A TROTTER PRODUCT FORMULA FOR GRADIENT FLOWS IN METRIC SPACES

PHILIPPE CLÉMENT AND JAN MAAS

ABSTRACT. We prove a Trotter product formula for gradient flows in metric spaces. This result is applied to establish convergence in the L^2-Wasserstein metric of the splitting method for some Fokker-Planck equations and porous medium type equations perturbed by a potential.

1. INTRODUCTION AND STATEMENT OF THE MAIN RESULTS

In the first part of [1] Ambrosio, Gigli and Savaré developed a rich theory of gradient flows in metric spaces. In particular they studied in great detail the following situation.

Let (X, d) be a complete metric space, and let $\varphi : X \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous functional which is not identically $+\infty$. Associated with φ is the *Moreau-Yosida functional* Φ defined for $h > 0$ and $x \in X$ by

$$\Phi(h, x; y) := \begin{cases} \varphi(y) + \frac{1}{2h} d^2(x, y), & y \in \mathsf{D}(\varphi), \\ +\infty, & \text{otherwise}, \end{cases}$$

where $\mathsf{D}(\varphi) := \{x \in X : \varphi(x) < \infty\}$. In [1] conditions are given which guarantee

(1) existence and uniqueness of a global minimizer of $\Phi(h, x; \cdot)$, which is denoted by $J_h x$ and called the *resolvent* of φ at x;
(2) convergence of sequences of iterated resolvents $\{(J_{t/n})^n x\}_{n \geq 1}$;
(3) the validity of a certain *evolution variational inequality* (EVI) for the limit.

In the second part of [1] the theory is applied to problems in the space of probability measures where the functional φ can be naturally written as the sum of two (or more) functionals φ^i, $i = 1, 2$. It appears that in most cases one can associate with each φ^i a resolvent J_h^i. It is therefore natural to consider the problem of convergence of sequences of iterates of the form $\{(J_{t/n}^2 J_{t/n}^1)^n x\}_{n \geq 1}$ provided they are well-defined and to investigate whether the limit satisfies the EVI associated with φ. In this paper we give sufficient conditions for this to be true (Theorem 1.1).

We apply our abstract results to establish convergence with respect to the L^2-Wasserstein metric of the splitting method for Fokker-Planck equations and porous medium equations with a potential satisfying appropriate conditions.

Let us now present the setting of the paper and state the main results. Throughout the paper, we let (X, d) be a complete metric space. For $i = 1, 2$, let $\varphi^i : X \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous (lsc) functional satisfying

$$\mathsf{D}(\varphi^1) \cap \mathsf{D}(\varphi^2) \neq \varnothing.$$

We consider the functional $\varphi := \varphi^1 + \varphi^2$ defined by

$$\mathsf{D}(\varphi) := \mathsf{D}(\varphi^1) \cap \mathsf{D}(\varphi^2),$$
$$\varphi(x) := \varphi^1(x) + \varphi^2(x), \quad x \in \mathsf{D}(\varphi),$$

and note that $\mathsf{D}(\varphi) \neq \varnothing$, and φ is lower semicontinuous.

We shall impose three assumptions:

(A_1) For $i = 1, 2$, for any $h > 0$ and any $x \in \overline{\mathsf{D}(\varphi^i)}$, the following variational inequality has a solution:

find $y \in \mathsf{D}(\varphi^i)$ satisfying

$$\frac{1}{2h}[d^2(y, z) - d^2(x, z)] + \frac{1}{2h}d^2(y, x) + \varphi^i(y) \leq \varphi^i(z) \qquad (1.1)$$

for all $z \in \mathsf{D}(\varphi^i)$.

Clearly, if $y \in \mathsf{D}(\varphi^i)$ satisfies (1.1), then y is a global minimizer of $\Phi^i(h, x; \cdot)$. Since $\frac{1}{2h}d^2(y, z) + \Phi^i(h, x; y) \leq \Phi^i(h, x; z)$ for every $z \in \mathsf{D}(\varphi^i)$, this global minimizer is unique. We will denote the minimizer by $J_h^i x$. Notice that for $x \in \mathsf{D}(\varphi^i)$ and $h > 0$ we have

$$\varphi^i(J_h^i x) \leq \varphi^i(x),$$

as can be seen by setting $z = x$ in (1.1).

(A_2) For any $h > 0$ we have

$$(i) \qquad J_h^1\big(\overline{\mathsf{D}(\varphi^1)} \cap \mathsf{D}(\varphi^2)\big) \subseteq \overline{\mathsf{D}(\varphi^2)},$$
$$(ii) \qquad J_h^2\big(\mathsf{D}(\varphi^1) \cap \overline{\mathsf{D}(\varphi^2)}\big) \subseteq \mathsf{D}(\varphi^1).$$

For $h > 0$ and $x \in \overline{\mathsf{D}(\varphi^1)} \cap \mathsf{D}(\varphi^2)$ we define

$$K_h x := J_h^2 J_h^1 x.$$

It follows from (A_2) that $K_h x \in \mathsf{D}(\varphi)$. In particular, it follows that K_h maps $\mathsf{D}(\varphi)$ into itself.

A *discretisation* $\mathbf{h}$ is a finite sequence of positive numbers $(h_i)_{i=1}^n \subseteq (0, \infty)$. For $k = 1, \ldots, n$ we set

$$|\mathbf{h}| := \sup_{1 \leq k \leq n} h_k, \qquad t_{\mathbf{h}}^0 := 0, \qquad t_{\mathbf{h}}^k := 2\sum_{j=1}^k h_j.$$

Given $x \in \mathsf{D}(\varphi)$ and a discretisation $\mathbf{h} = (h_i)_{i=1}^n$, the discrete scheme is defined for $k = 1, \ldots, n$ by

$$x_{\mathbf{h}}^0 := x, \qquad \widehat{x}_{\mathbf{h}}^k := J_{h_k}^1 x_{\mathbf{h}}^{k-1}, \qquad x_{\mathbf{h}}^k := J_{h_k}^2 \widehat{x}_{\mathbf{h}}^k = K_{h_k} x_{\mathbf{h}}^{k-1}.$$

We shall associate with the discretisation the piecewise constant function $\underline{x}_{\mathbf{h}}$ (resp. $\overline{x}_{\mathbf{h}}$) : $[0, t_{\mathbf{h}}^n] \to X$ which takes the values $x_{\mathbf{h}}^k$ at $t_{\mathbf{h}}^k$, $0 \le k \le n$, is constant on the intervals $(t_{\mathbf{h}}^{k-1}, t_{\mathbf{h}}^k)$, $1 \le k \le n$, and is right-continuous (resp. left-continuous).

To motivate the next assumption, let us remark that, as we have seen above, for $x \in \mathsf{D}(\varphi^i)$ we have $\varphi^i(J_h^i x) - \varphi^i(x) \le 0$, $i = 1, 2$, but in general we do not have any bound for $\varphi^1(J_h^2 x) - \varphi^1(x)$ with $x \in \mathsf{D}(\varphi^1)$. The next assumption provides some control on this quantity. For $k = 1, \dots, n$ we set

$$\delta_{\mathbf{h},x}^k := [\varphi^1(x_{\mathbf{h}}^k) - \varphi^1(\widehat{x}_{\mathbf{h}}^k)]^+, \qquad \Delta_{\mathbf{h},x}^k := \sum_{j=1}^k \delta_{\mathbf{h},x}^j.$$

We can now state the assumption:

(A_3) There exists a lsc functional $\chi : X \to \mathbb{R} \cup \{+\infty\}$, not identically $+\infty$, such that the following holds: for any $w \in \overline{\mathsf{D}(\varphi)}$ and any $h_*, T, R, U > 0$ there exists $K \in (0, \infty)$ such that for any discretisation $\mathbf{h} = (h_i)_{i=1}^n$ satisfying

$$|\mathbf{h}| \le h_*, \qquad t_{\mathbf{h}}^n \le T, \tag{1.2}$$

and any $x \in \mathsf{D}(\varphi)$ satisfying

$$d^2(x, w) \le R, \qquad \chi(x) \le U, \tag{1.3}$$

we have

$$\Delta_{\mathbf{h},x}^n \le K.$$

Now we are ready to state the main result of this paper.

Theorem 1.1. *Assume that* $(A_1), (A_2), (A_3)$ *hold. Let* $T_* > 0$ *and let* $(\mathbf{h}^i)_{i \ge 1}$ *be a sequence of discretisations, where* $\mathbf{h}^i = (h_k^i)_{k=1}^{n^i}$, *such that*

$$(i) \quad \inf_{i \ge 1} t_{\mathbf{h}^i}^{n^i} \ge T_* \qquad and \qquad (ii) \quad \lim_{i \to \infty} |\mathbf{h}^i| = 0.$$

Let $x \in \mathsf{D}(\varphi)$ *and let* $(x^i)_{i \ge 1} \subseteq \mathsf{D}(\varphi)$ *be a sequence satisfying*

$$(iii) \quad \lim_{i \to \infty} x^i = x, \qquad (iv) \quad \sup_{i \ge 1} \chi(x^i) < \infty, \qquad (v) \quad \sup_{i \ge 1} \varphi(x^i) < \infty.$$

Then the sequences $\{\underline{x}^i_{\mathbf{h}^i}\}_{i \ge 1}$ *and* $\{\overline{x}^i_{\mathbf{h}^i}\}_{i \ge 1}$ *converge uniformly on* $[0, T_*]$ *to a continuous function* $u : [0, T_*] \to X$ *which satisfies* $u(0) = x$, $\varphi \circ u \in L^1((0, T_*); \mathbb{R})$ *and*

$$\frac{d}{dt} \frac{1}{2} d^2(u(t), y) \le \varphi(y) - \varphi(u(t)) \tag{1.4}$$

in the sense of distributions on $(0, T_*)$ *for any* $y \in \mathsf{D}(\varphi)$.

