Distributed Online Optimization for Multi-Agent Optimal Transport *

Vishaal Krishnan Sonia Martínez †

Abstract

In this work, we propose and investigate a scalable, distributed iterative algorithm for large-scale optimal transport of collectives of autonomous agents. We formulate the problem as one of steering the collective towards a target probability measure while minimizing the total cost of transport, with the additional constraint of distributed implementation imposed by a range-limited network topology. Working within the framework of optimal transport theory, we realize the solution as an iterative transport based on a proximal point algorithm. At each stage of the transport, the agents implement an online, distributed primal-dual algorithm to obtain local estimates of the Kantorovich potential for optimal transport from the current distribution of the collective to the target distribution. Using these estimates as their local objective functions, the agents then implement the transport by a proximal point algorithm. This two-step process is carried out recursively by the collective to converge asymptotically to the target distribution. We analyze the behavior of the algorithm via a candidate system of feedback interconnected PDEs for the continuous time and $N \to \infty$ limit, and establish the asymptotic stability of this system of PDEs. We then test the behavior of the algorithm in simulation.

1 Introduction

The problem of transport of multi-agent collectives arises naturally in various settings, from the modeling of cell populations in biology, to engineering applications of coverage control and deployment in robotics and mobile sensing networks [11, 14, 34]. As these scenarios involve physical transport of resources, there is an associated cost of transport owing to energy considerations. Optimal transport theory [44], which deals with the problem of rearranging probability measures while minimizing the cost of transport, presents an appropriate theoretical framework for these problems. Another consideration in the multi-agent setting is that of scalability of implementation when the size of the collective increases, which underlines the need for distributed algorithms. In the engineering context, the development of low-cost sensor, communication and computational systems makes foreseeable in the near future the deployment of large collectives of robots in diverse areas such as remote monitoring, manufacturing, and construction. Consequently, the emphasis on distributed implementation is

*A preliminary version of this work appeared in the proceedings of the IEEE Conference on Decision and Control 2018, as [29]

†The authors are with the Department of Mechanical and Aerospace Engineering, University of California at San Diego, La Jolla CA 92093 USA (email: v6krishn@ucsd.edu; soniamd@ucsd.edu).

rather stringent, as such a drastic increase in network size would render centralized implementations unsuitable. Examples of this are also pervasive in biology, and the exact mechanisms used by biological systems to achieve scalability remains an active area of research. From a theoretical perspective, as the number of agents increases, the design and analysis of efficient distributed transport laws poses new challenges, starting with the choice of appropriate mathematical abstractions. The need for parsimonious descriptions of the collectives, along with the fact that tasks for these systems are more likely to be specified at a high level, calls for the use of macroscopic models. We place this work in the above context as an effort to address the difficulties arising in control design and analysis for large-scale optimal transport of multi-agent collectives.

The applications of optimal transport in image processing and various engineering domains has motivated a search for efficient computational methods for the optimal transport problem, and we refer the reader to [38] for a comprehensive account. Entropic regularization of the Kantorovich formulation has been an efficient tool for approximate computation of the optimal transport cost using the Sinkhorn algorithms [15], [17]. Data-driven approaches to the computation of the optimal transport cost between two distributions from their samples have been investigated in [30, 43], and with an eye towards large-scale problems in [24], [42], [33]. A related problem of computation of Wasserstein barycenters was addressed in [16]. Optimal transport from continuous to discrete probability distributions has been studied under the name of semi-discrete optimal transport, with connections to the problem of optimal quantization of probability measures, in [10]. While computational approaches to optimal transport often work with the static, Monge or Kantorovich formulations of the problem, investigations involving dynamical formulations was initiated by [7], where the authors recast the L^2 Monge-Kantorovich mass transfer problem in a fluid mechanics framework. This largely owes to notion of displacement interpolation originally introduced in [32]. The underlying rationale is that the optimal transport cost defines a metric in the space of probability measures, which allows for the interpretation of optimal transport between two probability measures as transport along distance-minimizing geodesics connecting them. [37] and [8] are other works in this vein. The problem of optimal transport was also explored from a stochastic control perspective in [36] and [12], where the latter further explored connections to Schrodinger bridges. However, there has remained a gap in this literature with regard to distributed computation of optimal transport, which arises as a rather stringent constraint in multi-agent transport scenarios.

Markov Chain Monte Carlo (MCMC) methods [3, 25, 39] present another framework for the problem of rearranging probability measures, and can be traced back to early works by Metropolis [35] and Hastings [26]. MCMC methods involve the construction of a Markov chain with the target probability measure as its equilibrium measure, and yield samples of the target measure as $t \to \infty$. From a computational perspective, MCMC methods allow for the agents to be transported independently of one another, which results in a fully decentralized implementation. However, MCMC methods are inefficient with respect to the cost of transport. On the other hand, an optimal transport-based approach suffers from the need for a centralized implementation, as the optimal transport plan is computed using global information of the intial and target probability measures. This further motivates our search for scalable, distributed iterative algorithms that occupy the middle ground. We attempt to improve the cost of transport by imposing more structure to the set of agents in the form of a nearest-neighbor network and using the information from the neighbors to compute the successive iterates. From a computational standpoint, such an approach would neither be decentralized to the point of complete independence between agents as in the case of MCMC, nor would it be centralized as is typical of conventional optimal transport-based methods.

Transport problems in robotics and mobile sensing network applications arise in the form of coverage control and deployment objectives, where the underlying goal is to steer a group of robots towards a target coverage profile over a spatial region. Among the approaches to the coverage control and deployment problem for large-scale multi-agent systems are transport by synthesis of Markov transition matrices [4, 5, 18], the use of continuum models [20, 28] for transport, and coverage control by parameter tuning and/or boundary control of the reaction-advection-diffusion PDE [19, 23, 47]. We note, however, that despite the potential for the application of optimal transport ideas to the multi-agent setting, it has hitherto largely remained unsuccessful. The papers [22] and [6] represent attempts in this direction, while in the first paper the problem is formulated as one of optimal control, the second is placed in the framework of optimal transport. These works, however, present significant limitations either because they require centralized offline planning [22], or because of a need for costly computation and information exchange between agents [6]. This serves as a strong motivation for the development of a distributed iterative scheme for optimal transport in this paper.

In this work, we propose and investigate a scheme for large-scale optimal transport of multi-agent collectives based on a scalable, distributed online optimization. Working with a reduction of the Kantorovich duality for metric costs conformal to the Euclidean metric, we note that the Kantorovich potential is almost everywhere differentiable and obtain a bound on the norm of its gradient. We then obtain an iterative scheme for optimal transport of probability measures based on Kantorovich duality, showing it to be equivalent to optimal transport along geodesics. We propose a distributed primal-dual algorithm to be implemented online by the agents to obtain local estimates of the Kantorovich potential, which are then used as local objectives in a proximal algorithm for transport. In the continuous-time limit and as $N \to \infty$, we derive a PDE-based flow for optimal transport, and obtain convergence results for an online implementation of the transport. The paper contributes not only to the vast literature on computational methods for the optimal transport problem, but also presents a novel scalable, distributed approach to multi-agent optimal transport addressing a longstanding concern in the research on multi-agent systems. A preliminary version of this work appeared in [29].

The paper is organized as follows: Section 2 introduces the notation and mathematical preliminaries used in the paper. Section 3 provides a brief description of the Monge and Kantorovich formulations of optimal transport. The multi-stage iterative transport scheme is introduced in Section 4. In Section 5, we develop the algorithm for multi-agent optimal transport, based on distributed online optimization. In Section 6, we study analytically the behavior of the candidate PDE model for the $N \to \infty$ and continuous-time limit of the multi-agent optimal transport algorithm. Simulation results and a discussion of the numerical behavior of the multi-agent optimal transport algorithm are presented in Section 7. We conclude with a summary of this work and scope for future work in Section 8.

2 Notation and preliminaries

Let $\|\cdot\|:\mathbb{R}^d \to \mathbb{R}_{\geq 0}$ denote the Euclidean norm on $\mathbb{R}^d$ and $|\cdot|:\mathbb{R}\to\mathbb{R}_{\geq 0}$ the absolute value function. We denote by $\nabla = \left(\frac{\partial}{\partial x_1}, \dots \frac{\partial}{\partial x_n}\right)$ the gradient operator in $\mathbb{R}^d$. As a shorthand, we let $\frac{\partial}{\partial z}(\cdot) = \partial_z(\cdot)$ for a variable z. We denote by $C^k(\Omega)$ the space of k-times continuously differentiable functions on Ω. For any $x \in \Omega \subset \mathbb{R}^d$, we denote by $B_r^m(x)$ the closed d-ball of radius $r > 0$, with respect to a metric m, centered at x. Let $\partial\Omega$ denote the boundary of Ω, $\bar{\Omega} = \Omega \cup \partial\Omega$ its closure and $\mathring{\Omega} = \Omega \setminus \partial\Omega$ its interior with respect to the standard Euclidean topology. For $M \subseteq \Omega$, let the distance $d(x, M)$ of a point $x \in \Omega$ to the set M be given by $d(x, M) = \inf_{y\in M} \|x - y\|$. Let $1_M : \Omega \to \{0, 1\}$ be the indicator function on Ω for the subset M. We denote by $\langle f, g\rangle$ the inner product of functions $f, g : \Omega \to \mathbb{R}$ with respect to the Lebesgue measure, given by $\langle f, g\rangle = \int_\Omega fg\,\mathrm{dvol}$. Let $\mu \in \mathcal{P}(\Omega)$ be an absolutely continuous probability measure on $\Omega \subset \mathbb{R}^d$, with ρ the corresponding density function (where $\mathrm{d}\,\mu = \rho\,\mathrm{dvol}$), with vol being the Lebesgue measure. We denote by $\mathbb{E}_\mu$ the expectation w.r.t. the measure μ. Given a map $\mathcal{T} : \Omega \to \Gamma$ and a measure $\mu \in \mathcal{P}(\Omega)$, we let $\nu = \mathcal{T}_\#\mu$ denote the pushforward measure of μ by $\mathcal{T}$, where for a measurable set $\mathcal{B} \subset \mathcal{T}(\Omega)$, we have $\nu(\mathcal{B}) = \mathcal{T}_\#\mu(\mathcal{B}) = \mu(\mathcal{T}^{-1}(\mathcal{B}))$. Let $F : \mathcal{P}(\Omega) \to \mathbb{R}$ be a smooth real-valued function on the space of probability measures on $\Omega \subset \mathbb{R}^d$. We denote by $\frac{\delta F}{\delta\mu}(x)$ the derivative of F with respect to the measure μ, see [21], such that a perturbation $\delta\mu$ of the measure results in a perturbation $\delta F = \int_{\mathcal{X}} \frac{\delta F}{\delta \mu} d(\delta\mu)$. The L^p space of functions on a measurable space U is given by $L^p(U) = \{f : U \to \mathbb{R} \,|\, \|f\|_{L^p(U)} = \left(\int_U |f|^p\,\mathrm{dvol}\right)^{1/p} < \infty\}$, where $\|\cdot\|_{L^p(U)}$ is the L^p norm. Of particular interest is the L^2 space, or the space of square-integrable functions. In this paper, we denote by $\|f\|_{L^2(\Omega)}$ the L^2 norm of f with respect to the Lebesgue measure, and by $\|f\|_{L^2(\Omega,\mu)} = \left(\int_\Omega |f|^2\,\mathrm{d}\,\mu\right)^{1/2}$ the weighted L^2 norm. The Sobolev space $W^{1,p}(\Omega)$ is defined as $W^{1,p}(\Omega) = \{f : \Omega \to \mathbb{R} \,|\, \|f\|_{W^{1,p}} = \left(\int_\Omega |f|^p + \int_\Omega |\nabla f|^p\right)^{1/p} < \infty\}$. For two functions $f : \mathbb{R}\times\Omega \to \mathbb{R}$ and $g : \Omega \to \mathbb{R}$, denote $f(t, \cdot) \equiv f_t$ and further denote $f \to_{L^2} g$ the convergence in L^2 norm of f_t to g as $t \to \infty$, that is, $\lim_{t\to\infty} \|f_t - g\|_{L^2} = 0$. Convergence in H^1 norm is denoted similarly by $f \to_{H^1} g$.

