



Quantitative Convergence of a Discretization of Dynamic Optimal Transport Using the Dual Formulation

Sadashige Ishida¹ · Hugo Lavenant²

Received: 19 December 2023 / Revised: 26 July 2024 / Accepted: 26 August 2024 /

Published online: 11 November 2024

© The Author(s) 2024

Abstract

We present a discretization of the dynamic optimal transport problem for which we can obtain the convergence rate for the value of the transport cost to its continuous value when the temporal and spatial stepsize vanish. This convergence result does not require any regularity assumption on the measures, though experiments suggest that the rate is not sharp. Via an analysis of the duality gap we also obtain the convergence rates for the gradient of the optimal potentials and the velocity field under mild regularity assumptions. To obtain such rates, we discretize the dual formulation of the dynamic optimal transport problem and use the mature literature related to the error due to discretizing the Hamilton–Jacobi equation.

Keywords Optimal transport · Hamilton–Jacobi equation · Convex optimization.

Mathematics Subject Classification Primary 49Q22 · Secondary 65K10 · 49L12.

1 Introduction

The dynamic optimal transport problem and its discretization

In this work, we are interested in the dynamic optimal transport problem. Given two probability measures μ, ν over a spatial domain Ω , it reads

$$\inf_{(\rho_t, v_t)} \int_0^1 \int_{\Omega} L(v_t) \, d\rho_t \, dt, \quad (1.1)$$

Communicated by Lénaïc Chizat.

✉ Sadashige Ishida
sadashige.ishida@ist.ac.at

Hugo Lavenant
hugo.lavenant@unibocconi.it

¹ Institute of Science and Technology Austria, Klosterneuburg, Austria

² Department of Decision Sciences and BIDSa, Bocconi University, Milan, Italy

where $L : \mathbb{R}^d \rightarrow \mathbb{R}$ is a given convex function and the infimum is taken over all pairs (ρ_t, v_t) of a time-dependent probability distribution and velocity field which are solutions of

$$\begin{cases} \partial_t \rho_t + \nabla \cdot (\rho_t v_t) = 0, \\ \rho_0 = \mu, \quad \rho_1 = \nu, \end{cases} \quad (1.2)$$

that is of the continuity equation with temporal boundary conditions μ and ν . Here ρ_t, v_t are indexed by a temporal variable $t \in [0, 1]$ and $\nabla \cdot$ stands for the divergence with respect to the spatial variable. The interpretation is that (ρ_t) must join μ and ν while being transported by the flow of (v_t) , at a minimal cost. Originally introduced by Benamou and Brenier in [5] for numerical purposes as it is linked to the (static) optimal transport problem with cost $c(x, y) = L(y - x)$, this formulation turned out to be very fruitful. From a theoretical point of view, it is a robust formulation which enables to extend and generalize the optimal transport problem: it is used for optimal transport on graphs [36], for unbalanced optimal transport [19, 30, 35], for optimal transport of matrix-valued measures [16] to cite a few extensions. From the numerical point of view, in addition to being one of the first methods proposed to solve the optimal transport problem in dimension more than one, it can be adapted to a great variety of related problems: Wasserstein gradient flow [7, 17], mean field games [6, 8], and trajectory inference [48] to mention a few.

The usual road to solve (1.1) with the constraints (1.2) is to first rewrite it as a convex problem by using the momentum $m_t = \rho_t v_t$ as an unknown rather than v_t . In this case both m_t and ρ_t are measures, and v_t is recovered as the Radon-Nikodym derivative $dm_t/d\rho_t$ of m_t with respect to ρ_t . This leads to

$$\inf_{(\rho_t, m_t)} \int_0^1 \int_{\Omega} L \left(\frac{dm_t}{d\rho_t} \right) d\rho_t dt, \quad \text{with constraints} \quad \begin{cases} \partial_t \rho_t + \nabla \cdot m_t = 0, \\ \rho_0 = \mu, \quad \rho_1 = \nu. \end{cases} \quad (1.3)$$

The constraint $\partial_t \rho_t + \nabla \cdot m_t = 0$ is now linear, and the functional to be minimized is convex as $(x, y) \mapsto L(x/y)y$ is a convex function both in x and y when extended to $+\infty$ for $y \leq 0$, except at $(0, 0)$ where it is 0. Then one proposes a finite dimensional version of the problem (1.3) where both the time and space variables are discretized, and solves the resulting finite dimensional convex problem with standard methods in non-smooth convex optimization. We refer to [17, 32, 38, 39, 42] for instantiations of this approach with plain optimal transport. This comes with two challenges from the viewpoint of numerical analysis:

1. Guarantee (quantitatively) the convergence of the convex optimization solver used to solve the discretized problem.
2. Guarantee (quantitatively) the convergence of the value and solutions to the discretized problem to the original one (1.3) when the temporal and spatial stepsizes vanish.

The first point is a question of convex optimization which is quite well understood and that we will not address referring to [29, 42]. *Our main concern is rather the second*

question. The difficulty lies in the roughness of the functional to optimize: $(x, y) \mapsto L(x/y)y$ is only lower semi-continuous and discontinuous at $(0, 0)$. Moreover a priori the data μ, ν could be allowed to be arbitrary probability distributions (even Dirac masses), and there is no regularizing effect, so that ρ_t could be a Dirac mass for every $t \in [0, 1]$. Thus one really has to face the discontinuity of the functional to optimize. Nevertheless, numerical examples in the aforementioned works suggest that convergence holds even when the measures μ, ν are quite rough.

In [31] building on the previous works [17, 24, 28], the second author gave an answer to the second question by providing sufficient conditions for any discretization of the dynamic optimal transport problem to indeed converge to the original problem when the temporal and spatial stepsizes vanish. This framework was later used in [39] and extended to matrix-valued optimal transport in [33, 34]. However the arguments were inspired by the theory of Γ -convergence and were *not* quantitative, that is, they did not come with a convergence rate.

The goal of this article is to provide *quantitative* convergence rates of a discretized version of (1.3) to the original problem, as the temporal and spatial stepsize vanish. There are few results already available: the works [17] and [38] show that their respective methods reach first order of convergence. Specifically, for h the (common) temporal and spatial stepsize, the error in the value of problem (1.3), that is, the transport cost, is of order h . The article [38] also extends this convergence to the solutions of the problem via an analysis of the dual gap. However, in both cases this rate is only available when the data μ, ν have smooth densities which are bounded from below, and moreover the solution ρ to the optimal transport problem needs to have a smooth density bounded from below. The latter assumption (positivity of the solution ρ everywhere) is very strong, not directly implied by μ, ν smooth and bounded from below [47]. Moreover, as we said above, in practice the discretizations behave well even when μ, ν or ρ are not bounded from below.

In this work we propose a new discretization of the dynamic optimal transport problem (see Sect. 3) for which we are able to quantify the convergence, at least of the value of the transport cost. With h the (common) temporal and spatial stepsize the error is of order $\sqrt{h}$ but our result does not require *any* assumption on the probability distributions μ, ν , besides that they have bounded support (see Theorem 4.1).

Let us already emphasize that this result comes with two important limitations. First we are not able to show that the rate improves if μ, ν have a smooth density, and the numerical experiments we conducted suggest that our rate is not sharp. To prove better convergence rates for more regular inputs, we would need a refinement of the previous result [23], which we largely rely on (see Remark 5.6).

The second point is about the efficiency of numerical computation. Our discretization is restricted to periodic boundary conditions although we can treat the transport between μ, ν on a bounded domain without any theoretical limitations by taking the entire periodic domain large enough. A full extension to bounded domains for more numerical efficiency may be potentially done, but is out of the scope of the present paper (see Remark 4.4).

Intuition of our proposed discretization

Our strategy to get such rates is the following. Rather than solving (1.3), we look at the dual problem which is known to be

$$\sup_{\varphi} \int_{\Omega} \varphi(1, \cdot) \, d\nu - \int_{\Omega} \varphi(0, \cdot) \, d\mu \quad \text{with constraint} \quad \partial_t \varphi + H(\nabla \varphi) \leq 0, \quad (1.4)$$

where $\varphi = \varphi(t, x)$ for $t \in [0, 1]$ and $x \in \Omega$ is a time-space dependent scalar function which is subsolution of the *Hamilton–Jacobi* equation $\partial_t \varphi + H(\nabla \varphi) = 0$, being H the Legendre transform of L .

Discretizations of the Hamilton–Jacobi equation in a finite difference fashion have been studied in the context of viscosity solutions starting with the seminal work of Crandall and Lions [23] and the rate of convergence is proved $\sqrt{h}$ under fairly general assumptions. This holds specifically for the initial value problem

$$\begin{cases} \partial_t \varphi + H(\nabla \varphi) = 0, \\ \varphi(0, \cdot) \text{ given,} \end{cases}$$

for the Hamilton–Jacobi equation, importantly under an a priori Lipschitz bound on $\varphi(0, \cdot)$. In the context of optimal transport, an optimal solution φ to the dual problem (1.4) is actually known to be Lipschitz in space, with Lipschitz constant depending on the diameter of Ω but independent of μ, ν (see Proposition 2.2). In our analysis, we also need to ensure that an optimal solution to the discretized version of (1.4) is also Lipschitz. As it cannot be guaranteed a priori, we add Lipschitz continuity at $t = 0$ as an additional constraint in our discretized problem; see Definition 3.2 and in particular the constraint (3.4).

In short we propose a discretization of the dual problem (1.4) where we use a finite difference discretization of the Hamilton–Jacobi equation for which convergence rates are already studied, adding a Lipschitz constraint to $\varphi(0, \cdot)$ in the discrete dual problem.

A brief comment on rates for other numerical methods

Dynamic optimal transport is not the only way to solve the optimal transport problem, we refer to [43] and references therein for a comprehensive introduction. When faced with the linear programming formulation and its entropic regularization, the standard setting is to assume that measures are approximated via i.i.d. samples and rates should be understood in a statistical setting (that is, written in probability or in expectation as a function of the sample size) rather than a numerical analysis one (that is, written as a deterministic function of the stepsizes). We refer to [40] and [37] as well as references therein for results in this direction. In semi-discrete optimal transport, that is, when only one of the two measures is discrete, rates have been investigated both in a statistical setting [2], but also in a more standard setting of numerical analysis, related to the stability of the Monge–Ampère equation [12]. As far as PDEs methods are concerned, in addition to dynamic optimal transport, one could also solve the Monge–Ampère equation to compute a transport map [11]. The convergence when

the resolution of discretization is reduced has been established in the context of viscosity solutions [10]; see also [9, 13]. Note that these convergence results concern a different method, so comparison with the present work is hard to do. We only point out that the results are not quantitative and typically need to assume some regularity of the input measures μ, ν (like absolute continuity) in order to be able to write the Monge–Ampère equation in the first place.

Organization

We first define properly the dynamic optimal transport problem, and we present the Hamilton–Jacobi equation together with its finite difference discretization: this is a well-understood theory that we summarize in Sect. 2. We move to our proposed discretization in Sect. 3. Our main result, the quantitative convergence rate of the optimal transport cost, is presented in Sect. 4. With a standard analysis of the duality gap, we prove, under an additional regularity assumption, that we obtain quantitative convergence of some variables (the velocity field v and $\nabla\varphi$) from their discrete to their continuous counterpart in Sect. 5. We numerically illustrate our results on simple one-dimensional test cases in Sect. 6: this shows that our rates are likely not sharp.

2 Settings and Preliminaries

In this section, we present our setting and review the previous works that are going to be ingredients of our discretization of dynamic optimal transport.

Assumptions We assume the following in the rest of the article.

- We restrict to $\Omega := \mathbb{R}^d / (D\mathbb{Z}^d)$ to be a d -dimensional torus with diameter

$$\text{diam}(\Omega) = \frac{\sqrt{d}}{2} D.$$

- The measures μ, ν are Borel probability measures on Ω and we do not make any assumption such as absolute continuity unless otherwise stated.
- We take $L : \mathbb{R}^d \rightarrow [0, +\infty)$, the “Lagrangian” which is a non-negative, strictly convex and superlinear function.

As for the first point about the domain, we could assume $D = 1$, or $\text{diam}(\Omega) = 1$ without loss of generality. However we prefer to keep it that way to emphasize how some constants depend on $\text{diam}(\Omega)$. The last condition for the Lagrangian includes the most common choice $L(v) = |v|^p/p$ for $p \in (1, +\infty)$.

2.1 Dynamic (and Static) Optimal Transport

We briefly recall some ingredients of the dynamic optimal transport problem. We refer to the textbooks [43, 45, 49] for additional details.

We directly move to the convex formulation already mentioned in (1.3) where we now define each term. The infimum will run over all the pairs $(\rho_t, m_t)_{t \in [0,1]}$ of

probability distributions $\rho_t \in \mathcal{P}(\Omega)$ and vector-valued measure $m_t \in \mathcal{M}(\Omega)^d$ such that $t \mapsto \rho_t$ is continuous for the weak topology with $\rho_0 = \mu$ and $\rho_1 = \nu$. The continuity equation

$$\partial_t \rho_t + \nabla \cdot m_t = 0,$$

is meant in the sense of distributions over the space $[0, 1] \times \Omega$. For the functional to be optimized, we need a bit more of notations. The so-called Benamou-Brenier functional is defined as a functional for measures $\rho \in \mathcal{M}(\Omega)$ and $m \in \mathcal{M}^d(\Omega)$ by,

$$\mathcal{B}(\rho, m) := \sup_{(a,b)} \int_{\Omega} a \, d\rho + \int_{\Omega} b \cdot dm$$

where the pair (a, b) runs through $C_b(\Omega; K)$ the space of continuous bounded functions valued in the convex domain K given by

$$K := \left\{ (s, w) \in \mathbb{R} \times \mathbb{R}^d \text{ such that } s + H(w) \leq 0 \right\}.$$

Here the function $H : \mathbb{R}^d \rightarrow \mathbb{R}$ is called the Hamiltonian, given as the Legendre transform of L by $H(w) = \sup_v \langle v, w \rangle - L(v)$. With a slight variation of [45, Proposition 7.7], it can be proved that if $\mathcal{B}(\rho, m) < +\infty$ then ρ is a non-negative measure and the measure m is absolutely continuous with respect to ρ , and in this case

$$\mathcal{B}(\rho, m) = \int_{\Omega} L(v) \, d\rho,$$

being $v \in L^1(\Omega, \rho)^d$ the Radon-Nikodym density of m with respect to ρ . With these notations the problem (1.3) now reads

$$\mathcal{K}(\mu, \nu) = \inf_{(\rho_t, m_t)} \int_0^1 \mathcal{B}(\rho_t, m_t) \, dt \quad \text{such that} \quad \begin{cases} \partial_t \rho_t + \nabla \cdot m_t = 0 \text{ weakly,} \\ \rho_0 = \mu, \quad \rho_1 = \nu. \end{cases} \quad (2.1)$$

The existence of a minimizer to the problem can be shown by the direct method of calculus of variations (see e.g. [19, Sect. 2] for a proof in a more general context), or alternatively by building it via an optimal solution for the static primal problem (2.4) introduced below as in (2.5). We do not include the proof as it is not our main concern.

