C. R. Acad. Sci. Paris, Ser. I 355 (2017) 345–353

Contents lists available at ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Numerical analysis

Second-order in time schemes for gradient flows in Wasserstein and geodesic metric spaces

Schémas d'ordre deux en temps pour des flots de gradient dans des espaces métriques géodésiques et de Wasserstein

Guillaume Legendre [a], Gabriel Turinici [a,b]

[a] *Université Paris-Dauphine, PSL Research University, CNRS, UMR 7534, CEREMADE, 75016 Paris, France*
[b] *Institut universitaire de France, France*

A R T I C L E   I N F O

Article history:
Received 17 November 2016
Accepted 2 February 2017
Available online 20 February 2017

Presented by Olivier Pironneau

A B S T R A C T

The time discretization of gradient flows in metric spaces uses variants of the celebrated implicit Euler-type scheme of Jordan, Kinderlehrer, and Otto [9]. We propose in this Note a different approach, which allows us to construct two second-order in time numerical schemes. In a metric space framework, we show that the schemes are well defined and prove the convergence for one of them under some regularity assumptions. For the particular case of a Fokker–Planck gradient flow in the Wasserstein space, we obtain (theoretically and numerically) the second-order convergence.

R É S U M É

La discrétisation temporelle des flots de gradient dans des espaces métriques utilise des variantes du schéma d'Euler implicite issu du travail séminal de Jordan, Kinderlehrer et Otto [9]. Nous proposons dans cette Note une approche différente, permettant de construire deux schémas numériques d'ordre deux en temps. Dans le cadre d'un espace métrique, nous montrons que les schémas sont bien définis et prouvons la convergence de l'un d'entre eux sous des hypothèses de régularité. Pour le cas particulier d'un flot de gradient Fokker–Planck dans l'espace de Wasserstein, nous obtenons (théoriquement et numériquement) la convergence à l'ordre deux.

1. Background on gradient flows and relationship with the implicit Euler scheme

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a smooth convex function and $\bar{x}$ be a point in $\mathbb{R}^d$; a gradient flow starting from $\bar{x}$ is a curve $(x_t)_{t\geq 0}$ solution to the Cauchy problem $x_t' = -\nabla f(x_t)$ for $t > 0$, $x_0 = \bar{x}$. A topic that has received considerable attention lately is

E-mail addresses: Guillaume.Legendre@dauphine.fr (G. Legendre), Gabriel.Turinici@dauphine.fr (G. Turinici).

http://dx.doi.org/10.1016/j.crma.2017.02.001

the extension of this definition to a Polish metric space $(\mathcal{X}, d)$ and a functional $F : (\mathcal{X}, d) \to \mathbb{R} \cup \{+\infty\}$ (see [3,14,2,13] for instance). An important advance was in particular the work of Jordan, Kinderlehrer and Otto [9], who introduced the following numerical scheme: given a time step $\tau > 0$ and $\bar{x}$ in $\mathcal{X}$, define recursively the sequence $(x_n^\tau)_{n\in\mathbb{N}}$ such that $x_0^\tau = \bar{x}$ and, for any natural integer n, x_{n+1}^τ is selected as a minimizer of the functional

$$x \mapsto P_F^{\mathrm{JKO}}(x; x_n^\tau, \tau) := \frac{1}{2\tau}\,\mathrm{d}^2(x_n^\tau, x) + F(x). \tag{1}$$

When the metric space is a Hilbert space and the functional F is smooth enough, the above scheme amounts to the implicit Euler scheme, i.e. $\frac{x_{n+1}^\tau - x_n^\tau}{\tau} = -\nabla F(x_{n+1}^\tau)$. As a consequence, the Jordan–Kinderlehrer–Otto method (abbreviated as JKO from now on) can be seen as a variational generalization of the implicit Euler scheme in a metric space. Coupled with a suitable interpolation, the sequence constructed by this scheme defines, as the time step τ tends to zero, a (metric) gradient flow, the convergence being of order one in τ (see [3, Theorem 4.0.4] for a sharp estimate in a particular setting). As a natural numerical counterpart of the theoretical procedure, the JKO scheme was subsequently used to compute the gradient flow in applications (see [10,5,4] for instance).

However, while the JKO scheme has theoretical advantages, its first-order convergence may be judged insufficient in practice. When the gradient flow is regular enough, one may seek to replace it with a second-order alternative, or even combine a second-order scheme with the JKO scheme in order to construct a time-step selection mechanism (see [12, Chapter 17, Section 17.2] for instance). The purpose of this Note is to propose two second-order schemes, which are both easy to implement (one of them directly uses the JKO scheme), and provide several preliminary theoretical and numerical results pertaining to them.

2. Second-order numerical schemes on metric spaces

Let $(\mathcal{X}, d)$ be a Polish metric space. We recall (see [2, page 37]) that a curve $\gamma : [0,1] \to \mathcal{X}$ is called a (constant speed) geodesic provided that $d(\gamma(t), \gamma(s)) = |t-s| \cdot d(\gamma(0), \gamma(1))$, and that the space $(\mathcal{X}, d)$ is called geodesic if, for any couple of points $(x, y) \in \mathcal{X}^2$, there exists a geodesic γ connecting them, that is, such that $\gamma(0) = x$ and $\gamma(1) = y$. We will denote by $Geod_{x,y}$ the set of all geodesics connecting x and y. Moreover, the set $\{\gamma(\frac{1}{2}) \mid \gamma \in Geod_{x,y}\}$ will be called the set of midpoints of the couple (x, y) and will be denoted $\frac{x+y}{2}$ (see [7, Chapter 2] for a definition in general metric spaces and [1] in the particular case of Wasserstein spaces).

2.1. Definition of the schemes

Given a time step τ, we define the *Variational Implicit Midpoint (VIM) scheme*, starting from $\bar{x}$ in $\mathcal{X}$, for the (gradient flow of the) functional F by setting $x_0^\tau = \bar{x}$ and, for any natural integer n, recursively choosing

$$x_{n+1}^\tau \in \operatorname*{argmin}_{y\in\mathcal{X}} P_F^{\mathrm{VIM}}(y; x_n^\tau, \tau), \tag{2}$$

where

$$\forall (x, y) \in \mathcal{X}^2,\ P_F^{\mathrm{VIM}}(y; x, \tau) = \inf\left\{ \frac{d^2(x, y)}{2\tau} + 2F(u) \,\middle|\, u \in \frac{x+y}{2} \right\}. \tag{3}$$

In the particular case of a Hilbert space and a smooth functional F, the point x_{n+1}^τ is effectively a solution to the equation $\frac{x - x_n^\tau}{\tau} + \nabla F(\frac{x_n^\tau + x}{2}) = 0$, which characterizes the implicit midpoint rule method, a one-stage implicit Runge–Kutta method of second order (see [8, Chapter II, Section 7]).

