**Noname manuscript No.**
(will be inserted by the editor)

# Discretization of functionals involving the Monge-Ampère operator

**Jean-David Benamou · Guillaume Carlier · Quentin Mérigot · Édouard Oudet**

the date of receipt and acceptance should be inserted later

**Abstract** Gradient flows in the Wasserstein space have become a powerful tool in the analysis of diffusion equations, following the seminal work of Jordan, Kinderlehrer and Otto (JKO). The numerical applications of this formulation have been limited by the difficulty to compute the Wasserstein distance in dimension $\geqslant 2$. One step of the JKO scheme is equivalent to a variational problem on the space of convex functions, which involves the Monge-Ampère operator. Convexity constraints are notably difficult to handle numerically, but in our setting the internal energy plays the role of a barrier for these constraints. This enables us to introduce a consistent discretization, which inherits convexity properties of the continuous variational problem. We show the effectiveness of our approach on nonlinear diffusion and crowd-motion models.

## 1 Introduction

### 1.1 Context

*Optimal transport and displacement convexity*

In the following, we consider two probability measures $\mu$ and $\nu$ on $\mathbb{R}^d$ with finite second moments, the first of which is absolutely continuous with respect

J.-D. Benamou
Inria Rocquencourt

G. Carlier
Ceremade, Université Paris-Dauphine

Q. Mérigot
Laboratoire Jean Kuntzmann, Université Grenoble-Alpes / CNRS

É. Oudet
Laboratoire Jean Kuntzmann, Université Grenoble-Alpes

to the Lebesgue measure. We are interested in the quadratic optimal transport problem between $\mu$ and $\nu$:

$$\min\left\{\int_X \|T(x)-x\|^2 \,\mathrm{d}\mu(x);\, T : X \to \mathbb{R}^d,\ T_{\#}\mu = \nu\right\} \tag{1.1}$$

where $T_{\#}\mu$ denotes the pushforward of $\mu$ by $T$. A theorem of Brenier shows that the optimal map in (1.1) is given by the gradient of a convex function [8]. Define the Wasserstein distance between $\mu$ and $\nu$ as the square root of the minimum in (1.1), and denote it $\mathrm{W}_2(\mu,\nu)$. Denoting by $\mathcal{K}$ the space of convex functions, Brenier's theorem implies that for $\varphi \in \mathcal{K}$,

$$\mathrm{W}_2^2(\mu, \nabla\varphi_{\#}\mu) = \int_{\mathbb{R}^d} \|x - \nabla\varphi(x)\|^2 \,\mathrm{d}\mu(x) \tag{1.2}$$

and that the map defined

$$\varphi \in \mathcal{K} \mapsto \nabla\varphi_{\#}\mu \in \mathcal{P}(\mathbb{R}^d), \tag{1.3}$$

is onto. This map can be seen as a parameterization, depending on $\mu$, of the space of probability measures by the set of convex potentials $\mathcal{K}$. This idea has been exploited by McCann [20] to study the steady states of gases whose energy $\mathcal{F} : \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$ is the sum of an internal energy $\mathcal{U}$, such as the negative entropy, and an interaction energy $\mathcal{E}$. McCann gave sufficient conditions for such a functional $\mathcal{F}$ to be convex along minimizing Wasserstein geodesics. These conditions actually imply a stronger convexity property for the functional $\mathcal{F}$: this functional is convex under generalized displacement: for any absolutely continuous probability measure $\mu$, the composition of $\mathcal{F}$ with the parameterization given in Eq. (1.3), $\varphi \in \mathcal{K} \mapsto \mathcal{F}(\nabla\varphi_{\#\mu})$, is convex. Generalized displacement convexity allows one to turn a non-convex optimization problem over the space of probability measures into a convex optimization problem on the space of convex functions.

*Gradient flows in Wasserstein space and JKO scheme*

Our goal is to simulate numerically non-linear evolution PDEs which can be formulated as gradient flows in the Wasserstein space. The first formulation of this type has been introduced in the seminal article of Jordan, Kinderlehrer and Otto [16]. The authors considered the linear Fokker-Planck equation

$$\begin{cases} \dfrac{\partial\rho}{\partial t} = \Delta\rho + \mathrm{div}(\rho\nabla V) \\ \rho(0,.) = \rho_0 \end{cases}, \tag{1.4}$$

where $\rho(t,.)$ is a time-varying probability density on $\mathbb{R}^d$ and $V$ is a potential energy. The main result of the article is that (1.4) can be reinterpreted as the gradient flow in the Wasserstein space of the energy functional

$$\mathcal{F}(\rho) = \int_{\mathbb{R}^d} (\log\rho(x) + V(x))\rho(x)\mathrm{d}x. \tag{1.5}$$

Jordan, Kinderlehrer and Otto showed how to construct such a gradient flow through a time-discretization, using a generalization of the backward Euler scheme. Given a timestep $\tau$, one defines recursively a sequence of probability densities $(\rho_k)_{k\geqslant 0}$ :

$$\rho_{k+1} = \arg\min_{\rho\in\mathcal{P}^{\mathrm{ac}}(\mathbb{R}^d)} \frac{1}{2\tau}\mathrm{W}_2^2(\rho_k,\rho) + \mathcal{F}(\rho). \tag{1.6}$$

The main theorem of [16] is that the discrete gradient flow constructed by (1.6) converges to the solution of the Fokker-Planck equation (1.4) in a suitable weak sense as $\tau$ tends to zero. Similar formulations have been proposed for other non-linear partial differential equations : the porous medium equation [25] and more general degenerate parabolic PDEs [3], the sub-critical Keller-Segel equation [6], macroscopic models of crowds [19], to name but a few. The construction and properties of gradient flows in the Wasserstein space have been studied systematically in [5]. Finally, even solving for a single step of the JKO scheme leads to nontrivial nonlocal PDEs of Monge-Ampère type which appear for instance in the Cournot-Nash problem in game theory [7].

## 1.2 Previous work

*Numerical resolution of gradient flows.*

Despite the potential applications, there exists very few numerical simulations that use the Jordan-Kinderlehrer-Otto scheme and its generalizations. The main reason is that the first term of the functional that one needs to minimize at each time step, e.g. Eq. (1.6), is the Wasserstein distance. Computing the Wasserstein distance and its gradient is notably difficult in dimension two or more. In dimension one however, the optimal transport problem is much simpler because of its relation to monotone rearrangement. This remark has been used to implement discrete gradient flows for the quadratic cost [17,6,7] or for more general convex costs [4]. In 2D, the Lagragian method proposed in [12,9] is inspired by the JKO formulation but the convexity of the potential is not enforced.

*Calculus of variation under convexity constraints.*

When the functional $\mathcal{F}$ is convex under generalized displacement, one can use the parameterization Eq. (1.3) to transform the problem into a convex optimization problem over the space of convex functions. Optimization problems over the space of convex functions are also frequent in economy and geometry, and have been studied extensively, from a numerical viewpoint, when $\mathcal{F}$ is an integral functional that involve function values and gradients:

$$\min_{\varphi\in\mathcal{K}} \int_\Omega F(x,\varphi(x),\nabla\varphi(x))\mathrm{d}x \tag{1.7}$$

The main difficulty to solve this minimization problem numerically is to construct a suitable discretization of the space of convex functions over $\Omega$. The first approach that has been considered is to approximate $\mathcal{K}$ by piecewise linear functions over a fixed mesh. This approach has an important advantage: the number of linear constraints needed to ensure that a piecewise linear function over a mesh is convex is proportional to the size of the mesh. Unfortunately, Choné and Le Meur [13] showed that there exists convex functions on the unit square that cannot be approximated by piecewise-linear convex functions on the regular grid with edgelength $\delta$, even as $\delta$ converges to zero. This difficulty has generated an important amount of research in the last decade.

Finite difference approaches have been proposed by Carlier, Lachand-Robert and Maury [11], based on the notion of convex interpolate, and by Ekeland and Moreno-Bromberg using the representation of a convex function as a maximum of affine functions [14], taking inspiration from Oudet and Lachand-Robert [18]. In both methods, the number of linear inequality constraints used to discretize the convexity constraints is quadratic in the number of input points, thus limiting the applicability of these methods. More recently, Mirebeau proposed a refinement of these methods in which the set of active constraints is learned during the optimization process [22]. Oberman used the idea of imposing convexity constraints on a wide-stencil [24], which amounts to only selecting the constraints that involve nearby points in the formulation of [11]. Oudet and Mérigot [21] used interpolation operators to approximate the solutions of (1.7) on more general finite-dimensional spaces of functions. All these methods can be used to minimize functionals that involve the value of the function and its gradient only. They are not able to handle terms that involve the Monge-Ampère operator $\det \mathrm{D}^2\varphi$ of the function, which appears when e.g. considering the negative entropy of $\nabla\varphi|_{\#}\rho$. It is worth mentioning here that convex variational problems with a convexity constraint and involving the Monge-Ampère operator $\det \mathrm{D}^2\varphi$ appear naturally in geometric problems such as the affine Plateau problem, see Trudinger and Wang [27] or Abreu's equation, see Zhou [28]. The Euler-Lagrange equations of such problems are fully nonlinear fourth-order PDEs and looking numerically for convex solutions can be done by similar methods as the ones developed in the present paper.

### 1.3 Contributions.

In this article, we construct a discretization in space of the type of variational problems that appear in the definition of the JKO scheme. More precisely, given two bounded convex subsets $X, Y$ of $\mathbb{R}^d$, and an absolutely continuous measure $\mu$ on $X$, we want to discretize in space the minimization problem

$$\min_{\nu\in\mathcal{P}(Y)} \mathrm{W}_2^2(\mu,\nu) + \mathcal{E}(\nu) + \mathcal{U}(\nu), \tag{1.8}$$

where $\mathcal{P}(Y)$ denotes the set of probability measures on $Y$, and where the potential energy $\mathcal{E}$ and the internal energy $\mathcal{U}$ are defined as follows:

$$\mathcal{E}(\nu) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} W(x,y)\mathrm{d}[\nu\nu](x,y) + \int_{\mathbb{R}^d} V(x)\mathrm{d}\nu(x) \tag{1.9}$$

$$\mathcal{U}(\nu) = \begin{cases} \displaystyle\int_{\mathbb{R}^d} U(\sigma(x))\mathrm{d}x \text{ if } \mathrm{d}\nu = \sigma\mathrm{d}\mathcal{H}^d, \sigma \in \mathrm{L}^1(\mathbb{R}^d) \\ +\infty \text{ if not} \end{cases} \tag{1.10}$$

We assume McCann's sufficient conditions [20] for the generalized displacement convexity of the functional $\mathcal{F} = \mathrm{W}_2^2(\mu,.) + \mathcal{E} + \mathcal{U}$, namely:

(HE) the potential $V : \mathbb{R}^d \to \mathbb{R}$ and interaction potential $W : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ are convex functions. (If in addition $V$ or $W$ is strictly convex, we denote this assumption (HE$^+$))

(HU) The function $U : \mathbb{R}^+ \to \mathbb{R}$ is such that the map $r \mapsto r^d U(r^{-d})$ is convex and non-increasing, and $U(0) = 0$. (If the convexity of $r \mapsto r^d U(r^{-d})$ is strict, we denote this assumption (HU$^+$).)