Theorem 2.1 (*Existence of a solution in the primal problem*) *Under our assumptions the infimum in (2.1) is attained.*

The question of uniqueness is more subtle and one would go to the static problem introduced below to analyze it. The outcome is that the minimum may not be unique, but it is so if at least one of the two measures μ, ν is absolutely continuous with respect to the Lebesgue measure [45, Theorem 1.25].

Being a convex optimization problem under constraint, the dynamic optimal transport problem has a dual form. It reads

$$\mathcal{K}(\mu, \nu) = \sup_{\varphi} \int_{\Omega} \varphi(1, \cdot) \, d\nu - \int_{\Omega} \varphi(0, \cdot) \, d\mu \quad \text{such that} \quad \partial_t \varphi + H(\nabla \varphi) \leq 0 \quad (2.2)$$

where φ runs over Lipschitz functions defined over $[0, 1] \times \Omega$, and the Hamilton–Jacobi constraint $\partial_t \varphi + H(\nabla \varphi) \leq 0$ means that φ is a viscosity subsolution of $\partial_t \varphi + H(\nabla \varphi) = 0$ as we explain below in Sect. 2.2. Here φ should be interpreted as a Lagrange multiplier for the continuity equation. The key result is that, not only a solution to the dual problem exists, but it has some Lipschitz regularity.

As L is convex, it is Lipschitz on bounded domains and we denote by $\text{Lip}(L, B_r)$ its Lipschitz constant on the ball B_r centered at 0 and of radius r . On the other hand, $\text{Lip}(\psi)$ for a real-valued function ψ stands for its Lipschitz constant on its whole domain of definition.

Proposition 2.2 (*Existence of a solution in the dual problem*) *Under our assumptions there exists an optimal potential $\bar{\varphi}$ in the dual dynamic transport problem satisfying*

$$\text{Lip}(\bar{\varphi}(t, \cdot)) \leq \text{Lip}(L, B_{\text{diam}(\Omega)}), \quad \forall t \in [0, 1]. \quad (2.3)$$

We delay the proof of this result until the next section as we need additional preliminary results, including the static formulation.

The *dynamic* formulation of optimal transport is to be contrasted with its *static* one, which we introduce for the sake of completeness, and with which the reader may be more familiar. The cost function $c : \Omega \times \Omega \rightarrow [0, +\infty)$ associated to our Lagrangian L is

$$c(x, y) := \inf_k \{L(y - x - k), k \in D\mathbb{Z}^d\}$$

as we are on the torus. The transport cost (2.1) is actually equal to:

$$\mathcal{K}(\mu, \nu) := \inf_{\pi \in \text{ADM}(\mu, \nu)} \int_{\Omega \times \Omega} c(x, y) \, d\pi(x, y) \quad (2.4)$$

where $\text{ADM}(\mu, \nu)$ is the set of the probability measures on $\Omega \times \Omega$ whose marginals are μ and ν . Given a solution π to the static optimal transport problem, one can construct a solution to the dynamic optimal transport problem as follows: for any $t \in [0, 1]$, being $\Gamma_t : (x, y) \in \Omega \rightarrow (1-t)x + ty \in \Omega$ (where the addition is understood modulo $D\mathbb{Z}^d$), one sets for any Borel set A ,

$$\rho_t(A) = \pi(\Gamma_t^{-1}(A)), \quad m_t(A) = \int_{\Gamma_t^{-1}(A)} (y - x) \, d\pi(x, y). \quad (2.5)$$

The intuition is that, once its initial and final position are chosen via the coupling π , each particle travels at constant speed on a straight line. Finally, we introduce the dual formulation of the static optimal transport problem. It reads

$$\mathcal{K}(\mu, \nu) = \sup_{\psi} \int_{\Omega} \psi^c \, d\nu + \int_{\Omega} \psi \, d\mu \quad (2.6)$$

where ψ^c is the c -transform of ψ given by $\psi^c(y) := \inf_x c(x, y) - \psi(x)$ and ψ runs through functions on Ω such that $\psi = \phi^c$ for some $\phi \in C(\Omega)$. When a function $\bar{\psi}$ attains the maximum, it is called a Kantorovich potential. The link between the static dual problem and the dynamic dual problem will be made clear in the next section.

2.2 Viscosity Solutions to the Hamilton–Jacobi Equation

We briefly review the notion of viscosity solutions as it plays a central role in this article and highlights PDE aspects of optimal transport. The Hamilton–Jacobi equations often do not admit a classical solution, but naive notions of weak solution such as continuous functions differentiable almost Lebesgue everywhere can be too weak so that infinitely many solutions may exist. To ensure both existence and weakness, Evans [25] and Crandall and Lions [22] independently introduced so-called viscosity solutions. For a Hamilton–Jacobi equation of the form

$$\partial_t \varphi + H(\nabla \varphi) = 0$$

on a domain $[0, 1] \times \Omega$, a Lipschitz function φ is said to be a viscosity subsolution (resp. supersolution) if, for any test function $f \in C^1([0, 1] \times \Omega)$, any local maximum x_0 of $f - \varphi$ (resp. local minimum) satisfies

$$\partial_t f(x_0) + H(\nabla f(x_0)) \leq 0 \quad (\text{resp. } \partial_t f(x_0) + H(\nabla f(x_0)) \geq 0).$$

The inequality constraint of the dynamic dual optimal transport (2.2) means that a competitor φ is required to be a viscosity subsolution.

When φ is both a viscosity subsolution and a viscosity supersolution, it is called a viscosity solution. This definition indeed guarantees the existence of a unique solution. We present here the statement of [21] summarizing the results of [22]: even though it is phrased in $\mathbb{R}^d$, it can be adapted at no cost on $\Omega = \mathbb{R}^d / (D\mathbb{Z}^d)$ by simply identifying a function on Ω with a D -periodic function on $\mathbb{R}^d$.

Theorem 2.3 (*Existence and uniqueness of viscosity solution*) *Let $H : \mathbb{R}^d \rightarrow \mathbb{R}$ be continuous and $\varphi_0 : \Omega \rightarrow \mathbb{R}$ be Lipschitz. Then there exists exactly one viscosity solution to the initial value problem,*

$$\begin{cases} \partial_t \varphi + H(\nabla \varphi) = 0 & \text{in } [0, \infty) \times \Omega, \\ \varphi(0, \cdot) = \varphi_0 & \text{in } \Omega. \end{cases}$$

Moreover, the Lipschitz constant does not increase in time i.e.

$$\text{Lip}(\varphi(t', \cdot)) \leq \text{Lip}(\varphi(t, \cdot)) \quad \text{for } t' > t.$$

Actually, in this simple setting where the Hamiltonian is not dependent on the spatial variable, the unique viscosity solution is characterized by the Hopf-Lax formula [26, Theorem 3 in Chapter 10.3]: with the assumptions and notations of the theorem it is equal to

$$\varphi(t, x) := \inf_y \varphi_0(y) + tL\left(\frac{x-y}{t}\right). \quad (2.7)$$

The Hopf-Lax formula is a key ingredient to bridge the dual formulations of optimal transport in its static (2.6) and dynamic (2.2) form. This is also what enables us to prove Proposition 2.2.

Proof of Proposition 2.2 We can choose an optimal Kantorovich potential $\bar{\psi}$ for the static dual problem (2.6) so that it is Lipschitz continuous with $\text{Lip}(\bar{\psi}) \leq \sup_y \text{Lip}(c(\cdot, y)) \leq \text{Lip}(L, B_{\text{diam}(\Omega)})$ [45, Sect. 1.2.]. Then we define $\bar{\varphi}$ as the unique viscosity solution of the Hamilton–Jacobi equation with initial data $-\bar{\psi}$. Observe by (2.7) that $\bar{\varphi}(1, \cdot) = \bar{\psi}^c$ and hence

$$\int_{\Omega} \bar{\varphi}(1, \cdot) \, dv - \int_{\Omega} \bar{\varphi}(0, \cdot) \, d\mu = \int_{\Omega} \bar{\psi}^c \, dv + \int_{\Omega} \bar{\psi} \, d\mu = \mathcal{K}(\mu, \nu).$$

Therefore, $\bar{\varphi}$ is not merely an admissible competitor but is an optimal potential for the dynamic dual problem (2.2). Finally Theorem 2.3 asserts that $\text{Lip}(\varphi(t, \cdot)) \leq \text{Lip}(\varphi(0, \cdot)) = \text{Lip}(\bar{\psi})$ for $t \in [0, 1]$. $\square$

2.3 Discretization of the Hamilton–Jacobi Equation

As we mentioned in the introduction, our key idea is to discretize not the primal formulation (2.1) but rather the dual formulation (2.2). Indeed discretization of the Hamilton–Jacobi equations is a widely studied topic, in particular since the seminal work of Crandall and Lions [23].

We adapt Crandall and Lions’s original setting of the domain $[0, \infty) \times \mathbb{R}^d$ for $Q = [0, 1] \times (\mathbb{R}^d / (D\mathbb{Z}^d))$. Their results are still valid on Q without additional conditions as we can simply extend functions on $\Omega = \mathbb{R}^d / (D\mathbb{Z}^d)$ by copying it infinite times for $\mathbb{R}^d$.

The time domain $[0, 1]$ is discretized by uniformly sampling points: we define $T_D := \{0, \Delta t, \dots, (N_T - 1)\Delta t, 1\}$ with some positive integer N_T and $\Delta t = 1/N_T$. The space domain $\Omega = \mathbb{R}^d / (D\mathbb{Z}^d)$ is discretized in the same way, that is we write $\Omega_D := \{(j_1 \Delta x, \dots, j_d \Delta x)\}$ with indices $(j_1, \dots, j_d) \in (\mathbb{Z}/N_X \mathbb{Z})^d := \{0, 1, \dots, N_X - 1\}^d$ with some positive integer N_X and $\Delta x = D/N_X$. In the sequel, we write $j := (j_1, \dots, j_d)$ and $j \Delta x := (j_1 \Delta x, \dots, j_d \Delta x)$ for simplicity. Hence the discretization of Q is given by $Q_D := T_D \times \Omega_D$.

Each continuous function φ is approximated by a discrete function $\Phi : Q_D \rightarrow \mathbb{R}$ given as the evaluation of φ at the grid points. For the value at $(i \Delta t, j \Delta x)$, we write Φ_j^i or $\Phi_{j_1, \dots, j_d}^i$, but we will occasionally omit subscripts i and j when we mean the collection $\{\Phi_j^i\}_j$ or we do not need to specify i or j . We denote by $\mathbb{R}^Z$ the collection of functions from a set $Z \subseteq Q_D$ to $\mathbb{R}$, and define

$$\|\Psi\|_{L^\infty(Z)} := \max_{z \in Z} |\Psi(z)|.$$

for function $\Psi \in \mathbb{R}^Z$.

Discretization of the Hamilton–Jacobi equation of the form $\partial_t \varphi + H(\nabla \varphi) = 0$ is given as follows. The initial state of the discrete function Φ^0 is given on the grid points $\{0\} \times \Omega_D \subset Q_D$ and it is time-updated as

$$\Phi^{i+1} = \mathcal{S}(\Phi^i),$$

by a map $\mathcal{S} : \mathbb{R}^{\Omega_D} \rightarrow \mathbb{R}^{\Omega_D}$ called a *scheme*. We will deal with a certain class of schemes following a standard finite difference setting as in [23]. A scheme $\mathcal{S}$ is said of *difference form* if there exists a function $\mathcal{G} : \mathbb{R}^{2d} \rightarrow \mathbb{R}$ such that

$$\mathcal{S}_j(\Psi) = \Psi_j - \Delta t \mathcal{G} \left(\frac{\Delta_{-,j} \Psi}{\Delta x}, \frac{\Delta_{+,j} \Psi}{\Delta x} \right),$$

for each j where $\Delta_{-,j} := (\Delta_{-,j}^1, \dots, \Delta_{-,j}^d)$ takes the spatially backward difference

$$\Delta_{-,j}^k \Psi := \Psi_{j_1, \dots, j_k, \dots, j_d} - \Psi_{j_1, \dots, j_k - 1, \dots, j_d},$$

for each k and $\Delta_{+,j} := (\Delta_{+,j}^1, \dots, \Delta_{+,j}^d)$ takes the forward difference

$$\Delta_{+,j}^k \Psi := \Psi_{j_1, \dots, j_k + 1, \dots, j_d} - \Psi_{j_1, \dots, j_k, \dots, j_d},$$

for each $k = \{1, \dots, d\}$. Of course the expression $j_k \pm 1$ is understood modulo N_X as we have periodic boundary conditions. We will omit the subscript j for $\Delta_{\pm}$ when there is no confusion. The two important properties that a scheme can have are the following:

- *Consistency*. A scheme $\mathcal{S}$ of difference form is *consistent* with H if $\mathcal{G}$ satisfies $\mathcal{G}(a, a) = H(a)$ for any $a \in \mathbb{R}^d$. It is equivalent to say that $\mathcal{S}$ gives the exact solution on grid points for any time-space affine function.
- *Monotonicity*. Let us define a subset of discrete space functions

$$\mathcal{C}_R := \left\{ \Psi \in \mathbb{R}^{\Omega_D} \text{ such that } \forall k, j \left| \frac{\Delta_{+,j}^k \Psi}{\Delta x} \right| \leq R \right\}$$

for a fixed $R > 0$. The space $\mathcal{C}_R$ can be interpreted as the discrete counterpart of the functions which are R -Lipschitz in each coordinate. A scheme is *monotone* on

$[-R, R]$ if for each grid point j the restriction of $\mathcal{S}_j$ to $\mathcal{C}_R$ is non-decreasing with respect to any of its variables.