We now prove that the VIM scheme is well defined by establishing the existence of a solution to problem (2).

Theorem 2.1. *Let $(\mathcal{X}, d)$ be a Polish metric space such that,*

$$\text{for any } x \text{ in } \mathcal{X},\ \text{the set } \bigcup_{y\in\mathcal{X}} \frac{x+y}{2} \text{ is closed.} \tag{4}$$

Assume that the functional F is lower semicontinuous, bounded from below, and such that, for all $r > 0$, and $c \in \mathbb{R}$, the set $\{y \in \mathcal{X} \mid F(y) \le c,\ d(\tilde{x}, y) \le r\}$ is compact for some $\tilde{x}$ in $\mathcal{X}$. Then, there exists x in $\mathcal{X}$ such that $P_F^{\mathrm{VIM}}(x; \tilde{x}, \tau) = \inf_{y\in\mathcal{X}} P_F^{\mathrm{VIM}}(y; \tilde{x}, \tau)$.

Proof. Let $(x_n, u_n)_{n\in\mathbb{N}}$ be a minimizing sequence for $\inf_{y\in\mathcal{X}} P_F^{\mathrm{VIM}}(y; \tilde{x}, \tau)$. In particular, the sequence $\left(\frac{1}{2}\, d^2(\tilde{x}, x_n)\right)_{n\in\mathbb{N}}$ is bounded and so is $\left(d(\tilde{x}, u_n)\right)_{n\in\mathbb{N}}$. On the other hand, the sequence $(F(u_n))_{n\in\mathbb{N}}$ is also bounded, so that, by hypothesis, the sequence $(u_n)_{n\in\mathbb{N}}$ lives in a compact set and thus converges, up to a subsequence, to some u. In addition, since $\cup_{y\in\mathcal{X}} \frac{\tilde{x}+y}{2}$ is closed, there exists x in $\mathcal{X}$ such that $u \in \frac{\tilde{x}+x}{2}$. Recalling that $u_n \in \frac{\tilde{x}+x_n}{2}$, we obtain that $d(\tilde{x}, x) = 2d(\tilde{x}, u) = 2\lim_{n\to\infty} d(\tilde{x}, u_n) = \lim_{n\to\infty} d(\tilde{x}, x_n)$. The lower semicontinuity property of F then allows us to conclude. $\square$

Remark 1. The assumptions on the functional F in the above theorem are classical (see [2, Section 4.2.2] for instance), but can be weakened. In particular, we may not suppose F to be bounded from below.

One may observe that, for any natural integer n, any point x in $\mathcal{X}$ and any midpoint u in $\frac{x_n^\tau+x}{2}$, it holds

$$\frac{1}{2\tau}\,d^2(x_n^\tau,x)+2\,F(u)=\frac{2}{\tau}\,d^2(x_n^\tau,u)+2\,F(u)=2\,P_F^{\mathrm{JKO}}\left(u;x_n^\tau,\frac{\tau}{2}\right). \tag{5}$$

As a consequence, another interpretation of the VIM scheme is to view the midpoint u as computed by a JKO scheme with a halved time step $\frac{\tau}{2}$ and to *extrapolate* it as follows.

Recall that a geodesic space $(\mathcal{X},d)$ is called non-branching (see [2, Definition 3.15] or [3, Theorem 11.2.10]) if, for any $t\in]0,1[$, a constant speed geodesic γ is uniquely determined by its initial point $\gamma(0)$ and by the point $\gamma(t)$. In particular, if $u\in\frac{x+y_1}{2}\cap\frac{x+y_2}{2}$ then $y_1=y_2$. By definition, the 2-extrapolate of the couple (x,u) (denoted $2u-x$ hereafter) is the unique point y such that $u\in\frac{x+y}{2}$. On the contrary, when there exists no such point, the 2-extrapolate is defined as the point y with largest distance to x such that u is on a geodesic from x to y. When the space is non-branching and complete, using the properties of the distance between two points on a geodesic and the fact that any Cauchy sequence converges, we obtain that, for any fixed points x and u, the set $\{d(x,y)\,|\,\exists\gamma\in Geod_{x,y},\ u\in\gamma\}$ is closed. The 2-extrapolate is thus well defined and unique. It moreover satisfies $d(x,u)\le d(x,2\,u-x)\le 2\,d(x,u)$.

These considerations lead to the definition of the *Extrapolated Variational Implicit Euler (EVIE) scheme*, in which x_{n+1}^τ is now the 2-extrapolate $2u-x_n^\tau$ of the point u computed by the JKO scheme with time step $\frac{\tau}{2}$. Implementation-wise, the last scheme has the obvious advantage of using the JKO scheme. From a theoretical point of view, the existence and uniqueness of its solution in a non-branching space is assured under the same hypotheses that those required for the JKO scheme. While adaptations would certainly allow us to define it in branching spaces, uniqueness may be lost, even when the JKO scheme possesses a unique solution at each step.