Remark 1.2. More explicitly, (1.4) means that

$$-\frac{1}{2} \int_0^\infty d^2(u(t), y) \zeta'(t) \, dt \le \int_0^\infty \left(\varphi(y) - \varphi(u(t)) \right) \zeta(t) \, dt$$

for any non-negative test function $\zeta \in C_c^\infty((0, \infty); \mathbb{R})$. Equivalently, see, e.g, [4], for any $0 < a < b < \infty$,

$$\frac{1}{2} d^2(u(b), y) - \frac{1}{2} d^2(u(a), y) \le (b - a) \varphi(y) - \int_a^b \varphi(u(t)) \, dt. \tag{1.5}$$

Remark 1.3. Existence and uniqueness of a solution to (1.4) for $x \in \overline{\mathsf{D}(\varphi)}$ has been proved in [1, Theorem 4.0.4] under suitable coercivity and convexity assumptions which imply (A_1) for φ. Here we do not assume that (A_1) holds for φ. Therefore the existence of a gradient flow for φ does not follow from the results in [1].

Remark 1.4. As we observed before, in (A_3) we impose a bound for $\varphi^1(J_h^2 x) - \varphi^1(x)$ with $x \in \mathsf{D}(\varphi^1)$. Note however that we do *not* assume any bound for $\varphi^2(J_h^1 x) - \varphi^2(x)$ with $x \in \mathsf{D}(\varphi^2)$.

Remark 1.5. Convergence of the splitting method is well-known in the case where X is a Hilbert space and each φ^i is a convex functional [2, 7]. If X is a Hilbert space, then our assumptions are more restrictive than the ones in [7].

Remark 1.6. It follows from the theory presented in [1], see also [3], that for $i = 1, 2$, (A_1) implies the existence of a semigroup of operators $S_t^i : \overline{\mathsf{D}(\varphi^i)} \to \overline{\mathsf{D}(\varphi^i)}$, $t \geq 0$, such that for any $x \in \overline{\mathsf{D}(\varphi^i)}$ the function $u^i(t) := S_t^i x$ satisfies (1.4) with $\varphi = \varphi^i$. It appears from its proof that Theorem 1.1 remains valid if we replace one or both of the resolvents with the associated semigroup.

At first sight (A_3) may seem difficult to verify in concrete situations. However, the next result provides some sufficient conditions for (A_3) which are easier to state and which will be shown to be fulfilled in a number of examples in Section 3.

Proposition 1.7. *Assume that (A_1) and (A_2) hold and suppose that φ^1 and φ^2 satisfy at least one of the following conditions:*

(1) *There exists $c \geq 0$ such that for any $h > 0$ and $x \in \mathsf{D}(\varphi)$ we have*
$$\varphi^1(J_h^2 x) \leq \varphi^1(x) + ch;$$

(2) *Assume that $\varphi^1[X] \subseteq [0, \infty]$. Moreover, assume that there exists $\alpha \geq 0$ such that for any $h > 0$ and $x \in \mathsf{D}(\varphi)$ we have*
$$\varphi^1(J_h^2 x) \leq e^{\alpha h} \varphi^1(x);$$

(3) *Assume that $\varphi^2[X] \subseteq [0, \infty]$. Moreover, assume that there exist $\alpha, c \geq 0$ such that for any $h > 0$ and $x \in \mathsf{D}(\varphi)$ we have*
$$\begin{aligned}(i) \qquad & \varphi^1(J_h^2 x) \leq \varphi^1(x) + ch\varphi^2(J_h^2 x), \\ (ii) \qquad & \varphi^2(J_h^1 x) \leq e^{\alpha h} \varphi^2(x).\end{aligned}$$

Then (A_3) is satisfied with
$$(1)\ \chi = constant, \qquad (2)\ \chi = \varphi^1, \qquad (3)\ \chi = \varphi^2.$$

We apply our results to the case where $X = \mathscr{P}_2(\mathbb{R}^d)$, the space of probability measures on $\mathbb{R}^d$ with finite second moment endowed with the L^2-Wasserstein metric. On this space we consider the sum of the (negative) Boltzmann entropy and a potential energy. The associated gradient flow corresponds to the Fokker-Planck equation [6]. We show that the conditions of Proposition 1.7(1) are satisfied under suitable assumptions on the potential, and therefore the splitting method converges in this setting. Using (2) and (3) of Proposition 1.7, we obtain similar results by replacing the

Boltzmann entropy by the Rényi entropy, which corresponds to the porous medium equation [9].

The paper is organised as follows. In Section 2 we shall work in the abstract setting of a metric space and give the proof of Theorem 1.1 and Proposition 1.7. The applications to gradient flows in the Wasserstein space are presented in Section 3.

Acknowledgement. We are grateful to the anonymous referee for his useful comments.

2. Proof of Theorem 1.1 and of Proposition 1.7

We continue working in the setting of Section 1. In particular, we assume throughout this section (with the exception of the proof of Proposition 1.7) that (A_1), (A_2) and (A_3) hold.

We will adapt the arguments from [1] where a single functional has been considered. First we state a simple analogue of (1.1).

Lemma 2.1 (Discrete Evolution Variational Inequality). *Let $x \in \mathsf{D}(\varphi)$ and let $\mathbf{h} := (h_k)_{k=1}^n$ be a discretisation. For $w \in \mathsf{D}(\varphi)$ and $k = 1, \ldots, n$ we have*

$$\frac{1}{2h_k}\Big(d^2(x_\mathbf{h}^k, w) - d^2(x_\mathbf{h}^{k-1}, w)\Big) \le \varphi(w) - \varphi(x_\mathbf{h}^k) - \frac{1}{4h_k}d^2(x_\mathbf{h}^k, x_\mathbf{h}^{k-1}) + \delta_{\mathbf{h},x}^k. \tag{2.1}$$

In particular,

$$\frac{3}{4h_k}d^2(x_\mathbf{h}^k, x_\mathbf{h}^{k-1}) \le \varphi(x_\mathbf{h}^{k-1}) - \varphi(x_\mathbf{h}^k) + \delta_{\mathbf{h},x}^k. \tag{2.2}$$

Proof. Recall that $\widehat{x}_\mathbf{h}^k = J_{h_k}^1 x_\mathbf{h}^{k-1}$. Using (1.1) we find that

$$\frac{1}{2h_k}\Big(d^2(\widehat{x}_\mathbf{h}^k, w) - d^2(x_\mathbf{h}^{k-1}, w)\Big) \le \varphi^1(w) - \varphi^1(\widehat{x}_\mathbf{h}^k) - \frac{1}{2h_k}d^2(\widehat{x}_\mathbf{h}^k, x_\mathbf{h}^{k-1}),$$

$$\frac{1}{2h_k}\Big(d^2(x_\mathbf{h}^k, w) - d^2(\widehat{x}_\mathbf{h}^k, w)\Big) \le \varphi^2(w) - \varphi^2(x_\mathbf{h}^k) - \frac{1}{2h_k}d^2(x_\mathbf{h}^k, \widehat{x}_\mathbf{h}^k).$$

Adding these inequalities, we obtain

$$\frac{1}{2h_k}\Big(d^2(x_\mathbf{h}^k, w) - d^2(x_\mathbf{h}^{k-1}, w)\Big) \le \varphi(w) - \varphi^2(x_\mathbf{h}^k) - \varphi^1(x_\mathbf{h}^k)$$
$$+ \varphi^1(x_\mathbf{h}^k) - \varphi^1(\widehat{x}_\mathbf{h}^k) - \frac{1}{2h_k}\Big(d^2(\widehat{x}_\mathbf{h}^k, x_\mathbf{h}^{k-1}) + d^2(x_\mathbf{h}^k, \widehat{x}_\mathbf{h}^k)\Big).$$

Finally, observe that

$$d^2(x_\mathbf{h}^k, x_\mathbf{h}^{k-1}) \le 2d^2(x_\mathbf{h}^k, \widehat{x}_\mathbf{h}^k) + 2d^2(\widehat{x}_\mathbf{h}^k, x_\mathbf{h}^{k-1}).$$

$\square$

Our next goal is to prove some a priori estimates in Proposition 2.3. We will use the following discrete version of Gronwall's lemma, taken from [1, Lemma 3.2.4]. For the sake of completeness we include the proof.

Lemma 2.2. *Let $A \geq 0$, and let $\{a_n\}_{n \geq 1}, \{\tau_n\}_{n \geq 1}$ be sequences of positive numbers satisfying $m := \sup_{n \geq 1} \tau_n < 1$ and*

$$a_n \leq A + \sum_{k=1}^{n} \tau_k a_k, \qquad n \geq 1.$$

Then, writing $\beta := \frac{1}{1-m}$, $t_0 := 0$, and $t_n := \sum_{k=1}^{n} \tau_k$ for $n \geq 1$, we have

$$a_n \leq A\beta \exp(\beta t_{n-1}). \tag{2.3}$$

Proof. We argue by induction and observe that (2.3) clearly holds for $n = 1$.

Let $n \geq 1$ and suppose that (2.3) holds for all $1 \leq k \leq n$. Since, for any $n \geq 1$,

$$a_n \leq \frac{A}{1 - \tau_n} + \frac{1}{1 - \tau_n} \sum_{j=1}^{n-1} \tau_j a_j,$$

we obtain

$$a_{n+1} \leq A\beta + \beta \sum_{j=1}^{n} \tau_j a_j \leq A\beta + A\beta^2 \sum_{j=1}^{n} \tau_j e^{\beta t_{j-1}}$$

$$\leq A\beta + A\beta^2 \sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} e^{\beta t} \, dt = A\beta + A\beta^2 \int_0^{t_n} e^{\beta t} \, dt$$

$$= A\beta e^{\beta t_n},$$

which completes the induction step. $\qquad \square$

Proposition 2.3 (A priori estimates). *Let $w \in \overline{\mathsf{D}(\varphi)}$ and $\widetilde{h}, K, R, S, T > 0$ be given. There exist constants $C, \widetilde{C} \in (0, \infty)$ such that for every $x \in \mathsf{D}(\varphi)$ and any discretisation $\mathbf{h} := (h_k)_{k=1}^{n}$ satisfying*

$$|\mathbf{h}| \leq \frac{1}{8}\widetilde{h}, \qquad \Delta_{\mathbf{h},x}^{n} \leq K, \qquad d^2(x, w) \leq R, \qquad \varphi(x) \leq S, \qquad t_{\mathbf{h}}^{n} \leq T, \tag{2.4}$$

we have

$$d^2(x_{\mathbf{h}}^{n}, w) \leq C, \tag{2.5}$$

$$\frac{3}{4} \sum_{k=1}^{n} \frac{1}{h_k} d^2(x_{\mathbf{h}}^{k}, x_{\mathbf{h}}^{k-1}) \leq \varphi(x_{\mathbf{h}}^{0}) - \varphi(x_{\mathbf{h}}^{n}) + \Delta_{\mathbf{h},x}^{n} \leq \widetilde{C}. \tag{2.6}$$

Notice that, by applying the first inequality of (2.6) to the subdiscretisation $(h_j)_{j=1}^{k}$, $k = 1, \ldots, n$, we have

$$\varphi(x_{\mathbf{h}}^{k}) \leq \varphi(x_{\mathbf{h}}^{0}) + \Delta_{\mathbf{h},x}^{k} \leq \varphi(x_{\mathbf{h}}^{0}) + \Delta_{\mathbf{h},x}^{n} \tag{2.7}$$

for any $k = 1, \ldots, n$.