We now state some well-known results that we will be used in the subsequent sections of this paper.

Lemma 1 (Divergence Theorem [13]). *For a smooth vector field* $\mathbf{F}$ *over a bounded open set* $\Omega \subseteq \mathbb{R}^d$ *with boundary* $\partial\Omega$*, the volume integral of the divergence* $\nabla\cdot\mathbf{F}$ *of* $\mathbf{F}$ *over* Ω *is equal to the surface integral of* $\mathbf{F}$ *over* $\partial\Omega$*:*

$$\int_\Omega (\nabla\cdot\mathbf{F})\ d\mu = \int_{\partial\Omega} \mathbf{F}\cdot\mathbf{n}\ dS, \tag{1}$$

where $\mathbf{n}$ *is the outward normal to the boundary and* dS *the measure on the boundary. For a scalar field* ψ *and a vector field* $\mathbf{F}$ *defined over* $\Omega \subseteq \mathbb{R}^d$*:*

$$\int_\Omega (\mathbf{F}\cdot\nabla\psi)\ d\mu = \int_{\partial\Omega} \psi(\mathbf{F}\cdot\mathbf{n})\ dS - \int_\Omega \psi(\nabla\cdot\mathbf{F})\ d\mu.$$

Lemma 2 (Rademacher's Theorem [31]). *Let* $\Omega \subset \mathbb{R}^d$ *be open and* $f : \Omega \to \mathbb{R}^m$ *be Lipschitz continuous. Then* f *is differentiable at almost every* $x \in \Omega$*.*

Lemma 3 (Poincaré-Wirtinger Inequality [31]). *For $p \in [1, \infty]$ and Ω, a bounded connected open subset of $\mathbb{R}^d$ with a Lipschitz boundary, there exists a constant C depending only on Ω and p such that for every function u in the Sobolev space $W^{1,p}(\Omega)$:*

$$\|u - u_\Omega\|_{L^p(\Omega)} \leq C \|\nabla u\|_{L^p(\Omega)},$$

where $u_\Omega = \frac{1}{|\Omega|} \int_\Omega u d\mu$, and $|\Omega|$ is the Lebesgue measure of Ω.

Lemma 4 (Rellich-Kondrachov Compactness Theorem [21]). *Let $\Omega \subset \mathbb{R}^d$ be open, bounded and such that $\partial\Omega$ is C^1. Suppose $1 \leq p < n$, then $W^{1,p}(\Omega)$ is compactly embedded in $L^q(\Omega)$ for each $1 \leq q < \frac{pn}{n-p}$. In particular, we have $W^{1,p}(\Omega)$ is compactly contained in $L^p(\Omega)$.*

Lemma 5 (LaSalle Invariance Principle [27, 45, 46]). *Let $\{\mathcal{P}(t) \,|\, t \in \mathbb{R}_{\geq 0}\}$ be a continuous semigroup of operators on a Banach space U (closed subset of a Banach space with norm $\|\cdot\|$), and for any $u \in U$, define the positive orbit starting from u at $t = 0$ as $\Gamma_+(u) = \{\mathcal{P}(t)u \,|\, t \in \mathbb{R}_{\geq 0}\} \subseteq U$. Let $V : U \to \mathbb{R}$ be a continuous Lyapunov functional on $\mathcal{G} \subset U$ for $\mathcal{P}$ (such that $\dot{V}(u) = \frac{d}{dt} V(\mathcal{P}(t)u) \leq 0$ in $\mathcal{G}$). Define $E = \{u \in \bar{\mathcal{G}} \,|\, \dot{V}(u) = 0\}$, and let $\tilde{E}$ be the largest invariant subset of E. If for $u_0 \in \mathcal{G}$, the orbit $\Gamma_+(u_0)$ is pre-compact (lies in a compact subset of U), then $\lim_{t \to +\infty} d_U(\mathcal{P}(t)u_0, \tilde{E}) = 0$, where $d_U(y, \tilde{E}) = \inf_{x \in \tilde{E}} \|y - x\|_U$ (where d_U is the distance in U).*

Macroscopic model of multi-agent collectives

Let the configuration of a collective be denoted by the tuple $(\mathcal{I}, \{x_i\}_{i \in \mathcal{I}}, \{\mathbf{v}_i\}_{i \in \mathcal{I}})$ consisting of agent indices, their positions and velocities (with $|\mathcal{I}| = N$). We assume that agents are distributed across $\Omega = \bar{D}$, where $D \subset \mathbb{R}^d$ is open, bounded and connected. That is, $x_i \in \Omega$ and $\mathbf{v}_i \in \mathbb{R}^d$ for all $i \in \mathcal{I}$. The dynamics for an agent i at $x_i \in \Omega$ is given by $\dot{x}_i = \mathbf{v}_i$.

We abstract a multi-agent collective at any instant t by means of a probability distribution $\mu_t \in \mathcal{P}(\Omega)$, where $\mathcal{P}(\Omega)$ is the space of absolutely continuous probability distributions on Ω. In the limit $N \to \infty$, it follows from the Glivenko-Cantelli theorem [9] that the discrete probability measure generated by the N samples $\{x_i(t)\}_{i=1}^N$ converges uniformly, almost surely to the underlying measure μ_t. This allows us to represent the configuration of the system by the distribution μ_t.

We now let $\mathbf{v}_i = \mathbf{v}(x_i)$, where $\mathbf{v}$ is a velocity field over Ω (with a no-flux boundary condition $\mathbf{v} \cdot \mathbf{n} = 0$ over $\partial\Omega$), and let Φ be the flow associated with this field, such that $\frac{\mathrm{d}}{\mathrm{d}\,t} \Phi_t(x) = \mathbf{v}(t, \Phi_t(x))$ and $\Phi_0(x) = x$, for $x \in \Omega$. The position at time t of a point starting from x at $t = 0$ and transported by the flow is represented by $\Phi_t(x) \in \Omega$. Let $\mu_t \in \mathcal{P}(\Omega)$ (with $\mathrm{d}\,\mu_t = \rho_t$ dvol) be a one-parameter family of probability measures generated by the flow Φ starting from μ_0 at $t = 0$. The evolution of the density function is then given by the continuity equation:

$$\partial_t \rho + \nabla \cdot (\rho \mathbf{v}) = 0. \tag{2}$$

We now introduce the notion of gradient flows in the space of probability measures.

Definition 1 (Gradient flows in the space of probability measures). *For a C^1 function $F : \mathcal{P}(\Omega) \to \mathbb{R}$, the transport* (2) *with $\mathbf{v} = -\nabla \left(\frac{\delta F}{\delta \mu} \right)$ is called a* gradient flow *on F.*

We refer the reader to [41, 44] for detailed treatments of gradient flows in the space of probability measures.

3 On the Monge and Kantorovich formulations of optimal transport

We begin this section with an overview of the Monge and Kantorovich formulations of optimal transport, followed by preliminary results used later in the paper.

Let $\mu, \nu \in \mathcal{P}(\Omega)$ be absolutely continuous probability measures on Ω. Let $c : \Omega \times \Omega \to \mathbb{R}_{\geq 0}$ be such that for $x, y \in \Omega$, $c(x, y)$ is the unit cost of transport from x to y. We now make the following assumptions on the cost c:

Assumption 1. *The cost c is continuous and is a metric on Ω conformal to the Euclidean metric (with strictly positive conformal factor $\xi \in C^1(\Omega)$).*

In the Monge (deterministic) formulation, the optimal cost of transporting the probability measure μ onto ν is defined as the infimum of the transport cost over the set of all maps for which ν is obtained as the pushforward measure of μ, as given below:

$$C_M(\mu, \nu) = \inf_{\substack{T:\Omega\to\Omega \\ T_\#\mu=\nu}} \int_{\Omega\times\Omega} c(x, T(x)) d\mu(x). \tag{3}$$

The Kantorovich formulation relaxes the above formulation by minimizing the transport cost over the set of joint probability measures $\Pi(\mu, \nu) \subset \mathcal{P}(\Omega \times \Omega)$, for which μ and ν are the respective marginals over Ω. The optimal transport cost from μ to ν is defined as follows:

$$C_K(\mu, \nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int_{\Omega\times\Omega} c(x, y)\ d\pi(x, y). \tag{4}$$

We now present the following lemma on the existence of minimizers to the Monge and Kantorovich formulations and the equivalence between them. We refer the reader to [1] for proofs.

Lemma 6 (Existence of minimizers)**.** *Under Assumption 1, there exists a minimizer π^* to the Kantorovich problem. Moreover, if the measure μ is atomless (i.e., $\mu(\{x\}) = 0$ for all $x \in \Omega$), the Monge formulation has a minimizer T^* and it holds that $\pi^* = (id, T^*)_\#\mu$.*

Following Lemma 6, we denote by $C(\mu, \nu) = C_M(\mu, \nu) = C_K(\mu, \nu)$ the optimal transport cost from μ to ν. We now present the following key result that the optimal transport cost C defines a metric on the space of probability measures $\mathcal{P}(\Omega)$:

Lemma 7 (Corollary 3.2, 3.3 [40])**.** *Under Assumption 1, the optimal transport cost $C : \mathcal{P}(\Omega) \times \mathcal{P}(\Omega) \to \mathbb{R}_{\geq 0}$ defines a metric on $\mathcal{P}(\Omega)$.*

Kantorovich duality

The Kantorovich formulation (4) allows the following dual formulation [44]:

$$\begin{aligned} K(\mu, \nu) = \sup_{\phi\in L^1(\Omega);\psi\in L^1(\Omega)} & \int_\Omega \phi(x) d\mu(x) + \int_\Omega \psi(y) d\nu(y) \\ \text{s.t} \quad & \phi(x) + \psi(y) \leq c(x, y), \quad \forall x, y \in \Omega. \end{aligned} \tag{5}$$

The maximizers of the above dual formulation are pairs of functions (ϕ, ψ), called Kantorovich potentials, which occur at the boundary of the inequality constraint, and satisfy the equations:

$$\phi(x) = \inf_{y \in \Omega} \left(c(x,y) - \psi(y) \right), \quad \psi(y) = \inf_{z \in \Omega} \left(c(z,y) - \phi(z) \right). \tag{6}$$

We refer to (ϕ, ψ) defined above as a c-conjugate pair, and write $\psi = \phi^c$ to denote that ψ is the conjugate of ϕ. We therefore have:

$$\phi(x) = \inf_{y \in \Omega} \left[c(x,y) - \inf_{z \in \Omega} \left(c(z,y) - \phi(z) \right) \right]. \tag{7}$$

The Kantorovich duality (5) can now be rewritten as:

$$K(\mu, \nu) = \sup_{\phi \in L^1(\Omega)} \int_\Omega \phi(x) d\mu(x) + \int_\Omega \phi^c(y) d\nu(y). \tag{8}$$

We recall the following lemma on the strong duality property of the Kantorovich formulation. We refer the reader to Theorem 5.10 in [44] for a detailed proof.

