

Lipschitz estimates on the JKO scheme for the Fokker-Planck equation on bounded convex domains

Vincent Ferrari* and Filippo Santambrogio^{†‡}

July 15, 2020

Abstract

Given a semi-convex potential V on a convex and bounded domain Ω , we consider the Jordan-Kinderlehrer-Otto scheme for the Fokker-Planck equation with potential V , which defines, for fixed time step $\tau > 0$, a sequence of densities $\rho_k \in \mathcal{P}(\Omega)$. Supposing that V is α -convex, i.e. $D^2V \geq \alpha I$, we prove that the Lipschitz constant of $\log \rho + V$ satisfies the following inequality: $\text{Lip}(\log(\rho_{k+1}) + V)(1 + \alpha\tau) \leq \text{Lip}(\log(\rho_k) + V)$. This provides exponential decay if $\alpha > 0$, Lipschitz bounds on bounded intervals of time, which is coherent with the results on the continuous-time equation, and extends a previous analysis by Lee in the periodic case.

1 Introduction

This short paper is concerned with the parabolic PDE

$$\partial_t \rho - \Delta \rho - \nabla \cdot (\rho \nabla V) = 0, \quad (1)$$

which is known as Fokker-Planck equation. It includes a linear diffusion effect (corresponding to the Laplacian in the above equation) and advection by a drift, which is supposed to be of gradient type. The function V is supposed here to be given, independent of time and of ρ , and sufficiently smooth. We consider this equation on a bounded domain Ω , and we complete it with no-flux boundary conditions

$$(\nabla \rho + \rho \nabla V) \cdot \mathbf{n} = 0.$$

Starting from the seminal paper [10], it is well-known that this equation is the gradient flow of the functional

$$\rho \mapsto \mathcal{E}(\rho) := \int \rho \log \rho + \int V d\rho$$

with respect to the Wasserstein distance W_2 . This distance is the one defined from the minimal cost in the optimal transport with quadratic cost. The reader can refer to [15, 13] for more details about optimal transport and Wasserstein distances, and [1, 14] for gradient flows in this setting.

The same paper [10] also provides a time-discretized approach to the above equation, now known as the Jordan-Kinderlehrer-Otto scheme (JKO). It consists in fixing a time step $\tau > 0$ and then iteratively solving

$$\min \left\{ \frac{W_2^2(\rho, \rho_k)}{2\tau} + \mathcal{E}(\rho) : \rho \in \mathcal{P}(\Omega) \right\},$$

*École Normale Supérieure; 45 Rue d'Ulm, 75005 Paris, FRANCE, email: vferrari@clipper.ens.fr

†Institut Camille Jordan, Université Claude Bernard - Lyon 1; 43 boulevard du 11 novembre 1918, 69622 Villeurbanne cedex, FRANCE, email: santambrogio@math.univ-lyon1.fr

‡Institut Universitaire de France (IUF)

defining ρ_{k+1} as the unique minimizer of the above optimization problem. The sequence obtained in this way represents a good approximation of the continuous-time solution, in the sense that ρ_k is close to $\rho(k\tau, \cdot)$.

For this equation, which is linear and quite simple, using the JKO scheme as a tool to approximate the solution (for numerical or theoretical purposes) is not necessary, but the same scheme can be useful for equations which are gradient flow of more involved functionals. In order to prove convergence, it is often necessary to prove compactness estimates on the solutions of the scheme. Among the bounds that one can prove, the most striking ones are those where a certain norm decreases when passing from ρ_k to ρ_{k+1} . The case where a same norm increases when passing from ρ_k to ρ_{k+1} , but no more than a quantity of the order of τ , so that the bound can be iterated, is also interesting. We cite for instance L^∞ bounds (see for instance Section 7.4.1 in [13]), BV bounds (see [6]), which are valid when $V = 0$ and the entropy $\int \rho \log \rho$ is replaced by an arbitrary convex penalization $\int f(\rho)$.