Under assumptions (HE) and (HU), the problem (1.8) can be rewritten as a convex optimization problem. Introducing the space $\mathcal{K}_Y$ of convex functions on $\mathbb{R}^d$ whose gradient lie in $Y$ almost everywhere, (1.8) is equivalent to

$$\min_{\varphi\in\mathcal{K}_Y} \mathrm{W}_2^2(\mu, \nabla\varphi_{\#}\mu) + \mathcal{E}(\nabla\varphi_{\#}\mu) + \mathcal{U}(\nabla\varphi_{\#}\mu). \tag{1.11}$$

Our contributions are the following:

- In Section 2, we discretize the space $\mathcal{K}_Y$ of convex functions with gradients contained in $Y$ by associating to every finite subset $P$ of $\mathbb{R}^d$ a finite-dimensional convex subset $\mathcal{K}_Y(P)$ contained in the space of real-valued functions on the finite-set $P$. We construct a discrete Monge-Ampère operator, in the spirit of Alexandrov, which satisfies some structural properties of the operator $\varphi \mapsto \det(\mathrm{D}^2\varphi)$, such as Minkowski's determinant inequality. Moreover, we show how to modify the construction of $\mathcal{K}_Y(P)$ so as to get a linear gradient operator, following an idea of Ekeland and Moreno-Bromberg [14].
- In Section 3, we construct a convex discretization of the problem (1.11). In order to do so, we need to define an analogue of $\nabla\varphi_{\#}\mu$, where $\varphi$ is a function in our discrete space $\mathcal{K}_Y(P)$ and where $\mu_P$ is a measure supported on $P$. It turns out that in order to maintain the convexity of the discrete problem, one needs to define two such notions: the pushforward $G^{\mathrm{ac}}_{\varphi\#}\mu_P$ which is absolutely continous on $Y$ and whose construction involves the discrete Monge-Ampère operator, and $G_{\varphi\#}\mu_P$ which is supported on a finite set and whose construction involves the discrete gradient. The discretization of (1.11) is given by

$$\min_{\varphi\in\mathcal{K}_Y(P)} \mathrm{W}_2^2(\mu, G_{\varphi\#}\mu_P) + \mathcal{E}(G_{\varphi\#}\mu_P) + \mathcal{U}(G^{\mathrm{ac}}_{\varphi\#}\mu_P). \tag{1.12}$$

- In Section 4, we show that if $(\mu_{P_n})_{n\geqslant 0}$ is a sequence of probability measures on $X$ that converge to $\mu$ in the Wasserstein sense, minimizers of the discretized problem (1.12) with $P = P_n$ converge, in a sense to be made precise, to minimizers of the continuous problem. In order to prove this result, we need a few additional assumptions: the density of $\mu$ should be bounded from above and below on the convex domain $X$, and the integrand in the definition of the internal energy (1.10) should be convex.
- Finally, in Section 5 we present two numerical applications of the space-discretization (1.12). Our first simulation is a meshless Lagrangian simulation of the porous medium equation and the fast-diffusion equation using the gradient flow formulation of Otto [25]. The second simulation concerns the gradient-flow model of crowd motion introduced by Maury, Roudneff-Chupin and Santambrogio [19].

Notation

The Lebesgue measure is denoted $\mathcal{H}^d$. The space of probability measures on a domain $X$ of $\mathbb{R}^d$ is denoted $\mathcal{P}(X)$, while $\mathcal{P}^{\mathrm{ac}}(X)$ denotes the space of probability measures that are absolutely continuous with respect to the Lebesgue measure.

## 2 Discretization of the space of convex functions

The first goal of this section is to discretize the space of convex functions whose gradients lie in a prescribed convex set $Y$. Then, we will define a notion of discrete Monge-Ampère operator for functions in this space. We will consider functions from $\mathbb{R}^d$ to the set of extended reals $\overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$.

**Definition 1 (Legendre-Fenchel transform)** The Legendre transform $\psi^*$ of a function $\psi : Y \to \overline{\mathbb{R}}$, is the function $\psi^* : \mathbb{R}^d \to \overline{\mathbb{R}}$ defined by the formula

$$\psi^*(x) := \sup_{y\in Y} \langle x|y\rangle - \psi(y). \tag{2.13}$$

The space of Legendre-Fenchel transforms of functions defined over a convex set $Y$ is denoted by $\mathcal{K}_Y := \{\psi^*; \psi : Y \to \overline{\mathbb{R}}\}$. A function on $\mathbb{R}^d$ is called *trivial* if it is constant and equal to $+\infty$. The space of non-trivial functions in $\mathcal{K}_Y$ is denoted $\mathcal{K}_Y^0$.

**Lemma 1** *Assume that $Y$ is a bounded convex subset of $\mathbb{R}^d$. Then,*

*(i) functions in $\mathcal{K}_Y$ are trivial or finite everywhere: $\mathcal{K}_Y = \mathcal{K}_Y^0 \cup \{+\infty\}$;*
*(ii) a convex function $\varphi$ belongs to $\mathcal{C}^1 \cap \mathcal{K}_Y^0$ if and only if $\nabla\varphi(\mathbb{R}^d) \subseteq Y$ ;*
*(iii) the set $\mathcal{C}^1 \cap \mathcal{K}_Y^0$ is dense in the set $\mathcal{K}_Y^0$ for $\|.\|_\infty$;*
*(iv) the space $\mathcal{K}_Y$ is convex;*
*(v) (stability by maximum) given a family of functions $(\varphi_i)_{i\in I}$ in $\mathcal{K}_Y$, the function $\varphi(x) := \sup_{i\in I} \varphi_i(x)$ is also in $\mathcal{K}_Y$.*

*Proof* (i) We assume that $\varphi$ belongs to $\mathcal{K}_Y$, i.e. $\varphi = \psi^*$, where $\psi$ is a function from $Y$ to $\overline{\mathbb{R}}$. We will first show that if $\varphi$ is non-trivial, then $\psi$ is lower bounded by a constant on $Y$. By contradiction, assume that there exists a set of points $y_k$ in $Y$ such that $\psi(y_k) \to -\infty$. In this case, given any point $x$ in $X$ we have

$$\psi^*(x) \geqslant \max_k \langle x|y_k\rangle - \psi(y_k) \geqslant \max_k - \|x\| \, \|y_k\| - \psi(y_k) = +\infty,$$

so that $\varphi$ is trivial.

(iii) Assume that $\varphi$ belongs to $\mathcal{K}_Y^0$, so that there exists a convex function $\psi : Y \to \mathbb{R}$ lower bounded by a constant and such that $\psi^* = \varphi$. Then, we can approximate $\psi$ by uniformly convex functions $\psi_\varepsilon(y) := \psi(y) + \varepsilon \|y\|^2$ on $Y$. The functions $\varphi_\varepsilon := \psi_\varepsilon^*$ belong to $\mathcal{K}_Y$, are smooth, and uniformly converge to the function $\varphi$.

**Definition 2 ($\mathcal{K}_Y$-envelope and $\mathcal{K}_Y$-interpolate)** The *$\mathcal{K}_Y$–envelope* of a function $\varphi$ defined on a subset $P$ of $\mathbb{R}^d$ is the largest function in $\mathcal{K}_Y$ whose restriction to $P$ lies below $\varphi$. In other words,

$$\varphi_{\mathcal{K}_Y} := \max\{\psi \in \mathcal{K}_Y;\, \psi|_P \leqslant \varphi|_P\}. \tag{2.14}$$

A function $\varphi$ on a set $P \subseteq \mathbb{R}^d$ is a *$\mathcal{K}_Y$-interpolate* if it coincides with the restriction to $P$ of its $\mathcal{K}_Y$–envelope. The space of $\mathcal{K}_Y$–interpolates is denoted

$$\mathcal{K}_Y(P) := \{\varphi : P \to \mathbb{R}; \varphi = \varphi_{\mathcal{K}_Y}|_P\}. \tag{2.15}$$

## 2.1 Subdifferential and Laguerre cells

Consider a convex function $\varphi$ on $\mathbb{R}^d$, and a point $x$. A vector $y \in \mathbb{R}^d$ is a *subgradient* of $\varphi$ at $x$ if for every $z$ in $\mathbb{R}^d$, the inequality $\varphi(z) \geqslant \varphi(x) + \langle z - x|y\rangle$ holds. The *subdifferential* of $\varphi$ at $x$ is the set of subgradients to $\varphi$ at $x$, i.e.

$$\partial\varphi(x) := \{y \in \mathbb{R}^d; \forall z \in \mathbb{R}^d, \varphi(z) \geqslant \varphi(x) + \langle z - x|y\rangle\} \tag{2.16}$$

The following lemma allows one to compute the subdifferential of the $\mathcal{K}_Y$–envelope of a function in $\mathcal{K}_Y(P)$.

**Definition 3 (Laguerre cell)** Given a finite point set $P$ contained in $\mathbb{R}^d$, a function $\varphi$ on $P$, we denote the *Laguerre cell* of a point $p$ in $P$ the polyhedron

$$\mathrm{Lag}_P^\varphi(p) := \{y \in \mathbb{R}^d; \forall q \in P, \varphi(q) \geqslant \varphi(p) + \langle q - p|y\rangle\}.$$

Note that the union of the Laguerre cells covers the space, while the intersection of the interior of two Laguerre cells is always empty.