Proof. It follows from (2.2) that, for $k = 1, \ldots, n$,

$$\frac{3}{4h_k} d^2(x_{\mathbf{h}}^{k}, x_{\mathbf{h}}^{k-1}) \leq \varphi(x_{\mathbf{h}}^{k-1}) - \varphi(x_{\mathbf{h}}^{k}) + \delta_{\mathbf{h},x}^{k}.$$

Summation over k yields the first inequality in (2.6).

For $\varepsilon > 0$ we obtain, setting $h_0 = h_{n+1} = 0$,

$$d^2(x_{\mathbf{h}}^n, w) - d^2(x_{\mathbf{h}}^0, w)$$

$$= \sum_{k=1}^{n} d^2(x_{\mathbf{h}}^k, w) - d^2(x_{\mathbf{h}}^{k-1}, w)$$

$$= \sum_{k=1}^{n} \big(d(x_{\mathbf{h}}^k, w) - d(x_{\mathbf{h}}^{k-1}, w)\big)\big(d(x_{\mathbf{h}}^k, w) + d(x_{\mathbf{h}}^{k-1}, w)\big)$$

$$\leq \sum_{k=1}^{n} d(x_{\mathbf{h}}^k, x_{\mathbf{h}}^{k-1})\big(d(x_{\mathbf{h}}^k, w) + d(x_{\mathbf{h}}^{k-1}, w)\big)$$

$$\leq \sum_{k=1}^{n} \frac{\varepsilon}{h_k} d^2(x_{\mathbf{h}}^k, x_{\mathbf{h}}^{k-1}) + \sum_{k=1}^{n} \frac{h_k}{4\varepsilon} \big(d(x_{\mathbf{h}}^k, w) + d(x_{\mathbf{h}}^{k-1}, w)\big)^2$$

$$\leq \sum_{k=1}^{n} \frac{\varepsilon}{h_k} d^2(x_{\mathbf{h}}^k, x_{\mathbf{h}}^{k-1}) + \sum_{k=1}^{n} \frac{h_k}{2\varepsilon} \big(d^2(x_{\mathbf{h}}^k, w) + d^2(x_{\mathbf{h}}^{k-1}, w)\big)$$

$$= \sum_{k=1}^{n} \frac{\varepsilon}{h_k} d^2(x_{\mathbf{h}}^k, x_{\mathbf{h}}^{k-1}) + \sum_{k=0}^{n} \frac{h_k + h_{k+1}}{2\varepsilon} d^2(x_{\mathbf{h}}^k, w).$$

Combining this estimate with the first inequality in (2.6) (which we already proved), we arrive at

$$\frac{3}{4}(d^2(x_{\mathbf{h}}^n, w) - d^2(x_{\mathbf{h}}^0, w)) \leq \varepsilon(\varphi(x_{\mathbf{h}}^0) - \varphi(x_{\mathbf{h}}^n) + \Delta_{\mathbf{h},x}^n)$$
$$+ \frac{3}{4}\sum_{k=0}^{n} \frac{h_k + h_{k+1}}{2\varepsilon} d^2(x_{\mathbf{h}}^k, w). \tag{2.8}$$

For $i = 1, 2$, $h > 0$, and $z \in X$ we set

$$\Phi^i(h, z; y) := \begin{cases} \varphi^i(y) + \frac{1}{2h} d^2(z, y), & y \in \mathsf{D}(\varphi^i), \\ +\infty, & \text{otherwise,} \end{cases}$$

and

$$\widehat{\varphi}_h(w) := \Phi^1(h, w; J_h^1 w) + \Phi^2(h, w; J_h^2 w).$$

The defining property of $J_h^i w$ implies that

$$\Phi^i(h, w; J_h^i w) \leq \Phi^i(h, w; x_{\mathbf{h}}^n).$$

Adding these inequalities for $i = 1, 2$, it follows that

$$\widehat{\varphi}_h(w) \leq \varphi(x_{\mathbf{h}}^n) + \frac{1}{h} d^2(x_{\mathbf{h}}^n, w). \tag{2.9}$$

Substituting $\varepsilon := \frac{\widetilde{h}}{2}$ in (2.8), and using (2.4) and (2.9), we obtain

$$\frac{3}{4}\big(d^2(x_{\mathbf{h}}^n, w) - d^2(x_{\mathbf{h}}^0, w)\big) \leq \frac{\widetilde{h}}{2}\Big(S - \widehat{\varphi}_{\widetilde{h}}(w) + \frac{1}{\widetilde{h}} d^2(x_{\mathbf{h}}^n, w) + \Delta_{\mathbf{h},x}^n\Big)$$
$$+ \frac{3}{4}\sum_{k=0}^{n} \frac{h_k + h_{k+1}}{\widetilde{h}} d^2(x_{\mathbf{h}}^k, w).$$

Rearranging terms, using (2.4), and multiplying the inequality by 4, yields

$$d^2(x_{\mathbf{h}}^n, w) \leq A + \sum_{k=0}^{n} \tau_k d^2(x_{\mathbf{h}}^k, w),$$

where $A := [3R + 2\widetilde{h}(S - \widehat{\varphi}_{\widetilde{h}}(w) + K)]^+$ and $\tau_k := 3\frac{h_k + h_{k+1}}{\widetilde{h}} \leq \frac{3}{4}$. Applying Lemma 2.2 we obtain

$$d^2(x_{\mathbf{h}}^n, w) \leq 4A \exp\Big(12 \sum_{k=1}^{n-1} \frac{h_k + h_{k+1}}{\widetilde{h}} \Big) \leq 4A \exp\Big(\frac{24}{\widetilde{h}} \sum_{k=1}^{n} h_k \Big)$$

$$\leq 4A \exp\Big(\frac{12T}{\widetilde{h}} \Big) =: C,$$

which proves (2.5).

Finally, using (2.4), (2.5), and (2.9), we obtain

$$\varphi(x_{\mathbf{h}}^0) - \varphi(x_{\mathbf{h}}^n) + \Delta_{\mathbf{h},x}^n \leq S - \widehat{\varphi}_{\widetilde{h}}(w) + \frac{1}{\widetilde{h}}d^2(x_{\mathbf{h}}^n, w) + K$$

$$\leq S - \widehat{\varphi}_{\widetilde{h}}(w) + \frac{C}{\widetilde{h}} + K =: \widetilde{C}.$$

which proves the second inequality in (2.6). $\square$

Let $x \in \mathsf{D}(\varphi)$ and a discretisation $\mathbf{h} := (h_k)_{k=1}^n \subseteq (0, \infty)$ be given. As in [1] it will be useful to consider continuous interpolants of some relevant quantities which are originally defined only on the discrete set $(t_{\mathbf{h}}^j)_{j=0}^n$. For this purpose we will use the (unique) function $\ell_{\mathbf{h}} : [0, t_{\mathbf{h}}^n] \to [0, 1]$ which is affine on each interval $[t_{j-1}, t_j)$, $j = 1, \ldots, n$, and satisfies $\ell_{\mathbf{h}}(t_{\mathbf{h}}^j) = 0$ for $j = 0, \ldots, n$, and $\lim_{t \uparrow t_{\mathbf{h}}^j} \ell_{\mathbf{h}}(t) = 1$ for $j = 1, \ldots, n$. We consider the function $d_{\mathbf{h},x} : [0, t_{\mathbf{h}}^n] \times X \to \mathbb{R}$ defined by

$$d_{\mathbf{h},x}^2(t, y) := \big(1 - \ell_{\mathbf{h}}(t)\big)d^2(\underline{x}_{\mathbf{h}}(t), y) + \ell_{\mathbf{h}}(t)d^2(\overline{x}_{\mathbf{h}}(t), y)$$

and the function $\varphi_{\mathbf{h},x} : [0, t_{\mathbf{h}}^n] \to \mathbb{R}$ defined by

$$\varphi_{\mathbf{h},x}(t) := \big(1 - \ell_{\mathbf{h}}(t)\big)\varphi(\underline{x}_{\mathbf{h}}(t)) + \ell_{\mathbf{h}}(t)\varphi(\overline{x}_{\mathbf{h}}(t)).$$

Note that $\underline{x}_{\mathbf{h}}(t), \overline{x}_{\mathbf{h}}(t) \in \mathsf{D}(\varphi)$, since K_h maps $\mathsf{D}(\varphi)$ into itself, as has already been observed before. Finally, we define the function $R_{\mathbf{h},x} : [0, t_{\mathbf{h}}^n] \to \mathbb{R}$ by

$$R_{\mathbf{h},x}(t) := \sum_{k=1}^{n} \mathbf{1}_{[t_{\mathbf{h}}^{k-1}, t_{\mathbf{h}}^k)}(t)\bigg((1 - \ell_{\mathbf{h}}(t))\Big(\varphi(x_{\mathbf{h}}^{k-1}) - \varphi(x_{\mathbf{h}}^k) + \delta_{\mathbf{h},x}^k$$

$$- \frac{1}{4h_k}d^2(x_{\mathbf{h}}^k, x_{\mathbf{h}}^{k-1}) \Big) + \ell_{\mathbf{h}}(t)\Big(\delta_{\mathbf{h},x}^k - \frac{1}{4h_k}d^2(x_{\mathbf{h}}^k, x_{\mathbf{h}}^{k-1}) \Big) \bigg).$$

The following result is an analogue of [1, Theorem 4.1.4].

