A Multiscale Analysis of Multi-Agent Coverage Control Algorithms

Vishaal Krishnan and Sonia Martínez

Abstract—**This paper presents a theoretical framework for the design and analysis of gradient descent-based algorithms for coverage control tasks involving robot swarms. We adopt a multiscale approach to analysis and design to ensure consistency of the algorithms in the large-scale limit. First, we represent the macroscopic configuration of the swarm as a probability measure and formulate the macroscopic coverage task as the minimization of a convex objective function over probability measures. We then construct a macroscopic dynamics for swarm coverage, which takes the form of a proximal descent scheme in the L^2-Wasserstein space. Our analysis exploits the generalized geodesic convexity of the coverage objective function, proving convergence in the L^2-Wasserstein sense to the target probability measure. We then obtain a consistent gradient descent algorithm in the Euclidean space that is implementable by a finite collection of agents, via a "variational" discretization of the macroscopic coverage objective function. We establish the convergence properties of the gradient descent and its behavior in the continuous-time and large-scale limits. Furthermore, we establish a connection with well-known Lloyd-based algorithms, seen as a particular class of algorithms within our framework, and demonstrate our results via numerical experiments.**

Index Terms—**Multi-agent systems, coverage control, multi-scale analysis, proximal descent, Lloyd's algorithm.**

I. Introduction

Multi-agent systems are groups of autonomous agents with sensing, communication, and computational capabilities. It is often necessary to achieve a desired coverage of a spatial region before these systems can be deployed for specific purposes. This has spurred intense research activity on the design of multi-agent coverage control algorithms [1]–[4]. In spatial coverage control problems involving large-scale multi-agent systems, it is often more appropriate and convenient to specify the task objective at the macroscopic scale for the distribution of agents over the spatial region. However, actuation still rests at the microscopic scale at the level of the individual agents, and faces a multitude of constraints imposed by the multi-agent setting. These include information constraints from limitations on sensing, communication and localization, and physical constraints such as collision and obstacle avoidance. This separation of scales poses a problem for the analysis and design of algorithms with performance guarantees. While *mechanistic* models relying on theoretical tools from infinite-dimensional analysis are often more appropriate for macro scales, an *algorithmic* approach that relies on tools from finite dimensional analysis is more effective in addressing the above microscopic constraints. This underscores the need for a formal theory bridging the two scales. Such a bridge theory is crucial for integrating the mechanistic and algorithmic paradigms and in understanding how macroscopic coverage objectives translate to the microscopic level of individual agents and conversely, how the microscopic algorithms shape macroscopic behavior.

This material is based upon work supported by grant AFOSR FA-9550-18-1-0158. V. Krishnan is with the Mechanical Engineering Department, University of California, Riverside, Riverside CA 92521 USA. S. Martínez is with the Department of Mechanical and Aerospace Engineering, University of California, San Diego, La Jolla CA 92093 USA. Email: `vishaalk@ucr.edu`, `soniamd@ucsd.edu`.

Related work. Multi-agent coverage control algorithms have been widely studied over the past two decades and have a rich literature. For an (inexhaustive) overview of the literature, we adopt the classification into *mechanistic* vs *algorithmic* models, as introduced earlier. The algorithmic perspective is predominantly based on tools from distributed optimization. Initial works combined distributed optimization with ideas from computational geometry and dynamic systems [1], [5]–[7]. These were then extended to include sensing, energy, and, obstacle, and dynamic constraints encountered in the multi-agent setting [3], [8], [9]. Interest in the mechanistic perspective was fueled by efforts to scale up the size of these systems, which emphasized the need for tools of macroscopic analysis. This led to the application of mathematical tools from probability, stochastic processes and partial differential equations. For large-scale multi-agent systems, one such approach involves the design of coverage by synthesis of Markov transition matrices [10]–[13]. Another approach involves the use of continuum/PDE-based models, applying ideas of diffusion and heat flow to coverage control [14]–[16]. Tools from parameter tuning and boundary control of PDEs [17]–[19] have been used in this context. Statistical physics-based approaches, including the application of mean-field theory, have also been recently explored [20], [21]. Some works at the intersection of the microscopic and macroscopic perspectives include [19], where the authors obtain performance bounds for spatial coverage by multi-agent swarms, characterizing coverage performance as a function of the number of robots and robot sensing radius.

More recently, tools from optimal transport theory have been applied to multi-agent coverage. Interest in optimal transport and optimal control is motivated by energy considerations, and constitutes another active area of research [22]–[26]. Furthermore, coverage algorithms often work with a quantization of the underlying spatial domain. Recently [27]–[30] explores the underlying connections of quantization to optimal transport. Some well-known transport PDEs can be formulated as gradient flows on functionals in the space of probability measures [31]. Furthermore, from a computational perspective, gradient flows in the space of probability measures are often discretized into particle gradient flows. The gradient flow structure underlying these PDEs allows for their discretization by formulating proximal gradient descent schemes in the space of probability measures. For instance, in [32] the authors discretize the well-known Fokker-Planck equation by a proximal recursion. In [33], the authors investigate the convergence of such particle gradient flows to global minima in the limit $N \to \infty$. In [34], the authors apply proximal descent schemes to study uncertainty propagation in stochastic systems.

Contributions. This paper contributes a multi-scale analysis of gradient descent-based coverage algorithms for multi-agent systems, with three main goals in mind: (i) the formalization of coverage objectives for large-scale multi-agent systems via meaningful macroscopic metrics, (ii) the systematic design of provable correct algorithms that are consistent across the macroscopic and microscopic scales, and (iii) to gain a fundamental understanding of widely studied coverage algorithms for large-scale multi-agent systems and shed new light on their behavior as the number of agents $N \to \infty$. A suitable theoretical framework for the above is largely missing in the literature and this work addresses the gap.

We formulate the coverage task as a minimization in the space of probability measures and define a proximal gradient descent on the aggregate objective function. The multi-agent configuration is specified by discretizing the underlying probability measure and we obtain implementable coverage algorithms as a proximal gradient descent on the discretized aggregate objective function w.r.t. agent positions. This leads to a new class of "variational" gradient algorithms, and we show that this class of algorithms subsumes previously defined coverage algorithms based on distortion metrics. This allows us to establish a connection between the macroscopic and microscopic perspectives and present a unified theory of multi-agent coverage algorithms.

Paper outline. The rest of the paper is organized as follows. Section II contains a description of the coverage optimization problem setting. In Section III, we present an iterative descent scheme in the space of probability measures and establish convergence results for such a scheme. Building on these results, we propose multi-agent coverage algorithms in Section IV as the discretization of the iterative descent scheme from Section III, establish convergence results and study their behavior in the continuous-time and $N \to \infty$ limits. Section V contains a case study of the well-known Lloyd's algorithm within the theoretical framework developed in the prior sections and results from numerical experiments. An overview of the mathematical preliminaries is presented in Appendix A.

II. Coverage optimization problem

In this section, we formulate the multi-agent coverage problem as an optimization of a macroscopic coverage objective, which forms the focus of our analysis and algorithm design in the subsequent sections. We begin by specifying the problem setting. Let $\Omega \subset \mathbb{R}^d$ be compact and convex (see additional notation here[1]), and $\mathbf{x} = (x_1, \ldots, x_N)$ (with $x_i \in \Omega$ for $i \in \mathcal{I} = \{1, \ldots, N\}$ being the agent positions) denote the microscopic state of the multi-agent system. In specifying the macroscopic configuration, we look for a representation that satisfies two key properties, (i) *Permutation-invariance:* Assuming that the agents are identical, we note that every microscopic configuration $\mathbf{x} \in \Omega^N$ is equivalent to $(P \otimes I_d)\,\mathbf{x}$ for any permutation $P \in \mathbb{R}^{N \times N}$. The representation must be invariant under such permutations, and (ii) *Consistency in the $N \to \infty$ limit*: The space of representations must contain the "representation limit" as $N \to \infty$, to enable the study of large-scale properties of coverage algorithms. This leads us to specifying the macroscopic configuration of the multi-agent system by probability measures over the underlying space Ω. For the microscopic configuration $\mathbf{x} = (x_1, \ldots, x_N)$, we specify the corresponding macroscopic configuration by the probability measure $\widehat{\mu}_{\mathbf{x}}^N = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$. We note that $\widehat{\mu}_{\mathbf{x}}^N$ is invariant under permutations of agent positions. Furthermore, if the positions x_i are independently and identically distributed according to an (absolutely continuous) probability measure $\mu \in \mathcal{P}(\Omega)$, it follows from the Glivenko-Cantelli theorem [35] that as $N \to \infty$, the discrete probability measure $\widehat{\mu}_{\mathbf{x}}^N$ converges uniformly, and almost surely, to μ. In this way, probability measures over Ω are a suitable space of macroscopic representations that combine the desired properties of permutation-invariance and consistency in the $N \to \infty$ limit.

With the microscopic and macroscopic representations of the multi-agent system in place, we now move to the specification of the coverage task as the minimization of a macroscopic coverage objective function $F : \mathcal{P}(\Omega) \to \mathbb{R}$. We let F be l-smooth and strictly (generalized) geodesically convex[2], with a unique minimizer $\mu^* \in \mathcal{P}(\Omega)$. The coverage problem can then be described as follows: Given an initial macroscopic configuration $\mu_0 \in \mathcal{P}(\Omega)$ of the multi-agent system (with μ_0 being an absolutely continuous probability measure), specify a descent scheme in $\mathcal{P}(\Omega)$ that minimizes the coverage objective function F, generating a sequence $\{\mu_k\}_{k \in \mathbb{N}}$ that converges weakly to μ^* as $k \to \infty$. In Section III, we propose a proximal descent scheme that exploits the (generalized) convexity of F to solve the coverage task. Furthermore, in Section IV we obtain an implementable multi-agent coverage algorithm that

[1] We 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. The gradient operator in $\mathbb{R}^d$ is denoted as $\nabla = (\partial/\partial x_1, \ldots \partial/\partial x_n)$, where, as a shorthand, we use $\partial/\partial z \equiv \partial_z$ to denote the partial derivative w.r.t. a variable z and $\frac{\partial}{\partial x_i} \equiv \partial_i$. Consider a set $\Omega \subseteq \mathbb{R}^d$. In what follows, $\partial\Omega \subseteq \mathbb{R}^d$ denotes its boundary, $\bar{\Omega} = \Omega \cup \partial\Omega$ its closure, and $\mathring{\Omega} = \Omega \setminus \partial\Omega$ its interior with respect to the standard Euclidean topology. For $M \subseteq \Omega$, we define the distance $d(x, M)$ of a point $x \in \Omega$ to M as $d(x, M) = \inf_{y \in M} \|x - y\|$. Given any $x \in \Omega \subset \mathbb{R}^d$, we denote by $B_r(x)$ the closed d-ball of radius $r > 0$, centered at x. The indicator function on Ω for the subset M will be denoted as $\mathbf{1}_M : \Omega \to \{0, 1\}$. We use $\langle f, g\rangle$ to represent the inner product of functions $f, g : \Omega \to \mathbb{R}$ w.r.t. the Lebesgue measure, given by $\langle f, g\rangle = \int_\Omega fg\,\mathrm{dvol}$. We denote by $\mathrm{Lip}(\Omega)$ the space of Lipschitz continuous functions on Ω. A function $p : \Omega \to \mathbb{R}$ is called l-smooth (or Lipschitz differentiable) if for any $x, y \in \Omega$, we have $|\nabla p(y) - \nabla p(x)| \leq l\|y - x\|$. It can be shown that for an l-smooth function $p : \Omega \to \mathbb{R}$ and any $x, y \in \Omega$, we have $|p(y) - p(x) - \langle \nabla p(x), y - x\rangle| \leq \frac{l}{2}\|y - x\|^2$. We denote by $\mathcal{P}(\Omega)$ the space of probability measures over Ω. For a measurable mapping $\mathcal{T} : \Omega \to \Theta$, where Ω and Θ are measurable, we denote by $\mathcal{T}_{\#}\mu \in \mathcal{P}(\Theta)$ the pushforward measure of $\mu \in \mathcal{P}(\Omega)$ and we have $\mathcal{T}_{\#}\mu(B) = \mu(\mathcal{T}^{-1}(B))$. for all measurable $B \subseteq \Theta$.

[2] in the sense of Definition 7 in Appendix A.

updates agent positions in Ω and performs consistently (in the $N \to \infty$ limit) with the macroscopic descent scheme. That is, we design a provably-correct, discrete-time, agent-based algorithm that generates microscopic sequences $\{\mathbf{x}_k\}_{k\in\mathbb{N}} \subseteq \Omega^N$ such that $\lim_{k,N\to\infty} \widehat{\mu}^N_{\mathbf{x}_k} = \mu^\star$. We address this question in Section IV by tying the macroscopic descent scheme with the microscopic coverage algorithm by means of a variational approach.

Example coverage objective functions. We introduce a class of coverage objective functions, whose convexity properties will be analyzed in Section V. Furthermore, in Section V we also establish a relationship between the macroscopic descent scheme corresponding to these objective functions and the well-known Lloyd's algorithm [1]. Let $f : \mathbb{R} \to \mathbb{R}$ be a strictly convex, non-decreasing and l-smooth function with $f(0) = 0$, and let:

$$C_f(\mu, \nu) = \inf_{\substack{T:\Omega\to\Omega \\ T_\#\mu=\nu}} \int_\Omega f(|x - T(x)|)\, d\mu(x), \tag{1}$$

be defined for two probability measures μ and ν. In the quadratic case $f(x) = x^2$, we get $C_f \equiv W_2^2$, the so-called L^2-Wasserstein distance, which is a metric over $\mathcal{P}(\Omega)$. Conversely, this suggests the design of a coverage objective function given a target macroscopic configuration $\mu^\star$, as $F(\mu) = W_2^2(\mu, \mu^\star)$, which quantifies how far μ is from the target $\mu^\star$.

III. Macroscopic and particle descent schemes

In this section, we present a (macroscopic) iterative descent scheme in the space of probability measures $\mathcal{P}(\Omega)$ and establish weak convergence to the minimizer under certain conditions. Furthermore, we derive an equivalent (microscopic) characterization of the descent scheme in Ω. We refer to Appendix A for additional definitions and supporting results.

We consider the following proximal recursion in $\mathcal{P}(\Omega)$ starting from any absolutely continuous $\mu_0 \in \mathcal{P}(\Omega)$:

$$\mu_{k+1} \in \arg\min_{\nu\in\mathcal{P}(\Omega)} \frac{1}{2\tau} W_2^2(\mu_k, \nu) + F(\nu). \tag{2}$$

We assume that F satisfies the Neumann boundary condition $\nabla\left(\frac{\delta F}{\delta \nu}\right) \cdot \mathbf{n} \geq 0$ on $\partial\Omega$ (where $\mathbf{n}$ is the outward normal to $\partial\Omega$) for any $\nu \in \mathcal{P}(\Omega)$. This ensures conservation of mass and that the solutions of the gradient descent w.r.t. F, which are sequences of measures, are contained in $\mathcal{P}(\Omega)$.

Lemma 1 (***Compactness and convexity of sublevel sets***)**.** *Let F be an l-smooth, geodesically convex functional (in the sense of Definition 7 in Appendix A) over Ω. The F-sublevel set of any absolutely continuous probability measure $\mu \in \mathcal{P}(\Omega)$ is compact and geodesically convex in the L^2-Wasserstein space $(\mathcal{P}(\Omega), W_2)$.*

Proof. For any $\mu \in \mathcal{P}(\Omega)$, the sublevel set $\mathcal{S}(\mu) = \{\nu \in \mathcal{P}(\Omega) | F(\nu) \leq F(\mu)\}$ is closed in $(\mathcal{P}(\Omega), W_2)$, since F is continuous and $\mathcal{P}(\Omega)$ is closed and compact (see Corollary 3 in Appendix A on the compactness of $\mathcal{P}(\Omega)$). This implies that $\mathcal{S}(\mu)$ is also compact since it is a closed subset of a compact set.