**Lemma 2** *Let $P$ be a finite point set. A function $\varphi$ on $P$ belongs to $\mathcal{K}_Y(P)$ if and only if for every $p$ in $P$, the intersection $\mathrm{Lag}_P^\varphi(p) \cap Y$ is non-empty. Moreover, if this is the case, then*

$$\partial\varphi_{\mathcal{K}_Y}(p) = \mathrm{Lag}_P^\varphi(p) \cap Y. \tag{2.17}$$

*Proof* Denote $\mathcal{K} := \mathcal{K}_{\mathbb{R}^d}$ and $\varphi_{\mathcal{K}}$ the convex envelope of $\varphi$. It is then easy to see that for every point $p$ in $P$ such that $\varphi_{\mathcal{K}}(p) = \varphi(p)$,

$$\partial\varphi_{\mathcal{K}}(p) = \mathrm{Lag}_P^{\varphi}(p).$$

Since $\mathcal{K}_Y \subseteq \mathcal{K}$ and by definition, one has $\varphi_{\mathcal{K}_Y}(x) \leqslant \varphi_{\mathcal{K}}(x)$, with equality when $x$ is a point in $P$. This implies the inclusion $\partial\varphi_{\mathcal{K}_Y}(p) \subseteq Y \cap \partial\varphi_{\mathcal{K}}(p)$. In order to show that the converse also holds, one only needs to remark that

$$Y \subseteq \bigcup_{p\in P} \partial\varphi_{\mathcal{K}_Y}(p).$$

**Lemma 3** *Let $\varphi_0, \varphi_1$ in $\mathcal{K}_Y(P)$, let $\varphi_t = (1-t)\varphi_0 + t\varphi_1$ be the linear interpolation on $P$ between these functions, and denote $\hat{\varphi}_t := [\varphi_t]_{\mathcal{K}_Y}$. Then for any $p$ in $P$,*

$$\partial\hat{\varphi}_t(p) \supseteq (1-t)\partial\hat{\varphi}_0(p) + t\partial\hat{\varphi}_1(p) \tag{2.18}$$

$$\mathrm{Lag}_P^{\varphi_t}(p) \cap Y \supseteq (1-t)(\mathrm{Lag}_P^{\varphi_0}(p) \cap Y) + t(\mathrm{Lag}_P^{\varphi_1}(p) \cap Y) \tag{2.19}$$

*Proof* Thanks to the previous lemma, the two inclusions are equivalent. Now, let $y_i$ be a point in $\mathrm{Lag}_P^{\varphi_i}(p) \cap Y$, so that

$$\forall q \in P,\ \varphi_i(q) \geqslant \varphi_i(p) + \langle q - p | y_i \rangle.$$

Taking a linear combination of these inequalities, we get

$$\forall q \in P,\ (1-t)\varphi_0(q) + t\varphi_1(q) \geqslant (1-t)\varphi_0(p) + t\varphi_1(p) + \langle q - p | y_t \rangle,$$

with $y_t = (1-t)y_0 + ty_1$. In other words, the point $y_t$ belongs to the Laguerre cell $\mathrm{Lag}_P^{\varphi_t}(p)$. Since this holds for any pair of points $y_0$ in $\mathrm{Lag}_P^{\varphi_0}(p)$ and $y_1$ in $\mathrm{Lag}_P^{\varphi_1}(p)$, we get the desired inclusion.

*Remark 1* A corollary of the two previous lemmas is the convexity of the space $\mathcal{K}_Y(P)$ of $\mathcal{K}_Y$-interpolates, a fact that does not obviously follow from the definition.

*Remark 2* The convex envelope of a function defined on a finite set is always piecewise-linear. In contrast, when the domain $Y$ is bounded, the $\mathcal{K}_Y$-envelope of an element $\varphi$ of the polyhedron $\mathcal{K}_Y(P)$ does not need to be piecewise linear, even when restricted to the convex hull of $P$. Fortunately, for the applications that we are targeting, we will never need to compute this envelope explicitely, and we will only use formula (2.17) giving the explicit expression of the subdifferential.

### 2.2 Monge-Ampère operator

In this paragraph, we introduce a notion of discrete Monge-Ampère operator of $\mathcal{K}_Y$-interpolates on a finite set. This definition is closely related to the notion of Monge-Ampère measure introduced by Alexandrov. Given a smooth uniformly convex function $\varphi$ on $\mathbb{R}^d$, a change of variable gives

$$\int_B \det(\mathrm{D}^2\varphi(x))\mathrm{d}x = \int_{\nabla\varphi(B)} 1\mathrm{d}x = \mathcal{H}^d(\nabla\varphi(B)). \tag{2.20}$$

This equation allows one to define a measure on the source domain $X \subseteq \mathbb{R}^d$, called the Monge-Ampère measure and denoted $\mathrm{MA}[\varphi]$. Using the right-hand side of the equality, it is possible to extend the notion of Monge-Ampère measure to convex functions that are not necessarily smooth (see e.g. [15]):

$$\mathrm{MA}[\varphi](B) := \mathcal{H}^d(\partial\varphi(B)). \tag{2.21}$$

**Definition 4** The discrete Monge-Ampère operator of a $\mathcal{K}_Y$-interpolate $\varphi : P \to \mathbb{R}$ at a point $p$ in $P$ is defined by the formula:

$$\mathrm{MA}_Y[\varphi](p) := \mathcal{H}^d(\partial\varphi_{\mathcal{K}_Y}(p)), \tag{2.22}$$

where $\mathcal{H}^d$ denotes the $d$-dimensional Lebesgue measure.

The relation between the discrete Monge-Ampère operator and the Monge-Ampère measure is given by the formula:

$$\forall\varphi \in \mathcal{K}_Y(P),\ \mathrm{MA}[\varphi_{\mathcal{K}_Y}] = \sum_{p\in P} \mathrm{MA}_Y[\varphi](p)\delta_p. \tag{2.23}$$

In other words, the Monge-Ampère operator can be seen as the density of the Monge-Ampère measure of $\varphi_{\mathcal{K}_Y}$ with respect to the counting measure on $P$. The next lemma is crucial to the proof of convexity of our discretized energies. It is also interesting in itself, as it shows that the interior of the set $\mathcal{K}_Y(P)$ of convex interpolates can be defined by $|P|$ explicit non-linear convex constraints.

**Lemma 4** *For any point $p$ in $P$, the following map is convex:*

$$\varphi \in \mathcal{K}_Y(P) \mapsto -\log(\mathrm{MA}_Y[\varphi](p)). \tag{2.24}$$

*Proof* Let $\varphi_0, \varphi_1$ in $\mathcal{K}_Y(P)$, let $\varphi_t = (1-t)\varphi_0 + t\varphi_1$ be the linear interpolation between these functions, and denote $\hat{\varphi}_t := [\varphi_t]_{\mathcal{K}_Y}$. Using Lemma 3, and with the convention $\log(0) = -\infty$, we have

$$\begin{aligned}\log(\mathcal{H}^d(\partial\hat{\varphi}_t(p))) &\geqslant \log(\mathcal{H}^d((1-t)\partial\hat{\varphi}_0(p) + t\partial\hat{\varphi}_1(p)))\\ &\geqslant (1-t)\log(\mathcal{H}^d(\partial\hat{\varphi}_0(p))) + t\log(\mathcal{H}^d(\partial\hat{\varphi}_1(p))),\end{aligned}$$

where the second inequality is the logarithmic version of the Brunn-Minkowski inequality.

### 2.3 Convex interpolate with gradient

In applications, we want to minimize energy functionals over the space $\mathcal{K}_Y$, which involve potential energy terms such as

$$\varphi \mapsto \int_X V(\nabla\varphi(x))\mathrm{d}\mu(x), \tag{2.25}$$

where $V$ is a convex potential on $\mathbb{R}^d$. Any functional defined this way is convex in $\varphi$, and one would like to be able to define a discretization of this functionals that preserves this property. Given a function $\varphi$ in the space $\mathcal{K}_Y(P)$ and a point $p$ in $P$, one wants to select a vector in the subdifferential $\partial\varphi_{\mathcal{K}_Y}(p)$, and this vector needs to depend linearly on $\varphi$. A way to achieve this is to increase the dimension of the space of variables, and to include the chosen subgradients as unknown of the problem. This can be done as in Ekeland and Moreno-Bromberg [14].

**Definition 5 (Convex interpolate with gradient)** A $\mathcal{K}_Y$*-interpolate with gradient* on a finite subset $P$ of $\mathbb{R}^d$ is a couple $(\varphi, G_\varphi)$ consisting of a function $\varphi$ in the space of $\mathcal{K}_Y$-interpolates $\mathcal{K}_Y(P)$ and a gradient map $G_\varphi : P \to \mathbb{R}^d$ such that

$$\forall p \in P,\ G_\varphi(p) \in \partial\varphi_{\mathcal{K}_Y}(p). \tag{2.26}$$

The space of convex interpolates with gradients is denoted $\mathcal{K}_Y^G(P)$.

Note that the space $\mathcal{K}_Y^G(P)$ can be considered as a convex subset of the vector space of functions from $P$ to $\mathbb{R} \times \mathbb{R}^d$. Lemma 5 below implies that $\mathcal{K}_Y^G(P)$ forms a convex subset of this vector space, for which one can construct explicit convex barriers. Given a closed subset $A$ of $\mathbb{R}^d$ and $x$ a point of $\mathbb{R}^d$, $\mathrm{d}(x, A)$ denotes the minimum distance between $x$ and any point in $A$.

**Lemma 5** *Let $\varphi_0$ and $\varphi_1$ be two functions in $\mathcal{K}_Y(P)$ and let $v_i$ a vector in the subdifferential $\partial\hat{\varphi}_i(p)$ for a certain point $p$ in $P$. Then,*

*(i) the vector $v_t = (1-t)v_0 + tv_1$ lies in $\partial\hat{\varphi}_t(p)$;*
*(ii) the map $t \mapsto \mathrm{d}(v_t, \mathbb{R}^d \setminus \partial\hat{\varphi}_t(p))$ is concave;*

*Moreover, a function $\varphi$ belongs to the interior of $\mathcal{K}_Y(P)$ if and only if*

$$\forall p \in P,\ \mathrm{MA}_Y[\varphi](p) > 0. \tag{2.27}$$

*Proof* The first item is a simple consequence of Lemma 3. In order to prove the second item, we first remark that setting $R_i := \mathrm{d}(v_i, \mathbb{R}^d \setminus \partial\hat{\varphi}_i(p))$, one has: $\mathrm{B}(v_i, R_i) \subseteq \partial\hat{\varphi}_i(p)$. Using the second inclusion from Lemma 3 and the explicit formula for the Minkowski sum of balls, we get:

$$(1-t)\mathrm{B}(v_0, R_0) + t\mathrm{B}(v_1, R_1) = \mathrm{B}(v_t, (1-t)R_0 + tR_1) \subseteq \partial\hat{\varphi}_t(p).$$

This implies the desired concavity property:

$$\mathrm{d}(v_t, \mathbb{R}^d \setminus \partial\hat{\varphi}_t(p)) \geqslant (1-t)R_0 + tR_1.$$

As for the last assertion, assume by contradiction that there is a $p \in P$ such that $\partial\varphi_{\mathcal{K}_Y}(p)$ has empty interior, and let $y \in \partial\varphi_{\mathcal{K}_Y}(p)$. Since the Laguerre cells cover the space, this means that $y$ also belongs to (the boundary) of Laguerre cells with nonempty interior corresponding to points $p_1, \dots, p_k \in P^k$ for some $k \geq 2$. In this case, $p$ necessarily is in the relative interior of the convex hull of $\{p_1, \dots, p_k\}$ and $\varphi_{\mathcal{K}_Y}$ is affine on this convex hull, contradicting interiority of $\varphi$.

## 3 Convex discretization of displacement-convex functionals

In the discrete setting, the reference probability density $\rho$ is replaced by a probability measure $\mu$ on a finite point set. Since the subdifferential of a convex function $\varphi$ can be multi-valued, the pushforward $\nabla\varphi_{\#}\mu$ is not uniquely defined in general. In order to maintain the convexity properties of the three functionals in our discrete setting, we will need to consider two different types of push-forwards.

**Definition 6 (Push-forwards)** Let $\mu$ be a probability measure supported on a finite point set $P$, i.e. $\mu = \sum_{p\in P} \mu_p \delta_p$. We consider a convex interpolate with gradient $(\varphi, G_\varphi)$ in $\mathcal{K}_Y^G(P)$ and we define two ways of pushing forward the measure $\mu$ by the gradient of $\varphi_{\mathcal{K}_Y}$.