Lemma 2.4 (Gradient flow approximation). *Let $x \in \mathsf{D}(\varphi)$ and a discretisation $\mathbf{h} := (h_k)_{k=1}^n$ be given. For every $y \in \mathsf{D}(\varphi)$ and every $t \in [0, t_{\mathbf{h}}^n] \setminus \{t_{\mathbf{h}}^0, \ldots, t_{\mathbf{h}}^n\}$ we have*

$$\frac{d}{dt}d_{\mathbf{h},x}^2(t, y) + \varphi_{\mathbf{h},x}(t) - \varphi(y) \leq R_{\mathbf{h},x}(t), \qquad (2.10)$$

where $\frac{d}{dt}$ denotes the pointwise derivative.

Proof. First we remark that $t \mapsto d_{\mathbf{h},x}^2(t,y)$ is a piecewise affine function. As a consequence, the derivative in (2.10) exists for t in the above-mentioned set. For $k = 1, \ldots, n$ and $t \in (t_{\mathbf{h}}^{k-1}, t_{\mathbf{h}}^k)$ we obtain using (2.1),

$$
\begin{aligned}
\frac{d}{dt} d_{\mathbf{h},x}^2(t,y) &+ \varphi_{\mathbf{h},x}(t) - \varphi(y) \\
&= \frac{1}{2h_k}(d^2(x_{\mathbf{h}}^k, y) - d^2(x_{\mathbf{h}}^{k-1}, y)) + \varphi_{\mathbf{h},x}(t) - \varphi(y) \\
&\leq \delta_{\mathbf{h},x}^k - \frac{1}{4h_k} d^2(x_{\mathbf{h}}^k, x_{\mathbf{h}}^{k-1}) + \varphi_{\mathbf{h},x}(t) - \varphi(x_{\mathbf{h}}^k) \\
&= \delta_{\mathbf{h},x}^k - \frac{1}{4h_k} d^2(x_{\mathbf{h}}^k, x_{\mathbf{h}}^{k-1}) + (1 - \ell_{\mathbf{h}}(t))(\varphi(x_{\mathbf{h}}^{k-1}) - \varphi(x_{\mathbf{h}}^k)) \\
&= R_{\mathbf{h},x}(t).
\end{aligned}
$$

$\square$

The following estimate will be useful in the proof of Proposition 2.7 below.

Lemma 2.5. *Let $x \in \mathsf{D}(\varphi)$ and a discretisation $\mathbf{h} := (h_k)_{k=1}^n$ be given. For $1 \leq k \leq n$,*

$$
\int_0^{t_{\mathbf{h}}^k} [R_{\mathbf{h},x}(s)]^+ \, ds \leq |\mathbf{h}| \big(\varphi(x_{\mathbf{h}}^0) - \varphi(x_{\mathbf{h}}^k) + 2\Delta_{\mathbf{h},x}^k\big).
$$

Proof. For $i = 1, \ldots, k$ and $s \in [t_{\mathbf{h}}^{i-1}, t_{\mathbf{h}}^i)$ we claim that

$$
[R_{\mathbf{h},x}(s)]^+ \leq (1 - \ell_{\mathbf{h}}(s))\Big(\varphi(x_{\mathbf{h}}^{i-1}) - \varphi(x_{\mathbf{h}}^i) + \delta_{\mathbf{h},x}^i\Big) + \ell_{\mathbf{h}}(s)\delta_{\mathbf{h},x}^i. \qquad (2.11)
$$

Indeed, if $R_{\mathbf{h},x}(s)$ is non-negative, this is clear from the definition of $R_{\mathbf{h},x}(s)$. Otherwise, if $R_{\mathbf{h},x}(s)$ is negative, (2.11) follows from the non-negativity of $\delta_{\mathbf{h},x}^i$ and $\varphi(x_{\mathbf{h}}^{i-1}) - \varphi(x_{\mathbf{h}}^i) + \delta_{\mathbf{h},x}^i$, the latter being a consequence of (2.2).

Combining (2.11) with the identities

$$
\int_{t_{\mathbf{h}}^{i-1}}^{t_{\mathbf{h}}^i} \ell_{\mathbf{h}}(t) \, dt = \int_{t_{\mathbf{h}}^{i-1}}^{t_{\mathbf{h}}^i} 1 - \ell_{\mathbf{h}}(t) \, dt = h_i,
$$

we obtain

$$
\begin{aligned}
\int_0^{t_{\mathbf{h}}^k} [R_{\mathbf{h},x}(s)]^+ \, ds &= \sum_{i=1}^k \int_{t_{\mathbf{h}}^{i-1}}^{t_{\mathbf{h}}^i} [R_{\mathbf{h},x}(s)]^+ \, ds \\
&\leq \sum_{i=1}^k h_i\big(\varphi(x_{\mathbf{h}}^{i-1}) - \varphi(x_{\mathbf{h}}^i) + \delta_{\mathbf{h},x}^i\big) + h_i\delta_{\mathbf{h},x}^i \\
&\leq |\mathbf{h}| \sum_{i=1}^k \big(\varphi(x_{\mathbf{h}}^{i-1}) - \varphi(x_{\mathbf{h}}^i) + \delta_{\mathbf{h},x}^i\big) + |\mathbf{h}|\delta_{\mathbf{h},x}^i \\
&\leq |\mathbf{h}|\big(\varphi(x_{\mathbf{h}}^0) - \varphi(x_{\mathbf{h}}^k) + 2\Delta_{\mathbf{h},x}^k\big).
\end{aligned}
$$

$\square$

We will now compare the discrete scheme induced by $(\mathbf{h}, x)$ to another discrete scheme induced by $(\mathbf{r}, y)$, where $y \in \mathsf{D}(\varphi)$ and $\mathbf{r} = (r_j)_{j=1}^m$ is a

discretisation. For this purpose we consider the continuous function $d^2_{\mathbf{hr}}$: $[0, t^n_{\mathbf{h}}] \times [0, t^m_{\mathbf{r}}] \to \mathbb{R}$ defined by

$$d^2_{\mathbf{hr}}(t, s) := (1 - \ell_{\mathbf{r}}(s))d^2_{\mathbf{h},x}(t, \underline{y}_{\mathbf{r}}(s)) + \ell_{\mathbf{r}}(s)d^2_{\mathbf{h},x}(t, \overline{y}_{\mathbf{r}}(s)).$$

In this formula the dependence of $d^2_{\mathbf{hr}}(t, s)$ on x and y is suppressed in the notation. With this notation we have the following result:

Corollary 2.6. *For all $t \in [0, \min\{t^n_{\mathbf{h}}, t^m_{\mathbf{r}}\}]$ we have*

$$d^2_{\mathbf{hr}}(t, t) \leq d^2(x^0_{\mathbf{h}}, y^0_{\mathbf{r}}) + \int_0^t R_{\mathbf{h},x}(s) + R_{\mathbf{r},y}(s)\,ds.$$

Proof. For each fixed $s \in [0, t^m_{\mathbf{r}}]$ we obtain for all $t \in [0, t^n_{\mathbf{h}}] \setminus \{t^0_{\mathbf{h}}, \ldots, t^n_{\mathbf{h}}\}$ by (2.10),

$$\frac{\partial}{\partial t}d^2_{\mathbf{hr}}(t, s) + \varphi_{\mathbf{h},x}(t) - \varphi_{\mathbf{r},y}(s) \leq R_{\mathbf{h},x}(t).$$

Similarly, reversing the roles of $(\mathbf{h}, x)$ and $(\mathbf{r}, y)$, yields for fixed $t \in [0, t^m_{\mathbf{r}}]$ and for all $s \in [0, t^m_{\mathbf{r}}] \setminus \{t^0_{\mathbf{r}}, \ldots, t^m_{\mathbf{r}}\}$,

$$\frac{\partial}{\partial s}d^2_{\mathbf{rh}}(s, t) + \varphi_{\mathbf{r},y}(s) - \varphi_{\mathbf{h},x}(t) \leq R_{\mathbf{r},y}(s).$$

Noting that $d^2_{\mathbf{hr}}(t, s) = d^2_{\mathbf{rh}}(s, t)$, we obtain by adding these inequalities, for each $t \in [0, \min\{t^n_{\mathbf{h}}, t^m_{\mathbf{r}}\}] \setminus \{t^0_{\mathbf{h}}, \ldots, t^n_{\mathbf{h}}, t^0_{\mathbf{r}}, \ldots, t^m_{\mathbf{r}}\}$,

$$\frac{d}{dt}d^2_{\mathbf{hr}}(t, t) \leq R_{\mathbf{h},x}(t) + R_{\mathbf{r},y}(t).$$

Taking into account that $t \mapsto d^2_{\mathbf{hr}}(t, t)$ is continuous and piecewise C^1, the result follows by integrating this inequality. $\qquad\square$

The next result contains the main estimate for the proof of the Trotter product formula in Theorem 1.1 below.

Proposition 2.7. *Let $w \in \overline{\mathsf{D}(\varphi)}$ and $\widetilde{h}, R, S, U, T > 0$ be given. There exists a constant $\widetilde{K} \in (0, \infty)$ such that for all $x, y \in \mathsf{D}(\varphi)$ and all discretisations $\mathbf{h} := (h_k)^n_{k=1}$ and $\mathbf{r} := (r_j)^m_{j=1}$ satisfying*

$$|\mathbf{h}| \leq \frac{1}{8}\widetilde{h}, \qquad d^2(x^0_{\mathbf{h}}, w) \leq R, \qquad \varphi(x^0_{\mathbf{h}}) \leq S, \qquad \chi(x^0_{\mathbf{h}}) \leq U, \qquad t^n_{\mathbf{h}} \leq T;$$

$$|\mathbf{r}| \leq \frac{1}{8}\widetilde{h}, \qquad d^2(y^0_{\mathbf{r}}, w) \leq R, \qquad \varphi(y^0_{\mathbf{r}}) \leq S, \qquad \chi(y^0_{\mathbf{r}}) \leq U, \qquad t^m_{\mathbf{r}} \leq T,$$

we have, for $t \in [0, \min\{t^n_{\mathbf{h}}, t^m_{\mathbf{r}}\}]$,

$$d^2(\overline{x}_{\mathbf{h}}(t), \overline{y}_{\mathbf{r}}(t)) \leq \widetilde{K}\big(d^2(x, y) + |\mathbf{h}| + |\mathbf{r}|\big).$$