Lemma 8 (Theorem 5.10, [44])**.** *Strong duality holds for the Kantorovich formulation. In other words, the gap between the costs defined in the Kantorovich formulation* (4) *and its dual* (5) *is zero, i.e.,* $C(\mu, \nu) = K(\mu, \nu)$*.*

Under Assumption 1 on the transport cost function c, we can obtain a further reduction of the Kantorovich duality (8). The following key lemma allows for such a reduction:

Lemma 9. *Under Assumption 1 and from* (7)*, the conjugate of the Kantorovich potential satisfies* $\phi^c = -\phi$ *and* $|\phi(x) - \phi(y)| \le c(x,y)$ *for all* $x, y \in \Omega$*.*

Proof. From (6), we have:

$$\begin{aligned}
\phi(x) &= \inf_{y \in \Omega} \Big(c(x,y) - \inf_{z \in \Omega} \left(c(z,y) - \phi(z) \right) \Big) = \inf_{y \in \Omega} \sup_{z \in \Omega} \Big(c(x,y) - c(z,y) + \phi(z) \Big) \\
&\ge \inf_{y \in \Omega} \Big(c(x,y) - c(z,y) + \phi(z) \Big) = \inf_{y \in \Omega} \Big(c(x,y) - c(z,y) \Big) + \phi(z) \\
&\ge -c(x,z) + \phi(z),
\end{aligned}$$

where we have used the fact that c is a metric to obtain the final inequality (for any y, we have $c(x,y) - c(z,y) = c(x,y) - c(y,z) \ge -c(x,z)$, which implies that $\inf_{y \in \Omega} \left(c(x,y) - c(z,y) \right) \ge -c(x,z)$). Moreover, since the above inequality holds for any $x, z \in \Omega$, we have $|\phi(x) - \phi(z)| \le c(x,z)$.

Now, when $|\phi(x) - \phi(y)| \le c(x,y)$, we have that $-\phi(x) \le c(x,y) - \phi(y)$, which implies that $-\phi(x) \le \inf_y \left(c(x,y) - \phi(y) \right) = \phi^c(x)$. Equivalently, we obtain the relation $\phi(x) \ge -\phi^c(x)$.

Similarly, from (6) $\phi^c(x) = \inf_y c(x,y) - \phi(y)$, we obtain $\phi^c(x) \le c(x,y) - \phi(y)$. By setting $y = x$ in the above inequality, and using $c(x,x) = 0$ we get $\phi(x) \le -\phi^c(x)$.

In all, we have that $\phi^c(x) = -\phi(x)$ when $|\phi(x) - \phi(y)| \le c(x,y)$. □

Following Lemma 9, we can now reduce the Kantorovich duality (8) to obtain:

$$\begin{aligned}
K(\mu, \nu) &= \sup_{\phi \in \mathcal{L}(\Omega)} \mathbb{E}_\mu [\phi] - \mathbb{E}_\nu [\phi], \\
\text{where } \; \mathcal{L}(\Omega) &= \{ \phi \in L^1(\Omega) : |\phi(x) - \phi(y)| \le c(x,y), \;\; \forall\, x, y \in \Omega \}.
\end{aligned} \tag{9}$$

Remark 1. *We note from* (9) *that functions* $\phi \in \mathcal{L}(\Omega)$ *are Lipschitz continuous (since* c *is conformal to the Euclidean metric from Assumption 1, and* Ω *is compact). It then follows from Rademacher's theorem (in Lemma 2) that* ϕ *is differentiable* μ*-almost everywhere in* Ω*. Moreover, its (pointwise a.e.) derivative is equal to its weak derivative, and we interpret the derivative of the Kantorovich potential in the weak sense in the rest of the paper. Moreover, we have that the Kantorovich potential* ϕ *is differentiable at every* $x \in \Omega$ *that is not a fixed point of the optimal transport map* T^* *[40].*

Furthermore, we would like to obtain a bound on the gradient of functions in the set $\mathcal{L}(\Omega)$, with the added assumption that they are everywhere differentiable. To this end, we characterize the set $\mathcal{L}(\Omega) \cap \mathcal{C}^1(\Omega)$ through the following lemma:

Lemma 10. *Under Assumption 1, the set* $\mathcal{L}(\Omega) \cap \mathcal{C}^1(\Omega)$ *is given by:*

$$\mathcal{L}(\Omega) \cap \mathcal{C}^1(\Omega) = \left\{ \phi \in L^1(\Omega) \cap \mathcal{C}^1(\Omega) : |\nabla \phi| \leq \xi \text{ in } \Omega \right\}. \tag{10}$$

Proof. Let $\phi \in \mathcal{L}(\Omega) \cap \mathcal{C}^1(\Omega)$, and $x, y \in \mathring{\Omega}$ with $x \neq y$ such that the line segment joining x and y is contained in Ω. By the Mean Value Theorem and the definition of $\mathcal{L}(\Omega)$ in (9), for some $m \in [0, 1]$, we get:

$$\frac{|\phi(y) - \phi(x)|}{|y - x|} = \frac{|\nabla \phi((1 - m)x + my) \cdot (y - x)|}{|y - x|} \leq \frac{c(x, y)}{|y - x|}.$$

With $y = x + tv$, where $v \in \mathcal{T}_x\Omega$ (tangent space of Ω at $x \in \Omega$), in the limit $t \to 0$, we get:

$$\frac{|\nabla \phi(x) \cdot v|}{|v|} \leq \xi(x),$$

where the above inequality holds for all $v \in T_x\Omega$ and ξ is the conformal factor for the metric c w.r.t the Euclidean metric, which implies that $|\nabla \phi(x)| \leq \xi(x)$ for any $x \in \mathring{\Omega}$.

Now, to prove the converse, we suppose that $|\nabla \phi(x)| \leq \xi(x)$ for any $x \in \mathring{\Omega}$. For $x, y \in \mathring{\Omega}$ with $x \neq y$, along the geodesic γ (w.r.t the metric c) joining x and y, we have:

$$\begin{aligned} |\phi(y) - \phi(x)| &= \left| \int_0^1 \nabla \phi(\gamma(t)) \cdot \dot{\gamma}(t)\ dt \right| \leq \int_0^1 |\nabla \phi(\gamma(t))|\ |\dot{\gamma}(t)|\ dt \\ &\leq \int_0^1 \xi(\gamma(t))\ |\dot{\gamma}(t)|\ dt = c(x, y) \end{aligned}$$

□

We now define the restricted Kantorovich duality as follows:

$$K(\mu, \nu) = \sup_{\phi \in \mathcal{L}(\Omega) \cap \mathcal{C}^1(\Omega)} \mathbb{E}_\mu[\phi] - \mathbb{E}_\nu[\phi], \tag{11}$$

where it is restricted in the sense that the constraint set is $\mathcal{L}(\Omega) \cap \mathcal{C}^1(\Omega)$ as opposed to $\mathcal{L}(\Omega)$ as given in (9).

4 Iterative scheme for multi-stage optimal transport

In this section, we establish a framework for multi-stage optimal transport of probability measures.

Let $\mu_0 \in \mathcal{P}(\Omega)$ be a given initial probability measure and $\mu^* \in \mathcal{P}(\Omega)$ the target probability measure. Our objective is to optimally transport μ_0 onto μ^* by an iterative scheme. To this end, we begin by constructing a finite sequence $\{\mu_k\}_{k=1}^K$ such that $\mu_K = \mu^*$, and carrying out optimal transport in stages $\{\mu_{k-1} \to \mu_k\}_{k=1}^K$. The net cost of transport along the sequence would then be given by $\sum_{k=1}^K C(\mu_{k-1}, \mu_k)$, the sum of the (optimal) stage costs. We now have the following lemma on the retrieval of the optimal transport cost:

Lemma 11. *Given atomless probability measures $\mu_0, \mu^* \in \mathcal{P}(\Omega)$, the cost of optimal transport from μ_0 to μ^* satisfies:*

$$C(\mu_0, \mu^*) = \min_{\substack{(\mu_1,\ldots,\mu_K) \\ \mu_k \in \mathcal{P}(\Omega) \\ \mu_K = \mu^*}} \sum_{k=1}^K C(\mu_{k-1}, \mu_k) \tag{12}$$

Proof. We begin by noting that there clearly exists at least one minimizing sequence for the optimization problem (12) (the trivial sequence $\mu_k = \mu^*$ for all $k = 1, \ldots, T$, minimizes the cost).

From the Monge formulation (3) and Lemma 6, we have:

$$C(\mu_0, \mu^*) = \min_{\substack{T:\Omega \to \Omega \\ T_\# \mu_0 = \mu^*}} \int_\Omega c(x, T(x)) \, d\mu_0(x),$$

and let T^* be a minimizing map above. Let T_0 be the identity map on Ω and let $\{T_k\}_{k=1}^K$ be a sequence of maps on Ω such that $T_K \circ \ldots \circ T_0 = T^*$, with $\mu_k = (T_k \circ \ldots \circ T_0)_\# \mu_0 = T_{k\#} \ldots T_{0\#} \mu_0$. Since c is a metric, we have:

$$c(x, T^*(x)) \leq \sum_{k=1}^K c(T_{k-1} \circ \ldots \circ T_0(x), T_k \circ \ldots \circ T_0(x)).$$

It then follows that:

$$\begin{aligned}
C(\mu_0, \mu^*) &= \int_\Omega c(x, T^*(x)) \, d\mu_0(x) \\
&\leq \int_\Omega \sum_{k=1}^K c(T_{k-1} \circ \ldots \circ T_0(x), T_k \circ \ldots \circ T_0(x)) d\mu_0(x) \\
&= \sum_{k=1}^K \int_\Omega c(T_{k-1} \circ \ldots \circ T_0(x), T_k \circ \ldots \circ T_0(x)) d\mu_0(x) \\
&= \sum_{k=1}^K \int_\Omega c(x, T_k(x)) d\mu_{k-1}(x) \\
&= \sum_{k=1}^K C(\mu_{k-1}, \mu_k).
\end{aligned}$$

We also have:

$$c(x, T^*(x)) = \min_{\substack{T_1,\dots,T_K \\ T_k:\Omega\to\Omega \\ T_K\circ\ldots\circ T_0 = T^*}} \sum_{k=1}^{K} c(T_{k-1}\circ\ldots\circ T_0(x), T_k\circ\ldots\circ T_0(x)),$$

where the minimum is attained when the point $T_k \circ \ldots \circ T_0(x)$ lies on the geodesic from $T_{k-1}\circ \ldots \circ T_0(x)$ to $T^*(x)$. This can be seen from the fact that for any $x_1, x_2 \in \Omega$, $z^* \in \arg\min_{z\in\Omega} c(x_1, z) + c(z, x_2)$ lies on the geodesic from x_1 to x_2. Thus, we get:

$$C(\mu_0, \mu^*) = \min_{\substack{T_1,\dots,T_K \\ T_k:\Omega\to\Omega \\ T_K\circ\ldots\circ T_0 = T^* \\ T_{k\#}\mu_{k-1}=\mu_k}} \sum_{k=1}^{K} C(\mu_{k-1}, \mu_k).$$

We further note that any minimizing sequence $\{\mu_k\}_{k=1}^K$ must be generated by a sequence of maps $\{T_k\}_{k=1}^K$ such that for any $x \in \Omega$, $T_k\circ\ldots\circ T_0(x)$ lies on the geodesic from $T_{k-1}\circ\ldots\circ T_0(x)$ to $T^*(x)$, which yields (12). □

