



Asymptotic analysis of domain decomposition for optimal transport

Mauro Bonafini¹ · Ismael Medina² · Bernhard Schmitzer²

Received: 25 June 2021 / Revised: 17 January 2023 / Accepted: 28 January 2023 /
Published online: 28 February 2023

© The Author(s) 2023

Abstract

Large optimal transport problems can be approached via domain decomposition, i.e. by iteratively solving small partial problems independently and in parallel. Convergence to the global minimizers under suitable assumptions has been shown in the unregularized and entropy regularized setting and its computational efficiency has been demonstrated experimentally. An accurate theoretical understanding of its convergence speed in geometric settings is still lacking. In this article we work towards such an understanding by deriving, via Γ -convergence, an asymptotic description of the algorithm in the limit of infinitely fine partition cells. The limit trajectory of couplings is described by a continuity equation on the product space where the momentum is purely horizontal and driven by the gradient of the cost function. Global optimality of the limit trajectories remains an interesting open problem, even when global optimality is established at finite scales. Our result provides insights about the efficiency of the domain decomposition algorithm at finite resolutions and in combination with coarse-to-fine schemes.

Mathematics Subject Classification 65K10 · 49M27

✉ Bernhard Schmitzer
schmitzer@cs.uni-goettingen.de

Mauro Bonafini
mauro.bonafini@univr.it

Ismael Medina
ismael.medina@cs.uni-goettingen.de

¹ Department of Computer Science, University of Verona, Verona, Italy

² Campus Institute Data Science, University of Göttingen, Göttingen, Germany

1 Introduction

1.1 Overview

(Computational) optimal transport. Optimal transport (OT) is an ubiquitous optimization problem with applications in various branches of mathematics, including stochastics, PDE analysis and geometry. Let μ and ν be probability measures over spaces X and Y and let $\Pi(\mu, \nu)$ be the set of transport plans, i.e. probability measures on $X \times Y$ with μ and ν as first and second marginal. Intuitively, for some $\pi \in \Pi(\mu, \nu)$ and measurable $A \subset X$, $B \subset Y$, $\pi(A \times B)$ gives the amount of mass transported from A to B . Further, let $c : X \times Y \rightarrow \mathbb{R}$ be a *cost function*, such that $c(x, y)$ gives the cost of moving one unit of mass from x to y . For a transport plan $\pi \in \Pi(\mu, \nu)$, $\int_{X \times Y} c(x, y) d\pi(x, y)$ then gives the total cost associated with this plan. The Kantorovich optimal transport problem consists of finding the plan with minimal cost,

$$\inf \left\{ \int_{X \times Y} c(x, y) d\pi(x, y) \mid \pi \in \Pi(\mu, \nu) \right\}. \quad (1.1)$$

We refer to the monographs [23, 26] for a thorough introduction and historical context. Due to its geometric intuition and robustness it is becoming particularly popular in data analysis and machine learning. Therefore, the development of efficient numerical methods is of immense importance, and considerable progress was made in recent years, such as solvers for the Monge–Ampère equation [5], semi-discrete methods [17, 19], entropic regularization [12], and multi-scale methods [20, 25]. An introduction to computational optimal transport, an overview on available efficient algorithms, and applications can be found in [21].

Domain decomposition. Benamou introduced a domain decomposition algorithm for Wasserstein-2 optimal transport on \mathbb{R}^d [4], based on Brenier’s polar factorization [9]. The case of entropic transport was studied in [8]. The algorithm works as follows: X is divided into two ‘staggered’ partitions $\{X_J \mid J \in \mathcal{J}_A\}$ and $\{X_{\hat{J}} \mid \hat{J} \in \mathcal{J}_B\}$. In the first iteration, an initial coupling π^0 is optimized separately on the cells $X_J \times Y$ for $J \in \mathcal{J}_A$, yielding π^1 . Then π^1 is optimized separately on the cells $X_{\hat{J}} \times Y$ for $\hat{J} \in \mathcal{J}_B$, yielding π^2 . Subsequently, one continues alternating optimizing on the two partitions. This is illustrated in the first row of Fig. 1, a pseudo code formulation is given in Algorithm 1. In each iteration the problems on the individual cells can be solved in parallel, thus making the algorithm amenable for large-scale parallelization.

In [4] it was shown that the algorithm converges to the global minimizer of (1.1) for X, Y being bounded subsets of \mathbb{R}^d , μ being Lebesgue-absolutely continuous and $c(x, y) = \|x - y\|^2$, if each partition contains two cells that satisfy a ‘convex overlap principle’ which roughly requires that a function $f : X \rightarrow \mathbb{R}$ which is convex on each of the cells of \mathcal{J}_A and \mathcal{J}_B must be convex on X . If one is careful about consistency around ‘cycles’ in the partition, the proof can be extended to more general partitions with more cells.

Algorithm 1 Domain decomposition for optimal transport [8, Algorithm 1]. See Figure 1 for illustration and Section 2 for full notation.

```

Input: initial coupling  $\pi_{\text{init}} \in \Pi(\mu, \nu)$ 
Output: a sequence  $(\pi^k)_k$  of feasible couplings in  $\Pi(\mu, \nu)$ 

1:  $\pi^0 \leftarrow \pi_{\text{init}}$ 
2:  $k \leftarrow 0$ 
3: loop
4:    $k \leftarrow k + 1$ 
5:   if ( $k$  is odd) then  $\mathcal{J}^k \leftarrow \mathcal{J}_A$  else  $\mathcal{J}^k \leftarrow \mathcal{J}_B$  // select the partition
6:   for all  $J \in \mathcal{J}^k$  do // iterate over each composite cell
7:      $\mu_J^k \leftarrow P_X(\pi^{k-1} \llcorner (X_J \times Y))$  // get X-marginal on cell
8:      $\nu_J^k \leftarrow P_Y(\pi^{k-1} \llcorner (X_J \times Y))$  // get Y-marginal on cell
9:      $\pi_J^k \leftarrow \arg \min \left\{ \int_{X_J \times Y} c \, d\pi + \varepsilon \text{KL}(\pi | \mu_J \otimes \nu) \mid \pi \in \Pi(\mu_J, \nu_J^k) \right\}$  // get best local coupling for local marginals
10:  end for
11:   $\pi^k \leftarrow \sum_{J \in \mathcal{J}^k} \pi_J^k$  // combine local couplings
12: end loop
    
```

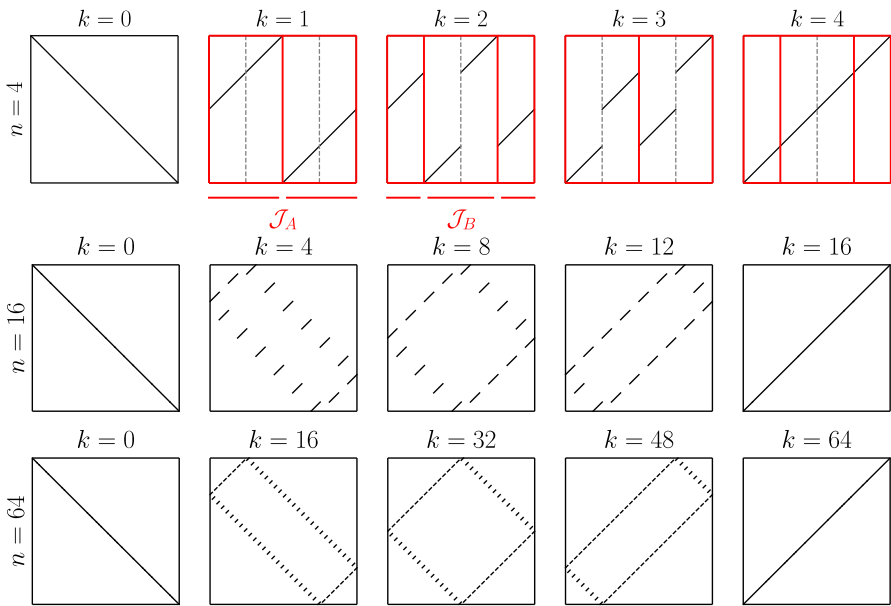


Fig. 1 Iterations of the domain decomposition algorithm for $X = Y = [0, 1]$, $\mu = \nu = \mathcal{L}_\llcorner[0, 1]$, $c(x, y) = (x - y)^2$ and a ‘flipped’ initialization (the diagonal plan is optimal, we start with the flipped ‘anti-diagonal’), for several resolution levels. At each level X is divided into n equal cells, which are then grouped into two staggered partitions \mathcal{J}_A and \mathcal{J}_B . Partition cells of \mathcal{J}_A and \mathcal{J}_B are shown in red in the first row. As the number of cells n is increased, the trajectories of the algorithm seem to converge to an asymptotic limit where each iteration corresponds to a time-step of size $1/n$

In [8] convergence to the minimizer for the entropic setting was shown under rather mild conditions: c needs to be bounded and the two partitions need to be ‘connected’, indicating roughly that it is possible to traverse X by jumping between overlapping

cells of \mathcal{J}_A and \mathcal{J}_B . Convergence was shown to be linear in the Kullback–Leibler (KL) divergence. In addition, an efficient numerical implementation with various features such as parallelization, coarse-to-fine optimization, adaptive sparse truncation and gradual reduction of the regularization parameter was introduced and its favourable performance was demonstrated on numerical examples. The convergence mechanism used in the proof is based on the entropic smoothing and the obtained convergence rate is exponentially slow as regularization goes to zero. It was shown to be approximately accurate on carefully designed worst-case problems.

For the squared distance cost, as considered by Benamou, the algorithm was observed to converge much faster in the experiments of [8]. In combination with the coarse-to-fine scheme even a logarithmic number of iterations (in the image pixel number) was sufficient. In these cases the dominating mechanism that drives convergence seems to be the geometric structure of the squared distance cost. This is not captured accurately by the convergence analysis of [8] and there are no rate estimates given in [4]. The relation of the one-dimensional case to the odd-even-transposition sort was discussed in [8], but the argument does not extend to higher dimensions. An accurate description of this setting is therefore still missing.

Asymptotic dynamic of the Sinkhorn algorithm. The celebrated Sinkhorn algorithm has advanced to an ubiquitous numerical method for optimal transport by means of entropic regularization [12, 21]. Linear convergence of the algorithm in Hilbert’s projective metric is established in [13]. As in [8] the convergence analysis of [13] is solely based on the entropic smoothing and the convergence rate tends to 1 exponentially as regularization decreases (in fact, the former article was inspired by the latter). It was observed numerically (e.g. [24]) that for the squared distance (and similar strictly convex increasing functions of the distance) the Sinkhorn algorithm tends to converge much faster, in particular with appropriate auxiliary techniques such as coarse-to-fine optimization and gradual reduction of the regularization parameter.

In [6] the asymptotic dynamic of the Sinkhorn algorithm for the squared distance cost on the torus is studied in the joint limit of decreasing regularization and refined discretization. The dynamic is fully characterized by the evolution of the dual variables (corresponding to the scaling factors in the Sinkhorn algorithm) which were shown to converge towards the solution of a parabolic PDE of Monge–Ampère type. This PDE had already been studied by [15, 16] and thus allowed estimates on the required number of iterations of the Sinkhorn algorithm for convergence within a given accuracy. This bound is much more accurate for the squared distance, providing a theoretical explanation for the efficiency of numerical methods.

1.2 Contribution and outline

Motivation. For transport problems with squared distance, domain decomposition was empirically shown to be a robust and efficient numerical method for optimal transport, amenable for large-scale parallelization. We expect this to generalize to other costs that are strictly convex increasing functions of the distance. But an accurate theoretical analysis of the convergence speed for this setting is still missing. This article provides some steps in this direction. In this article we work towards such a convergence analysis

in the asymptotic limit as the number of partition cells tends to ∞ . The conjecture for the existence of the asymptotic limit is motivated by Fig. 1 (and additional illustrations throughout the article). The asymptotic regime is also motivated by experiments in [8] where problems on images of pixel size 2048×2048 were solved on a very fine partition into 256×256 cells (of 8×8 pixels each). Finally, an asymptotic analysis in [6] provided a very elegant and novel interpretation of the behaviour of the Sinkhorn algorithm and we hope to establish a similar perspective for the domain decomposition algorithm.

Preview of the main result. For simplicity we consider the case $X = [0, 1]^d$, $Y \subset \mathbb{R}^d$ compact and $c \in C^1(X \times Y)$ with partition cells of X being staggered regular d -dimensional cubes (see Fig. 1 for an illustration in one dimension). At discretization scale n , during iteration k , the domain decomposition algorithm applied to (1.1) requires the solution of the cell problem

$$\inf \left\{ \int_{X_J^n \times Y} c \, d\pi + \varepsilon^n \cdot \text{KL}(\pi | \mu \otimes \nu) \mid \pi \in \Pi(\mu_J^n, \nu_J^{n,k}) \right\}. \quad (1.2)$$

Here, X_J^n is a cell of the relevant partition \mathcal{J}_A^n or \mathcal{J}_B^n (depending on k), ε^n is the entropic regularization parameter at scale n (we consider the cases $\varepsilon^n = 0$ and $\varepsilon^n > 0$, and in the latter case a dependency on n will turn out to be essential), μ_J^n is the restriction of μ to X_J^n and $\nu_J^{n,k}$ is the Y -marginal of the previous iterate $\pi^{n,k-1}$, restricted to $X_J^n \times Y$. For each n this generates a sequence of iterates $(\pi^{n,k})_k$, which we interpret as time-dependent piecewise constant trajectories $\mathbb{R}_+ \ni t \mapsto \pi_t^n := \pi^{n, \lfloor n \cdot t \rfloor}$. That is, at scale n , one iteration corresponds to a time-step $1/n$.

Remark 1 (*Convergence assumption*) In this article, we will assume that there is a limit trajectory $\mathbb{R}_+ \ni t \mapsto \pi_t$ such that (up to a smoothing step) the disintegration of the discrete trajectories $\pi_{t,x}^n$ against the X -marginal converges weak* to $\pi_{t,x}$ for almost all (t, x) . This is of course stronger than mere weak* convergence of the whole trajectories π^n to a limit π (which follows relatively easily from standard compactness arguments). This strong convergence is required for a meaningful pointwise limit of the domain decomposition scheme.

The validity of the assumption is studied in more detail in a longer version of this manuscript that is available at <https://arxiv.org/abs/2106.08084v1>. A notion of bounded total variation for metric space-valued functions, as introduced by Ambrosio in [1], can be shown to be sufficient for the assumption. It can fail when the initial coupling π^0 has unbounded variation in this particular sense. This is however a rather pathological setting that is not relevant in practice. Otherwise, by a monotonicity argument the assumption can be shown to hold in the unregularized one-dimensional case. The argument does not extend to higher dimensions and entropic regularization (although we expect that entropic regularization has a damping effect on the oscillations), so in general it is not yet clear whether oscillations can grow in an uncontrolled way during iterations (and as n increases). A potential counter example is illustrated in the extended manuscript, but it is unclear whether or not its oscillations are really growing without bound as the numerical resolution increases.

The necessity of the smoothing step can be seen from Figs. 1 and 3: within the partition cells the iterates $\pi^{n, \lfloor n \cdot t \rfloor}$ are very oscillatory, but after an averaging over the cells a much more stable and smoother behaviour seems to emerge.

Disintegration notation and smoothing are detailed in Sect. 3.2, the desired convergence is illustrated in Fig. 5. The assumption is formally stated in Sect. 4, Assumption 1.

Under this assumption we will show in this article that there is a momentum field $\mathbb{R}_+ \ni t \mapsto \omega_t \in \mathcal{M}(X \times Y)^d$ such that π_t and ω_t solve a ‘horizontal’ continuity equation on $X \times Y$,

$$\partial_t \pi_t + \operatorname{div}_X \omega_t = 0 \tag{1.3}$$

for $t \geq 0$ with an initial-time boundary condition, in a distributional sense. Here div_X is the divergence of vector fields on $X \times Y$ that only have a ‘horizontal’ component along X . We find that $\omega_t \ll \pi_t$ and the velocity $v_t := \frac{d\omega_t}{d\pi_t}$ has entries bounded by 1, i.e. mass moves at most with unit speed along each spatial axis, corresponding to the fact that particles can at most move by one cell per iteration.

The momentum field ω_t is generated from a family of measures $(\lambda_{t,x})_{x \in X}$ which are minimizers of

$$\inf \left\{ \int_{Z \times Y} \langle \nabla_X c(x, y), z \rangle d\lambda(z, y) + \eta \cdot \operatorname{KL}(\lambda | \sigma \otimes \pi_{t,x}) \mid \lambda \in \Pi(\sigma, \pi_{t,x}) \right\}. \tag{1.4}$$

which can be shown to be the Γ -limit of problem (1.2), which can be anticipated by a careful comparison of the two problems: $Z = [-1, 1]^d$ represents the asymptotic infinitesimal partition cell X^n_j (blown up by a factor n), we find that the transport cost is linearly expanded in X -direction by the gradient, $\eta := \lim_{n \rightarrow \infty} \varepsilon^n \cdot n$ is the asymptotic entropic contribution (which we assume to be finite for now, but the case $\eta = \infty$ is also discussed), σ is the asymptotic infinitesimal restriction of μ to the partition cells (which may be the Lebesgue measure on Z , but we may also obtain different measures if μ is approximated by discretization) and $\pi_{t,x}$ is the disintegration of π_t with respect to the X marginal at x , which corresponds to the asymptotic infinitesimal Y -marginal of π_t , when restricted to the ‘point-like’ cell at x . It is this pointwise Γ -convergence that requires the particular notion of convergence of the trajectories $(\pi_t^n)_n$.

In one dimension, the disintegration $\omega_{t,x}$ of ω_t is then obtained from $\lambda_{t,x}$ via

$$\omega_{t,x}(A) := \lambda_{t,x}(\{z > 0\} \times A) - \lambda_{t,x}(\{z < 0\} \times A) \tag{1.5}$$

for measurable $A \subset Y$. That is, particles sitting in the left half of the cell ($z < 0$) move left with velocity -1 , particles in the right half move right with velocity $+1$.

In a nutshell, the limit of the trajectories generated by the domain decomposition algorithm is described by a flow field which is generated by a limit version of the algorithm.

Compared to the Sinkhorn algorithm studied in [6], the state of the domain decomposition algorithm cannot be described by a scalar potential $X \rightarrow \mathbb{R}$, but requires

the full (generally non-deterministic) coupling $\pi^{n,k}$. Consequently, the limit system (1.3)–(1.5) is not ‘merely’ PDE for a scalar function but formally a non-local PDE for a measure. This system has not been studied previously. Consequently, after having established the convergence to this system, we cannot use existing results to conclude our convergence analysis. Instead, the behaviour of the limit system remains an open question.

Some implications. However, from the result one can already deduce that as n increases, the number of iterations required to approximate the asymptotic stationary state of the algorithm (which need not be a global minimizer) increases linearly in n , which is much faster than the exponential bound in [8].