We are interested in this paper in Lipschitz bounds and more precisely in the quantity $\text{Lip}(\log \rho + V)$. Using for instance the well-known Bakry-Emery theory (see [2]) this quantity decreases in time along solutions of the continuous equation (which is a re-writing of the transport-diffusion equation $\partial_t u = \Delta u - \nabla V \cdot \nabla u$ if one sets $u = \rho e^V$) if V is convex. If $D^2V \geq \alpha I$ one can see that $e^{\alpha t} \text{Lip}(\log \rho_t + V)$ decreases in time, which implies exponential convergence as $t \rightarrow \infty$ if $\alpha > 0$ or controlled growth if $\alpha < 0$.

It is highly remarkable that these bounds can exactly be translated into the JKO setting, i.e. they also holds along the discrete-time iterations of the scheme. This was already observed in [11], and the analysis is very similar to ours, but was only restricted to the periodic setting. The main object of the present paper is to extend the analysis of [11] to a case with boundary. This recalls what is done in [7], where a similar estimate is performed on the quantity

$$\int \left| \frac{\nabla \rho}{\rho} + \nabla V \right|^p d\rho,$$

which is proven to decrease along iterations if V is convex and the domain is also convex (this was obtained thanks to the so-called *five gradients inequality* introduced in [6]). Taking the power $1/p$ and sending $p \rightarrow \infty$ also provides an estimate on $\|\frac{\nabla \rho}{\rho} + \nabla V\|_{L^\infty}$. The same computation unfortunately does not work if we only have $D^2V \geq \alpha I$ with $\alpha < 0$, since a similar estimate is only possible as soon as $1 + \alpha\tau p > 0$, which prevents to send $p \rightarrow \infty$ without sending $\tau \rightarrow 0$. In this sense, it is not an estimate on the JKO scheme. Moreover, obtaining an L^∞ bound (i.e. an estimate on a maximal value) out of a limit of integral quantities and using the five gradients inequality is a complicated procedure which can be highly simplified, which is the scope of this note.

As in [11], we will obtain the desired estimate by taking the maximal point of the squared norm of a gradient and using the optimality conditions together with the Monge-Ampère equation on the optimal transport map from ρ_{k+1} to ρ_k . Even if we believe that the presentation that we give here is lighter and if the regularity assumptions on the data have been reduced to the minimal ones, the computations that we present are exactly the same as those of [11] in the case of a torus. The added value of our work is the attention to the boundary. The key point are

- a lemma which guarantees that $\max |\nabla \varphi|$ is not attained on the boundary when φ is the Kantorovich potential between two smooth densities on a ball;
- an extension-approximation procedure to pass from an arbitrary convex domain to a larger ball, and to regularize the densities.

2 Main result

Let Ω be a bounded closed convex domain in \mathbb{R}^d , V a non-negative Lipschitz function on Ω , $\tau > 0$ and g a probability density on Ω (i.e. $g \in L^1(\Omega)$, $g \geq 0$ and $\int_\Omega g = 1$). We study the minimizer of the functional $\mathcal{F}_{g,V}^\tau$ on $\mathcal{P}(\Omega)$ (the set of probabilities measures on Ω) given by :

$$\mathcal{F}_{g,V}^\tau(\rho) = \frac{W_2^2(\rho, g)}{2\tau} + \mathcal{E}(\rho) = \frac{W_2^2(\rho, g)}{2\tau} + \int_\Omega \rho \log \rho + \int_\Omega V d\rho \quad (2)$$

where W_2 is the Wasserstein distance on $\mathcal{P}(\Omega)$.

