ON THE JORDAN-KINDERLEHRER-OTTO SCHEME

PAUL W.Y. LEE

Abstract. In this paper, we prove that the Jordan-Kinderlehrer-Otto scheme for a family of linear parabolic equations on the flat torus converges uniformly in space.

1. Introduction

Diffusion equations including the heat equation, the Fokker-Planck equation, and the porous medium equation are gradient flows in the Wasserstein space, the space of all probability measures, with respect to the $L^2$-Wasserstein distance from the theory of optimal transportation. One way to make the above statement precise is to use a time-discretization scheme introduced in [7] which is now called the Jordan-Kinderlehrer-Otto (JKO) scheme. An interesting and insightful formal Riemannian structure of the $L^2$-Wasserstein distance was also found in [10]. It was also shown in [2] that the above equations are examples of generalized gradient flows on abstract metric spaces (see [1] for a quick overview, further details, and some recent developments, see also [11, 12]).

We consider, in this paper, the JKO scheme for the equation

$$
\partial_t \phi_t = \Delta \phi_t + \langle \nabla \phi_t, \nabla \Psi \rangle + f \phi_t
$$
on the flat torus $\mathbb{T}^n$, assuming that we have a positive solution $v_0$ of the equation

$$
\Delta v_0 = \langle \nabla v_0, \nabla \Psi \rangle + (\Delta \Psi - f)v_0,
$$

where $\Psi$ and $f$ are smooth functions on the flat torus $\mathbb{T}^n$.

For this, let $\mu_0$ and $\mu_1$ be two Borel probability measure on $\mathbb{T}^n$. The $L^2$ Wasserstein distance $d$ between $\mu_0$ and $\mu_1$ is defined as follows

$$
d^2(\mu_0, \mu_1) = \inf_{\varphi: \mu_0 = \mu_1} \int_M d^2(x, \varphi(x))d\mu_0(x),
$$

where the infimum is taken over all Borel maps $\varphi: \mathbb{T}^n \to \mathbb{T}^n$ pushing $\mu_0$ forward to $\mu_1$. Minimizers of (1.2) are called optimal maps.
Let $dx^n$ be the Lebesgue measure on $\mathbb{T}^n$, let $\mu = v_0 dx^n$, let $\rho_0 \mu$ be a probability measure $\mathbb{T}^n$, let $\rho_0^N = \rho_0$, and let $K > 0$ be a constant. The following minimization problem has a unique solution $\rho_k^N$ for each $k = 1, \ldots, N$.

\[(1.3) \inf \left[ \frac{1}{2} d^2(\rho_{k-1}^N, \rho \mu) + \frac{K}{N} \int_{\mathbb{T}^n} (\log \rho - \log v_0 + \Psi) \rho d\mu \right],\]

where the infimum is taken over the set of $L^1$ functions $\rho : \mathbb{T}^n \to [0, \infty)$ satisfying $\int \rho d\mu = 1$.

This defines, for each positive integer $N$, a sequence of functions $\{\rho_k^N | k = 0, 1, \ldots\}$. This discrete scheme is the so-called JKO scheme.

Let $\phi_t^N : [0, K] \times \mathbb{T}^n \to [0, \infty)$ be defined by

$\phi_t^N = \rho_k^N$

if $t$ is in $\left[ \frac{kK}{N}, \frac{(k+1)K}{N} \right)$ and $k = 0, \ldots, N - 1$.

It follows as in [7] that a subsequence of $\{\phi_t^N | N = 1, 2, \ldots\}$ converges in $L^1$ to the solution of the initial value problem

\[(1.4) \quad \partial_t \phi_t = \Delta \phi_t + \langle \nabla \phi_t, \nabla \Psi \rangle + f \phi_t, \quad \phi_0 = \rho_0.\]

In the case of the flat torus $\mathbb{T}^n$, we show that the convergence is uniform in space. More precisely,

**Theorem 1.1.** For each fixed $t$ in the interval $[0, K]$ and each $\alpha$ satisfying $0 < \alpha < 1$, a subsequence of $\{\phi_t^N | N = 1, 2, \ldots\}$ converges in $C^{0,\alpha}$ to the solution of the initial value problem (1.4).

The rest of the sections are devoted to the proof of Theorem 1.1.

### 2. The optimal Transportation problem and the JKO Scheme

In this section, we recall basic results in the theory of optimal transportation and discuss the JKO scheme. Let $\mathbb{T}^n$ be the n-dimensional torus equipped with the flat distance $d$. Let $\mu_0$ and $\mu_1$ be two Borel probability measures on $\mathbb{T}^n$. Recall that the $L^2$ Wasserstein distance $d$ between $\mu_0$ and $\mu_1$ is the following minimization problem

\[(2.1) \quad d^2(\mu_0, \mu_1) = \inf_{\varphi : \mu_0 = \mu_1} \int_M d^2(x, \varphi(x)) d\mu_0(x),\]

where the infimum is taken over all Borel maps $\varphi : \mathbb{T}^n \to \mathbb{T}^n$ pushing $\mu_0$ forward to $\mu_1$. Minimizers of (2.1) are called optimal maps.
The minimization problem (2.1) admits the following dual problem

\[
\sup_{f(x) + g(y) \leq \frac{1}{2}d^2(x,y)} \left[ \int_{\mathbb{T}^n} f d\mu_0 + \int_{\mathbb{T}^n} g d\mu_1 \right],
\]

where the supremum is taken over all pairs \((f, g)\) of continuous functions satisfying \(f(x) + g(y) \leq \frac{1}{2}d^2(x, y)\) for all \(x\) and \(y\) in \(\mathbb{T}^n\).