This also sheds preliminary light on the efficiency of the coarse-to-fine approach of [8]. Assume that from an iterate $\pi^{n,k}$ at scale n an approximation of the iterate $\pi^{2n,2k}$ at scale $2n$ can be obtained by refinement and that the approximation error can be remedied by running a finite, fixed number of additional iterations at scale $2n$. Then, by starting at low n , and then repeatedly running a fixed number of iterations, and refining to $2n$, one can obtain an approximation of the limit point π_t in a number of iterations that is logarithmic in t . This is in agreement with the numerical results of [8].

Open questions. The results of this article do not yet give a complete picture of the dynamics but raise several intriguing follow-up questions.

- We already discussed Assumption 1 and whether one can give sufficient conditions for it that can be checked a priori (see Remark 1).
- We conjecture that under suitable conditions the limit trajectory $t \mapsto \pi_t$ converges to a stationary coupling π_∞ as $t \rightarrow \infty$ and that this coupling is concentrated on the graph of a map (for suitable cost functions c , for instance of ‘McCann-type’ [14], $c(x, y) = h(x - y)$ for strictly convex $h : \mathbb{R}^d \rightarrow \mathbb{R}$).
- Under what conditions is π_∞ a minimizer of the transport problem? Several counter-examples with different underlying mechanisms that prevent optimality are presented or discussed in this article:
 - In the discretized case, with a fixed number of points per basic cell and with $\varepsilon^n = 0$, local optimality on the composite cells does not necessarily induce global optimality of the problem (Sect. 5.1).
 - In this case, for $\varepsilon^n > 0$, convergence to the global minimizer was shown in [8] for each fixed n . But we conjecture that if $\varepsilon^n \rightarrow 0$ too fast, the asymptotic trajectory may still get stuck in a sub-optimal position.
 - Conversely, if ε^n does not tend to zero sufficiently fast ($\eta = \infty$), asymptotically the algorithm freezes in the initial configuration (Fig. 6, Theorem 3).
 - Finally, even in the non-discretized, unregularized setting, where convergence of the algorithm for finite n follows (essentially) from Benamou’s work [4], the asymptotic trajectory may become stuck in a sub-optimal configuration if the sub-optimality is ‘concentrated on an interface’ and almost all cells are locally optimal (Sect. 5.2).
- In cases where convergence to the minimizer can be established, how fast is this convergence with respect to t ? Via the convergence result this then provides an

estimate on the number of required iterations. Intuitively, domain decomposition (and its asymptotic limit dynamics) resembles a minimizing movement scheme on the set $\Pi(\mu, \nu)$ with respect to a ‘horizontal W_∞ metric’, where particles are only allowed to move in X direction by at most distance 1 along each spatial axis per unit time. Can this interpretation be made rigorous?

Outline. Notation is established in Sect. 2. The detailed setting for the algorithm and the discrete trajectories are introduced in Sect. 3. In Sect. 4 we proof convergence of the cell problems, the continuity equation and state the complete main result, Theorem 3. Some examples that illustrate limit cases are presented in Sect. 5. Conclusion and open questions are given in Sect. 6.

2 Background

2.1 Notation and setting

- Let $\mathbb{R}_+ := [0, \infty)$.
- Let $X := [0, 1]^d$, Y be a compact subset of \mathbb{R}^d . We assume compactness to avoid overly technical arguments while covering the numerically relevant setting. We conjecture that the results of this paper can be generalized to compact X with Lipschitz boundaries.
- For a metric space Z denote by $\mathcal{M}(Z)$ the σ -finite measures over Z . If Z is compact, then measures in $\mathcal{M}(Z)$ are finite. Further, denote by $\mathcal{M}_+(Z)$ the subset of non-negative σ -finite measures and by $\mathcal{M}_1(Z)$ the subset of probability measures.
- The Lebesgue measure of any dimension is denoted by \mathcal{L} . The dimension will be clear from context.
- For $\rho \in \mathcal{M}_+(Z)$ and a measurable $S \subset Z$ we denote by $\rho \llcorner S$ the restriction of ρ to S .
- The maps $P_X : \mathcal{M}_+(X \times Y) \rightarrow \mathcal{M}_+(X)$ and $P_Y : \mathcal{M}_+(X \times Y) \rightarrow \mathcal{M}_+(Y)$ denote the projections of measures on $X \times Y$ to their marginals, i.e.

$$(P_X \pi)(S_X) := \pi(S_X \times Y) \quad \text{and} \quad (P_Y \pi)(S_Y) := \pi(X \times S_Y)$$

for $\pi \in \mathcal{M}_+(X \times Y)$, $S_X \subset X$, $S_Y \subset Y$ measurable. We will use the projection notation analogously for other product spaces.

- For a compact metric space Z and $\mu \in \mathcal{M}(Z)$, $\nu \in \mathcal{M}_+(Z)$ the *Kullback–Leibler divergence* (or relative entropy) of μ with respect to ν is given by

$$\text{KL}(\mu|\nu) := \begin{cases} \int_Z \varphi \left(\frac{d\mu}{d\nu} \right) d\nu & \text{if } \mu \ll \nu, \mu \geq 0, \\ +\infty & \text{else,} \end{cases} \quad \text{with}$$

$$\varphi(s) := \begin{cases} s \log(s) - s + 1 & \text{if } s > 0, \\ 1 & \text{if } s = 0, \\ +\infty & \text{else.} \end{cases}$$

2.2 Optimal transport

For $\mu \in \mathcal{M}_+(X)$, $\nu \in \mathcal{M}_+(Y)$ with the same mass, denote by

$$\Pi(\mu, \nu) := \left\{ \pi \in \mathcal{M}_+(X \times Y) \mid P_X \pi = \mu, P_Y \pi = \nu \right\} \quad (2.1)$$

the set of *transport plans* between μ and ν . Note that $\Pi(\mu, \nu)$ is non-empty if and only if $\mu(X) = \nu(Y)$.

Let $c \in \mathcal{C}(X \times Y)$ and $\varepsilon \in \mathbb{R}_+$. Pick $\hat{\mu} \in \mathcal{M}_+(X)$ and $\hat{\nu} \in \mathcal{M}_+(Y)$ such that $\mu \ll \hat{\mu}$ and $\nu \ll \hat{\nu}$. The (entropic) optimal transport problem between μ and ν with respect to the cost function c , with regularization strength ε and with respect to the reference measure $\hat{\mu} \otimes \hat{\nu}$ is given by

$$\inf \left\{ \int_{X \times Y} c(x, y) d\pi(x, y) + \varepsilon \text{KL}(\pi | \hat{\mu} \otimes \hat{\nu}) \mid \pi \in \Pi(\mu, \nu) \right\}. \quad (2.2)$$

Often one simply considers $\hat{\mu} = \mu$ and $\hat{\nu} = \nu$, but other choices are admissible, and one can also replace $\hat{\mu} \otimes \hat{\nu}$ by a more general measure on the product space, as long as μ and ν are absolutely continuous with respect to its marginals.

For $\varepsilon = 0$ this is the (unregularized) Kantorovich optimal transport problem. The existence of minimizers follows from standard compactness and lower-semicontinuity arguments. Of course, more general cost functions (e.g. lower-semicontinuous) can be considered. We refer, for instance, to [23, 26] for in-depth introductions of unregularized optimal transport. Common motivations for choosing $\varepsilon > 0$ are the availability of efficient numerical methods and increased robustness in machine learning applications, see [21] for a broader discussion of entropic regularization. In this article, the above setting is entirely sufficient.

For a compact metric space (Z, d) we denote by W_Z the Wasserstein-1 metric on $\mathcal{M}_1(Z)$ (or more generally, subsets of $\mathcal{M}_+(Z)$ with a prescribed mass). By the Kantorovich–Rubinstein duality [26, Remark 6.5] one has for $\mu, \nu \in \mathcal{M}_+(Z)$, $\mu(Z) = \nu(Z)$,

$$W_Z(\mu, \nu) = \sup_{\phi \in \text{Lip}_1(Z)} \int_Z \phi d(\mu - \nu) \quad (2.3)$$

where $\text{Lip}_1(Z) \subset \mathcal{C}(Z)$ denotes the Lipschitz continuous functions over Z with Lipschitz constant at most 1.

3 Domain decomposition algorithm for optimal transport and discrete trajectories

3.1 Problem setup

In this article, we are concerned with applying the domain decomposition algorithm to (discretizations) of the (possibly entropy regularized) optimal transport problem (2.2),

$$\inf \left\{ \int_{X \times Y} c(x, y) \, d\pi(x, y) + \varepsilon \, \text{KL}(\pi | \mu \otimes \nu) \mid \pi \in \Pi(\mu, \nu) \right\}, \quad (3.1)$$

with increasingly finer cells and with studying its asymptotic behaviour as the cell size tends to zero. Here for simplicity we regularize with respect to $\mu \otimes \nu$, but for the cell problems within Algorithm 1, in line 9, the marginals of the problem differ from those of the regularization measure. In the following we outline the adopted setting and notation for the subsequent analysis.

1. $\mu \in \mathcal{M}_1(X)$, $\nu \in \mathcal{M}_1(Y)$ with $\mu \ll \mathcal{L}$. We assume that $\frac{d\mu}{d\mathcal{L}} > 0$ almost everywhere.

We will build the two composite partitions \mathcal{J}_A and \mathcal{J}_B from the following more basic partition of X , see [8] for more details and Fig. 2 for an illustration.

2. For a discretization level $n \in 2\mathbb{N}$, that we assume *even* for simplicity, the index set I^n for the *basic partition* is given by a uniform Cartesian grid with n points along each axis, with the basic cells $(X_i^n)_{i \in I^n}$ being the corresponding hypercubes between lattice points:

$$I^n := \left\{ (i_1, \dots, i_d) \mid i_1, \dots, i_d \in \{0, \dots, n-1\} \right\} \quad \text{and} \\ X_i^n := i/n + [0, 1/n]^d \quad \text{for } i \in I^n.$$

Remark 2 The set $(X_i^n)_{i \in I^n}$ of closed hypercubes is not a partition of X , since adjacent cubes overlap on their common boundary. However, due to the assumption $\mu \ll \mathcal{L}$, these overlaps do not carry any mass, and may thus be ignored.

3. The *mass* and the *center* of basic cell $i \in I^n$ are given by

$$m_i^n := \mu(X_i^n), \quad x_i^n := n^d \int_{X_i^n} x \, dx = (i + (\frac{1}{2}, \dots, \frac{1}{2}))/n.$$

Note that $m_i^n > 0$ for all $n \in 2\mathbb{N}$ and $i \in I^n$ by positivity of $\frac{d\mu}{d\mathcal{L}}$.

4. At level n we approximate the original marginal μ by $\mu^n \in \mathcal{M}_1(X)$. For example, μ^n could be a discretization of μ . We assume that $\mu^n(X_i^n) = m_i^n$ for each basic cell $i \in I^n$, which in particular implies that $(\mu^n)_n$ converges weak* to μ . We also assume that μ^n assigns no mass to any basic cell boundary, so Remark 2 remains applicable. Setting $\mu_i^n := \mu^n \llcorner X_i^n$, it holds $\sum_{i \in I^n} \mu_i^n = \mu^n$. Further regularity conditions on the sequence $(\mu^n)_n$ will be required in Definition 4.
5. Analogously, let $(\nu^n)_n$ be a sequence in $\mathcal{M}_1(Y)$, converging weak* to ν , and $(\pi_{\text{init}}^n)_n$ a sequence in $\mathcal{M}_1(X \times Y)$ with $\pi_{\text{init}}^n \in \Pi(\mu^n, \nu^n)$, converging weak* to some $\pi_{\text{init}} \in \Pi(\mu, \nu)$. Again, ν^n can slightly differ from ν to allow for potential discretization or approximation steps. There are various ways how a corresponding sequence π_{init}^n could be generated, for instance via an adaptation of the *block approximation* [10] from some $\pi_{\text{init}} \in \Pi(\mu, \nu)$.
6. The cells of the composite partition \mathcal{J}_A^n are generated by forming groups of 2^d adjacent basic cells; the cells of \mathcal{J}_B^n are generated analogously, but with an offset

of 1 basic cell in every direction. (Of course, composite cells may contain less basic cells at the boundaries of X). Then we set

$$X_J^n := \bigcup_{i \in J} X_i^n, \quad \mu_J^n := \sum_{i \in J} \mu_i^n = \mu^n \llcorner X_J^n,$$

for $J \in \mathcal{J}_A^n$ or $J \in \mathcal{J}_B^n$. Again, Remark 2 remains applicable. See again Fig. 2 for an illustration.

7. The *mass* of a composite cell J is $m_J^n := \sum_{i \in J} m_i^n$. For an A (resp. B) composite cell J , we will define its center x_J^n as the unique point on the regular grid

$$\left\{ \frac{1}{n}, \frac{3}{n}, \dots, \frac{n-1}{n} \right\}^d, \quad \left(\text{respectively } \left\{ 0, \frac{2}{n}, \dots, \frac{n-2}{n}, 1 \right\}^d \right), \quad (3.2)$$

that is contained in X_J^n . For A composite cells and B composite cells that do not lie at the boundary of X , it coincides with the average of the centers of their basic cells.

8. Two distinct composite cells $J \in \mathcal{J}_A^n$ and $\hat{J} \in \mathcal{J}_B^n$ are said to be *neighboring* if they share a basic cell. The set of neighbouring composite cells for a given composite cell J is denoted by $\mathcal{N}(J)$. By construction, the shared basic cell is unique, and we denote it by $i(J, \hat{J})$. For compactness, instead of writing, for instance, $m_{i(J, \hat{J})}^n$, we often merely write $m_{J, \hat{J}}^n$.
9. Two composite cells $J, \hat{J} \in \mathcal{J}_A^n$ or $J, \hat{J} \in \mathcal{J}_B^n$ are *adjacent* if the sets X_J^n and $X_{\hat{J}}^n$ share a boundary.
10. The *transport cost function* c is some C^1 function on $X \times Y$.
11. The entropic *regularization parameter* at level n is $\varepsilon^n \in [0, \infty)$. While [4] was restricted to $\varepsilon = 0$ and [8] to $\varepsilon > 0$, in this article we consider both cases. We assume that the sequence $(n \cdot \varepsilon^n)_n$ is convergent as $n \rightarrow \infty$ and set

$$\eta := \lim_{n \rightarrow \infty} n \cdot \varepsilon^n, \quad (3.3)$$

where we explicitly allow for the case $\eta = \infty$.

Now, for given $n \in 2\mathbb{N}$, we apply Algorithm 1 to the discrete problem

$$\inf \left\{ \int_{X \times Y} c(x, y) \, d\pi(x, y) + \varepsilon^n \text{KL}(\pi | \mu^n \otimes \nu^n) \mid \pi \in \Pi(\mu^n, \nu^n) \right\}, \quad (3.4)$$

where we use the composite partitions \mathcal{J}_A^n and \mathcal{J}_B^n and the initial coupling π_{init}^n .

12. At scale n , the k -th *iterate* will be denoted by $\pi^{n,k}$ for $k \in \mathbb{N}$ and one has $\pi^{n,k} \in \Pi(\mu^n, \nu^n)$. The composite partition used during that iteration will be denoted by $\mathcal{J}^{n,k}$, which is either \mathcal{J}_A^n or \mathcal{J}_B^n , depending on whether k is odd or even.

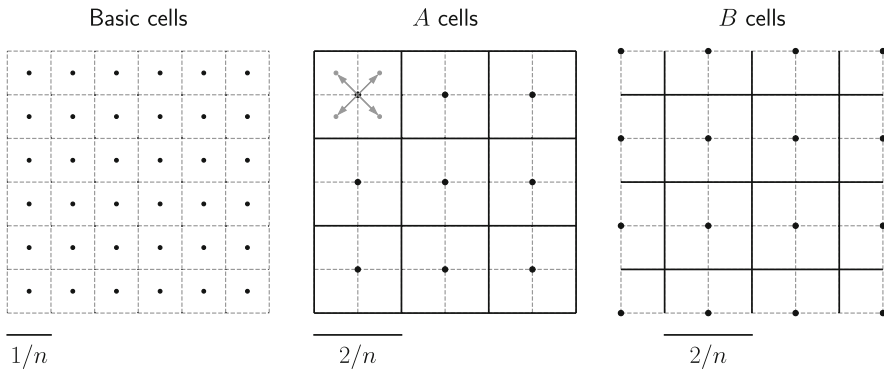


Fig. 2 Close-up of the decomposition of $X = [0, 1]^2$ into basic and composite cells. Left, basic cells and their centers. Center and right, respectively A and B composite cells and their centers. The top left A composite cell shows that the basic cell center of cell $j \in J$ is of the form $x_j^n + b/2n$, for some $b \in \{-1, 1\}^d$

13. Based on line 8 of Algorithm 1 we introduce, for any given $k \geq 0$, the partial Y -marginals of the iterates $\pi^{n,k}$ when restricted to basic or composite cells:

$$v_i^{n,k} := P_Y(\pi^{n,k} \llcorner (X_i^n \times Y)) \quad \text{for } i \in I^n,$$

$$v_J^{n,k} := \sum_{i \in J} v_i^{n,k} = P_Y(\pi^{n,k} \llcorner (X_J^n \times Y)) \quad \text{for } J \in \mathcal{J}_A^n \cup \mathcal{J}_B^n.$$

From Algorithm 1, lines 8 and 9, since we are optimizing among couplings with fixed marginals, we get

$$P_Y(\pi^{n,k} \llcorner (X_J^n \times Y)) = P_Y(\pi^{n,k-1} \llcorner (X_J^n \times Y)) \quad \text{for } J \in \mathcal{J}^{n,k}$$

and we conclude that composite cell marginals are preserved during an iteration:

$$v_J^{n,k} = v_J^{n,k-1} \quad \text{for } J \in \mathcal{J}^{n,k}. \tag{3.5}$$

3.2 Discrete trajectories and momenta

At discretization level $n \in 2\mathbb{N}$, we now associate one iteration with a time-step of $\Delta t = 1/n$, i.e. iterate k is associated with the time $t = k/n$. Loosely speaking, we now want to consider the family of trajectories $(\mathbb{R}_+ \ni t \mapsto \pi^{n, \lfloor t \cdot n \rfloor})_{n \in 2\mathbb{N}}$ and then study their asymptotic behaviour as $n \rightarrow \infty$. However, we find that the measures $\pi^{n,k}$ can oscillate very strongly at the level of basic cells, allowing at best for weak* convergence to some limit coupling, cf. Fig. 3, top. In contrast, our hypothesis for the dynamics of the limit trajectory requires a stronger ‘fiber-wise’ convergence of the disintegration against the X -marginal. We observe that the oscillations in the couplings become much weaker if we average the couplings $\pi^{n,k}$ at the level of composite cells first, cf. Fig. 3, bottom. This averaging is merely intended as an instrument for theoretical analysis of

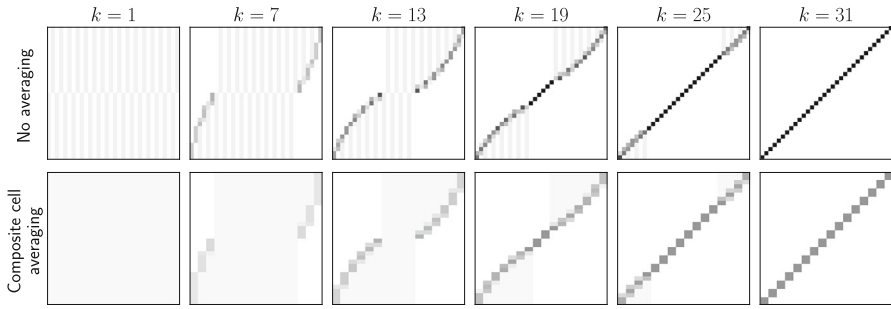


Fig. 3 Top, discrete iterates $\pi^{n,k}$ for $n = 32$, $\mu^n = \nu^n$ discretized Lebesgue, $\pi_{\text{init}}^n = \mu^n \otimes \nu^n$, the quadratic cost, $\varepsilon^n = 0$. Bottom, same iterates but averaged at the composite cell level. Note how the oscillations in space become much weaker

the iterates, not as an actual step in the numerical algorithm. (However, in principle one could add such an averaging step before each iteration to the actual algorithm since the result of each composite cell problem only depends on the Y -marginal of the entire composite cell, and not on the distribution within its constituent basic cells.) Motivated by this we now introduce discrete trajectories of approximate couplings, averaged over composite cells. We rely on the following conventions for notation.