We know that this functional admits a unique minimizer, denoted ρ in the rest of the paper. Indeed, \mathcal{F} is the sum of three convex functionals, and the entropy part is strictly convex, which provides uniqueness. We denote $T(x) = x - \nabla\varphi(x)$ the optimal transport map between ρ and g (which is the gradient of the convex function $x \mapsto \frac{|x|^2}{2} - \varphi(x)$, see [3]), where φ is the corresponding Kantorovitch potential. We recall that the minimizer ρ satisfies the following properties (we refer to Chapter 8 in [13]):

ρ is Lipschitz and strictly positive on Ω

$$\log \rho + V + \frac{\varphi}{\tau} = \text{constant on } \Omega \quad (3)$$

$$\det(I - D^2\varphi(x)) = \frac{\rho(x)}{g(T(x))} \quad (4)$$

Equation (4) is the Monge-Ampère equation for the transport between ρ and g and (3) is the optimality condition for ρ . We will prove the following:

Theorem 2.1. *Suppose that $\Omega \subset \mathbb{R}^d$ is a bounded convex set, that g is bounded from below by a strictly positive constant and Lipschitz continuous on Ω , and that V is Lipschitz continuous and uniformly α -convex on the same domain for some $\alpha \in \mathbb{R}$. Then we have*

$$\text{Lip}_\Omega(\log \rho + V)(1 + \alpha\tau) \leq \text{Lip}_\Omega(\log g + V) \quad (5)$$

where $\text{Lip}_\Omega(f) = \sup_{x \neq y \in \Omega} \frac{|f(x) - f(y)|}{|x - y|}$

3 The proof

Proof. We follow the same idea as in [11] (Lemma 3.2), except that we need to get rid of a possible maximum on the boundary. We eliminate this case when the domain is a ball and we do the general case by approximation.

3.1 The computation

Suppose that V and $\log g$ are $C^{1,\beta}$ functions for some $\beta \in (0, 1)$ and that Ω is smooth and uniformly convex. In this case, the regularity theory for the Monge-Ampère equation by Caffarelli ([4, 5, 8]) provides $\varphi \in C^{3,\beta}$. Indeed, we first observe that ρ is Lipschitz and bounded from below, so we can apply Caffarelli's theory to obtain $\varphi \in C^{2,\beta}$ (using $g, \rho \in C^{0,\beta}$ and the assumptions on the boundary of Ω). This implies $\log \rho \in C^{1,\beta}$ and allows to obtain $\varphi \in C^{3,\beta}$, which is the regularity we need for our next computations.

We proceed by a maximum principle argument : let x_0 be a point where $|\nabla\varphi|^2$ achieves its maximum. If x_0 is in the interior of Ω , the optimality conditions are :

$$\sum_{i=1}^d \varphi_{ij} \varphi_i(x_0) = 0 \text{ for all } j \quad (6)$$

$$(\varphi_{ijk} \varphi_i + \varphi_{ij} \varphi_{ik})_{(j,k)}(x_0) \leq 0 \text{ for all } i \quad (7)$$

where the indices correspond to partial derivatives in the associated direction and the inequality in the second line means that the matrix (indexed by j and k) is negative semidefinite for every i .

We differentiate the logarithm of (4) in the i direction to obtain :

$$-\text{Tr}(B^{jk} \varphi_{ijk}(x_0)) = (\log(\rho))_i(x_0) - (\log g)_i(T(x_0)) + \sum_{k=1}^d (\log g)_k(T(x_0)) \varphi_{ik}(x_0) \quad (8)$$

where the matrix B^{jk} is given by $B = (I - D^2\varphi(x_0))^{-1}$.

By using that B is a symmetric and non-negative matrix (as it is the inverse of the Hessian of a convex function) and the optimality conditions (7), we get :

$$-\text{Tr}(B^{jk}\varphi_{ijk}\varphi_i) \geq \text{Tr}(B^{jk}\varphi_{ij}\varphi_{ik}) \geq 0 \quad (9)$$

where we use $A \geq 0, B \leq 0 \Rightarrow \text{Tr}(AB) \leq 0$, with $A = (D^2\varphi)^2 \geq 0$. Then, by multiplying (8) by φ_i and using (6),(9),

$$\nabla\varphi(x_0) \cdot \nabla(\log \rho)(x_0) \geq \nabla(\log g)(T(x_0)) \cdot \nabla\varphi(x_0) \quad (10)$$