The striking result is that these two properties guarantee not only convergence of the discrete solutions to the continuous ones, but also a quantitative rate of convergence. Crandall and Lions showed a quantitative estimate for the approximation error of such discrete solutions on a family of time-space discretizations of domain Q [23, Theorem 1]. In their work, the ratio $\zeta := \Delta t / \Delta x$ within a family of discretizations is fixed. Throughout this article, we follow their setting, and for each discretization Q_D we denote its resolution by $h := \Delta t$.

Theorem 2.4 (Convergence of discrete Hamilton–Jacobi equation) *Let $H : \mathbb{R}^d \rightarrow \mathbb{R}$ be continuous and φ be the unique viscosity solution to the initial value problem,*

$$\begin{cases} \partial_t \varphi + H(\nabla \varphi) = 0 & \text{in } Q = [0, 1] \times \Omega, \\ \varphi(0, \cdot) = \varphi_0, \end{cases}$$

with initial data φ_0 Lipschitz in each coordinate with Lipschitz constant $R > 0$. Let $Q_D = T_D \times \Omega_D$ be a discretization of Q with resolution $\Delta t = h$ and ratio $\zeta = \Delta t / \Delta x$ fixed. Define a discrete solution $\Phi \in \mathbb{R}^{Q_D}$ by

$$\begin{cases} \Phi^{i+1} = \mathcal{S}(\Phi^i), & i \in \{0, \dots, N_T - 1\}, \\ \Phi_j^0 := \varphi_0(j \Delta x), & \forall j, \end{cases}$$

with a scheme $\mathcal{S}$ of difference form which is consistent with H and monotone on $[-R - \delta, R + \delta]$ for some $\delta > 0$.

Then we have the estimate,

$$\|\Phi - \varphi\|_{L^\infty(Q_D)} \leq C \sqrt{h}$$

with a constant C depending only on the scheme $\mathcal{S}$, $\|\varphi_0\|_{L^\infty(Q)}$, $R + \delta$, and H .

This theorem will be the main result we will use to get our quantitative convergence rates for the optimal transport problem.

Remark 2.5 The theorem is usually stated with $\delta = 1$, that is, the scheme should be monotone on $[-R - 1, R + 1]$, but a close inspection of the proof in [23] reveals that any $\delta > 0$ is possible.

Remark 2.6 (Lipschitz in each coordinate and why we restrict to a periodic setting) We have done another modification compared to the classical statement of the theorem. We assume that φ_0 is R -Lipschitz in each coordinate instead of simply R -Lipschitz, that is, for every k and $(x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_d) \in \mathbb{R}^{d-1} / (D\mathbb{Z}^{d-1})$ the function

$$x \mapsto \varphi_0(x_1, \dots, x_{k-1}, x, x_{k+1}, \dots, x_d)$$

is R -Lipschitz. A R -Lipschitz function is R -Lipschitz in each coordinate, and a function which is R -Lipschitz in each coordinate is $\sqrt{d}R$ -Lipschitz in the classical sense.

Functions which are R -Lipschitz in each coordinate are the perfect analogue of the space $\mathcal{C}_R$ of discrete functions, and this will be useful in the proof of our Theorem 4.1.

An important step in the proof of the theorem is the propagation of the regularity for the discrete solution: one can prove that if $\Phi^0 \in \mathcal{C}_R$ then

$$\Phi^i \in \mathcal{C}_R, \quad \forall i. \quad (2.8)$$

It is obtained by combining two simple arguments. The first one is that, as the scheme commutes with the addition with constant functions and is monotone, it is non-expansive in L^∞ :

$$\|\mathcal{S}(\Psi) - \mathcal{S}(\Psi')\|_{L^\infty(\Omega_D)} \leq \|\Psi - \Psi'\|_{L^\infty(\Omega_D)}$$

for any pair of Ψ and Ψ' in $\mathcal{C}_R$ [23, Proposition 3.1]. The second one is to apply the non-expansiveness to Ψ' , a shifted version of Ψ . As a shift in space commutes with the discrete differential operators Δ_+^k, Δ_-^k for fixed k , we obtain

$$\|\Delta_+^k \mathcal{S}(\Psi)\|_{L^\infty(\Omega_D)} \leq \|\Delta_+^k \Psi\|_{L^\infty(\Omega_D)},$$

and from there an immediate induction gives (2.8). Note that the same argument at the continuous level gives that, if φ_0 is R -Lipschitz in each coordinate, then so is $\varphi(t, \cdot)$ for any $t \geq 0$.

Importantly, this regularizing effect of the scheme does not work if the spatial domain is no longer the torus nor the whole Euclidean space: the first estimate stays valid but the second argument about the commutativity breaks down on the boundary. For this reason, our work is phrased on a periodic domain rather than on a bounded domain: we will not rely on the estimate (2.8) in itself, but this estimate is necessary in order for Theorem 2.4 to be true.

2.3.1 Vanishing Viscosity Scheme

An example of monotone and consistent schemes presented in [23] is the so-called vanishing viscosity scheme, which is a discrete analogue of the vanishing viscosity method for the continuous Hamilton–Jacobi equation. For a discrete space function $\Psi \in \mathbb{R}^{\Omega_D}$, it is given by

$$\mathcal{S}(\Psi) = \Psi - \Delta t \{H(\nabla_D \Psi) - \varepsilon \Delta_D \Psi\}, \quad (2.9)$$

with some $\varepsilon > 0$. Here ∇_D, Δ_D are respectively the discrete centered gradient and the discrete Laplacian given by,

$$\nabla_{D,j} \Psi := \frac{\Delta_{-,j} \Psi + \Delta_{+,j} \Psi}{2\Delta x}, \quad \Delta_{D,j} \Psi := \sum_k \frac{\Delta_{+,j}^k \Psi - \Delta_{-,j}^k \Psi}{(\Delta x)^2}$$

at each location j . With the (forward) discrete time derivative

$$\partial_{\Delta t}^i \Phi := \frac{\Phi^{i+1} - \Phi^i}{\Delta t},$$

the scheme (2.9) reads

$$\partial_{\Delta t} \Phi + H(\nabla_D \Phi) - \varepsilon \Delta_D \Phi = 0,$$

for a discrete time-space function $\Phi \in \mathbb{R}^{Q_D}$, which is a discrete analogue of the Hamilton–Jacobi equation with the viscosity term.

The consistency is immediate: for $\Psi \in \mathbb{R}^{\Omega_D}$ satisfying $\Delta_+ \Psi / \Delta x = \Delta_- \Psi / \Delta x = (a_1, \dots, a_d)$, we have $H(\nabla_D \Psi) = H(a_1, \dots, a_d)$ and $\Delta_D \Psi = 0$. The monotonicity is not attained only with the Hamiltonian term, but can be ensured together with the viscosity term. A simple computation reveals that the scheme is monotone on $\mathcal{C}_R$ when ε satisfies

$$\frac{\text{Lip}(H, B_R)}{2} \leq \frac{\varepsilon}{\Delta x} \leq \frac{\Delta x}{2d \Delta t} \quad (2.10)$$

where $\text{Lip}(H, B_R)$ is the Lipschitz constant of H on the centered ball of radius R . Recalling that we assume that the ratio $\zeta = \Delta t / \Delta x$ is fixed among a family of discretizations of Q , we need to choose $\zeta \geq d \text{Lip}(H, B_R)$, and in this case ε will go to zero at rate $h = \Delta t \propto \Delta x$: the viscosity ε vanishes together with the stepsize.

Several alternatives can be found in the literature, such as the upwind scheme also in [23] (for $d = 1$), higher order finite difference schemes [41], and also discontinuous Galerkin methods [18]. We, however, focus on the vanishing viscosity scheme in this article because of the following reasons: it is simple; we are able to solve efficiently the discrete dynamic optimal transport we build on it; and the function $\bar{\varphi}$ is a priori not expected to be better than Lipschitz uniformly over the space-time domain so using high order schemes seems less helpful as far as the theoretical analysis is concerned.

3 Discrete Dynamic Optimal Transport

In this section, we introduce a discrete formulation of dynamic optimal transport. We first do so using a general monotone and consistent scheme, and then specifically with the vanishing viscosity scheme explained in the previous section.

3.1 Discrete Problem for a General Scheme

We discretize the dynamic dual problem (2.2) on $Q = [0, 1] \times \Omega$. For the discretization of the domain, we use $Q_D = T_D \times \Omega_D$ defined in Sect. 2.3.

Regarding the discretization of probability measures, we follow a standard approach. For a given $\mu \in \mathcal{P}(\Omega)$, we define its discretization $\Pi \mu$ in a way that $\Pi \mu \xrightarrow{*} \mu$ as $\Delta x \rightarrow 0$. A simple choice is the projection onto the Dirac measures on

grid points, given by

$$\Pi\mu = \sum_j \mu(B_{j\Delta x})\delta_{j\Delta x}, \quad (3.1)$$

where $B_{j\Delta x}$ is the d -dimensional half-open box centered at $j\Delta x$ with edge length Δx given by,

$$B_{j\Delta x} = [(j_1 - 1/2)\Delta x, (j_1 + 1/2)\Delta x) \times \cdots \times [(j_d - 1/2)\Delta x, (j_d + 1/2)\Delta x).$$

Remark 3.1 (Choice of discrete measure) In this article, we stick to this specific discretization of measures, but this is not the unique choice. For example, piecewise uniform measures in the boxes $\{B_{j\Delta x}\}_j$ is also a reasonable option. All our results easily extend to this discretization of measures as soon as Lemma 4.2 is valid.

Clamped discrete gradient With the above settings, we are almost ready to introduce our discrete optimal transport problem. We finally impose a constraint on discrete functions as follows. We will later guarantee the convergence of the discrete transport cost to the continuous one. For this result, we will need Theorem 2.4 which requires a monotone scheme on $[-R - \delta, R + \delta]$ with $\delta > 0$ and an initial condition $\Phi^0 \in \mathcal{C}_R$. However, the boundedness of $\Delta_+ \Phi^0 / \Delta x$ is not a priori guaranteed in our upcoming formulation of discrete optimal transport, in contrast to that, the gradient of an optimal potential can be chosen to be bounded by $\text{Lip}(L, B_{\text{diam}(\Omega)})$ in the continuous setting stated as Proposition 2.2. To cope with this problem, we explicitly constrain

$$\forall j, k, \quad \left| \frac{\Delta_{+,j}^k \Phi^0}{\Delta x} \right| \leq R$$

for a parameter $R \geq \text{Lip}(L, B_{\text{diam}(\Omega)})$. Said differently we impose $\Phi^0 \in \mathcal{C}_{\text{Lip}(L, B_R)}$.

Definition 3.2 (Discrete dynamic optimal transport) Let $\mu, \nu \in \mathcal{P}(\Omega)$ and choose a parameter $R \geq \text{Lip}(L, B_{\text{diam}(\Omega)})$. Let us assume a time-space discretization as defined so far, and $\mathcal{S}$ be a scheme of difference form that is monotone on $[-R - \delta, R + \delta]$ for some $\delta > 0$ and consistent with the Hamiltonian H of the continuous optimal transport problem. We say that the discrete optimal transport cost between μ and ν is

$$\mathcal{K}_D(\mu, \nu) := \max_{\Phi} \int_{\Omega} \Phi^{N_T} d\Pi\nu - \int_{\Omega} \Phi^0 d\Pi\mu \quad (3.2)$$

where $\Phi \in \mathbb{R}^{\mathcal{Q}_D}$ runs over the discrete functions satisfying,

$$\Phi^{i+1} - \mathcal{S}(\Phi^i) \leq 0, \quad \forall i \in \{0, \dots, N_T - 1\}, \quad (3.3)$$

$$\left| \frac{\Delta_{+,j}^k \Phi^0}{\Delta x} \right| \leq R \quad \forall j \in \{0, \dots, N_X - 1\}^d, \quad \forall k \in \{1, \dots, d\}. \quad (3.4)$$

If a discrete function $\bar{\Phi}$ attains $\mathcal{K}_D(\mu, \nu)$, we call $\bar{\Phi}$ an *optimal potential* for the discrete problem.

This definition clearly mimics (2.2) which is the continuous dual problem. At this level of generality, this is not necessarily a concave maximization problem. It would depend on the precise choice of the scheme $\mathcal{S}$. We emphasize that our convergence result for the transport cost (Theorem 4.1) holds even in the absence of concavity.

Remark 3.3 (Break of the symmetry between μ and ν) This discrete optimal transport is in general not symmetric between μ and ν , that is $\mathcal{K}_D(\mu, \nu) \neq \mathcal{K}_D(\nu, \mu)$ as can be seen both from the Hamilton–Jacobi constraint (3.3) and the constraint (3.4) on the initial potential. This is in contrast with other discretizations like [32, 42] which are symmetric in μ and ν .

Remark 3.4 The threshold R for clamping in Definition 3.2 is allowed to be greater than $\text{Lip}(L, B_{\text{diam}(\Omega)})$ as far as the scheme $\mathcal{S}$ is monotone. But choosing a larger R makes room for the scheme to be monotone smaller (see (2.10)), and makes the convergence a bit slower in the sense that the multiplicative constant in front of $\sqrt{h}$ in Theorem 4.1 increases with R . Thus choosing a larger R would make sense only if one does not have an exact access to $\text{Lip}(L, B_{\text{diam}(\Omega)})$.

3.2 Discrete Problem with Vanishing Viscosity

We have introduced a discretization of dynamic optimal transport. Notice it is a discrete counterpart of the dynamic *dual* formulation. In order to obtain quantities such as optimal measures and optimal velocities, we need a *primal* formulation as well. This will not be needed for our main convergence result about the transport cost (Theorem 4.1), but will be necessary for the convergence of optimizers (Theorem 5.1) to make sense.

We focus on the vanishing viscosity scheme explained in Sect. 2.9 as this is a simple and practical example of schemes that make the discrete optimal transport a convex problem. With this scheme the constraint (3.3) is written as

$$\partial_{\Delta t} \Phi + H(\nabla_D \Phi) - \varepsilon \Delta_D \Phi \leq 0.$$

which is a convex constraint.