From the set of minimizing sequences characterized by Lemma 11, we are interested in those sequences for which the individual stage costs are upper bounded by an $\epsilon > 0$. We thereby consider the following optimization-based iterative scheme to generate a minimizing sequence:

$$\begin{aligned} \mu_{k+1} \in \arg\min_{\nu\in\mathcal{P}(\Omega)} \quad & C(\mu_k, \nu) + C(\nu, \mu^*) \\ \text{s.t.} \quad & C(\mu_k, \nu) \le \epsilon, \end{aligned} \tag{13}$$

where the iterative scheme (13) additionally satisfies the constraint $\lim_{k\to\infty} \mu_k = \mu^*$. Now, let T_k^* be an optimal transport map from μ_k to μ^*. We now construct the following optimization-based iterative process:

$$\begin{aligned} x(k+1) \in \arg\min_{z\in\Omega} \quad & c(x(k), z) + c(z, T_k^*(x(k))) \\ \text{s.t.} \quad & c(x(k), z) \le \epsilon, \end{aligned} \tag{14}$$

where $x(k+1)$ obtained from the above process lies on the geodesic connecting $x(k)$ and $T_k^*(x(k))$. We now have the following lemma on the connection between the process (14) and the iterative scheme (13):

Lemma 12. *The law of the process* (14)*, when $x(0) \sim \mu_0$, evolves according to* (13)*.*

Proof. This result follows from the arguments in the proof of Lemma 11. □

Following Lemma 12, it is clear that if we can compute the optimal transport map T_k^*, then (14) defines an iterative scheme for multi-stage optimal transport from an initial μ_0 to μ^*. We achieve this equivalently using the Kantorovich duality via the following process:

$$x(k+1) \in \arg\min_{z\in B_\epsilon^c(x(k))} c(x(k), z) + \phi_{\mu_k\to\mu^*}(z), \tag{15}$$

We recall that $B_\epsilon^c(x(k))$ is the closed ϵ-ball with respect to the metric c, centered at $x(k)$. The following lemma establishes that the processes (14) and (15) are equivalent.

Lemma 13. *The processes* (14) *and* (15) *are equivalent. The equivalence is in the sense that the sets of minimizers in* (14) *and* (15) *are equal.*

Proof. We recall from (6) and Lemma 9 that for the transport $\mu_k \to \mu^*$, and for any $x \in \Omega$, we have:

$$\phi_{\mu_k \to \mu^*}(x) = \inf_{y \in \Omega} c(x, y) + \phi_{\mu_k \to \mu^*}(y). \tag{16}$$

Also, for any $x, y \in \Omega$, we have the inequality $\phi_{\mu_k \to \mu^*}(x) \leq c(x, y) + \phi_{\mu_k \to \mu^*}(y)$. This implies in particular that for any transport map T_k from μ_k to μ^*, we get $\phi_{\mu_k \to \mu^*}(x) \leq c(x, T_k(x)) + \phi_{\mu_k \to \mu^*}(T_k(x))$. It then follows that:

$$\begin{aligned}\int_\Omega (\phi_{\mu_k \to \mu^*}(x) - \phi_{\mu_k \to \mu^*}(T_k(x)))\, d\mu_k(x) &= \int_\Omega \phi_{\mu_k \to \mu^*} d\mu_k - \int_\Omega \phi_{\mu_k \to \mu^*} d\mu^* \\ &\leq \int_\Omega c(x, T_k(x)) d\mu_k(x).\end{aligned}$$

We see that the LHS is the optimal transport cost obtained from the Kantorovich dual formulation, while an infimum over the RHS w.r.t. T_k would again yield the optimal transport cost from the Monge formulation and an equality would then be attained. Therefore, we get that the equality is attained when $T_k = T_k^*$, the corresponding optimal transport map from μ_k to μ^*. Thus, we infer that $\phi_{\mu_k \to \mu^*}(x) = c(x, T_k^*(x)) + \phi_{\mu_k \to \mu^*}(T_k^*(x))$ μ_k-almost everywhere in Ω. Since $c(x, T_k^*(x)) = c(x, z) + c(z, T_k^*(x))$ for any (and only) z on the geodesic from x to $T_k^*(x)$, we can write:

$$\phi_{\mu_k \to \mu^*}(x) - \phi_{\mu_k \to \mu^*}(z) + \phi_{\mu_k \to \mu^*}(z) - \phi_{\mu_k \to \mu^*}(T_k^*(x)) = c(x, z) + c(z, T_k^*(x)),$$

which implies that:

$$\begin{aligned}[\phi_{\mu_k \to \mu^*}(x) - \phi_{\mu_k \to \mu^*}(z) - c(x, z)] + [\phi_{\mu_k \to \mu^*}(z) - \phi_{\mu_k \to \mu^*}(T_k^*(x)) - c(z, T_k^*(x))]& \\ = 0.&\end{aligned}$$

Moreover, since the expressions on the LHS are each non-positive, and their sum is zero, we get that they are individually zero. In other words, for any (and only) z on the geodesic from x to $T_k^*(x)$ we get $\phi_{\mu_k \to \mu^*}(x) - \phi_{\mu_k \to \mu^*}(z) - c(x, z) = 0$, and these $z \in \Omega$ are in fact minimizers in (16). Therefore, set of minimizers obtained from (15) is essentially the segment of the geodesic from $x(k)$ to $T_k^*(x(k))$ contained in the ball $B_\epsilon^c(x(k))$ which is also the set of minimizers obtained from (14), establishing equivalence in this sense between the processes (14) and (15). □

5 Multi-agent optimal transport

Working within the framework established in Section 4, we develop in this section the algorithm for multi-agent optimal transport based on distributed online optimization.

Let $\{x_i(0)\}_{i=1}^N$ be the positions of the N agents, distributed independently and identically according to a probability measure μ_0. The idea is to transport the agents by the iterative scheme (15) to obtain $\{x_i(k)\}_{i=1}^N$ at any time k. Let $\widehat{\mu}_N(k) = \frac{1}{N}\sum_{i=1}^N \delta_{x_i(k)}$ be the empirical measure generated by the agents $\{x_i(k)\}_{i=1}^N$ at time k. To this end, we formulate a (finite) N-dimensional distributed optimization to be implemented by the agents to obtain local estimates of the Kantorovich potential. We approximate the true Kantorovich potential by a $\Phi^d : \mathbb{N} \times \Omega \to \mathbb{R}$ generated by an (finite) N-dimensional vector $\phi(k) = (\phi^1(k), \ldots, \phi^N(k)) \in \mathbb{R}^N$, such that $\Phi^d(k, x_i(k)) = \phi^i(k)$ for $i \in \{1, \ldots, N\}$ and $\Phi^d(k, x)$ for $x \in \Omega \backslash \{x_1(k), \ldots, x_N(k)\}$ is defined by a suitable multivariate interpolation. In particular, let $\{\mathcal{V}_i(k)\}_{i=1}^N$ be the Voronoi partition of Ω generated by $\{x_1(k), \ldots, x_N(k)\}$ w.r.t. the metric c, and $\Phi^d = \sum_{i=1}^N \phi^{\mathcal{V}_i(k)}$ (decomposed into a sum of N functions $\phi^{\mathcal{V}_i(k)}$ with supports $\mathcal{V}_i(k)$). We assume that at time k, the agents i, j corresponding to neighboring cells $\mathcal{V}_i(k)$ and $\mathcal{V}_j(k)$ are connected by an edge, which defines a connected graph $G(k) = \big(\{x_i(k)\}_{i=1}^N, E(k)\big)$ (where $E(k)$ is the edge set of the graph $G(k)$ at time k).

Dropping the index k (as is clear from context), the finite dimensional approximation of the Kantorovich duality (9) for the transport between $\widehat{\mu}_N$ and μ^*, restricted to the graph G, is given by:

$$\begin{aligned} \max_{(\phi^1, \ldots, \phi^N)} \quad & \sum_{i=1}^N \left(\frac{1}{N} \cdot \phi^i - \mathbb{E}_{\mu^*}[\phi^{\mathcal{V}_i}] \right) \\ \text{s.t.} \quad & |\phi^i - \phi^j| \le c(x_i, x_j), \quad \forall (i,j) \in E. \end{aligned} \tag{17}$$

We call (17) a restriction of (9) to the graph G because we only impose the constraint $|\phi^i - \phi^j| \le c(x_i, x_j)$ on neighbors i, j on the graph.

We solve the optimization problem (17) by a primal-dual algorithm, and its solution is used to update the agent positions by (15). We take Φ^d here to be a simple function, such that $\phi^{\mathcal{V}_i}(x) = \phi^i$ for $x \in \mathcal{V}_i$. The Lagrangian for the problem (17), with Φ^d a simple function and $c(x_i, x_j) = c_{ij}$, is given by:

$$L_d = \sum_{i=1}^N \phi^i \left(\frac{1}{N} - \mu^*(\mathcal{V}_i) \right) - \frac{1}{2} \sum_{i=1}^N \sum_{j \in \mathcal{N}_i} \lambda_{ij} \left(\left|\phi^i - \phi^j\right|^2 - c_{ij}^2 \right).$$

The primal-dual (primal-ascent, dual-descent) algorithm (with step size τ) is then given by:

$$\begin{aligned} \phi^i(l+1) &= \phi^i(l) - \tau \sum_{j \in \mathcal{N}_i} \lambda_{ij}(l) \left(\phi^i(l) - \phi^j(l) \right) + \left(\frac{1}{N} - \mu^*(\mathcal{V}_i) \right), \\ \lambda_{ij}(l+1) &= \max \left\{ 0, \lambda_{ij}(l) + \tau \left(\frac{1}{2} \left|\phi^i(l) - \phi^j(l)\right|^2 - c_{ij}^2 \right) \right\}, \quad \text{where } j \in \mathcal{N}_i. \end{aligned} \tag{18}$$

We note from the structure of the above algorithm that it renders itself to a distributed implementation by the agents, where agent i uses information from its neighbors $j \in \mathcal{N}_i$ to update ϕ^i and $\{\lambda_{ij}\}_{j \in \mathcal{N}_i}$. The primal algorithm is in fact a weighted Laplacian-based update.

At the end of every step $x_i(k) \mapsto x_i(k+1)$ from (15), the agent i assigns $\phi^i \leftarrow \Phi_k^d(x_i(k+1))$ as the initial condition for the primal algorithm (18) at the time step $k+1$ of the transport. Moreover, we are interested in an on-the fly implementation of the transport, in that the agents do not wait for convergence of the distributed primal-dual algorithm but carry out n iterations of it for every update step (15), as outlined formally in the algorithm below..

Algorithm 1 Multi-agent (on-the-fly) optimal transport

Input: Target measure μ^*, Transport cost $c(x, y)$, Bound on step size ϵ, Time step τ
For each agent i at time instant k of transport:

1: Obtain: Positions $x_j(k)$ of neighbors within communication/sensing radius r ($r \leq \text{diam}(\Omega)$, large enough to cover Voronoi neighbors)
2: Compute: Voronoi cell $\mathcal{V}_i(k)$, Mass of cell $\mu^*(\mathcal{V}_i(k))$, Voronoi neighbors $\mathcal{N}_i(k)$
3: Initialize: $\phi^i \leftarrow \Phi^d_{k-1}(x_i(k))$, $\lambda_{ij} \leftarrow \lambda_{ij}(k-1)$ (with $\Phi^d_0 = 0$, $\lambda_{ij}(0) = 0$)
4: Implement n iterations of primal-dual algorithm (18) (synchronously, in communication with neighbors $j \in \mathcal{N}_i$) to obtain $\phi^i(k)$, $\lambda_{ij}(k)$
5: Communicate with neighbors $j \in \mathcal{N}_i$ to obtain $\phi^j(k)$, construct local estimate of Φ^d_k by multivariate interpolation
6: Implement transport step (15) with local estimate of Φ^d_k (which approximates $\phi_{\mu_k \to \mu^*}$)

6 Analysis of PDE model

We investigate the behavior of the multi-agent transport by the update scheme (15) by studying the candidate system of PDEs for the continuous time and $N \to \infty$ limit. The results contained in this section are summarized below:

1. The candidate PDE model for transport in the continuous-time and $N \to \infty$ limit of the transport scheme (15) is derived formally in Section 6.1. The transport is described by the continuity equation (2) with the transport vector field (19).