Remark 3 (Disintegration notation and measure trajectories)

- (i) We will represent trajectories of measures on $X \times Y$ as measures on $\mathbb{R}_+ \times X \times Y$ and use bold symbols to denote them. For a non-negative measure $\lambda \in \mathcal{M}_+(\mathbb{R}_+ \times X \times Y)$ with $P_{\mathbb{R}_+} \lambda \ll \mathcal{L}$ we write $(\lambda_t)_{t \in \mathbb{R}_+}$ for its disintegration with respect to \mathcal{L} on the time-axis such that

$$\int_{\mathbb{R}_+ \times X \times Y} \phi(t, x, y) \cdot d\lambda(t, x, y) = \int_{\mathbb{R}_+} \left[\int_{X \times Y} \phi(t, x, y) \cdot d\lambda_t(x, y) \right] dt,$$

for $\phi \in C_c(\mathbb{R}_+ \times X \times Y)$. This disintegration is well defined for σ -finite measures.

- (ii) When $P_X \lambda_t \ll \mu$ for \mathcal{L} -a.e. $t \in \mathbb{R}_+$, we write $(\lambda_{t,x})_{(t,x) \in \mathbb{R}_+ \times X}$ for the disintegration in time and X such that

$$\int_{\mathbb{R}_+ \times X \times Y} \phi(t, x, y) \cdot d\lambda(t, x, y) = \int_{\mathbb{R}_+} \left[\int_X \left[\int_Y \phi(t, x, y) \cdot d\lambda_{t,x}(y) \right] d\mu(x) \right] dt,$$

for $\phi \in C_c(\mathbb{R}_+ \times X \times Y)$.

- (iii) Conversely, a (measurable) family of signed measures $(\lambda_x)_{x \in X}$ in $\mathcal{M}(Y)$ with uniformly bounded variation (i.e. $\sup_{x \in X} |\lambda_x|(Y) < \infty$) can be glued together along X with respect to μ to obtain a measure $\lambda \in \mathcal{M}(X \times Y)$ via

$$\int_{X \times Y} \phi(x, y) d\lambda(x, y) := \int_X \left[\int_Y \phi(x, y) d\lambda_x(y) \right] d\mu(x)$$

for $\phi \in \mathcal{C}(X \times Y)$. We will denote λ as $\mu \otimes \lambda_x$. Similarly, we can glue families over $(t, x) \in \mathbb{R}_+ \times X$ to obtain a σ -finite measure on $\mathbb{R}_+ \times X \times Y$, which we denote by $\mathcal{L} \otimes \mu \otimes \lambda_{t,x}$.

(iv) The above points extend to vector measures by component-wise application.

Definition 1 (*Discrete trajectories and momenta*) The discrete trajectory and momentum, $\pi^n \in \mathcal{M}_+(\mathbb{R}_+ \times X \times Y)$ and $\omega^n \in \mathcal{M}(\mathbb{R}_+ \times X \times Y)^d$ are defined via their disintegration with respect to $\mathcal{L} \otimes \mu$ at $t \in \mathbb{R}_+, x \in X$ as:

$$\begin{aligned} \pi^n_{t,x} &:= \frac{1}{m^n_J} v^{n,k}_J, & \omega^n_{t,x} &:= \frac{1}{m^n_J} \sum_{\hat{j} \in \mathcal{N}(J)} v^{n,k}_{J,\hat{j}} \cdot (x^n_{\hat{j}} - x^n_J) \cdot n \\ &= \frac{1}{m^n_J} \sum_{b \in \{-1,+1\}^d} v^{n,k}_{i(J,b)} \cdot b, \end{aligned} \tag{3.6}$$

where we set $k = \lfloor nt \rfloor$, J is the (μ -a.e. unique) composite cell in $\mathcal{J}^{n,k}$ such that $x \in X^n_J$ and $i(J, b)$ is the basic cell contained in J whose center sits at $x^n_J + b/2n$. The vectors b are illustrated in Fig. 2. For composite B cells at the boundary some of the basic cells $i(J, b)$ might lie outside of X and we ignore the corresponding terms in the sum.

Note that we use the composite cell marginals $v^{n,k}_J$ in the definition of $\pi^n_{t,x}$, hence this implements the aforementioned averaging over composite cells. An intuitive interpretation of the discrete momentum ω^n is that mass in the basic cell $i(J, \hat{j})$ will travel from $x^n_{\hat{j}}$ to x^n_J during iteration k within a time span of $1/n$.

Remark 4 π^n_t is generated from $\pi^{n,\lfloor nt \rfloor}$ by averaging the Y -marginals over the composite cells. That is, by the averaging each mass particle is moved only horizontally, and only within its current composite cell. Therefore, for all $n \in 2\mathbb{N}$ and $t \in \mathbb{R}_+$, it holds $W_{X \times Y}(\pi^n_t, \pi^{n,\lfloor nt \rfloor}) \leq 2\sqrt{d}/n$ where the latter is the diameter of composite cells. Consequently the sequences $(\pi^n_t)_n$ and $(\pi^{n,\lfloor nt \rfloor})_n$ have the same weak* limits or cluster points.

The discrete trajectory π^n_t is illustrated in the bottom row of Fig. 3, and its corresponding momentum ω^n_t is visualized in Fig. 4. We will show in Sect. 4.4 that π^n and ω^n approximately solve a continuity equation on the product space $X \times Y$ in a distributional sense, where ω^n encodes a ‘horizontal’ flow (i.e. only along the X -component). Formally, this can be written as

$$\partial_t \pi^n_t + \operatorname{div}_X \omega^n_t = o(1) \quad \text{for } t > 0 \text{ and } \pi^n_{t=0} \rightarrow \pi_{\text{init}} \text{ as } n \rightarrow \infty \tag{3.7}$$

and it is to be interpreted via integration against test functions in $\mathcal{C}^1_c(\mathbb{R}_+ \times X \times Y)$ where the \mathbb{R}_+ factor corresponds to time t (cf. Proposition 4).

In Sect. 4 we study the limit ω (up to subsequences) of the ω^n and show that it can be constructed from solutions to a problem that is the Γ -limit of the cell-wise domain decomposition problems where the cells have collapsed to single points $x \in X$

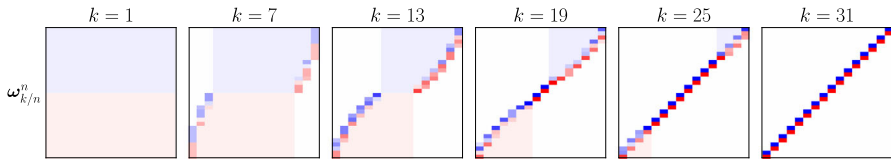


Fig. 4 Discrete momentum field for the trajectory shown in Fig. 3. Blue shade indicates positive velocity (i.e., mass moving towards the right); red shade indicates a negative velocity. The intensity marks the amount of mass that is transported. Note that the momentum does not vanish completely even when the algorithm has converged. This is related to the finite discretization scale. For $n \rightarrow \infty$ we anticipate that the momentum (after convergence of the algorithm) converges weak* (but not in norm) to zero (color figure online)

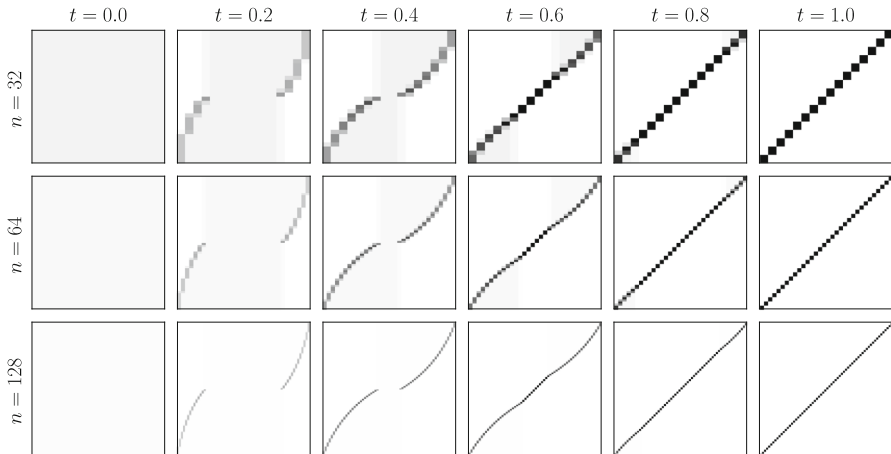


Fig. 5 Comparison of discrete trajectories π^n for increasing n for the same setting as in Fig. 3. It is tempting to conjecture that the disintegrations $\pi_{t,x}^n$ converge weak* for almost all (t, x)

(Proposition 3). The limit pair (π, ω) is an exact solution of the ‘horizontal’ continuity equation (3.7) (Proposition 4). In summary, the limit of trajectories generated by the domain decomposition algorithm can be associated with a limit notion of the domain decomposition algorithm (Theorem 3).

Role of regularization parameter ε^n . We expect that the behaviour of the limit trajectory and momentum (π, ω) depends on the behaviour of the sequence of regularization parameters ε^n . This is motivated by the numerical simulations illustrated in Fig. 6. The upper rows show the evolution under the domain decomposition algorithm on two kinds of initial data, for $d = 1, n = 64, \varepsilon^n = 0$. The setting dubbed ‘flipped’ has again $\mu^n = \nu^n =$ discretized Lebesgue, but the initial plan is the ‘flipped’ version of the optimal (diagonal) plan. The setting named ‘bottleneck’ has as μ^n a measure with piecewise constant density that features a low-density region (the bottleneck) around $x = 0.5$, while ν^n is again discretized Lebesgue and $\pi_{\text{init}}^n = \mu^n \otimes \nu^n$. This bottleneck slows down exchange of mass between the two sides of the X domain, and thus two shocks appear (cf. $t = 0.4$), which slowly merge as mass traverses the bottleneck.

These evolution examples for $\varepsilon^n = 0$ serve as reference for comparison with the regularized cases that are illustrated in the bottom plots of Fig. 6 for fixed t and various

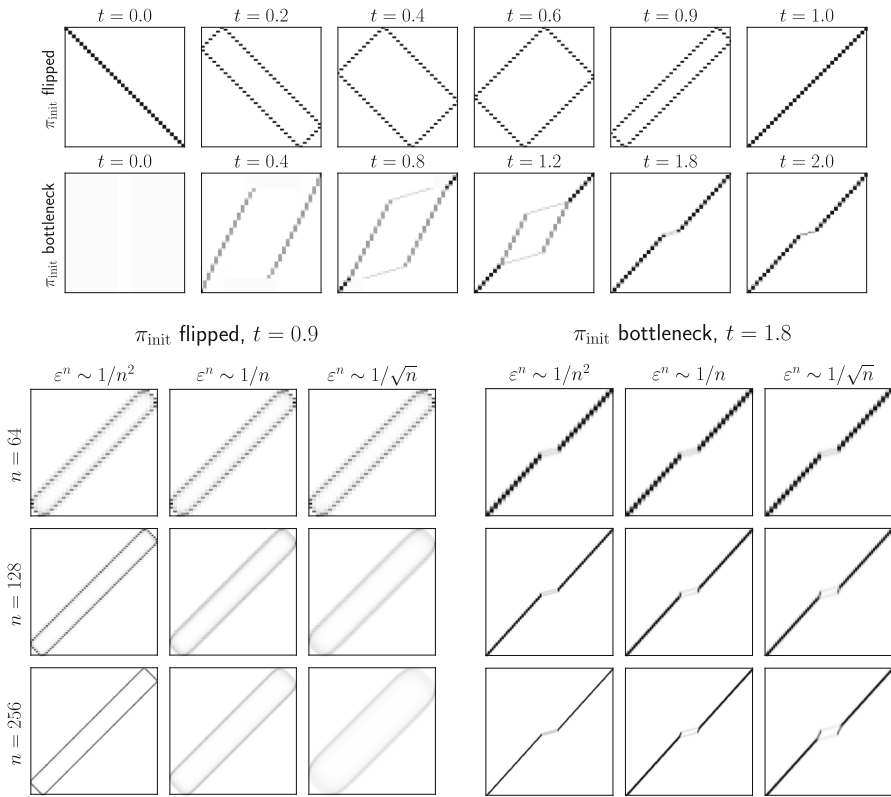


Fig. 6 Top, evolution of the discrete trajectories for π_{init} of type ‘flipped’ and ‘bottleneck’, $\epsilon = 0$. Bottom, snapshot of the discrete trajectories at a fixed time (left $t = 0.9$, right $t = 1.8$), for different values of n and scaling behaviour of the regularization parameter ϵ^n

n . We examine three different ‘schedules’ for the regularization parameter: $\epsilon^n = 2/n^2$ (left), $\epsilon^n = 2/(64n)$ (middle) and $\epsilon^n = 1/(256\sqrt{n})$ (right). Note that in all the schedules the regularization converges to zero. The values were chosen so that the regularization at scale $n = 64$ is the same for all schedules.

- For $\epsilon^n \sim 1/n^2$, the trajectories become increasingly ‘crisp’ and are very close to the unregularized ones.
- For $\epsilon^n \sim 1/n$, the trajectories are slightly blurred and lag a little behind the unregularized ones, but still evolve consistently as $n \rightarrow \infty$.
- For $\epsilon^n \sim 1/\sqrt{n}$ blur and lag increase with n .

The three schedules yield $\eta = \lim_{n \rightarrow \infty} n \cdot \epsilon^n = 0$, $\eta \in (0, \infty)$ and $\eta = +\infty$ respectively, see (3.3). Based on Fig. 6 we conjecture that for $\eta = 0$, the problem describing the limit dynamics of (π, ω) does not contain an entropic term. For $\eta \in (0, \infty)$, there will be an entropic term. For $\eta = \infty$ the entropic term will dominate and the trajectory $t \mapsto \pi_t = \pi_{\text{init}}$ will be stationary. This conjecture will be confirmed in Sect. 4.

4 Asymptotic analysis

In this section we derive the asymptotic description of the domain decomposition trajectories and momenta. It is structured as follows: In Sect. 4.1 we introduce the re-scaled versions of the domain decomposition cell problems which have a meaningful Γ -limit. In Sect. 4.2 we state the supposed limit functional and prepare the proof. Liminf and limsup conditions are provided in Sect. 4.3. The continuity equation is addressed in Sect. 4.4.

The results hinge on the following assumption on the pointwise convergence of the discrete trajectories. For additional context, see Remark 1.

Assumption 1 Assume that there is a trajectory $\pi = \mathcal{L} \otimes \pi_t \in \mathcal{M}_+(\mathbb{R}_+ \times X \times Y)$, $\pi_t \in \Pi(\mu, \nu)$ for a.e. $t \in \mathbb{R}_+$, such that, up to extraction of a subsequence $\mathcal{Z} \subset 2\mathbb{N}$, the discrete trajectories $(\pi^n)_n$, defined in Definition 1, converge for a.e. $t \in \mathbb{R}_+$ and μ -a.e. $x \in X$ to π in W_Y . More precisely,

$$\lim_{n \in \mathcal{Z}, n \rightarrow \infty} W_Y(\pi^n_{t,x}, \pi_{t,x}) = 0, \quad \text{for } \mathcal{L}\text{-a.e. } t \text{ and } \mu\text{-a.e. } x \in X. \tag{4.1}$$

4.1 Re-scaled discrete cell problems

Recall that at resolution $n \in 2\mathbb{N}$, during iteration $k \in \mathbb{N}$, in a composite cell $J \in \mathcal{J}^{n,k}$ we need to solve the following regularized optimal transport problem (Algorithm 1, line 9):

$$\inf \left\{ \int_{X^n_J \times Y} c \, d\pi + \varepsilon^n \cdot \text{KL}(\pi | \mu^n_J \otimes \nu^n) \mid \pi \in \Pi(\mu^n_J, \nu^n_{J,k}) \right\} \tag{4.2}$$

For the limiting procedure we will map X^n_J to a reference hyper-cube $Z = [-1, 1]^d$, normalize the cell marginals μ^n_J and $\nu^n_{J,k}$ (the latter will then become $\pi^n_{k/n,x}$ for $x \in X^n_J$). We will subtract some constant contributions from the transport and regularization terms and re-scale the objective such that the dominating contribution is finite in the limit (the proper scaling will depend on whether η is finite). In addition, as $n \rightarrow \infty$, the cells X^n_J become increasingly finer, we thus expect that we can replace the cost function c by a linear expansion along x . These transformations are implemented in the two following definitions, yielding the functional (4.3). Equivalence with (4.2) is then established in Proposition 1.

Definition 2 We define the *scaled composite reference cell* as $Z = [-1, 1]^d$. Let $n \in \mathcal{Z}$ (recall that \mathcal{Z} is the subsequence from Assumption 1).