Remark 3.5 (Range of admissible parameters) Let’s focus on the case where L (and thus H) are radial. Note that the scheme must be monotone on $[-R - \delta, R + \delta]$ with $\delta > 0$ and $R := \text{Lip}(L, B_{\text{diam}(\Omega)})$. But as ∇L and ∇H are the inverse to each other, the constraint (2.10) reads

$$\frac{\text{diam}(\Omega)}{2} < \frac{\varepsilon}{\Delta x} \leq \frac{\Delta x}{2d \Delta t}.$$

Interestingly it does *not* depend on L .

To obtain the primal formulation, we begin with writing the dual problem (Definition 3.2) more concretely for the vanishing viscosity scheme. To make the exposition simple, we introduce some notations. We fix $\mu, \nu \in \mathcal{P}(\Omega)$ in the rest of the section and define a functional $F_D : \mathbb{R}^{\mathcal{Q}_D} \rightarrow \mathbb{R}$ by

$$F_D \Phi = \int_{\Omega} \Phi^{N_T} d\Pi \nu - \int_{\Omega} \Phi^0 d\Pi \mu, \quad (3.5)$$

and set a constant $R \geq \text{Lip}(L, B_{\text{diam}(\Omega)})$. Next, we concatenate discrete differential operators as follows. We define $T'_D := \{0, \Delta t, \dots, (N_T - 1)\Delta t\}$ by dropping the last time step $N_T = 1$ from T_D and define a subset of the entire discrete space $\mathcal{Q}_D$ by $\mathcal{Q}'_D := T'_D \times \Omega_D$. We then define an operator $A = (A_t, A_x, A_R) : \mathbb{R}^{\mathcal{Q}_D} \rightarrow \mathbb{R}^{\mathcal{Q}'_D + d \times \mathcal{Q}'_D + d \times \Omega_D}$ by

$$\begin{aligned} A_t \Phi &= (\partial_{\Delta t}^0 \Phi - \varepsilon \Delta_D \Phi^0, \dots, \partial_{\Delta t}^{N_T-1} \Phi - \varepsilon \Delta_D \Phi^{N_T-1}) \in \mathbb{R}^{\mathcal{Q}'_D}, \\ A_x \Phi &= (\nabla_D \Phi^0, \dots, \nabla_D \Phi^{N_T-1}) \in \mathbb{R}^{d \times \mathcal{Q}'_D}, \\ A_R \Phi &= (\Delta_+ \Phi^0 / \Delta x) \in \mathbb{R}^{d \times \Omega_D}. \end{aligned}$$

We rewrite the problem with these settings.

Definition 3.6 (Dual problem with vanishing viscosity) Let $\mu, \nu \in \mathcal{P}(\Omega)$, H be the Hamiltonian of the continuous problem, and $\mathcal{S}$ be the vanishing viscosity scheme. The dual formulation of discrete transport problem is defined as

$$\sup_{\Phi} F_D \Phi, \quad (3.6)$$

with constraints

$$A_t \Phi + H(A_x \Phi) \leq 0, \quad (3.7)$$

$$|A_R \Phi| \leq R. \quad (3.8)$$

To derive the expression of the dual of this problem, which we call the *primal* problem, we use the method of Lagrange multipliers. We introduce a new variable Σ in the dual which at optimality coincides with $A\Phi$, and take a Lagrange multiplier Λ to enforce the constraint $A\Phi = \Sigma$. Specifically the problem reads as the saddle point problem

$$\mathcal{K}_D(\mu, \nu) = \sup_{\Phi, \Sigma} \inf_{\Lambda} \mathcal{L}(\Phi, \Sigma, \Lambda), \quad (3.9)$$

with the functional

$$\mathcal{L}(\Phi, \Sigma, \Lambda) = F_D \Phi - I_{\Sigma_t + H(\Sigma_x) \leq 0} - I_{|\Sigma_R| \leq R} - \Lambda \cdot (A\Phi - \Sigma).$$

The newly introduced symbols are defined as follows. We set the variable $\Sigma := (\Sigma_t, \Sigma_x, \Sigma_R) \in \mathbb{R}^{Q'_D + d \times Q'_D + d \times \Omega_D}$. Each indicator function I takes 0 if the constraint is satisfied, and otherwise $+\infty$. The variables $\Lambda := (\Lambda_\rho, \Lambda_m, \Lambda_\eta) \in \mathbb{R}^{Q'_D + d \times Q'_D + d \times \Omega_D}$ are the Lagrange multipliers. Finally by “ $\cdot$ ” we denote the standard Euclidean product between vectors in $\mathbb{R}^N$, with a dimension N which should be clear from context. It should be interpreted as the integral of a discrete scalar or vector field by a discrete scalar or vector-valued measure respectively, which is just the summation of element-wise products.

Now let us consider the formal exchange of the infimum and the supremum. By a direct computation, we have

$$\begin{aligned} \inf_{\Lambda} \sup_{\Phi, \Sigma} \mathcal{L}(\Phi, \Sigma, \Lambda) &= \inf_{\Lambda} \sup_{\Phi, \Sigma} (F_D - A^\top \Lambda) \cdot \Phi - I_{\Sigma_t + H(\Sigma_x) \leq 0} - I_{|\Sigma_R| \leq R} + \Lambda \cdot \Sigma \\ &= \inf_{\Lambda} \sup_{\Phi, \Sigma_t, \Sigma_x} (F_D - A^\top \Lambda) \cdot \Phi - I_{\Sigma_t + H(\Sigma_x) \leq 0} \\ &\quad + \Lambda_\rho \cdot \Sigma_t + \Lambda_m \cdot \Sigma_m + R \cdot |\Lambda_\eta| \end{aligned}$$

where in the last line we have performed the maximization over Σ_R by taking $\Sigma_R = R \Lambda_\eta / |\Lambda_\eta|$ elementwise. If Λ_ρ is not element-wise non-negative, then the supremum in $\Sigma_t \rightarrow -\infty$ would yield $+\infty$. But once we know $\Lambda_\rho \geq 0$, we see that the supremum in Σ is attained in the boundary of the convex constraints, namely $\Sigma_t = -H(\Sigma_x)$. When we take the supremum in Σ_x we see that Λ_ρ should be strictly positive if Λ_m is non-zero and that in this case

$$\sup_{\Sigma_x} \Lambda_m \cdot \Sigma_x - \Lambda_\rho \cdot H(\Sigma_x) = L \left(\frac{\Lambda_m}{\Lambda_\rho} \right) \cdot \Lambda_\rho$$

Note that in this case we also have at optimality

$$\frac{\Lambda_m}{\Lambda_\rho} = \nabla H(\Sigma_x) = \nabla H(\nabla_D \Phi) \quad (3.10)$$

for element-wise non-zero Λ_ρ . Thus the problem boils down to

$$\inf_{\Lambda} \sup_{\Phi, \Sigma} \mathcal{L}(\Phi, \Sigma, \Lambda) = \inf_{\Lambda} L \left(\frac{\Lambda_m}{\Lambda_\rho} \right) \cdot \Lambda_\rho + R \cdot |\Lambda_\eta| + \sup_{\Phi} \left\{ (F_D - A^\top \Lambda) \cdot \Phi \right\}$$

Writing $R \cdot |\Lambda_\eta| = R \|\Lambda_\eta\|_{L^1(\Omega_D)}$ and interpreting Φ as a Lagrange multiplier we obtain the following minimization problem.

Definition 3.7 (Primal problem with vanishing viscosity) Let $\mu, \nu \in \mathcal{P}(\Omega)$, L be the Lagrangian of the continuous problem, and $\mathcal{S}$ be the vanishing viscosity scheme. The primal formulation of discrete transport problem is defined as

$$\inf_{\Lambda} L \left(\frac{\Lambda_m}{\Lambda_\rho} \right) \cdot \Lambda_\rho + R \|\Lambda_\eta\|_{L^1(\Omega_D)} \quad (3.11)$$

where Λ runs through discrete functions satisfying the discrete continuity equation,

$$A^\top \Lambda = F_D \quad (3.12)$$

together with the element-wise constraints $\Lambda_\rho \geq 0$ and $\Lambda_m = 0$ as soon as $\Lambda_\rho = 0$. When $\bar{\Lambda} = (\bar{\Lambda}_\rho, \bar{\Lambda}_m, \bar{\Lambda}_\eta)$ attains the minimum, we call $\bar{\Lambda}_\rho$ an optimal measure and $\bar{\Lambda}_m$ an optimal momentum.

Remark 3.8 Note that (3.12) is a discrete version of the continuity equation with a viscosity term. The temporal boundary conditions are encoded in the equation as expanding (3.12) reads

$$\begin{cases} \partial_{-, \Delta t} \Lambda_\rho - \nabla_D^\top \Lambda_m + \varepsilon \Delta_D \Lambda_\rho = 0, \\ \Lambda_\rho^{-1} = \Pi \mu - \frac{\Delta_- \Lambda_\eta}{\Delta x}, \quad \Lambda_\rho^{N_T-1} = \Pi v. \end{cases}$$

Here $\nabla_D^\top$ is the adjoint of the discrete gradient ∇_D with respect to the standard Euclidean product, hence $-\nabla_D^\top$ is a discrete analogue of divergence. The operator $\partial_{-, \Delta t}$ is the backward discrete time derivative given as $\partial_{-, \Delta t}^i \Lambda_\rho := (\Lambda_\rho^i - \Lambda_\rho^{i-1})/\Delta t$ for $i \in \{0, 1, \dots, N_T - 1\}$ provided that Λ_ρ^{-1} is defined as in the second line above.

Thus it is tempting to think of the primal problem (Definition 3.7) as the discretization of

$$\inf_{\rho, m} \int_0^1 \int_\Omega L\left(\frac{dm}{d\rho}\right) d\rho dt \quad \text{such that} \quad \begin{cases} \partial_t \rho + \nabla \cdot m = -\varepsilon \Delta \rho \\ \rho_0 = \mu, \quad \rho_1 = v. \end{cases}$$

Said differently: from the beginning the parameter ε was interpreted in the dual problem as a regularization parameter, as it introduces a diffusive term $\varepsilon \Delta \varphi$ in the Hamilton–Jacobi equation. The computation we made shows that, in the primal problem, the same parameter ε also adds diffusion, but this time in the continuity equation which becomes $\partial_t \rho + \nabla \cdot m = -\varepsilon \Delta \rho$. In particular, the case of $L(v) = |v|^2/2$ amounts to the entropic regularized optimal transport [27].

However, this analogy is not perfect. First, because of the additional variable Λ_η used to constrain the discrete gradient in the dual formulation, but also because ε depends on Δx and vanishes as Δx goes to 0. Nevertheless, this analogy explains why our discrete solutions would be slightly more smooth than the continuous solutions (and less and less as the stepsize decreases), something that we observe numerically in Sect. 6.

We conclude this section by stating that our exchange between infimum and supremum was formal, but can be made rigorous.

Proposition 3.9 (Strong duality of discrete optimal transport) *There is no duality gap between the dual problem (Definition 3.6) and the primal problem (Definition 3.7) i.e.*

we have

$$F_D \bar{\Phi} = L \left(\frac{\bar{\Lambda}_m}{\bar{\Lambda}_\rho} \right) \cdot \bar{\Lambda}_\rho + R \|\bar{\Lambda}_\eta\|_{L^1(\Omega_D)},$$

for a maximizer $\bar{\Phi}$ of the dual problem and a minimizer $\bar{\Lambda}$ of the primal problem.

Proof Slater’s condition states that the strong duality holds if the feasible region has non-empty interior [15, Sects. 5.2.3 & 5.3.2]. In our case, it suffices to show the existence of a strictly admissible competitor of the dual problem *i.e.* $\Phi \in \mathbb{R}^{\mathcal{Q}_D}$ satisfying,

$$A_t \Phi + H(A_x \Phi) < 0, \quad |A_R \Phi| < R.$$

It is attained for instance by defining $\Phi_j^i := -i \Delta t H(0) - \epsilon$ with some $\epsilon > 0$. $\square$

4 Convergence of the Optimal Transport Cost

We now state and prove our main result on the quantitative convergence of the optimal transport cost. This result is valid for a general monotone and consistent scheme.

Theorem 4.1 (Convergence of transport cost) *Let $\mu, \nu \in \mathcal{P}(\Omega)$ and let $\mathcal{K}(\mu, \nu)$ and $\mathcal{K}_D(\mu, \nu)$ be the continuous and the discrete optimal transport cost; see respectively (2.2) and Definition 3.2. Then there is a constant C depending only on Ω , L and R such that*

$$|\mathcal{K}(\mu, \nu) - \mathcal{K}_D(\mu, \nu)| \leq C \sqrt{h}.$$

To make the exposition of the proof simple, let us introduce the following notations: for a function $\varphi : \mathcal{Q} \rightarrow \mathbb{R}$ we define the functionals F and F_D via

$$F\varphi := \int_{\Omega} \varphi(1, \cdot) \, d\nu - \int_{\Omega} \varphi(0, \cdot) \, d\mu, \quad F_D\varphi := \int_{\Omega} \varphi(1, \cdot) \, d\Pi\nu - \int_{\Omega} \varphi(0, \cdot) \, d\Pi\mu.$$

Our goal is to control the gap between $F\bar{\varphi}$ and $F_D\bar{\Phi}$ for the continuous and discrete optimal potentials $\bar{\varphi}$ and $\bar{\Phi}$ respectively. However, the relation between $\bar{\varphi}$ and $\bar{\Phi}$ is unclear as they are solutions to two different problems. To circumvent this issue, we replace them with functions that can be easily compared with each other. For this purpose, we exploit not only a solution to the discrete Hamilton–Jacobi equation, but also a viscosity solution to the continuous Hamiltonian–Jacobi equation. Before proving the theorem, we review elementary results about errors caused by discretizing measures.

Lemma 4.2 (*Gap between continuous and discrete measures*) Let $\varphi : Q \rightarrow \mathbb{R}$ be a Lipschitz function. Then we have a control,

$$|F\varphi - F_D\varphi| \leq \frac{\sqrt{d}}{2} \Delta x (\text{Lip}(\varphi(0, \cdot)) + \text{Lip}(\varphi(1, \cdot))).$$

Proof First note that the Wasserstein-1 distances $W_1(\Pi\mu, \mu)$ and $W_1(\Pi\nu, \nu)$ are bounded by $\sqrt{d}\Delta x/2$ since each mass moves up to $\sqrt{d}\Delta x/2$ via discretization. The claim follows from the fact that, for any Lipschitz function $f : \Omega \rightarrow \mathbb{R}$ and $\rho_1, \rho_2 \in \mathcal{P}(\Omega)$ we have,

$$\int_{\Omega} f \, d(\rho_1 - \rho_2) \leq \text{Lip}(f) W_1(\rho_1, \rho_2).$$

□

Lemma 4.3 (*Gap between discrete integral of functions*) The functional F_D is 2-Lipschitz with respect to $L^\infty(Q_D)$ norm.