2. In Section 6.2, we first derive the candidate PDE model (22) for the primal-dual algorithm (18) in the continuous-time and $N \to \infty$ limit. We then establish analytically that the solutions to (22) converge as $t \to \infty$ to the optimality condition of the Kantorovich duality, in Lemma 15.

3. Section 6.3 deals with the stability of the feedback interconnection between the transport PDE (continuity equation with the transport vector field (19)) and the primal-dual flow (22). Convergence of the probability density (as solutions to the transport PDE) to the target density in the limit $t \to \infty$, provided that the primal-dual flow is always at steady state, is first established in Theorem 1. On-the-fly implementation is considered next, and a convergence result is obtained under a second-order relaxation of the dual flow in Theorem 2. Although the primal-dual flow is asymptotically stable and the transport PDE under the action of the field (19) is asymptotically stable, the stability of the feedback interconnected system of PDEs, in general, does not follow. This motivates the second-order relaxation of the dual flow, and we are able to establish asymptotic stability of the feedback interconnected system through a backstepping control of the dual flow. We reserve the feedback interconnection of the transport PDE directly with the primal-dual flow (22) for investigated by numerical simulations in Section 7. Although we are only able to obtain analytical stability results for the feedback interconnection with a relaxed dual flow, we observe convergence in simulation of the original feedback interconnected system, which motivates us to conjecture that it is indeed stable.

6.1 Transport PDE

We recall that the continuous-time evolution of a probability density function is described by the continuity equation (2):

$$\partial_t \rho + \nabla \cdot (\rho \mathbf{v}) = 0,$$

where $\mathbf{v}$ is the underlying transport vector field. In what follows, we derive the transport vector field which is the candidate for the continuous-time limit of the update scheme (15). We assume that all the probability measures considered have the same support Ω.

Let $x \in \Omega$, $\mu \in \mathcal{P}(\Omega)$ and $x^+ \in \arg\min_{z \in B_\epsilon^c(x)} \ c(x,z) + \phi_{\mu\to\mu^*}(z)$, where x^+ is the update by the scheme (15). It then follows that:

$$c(x, x^+) + \phi_{\mu\to\mu^*}(x^+) \le \phi_{\mu\to\mu^*}(x),$$

where $x^+ \in B_\epsilon^c(x)$. We interpret the iterative scheme as a discrete-time dynamical system with uniform timestep Δt between successive instants, and derive the continuous-time limit by letting $\Delta t = g(\epsilon) \to 0$ (where $g : \mathbb{R} \to \mathbb{R}$ is a monotonically increasing function). We have that $\phi_{\mu\to\mu^*}$ is bounded and continuously differentiable, which implies that $\lim_{\epsilon\to 0} x^+ = x$ and $\lim_{x^+ \to x} \nabla\phi_{\mu\to\mu^*}(x^+) = \nabla\phi_{\mu\to\mu^*}(x)$. Let $\mathbf{v}(x) = \lim_{\Delta t\to 0} \frac{1}{\Delta t}(x^+ - x)$ (we note that this limit indeed exists), and we have:

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} c(x, x^+) \le \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left(\phi_{\mu\to\mu^*}(x) - \phi_{\mu\to\mu^*}(x^+) \right),$$

and it follows that:

$$\xi(x) \ |\mathbf{v}(x)| \le -\nabla\phi_{\mu\to\mu^*}(x) \cdot \mathbf{v}(x).$$

The above inequality is satisfied only if $\mathbf{v}(x) = -\alpha \nabla\phi_{\mu\to\mu^*}(x)$ when $\lambda \neq 0$ (the Lagrangian dual function corresponding to the constraint $|\nabla\phi_{\mu\to\mu^*}| \le \xi$) and for any $\alpha \ge 0$. This follows from the fact that $|\nabla\phi_{\mu\to\mu^*}| = \xi$, as $\phi_{\mu\to\mu^*}$ is the solution to (11) for the transport from μ to μ^*, satisfies (10) and occurs at the boundary of the constraint. Therefore, as $\Delta t \to 0$, we have the candidate velocity field:

$$\mathbf{v} = -\alpha \nabla\phi_{\mu\to\mu^*}, \tag{19}$$

where α can be any non-negative function on Ω. The implementation of the transport with the vector field (19) requires the computation of the Kantorovich potential $\phi_{\mu_t\to\mu^*}$ at any time t. Thus, we set up a primal-dual flow to obtain the Kantorovich potential as the solution to (11) (to which (17) is seen as the discrete counterpart as noted earlier).

6.2 Primal-Dual flow

The Lagrangian functional corresponding to the restricted Kantorovich duality (11) (to which (17) is seen as the discrete counterpart as noted earlier) for the optimal transport from μ to μ^* is given by:

$$L(\phi, \lambda) = \int_\Omega \phi(\rho - \rho^*) - \frac{1}{2}\int_\Omega \lambda(|\nabla\phi|^2 - |\xi|^2), \tag{20}$$

where all the integrals are with respect to the Lebesgue measure, and $\lambda \ge 0$ is the Lagrange multiplier function for the constraint $|\nabla\phi| \le \xi$ (which corresponds to the set $\mathcal{L}(\Omega) \cap \mathcal{C}^1(\Omega)$), as specified in (10), which we have rewritten here as $|\nabla\phi|^2 \le |\xi|^2$.

Lemma 14 (Optimality conditions)**.** *The necessary and sufficient conditions for a feasible solution $\bar{\phi}$ of* (11) *to be optimal are:*

$$\begin{aligned} -\nabla \cdot \left(\bar{\lambda}\nabla\bar{\phi}\right) &= \rho - \rho^*, \quad \textit{(in } \Omega) \\ \bar{\lambda}\nabla\bar{\phi} \cdot \mathbf{n} &= 0, \quad \textit{(on } \partial\Omega) \\ \bar{\lambda} \geq 0, \quad |\nabla\bar{\phi}| &\leq \xi, \quad \textit{(Feasibility)} \\ \bar{\lambda}(|\nabla\bar{\phi}| - \xi) &= 0 \ \ \textit{a.e.}, \quad \textit{(Complementary slackness)} \end{aligned} \tag{21}$$

where $\bar{\lambda}$ is the optimal Lagrange multiplier function.

Proof. We consider the Lagrangian (20), for which the first variation with respect to a variation $\delta\phi$, is given by:

$$\begin{aligned} \left\langle \frac{\delta L}{\delta \phi}, \delta\phi \right\rangle &= \int_\Omega (\rho - \rho^*)\delta\phi - \int_\Omega \lambda\nabla\phi \cdot \nabla\delta\phi \\ &= \int_\Omega (\rho - \rho^*)\delta\phi + \int_\Omega \nabla \cdot (\lambda\nabla\phi)\delta\phi - \int_{\partial\Omega} \lambda\nabla\phi \cdot \mathbf{n}\delta\phi, \end{aligned}$$

where we have used the divergence theorem to obtain the final equality. We have $\left\langle \frac{\delta L}{\delta \phi}, \delta\phi \right\rangle = 0$ for any variation $\delta\phi$ around the stationary point $(\bar{\phi}, \bar{\lambda})$. Therefore, we obtain $-\nabla \cdot \left(\bar{\lambda}\nabla\bar{\phi}\right) = \rho - \rho^*$ in Ω and $\bar{\lambda}\nabla\bar{\phi} \cdot \mathbf{n} = 0$ on $\partial\Omega$. Also, $\bar{\lambda} \geq 0$ is the feasibility condition for the Lagrange multiplier, $|\nabla\bar{\phi}| \leq \xi$ is the feasibility condition on ϕ and $\bar{\lambda}(|\nabla\bar{\phi}| - \xi) = 0$ is the complementary slackness condition. These correspond to the necessary KKT conditions, which for this problem (linear objective function and a convex constraint) are also the sufficient conditions for optimality. □

We now define a primal-dual flow to converge to the saddle point of the Lagrangian (20). For this, we henceforth consider the functions ϕ and λ to be additionally parametrized by time t. The primal-dual flow for the Lagrangian (20) is given by:

$$\begin{aligned} \partial_t \phi &= \nabla \cdot (\lambda\nabla\phi) + \rho - \rho^*, \\ \nabla\phi \cdot \mathbf{n} &= 0, \quad \text{on } \partial\Omega, \\ \partial_t \lambda &= \frac{1}{2}\left[|\nabla\phi|^2 - |\xi|^2\right]^+_\lambda, \\ \phi(0, x) &= \phi_0(x), \quad \lambda(0, x) = \lambda_0(x), \end{aligned} \tag{22}$$

where $[f]^+_\lambda = \begin{cases} f & \text{if } \lambda > 0 \\ \max\{0, f\} & \text{if } \lambda = 0 \end{cases}$ is a projection operator.

We note that $\partial_t\phi = \frac{\delta L}{\delta \phi}$ and $\partial_t\lambda = \left[-\frac{\delta L}{\delta \lambda}\right]^+_\lambda$, and we have a gradient ascent on $L(\phi, \lambda)$ w.r.t. ϕ and a projected gradient descent on $L(\phi, \lambda)$ w.r.t. λ.