- For a composite cell $J \in \mathcal{J}^{n,k}$, $k > 0$, the *scaling map* of cell J is given by

$$S^n_J : X^n_J \rightarrow Z, \quad x \mapsto n(x - x^n_J).$$

- For $J \in \mathcal{J}^{n,k}$, $k > 0$, we define the *scaled X -marginal* as

$$\sigma^n_J := (S^n_J)_\# \mu^n_J / m^n_J \in \mathcal{M}_1(Z).$$

– For $t > 0, x \in X$, let $k = \lfloor tn \rfloor$ and let $J \in \mathcal{J}^{n,k}$ be the μ -a.e. unique composite cell in $\mathcal{J}^{n,k}$ such that $x \in X_J^n$. We will write

$$J_{t,x}^n := J, \quad \bar{x}_{t,x}^n := x_J^n, \quad S_{t,x}^n := S_J^n, \quad \sigma_{t,x}^n := \sigma_J^n.$$

This will allow us to reference more easily between the continuum limit problem in fiber $x \in X$ and its corresponding family of discrete problems at finite scale n .

Definition 3 (*Discrete fiber problem*) For each $n \in 2\mathbb{N}, t \in \mathbb{R}_+$ and $x \in X$, we define the following functional over $\mathcal{M}_1(Z \times Y)$:

$$F_{t,x}^n(\lambda) := \begin{cases} C_{t,x}^n(\lambda) & \text{if } \lambda \in \Pi(\sigma_{t,x}^n, \pi_{t,x}^n), \\ +\infty & \text{otherwise,} \end{cases} \tag{4.3}$$

where for $\eta < \infty$,

$$C_{t,x}^n(\lambda) := \int_{Z \times Y} n \cdot [c(\bar{x}_{t,x}^n + z/n, y) - c(\bar{x}_{t,x}^n, y)] d\lambda(z, y) + n \varepsilon^n \cdot \text{KL}(\lambda | \sigma_{t,x}^n \otimes \pi_{t,x}^n), \tag{4.4}$$

and if $\eta = \infty$,

$$C_{t,x}^n(\lambda) := \text{KL}(\lambda | \sigma_{t,x}^n \otimes \pi_{t,x}^n) + \frac{1}{n \varepsilon^n} \int_{Z \times Y} n \cdot [c(\bar{x}_{t,x}^n + z/n, y) - c(\bar{x}_{t,x}^n, y)] d\lambda(z, y). \tag{4.5}$$

We now establish equivalence between (4.2) and minimizing (4.3).

Proposition 1 (Domain decomposition algorithm generates a minimizer of $F_{t,x}^n$) *Let $n \in 2\mathbb{N}, t > 0, x \in X, k = \lfloor tn \rfloor$ and $J = J_{t,x}^n$. Then problem (4.2) is equivalent to minimizing $F_{t,x}^n$, (4.3), over $\mathcal{M}_1(Z \times Y)$ in the sense that the latter is obtained from the former by a coordinate transformation, a positive re-scaling and subtraction of constant terms. The minimizers $\pi_J^{n,k}$ of (4.2) are in one-to-one correspondence with minimizers $\lambda_{t,x}^n \in \mathcal{M}_1(Z \times Y)$ of (4.3) via the bijective transformation*

$$\lambda_{t,x}^n := (S_J^n, \text{id})_{\#} \pi_J^{n,k} / m_J^n. \tag{4.6}$$

Note that $m_J^n > 0$ is a consequence of $m_i^n > 0$ for all $n \in \mathbb{N}$ and $i \in I^n$ (Sect. 2.1, item 3.).

Proof We subsequently apply equivalent transformations to (4.2) to turn it into (4.3), while keeping track of the corresponding transformation of minimizers. We start with the case $\eta < \infty$.

First, we multiply the objective of (4.2) by n and re-scale the mass of π by $1/m_J^n$ such that it becomes a probability measure. We obtain that (4.2) is equivalent to

$$m_J^n \cdot \inf \left\{ \int_{X_J^n \times Y} (n \cdot c) \, d\pi + n \cdot \varepsilon^n \cdot \text{KL}(\pi | \frac{\mu_J^n}{m_J^n} \otimes \nu^n) \mid \pi \in \Pi(\frac{\mu_J^n}{m_J^n}, \pi_{t,x}^n) \right\}, \tag{4.7}$$

where we used that $v_J^{n,k}/m_J^n = \pi_{t,x}^n$ by (3.6) (and the relation of t, x, n, k and J) and that $\text{KL}(\cdot|\cdot)$ is positively 1-homogeneous under joint re-scaling of both arguments, i.e. $m_J^n \cdot \text{KL}(\frac{\pi}{m_J^n} | \frac{\mu_J^n}{m_J^n} \otimes \nu^n) = \text{KL}(\pi | \mu_J^n \otimes \nu^n)$. Minimizers of (4.7) are obtained from minimizers π of (4.2) as π/m_J^n . The factor m_J^n in front can be ignored.

Second, we transform the cell X_J^n to the reference cell Z via the map S_J^n . For the transport term in (4.7) we find

$$\int_{X_J^n \times Y} (n \cdot c) \, d\pi = \int_{Z \times Y} (n \cdot c) \circ (S_J^n, \text{id})^{-1} \, d(S_J^n, \text{id})_{\#}\pi.$$

Using that (S_J^n, id) is a homeomorphism one gets that

$$\frac{d(S_J^n, \text{id})_{\#}\pi}{d\left((S_J^n)_{\#}\frac{\mu_J^n}{m_J^n} \otimes \nu^n\right)} \circ (S_J^n, \text{id}) = \frac{d\pi}{d\left(\frac{\mu_J^n}{m_J^n} \otimes \nu^n\right)} \left(\frac{\mu_J^n}{m_J^n} \otimes \nu^n\right)\text{-almost everywhere.}$$

With this we can transform the entropy term of (4.7) to

$$\text{KL}(\pi | \frac{\mu_J^n}{m_J^n} \otimes \nu^n) = \text{KL}((S_J^n, \text{id})_{\#}\pi | (S_J^n)_{\#}\frac{\mu_J^n}{m_J^n} \otimes \nu^n) = \text{KL}((S_J^n, \text{id})_{\#}\pi | \sigma_J^n \otimes \nu^n).$$

Finally, using once more that (S_J^n, id) is a homeomorphism one finds

$$\left[\pi \in \Pi\left(\frac{\mu_J^n}{m_J^n}, \pi_{t,x}^n\right) \right] \Leftrightarrow \left[(S_J^n, \text{id})_{\#}\pi \in \Pi(\sigma_J^n, \pi_{t,x}^n) \right].$$

Consequently, (4.7) is equivalent to

$$\inf \left\{ \int_{Z \times Y} (n \cdot c) \circ (S_J^n, \text{id})^{-1} \, d\pi + n \cdot \varepsilon^n \cdot \text{KL}(\pi | \sigma_J^n \otimes \nu^n) \mid \pi \in \Pi(\sigma_J^n, \pi_{t,x}^n) \right\}, \tag{4.8}$$

with minimizers to the latter obtained from minimizers π of the former as $(S_J^n, \text{id})_{\#}\pi$.

Third, observe that $[(n \cdot c) \circ (S_J^n, \text{id})^{-1}](z, y) = n \cdot c(\bar{x}_{t,x}^n + z/n, y)$ and that $\int_{Z \times Y} n \cdot c(\bar{x}_{t,x}^n, y) \, d\lambda(z, y)$ depends just on the second marginal of λ , which is fixed, hence it is a constant contribution. So the transport cost term in (4.8) is equivalent to that in (4.4).

Fourth, we subtract constant parts of the entropy term. Recall that $\pi_{t,x}^n = v_J^{n,k}/m_J^n$ with $\sum_{J' \in \mathcal{J}^{n,k}} v_{J'}^{n,k} = v^n$ and all partial marginals are non-negative. This implies

that $\pi^n_{t,x} \ll v^n$ with the density $\frac{d\pi^n_{t,x}}{dv^n}$ lying in $[0, 1/m^n_J]$ v^n -almost everywhere. Consequently, if $\lambda \ll \sigma^n_J \otimes \pi^n_{t,x}$ then $\lambda \ll \sigma^n_J \otimes v^n$ and the densities satisfy

$$\frac{d\lambda}{d\sigma^n_J \otimes v^n}(z, y) = \frac{d\lambda}{d\sigma^n_J \otimes \pi^n_{t,x}}(z, y) \cdot \frac{d\pi^n_{t,x}}{dv^n}(y) \text{ for } \sigma^n_J \otimes v^n\text{-a.e. } (z, y).$$

Since $P_Y \lambda = \pi^n_{t,x}$ for all feasible $\lambda \in \Pi(\sigma^n_J, \pi^n_{t,x})$, one also has $[\lambda \ll \sigma^n_J \otimes v^n] \Rightarrow [\lambda \ll \sigma^n_J \otimes \pi^n_{t,x}]$ and the same relation between the densities. Using this one finds that when either of the two entropic terms in (4.4) or (4.8) is finite, so is the other one where one has the relation

$$KL(\lambda | \sigma^n_J \otimes v^n) = KL(\lambda | \sigma^n_J \otimes \pi^n_{t,x}) + KL(\pi^n_{t,x} | v^n).$$

Here, the second term on the right hand side is finite (due to the bound on the density $\frac{d\pi^n_{t,x}}{dv^n}$) and does not depend on λ . Hence, the entropic terms in (4.4) and (4.8) are identical up to a constant and in conclusion, for $\eta < \infty$, both minimization problems are equivalent with the prescribed relation between minimizers. The adaption to the case $\eta = \infty$ is trivial since (4.5) is just a positive re-scaling of (4.4). \square

Remark 5 For $\lambda^n_{t,x}$ a minimizer of $F^n_{t,x}$ constructed as in (4.6), the discrete momentum field disintegration $\omega^n_{t,x}$ (3.6) can be written in terms of $\lambda^n_{t,x}$ as:

$$\begin{aligned} (\omega^n_{t,x})_\ell &= P_Y(\lambda^n_{t,x} \llcorner Z^{\ell}_+ \times Y) - P_Y(\lambda^n_{t,x} \llcorner Z^{\ell}_- \times Y) \text{ for } \ell = 1, \dots, d, \text{ with} \\ Z^{\ell}_{\pm} &= \{z \in Z \mid \pm z_\ell > 0\}. \end{aligned} \tag{4.9}$$

To see this, fix $J = J^n_{t,x}$. Then, for each $b \in B := \{-1, +1\}^d$ define $Z_b := \{z \in Z \mid \text{sign}(z_\ell) = b_\ell \text{ for all } \ell = 1, \dots, d\}$. Further define $i(J, b)$ as the basic cell in composite cell J whose center lies at $x^n_J + b/2n$. Then

$$P_Y(\lambda^n_{t,x} \llcorner Z^{\ell}_+ \times Y) - P_Y(\lambda^n_{t,x} \llcorner Z^{\ell}_- \times Y) = \sum_{b \in B} b_\ell \cdot P_Y(\lambda^n_{t,x} \llcorner Z_b \times Y) = \sum_{b \in B} b_\ell \cdot \frac{v^{n,k}_{i(J,b)}}{m^n_{i(J,b)}}, \tag{4.10}$$

which is precisely the ℓ -th component in (3.6).

4.2 Limit fiber problems and problem gluing

In this section we state the expected limit of the discrete fiber problems (4.3) as $n \rightarrow \infty$. For this we need a sufficiently regular sequence of first marginals $(\sigma^n_{t,x})_n$, which will be dealt with in the first part of this section (Definition 4, Assumption 2, Lemma 1). Sufficient regularity of the second marginal constraint will be provided by the pointwise convergence of $(\pi^n)_n$ (Assumption 1). The conjectured limit problem is introduced in Definition 5. Instead of proving Γ -convergence on the level of single

fibers, we first ‘glue’ the problems together (Definition 6) along $t \in \mathbb{R}_+$ and $x \in X$ and then establish Γ -convergence for the glued problems (Proposition 2). This avoids issues with measurability and the selection of convergent subsequences. Finally, from this we can deduce the convergence of the momenta to a suitable limit (Proposition 3).

Definition 4 We say that $(\mu^n)_n$ is a *regular discretization sequence for the X -marginal* if there is some $\sigma \in \mathcal{M}_1(Z)$ such that for $\mathcal{L} \otimes \mu$ almost all $(t, x) \in \mathbb{R}_+ \times X$ the sequence $(\sigma_{t,x}^n)_n$ converges weak* to σ and σ does not give mass to any coordinate axis, i.e.,

$$\sigma(\{z \in Z \mid z_\ell = 0\}) = 0 \quad \text{for } \ell = 1, \dots, d. \tag{4.11}$$

The motivation behind this definition is that mass on the coordinate axes cannot be uniquely assigned to a basic cell (‘left or right’) and thus we desire to avoid this case at each n and also in the limit. The definition clarifies how the discretizations μ^n should be chosen: a malicious or clumsy choice could yield a limit problem where the association of mass to basic cells may not longer be well-defined. The following Lemma states that several canonical ways of choosing μ^n are all regular. The first corresponds to no discretization at all, the second one to simply approximating μ^n by a single Dirac measure on each basic cell, and the third to a discretization with a grid that is refined faster than the partitions, i.e. within each basic cell the measure is approximated by increasingly more Dirac masses.

Lemma 1 (Regularity of discretization schemes) *Prototypical choices for μ^n are:*

- (i) *Using the measure μ itself, without discretization, $\mu^n = \mu$. In this case $\sigma_{t,x}^n = (S_j^n)_\# \mu_j^n / m_j^n$ is a re-scaled restriction of the original measure, still absolutely continuous and one obtains $\sigma = 2^{-d} \cdot \mathcal{L} \llcorner Z$.*
- (ii) *Collapsing all the mass within each basic cell to a Dirac at its center, $\mu^n = \sum_{i \in I^n} m_i^n \delta_{x_i^n}$. In this case, $\sigma_{t,x}^n = (S_j^n)_\# \mu_j^n / m_j^n$ contains one Dirac mass per basic cell and one obtains $\sigma = 2^{-d} \sum_{b \in \{-1, +1\}^d} \delta_{b/2}$.*
- (iii) *At every $n \in 2\mathbb{N}$ we collapse the mass of μ onto Diracs on a regular Cartesian grid such that every basic cell contains a sub-grid of s^n points along each dimension, for a sequence $(s^n)_n$ in \mathbb{N} with $s^n \rightarrow +\infty$. Thus at finite n , $\sigma_{t,x}^n$ is becoming an increasingly finer measure supported on a regular grid and one obtains $\sigma = 2^{-d} \cdot \mathcal{L} \llcorner Z$. A related refined discretization scheme was considered in [4, Section 5.3] where n was kept fixed but s^n was sent to $+\infty$ and it was shown that the sequence of fixed-points of the algorithm converges to the globally optimal solution.*

For $\mu \ll \mathcal{L}$, the above schemes yield regular discretization sequences in the sense of Definition 4. For a regular discretization sequence (not just the examples above) the limit σ assigns mass $1/2^d$ to each of the ‘quadrants’

$$Z_b := \{z \in Z \mid \text{sign}(z_\ell) = b_\ell \text{ for all } \ell = 1, \dots, d\} \quad \text{for } b \in \{-1, 1\}^d. \tag{4.12}$$

The proof is presented in Appendix A. It hinges on $\mu \ll \mathcal{L}$ and the Lebesgue differentiation theorem for L^1 functions [22, Theorem 7.10]. The last part will be relevant for the case $\eta = \infty$.

Assumption 2 From now on, we assume that $(\mu^n)_n$ is a regular discretization sequence.

Definition 5 (*Limiting fiber problem*) For each $t \in \mathbb{R}_+, x \in X$ we define the following functional over $\mathcal{M}_1(Z \times Y)$:

$$F_{t,x}(\lambda) := \begin{cases} C_{t,x}(\lambda) & \text{if } \lambda \in \Pi(\sigma, \pi_{t,x}), \\ +\infty & \text{otherwise,} \end{cases} \tag{4.13}$$

where

$$C_{t,x}(\lambda) := \begin{cases} \int_{Z \times Y} \langle \nabla_X c(x, y), z \rangle d\lambda(z, y) + \eta \cdot \text{KL}(\lambda | \sigma \otimes \pi_{t,x}) & \text{if } \eta < \infty, \\ \text{KL}(\lambda | \sigma \otimes \pi_{t,x}) & \text{if } \eta = \infty. \end{cases} \tag{4.14}$$

Definition 6 (*Glued problems (discrete and limiting)*) Fix $T > 0$. We define

$$\mathcal{V}_T := \{ \lambda \in \mathcal{M}_+([0, T] \times X \times Z \times Y) \mid P_{(\mathbb{R}_+ \times X)} \lambda = (\mathcal{L}_\perp[0, T]) \otimes \mu \}, \tag{4.15}$$

where $P_{(\mathbb{R}_+ \times X)}$ takes non-negative measures on $\mathbb{R}_+ \times X \times Z \times Y$ to their marginal on $\mathbb{R}_+ \times X$ (cf. Sect. 2.1). In particular, any $\lambda \in \mathcal{V}_T$ can be disintegrated with respect to $\mathcal{L} \otimes \mu$, i.e. there is a measurable family of probability measures $(\lambda_{t,x})_{t,x}$ such that $\lambda = (\mathcal{L}_\perp[0, T]) \otimes \mu \otimes \lambda_{t,x}$ and

$$\int_{[0,T] \times X \times Z \times Y} \phi(t, x, z, y) d\lambda(t, x, z, y) = \int_{[0,T] \times X} \int_{Z \times Y} \phi(t, x, z, y) d\lambda_{t,x}(z, y) d\mu(x) dt$$

for all measurable ϕ from $[0, T] \times X \times Z \times Y \rightarrow \mathbb{R}_+$ (see Remark 3).

For $\lambda \in \mathcal{V}_T, n \in 2\mathbb{N}$ we define the glued discrete and limiting functionals

$$F_T^n(\lambda) := \int_0^T \int_X F_{t,x}^n(\lambda_{t,x}) d\mu(x) dt, \quad F_T(\lambda) := \int_0^T \int_X F_{t,x}(\lambda_{t,x}) d\mu(x) dt. \tag{4.16}$$

The finite time horizon T ensures that the infima of the glued functionals (4.16) are finite.

Remark 6 For any $n \in 2\mathbb{N}, T > 0$ a minimizer $\lambda^n \in \mathcal{V}$ for F_T^n can be obtained via Proposition 1 (and hence via the domain decomposition algorithm) by gluing together discrete fiber-wise minimizers $\lambda_{t,x}^n$ of (4.3) given by (4.6) to obtain $\lambda^n := (\mathcal{L}_\perp[0, T]) \otimes \mu \otimes \lambda_{t,x}^n$ (see Remark 3). The obtained λ^n clearly lies in \mathcal{V}_T and minimizes F_T^n because each $\lambda_{t,x}^n$ minimizes the fiberwise functional $F_{t,x}^n$. Due to the discreteness at scale n , only a finite number of minimizers must be chosen (one per discrete time-step and composite cell) and thus no measurability issues arise.

Proposition 2 Under Assumptions 1 and 2, for any $T > 0$, F_T^n Γ -converges to F_T with respect to weak* convergence on \mathcal{V}_T on the subsequence $n \in \hat{\mathcal{Z}}$.