2.2. Convergence of the VIM scheme

In this section, we prove the convergence of the VIM scheme in a metric space setting by showing that a natural interpolation of its discrete solution converges to a gradient flow written in the integral form of an Energy Dissipation Inequality (EDI) (see [2, Section 4.2]). More precisely, we recall that, given a functional $F:(\mathcal{X},d)\to\mathbb{R}\cup\{+\infty\}$ with domain $D(F)=\{x\in\mathcal{X},\ F(x)<+\infty\}$, and a point $\bar{x}\in D(F)$, the curve $(x_t)_{t\ge0}$ in $\mathcal{X}$ is a gradient flow in the EDI sense starting at $\bar{x}$ provided that it is a locally absolutely continuous curve, $x_0=\bar{x}$ and

$$\forall s\ge0,\ F(x_s)+\frac{1}{2}\int_0^s\left|x_r'\right|\,\mathrm{d}r+\frac{1}{2}\int_0^s|\nabla F|^2\,(x_r)\,\mathrm{d}r\le F(\bar{x}), \tag{6}$$

$$\text{a.e. } t>0,\ \forall s\ge t,\ F(x_s)+\frac{1}{2}\int_t^s\left|x_r'\right|\,\mathrm{d}r+\frac{1}{2}\int_t^s|\nabla F|^2\,(x_r)\,\mathrm{d}r\le F(x_t), \tag{7}$$

where, for any point x in $D(F)$, the slope of F at x is

$$|\nabla F|\,(x)=\limsup_{z\to x}\frac{(F(x)-F(z))^+}{d(x,z)}=\max\left\{\limsup_{z\to x}\frac{F(x)-F(z)}{d(x,z)},0\right\},$$

and the metric derivative of x at r is

$$\left|x_r'\right|=\lim_{h\to0}\frac{d(x_{r+h},x_r)}{|h|}. \tag{8}$$

The time step τ being fixed with $0<\tau<\bar{\tau}$, we introduce a variational interpolation *à la* De Giorgi of the discrete solution to the VIM scheme by defining the curve $(x_t^\tau)_{t\ge0}$ as follows:

- $x_0^\tau=\bar{x}$,
- for $n\in\mathbb{N}$, $x_{(n+1)\tau}^\tau\in\operatorname{argmin}_{y\in\mathcal{X}}P_F^{\mathrm{VIM}}(y;x_{n\tau}^\tau,\tau)$,
- for $n\in\mathbb{N}$ and $t\in]n\tau,(n+1)\tau[$, $x_t^\tau\in\operatorname{argmin}_{y\in\mathcal{X}}P_F^{\mathrm{VIM}}(y;x_{n\tau}^\tau,t-n\tau)$.

For such a map, we define the discrete speed $\mathrm{Dsp}^\tau:[0,+\infty)\to[0,+\infty)$ by

$$\mathrm{Dsp}_t^\tau=\frac{d(x_{n\tau}^\tau,x_{(n+1)\tau}^\tau)}{\tau}\ \text{for } t \text{ in } (n\tau,(n+1)\tau),$$

and the discrete slope $\mathrm{Dsl}^\tau:[0,+\infty)\to[0,+\infty)$ by

$$\mathrm{Dsl}_t^\tau = \frac{d(x_{n\tau}^\tau, x_t^\tau)}{t - n\tau} \text{ for } t \text{ in } (n\tau, (n+1)\tau).$$

In order to avoid unnecessary technicalities in the forthcoming analysis, we will suppose from now on that midpoints are unique, that is

$$\forall (x,y) \in \mathcal{X}^2, \ \text{ the set } \frac{x+y}{2} \text{ is a singleton.} \tag{9}$$

We finally introduce a notion of slope of F at the midpoint of (x, y) by setting

$$\left|\nabla^M F\right|(x,y) = \limsup_{z \to y} \frac{\left(F\left(\frac{x+y}{2}\right) - F\left(\frac{x+z}{2}\right)\right)^+}{d\left(\frac{x+y}{2}, \frac{x+z}{2}\right)}.$$

Theorem 2.2. *Let $T > 0$ be fixed and $(\mathcal{X}, d)$ be a Polish metric space satisfying hypotheses (4) and (9). We moreover assume that*

(i) *F is lower semicontinuous, bounded from below, and such that*

$$\forall r > 0, \ \forall c \in \mathbb{R}, \ \forall x \in \mathcal{X} \text{ the set } \{y \in \mathcal{X} \mid F(y) \le c, \ d(x,y) \le r\} \text{ is compact,} \tag{10}$$

(ii) *F has the following continuity property*

$$\text{if } x_n \to x, \text{ and } \sup\{|\nabla F|(x_n), E(x_n)\} < \infty \text{ then } F(x_n) \to F(x); \tag{11}$$

(iii) *$\left|\nabla^M F\right| : D(F) \times D(F) \to [0, \infty)$ has the following semicontinuity property: for any x in $D(F)$ and any two sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ in $D(F)$ converging to x, it holds*

$$\left|\nabla^M F\right|(x,x) \le \liminf_{n\to\infty} \left|\nabla^M F\right|(x_n, y_n),$$

(iv) *if any two of the elements x, y, $\frac{x+y}{2}$ belong to $D(F)$, then the third also does and*

$$\left| \frac{F(x) + F(y) - 2F(\frac{x+y}{2})}{d^2(x,y)} \right| \le H, \tag{12}$$

where H is a constant independent of x and y.

Then, for some $\bar\tau > 0$, the set of curves $\{(x_t^\tau)_{t\in[0,T]}; 0 \le \tau \le \bar\tau\}$ is relatively compact (with respect to the local uniform convergence) and any limit curve is a gradient flow in the EDI formulation (6)–(7).

Remark 2. Assumptions (i) and (ii) are classical and used in [2, Assumption 4.13], whereas (iii) is a generalization of the second assumption in [2, Assumption 4.13]. Item (iv) is however a regularity property specific to our setting; in particular it is satisfied when both F and $-F$ are λ-convex in the sense of [3, Section 2.4 page 49]. On the other hand, it can be weakened to $F(x) + F(y) - 2F(\frac{x+y}{2}) = o(d(x,y))$.

Proof. We will follow closely and adapt where necessary the method of proof given in [2, Subsection 4.2.2]. The proof is divided into several steps, starting with the derivation of some key properties of the variational interpolation $(x_t^\tau)_{t\ge0}$ introduced above.