Proof. Let $t \in [0, \min\{t^n_{\mathbf{h}}, t^m_{\mathbf{r}}\}]$ and let $k, j \geq 1$ be such that $t \in [t^{k-1}_{\mathbf{h}}, t^k_{\mathbf{h}}) \cap [t^{j-1}_{\mathbf{r}}, t^j_{\mathbf{r}})$, where we use the convention that $t^{n+1}_{\mathbf{h}} := t^n_{\mathbf{h}} + 1$ and $t^{m+1}_{\mathbf{r}} := t^m_{\mathbf{r}} + 1$. To simplify notation we write

$$a_{\overline{k}\underline{k}} := d^2(x^k_{\mathbf{h}}, x^{k-1}_{\mathbf{h}}), \qquad a_{\overline{j}\underline{j}} := d^2(y^j_{\mathbf{r}}, y^{j-1}_{\mathbf{r}}),$$

$$a_{\underline{k}\underline{j}} := d^2(x^{k-1}_{\mathbf{h}}, y^{j-1}_{\mathbf{r}}), \qquad a_{\overline{k}\underline{j}} := d^2(x^k_{\mathbf{h}}, y^{j-1}_{\mathbf{r}}),$$

$$a_{\underline{k}\overline{j}} := d^2(x^{k-1}_{\mathbf{h}}, y^j_{\mathbf{r}}), \qquad a_{\overline{k}\overline{j}} := d^2(x^k_{\mathbf{h}}, y^j_{\mathbf{r}}).$$

With this notation we have (using that $(\sum_{i=1}^{n} b_i)^2 \leq n \sum_{i=1}^{n} b_i^2$),

$$
\begin{aligned}
d^2(\overline{x}_{\mathbf{h}}(t), \overline{y}_{\mathbf{r}}(t)) &= d^2(x_{\mathbf{h}}^k, y_{\mathbf{r}}^j) \\
&= \big((1 - \ell_{\mathbf{h}}(t))(1 - \ell_{\mathbf{r}}(t)) + (1 - \ell_{\mathbf{h}}(t))\ell_{\mathbf{r}}(t) \\
&\quad + \ell_{\mathbf{h}}(t)(1 - \ell_{\mathbf{r}}(t)) + \ell_{\mathbf{h}}(t)\ell_{\mathbf{r}}(t)\big)a_{\overline{kj}} \\
&\leq 3(1 - \ell_{\mathbf{h}}(t))(1 - \ell_{\mathbf{r}}(t))(a_{\overline{k}\underline{k}} + a_{\underline{kj}} + a_{\overline{j}\underline{j}}) \\
&\quad + 2(1 - \ell_{\mathbf{h}}(t))\ell_{\mathbf{r}}(t)(a_{\overline{k}\underline{k}} + a_{k\overline{j}}) \\
&\quad + 2\ell_{\mathbf{h}}(t)(1 - \ell_{\mathbf{r}}(t))(a_{\overline{k}\underline{j}} + a_{\overline{j}\underline{j}}) + \ell_{\mathbf{h}}(t)\ell_{\mathbf{r}}(t)a_{\overline{kj}} \\
&\leq 3(1 - \ell_{\mathbf{h}}(t))(1 - \ell_{\mathbf{r}}(t))(a_{\overline{k}\underline{k}} + a_{\underline{kj}} + a_{\overline{j}\underline{j}}) \\
&\quad + 3(1 - \ell_{\mathbf{h}}(t))\ell_{\mathbf{r}}(t)(a_{\overline{k}\underline{k}} + a_{k\overline{j}}) \\
&\quad + 3\ell_{\mathbf{h}}(t)(1 - \ell_{\mathbf{r}}(t))(a_{\overline{k}\underline{j}} + a_{\overline{j}\underline{j}}) + 3\ell_{\mathbf{h}}(t)\ell_{\mathbf{r}}(t)a_{\overline{kj}} \\
&= 3d_{\mathbf{hr}}^2(t,t) + 3(1 - \ell_{\mathbf{h}}(t))d^2(x_{\mathbf{h}}^k, x_{\mathbf{h}}^{k-1}) + 3(1 - \ell_{\mathbf{r}}(t))d^2(y_{\mathbf{r}}^j, y_{\mathbf{r}}^{j-1}).
\end{aligned}
$$

To complete the proof, we will estimate each of the terms at the right-hand side. Using Corollary 2.6, Lemma 2.5, Proposition 2.3 and (A_3) with $h_* = \frac{1}{8}\widetilde{h}$,

$$
\begin{aligned}
d_{\mathbf{hr}}^2(t,t) \\
&\leq d^2(x_{\mathbf{h}}^0, y_{\mathbf{r}}^0) + \int_0^t R_{\mathbf{h},x}(s) + R_{\mathbf{r},y}(s)\, ds \\
&\leq d^2(x_{\mathbf{h}}^0, y_{\mathbf{r}}^0) + \int_0^{t_{\mathbf{h}}^k}[R_{\mathbf{h},x}(s)]^+\, ds + \int_0^{t_{\mathbf{r}}^j}[R_{\mathbf{r},y}(s)]^+\, ds \\
&\leq d^2(x_{\mathbf{h}}^0, y_{\mathbf{r}}^0) + |\mathbf{h}|\big(\varphi(x_{\mathbf{h}}^0) - \varphi(x_{\mathbf{h}}^k) + 2\Delta_{\mathbf{h},x}^k\big) + |\mathbf{r}|\big(\varphi(y_{\mathbf{r}}^0) - \varphi(y_{\mathbf{r}}^j) + 2\Delta_{\mathbf{r},y}^j\big) \\
&\leq d^2(x_{\mathbf{h}}^0, y_{\mathbf{r}}^0) + (\widetilde{C} + K)(|\mathbf{h}| + |\mathbf{r}|),
\end{aligned}
$$

where $\tilde{C}$ and K are the constants from (2.6) and (A_3) respectively. By another application of (2.6),

$$
d^2(x_{\mathbf{h}}^k, x_{\mathbf{h}}^{k-1}) \leq |\mathbf{h}|\sum_{j=1}^{n}\frac{1}{h_j}d^2(x_{\mathbf{h}}^j, x_{\mathbf{h}}^{j-1}) \leq \frac{4}{3}\widetilde{C}|\mathbf{h}|,
$$

and similarly,

$$
d^2(y_{\mathbf{r}}^j, y_{\mathbf{r}}^{j-1}) \leq \frac{4}{3}\widetilde{C}|\mathbf{r}|,
$$

which completes the proof. $\qquad\square$

The following elementary lemma will be used in the proof of Theorem 1.1 below.

Lemma 2.8. *Let $\{a_i\}_{i\geq 1}$ and $\{b_i\}_{i\geq 1}$ be sequences in X converging to the same limit $c \in X$. Let $\{\lambda_i\}_{i\geq 1}$ be a sequence in $[0,1]$, and let $\psi : X \to \mathbb{R} \cup \{+\infty\}$ be a lsc functional which is not identically $+\infty$. Then*

$$
\psi(c) \leq \liminf_{i\to\infty}\big((1 - \lambda_i)\psi(a_i) + \lambda_i\psi(b_i)\big).
$$

Proof. Suppose that $\psi(c) \in \mathbb{R}$ (resp. $\psi(c) = +\infty$.) Let $\varepsilon > 0$ (resp. let $M > 0$). Since ψ is lsc, we can find $\delta > 0$ such that $\psi(x) \geq \psi(c) - \varepsilon$ (resp. $\psi(x) \geq M$) whenever $d(c, x) < \delta$. Since $a_i, b_i \to x$ as $i \to \infty$, we can take $N \geq 1$ such that $d(a_i, x) < \delta$ and $d(b_i, x) < \delta$ for all $i \geq N$. Consequently, for all $i \geq N$ we have $\psi(a_i) \geq \psi(c) - \varepsilon$ (resp. $\psi(a_i) \geq M$) and $\psi(b_i) \geq \psi(c) - \varepsilon$ (resp. $\psi(b_i) \geq M$). It follows that $(1 - \lambda_i)\psi(a_i) + \lambda_i\psi(b_i) \geq \psi(c) - \varepsilon$ (resp. $(1 - \lambda_i)\psi(a_i) + \lambda_i\psi(b_i) \geq M$), which implies the result. $\qquad\square$

Proof of Theorem 1.1. Note that the sequences $\{\underline{x^i}_{\mathbf{h}^i}(t)\}_{i\geq 1}$ and $\{\overline{x^i}_{\mathbf{h}^i}(t)\}_{i\geq 1}$ are well-defined on $[0, T_*]$ as a consequence of (i). Proposition 2.7, (ii) and (iii) imply that the sequence $\{\overline{x^i}_{\mathbf{h}^i}(t)\}_{i\geq 1}$ is a Cauchy sequence, even uniformly in $t \in [0, T_*]$. Using the completeness of (X, d) there exists a limit $u(t)$, $t \in [0, T_*]$, which is right-continuous. From (2.6) we have $d^2(x_{\mathbf{h}}^k, x_{\mathbf{h}}^{k-1}) \leq \frac{4}{3}\widetilde{C}h_k$. This implies that the sequence $\{\underline{x^i}_{\mathbf{h}^i}(t)\}_{i\geq 1}$ converges to the same limit $u(t)$, $t \geq 0$, which is also left-continuous on $[0, T_*]$. Assumption (iii) implies that $u(0) = x$.

Next we show that $\varphi \circ u \in L^1((0, T_*); \mathbb{R})$. Since φ is lsc and u is continuous, the function $\varphi \circ u : [0, T_*] \to \mathbb{R} \cup \{+\infty\}$ is lsc, hence Borel measurable and bounded from below. From (2.7) we obtain for $t \in [0, T_*]$,

$$\varphi(\overline{x^i}_{\mathbf{h}^i}(t)) \leq \varphi(x^i) + \Delta_{\mathbf{h}^i, x^i}^{n^i}.$$

Assumption (v) implies $K_1 := \sup_{i\geq 1} \varphi(x^i) < \infty$. Using assumption (A_3) with $h_* := \sup_{i\geq 1}|\mathbf{h}^i|$, $T := T_* + h_*$, $w := x$, $R := \sup_{i\geq 1} d^2(x^i, x)$, $U := \sup_{i\geq 1}\chi(x^i)$, which is finite by (iv), we have $K_2 := \sup_{i\geq 1}\Delta_{\mathbf{h}^i, x^i}^{n^i} < \infty$. Therefore

$$\varphi(u(t)) \leq \liminf_{i\to\infty} \varphi(\overline{x^i}_{\mathbf{h}^i}(t)) \leq \sup_{i\geq 1} \varphi(x^i) + \sup_{i\geq 1} \Delta_{\mathbf{h}^i, x^i}^{n^i} \leq K_1 + K_2.$$

It follows that $\varphi \circ u$ is bounded from above, which implies the claim.