- The first way consists in moving each Dirac mass $\mu_p\delta_p$ to the selected subgradient $G_\varphi(p)$, thus defining

$$G_{\varphi\#}\mu := \sum_{p\in P} \mu_p \delta_{G_\varphi(p)}. \tag{3.28}$$

- The second possibility, is to spread each Dirac mass $\mu_p\delta_p$ on the whole subdifferential $\partial\varphi_{\mathcal{K}_Y}(p)$. This defines, when $\varphi$ is in the interior of $\mathcal{K}_Y(P)$, an absolutely continuous measure:

$$G^{\mathrm{ac}}_{\varphi\#}\mu := \sum_{p\in P} \mu_p \frac{\mathcal{H}^d\big|_{\partial\varphi_{\mathcal{K}_Y}(p)}}{\mathcal{H}^d(\partial\varphi_{\mathcal{K}_Y}(p))}. \tag{3.29}$$

*Remark 3* Note that in both cases, the mass of $\mu$ located at $p$ is transported into the subdifferential $\partial\varphi_{\mathcal{K}_Y}(p)$. This implies that the transport plan between $\mu$ and $G_{\varphi\#}\mu$ induced by this definition is optimal, and similarly for $G^{\mathrm{ac}}_{\varphi\#}\mu$. We therefore have an explicit expression for the squared Wasserstein distance between $\mu$ and these pushforwards:

$$\mathrm{W}_2^2(\mu, G_{\varphi\#}\mu) = \sum_{p\in P} \mu_p \left\|p - G_\varphi(p)\right\|^2 \tag{3.30}$$

$$\mathrm{W}_2^2(\mu, G^{\mathrm{ac}}_{\varphi\#}\mu) = \sum_{p\in P} \frac{\mu_p}{\mathcal{H}^d(\partial\varphi_{\mathcal{K}_Y}(p))} \int_{\partial\varphi_{\mathcal{K}_Y}(p)} \left\|p - x\right\|^2 \mathrm{d}x \tag{3.31}$$

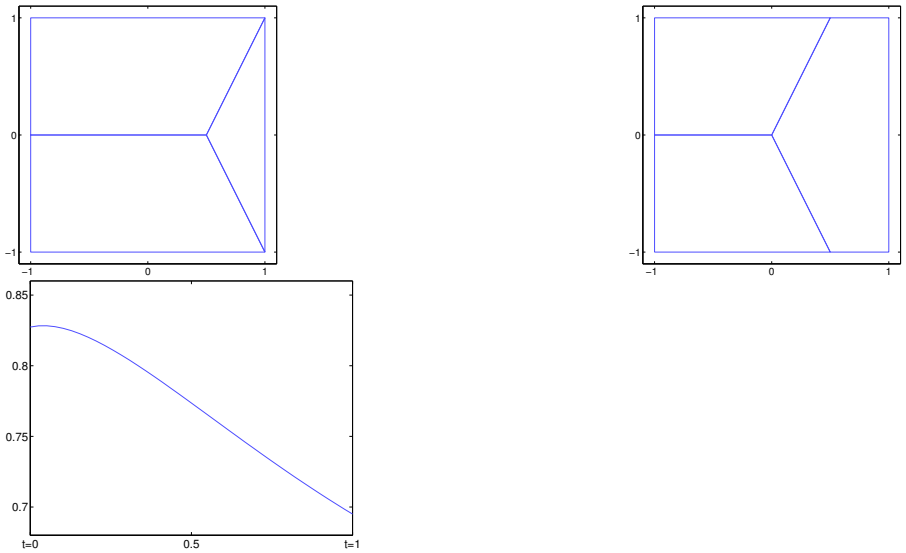


**Fig. 1** We consider a point set $P = \{q, p_\pm\}$, with $q = (2, 0)$ and $p_\pm = (0, \pm 1)$, and a function $\varphi_t$ which linearly interpolates between $\varphi_0 = \chi_{\{q\}}$ and $\varphi_1 = 0$. (Left) Laguerre cells $\mathrm{Lag}_P^{\varphi_t}(p)$ intersected with the square $[-1, 1]^2$ at $t = 0$. (Middle) Laguerre cells at $t = 1$. (Right) Graph of the second moment of the measure $G^{\mathrm{ac}}_{\varphi_t \#}\mu$ as a function of $t$, showing the lack of convexity of a discretized energy.

**Theorem 1** *Given a bounded convex set $Y$ and a measure $\mu$ supported on a finite set $P$, and under hypothesis (HE) and (HU), the maps*

$$(\varphi, G_\varphi) \in \mathcal{K}_Y^G(P) \mapsto \mathcal{E}(G_{\varphi\#}\mu) \tag{3.32}$$

$$\varphi \in \mathcal{K}_Y(P) \mapsto \mathcal{U}(G^{\mathrm{ac}}_{\varphi\#}\mu) \tag{3.33}$$

*are convex. Moreover, under assumptions (HE$^+$) and (HU$^+$) the functional*

$$(\varphi, G_\varphi) \in \mathcal{K}_Y^G(P) \mapsto \mathcal{F}(\varphi) := \mathcal{E}(G_{\varphi\#}\mu) + \mathcal{U}(G^{\mathrm{ac}}_{\varphi\#}\mu) \tag{3.34}$$

*has the following strict convexity property: given two functions $\varphi_0$ and $\varphi_1$ in $\mathcal{K}_Y^G(P)$, and $\varphi_t = (1-t)\varphi_0 + t\varphi_1$ with $t \in (0, 1)$, then*

$$\mathcal{F}(\varphi_t) \leqslant (1-t)\mathcal{F}(\varphi_0) + t\mathcal{F}(\varphi_1),$$

*with equality only if $\varphi_0 - \varphi_1$ is a constant. In particular, there is at most one minimizer of $\mathcal{F}$ up to an additive constant.*

*Proof* The proof of (3.33) uses the log-concavity of the discrete Monge-Ampère operator as in Lemma 4 and McCann's condition [20]. The proof of (3.32) is direct: if $(\varphi_0, G_{\varphi_0})$ and $(\varphi_0, G_{\varphi_0})$ belong to $\mathcal{K}_Y^G(P)$, and $G_{\varphi_t} := (1-t)G_{\varphi_0} + tG_{\varphi_1}$, then the convexity of $\mathcal{E}$ follows from that of $V$ and $W$.

*Remark 4* The convexity of the internal energy (3.33) also holds when considering the monotone discretization of the Monge-Ampère operator introduced by Oberman in [23].

*Remark 5* It seems necessary to consider two notions of push-forward of a given measure $\mu$. Indeed, the internal energy of a measure that is not absolutely continuous is $+\infty$, so that it only makes sense to compute the map $\mathcal{U}$ on the absolutely continuous measure $G^{\mathrm{ac}}_{\varphi\#}\mu$. On the other hand, condition (HE) is not sufficient to make the potential energy functional $\varphi \in \mathcal{K}^G_Y(P) \mapsto \mathcal{E}(G^{\mathrm{ac}}_{\varphi\#}\mu)$ convex. This can be seen on the example given in Figure 1: let $Y = [-1,1]^2$ and $P = \{q, p_\pm\}$ with $q = (2,0)$ and $p_\pm = (0, \pm 1)$. We let $\varphi_t$ be the linear interpolation between $\varphi_0 := \mathbf{1}_{\{q\}}$, and $\varphi_1 = 0$, and We let $\mu = 0.8\delta_q + 0.1\delta_{p_+} + 0.1\delta_{p_-}$. The third column of Figure 1 displays the graph of the second moment of the absolutely continuous push-forward, i.e.

$$t \mapsto \mathcal{E}(G^{\mathrm{ac}}_{\varphi_t\#}\mu), \text{ where } \mathcal{E}(\nu) = \int_{\mathbb{R}^d} \|x\|^2 \,\mathrm{d}\nu(x), \tag{3.35}$$

The graph shows that this function is not convex in $t$, even though $\mathcal{E}$ is convex under generalized displacement since it satisfies McCann's condition (HE).

*Remark 6* The two maps considered in the Theorem can be computed more explicitely:

$$\mathcal{E}(G_{\varphi\#}\mu) = \sum_{p\in P} \mu_p V(G_\varphi(p)) + \sum_{p,q\in P} \mu_p\mu_q W(G_{\varphi(p)}, G_{\varphi(q)}) \tag{3.36}$$

$$\mathcal{U}(G^{\mathrm{ac}}_{\varphi\#}\mu) = \sum_{p\in P} U\left(\frac{\mu_p}{\mathrm{MA}_Y[\varphi](p)}\right) \mathrm{MA}_Y[\varphi](p) \tag{3.37}$$

In particular, when $\mathcal{U}$ is the negative entropy ($U(r) = r\log r$), one has:

$$\mathcal{U}(G^{\mathrm{ac}}_{\varphi\#}\mu) = -\sum_{p\in P} \mu_p \log(\mathrm{MA}_Y[\varphi](p)). \tag{3.38}$$

Consequently the internal energy term plays the role of a barrier for the constraint set $\mathcal{K}_Y(P)$, that is: if $\mathcal{U}(G^{\mathrm{ac}}_{\varphi\#}\mu)$ is finite, then $\varphi$ belongs to the interior of $\mathcal{K}_Y(P)$. The same behavior remains true if the function $U$ has super-linear growth at infinity. This enables us to extend $\mathcal{U}(G^{\mathrm{ac}}_{\varphi\#}\mu)$ to the whole space $\mathbb{R}^P$, by setting it to $+\infty$ when $\mathrm{MA}_Y[\varphi](p) = 0$ for some $p \in P$.

## 4 A convergence theorem

Let $X, Y$ be two convex domains in $\mathbb{R}^d$, and $\mu$ be a probability measure on $X$ which is absolutely continuous with respect to the Lebesgue measure on

$X$, and whose density $\rho$ is bounded from above and below: $\rho \in [r, 1/r]$, with $r > 0$. We are interested in the minimization problem

$$\min_{\nu \in \mathcal{P}(Y)} \mathcal{F}(\nu) = \min\{\mathcal{F}(\nabla\varphi|_{\#}\,\mu); \varphi \in \mathcal{K}_Y\}, \tag{4.39}$$

$$\text{where } \mathcal{F}(\nu) := \mathrm{W}_2^2(\mu, \nu) + \mathcal{E}(\nu) + \mathcal{U}(\nu), \tag{4.40}$$

and where the terms of the functional $\mathcal{F}$ satisfy the following assumptions:

(C1) the energy $\mathcal{E}$ (resp $\mathcal{U}$) is weakly continuous (resp. lower semicontinuous) on $\mathcal{P}(Y)$;

(C2) $\mathcal{U}$ is an internal energy, defined as in (1.10), where the integrand $U : \mathbb{R} \to \mathbb{R}$ is convex, $U(0) = 0$ and $U$ has superlinear growth at infinity i.e. $\lim_{s\to\infty} s^{-1}U(s) = +\infty$.