The proof is divided into liminf and limsup condition that are given in Sect. 4.3.

Based on this we can now extract cluster points from the minimizers to the discrete fiber problems and also get convergence for the associated momenta. Besides, under Assumption 1, the cluster points will be minimizers of the limit problem.

Proposition 3 (Convergence of fiber-problem minimizers and momenta) Let $\lambda^n \in \mathcal{M}_+(\mathbb{R}_+ \times X \times Z \times Y)$ be constructed from the discrete iterates $\pi^{n,k}$ as shown in Proposition 1. Under Assumption 2 there is a subsequence $\hat{\mathcal{Z}} \subset \mathcal{Z} \subset 2\mathbb{N}$ and a measure $\lambda \in \mathcal{M}_+(\mathbb{R}_+ \times X \times Z \times Y)$ such that for all $T \in (0, \infty)$,

$$\lambda^n \llcorner ([0, T] \times X \times Z \times Y) \xrightarrow{*} \lambda \llcorner ([0, T] \times X \times Z \times Y) \in \mathcal{V}_T$$

In addition, analogous to Remark 5, we introduce the limit momentum field $\omega := \mathcal{L} \otimes \mu \otimes \omega_{t,x} \in \mathcal{M}(\mathbb{R}_+ \times X \times Y)^d$ via

$$\begin{aligned} (\omega_{t,x})_\ell &:= P_Y(\lambda_{t,x} \llcorner Z_+^\ell \times Y) - P_Y(\lambda_{t,x} \llcorner Z_-^\ell \times Y) \\ &\text{for } \ell = 1, \dots, d, \quad \text{with } Z_\pm^\ell = \{z \in Z \mid \pm z_\ell > 0\}. \end{aligned}$$

Then ω^n , $n \in \hat{\mathcal{Z}}$, converges weak* to ω on any finite time interval $[0, T]$.

Furthermore, if we allow Assumption 1, λ is a minimizer of F_T for all $T \in (0, \infty)$.

Proof For any $T > 0$ the sequence $(\lambda^n \llcorner ([0, T] \times X \times Z \times Y))_{n \in \mathcal{Z}}$ is tight (by compactness of the domain) and therefore weak* precompact, so we can extract a subsequence $\mathcal{Z}' \subset \mathcal{Z}$ such that $(\lambda^n \llcorner ([0, T] \times X \times Z \times Y))_{n \in \mathcal{Z}'}$ converges weak* to some $\lambda_T \in \mathcal{M}_+([0, T] \times X \times Z \times Y)$. By a diagonal argument, we can choose a further subsequence $\hat{\mathcal{Z}} \subset \mathcal{Z}$ such that $(\lambda^n)_{n \in \hat{\mathcal{Z}}}$ converges to some $\lambda \in \mathcal{M}_+(\mathbb{R}_+ \times X \times Z \times Y)$ when restricted to $[0, T] \times X \times Z \times Y$ for any $T > 0$. By construction (Proposition 1, Remark 6), λ^n , $n \in \hat{\mathcal{Z}}$, is a minimizer of F_T^n for any choice of T . Thus, by Γ -convergence (Proposition 2, if Assumption 1 is verified) λ is also a minimizer of F_T for all choices of T .

It remains to be shown that the construction of ω^n from λ^n (for the discrete and the limit case) is a weak* continuous operation. The fiber-wise construction can be written at the level of the whole measures as

$$(\omega^n)_\ell = P_{\mathbb{R}_+ \times X \times Y}(\lambda^n \llcorner (\mathbb{R}_+ \times X \times Z_+^\ell \times Y)) - P_{\mathbb{R}_+ \times X \times Y}(\lambda^n \llcorner (\mathbb{R}_+ \times X \times Z_-^\ell \times Y))$$

for $\ell = 1, \dots, d$. Marginal projection is a weak* continuous operation, and so is the addition (subtraction) of two measures. Let us therefore focus on the restriction operation. In general, restriction is not weak* continuous, but it is under our regularity assumptions on $\sigma_{t,x}^n$ and σ (Sect. 3.1, item 4. and Assumption 2). None of these measures carry mass on the boundaries between the Z_\pm^ℓ (and the sets Z_\pm^ℓ are relatively open in Z). For simplicity, we will now show that under these conditions, $[\sigma_{t,x}^n \xrightarrow{*} \sigma]$

$\Rightarrow [\sigma_{t,x}^n \llcorner Z_{\pm}^{\ell} \xrightarrow{*} \sigma \llcorner Z_{\pm}^{\ell}]$ for any $\ell \in \{1, \dots, d\}$. The same argument (but with heavier notation) will then apply to the convergence of the restrictions of λ^n . By weak* compactness we can select a subsequence such that

$$\sigma_{t,x}^n \llcorner Z_{\pm}^{\ell} \xrightarrow{*} \sigma_{\pm}$$

for two measures $\sigma_{\pm} \in \mathcal{M}_+(Z)$, and by the Portmanteau theorem for weak convergence of measures [7, Theorem 2.1] we have

$$\sigma_{\pm}(Z_{\mp}^{\ell}) \leq \liminf_n (\sigma_{t,x}^n \llcorner Z_{\pm}^{\ell})(Z_{\mp}^{\ell}) = 0. \tag{4.17}$$

Now observe

$$\sigma \llcorner Z_+^{\ell} + \sigma \llcorner Z_-^{\ell} = \sigma = \lim_{n \rightarrow \infty} \sigma_{t,x}^n = \lim_{n \rightarrow \infty} (\sigma_{t,x}^n \llcorner Z_+^{\ell} + \sigma_{t,x}^n \llcorner Z_-^{\ell}) = \sigma_+ + \sigma_-$$

where in the first and third equality we used that σ and $\sigma_{t,x}^n$ do not carry mass on the set $\{z_{\ell} = 0\}$. With (4.17) we conclude that $\sigma_{\pm} = \sigma \llcorner Z_{\pm}^{\ell}$. This holds for any convergent subsequence and thus by weak* compactness the whole sequences of restrictions converge to $\sigma \llcorner Z_{\pm}^{\ell}$. As indicated, the same argument will apply to the restriction of λ^n to the sets $[0, T] \times X \times Z_{\pm}^{\ell} \times Y$ for any finite horizon $T \in (0, \infty)$. \square

4.3 Liminf and limsup condition

We start by establishing that the transport cost contribution in F_T^n converges to that of F_T .

Lemma 2 (Convergence of the transport cost) *Let $T > 0$ and $(\lambda^n)_n$ be a weak* convergent sequence in \mathcal{V}_T with limit $\lambda \in \mathcal{V}_T$. Then the transport part of the glued functional F_T^n converges to that of F_T . More specifically,*

$$\begin{aligned} & \lim_{n \rightarrow \infty} \int_{[0,T] \times X \times Z \times Y} n \cdot [c(\bar{x}_{t,x}^n + z/n, y) - c(\bar{x}_{t,x}^n, y)] d\lambda^n(t, x, z, y) \\ &= \int_{[0,T] \times X \times Z \times Y} \langle \nabla_X c(x, y), z \rangle d\lambda(t, x, z, y). \end{aligned} \tag{4.18}$$

Proof By the mean value theorem, for each t, x, z, y, n , there exists a point ξ on the segment $[\bar{x}_{t,x}^n, \bar{x}_{t,x}^n + z/n]$ such that

$$n \cdot [c(\bar{x}_{t,x}^n + z/n, y) - c(\bar{x}_{t,x}^n, y)] = \langle \nabla_X c(\xi, y), z \rangle.$$

Since $z \in Z$, all points ξ converge uniformly to $\bar{x}_{t,x}^n$ as $n \rightarrow \infty$ and thus $\langle \nabla_X c(\xi, y), z \rangle$ converges uniformly to $\langle \nabla_X c(\bar{x}_{t,x}^n, y), z \rangle$. Paired with weak* convergence of λ^n to λ this implies convergence of the integral. \square

Lemma 3 (Liminf inequality) *Let $T > 0$ and $(\lambda^n)_{n \in \mathbb{Z}}$ be a weak* convergent sequence in \mathcal{V}_T with limit $\lambda \in \mathcal{V}_T$. Then, under Assumptions 1 and 2,*

$$\liminf_{n \in \mathbb{Z}, n \rightarrow \infty} F_T^n(\lambda^n) \geq F_T(\lambda). \tag{4.19}$$

Proof By disintegration λ^n and λ can be written as

$$\lambda^n = (\mathcal{L} \llcorner [0, T]) \otimes \mu \otimes \lambda^n_{t,x}, \quad \lambda = (\mathcal{L} \llcorner [0, T]) \otimes \mu \otimes \lambda_{t,x}$$

for suitable families $(\lambda^n_{t,x})_{t,x}$ and $(\lambda_{t,x})_{t,x}$. If the liminf is $+\infty$ there is nothing to prove. So we may limit ourselves to study subsequences with finite limit (and assume that we have extracted and relabeled such a sequence as $\mathbb{Z}' \subset \mathbb{Z}$). Unless otherwise stated, all limits in the proof are taken on this subsequence \mathbb{Z}' , though we may not always state it to avoid overloading the notation.

Step 1: marginal constraints. $F_T^n(\lambda^n)$ can only be finite if $\lambda^n_{t,x} \in \Pi(\sigma^n_{t,x}, \pi^n_{t,x})$ for $\mathcal{L} \otimes \mu$ almost all $(t, x) \in [0, T] \times X$. We find that this implies $P_Z \lambda_{t,x} = \sigma$ for almost all (t, x) by observing that for any $\phi \in \mathcal{C}([0, T] \times X \times Z)$ one has

$$\begin{aligned} & \int_{[0, T] \times X} \int_{Z \times Y} \phi(t, x, z) \, d\lambda_{t,x}(z, y) \, d\mu(x) \, dt \\ &= \lim_{n \rightarrow \infty} \int_{[0, T] \times X} \int_{Z \times Y} \phi(t, x, z) \, d\lambda^n_{t,x}(z, y) \, d\mu(x) \, dt \\ &= \lim_{n \rightarrow \infty} \int_{[0, T] \times X} \int_Z \phi(t, x, z) \, d\sigma^n_{t,x}(z) \, d\mu(x) \, dt \\ &= \int_{[0, T] \times X} \int_Z \phi(t, x, z) \, d\sigma(z) \, d\mu(x) \, dt \end{aligned}$$

where the first equality follows from $\lambda^n \xrightarrow{*} \lambda$ and the third one from dominated convergence since the inner integral converges pointwise almost everywhere (see Assumption 2) and is uniformly bounded. A symmetric argument provides $P_Y \lambda_{t,x} = \pi_{t,x}$: for any $\phi \in \mathcal{C}([0, T] \times X \times Y)$ one has

$$\begin{aligned} & \int_{[0, T] \times X} \int_{Z \times Y} \phi(t, x, y) \, d\lambda_{t,x}(z, y) \, d\mu(x) \, dt \\ &= \lim_{n \rightarrow \infty} \int_{[0, T] \times X} \int_{Z \times Y} \phi(t, x, y) \, d\lambda^n_{t,x}(z, y) \, d\mu(x) \, dt \\ &= \lim_{n \rightarrow \infty} \int_{[0, T] \times X} \int_Y \phi(t, x, y) \, d\pi^n_{t,x}(y) \, d\mu(x) \, dt \\ &= \int_{[0, T] \times Y} \int_Y \phi(t, x, y) \, d\pi_{t,x}(y) \, d\mu(x) \, dt \end{aligned}$$

where we argue again via dominated convergence combined with convergence pointwise almost everywhere (see Assumption 1) and uniform boundedness. Hence, $\lambda_{t,x} \in \Pi(\sigma, \pi_{t,x})$ for almost all (t, x) .

Step 2: transport contribution. The transport cost contributions converge by Lemma 2.

Step 3: entropy contribution. Assume first $\eta < \infty$. Introducing the measures

$$\lambda^n_{\otimes} := (\mathcal{L}_-[0, T]) \otimes \mu \otimes \sigma^n_{t,x} \otimes \pi^n_{t,x} \quad \text{and} \quad \lambda_{\otimes} := (\mathcal{L}_-[0, T]) \otimes \mu \otimes \sigma \otimes \pi_{t,x}, \tag{4.20}$$

the entropic terms of F^n_T and F_T can be written as $n \varepsilon^n \text{KL}(\lambda^n | \lambda^n_{\otimes})$ and $\eta \text{KL}(\lambda | \lambda_{\otimes})$ respectively, since

$$\begin{aligned} & \int_{[0,T] \times X} \text{KL}(\lambda^n_{t,x} | \sigma^n_{t,x} \otimes \pi^n_{t,x}) d\mu(x) dt \\ &= \int_{[0,T] \times X \times Z \times Y} \varphi \left(\frac{d\lambda^n_{t,x}(z, y)}{d(\sigma^n_{t,x} \otimes \pi^n_{t,x})} \right) d\sigma^n_{t,x}(z) d\pi^n_{t,x}(y) d\mu(x) dt \\ &= \int_{[0,T] \times X \times Z \times Y} \varphi \left(\frac{d\lambda^n(t, x, z, y)}{d\lambda^n_{\otimes}} \right) d\lambda^n_{\otimes}(t, x, z, y) = \text{KL}(\lambda^n | \lambda^n_{\otimes}). \end{aligned}$$

In the second equality we used that λ^n and λ^n_{\otimes} have the same $[0, T] \times X$ -marginal and only differ in their disintegrations along $Z \times Y$. Hence the relative density of these disintegrations determines their full relative density. Analogously, the entropic contribution in F_T is $\eta \text{KL}(\lambda | \lambda_{\otimes})$. Note that since we have restricted ourselves to a subsequence with finite objective at the beginning of the proof, we have that the integrals above are finite (and since φ is non-negative, even the divergence to $+\infty$ would be unambiguously defined). Thus, by joint lower semicontinuity of KL (where we use $\lambda^n_{\otimes} \xrightarrow{*} \lambda_{\otimes}$, which follows from Assumptions 1 and 2), convergence of $n\varepsilon^n$ to η and the fact that we selected a subsequence with finite limit (such that $\text{KL}(\lambda^n | \lambda^n_{\otimes})$ is uniformly bounded) we find

$$\liminf_{n \in \mathcal{Z}', n \rightarrow \infty} n \varepsilon^n \text{KL}(\lambda^n | \lambda^n_{\otimes}) \geq \eta \text{KL}(\lambda | \lambda_{\otimes}). \tag{4.21}$$

This shows that, for any subsequence \mathcal{Z}' with finite limit, $\liminf_{n \in \mathcal{Z}', n \rightarrow \infty} F^n_T(\lambda^n) \geq F_T(\lambda)$, so

$$\liminf_{n \in \mathcal{Z}, n \rightarrow \infty} F^n_T(\lambda^n) = \inf_{\mathcal{Z}' \subset \mathcal{Z}} \liminf_{n \in \mathcal{Z}', n \rightarrow \infty} F^n_T(\lambda^n) \geq F_T(\lambda). \tag{4.22}$$

This concludes the proof for $\eta < \infty$. The case $\eta = \infty$ is analogous. □

Lemma 4 (Limsup inequality) *Let $T > 0$, $\lambda \in \mathcal{V}_T$. Under Assumptions 1 and 2, there exists a sequence $(\lambda^n)_{n \in \mathcal{Z}}$ in \mathcal{V}_T , converging weak* to λ such that*

$$\limsup_{n \in \mathcal{Z}, n \rightarrow \infty} F^n_T(\lambda^n) \leq F_T(\lambda). \tag{4.23}$$

Proof Since $\lambda \in \mathcal{V}_T$ it can be disintegrated into $\lambda = \mathcal{L} \otimes \mu \otimes \lambda_{t,x}$ for a family $(\lambda_{t,x})_{t,x}$ in $\mathcal{M}_1(Z \times Y)$. We may assume that $F_T(\lambda) < \infty$, as otherwise there is nothing to prove. Hence, $\lambda_{t,x} \in \Pi(\sigma, \pi_{t,x})$ for $\mathcal{L} \otimes \mu$ -almost all $(t, x) \in [0, T] \times X$. We will

build our recovery sequence by gluing, setting $\lambda^n := \mathcal{L} \otimes \mu \otimes \lambda^n_{t,x}$ where we construct the fibers $\lambda^n_{t,x} \in \Pi(\sigma^n_{t,x}, \pi^n_{t,x})$ by tweaking the measures $\lambda_{t,x}$.

Step 1: construction of the recovery sequence. For every $n \in \mathcal{Z}$, let $(\gamma^n_{t,x})_{t,x}$ be a (measurable) family of measures in $\mathcal{M}_1(Y \times Y)$ where $\gamma^n_{t,x} \in \Pi(\pi^n_{t,x}, \pi_{t,x})$ is an optimal transport plan for $W_Y(\pi^n_{t,x}, \pi_{t,x})$. This family can be obtained, for instance, by disintegration of a minimizer of a ‘vertical’ transport problem written as

$$\inf \left\{ \int_{([0,T] \times X \times Y)^2} \|(t, x, y) - (t', x', y')\| \, d\gamma((t, x, y), (t', x', y')) \right\}$$

$$\gamma \in \Pi(\pi^n \llcorner ([0, T] \times X \times Y), \pi \llcorner ([0, T] \times X \times Y))$$

$$\text{with } x = x' \text{ and } t = t' \gamma((t, x, y), (t', x', y'))\text{-a.e.} \Big\}.$$

Likewise, let $(\bar{\gamma}^n_{t,x})_{t,x}$ be a (measurable) family of measures in $\mathcal{M}_1(Z \times Z)$ where $\bar{\gamma}^n_{t,x} \in \Pi(\sigma^n_{t,x}, \sigma)$ is an optimal transport plan for $W_Z(\sigma^n_{t,x}, \sigma)$. We then define $\lambda^n_{t,x}$ for $n \in \mathcal{Z}$ by integration against $\phi \in \mathcal{C}(Z \times Y)$ via

$$\int_{Z \times Y} \phi(z, y) \, d\lambda^n_{t,x}(z, y) := \int_{Z^2 \times Y^2} \phi(z, y) \, d(\bar{\gamma}^n_{t,x})_{z'}(z) \, d(\gamma^n_{t,x})_{y'}(y) \, d\lambda_{t,x}(z', y'), \tag{4.24}$$

where $(\bar{\gamma}^n_{t,x})_{z'}$ denotes the disintegration of $\bar{\gamma}^n_{t,x}$ with respect to its second marginal (namely σ) at point z' (and analogously for $(\gamma^n_{t,x})_{y'}$).