Remark 2 (On the connection between (22) and (18))**.** *The primal-dual algorithm* (18) *is the discretization of the primal-dual flow* (22) *over a graph G (as defined in the previous subsection) with a step size τ. The term $-\sum_{j\in\mathcal{N}_i} \lambda_{ij}(l)\left(\phi^i(l) - \phi^j(l)\right)$ in* (18) *is the action of the weighted Laplacian matrix (with weights $\lambda_{ij}(l)$) on $\phi(l)$, which is the discretization over the graph of the term $\nabla \cdot (\lambda\nabla\phi)$ in* (22)*.*

Remark 3 (Existence and Uniqueness of solutions to (22))**.** *We first note that* (22) *generates a strongly continuous semigroup of operators and we interpret any solution of* (22) *as generated by this operator semigroup. We now consider the evolution of the Lagrange multiplier function λ. Letting $\lambda_0 \equiv 0$ and $h = \frac{1}{2}\left[|\nabla\phi|^2 - |\xi|^2\right]^+_\lambda$, we note that at any $x \in \Omega$, we have $\lambda(t,x) = \int_0^t h(\tau,x)d\tau$. Thus, a unique solution λ exists if $h(t,x) = \frac{1}{2}\left[|\nabla\phi|^2 - |\xi|^2\right]^+_\lambda$ is integrable in time at every $x \in \Omega$, which depends on the regularity of the solution ϕ. However, we do not apriori characterize or establish the desired level of regularity of the solutions ϕ, but instead assume that $\lambda \in L^\infty(0,T;L^\infty(\Omega))$ for any given $T > 0$. For any given $T > 0$, under the assumptions that $\lambda \in L^\infty(0,T;L^\infty(\Omega))$ and $\rho,\rho^* \in L^2(0,T;L^2(\Omega))$, there exists a unique weak solution $\phi \in L^2(0,T;H^1(\Omega))$ to the primal-dual flow* (22) *(we recall that we impose the Neumann boundary condition as $\nabla\phi\cdot\mathbf{n} = 0$ on $\partial\Omega$). The existence and uniqueness results follow by adapting the arguments presented in [21], Section 7.1 to the current problem (a homogenous second order parabolic PDE with a Neumann boundary condition). We note that the solution ϕ completely determines λ. To guarantee that $\lambda \in L^\infty(0,T;L^\infty(\Omega))$ is consistent with the solution ϕ, it may be necessary to add further regularity assumptions on ρ,ρ^*. However, further investigation into the regularity of solutions of the primal-dual flow is beyond the scope of this present work.*

Assumption 2 (Well-posedness of primal-dual flow)**.** *We assume that* (22) *is well-posed, with solution (ϕ,λ) such that $\phi \in L^\infty(0,\infty;H^1(\Omega))$ and the Lagrange multiplier function $\lambda \in L^\infty(0,\infty;L^\infty(\Omega))$ and is precompact in $L^2(\Omega)$.*

The following lemma establishes the convergence of solutions of (22) to the optimality conditions (21):

Lemma 15 (Convergence of primal-dual flow)**.** *The solutions (ϕ_t,λ_t) to the primal-dual flow* (22)*, under Assumption 2 on the well-posedness of the primal-dual flow, converge to an optimizer $(\tilde{\phi},\tilde{\lambda})$ given in* (21) *in the L^2 norm as $t \to \infty$, for any fixed $\rho,\rho^* \in L^2(\Omega)$.*

Proof. Let $(\tilde{\phi},\tilde{\lambda})$ be an optimizer of (21) and let $V(\phi,\lambda) = \frac{1}{2}\int_\Omega \left|\phi - \bar{\phi}\right|^2 \mathrm{dvol} + \frac{1}{2}\int_\Omega \left|\lambda - \bar{\lambda}\right|^2 \mathrm{dvol}$. Clearly, $V(\phi,\lambda) \geq 0$ for all $\phi,\lambda \in L^2(\Omega)$. The time-derivative of V along the solutions of the primal-dual flow (22) is given by:

$$\begin{aligned}\dot{V} &= \left\langle \frac{\delta L}{\delta\phi}, \phi - \bar{\phi}\right\rangle + \left\langle \left[-\frac{\delta L}{\delta\lambda}\right]^+_\lambda, \lambda - \bar{\lambda}\right\rangle \\ &= \left\langle \frac{\delta L}{\delta\phi}, \phi - \bar{\phi}\right\rangle - \left\langle \frac{\delta L}{\delta\lambda}, \lambda - \bar{\lambda}\right\rangle + \left\langle \frac{\delta L}{\delta\lambda} + \left[-\frac{\delta L}{\delta\lambda}\right]^+_\lambda, \lambda - \bar{\lambda}\right\rangle.\end{aligned}$$

Since L is concave in ϕ and convex in λ, we get:

$$\begin{aligned}\dot{V} &\leq L(\phi,\lambda) - L(\bar{\phi},\lambda) + L(\phi,\bar{\lambda}) - L(\phi,\lambda) + \left\langle \frac{\delta L}{\delta\lambda} + \left[-\frac{\delta L}{\delta\lambda}\right]^+_\lambda, \lambda - \bar{\lambda}\right\rangle \\ &= L(\bar{\phi},\bar{\lambda}) - L(\bar{\phi},\lambda) + L(\phi,\bar{\lambda}) - L(\bar{\phi},\bar{\lambda}) + \left\langle \frac{\delta L}{\delta\lambda} + \left[-\frac{\delta L}{\delta\lambda}\right]^+_\lambda, \lambda - \bar{\lambda}\right\rangle.\end{aligned}$$

We have that $L(\bar{\phi},\bar{\lambda}) - L(\bar{\phi},\lambda) \leq 0$ and $L(\phi,\bar{\lambda}) - L(\bar{\phi},\bar{\lambda}) \leq 0$ (recall that $(\bar{\phi},\bar{\lambda})$ is a saddle point of L). Moreover, by definition, when $\lambda(t,x) > 0$, we have $\left[-\frac{\delta L}{\delta\lambda}\right]^+_\lambda = -\frac{\delta L}{\delta\lambda}$ at (t,x), and when $\lambda(t,x)=0$ (which implies that $\lambda-\bar{\lambda}\le 0$), we have $\left[-\frac{\delta L}{\delta\lambda}\right]^+_\lambda \ge -\frac{\delta L}{\delta\lambda}$ at (t,x). This implies that $\left\langle \frac{\delta L}{\delta\lambda} + \left[-\frac{\delta L}{\delta\lambda}\right]^+_\lambda, \lambda-\bar{\lambda}\right\rangle \le 0$ at any (t,x). We therefore can say that $\dot{V}\le 0$. Moreover, by Assumption 2, it holds that the orbit ϕ is bounded in $H^1(\Omega)$ which, by Lemma 4, is compactly embedded in $L^2(\Omega)$. It then follows that the orbit is precompact in $L^2(\Omega)$. Moreover, by Assumption 2, we have that λ is precompact in $L^2(\Omega)$. We get that $\dot{V}=0$ only at an optimizer $(\tilde{\phi},\tilde{\lambda})$, which implies that the flow converges asymptotically to a $(\tilde{\phi},\tilde{\lambda})$. □

6.3 Convergence of PDE-based transport

Following the outline from earlier in the section, we now investigate the convergence properties of transport by the vector field $\mathbf{v}=-\alpha\nabla\phi_{\mu\to\mu^*}$. However, as discussed earlier, the scenario of particular interest to us is that of an on-the-fly implementation of the transport, where we do not wait for the convergence of the primal-dual flow to its steady state to obtain $\phi_{\mu\to\mu^*}$. This results in a coupling between the transport PDE and the primal-dual flow, and we investigate the convergence of solutions of this system of PDEs later in this section.

Lemma 16. *The transport* (2) *by the vector field* (19) *is a gradient flow, in the sense of Definition 1, on the optimal transport cost* $C(\cdot,\mu^*):\mathcal{P}(\Omega)\to\mathbb{R}_{\ge 0}$.

Proof. From Lemma 8, we have the strong duality $K(\mu,\mu^*)=C(\mu,\mu^*)$. The Kantorovich potential $\phi_{\mu\to\mu^*}$ is such that $\nabla\phi_{\mu\to\mu^*}=\nabla\left(\frac{\delta K}{\delta\mu}\right)$ (since $\phi_{\mu\to\mu^*}=\frac{\delta K}{\delta\mu}$, and we refer the reader to Chapter 7 in [40] for a proof). Therefore, the transport vector field $\mathbf{v}=-\alpha\nabla\phi_{\mu\to\mu^*}$ yields a gradient flow on the optimal transport cost $C(\mu,\mu^*)$. □

Remark 4 (Existence and uniqueness of solutions to the transport PDE)**.** *We refer the reader to [2] for a detailed treatment of existence and uniqueness results for the continuity equation, for transport vector fields with Sobolev regularity. We make the necessary well-posedness assumption for our purposes.*

Assumption 3 (Well-posedness of gradient flow on optimal transport cost)**.** *We assume that the desired distribution* μ^* *is absolutely continuous (with density function* ρ^* *in* $H^1(\Omega)$*) with* $\mathrm{supp}(\mu^*)=\Omega$. *Further, we assume that* (2) *is well-posed for the gradient flow on the optimal transport cost, with solution* $\rho\in L^\infty(0,\infty;H^1(\Omega))$.

Theorem 1. *Under Assumption 3 on the well-posedness of the gradient flow on the optimal transport cost and for absolutely continuous initial distributions* μ_0 *with* $\mathrm{supp}(\mu_0)=\Omega$*, the solutions* ρ *to the transport* (2) *by the vector field* $\mathbf{v}=-\frac{\lambda_{\mu\to\mu^*}}{\rho}\nabla\phi_{\mu\to\mu^*}$ *(where* $\phi_{\mu\to\mu^*}$ *and* $\lambda_{\mu\to\mu^*}$ *are the Kantorovich potential and the optimal Lagrange multiplier function for the transport* $\mu\to\mu^*$*) converge exponentially to* ρ^* *in the* L^2 *norm as* $t\to\infty$.

Proof. From the optimality conditions (21), we have that $\nabla\cdot(\lambda_{\mu\to\mu^*}\nabla\phi_{\mu\to\mu^*})=\rho^*-\rho$, which implies that $\partial_t\rho=-\nabla\cdot(\rho\mathbf{v})=\nabla\cdot(\lambda_{\mu\to\mu^*}\nabla\phi_{\mu\to\mu^*})=\rho^*-\rho$ when $\rho>0$. Moreover, we have that ρ_0 and ρ^* are strictly positive in Ω. Therefore, for any $t\in[0,\infty]$ and $x\in\mathring{\Omega}$, we have $\rho(t,x)>0$. Consequently, since $\rho(t,x)>0$, the transport vector field $\mathbf{v}=-\frac{\lambda_{\mu\to\mu^*}}{\rho}\nabla\phi_{\mu\to\mu^*}$ is well-defined on Ω. Let $V:L^2(\Omega)\to\mathbb{R}_{\ge 0}$ be defined by $V(\rho)=\frac{1}{2}\int_\Omega|\rho-\rho^*|^2\ \mathrm{dvol}$, where ρ is the density function of the absolutely continuous probability measure μ. The time derivative $\dot{V}$, under the transport (2) by $\mathbf{v} = -\frac{\lambda_{\mu\to\mu^*}}{\rho}\nabla\phi_{\mu\to\mu^*}$ is given by:

$$\begin{aligned}\dot{V} &= \int_\Omega (\rho - \rho^*)\partial_t \rho = -\int_\Omega (\rho-\rho^*)\nabla\cdot(\rho\mathbf{v})\\ &= \int_\Omega (\rho-\rho^*)\nabla\cdot(\lambda_{\mu\to\mu^*}\nabla\phi_{\mu\to\mu^*}).\end{aligned}$$

Further, from (21), we get:

$$\dot{V} = -\int_\Omega |\rho-\rho^*|^2 = -2V,$$

which implies that V is a Lyapunov functional for the transport by the vector field $\mathbf{v} = -\frac{\lambda_{\mu\to\mu^*}}{\rho}\nabla\phi_{\mu\to\mu^*}$. Moreover, by Assumption 3, we have that the solution ρ is bounded in $H^1(\Omega)$, which by the Rellich-Kondrachov theorem 4 is compactly contained in $L^2(\Omega)$. We then infer that the solution ρ to the transport (2) by the vector field $\mathbf{v} = -\frac{\lambda_{\mu\to\mu^*}}{\rho}\nabla\phi_{\mu\to\mu^*}$ is precompact, and therefore by the invariance principle in Lemma 5, converges to ρ^* in the L^2-norm in the limit $t\to\infty$, i.e. $\lim_{t\to\infty}\|\rho-\rho^*\|_{L^2} = 0$. Moreover, since we have $\dot{V} = -2V$, we note that the convergence is exponential. □

Remark 5 (Adaptation to tracking of time-varying target distributions)**.** *The exponential convergence result in the above theorem permits adaptation of the transport scheme to multi-agent tracking scenarios involving target distributions that evolve on a much slower timescale.*

We now present an on-the-fly implementation of the transport, where we do not wait for the primal-dual flow to reach steady state, but instead set the transport vector field as $\mathbf{v} = -\frac{\lambda}{\rho}\nabla\phi$, where ϕ and λ are supplied by (22). This results in a coupling between the transport PDE (2) and the primal-dual flow (22), and we investigate the behavior of the transport in simulation in Section 7.