Equation (3) gives $\nabla(\log \rho) = -\nabla V - \frac{\nabla\varphi}{\tau}$ and thus we get :

$$\begin{aligned} \frac{|\nabla\varphi(x_0)|^2}{\tau} &\leq \nabla\varphi(x_0) \cdot (-\nabla V(x_0) - (\nabla(\log g))(T(x_0))) \\ \frac{|\nabla\varphi(x_0)|^2}{\tau} + \nabla\varphi(x_0) \cdot (\nabla V(x_0) - \nabla V(T(x_0))) &\leq -\nabla\varphi(x_0) \cdot (\nabla(V + \log g)(T(x_0))) \\ \frac{|\nabla\varphi(x_0)|^2}{\tau} + (x_0 - T(x_0)) \cdot (\nabla V(x_0) - \nabla V(T(x_0))) &\leq |\nabla\varphi(x_0)| \|\nabla(V + \log g)\|_\infty \\ \frac{|\nabla\varphi(x_0)|^2}{\tau} (1 + \tau\alpha) &\leq |\nabla\varphi(x_0)| \|\nabla(V + \log g)\|_\infty \end{aligned} \quad (11)$$

$$(12)$$

The last expression can be simplified in order to obtain

$$\|\nabla(V + \log \rho)\|_\infty(1 + \alpha\tau) = \frac{|\nabla\varphi(x_0)|}{\tau}(1 + \tau\alpha) \leq \|\nabla(V + \log g)\|_\infty$$

since x_0 is the point where $|\nabla\varphi|^2$ and also $|\nabla(\log \rho + V)|$ achieve their maximal values. We used the fact that, if V is C^1 and α uniformly convex, then for all x, y we have $(\nabla V(x) - \nabla V(y)) \cdot (x - y) \geq \alpha|x - y|^2$.

3.2 Treating the boundary

The computations above required the maximum point x_0 to be an interior point of Ω . We will remove this difficulty in two steps: first we show that if Ω is a ball the maximum can't be achieved on the boundary. Then we reduce the general case to this one by approximation.

3.2.1 If the domain is a ball

In this subsection, we suppose $\Omega = B_R(0)$ and prove the following

Lemma 3.1. *Suppose that $\Omega = B_R(0)$ is a ball and that φ is the Kantorovich potential for the quadratic cost between two densities $\rho, g \in C^{0,\beta}(\Omega)$ which are both bounded from below. Then $\phi \in C^1$ and the maximum of $|\nabla\varphi|$ cannot be achieved on $\partial\Omega$.*

Proof. Let us define $f(x) = \frac{1}{2}|\nabla\varphi(x)|^2$. From the regularity on the densities and on the domain, Caffarelli's theory implies $\varphi \in C^{2,\beta}$ and T is a homeomorphism of Ω . Therefore it sends the boundary to the boundary (an observation already used for similar purposes in Lemma 2.4 in [9]). Now, for $x \in \partial\Omega$ we have

$$|T(x)|^2 = |x - \nabla\varphi(x)|^2 = R^2 \quad \Rightarrow \quad f(x) = x \cdot \nabla\varphi(x)$$

Suppose that f achieve its maximum in Ω at $x_0 \in \partial\Omega$. Since $f(x) = x \cdot \nabla\varphi$ in $\partial\Omega$, $x \cdot \nabla\varphi$ also achieve its maximum in $\partial\Omega$ at x_0 . If we denote $\mathbf{n}(x) = \frac{x}{R}$ the normal at $x \in \partial\Omega$, there exists $\lambda \in \mathbb{R}^+$ (global maximum) and $\mu \in \mathbb{R}$ (maximum on the boundary) so that :

$$D^2\varphi(x_0) \cdot \nabla\varphi(x_0) = \lambda\mathbf{n}(x_0) \quad (13)$$

$$D^2\varphi(x_0) \cdot x_0 + \nabla\varphi(x_0) = \mu\mathbf{n}(x_0) \quad (14)$$