Step 2: correct marginals along the recovery sequence. Let us check that $\lambda^n_{t,x} \in \Pi(\sigma^n_{t,x}, \pi^n_{t,x})$. First, for any $\phi \in \mathcal{C}(Y)$,

$$\begin{aligned} \int_{Z \times Y} \phi(y) \, d\lambda^n_{t,x}(z, y) &= \int_{Z^2 \times Y^2} \phi(y) \, d(\bar{\gamma}^n_{t,x})_{z'}(z) \, d(\gamma^n_{t,x})_{y'}(y) \, d\lambda_{t,x}(z', y') \\ &= \int_{Y^2} \phi(y) \, d(\gamma^n_{t,x})_{y'}(y) \, d\pi_{t,x}(y') = \int_{Y^2} \phi(y) \, d\gamma^n_{t,x}(y, y') \\ &= \int_Y \phi(y) \, d\pi^n_{t,x}(y) \end{aligned}$$

where we used that $(\bar{\gamma}^n_{t,x})_{z'}$ is a probability measure. The same argument applies to the Z -marginal.

Step 3: convergence of the recovery sequence. Now we show that $\lambda^n \xrightarrow{*} \lambda$ for $n \in \mathcal{Z}$. For this we will use the Kantorovich–Rubinstein duality for the Wasserstein-1 distance (2.3)

$$W_{[0,T] \times X \times Z \times Y}(\lambda^n, \lambda) = \sup_{\phi \in \text{Lip}_1} \int_{[0,T] \times X \times Z \times Y} \phi \, d(\lambda^n - \lambda). \tag{4.25}$$

where we abbreviate $\text{Lip}_1 := \text{Lip}_1([0, T] \times X \times Z \times Y)$. In the following, let $\phi \in \text{Lip}_1$. We find

$$\begin{aligned} & \left| \int_{[0,T] \times X \times Z \times Y} \phi \, d(\lambda^n - \lambda) \right| \\ & \leq \int_{[0,T] \times X \times (Z \times Y)^2} |\phi(t, x, z, y) - \phi(t, x, z', y')| \, d(\bar{\gamma}_{t,x}^n)_{z'}(z) d(\gamma_{t,x}^n)_{y'}(y) d\lambda_{t,x}(z', y') d\mu(x) dt \end{aligned} \tag{4.26}$$

Using the Lipschitz continuity of ϕ we can bound

$$\begin{aligned} & |\phi(t, x, z, y) - \phi(t, x, z', y')| \\ & \leq |\phi(t, x, z, y) - \phi(t, x, z', y)| + |\phi(t, x, z', y) - \phi(t, x, z', y')| \\ & \leq \|z - z'\| + \|y - y'\|. \end{aligned}$$

Thus, we can continue

$$\begin{aligned} (4.26) & \leq \int_{[0,T] \times X \times Z^2} \|z - z'\| d(\bar{\gamma}_{t,x}^n)_{z'}(z) \, d\sigma(z') \, d\mu(x) \, dt \\ & \quad + \int_{[0,T] \times X \times Y^2} \|y - y'\| d(\gamma_{t,x}^n)_{y'}(y) \, d\pi_{t,x}(y') \, d\mu(x) \, dt \\ & \leq \int_{[0,T] \times X} W_Z(\sigma_{t,x}^n, \sigma) \, d\mu(x) \, dt + \int_{[0,T] \times X} W_Y(\pi_{t,x}^n, \pi_{t,x}) \, d\mu(x) \, dt \end{aligned}$$

where we have used optimality of the plans $\bar{\gamma}_{t,x}^n$ and $\gamma_{t,x}^n$. Using Assumptions 1 and 2 and dominated convergence (where we exploit that Z and Y are compact, hence W_Z and W_Y are bounded), we find that this tends to zero as $\mathcal{Z} \ni n \rightarrow \infty$. Plugging this into (4.25), we find that $W_{[0,T] \times X \times Z \times Y}(\lambda^n, \lambda) \rightarrow 0$ and since Wasserstein distances metrize weak* convergence on compact spaces, we obtain $\lambda^n \xrightarrow{*} \lambda$ for $n \in \mathcal{Z}$.

Step 4: lim sup inequality. Now we have to distinguish between different behaviors of $(n \cdot \varepsilon^n)_n$.

- [$\eta = 0, \varepsilon^n = 0$ for all $n \in \mathcal{Z}$, with only a finite number of exceptions] The exceptions have no effect on the lim sup, hence we may skip them. By Lemma 2 the transport contribution to the functional converges, and so we obtain that $\lim_{n \in \mathcal{Z}, n \rightarrow \infty} F_T^n(\lambda^n) = F_T(\lambda)$.
- [$\eta > 0$] We have that $\varepsilon^n > 0$ for all n up to a finite number of exceptions, which we may again skip. In this case, the limit cost has an entropic contribution, and thus for a.e. $t \in [0, T]$ and μ -a.e. $x \in X$, $\lambda_{t,x}$ has a density with respect to $\sigma \otimes \pi_{t,x}$, that we denote by $u_{t,x}$. Then, as we will show below, $\lambda_{t,x}^n$ also has a density with respect to $\sigma_{t,x}^n \otimes \pi_{t,x}^n$, that is given by:

$$u_{t,x}^n(z, y) := \int_{Z \times Y} u_{t,x}(z', y') d(\bar{\gamma}_{x,t}^n)_{z'}(z') d(\gamma_{x,t}^n)_{y'}(y'), \tag{4.27}$$

where we use again the transport plans $\bar{\gamma}_{x,t}^n \in \Pi(\sigma_{t,x}^n, \sigma)$ and $\gamma_{x,t}^n \in \Pi(\pi_{t,x}^n, \pi_{t,x})$ and this time their disintegrations against the first marginals. Let us prove that $u_{t,x}^n$ is indeed the density of $\lambda_{t,x}^n$ with respect to $\sigma_{t,x}^n \otimes \pi_{t,x}^n$:

$$\begin{aligned} & \int_{Z \times Y} \phi(z, y) u_{t,x}^n(z, y) d\sigma_{t,x}^n(z) d\pi_{t,x}^n(y) \\ &= \int_{Z^2 \times Y^2} \phi(z, y) u_{t,x}(z', y') \underbrace{d(\bar{\gamma}_{x,t}^n)_z(z')}_{=d(\bar{\gamma}_{x,t}^n)_{z'}(z) d\sigma(z')} \underbrace{d(\gamma_{x,t}^n)_y(y')}_{=d(\gamma_{x,t}^n)_{y'}(y) d\pi_{t,x}(y')} d\pi_{t,x}^n(y), \end{aligned}$$

where we switched the disintegration from the first to the second marginals. Now use that $\lambda_{t,x} = u_{t,x} \cdot (\sigma \otimes \pi_{t,x})$, and finally we use (4.24),

$$= \int_{Z^2 \times Y^2} \phi(z, y) d(\bar{\gamma}_{x,t}^n)_{z'}(z) d(\gamma_{x,t}^n)_{y'}(y) d\lambda_{t,x}(z', y') = \int_{Z \times Y} \phi(z, y) d\lambda_{t,x}^n(z, y).$$

Regarding the entropic regularization, notice that $\varphi(s) = s \log(s) - s + 1$ is a convex function, so using Jensen's inequality we obtain:

$$\begin{aligned} \varphi(u_{t,x}^n(z, y)) &= \varphi\left(\int_{Z \times Y} u_{t,x}(z', y') d(\bar{\gamma}_{x,t}^n)_z(z') d(\gamma_{x,t}^n)_y(y')\right) \\ &\leq \int_{Z \times Y} \varphi(u_{t,x}(z', y')) d(\bar{\gamma}_{x,t}^n)_z(z') d(\gamma_{x,t}^n)_y(y'), \end{aligned}$$

so the entropic term can be bounded as

$$\begin{aligned} & \int_{[0,T] \times X} \text{KL}(\lambda_{t,x}^n | \sigma_{t,x}^n \otimes \pi_{t,x}^n) d\mu(x) dt \\ &= \int_{[0,T] \times X} \int_{Z \times Y} \varphi(u_{t,x}^n(z, y)) d\sigma_{t,x}^n(z) d\pi_{t,x}^n(y) d\mu(x) dt \\ &\leq \int_{[0,T] \times X} \int_{Z^2 \times Y^2} \varphi(u_{t,x}(z', y')) d\bar{\gamma}_{x,t}^n(z, z') d\gamma_{x,t}^n(y, y') d\mu(x) dt \\ &= \int_{[0,T] \times X} \int_{Z \times Y} \varphi(u_{t,x}(z', y')) d\sigma(z') d\pi_{t,x}(y') d\mu(x) dt \\ &= \int_{[0,T] \times X} \text{KL}(\lambda_{t,x} | \sigma \otimes \pi_{t,x}) d\mu(x) dt. \end{aligned}$$

Adding to this the convergence of the transport contribution along weak* converging sequences (Lemma 2) and that $n\varepsilon^n$ converges to η it follows that, for both $\eta < \infty$ and $\eta = \infty$, $\limsup_{n \in \mathcal{Z}, n \rightarrow \infty} F_T^n(\lambda^n) \leq F_T(\lambda)$.

- [$\eta = 0$, $\varepsilon_n > 0$ for an infinite number of indices n] This case is slightly more challenging since the reconstructed $\lambda_{t,x}^n$ may not have a density with respect to $\sigma_{t,x}^n \otimes \pi_{t,x}^n$, and thus the KL term at finite n may explode for $\varepsilon^n > 0$. Hence, for those n the recovery sequence needs to be adjusted. We apply the *block approximation* technique as in [10], which is summarized in Lemma 5. We set $\hat{\lambda}_{t,x}^n$ to be the block approximation of $\lambda_{t,x}^n$ at scale $L_n := n\varepsilon^n$ (where we set $\Omega := Y \cup Z$). Lemma 5 provides that the marginals are preserved, i.e. $\hat{\lambda}_{t,x}^n \in \Pi(\sigma_{t,x}^n, \pi_{t,x}^n)$. In

addition we find that $W_{Z \times Y}(\lambda_{t,x}^n, \hat{\lambda}_{t,x}^n) \leq L_n \cdot \sqrt{2d}$ and thus $\hat{\lambda}^n \xrightarrow{*} \lambda$ (arguing as above, e.g. via dominated convergence). So by Lemma 2 the transport contribution still converges. Finally, for the entropic contribution we get from Lemma 5,

$$n\varepsilon^n \text{KL}(\hat{\lambda}_{t,x}^n | \sigma_{t,x}^n \otimes \pi_{t,x}^n) \leq Cn\varepsilon^n - 2dn\varepsilon^n \log(n\varepsilon^n) \xrightarrow{n \rightarrow \infty} 0. \tag{4.28}$$

Wrapping up, this means that $\limsup_{n \in \mathcal{Z}, n \rightarrow \infty} F_T^n(\hat{\lambda}^n) \leq F_T(\lambda)$, and $(\hat{\lambda}^n)_n$ represents a valid recovery sequence. □

4.4 Continuity equation

The discrete momenta ω^n , (3.6) have been introduced to approximately describe the ‘horizontal’ mass movement in the discrete trajectories π^n , (3.6) via a continuity equation on $X \times Y$. We now establish that in the limit the relation becomes exact for any cluster point $\hat{\omega}$ of $(\omega_n)_n$ (i.e., we do not use the structure given by Proposition 1, which is only valid under Assumptions 1 and 2).

Proposition 4 *Let $\omega^n \in \mathcal{M}(\mathbb{R}_+ \times X \times Y)^d$ defined as in (3.6), let $\hat{\omega} \in \mathcal{M}(\mathbb{R}_+ \times X \times Y)^d$ and $\hat{Z} \subset \mathcal{Z} \subset 2\mathbb{N}$ be such that $\omega^n \llcorner [0, T] \xrightarrow{*} \hat{\omega} \llcorner [0, T]$ for any $T \in (0, \infty)$. Then π and $\hat{\omega}$ solve the horizontal continuity equation*

$$\partial_t \pi_t + \text{div}_X \hat{\omega}_t = 0 \text{ for } t > 0 \text{ and } \pi_{t=0} = \pi_{init}$$

in a distributional sense. More precisely, for any $\phi \in \mathcal{C}_c^1(\mathbb{R}_+ \times X \times Y)$ one has

$$\int_{\mathbb{R}_+ \times X \times Y} \partial_t \phi \, d\pi + \int_{\mathbb{R}_+ \times X \times Y} \nabla_X \phi \cdot d\hat{\omega} = - \int_{X \times Y} \phi(0, x, y) \, d\pi_{init}(x, y). \tag{4.29}$$

Proof Let $\phi \in \mathcal{C}_c^1(\mathbb{R}_+ \times X \times Y)$. We will show that

$$\int_{\mathbb{R}_+ \times X \times Y} \partial_t \phi \, d\pi^n + \int_{\mathbb{R}_+ \times X \times Y} \nabla_X \phi \cdot d\omega^n = - \int_{X \times Y} \phi(0, x, y) \, d\pi_{init}^n(x, y) + o(1) \tag{4.30}$$

for $\hat{Z} \ni n \rightarrow \infty$ and then (4.29) will follow by weak* convergence of $(\pi^n)_n$ to π and of $(\omega^n)_n$ to $\hat{\omega}$ on compact time intervals.

Since ϕ has compact support, there exists some $T \in \mathbb{R}_+$ such that $\phi(t, \cdot, \cdot) = 0$ for all $t \geq T$. Now fix some n , and note that $\partial_t \phi$ and $\nabla_X \phi$ are uniformly continuous, π_t^n is a probability measure for almost all $t \in [0, T]$, and similarly, the mass of each spatial component of ω_t^n is bounded by 1 (intuitively, since mass can only move by one cell per iteration, see (3.6) for the definition). Thus, replacing $\partial_t \phi(t, x, y)$ and $\nabla_X \phi(t, x, y)$ on the left hand side of (4.30) by $\partial_t \phi(t, \bar{x}_{t,x}^n, y)$ and $\nabla_X \phi(t, \bar{x}_{t,x}^n, y)$ only introduces an error of $o(1)$ in the first two terms, since $\|x - \bar{x}_{t,x}^n\| \leq \sqrt{d}/n$. Thus the first term of (4.30) becomes

$$\int_0^\infty \int_{X \times Y} \partial_t \phi(t, x, y) \, d\pi_t^n(x, y) \, dt = \int_0^T \int_{X \times Y} \partial_t \phi(t, \bar{x}_{t,x}^n, y) \, d\pi_t^n(x, y) \, dt + o(1).$$

Now take $K = \lceil nT \rceil$, and use that π_t^n is constant on time intervals of length $1/n$ and on composite cells, so we can continue

$$\begin{aligned} &= \sum_{k=0}^{K-1} \sum_{J \in \mathcal{J}^{n,k}} \int_0^{1/n} \int_Y \partial_t \phi\left(\frac{k}{n} + s, x_J^n, y\right) \, dv_{J,J}^{n,k}(y) \, ds + o(1) \\ &= \sum_{k=0}^{K-1} \sum_{J \in \mathcal{J}^{n,k}} \int_Y \left[\phi\left(\frac{k+1}{n}, x_J^n, y\right) - \phi\left(\frac{k}{n}, x_J^n, y\right) \right] \, dv_{J,J}^{n,k}(y) + o(1). \end{aligned} \tag{4.31}$$

Likewise, in the second term of (4.30) we replace again x by $\bar{x}_{t,x}^n$, and also t by $\frac{\lceil nt \rceil}{n}$, which yields again an error of order $o(1)$. The second term then becomes

$$\begin{aligned} &\int_{\mathbb{R}_+} \int_{X \times Y} \nabla_X \phi(t, x, y) \cdot d\omega_t^n(x, y) \, dt = \int_0^T \int_{X \times Y} \nabla_X \phi\left(\frac{\lceil nt \rceil}{n}, \bar{x}_{t,x}^n, y\right) \cdot d\omega_t^n(x, y) \, dt + v o(1) \\ &= o(1) + \sum_{k=0}^{K-1} \sum_{J \in \mathcal{J}^{n,k}} \sum_{\hat{J} \in \mathcal{N}(J)} \int_0^{1/n} \int_Y \nabla_X \phi\left(\frac{k+1}{n}, x_J^n, y\right) \cdot (x_J^n - x_{\hat{J}}^n) \cdot n \, dv_{J,\hat{J}}^{n,k}(y) \, ds. \end{aligned}$$

Now the integral over $[0, 1/n]$ cancels with the factor n , and $\nabla_X \phi\left(\frac{k+1}{n}, x_J^n, y\right) \cdot (x_J^n - x_{\hat{J}}^n) = \phi\left(\frac{k+1}{n}, x_J^n, y\right) - \phi\left(\frac{k+1}{n}, x_{\hat{J}}^n, y\right) + o(1/n)$. We get

$$= o(1) + \sum_{k=0}^{K-1} \sum_{J \in \mathcal{J}^{n,k}} \sum_{\hat{J} \in \mathcal{N}(J)} \int_Y \left[\phi\left(\frac{k+1}{n}, x_J^n, y\right) - \phi\left(\frac{k+1}{n}, x_{\hat{J}}^n, y\right) + o(1/n) \right] \, dv_{J,\hat{J}}^{n,k}(y).$$

The sum of all $v_{J,\hat{J}}^{n,k}$ over J and \hat{J} has unit mass, so the total contribution of the $o(1/n)$ errors scales like $K \cdot o(1/n) = T \cdot n \cdot o(1/n)$, which is $o(1)$. Thus, we can absorb this error term into the global $o(1)$ error:

$$\begin{aligned} &= o(1) + \sum_{k=0}^{K-1} \left[\sum_{J \in \mathcal{J}^{n,k}} \sum_{\hat{J} \in \mathcal{N}(J)} \int_Y \phi\left(\frac{k+1}{n}, x_J^n, y\right) \, dv_{J,\hat{J}}^{n,k}(y) \right. \\ &\quad \left. - \sum_{J \in \mathcal{J}^{n,k}} \sum_{\hat{J} \in \mathcal{N}(J)} \int_Y \phi\left(\frac{k+1}{n}, x_{\hat{J}}^n, y\right) \, dv_{J,\hat{J}}^{n,k}(y) \right] \end{aligned}$$

Then, in the second term we can regroup all $v_{J,\hat{J}}^{n,k}$ with the same J into $v_J^{n,k}$. In the first term we can first reverse the order of the sums, and then use that adding up all

$v_{J,\hat{J}}^{n,k}$ with the same \hat{J} results in $v_{\hat{J}}^{n,k} = v_{\hat{J}}^{n,k+1}$ (see (3.5) for the equality). This leaves us with

$$= o(1) + \sum_{k=0}^{K-1} \left[\sum_{\hat{J} \in \mathcal{J}^{n,k+1}} \int_Y \phi\left(\frac{k+1}{n}, x_{\hat{J}}^n, y\right) dv_{\hat{J}}^{n,k+1}(y) - \sum_{J \in \mathcal{J}^{n,k}} \int_Y \phi\left(\frac{k+1}{n}, x_J^n, y\right) dv_J^{n,k}(y) \right]. \tag{4.32}$$

Now we can combine the temporal and spatial parts, noticing that the first term in (4.31) cancels with the second term in (4.32), so the left hand side of (4.30) equals

$$\begin{aligned} & \int_{\mathbb{R}_+} \int_{X \times Y} \partial_t \phi(t, x, y) d\pi_t^n(x, y) dt + \int_{\mathbb{R}_+} \int_{X \times Y} \nabla_X \phi(t, x, y) \cdot d\omega_t^n(x, y) dt \\ &= o(1) + \sum_{k=0}^{K-1} \left[\sum_{\hat{J} \in \mathcal{J}^{n,k+1}} \int_Y \phi\left(\frac{k+1}{n}, x_{\hat{J}}^n, y\right) dv_{\hat{J}}^{n,k+1}(y) - \sum_{J \in \mathcal{J}^{n,k}} \int_Y \phi\left(\frac{k}{n}, x_J^n, y\right) dv_J^{n,k}(y) \right] \end{aligned}$$

which is a telescopic sum. The surviving terms are just

$$= o(1) + \sum_{\hat{J} \in \mathcal{J}^{n,K}} \int_Y \phi\left(\frac{K}{n}, x_{\hat{J}}^n, y\right) dv_{\hat{J}}^{n,K}(y) - \sum_{J \in \mathcal{J}^{n,0}} \int_Y \phi(0, x_J^n, y) dv_J^{n,0}(y)$$

The first integral vanishes, since $\phi(K/n, \cdot, \cdot) = 0$. In the second integral we first integrate again in space:

$$\begin{aligned} &= o(1) - \sum_{J \in \mathcal{J}^{n,0}} \int_{X_J^n \times Y} \phi(0, x_J^n, y) d\mu(x) \frac{1}{m_J^n} dv_J^{n,0}(y) \\ &= o(1) - \sum_{J \in \mathcal{J}^{n,0}} \int_{X_J^n \times Y} \phi(0, x_J^n, y) d\pi_0^n(x, y) \end{aligned}$$

We use for the last time that replacing x_J^n by x in X_J^n introduces a global $o(1)$ error (since π_0^n has unit mass):

$$= o(1) - \sum_{J \in \mathcal{J}^{n,0}} \int_{X_J^n \times Y} \phi(0, x, y) d\pi_0^n(x, y) = o(1) - \int_{X \times Y} \phi(0, x, y) d\pi_0^n(x, y)$$

and finally, since $(\pi_0^n)_n$ converges weak* to π_{init} , which yields precisely (4.30):

$$= - \int_{X \times Y} \phi(0, x, y) d\pi_{\text{init}}(x, y) + o(1).$$

□

The following observation may help in the interpretation of the trajectories generated by domain decomposition.