The maximizers of the above problem are given by pairs of \(c\)-concave functions. If \(f : \mathbb{T}^n \to \mathbb{R}\) is a continuous function, then the \(c\)-transform \(f^c\) of \(f\) is defined by

\[
f^c(x) = \inf_{y \in \mathbb{T}^n} \left[ \frac{1}{2}d^2(x, y) - f(y) \right].
\]

The function \(f\) is \(c\)-concave if \(f^c = f\).

For the proof of following result, see, for instance, [3, 9].

**Theorem 2.1.** The infimum in (1.2) and the supremum in (2.2) coincide. Moreover, the supremum in (2.2) is achieved by a pair \((f, f^c)\), where \(f\) is a \(c\)-concave function.

The following existence and uniqueness result can be found in [3, 9]. Note that \(c\)-concave functions are locally semi-concave and hence twice differentiable almost everywhere (see [5]).

**Theorem 2.2.** Assume that the measure \(\mu_0\) is absolutely continuous with respect to the Lebesgue measure \(dx^n\). Then there is an optimal map \(\varphi\) of (1.2) pushing \(\mu_0\) forward to \(\mu_1\) which is of the form

\[
\varphi(x) = x - \nabla f(x),
\]

where \(f\) is a \(c\)-concave function. This map is unique up to a set of \(\mu_0\) measure zero.

Moreover, if \(\mu_1\) is also absolutely continuous with respect to \(dx^n\), then the map

\[
\varphi^c(x) := x - \nabla f^c(x)
\]

is the optimal map pushing \(\mu_1\) forward to \(\mu_0\), where \(f^c\) is the \(c\)-transform of \(f\).

Next, let us fix a positive constant \(h\) and a positive function \(\rho_0\) satisfying \(\int_{\mathbb{T}^n} \rho_0 d\mu = 1\) and consider the following minimization problem

\[
\inf \left[ \frac{1}{2}d^2(\rho_0 \mu, \rho \mu) + h \int_{\mathbb{T}^n} (\log \rho - \log v_0 + \Psi) \rho d\mu \right],
\]

where the infimum is taken over the set of \(L^1\) functions \(\rho : \mathbb{T}^n \to [0, \infty)\) satisfying \(\int_{\mathbb{T}^n} \rho d\mu = 1\).
Theorem 2.3. There is a unique minimizer $\rho$ of (2.4) which is locally semi-convex on the set 
\[ \{ x | \rho_0(x) > 0 \} \]
and satisfying the following
(2.5) \[ \varphi_*(\rho \mu) = \rho_0 \mu, \]
where
\[ \varphi(x) = x + h \nabla F(x) \]
is the optimal map which pushes forward $\rho \mu$ to $\rho_0 \mu$ and
\[ F := \log \rho - \log v_0 + \Psi. \]
Moreover, if there is a $C^2$ positive solution of (2.5), then it coincides with $\rho$.

Proof. By the convexity of the functional in (2.4), any minimizing sequence of (2.4) converges weakly in $L^1$ to a unique minimizer $\rho$. Following [7], we let $\psi_s$ be the flow of a vector field $X$ and let $(\psi_s)_*(\rho \mu) = \sigma_s \mu$. It follows that
(2.6) \[ \frac{d}{ds} \left[ \frac{1}{2} \int_{T^n} d^2(x, \psi_s(\varphi(x)))\rho_0(x) + h \log \sigma_s(\psi_s(x))\rho(x) \right] \bigg|_{s=0} = 0, \]
where $\varphi$ is the optimal map pushing $\rho_0 \mu$ forward to $\rho \mu$.

Since $\sigma_s(\psi_s)v_0(\psi_s)\det(d\psi_s) = \rho v_0$, we have
(2.7) \[ \frac{d}{ds} \int_{T^n} \left[ \log \sigma_s - \log v_0 + \Psi(\psi_s(x)) \right] \rho(\varphi(x))d\mu(x) \bigg|_{s=0} \]
\[ = \int_{T^n} \left[ -\text{div}(X(x)) + \langle X(x), \nabla(\Psi - 2 \log v_0(x)) \rangle \right] \rho(\varphi(x))d\mu(x). \]

On the other hand, we have
\[ \frac{d^2(x, \psi_s(\varphi(x))) - d^2(x, \varphi(x))}{s} \]
\[ \leq 2D \frac{d(x, \psi_s(\varphi(x))) - d(x, \varphi(x))}{s} \]
\[ \leq 2D \frac{d(\varphi(x), \psi_s(\varphi(x)))}{s} \leq 2D ||X||c_0 \]
where $D$ is the diameter of the torus $T^n$.

By Theorem 2.1, we also have
\[ \int_{T^n} \left[ \frac{1}{2} d^2(\varphi(x)) - f(x) - f^c(\varphi(x)) \right] \rho_0(x)d\mu(x) = 0 \]
with \( f(x) + f^c(y) \leq \frac{1}{2} d^2(x, y) \) for all \( x, y \) in \( M \). Therefore,

\[
(2.8) \quad f(x) + f^c(\bar{\phi}(x)) = \frac{1}{2} d(x, \bar{\phi}(x))^2
\]

for \( \rho_0 \mu \) almost all \( x \).