Recall from Lemma 16 that $(\mathcal{P}(\Omega), W_2)$ is geodesically convex, and consider, for any $\nu_0, \nu_1 \in \mathcal{S}(\mu)$, and $\nu_t \in \mathcal{P}(\Omega)$, for $t \in [0, 1]$, the generalized geodesic between ν_0 to ν_1 with μ as the reference measure (from Lemma 13 it follows that unique optimal transport maps from μ to ν_0 and μ to ν_1 exist, since μ is absolutely continuous, and therefore so does a unique generalized geodesic in $(\mathcal{P}(\Omega), W_2)$ between ν_0 and ν_1 as in Definition 6). From the (generalized) geodesic convexity of F we have that $F(\nu_t) \leq (1 - t)F(\nu_0) + tF(\nu_1) \leq F(\mu)$ (since $F(\nu_0) \leq F(\mu)$ and $F(\nu_1) \leq F(\mu)$ by definition of $\mathcal{S}(\mu)$). This implies that $\nu_t \in \mathcal{S}(\mu)$ for any $t \in [0, 1]$, from which we infer the geodesic convexity of $\mathcal{S}(\mu)$. □

Lemma 2 (***Strong convexity of objective functional***)**.** *Let F be an l-smooth, geodesically convex functional over $\mathcal{P}(\Omega)$. For any absolutely continuous probability measure $\mu \in \mathcal{P}(\Omega)$, the functional $G(\nu) = \frac{1}{2\tau} W_2^2(\mu, \nu) + F(\nu)$ is $\left(\frac{1}{\tau} - l\right)$-strongly geodesically convex (in the sense of Definition 8 in Appendix A) over $\mathcal{P}^r(\Omega)$ for $0 < \tau < 1/l$.*

Proof. Since F is l-smooth, applying Lemma 15 for two atomless measures ν_1 and ν_2, we get:

$$\left| \int_\Omega \langle \xi_2 - \xi_1, T_{\mu\to\nu_2} - T_{\mu\to\nu_1} \rangle\, d\mu \right| \leq l W_2^2(\nu_1, \nu_2), \tag{3}$$

where ξ_1 and ξ_2 are the Fréchet derivatives of F evaluated at ν_1 and ν_2, respectively, and $T_{\mu\to\nu_1}$ and $T_{\mu\to\nu_2}$ are the optimal transport maps from μ to ν_1 and ν_2, respectively. Let $\eta_i = \nabla\left(\frac{\delta G}{\delta \nu}\right)\big|_{\nu_i}$, for $i = 1, 2$, and let $\phi_i = \frac{1}{2} \left.\frac{\delta W_2^2(\mu,\nu)}{\delta \nu}\right|_{\nu_i}$ be the so-called Kantorovich potential for the transport from ν_1 to μ, for $i = 1, 2$. We now have:

$$\begin{aligned}
&\int_\Omega \langle \eta_2 - \eta_1, T_{\mu\to\nu_2} - T_{\mu\to\nu_1} \rangle\, d\mu \\
&= \int_\Omega \left\langle \frac{1}{\tau}\nabla\phi_2 - \frac{1}{\tau}\nabla\phi_1 - \xi_1 + \xi_2, T_{\mu\to\nu_2} - T_{\mu\to\nu_1} \right\rangle d\mu \\
&= \frac{1}{\tau} \int_\Omega \langle \nabla\phi_2 - \nabla\phi_1, T_{\mu\to\nu_2} - T_{\mu\to\nu_1} \rangle\, d\mu \\
&\quad + \int_\Omega \langle \xi_2 - \xi_1, T_{\mu\to\nu_2} - T_{\mu\to\nu_1} \rangle\, d\mu \\
&\geq \frac{1}{\tau} \int_\Omega |T_{\mu\to\nu_2} - T_{\mu\to\nu_1}|^2\, d\mu - l W_2^2(\nu_1, \nu_2) \\
&\geq \left(\frac{1}{\tau} - l\right) W_2^2(\nu_1, \nu_2),
\end{aligned}$$

where the penultimate inequality above follows from (3). We have also used the fact that $\int_\Omega \langle \nabla\phi_2 - \nabla\phi_1, T_{\mu\to\nu_2} - T_{\mu\to\nu_1} \rangle\, d\mu = \int_\Omega |T_{\mu\to\nu_2} - T_{\mu\to\nu_1}|^2\, d\mu \geq W_2^2(\nu_1, \nu_2)$ (this follows from an application of Lemma 13 in Appendix A), which implies that $\int_\Omega \langle \nabla\phi_2 - \nabla\phi_1, T_{\nu_1\to\nu_2} - \mathrm{id} \rangle\, d\nu_1 = W_2^2(\nu_1, \nu_2)$. Since $\tau < \frac{1}{l}$, we get that the functional G is strongly convex with parameter $\frac{1}{\tau} - l$. □

Assumption 1 (***Atomless proximal descent sequence***)**.** *We assume that the sequence $\{\mu_k\}_{k\in\mathbb{N}}$ generated by* (2) *is such that $\mu_k \in \mathcal{P}^r(\Omega)$ for all $k \in \mathbb{N}$.*

We remark here that sufficient regularity of the functional F and the atomlessness of μ_0 should guarantee validity of Assumption 1. Since we do not offer a characterization of the regularity of F to this end, we retain Assumption 1 in establishing the following theorem:

Theorem 1 (***Convergence of proximal recursion*** (2))**.** *Let $\Omega \subseteq \mathbb{R}^d$ be a compact, convex set, and let $F : \mathcal{P}(\Omega) \to \mathbb{R}$ be an l-smooth, strictly geodesically convex functional satisfying the Neumann boundary condition $\nabla\left(\frac{\delta F}{\delta \mu}\right) \cdot \mathbf{n} \geq 0$ on $\partial\Omega$. Let μ_0 be an absolutely continuous measure. Under Assumption 1 on the generation of a proximal descent atomless sequence, and for $0 < \tau < 1/l$, the sequence $\{\mu_k\}_{k\in\mathbb{N}}$, generated by the proximal recursion* (2)*, converges weakly to $\mu^\star = \arg\min_{\nu\in\mathcal{P}(\Omega)} F(\nu)$ as $k \to \infty$.*

Proof. It follows that:

$$\begin{aligned}&\frac{1}{2\tau}W_2^2(\mu_k, \mu_{k+1}) + F(\mu_{k+1}) \leq F(\mu_k)\\ &\iff F(\mu_{k+1}) \leq F(\mu_k) - \frac{1}{2\tau}W_2^2(\mu_k, \mu_{k+1}).\end{aligned}$$

This implies that for $\mu_k \neq \mu_{k+1}$, we have $F(\mu_{k+1}) < F(\mu_k)$ and the sequence $\{F(\mu_k)\}_{k\in\mathbb{N}}$ is monotonically strictly decreasing. In addition, $\{\mu_k\}_{k\in\mathbb{N}}$ is contained in the sublevel set $\mathcal{S}(\mu_0)$ of $F(\mu_0)$.

From Lemma 1, $\mathcal{S}(\mu_0)$ is convex and compact in the L^2-Wasserstein space $(\mathcal{P}(\Omega), W_2)$. Thus, there is a weakly convergent subsequence $\{\mu_{k_\ell}\} \to_\ell \overline{\mu} \in \mathcal{S}(\mu_0)$. Consider the functional G_μ from (2), for $\mu \in \mathcal{P}(\Omega)$, such that $G_\mu(\nu) = \frac{1}{2\tau}W_2^2(\mu, \nu) + F(\nu)$. First, note that

$$\begin{aligned}&|G_{\mu_{k_\ell}}(\nu) - G_{\overline{\mu}}(\nu)| = \frac{1}{2\tau}|W_2^2(\overline{\mu}, \nu) - W_2^2(\mu_{k_\ell}, \nu)|\\ &= \frac{1}{2\tau}(W_2(\overline{\mu}, \nu) + W_2(\mu_{k_\ell}, \nu))|W_2(\overline{\mu}, \nu) - W_2(\mu_{k_\ell}, \nu)|,\end{aligned}$$

for all ℓ. Due to the triangular inequality, for all ν, $|W_2(\overline{\mu}, \nu) - W_2(\mu_{k_\ell}, \nu)| \leq W_2(\mu_{k_\ell}, \overline{\mu})$. Therefore,

$$|G_{\mu_{k_\ell}}(\nu) - G_{\overline{\mu}}(\nu)| \leq \frac{1}{2\tau}(W_2(\overline{\mu}, \nu) + W_2(\mu_{k_\ell}, \nu))W_2(\overline{\mu}, \mu_{k_\ell}).$$

In addition, $\mathcal{S}(\mu_0)$ is a compact set and W_2 is a continuous functional, then there is a constant M such that

$$|G_{\mu_{k_\ell}}(\nu) - G_{\overline{\mu}}(\nu)| \leq MW_2(\overline{\mu}, \mu_{k_\ell}),$$

for all ν. Since $\mu_{k_\ell} \to_\ell \overline{\mu}$, this implies the uniform convergence of the functionals $G_{\mu_{k_\ell}}(\nu)$ to $G_{\overline{\mu}}(\nu)$. In particular, this implies that for all $\epsilon > 0$, there is an ℓ_0 such that for all $\ell \geq \ell_0$, we have

$$|G_{\mu_{k_\ell}}(\nu) - G_{\overline{\mu}}(\nu)| < \epsilon,$$

for all ν. Let $\overline{\mu}^+ = \arg\min_\nu G_{\overline{\mu}}(\nu)$, and recall that $\mu_{k_\ell+1} = \arg\min_\nu G_{\mu_{k_\ell}}(\nu)$. Then, by the min properties:

$$\begin{aligned}&G_{\mu_{k_\ell}}(\mu_{k_\ell+1}) \leq G_{\mu_{k_\ell}}(\nu) < G_{\overline{\mu}}(\nu) + \epsilon\\ &\qquad\implies G_{\mu_{k_\ell}}(\mu_{k_\ell+1}) \leq G_{\overline{\mu}}(\overline{\mu}^+) + \epsilon,\\ &G_{\overline{\mu}}(\overline{\mu}^+) - \epsilon < G_{\overline{\mu}}(\nu) - \epsilon \leq G_{\mu_{k_\ell}}(\nu)\\ &\qquad\implies G_{\overline{\mu}}(\overline{\mu}^+) - \epsilon \leq G_{\mu_{k_\ell}}(\mu_{k_\ell+1}).\end{aligned}$$

That is, we have $|G_{\mu_{k_\ell}}(\mu_{k_\ell+1}) - G_{\overline{\mu}}(\overline{\mu}^+)| \leq \epsilon$ for all $\ell \geq \ell_0$. The fact that $\overline{\mu}$ is a fixed point for $G_{\overline{\mu}}(\nu)$ now follows from the set of inequalities:

$$G_{\overline{\mu}}(\overline{\mu}^+) \leq G_{\overline{\mu}}(\overline{\mu}) = F(\overline{\mu}) \leq G_{\mu_{k_\ell}}(\mu_{k_\ell+1}) < F(\mu_{k_\ell})$$

The gap $G_{\mu_{k_\ell}}(\mu_{k_\ell+1}) - G_{\overline{\mu}}(\overline{\mu}^+)$ can be made arbitrarily small by increasing ℓ, so it must be that $G_{\overline{\mu}}(\overline{\mu}) = F(\overline{\mu}) = G_{\overline{\mu}}(\overline{\mu}^+)$, which implies $\overline{\mu}^+ = \overline{\mu}$ is the solution to the minimization problem of $G_{\overline{\mu}}$ and satisfies $\nabla\left(\frac{\delta G}{\delta \nu}\right)_{\overline{\mu}} = 0$. The equation $\nabla\left(\frac{\delta G}{\delta \nu}\right)_{\overline{\mu}} = 0$ is equivalent to $\frac{1}{\tau}\nabla\phi_{\overline{\mu}\to\overline{\mu}} + \nabla\left(\frac{\delta F}{\delta \nu}\right)_{\overline{\mu}} = 0$. Since $\nabla\phi_{\overline{\mu}\to\overline{\mu}} = 0$, then $\overline{\mu}$ is a minimizer of F, and from the strict geodesic convexity of F we get that the minimizer is unique and $\overline{\mu} = \mu^\star$. Note that we can apply this reasoning to all the accumulation points $\tilde{\mu}$ of the sequence $\{\mu_k\}$. Since all the convergent subsequences of $\{\mu_k\}$ have the same limit $\mu^\star$ and $\{\mu_k\}$ is contained in $\mathcal{S}(\mu_0)$ which is compact, we conclude that the whole sequence $\{\mu_k\}$ converges to $\mu^\star$ in W_2, i.e., weakly as $k \to \infty$. □

The implementation of (2) can be challenging because involves the solution of an infinite-dimensional optimization problem. To address this, we determine the stochastic process in Ω that equivalently describes the recursion (2). More precisely, consider a proximal recursion in Ω from an initial condition $x_0 \in \Omega$:

$$x_{k+1} \in \arg\min_{z\in\Omega} \frac{1}{2\tau}|x_k - z|^2 + f_k(z), \tag{4}$$

where $\{f_k\}_{k\in\mathbb{N}}$ is a sequence of functions on Ω. Suppose that the initial condition x_0 is in fact a random variable distributed according to μ_0 (denoted $x_0 \sim \mu_0$). We are interested in defining the process in Ω, through an appropriate choice of $\{f_k\}_{k\in\mathbb{N}}$, which results in a consistent transport of the initial measure μ_0 according to the recursion (2).

Theorem 2 (***Target dynamics in*** Ω)**.** *Let $\Omega \subseteq \mathbb{R}^d$ be a compact, convex set, and let $F : \Omega \longrightarrow \mathbb{R}$ satisfy the conditions of Theorem 1. Under Assumption 1, the proximal recursion* (2)*, for $0 < \tau < 1/l$, starting from $\mu_0 \in \mathcal{P}^r(\Omega)$ is obtained as the transport of μ_0 by* (4) *with $x_0 \sim \mu_0$ and $f_k = \left.\frac{\delta F}{\delta \nu}\right|_{\mu_{k+1}}$, for all $k \in \mathbb{N}$.*

Proof. We rewrite the single-step update in (2) from an absolutely continuous probability measure $\mu \in \mathcal{P}(\Omega)$ as follows:

$$\mu^+ = \arg\min_{\nu\in\mathcal{P}(\Omega)} \frac{1}{2\tau}W_2^2(\mu, \nu) + F(\nu). \tag{5}$$

From Lemma 2 the minimizer μ^+ in (5) is unique. Let $\{\mathbf{v}_\epsilon\}$ be a smooth one-parameter family of vector fields such that $\mathbf{v}_0 = \mathbf{v}$, where $\mathbf{v}$ is any vector field on Ω. Now, define a one-parameter family of absolutely continuous probability measures $\{\nu_\epsilon\}_{\epsilon\in\mathbb{R}}$ by means of $\partial_\epsilon\nu_\epsilon + \nabla\cdot(\nu_\epsilon\mathbf{v}_\epsilon) = 0$, subject to $\mathbf{v}_\epsilon \cdot \mathbf{n} = 0$, and such that $\nu_0 = \mu^+$. Since μ^+ is a critical point of the objective function in (5), we have:

$$
\begin{aligned}
0 &= \frac{d}{d\epsilon}\left(\frac{1}{2\tau}W_2^2(\mu,\nu_\epsilon)+F(\nu_\epsilon)\right)\Big|_{\epsilon=0} \\
&= \frac{1}{\tau}\int_\Omega \left\langle \nabla\phi_{\mu^+\to\mu},\mathbf{v}\right\rangle d\mu^+ + \int_\Omega \langle \xi,\mathbf{v}\rangle\, d\mu^+ \\
&= \int_\Omega \left\langle \frac{1}{\tau}\nabla\phi_{\mu^+\to\mu}+\xi,\mathbf{v}\right\rangle d\mu^+,
\end{aligned}
$$

where $\xi = \nabla\left(\frac{\delta F}{\delta\nu}\right)\big|_{\nu=\mu^+}$ and $\nabla\phi_{\mu^+\to\mu} = \mathrm{id} - T_{\mu^+\to\mu}$, with $T_{\mu^+\to\mu}:\Omega\to\Omega$ being the optimal transport map from μ^+ to μ. Since $\int_\Omega \left\langle \frac{1}{\tau}\nabla\phi_{\mu^+\to\mu}+\xi,\mathbf{v}\right\rangle d\mu^+ = 0$ for all $\mathbf{v}$, it implies that $\frac{1}{\tau}\nabla\phi_{\mu^+\to\mu}+\xi = 0$ (μ^+ a.e. in Ω), and we obtain:

$$
\frac{1}{\tau}\nabla\phi_{\mu^+\to\mu}+\xi = \frac{1}{\tau}\left(\mathrm{id} - T_{\mu^+\to\mu}\right)+\xi = 0,
$$

which implies that:

$$
T_{\mu^+\to\mu} = \mathrm{id} + \tau\xi. \tag{6}
$$

Let $\varphi = \left(\frac{\delta F}{\delta\nu}\right)\big|_{\nu=\mu^+}$. For any $y\in\Omega$ and $\tau < 1/l$, consider:

$$
y^+ = \arg\min_{z\in\Omega} \underbrace{\frac{1}{2\tau}|y-z|^2+\varphi(z)}_{\triangleq g_y(z)}. \tag{7}
$$

The uniqueness of the minimizer above follows from the strong convexity of g_y for $\tau < 1/l$ (this can be verified by following a similar procedure as in the proof of Lemma 2, but now in the Euclidean space). If $y^+\in\mathring{\Omega}$ is a critical point of g_y in (7), then it satisfies $y^+ = y - \tau\nabla\varphi(y^+)$. Since $\xi = \nabla\varphi$, we can equivalently write $y^+ = (\mathrm{id}+\tau\xi)^{-1}(y)$. That is, when the image of $y\in\Omega$ under the $\arg\min$ map in (7) is a critical point in the interior of Ω, then it is also the inverse image of y under the optimal transport map $T_{\mu^+\to\mu}$.