It remains to show that $(u(t))_{t>0}$ is a solution to (1.4). This will be done by passing to the limit in (2.10). Let $\zeta \in C_c^\infty((0, T_*); \mathbb{R})$ be non-negative. Take $y \in \mathsf{D}(\varphi)$ and note that, by what we just proved,

$$\lim_{i\to\infty} d_{\mathbf{h}^i, x^i}^2(t, y) = d^2(u(t), y), \quad \text{uniformly on } [0, T_*].$$

Consequently, the mapping $t \mapsto \zeta'(t)d^2(u(t), y)$ is continuous (hence integrable) on $[0, T_*]$ and

$$\lim_{i\to\infty} \int_0^{T_*} \zeta'(t)d_{\mathbf{h}^i, x^i}^2(t, y)\, dt = \int_0^{T_*} \zeta'(t)d^2(u(t), y)\, dt. \qquad (2.12)$$

Using the second inequality of (2.6) and the lower semicontinuity of φ, we infer that there exist constants $C, \widetilde{C} \in \mathbb{R}$ not depending on $t \in [0, T_*]$ and $i \geq 1$ such that

$$\varphi_{\mathbf{h}^i, x^i}(t) \geq \varphi(x^i) - \widetilde{C} \geq C.$$

Since ζ is non-negative, it thus follows that the functions $t \mapsto \zeta(t)\varphi_{\mathbf{h}^i, x^i}(t)$ are bounded from below, uniformly in i. Therefore we may apply Fatou's

Lemma to obtain

$$\liminf_{i\to\infty} \int_0^{T_*} \zeta(t)\varphi_{\mathbf{h}^i,x^i}(t)\,dt \geq \int_0^{T_*} \liminf_{i\to\infty} \zeta(t)\varphi_{\mathbf{h}^i,x^i}(t)\,dt.$$

Applying Lemma 2.8 with $a_i := \underline{x^i}_{\mathbf{h}^i}(t)$, $b_i := \overline{x^i}_{\mathbf{h}^i}(t)$, $\lambda_i := \ell_{\mathbf{h}^i}(t)$, and $\psi := \varphi$, we infer that

$$\liminf_{i\to\infty} \int_0^{T_*} \zeta(t)\varphi_{\mathbf{h}^i,x^i}(t)\,dt \geq \int_0^{T_*} \zeta(t)\varphi(u(t))\,dt. \qquad (2.13)$$

Combining (2.12) and (2.13), integrating by parts, using Lemma 2.4, Lemma 2.5, (2.6) and (A_3), we arrive at

$$\int_0^{T_*} -\zeta'(t)d^2(u(t),y) + \zeta(t)\varphi(u(t))\,dt$$

$$\leq \liminf_{i\to\infty} \int_0^{T_*} -\zeta'(t)d^2_{\mathbf{h}^i,x^i}(t,y) + \zeta(t)\varphi_{\mathbf{h}^i,x^i}(t)\,dt$$

$$= \liminf_{i\to\infty} \int_0^{T_*} \zeta(t)\left(\frac{d}{dt}d^2_{\mathbf{h}^i,x^i}(t,y) + \varphi_{\mathbf{h}^i,x^i}(t)\right)dt$$

$$= \liminf_{i\to\infty} \int_0^{T_*} \zeta(t)\big(\varphi(y) + R_{\mathbf{h}^i,x^i}(t)\big)\,dt$$

$$\leq \liminf_{i\to\infty} \int_0^{T_*} \zeta(t)\varphi(y)\,dt + (\widetilde{C}+K)\|\zeta\|_\infty|\mathbf{h}^i|$$

$$= \int_0^{T_*} \zeta(t)\varphi(y)\,dt,$$

which shows in view of Remark 1.2 that $(u(t))_{t>0}$ satisfies (1.4). $\qquad \square$

Proof of Proposition 1.7. Let $w \in \overline{\mathsf{D}(\varphi)}$ and $h_*, T, R, U > 0$ be given, and take $x \in \mathsf{D}(\varphi)$ and a discretisation $\mathbf{h} \subseteq (0,\infty)$ satisfying (1.2) and (1.3).

(1) Since $c, h_k \geq 0$ for $k = 1,\ldots,n$, we have

$$\Delta^n_{\mathbf{h},x} = \sum_{k=1}^n [\varphi^1(x_{\mathbf{h}}^k) - \varphi^1(\widehat{x}_{\mathbf{h}}^k)]^+ \leq c\sum_{k=1}^n h_k \leq \frac{1}{2}cT,$$

which implies (A_3) with $\chi = constant$.

(2) By assumption we have $\varphi^1(x_{\mathbf{h}}^k) \leq e^{\alpha h_k}\varphi^1(\widehat{x}_{\mathbf{h}}^k)$ for $k = 1,\ldots,n$, and therefore

$$\varphi^1(x_{\mathbf{h}}^k) - \varphi^1(\widehat{x}_{\mathbf{h}}^k) \leq (e^{\alpha h_k} - 1)\varphi^1(\widehat{x}_{\mathbf{h}}^k).$$

Since the right-hand side is non-negative, we have

$$[\varphi^1(x_{\mathbf{h}}^k) - \varphi^1(\widehat{x}_{\mathbf{h}}^k)]^+ \leq (e^{\alpha h_k} - 1)\varphi^1(\widehat{x}_{\mathbf{h}}^k).$$

Moreover, it follows by induction that $\varphi^1(\widehat{x}_{\mathbf{h}}^k) \leq e^{\frac{1}{2}\alpha t_{\mathbf{h}}^{k-1}}\varphi^1(x)$ for $k = 1, \ldots, n$. Consequently,

$$
\begin{aligned}
\Delta_{\mathbf{h},x}^n &= \sum_{k=1}^n [\varphi^1(x_{\mathbf{h}}^k) - \varphi^1(\widehat{x}_{\mathbf{h}}^k)]^+ \leq \sum_{k=1}^n (e^{\alpha h_k} - 1)\varphi^1(\widehat{x}_{\mathbf{h}}^k) \\
&\leq \sum_{k=1}^n (e^{\alpha h_k} - 1)e^{\frac{1}{2}\alpha t_{\mathbf{h}}^{k-1}}\varphi^1(x) = \sum_{k=1}^n (e^{\frac{1}{2}\alpha t_{\mathbf{h}}^k} - e^{\frac{1}{2}\alpha t_{\mathbf{h}}^{k-1}})\varphi^1(x) \\
&\leq (e^{\frac{1}{2}\alpha T} - 1)\varphi^1(x),
\end{aligned}
$$

which proves (A_3) with $\chi = \varphi^1$.

(3) Note that (i) implies that $\delta_{\mathbf{h},x}^k = [\varphi^1(x_{\mathbf{h}}^k) - \varphi^1(\widehat{x}_{\mathbf{h}}^k)]^+ \leq ch_k\varphi^2(x_{\mathbf{h}}^k)$, since $ch_k\varphi^2(x_{\mathbf{h}}^k)$ is non-negative. Moreover, using (ii) and induction it follows that $\varphi^2(x_{\mathbf{h}}^k) \leq e^{\frac{1}{2}\alpha t_{\mathbf{h}}^k}\varphi^2(x)$. Using these estimates we obtain

$$
\begin{aligned}
\Delta_{\mathbf{h},x}^n &\leq c\sum_{k=1}^n h_k\varphi^2(x_{\mathbf{h}}^k) \leq c\varphi^2(x)\sum_{k=1}^n h_k e^{\frac{1}{2}\alpha t_{\mathbf{h}}^k} \\
&\leq c\varphi^2(x)\int_{h_1}^{\frac{1}{2}t_{\mathbf{h}}^n + h_n} e^{\alpha s}\, ds \leq c\varphi^2(x)\int_0^T e^{\alpha s}\, ds,
\end{aligned}
$$

which proves (A_3) with $\chi = \varphi^2$. □

3. Applications

In this section we will apply the Trotter product formula from Theorem 1.1 in several concrete situations.

The Wasserstein space. Let $\mathscr{P}_2(\mathbb{R}^d)$ denote the set of all Borel probability measures μ on $\mathbb{R}^d$, $d \geq 1$, satisfying $\int_{\mathbb{R}^d} |x|^2\, d\mu(x) < \infty$. We consider the L^2-Wasserstein distance defined for $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ by

$$
W_2(\mu,\nu) := \inf\left\{ \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x_1 - x_2|^2\, d\Sigma(x_1, x_2) \right)^{1/2} : \Sigma \in \Gamma(\mu,\nu) \right\}. \quad (3.1)
$$

Here $\Gamma(\mu,\nu)$ denotes the collection of probability measures Σ on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν, i.e., for all Borel sets $A, B \subseteq \mathbb{R}^d$,

$$
\Sigma(A \times \mathbb{R}^d) = \mu(A), \qquad \Sigma(\mathbb{R}^d \times B) = \nu(B).
$$

Endowed with the metric W_2, $\mathscr{P}_2(\mathbb{R}^d)$ is a complete separable metric space. For a Borel mapping $T : \mathbb{R}^d \to \mathbb{R}^n$ and a Borel probability measure μ on $\mathbb{R}^d$ we write $T_\#\mu$ to denote the image measure on $\mathbb{R}^n$ defined by $T_\#\mu(B) := \mu(T^{-1}(B))$ for a Borel set $B \subseteq \mathbb{R}^n$.

The infimum in (3.1) is attained (see, e.g., [11, Theorem 4.1]). Moreover, a celebrated result by Brenier, independently due to Rachev and Rüschendorf and later refined by McCann, asserts that if μ is absolutely continuous with respect to the Lebesgue measure $\mathscr{L}^d$, the minimizer $\Sigma \in \Gamma(\mu,\nu)$ is unique and can be written as $\Sigma = (I \times \nabla f)_\#\mu$ for some convex function $f : \mathbb{R}^d \to \mathbb{R}$. We refer to ∇f as the optimal map pushing μ to ν. Detailed proofs of these results can be found in [11, Theorems 9.4 and 10.41].