Let \( x \) be a point where \( f \) is twice differentiable at \( x \), \( f^c \) is twice differentiable at \( \bar{\phi}(x) \), and \( (2.8) \) holds. Then \( \bar{\phi}(x) \) is a minimizer of \( g_1(y) := \frac{1}{2} d(x, y)^2 - f^c(y) \). On the other hand, if \( y_0 \) is another minimizer of \( g_1 \), then \( x \) is a minimizer of \( g_2(y) := \frac{1}{2} d^2(y, y_0) - f(y) \). Note that

\[
\inf_{y \in M} \left[ \frac{1}{2} d^2(y, y_0) - f(y) \right] = \inf_{\gamma(1) = y_0} \left[ \int_0^1 \frac{1}{2} |\gamma(t)|^2 dt - f(\gamma(0)) \right],
\]

where the infimum on the right side is taken over all smooth paths \( \gamma : [0, 1] \to M \) such that \( \gamma(1) = y_0 \). Moreover, if \( \gamma \) is a minimizer of the infimum on the right side, then the Euler-Lagrange equation implies that \( \gamma \) is a geodesic and \( \dot{\gamma}(0) = -\nabla f(\gamma(0)) \). It follows that \( y_0 = x - \nabla f(x) = \bar{\phi}(x) \). Therefore, \( \bar{\phi}(x) \) is the unique minimizer of \( g_1 \).

Moreover, \( t \mapsto x - t \nabla f(x) \) is the unique geodesic joining \( x \) and \( \bar{\phi}(x) \). Hence, for \( \rho_0 dx^n \) almost all \( x, y \mapsto d^2(x, y) \) is differentiable at \( \bar{\phi}(x) \).

Therefore,

\[
\frac{d^2(x, \psi_s(\bar{\phi}(x))) - d^2(x, \bar{\phi}(x))}{s}
\]

converges, as \( s \to 0 \), to \( \langle \nabla d^2_x, X \rangle_{\bar{\phi}(x)} \) for \( \rho_0 dx^n \) almost all \( x \), where \( d_z(\cdot := d(z, \cdot) \). Therefore, by the bounded convergence theorem, we have

\[
\frac{d}{ds} \left[ \frac{1}{2} \int_{\mathbb{T}^n} d^2(x, \psi_s(\bar{\phi}(x))) \rho_0(x) d\mu(x) \right] \bigg|_{s=0} = \int_{\mathbb{T}^n} \left\langle \nabla d^2_x, X \right\rangle_{\bar{\phi}(x)} \rho_0(x) d\mu(x).
\]

By combining this with \( (2.6) \) and \( (2.7) \), we have

\[
\int_{\mathbb{T}^n} \text{div}(X) \rho d\mu
\]

\[
= \int_{\mathbb{T}^n} \left\langle \frac{1}{2h} \nabla d^2_{\bar{\phi}^{-1}}(x) + \nabla (\Psi - 2 \log v_0), X \right\rangle_x \rho(x) d\mu(x).
\]

It follows that \( \rho \) is Lipschitz and

\[
-h \nabla \log \rho(x) + h \nabla \log v_0 - h \nabla \Psi(x) = \frac{1}{2} \nabla d^2_{\bar{\phi}^{-1}}(x)
\]

holds \( \rho \mu \) almost everywhere. Therefore,

\[
x + h \nabla F(x) = \bar{\phi}^{-1}(x)
\]
\( \rho \mu \) almost everywhere. (2.5) follows from this and Theorem 2.2.

It also follows from the above discussion that
\[
t \mapsto x + th \nabla F(x)
\]
is the unique minimizing geodesic between its endpoints for \( \rho \mu \) almost all \( x \). Therefore,
\[
h \nabla F(x) = -\nabla f^c(x)
\]
for \( \rho \mu \) almost all \( x \).

Since \( c \)-concave functions are locally semi-concave, \( \log \rho \) is locally semi-convex on the open set \( \{ x \in M | \rho(x) > 0 \} \). It follows that the following equation holds Lebesgue almost everywhere on the same set
\[
(2.11) \quad \rho_0(\varphi(x))v_0(\varphi(x)) \det(d\varphi(x)) = \rho(x)v_0(x)
\]
where \( \varphi(x) = x + h \nabla F(x) \).

Let \( \rho \) be a positive \( C^2 \) solution of (2.11). Let \( \bar{\rho} = \log \rho - \log \rho_0 \) and let \( x' \) be the maximum point of \( g \). Since \( g \) is locally semi-convex, there is a sequence of points \( x_i \) where \( g \) is twice differentiable and the followings hold
\[
\lim_{i \to \infty} x_i = x', \quad \lim_{i \to \infty} \nabla g(x_i) = 0, \quad \nabla^2 g(x_i) \leq \epsilon_i I
\]
for some sequence \( \epsilon_i \to 0 \) as \( i \to \infty \). It follows that
\[
\rho(x_i)v_0(x_i) = \rho_0(\varphi(x_i))v_0(\varphi(x_i)) \det(I + h \nabla^2 F(x_i))
\]
\[
\leq \rho_0(\varphi(x_i))v_0(\varphi(x_i)) \det((1 - \epsilon_i)I + h \nabla^2 (\log \rho - \log v_0 + \Psi)(x_i)).
\]

By letting \( i \to \infty \) and using (2.11), we obtain \( \rho \leq \bar{\rho} \) everywhere. On the other hand, \( \int \rho v_0 = \int \bar{\rho} v_0 \). Therefore, \( \rho \equiv \bar{\rho} \). \( \square \)

3. Regularity of Minimizers

In this section, we show \( C^2 \) regularity and positivity of the minimizers in (1.3). More precisely, we have the following result.