We now establish the convergence of the on-the-fly transport under the primal flow and a fixed dual function $\lambda > 0$, which we define as follows:

$$\begin{aligned}\partial_t\phi &= \nabla\cdot(\lambda\nabla\phi) + \rho - \rho^*,\\ \nabla\phi\cdot\mathbf{n} &= 0, \quad \text{on } \partial\Omega,\\ \lambda &= \lambda(x) > 0.\end{aligned}\tag{23}$$

We note that transport under the relaxed primal-dual flow differs from the transport under (22) only in that the Lagrange multiplier function λ that weights the primal flow is fixed and does not vary in time.

Remark 6 (Existence and uniqueness of solutions to on-the-fly transport)**.** *We first note that* (23) *generates a strongly continuous semigroup of operators and we interpret any solution of* (23) *as generated by this operator semigroup. We recall from Remark 3 that a unique weak solution* $\phi \in L^2(0,T;H^1(\Omega))$ *to the primal flow exists if* $\rho,\rho^* \in L^2(0,T;H^1(\Omega))$ *and* $\lambda \in L^\infty(\Omega)$.

Assumption 4. *We assume that the desired distribution μ^* is absolutely continuous (with density function ρ^* in $H^1(\Omega)$) and supported on Ω. Further, we assume that the primal flow* (23) *and the transport* (2) *are well-posed, with solutions ϕ and ρ such that $\phi \in L^\infty(0,\infty; H^1(\Omega))$, and strictly positive $\rho \in L^\infty(0,\infty; H^1(\Omega))$.*

Theorem 2 (Convergence of on-the-fly transport)**.** *Under Assumption 4, the solutions ρ to* (2) *with the transport vector field $\mathbf{v} = -\frac{\lambda}{\rho}\nabla\phi$, with ϕ from* (23)*, converge in the L^2-norm to ρ^* as $t \to \infty$, while the solutions to the primal flow* (23) *converge to the optimality condition* (21) *corresponding to $\rho = \rho^*$.*

Proof. We first note that since $\rho > 0$ from Assumption 4, the transport vector field $\mathbf{v} = -\frac{\lambda}{\rho}\nabla\phi$ is well-defined on Ω. We now consider the following Lyapunov functional:

$$E = \frac{1}{2}\int_\Omega \lambda|\nabla\phi|^2 + \frac{1}{2}\int_\Omega |\rho - \rho^*|^2,$$

where all the integrals are with respect to dvol. The time derivative of E under the flow (23) and $\mathbf{v} = -\frac{\lambda}{\rho}\nabla\phi$ is given by:

$$\dot{E} = \int_\Omega \lambda\nabla\phi \cdot \nabla\partial_t\phi + \int_\Omega (\rho - \rho^*)\partial_t\rho.$$

Applying the divergence theorem and using the boundary condition for the first term, and the continuity equation (2) for the second, we obtain:

$$\begin{aligned}\dot{E} &= -\int_\Omega \nabla\cdot(\lambda\nabla\phi)\partial_t\phi - \int_\Omega (\rho-\rho^*)\nabla\cdot(\rho\mathbf{v}) \\ &= -\int_\Omega |\nabla\cdot(\lambda\nabla\phi)|^2 - \int_\Omega \nabla\cdot(\lambda\nabla\phi)(\rho-\rho^*) \\ &\quad + \int_\Omega (\rho-\rho^*)\nabla\cdot(\lambda\nabla\phi) \\ &= -\int_\Omega |\nabla\cdot(\lambda\nabla\phi)|^2.\end{aligned}$$

By Assumption 4 we have that the orbits ϕ and ρ are bounded in $H^1(\Omega)$ and by Lemma 4 (Rellich-Kondrachov theorem) we have that the orbits are precompact in $L^2(\Omega)$. Now, from the invariance principle in Lemma 5, we infer that the orbits of the system converge to the largest invariant set in $\dot{E}^{-1}(0)$. We have that $\dot{E} = 0$ implies $\|\nabla\cdot(\lambda\nabla\phi)\|_{L^2(\Omega)} = 0$, from which it follows by substitution in (23) that the transport (2) with $\mathbf{v} = -\frac{\lambda}{\rho}\nabla\phi$ yields $\rho \to_{L^2} \rho^*$, while ϕ converges to the optimality conditions corresponding to $\rho = \rho^*$. □

This leads us to the following algorithm for on-the-fly multi-agent transport under fixed, positive dual weighting:

7 Simulation studies and discussion

In this section, we present simulation results for multi-agent optimal transport in $\mathbb{R}^2$, based on the the iterative multi-stage transport scheme (15) (with c being the Euclidean metric

Algorithm 2 Multi-agent (on-the-fly) optimal transport with fixed (dual) weighting

Input: Target measure μ^*, Weights (dual variable) λ_{ij}, Bound on step size ϵ, Time step τ

For each agent i at time instant k of transport:

1: Obtain: Positions $x_j(k)$ of neighbors within communication/sensing radius r ($r \leq \mathrm{diam}(\Omega)$, large enough to cover Voronoi neighbors)

2: Compute: Voronoi cell $\mathcal{V}_i(k)$, Mass of cell $\mu^*(\mathcal{V}_i(k))$, Voronoi neighbors $\mathcal{N}_i(k)$

3: Initialize: $\phi^i \leftarrow \Phi^d_{k-1}(x_i(k))$ (with $\Phi^d_0 = 0$)

4: Implement n iterations of primal algorithm (23) (synchronously, in communication with neighbors $j \in \mathcal{N}_i$) to obtain $\phi^i(k)$

5: Communicate with neighbors $j \in \mathcal{N}_i$ to obtain $\phi^j(k)$, construct local estimate of Φ^d_k by multivariate interpolation

6: Implement transport step (15) with local estimate of Φ^d_k (which approximates $\phi_{\mu_k \to \mu^*}$)

and $\epsilon = 0.02$), where the local estimates of the Kantorovich potential are computed by the distributed online algorithm (18) with a step size $\tau = 1$. We also present simulation results for the PDE-based transport (2) under the primal-dual flow (22).

We considered a bivariate Gaussian distribution with covariance $\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ and mean randomly chosen in $[0,1]^2$ as the target probability measure, and $N = 30$ agents for the transport. Figure 1 shows the agents along with the corresponding Voronoi partition of the domain, at three different stages (time instants $k = 0, 5, 10$) during the course of their transport. We observe that the agents are transported towards the target probability measure and that a quantization of the target measure is obtained. This is clarified further in Figure 2, as described below.

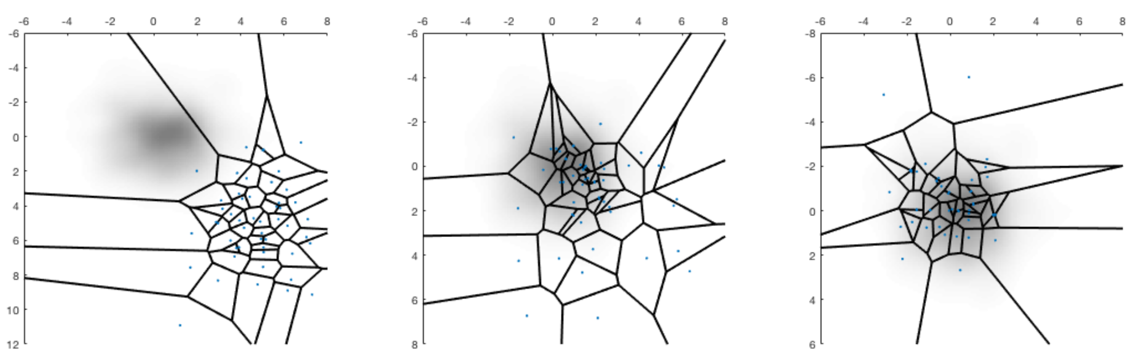

Figure 1: Positions of agents along with the Voronoi partition generated by them at three different stages (time instants $k = 0, 5, 10$) of transport by the iterative scheme (15) with local estimates of Kantorovich potential supplied by (18). Target probability measure shown in grayscale with a darker shade indicating a region of higher target density. The plots show convergence in time of the agents to full coverage of the target coverage profile (represented by the target probability distribution).

As we had noted in the previous section, there exists a fundamental trade-off between optimality and an on-the-fly implementation of the distributed optimal transport. We sought to investigate the extent of this trade-off in simulation by running multiple iterations n of the primal-dual algorithm (18) for every iteration of the transport (15). The underlying rationale is that the distributed computation is many times faster than the transport. Figure 2 shows the rate of convergence (w.r.t. the variance in target mass $\mu^*(\mathcal{V}_i)$ across the partition) for various values of n. Figure 3 is a plot of the net cost of transport w.r.t. n.

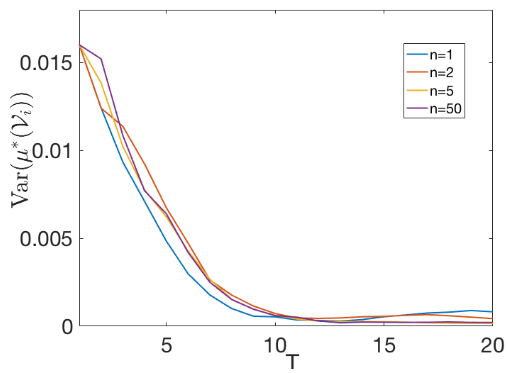

Figure 2: Variance in target mass $\mu^*(\mathcal{V}_i)$ across the partition vs time for various iteration steps n of the primal-dual algorithm (18) for every step of the transport (15).

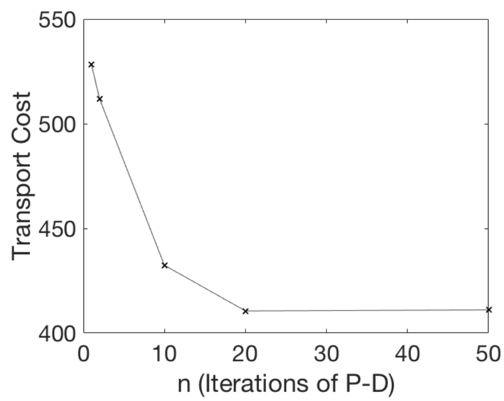

Figure 3: Net cost of transport for various iteration steps n of the primal-dual algorithm (18) for every step of the transport (15).

Figure 4 shows the evolution of the distribution of the agents over time. The grayscale images show the distribution of the agents in the domain, with darker shades representing higher density of agents at any given location. The domain is a 50×50 grid, and the PDE (22) was discretized over the grid. The initial distribution value was randomly generated (a random number was generated by the *rand* function in MATLAB for each cell of the grid and then normalized to obtain the probability distribution over the grid). The target density was defined by a grayscale image, as seen in the final subfigure in Figure 4. The cost of transport was chosen to be $c_{ij} = 1$ between neighboring cells i and j in the grid.

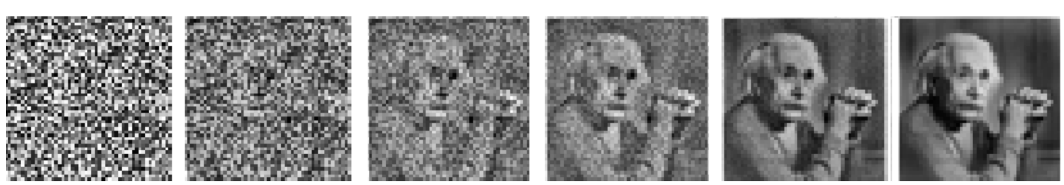

Figure 4: Distribution at various stages of the PDE-based transport (2) under the primal-dual flow (22). The figure shows convergence in time of the distribution to the target distribution represented by the final image.