(I) First, we shall show that the positive real number τ and the natural integer n being fixed, the function $(0,1] \ni \theta \mapsto \frac{1}{2\theta\tau} d^2(x_{n\tau}^\tau, x_{(n+\theta)\tau}^\tau) + 2F\left(\frac{x_{n\tau}^\tau + x_{(n+\theta)\tau}^\tau}{2}\right)$ is locally Lipschitz and that its derivative is given by

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{1}{2\theta\tau} d^2(x_{n\tau}^\tau, x_{(n+\theta)\tau}^\tau) + 2F\left(\frac{x_{n\tau}^\tau + x_{(n+\theta)\tau}^\tau}{2}\right)\right) = -\frac{1}{2\theta^2\tau} d^2(x_{n\tau}^\tau, x_{(n+\theta)\tau}^\tau). \tag{13}$$

Indeed, let $0 < \theta_0 < \theta < \theta_1 \le 1$. By the definition of the variational interpolation, one has

$$\frac{1}{2\theta_0\tau} d^2(x_{n\tau}^\tau, x_{(n+\theta_0)\tau}^\tau) + 2F\left(\frac{x_{n\tau}^\tau + x_{(n+\theta_0)\tau}^\tau}{2}\right) \le \frac{1}{2\theta_0\tau} d^2(x_{n\tau}^\tau, x_{(n+\theta_1)\tau}^\tau) + 2F\left(\frac{x_{n\tau}^\tau + x_{(n+\theta_1)\tau}^\tau}{2}\right),$$

so that

$$\frac{1}{2\theta_0\tau}\,d^2(x^\tau_{n\tau},x^\tau_{(n+\theta_0)\tau})+2\,F\left(\frac{x^\tau_{n\tau}+x^\tau_{(n+\theta_0)\tau}}{2}\right)-\frac{1}{2\theta_0\tau}\,d^2(x^\tau_{n\tau},x^\tau_{(n+\theta_1)\tau})-2\,F\left(\frac{x^\tau_{n\tau}+x^\tau_{(n+\theta_1)\tau}}{2}\right)$$
$$\leq\frac{(\theta_1-\theta_0)\tau}{2\theta_0\theta_1\tau^2}\,d^2(x^\tau_{n\tau},x^\tau_{(n+\theta_1)\tau}).$$

Arguing symmetrically, one also has

$$\frac{1}{2\theta_0\tau}\,d^2(x^\tau_{n\tau},x^\tau_{(n+\theta_0)\tau})+2\,F\left(\frac{x^\tau_{n\tau}+x^\tau_{(n+\theta_0)\tau}}{2}\right)-\frac{1}{2\theta_0\tau}\,d^2(x^\tau_{n\tau},x^\tau_{(n+\theta_1)\tau})-2\,F\left(\frac{x^\tau_{n\tau}+x^\tau_{(n+\theta_1)\tau}}{2}\right)$$
$$\geq\frac{(\theta_1-\theta_0)\tau}{2\theta_0\theta_1\tau^2}\,d^2(x^\tau_{n\tau},x^\tau_{(n+\theta_0)\tau}),$$

and the function is thus locally Lipschitz. Moreover, dividing by $\theta_1-\theta_0$ the last two inequalities and passing to the limits $\theta_0\uparrow\theta$ and $\theta_1\downarrow\theta$, we obtain the proposed expression for the derivative.

(II) Next, we establish that, with the same notations as above, $\theta\mapsto d(x^\tau_{n\tau},x^\tau_{(n+\theta)\tau})$ is non-decreasing, $\theta\mapsto F\left(\frac{x^\tau_{n\tau}+x^\tau_{(n+\theta)\tau}}{2}\right)$ is non-increasing and that it holds

$$\forall\theta\in(0,1],\ \left|\nabla^M F\right|(x^\tau_{n\tau},x^\tau_{(n+\theta)\tau})\leq\frac{d(x^\tau_{n\tau},x^\tau_{(n+\theta)\tau})}{\theta\tau}.\tag{14}$$

Let $0<\theta_0<\theta_1\leq 1$. By the respective minimality properties of $x^\tau_{(n+\theta_0)\tau}$ and $x^\tau_{(n+\theta_1)\tau}$, we have

$$\frac{1}{2\theta_0\tau}\,d^2(x^\tau_{n\tau},x^\tau_{(n+\theta_0)\tau})+2\,F\left(\frac{x^\tau_{n\tau}+x^\tau_{(n+\theta_0)\tau}}{2}\right)\leq\frac{1}{2\theta_0\tau}\,d^2(x^\tau_{n\tau},x^\tau_{(n+\theta_1)\tau})+2\,F\left(\frac{x^\tau_{n\tau}+x^\tau_{(n+\theta_1)\tau}}{2}\right),$$
$$\frac{1}{2\theta_1\tau}\,d^2(x^\tau_{n\tau},x^\tau_{(n+\theta_1)\tau})+2\,F\left(\frac{x^\tau_{n\tau}+x^\tau_{(n+\theta_1)\tau}}{2}\right)\leq\frac{1}{2\theta_1\tau}\,d^2(x^\tau_{n\tau},x^\tau_{(n+\theta_0)\tau})+2\,F\left(\frac{x^\tau_{n\tau}+x^\tau_{(n+\theta_0)\tau}}{2}\right).$$

Adding the last two inequalities, we get

$$\left(\frac{1}{\theta_0}-\frac{1}{\theta_1}\right)\left(d^2(x^\tau_{n\tau},x^\tau_{(n+\theta_0)\tau})-d^2(x^\tau_{n\tau},x^\tau_{(n+\theta_1)\tau})\right)\leq 0,$$

so that $d(x^\tau_{n\tau},x^\tau_{(n+\theta_0)\tau})\leq d(x^\tau_{n\tau},x^\tau_{(n+\theta_1)\tau})$. From this, we now have

$$\frac{1}{2\theta_1\tau}\,d^2(x^\tau_{n\tau},x^\tau_{(n+\theta_0)\tau})+2\,F\left(\frac{x^\tau_{n\tau}+x^\tau_{(n+\theta_1)\tau}}{2}\right)\leq\frac{1}{2\theta_1\tau}\,d^2(x^\tau_{n\tau},x^\tau_{(n+\theta_1)\tau})+2\,F\left(\frac{x^\tau_{n\tau}+x^\tau_{(n+\theta_1)\tau}}{2}\right),$$

which implies, using a previous inequality, that $F\left(\frac{x^\tau_{n\tau}+x^\tau_{(n+\theta_1)\tau}}{2}\right)\leq F\left(\frac{x^\tau_{n\tau}+x^\tau_{(n+\theta_0)\tau}}{2}\right)$.
Finally, by the definition of $x^\tau_{(n+\theta)\tau}$, one has

$$\forall y\in\mathcal{X},\ \frac{1}{2\theta\tau}\,d^2(x^\tau_{n\tau},x^\tau_{(n+\theta)\tau})+2\,F\left(\frac{x^\tau_{n\tau}+x^\tau_{(n+\theta)\tau}}{2}\right)\leq\frac{1}{2\theta\tau}\,d^2(x^\tau_{n\tau},y)+2\,F\left(\frac{x^\tau_{n\tau}+y}{2}\right),$$

so that, by the very definition of the set of midpoints between two points,

$$\frac{2}{\theta\tau}\,d^2\left(x^\tau_{n\tau},\frac{x^\tau_{n\tau}+x^\tau_{(n+\theta)\tau}}{2}\right)+2F\left(\frac{x^\tau_{n\tau}+x^\tau_{(n+\theta)\tau}}{2}\right)\leq\frac{2}{\theta\tau}\,d^2\left(x^\tau_{n\tau},\frac{x^\tau_{n\tau}+y}{2}\right)+2\,F\left(\frac{x^\tau_{n\tau}+y}{2}\right).$$