Proof The claim follows from the definition as $\Pi\mu, \Pi\nu$ are probability measures. □

Proof of Theorem 4.1 Let $\bar{\varphi}$ and $\bar{\Phi}$ be respectively the solutions to the continuous and discrete problem. Note that we can take $\bar{\varphi}$ to satisfy the Lipschitz bound (2.3) thanks to Proposition 2.2. On the other hand $\bar{\Phi}$ satisfies a similar bound (3.4) by design.

We first show $F_D\bar{\Phi} \leq F\bar{\varphi} + C\sqrt{h}$. Let us take the solution to the discrete initial value problem,

$$\begin{cases} \tilde{\Phi}^{i+1} = \mathcal{S}(\tilde{\Phi}^i), \\ \tilde{\Phi}^0 = \bar{\Phi}^0. \end{cases}$$

Note that $\bar{\Phi}$ and $\tilde{\Phi}$ have the same initial data, and that $\bar{\Phi} \leq \tilde{\Phi}$ due to the inequality constraint (3.3) and the monotonicity of the scheme, hence $F_D\bar{\Phi} \leq F_D\tilde{\Phi}$. To estimate $F_D\tilde{\Phi}$ let us consider $\tilde{\varphi}$ the unique viscosity solution to the the continuous initial value problem

$$\begin{cases} \partial_t \tilde{\varphi} + H(\nabla \tilde{\varphi}) = 0, \\ \tilde{\varphi}(0, \cdot) = \text{LI}(\bar{\Phi}^0), \end{cases}$$

where LI gives the piecewise linear interpolation of a discrete function. When $d = 1$, it is given for a discrete space function Ψ by

$$\text{LI}(\Psi)(x) := \frac{(j+1)\Delta x - x}{\Delta x} \Psi_j + \frac{x - j\Delta x}{\Delta x} \Psi_{j+1}, \quad x \in [j\Delta x, (j+1)\Delta x].$$

For higher dimensions, it is given by bilinear interpolation, trilinear interpolation, and so on. As $\bar{\Phi}^0 \in \mathcal{C}_R$ we can check that $\tilde{\varphi}(0, \cdot)$ is R -Lipschitz in each coordinate

and moreover $\text{Lip}(\tilde{\varphi}(0, \cdot)) \leq \sqrt{d}R$. Thus in particular $\text{Lip}(\tilde{\varphi}(t, \cdot)) \leq \sqrt{d}R$ for any $t \in [0, 1]$ (see Theorem 2.3). As $\tilde{\varphi}$ is an admissible competitor, we have $F\tilde{\varphi} \leq F\bar{\varphi}$. Thus we see

$$F_D\bar{\Phi} \leq F_D\tilde{\Phi} \leq F_D\tilde{\Phi} + F\tilde{\varphi} - F\tilde{\varphi} \leq F\bar{\varphi} + (F_D\tilde{\Phi} - F_D\tilde{\varphi}) + (F_D\tilde{\varphi} - F\tilde{\varphi}).$$

For the term $F_D\tilde{\Phi} - F_D\tilde{\varphi}$ we use here Lemma 4.3 followed by Theorem 2.4, as our constraint (3.4) guarantees that $\tilde{\varphi}(0, \cdot)$ is R -Lipschitz in each coordinate:

$$F_D\tilde{\Phi} - F_D\tilde{\varphi} \leq 2\|\tilde{\Phi} - \tilde{\varphi}\|_{L^\infty(Q_D)} \leq C\sqrt{h}.$$

On the other hand, it follows from Lemma 4.2 and the regularization effect of the continuous Hamilton–Jacobi equation (Theorem 2.3) that

$$F_D\tilde{\varphi} - F\tilde{\varphi} \leq \frac{\sqrt{d}\Delta x}{2} (\text{Lip}(\tilde{\varphi}(1, \cdot)) + \text{Lip}(\tilde{\varphi}(0, \cdot))) \leq \sqrt{d}\Delta x \text{Lip}(\tilde{\varphi}(0, \cdot)) \leq Ch,$$

hence we obtained the claimed inequality as C can be taken independent on μ, ν thanks to (3.4).

For the other inequality $F\bar{\varphi} \leq F_D\bar{\Phi} + C\sqrt{h}$, we start from an optimal potential $\bar{\varphi}$ to the continuous dual transport problem which satisfies the estimate (2.3) and is a viscosity solution to the continuous Hamilton–Jacobi equation. To transform it into a discrete competitor we consider the discrete system

$$\begin{cases} \tilde{\Phi}^{i+1} = \mathcal{S}(\tilde{\Phi}^i), \\ \tilde{\Phi}_j^0 = \bar{\varphi}(0, j\Delta x), \end{cases}$$

using the point sample of the continuous potential $\bar{\varphi}$ as initial data. Then its solution $\tilde{\Phi}$ is an admissible competitor for the discrete transport problem: this is because the discrete constraint (3.4) is automatically satisfied thanks to (2.3). Thus we have $F_D\tilde{\Phi} \leq F_D\bar{\Phi}$ that we use in

$$F\bar{\varphi} = F\bar{\varphi} - F_D\bar{\varphi} + F_D\bar{\varphi} + F_D\tilde{\Phi} - F_D\tilde{\varphi} \leq F_D\bar{\Phi} + (F\bar{\varphi} - F_D\bar{\varphi}) + (F_D\bar{\varphi} - F_D\tilde{\varphi}).$$

In a similar way, we have $F\bar{\varphi} - F_D\bar{\varphi} \leq Ch$ thanks to Lemma 4.2 and the bound (2.3). On the other hand $F_D\bar{\varphi} - F_D\tilde{\varphi} \leq C\sqrt{h}$: for this estimate we rely again on Lemma 4.3 followed by Theorem 2.4. $\square$

Remark 4.4 (Optimal transport on a bounded domain) We built a discrete formulation on the spatial domain $\Omega = \mathbb{R}^d/(D\mathbb{Z}^d)$ to avoid boundary conditions, and we prove our convergence results in this setting. At least from a theoretical point of view, given measures μ, ν with compact support, it is possible to embed them in $\mathbb{R}^d/(D\mathbb{Z}^d)$ with D large enough such that no mass crosses the “periodic boundary”, and thus the optimal transport on the torus between μ and ν coincides with the optimal transport on $\mathbb{R}^d$. However from a numerical point of view this is problematic, as most of the domain

Ω would be empty, and thus most of the nodes of the discretized domain Ω_D do not record any motion of mass. This leads to a lot of wasted computational resources.

One way to address this issue may be to use a discretization of the Hamilton–Jacobi equation on a bounded domain. We should then impose normal boundary conditions $\partial\varphi/\partial n = 0$, being n the outward normal to the domain. We refer to [44] and [1] for discussions of possible discretization. However, at this point the challenge would be to adapt the proof of Theorem 4.1. As we discussed in Remark 2.6, the main problem is that the regularizing effect (2.8) is no longer available automatically in such schemes, so that even clamping the value of the gradient at the initial time is not enough.

We leave for future work the proposal of a discretization of dynamic optimal transport via its dual formulation which would also handle the case of convex domains of $\mathbb{R}^d$.

Remark 4.5 By looking into the proof of Theorem 4.1, we see that a quantitative convergence of the transport cost in our discrete problem is provable if, for the chosen scheme $\mathcal{S}$, a) solutions of the discrete Hamilton–Jacobi equation converge to continuous viscosity solutions at a known rate, and b) the discrete solutions have controlled Lipschitz constants at time $t = 0$ and $t = 1$. This may leave the room for other scheme. The challenge though would be to design an optimization method to solve the resulting discrete problem. Typically, certain types of upwind schemes define $\mathcal{S}$ piecewise, making it challenging to optimize over constraints such as $\Phi^{i+1} - \mathcal{S}(\Phi^i) \leq 0$. We stick to the vanishing viscosity scheme because of the simplicity of its implementation.

5 Convergence of the Optimizers

We can also obtain a quantitative convergence of solutions to discrete problems using the convergence of transport cost we showed in the previous section. Note that if the functionals to optimize were uniformly strictly convex, then this would be a direct consequence of the convergence of the value, that is, of the transport cost. This is not possible here: the functional in the dual problem (2.2) is only linear in φ ! Actually without further assumptions on μ, ν there is not necessarily a unique solution to the continuous problem so it is very unlikely we can prove any convergence of the optimizers.

Thus in this section we impose additional assumptions that are typically satisfied when the measures μ, ν have some smoothness. Then, by a (now classical) study of the duality gap between the primal and dual problem, it is possible to recover some convergence of the optimizers.

Assumption on the scheme We give our result specifically for the discrete problem with the vanishing viscosity scheme we introduced in Sect. 3.2. At this point we are not aware of the exact conditions for extending the results to a general scheme as it requires also the primal formulation of the discrete problem.

Assumption on the Lagrangian and the Hamiltonian We suppose that the Lagrangian L and the Hamiltonian H satisfy for any $v, w \in \mathbb{R}^d$,

$$L(v) + H(w) \geq v \cdot w + |f_L(v) - f_H(w)|^2, \quad (5.1)$$

for some functions $f_L, f_H : \mathbb{R}^d \rightarrow \mathbb{R}^d$. This is an improvement of Young's inequality $L(v) + H(w) \geq v \cdot w$ and we refer to [46] for a discussion of this assumption, including the link with Bregman divergences. We refer to two examples of such L and H from [46, Lemma 3.3 and Lemma 3.2]. If L is uniformly convex *i.e.* $D^2L \geq \lambda I$ for some $\lambda > 0$, we have

$$L(v) + H(w) \geq v \cdot w + \frac{\lambda}{2} |v - \nabla H(w)|^2.$$

On the other hand, for $L(v) = |v|^p/p$ and $H(w) = |w|^q/q$ with some $p, q > 1$ such that $1/p + 1/q = 1$, we have

$$L(v) + H(w) \geq v \cdot w + \frac{1}{2 \max\{p, q\}} |v^{p/2} - w^{q/2}|^2,$$

where the power of a vector is defined as $v^p = v|v|^{p-1}$. Note in any case, as we know that there is equality in Young's inequality when $v = \nabla H(w)$, that we must necessarily have for $w \in \mathbb{R}^d$:

$$f_H(w) = f_L(\nabla H(w)). \quad (5.2)$$

Assumption on the optimal potential Finally, we assume that an optimal potential of the continuous dual problem $\bar{\varphi}$ is of class $C^{1,1}$ on $[0, 1] \times \Omega$. That is, $\bar{\varphi}$ is time-space differentiable and its derivatives are time-space Lipschitz. If the measures μ, ν are in $\mathbb{R}^d$, by Caffarelli's regularity theory a sufficient condition for this is that they are supported on a uniformly convex C^2 domains X_μ, X_ν and have Lebesgue densities f_μ, f_ν that are Hölder continuous and bounded from below by a strictly positive constant on $\overline{X_\mu}, \overline{X_\nu}$ respectively [49, Theorem 4.14]. On the periodic domain Ω , it is still the case if X_μ, X_ν are small enough and close to each other so that the situation reduces to the case of $\Omega = \mathbb{R}^d$ and the optimal velocity of mass $\bar{v} = \nabla H(\nabla \bar{\varphi})$ has no discontinuity on the support of the measures. Alternatively on the torus a sufficient condition for $\bar{\varphi}$ of class $C^{1,1}$ is f_μ, f_ν being nowhere vanishing and Hölder continuous [20, Theorem 1], [3, Theorem 2.2 (iii)].

Such a condition on μ, ν is still weaker than the one required for similar results in the recent work [38], which has to assume that the optimal density $\bar{\rho}$ solving the continuous primal problem is uniformly bounded below in the entire domain Q .

Statement of the result Now we state our convergence result for optimizers. We quantify the result in terms of the discrete L^2 norm $\|\cdot\|_{L^2_{\bar{\Lambda}_\rho}(Q'_D)}$ on Q'_D given by

$$\|M\|_{L^2_{\bar{\Lambda}_\rho}(Q'_D)}^2 = \sum_{z \in Q'_D} |M_z|^2 \bar{\Lambda}_{\rho,z}$$

for $M \in \mathbb{R}^{d \times Q'_D}$: it is norm weighted by $\bar{\Lambda}_\rho$ which is solution to the discrete primal problem (see Definition 3.7). Eventually, for a continuous function φ defined on our domain Q , we denote by $\Pi\varphi$ the function defined on Q_D which corresponds to the pointwise evaluation of the function on grid points, that is, $(\Pi\varphi)_j^i = \varphi(i\Delta t, j\Delta x)$.

Theorem 5.1 (Norm convergence of optimizers) Suppose that the Lagrangian L and the Hamiltonian H of the continuous problem satisfy the inequality (5.1) with some functions f_L and f_H . Let $\bar{\varphi}$ and $\bar{\Phi}$ be optimal potentials for the continuous and the discrete dual problems, and let $\bar{\Lambda} = (\bar{\Lambda}_\rho, \bar{\Lambda}_m, \bar{\Lambda}_\eta)$ be an optimizer for the discrete primal problem. Assume that $\bar{\varphi} \in C^{1,1}(Q)$. Then we have estimate,

$$\|f_H(\nabla_D \bar{\Phi}) - f_H(\nabla_D \Pi \bar{\varphi})\|_{L^2_{\bar{\Lambda}_\rho}(Q'_D)}^2 \leq C\sqrt{h}, \quad (5.3)$$

where the constant C depends only on $\text{diam}(\Omega)$, L and R . If f_H is Lipschitz on bounded sets we also have an estimate

$$\|f_L(\bar{V}) - f_L(\bar{v})\|_{L^2_{\bar{\Lambda}_\rho}(Q'_D)}^2 \leq C\sqrt{h}, \quad (5.4)$$

where $\bar{v} = \nabla H(\nabla \bar{\varphi})$ is the optimal velocity for an optimal pair $(\bar{\rho}, \bar{v})$ of the continuous problem and the optimal discrete velocity $\bar{V} := \bar{\Lambda}_m / \bar{\Lambda}_\rho$ defined on $\text{support}(\bar{\Lambda}_\rho)$.