Now, for a $y\in\mathring{\Omega}$, the inner product of the gradient of g_y at any point $z\in\partial\Omega$ on the boundary of Ω with the outward normal $\mathbf{n}$ to $\partial\Omega$ at z is given by $\nabla g_y\cdot\mathbf{n} = \left(\frac{1}{\tau}(z-y)+\nabla\varphi(z)\right)\cdot\mathbf{n} = \frac{1}{\tau}(z-y)\cdot\mathbf{n} > 0$, since $\nabla\varphi\cdot\mathbf{n}=0$ and $z-y$ points outward to Ω (as $z\in\partial\Omega$ and $y\in\mathring{\Omega}$ and Ω is convex). This implies that there exists a point $\tilde{z}$ in the interior of Ω in a neighborhood of z such that $g_y(\tilde{z}) < g_y(z)$, which implies that z cannot be the minimizer. Thus, for any $y\in\mathring{\Omega}$, the minimizer of $g_y(z) = \frac{1}{2\tau}|y-z|^2+\varphi(z)$ cannot lie on the boundary $\partial\Omega$, and must therefore lie in the interior of Ω and be a critical point of the objective function g_y. Now, when $y\in\partial\Omega$, if $y^+\notin\mathring{\Omega}$, it must be that $y^+=y$ (otherwise we obtain a contradiction for the same reason as above, the inner product of ∇g_y with the outward normal would be strictly positive) and the argmin map (and the optimal transport map) coincides with the identity map in this case.

It then follows that for any $y\in\Omega$, its image y^+ under the argmin map is exactly its inverse image under the optimal transport map $T_{\mu^+\to\mu}$. That is, the map in (7) is the inverse of the optimal transport map $T_{\mu^+\to\mu}$. Thus, we have that the map $T_{\mu^+\to\mu} = \mathrm{id}+\tau\xi$ is well-defined and so is its inverse, it holds that $\left(T_{\mu^+\to\mu}\right)^{-1}_{\#}\mu = (\mathrm{id}+\tau\xi)^{-1}_{\#}\mu = \mu^+$, and (5) is the lift to the space of probability measures of (7).

We therefore conclude that the proximal recursion (2) starting from μ_0 is the transport of μ_0 by (4) with $x_0\sim\mu_0$. $\square$

From a computational perspective, Theorem 2 still requires the evaluation of the first variation $\frac{\delta F}{\delta\nu}$ at μ_{k+1}, the transported measure at the future time instant $k+1$. To circumvent this problem, we can alternatively consider the dynamics (4) with the choice of $\tilde{f}_k = \frac{\delta F}{\delta\nu}\big|_{\mu_k}$, which only requires the evaluation, at time instant k, of the first variation $\frac{\delta F}{\delta\nu}$ at μ_k. Consider the l-smooth, geodesically-convex (linear) $\widetilde{F}(\nu) = \mathbb{E}_\nu\left[\frac{\delta F}{\delta\mu}\Big|_{\mu_k}\right]$, for $\nu\in\Omega$, which satisfies $\frac{\delta\widetilde{F}}{\delta\nu} = \frac{\delta F}{\delta\mu}\Big|_{\mu_k}$. It follows from Theorem 2 that the descent in $\mathcal{P}(\Omega)$ corresponding to (4) with $\tilde{f}_k = \frac{\delta\widetilde{F}}{\delta\nu}\Big|_{\mu_k}$ is given by:

$$
\mu_{k+1}\in\arg\min_{\nu\in\mathcal{P}(\Omega)} \frac{1}{2\tau}W_2^2(\mu_k,\nu)+\mathbb{E}_\nu\left[\frac{\delta F}{\delta\mu}\Big|_{\mu_k}\right]. \tag{8}
$$

The convergence of (8) can also be established as follows:

Theorem 3 (***Convergence of recursion*** (8))**.** *Let $F:\Omega\to\mathbb{R}$ satisfy the conditions of Theorem 1. The sequence $\{\mu_k\}_{k\in\mathbb{N}}$ obtained as the transport of measure $\mu_0\in\mathcal{P}^r(\Omega)$ by* (8) *with $\tau<1/l$, $x_0\sim\mu_0$ and the choice $\tilde{f}_k = \frac{\delta\widetilde{F}}{\delta\nu}\Big|_{\mu_k}$, converges weakly to $\mu^\star = \arg\min_{\nu\in\mathcal{P}(\Omega)} F(\nu)$ as $k\to\infty$.*

Proof. Suppose that $\{\mu_k\}$ is a sequence derived from (8). From the l-smoothness of F and Lemma 15 (with μ_{k+1} as the reference measure), we have:

$$
\begin{aligned}
&\int_\Omega \left\langle \nabla\left(\frac{\delta F}{\delta\nu}\Big|_{\mu_k} - \frac{\delta F}{\delta\nu}\Big|_{\mu_{k+1}}\right), T_{\mu_{k+1}\to\mu_k}-\mathrm{id}\right\rangle d\mu_{k+1} \\
&\le l W_2^2(\mu_k,\mu_{k+1}).
\end{aligned}
$$

By Lemma 2, we have that the objective functional in (8) is strongly convex and therefore has a unique minimizer, since $\mathbb{E}_\nu\left[\frac{\delta F}{\delta\mu}\Big|_{\mu_k}\right]$ is linear in ν for a given μ_k. Following similar steps as in the proof of Theorem 1 to characterize the critical point of (8), we get that $T_{\mu_{k+1}\to\mu_k} = \mathrm{id}+\tau\nabla\left(\frac{\delta F}{\delta\nu}\big|_{\mu_k}\right)$, and by substitution in the above, we obtain:

$$
\begin{aligned}
&\tau\int_\Omega \left\langle \nabla\left(\frac{\delta F}{\delta\nu}\Big|_{\mu_k} - \frac{\delta F}{\delta\nu}\Big|_{\mu_{k+1}}\right), \nabla\left(\frac{\delta F}{\delta\nu}\Big|_{\mu_k}\right)\right\rangle d\mu_{k+1} \\
&\le l W_2^2(\mu_k,\mu_{k+1}).
\end{aligned}
$$

Therefore, it follows that:

$$
\begin{aligned}
&\tau\int_\Omega \left\langle \nabla\left(\frac{\delta F}{\delta\nu}\Big|_{\mu_{k+1}}\right), \nabla\left(\frac{\delta F}{\delta\nu}\Big|_{\mu_k}\right)\right\rangle d\mu_{k+1} \\
&\ge \left(\frac{1}{\tau}-l\right) W_2^2(\mu_k,\mu_{k+1}),
\end{aligned}
$$

where we have used the fact that:

$$
\begin{aligned}
&\tau\int_\Omega \left\langle \nabla\left(\frac{\delta F}{\delta\nu}\Big|_{\mu_k}\right), \nabla\left(\frac{\delta F}{\delta\nu}\Big|_{\mu_k}\right)\right\rangle d\mu_{k+1} \\
&= \frac{1}{\tau}\int_\Omega \left\langle T_{\mu_{k+1}\to\mu_k}-\mathrm{id}, T_{\mu_{k+1}\to\mu_k}-\mathrm{id}\right\rangle d\mu_{k+1} \\
&= \frac{1}{\tau}W_2^2(\mu_k,\mu_{k+1}).
\end{aligned}
$$

Moreover, from the convexity of F and Lemma 17 (with μ_{k+1} as the reference measure) we have:

$$F(\mu_k) \geq F(\mu_{k+1}) + \int_\Omega \left\langle \nabla \left(\frac{\delta F}{\delta \nu}\bigg|_{\mu_{k+1}} \right), T_{\mu_{k+1}\to\mu_k} - \mathrm{id} \right\rangle d\mu_{k+1}.$$

Substituting in the latest inequality, we obtain:

$$F(\mu_k) \geq F(\mu_{k+1}) + \left(\frac{1}{\tau} - l\right) W_2^2(\mu_k, \mu_{k+1}).$$

From this inequality, we deduce that μ_{k+1} belongs to the F-sublevel set of μ_k, and consequently that the sequence $\{\mu_k\}_{k\in\mathbb{N}}$ is contained in $\mathcal{S}(\mu_0)$, the F-sublevel set of μ_0. From here, following similar steps as in the proof of Theorem 1, we conclude that the sequence $\{\mu_k\}_{k\in\mathbb{N}}$ is convergent and $\lim_{K\to\infty} W_2^2(\mu_K, \bar{\mu}) = 0$ for some $\bar{\mu} \in \mathcal{S}(\mu_0)$. As the sequence $\{\mu_k\}_{k\in\mathbb{N}}$ is generated by (8), the limit $\bar{\mu}$ must be one of its fixed points, again following similar reasoning as in Theorem 1. Since F is strictly convex, we get that the only fixed point of (8) is $\mu^\star$. We therefore have $\bar{\mu} = \mu^\star$. □

Now Theorem 2 allows us to consider the transport in $\mathcal{P}(\Omega)$ given by the following proximal scheme in Ω:

$$x^+ = \arg\min_{z\in\Omega} \frac{1}{2\tau}|x - z|^2 + f(z), \tag{9}$$

where $x \sim \mu$ and $f = \frac{\delta F}{\delta \nu}\big|_\mu$. This scheme is convergent according to Theorem 3.

IV. Multi-agent proximal descent algorithms

In this section, we bring the sample-based, proximal descent schemes of the previous section to a form that is closer to the more familiar multi-agent cooperative control algorithms. We achieve this by a direct discretization of the functional. By doing so, we are able to retain some convergence properties of the algorithms, as shown in this section. We then show that, in the limit of space and time discretizations, the corresponding algorithm recovers the lost properties.

We start by describing the multi-agent system by an appropriate probability distribution. Recall that the configuration of the collective is given by $\mathbf{x} = (x_1, \ldots, x_N)$, with $x_i \in \Omega$ for $i \in \{1, \ldots, N\}$. Let $\widehat{\mu}_{\mathbf{x}}^N = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$, be the discrete measure in $\mathcal{P}(\Omega)$ corresponding to the configuration $\mathbf{x}$. For a macroscopic description of the transport, we first let the macroscopic configuration be specified by an absolutely continuous probability measure, and since $\widehat{\mu}_{\mathbf{x}}^N$ is is not absolutely continuous, we consider an alternative absolutely continuous probability measure $\widehat{\mu}_{\mathbf{x}}^{h,N}$ through its density function using a smooth kernel, as follows:

$$\widehat{\mu}_{\mathbf{x}}^{h,N}(x) = \frac{1}{N}\sum_{i=1}^N K_h(x - x_i), \tag{10}$$

where $h > 0$ is the bandwidth of the kernel. With a slight abuse of notation, we allow $\widehat{\mu}_{\mathbf{x}}^{h,N}$ to denote both the absolutely continuous measure and its corresponding density function. We also denote, for $x \in \Omega$, $\widehat{\mu}_x^{h,1}$ simply by $\widehat{\mu}_x^h$. Thus, we have $\widehat{\mu}_{\mathbf{x}}^{h,N} = \frac{1}{N}\sum_{i=1}^N \widehat{\mu}_{x_i}^h$, for $\mathbf{x} \in \Omega^N$.

Assumption 2 (***Properties of kernel and kernel-based measures***)**.** *For $h > 0$ and $z \in \Omega$ and a kernel-based probability measure $\widehat{\mu}_z^h$ defined as in* (10) *for $N = 1$, the following hold:*
(i) Smoothness: *The kernel K_h is smooth, $K_h \in C^\infty(\Omega)$, for every $h > 0$.*
(ii) Monotonicity of support: *For any $z \in \Omega$ and $h_1 < h_2$, we let* $\mathrm{supp}\left(\widehat{\mu}_z^{h_1}\right) \subset \mathrm{supp}\left(\widehat{\mu}_z^{h_2}\right)$.
(iii) Containment: *For every $h > 0$, there exists a set $\tilde{\Omega}_h \subset \Omega$ (relatively) open, such that for $z \in \tilde{\Omega}_h$, the support of the measure $\widehat{\mu}_z^h$ satisfies* $\mathrm{supp}(\widehat{\mu}_z^h) \subset \Omega$. *Moreover,* $\lim_{h\to 0}\tilde{\Omega}_h = \Omega$ *in Hausdorff distance.*
(iv) Total variation convergence: *Let $\mathcal{M}$ be the space of all measureable functions over Ω. It holds that* $\lim_{h\to0}\sup_{f\in\mathcal{M}}\left\{\int_\Omega f(z)K_h(x-z)\,\mathrm{dvol}(z) - f(x)\right\} = 0$, *that is, the kernel-based measure converges uniformly to the Dirac measure as $h \to 0$.*

An example kernel for (10) that satisfies Assumption 2 is the truncated Gaussian kernel restricted to an open ball $B_h(x_i)$ of radius h centered at x_i, given by $K_h(x - x_i) = \frac{1}{C}\exp\left(\frac{-|x-x_i|^2}{2h^2}\right)\mathbf{1}_{B_h(x_i)}(x)$, where $C = \int_{B_h(x_i)} \exp\left(\frac{-|x-x_i|^2}{2h^2}\right)\mathrm{dvol}(x)$ is the normalizing constant.

A. Discretization of functional F and its properties

We define an aggregate objective function $F^{h,N}$ for the multi-agent system as the discretization of the functional F, for $h > 0$, as follows:

$$F^{h,N}(\mathbf{x}) = F(\widehat{\mu}_{\mathbf{x}}^{h,N}), \tag{11}$$

and, subsequently, analyze its properties. First note that $F^{h,N}$ is invariant under permutations, that is, for $\mathbf{x} \in \tilde{\Omega}_h^N$ and $P \in \mathbb{R}^{N\times N}$ a permutation, we have $F^{h,N}(\mathbf{x}) = F^{h,N}((P\otimes I_d)\,\mathbf{x})$. The following lemma establishes the almost sure convergence of the $F^{h,N}$ to F as $h \to 0, N \to \infty$:

Lemma 3 (***Convergence as*** $h \to 0$, $N \to \infty$)**.** *Let Assumption 2 and the Fréchet differentiability of the functional F hold, and let $x_i \sim \mu$ for $i \in \{1, \ldots, N\}$, independent and identically distributed. Then, we have* $\lim_{h\to0}\lim_{N\to\infty} F^{h,N}(x_1, \ldots, x_N) = F(\mu)$, *$\mu$-almost surely.*

Proof. We first recall that $F^{h,N}(\mathbf{x}) = F(\widehat{\mu}_{\mathbf{x}}^{h,N})$. By the Glivenko-Cantelli Theorem [36] and Assumption 2-(iv), we have:

$$\lim_{\substack{h\to0,\\ N\to\infty}} \sup_{f\in\mathcal{M}} \left\{\mathbb{E}_{\widehat{\mu}_{\mathbf{x}}^{h,N}}[f] - \mathbb{E}_\mu[f]\right\} = 0, \quad a.s.$$

We denote the above as $\widehat{\mu}_{\mathbf{x}}^{h,N} \to_{\mathrm{u.a.s}} \mu$, i.e., $\widehat{\mu}_{\mathbf{x}}^{h,N}$ converges uniformly almost surely to μ as $h \to 0$ and $N \to \infty$. Note that this implies the (almost sure) weak convergence of $\{\widehat{\mu}_{\mathbf{x}}^{h,N}\}$ to μ. Therefore, by continuity of F in the topology of weak convergence (which follows from the fact that F is Frećhet differentiable in the L^2-Wasserstein space), we have $\lim_{h\to0,N\to\infty} F^{h,N}(\mathbf{x}) = \lim_{h\to0,N\to\infty} F(\widehat{\mu}_{\mathbf{x}}^{h,N}) = F(\lim_{h\to0,N\to\infty}\widehat{\mu}_{\mathbf{x}}^{h,N}) = F(\mu)$, almost surely. □

The following lemma relates the derivative of the function $F^{h,N}$ to the Fréchet derivative of the functional F:

Lemma 4 (***Derivative of*** $F^{h,N}$). *Let Assumption 2 and the Fréchet differentiability of the functional F hold, and let $h > 0$ with set $\tilde{\Omega}_h$ as in Assumption 2-(iii). For $\mathbf{x} = (z, \eta) \in \tilde{\Omega}_h \times \tilde{\Omega}_h^{N-1}$, we have that the derivative of the function $F^{h,N}$ satisfies:*

$$\partial_1 F^{h,N}(z, \eta) = \frac{1}{N} \int_{\text{supp}(\widehat{\mu}_z^h)} \nabla \varphi_{\mathbf{x}}^{h,N} \, d\widehat{\mu}_z^h,$$

where $d\widehat{\mu}_z^h = \rho_z^h$ dvol with $\rho_z^h(x) = K(x - z, h)$, $\varphi_{\mathbf{x}}^{h,N} = \frac{\delta F}{\delta \nu} \mid_{\widehat{\mu}_{\mathbf{x}}^{h,N}}$ and ∂_1 denotes the derivative w.r.t the first argument.