Hence, we get

$$\frac{F\left(\frac{x^\tau_{n\tau}+x^\tau_{(n+\theta)\tau}}{2}\right)-F\left(\frac{x^\tau_{n\tau}+y}{2}\right)}{d\left(\frac{x^\tau_{n\tau}+x^\tau_{(n+\theta)\tau}}{2},\frac{x^\tau_{n\tau}+y}{2}\right)}\leq\frac{1}{\theta\tau}\,\frac{\left(d\left(x^\tau_{n\tau},\frac{x^\tau_{n\tau}+x^\tau_{(n+\theta)\tau}}{2}\right)-d\left(x^\tau_{n\tau},\frac{x^\tau_{n\tau}+y}{2}\right)\right)\left(d\left(x^\tau_{n\tau},\frac{x^\tau_{n\tau}+x^\tau_{(n+\theta)\tau}}{2}\right)+d\left(x^\tau_{n\tau},\frac{x^\tau_{n\tau}+y}{2}\right)\right)}{d\left(\frac{x^\tau_{n\tau}+x^\tau_{(n+\theta)\tau}}{2},\frac{x^\tau_{n\tau}+y}{2}\right)}$$

Using the reverse triangle inequality and taking the upper limit as y tends to $x^\tau_{(n+\theta)\tau}$, we obtain that

$$\left|\nabla^M F\right|(x^\tau_{n\tau}, x^\tau_{(n+\theta)\tau}) \le \frac{1}{\theta\tau}\limsup_{y\to x^\tau_{(n+\theta)\tau}} \left(d\left(x^\tau_{n\tau}, \frac{x^\tau_{n\tau}+x^\tau_{(n+\theta)\tau}}{2}\right) + d\left(x^\tau_{n\tau}, \frac{x^\tau_{n\tau}+y}{2}\right)\right)$$
$$= \frac{2}{\theta\tau} d\left(x^\tau_{n\tau}, \frac{x^\tau_{n\tau}+x^\tau_{(n+\theta)\tau}}{2}\right) = \frac{1}{\theta\tau} d(x^\tau_{n\tau}, x^\tau_{(n+\theta)\tau}).$$

(III) We now prove that for any integers n and m such that $0 \le n < m$, one has

$$\frac{1}{2}\int_{n\tau}^{m\tau} (\mathrm{Dsp}^\tau_r)^2\,\mathrm{d}r + 2\sum_{k=n}^{m-1} F\left(\frac{x^\tau_{k\tau}+x^\tau_{(k+1)\tau}}{2}\right) - 2\sum_{k=n}^{m-1} F(x^\tau_{k\tau}) = -\frac{1}{2}\int_{n\tau}^{m\tau} (\mathrm{Dsl}^\tau_r)^2\,\mathrm{d}r. \tag{15}$$

From (13), we infer that

$$\forall k\in\mathbb{N},\ \frac{1}{2\tau} d(x^\tau_{k\tau}, x^\tau_{(k+1)\tau}) + 2\,F\left(\frac{x^\tau_{k\tau}+x^\tau_{(k+1)\tau}}{2}\right) - 2\,F(x^\tau_{k\tau}) = -\int_0^1 \frac{d^2(x^\tau_{k\tau}, x^\tau_{(k+\theta)\tau})}{2\theta^2\tau}\,\mathrm{d}\theta,$$

which yields, by a change of variable and the respective definitions of the discrete speed and slopes,

$$\forall k\in\mathbb{N},\ \forall t\in(k\tau,(k+1)\tau),\ \frac{\tau}{2}(\mathrm{Dsp}^\tau_t)^2 + 2\,F\left(\frac{x^\tau_{k\tau}+x^\tau_{(k+1)\tau}}{2}\right) - 2\,F(x^\tau_{k\tau}) = -\frac{1}{2}\int_{k\tau}^{(k+1)\tau} (\mathrm{Dsl}^\tau_r)^2\,\mathrm{d}r. \tag{16}$$

The result follows from summing from $k = n$ to $m-1$.

(IV) Combining relation (15) with hypothesis (12), we obtain

$$\left|\frac{1}{2}\int_{n\tau}^{m\tau} (\mathrm{Dsp}^\tau_r)^2\,\mathrm{d}r + F(x^\tau_{m\tau}) - F(x^\tau_{n\tau}) + \frac{1}{2}\int_{n\tau}^{m\tau} (\mathrm{Dsl}^\tau_r)^2\,\mathrm{d}r\right| \le H\tau\int_{n\tau}^{m\tau} (\mathrm{Dsp}^\tau_r)^2\,\mathrm{d}r. \tag{17}$$

Therefore, for $t \le T$ and $T = n\tau$, it holds

$$d^2(x^\tau_t, x^\tau_0) \le \left(\int_0^T \mathrm{Dsp}^\tau_r\,\mathrm{d}r\right)^2 \le T\int_0^T (\mathrm{Dsp}^\tau_r)^2\mathrm{d}r \le \frac{2T}{1-2H\tau}(F(x^\tau_0) - \inf F), \tag{18}$$

which shows that, for any T, the set $\{x^\tau_t\}_{0\le t\le T}$ is bounded (uniformly with respect to τ). Using (10), we conclude that it is relatively compact. Moreover, further exploitation of inequality (17) shows that for $t = n\tau < m\tau = s$, one has

$$d^2(x^\tau_t, x^\tau_s) \le \left(\int_t^s \mathrm{Dsp}^\tau_r\,\mathrm{d}r\right)^2 \le \frac{2(s-t)}{1-2H\tau}(F(x^\tau_0) - \inf F), \tag{19}$$

which gives equicontinuity and, by the Arzelà–Ascoli theorem, the relative compactness of the set of curves $\{(x^\tau_t)_{0\le t\le T}; 0 \le \tau \le \bar\tau\}$ with respect to the local uniform convergence.