**Theorem 3.1.** Let
\[
\lambda := \sup_{\{w \in T^n \mid |w| = 1\}} \nabla^2 F(w, w).
\]
Assume that \( \rho_0 \) is a positive \( C^{r,\alpha} \) function for some \( r \geq 2 \) and \( \alpha > 0 \). Assume also that \( h\lambda \leq \frac{1}{8} \). Then there is a constant \( h_0 > 0 \) depending on \( \Psi, v_0, \) and \( \rho_0 \) such that the minimizer of (2.4) with \( h < h_0 \) is a \( C^{r+2,\alpha} \) solution of (2.11).

First, we prove the following a priori estimates for solutions of the equation (2.11).
Lemma 3.2. Let $\rho$ be a $C^4$ positive solution of the equation (2.11). If $I + h\nabla^2 F \geq 0$, then

\begin{enumerate}
  \item \[(1 + h||\nabla^2(\Psi - \log v_0)||_{\infty})^n \sup_{T^n}(\rho_0v_0)\geq \rho_0v_0 \geq (1 - h||\nabla^2(\Psi - \log v_0)||_{\infty})^n \inf_{T^n}(\rho_0v_0),\]
  \item \[(1 - h||\nabla^2(\Psi - 2\log v_0)||_{\infty})||\nabla F||_{\infty} \leq ||\nabla F_{0}||_{\infty},\]
  \item \[0 \leq h||\nabla^3(\Psi - 2\log v_0)||_{\infty}||\nabla F||_{\infty} + \lambda_0\]
\end{enumerate}

where $|S|$ is the norm of the tensor $S$, $||S||_{\infty} = \sup_{T^n} |S|$, 

\[\lambda_0 = \sup_{\{w \in T^n \mid \|w\| = 1\}} \nabla^2 F_0(w, w)\]

and $F_0 = \log \rho_0 - \log v_0 + \Psi$.

Proof. The first assertion follows immediately from (2.11) and the maximum principle.

For each fixed $x$ in the torus $T^n$ and each vector $w$ in $\mathbb{R}^n$, we let $\gamma(s) = x + sw$ and let $\varphi(x) = x + h\nabla F(x)$. It follows from (2.11) that

\[
\log \rho_0(\varphi(\gamma(s))) + \log v_0(\varphi(\gamma(s))) + \log \det(I + h\nabla^2 F(\gamma(s)))
\]

\[
= \log \rho(\gamma(s)) + \log v_0(\gamma(s)).
\]

By differentiating this equation with respect to $s$, we obtain

\[
\langle \nabla(\log \rho + \log v_0)(\gamma(s)), w \rangle = \langle \nabla(\log \rho_0 + \log v_0)(\varphi(\gamma(s))), w + h\nabla^2 F(\gamma(s))w \rangle
\]

\[
+ h\text{tr}((I + hS(s))^{-1}S'(s)).
\]

where $S(s)$ is the matrix defined by $S_{ij}(s) = \nabla^2 F(\gamma(s))(\partial_x^i, \partial_x^j)$.

If $x$ is a point where $||\nabla F||^2$ achieves its maximum, then $\nabla^2 F(x)(\nabla F(x)) = 0$ and $\nabla^3 F(x)(\nabla F(x), v, v) \leq -||\nabla^2 F(x)||^2$ for any vector $v$.

By combining this with (3.2), we obtain

\[
\langle \nabla(\log \rho + \log v_0)(x), \nabla F(x) \rangle - \langle \nabla(\log \rho_0 + \log v_0)(\varphi(x)), \nabla F(x) \rangle
\]

\[
\leq -h\text{tr}((I + hS(0))^{-1}S(0)^2) \leq 0.
\]

Therefore,

\[
||\nabla F(x)|| - \left\langle \nabla(\log \rho_0 - \log v_0 + \Psi)(\varphi(x)), \frac{\nabla F(x)}{||\nabla F(x)||} \right\rangle
\]

\[
\leq \left\langle \nabla(-2\log v_0 + \Psi)(x) - \nabla(-2\log v_0 + \Psi)(\varphi(x)), \frac{\nabla F(x)}{||\nabla F(x)||} \right\rangle
\]

\[
\leq h||\nabla^2(-2\log v_0 + \Psi)||_{\infty}||\nabla F(x)||
\]
and the second assertion follows from this.

By differentiating (3.2), we obtain
\[
\nabla^2 (\log \rho + \log v_0)(x)(w, w) \leq \nabla^2 (\log \rho_0 + \log v_0)(\varphi(x))(d\varphi(w), d\varphi(w)) + h\nabla^3 F(x)(w, w, \nabla (\log \rho_0 + \log v_0)(\varphi(x))) - h^2 \text{tr}
((I + hS(0))^{-1}S'(0))^2) + h^2 \text{tr}
((I + hS(0))^{-1}S''(0)).
\]

(3.3)

If \((x, w)\) achieves the supremum in (3.1), then
\[
\nabla^3 F(x)(w, w, v) = 0,
\]
\[
\langle S''(0)v, v \rangle = \nabla^4 F(x)(w, w, v, v) \leq 0
\]
for any vector \(v\).