Proof. Let $\mathbf{x}(t) = (x_1(t), \ldots, x_N(t))$ be a curve in $\tilde{\Omega}_h^N$ parametrized by $t \in \mathbb{R}$, with $\dot{\mathbf{x}}(0) = \mathbf{v} = (\mathbf{v}_1, \ldots, \mathbf{v}_N)$, where $\mathbf{v}_i \in \mathbb{R}^d$ for all $i \in \{1, \ldots, N\}$. As $F^{h,N}$ is differentiable, partial derivatives exist and we can write:

$$\frac{d}{dt} F^{h,N}(\mathbf{x}(0)) = \sum_{i=1}^{N} \left\langle \partial_i F^{h,N}(\mathbf{x}(0)), \mathbf{v}_i \right\rangle.$$

Since $F^{h,N}(\mathbf{x}) = F(\widehat{\mu}_{\mathbf{x}}^{h,N})$, using the Fréchet derivative of F, we can write:

$$\begin{aligned} \frac{d}{dt} F^{h,N}(\mathbf{x}(0)) &= \frac{1}{N} \sum_{i=1}^{N} \int_{\Omega} \left\langle \nabla \varphi_{\mathbf{x}(0)}^{h,N}, \mathbf{v}_i \right\rangle \, d\widehat{\mu}_{x_i(0)}^h \\ &= \frac{1}{N} \sum_{i=1}^{N} \left\langle \int_{\Omega} \nabla \varphi_{\mathbf{x}(0)}^{h,N} \, d\widehat{\mu}_{x_i(0)}^h, \mathbf{v}_i \right\rangle. \end{aligned}$$

This holds for all $\mathbf{v} = (\mathbf{v}_1, \ldots, \mathbf{v}_N)$ and $\mathbf{x}(0) \in \tilde{\Omega}_h^N$, thus, by uniqueness of the partial derivatives, it holds that:

$$\partial_i F^{h,N}(\mathbf{x}) = \frac{1}{N} \int_{\Omega} \nabla \varphi_{\mathbf{x}}^{h,N} \, d\widehat{\mu}_{x_i(0)}^h,$$

where ∂_i denotes the derivative w.r.t. the i^{th} argument, and we consider any $\mathbf{x}(0) \in \tilde{\Omega}_h^N$. From the previous expression:

$$\begin{aligned} \partial_1 F^{h,N}(z, \eta) &= \frac{1}{N} \int_{\Omega} \nabla \varphi_{\mathbf{x}}^{h,N} \, d\widehat{\mu}_z^h \\ &= \frac{1}{N} \int_{\text{supp}(\widehat{\mu}_z^h)} \nabla \varphi_{\mathbf{x}}^{h,N} \, d\widehat{\mu}_z^h, \end{aligned}$$

where $z \in \tilde{\Omega}_h$, $\eta \in \tilde{\Omega}_h^{N-1}$, $d\widehat{\mu}_z^h = \rho_z^h$ dvol with $\rho_z^h(x) = K(x - z, h)$, and $\varphi_{\mathbf{x}}^{h,N} = \frac{\delta F}{\delta \nu} \mid_{\widehat{\mu}_{\mathbf{x}}^{h,N}}$, and the result follows. □

From the invariance of $F^{h,N}$ under permutations, the expression in Lemma 4 holds for the partial derivative of $F^{h,N}$ w.r.t every component of $\mathbf{x}$.

Lemma 5 (α***-smoothness of*** $F^{h,N}$). *Let Assumption 2 and l-smoothness of F hold. Then there exists an $\alpha > 0$ such that $F^{h,N}$ is α-smooth.*

Proof. From l-smoothness of F, we have that the function $\varphi = \frac{\delta F}{\delta \nu}\Big|_{\mu}$ is continuously differentiable on Ω for all μ. We note that for $x, y \in \tilde{\Omega}_h$, $\widehat{\mu}_y^h(z) = \widehat{\mu}_x^h(z + (x - y))$ for all $z \in \text{supp}\left(\widehat{\mu}_y^h\right)$. For any $\mathbf{x} \in \tilde{\Omega}_h^N$, we use $(x_i, \mathbf{x}_{-i}) \in \tilde{\Omega}_h \times \tilde{\Omega}_h^{N-1}$ to denote the vector with its first entry equal to the i^{th} component of $\mathbf{x}$ and all others equal to the remaining $N - 1$ components of $\mathbf{x}$. We now have:

$$\begin{aligned} &\left\| \nabla F^{h,N}(\mathbf{y}) - \nabla F^{h,N}(\mathbf{x}) \right\| \\ &= \sqrt{\sum_{i=1}^{N} \left| \partial_1 F^{h,N}(y_i, \mathbf{y}_{-i}) - \partial_1 F^{h,N}(x_i, \mathbf{x}_{-i}) \right|^2} \\ &= \frac{1}{N} \sqrt{\sum_{i=1}^{N} \left| \int_{\Omega} \nabla \varphi_{\mathbf{y}}^{h,N}(z) d\widehat{\mu}_{y_i}^h(z) - \int_{\Omega} \nabla \varphi_{\mathbf{x}}^{h,N}(z) d\widehat{\mu}_{x_i}^h(z) \right|^2} \\ &= \frac{1}{N} \sqrt{\sum_{i=1}^{N} \left| \int_{\Omega} \left[\nabla \varphi_{\mathbf{y}}^{h,N}(z + (y_i - x_i)) - \nabla \varphi_{\mathbf{x}}^{h,N}(z) \right] d\widehat{\mu}_{x_i}^h(z) \right|^2} \\ &\leq \int_{\Omega} \left| \nabla \varphi_{\mathbf{y}}^{h,N}(z) - \nabla \varphi_{\mathbf{x}}^{h,N}(z) \right| d\widehat{\mu}_{\mathbf{x}}^{h,N}(z) \\ &\quad + \frac{1}{N} \sum_{i=1}^{N} \int_{\Omega} \left| \nabla \varphi_{\mathbf{y}}^{h,N}(z + (y_i - x_i)) - \nabla \varphi_{\mathbf{y}}^{h,N}(z) \right| d\widehat{\mu}_{x_i}^h(z) \\ &\leq l W_2(\widehat{\mu}_{\mathbf{x}}^{h,N}, \widehat{\mu}_{\mathbf{y}}^{h,N}) + M \|\mathbf{y} - \mathbf{x}\| \\ &\leq \alpha \|\mathbf{y} - \mathbf{x}\|, \end{aligned}$$

where the penultimate inequality results from the l-smoothness of F (which implies φ has a Lipschitz-continuous gradient in expectation). Moreover, the final inequality results from the fact that $W_2(\widehat{\mu}_{\mathbf{x}}^{h,N}, \widehat{\mu}_{\mathbf{y}}^{h,N}) \leq \|\mathbf{y} - \mathbf{x}\|$. □

In what follows, we will make the following assumption characterize the behavior of the discretization $F^{h,N}$ along the boundary through the following assumption:

Assumption 3 (***Boundary conditions***). *The function $F^{h,N}$ is Fréchet differentiable and its derivative satisfies the boundary condition $\partial_1 F^{h,N}(z, \xi) \cdot \mathbf{n}(z) = 0$ for $z \in \partial \tilde{\Omega}_h$ and all $\xi \in \tilde{\Omega}_h^{N-1}$.*

In general, note that $F^{h,N} : \Omega^N \to \mathbb{R}$ is nonconvex in spite of being the discretization of a strictly geodesically convex functional $F : \mathcal{P}(\Omega) \to \mathbb{R}$. This is because the notion of convexity of functions over Ω^N, which is the domain of the function $F^{h,N}$, is not implied by the notion of geodesic convexity over the space of probability measures over Ω. In this way, for $\mathbf{x}, \mathbf{y} \in \Omega^N$ with $\sum_{i=1}^{N} \frac{1}{N} \delta_{x_i}, \sum_{i=1}^{N} \frac{1}{N} \delta_{y_i} \in \mathcal{P}(\Omega)$ being the corresponding discrete measures, the supports of the geodesics (when they exist) between $\sum_{i=1}^{N} \frac{1}{N} \delta_{x_i}$ and $\sum_{i=1}^{N} \frac{1}{N} \delta_{y_i}$ in $\mathcal{P}(\Omega)$ do not necessarily correspond to the straight line segment between $\mathbf{x}$ and $\mathbf{y}$ in Ω^N. In what follows, we identify a condition that can guarantee convexity of the discretized functional. We note that this condition is employed later to prove the convergence of the discrete algorithms to local minimizers.

Definition 1 (***Cyclical monotonicity***). *A set $\Gamma \subset \Omega \times \Omega$ is cyclically monotone if any sequence $\{(x_i, y_i)\}_{i=1}^N$, with $(x_i, y_i) \in \Gamma$, satisfies:*

$$\sum_{i=1}^{N} |x_i - y_i|^2 \leq \sum_{i=1}^{N} |x_i - y_{\sigma(i)}|^2,$$

where σ is any permutation, $\sigma \in \Sigma_N$.

For $\delta > 0$, we define a subset $\Delta_\delta \subset \Omega^N$ as follows:

$$\Delta_\delta = \left\{ \mathbf{z} = (z_1, \ldots, z_N) \in \mathring{\Omega}^N \,\middle|\, |z_i - z_j| > \delta, \ \forall\, i \neq j \right\}.$$

For every $\mathbf{x} \in \Delta_\delta$, we now define a set $\Gamma_\mathbf{x} \subset \Omega^N$ such that for all $\mathbf{y} \in \Gamma_\mathbf{x}$, we have:

$$\sum_{i=1}^N |x_i - y_i|^2 \leq \sum_{i=1}^N |x_i - y_{\sigma(i)}|^2,$$

for any permutation σ. In other words, $\Gamma_\mathbf{x}$ is the subset of Ω^N such that for any $\mathbf{y} \in \Gamma_\mathbf{x}$, $\{(x_i, y_i)\}_{i=1}^N$ is cyclically monotone. We now establish through the following lemma that the set $\Gamma_\mathbf{x}$ contains an open neighborhood of $\mathbf{x}$:

Lemma 6 ($\Gamma_\mathbf{x}$ ***contains an open neighborhood of*** $\mathbf{x}$)**.** *For any $\delta > 0$ and $\mathbf{x} \in \Delta_\delta$, there exists an open neighborhood $\mathcal{N}(\mathbf{x}) \subset \Omega^N$ of $\mathbf{x}$ such that $\mathcal{N}(\mathbf{x}) \subset \Gamma_\mathbf{x}$.*

Proof. For $\mathbf{x} \in \Delta_\delta \subset \Omega^N$, let $\mathbf{y} \in \mathring{\Omega}^N$ such that for all $i \in \{1, \ldots, N\}$, we have $y_i \in B_{\delta/2}(x_i)$, where $B_{\delta/2}(x_i)$ is the open $\delta/2$-ball centered at $x_i \in \Omega$. Now for any $j \in \{1, \ldots, N\}$ with $j \neq i$, we have $|y_i - x_j| = |y_i - x_i + x_i - x_j| \geq |x_i - x_j| - |y_i - x_i| > \delta - \delta/2 > \delta/2$, since $|x_i - x_j| > \delta$ as $\mathbf{x} \in \Delta_\delta$ and $|y_i - x_i| < \delta/2$. Thus, among all (non-identity) permutations σ, we have:

$$\frac{1}{N}\sum_{i=1}^N |x_i - y_{\sigma(i)}|^2 > \frac{\delta^2}{4} > \frac{1}{N}\sum_{i=1}^N |x_i - y_i|^2.$$

Thus, we infer that $\mathbf{y} \in \Gamma_\mathbf{x}$ for an arbitrary $\mathbf{y} \in \Omega^N \cap \Pi_{i=1}^N B_{\delta/2}(x_i)$, and the result follows. ☐

From Lemma 6 that for $\mathbf{x} \in \Delta_\delta$ with a given $\delta > 0$, there is a $\bar{h}_\delta$ such that for all $0 < h < \bar{h}_\delta$, the supports of the components $\widehat{\mu}_{x_i}^h$ of the measure $\widehat{\mu}_\mathbf{x}^{h,N}$ can be made disjoint.

Lemma 7 (***Relaxation to atomless measures***)**.** *For any $\delta > 0$ and $\mathbf{x} \in \Delta_\delta$ and $\mathbf{y} \in \Gamma_\mathbf{x}$, there is $\bar{h}_\delta > 0$ such that for $0 \leq h \leq \bar{h}_\delta$ and the measures $\widehat{\mu}_\mathbf{x}^{h,N}, \widehat{\mu}_\mathbf{y}^{h,N}$ defined in* (10)*, the optimal transport map $T_{\widehat{\mu}_\mathbf{x}^{h,N} \to \widehat{\mu}_\mathbf{y}^{h,N}}$ from $\widehat{\mu}_\mathbf{x}^{h,N}$ to $\widehat{\mu}_\mathbf{y}^{h,N}$ satisfies:*

$$\left(T_{\widehat{\mu}_\mathbf{x}^{h,N} \to \widehat{\mu}_\mathbf{y}^{h,N}} - \mathrm{id}\right)(z) = y_i - x_i, \quad \forall\, z \in \mathrm{supp}\left(\widehat{\mu}_{x_i}^h\right).$$

Proof. The proof applies a generalization of Brenier's Theorem in [37]. We consider convex functions $\chi_i : \Omega \to \mathbb{R}$, for $i \in \{1, \ldots, N\}$ defined by:

$$\chi_i(z) = \frac{1}{2}\left|z + y_i - x_i\right|^2.$$

We note that the gradient of χ_i, $\nabla\chi_i(z) = z + y_i - x_i$ defines a map that transports the measure $\widehat{\mu}_{x_i}^h$ to $\widehat{\mu}_{y_i}^h$ simply by translation. In addition, this mapping defines a measure with cyclically monotone support and marginals $\widehat{\mu}_\mathbf{x}^{h,N}$ and $\widehat{\mu}_\mathbf{y}^{h,N}$. By the generalization of Brenier's Theorem [37] (c.f. Theorem 12 and extensions on uniqueness) a measure that has cyclic monotone support is both unique and optimal in the Monge-Kantorovich sense. Thus it coincides with the measure defined by the χ_i and the statement of the lemma follows. ☐

In this way, Lemma 7 essentially establishes that for $\mathbf{x} \in \Delta_\delta$ and any $\mathbf{y} \in \Gamma_\mathbf{x}$, the optimal transport from $\widehat{\mu}_\mathbf{x}^{h,N}$ to $\widehat{\mu}_\mathbf{y}^{h,N}$ is simply achieved by the translation of components $\widehat{\mu}_{x_i}^h$ along the rays $y_i - x_i$ to $\widehat{\mu}_{y_i}^h$ for each $i \in \{1, \ldots, N\}$.

Corollary 1 (L^2***-Wasserstein distance***)**.** *For any $\delta > 0$ and $\mathbf{x} \in \Delta_\delta$ and $\mathbf{y} \in \Gamma_\mathbf{x}$, there is a $\bar{h}_\delta > 0$ such that for any $0 < h \leq \bar{h}_\delta$:*

$$W_2^2\left(\widehat{\mu}_\mathbf{x}^{h,N}, \widehat{\mu}_\mathbf{y}^{h,N}\right) = \frac{1}{N}\sum_{i=1}^N |x_i - y_i|^2.$$

With the above results we now establish the following:

Lemma 8 (***Comparison lemma for*** $F^{h,N}$ ***on cyclically monotone sets***)**.** *Let F be a Fréchet differentiable and geodesically convex functional (in the sense of Definition 7). For any $\delta > 0$, $\mathbf{x} \in \Delta_\delta$, $h \in (0, \bar{h}_\delta]$ and $\mathbf{y} \in \Gamma_\mathbf{x}$:*

$$F^{h,N}(\mathbf{y}) \geq F^{h,N}(\mathbf{x}) + \left\langle \nabla F^{h,N}(\mathbf{x}), \mathbf{y} - \mathbf{x}\right\rangle.$$

Proof. For $\mathbf{x} \in \Delta_\delta$ and $\mathbf{y} \in \Gamma_\mathbf{x}$, using the geodesic convexity of the functional F and Lemma 17 with $\widehat{\mu}_\mathbf{x}^{h,N}$ as the reference measure, it follows that:

$$\begin{aligned}
&F^{h,N}(\mathbf{y}) = F(\widehat{\mu}_\mathbf{y}^{h,N}) \\
&\geq F(\widehat{\mu}_\mathbf{x}^{h,N}) + \int_\Omega \left\langle \nabla\varphi_\mathbf{x}^{h,N}, T_{\widehat{\mu}_\mathbf{x}^{h,N} \to \widehat{\mu}_\mathbf{y}^{h,N}} - \mathrm{id}\right\rangle d\widehat{\mu}_\mathbf{x}^{h,N} \\
&= F(\widehat{\mu}_\mathbf{x}^{h,N}) + \frac{1}{N}\sum_{i=1}^N \int_\Omega \left\langle \nabla\varphi_\mathbf{x}^{h,N}, T_{\widehat{\mu}_\mathbf{x}^{h,N} \to \widehat{\mu}_\mathbf{y}^{h,N}} - \mathrm{id}\right\rangle d\widehat{\mu}_{x_i}^h \\
&= F(\widehat{\mu}_\mathbf{x}^{h,N}) + \frac{1}{N}\sum_{i=1}^N \int_{\mathrm{supp}(\widehat{\mu}_{x_i}^h)} \left\langle \nabla\varphi_\mathbf{x}^{h,N}, T_{\widehat{\mu}_\mathbf{x}^{h,N} \to \widehat{\mu}_\mathbf{y}^{h,N}} - \mathrm{id}\right\rangle d\widehat{\mu}_{x_i}^h \\
&= F(\widehat{\mu}_\mathbf{x}^{h,N}) + \frac{1}{N}\sum_{i=1}^N \int_{\mathrm{supp}(\widehat{\mu}_{x_i}^h)} \left\langle \nabla\varphi_\mathbf{x}^{h,N}, y_i - x_i\right\rangle d\widehat{\mu}_{x_i}^h \\
&= F(\widehat{\mu}_\mathbf{x}^{h,N}) + \frac{1}{N}\sum_{i=1}^N \left\langle \int_{\mathrm{supp}(\widehat{\mu}_{x_i}^h)} \nabla\varphi_\mathbf{x}^{h,N} d\widehat{\mu}_{x_i}^h\ ,\ y_i - x_i\right\rangle \\
&= F^{h,N}(\mathbf{x}) + \sum_{i=1}^N \left\langle \partial_1 F^{h,N}(x_i, \mathbf{x}_{-i}), y_i - x_i\right\rangle,
\end{aligned}$$

thereby establishing the claim. ☐

We remark here that $F^{h,N}$ is convex in the limited sense established by the comparison result in Lemma 8, and this does not necessarily generalize to the entire domain Ω^N, due to which the function $F^{h,N}$ can be non-convex in general.