(V) We finally pass to the limit. Let $(\tau_n)_{n\in\mathbb{N}}$ be a decreasing sequence tending to zero such that $(x^{\tau_n}_t)$ converges to a limit curve x_t locally uniformly as n tends to infinity. Using inequality (19), this curve is absolutely continuous and satisfies

$$\forall 0\le t<s,\ \int_t^s \left|x'_r\right|\,\mathrm{d}r \le \liminf_{n\to+\infty}\int_t^s (\mathrm{Dsp}^{\tau_n}_r)^2\,\mathrm{d}r. \tag{20}$$

In addition, set t and, $\forall n\in\mathbb{N}$, let $N(n)\in\mathbb{N}$ be such that $N(n)\tau_n < t \le (N(n)+1)\tau_n$, so that $N(n)\tau_n$ tends to t as n tends to infinity. Using assumption (iii) on the lower semicontinuity of $\left|\nabla^M F\right|$ and (14), we have, on the one hand,

$$\left|\nabla^M F\right|(x_t, x_t) \le \liminf_{n\to+\infty}\left|\nabla^M F\right|(x^{\tau_n}_{N(n)\tau_n}, x^{\tau_n}_t) \le \liminf_{n\to+\infty}\frac{d(x^{\tau_n}_{N(n)\tau_n}, x^{\tau_n}_t)}{t-N(n)\tau_n} = \liminf_{n\to+\infty}\mathrm{Dsl}^{\tau_n}_t.$$

On the other hand, it follows from (12) that

$$\forall t>0,\ \left|\nabla^M F\right|(x_t,x_t)=\limsup_{z\to x_t}\frac{\left(F(x_t)-F\left(\frac{x_t+z}{2}\right)\right)^+}{d\left(x_t,\frac{x_t+z}{2}\right)}=\limsup_{z\to x_t}\frac{2\left(F(x_t)-F\left(\frac{x_t+z}{2}\right)\right)^+}{d(x_t,z)}$$
$$=\limsup_{z\to x_t}\frac{(2F(x_t)-(F(x_t)+F(z)))^+}{d(x_t,z)}=\limsup_{z\to x_t}\frac{(F(x_t)-F(z))^+}{d(x_t,z)}=|\nabla F|(x_t).$$

By Fatou's lemma and (17), we thus obtain that, for $\tau_n\le 1/(2H)$,

$$\forall t<s,\ \int_t^s |\nabla F|^2(x_r)\,\mathrm{d}r\le\liminf_{n\to+\infty}\int_t^s(\mathrm{Dsl}_r^{\tau_n})^2\,\mathrm{d}r\le 2\,(F(x_0^\tau)-\inf F). \tag{21}$$

Passing to the limit in (17) using (20) and (21) for $t=0$ and s arbitrary yields (6). Note that the same technique does not work for $t>0$, since F is not necessarily continuous. However, from (21), it follows that, for almost every $t>0$ (taken as fixed from now on), there exists a subsequence τ_{n_k} with $\sup_k|\nabla F|(x_t^{\tau_{n_k}})<\infty$. By assumption (11), we then obtain (7). $\square$

3. Second-order schemes for the Fokker–Planck equation as a gradient flow in the Wasserstein space

As a complement to the above results, we consider in this final section a particular gradient flow in the one-dimensional case, for which the results of Section 2.2 do not apply, but which provides evidence that the VIM and EVIE schemes have good numerical properties. We refer to [11,4] for possible approaches in higher dimensions compatible with these schemes.

The space $\mathcal{X}$ is now set to be the Wasserstein space of order 2, that is the set of probability measures on $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ with finite second-order moment, $\mathcal{P}_2(\mathbb{R})$, endowed with the 2-Wasserstein distance $\mathcal{W}_2$ (see [13, Chapter 5] or [9,14]). Note that, from now on, the variable x will denote the position over the real line.

Consider a probability measure ν in $\mathcal{P}_2(\mathbb{R})$. When ν is absolutely continuous with respect to the Lebesgue measure, with probability density ρ, the functional F is defined by

$$F(\nu)=E(\rho)+S(\rho)\text{ with }E(\rho)=\int_{\mathbb{R}}V(x)\rho(x)\,\mathrm{d}x\text{ and }S(\rho)=\frac{\sigma^2}{2}\int_{\mathbb{R}}\rho(x)\log(\rho(x))\,\mathrm{d}x. \tag{22}$$

Here, the term $E(\rho)$ corresponds to a potential energy, the function V being the potential in question, and the term $S(\rho)$ corresponds to an internal energy, the scalar σ being a positive real number. Otherwise, we set $F(\nu)=+\infty$. Correspondingly, if the measures ν_1 and ν_2 are absolutely continuous with respect to the Lebesgue measure, with respective densities ρ_1 and ρ_2, we denote by $\mathcal{W}_2(\rho_1,\rho_2)$ the Wasserstein distance between ν_1 and ν_2 and by $Geod_{\rho_1,\rho_2}$ the set of geodesics connecting ν_1 to ν_2.

For any smooth potential V, it is well known (see [9]) that the gradient flow in $\mathcal{P}_2(\mathbb{R})$ of the functional F is a curve $t\mapsto\nu(t)$ in $\mathcal{P}_2(\mathbb{R})$, such that, at almost any time t, the measure $\nu(t)$ is absolutely continuous with respect to the Lebesgue measure, with probability density $\rho(t,\cdot)$, and that ρ is solution to the Fokker–Planck partial differential equation of the form

$$\frac{\partial\rho}{\partial t}(t,x)=\frac{\partial}{\partial x}[V'(x)\rho(t,x)]+\frac{\sigma^2}{2}\frac{\partial^2\rho}{\partial x^2}(t,x), \tag{23}$$

which is itself related to the following stochastic differential equation

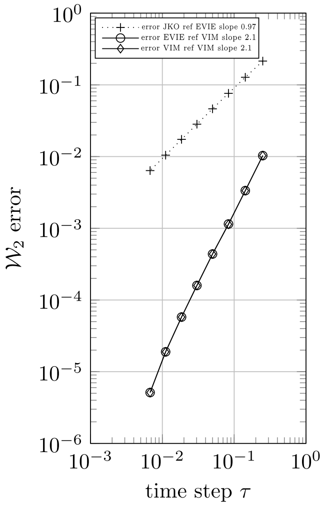
$$\mathrm{d}X(t)=-V'(X(t))\,\mathrm{d}t+\sigma\,\mathrm{d}W(t), \tag{24}$$

the stochastic process W being a standard one-dimensional Wiener process.