*Remark 7* Note that the condition (C2) is different from McCann's condition (HU) for the displacement convexity of an internal energy. Among the internal energies that satisfy both McCann's conditions and (C1)–(C2), one can cite those that occur in the gradient flow formulation of the heat equation, where $U(r) = r\log r$, and of the porous medium equation, for which $U(r) = \frac{1}{m-1}r^m$, with $m > 1$. The superlinear growth assumption in (C2) ensures that the internal energy acts as a barrier for the convexity constraint in the approximated problem (4.41).

**Theorem 2 ($\Gamma$-convergence)** *Assume (C1)– (C2). Let $\mu_n$ be a sequence of probability measures supported on finite subsets $P_n \subseteq X$, converging weakly to the probability density $\rho$, and consider the discretized problem*

$$\min_{(\varphi, G_\varphi) \in \mathcal{K}_Y^G(P_n)} \mathrm{W}_2^2(\mu_n, G_{\varphi_n\#}\mu_n) + \mathcal{E}(G_{\varphi_n\#}\mu_n) + \mathcal{U}(G^{\mathrm{ac}}_{\varphi_n\#}\mu_n). \tag{4.41}$$

*Then, there exists a minimizer $\varphi_n$ of (4.41). Moreover, the sequence of absolutely continuous measure $\sigma_n := G^{\mathrm{ac}}_{\varphi_n\#}\mu_n$ is a minimizing sequence for the problem (4.39). If $\mathcal{F}$ has a unique minimizer $\nu$ on $\mathcal{P}(Y)$, then $\sigma_n$ converges weakly to $\nu$.*

**Step 1** *There exists a minimizer to* (4.41).

*Proof* Let $(\varphi_n^k)_k$ be a minimizing sequence (which we can normalize by imposing $\varphi_n^k(p) = 0$ at a fixed $p \in P_n$). Since $Y$ is bounded, we may assume that, up to some not relabeled subsequences $\varphi_n^k$ and $G_{\varphi_n^k}$ converge to some $(\varphi_n, G_{\varphi_n})$. We can also assume that $\hat{\varphi}_n^k := [\varphi_n^k]_{\mathcal{K}_Y}$ converges uniformly to $\hat{\varphi}_n = [\varphi_n]_{\mathcal{K}_Y}$. The convergence in the Wasserstein term and in $\mathcal{E}$ is then obvious, it remains to prove a liminf inequality for the discretized internal energy. First note that thanks to (C2), we also have that there is a $\nu > 0$ such that $\mathrm{MA}_Y[\varphi_n^k](p) \geq \nu$ for every $k$ and every $p \in P_n$. Then observe that the internal energy can be written as

$$\mathcal{U}(G^{\mathrm{ac}}_{\varphi_n^k\#}\mu_n) := \sum_{p\in P_n} F(p, \mathrm{MA}_Y[\varphi_n^k](p)),\ F(p,t) := tU\left(\frac{\mu_p}{t}\right)$$

so that $F(p,.)$ is nonincreasing thanks to (C2). It is then enough to prove that for every $p \in P_n$ one has:

$$\limsup_k \mathrm{MA}_Y[\varphi_n^k](p) = \limsup_k \mathcal{H}^d(\partial\hat{\varphi}_n^k(p)) \leq \mathcal{H}^d(\partial\hat{\varphi}_n(p)) \tag{4.42}$$

but the latter inequality follows at once from Fatou's Lemma and the fact that if $y$ belongs to $\partial\hat{\varphi}_n^k(p)$ for infinitely many $k$ then it also necessarily belongs to $\partial\hat{\varphi}_n(p)$. This proves that $\varphi_n$ solves (4.41).

Let $m$ and $m_n$ be the minima of (4.39) and (4.41) respectively. Our goal now is to show that $\lim_{n\to\infty} m_n = m$. In order to simplify the proof, we will keep the same notation for an absolutely continuous probability measure and its density.

**Step 2** $\liminf_{n\to\infty} m_n \geqslant m$

*Proof* For every $n$, let $\varphi_n \in \mathcal{K}_Y(P_n)$ be a minimizer of the discretized problem (4.41). By compactness of the set $\mathcal{K}_Y$ (up to an additive constant), and taking a subsequence if necessary, we can assume that $\hat{\varphi}_n := [\varphi_n]_{\mathcal{K}_Y}$ converges uniformly to a function $\varphi$ in $\mathcal{K}_Y$. We can also assume that both sequence of measures $\sigma_n := G^{\mathrm{ac}}_{\varphi_n\#}\mu_n$ and $\nu_n := G_{\varphi_n\#}\mu_n$ converge to two measures $\sigma$ and $\nu$ for the Wasserstein distance. The difficulty is to show that these two measures $\nu$ and $\sigma$ must coincide. Indeed, let $\pi_n$ (resp. $\pi'_n$) be optimal transport plans between $\mu_n$ and $\nu_n$ (resp. $\mu_n$ and $\sigma_n$). Taking subsequences if necessary, these optimal transport plans converge to two transport plans $\pi$ (resp. $\pi'$) between $\rho$ and $\nu$ (resp. $\rho$ and $\sigma$) that are supported on the graph of the gradient of $\varphi$. Since the first marginal of $\pi$ and $\pi'$ coincide, one must have $\pi = \pi'$ and therefore $\nu = \sigma$. The result then follows from the weak lower semicontinuity of $\mathcal{U}$, and the continuity of $(\mu,\nu) \mapsto \mathrm{W}_2^2(\mu,\nu) + \mathcal{E}(\nu)$.

We now proceed to the proof that $\limsup_{n\to\infty} m_n \leqslant m$. Our first step is to show that probability measures with a smooth density bounded from below and above are dense in energy. More precisely, we have:

**Step 3** $m = \min_{\varepsilon>0} \min\{\mathcal{F}(\sigma); \sigma \in \mathcal{P}^{\mathrm{ac}}(Y) \cap \mathcal{C}^0(Y), \varepsilon \leqslant \sigma \leqslant 1/\varepsilon\}$

*Proof* Let $\sigma$ be a probability density on $Y$ such that $\mathcal{F}(\sigma) < +\infty$. Then, according to Corollary 1.4.3 in [2], there exists a sequence of probability densities $\sigma_n$ on $Y$ that satisfy the three properties:

(a) For every $n > 0$, $\sigma_n$ is bounded from above and below:

$$0 < \inf_{y\in Y} \sigma_n(y) < \sup_{y\in Y} \sigma_n(y) < +\infty;$$

(b) $\sigma_n$ converges to $\sigma$ in $\mathrm{L}^1(Y)$;
(c) $\mathcal{U}(\sigma_n) \leqslant \mathcal{U}(\sigma)$.

Moreover the proof of Corollary 1.4.3 in [2] can be modified by taking a smooth convolution operator so as to ensure that each $\sigma_n$ is continuous on $Y$. Our task is then to show that

$$\liminf_{n\to\infty} \mathcal{F}(\sigma_n) \leqslant \mathcal{F}(\sigma),$$

where $\mathcal{F}(\sigma) = \mathrm{W}_2^2(\mu,\sigma) + \mathcal{E}(\sigma) + \mathcal{U}(\sigma)$. Thanks to (C1), and thanks to the Wasserstein continuity of the terms $\sigma \mapsto \mathrm{W}_2^2(\mu,\sigma) + \mathcal{E}(\sigma)$, we only need to show that $\sigma_n$ converges to $\sigma$ in the Wasserstein sense. This follows from the easy inequality

$$\mathrm{W}_2^2(\sigma,\sigma') \leqslant \|\sigma - \sigma'\|_{\mathrm{L}^1(Y)} \operatorname{diam}(Y)^2.$$

**Step 4** *Let $\sigma \in \mathcal{P}(Y) \cap \mathcal{C}^0(Y)$, with $\varepsilon \leqslant \sigma \leqslant 1/\varepsilon$. Then, for every $n \geqslant 0$, there exists a convex interpolate $\varphi_n \in \mathcal{K}_Y(P_n)$ such that*

$$\forall p \in P_n,\ \sigma(\partial[\varphi_n]_{\mathcal{K}_Y}(p)) = \mu_n(\{p\}). \tag{4.43}$$

*Proof* By Breniers' theorem, there is a convex potential $\psi_n$ on $Y$ such that $\nabla\psi_n \# \sigma = \mu_n$, so that $\varphi_n := \psi_n^*$ has the desired property.

**Step 5** *Assuming that the functions $\varphi_n$ in $\mathcal{K}_Y(P_n)$ are constructed as above, we can bound the diameter of their subdifferentials:*

$$\lim_{n\to\infty} \max_{p\in P_n} \operatorname{diam}(\partial[\varphi_n]_{\mathcal{K}_Y}(p)) = 0. \tag{4.44}$$

*Proof* Let $\hat{\varphi} \in \mathcal{K}_Y$ be a potential for the quadratic optimal transport problem between $\rho$ and $\sigma$. Let $\hat{\varphi}_n := [\varphi_n]_{\mathcal{K}_Y}$ and $\psi = \hat{\varphi}^*$ and $\psi_n = \hat{\varphi}_n^*$. First, we add a constant to $\varphi$ and $\varphi_n$ such that the integral of $\psi$ and $\psi_n$ over $\sigma$ is zero,

$$\int_Y \psi(y)\sigma(y)\mathrm{d}y = \int_Y \psi_n(y)\sigma(y)\mathrm{d}y = 0.$$

Poincaré's inequality on $Y$ with density $\sigma$ gives us

$$\int_Y |\psi_n(y) - \psi(y)|^2 \sigma(y)\mathrm{d}y \leqslant \mathrm{const}(p, Y, \sigma) \int_Y \|\nabla\psi_n - \nabla\psi(y)\|^2 \sigma(y)\mathrm{d}y,$$

and the weak continuity of optimal transport plans then ensures that the right-hand term converges to zero. Noting that $\psi_n$ and $\psi$ are convex on $Y$ and have a bounded Lipschitz constant, because the gradients $\nabla\psi, \nabla\psi_n$ belong to $X$, this implies that $\psi_n$ converge uniformly to $\psi$. Taking the Legendre transform, this shows that $\hat{\varphi}_n$ converges uniformly to $\hat{\varphi}$ on the compact domain $X$.

We now prove (4.44) by contradiction, and we assume that there exists a positive constant $r$, and a sequence of points $(p_n)$, with $p_n \in P_n$ and such that there exists $y_n, y_n' \in \partial\hat{\varphi}_n(p_n)$ with $\|y_n - y_n'\| \geqslant r$. By compactness, and taking subsequences if necessary, we can assume that $p_n$ converges to a point $p$ in $X$ and that the sequences $(y_n)$ and $(y_n')$ converge to two points $y, y'$ in $Y$ with $\|y - y'\| \geqslant r$. Since the point $y_n$ belongs to $\partial\hat{\varphi}_n(p_n)$, one has:

$$\forall x \in X,\ \hat{\varphi}_n(x) \geqslant \hat{\varphi}_n(p_n) + \langle y_n | x - p\rangle.$$

Taking the limit as $n$ goes to $\infty$, this shows us that $y$ (and similarly $y'$) belongs to $\partial\hat{\varphi}(p)$, so that $\mathrm{diam}(\partial\hat{\varphi}(p)) \geqslant r$. The contradiction then follows from Caffarelli's regularity result [10]: under the assumptions on the supports and on the densities, the map $\hat{\varphi}$ is $\mathcal{C}^{1,\beta}$ up to the boundary of $X$. In particular, the subdifferential of $\hat{\varphi}$ must be a singleton at every point of $X$, thus contradicting the lower bound on its diameter.