B. Multi-agent proximal descent algorithms

We formulate the proximal descent algorithm on the function $F^{h,N}$ as follows:

$$\mathbf{x}^+ \in \arg\min_{\mathbf{z} \in \tilde{\Omega}_h^N} \frac{1}{2\tau}\|\mathbf{x} - \mathbf{z}\|^2 + F^{h,N}(\mathbf{z}). \tag{12}$$

Even though $F^{h,N}$ is in general nonconvex, we can establish strong convexity of the proximal descent objective function in (12) under some conditions through the following lemma:

Lemma 9 (***Strong convexity of objective function***)**.** *For an α-smooth function $F^{h,N}$, the function $G_\mathbf{x}^{h,N}(\mathbf{z}) = \frac{1}{2\tau}\|\mathbf{x} - \mathbf{z}\|^2 + F^{h,N}(\mathbf{z})$ is $\left(\frac{1}{\tau} - \alpha\right)$-strongly convex for $0 < \tau < \frac{1}{\alpha}$.*

Proof. From Lemma 5 on α-smoothness of $F^{h,N}$, we have:

$$\left|\left\langle \nabla F^{h,N}(\mathbf{y}) - \nabla F^{h,N}(\mathbf{x}), \mathbf{y} - \mathbf{x}\right\rangle\right| \leq \alpha\|\mathbf{y} - \mathbf{x}\|^2.$$

With $G^{h,N}_{\mathbf{x}}(\mathbf{z}) = \frac{1}{2\tau}\|\mathbf{x}-\mathbf{z}\|^2 + F^{h,N}(\mathbf{z})$, we have:

$$\begin{aligned}&\left\langle \nabla G^{h,N}_{\mathbf{x}}(\mathbf{z}_1) - \nabla G^{h,N}_{\mathbf{x}}(\mathbf{z}_2), \mathbf{z}_1 - \mathbf{z}_2 \right\rangle \\ &= \left\langle \frac{1}{\tau}(\mathbf{z}_1 - \mathbf{z}_2) + \nabla F^{h,N}(\mathbf{z}_1) - \nabla F^{h,N}(\mathbf{z}_2)\ ,\ \mathbf{z}_1 - \mathbf{z}_2 \right\rangle \\ &= \frac{1}{\tau}\|\mathbf{z}_1 - \mathbf{z}_2\|^2 + \left\langle \nabla F^{h,N}(\mathbf{z}_1) - \nabla F^{h,N}(\mathbf{z}_2)\ ,\ \mathbf{z}_1 - \mathbf{z}_2 \right\rangle \\ &\geq \frac{1}{\tau}\|\mathbf{z}_1 - \mathbf{z}_2\|^2 - \alpha\|\mathbf{z}_1 - \mathbf{z}_2\|^2 \\ &= \left(\frac{1}{\tau} - \alpha\right)\|\mathbf{z}_1 - \mathbf{z}_2\|^2,\end{aligned}$$

thereby establishing the claim. □

It follows from Lemma 9 that the minimizer in (12) is unique for α-smooth $F^{h,N}$ and sufficiently small τ. Now, with $\mathbf{x}_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_N) \in \tilde{\Omega}_h^{N-1}$, we can write $F^{h,N}(x_1, \ldots, x_N) = \frac{1}{N}\sum_{i=1}^N F^{h,N}(x_1, \ldots, x_N) = \frac{1}{N}\sum_{i=1}^N F^{h,N}(x_i, \mathbf{x}_{-i})$. By means of this decomposition, the proximal gradient descent (12) can be decomposed into the following agent-wise update, for $i \in \{1, \ldots, N\}$:

$$x_i^+ = \arg\min_{z \in \overline{\Omega}_h} \frac{1}{2\tau}|x_i - z|^2 + F^{h,N}(z, \mathbf{x}_{-i}^+).$$

where $\overline{\Omega}_h$ is the closure of $\tilde{\Omega}_h$. Note that the above scheme requires $\mathbf{x}_{-i}^+$. In other words, to implement the above algorithm, every agent i, at time k, requires the positions of the other agents at a future time $k+1$, posing a hurdle for implementation. To avoid this problem, we consider the following proximal descent scheme:

$$x_i^+ = \arg\min_{z \in \overline{\Omega}_h} \frac{1}{2\tau}|x_i - z|^2 + F^{h,N}(z, \mathbf{x}_{-i}), \tag{13}$$

for every $i \in \{1, \ldots, N\}$. It follows from Lemma 9 that the objective function in (13) is also strongly convex, and thereby has a unique minimizer. We now present the following result on the convergence of (13) to the local minimizers of $F^{h,N}$:

Theorem 4 (***Convergence of*** (13) ***to critical points of*** $F^{h,N}$**).** *Let $F^{h,N}$ be α-smooth and satisfy Assumption 3. For $\tau < \frac{2}{3\alpha}$, the sequence $\{\mathbf{x}(k)\}_{k\in\mathbb{N}}$ generated by the update scheme* (13) *converges to a critical point $\mathbf{x}^*$ of $F^{h,N}$ that is not a local maximizer, for all initial conditions $\mathbf{x}(0) \in \overline{\Omega}_h^N$. Moreover, if the critical point $\mathbf{x}^* \in \Delta_\delta$ for some $\delta > 0$ and $h \in (0, \bar{h}_\delta]$, then $\mathbf{x}^*$ is a local minimizer.*

Proof. We first consider the objective function in (13), $J_i(z) = \frac{1}{2\tau}|x_i - z|^2 + F^{h,N}(z, \mathbf{x}_{-i})$, with $z \in \overline{\Omega}_h$. The inner product of the gradient of J_i on $z \in \partial\overline{\Omega}_h$ with the outward normal $\tilde{\mathbf{n}}$ to $\partial\overline{\Omega}_h$, is given by:

$$\begin{aligned}\nabla J_i(z) \cdot \tilde{\mathbf{n}}(z) &= \frac{1}{\tau}(z - x_i) \cdot \tilde{\mathbf{n}}(z) + \partial_1 F^{h,N}(z, \mathbf{x}_{-i}) \cdot \tilde{\mathbf{n}}(z) \\ &= \frac{1}{\tau}(z - x_i) \cdot \tilde{\mathbf{n}}(z) \geq 0,\end{aligned}$$

with the inequality being strict when $x_i \notin \partial\overline{\Omega}_h$. This implies that the $x_i^+ \in \partial\overline{\Omega}_h$ cannot be a minimizer if $x_i \notin \partial\overline{\Omega}_h$, and if $x_i \in \partial\overline{\Omega}_h$, we will have $x_i^+ = x_i$. In both cases, the minimizer x_i^+ is also a critical point of the function J_i. This allows us to express (13) equivalently by:

$$x_i^+ = x_i - \tau\partial_1 F^{h,N}(x_i^+, \mathbf{x}_{-i}). \tag{14}$$

We note that in the limit $\tau \to 0$, we get a gradient flow that can be shown to converge to a critical point of $F^{h,N}$. We therefore hope that this property is preserved over a neighborhood of $\tau = 0$. In what follows, we establish that this is indeed the case and provide a sufficient strict upper bound on τ for which the property is preserved. From α-smoothness of $F^{h,N}$, we get:

$$\begin{aligned}&\left| F^{h,N}(\mathbf{x}^+) - F^{h,N}(\mathbf{x}) - \sum_{i=1}^N \left\langle \partial_1 F^{h,N}(x_i, \mathbf{x}_{-i}), x_i^+ - x_i \right\rangle \right| \\ &\leq \frac{\alpha}{2}\|\mathbf{x}^+ - \mathbf{x}\|^2.\end{aligned}$$

We can rewrite the above as:

$$\begin{aligned}&\Bigg| F^{h,N}(\mathbf{x}^+) - F^{h,N}(\mathbf{x}) - \sum_{i=1}^N \left\langle \partial_1 F^{h,N}(x_i^+, \mathbf{x}_{-i}), x_i^+ - x_i \right\rangle \\ &- \sum_{i=1}^N \left\langle \partial_1 F^{h,N}(x_i, \mathbf{x}_{-i}) - \partial_1 F^{h,N}(x_i^+, \mathbf{x}_{-i}), x_i^+ - x_i \right\rangle \Bigg| \\ &\leq \frac{\alpha}{2}\|\mathbf{x}^+ - \mathbf{x}\|^2.\end{aligned}$$

By (14), we now have $-\sum_{i=1}^N \left\langle \partial_1 F^{h,N}(x_i^+, \mathbf{x}_{-i}), x_i^+ - x_i \right\rangle = \frac{1}{\tau}\|\mathbf{x}^+ - \mathbf{x}\|^2$ and by the α-smoothness of $F^{h,N}$:

$$\begin{aligned}&\left| \sum_{i=1}^N \left\langle \partial_1 F^{h,N}(x_i, \mathbf{x}_{-i}) - \partial_1 F^{h,N}(x_i^+, \mathbf{x}_{-i}), x_i^+ - x_i \right\rangle \right| \\ &\leq \alpha\|\mathbf{x}^+ - \mathbf{x}\|^2.\end{aligned}$$

From the above inequalities, we therefore obtain:

$$F^{h,N}(\mathbf{x}^+) \leq F^{h,N}(\mathbf{x}) - \left(\frac{1}{\tau} - \frac{3\alpha}{2}\right)\|\mathbf{x}^+ - \mathbf{x}\|^2.$$

Thus, for $\tau < \frac{2}{3\alpha}$, when every agent follows the update (13), we get a descent in $F^{h,N}$, and $\mathbf{x}^+$ belongs to the $F^{h,N}$-sublevel set of $\mathbf{x}$. We can express the above inequality for any time instant $k \in \mathbb{N}$ as:

$$F^{h,N}(\mathbf{x}(k+1)) \leq F^{h,N}(\mathbf{x}(k)) - \left(\frac{1}{\tau} - \frac{3\alpha}{2}\right)\|\mathbf{x}(k+1) - \mathbf{x}(k)\|^2.$$

Summing over $k = 0, \ldots, K-1$, we obtain:

$$F^{h,N}(\mathbf{x}(K)) \leq F^{h,N}(\mathbf{x}(0)) - \left(\frac{1}{\tau} - \frac{3\alpha}{2}\right)\sum_{k=1}^K \|\mathbf{x}(k) - \mathbf{x}(k-1)\|^2,$$

and it follows that:

$$\begin{aligned}&\sum_{k=1}^K \|\mathbf{x}(k) - \mathbf{x}(k-1)\|^2 \\ &\leq \left(\frac{1}{\frac{1}{\tau} - \frac{3\alpha}{2}}\right)\left(F^{h,N}(\mathbf{x}(0)) - F^{h,N}(\mathbf{x}(K))\right).\end{aligned}$$

Since the sequence $\{\mathbf{x}(k)\}_{k\in\mathbb{N}}$ belongs to the $F^{h,N}$-sublevel set of $\mathbf{x}(0)$ (for all $\mathbf{x}(0) \in \overline{\Omega}_h^N$), which is a subset of $\overline{\Omega}_h^N$ (compact), it is precompact. By the boundedness above, in the limit $K \to \infty$, we get $\lim_{K\to\infty}\|\mathbf{x}(K) - \mathbf{x}(K-1)\|^2 = 0$.

Since $\overline{\Omega}_h$ is compact, there is a convergent subsequence $\{\mathbf{x}(k_\ell)\}$ to a point $\overline{\mathbf{x}} \in \overline{\Omega}_h^N$. Given $\mathbf{x}$, define the mapping

$$G^{h,N}_{\mathbf{x}}(\mathbf{z}) = \left(\frac{1}{\tau} - \frac{3\alpha}{2}\right)\|\mathbf{x} - \mathbf{z}\|^2 + F^{h,N}(\mathbf{z}), \quad \mathbf{z} \in \overline{\Omega}_h^N.$$

Let $\overline{\mathbf{x}}^+$ be the next iteration of (14) from $\overline{\mathbf{x}}$. Then, from the above, $G^{h,N}_{\overline{\mathbf{x}}}(\overline{\mathbf{x}}^+) \leq F^{h,N}(\overline{\mathbf{x}}) = G^{h,N}_{\overline{\mathbf{x}}}(\overline{\mathbf{x}})$. Due to the fact that $\mathbf{x}(k_\ell)$ converges to $\overline{\mathbf{x}}$, we also have that $G^{h,N}_{\overline{\mathbf{x}}}(\overline{\mathbf{x}}) = F^{h,N}(\overline{\mathbf{x}}) \leq G^{h,N}_{\mathbf{x}(k_\ell)}(\mathbf{x}(k_\ell+1))$, for all ℓ. Following similar steps as in the proof of Theorem 1, one can find a constant M such that $|G^{h,N}_{\overline{\mathbf{x}}}(\mathbf{z}) - G^{h,N}_{\mathbf{x}(k_\ell)}(\mathbf{z})| \leq M\|\overline{\mathbf{x}} - \mathbf{x}(k_\ell)\|$ for all $\mathbf{z} \in \overline{\Omega}^N_h$. This implies that $|G^{h,N}_{\overline{\mathbf{x}}}(\overline{\mathbf{x}}^+) - G^{h,N}_{\mathbf{x}(k_\ell)}(\mathbf{x}(k_\ell+1))| \leq \epsilon$, for all $\ell \geq \ell_0$. It is easy to see that $G^{h,N}_{\overline{\mathbf{x}}}(\overline{\mathbf{x}}^+) \leq G^{h,N}_{\overline{\mathbf{x}}}(\overline{\mathbf{x}}) \leq G^{h,N}_{\mathbf{x}(k_\ell)}(\mathbf{x}(k_\ell+1))$ holds, and thus $G^{h,N}_{\overline{\mathbf{x}}}(\overline{\mathbf{x}}^+) = F^{h,N}(\overline{\mathbf{x}})$, which can only happen when $\overline{\mathbf{x}}^+ = \overline{\mathbf{x}}$. In other words, $\overline{\mathbf{x}}$ is a fixed point of (14), and we thereby get:

$$\partial_1 F^{h,N}(\overline{x}_i, \overline{\mathbf{x}}_{-i}) = 0, \quad \forall\, i \in \{1, \ldots, N\},$$

and $\nabla F^{h,N}(\overline{\mathbf{x}}) = 0$. From here, the point $\overline{\mathbf{x}}$ cannot be a local maximizer since $\{F^{h,N}(\mathbf{x}(k_\ell))\}_{k\in\mathbb{N}}$ is decreasing and lower-bounded by $F^{h,N}(\overline{\mathbf{x}})$ and, consequently, every neighborhood of $\overline{\mathbf{x}}$ contains at least one point with a higher value of $F^{h,N}$. Note that this conclusion applies for every accumulation point of the entire sequence $\{\mathbf{x}(k)\}_{k\in\mathbb{N}}$.

Finally, suppose that an accumulation point $\overline{\mathbf{x}}$ satisfies $\overline{\mathbf{x}} \in \Delta_\delta$, for some $\delta > 0$ and $h \in (0, \bar{h}_\delta]$. From Lemmas 8 and 6, we conclude that there exists an open ball $B(\overline{\mathbf{x}}) \subset \Omega^N$ such that for all $\mathbf{x} \in B(\overline{\mathbf{x}})$, we have $F^{h,N}(\mathbf{x}) \geq F^{h,N}(\overline{\mathbf{x}})$, which implies that $\overline{\mathbf{x}}$ must be a local minimizer. □

Theorem 4 establishes the convergence of (13) to critical points of the function $F^{h,N}$, which are not necessarily local maximizers. This is a weaker result than Theorem 2, which established convergence of the transport scheme (9) to the global minimizer $\mu^\star$ of F. The guarantee is weakened after the discretization of F, which is involved in defining the multi-agent transport scheme (the convergence results for F employ the convexity properties of F, which are lost by $F^{h,N}$.) However, we can still hope to achieve the convergence to the global minimizer in the limit of particle and time discretizations, thereby guaranteeing best performance asymptotically. In the section that follows, we evaluate this possibility.