By multiplying equation (14) by $\mathbf{n}(x_0) = \frac{x_0}{R}$, we get :

$$R\mathbf{n}(x_0) \cdot D^2\varphi(x_0) \cdot \mathbf{n}(x_0) + \frac{f(x_0)}{R} = \mu \quad \Rightarrow \quad \frac{f(x_0)}{R} > \mu - R \quad (15)$$

where we used the inequality $D^2\varphi < I$, which is strict since the Monge-Ampere equation (4) tells us that $I - D^2\varphi$ can't have a zero eigen-value, since $\rho > 0$. If we multiply the same equation (14) by $\nabla\varphi(x_0)$ and use (13), we get :

$$\nabla\varphi(x_0) \cdot D^2\varphi(x_0) \cdot x_0 + |\nabla\varphi(x_0)|^2 = \mu\mathbf{n}(x_0) \cdot \nabla\varphi(x_0) \quad \rightarrow \quad \lambda x_0 \cdot \mathbf{n}(x_0) + 2f(x_0) = \mu \frac{x_0}{R} \cdot \nabla\varphi(x_0),$$

which can be re-written as

$$f(x_0)\left(2 - \frac{\mu}{R}\right) = -R\lambda \leq 0 \quad \Rightarrow \quad \mu \geq 2R, \quad (16)$$

since $\lambda \geq 0$. Putting (15) and (16) together, we obtain $f(x_0) > R^2$.

However we know that if Ω is a strictly convex domain, the transport plan shall verify for all $x \in \partial\Omega$, $\mathbf{n}(T(x)) \cdot \mathbf{n}(x) \geq 0$ (see the proof of Lemma 2.4 in [9]). In the case of a ball this implies that $T(x)$ is in the same hemisphere as x , thus in term of the distance between the two points we have $|T(x) - x| = |\nabla\varphi(x)| \leq \sqrt{2}R$. Therefore we have $f(x_0) \leq R^2$ on $\partial\Omega$, which leads to a contradiction if the maximum is attained on the boundary. \square

3.2.2 Approximation

We come back to the general case and we fix $R > 0$ such that $\Omega \subset B_R$. The idea is to approximate our problem by a sequence of regularized problems defined on B_R with a sequence of potentials V_n diverging to infinity out of Ω . First we extend V to \mathbb{R}^d by defining a function \tilde{V} :

$$\tilde{V}(x) = \sup_{y \in \Omega} \left(V(y) + \nabla V(y) \cdot (x - y) + \frac{\alpha}{2}|x - y|^2 \right)$$

This extension is non-negative and uniformly α -convex in \mathbb{R}^d . It coincides with V on Ω since V itself is uniformly α -convex. We consider now the function $h := \log g + V$ and extend it to \mathbb{R}^d by defining \tilde{h} :

$$\tilde{h}(x) = \sup_{y \in \Omega} (h(y) + \text{Lip}(h)|x - y|)$$

where $\text{Lip}(h)$ is the Lipschitz constant of h . We have $\text{Lip}(h) = \text{Lip}(\tilde{h})$. We then take convolutions in order to enforce higher regularity. Let $\xi \in C^\infty(\mathbb{R}^d, \mathbb{R}^+)$ be a convolution kernel, supported in B_1 , and satisfying $\int_{\mathbb{R}^d} \xi dx = 1$. We then rescale ξ defining $\xi_n(x) = n^d \xi(nx)$ and set

$$\begin{aligned} V_n(x) &= \int_{\mathbb{R}^d} (\tilde{V}(y) + nd(y, \Omega)^2) \xi_n(x - y) dy \\ h_n(x) &= \int_{\mathbb{R}^d} \tilde{h}(y) \xi_n(x - y) dy \\ g_n(x) &= \lambda_n e^{h_n(x) - V_n(x)} \end{aligned}$$

where λ_n is a normalisation constant chosen so as to ensure $\int_{B_R} g_n = 1$. We note that for every n we have $\|\nabla h_n\|_\infty = \text{Lip}(h_n) \leq \text{Lip}(\tilde{h}) = \text{Lip}(h)$ and that V_n is α -uniformly convex and non-negative in \mathbb{R}^d . We also define a sequence of functional \mathcal{F}_n on $\mathcal{P}(B_R)$:

$$\mathcal{F}_n(\rho) := \mathcal{F}_{g_n, V_n}^\tau(\rho) = \frac{W_2^2(\rho, g_n)}{2\tau} + \int_{B_R} \rho \log \rho + \int_{B_R} V_n d\rho \quad (17)$$