In what follows, we consider quadratic potentials of the form $V(x)=\theta\frac{(x-\mu)^2}{2}$, where θ and μ are given constants. We denote by $\mathcal{N}(a,b^2)$ the normal (Gaussian) distribution with mean a and variance b^2 and also, by abusing the notation, the associated density (with respect to the Lebesgue measure) $\rho(x)=\frac{1}{\sqrt{2\pi b^2}}\,\mathrm{e}^{-\frac{(x-a)^2}{2b^2}}$.

Proposition 3.1. *If $\rho(0,\cdot)=\mathcal{N}(\mu_0,\sigma_0^2)$, any intermediary state of the semi-discrete VIM and EVIE schemes has a Gaussian distribution and the error of both schemes has order $O(\tau^2)$.*

Proof. We will prove the assertion for the VIM scheme, the proof for the EVIE one being similar. The potential V being quadratic, the solution to (24) is an Ornstein–Uhlenbeck process, given by $X(t)=X(0)\,\mathrm{e}^{-\theta t}+\mu(1-\mathrm{e}^{-\theta t})+\frac{\sigma}{\sqrt{2\theta}}\mathrm{e}^{-\theta t}W(\mathrm{e}^{2\theta t}-1)$. In particular, the exact evolution from $\rho_n=\mathcal{N}\left(\mu_n,\sigma_n^2\right)$ after τ time units is



No. of time steps	JKO	VIM	EVIE
4	0.56	0.14	0.12
7	0.21	0.22	0.20
12	0.35	0.37	0.29
20	0.57	0.45	0.41
33	0.69	0.73	0.66
54	1.11	1.17	1.07
90	1.81	1.95	1.79
148	2.95	3.21	2.91
244	5.47	5.28	4.79

Fig. 1. Results of numerical simulation for the case considered in Section 3. **Left:** errors of the JKO scheme (dotted line) and of the VIM and EVIE schemes (solid lines) as functions of the length of the time step τ. The error is the $\mathcal{W}_2$ distance between the numerical solution computed by the schemes on grids containing respectively 4, 7, 12, 20, 33, 54, 90, and 148 time steps and a numerical reference solution obtained using the VIM scheme on a finer grid (244 time steps). The solutions to the VIM and EVIE schemes are indistinguishable to machine precision. We observe that any of the second-order schemes on a grid with four steps is as accurate as the standard JKO scheme on a grid with 90 steps. Similarly, the VIM or EVIE scheme on a grid with seven steps is as good as the JKO scheme on a grid with 148 steps. The numerical estimation of the order of convergence of the JKO scheme and of the VIM and EVIE schemes are given in the legend. **Right:** CPU times (in seconds) with respect to the number of time steps (the implementation uses MATLAB Version: 8.6.0.267246 (R2015b) on an 8-core Intel(R) Core(TM) i7-3770 CPU @ 3.40 GHz HP Thinkstation).

$$\rho_{n+1}^{\text{exact}} = \mathcal{N}\left(\mu_n \mathrm{e}^{-\theta\tau} + \mu(1-\mathrm{e}^{-\theta\tau}), \mathrm{e}^{-2\theta\tau}\sigma_n^2 + \frac{\sigma^2}{2\theta}\left(1-\mathrm{e}^{-2\theta\tau}\right)\right).$$

On the other hand, $\rho_n(x)\mathrm{d}x$ being absolutely continuous with respect to the Lebesgue measure, the set $Geod_{\rho_n,\rho}$ is a singleton for any ρ in $\mathcal{P}_2(\mathbb{R})$ (and so is the set $\frac{\rho_n+\rho}{2}$). Hence, the unique geodesic γ_ρ in $Geod_{\rho_n,\rho}$ is determined by a map $T : \mathbb{R} \to \mathbb{R}$ (which is the gradient of a convex function, see [6,14]), $\gamma_\rho(t)$ being the push-forward of ρ_n by $x \mapsto (1-t)x + tT(x)$. The minimization in (2) can thus be expressed in terms of T, for instance $\frac{1}{2}d^2(\rho_n,\rho) = \int_{\mathbb{R}} \frac{(T(x)-x)^2}{2}\rho_n(x)\,\mathrm{d}x$. After some tedious (but rather straightforward) computations, one obtains that the unique element ρ_{n+1}^{VIM} minimizing $\frac{1}{2\tau}d^2(\rho_n,\rho) + 2F\left(\frac{\rho_n+\rho}{2}\right)$ also has a normal distribution, that is $\rho_{n+1}^{\text{VIM}} = \mathcal{N}\left(\mu_{n+1}, \sigma_{n+1}^2\right)$, with $\mu_{n+1} = \left(1+\frac{\theta\tau}{2}\right)^{-1}\left(\mu_n\left(1-\frac{\theta\tau}{2}\right)+\frac{\mu\theta\tau}{2}\right)$ and $\sigma_{n+1} = \left(1+\frac{\theta\tau}{2}\right)^{-1}\sqrt{\sigma_n^2+\sigma^2\tau\left(1+\frac{\theta\tau}{2}\right)} - \frac{\sigma_n\theta\tau}{2}$. Consequently, ρ_{n+1}^{VIM} is an approximation of $\rho_{n+1}^{\text{exact}}$ exact to the second order in τ and, by the usual Gronwall inequalities, the error of the VIM scheme is globally of order $O(\tau^2)$. $\square$

We performed numerical simulations to support these results. Roughly speaking, we employed a continuous piecewise affine discretization for the cumulative distribution functions of elements in $\mathcal{P}_2(\mathbb{R})$, which amounts to replace the space $\mathcal{P}_2(\mathbb{R})$ by a discrete one, denoted $\mathcal{P}_2(\mathbb{R})_m$ and defined as follows. Let $M \in \mathbb{N}^*$ and $m = (m_1, \ldots, m_M)$ be an (ordered) mass distribution such that $m_k > 0$ for any integer k in $\{1, \ldots, M\}$ and $\sum_{k=1}^M m_k = 1$, fixed once and for all. An element in $\mathcal{P}_2(\mathbb{R})_m$ with density ρ_X is then represented by a vector $X = (x_0, \ldots, x_M) \in \mathbb{R}^{M+1}$, such that $x_0 < x_1 < \cdots < x_{M-1} < x_M$, by setting ρ_X constant and equal to $\frac{m_k}{x_{k+1}-x_k}$ on the segment $[x_k, x_{k+1})$.