We shall present four applications to gradient flows in the metric space $(\mathscr{P}_2(\mathbb{R}^d), W_2)$.

The functionals. In all of our examples below we shall consider a convex potential $V \in C^2(\mathbb{R}^d; \mathbb{R})$ satisfying the (strong) assumption that the Hessian $D^2 V$ is bounded, or equivalently since V is convex, that there exists $c \geq 0$ such that

$$\Delta V(x) \leq c, \qquad x \in \mathbb{R}^d. \tag{3.2}$$

The *potential energy* $\mathcal{V} : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ given by

$$\mathcal{V}(\mu) := \int_{\mathbb{R}^d} V(x) \, d\mu(x)$$

is well-defined, since the assumption on the Hessian implies that $|V|$ is of at most quadratic growth.

We shall also consider the (negative of the) *Boltzmann entropy* $\mathcal{H} : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ given by

$$\mathcal{H}(\mu) := \begin{cases} \int_{\mathbb{R}^d} \rho(x) \log \rho(x) \, dx, & \mu = \rho \mathscr{L}^d, \\ +\infty, & \text{otherwise}, \end{cases}$$

and the *Rényi entropy* $\mathcal{F} : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ defined for $m > 1$, by

$$\mathcal{F}(\mu) := \begin{cases} \frac{1}{m-1} \int_{\mathbb{R}^d} \rho^m(x) \, dx, & \mu = \rho \mathscr{L}^d, \\ +\infty, & \text{otherwise}. \end{cases} \tag{3.3}$$

The functionals $\mathcal{V}$, $\mathcal{H}$, and $\mathcal{F}$ are lower semicontinuous on $\mathscr{P}_2(\mathbb{R}^d)$ (see, e.g., [1, Lemma 5.1.7 and Remark 9.3.8]). A famous result by McCann [8] asserts that these functionals are displacement convex, that is, convex along geodesics in $\mathscr{P}_2(\mathbb{R}^d)$. In the first part of [1], an abstract theory of gradient flows in metric spaces has been developed for functionals which are convex along interpolating curves, not necessarily geodesics, along which the squared distance function satisfies an appropriate convexity condition. This condition fails for W_2-geodesics, but it holds for a different class of interpolating curves along which the functionals $\mathcal{V}$, $\mathcal{H}$, and $\mathcal{F}$ are convex as well [1, Propositions 9.3.2 and 9.3.9]. For our purpose, it is important to note that as a consequence (see [1, Theorem 4.1.2]), (A_1) is satisfied for any pair of functionals chosen from $\mathcal{V}$, $\mathcal{H}$, and $\mathcal{F}$.

Moreover, the first inclusion of (A_2) is satisfied in this situation, since

$$\mathsf{D}(\mathcal{V}) = \overline{\mathsf{D}(\mathcal{H})} = \overline{\mathsf{D}(\mathcal{F})} = \mathscr{P}_2(\mathbb{R}^d).$$

In order to prove that the second inclusion of (A_2) and (A_3) hold in the examples below, we shall use the following known result which provides formulas for the densities of the resolvents. We let J_h^φ denote the resolvent associated with a functional φ defined on $\mathbb{R}^d$ or $\mathscr{P}_2(\mathbb{R}^d)$.

Lemma 3.1. *Let $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ and $h > 0$.*

(1) We have $J_h^{\mathcal{V}} \mu = (J_h^V)_\# \mu$. If $\mu = \rho \mathscr{L}^d$, then $J_h^{\mathcal{V}} \mu = \rho^1 \mathscr{L}^d$, where

$$\rho^1(x) = \rho(x + h\nabla V(x)) \det(I + hD^2 V(x)), \qquad x \in \mathbb{R}^d, \tag{3.4}$$

(2) *We have $J_h^{\mathcal{H}}\mu = \rho^1 \mathscr{L}^d$ for some $\rho^1 \in W^{1,1}(\mathbb{R}^d)$. Let T be the optimal map pushing $J_h^{\mathcal{H}}\mu$ to μ. Then we have*

$$h\nabla\rho^1(x) = (T(x) - x)\rho^1(x), \quad \mathscr{L}^d\text{-a.e.} \tag{3.5}$$

(3) *We have $J_h^{\mathcal{F}}\mu = \rho^1 \mathscr{L}^d$ for some $\rho^1 \in W^{1,1}(\mathbb{R}^d)$. Let T be the optimal map pushing $J_h^{\mathcal{F}}\mu$ to μ. Then we have*

$$h\nabla(\rho^1)^m(x) = (T(x) - x)\rho^1(x), \quad \mathscr{L}^d\text{-a.e.} \tag{3.6}$$

Proof. The result follows from [1, Lemma 10.1.2. and Theorem 10.4.6]; see also [10, Proposition 3]. For the convenience of the reader we provide a simple direct proof of (1).

For $x \in \mathbb{R}^d$ we use the fact that $J_h^V(x) = (I + h\nabla V)^{-1}(x)$ to obtain

$$V(J_h^V(x)) + \frac{1}{2h}|x - J_h^V(x)|^2 \leq V(y) + \frac{1}{2h}|x - y|^2,$$

for any $y \in \mathbb{R}^d$. Take $\nu \in \mathscr{P}_2(\mathbb{R}^d)$ and let T be the optimal map pushing μ to ν. Using the estimate above we obtain

$$\mathcal{V}((J_h^V)_\#\mu) + \frac{1}{2h}W_2^2(\mu, (J_h^V)_\#\mu) \leq \int_{\mathbb{R}^d} V(J_h^V(x)) + \frac{1}{2h}|x - J_h^V(x)|^2 \, d\mu(x)$$

$$\leq \int_{\mathbb{R}^d} V(T(x)) + \frac{1}{2h}|x - T(x)|^2 \, d\mu(x)$$

$$= \mathcal{V}(\nu) + \frac{1}{2h}W_2^2(\mu, \nu),$$

which implies that $J_h^{\mathcal{V}}\mu = (J_h^V)_\#\mu$. This proves the first assertion.

To complete the proof, we note that, $(I + h\nabla V)_\#(J_h^{\mathcal{V}}\mu) = \mu$. The convexity of V implies that $\det(I + hD^2V(x)) > 0$ for all $x \in \mathbb{R}^d$, and therefore (3.4) follows from the change of variable formula. $\qquad\square$

Compatibility of the functionals. It has been shown in [6] that the L^2-Wasserstein gradient flow associated with the sum $\mathcal{H} + \mathcal{V}$ solves the Fokker-Planck equation

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

in an appropriate sense. Similarly, in [9] it has been shown that the gradient flow associated with $\mathcal{F}$ is a solution to the porous medium equation

$$\partial_t \rho = \Delta\rho^m.$$

The following result shows that the assumptions of Proposition 1.7 are satisfied in several examples.

Proposition 3.2. *Let $h > 0$ and let $c \geq 0$ be as in (3.2).*

(1) *For $\mu \in \mathsf{D}(\mathcal{H})$ we have*

$$\mathcal{H}(J_h^{\mathcal{V}}\mu) \leq \mathcal{H}(\mu) + ch. \tag{3.7}$$

(2) *For $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ we have*

$$\mathcal{V}(J_h^{\mathcal{H}}\mu) \leq \mathcal{V}(\mu) + ch. \tag{3.8}$$

(3) *For $\mu \in \mathsf{D}(\mathcal{F})$ we have*

$$\mathcal{F}(J_h^{\mathcal{V}}\mu) \leq e^{(m-1)ch}\mathcal{F}(\mu). \tag{3.9}$$

(4) For $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ we have

$$\mathcal{V}(J_h^{\mathcal{F}}\mu) \leq \mathcal{V}(\mu) + c(m-1)h\mathcal{F}(J_h^{\mathcal{F}}\mu). \tag{3.10}$$

Proof. (1) We write $\mu^1 := J_h^{\mathcal{V}}\mu$ and denote the densities of μ and μ^1 by ρ and ρ^1 respectively. Using the convexity of V, the inequality $\log \det(I + A) \leq \operatorname{tr}(A)$ which holds for any non-negative symmetric matrix A, and the assumption on ΔV, we obtain

$$0 \leq \log \det(I + hD^2V(x)) \leq h\Delta V(x) \leq ch, \qquad x \in \mathbb{R}^d. \tag{3.11}$$

Lemma 3.1 implies that

$$\mathcal{H}(\mu) + \int_{\mathbb{R}^d} \log \det(I + hD^2V(x))\,d\mu^1(x)$$

$$= \int_{\mathbb{R}^d} \log \rho(x)\,d\mu(x) + \int_{\mathbb{R}^d} \log \det(I + hD^2V(x))\,d\mu^1(x)$$

$$= \int_{\mathbb{R}^d} \log \rho(x + h\nabla V(x))\,d\mu^1(x) + \int_{\mathbb{R}^d} \log \det(I + hD^2V(x))\,d\mu^1(x)$$

$$= \int_{\mathbb{R}^d} \log \rho^1(x)\,d\mu^1(x)$$

$$= \mathcal{H}(\mu^1).$$

The conclusion follows by combining this identity with (3.11).

(2) Let V^+ and V^- denote the positive and negative part of V respectively. Since V is convex, there exist $k_1, k_2 \geq 0$ such that $|V^-(x)| \leq k_1(1 + k_2|x|)$ for all $x \in \mathbb{R}^d$. Since $\mu^1 := J_h^{\mathcal{H}}\mu \in \mathscr{P}_2(\mathbb{R}^d)$, this implies that $V^- \in L^1(\mathbb{R}^d; \mu^1)$.

To show that $V^+ \in L^1(\mathbb{R}^d; \mu^1)$ as well, let T denote the optimal map pushing $\mu^1 := J_h^{\mathcal{H}}\mu$ to μ, and note that by the convexity of V,

$$V^+(x) = V(x) + V^-(x) \leq V(T(x)) + \langle \nabla V(x), x - T(x)\rangle + V^-(x).$$

We claim that the first two summands at the right-hand side are contained in $L^1(\mathbb{R}^d, \mu^1)$.