C. Continuous-time and many-particle limits

We now derive the continuous-time and many-particle limits for the multi-agent transport scheme (13), retrieving (9) from (13) as $N \to \infty$ and $h \to 0$ limit. We know from Theorem 2 that transport of a probability measure μ_0 by (9), which is identical to the following:

$$\begin{aligned} x^+ = \arg\min_{z\in\Omega}\ \frac{1}{2\tau}|x - z|^2 + \varphi(z), \\ x \sim \mu. \end{aligned} \tag{15}$$

with $\varphi \equiv \frac{\delta F}{\delta \nu}\big|_\mu$, is guaranteed to converge to the global minimizer μ^* of F. Informally, we see that as $\tau \to 0$ in (15), we have that $x^+ \to x$ and we let $\mathbf{v}(x) = \lim_{\tau\to 0} \frac{x^+ - x}{\tau} = -\nabla\varphi(x)$. We can thus expect the solutions to (15) converge to the solution of the gradient flow under the vector field $\mathbf{v} = -\nabla\varphi$. We now show, in a weak sense, that the above reasoning holds. We observe that the vector field $\mathbf{v} = -\nabla\varphi$ satisfies a zero-flux boundary condition $\mathbf{v}\cdot\mathbf{n} = \nabla\varphi\cdot\mathbf{n} = 0$ on $\partial\Omega$ owing to the definition of the functional F.

Proposition 1 (***Model of transport in the continuous time and many-particle limits***)**.** *Let Ω and F satisfy the assumptions of Theorem 1. The following hold:*
(i) Convergence of update scheme: *The scheme* (13) *converges in distribution to* (15) *in the limit $N \to \infty$.*
(ii) Gradient flow: *For every decreasing sequence $\{\tau_n\}_{n\in\mathbb{N}}$ satisfying $\tau_0 < \frac{1}{l}$ and $\lim_{n\to\infty}\tau_n = 0$, the sequence of solutions $\{x^n\}_{n\in\mathbb{N}}$ to* (15) *with corresponding $\{\tau_n\}_{n\in\mathbb{N}}$ contains a convergent subsequence, and the limit is a weak solution to the gradient flow:*

$$\partial_t X^t(x) = -\nabla\varphi_t(X^t(x)), \tag{16}$$

with $X^0(x) = x$, $\mu(t) = X^t_\# \mu_0$ and $\varphi_t = \frac{\delta F}{\delta\nu}\big|_{\mu(t)}$.
(iii) Continuity equation: *Let $T > 0$ and $\mathbf{v} \in L^\infty([0,T] \times \mathrm{Lip}(\Omega)^d)$, and $\dot{x}_i(t) = \mathbf{v}(t, x_i(t))$ for any $t \in [0,T]$ and $i \in \mathbb{N}$, with $x_i(0) \sim_{i.i.d}\mu_0$. Then, for $\mathbf{x}^N = (x_1, \ldots, x_N)$ for any $N \in \mathbb{N}$, the sequence $\{\mathbf{x}^N\}_{N\in\mathbb{N}}$ converges in a distributional sense to a solution μ of the continuity equation:*

$$\frac{\partial\mu}{\partial t} + \nabla\cdot(\mu\mathbf{v}) = 0, \qquad \mu(0) = \mu_0. \tag{17}$$

Owing to space constraints, we skip here the proof of Proposition 1. The gradient flow on the functional F is defined here as the transport (17) with $\mathbf{v} = -\nabla\varphi$ as in (16). Recall that the gradient flow satisfies the boundary condition $\nabla\varphi\cdot\mathbf{n} = 0$ on $\partial\Omega$. The following theorem establishes the asymptotic stability of the gradient flow on F, with convergence to $\mu^* \in \mathcal{P}(\Omega)$, the global minimizer of F as $t \to \infty$:

Theorem 5 (***Asymptotic stability of gradient flow***)**.** *Let $\Omega \subseteq \mathbb{R}^d$ be a compact, convex set and $F : \Omega \to \mathbb{R}$ be an l-smooth and strictly geodesically convex functional with minimizer $\mu^\star$. Then the solutions to the gradient flow w.r.t. F converge to $\mu^\star$ in the limit $t \to \infty$.*

Proof. Let $\{\mu_t\}_{t\geq 0}$ be a solution gradient flow w.r.t. F in $\mathcal{P}(\Omega)$. We have:

$$\frac{d}{dt}F(\mu_t) = \int_\Omega \left\langle \nabla\left(\frac{\delta F}{\delta\mu}\right), \mathbf{v}\right\rangle d\mu_t = -\int_\Omega \left|\nabla\left(\frac{\delta F}{\delta\mu}\right)\right|^2 d\mu_t \leq 0.$$

This implies that $F(\mu_t) \leq F(\mu_0)$ for all $t \geq 0$, and therefore $\{\mu_t\}_{t\geq 0}$ is contained in the sublevel set $\mathcal{S}(\mu_0) = \{\nu \in \mathcal{P}(\Omega) | F(\nu) \leq F(\mu_0)\}$. From Lemma 1, we have that $S(\mu_0)$ is compact in $(\mathcal{P}(\Omega), W_2)$, which implies that the orbit $\{\mu_t\}_{t\geq 0}$ is precompact. Moreover, the functional F is lower bounded in $S(\mu_0)$ by $F(\mu^*)$. By the LaSalle invariance principle for Banach spaces [38]–[40], we have that the orbit converges in $(\mathcal{P}(\Omega), W_2)$ (also weakly, from Lemma 14) asymptotically to the largest invariant set contained in $\dot{F}^{-1}(0)$. We have:

$$\dot{F}^{-1}(0) = \left\{\mu \in \mathcal{P}(\Omega) \,\middle|\, \nabla\left(\frac{\delta F}{\delta\mu}\right) = 0, \text{ a.e. in } \Omega\right\},$$

which implies that the Fréchet derivative of F is zero in the set $\dot{F}^{-1}(0)$. This corresponds to the set of critical points of F and from the strict geodesic convexity of F, we therefore get that $\dot{F}^{-1}(0) = \{\mu^\star\}$. □

V. Multi-agent coverage control algorithms

In this section, we aim to place well-known multi-agent coverage control algorithms in the literature [1], [4] within the multiscale theoretical framework established in the previous sections, in an effort to understand the macroscopic behavior of the coverage algorithms. To do this, we first relate the corresponding coverage objective functions used in both formulations and then apply our results to analyze their behavior in the limit $N \to \infty$. We begin with a widely-used aggregate objective function for coverage control of multi-agent systems, the multi-center distortion function, and then obtain its functional counterpart in the space of probability measures. The multi-center distortion function $\mathcal{H}_f : \Omega^N \to \mathbb{R}_{\geq 0}$ [1] is given by:

$$\mathcal{H}_f(\mathbf{x}) = \int_\Omega \min_{i \in \{1,\dots,N\}} f(|x - x_i|) d\mu^*(x). \tag{18}$$

where $f : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is a non-decreasing function and $\mu^\star(x) = \rho^\star(x)\,\mathrm{dvol}$, with $\rho^\star$ a target density in Ω. The Voronoi partition of Ω, $\{\mathcal{V}_i\}_{i=1}^N$, generated by $\mathbf{x} \in \Omega^N$ facilitates the analysis of $\mathcal{H}_f$ and is defined is as follows:

$$\mathcal{V}_i = \{x \in \Omega \mid |x - x_i| \leq |x - x_j| \; \forall j \in \{1,\dots,N\}\}, \forall\, i.$$

The following proposition establishes the relationship between $\mathcal{H}_f$ and the optimal transport cost C_f in (1):

Proposition 2 (***Optimal transport formulation of coverage objective***)**.** *The aggregate objective function $\mathcal{H}_f$ as defined in* (18)*, satisfies:*

$$\mathcal{H}_f(\mathbf{x}) = \min_{\mathbf{w} \in \Delta^{N-1}} C_f\left(\sum_{i=1}^N w_i \delta_{x_i} \;,\; \mu^\star\right).$$

where $\Delta^{N-1} = \{\mathbf{w} \in \mathbb{R}^N_{\geq 0} \;\big|\; \sum_{i=1}^N w_i = 1\}$ is the $(N-1)$-simplex. Furthermore, the minimizing weights $\mathbf{w}^\star = (w_1^\star, \dots, w_N^\star)$ are given by $w_i^\star = \mu^\star(\mathcal{V}_i)$, where $\{\mathcal{V}_i\}_{i=1}^N$ is the Voronoi partition of Ω.

We skip the proof of Proposition 2 here owing to space constraints. The following corollary applies Proposition 2 to the special case of $f(x) = x^2$:

Corollary 2 (***L^2-Wasserstein distance as aggregate objective function***)**.** *Applying Proposition 2 with a quadratic cost $f(x) = x^2$ (and the corresponding aggregate objective function $\mathcal{H}_2$), we have:*

$$\mathcal{H}_2(\mathbf{x}) = W_2^2\left(\sum_{i=1}^N \mu^\star(\mathcal{V}_i)\delta_{x_i}, \mu^\star\right).$$

We now investigate the properties of the aggregate objective function $\mathcal{H}_f$ in the limit $N \to \infty$.

Lemma 10. *Let $\mu^\star \in \mathcal{P}(\Omega)$ be an absolutely continuous measure defining $\mathcal{H}_f$. Let $x_i \sim_{i.i.d} \mu$, for $i \in \{1,\dots,N\}$, where $\mu \in \mathcal{P}(\Omega)$ is any absolutely continuous probability measure such that* $\mathrm{supp}(\mu) \supseteq \mathrm{supp}(\mu^\star)$*. It holds almost surely that* $\lim_{N\to\infty} \mathcal{H}_f(\mathbf{x}) = 0$.

Proof. From the Glivenko-Cantelli theorem, it follows that, as $N \to \infty$, the limit $\sum_{i=1}^N \mu^\star(\mathcal{V}_i)\delta_{x_i} \to \mu^\star$ holds almost surely, in the weak sense (from the expectation w.r.t. $\sum_{i=1}^N \mu^\star(\mathcal{V}_i)\delta_{x_i}$ of any simple function). Thus, by the continuity of C_f:

$$\lim_{N\to\infty} \mathcal{H}_f(\mathbf{x}) = \lim_{N\to\infty} C_f\left(\sum_{i=1}^N \mu^\star(\mathcal{V}_i)\delta_{x_i}, \mu^\star\right) = 0.$$

□

The previous result holds for any configuration of the points $\{x_i\}_{i=1}^N$ as long as they are sampled from a distribution whose support contains that of $\mu^\star$. Note that this is consistent with what happens in the discrete particle case, in the coverage control problem. In this case, critical point configurations are given by the so-called centroidal Voronoi configurations [1]. However, as the number of agents goes to infinity, any configuration of points asymptotically become centroids of their Voronoi regions. Thus, those positions correspond to local optimizers of the discrete coverage control problem. In this way, while the empirical measure $\frac{1}{N}\sum_{i=1}^N \delta_{x_i}$ corresponding to the points $\{x_i\}_{i=1}^N$ samples from μ converges uniformly almost surely to μ (Glivenko-Cantelli theorem), the quantization energy $\mathcal{H}_f$, converges to zero, which does not really reflect the discrepancy between the measures μ and $\mu^\star$. Thus, the functional $\mathcal{H}_f$ suffers from this deficiency as a candidate aggregate function for coverage control in the large scale limit.

Consider instead the following aggregate objective function:

$$\bar{\mathcal{H}}_f(\mathbf{x}) = C_f\left(\frac{1}{N}\sum_{i=1}^N \delta_{x_i} \;,\; \mu^\star\right). \tag{19}$$

This performance metric has been used before in the so-called area (weight)-constrained coverage control problem [4] (the weights $w_i = 1/N$ are balanced in the case of (19)).

Lemma 11. *Let $\mu^\star \in \mathcal{P}(\Omega)$ be an absolutely continuous measure and let $\bar{\mathcal{H}}_f$ be defined as in* (19)*. Let $x_i \sim_{i.i.d} \mu$, for $i \in \{1,\dots,N\}$, where $\mu \in \mathcal{P}(\Omega)$ is any absolutely continuous probability measure. It holds almost surely that* $\lim_{N\to\infty} \bar{\mathcal{H}}_f(\mathbf{x}) = C_f(\mu, \mu^\star)$.

Proof. This can be seen from the following:

$$\begin{aligned}
\bar{\mathcal{H}}_f(\mathbf{x}) &= C_f\left(\frac{1}{N}\sum_{i=1}^N \delta_{x_i} \;,\; \mu^\star\right) \\
&= \min_{\substack{T:\Omega \to \{x_i\}_{i=1}^N \\ T_\# \mu^\star = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}}} \int_\Omega f(|x - T(x)|) d\mu^\star(x) \\
&= \min_{\substack{T:\Omega \to \{x_i\}_{i=1}^N \\ \mu^\star(T^{-1}(\{x_i\})) = \frac{1}{N}}} \int_\Omega f(|x - T(x)|) d\mu^\star(x).
\end{aligned}$$

Similar to $\mathcal{H}_f(\mathbf{x})$, the functional $\bar{\mathcal{H}}_f$ can be expressed as the sum of integrals over certain space partition. However, this case involves a generalized Voronoi partition $\{\mathcal{W}_i\}_{i=1}^N$:

$$\mathcal{W}_i = \{x \in \Omega \,|\, f(|x - x_i|) - \omega_i \leq f(|x - x_j|) - \omega_j\},$$

where $\{\omega_1, \ldots, \omega_N\}$ are chosen such that $\mu^\star(\mathcal{W}_i) = 1/N$ for all $i \in \{1, \ldots, N\}$. We refer the reader to [4] for a detailed treatment. We can now write:

$$\bar{\mathcal{H}}_f(\mathbf{x}) = \sum_{i=1}^{N} \int_{\mathcal{W}_i} f(|x - x_i|)\; d\mu^\star(x).$$

Now, by letting $x_i \sim_{i.i.d} \mu$, where $\mu \in \mathcal{P}(\Omega)$ is any absolutely continuous probability measure, in the limit $N \to \infty$, we have $\frac{1}{N}\sum_{i=1}^{N} \delta_{x_i}$ converging uniformly almost surely to μ. In this way, by the continuity of C_f, we have:

$$\lim_{N\to\infty} \bar{\mathcal{H}}_f(\mathbf{x}) = C_f\left(\mu, \mu^\star\right), \quad a.s.$$

□

Similarly to (12), we can formulate a multi-agent proximal descent algorithm on the aggregate objective function $\bar{\mathcal{H}}_f$, with $f(x) = x^2$, as follows, for every $i \in \{1, \ldots, N\}$:

$$x_i^+ = \arg\min_{z\in\Omega} \frac{1}{2\tau}|x_i - z|^2 + \bar{\mathcal{H}}_f(z, \mathbf{x}_{-i}). \tag{20}$$

Note that this is a proximal formulation of the load-balancing variant of the Lloyd's algorithm in [4].

Theorem 6 (***Convergence to generalized centroidal Voronoi configuration and $\mu^\star$***)**.** *The Lloyd proximal descent* (20)*, with $f(x) = x^2$, converges to a local minimizer of $\bar{\mathcal{H}}_f$. Furthermore, as $N \to \infty$, the proximal descent scheme* (20) *converges to:*

$$x^+ = \arg\min_{z\in\Omega} \frac{1}{2\tau}|x - z|^2 + \phi(z), \tag{21}$$

with $x \sim \mu$ and $\phi = \left.\frac{\delta W_2^2(\nu,\mu^\star)}{\delta\nu}\right|_{\mu}$, the Kantorovich potential for optimal transport from μ to $\mu^\star$. The sequence $\{\mu_k\}_{k\in\mathbb{N}}$ obtained as the transport of an absolutely continuous probability measure $\mu_0 \in \mathcal{P}(\Omega)$ by (21)*, with $x_0 \sim \mu_0$, converges weakly to $\mu^\star$ as $k \to \infty$.*

Proof. Let $\widehat{\mu}_{\mathbf{x}}^{h,N}$ be defined as in (10) with a kernel satisfying Assumption 2. We see that $C_f(\widehat{\mu}_{\mathbf{x}}^{h,N}, \mu^\star)$ as a function of $\mathbf{x}$ is α-smooth for some $\alpha > 0$ (from Proposition 4 in Appendix B and an application of Lemma 5). Further, we note that $\bar{\mathcal{H}}_f(\mathbf{x}) = \lim_{h\to 0} C_f(\widehat{\mu}_{\mathbf{x}}^{h,N}, \mu^\star)$ and the α-smoothness property carries over to the limit, as well as the comparison Lemma 8 for $\bar{\mathcal{H}}_f(\mathbf{x})$. The convergence of (20) with $f(x) = x^2$ to a local minimizer of $\bar{\mathcal{H}}_f$ then follows from a similar version of Theorem 4 applied to $\bar{\mathcal{H}}_f(\mathbf{x})$. It is easy to see that these local minima correspond to generalized centroidal Voronoi configurations as in [4].