Approximating the integral defining the discrete potential energy $E(\rho_X)$ by using the composite Simpson quadrature rule (see [12, Chapter 4, Subsection 4.1.3]), one obtains

$$E_m(X) = \sum_{k=0}^{M} \frac{m_k}{6}\left[V(x_k) + 4V\left(\frac{x_k+x_{k+1}}{2}\right) + V(x_{k+1})\right]. \tag{25}$$

On the other hand, the discrete internal energy $S(\rho_X)$ can be computed exactly and has the simple form

$$S_m(X) = \sum_{k=0}^{M} \frac{m_k}{x_{k+1}-x_k}(x_{k+1}-x_k)\log\left[\frac{m_k}{x_{k+1}-x_k}\right] = \sum_{k=0}^{M} m_k \log(m_k) - \sum_{k=0}^{M} m_k \log(x_{k+1}-x_k). \tag{26}$$

To compute the Wasserstein distance $\mathcal{W}_2$ between the discrete densities ρ_{X_1} and ρ_{X_2}, we note that the map T is the unique piecewise linear function on any interval $[x_k^1, x_{k+1}^1]$, which associates x_k^1 with x_k^2 and x_{k+1}^1 with x_{k+1}^2, respectively. Straightforward computations then show that the squared distance $\mathcal{W}_2{}^2(\rho_{X_1}, \rho_{X_2}) = \int_{\mathbb{R}} (T(x) - x)\rho_{X_1}(x)\,\mathrm{d}x$ reduces to

$$\mathcal{W}_2{}^2(\rho_{X_1}, \rho_{X_2}) = \frac{1}{3}\sum_{k=0}^{M} m_k \left[\left(x_k^2 - x_k^1\right)^2 + \left(x_{k+1}^2 - x_{k+1}^1\right)\left(x_k^2 - x_k^1\right) + \left(x_{k+1}^2 - x_{k+1}^1\right)^2 \right]. \tag{27}$$

For the numerical results presented in Fig. 1, we have set $\sigma = 1$, $\theta = \frac{1}{2}$, $\mu = 5$, $M = 32$, a uniform mass step $m_k = \frac{1}{M}$, the final time equal to 1, and the projection on the discrete space of a standard Gaussian distribution as the initial datum. Second-order in time convergence is confirmed for both schemes.

The internal energy $S(\rho_X)$ constitutes a barrier enforcing the constraint $x_k < x_{k+1}$. As such the (continuous in time) dynamics of the probability density can also be viewed as a dynamics of the vector X that follows a cumbersome but regular system of ordinary differential equations. The proposed schemes provide a consistent and second-order in time discretization of this system, and we have obtained the following result.

Corollary 3.1. *The scheme (2) discretized using (25), (26), and (27) converges at second order in time to the exact solution to the semi-discrete gradient flow of the functional F in the Wasserstein space of order* 2.

Acknowledgements

G. Turinici acknowledges support from the Agence Nationale de la Recherche (ANR), projects EMAQS ANR-2011-BS01-017-01 and CINE-PARA ANR-15-CE23-0019.

References

[1] M. Agueh, G. Carlier, Barycenters in the Wasserstein space, SIAM J. Math. Anal. 43 (2) (2011) 904–924.
[2] L. Ambrosio, N. Gigli, A user's guide to optimal transport, in: B. Piccoli, M. Rascle (Eds.), Modelling and Optimisation of Flows on Networks, Cetraro, Italy, 2009, in: Lecture Notes in Mathematics, Springer, 2008, pp. 1–155.
[3] 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, Basel, 2008.
[4] J.-D. Benamou, G. Carlier, Q. Mérigot, É. Oudet, Discretization of functionals involving the Monge–Ampère operator, Numer. Math. 134 (3) (2016) 611–636.
[5] A. Blanchet, V. Calvez, J.A. Carrillo, Convergence of the mass-transport steepest descent scheme for the subcritical Patlak–Keller–Segel model, SIAM J. Numer. Anal. 46 (2) (2008) 691–721.
[6] Y. Brenier, Polar factorization and monotone rearrangement of vector-valued functions, Commun. Pure Appl. Math. 44 (4) (1991) 375–417.
[7] D. Burago, Y. Burago, S. Ivanov, A Course in Metric Geometry, Graduate Studies in Mathematics, vol. 33, American Mathematical Society, 2001.
[8] E. Hairer, S.P. Nørsett, G. Wanner, Solving Ordinary Differential Equations I. Nonstiff Problems, Springer Series in Computational Mathematics, vol. 8, Springer, 1993.
[9] R. Jordan, D. Kinderlehrer, F. Otto, The variational formulation of the Fokker–Planck equation, SIAM J. Math. Anal. 29 (1) (1998) 1–17.
[10] D. Kinderlehrer, N.J. Walkington, Approximation of parabolic equations using the Wasserstein metric, ESAIM Math. Model. Numer. Anal. 33 (4) (1999) 837–852.
[11] J.-M. Mirebeau, Discretization of the 3D Monge–Ampere operator, between wide stencils and power diagrams, ESAIM Math. Model. Numer. Anal. 49 (5) (2015) 1511–1523.
[12] W.H. Press, S.A. Teukolsky, W.T. Veterling, B.P. Flannery, Numerical Recipes: The Art of Scientific Computing, third edition, Cambridge University Press, 2007.
[13] F. Santambrogio, Optimal Transport for Applied Mathematicians. Calculus of Variations, PDEs, and Modeling, Progress in Nonlinear Differential Equations and Their Applications, vol. 87, Birkhäuser, 2015.
[14] C. Villani, Optimal Transport. Old and New, Grundlehren der Mathematischen Wissenschaften, vol. 338, Springer, 2009.