Corollary 5.2 The estimates in Theorem 5.1 hold for the Lagrangians $L(v) = |v|^p/p$ with $p \leq 2$, and in this case they read, with $1/p + 1/q = 1$,

$$\|(\nabla_D \bar{\Phi})^{q/2} - (\nabla_D \Pi \bar{\varphi})^{q/2}\|_{L^2_{\bar{\Lambda}_\rho}(Q'_D)}^2 \leq C\sqrt{h}, \quad \|\bar{V}^{p/2} - \bar{v}^{p/2}\|_{L^2_{\bar{\Lambda}_\rho}(Q'_D)}^2 \leq C\sqrt{h}.$$

In particular for $p = 2$, it simplifies to

$$\|\nabla_D \bar{\Phi} - \nabla_D \Pi \bar{\varphi}\|_{L^2_{\bar{\Lambda}_\rho}(Q'_D)}^2 \leq C\sqrt{h}, \quad \|\bar{V} - \bar{v}\|_{L^2_{\bar{\Lambda}_\rho}(Q'_D)}^2 \leq C\sqrt{h}.$$

Note that a quantity such as $\|\bar{V}^{p/2} - \bar{v}^{p/2}\|_{L^2_{\bar{\Lambda}_\rho}(Q'_D)}^2$ is not a discrete L^p norm, it is a discrete L^2 norm between (vectorial) powers of quantities of interest.

Proof The map $f_H(w) = w^{q/2}/\sqrt{2q}$ for $q = p/(p-1) \geq 2$ is Lipschitz on bounded sets, while we can take $f_L(v) = v^{p/2}/\sqrt{2q}$, and the factor $\sqrt{2q}$ can be absorbed in the constant C . The second claim follows from $f_L(v) = v/2$ and $f_H(w) = w/2$ for $p = 2$. $\square$

To prove the theorem we follow an argument already present in the literature. It consists in quantifying the duality gap between an optimal primal solution and a (non-optimal) dual competitor: see [8, 46] and references therein for this argument applied to get regularity estimates at the continuous level, and [38] where this method is used to analyze discretization of dynamic optimal transport. We first prove some primal-dual relationship at the continuous level, then estimate the duality gap at the discrete level, and finally proceed with the proof.

Proposition 5.3 (*Primal-dual relation at optimality in the continuous problem*) Let L, H, f_L, f_H be assumed in Theorem 5.1. Let $(\bar{\rho}, \bar{v})$ be an optimizer for the continuous primal problem and $\bar{\varphi}$ be an optimizer in the dual problem. Then at least $\bar{\rho}$ -a.e.

$$\bar{v} = \nabla H(\bar{\varphi}), \quad (5.5)$$

and in particular combining with (5.2) we obtain

$$f_L(\bar{v}) = f_H(\nabla \bar{\varphi}). \quad (5.6)$$

When $\bar{\varphi}$ is $C^{1,1}$ the relation (5.5) is actually a way to define $\bar{v}$ in an unambiguous manner outside of the support of $\bar{\rho}$. The analogue for the discrete problem would be (3.10) that we derived formally when doing the inf-sup exchange. For the proof one could use arguments in the style of [8, Lemma 4.1] and even quantify a duality gap but we would have to be careful between the pairing of the continuity equation in a weak sense and the Hamilton–Jacobi equation in a viscosity sense. Rather we exploit the precise structure of the optimizers that come from the static problem, at the price of more obscure proof for the non-expert reader.

Proof Recall that $-\bar{\varphi}(0, \cdot)$ is an optimal Kantorovich potentials in the static dual problem (2.6); see the proof of Proposition 2.2. Thus the optimal transport map T between μ and ν is given by $T(x) = x + \nabla H(\nabla \varphi(0, x))$ at least μ -a.e. (see e.g. [45, Sect. 1.3]). Let us write $T_t = (1 - t)\text{Id} + tT$ which is the optimal transport between μ and $\bar{\rho}_t$ and is invertible [45, Lemma 5.29]. Proposition 5.30 in [45] yields $\bar{v}_t = (T - \text{Id}) \circ T_t^{-1}$.

On the other hand as $\bar{\varphi}$ is smooth and solves the Hamilton–Jacobi equation we can use the method of characteristics [26, Sect. 3.3]. The characteristics are given by $t \mapsto T_t(x)$, and we know that $\nabla \bar{\varphi}$ is constant along characteristics. That gives $\nabla \bar{\varphi}(t, T_t(x)) = \nabla \bar{\varphi}(0, x)$. Composing with ∇H and using the definition of T we indeed obtain $\nabla H(\nabla \bar{\varphi})(t, T_t(x)) = \nabla H(\nabla \bar{\varphi})(0, x) = T(x) - x = \bar{v}_t(T_t(x))$. Eventually we compose on the left with T_t^{-1} to get the final result. $\square$

In the next lemma, we show the key estimate: the quantification of the duality gap for the discrete problem.

Lemma 5.4 (*Quantification of the duality gap in the discrete problem*) Let $L, H, f_L, f_H, \bar{\Phi}, \bar{\Lambda} = (\bar{\Lambda}_\rho, \bar{\Lambda}_m, \bar{\Lambda}_\eta), \bar{V} = \bar{\Lambda}_m/\bar{\Lambda}_\rho$ be assumed to be as in Theorem 5.1. Then for any admissible discrete competitor Φ , we have,

$$\|f_H(\nabla_D \bar{\Phi}) - f_H(\nabla_D \Phi)\|_{L^2_{\bar{\Lambda}_\rho}(Q'_D)}^2 \leq F_D \bar{\Phi} - F_D \Phi. \quad (5.7)$$

Proof We will actually prove

$$\|f_L(\bar{V}) - f_H(\nabla_D \Phi)\|_{L^2_{\bar{\Lambda}_\rho}(Q'_D)}^2 \leq F_D \bar{\Phi} - F_D \Phi, \quad (5.8)$$

and by setting $\Phi = \bar{\Phi}$ we obtain $f_L(\bar{V}) = f_H(\nabla_D \bar{\Phi})$ on support $(\bar{\Lambda}_\rho)$ (that we already derived combining (3.10) and (5.2)), thus the stated inequality.

Thanks to the absence of duality gap at optimality (see Proposition 3.9), we obtain with the discrete continuity equation $F_D = A^\top \Lambda$ (3.12),

$$\begin{aligned} F_D \bar{\Phi} - F_D \Phi &= L \left(\frac{\bar{\Lambda}_m}{\bar{\Lambda}_\rho} \right) \cdot \bar{\Lambda}_\rho + R \|\bar{\Lambda}_\eta\|_{L^1(\Omega_D)} - F_D \Phi \\ &= L \left(\frac{\bar{\Lambda}_m}{\bar{\Lambda}_\rho} \right) \cdot \bar{\Lambda}_\rho + R \|\bar{\Lambda}_\eta\|_{L^1(\Omega_D)} - (A^\top \bar{\Lambda}) \cdot \Phi \\ &= L(\bar{V}) \cdot \bar{\Lambda}_\rho + R \|\bar{\Lambda}_\eta\|_{L^1(\Omega_D)} - \bar{\Lambda}_\rho \cdot (A_t \Phi) \\ &\quad - \bar{\Lambda}_m \cdot (A_x \Phi) - \bar{\Lambda}_\eta \cdot (A_R \Phi) \\ &\geq L(\bar{V}) \cdot \bar{\Lambda}_\rho - \bar{\Lambda}_\rho \cdot (A_t \Phi) - \bar{\Lambda}_m \cdot (A_x \Phi) \\ &= \bar{\Lambda}_\rho \cdot \{L(\bar{V}) - \bar{V} \cdot A_x \Phi - A_t \Phi\} \\ &\geq \bar{\Lambda}_\rho \cdot \left\{ -A_t \Phi - H(A_x \Phi) + |f_L(\bar{V}) - f_H(A_x \Phi)|^2 \right\} \\ &\geq \|f_L(\bar{V}) - f_H(\nabla_D \Phi)\|_{L^2_{\bar{\Lambda}_\rho}(Q'_D)}^2, \end{aligned}$$

where we have used in the last inequality the constraint (3.7) coming from the admissibility of Φ . $\square$

We next quantify the violation of the discrete constraint that occurs by discretizing a continuous admissible competitor. This is where we crucially rely on the $C^{1,1}$ regularity of φ .

Lemma 5.5 (*Violation of the Hamilton–Jacobi constraint due to discretization*) *Let φ be an admissible potential for the continuous dual problem and $\Pi\varphi$ be its discretization. Assume that $\varphi \in C^{1,1}(Q)$. Then the maximum violation of the discrete Hamilton–Jacobi inequality by $\Pi\varphi$ is controlled as,*

$$\| [\partial_{\Delta t} \Pi\varphi + H(\nabla_D \Pi\varphi) - \varepsilon \Delta_D \Pi\varphi]^+ \|_{L^\infty(Q'_D)} \leq Ch,$$

with a constant C depending only on H and $\|D^2\varphi\|_{L^\infty(Q)}$. Here $[\cdot]^+$ denotes the positive part and the parameter ε satisfies the monotonicity condition (2.10) for the vanishing viscosity scheme.

Proof Using the admissibility of φ and a standard argument by the mean value theorem, we have

$$\begin{aligned} &\| [\partial_{\Delta t} \Pi\varphi + H(\nabla_D \Pi\varphi) - \varepsilon \Delta_D \Pi\varphi]^+ \| \\ &\leq \| \partial_{\Delta t} \Pi\varphi - \partial_t \varphi \| + \| H(\nabla_D \Pi\varphi) - H(\nabla \varphi) \| + \varepsilon \| \Delta_D \Pi\varphi \| + \| [\partial_t \varphi + H(\nabla \varphi)]^+ \| \\ &\leq \Delta t \| \partial_{tt} \varphi \| + (\Delta x \operatorname{Lip}(H, B_R) + \varepsilon) \sum_{k,l} \| \partial_{kl} \varphi \|, \end{aligned}$$

where we denoted $\|\cdot\|_{L^\infty(Q'_D)}$ by $\|\cdot\|$ for simplicity. Finally, the condition (2.10) ensures that ε scales like Δx , so that together with the assumption $\varphi \in C^{1,1}(Q)$, we get the stated inequality. $\square$

Proof of Theorem 5.1 Let $\delta_{\bar{\varphi}} := \|\partial_{\Delta t} \Pi \bar{\varphi} + H(\nabla_D \Pi \bar{\varphi}) - \varepsilon \Delta_D \Pi \bar{\varphi}\|_{L^\infty(Q'_D)}$ be the maximum violation of the discrete Hamilton–Jacobi inequality by $\Pi \bar{\varphi}$. We define a continuous potential $\tilde{\varphi}$ by

$$\tilde{\varphi}(t, x) := \bar{\varphi}(t, x) - t \delta_{\bar{\varphi}}$$

so that it is an admissible competitor for the continuous problem and $\Pi \tilde{\varphi}$ is an admissible competitor for the discrete problem by construction. This allows us to apply Lemma 5.4 to obtain

$$\begin{aligned} \|f_H(\nabla_D \bar{\Phi}) - f_H(\nabla_D \Pi \bar{\varphi})\|_{L^2_{\Lambda_\rho}(Q'_D)}^2 &= \|f_H(\nabla_D \bar{\Phi}) - f_H(\nabla_D \Pi \tilde{\varphi})\|_{L^2_{\Lambda_\rho}(Q'_D)}^2 \\ &\leq F_D \bar{\Phi} - F_D \Pi \tilde{\varphi} = F_D \bar{\Phi} - F_D \bar{\varphi} + \delta_{\bar{\varphi}}. \end{aligned}$$

As $F_D \bar{\varphi} \geq F \bar{\varphi} - Ch$ by Lemma 4.2, and that $F \bar{\varphi} \geq F_D \bar{\Phi} - C\sqrt{h}$ thanks to Theorem 4.1, we obtain indeed

$$\|f_H(\nabla_D \bar{\Phi}) - f_H(\nabla_D \Pi \bar{\varphi})\|_{L^2_{\Lambda_\rho}(Q'_D)}^2 \leq Ch + C\sqrt{h} + \delta_{\bar{\varphi}}.$$

Then we use Lemma 5.5 to bound $\delta_{\bar{\varphi}}$ and get (5.3).

The estimate (5.4) is a simple consequence. We have

$$\|\nabla \bar{\varphi} - \nabla_D \Pi \bar{\varphi}\|_{L^\infty(Q'_D)} \leq \Delta x \|D^2 \bar{\varphi}\|_{L^\infty(Q)} \leq C'h,$$

with some constant $C' \geq 0$ by a standard argument using the mean value theorem. With Proposition 5.3 and the regularity assumption for f_H we have the inequality,

$$\begin{aligned} \|f_L(\bar{V}) - f_L(\bar{v})\|_{L^2_{\Lambda_\rho}(Q'_D)} &= \|f_H(\nabla_D \bar{\Phi}) - f_H(\nabla \bar{\varphi})\|_{L^2_{\Lambda_\rho}(Q'_D)} \\ &\leq \|f_H(\nabla_D \bar{\Phi}) - f_H(\nabla_D \Pi \bar{\varphi})\|_{L^2_{\Lambda_\rho}(Q'_D)} + C''h. \end{aligned}$$

with some $C'' \geq 0$ depending on the Lipschitz constant of f_H on $B_{\text{Lip}(L, \text{diam}(\Omega))}$. Applying the estimate (5.3) concludes the proof. $\square$

Remark 5.6 (Comments on the rate $\sqrt{h}$) We see from the proofs of Theorem 4.1 and Theorem 5.1 that the lowest powers of convergence come from Theorem 2.4 (Theorem 1 in [23]). This means that the convergence rate $\sqrt{h}$ for the solutions of the discrete Hamilton–Jacobi equations to the continuous one determines the convergence rate of both the cost and the optimizer of the optimal transport problem. In our numerical experiments in Sect 6, we found no examples where the convergence rate is exactly $\sqrt{h}$. This suggests that the convergence for any pair of (possibly very rough) input

measures is better than our theoretical result $\sqrt{h}$. While the root $\sqrt{h}$ for a discrete Hamilton–Jacobi equation (Theorem 2.4) is sharp because viscosity solutions may develop shocks in finite time, it is known that solutions $\bar{\varphi}$ to the dual problem can only have shocks exactly at time $t = 0$ or $t = 1$ (see the proof of [49, Theorem 5.51]). It is however still unclear to us if this property can be leveraged to improve the rate of convergence. Some discretizations of Hamilton–Jacobi equation with higher order of convergence for smooth inputs can be found e.g. in [18, 50] but the convergence is in L^2 , and not in L^∞ as in Theorem 2.4. Note, however, that a different discretization of the Hamilton–Jacobi equation with higher rates of convergence does not automatically improve our result: one would have to make sure that the steps in the proof of Theorem 4.1 and Theorem 5.1 carry through, and also that the constraint (3.3) is easy to handle numerically in case the scheme is different.