**Step 6** *Let $\sigma_n := G^{\mathrm{ac}}_{\varphi_n\#}\mu_n$ and $\nu_n := G_{\varphi_n\#}\mu_n$, where $\varphi_n$ is defined above. Then,*

$$\lim_{n\to\infty} \|\sigma_n - \sigma\|_{\mathrm{L}^\infty(Y)} = 0. \tag{4.45}$$

$$\lim_{n\to\infty} \mathrm{W}_2(\nu_n, \sigma) = 0. \tag{4.46}$$

*Proof* First, note that since $\sigma$ is continuous on a compact set, it is also uniformly continuous. For any $\delta > 0$, there exists $\varepsilon > 0$ such that $\|x - y\| \leqslant \varepsilon$ implies $|\sigma(x) - \sigma(y)| \leqslant \delta$. Using Equation (4.44), for $n$ large enough, the sets $V_p := \partial[\varphi_n]_{\mathcal{K}_Y}(p)$ have diameter bounded by $\varepsilon$ for all point $p$ in $P_n$. By definition, the density $\sigma_n$ is equal to

$$\sigma_n = \sum_{p\in P} \hat{\sigma}_p \chi_{V_p} \text{ with } \hat{\sigma}_p := \frac{1}{\mathcal{H}^d(V_p)} \int_{V_p} \sigma(x)\mathrm{d}x. \tag{4.47}$$

By the uniform continuity property, on every cell $V_p$ one has $|\sigma(x) - \hat{\sigma}_p| \leqslant \delta$, thus proving $\|\sigma_n - \sigma\|_{\mathrm{L}^\infty(Y)} \leqslant \delta$ for $n$ large enough. This implies that $\sigma_n$ converges to $\sigma$ uniformly, and a fortiori that $\lim_{n\to\infty} \mathrm{W}_2(\sigma_n, \sigma) = 0$. Then,

$$\mathrm{W}_2(\nu_n, \sigma) \leqslant \mathrm{W}_2(\nu_n, \sigma_n) + \mathrm{W}_2(\sigma_n, \sigma). \tag{4.48}$$

Moreover, one can bound the Wasserstein distance explicitely between $\sigma_n$ and $\nu_n$ by considering the obvious transport plan on each of the subdifferentials $(\partial[\varphi_n]_{\mathcal{K}_Y}(p))_{p\in P_n}$:

$$\mathrm{W}_2^2(\nu_n, \sigma_n) \leqslant \sum_{p\in P_n} \mathrm{diam}(\partial[\varphi_n]_{\mathcal{K}_Y}(p))^2 \mu_p \leqslant \max_{p\in P_n} \mathrm{diam}(\partial[\varphi_n]_{\mathcal{K}_Y}(p))^2. \tag{4.49}$$

The second statement (4.46) follows from Eqs. (4.48), (4.49) and (4.44).

**Step 7** $\lim_{n\to\infty} \mathrm{W}_2^2(\mu_n, \nu_n) + \mathcal{E}(\nu_n) + \mathcal{U}(\sigma_n) = \mathrm{W}_2^2(\mu, \sigma) + \mathcal{E}(\sigma) + \mathcal{U}(\sigma)$

*Proof* The convergence of the first two terms follows from the Wasserstein continuity of the map $(\mu, \nu) \in \mathcal{P}(Y) \mapsto \mathrm{W}_2^2(\mu, \nu) + \mathcal{E}(\nu)$. In order to deal with the third term, we will assume that $n$ is large enough, so that the densities $\sigma, \sigma_n$ belong to the segment $S_{\varepsilon/2} = [\varepsilon/2, 2/\varepsilon]$. The integrand $U$ of the internal energy is convex on $\mathbb{R}$, and therefore Lipschitz with constant $L$ on $S_{\varepsilon/2}$, so that

$$\begin{aligned}\left|\int_Y U(\sigma_n)\mathrm{d}x - \int_Y U(\sigma(x))\mathrm{d}x\right| &\leqslant \int_Y |U(\sigma_n) - U(\sigma(x))|\,\mathrm{d}x \\ &\leqslant L\,\|\sigma_n - \sigma\|_{\mathrm{L}^\infty(Y)}.\end{aligned}$$

## 5 Numerical results

### 5.1 Computation of the Monge-Ampère operator

In this paragraph we explain how to evaluate the discretized internal energy of $G^{\mathrm{ac}}_{\varphi\#}\mu_P$, where $\varphi$ is a discrete convex function in $\mathcal{K}_Y(P)$, and $Y$ is a polygon in the euclidean plane. Thanks to the equation

$$\mathcal{U}(G^{\mathrm{ac}}_{\varphi\#}\mu_P) = \sum_{p\in P} U\left(\mu_p/\mathrm{MA}_Y[\varphi](p)\right)\mathrm{MA}_Y[\varphi](p), \tag{5.50}$$

one can see that the internal energy and its first and second derivatives can be easily computed if one knows how to evaluate the discrete Monge-Ampère operator and its derivatives with respect to $\varphi$. Our assumptions for performing this computation will be the following:

(G1) the domain $Y$ is a convex polygon and its boundary can be decomposed as a finite union of segments $S = \{s_1, \dots, s_k\}$

(G2) the points in $P$ are in generic position, i.e. (a) there does not exist a triple of collinear points in $P$ and (b) for any pair $p, q$ of distinct points in $P$, there is no segment $s$ in $S$ which is collinear to the bisector of $[pq]$.

The Jacobian matrix of the discrete Monge-Ampère operator is a square matrix denoted $(\mathrm{JMA}_Y[\varphi])_{p,q\in P}$, while its Hessian is a 3-tensor denoted $(\mathrm{HMA}_Y[\varphi])_{p,q,r\in P}$. The entries of this matrix and tensor are given by the formulas

$$\mathrm{JMA}_Y[\varphi]_{pq} := \frac{\partial \mathrm{MA}_Y[\varphi](p)}{\partial \mathbf{1}_q}, \tag{5.51}$$

$$\mathrm{HMA}_Y[\varphi]_{pqr} := \frac{\partial^2 \mathrm{MA}_Y[\varphi](p)}{\partial \mathbf{1}_r \partial \mathbf{1}_q}, \tag{5.52}$$

where $\mathbf{1}_p$ denotes the indicator function of a point $p$ in $P$. The goal of the remaining of this section is to show how the computation of the Jacobian matrix and the Hessian tensor are related to a triangulation which is defined from the Laguerre cells by duality.

*Abstract dual triangulation*

Given any function $\varphi$ on $P$, we introduce a notation for the intersection of the Laguerre cell of $P$ with $Y$, and we extend this notation to handle boundary segments as well. More precisely, we set:

$$\begin{aligned} &\forall p \in P,\ V^\varphi(p) := \mathrm{Lag}^\varphi_P(p) \cap Y, \\ &\forall s \in S,\ V^\varphi(s) := s. \end{aligned} \tag{5.53}$$

We also introduce a notation for the finite intersections of these cells:

$$\forall p_1, \dots, p_s \in P \cup S,\ V^\varphi(p_1 \dots p_s) := V^\varphi(p_1) \cap \dots \cap V^\varphi(p_s) \tag{5.54}$$

The decomposition of $Y$ given by the cells $V^\varphi(p)$ induces an abstract dual triangulation $T^\varphi$ of the set $P \cup S$, whose triangles and edges are characterized by:

(i) a pair $(p, q)$ in $P \cup S$ is an edge of $T^\varphi$ iff $V^\varphi(pq) \neq \emptyset$;
(ii) a triplet $(p, q, r)$ in $P \cup S$ is a *triangle* of $T^\varphi$ iff $V^\varphi(pqr) \neq \emptyset$.

An example of such an abstract dual triangulation is displayed in Figure 2.

The construction of this triangulation can be performed in time $\mathrm{O}(N \log N + k)$, where $N$ is the number of points and $k$ is the number of segments in the boundary of $Y$. The construction works by adapting the regular triangulation of the point set, which is the triangulation obtained when $Y = \mathbb{R}^d$, and for which there exists many algorithms, see e.g. [1].

*Jacobian of the Monge-Ampère operator*

By Lemma 3, for any point $p$ in $P$, the function $\varphi \mapsto \mathrm{MA}_Y[\varphi](p)$ is log-concave on the set $\mathcal{K}_Y(P)$. This function is therefore twice differentiable almost everywhere on the interior of $\mathcal{K}_Y(P)$, using Alexandrov's theorem. The first derivatives of the Monge-Ampère operator is easy to compute, and involves boundary terms: two points $p, q$ in $P$,

$$\mathrm{JMA}_Y[\varphi]_{pq} = \frac{\mathcal{H}^1(V^\varphi(pq))}{\|p - q\|} \text{ if } q \neq p \tag{5.55}$$

$$\mathrm{JMA}_Y[\varphi]_{pq} = - \sum_{\substack{q \in P \\ (qp) \in T^\varphi}} \frac{\mathcal{H}^1(V^\varphi(pq))}{\|p - q\|} \tag{5.56}$$

Note that every non-zero element in the square matrix corresponds to an edge in the dual triangulation $T^\varphi$.