We observe convergence of the on-the-fly transport under the primal-dual flow (22). Although we have established convergence of the transport analytically only under a primal flow with a fixed dual function, we conjecture that an on-the-fly transport under the primal-dual flow (22) also possesses the asymptotic stability property.

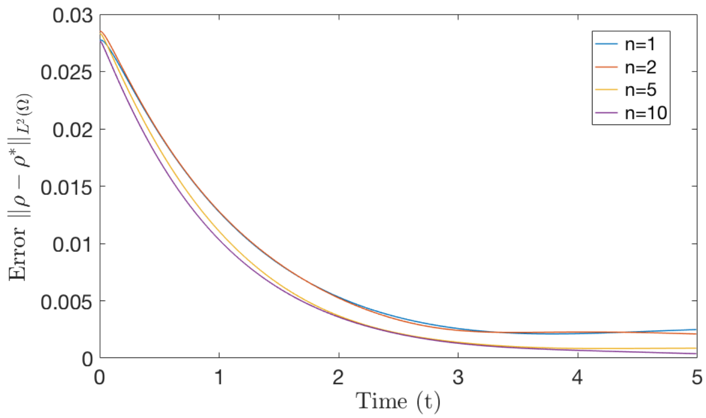

Figure 5: Density error $\|\rho-\rho^*\|_{L^2(\Omega)}$ vs time for various multiples n of the time scale of primal-dual flow (22) w.r.t the time scale of transport (2). The plot shows the rate of convergence to the optimal transport gradient flow (represented by $n = 10$).

Figure 5 is the plot of the L^2-density error $e(t) = \|\rho - \rho^*\|_{L^2(\Omega)}$ as a function of time, for various iteration steps of the primal-dual flow (to converge to the optimal gradient flow velocity) for every iteration of the transport PDE. We notice a significant improvement in the tracking performance (as measured by $e(t)$) within a few iterations of the primal-dual flow per iteration of the transport PDE, and the convergence to true optimal transport (in the sense of decay rate of the error $e(t)$) is obtained with approximately an order ($n \approx 10^1$) of magnitude time scale separation between computation and transport.

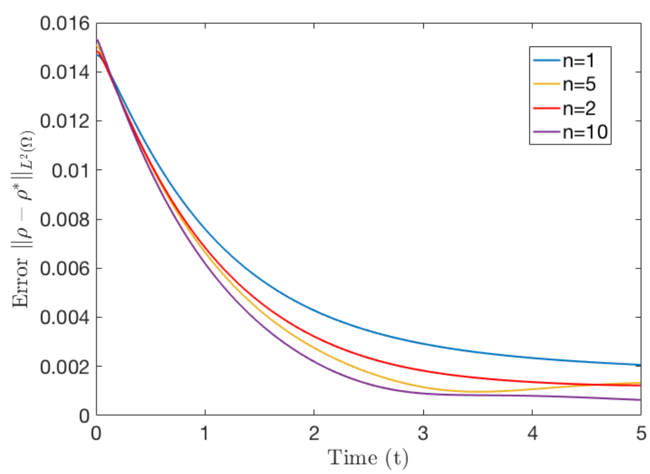

Figure 6: Density error $\|\rho - \rho^*\|_{L^2(\Omega)}$ vs time for various multiples n of the time scale of primal flow (23) with a fixed dual function, w.r.t the time scale of transport (2). The plot shows the rate of convergence to the optimal transport gradient flow (represented by $n = 10$).

8 Conclusion

In this work, we proposed a scalable, distributed iterative proximal point algorithm for large-scale optimal transport of multi-agent collectives. We obtained a dynamical formulation of optimal transport of agents, for metric transport costs that are conformal to the Euclidean distance. We proposed a distributed primal-dual algorithm to be implemented by the agents to obtain local estimates of the Kantorovich potential, which are then used as local objectives in a proximal point algorithm for transport. We studied the behavior of the transport in simulation and presented an analysis of the candidate PDE model for the continuous time and $N \to \infty$ limit, establishing asymptotic stability of the transport. We explored in simulation the suboptimality of the on-the-fly implementation. The analytic characterization of the extent of the trade-off between optimality and on-the-fly implementation is left for future work, as is the rigorous characterization of the convergence of solutions of the multi-agent optimal transport to the solutions of the PDE-based transport as $N \to \infty$.

References

[1] L. Ambrosio. Lecture notes on optimal transport problems. In *Mathematical aspects of evolving interfaces*, pages 1–52. Springer, 2003.

[2] L. Ambrosio. Transport equation and Cauchy problem for non-smooth vector fields. In *Calculus of variations and nonlinear partial differential equations*, pages 1–41. Springer, 2008.

[3] C. Andrieu, A. Doucet, and R. Holenstein. Particle Markov Chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3):269–342, 2010.

[4] S. Bandyopadhyay, S. J. Chung, and F. Hadaegh. Probabilistic and distributed control of a large-scale swarm of autonomous agents. *IEEE Transactions on Robotics*, 33(5):1103–1123, 2017.

[5] S. Bandyopadhyay, S. J. Chung, and F. Y. Hadaegh. Inhomogeneous Markov chain approach to probabilistic swarm guidance algorithms. In *Int. Conf. on Spacecraft Formation Flying Missions and Technologies*, 2013.

[6] S. Bandyopadhyay, S. J. Chung, and F. Y. Hadaegh. Probabilistic swarm guidance using optimal transport. In *IEEE Conf. on Control Applications*, pages 498–505, 2014.

[7] J. D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

[8] J. D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2):A1111–A1138, 2015.

[9] P. Billingsley. *Probability and measure.* John Wiley, 2008.

[10] D. Bourne, B. Schmitzer, and B. Wirth. Semi-discrete unbalanced optimal transport and quantization. *arXiv preprint arXiv:1808.01962*, 2018.

[11] F. Bullo, J. Cortés, and S. Martínez. *Distributed Control of Robotic Networks*. Applied Mathematics Series. Princeton University Press, 2009.

[12] Y. Chen, T. Georgiou, and M. Pavon. On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint. *Journal of Optimization Theory & Applications*, 169(2):671–691, 2016.

[13] A. J. Chorin and J. E. Marsden. *A mathematical introduction to fluid mechanics*, volume 3. Springer, 1990.

[14] J. Cortés, S. Martínez, T. Karatas, and F. Bullo. Coverage control for mobile sensing networks. *IEEE Transactions on Robotics and Automation*, 20(2):243–255, 2004.

[15] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, pages 2292–2300, 2013.

[16] M. Cuturi and A. Doucet. Fast computation of Wasserstein barycenters. In *International Conference on Machine Learning*, pages 685–693, 2014.

[17] M. Cuturi and G. Peyré. Semidual regularized optimal transport. *SIAM Review*, 60(4):941–965, 2018.

[18] N. Demir, U. Eren, and B. Açıkmeşe. Decentralized probabilistic density control of autonomous swarms with safety constraints. *Autonomous Robots*, 39(4):537–554, 2015.

[19] K. Elamvazhuthi, H. Kuiper, and S. Berman. PDE-based optimization for stochastic mapping and coverage strategies using robotic ensembles. *Automatica*, 95:356–367, 2018.

[20] U. Eren and B. Açıkmeşe. Velocity field generation for density control of swarms using heat equation and smoothing kernels. In *IFAC Papers Online*, volume 50, pages 9405–9411, 2017.

[21] L.C. Evans. *Partial differential equations*. Graduate studies in mathematics. American Mathematical Society, Providence (R.I.), 1998.

[22] S. Ferrari, G. Foderaro, P. Zhu, and T. A. Wettergren. Distributed optimal control of multiscale dynamical systems: a tutorial. *IEEE Control Systems*, 36(2):102–116, 2016.

[23] P. Frihauf and M. Krstic. Leader-enabled deployment onto planar curves: A PDE-based approach. *IEEE Transactions on Automatic Control*, 56, 2011.

[24] A. Genevay, M. Cuturi, G. Peyré, and F. Bach. Stochastic optimization for large-scale optimal transport. In *Advances in Neural Information Processing Systems*, pages 3440–3448, 2016.

[25] W. Gilks, S. Richardson, and D. Spiegelhalter. *Markov chain Monte Carlo in practice*. Chapman and Hall/CRC, 1995.

[26] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[27] D. Henry. *Geometric theory of semilinear parabolic equations.* Springer, 1981.

[28] V. Krishnan and S. Martínez. Distributed control for spatial self-organization of multiagent swarms. *SIAM Journal on Control and Optimization*, 56(5):3642–3667, 2018.

[29] V. Krishnan and S. Martínez. Distributed optimal transport for the deployment of swarms. In *IEEE Int. Conf. on Decision and Control*, pages 4583–4588, Miami Beach, USA, 2018.

[30] M. Kuang and E. Tabak. Sample-based optimal transport and barycenter problems. *Communications on Pure and Applied Mathematics*, 2017. Submitted.

[31] Giovanni Leoni. *A first course in Sobolev spaces*, volume 105. American Mathematical Society, 2009.

[32] R. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Mathematical Journal*, 80(2):309–324, 1995.

[33] Q. Mérigot. A multiscale approach to optimal transport. In *Computer Graphics Forum*, volume 30, pages 1583–1592. Wiley Online Library, 2011.

[34] M. Mesbahi and M. Egerstedt. *Graph Theoretic Methods in Multiagent Networks.* Princeton Series in Applied Mathematics. Princeton University Press, 2010.

[35] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.

[36] T. Mikami and M. Thieullen. Optimal transportation problem by stochastic optimal control. *SIAM Journal on Control and Optimization*, 47(3):1127–1139, 2008.

[37] N. Papadakis, G. Peyré, and E. Oudet. Optimal transport with proximal splitting. *SIAM Journal on Imaging Sciences*, 7(1):212–238, 2014.

[38] G. Peyré and M. Cuturi. Computational optimal transport. Technical report, 2017.

[39] C. Robert and G. Casella. *Monte Carlo statistical methods.* Springer, 2013.

[40] F. Santambrogio. *Optimal transport for applied mathematicians.* Springer, 2015.

[41] F. Santambrogio. {Euclidean, metric, and Wasserstein} gradient flows: an overview. *Bulletin of Mathematical Sciences*, 7(1):87–154, 2017.

[42] V. Seguy, B. Damodaran, R. Flamary, N. Courty, A. Rolet, and M. Blondel. Large-scale optimal transport and mapping estimation. *arXiv preprint arXiv:1711.02283*, 2017.

[43] E. Tabak and G. Trigila. Data-driven optimal transport. *Communications on Pure and Applied Mathematics*, 10:1002, 2014.

[44] C. Villani. *Optimal transport: old and new*, volume 338. Springer, 2008.

[45] J. A. Walker. Some results on Liapunov functions and generated dynamical systems. *Journal of Differential Equations*, 30(3):424–440, 1978.

[46] J.A. Walker. *Dynamical Systems and Evolution Equations: Theory and Applications*, volume 20. Springer, 2013.

[47] F. Zhang, A. Bertozzi, K. Elamvazhuthi, and S. Berman. Performance bounds on spatial coverage tasks by stochastic robotic swarms. *IEEE Transactions on Automatic Control*, 63(6):1563–1578, 2018.