6 Numerical Illustrations

We numerically examine our discrete optimal transport problems and convergence results.

6.1 Settings

We solve the discrete problem with the vanishing viscosity scheme introduced in Sect. 3.2. For the cost function, we choose the quadratic Lagrangian $L(v) = |v|^2/2$ following many works in the numerical computation of optimal transport although our theory is not limited to this specific cost function.

As for the domain, we discretize the $1 + 1$ dimensional time-space domain $Q = [0, 1] \times [-\frac{1}{2}, \frac{1}{2}]$ with identification $(t, -\frac{1}{2}) \sim (t, \frac{1}{2})$. We make a family of discrete domains by subdividing Q into grid points with subdivisions $N_T \in \{16 \times 2^n \mid n = 0, 1, 2, 3, 4, 5\}$ and $N_X = N_T$. Namely the resolution $h = \Delta t = \Delta x$ of each discrete domain is in $\{\frac{1}{16}, \dots, \frac{1}{16 \times 2^5}\}$.

For each input probability measure, we compute its discretization given as (3.1) that we evaluate with explicit expressions or via numerical integration.

Implementation of our discretization

For each test case, we fix a pair of probability measures μ and ν and numerically solve the transport problem on each discrete domain. This can be performed by any standard convex optimization framework. We can obtain an optimal potential $\bar{\Phi}$ by solving the dual problem (Definition 3.6) and a minimizer $\bar{\Lambda} = (\bar{\Lambda}_\rho, \bar{\Lambda}_m, \bar{\Lambda}_\eta)$ by solving the primal problem (Definition 3.7). As an example implementation, we solved the problems using the alternating direction method of multipliers (ADMM) [14]. As this method directly finds a saddle point of (3.9) the output of the algorithm contains both $\bar{\Phi}$ and $\bar{\Lambda}$. We ran ADMM until the L^2 norms of both the primal and dual residuals became smaller than 10^{-5} . Such stopping criteria are discussed in literature such as [15]. Similarly to observations made in other discretizations of optimal transport such as [32], when the mesh size h decreases in our discretization, the number of necessary iterations for ADMM is unaffected while the time per iteration increases, thus the total

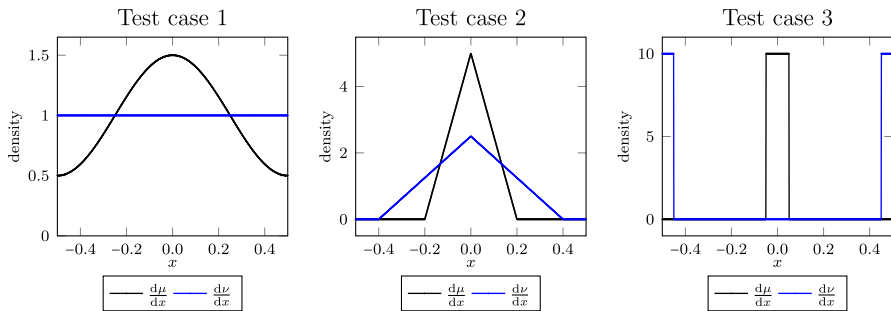


Fig. 1 Densities of μ and ν , the initial and final measures, with respect to the Lebesgue measure

computational times increases. For interested readers, we share the source code in Jupyter (Python 3); see <https://github.com/sdsgisd/DynamicOTwithDualFormulation>.

Comparison with standard finite differences

As a point of comparison, we also solve the same test cases with the finite difference discretization proposed by Papadakis, Peyré, and Oudet [42] while it does not guarantee any convergence result. For that end, we used the Julia code originally written by the second author to handle regularized unbalanced optimal transport [4]. In this code the resulting convex optimization problem is solved with proximal splitting which makes hard an exact comparison with our ADMM implementation. The Julia code is also shared in the same repository as above.

Monitoring of the errors

We consider the following gaps between continuous and discrete quantities:

$$\begin{aligned}\epsilon_{\mathcal{K}} &:= |\mathcal{K}(\mu, \nu) - \mathcal{K}_D(\mu, \nu)|, \\ \epsilon_{\varphi} &:= \|\nabla_D \Pi \bar{\varphi} - \nabla_D \bar{\Phi}\|_{L^2_{\bar{\Lambda}_\rho}(Q'_D)}^2, \\ \epsilon_v &:= \|\bar{v} - \bar{V}\|_{L^2_{\bar{\Lambda}_\rho}(Q'_D)}^2, \\ \epsilon_\rho &:= \|\Pi \bar{\rho}_t - \bar{\Lambda}_\rho\|_{L^1(Q'_D)}.\end{aligned}$$

Here $\bar{\rho}$, $\bar{v}$ are the optimal measure, velocity, and potential for the continuous problem. The error $\epsilon_{\mathcal{K}}$ is for the transport cost as in Theorem 4.1 and the error ϵ_v is for optimal velocities in Corollary 5.2 (a special case of Theorem 5.1). While we do not have a proven estimate between the continuous and discrete optimal measures, we also compute the error ϵ_ρ for a reference purpose.

We plotted these errors in Fig. 3, in which we also estimated the rates α such that $\epsilon \sim h^\alpha$ by a standard routine with linear regression in the log domain. We omitted the plots for the error ϵ_φ since the explicit expression for $\bar{\varphi}$ is unavailable in Test case 1, and ϵ_φ overlapped almost perfectly with ϵ_v in Test case 2 and 3 as $\bar{V} = \nabla_D \bar{\Phi}$ and the difference between ϵ_v and ϵ_φ originates from the discrepancy between $\nabla_D \Pi \bar{\varphi}$ and $\bar{v} = \nabla \bar{\varphi}$ which is always zero except at a few exceptional grid points in these test cases.

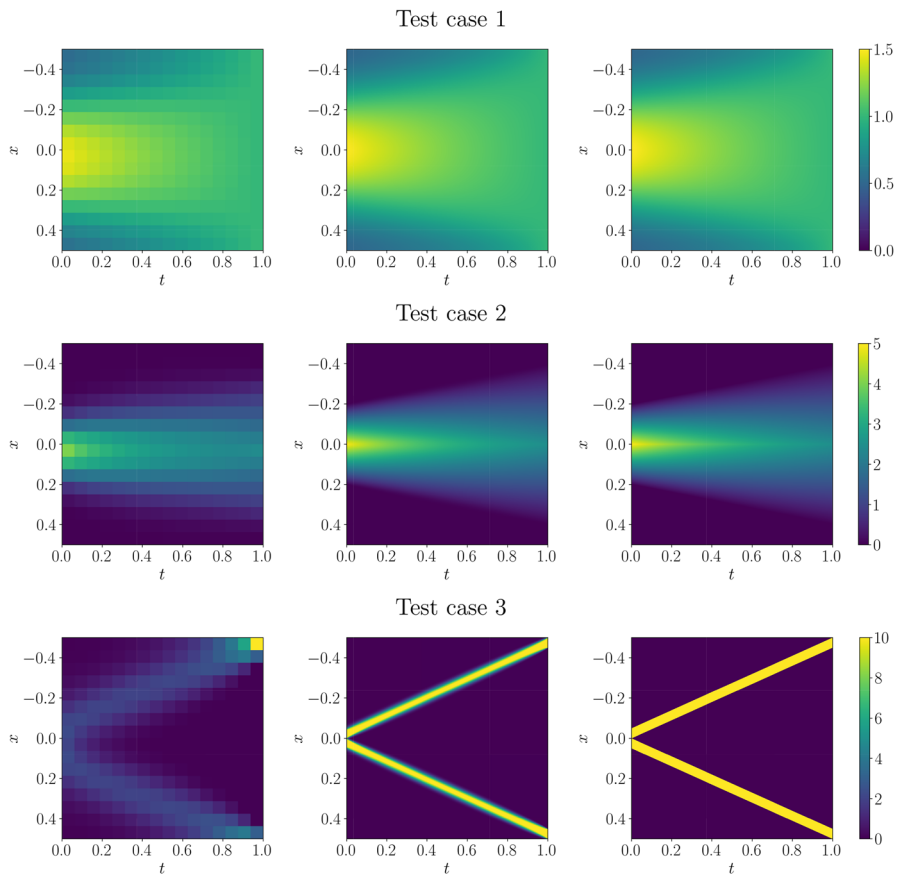


Fig. 2 Discrete and continuous optimal measures. From left to right: the discrete optimal measures $\bar{\Lambda}_\rho$ for $h = 1/16$ and $h = 1/16 \times 2^5$, and the continuous optimal measure $\bar{\rho}$

6.2 Test Cases

We test three simple examples with different levels of regularity. In the second and the third test cases, the measures μ, ν do not have full support, so they do not satisfy the assumptions of the previous works finding quantitative rates.

Test case 1 Our first example is a pair of smooth densities bounded from below, which is of the highest regularity among our test cases. We set

$$\frac{d\mu}{dx}(x) = 1 + \frac{1}{2} \cos(2\pi wx), \quad \frac{d\nu}{dx}(x) = 1,$$

with a parameter $w \in \mathbb{Z} \setminus 0$, which we set $w = 1$ (Fig. 1). The optimizers and the transport cost are

$$\frac{d\bar{\rho}}{dx}(t, x) = \frac{1 + \frac{1}{2} \cos(2w\pi T_t^{-1}(x))}{1 + \frac{t}{2} \cos(2w\pi T_t^{-1}(x))}, \quad \bar{v}(t, x) = \frac{1}{4\pi w} \sin(2\pi w T_t^{-1}(x)),$$

$$\mathcal{K}(\mu, \nu) = \frac{1}{64\pi^2 w^2}.$$

where the inverse of the time- t optimal transport map $T_t(y) = y + t \sin(2\pi wy)/4\pi w$ can be computed by the Newton-Raphson method, for instance. Note that the expression for $\bar{\rho}$ was obtained through the change of variable formula $d\bar{\rho}_t/dx(T_t(y), t) = d\mu/dx(y)(T_t'(y))^{-1}$. The input measures have absolutely continuous densities bounded from below, which is a condition required in the previous work for the quantitative convergence [38]. We observed that all the errors are decreasing linearly or faster than the stepsize h as in Fig. 3, thus faster than our upper bound.

Test case 2 We next test the transport between triangular densities. We set μ and ν by,

$$\frac{d\mu}{dx}(x) = \frac{1}{w^2} \max(w - |x|, 0), \quad \frac{d\nu}{dx}(x) = \frac{1}{4w^2} \max(2w - |x|, 0),$$

with a parameter $0 < w < 1/4$, which we set $w = 0.2$ (Fig. 1). The optimizers and the transport cost are

$$\frac{d\bar{\rho}}{dx}(t, x) = \frac{1}{(1+t)^2 w^2} \max((1+t)w - |x|, 0), \quad \bar{\varphi}(t, x) = \frac{x^2}{2(1+t)},$$

$$\bar{v}(t, x) = \frac{x}{1+t}, \quad \mathcal{K}(\mu, \nu) = \frac{w^2}{12}.$$

The measure $\bar{\rho}$ is expanded and flattened in time as in Fig. 2. Even though μ and ν are not supported on the whole space, $\bar{\varphi}$ is of class $C^{1,1}$ thus both Theorem 4.1 and Theorem 5.1 apply. All the errors are empirically decreasing linearly or faster as in Fig. 3.

Test case 3 Our last example is the breaking of a unimodal density toward another one, where the potential is not of class $C^{1,1}$. We set μ and ν as characteristic functions,

$$\frac{d\mu}{dx}(x) = \frac{1}{2w} \mathbb{1}_{|x| \leq w}, \quad \frac{d\nu}{dx}(x) = \frac{1}{2w} \mathbb{1}_{1/2 - |x| \leq w},$$

with a parameter $0 < w < 1/2$, which we set $w = 0.05$ (Fig. 1). The optimizers and the transport cost are

$$\frac{d\bar{\rho}}{dx}(t, x) = \frac{1}{2w} \mathbb{1}_{0 \leq |x| - ts \leq w}, \quad \bar{\varphi}(t, x) = |x|s - \frac{s^2 t}{2}, \quad \bar{v}(t, x) = s \operatorname{sign}(x),$$

$$\mathcal{K}(\mu, \nu) = \frac{s^2}{2}$$

where all the mass travel the same distance $s = \frac{1}{2} - w$. Note that the mass travels in two directions as seen in Fig. 2 and the velocity field $\bar{v}$ is discontinuous at $x = 0$. We