Indeed, since $\mu \in \mathsf{D}(\mathcal{V})$ and $\mu = T_\#\mu^1$ we have $V \circ T \in L^1(\mu^1)$ and

$$\int_{\mathbb{R}^d} V(T(x))\,d\mu^1(x) = \int_{\mathbb{R}^d} V(x)\,d\mu(x).$$

Furthermore, we note that the convexity of V and the upper bound on its Laplacian imply that the Hessian of V is bounded. As a consequence, ∇V is of at most linear growth, which in view of the fact that $\mu^1 \in \mathscr{P}_2(\mathbb{R}^d)$ implies that $\nabla V \in L^2(\mathbb{R}^d, \mu^1; \mathbb{R}^d)$. Since $\|I - T\|_{L^2(\mathbb{R}^d, \mu; \mathbb{R}^d)} = W_2(\mu, \mu^1) < \infty$, it follows that $\langle \nabla V, I - T\rangle \in L^1(\mathbb{R}^d, \mu^1)$ by the Cauchy-Schwarz inequality. This proves the claim and we conclude that $V^+ \in L^1(\mathbb{R}^d; \mu^1)$.

Lemma 3.1(2) implies that $\mu^1 = \rho^1 \mathscr{L}^d$ for some $\rho^1 \in W^{1,1}(\mathbb{R}^d)$. Using the convexity of V and (3.5),

$$\mathcal{V}(\mu^1) = \int_{\mathbb{R}^d} V(x)\,d\mu^1(x)$$

$$\leq \int_{\mathbb{R}^d} V(T(x)) - \langle \nabla V(x), T(x) - x\rangle\,d\mu^1(x) \tag{3.12}$$

$$= \mathcal{V}(\mu) - h\int_{\mathbb{R}^d} \langle \nabla V(x), \nabla \rho^1(x)\rangle\,dx.$$

We shall show that

$$\int_{\mathbb{R}^d} \langle \nabla V(x), \nabla \rho^1(x) \rangle \, dx = - \int_{\mathbb{R}^d} \Delta V(x) \rho^1(x) \, dx. \tag{3.13}$$

Then (3.8) follows by combining (3.12), (3.13), and the condition $\Delta V \leq c$.

To prove (3.13), we set $n(y) = \frac{y}{|y|}$ for $y \neq 0$, and apply the Gauss-Green theorem in the ball $B_R := \{x \in \mathbb{R}^d : |x| < R\}$ to write for $R > 0$,

$$\int_{\partial B_R} \rho^1(y) \langle \nabla V(y), n(y) \rangle \, dS(y) = \int_{B_R} \langle \nabla V(x), \nabla \rho^1(x) \rangle \, dx \\ + \int_{B_R} \Delta V(x) \rho^1(x) \, dx. \tag{3.14}$$

Now we observe that

$$\langle \nabla V, \nabla \rho^1 \rangle = \frac{1}{h} \langle \nabla V, T - I \rangle \rho^1 \in L^1(\mathbb{R}^d),$$

and $\rho^1 \Delta V \in L^1(\mathbb{R}^d)$ as a consequence of the assumption that $0 \leq \Delta V \leq c$. Therefore the dominated convergence theorem implies that the right hand side of (3.14) converges as $R \to \infty$. In particular it follows that

$$L := \lim_{R \to \infty} \int_{\partial B_R} \rho^1(y) \langle \nabla V(y), n(y) \rangle \, dS(y) \tag{3.15}$$

exists. On the other hand, since we already showed that $\nabla V \in L^2(\mathbb{R}, \mu^1; \mathbb{R}^d)$, it follows that $\rho^1 \langle \nabla V, n \rangle \in L^1(\mathbb{R}^d)$, hence the coarea formula (see, e.g., [5, Proposition 1, p.118]) implies that for a.e. $R > 0$,

$$\int_{\partial B_R} \rho^1(y) \langle \nabla V(y), n(y) \rangle \, dS(y) = \frac{d}{dR} \left(\int_{B_R} \rho^1(x) \langle \nabla V(x), n(x) \rangle \, dx \right). \tag{3.16}$$

Combining (3.15) and (3.16) we conclude (as in e.g. [10, Proof of Theorem 1]) that $L = 0$, and therefore (3.13) follows from (3.14).

(3) Write $\mu^1 := J_h^{\mathcal{V}} \mu$ and $\mu^1 = \rho^1 \mathscr{L}^d$. It follows from Lemma 3.1 that

$$\mathcal{F}(\mu^1) = \frac{1}{m-1} \int_{\mathbb{R}^d} (\rho^1(x))^{m-1} \, d\mu^1(x)$$
$$= \frac{1}{m-1} \int_{\mathbb{R}^d} \left(\rho(x + h\nabla V(x)) \det(I + hD^2 V(x)) \right)^{m-1} \, d\mu^1(x)$$
$$= \frac{1}{m-1} \int_{\mathbb{R}^d} \left(\rho(y) \det \left(I + hD^2 V(J_h^V(y)) \right) \right)^{m-1} \, d\mu(y). \tag{3.17}$$

Using the inequality $\det(I + A) \leq e^{\operatorname{tr}(A)}$, which holds for any non-negative symmetric matrix A, we obtain

$$\mathcal{F}(\mu^1) \leq \frac{e^{(m-1)ch}}{m-1} \int_{\mathbb{R}^d} \left(\rho(y) \right)^{m-1} \, d\mu(y) = e^{(m-1)ch} \mathcal{F}(\mu), \tag{3.18}$$

which proves (3.9).

(4) As the proof is very similar to the proof of (2), we will only give a sketch of the argument.

Set $\mu^1 := J_h^{\mathcal{F}}\mu$. Arguing as in the proof of (2), we infer that V^- and V^+ are contained in $L^1(\mathbb{R}^d; \mu^1)$. Furthermore, Lemma 3.1(3) implies that $\mu^1 = \rho^1 \mathscr{L}^d$ for some $\rho^1 \in W^{1,1}(\mathbb{R}^d)$. Using the convexity of V and (3.6),

$$
\begin{aligned}
\mathcal{V}(\mu^1) &= \int_{\mathbb{R}^d} V(x)\, d\mu^1(x) \\
&\leq \int_{\mathbb{R}^d} V(T(x)) - \langle \nabla V(x), T(x) - x \rangle\, d\mu^1(x) \\
&= \mathcal{V}(\mu) - h \int_{\mathbb{R}^d} \langle \nabla V(x), \nabla(\rho^1)^m(x) \rangle\, dx.
\end{aligned}
$$

As in the proof of (2), it follows that

$$
\int_{\mathbb{R}^d} \langle \nabla V(x), \nabla(\rho^1)^m(x) \rangle\, dx = -\int_{\mathbb{R}^d} \Delta V(x)\rho^1(x)^m\, dx, \tag{3.19}
$$

and consequently,

$$
\begin{aligned}
\mathcal{V}(\mu^1) &\leq \mathcal{V}(\mu) + h \int_{\mathbb{R}^d} \Delta V(x)(\rho^1(x))^m\, dx \\
&\leq \mathcal{V}(\mu) + ch \int_{\mathbb{R}^d} (\rho^1(x))^m\, dx = \mathcal{V}(\mu) + c(m-1)h\mathcal{F}(\mu^1).
\end{aligned}
$$

$\square$

Remark 3.3. The estimates (3.12)–(3.13) extend a recent result by Tudorascu [10, Theorem 1], who considered the special case $V(x) = \frac{1}{2}|x|^2$.

Theorem 3.4. *In each of the following four cases, the functionals φ^1 and φ^2 satisfy (A_1), (A_2), and (A_3):*

(1) $\varphi^1 = \mathcal{H}$ and $\varphi^2 = \mathcal{V}$,　　*(3) $\varphi^1 = \mathcal{F}$ and $\varphi^2 = \mathcal{V}$,*
(2) $\varphi^1 = \mathcal{V}$ and $\varphi^2 = \mathcal{H}$,　　*(4) $\varphi^1 = \mathcal{V}$ and $\varphi^2 = \mathcal{F}$.*

As a consequence, the Trotter product formula from Theorem 1.1 holds.

Proof. In case (1), Proposition 3.2(1) implies $(A_2)(ii)$ and Assumption (1) of Proposition 1.7, hence (A_3).

In case (2), Proposition 3.2(2) implies $(A_2)(ii)$ and Assumption (1) of Proposition 1.7, hence (A_3).

In case (3), Proposition 3.2(3) implies $(A_2)(ii)$ and Assumption (2) of Proposition 1.7, hence (A_3).

In case (4), Proposition 3.2(3 & 4) implies $(A_2)(ii)$ and Assumption (3) of Proposition 1.7, hence (A_3). $\square$

References

[1] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, second ed., Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2008.

[2] H. Brezis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland Publishing Co., Amsterdam, 1973, North-Holland Mathematics Studies, No. 5. Notas de Matemática (50).

[3] Ph. Clément, *Introduction to gradient flows in metric spaces (II)*, available at https://igk.math.uni-bielefeld.de/study-materials/notes-clement-part2.pdf.

[4] Ph. Clément and W. Desch, *Some remarks on the equivalence between metric formulations of gradient flows*, Boll. Unione Mat. Ital. (9) **3** (2010), no. 3, 583–588.

[5] L. C. Evans and R. F. Gariepy, *Measure theory and fine properties of functions*, Studies in Advanced Mathematics, CRC Press, Boca Raton, FL, 1992.

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

[7] T. Kato and K. Masuda, *Trotter's product formula for nonlinear semigroups generated by the subdifferentials of convex functionals*, J. Math. Soc. Japan **30** (1978), no. 1, 169–178.

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

[9] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations **26** (2001), no. 1-2, 101–174.

[10] A. Tudorascu, *On the Jordan-Kinderlehrer-Otto variational scheme and constrained optimization in the Wasserstein metric*, Calc. Var. Partial Differential Equations **32** (2008), no. 2, 155–173.

[11] C. Villani, *Optimal transport, old and new*, Grundlehren der Mathematischen Wissenschaften, vol. 338, Springer-Verlag, 2009.

DELFT INSTITUTE OF APPLIED MATHEMATICS, DELFT UNIVERSITY OF TECHNOLOGY, P.O. BOX 5031, 2600 GA DELFT, THE NETHERLANDS
   E-mail address: ppjeclement@aim.com

INSTITUTE FOR APPLIED MATHEMATICS, UNIVERSITY OF BONN, ENDENICHER ALLEE 60, 53115 BONN, GERMANY
   E-mail address: maas@iam.uni-bonn.de