Therefore, by combining this with (3.3), we obtain
\[
\nabla^2 (\log \rho + \log v_0)(x)(w, w) \leq \nabla^2 (\log \rho_0 + \log v_0)(\varphi(x))(d\varphi(w), d\varphi(w)) + h\nabla^3 F(x)(w, w, \nabla (\log \rho_0 + \log v_0)(\varphi(x))) - h^2 \text{tr}
((I + hS(0))^{-1}S'(0))^2) + h^2 \text{tr}
((I + hS(0))^{-1}S''(0)).
\]

Since \(d\varphi(w) = w + h\nabla^2 F(x)(w) = w + h\lambda w\), it also follows that
\[
\lambda - (1 + h\lambda)^2 \lambda_0 = \nabla^2 (\Psi - 2 \log v_0)(x)(w, w) - \nabla^2 (\Psi - 2 \log v_0)(\varphi(x))(d\varphi(w), d\varphi(w)) \leq \int_0^1 d\frac{dt}{dt} \nabla^2 (\Psi - 2 \log v_0)(\varphi(t))(\varphi(t) d\varphi(t)(\varphi(t)(w))(dt)
\]
\[
\leq h||\nabla^3 (\Psi - 2 \log v_0)||_\infty ||\nabla F||_\infty \left(1 + h\lambda + \frac{1}{3}h^2 \lambda^2\right)
\]
\[
+ h||\nabla^4 (\Psi - 2 \log v_0)||_\infty (2\lambda + h\lambda^2),
\]
where \(\Phi_t(x) = tx + (1 - t)\varphi(x)\).

The last assertion follows from this. \(\square\)

**Proof of Theorem 3.1.** Let \(C\) be a constant such that
\[
\int_M Ce^{\log v_0 - \Psi} \mu = 1.
\]

First, if \(\rho_0 = Ce^{\log v_0 - \Psi}\), then \(\rho = Ce^{\log v_0 - \Psi}\) is a smooth positive solution of (2.11). For more general \(\rho_0\), let us consider the following family of equations
\[
\text{det}(I + h\nabla^2 F(x)) = \frac{\rho(x)v_0(x)}{(Ce^{\log v_0 - \Psi})^{1-s}(\rho_0)^s v_0(x + h\nabla F(x))}.
\]

(3.4)

Let \(s_0\) be the supremum over the set of all \(s\) in \([0, 1]\) for which (3.4) has a \(C^1\) solution \(\rho_\ast\). By Theorem 2.3, \(\rho_\ast\) is a minimizer of (2.4) with \(\rho_0\) replaced by \((Ce^{\log v_0 - \Psi})^{1-s}(\rho_0)^s\). It follows that \(I + h\nabla^2 (\log \rho_\ast - \log v_0 + \Psi) \geq 0\). Let us choose \(h_0\) such that \(1 - h_0 ||\nabla^2 (\Psi - 2 \log v_0)||_\infty > 0\).
Then, by Theorem 3.2, the set of solutions \( \{ \log \bar{\rho}_s | 0 \leq s < s_0 \} \) has a uniform \( C^1 \) bound depending on \( \rho_0 \).

Let \( \lambda(s) = \sup_{x \in \mathbb{T}^n, |w|=1} \nabla^2 ((C e^{\log v_0 - \Psi})^{1-s}(\rho_0)^s) - \log v_0 + \Psi)(x)(w, w) \).

Since
\[
\sup_{x \in \mathbb{T}^n, |w|=1} \nabla^2 (\log \rho_0 - \log v_0 + \Psi)(x)(w, w) = s \lambda_0 < \lambda_0,
\]
we can apply Theorem 3.2 and conclude that there are positive constants \( h_0 \) and \( C \) depending only on \( \rho_0, v_0, \) and \( \Psi \) such that
\[
0 \leq h^2 C + h \lambda_0 + (hC + 2h \lambda_0 - 1) h \lambda(s) + (hC + h \lambda_0) (h \lambda(s))^2
\]
for all \( h < h_0 \).

By assumption, we have \( h \lambda_0 \leq \frac{1}{8} \). Therefore, by choosing a smaller \( h_0 \),
\[
 h^2 C + h \lambda_0 + (hC + 2h \lambda_0 - 1) x + (hC + h \lambda_0) x^2
\]
has real root.

Since \( \lambda(0) = 0 \), \( \lambda(s) \) has an upper bounded independent of \( s \). Therefore, \( \lambda(s) \) is bounded above independent on \( s \) (again by assuming \( h_0 \) small enough).

This, together with (3.4), also gives a uniform positive lower bound of \( I + h \nabla^2 (\log \bar{\rho}_s - \log v_0 + \Psi) \). Higher order estimates follow from standard elliptic theory [6]. Therefore, there is a solution \( \bar{\rho}_{s_0} \) of (3.4) with \( s = s_0 \). By elliptic theory and the implicit function theorem, we must have \( s_0 = 1 \).

**4. \( C^{0,\alpha} \)-Convergence of the JKO Scheme**

In this section, we show the \( C^{0,\alpha} \)-convergence of the JKO scheme. For each fixed positive integer \( N \), let \( h = \frac{K}{N} \), where \( K \) is a positive constant to be fixed later which depends only on \( \rho_0, v_0, \) and \( \Psi, \) and dimension. Let \( \rho_0 : \mathbb{T}^n \to (0, \infty) \) be a smooth function. Then (1.3) defines a sequence of functions \( \rho_N^0 := \rho_0, \rho_1^N, ..., \rho_N^N \). Let \( F^N_k := \log \rho_N^N - \log v_0 + \Psi \). Then the followings are consequences of Lemma 3.2.

