho_t^{\tau}(x) + V(x))|^2 \varrho_t^{\tau}(x)\, dxdt +$$

$$\int_0^T \int_\Omega (\nabla \log \varrho_t(x) + \nabla V(x))^T \cdot D^2 V(x) \cdot (\nabla \log \varrho_t(x) + \nabla V(x)) \varrho_t(x)\, dxdt +$$

$$\int_0^T \int_{\partial\Omega} \frac{\varrho_t(x)}{|\nabla h(x)|} (\nabla \log \varrho_t(x) + \nabla V(x))^T \cdot D^2 h(x) \cdot (\nabla \log \varrho_t(x) + \nabla V(x))\, d\mathcal{H}^{d-1} dt$$

The estimate above and Corollary 2.6 imply

$$\int_0^T \int_\Omega |D^2(\log \varrho_t(x) + V(x))|^2 \varrho_t(x)\, dxdt \geq \limsup_{\tau \to 0} \int_0^T \int_\Omega |D^2(\log \varrho_t^\tau(x) + V(x))|^2 \varrho_t^\tau(x)\, dxdt.$$

Because of weak convergence of $\log \varrho^\tau + V$ to $\log \varrho + V$ in $L^2([0,T]; H^2(\Omega))$ we have

$$\liminf_{\tau \to 0} \int_0^T \int_\Omega |D^2(\log \varrho_t^\tau(x) + V(x))|^2 \varrho_t^\tau(x)\, dxdt \geq \int_0^T \int_\Omega |D^2(\log \varrho_t(x) + V(x))|^2 \varrho_t(x)\, dxdt.$$

Therefore, we have

$$\lim_{\tau \to 0} \int_0^T \int_\Omega |D^2(\log \varrho_t^\tau(x) + V(x))|^2 \varrho_t^\tau(x)\, dxdt = \int_0^T \int_\Omega |D^2(\log \varrho_t(x) + V(x))|^2 \varrho_t(x)\, dxdt. \tag{53}$$

The limit above and the upper and lower boundson ϱ^τ show that $D^2 \log \varrho^\tau$ is bounded in $L^2([0,T]; L^2(\Omega))$. Since $L^2([0,T]; L^2(\Omega))$ is reflexive and ϱ^τ converges strongly in $L^2([0,T]; H^1(\Omega))$ to ϱ, then $D^2 \log \varrho^\tau$ converges weakly to $D^2 \log \varrho$ in $L^2([0,T]; L^2(\Omega))$. By adding $D^2 V$ and multiplying times $\sqrt{\varrho^\tau}$, which converges a.e. to $\sqrt{\varrho}$ and is bounded by a constant, we also have weak convergence in $L^2([0,T]; L^2(\Omega))$ of $\sqrt{\varrho^\tau} D^2(\log \varrho^\tau + V)$ to $\sqrt{\varrho} D^2(\log \varrho + V)$. Yet, this convergence becomes strong because of the convergence of the norm in (53). We can then multiply times $(\varrho^\tau)^{-1/2}$ and subtract $D^2 V$ and obtain strong convergence in $L^2([0,T]; H^2(\Omega))$ for $\log \varrho^\tau$ to $\log \varrho$.

We then apply Lemma 6.4 to obtain $\varrho^\tau \to \varrho$ in $L^2([0,T]; H^2(\Omega))$. □

References

[1] L.Ambrosio, N.Gigli, G.Savaré, *Gradient flows in Metric Flows and in the Space of Probability Measures,* Lectures in Mathematics, Birkhäuser, 2005.

[2] D. Bakry, I. Gentil, M. Ledoux, *Analysis and geometry of Markov diffusion operators*, Grundlehren der mathematischen Wissenschaften (GL, volume 348), Springer, 2014.

[3] H. Brezis, *Functional Analysis, Sobolev Spaces and Partial Differential Equations,* Springer-Verlag New York, 2011.

[4] G. Bouchitté, C. Jimenez, M. Rajesh, A new L^∞ estimate in optimal mass transport, *Proc.Am.Math.Soc.* 135, 3525-3535 (2007).

[5] L. A. Caffarelli, Interior $W^{2,p}$ estimates for solutions of the Monge-Ampere equation. *Ann. of Math.* (2) 131, 1 (1990), 135-150.