*Hessian of the Monge-Ampère operator*

We will not include the computation of the second order derivatives, but we will sketch how it can be performed using the triangulation $T^\varphi$. First, we remark that thanks to our genericity assumption, for every triangle $pqr$ of $T^\varphi$, the set $V^\varphi(pqr)$ consists of a single point, which we also denote $V^\varphi(pqr)$. For any edge $pq$ in the triangulation $T^\varphi$, where $p, q$ are two points in $P$, the intersection $V^\varphi(pq) = V^\varphi(q) \cap V^\varphi(q)$ is a segment $[x, y]$. The endpoint $x$ of this segment needs to be contained in a third cell $V^\varphi(r)$ for a certain element $r$ of $P \cup S \setminus \{p, q\}$, so that $x = V^\varphi(pqr)$. Similarly, there exists $r'$ in $P \cup S \setminus \{p, q\}$ such that $y = V^\varphi(pqr')$. One can therefore rewrite the length of $V^\varphi(pq)$ as

$$\mathcal{H}^1(V^\varphi(pq)) = \|V^\varphi(pqr) - V^\varphi(pqr')\| . \tag{5.57}$$

The expression of the Hessian can be deduced from Equations (5.55)–(5.56) and (5.57), and from an explicit computation for the point $V^\varphi(pqr)$. Moreover, to each nonzero element of the Hessian one can associate a point, an edge or a triangle in the triangulation $T^\varphi$. More precisely:

$$\begin{aligned} \mathrm{HMA}_Y[\varphi]_{pqr} \neq 0 \Longrightarrow p = q = r \text{ or } & (p = q \text{ and } (pr) \text{ is an edge of } T^\varphi) \\ & \text{or } (pqr) \text{ is a triangle of } T^\varphi. \end{aligned}$$

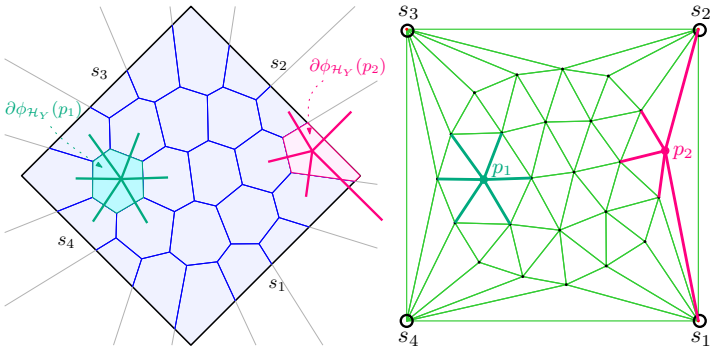


**Fig. 2** On the left, the intersection of power cells with a convex polygon (in red), and on the right, the dual triangulation.

In particular, the total number of non-zero elements of the tensor $\mathrm{HMA}_Y[\varphi]$ is at most proportional to the number $|P|$ of points plus the number $|S|$ of segments.

### 5.2 Non-linear diffusion on point clouds

The first application is non-linear diffusion in a bounded convex domain $X$ in the plane. We are interested in the following PDE, where the parameter $m$ is chosen in $[1-1/d, +\infty)$. A numerical application is displayed on Figure 3 .

$$\begin{cases} \dfrac{\partial \rho}{\partial t} = \Delta \rho^m & \text{on } X \\ \nabla \rho \perp \mathbf{n}_X & \text{on } \partial X \end{cases} \tag{5.58}$$

When $m = 1$, this PDE is the classical heat equation with Neumann boundary conditions. When $m < 1$, this PDE provides a model of fast diffusion, while for $m > 1$ it is a model for the evolution of a gas in a porous medium. Otto [25] reinterpreted this PDE as a gradient flow in the Wasserstein space for the internal energy

$$\mathcal{U}_m(\mu) = \begin{cases} \displaystyle\int_{\mathbb{R}^d} U_m(\rho(x))\mathrm{d}x \text{ if } \mu \ll \mathcal{H}^d, \rho := \frac{\mathrm{d}\mu}{\mathrm{d}\mathcal{H}^d}, \\ +\infty \text{ if not.} \end{cases} \tag{5.59}$$

where $U_m(r) = \frac{r^m(x)}{m-1}$ when $m \neq 1$ and $U_1(r) = r \log r$. A time-discretization of this gradient-flow model can be defined using the Jordan-Kinderlehrer-Otto scheme: given a timestep $\tau > 0$ and a probability measure $\mu_0$ supported on $X$, one defines a sequence of probability measures $(\mu_k)_{k \geqslant 1}$ recursively

$$\mu_{k+1} = \arg\min_{\mu \in \mathcal{P}(X)} \mathrm{W}_2^2(\mu_k, \mu) + \mathcal{U}_m(\mu) \tag{5.60}$$

The energies involved in this optimization problem satisfy McCann's assumption for displacement convexity, and our discrete framework is therefore able to provide a discretization in space of Equation (5.60) as a convex optimization problem. We use this discretization in order to construct the non-linear diffusion for a finite point set $P_0$ contained in the convex domain $X$. Note that for this experiment, we do not use the formulation involving the space of convex interpolates with gradient, $\mathcal{K}_X^G(P)$. For every function $\varphi$ in $\mathcal{K}_Y(P)$, and every point $p$ in $P$, we select explicitely a subgradient in the subdifferential $\partial\varphi_{\mathcal{K}_Y}(p)$ by taking its Steiner point [26].

We start with $\mu_0 = \sum_{p\in P} \delta_p / |P|$ the uniform measure on the set $P_0$, and we define recursively

$$\begin{cases} \varphi_k = \arg\min \left\{ \dfrac{1}{2\tau} \mathrm{W}_2^2(\mu_k, G_{\varphi\#}\mu_k) + \mathcal{U}(G_{\varphi\#}^{\mathrm{ac}}\mu_k); \varphi \in \mathcal{K}_X(P_k) \right\} \\ \mu_{k+1} = G_{\varphi_k\#}\mu_k, \\ P_{k+1} = \mathrm{spt}(\mu_{k+1}) \end{cases}, \quad (5.61)$$

where $G_\varphi(p)$ is the Steiner point of $\partial\varphi_{\mathcal{K}_Y}(p)$. This minimization problem is solved using a second-order Newton method. Note that, as mentioned in the remark following Theorem 1, the internal energy plays the role of a barrier for the convexity of the discrete function $\varphi$. When second-order methods fail, one could also resort to more robust first-order methods for the resolution of the optimization problem, using for instance a projected gradient algorithm.

### 5.3 Crowd-motion with congestion

As a second application, we consider the model of crowd motion with congestion introduced by Maury, Roudneff-Chupin and Santambrogio [19]. The crowd is represented by a probability density $\mu_0$ on a convex compact subset with nonempty interior $X$ of $\mathbb{R}^2$, which is bounded by a certain constant, which we assume normalized to one (so that we also naturally assume that $\mathcal{H}^d(X) > 1$). One is also given a potential $V : X \to \mathbb{R}$, which we assume to be $\lambda$-convex, i.e. $V(.) + \lambda \|.\|^2$ is convex. The evolution of the probability density describing the crowd is induced by the gradient flow of the potential energy

$$\mathcal{E}(\mu) = \int_{\mathbb{R}^d} V(x) \mathrm{d}\mu(x), \quad (5.62)$$

in the Wasserstein space, under the additional constraint that the density needs to remain bounded by one. We rely again on time-discretization of this gradient flow using the Jordan-Kinderlehrer-Otto scheme. This gives us the following formulation:

$$\mu_{k+1} = \arg\min_{\mu\in\mathcal{P}(X)} \frac{1}{2\tau} \mathrm{W}_2^2(\mu_k, \mu) + \mathcal{E}(\mu) + \mathcal{U}(\mu), \quad (5.63)$$

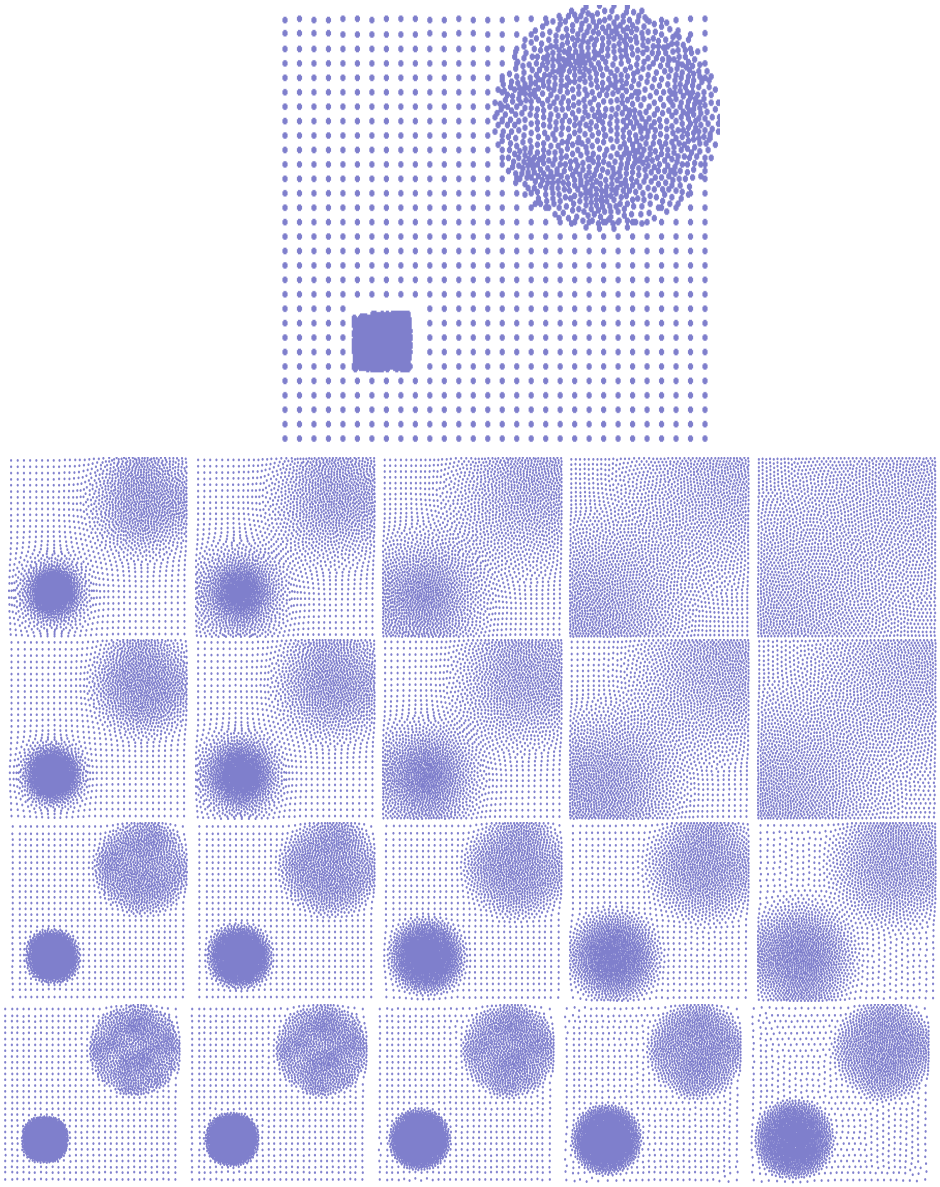

**Fig. 3** (Top) The original point set $P_0$ in the square domain $X = [-2, 2]^2$ contains 2 900 points. (Bottom rows) The evolution of the point cloud $P_k$ at timesteps $k = 5, 10, 20, 40, 80$ with $\tau = 0.01$ is defined using Eq. (5.61) for various values of the exponent $m$. From top to bottom, $m$ takes increasing values in $\{0.6, 1, 2, 3\}$. Note that the case $m = 0.6$ is outside of the scope of the $\Gamma$-convergence theorem

where $\mathcal{U}$ is the indicatrix function of the probability measures whose density is bounded by one:

$$\mathcal{U}(\mu) = \begin{cases} 0 \text{ if } \mu \ll \mathcal{H}^d \text{ and } \dfrac{\mathrm{d}\mu}{\mathrm{d}\mathcal{H}^d} \leqslant 1 \\ +\infty \text{ if not.} \end{cases} \tag{5.64}$$

In order to perform numerical simulations, we replace this indicatrix function by a smooth approximation.