**Lemma 4.1.** There are positive constants \( C \) and \( K \) depending only on \( \rho_0, v_0, \Psi, \) and dimension such that

1. \( \frac{1}{C} \leq \rho_k^N \leq C, \)
2. \( ||\nabla F_k^N||_\infty \leq C, \)
3. \( \lambda_k^N := \sup_{x \in M, |w|=1} \nabla^2 F_k^N(x)(w, w) \leq C, \)
4. \( K \lambda_k^N \leq \frac{1}{8}, \)
for all positive integer \( N \) and all \( k = 1, \ldots, N \).

**Proof.** We proceed by induction. Assume that the estimates hold and \( \frac{K_N}{N} \lambda_k \leq \frac{1}{8} \) for all \( k \leq m - 1 \). Then \( \rho_m^N \) is smooth. Assume that \( K \) satisfies

\[
K < \max \left\{ \frac{1}{\|\nabla^2(\Psi - 2 \log v_0)\|_\infty}, \frac{1}{\|\nabla^2(\Psi - \log v_0)\|_\infty} \right\}.
\]

It follows from Lemma 3.2 that

\[
e^{-C\|\nabla^2(\Psi - \log v_0)\|_\infty} \sup(\rho_0 v_0)
\leq \left( 1 + \frac{K_N}{N} \|\nabla^2(\Psi - \log v_0)\|_\infty \right)^{\frac{1}{nN}} \sup(\rho_0 v_0)
\]

\[
\geq \rho_m^N v_0 \geq \left( 1 - \frac{K_N}{N} \|\nabla^2(\Psi - \log v_0)\|_\infty \right)^{\frac{1}{nN}} \inf(\rho_0 v_0)
\]

\[
\geq e^{-K\|\nabla^2(\Psi - \log v_0)\|_\infty} \inf(\rho_0 v_0) > 0
\]

and

\[
\|\nabla F_m^N\|_\infty \leq \frac{1}{\left( 1 - \frac{K_N}{N} \|\nabla^2(\Psi - 2 \log v_0)\|_\infty \right)^{\frac{1}{nN}}} \|\nabla F_0^N\|_\infty
\leq e^{-K\|\nabla^2(\Psi - 2 \log v_0)\|_\infty} \|\nabla F_0\|_\infty.
\]

This proves the first two assertions.

By the above estimates and Lemma 3.2, there are positive constants \( C \) and \( h_0 \) depending only on \( \Psi, v_0 \), and \( \rho_0 \) such that

\[
0 \leq h^2 C + h \lambda_k^N + (hC + 2h \lambda_{k-1}^N - 1)h \lambda_k^N + \left( hC + h \lambda_{k-1}^N \right)(h \lambda_k^N)^2
\]

for all \( h < h_0 \).

Assume that \( K \leq \frac{1}{6C} \). It follows as in the proof of Theorem 3.1 that

\[
h \lambda_k^N \leq \frac{2(h^2 C + h \lambda_k^N)}{1 - hC - 2h \lambda_{k-1}^N + \sqrt{1 + h^2 C^2 - 2h C - 4h^3 C^2 - 4h \lambda_{k-1}^N - 4h^3 C \lambda_{k-1}^N}}
\]

\[
\leq \frac{2(h^2 C + h \lambda_k^N)}{1 - hC - 2h \lambda_{k-1}^N + \sqrt{1 - 3hC - 4h \lambda_{k-1}^N}}
\]

\[
\leq \frac{h^2 C + h \lambda_{k-1}^N}{1 - 2hC - 3h \lambda_{k-1}^N}.
\]

for all \( k \leq m - 1 \). Note that \( 1 - 3hC - 4h \lambda_{k-1}^N > 0 \) since \( K \leq \frac{1}{6C} \) and \( \frac{K_N}{N} \lambda_{k-1} = h \lambda_{k-1} \leq \frac{1}{8} \).
By assuming that the constant $C$ is large enough, we have
\[
ph^2 C + \frac{h^2 C + h \lambda_N}{1 - 2hC - 3h \lambda_N} \quad \leq \quad \frac{ph^2 C (1 - 2hC - 3h \lambda_N^k) + h^2 C + h \lambda_N^k}{1 - 2phC (1 - 2hC - 3h \lambda_N^k) - 3ph^2 C - 3ph \lambda_N^k} \quad \leq \quad \frac{(p + 1)h^2 C - 2ph^2 C + 3ph^2 C \lambda_N^N + h \lambda_N^N}{1 - 2(p + 1)hC - 2ph^2 C^2 - 3ph^2 C \lambda_N^N + h \lambda_N^N}.
\]

By iterating the two inequalities above, we obtain
\[
\frac{K \lambda_0^N}{N} \leq \frac{K}{N} \frac{\lambda_0 + \frac{km}{N} C}{1 - \frac{2CKm}{N} - \frac{3Km \lambda_0}{N}} \leq \frac{K}{N} \frac{\lambda_0 + KC}{1 - 2CK - 3K \lambda_0},
\]
where $\lambda_0 = \sup_{x \in M, |w| = 1} \nabla^2 F_0(x)(w, w) \leq C$ and $F_0 = \log \rho_0 - \log v_0 + \Psi$.
Assume that the constant $K$ satisfies $\frac{K(\lambda_0 + KC)}{1 - \frac{2CK - 3K \lambda_0}{N}} < \frac{1}{8}$ and $K < \frac{1}{2C + 3\lambda_0}$. Then the last two assertions follow.