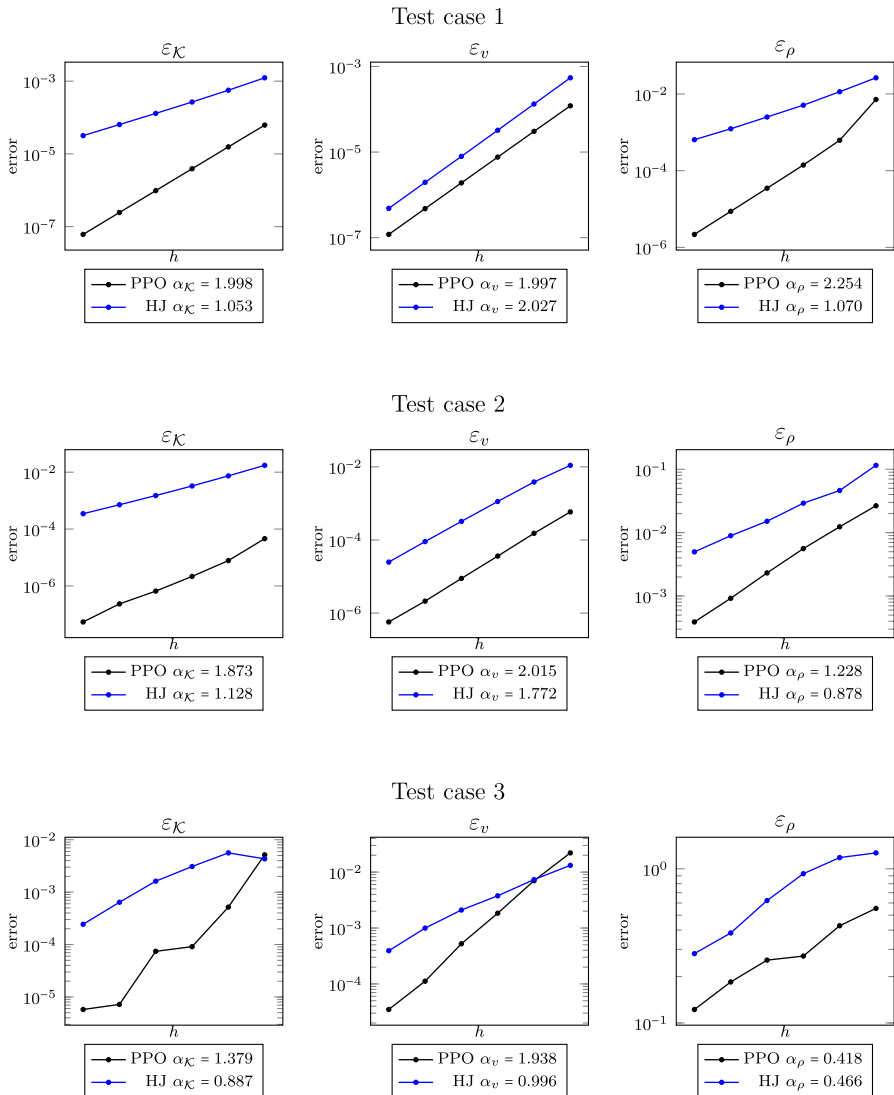


Fig. 3 Plots of the grid resolutions and the errors in the log-log domain. PPO denotes the discretization by [42] and HJ denotes our proposed one. The α values are the approximate rates of convergence

are in the situation where Theorem 4.1 applies, but the assumptions of Theorem 5.1 are not satisfied. Though our rates are again not sharp, the convergence of the transport cost seems to be a bit worse than in the first case, between $\sqrt{h}$ and h as can be seen in Fig. 3.

Remark 6.1 As we discussed in Sect. 3.2, the viscosity coefficient ε vanishes as the resolution $h \rightarrow 0$. In the both examples, we see that for moderately small values of

h (equivalently ε) there is a numerical smoothing whose effect becomes weaker as h decreases as in Fig. 2.

Remark 6.2 In both our test cases, we observed that our constraint (3.4), which here reads $|\Delta_+ \Phi^0 / \Delta x| \leq R := \text{Lip}(L, B_{\text{diam}(\Omega)}) = \text{diam}(\Omega) = 1/2$, was effective *i.e.* $\|\bar{\Lambda}_\eta\|_{L^1(\Omega_D)} > 0$. That is, the discrete dual optimal transport problem has no regularizing effect on the potential and we really need to encode it as a constraint (3.4). We can also run numerical computations with a larger R (including ∞) for which the constraint may become ineffective. As mentioned in Remark 3.4, the scheme may not be monotone when R is too large. In that case, we lose the theoretical guarantee for convergence as the control between a continuous solution and a discrete solution to the Hamilton–Jacobi equation (Theorem 2.4) is missing. We, however, have observed no cases so far without convergence of the cost or the optimizers while we lose control of their Lipschitz constants which may be quite large though seems independent of the stepsize. Said otherwise, the method behaves well even when the theory breaks, which happens quite often in the numerical study of dynamic optimal transport.

Remark 6.3 The finite difference discretization by Papadakis, Peyré, and Oudet [42] better performed in all the examples. We speculate that this is because a) this method uses staggered grids *i.e.* vector values and scalar values are stored in different locations, which more accurately discretizes vector quantities than the collocated grids that our discretization relies on, b) the discretization in [42] is symmetric in both time and space while ours is asymmetric in time as mentioned in Remark 3.1, and c) ours regularizes the solution of the Hamilton–Jacobi equation with viscosity.

Note, however, that convergence of the cost or the optimizer in [42] is not guaranteed. Indeed, though it would be possible to write the dual problem to their discretization and it looks like a discretization of the Hamilton–Jacobi equation, it is one for which no convergence is guaranteed due to the lack of properties such as monotonicity.

Acknowledgements The authors would like to thank Chris Wojtan for his continuous support and several interesting discussions. Part of this research was performed during two visits: one of SI to the BIDS research center at Bocconi University, and one of HL to the Institute of Science and Technology Austria. Both host institutions are warmly acknowledged for the hospitality. HL is partially supported by the MUR-Prin 2022-202244A7YL “Gradient Flows and Non-Smooth Geometric Structures with Applications to Optimization and Machine Learning”, funded by the European Union - Next Generation EU. SI is supported in part by ERC Consolidator Grant 101045083 “CoDiNA” funded by the European Research Council.

Funding Open access funding provided by Institute of Science and Technology (IST Austria).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Rémi Abgrall. Numerical discretization of boundary conditions for first order hamilton–jacobi equations. *SIAM Journal on Numerical Analysis*, 41(6):2233–2261, 2003.
2. Jason M Altschuler, Jonathan Niles-Weed, and Austin J Stromme. 2022 Asymptotics for semidiscrete entropic optimal transport. *SIAM Journal on Mathematical Analysis*, 54(2):1718–1741
3. Luigi Ambrosio, Maria Colombo, Guido Philippis, and Alessio Figalli. Existence of Eulerian Solutions to the Semigeostrophic Equations in Physical Space: The 2-Dimensional Periodic Case. *Communications in Partial Differential Equations*, 37, 11 2011.
4. Aymeric Baradat and Hugo Lavenant. Regularized unbalanced optimal transport as entropy minimization with respect to branching brownian motion. *Astérisque*, 2023.
5. Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. *Numerische Mathematik*, 84:375–393, 2000.
6. Jean-David Benamou and Guillaume Carlier. Augmented Lagrangian methods for transport optimization, mean field games and degenerate elliptic equations. *Journal of Optimization Theory and Applications*, 167:1–26, 2015.
7. Jean-David Benamou, Guillaume Carlier, and Maxime Laborde. An augmented Lagrangian approach to Wasserstein gradient flows and applications. *ESAIM: Proceedings and surveys*, 54:1–17, 2016.
8. Jean-David Benamou, Guillaume Carlier, and Filippo Santambrogio. Variational mean field games. *Active Particles, Volume 1: Advances in Theory, Models, and Applications*, pages 141–171, 2017.
9. Jean-David Benamou and Vincent Duval. Minimal convex extensions and finite difference discretization of the quadratic Monge–Kantorovich problem. *European Journal of Applied Mathematics*, 30(6):1041–1078, 2019.
10. Jean-David Benamou, Brittany D Froese, and Adam M Oberman. A viscosity solution approach to the Monge–Ampère formulation of the optimal transportation problem. arXiv preprint [arXiv:1208.4873](https://arxiv.org/abs/1208.4873), 2012.
11. Jean-David Benamou, Brittany D Froese, and Adam M Oberman. 2014 Numerical solution of the optimal transportation problem using the Monge–Ampère equation. *Journal of Computational Physics*, 260:107–126
12. Robert J. Berman. Convergence Rates for Discretized Monge–Ampère Equations and Quantitative Stability of Optimal Transport. *Foundations of Computational Mathematics*, 21(4):1099–1140, 2021.
13. Guillaume Bonnet and Jean-Marie Mirebeau. Monotone discretization of the Monge–Ampère equation of optimal transport. *ESAIM: Mathematical Modelling and Numerical Analysis*, 56(3):815–865, 2022.
14. Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.
15. Stephen Boyd and Lieven Vandenbergh. *Convex Optimization*. Cambridge University Press, March 2004.
16. Yann Brenier and Dmitry Vorotnikov. On optimal transport of matrix-valued measures. *SIAM Journal on Mathematical Analysis*, 52(3):2849–2873, 2020.
17. José A Carrillo, Katy Craig, Li Wang, and Chaozhen Wei 2022 Primal dual methods for Wasserstein gradient flows. *Foundations of Computational Mathematics*, 22:389–443
18. Yingda Cheng and Chi-Wang Shu. A discontinuous Galerkin finite element method for directly solving the Hamilton–Jacobi equations. *Journal of Computational Physics*, 223(1):398–415, 2007.
19. LÉNAÏC Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. Unbalanced optimal transport: Dynamic and Kantorovich formulations. *Journal of Functional Analysis*, 274(11):3090–3123, 2018.
20. Dario Cordero-Erausquin. Sur le transport de mesures périodiques. *Comptes Rendus de l’Académie des Sciences - Series I - Mathematics*, 329(3):199–202, 1999.
21. Michael G. Crandall, Lawrence C Evans, and Pierre-Louis Lions. 1984 Some Properties of Viscosity Solutions of Hamilton–Jacobi Equations. *Transactions of the American Mathematical Society*, 282(2):487–502,
22. Michael G. Crandall and Pierre-Louis Lions. Viscosity Solutions of Hamilton–Jacobi Equations. *Transactions of the American Mathematical Society*, 277(1):1–42, 1983.
23. Michael G. Crandall and Pierre-Louis Lions. Two Approximations of Solutions of Hamilton–Jacobi Equations. *Mathematics of Computation*, 43(167):1–19, 1984.

24. Matthias Erbar, Martin Rumpf, Bernhard Schmitzer, and Stefan Simon. Computation of optimal transport on discrete metric measure spaces. *Numerische Mathematik*, 144(1):157–200, 2020.
25. Lawrence C. Evans. On solving certain nonlinear partial differential equations by accretive operator methods. *Israel Journal of Mathematics*, 36:225–247, 1980.
26. Lawrence C. Evans. *Partial differential equations*. American Mathematical Society, Providence, R.I., 2010.
27. Ivan Gentil, Christian Léonard, and Luigi Ripani. About the analogy between optimal transport and minimal entropy. *Annales de la Faculté des Sciences de Toulouse. Mathématiques. Série 6*, 3:569–600, 2017.
28. Peter Gladbach, Eva Kopfer, and Jan Maas. Scaling limits of discrete optimal transport. *SIAM Journal on Mathematical Analysis*, 52(3):2759–2802, 2020.
29. Romain Hug, Emmanuel Maitre, and Nicolas Papadakis. On the convergence of augmented Lagrangian method for optimal transport between nonnegative densities. *Journal of Mathematical Analysis and Applications*, 485(2):123811, 2020.
30. Stanislav Kondratyev, Léonard Monsaingeon, and Dmitry Vorotnikov. A new optimal transport distance on the space of finite Radon measures. *Adv. Differential Equations*, 21(11/12):1117–1164, 2016.
31. Hugo Lavenant. Unconditional convergence for discretizations of dynamical optimal transport. *Mathematics of Computation*, 90(328):739–786, 2021.
32. Hugo Lavenant, Sebastian Clatici, Edward Chien, and Justin Solomon. Dynamical Optimal Transport on Discrete Surfaces. *ACM Trans. Graph.*, 37(6), dec 2018.
33. Bowen Li and Jun Zou. On a general matrix valued unbalanced optimal transport and its fully discretization: dynamic formulation and convergence framework. arXiv preprint [arXiv:2011.05845](https://arxiv.org/abs/2011.05845), 2020.
34. Bowen Li and Jun Zou. On the convergence of discrete dynamic unbalanced transport models. arXiv preprint [arXiv:2310.09420](https://arxiv.org/abs/2310.09420), 2023.
35. Matthias Liero, Alexander Mielke, and Giuseppe Savaré. Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures. *Inventiones mathematicae*, 211(3):969–1117, 2018.
36. Jan Maas. Gradient flows of the entropy for finite Markov chains. *Journal of Functional Analysis*, 261(8):2250–2292, 2011.
37. Gonzalo Mena and Jonathan Niles-Weed. Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem. *Advances in neural information processing systems*, 32, 2019.
38. Andrea Natale and Gabriele Todeschi. Computation of optimal transport with finite volumes. *ESAIM: Mathematical Modelling and Numerical Analysis*, 55(5):1847–1871, 2021.
39. Andrea Natale and Gabriele Todeschi. A mixed finite element discretization of dynamical optimal transport. *Journal of Scientific Computing*, 91(2):38, 2022.
40. Jonathan Niles-Weed and Francis Bach. Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli*, 25(4A):2620–2648, 2019.
41. Stanley Osher and Chi-Wang Shu. High-order essentially nonoscillatory schemes for Hamilton–Jacobi equations. *SIAM Journal on numerical analysis*, 28(4):907–922, 1991.
42. Nicolas Papadakis, Gabriel Peyré, and Edouard Oudet. Optimal transport with proximal splitting. *SIAM Journal on Imaging Sciences*, 7(1):212–238, 2014.
43. Gabriel Peyré and Marco Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.
44. Elisabeth Rouy. Numerical approximation of viscosity solutions of first-order Hamilton–Jacobi equations with Neumann type boundary conditions. *Mathematical Models and Methods in Applied Sciences*, 02:357–374, 1992.
45. Filippo Santambrogio. *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*. Progress in Nonlinear Differential Equations and Their Applications. Springer International Publishing, 2015.
46. Filippo Santambrogio. Regularity via duality in calculus of variations and degenerate elliptic PDEs. *Journal of Mathematical Analysis and Applications*, 457(2):1649–1674, 2018. Special Issue on Convex Analysis and Optimization: New Trends in Theory and Applications.
47. Filippo Santambrogio and Xu-Jia Wang. Convexity of the support of the displacement interpolation: Counterexamples. *Applied Mathematics Letters*, 58:152–158, 2016.

48. Bernhard Schmitzer, Klaus P. Schäfers, and Benedikt Wirth. Dynamic cell imaging in PET with optimal transport regularization. *IEEE Transactions on Medical Imaging*, 39(5):1626–1635, 2019.
49. Cedric Villani. *Topics in Optimal Transportation*. Graduate studies in mathematics. American Mathematical Society, 2003.
50. Tao Xiong, Chi-Wang Shu, and Mengping Zhang. A priori error estimates for semi-discrete discontinuous Galerkin methods solving nonlinear Hamilton-Jacobi equations with smooth solutions. *International Journal of Numerical Analysis & Modeling*, 10(1), 2013.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.