Proposition 5 *The sequence momentum fields $(\omega^n)_n$ and any cluster point $\hat{\omega}$ thereof are absolutely continuous with respect to their respective trajectories $(\pi^n)_n$ and π and the component-wise density is bounded by one. More precisely,*

$$\left| \frac{d(\omega^n_{t,x})_\ell}{d\pi^n_{t,x}} \right| \leq 1 \text{ for all } n \in 2\mathbb{N}, \text{ and } \left| \frac{d(\hat{\omega}_{t,x})_\ell}{d\pi_{t,x}} \right| \leq 1. \tag{4.33}$$

for a.e. t , μ -a.e. x and all $\ell \in \{1, \dots, d\}$.

Proof For $n \in 2\mathbb{N}$, this is a simple consequence of (4.9):

$$\begin{aligned} \left| \frac{d(\omega^n_{t,x})_\ell}{d\pi^n_{t,x}} \right| &= \left| \frac{d[P_Y(\lambda^n_{t,x} \llcorner Z^{\ell}_+ \times Y) - P_Y(\lambda^n_{t,x} \llcorner Z^{\ell}_- \times Y)]}{d\pi^n_{t,x}} \right| \leq \left| \frac{dP_Y(\lambda^n_{t,x} \llcorner Z^{\ell}_+ \times Y)}{d\pi^n_{t,x}} \right| \\ &\quad + \left| \frac{dP_Y(\lambda^n_{t,x} \llcorner Z^{\ell}_- \times Y)}{d\pi^n_{t,x}} \right| \\ &= \frac{dP_Y(\lambda^n_{t,x} \llcorner Z^{\ell}_+ \times Y)}{d\pi^n_{t,x}} + \frac{dP_Y(\lambda^n_{t,x} \llcorner Z^{\ell}_- \times Y)}{d\pi^n_{t,x}} = \frac{dP_Y(\lambda^n_{t,x} \llcorner Z \times Y)}{d\pi^n_{t,x}} \leq \frac{d\pi^n_{t,x}}{d\pi^n_{t,x}} = 1. \end{aligned}$$

where we used that λ^n is a positive measure. The claim for $\hat{\omega}$ follows from Reshetnyak lower semicontinuity [2, Theorem 2.38] □

Remark 7 (Interpretation) Proposition 5 suggests the interpretation of the changes in π^n_t and π_t over time as horizontal movement of mass, along the X -direction, where particles move at most with velocity 1 along each spatial axis. This corresponds to the fact that each mass particle can only travel by one basic cell along each axis per iteration.

4.5 Main result

We can now summarize and assemble the results from the two previous sections to arrive at the main result of the article.

Theorem 3 *Assume Assumptions 1 and 2 hold. Then, up to selection of a subsequence $\hat{\mathcal{Z}} \subset \mathcal{Z}$, the sequences of discrete trajectories $(\pi^n)_n$, (3.6), and momenta $(\omega^n)_n$, (3.6), that are generated by the domain decomposition algorithm at scales n , converge weak* on compact sets to a limit trajectory π and momentum ω as $n \rightarrow \infty$. The limits solve the horizontal continuity equation*

$$\partial_t \pi_t + \operatorname{div}_X \omega_t = 0 \text{ for } t > 0 \text{ and } \pi_{t=0} = \pi_{init}$$

on $X \times Y$ in a distributional sense, (4.29). The limit momentum ω is induced by an asymptotic version of the domain decomposition algorithm. More precisely, for $\mathcal{L} \otimes \mu$ -almost all (t, x) its disintegration is given by

$$(\omega_{t,x})_\ell := P_Y(\lambda_{t,x} \llcorner Z^{\ell}_+ \times Y) - P_Y(\lambda_{t,x} \llcorner Z^{\ell}_- \times Y)$$

$$\text{for } \ell = 1, \dots, d, \quad \text{with } Z_{\pm}^{\ell} = \{z \in Z \mid \pm z_{\ell} > 0\}$$

where for $\eta < \infty$ the measure $\lambda_{t,x}$ is given as a minimizer of the asymptotic cell problem

$$\inf \left\{ \int_{Z \times Y} \langle \nabla_X c(x, y), z \rangle d\lambda(z, y) + \eta \cdot \text{KL}(\lambda | \sigma \otimes \pi_{t,x}) \mid \lambda \in \Pi(\sigma, \pi_{t,x}) \right\}.$$

For $\eta = \infty$ one finds $\lambda_{t,x} = \sigma \otimes \pi_{t,x}$, which implies $\omega_{t,x} = 0$ and thus $\pi_t = \pi_{\text{init}}$ for all $t \in \mathbb{R}_+$. Hence, the algorithm asymptotically freezes.

Proof Assumptions 1 includes the weak* convergence on compact sets of π^n to a limit π for a subsequence \mathcal{Z} . Under Assumptions 1 and 2, Proposition 3 provides the existence of a subsequence $\hat{\mathcal{Z}} \subset \mathcal{Z}$ such that ω^n converges weak* on compact sets to a limit ω on $\hat{\mathcal{Z}}$, and this limit is of the prescribed form for (almost-everywhere) fiber-wise minimizers $\lambda_{t,x}$ of a limit fiber problem given in Definition 5.

For $\eta < \infty$ the limit fiber problem is as stated. For $\eta = \infty$, the unique minimizer of the limit fiber problem is given by $\lambda_{t,x} = \sigma \otimes \pi_{t,x}$. With Assumption 2 this implies $\omega_{t,x} = 0$ (see Lemma 1).

Solution to the continuity equation is provided by Proposition 4. For $\eta = \infty$, with $\omega = 0$ this implies that the limit trajectory π_t is constant and equal to π_{init} . □

Remark 8 (Discussion) We observe that solutions to the limit system given in Theorem 3 are not unique. For instance, if $\pi_{\text{init}} = (\text{id}, S)_\# \mu$ for some (non-optimal) transport map $S : X \rightarrow Y$, then a solution to the limit system is given by $\pi_t = \pi_{\text{init}}$, $\omega_t = 0$, $\lambda_{t,x} = \sigma \otimes \delta_{S(x)}$ for all $t \geq 0$ where we used that $\pi_{t,x} = (\pi_{\text{init}})_x = \delta_{S(x)}$. In contrast, limit solutions generated by the domain decomposition algorithm are usually able to leave such a point by making the coupling non-deterministic (i.e. mass from a point x will be split onto multiple targets y), since at each n the algorithm has a ‘non-zero range of vision’ (see Fig. 6, for an example). This situation is somewhat analogous to a gradient flow being stuck in a saddle point, whereas a minimizing movement scheme or a proximal point algorithm is able to move on.

The algorithm by Angenent, Haker, and Tannenbaum [3] solves the W_2 -optimal transport problem (for convex X and $\mu \ll \mathcal{L}$) by starting from some feasible transport map and then subsequently removing its curl in a suitable way. It therefore generates trajectories that lie solely in the subset of map couplings (i.e. concentrated on a map) and the algorithm breaks down when cusps or overlaps form. ([3] also discusses a regularized version). From the previous paragraph we deduce that in asymptotic domain decomposition trajectories, mass can only move when the coupling is not of map-type and the algorithm is well-defined on the whole space of Kantorovich transport plans.

5 Examples for asymptotic sub-optimality

Numerical examples for the practical efficiency of the method have already been illustrated in [8]. A basic intuition for the asymptotic behaviour as $n \rightarrow \infty$ on

‘well-behaved’ examples can be drawn from Figs. 1, 2, 3, 4, 5 6. The examples in this section demonstrate potential obstructions for the convergence of the asymptotic domain decomposition trajectory to the global minimizer.

5.1 Discretization

On discrete unregularized problems the domain decomposition algorithm may fail to converge to the globally optimal solution. Examples are given in [4, Section 5.2] and [8, Example 4.12]. We study such an example in the context of the asymptotic behaviour of the algorithm.

In the following, for $\beta \in S^1$ denote by e_β the unit vector in \mathbb{R}^2 with orientation β . Further, let

$$H := R_\theta^\top \begin{pmatrix} -(\tan \alpha)^2 & 0 \\ 0 & 1 \end{pmatrix} R_\theta \quad \text{where} \quad R_\theta := \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}$$

for some $\theta \in S^1$ and some $\alpha \in (0, \pi/2)$. Set $V : \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto \frac{1}{2}x^\top H x, S := \nabla V$. H is indefinite with eigenvalues $-(\tan \alpha)^2$ and 1 for eigenvectors e_θ and $e_{\theta+\pi/2}$. Consequently V is not convex and thus S is in general not an optimal transport map between $\mu \in \mathcal{M}_+(\mathbb{R}^2)$ and $S_\# \mu$ for the squared distance cost on \mathbb{R}^2 by virtue of Brenier’s polar factorization [9]. However, for a set $A \subset \mathbb{R}^2$ such that $(x_1 - x_2)^\top H(x_1 - x_2) \geq 0$ for all $x_1, x_2 \in A$ one quickly verifies that the graph of S over A is c -cyclically monotone for the squared distance and therefore S is an optimal transport map between μ and $S_\# \mu$ for $\mu \in \mathcal{M}_+(A)$. One has $e_\beta^\top H e_\beta < 0$ if and only if $\beta \in (\theta - \alpha, \theta + \alpha) \cup (\theta + \pi - \alpha, \theta + \pi + \alpha)$ on S^1 (see Fig. 7 for an illustration of this and the subsequent construction). Therefore, if $\mu^n = \sum_{i \in I^n} m_i^n \cdot \delta_{x_i^n}$ (see Lemma 1) such that each composite cell is essentially a small 2 by 2 Cartesian grid, and θ and α are chosen carefully, then S will be optimal on each composite cell. But for sufficiently large n , some grid points from another cell will eventually lie in the red cone and S is then not globally optimal on X . Hence, if we set $\nu^n := S_\# \mu^n$ and $\pi_{\text{init}}^n = (\text{id}, S)_\# \mu^n$,

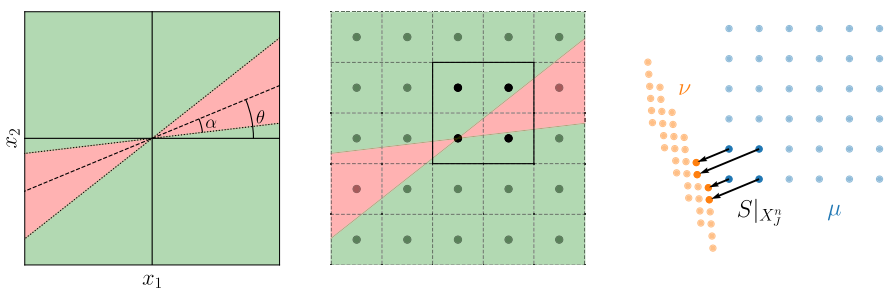


Fig. 7 Left, cone where $x^\top H x < 0$ in red. Center, for H with a sufficiently narrow, $S|_{X_j^n}$ is a monotone arrangement for all composite cells X_j^n . Right, the resulting coupling π_{init}^n is optimal on each composite cell, albeit not globally (color figure online)

$\varepsilon^n = 0$, then the discrete trajectory at each n will be stationary and so will be the limit trajectory. But it will not be globally optimal.

Fix now a scale n . If each basic cell contains more points, the space where the red cone ‘remains unnoticed’ becomes smaller and thus α must decrease, but it can always be chosen to be strictly positive, i.e. S will not be globally optimal on a sufficiently large grid. If we send the number of points per basic cell to infinity (being arranged on a regular Cartesian grid), for fixed n , it was shown in [4, Section 5.2] that in the limit one recovers a globally optimal coupling. The behaviour in the case where the number of points per basic cell and n tend to ∞ simultaneously remains open (see also Lemma 1).

If we set $\varepsilon^n > 0$, then for each fixed n we know by [8] that the algorithm converges to the global minimizer. If $\eta = \infty$, then asymptotically the algorithm will freeze in the initial configuration (Theorem 3). Entropy regularized optimal transport converges to the unregularized limit as $\varepsilon \rightarrow 0$ [10, 11, 18]. Therefore it seems possible that if $\varepsilon^n \rightarrow 0$ sufficiently fast, one may obtain the same asymptotic behaviour as for $\varepsilon^n = 0$, i.e. potentially we end up in a non-minimal configuration in the limit, even though at each n , eventually the globally optimal solution is found (after times that increase exponentially in n). An open question is therefore, if there is an intermediate regime of scaling $(\varepsilon^n)_n$ such that the global minimizer is obtained in the asymptotic trajectory. Preliminary numerical experiments (and data from [8]) suggest that such a regime exists.

5.2 Semi-discrete transport

Another asymptotic obstruction to global optimality occurs when a sub-optimal initial plan is chosen where sub-optimality is concentrated on an increasingly small subset of composite cells (as $n \rightarrow \infty$) and if this concentration is ‘stable’ under iterations. Then most of the cells will not induce any change in the plan and asymptotically the trajectory freezes. We illustrate this phenomenon with a semi-discrete example. Let $\mu = \mu^n = \mathcal{L} \llcorner X$, $\nu = \nu^n$ is the sum of two Diracs at $(\pm 1, 0)$ with equal mass, and the initialization $\pi_{\text{init}} = \pi_{\text{init}}^n$ takes all the mass to the left of an approximately vertical interface line to $(-1, 0)$ and everything to the right to $(+1, 0)$. The optimal coupling would be given by a vertical interface. The evolution of iterates for different resolutions is shown in Fig. 8.

All composite cells that do not touch the interface are locally optimal and will not change during an iteration, only cells that intersect the interface change. On a macroscopic level the effect is roughly as follows: mass for the left and right points in Y will essentially ‘travel along the interface’ in the appropriate direction, the rough structure of the interface itself remains stable. At the boundaries of X the interface will curve towards the right orientation and this will gradually propagate into the interior of the domain. For each fixed n convergence to the global minimizer follows from a slight extension of Benamou’s analysis [4] to more general partitions. However, the capacity of mass that can flow along the interface decreases with n and thus convergence will become gradually slower, freezing in the limit $n \rightarrow \infty$.

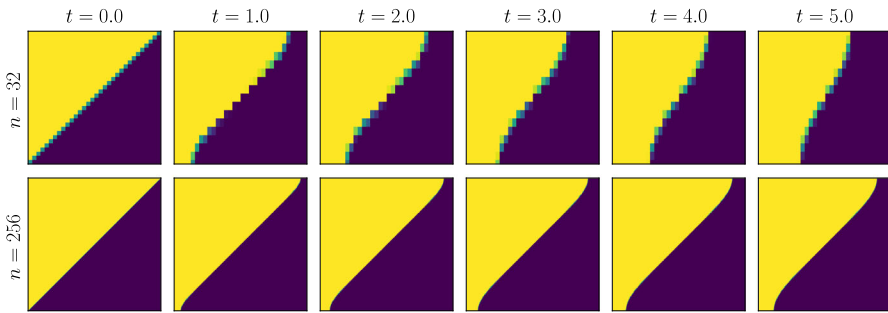


Fig. 8 Domain decomposition iterates $\pi^{n, \lfloor nt \rfloor}$ for the semi-discrete example. Colors in the images for couplings indicate the target point of mass in Y for given cells. Along the interface mass is sent to both points, resulting in a mixed color. Note that changes in the iterations occur only along the interface, which remains approximately stable during iterations. As $n \rightarrow \infty$ the dynamics freeze (color figure online)

6 Conclusion and outlook

Summary. In this article we derived a description of the asymptotic limit dynamics of a family of domain decomposition algorithms for optimal transport (regularized and unregularized, discretized and continuous) as the size of the individual cells tends to zero. To be able to analyze a ‘pointwise’ limit of the cell problems, a sufficiently strong convergence of trajectories at the level of the disintegrations along time and the first marginal had to be assumed, see Remark 1. Using this, we proved Γ -convergence of the cell problems to a limit problem where the cells have collapsed to single points. From the limit cell problems a ‘horizontal’ momentum field could be extracted which was shown to describe the evolution of the limit trajectory via a suitable continuity equation. Some examples were given for illustration.

Implications of the main result and open questions. Benamou [4] and Bonafini and Schmitzer [8] establish convergence of domain decomposition to the global minimizer in their respective settings (unregularized, regularized) at finite scale n . In this article we give examples for both where convergence fails asymptotically as $n \rightarrow \infty$. The former does not address the speed of convergence and the upper bound on the convergence speed given by the latter does not accurately describe the behaviour on the squared Euclidean distance. The limit dynamics derived in this article appear to be an appropriate description of the behaviour in this setting and suggests that at finite scale n the number of required iterations for approximate convergence to a stationary point will be proportional to n . Of course, it should be pointed out that this hinges on the strong Assumption 1, which can however be reduced to weaker assumptions that can be examined numerically. It will be studied elsewhere in more detail, see Remark 1.