Finally, we finish the proof of Theorem 1.1. The arguments are mild modifications of the ones in [7, 10, 2] combined with the estimates in Lemma 4.1.

Proof. Let $\phi^N_t : [0, K] \times M \to \mathbb{R}$ be defined by
\[
\phi^N_t = \rho^N_k
\]
if $t$ is in $[\frac{kk}{N}, \frac{(k+1)K}{N})$ and $k = 0, ..., N - 1$.

By (1.3),
\[
\frac{1}{2} \mathbf{d}^2(\rho^N_k \mu, \rho^N_{k+1} \mu) + \frac{K}{N} \int_M F^N_{k+1} d\mu \leq \frac{K}{N} \int_M F^N_k d\mu.
\]

Therefore,
\[
\frac{1}{2} \sum_{k=0}^{N-1} \mathbf{d}^2(\rho^N_k \mu, \rho^N_{k+1} \mu) \leq \frac{K}{N} \left( \int_M F^N_0 \rho_0 d\mu - \inf \int_M (\log \rho - \log v_0 + \Psi) \rho d\mu \right),
\]

(4.1)
where the infimum is taken over all $\rho : \mathbb{T}^n \to [0, \infty)$ satisfying
\[ \int_{\mathbb{T}^n} \rho \, d\mu = 1. \]

Therefore, we obtain
\[
\int_0^K \int_{\mathbb{T}^n} |\nabla \log \phi_t^N(x) - \nabla \log v_0 + \nabla \Psi(x)|^2 \phi_t^N(x) \, d\mu dt \\
= \frac{K}{N} \sum_{k=0}^{N-1} \int_{\mathbb{T}^n} |\nabla F_k^N|^2 \rho_k^N \, d\mu \\
= \sum_{k=0}^{N-1} d^2(\rho_k \mu, \rho_{k+1} \mu) \leq C.
\]

Hence, $t \mapsto \sqrt{\int_{\mathbb{T}^n} |\nabla \log \phi_t^N(x) - \nabla \log v_0 + \nabla \Psi(x)|^2 \phi_t^N(x) \, d\mu}$ converges weakly in $L^2$ to a function $A : [0, K] \to \mathbb{R}$. On the other hand, we have
\[
d(\phi_{\tau_0}^N, \phi_{\tau_1}^N) \leq \sum_{k=m_0}^{m_1-1} d(\rho_{k-1}^N, \rho_k^N) \\
= \frac{K}{N} \sum_{k=m_0}^{m_1-1} \sqrt{\int_{\mathbb{T}^n} |\nabla F_{k-1}^N|^2 \rho_{k-1}^N \, d\mu} \\
= \int_{\frac{(m_1-1)K}{N}}^{\frac{m_1K}{N}} \sqrt{\int_{\mathbb{T}^n} |\nabla \log \phi_t^N(x) - \nabla \log v_0 + \nabla \Psi(x)|^2 \phi_t^N(x) \, d\mu \, dt,
\]
where $\frac{(m_1-1)K}{N} \leq \tau_i < \frac{m_1K}{N}$, $i = 0, 1$.

By letting $N \to \infty$, we obtain
\[
\liminf_{N \to \infty} d(\phi_{\tau_0}^N, \phi_{\tau_1}^N) \leq \int_{\tau_0}^{\tau_1} A(t) \, dt.
\]

Let $D$ be a dense subset of $[0, K]$. By Lemma 4.1 and a diagonal argument, there is a subsequence $\phi_{t_k}^N$ such that $\phi_{t_k}^N$ converges uniformly on $\mathbb{T}^n$ to a continuous function $\phi_t$ for all $t$ in $D$. By Lemma 4.1 and (4.2), we can extend $\{\phi_t | t \in D\}$ to a unique curve $\{\phi_t | t \in [0, K]\}$ contained in $C^{0, \alpha}(\mathbb{T}^n)$ by continuity. Next, we show that $\phi_{t_k}^N$ converges to $\phi_t$ uniformly for all $t$ in $[0, K]$. By Lemma 4.1, it is enough to show that any convergence subsequence (say $\phi_{t_k}^{N_{km}}$) converges uniformly to
\( \phi_t \). Suppose that the uniform limit of \( \phi_t^{N_{km}} \) is \( \phi_t \). Then, by (4.2),
\[
\mathbf{d}(\phi_s \mu, \phi_t \mu) = \lim_{m \to \infty} \mathbf{d}(\phi_s^{N_{km}} \mu, \phi_t^{N_{km}} \mu) \leq \int_s^t A(\tau) d\tau
\]
for all \( s \) in \( \mathcal{D} \). It follows from this and the definition of \( \phi_t \) that \( \phi_t = \phi_t \).

We also know that \( T < \infty \) with compact support, then we have
\[
\sum_{k=1}^{\infty} \int_{T_n} \xi_t(\rho^N_k - \rho^N_{k-1}) d\mu + \frac{K}{N} \int_{T_n} \langle \nabla \xi_t, \nabla F^N_k \rangle \rho^N_k d\mu
\]
\[
= \left| \int_{T_n} \xi_t(\Phi^N_k) \rho^N_k d\mu - \int_{T_n} \xi_t \rho^N_k d\mu - \frac{K}{N} \int_{T_n} \langle \nabla \xi_t, \nabla F^N_k \rangle \rho^N_k d\mu \right|
\]
\[
\leq \frac{K^2}{2N^2} \int_{T_n} \langle \nabla^2 \xi_t(\nabla F^N_k), \nabla F^N_k \rangle \rho^N_k d\mu
\]
\[
\leq \frac{1}{2} \| \nabla^2 \xi \|_{C^0} \frac{K^2}{N^2} \int_{T_n} |\nabla F^N_k|^2 \rho^N_k d\mu.
\]