Following a similar reasoning as in Proposition 1 for $F = C_f$ and $F^{h,N} = C_f^{h,N}$, we have that, as $N \to \infty$, the proximal descent scheme (20) converges to (21).

With $F(\nu) = \frac{1}{2}W_2^2(\nu, \mu^*)$, let $G_{\mu_k}(\nu) = \frac{1}{2\tau}W_2^2(\mu_k, \nu) + F(\nu)$. The Fréchet derivative of G_{μ_k} is given by $\nabla\left(\left.\frac{\delta G_{\mu_k}(\nu)}{\delta\nu}\right|_{\nu}\right) = \frac{1}{\tau}\nabla\phi_{\nu\to\mu_k} + \nabla\phi_{\nu\to\mu^*}$. Moreover, at the critical point μ_{k+1} of G_{μ_k} we have $\frac{1}{\tau}\nabla\phi_{\mu_{k+1}\to\mu_k} + \nabla\phi_{\mu_{k+1}\to\mu^*} = \frac{1}{\tau}\left(\mathrm{id} - T_{\mu_{k+1}\to\mu_k}\right) + \left(\mathrm{id} - T_{\mu_{k+1}\to\mu^*}\right) = 0$, which implies that $\left(T_{\mu_{k+1}\to\mu_k} - \mathrm{id}\right) = \tau\left(\mathrm{id} - T_{\mu_{k+1}\to\mu^*}\right)$. We then have $W_2(\mu_k, \mu_{k+1}) = \tau W_2(\mu_{k+1}, \mu^*)$. For any (and only) ν on the geodesic between μ_k and μ^*, we have $W_2(\mu_k, \mu^*) = W_2(\mu_k, \nu) + W_2(\nu, \mu^*)$ (wherein the triangle inequality is an equality), and this is the case if and only if $\int_\Omega \langle \mathrm{id} - T_{\nu\to\mu_k}, T_{\nu\to\mu^*} - \mathrm{id}\rangle\, d\nu = 2W_2(\mu_k, \nu)W_2(\nu, \mu^*)$. We see that this is indeed the case for $\nu = \mu_{k+1}$, from which we infer that μ_{k+1} lies on the geodesic between μ_k and μ^*. We therefore get that $\{\mu_k\}_{k\in\mathbb{N}}$ lies on the geodesic connecting μ_0 and μ^*. Now, from Proposition 3 in Appendix B it follows that $W_2^2(\cdot, \mu_k)$ is generalized geodesically convex with reference measure μ_k, and similarly $W_2^2(\cdot, \mu^*)$ is generalized geodesically convex with reference measure μ^*, the two measures μ_k and μ^* are interchangeable as reference measures along the geodesic between them. It then follows that the function G_{μ_k} is generalized geodesically convex along the geodesic between μ_0 and μ^*, with reference measure μ_k. Then, weak convergence to $\mu^\star$ of the sequence $\{\mu_k\}_{k\in\mathbb{N}}$ obtained as the transport of an absolutely continuous probability measure $\mu_0 \in \mathcal{P}(\Omega)$ by (21) follows from an application of Theorem 3 and the strict (generalized) geodesic convexity and l-smoothness of $W_2^2(\cdot, \mu^\star)$ (by an application of Propositions 3 and 4 in Appendix B). □

It is known that the generalized Lloyd's algorithm results in convergence to generalized centroidal Voronoi configurations [4], where the generators $\{x_1, \ldots, x_N\}$ of the generalized Voronoi partition are also the centroids of their respective generalized Voronoi cells. The generalized centroidal Voronoi configuration is, however, not unique, and this relates to the fact that the convergence is to the local minimizers of $\bar{\mathcal{H}}_f$, which is typically nonconvex.

We now present results from numerical experiments for the coverage control algorithm (20) for the objective function $\bar{\mathcal{H}}_f$, with $f(x) = x^2$. We first sample i.i.d. from a multimodal Gaussian distribution and normalize the histogram of the samples over a discretization of the spatial domain to obtain a (quantized) target distribution over the domain. We then implement the coverage control algorithm (20) for various sizes N of the multi-agent system, from random initializations of the agent positions. We present the following: (i) the steady state distribution of agents (in Figure 1), and (ii) the value of the coverage objective function as a function of time (in Figure 2), for various sizes N of the multi-agent system.

VI. Conclusion

In this paper, we have introduced a multiscale framework for the analysis and design of multi-agent coverage algorithms that begins with a macroscopic specification of the target coverage behavior to derive provably-correct microscopic, agent-level algorithms that achieve the target macroscopic specification. Our class of macroscopic proximal descent schemes exploit convexity properties of coverage objective functionals to steer the macroscopic configuration, which are then translated into agent-level algorithms via a variational discretization. We uncover the relationship with previously studied coverage algorithms, and obtain insights into the large-scale behavior of these algorithms. Future work will consider the extension to a constrained optimization framework to include such

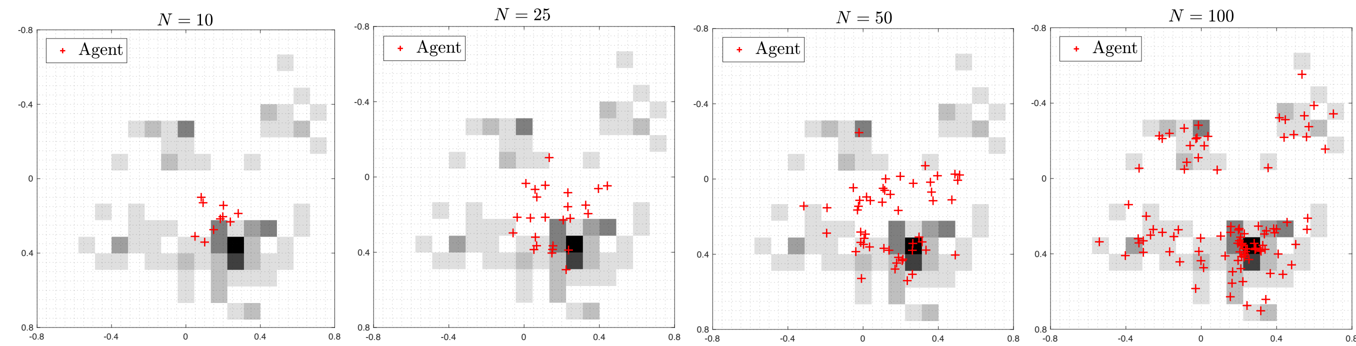

Fig. 1. The figure shows the steady state distribution of the agents implementing the coverage algorithm (20) with the target distribution depicted in grayscale, for $N = 10, 25, 50, 100$. We observe that the distribution of the agents more closely approximates the target distribution as the size N of the system increases.

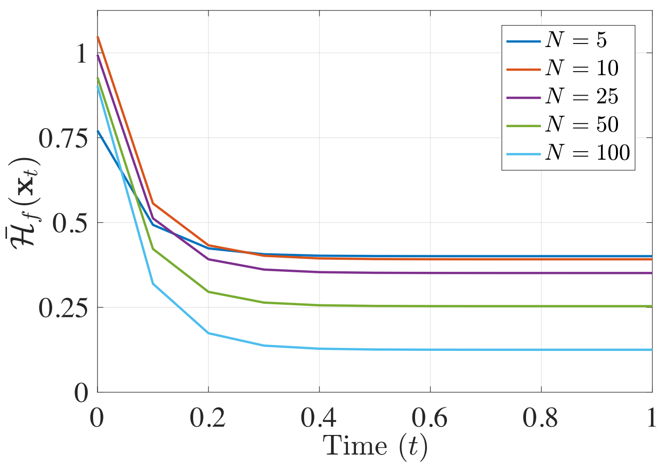

Fig. 2. The figure is a representative plot of the value of the aggregate objective function $\bar{\mathcal{H}}_f(\mathbf{x}_t)$ (with $f(x) = x^2$) vs. time t for various sizes N of the multi-agent system and random initializations of agent positions. We observe that the steady state value decreases with the size N of the system, in accordance with our theoretical results.

constraints as sensing limitations, dynamic and collision-avoidance constraints.

References

[1] 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.

[2] J. Cortés. Motion coordination algorithms resulting from classical geometric optimization problems. In K. Tas, D. Krupka, D. Baleanu, and O. Krupkova, editors, *Proceedings of the International Workshop on Global Analysis*, volume 729 of *AIP Conference Proceedings Series*, pages 54–68. American Institute of Physics, New York, 2004.

[3] J. Cortés, S. Martínez, and F. Bullo. Spatially-distributed coverage optimization and control with limited-range interactions. *ESAIM. Control, Optimisation & Calculus of Variations*, 11(4):691–719, 2005.

[4] J. Cortés. Coverage optimization and spatial load balancing by robotic sensor networks. *IEEE Transactions on Automatic Control*, 55(3):749–754, 2010.

[5] M. Zhong and C. G. Cassandras. Distributed coverage control and data collection with mobile sensor networks. *IEEE Transactions on Automatic Control*, 56(10):24452455, 2011.

[6] A. Breitenmoser, M. Schwager, J.-C. Metzger, R. Siegwart, and D. Rus. Voronoi coverage of non-convex environments with a group of networked robots. In *IEEE Int. Conf. on Robotics and Automation*, page 49824989, 2010.

[7] M. Pavone, A. Arsie, E. Frazzoli, and F. Bullo. Distributed algorithms for environment partitioning in mobile robotic networks. *IEEE Transactions on Automatic Control*, 56(8):1834–1848, 2011.

[8] Y. Ru and S. Martínez. Coverage control in constant flow environments based on a mixed energy-time metric. *Automatica*, 49(9):2632–2640, 2013.

[9] S. Bhattacharya, N. Michael, and V. Kumar. Distributed coverage and exploration in unknown non-convex environments. In *Int. Symposium on Distributed Autonomous Robotic Systems*, page 6175. Springer, 2013.

[10] 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*, page 113, 2013.

[11] N. Demir, U. Eren, and B. Acikmese. Decentralized probabilistic density control of autonomous swarms with safety constraints. *Autonomous Robots*, 39(4):537554, 2015.

[12] S. Bandyopadhyay, S. J. Chung, and F. Y. Hadaegh. Probabilistic and distributed control of a large-scale swarm of autonomous agents. *IEEE Transactions on Robotics*, 33(5):11031123, 2017.

[13] M. E. Chamie, Y. Yu, B. Açıkmeşe, and M. Ono. Controlled markov processes with safety state constraints. *IEEE Transactions on Automatic Control*, 64(3):1003–1018, 2018.

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

[15] U. Eren and B. Acikmese. Velocity field generation for density control of swarms using heat equation and smoothing kernels. In *IFAC Papers Online*, volume 50, page 94059411, 2017.

[16] T. Zheng, Q. Han, and H. Lin. PDE-based dynamic density estimation for large-scale agent systems. *IEEE Control Systems Letters*, 5(2):541–546, 2020.

[17] P. Frihauf and M. Krstic. Leader-enabled deployment onto planar curves: A PDE-based approach. *IEEE Transactions on Automatic Control*, 56(8):17911806, 2011.

[18] K. Elamvazhuthi, H. Kuiper, and S. Berman. PDE-based optimization for stochastic mapping and coverage strategies using robotic ensembles. *Automatica*, 95:356367, 2018.

[19] 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):15631578, 2018.

[20] K. Elamvazhuthi and S. Berman. Mean-field models in swarm robotics: A survey. *Bioinspiration & Biomimetics*, 15(1):015001, 2019.

[21] K. Elamvazhuthi, Z. Kakish, A. Shirsat, and S. Berman. Controllability and stabilization for herding a robotic swarm using a leader: A mean-field approach. *IEEE Transactions on Robotics*, 2020.

[22] 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, FL, USA, 2018.

[23] G. Foderaro, S. Ferrari, and T. A. Wettergren. Distributed optimal control for multi-agent trajectory optimization. *Automatica*, 50:149–154, 2014.

[24] S. Ferrari, G. Foderaro, P. Zhu, and T. A. Wettergren. Distributed optimal control of multiscale dynamical systems: a tutorial. *IEEE Control Systems*, 36(2):102116, 2016.

[25] S. Bandyopadhyay, S. J. Chung, and F. Y. Hadaegh. Probabilistic swarm guidance using optimal transport. In *IEEE Conf. on Control Applications*, page 498505, 2014.

[26] M. H. de Badyn, U. Eren, B. Açikmeşe, and M. Mesbahi. Optimal mass transport and kernel density estimation for state-dependent networked dynamic systems. In *IEEE Int. Conf. on Decision and Control*, pages 1225–1230, 2018.
[27] Q. Du, V. Faber, and M. Gunzburger. Centroidal voronoi tessellations: applications and algorithms. *SIAM Review*, 41(4):637676, 1999.
[28] D. Bourne, B. Schmitzer, and B. Wirth. Semi-discrete unbalanced optimal transport and quantization. *arXiv preprint arXiv:1808.01962*, 2018.
[29] D. Bourne and S. Roper. Centroidal power diagrams, lloyd's algorithm, and applications to optimal location problems. *SIAM Journal on Numerical Analysis*, 53(6):2545–2569, 2015.
[30] V. Hartmann and D. Schuhmacher. Semi-discrete optimal transport-the case p= 1. *arXiv preprint arXiv:1706.07650*, 2017.
[31] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer, 2008.
[32] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the fokker–planck equation. *SIAM Journal on Mathematical Analysis*, 29(1):1–17, 1998.
[33] L. Chizat and F. Bach. On the global convergence of gradient descent for over-parameterized models using optimal transport. In *Advances in Neural Information Processing Systems*, pages 3036–3046, 2018.
[34] K. Caluya and A. Halder. Proximal recursion for solving the fokker-planck equation. In *American Control Conference*, pages 4098–4103, 2019.
[35] P. Billingsley. *Probability and measure*. John Wiley, 2008.
[36] V. S. Varadarajan. On the convergence of sample probability distributions. *Sankhyā: The Indian Journal of Statistics (1933-1960)*, 19(1/2):23–26, 1958.
[37] R. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Mathematical Journal*, 80(2):309324, 1995.
[38] D. Henry. *Geometric theory of semilinear parabolic equations*. Springer, 1981.
[39] J.A. Walker. *Dynamical Systems and Evolution Equations: Theory and Applications*, volume 20. Springer, 2013.
[40] J. A. Walker. Some results on Liapunov functions and generated dynamical systems. *Journal of Differential Equations*, 30(3):424440, 1978.
[41] P. Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.
[42] F. Santambrogio. *Optimal transport for applied mathematicians*. Springer, 2015.
[43] C. Villani. *Optimal transport: old and new*, volume 338. Springer, 2008.

APPENDIX A
MATHEMATICAL PRELIMINARIES

We present here the mathematical preliminaries on convergence of measures, the L^2-Wasserstein space and smoothness and convexity notions for functions defined on the L^2-Wasserstein space.

A. Weak convergence of measures

The results of this manuscript rely on the notions of weak convergence in $\mathcal{P}(\Omega)$, the topology of weak convergence, its metrizability, and the compactness of sets of $\mathcal{P}(\Omega)$. We recall them here and refer the reader to [41] for more information.

Definition 2 (***Weak convergence***)**.** *Let $\Omega \subseteq \mathbb{R}^d$, and $\mathcal{P}(\Omega)$ be its set of probability measures. A sequence $\{\mu_k\}_{k\in\mathbb{N}} \subseteq \mathcal{P}(\Omega)$* converges weakly *to $\mu \in \mathcal{P}(\Omega)$ if for any bounded and continuous function f on Ω, $\lim_{k\to\infty} \int_\Omega f d\mu_k = \int_\Omega f d\mu$.*

Equivalently, in the definition above, the sequence $\{\mu_k\}_{k\in\mathbb{N}}$ in $\mathcal{P}(\Omega)$ is said to *converge to μ in $\mathcal{P}(\Omega)$ equipped with the topology of weak convergence*. The space of probability measures $\mathcal{P}(\Omega)$ equipped with the topology of weak convergence is *metrizable* [41]. In other words, there exists a metric on $\mathcal{P}(\Omega)$ such that the topology of weak convergence is obtained as the topology induced by the metric. One such metric is the Wasserstein distance, see Section A-B. We now state Prokhorov's theorem [41] on the equivalence between tightness and precompactness of a collection of probability measures over a separable and complete metric (Polish) space.