By applying our previous computations to the minimizer $\hat{\rho}_n$ of the functional \mathcal{F}_n (note that the regularity is enforced by the convolution and that we are on the ball, so that the maximum of the squared gradient of the Kantorovich potential cannot be achieved on the boundary) we obtain

$$\|\nabla(V_n + \log \hat{\rho}_n)\|_{L^\infty(B_R)}(1 + \alpha\tau) \leq \|\nabla(V_n + \log g_n)\|_{L^\infty(B_R)} \leq \text{Lip}_\Omega(V + \log g). \quad (18)$$

We now study the convergence of the V_n , f_n , g_n and \mathcal{F}_n . By the properties of the convolution and since \tilde{h} is extension of h , h_n converges uniformly to h in Ω . As for V_n , the situation is trickier, because we added a penalization of the form $nd(\cdot, \Omega)^2$ before convolving. Since the support of ξ_n is contained in the ball of radius $1/n$, when $x \in \Omega$ this penalization contributes at most $n \cdot (1/n)^2$ to the integral, so we have $\tilde{V} * \xi_n \leq V_n \leq \tilde{V} * \xi_n + 1/n$ on Ω . Hence V_n also converges uniformly to V on Ω . On the other hand, for all $x \in B_R \setminus \Omega$, we have $\lim_{n \rightarrow \infty} V_n(x) = +\infty$. We deduce that g_n converge uniformly to g in Ω and pointwise to 0 in $B_R \setminus \Omega$. Since g_n is a probability measure, we also have the weak convergence as probability measures of g_n to the measure supported on Ω with density g .

Then, a quick computation provides $\text{Lip}(V_n)_\Omega \leq \text{Lip}(V)_\Omega + 2$ (the additional term $+2$ comes from the gradient of the penalization $\nabla(nd(\cdot, \Omega)^2) = 2nd(\cdot, \Omega)\nabla d(\cdot, \Omega)$, whose modulus is almost 2 at points of $\Omega + \text{spt}(\xi_n)$), thus by (18) we get:

$$\text{Lip}(\log \rho_n)_\Omega \leq \frac{\text{Lip}(V + \log g)_\Omega}{1 + \alpha\tau} + \text{Lip}(V)_\Omega + 2. \quad (19)$$

This bound, together with the mass constraint $\int_\Omega \rho_n \leq \int_{B_R} \rho_n = 1$ implies a uniform upper bound on ρ_n . Composing with the exponential function, which is Lipschitz continuous on bounded intervals, we deduce that ρ_n is uniformly Lipschitz in Ω . We will now characterize the limit of ρ_n .

The sequence of functionals \mathcal{F}_n Γ -converges for the weak topology on measures to the functional \mathcal{F} defined as

$$\mathcal{F}(\rho) := \begin{cases} \mathcal{F}_{g, V}^\tau(\rho) & \text{if } \text{spt}(\rho) \subset \Omega, \\ +\infty & \text{if not.} \end{cases}$$

Indeed, if the sequence $\rho_n \in \mathcal{P}(B_R)$ weakly converges to $\rho \in \mathcal{P}(\Omega)$, we have

$$\mathcal{F}(\rho) \leq \liminf_{n \rightarrow \infty} \mathcal{F}_n(\rho_n). \quad (20)$$

This can be proven in the following way: $\rho \mapsto \int \rho \log \rho$ is lower-semicontinuous for the weak convergence of measure and W_2 metrizes the same convergence; if $\int V_n d\rho_n \leq C$ then $\rho_n(\{x : d(x, \Omega) > \varepsilon\}) \rightarrow 0$ as $n \rightarrow \infty$, hence $\text{spt}(\rho) \subset \Omega$; we also have $\int V_n d\rho_n \geq \int (\tilde{V} * \xi_n) d\rho_n \rightarrow \int \tilde{V} d\rho = \int V d\rho$ where the last equality is valid if ρ is concentrated on Ω .