On the other hand, by (2.6), (2.7), and (2.9), we have
\[
\int_M [-\text{div}(X) + \langle X, \nabla(\Psi - 2 \log v_0) \rangle - \langle \nabla F^N_k, X \rangle] \rho_k d\mu = 0.
\]
Therefore, by choosing \( X = \nabla \xi_t \), we obtain
\[
\left| \int_{T_n} [-\Delta \xi_t + \langle \nabla \xi_t, \nabla(\Psi - 2 \log v_0) \rangle] \rho^N_k d\mu
\]
\[
+ \int_{T_n} \frac{N}{K} \xi_t(\rho^N_k - \rho^N_{k-1}) d\mu \right| \leq \frac{1}{2} \| \nabla \xi \|_{C^0} \frac{K}{N} \int_{T_n} |\nabla F^N_k|^2 \rho^N_k d\mu.
\]

For each fix time \( T \) in \( [0, K] \), let \( k_N \) be the integer satisfying \( \frac{(k_N - 1)K}{N} \leq T < \frac{k_N K}{N} \). By applying (4.1), we obtain
\[
\sum_{k=1}^{k_N} \int_{T_n} \xi_t \frac{(\rho^N_k - \rho^N_{k-1})}{N} d\mu
\]
\[
= \sum_{k=1}^{k_N} \frac{K}{N} \int_{T_n} \left[ \xi_t \left( \rho^N_k - \rho^N_{k-1} \right) \right] d\mu + o \left( \frac{K}{N} \right)
\]
\[
\to \int_0^T \int_{T_n} \left[ \Delta \xi_t - \langle \nabla \xi_t, \nabla(\Psi - 2 \log v_0) \rangle \right] \phi_t d\mu dt
\]
as \( N \to \infty \).
On the other hand, we have

\[ \sum_{k=1}^{k_N} \int_{T^n} \xi^{(k-1)K} N^N \rho_k N - \xi^{(k-1)K} N^N \rho_k N' \, d\mu 
= \sum_{k=1}^{k_N} \int_{T^n} \xi^{(k-1)K} N^N \rho_k N \, d\mu - \sum_{k=0}^{k_N-1} \int_{T^n} \xi^{(k-1)K} N^N \rho_k N \, d\mu 
= \int_{T^n} \xi^{(k_N-1)K} N^N \rho_k N \, d\mu - \int_{T^n} \xi^{(K)N} N^N \rho_k N \, d\mu 
\rightarrow \int_{T^n} \xi_T \phi_T \, d\mu - \int_{T^n} \xi_0 \phi_0 \, d\mu - \int_0^T \int_{T^n} \partial_t \xi_t \phi_t \, d\mu \, dt \]

as \( N \to \infty \).

Finally, by the combining the above discussions with (1.1), we see that \( \phi \) is a weak solution (and hence the unique classical solution, see [8]) to the equation \( \dot{\phi} = \Delta \phi + \langle \nabla \phi, \nabla \Psi \rangle + f\phi \) on \([0, K] \times T^n\). This gives the result for a short time. The result for long time follows from \( C^2 \) estimate for linear parabolic equations (see [8]). \( \square \)

References

[1] L. Ambrosio, N. Gigli: A user's guide to optimal transport. Modelling and optimisation of flows on networks, 1-155, Lecture Notes in Math., 2062, Springer, Heidelberg, 2013
[2] L. Ambrosio, N. Gigli, G. Savaré: Gradient flows in metric spaces and in the space of probability measures. Second edition. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2008.
[3] Y. Brenier: Polar factorization and monotone rearrangement of vector-valued functions, Comm. Pure Appl. Math. 44 (1991), 375–417.
[4] P. Cannarsa, C. Sinestrari: Semiconcave functions, Hamilton-Jacobi equations, and optimal control. Progress in Nonlinear Differential Equations and their Applications, 58. Birkhäuser Boston, Inc., Boston, MA, 2004.
[5] L.C. Evans, R.F. Gariepy: Measure theory and fine properties of functions. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1992.
[6] D. Gilbarg, N. Trudinger: Elliptic partial differential equations of second order. Reprint of the 1998 edition. Classics in Mathematics. Springer-Verlag, Berlin, 2001.
[7] R. Jordan, D. Kinderlehrer, F. Otto: The variational formulation of the Fokker-Planck equation. SIAM J. Math. Anal. 29 (1998), no. 1, 1–17.
[8] G.M. Lieberman: Second order parabolic differential equations. World Scientific Publishing Co., Inc., River Edge, NJ, 1996.
[9] R.J. McCann: Polar factorization of maps on Riemannian manifolds. Geom. Funct. Anal. 11 (2001), no. 3, 589–608.
[10] F. Otto: The geometry of dissipative evolution equations: the porous medium equation. Comm. Partial Differential Equations 26 (2001), no. 1-2, 101-174.
[11] C. Villani: Topics in optimal transportation. Graduate Studies in Mathematics, 58. American Mathematical Society, Providence, RI, 2003.
[12] C. Villani: Optimal transport. Old and new. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 338. Springer-Verlag, Berlin, 2009

E-mail address: wylee@math.cuhk.edu.hk

Room 216, Lady Shaw Building, The Chinese University of Hong Kong, Shatin, Hong Kong