[6] L. A. Caffarelli, A localization property of viscosity solutions to the Monge-Ampere equation and their strict convexity. *Ann. of Math.* (2) 131, 1 (1990), 129-134.

[7] L. A. Caffarelli, Some regularity properties of solutions of Monge-Ampere equation. *Comm. Pure Appl. Math.* 44, 8-9 (1991), 965-969.

[8] L. A. Caffarelli, Boundary regularity of maps with convex potentials. *Comm.Pure Appl. Math.* 45,9 (1992), 1141-1151.

[9] L. A. Caffarelli, The regularity of mappings with a convex potential. *J. Amer.Math. Soc.* 5, 1 (1992), 99-104.

[10] L. A. Caffarelli, Boundary regularity of maps with convex potentials. II. *Ann. of Math.* (2) 144, 3 (1996), 453-496.

[11] G. De Philippis, A.-R. Mészáros, F. Santambrogio, B. Velichkov, BV estimates in optimal transportation and applications, *Archive for Rational Mechanics and Analysis* 219 (2), 829-860, 2016.

[12] S. Di Marino, F. Santambrogio, JKO estimates in linear and non-linear Fokker-Planck equations, and Keller-Segel: L^p and Sobolev bounds, *Ann. IHP, Anal. Non Linéaire*, 39 (2022), 1485-1517

[13] G. B. Folland, *Real Analysis: modern techniques and their applications.* Wiley, New York, second edition, 1999.

[14] M. Iacobelli, F. S. Patacchini, F. Santambrogio, Weighted ultrafast diffusion equations: from well-posedness to long-time behaviour, *Archive for Rational Mechanics and Analysis* 232 (3), 1165-1206, 2019.

[15] R. Jordan, D. Kinderlehrer, F. Otto, The variational formulation of the Fokker-Plank equation, *SIAM J. Math. Anal.* 29,1-17,1998.

[16] O.A. Ladyzhenskaja, V.A. Solonnikov, N.N. Ural'ceva,, *Linear and quasilinear equations of parabolic type,* Nauka, Moskow 1967 (Russian). English transl.: Transl. Math. Monographs, AMS, Providence (1968).

[17] G. Leoni, *A First Course in Sobolev Spaces,* American Mathematical Society, Volume 105, 2009.

[18] J.L. Lions, E. Magenes, *Problèmes aux limites non homogènes et applications,* Dunod, Paris, vol. 1, 2 (1968), vol. 3 (1970).

[19] A. Lunardi, *Analytic semigroups and optimal regularity in parabolic problems.* Birkhäuser Verlag, Basel (1995). xviii+424 pp. 2013 reprint of the 1995 original, Modern Birkhäuser Classics. Birkhäuser/Springer Basel AG, Basel, 1995. xviii+424 pp.

[20] R. J. McCann A convexity principle for interacting gases. *Adv. Math.* 128 (1), 153–159, 1997.

[21] F. Otto, The geometry of dissipative evolution equations: The porous medium equation, *Comm. Partial Differential Equations*, 26, 101–174, 2011.

[22] J.-D. Benamou, G. Carlier, M. Laborde, An augmented Lagrangian approach to Wasserstein gradient flows and applications, *ESAIM: Proocedings and surveys*, 2016, Vol. 54, p. 1–17

[23] G. Peyré. Entropic Wasserstein Gradient Flows. *SIAM Journal on Imaging Sciences,* 2015, 8 (4), 2323–2351.

[24] F. Santambrogio, *Optimal transport for Applied Mathematicians,* Progress in Nonlinear Partial Differential Equations and their Applications 87, Birkhäuser Basel 2015.

[25] F. Santambrogio {Euclidean, Metric, and Wasserstein} Gradient Flows: an overview, *Bull. Math. Sci.*, 7 (1), 87–154, 2017.

[26] J. Simon, Compact sets in the space $L^p(0,T;B)$. *Annali di Matematica Pura ed Applicata*, 146, pages 65–96, 1986.

[27] R. Temam, *Navier-Stokes equations* Studies in Mathematics ans Applications,,Vol 2.

[28] C. Villani, *Optimal transport. Old and New.* Grundlehren der Mathematischen Wissenschaften, 338.Springer-Verlag, Berlin, 2009.

[29] C. Villani, *Topics in Optimal Transportation.* Graduate Studies in Mathematics, American Mathematical Society, Providence,2003.