On the other hand, if $\rho \in \mathcal{P}(\Omega)$, then the constant sequence $\rho_n = \rho$ satisfies

$$\limsup_{n \rightarrow \infty} \mathcal{F}_n(\rho_n) \leq \mathcal{F}(\rho) \quad (21)$$

We conclude that the optimizers $\tilde{\rho}_n$ weakly converge to ρ , the only minimizer of \mathcal{F} . Since $\tilde{\rho}_n$ is uniformly Lipschitz in Ω , we deduce that the convergence is uniform in Ω . Finally taking the limit of (18) we obtain

$$\text{Lip}_\Omega(\log \rho + V)(1 + \alpha\tau) \leq \text{Lip}_\Omega(\log g + V). \quad \square$$

Acknowledgments The second author acknowledges the support of the Monge-Ampère et Géométrie Algorithmique project, funded by Agence nationale de la recherche (ANR-16-CE40-0014 - MAGA).

References

- [1] L. AMBROSIO, N. GIGLI, G. SAVARÉ, *Gradient flows in metric spaces and in the space of probability measures*, Lectures in Mathematics, Birkhäuser (2005).
- [2] D. BAKRY, I. GENTIL AND M. LEDOUX, *Analysis and Geometry of Markov Diffusion Operators*, Springer, 2014
- [3] Y. BRENIER, Décomposition polaire et réarrangement monotone des champs de vecteurs. (French) *C. R. Acad. Sci. Paris Sér. I Math.* 305 (19), 805–808, 1987.
- [4] L. CAFFARELLI, Some regularity properties of solutions of Monge Ampère equation, *Comm. Pure Appl. Math.* 44, 965-969, 1991.
- [5] G. DE PHILIPPIS AND A. FIGALLI, The Monge–Ampère equation and its link to optimal transportation, *Bulletin of the American Mathematical Society* 51, 527-580, 2014.
- [6] 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.
- [7] S. DI MARINO F. SANTAMBROGIO JKO estimates in linear and non-linear Fokker-Planck equations, and Keller-Segel: L^p and Sobolev bounds. Preprint, 2019.
- [8] A. FIGALLI, *The Monge-Ampère Equation and Its Applications*, Zurich Lectures in Advanced Mathematics, European Mathematical Society (EMS), Zürich, 2017.
- [9] M. IACOBELLI, F. S. PATACCHINI, F. SANTAMBROGIO Weighted ultrafast diffusion equations: from well-posedness to long-time behaviour, *Arch. Rati. Mech. An.*, 232 (3), 1165-1206, 2019.
- [10] R. JORDAN, D. KINDERLEHRER, F. OTTO. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. An.*, 1998.
- [11] P. LEE, On the Jordan-Kinderlehrer-Otto scheme, *J. Math. An. Appl.* Vol 429, Issue 1, 131-142, 2015.
- [12] P. LEE, A Harnack inequality for the Jordan-Kinderlehrer-Otto scheme, *Journal of Evolution Equations* volume 18, pages 143–152 (2018)
- [13] F. SANTAMBROGIO *Optimal Transport for Applied Mathematicians*, book, *Progress in Nonlinear Differential Equations and Their Applications* 87, Birkhäuser Basel (2015).
- [14] F. SANTAMBROGIO {Euclidean, Metric, and Wasserstein} Gradient Flows: an overview, *Bull. Math. Sci.*, Vol 7 (1), 87–154, 2017.
- [15] C. VILLANI *Topics in Optimal Transportation*. Graduate Studies in Mathematics, AMS, (2003).