$$\mathcal{U}_\alpha(\mu) = \begin{cases} \displaystyle\int_{\mathbb{R}^d} \rho^\alpha(x)(-\log(1-\rho(x)^{1/d})\mathrm{d}x \text{ if } \mu \ll \mathcal{H}^d \text{ and } \rho := \frac{\mathrm{d}\mu}{\mathrm{d}\mathcal{H}^d}, \\ +\infty \text{ if not.} \end{cases} \tag{5.65}$$

Note that if $\mathcal{U}_\alpha(\mu)$ is finite, then the density of $\mu$ is bounded by one almost everywhere. Moreover, we have the following convexity and $\Gamma$-convergence results:

**Proposition 1** *(i) The energy $\mathcal{U}_\alpha$ is convex under general displacement.*
*(ii) $\mathcal{U}_\alpha$ $\Gamma$-converges (for the weak convergence of measures on $X$) to $\mathcal{U}$ as $\alpha$ tends to $+\infty$ ;*
*(iii) $\beta\mathcal{U}_1$ $\Gamma$-converges to $\mathcal{U}$ as $\beta$ tends to $0$.*

*Proof* The proof of (i) uses McCann's theorem: one only needs $r^d U(r^{-d})$ to be convex non-increasing and $U(0) = 0$, which follows from a simple computation. (ii) The proof of the $\Gamma$-liminf inequality is obvious since $\mathcal{U}_\alpha \geq \mathcal{U}$ and $\mathcal{U}$ is lower semicontinuous. As for the $\Gamma$-limsup inequality, we proceed as follows: we first fix $\mu \in \mathcal{P}(X)$ such that $\mathcal{U}(\mu) = 0$ (otherwise, there is nothing to prove). Let us then fix a set $A \subset X$ such that $\mathcal{H}^d(A) > 1$ and let $m$ be the uniform probability measure on $A$. For $\varepsilon \in (0,1)$, let us then define $\mu_\varepsilon := (1-\varepsilon)\mu + \varepsilon m$ so that $\mu_\varepsilon$ has a density bounded by $1 - C\varepsilon$ where $C := 1 - \frac{1}{\mathcal{H}^d(A)} > 0$. Letting $\alpha \to \infty$ and setting $\varepsilon_\alpha \sim \alpha^{-1/2}$, one directly checks that $\mathcal{U}_\alpha(\mu_{\varepsilon_\alpha}) = O(e^{-\alpha^{1/2}}\log(\alpha))$. Therefore $\limsup_\alpha \mathcal{U}_\alpha(\mu_{\varepsilon_\alpha}) = 0 = \mathcal{U}(\mu)$ which proves the $\Gamma$-limsup inequality. For (iii), the proof is similar, choosing $\varepsilon_\beta \sim e^{-\beta^{-1/2}}$ as $\beta \to 0$ for the $\Gamma$-limsup inequality.

*Numerical result*

Figure 4 displays a numerical application, where we compute the Wasserstein gradient flow of a probability density whose energy is given by

$$\mathcal{F}(\rho) = \int_X V(x)\rho(x)\mathrm{d}x + \alpha\mathcal{U}_1(\rho), \tag{5.66}$$
$$\text{where } X = [-2,2]^2, \quad \text{and } V(x) = \|x-(2,0)\|^2 + 5\exp(-5\|x\|^2/2).$$

Note that the chosen potential is semi-convex. We track the evolution of a probability density on a fixed grid, which allows us to use a simple finite

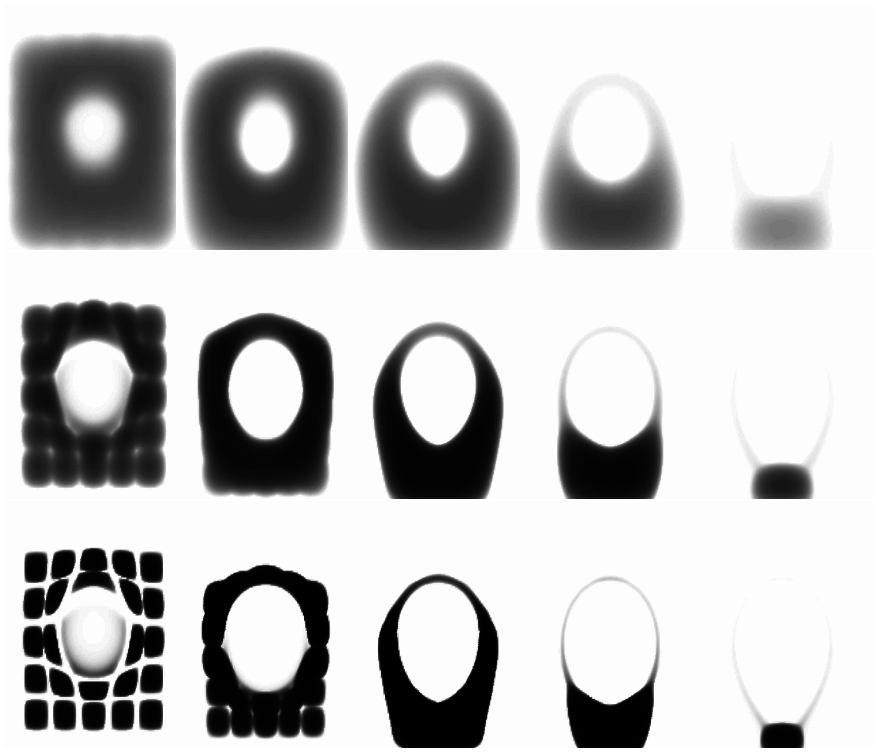

**Fig. 4** Simulation of crowd motion under congestion using the gradient flow formulation. The congestion term is given by $\beta\mathcal{U}_1$ (see Equation (5.65)), for various values of $\beta$. From top to bottom, $\beta$ is set to $10^{-k}$ for $0 \leqslant k \leqslant 2$.

difference scheme to evaluate the gradient of the transport potential. From one timestep to another, the mass of the absolutely continuous pushforward of the minimizer is redistributed on the fixed grid.

*Acknowledgements.* The authors gratefully acknowledge the support of the French ANR, through the projects ISOTACE (ANR-12-MONU-0013), OPTIFORM (ANR-12-BS01-0007) and TOMMI (ANR-11-BSO1-014-01).

## References

1. Cgal, Computational Geometry Algorithms Library. Http://www.cgal.org
2. Agueh, M.: Existence of solutions to degenerate parabolic equations via the Monge-Kantorovich theory. Ph.D. thesis, Georgia Institute of Technology, USA (2002)
3. Agueh, M.: Existence of solutions to degenerate parabolic equations via the Monge-Kantorovich theory. Advances in Differential Equations **10**(3), 309–360 (2005)
4. Agueh, M., Bowles, M.: One-dimensional numerical algorithms for gradient flows in the p-Wasserstein spaces. Acta applicandae mathematicae **125**(1), 121–134 (2013)
5. Ambrosio, L., Gigli, N., Savaré, G.: Gradient flows: in metric spaces and in the space of probability measures. Lectures in Mathematics ETH Zürich (2005)

6. Blanchet, A., Calvez, V., Carrillo, J.A.: Convergence of the mass-transport steepest descent scheme for the subcritical Patlak-Keller-Segel model. SIAM Journal on Numerical Analysis **46**(2), 691–721 (2008)
7. Blanchet, A., Carlier, G.: Optimal transport and Cournot-Nash equilibria. arXiv preprint arXiv:1206.6571 (2012)
8. Brenier, Y.: Polar factorization and monotone rearrangement of vector-valued functions. Communications on pure and applied mathematics **44**(4), 375–417 (1991)
9. Burger, M., Carrillo, J.A., Wolfram, M.T., et al.: A mixed finite element method for nonlinear diffusion equations. Kinetic and Related Models **3**(1), 59–83 (2010)
10. Caffarelli, L.A.: Boundary regularity of maps with convex potentials. Communications on pure and applied mathematics **45**(9), 1141–1151 (1992)
11. Carlier, G., Lachand-Robert, T., Maury, B.: A numerical approach to variational problems subject to convexity constraint. Numerische Mathematik **88**(2), 299–318 (2001)
12. Carrillo, J.A., Moll, J.S.: Numerical simulation of diffusive and aggregation phenomena in nonlinear continuity equations by evolving diffeomorphisms. SIAM Journal on Scientific Computing **31**(6), 4305–4329 (2009)
13. Choné, P., Le Meur, H.V.: Non-convergence result for conformal approximation of variational problems subject to a convexity constraint. Numer. Funct. Anal. Optim. **5-6**(22), 529–547 (2001)
14. Ekeland, I., Moreno-Bromberg, S.: An algorithm for computing solutions of variational problems with global convexity constraints. Numerische Mathematik **115**(1), 45–69 (2010)
15. Gutiérrez, C.E.: The Monge-Ampère equation, vol. 44. Birkhauser (2001)
16. Jordan, R., Kinderlehrer, D., Otto, F.: The variational formulation of the Fokker–Planck equation. SIAM journal on mathematical analysis **29**(1), 1–17 (1998)
17. Kinderlehrer, D., Walkington, N.J.: Approximation of parabolic equations using the Wasserstein metric. ESAIM: Mathematical Modelling and Numerical Analysis **33**(04), 837–852 (1999)
18. Lachand-Robert, T., Oudet, É.: Minimizing within convex bodies using a convex hull method. SIAM Journal on Optimization **16**(2), 368–379 (2005)
19. Maury, B., Roudneff-Chupin, A., Santambrogio, F.: A macroscopic crowd motion model of gradient flow type. Mathematical Models and Methods in Applied Sciences **20**(10), 1787–1821 (2010)
20. McCann, R.J.: A convexity principle for interacting gases. Advances in Mathematics **128**(1), 153–179 (1997)
21. Mérigot, Q., Oudet, E.: Handling convexity-like constraints in variational problems. arXiv preprint arXiv:1403.2340 (2014)
22. Mirebeau, J.M.: Adaptive, anisotropic and hierarchical cones of discrete convex functions. arXiv preprint arXiv:1402.1561 (2014)
23. Oberman, A.M.: Wide stencil finite difference schemes for the elliptic Monge-Ampère equation and functions of the eigenvalues of the hessian. Discrete Contin. Dyn. Syst. Ser. B **10**(1), 221–238 (2008)
24. Oberman, A.M.: A numerical method for variational problems with convexity constraints. SIAM Journal on Scientific Computing **35**(1), A378–A396 (2013)
25. Otto, F.: The geometry of dissipative evolution equations: the porous medium equation. Communications in partial differential equations **26**(1-2), 101–174 (2001)
26. Schneider, R.: Convex bodies: the Brunn-Minkowski theory, vol. 44. Cambridge University Press (1993)
27. Trudinger, N.S., Wang, X.J.: The affine Plateau problem. J. Amer. Math. Soc. **18**(2), 253–289 (2005)
28. Zhou, B.: The first boundary value problem for Abreu's equation. Int. Math. Res. Not. IMRN (7), 1439–1484 (2012)