In addition to Assumption 1, there are several open questions related to the behaviour of the trajectories π_t as $t \rightarrow \infty$, e.g. what the form of the limit (in time) will be, under what conditions it will be a minimizer of the problem, and how fast it will approach this limit (see Sect. 1.2 for some more details).

Conclusion. In summary we consider the results of this article as an important step in the geometric convergence analysis of the domain decomposition algorithm. The

obtained limit dynamics and their relation to global optimality should be studied further, both for their implications for numerical algorithms and in their own right as a new type of minimization-driven dynamics on measures.

Acknowledgements This work was supported by the Emmy Noether programme of the DFG.

Funding Open Access funding enabled and organized by Projekt DEAL.

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/>.

Appendix A: Proof of Lemma 1

The proof hinges on the Lebesgue differentiation theorem which we gather from [22].

Theorem 4 (Lebesgue points [22, Theorems 7.7, 7.10, Definition 7.9]) *For $x \in X$ we say a sequence $(E_n)_n$ of Borel sets in X shrinks nicely to x if there exists $\alpha > 0$ and radii $(r_n)_n$ such that*

$$\lim_{n \rightarrow \infty} r_n = 0, \quad E_n \subset B(x, r_n) \quad \text{and} \quad \mathcal{L}(E_n) \geq \alpha \cdot \mathcal{L}(B(x, r_n)).$$

Note that x may not belong to E_n .

Assume $f \in L^1(X)$. Then \mathcal{L} -a.e. $x \in X$ is a Lebesgue point of f and for every such x it holds

$$f(x) = \lim_{n \rightarrow \infty} \frac{1}{\mathcal{L}(E_n)} \int_{E_n} f(x') d\mathcal{L}(x').$$

if the sets $(E_n)_n$ shrink nicely to x .

Proof of Lemma 1 By assumption $\mu \ll \mathcal{L}$, so μ has a density with respect to \mathcal{L} and $\frac{d\mu}{d\mathcal{L}} \in L^1(X)$. Therefore

$$\lim_{n \rightarrow \infty} \frac{1}{\mathcal{L}(E_n)} \int_{E_n} \frac{d\mu}{d\mathcal{L}}(x') d\mathcal{L}(x') = \frac{d\mu}{d\mathcal{L}}(x) > 0 \quad (\text{A.1})$$

for μ -a.e. $x \in X$ and $(E_n)_n$ a corresponding sequence shrinking nicely to x . In particular, if $(A_n)_n$ and $(B_n)_n$ are two sequences shrinking nicely to x such that $\mathcal{L}(A_n) = a/n^d$

and $\mathcal{L}(B_n) = b/n^d$ for some $a, b > 0$, then

$$\lim_{n \rightarrow \infty} \frac{\mu(A_n)}{\mu(B_n)} = \lim_{n \rightarrow \infty} \frac{\overbrace{\mathcal{L}(A_n)}^{a/b}}{\mathcal{L}(B_n)} \overbrace{\frac{\frac{1}{\mathcal{L}(A_n)} \int_{A_n} \frac{d\mu}{d\mathcal{L}}(x') d\mathcal{L}(x')}{\frac{1}{\mathcal{L}(B_n)} \int_{B_n} \frac{d\mu}{d\mathcal{L}}(x') d\mathcal{L}(x')}}^{\text{Tends to 1}} = \frac{a}{b}. \tag{A.2}$$

Consider the first scheme, (i), with $\mu^n = \mu$ for all n . Choose a point $x \in X$ that is a Lebesgue point of $\frac{d\mu}{d\mathcal{L}}$ with $\frac{d\mu}{d\mathcal{L}}(x) > 0$ (μ -a.e. $x \in X$ fulfills this) and choose a time $t \in \mathbb{R}_+$. Then for any measurable set $A \subset Z$, the sequence $(A_n)_n$ defined by $A_n = (S_{t,x}^n)^{-1}(A)$ shrinks nicely to x , and so does the sequence $(B_n)_n$ defined by $B_n = (S_{t,x}^n)^{-1}(Z) = X_{J_{t,x}^n}$. Thus, by (A.2)

$$\lim_{n \rightarrow \infty} \sigma_{t,x}^n(A) = \lim_{n \rightarrow \infty} \frac{\mu((S_{t,x}^n)^{-1}(A))}{m_J^n} = \lim_{n \rightarrow \infty} \frac{\mu(A_n)}{\mu(B_n)} = \frac{1}{2^d} \mathcal{L}(A)$$

from which we conclude that $\sigma = \frac{1}{2^d} \mathcal{L} \llcorner Z$.

Consider now the second scheme, (ii), where the mass of each basic cell $i \in I^n$ is collapsed to its center x_i^n , so that $\mu^n = \sum_{i \in I^n} m_i^n \delta_{x_i^n}$. Let $x \in X \setminus \partial X$ be a Lebesgue point of $\frac{d\mu}{d\mathcal{L}}$ with $\frac{d\mu}{d\mathcal{L}}(x) > 0$ (this again holds for μ -a.e. $x \in X$). Fix $t \in \mathbb{R}_+$. For any $n \in 2\mathbb{N}$ denote $J^n := J_{t,x}^n$ (Definition 2). For n sufficiently large, J^n will contain 2^d basic cells: this holds whenever J^n is part of \mathcal{J}_A^n and it will eventually hold for $J^n \in \mathcal{J}_B^n$ since $x \in X \setminus \partial X$. Then

$$\sigma_{t,x}^n = \sigma_{J^n}^n = \sum_{b \in \{-1,1\}^d} \frac{m_{i(J^n,b)}^n}{m_{J^n}^n} \delta_{b/2} = \sum_{b \in \{-1,1\}^d} \frac{\mu(X_{i(J^n,b)}^n)}{\mu(X_{J^n}^n)} \delta_{b/2} \tag{A.3}$$

where $i(J^n, b)$ is the basic cell in composite cell J^n whose center x_i^n is at $x_J^n + b/2n$. Choose any $b \in \{-1, 1\}^d$ and define $A_n^b = X_{i(J^n,b)}^n$ and $B_n = X_{J^n}^n$. The sequences $(A_n^b)_n$ and $(B_n)_n$ shrink nicely to x and satisfy $\mathcal{L}(A_n^b) = 1/n^d$ and $\mathcal{L}(B_n) = 2^d/n^d$, so by (A.2)

$$\lim_{n \rightarrow \infty} \frac{\mu(X_{i(J^n,b)}^n)}{\mu(X_{J^n}^n)} = \frac{1}{2^d}, \tag{A.4}$$

and thus $(\sigma_{t,x}^n)_n$ converges as $n \rightarrow \infty$ to

$$\sigma = \sum_{b \in \{-1,+1\}^d} \frac{1}{2^d} \delta_{b/2}. \tag{A.5}$$

for μ -a.e. $x \in X$ and all $t \in \mathbb{R}_+$.

For the last scheme (iii), at μ -a.e. $x \in X$ and every $t \in \mathbb{R}_+$, the sequence $\sigma_{t,x}^n$ has the same weak* cluster points as in the scheme (i), since $s^n \rightarrow \infty$. Thus the result follows from the previous part.

Let us conclude showing that for μ^n a regular discretization sequence, the limit σ assigns mass $1/2^d$ to each of the ‘quadrants’

$$Z_b := \{z \in Z \mid \text{sign}(z_\ell) = b_\ell \text{ for all } \ell = 1, \dots, d\} \tag{A.6}$$

with $b \in \{-1, 1\}^d$. This follows quickly analogous to (A.4): for μ -a.e. $x \in X$ and all $t \in \mathbb{R}_+$, denoting $J^n = J(t, x, n)$, we obtain

$$\sigma(Z_b) = \frac{\sigma(Z_b)}{\sigma(Z)} = \lim_{n \rightarrow \infty} \frac{\sigma_{J^n}^n(Z_b)}{\sigma_{J^n}^n(Z)} = \lim_{n \rightarrow \infty} \frac{m_{J^n}^n \sigma_{J^n}^n(Z_b)}{m_{J^n}^n \sigma_{J^n}^n(Z)} = \lim_{n \rightarrow \infty} \frac{\mu(X_{i(J^n, b)}^n)}{\mu(X_{J^n}^n)} = \frac{1}{2^d}. \tag{A.7}$$

□

Appendix B: Block approximation

For the construction of recovery sequences in Sect. 4.3 we use the *block approximation* [10] of a transport plan. The idea is to cover the product space with a grid of small cubes and then replace the mass of the original coupling within any cube with a suitable product measure (on that cube). This preserves the marginals and by controlling the size of the cubes one may balance between the entropy and the transport term. For simplicity, we compactly gather the required results and provide a sketch of proof.

Lemma 5 (Block approximation) *Let $\Omega \subset \mathbb{R}^d$ be compact, $\mu, \nu \in \mathcal{M}_1(\Omega)$, $\gamma \in \Pi(\mu, \nu)$, $L > 0$. For each $k \in \mathbb{Z}^d$ let $Q_k := [k_1, k_1 + 1) \times \dots \times [k_d, k_d + 1)$ and $Q_k^L := L \cdot Q_k$. We define the block approximation of γ at scale L as*

$$\gamma_L := \sum_{\substack{j, k \in \mathbb{Z}^d: \\ \mu(Q_j^L) > 0, \nu(Q_k^L) > 0}} \frac{\gamma(Q_j^L \otimes Q_k^L)}{\mu(Q_j^L) \cdot \nu(Q_k^L)} \cdot (\mu \llcorner Q_j^L) \otimes (\nu \llcorner Q_k^L). \tag{B.1}$$

We find that

$$\gamma_L \in \Pi(\mu, \nu), \quad W_{\Omega \times \Omega}(\gamma, \gamma_L) \leq L \cdot \sqrt{2d}, \quad \text{KL}(\gamma_L | \mu \otimes \nu) \leq C - 2d \log L.$$

for a constant $C < \infty$ (only depending on d and the diameter of Ω).

Proof In [10], Ω was not compact, but μ and ν were assumed to be Lebesgue-absolutely continuous. Ω compact allows for a much shorter, self-contained proof. $\gamma_L \in \Pi(\mu, \nu)$ follows from a direct computation as in [10, Proposition 2.10]. For all measurable $A \subset \Omega$ one has

$$\gamma_L(A \times \Omega) = \sum_{\substack{j, k \in \mathbb{Z}^d: \\ \mu(Q_j^L) > 0, \nu(Q_k^L) > 0}} \frac{\gamma(Q_j^L \otimes Q_k^L)}{\mu(Q_j^L)} \cdot (\mu \llcorner Q_j^L)(A) = \sum_{\substack{j \in \mathbb{Z}^d: \\ \mu(Q_j^L) > 0}} (\mu \llcorner Q_j^L)(A) = \mu(A)$$

where we used $[\gamma(Q_j^L \otimes Q_k^L) > 0] \Rightarrow [v(Q_k^L) > 0]$ and that the first marginal of γ is μ in the second equality. The same argument applies for the second marginal.

Transforming γ into γ_L only requires rearrangement of mass within each hypercube $Q_j^L \otimes Q_k^L$, which has diameter $L \cdot \sqrt{2d}$ in \mathbb{R}^{2d} . This provides the bound $W_{\Omega \times \Omega}(\gamma, \gamma_L) \leq L \cdot \sqrt{2d}$, cf. [10, Corollary 2.12].

For the entropy bound we adopt [10, Lemma 2.15], accounting for more general μ, v and considerably simplifying the last part due to compactness of Ω :

$$\begin{aligned} \text{KL}(\gamma_L \mid \mu \otimes v) &= \int_{\Omega \times \Omega} \log \left(\frac{d\gamma_L}{d(\mu \otimes v)} \right) d\gamma_L - \gamma_L(\Omega \times \Omega) + (\mu \otimes v)(\Omega \times \Omega) \\ &= \sum_{\substack{j,k \in \mathbb{Z}^d: \\ \mu(Q_j^L) > 0, v(Q_k^L) > 0}} \log \left(\frac{\gamma(Q_j^L \otimes Q_k^L)}{\mu(Q_j^L)v(Q_k^L)} \right) \gamma(Q_j^L \otimes Q_k^L) \end{aligned}$$

where we deduce from (B.1) that $\frac{d\gamma_L}{d(\mu \otimes v)} = \frac{\gamma(Q_j^L \otimes Q_k^L)}{\mu(Q_j^L)v(Q_k^L)}$ on $Q_j^L \otimes Q_k^L$ and γ_L and γ carry the same mass on $Q_j^L \otimes Q_k^L$. We continue:

$$\begin{aligned} &= \sum_{\substack{j,k \in \mathbb{Z}^d: \\ \mu(Q_j^L) > 0, v(Q_k^L) > 0}} \gamma(Q_j^L \otimes Q_k^L) \cdot \log(\gamma(Q_j^L \otimes Q_k^L)) \\ &\quad - \sum_{j \in \mathbb{Z}^d} \mu(Q_j^L) \log(\mu(Q_j^L)) - \sum_{k \in \mathbb{Z}^d} v(Q_k^L) \log(v(Q_k^L)) \end{aligned} \tag{B.2}$$

where we use the convention $0 \log 0 = 0$. The first term is less than or equal to zero. For the other two we observe that by compactness of Ω at most a finite number $N_L \in \mathbb{N}$ of non-zero terms can appear in each sum with $N_L \leq \tilde{C} \cdot L^{-d}$ for some \tilde{C} depending on d and the diameter of Ω . Using Jensen’s inequality and convexity of $\mathbb{R}_+ \ni s \mapsto \phi(s) := s \log(s)$ we find

$$\begin{aligned} \sum_{j \in \mathbb{Z}^d} \phi(\mu(Q_j^L)) &\geq N_L \cdot \phi \left(\frac{1}{N_L} \sum_{j \in \mathbb{Z}^d} \mu(Q_j^L) \right) \\ &= N_L \phi(1/N_L) = -\log(N_L) \geq -\log(\tilde{C}) + d \log(L). \end{aligned}$$

Applying this bound on the second and third term in (B.2) we arrive at the desired entropy bound for $C := 2 \log(\tilde{C})$. □

References

1. Ambrosio, L.: Metric space valued functions of bounded variation. Ann. Sc. Norm. Super. Pisa Cl. Sci. Ser. 4 **17**(3), 439–478 (1990)

2. Ambrosio, L., Fusco, N., Pallara, D.: *Functions of Bounded Variation and Free Discontinuity Problems*. Courier Corporation (2000)
3. Angenent, S., Haker, S., Tannenbaum, A.: Minimizing flows for the Monge–Kantorovich problem. *SIAM J. Math. Anal.* **35**(1), 61–97 (2003)
4. Benamou, J.-D.: A domain decomposition method for the polar factorization of vector-valued mappings. *SIAM J. Numer. Anal.* **32**(6), 1808–1838 (1994)
5. Benamou, J.-D., Froese, B.D., Oberman, A.M.: Numerical solution of the optimal transportation problem using the Monge–Ampère equation. *J. Comput. Phys.* **260**(1), 107–126 (2014)
6. Berman, R.J.: The Sinkhorn algorithm, parabolic optimal transport and geometric Monge–Ampère equations. *Numer. Math.* **145**, 771–836 (2020)
7. Billingsley, P.: *Convergence of Probability Measures*. Wiley Series in Probability and Statistics, 2nd edn. Wiley (1999)
8. Bonafini, M., Schmitzer, B.: Domain decomposition for entropy regularized optimal transport. *Numer. Math.* **149**, 819–870 (2021). <https://doi.org/10.1007/s00211-021-01245-0>
9. Brenier, Y.: Polar factorization and monotone rearrangement of vector-valued functions. *Commun. Pure Appl. Math.* **44**(4), 375–417 (1991)
10. Carlier, G., Duval, V., Peyré, G., Schmitzer, B.: Convergence of entropic schemes for optimal transport and gradient flows. *SIAM J. Math. Anal.* **49**(2), 1385–1418 (2017)
11. Cominetti, R., San Martin, J.: Asymptotic analysis of the exponential penalty trajectory in linear programming. *Math. Program.* **67**, 169–187 (1992)
12. Cuturi, M.: Sinkhorn distances: lightspeed computation of optimal transportation distances. In: *Advances in Neural Information Processing Systems*, vol. 26 (NIPS 2013), pp. 2292–2300 (2013)
13. Franklin, J., Lorenz, J.: On the scaling of multidimensional matrices. *Linear Algebra Appl.* **114–115**, 717–735 (1989)
14. Gangbo, W., McCann, R.J.: The geometry of optimal transportation. *Acta Math.* **177**(2), 113–161 (1996)
15. Kim, Y.-H., Streets, J., Warren, M.: Parabolic optimal transport equations on manifolds. *Int. Math. Res. Not.* **2012**(19), 4325–4350 (2012)
16. Kitagawa, J.: A parabolic flow toward solutions of the optimal transportation problem on domains with boundary. *J. Reine Angew. Math.* **672**, 127–160 (2012)
17. Kitagawa, J., Mérigot, Q., Thibert, B.: Convergence of a Newton algorithm for semi-discrete optimal transport. *J. Eur. Math. Soc.* **21**, 2603 (2019)
18. Léonard, C.: From the Schrödinger problem to the Monge–Kantorovich problem. *J. Funct. Anal.* **262**(4), 1879–1920 (2012)
19. Lévy, B.: A numerical algorithm for L2 semi-discrete optimal transport in 3D. *ESAIM Math. Model. Numer. Anal.* **49**(6), 1693–1715 (2015)
20. Mérigot, Q.: A multiscale approach to optimal transport. *Comput. Gr. Forum* **30**(5), 1583–1592 (2011)
21. Peyré, G., Cuturi, M.: Computational optimal transport. *Found. Trends Mach. Learn.* **11**(5–6), 355–607 (2019)
22. Rudin, W.: *Real and Complex Analysis*, 3rd edn. McGraw-Hill Inc, USA (1987)
23. Santambrogio, F.: *Optimal Transport for Applied Mathematicians*, vol. 87 of *Progress in Nonlinear Differential Equations and Their Applications*. Birkhäuser, Boston (2015)
24. Schmitzer, B.: Stabilized sparse scaling algorithms for entropy regularized transport problems. *SIAM J. Sci. Comput.* **41**(3), A1443–A1481 (2019)
25. Schmitzer, B., Schnörr, C.: A hierarchical approach to optimal transport. In: *Scale Space and Variational Methods (SSVM 2013)*, pp. 452–464 (2013)
26. Villani, C.: *Optimal Transport: Old and New*, vol. 338 of *Grundlehren der mathematischen Wissenschaften*. Springer (2009)

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