Lemma 12 (***Prokhorov's theorem***)**.** *Let Ω be a complete metric space, and let $\mathcal{K} \subseteq \mathcal{P}(\Omega)$. The closure of $\mathcal{K}$ w.r.t. the topology of weak convergence in $\mathcal{P}(\Omega)$ is* compact *if and only if $\mathcal{K}$ is tight. That is, $\mathcal{K}$ is* tight *if for any $\epsilon > 0$ there exists a compact $K_\epsilon \subseteq \Omega$ such that $\mu(K_\epsilon) > 1 - \epsilon$, for all $\mu \in \mathcal{K}$.*

Corollary 3 (***Compactness of*** $\mathcal{P}(\Omega)$)**.** *Let $\Omega \subseteq \mathbb{R}^d$ a compact set. Then, the closure of $\mathcal{P}(\Omega)$ w.r.t. the topology of weak convergence in $\mathcal{P}(\Omega)$ is compact. This follows from Prokhorov's theorem in Lemma 12, since $\mathcal{P}(\Omega)$ is tight: for any $\epsilon > 0$, we choose Ω itself as the compact set and have $\mu(\Omega) = 1 > 1-\epsilon$ for any $\mu \in \mathcal{P}(\Omega)$. Moreover, since $\mathcal{P}(\Omega)$ is also closed w.r.t. the topology of weak convergence, it is therefore compact.*

B. The L^2-Wasserstein distance

The L^2-Wasserstein distance between two probability measures $\mu, \nu \in \mathcal{P}(\Omega)$ is given by:

$$W_2^2(\mu,\nu) = \min_{\pi\in\Pi(\mu,\nu)} \int_{\Omega\times\Omega} |x-y|^2 \, d\pi(x,y), \tag{22}$$

where $\Pi(\mu,\nu)$ is the space of joint probability measures over $\Omega \times \Omega$ with marginals μ and ν. The definition of L^2-Wasserstein distance in (22) follows from the so-called Kantorovich formulation of optimal transport. An alternative formulation of this problem, called the Monge formulation of optimal transport, is given below:

$$W_2^2(\mu,\nu) = \min_{\substack{T:\Omega\to\Omega \\ T_{\#}\mu=\nu}} \int_{\Omega} |x-T(x)|^2 \, d\mu(x). \tag{23}$$

In the Monge formulation (23), the minimization is carried out over the space of maps $T : \Omega \to \Omega$ for which the probability measure ν is obtained as the pushforward of μ. This can be viewed as a deterministic formulation of optimal transport, where the transport is carried out by a map, whereas the Kantorovich formulation (22) can be seen as a problem relaxation, where the transport plan is described by a joint probability measure π over $\Omega \times \Omega$, with μ and ν as its marginals. It is to be noted that the Monge formulation does not always admit a solution, while the Kantorovich problem does. Roughly speaking, the Kantorovich formulation is the "minimal" extension of the Monge formulation, as both problems attain the same infimum [42]. Further, the two formulations (22) and (23) are equivalent under certain conditions, and in the sense laid out in the ensuing lemma.

Lemma 13 (***Monge-Kantorovich optimal transport*, cf. [42], Theorem 1.17 for** $c(x,y) = |x-y|^2$)**.** *Assume that Ω is compact in $\mathbb{R}^d$. There exists a minimizer π^* to the Kantorovich problem* (22)*. Moreover, if the measure μ is atomless, and $\mu(\partial\Omega) = 0$, then the minimizer π^* is unique, the Monge formulation* (23) *admits a unique minimizer T^*, and it holds that $\pi^* = (\mathrm{id}, T^*)_{\#}\mu$, with* id $: \Omega \to \Omega$ *the identity mapping. Furthermore, there exists a Lipschitz continuous* *function $\phi : \Omega \to \mathbb{R}$, called the Kantorovich potential, such that $\nabla\phi = \mathrm{id} - T^*$.*

The space of probability measures $\mathcal{P}(\Omega)$ endowed with the L^2-Wasserstein distance W_2 will equivalently be referred to as the L^2-Wasserstein space $(\mathcal{P}(\Omega), W_2)$ over Ω. The following lemma, which follows from Theorem 6.9 in [43], establishes the equivalence between convergence in the sense of the topology of weak convergence and in the L^2-Wasserstein metric.

Lemma 14 (***Convergence in*** $(\mathcal{P}(\Omega), W_2)$**).** *For compact $\Omega \subset \mathbb{R}^d$, the L^2-Wasserstein distance W_2 metrizes the weak convergence in $\mathcal{P}(\Omega)$. That is, a sequence of measures $\{\mu_k\}_{k\in\mathbb{N}}$ in $\mathcal{P}(\Omega)$ converges weakly to $\mu \in \mathcal{P}(\Omega)$ if and only if $\lim_{k\to\infty} W_2(\mu_k, \mu) = 0$.*

C. Derivatives of functionals on atomless measures

We start by introducing the notion of first variation of a functional on $\mathcal{P}(\Omega)$ as follows:

Definition 3 (***First variation of a functional on*** $\mathcal{P}(\Omega)$**).** *Let $F : \mathcal{P}(\Omega) \to \mathbb{R}$, $\mu_0 \in \mathcal{P}(\Omega)$ and let $\{\mu_\epsilon\}_{\epsilon\in\mathbb{R}}$ be a smooth one-parameter family of probability measures. Suppose that there exists a unique $\frac{\delta F}{\delta\mu}(\mu_0)$ such that $\frac{d}{d\epsilon}F(\mu_\epsilon)\big|_{\epsilon=0} = \lim_{\epsilon\to 0}\frac{1}{\epsilon}\int_\Omega \frac{\delta F}{\delta\mu}(\mu_0)\,(d\mu_\epsilon - d\mu_0)$ for any smooth $\{\mu_\epsilon\}_{\epsilon\in\mathbb{R}}$. Then, $\frac{\delta F}{\delta\mu}(\mu_0)$ is the first variation of F evaluated at μ_0.*

For functionals for which the first variation exists as in the above definition, we can introduce the notion of Fréchet derivative on the L^2-Wasserstein space $(\mathcal{P}(\Omega), W_2)$:

Definition 4 (***Derivative of a functional on*** $(\mathcal{P}(\Omega), W_2)$**).** *A functional $F : \mathcal{P}(\Omega) \to \mathbb{R}$ is* Fréchet differentiable *with derivative ξ at an atomless measure $\mu_0 \in \mathcal{P}(\Omega)$, if for any smooth one-parameter family of probability measures $\{\mu_\epsilon\}_{\epsilon\in\mathbb{R}}$, the following limit exists:*

$$\lim_{\epsilon\to 0}\frac{F(\mu_\epsilon) - F(\mu_0) - \int_\Omega \langle \xi, T_{\mu_0\to\mu_\epsilon} - \mathrm{id}\rangle\, d\mu}{W_2(\mu_0, \mu_\epsilon)} = 0,$$

where $\xi = \nabla\varphi$, $\varphi = \frac{\delta F}{\delta\bar{\mu}}(\mu_0)$ and $T_{\mu_0\to\mu_\epsilon}$ is the optimal transport map from μ_0 to μ_ϵ.

We now introduce the notion of directional derivative of a functional over probability measures. For this, let $\mathbf{v} = \frac{1}{t}(T_{\mu\to\nu} - \mathrm{id})$, which implies that $\nu = (\mathrm{id} + t\mathbf{v})_\#\mu$. We have:

$$W_2(\mu,\nu) = \sqrt{\int_\Omega |T_{\mu\to\nu} - \mathrm{id}|^2\, d\mu} = t\sqrt{\int_\Omega |\mathbf{v}|^2 d\mu},$$

and we get:

$$\lim_{t\to 0}\frac{F((\mathrm{id} + t\mathbf{v})_\#\mu) - F(\mu) - t\int_\Omega \langle \xi, \mathbf{v}\rangle\, d\mu}{t\sqrt{\int_\Omega |\mathbf{v}|^2 d\mu}} = 0.$$

Therefore, the directional derivative of F along $\mathbf{v}$ is

$$\frac{d}{dt}\bigg|_{\mathbf{v}} F(\mu) = \lim_{t\to 0}\frac{F((\mathrm{id} + t\mathbf{v})_\#\mu) - F(\mu)}{t} = \int_\Omega \langle \xi, \mathbf{v}\rangle\, d\mu,$$

where ξ is the Fréchet derivative of F evaluated at μ.

1) Lipschitz-continuous derivatives: We now introduce the notion of l-smoothness that will be useful for the development of gradient descent-based transport schemes in the paper.

Definition 5 (l***-smoothness of functionals on*** $(\mathcal{P}(\Omega), W_2)$**).** *A functional $F : \mathcal{P}(\Omega) \to \mathbb{R}$ is called l-*smooth *(or* Lipschitz differentiable*) if for any $\mu, \nu \in \mathcal{P}(\Omega)$, we have:*

$$\sqrt{\int_\Omega |\xi_\mu - \xi_\nu|^2\, d\nu} \le l W_2(\mu,\nu),$$

where ξ_μ, ξ_ν are the Fréchet derivatives of F evaluated at μ and ν respectively.

From the above definition on l-smooth functionals, the following lemma can be easily verfied:

Lemma 15 (l***-smooth functionals***)**.** *A functional $F : \mathcal{P}(\Omega) \to \mathbb{R}$ that is l-smooth on $(\mathcal{P}(\Omega), W_2)$ satisfies:*

1) $\left|F(\nu) - F(\mu) - \int_\Omega \langle \xi_\mu, T_{\mu\to\nu} - \mathrm{id}\rangle\, d\mu\right| \le \frac{l}{2}W_2^2(\mu,\nu)$,
2) $\left|\int_\Omega \langle \xi_\mu - \xi_\nu, T_{\nu\to\mu} - \mathrm{id}\rangle\, d\nu\right| \le l W_2^2(\mu,\nu)$,

for any two atomless probability measures $\mu, \nu \in \mathcal{P}(\Omega)$, where ξ_μ and ξ_ν are the Fréchet derivatives of F evaluated at μ and ν respectively.

D. Convexity of functionals on the Wasserstein space

Results in convex analysis can be appropriately generalized to functionals on the L^2-Wasserstein space $(\mathcal{P}(\Omega), W_2)$, see [31] for a detailed treatment. In this section, we introduce and define notions related to the convexity of functionals on $(\mathcal{P}(\Omega), W_2)$ used to build the results in this paper. Before we can define any notion of convexity, we introduce an appropriate notion of interpolation:

Definition 6 (***Generalized displacement interpolation***)**.** *Let Ω be a compact subset of $\mathbb{R}^d$, $\mu, \nu \in \mathcal{P}(\Omega)$, and $\theta \in \mathcal{P}^r(\Omega)$ be an atomless probability measure. Let $T_{\theta\to\mu} : \Omega \to \Omega$ and $T_{\theta\to\nu} : \Omega \to \Omega$ are optimal transport maps from θ to μ, and θ to ν resp. in the L^2-Wasserstein space over Ω. A* (generalized) displacement interpolant *of μ and ν* w.r.t. *θ is given by $\gamma_t = ((1-t)T_{\theta\to\mu} + tT_{\theta\to\nu})_\#\,\theta$, for $t \in [0,1]$.*

It can be shown that for a compact and convex $\Omega \subset \mathbb{R}^d$, the space of probability measures $\mathcal{P}(\Omega)$ is geodesically convex w.r.t. the notion of (generalized) displacement interpolation in Definition 6.

Lemma 16 (***Geodesic convexity of*** $\mathcal{P}(\Omega)$**).** *Let $\Omega \subseteq \mathbb{R}^d$ be a compact, convex set. Then, the L^2-Wasserstein space $(\mathcal{P}(\Omega), W_2)$ is geodesically convex w.r.t. the notion of interpolations as in Definition 6.*

Now, we introduce the following standard definition on the (generalized) geodesic convexity of functionals on the L^2-Wasserstein space $(\mathcal{P}(\Omega), W_2)$:

Definition 7 (***Generalized geodesic convexity***)**.** *Let $\Omega \subseteq \mathbb{R}^d$ be a compact and convex set, and let $\mu, \nu \in \mathcal{P}(\Omega)$ and $\theta \in \mathcal{P}(\Omega)$ be an atomless probability measure, for which there exist $T_{\theta\to\mu} : \Omega \to \Omega$ and $T_{\theta\to\nu} : \Omega \to \Omega$ optimal transport maps from θ to μ and from θ to ν respectively, in the L^2-Wasserstein* *space over Ω. A functional $F : \mathcal{P}(\Omega) \to \mathbb{R}$ is* (generalized) geodesically convex *(resp.* (generalized) strictly geodesically convex*) if the following holds for every $t \in [0, 1]$:*

$$F(((1-t)T_{\theta\to\mu} + tT_{\theta\to\nu})_{\#}\,\theta) \leq (1-t)F(\mu) + tF(\nu).$$

(resp. the previous inequality holds with strict inequality).

Lemma 17 (***First-order convexity condition***)**.** *Let $\Omega \subseteq \mathbb{R}^d$ be compact and convex, $\mu, \nu, \theta \in \mathcal{P}(\Omega)$ be atomless probability measures. Let $F : \mathcal{P}(\Omega) \to \mathbb{R}$ be a Fréchet differentiable and (generalized) geodesically convex functional (in the sense of Definition 7). Then, we have:*

$$F(\nu) \geq F(\mu) + \int_\Omega \langle \xi_\mu, T_{\theta\to\nu} - T_{\theta\to\mu}\rangle \, d\theta, \qquad (24)$$

where ξ_μ is the Fréchet derivative of F at μ, and $T_{\theta\to\mu} : \Omega \to \Omega$ and $T_{\theta\to\nu} : \Omega \to \Omega$ are optimal transport maps from θ to μ and from θ to ν respectively.

We now define below the notion of strong geodesic convexity of Fréchet-differentiable functionals on $(\mathcal{P}(\Omega), W_2)$:

Definition 8 (***Strong geodesic convexity of a functional on*** $(\mathcal{P}(\Omega), W_2)$)**.** *Let $\Omega \subseteq \mathbb{R}^d$ be compact and convex and $\mu, \nu, \theta \in \mathcal{P}(\Omega)$ be atomless probability measures. Let $F : \mathcal{P}(\Omega) \to \mathbb{R}$ be a Frechét-differentiable functional. Let ξ_μ and ξ_ν be the Fréchet derivatives of F evaluated at measures μ and ν, respectively. Then, F is* strongly (geodesically) convex *if there exists an $m > 0$ such that:*

$$\int_\Omega \langle \xi_\nu - \xi_\mu, T_{\theta\to\nu} - T_{\theta\to\mu}\rangle \, d\theta \geq m W_2^2(\mu, \nu), \qquad (25)$$

where $T_{\theta\to\mu} : \Omega \to \Omega$ and $T_{\theta\to\nu} : \Omega \to \Omega$ are optimal transport maps from θ to μ and from θ to ν respectively.

APPENDIX B
AGGREGATE OBJECTIVE FUNCTIONS

Proposition 3 (***Strict geodesic convexity of*** $C_f(\cdot, \mu^*)$)**.** *Fix $\mu^* \in \mathcal{P}(\Omega)$ (absolutely continuous) as the reference measure and let $\mu_0, \mu_1 \in \mathcal{P}(\Omega)$. Let $T_{\mu^*\to\mu_0}$ and $T_{\mu^*\to\mu_1}$ be optimal transport maps from μ^* to μ_0 and μ^* to μ_1 respectively, corresponding to the optimal transport cost C_f, and let $T_t = (1-t)T_{\mu^*\to\mu_0} + tT_{\mu^*\to\mu_1}$ for $t \in [0, 1]$. For $\mu_t = T_{t\#}\mu^*$, we have:*

$$C_f(\mu_t, \mu^*) < (1-t)C_f(\mu_0, \mu^*) + tC_f(\mu_0, \mu^*).$$

Proof. We have:

$$\begin{aligned}
C_f(\mu_t, \mu^*) &\leq \int_\Omega f(|T_t(x) - x|) d\mu^*(x) \\
&= \int_\Omega f\left(|(1-t)T_{\mu^*\to\mu_0}(x) + tT_{\mu^*\to\mu_1}(x) - x|\right) d\mu^*(x) \\
&= \int_\Omega f\left(|(1-t)\left[T_{\mu^*\to\mu_0}(x) - x\right] \right. \\
&\qquad \left. + t\left[T_{\mu^*\to\mu_1}(x) - x\right]|\right) d\mu^*(x) \\
&\leq \int_\Omega f\left((1-t)\,|T_{\mu^*\to\mu_0}(x) - x| \right. \\
&\qquad \left. + t\,|T_{\mu^*\to\mu_1}(x) - x|\right) d\mu^*(x),
\end{aligned}$$

where the last inequality is a consequence of the fact that f is non-decreasing. Further, if f is strictly convex in Ω, we have:

$$\begin{aligned}
&C_f(\mu_t, \mu^*) \\
&< \int_\Omega \left[(1-t) f\left(|T_{\mu^*\to\mu_0}(x) - x|\right) \right. \\
&\qquad \left. + t f\left(|T_{\mu^*\to\mu_1}(x) - x|\right)\right] d\mu^*(x) \\
&= (1-t) \int_\Omega f\left(|T_{\mu^*\to\mu_0}(x) - x|\right) d\mu^*(x) \\
&\quad + t \int_\Omega f\left(|T_{\mu^*\to\mu_1}(x) - x|\right) d\mu^*(x) \\
&= (1-t)C_f(\mu_0, \mu^*) + tC_f(\mu_0, \mu^*).
\end{aligned}$$

□

We now establish the following result:

Proposition 4 (l***-smoothness of*** $C_f(\cdot, \mu^*)$)**.** *Let the Fréchet derivative of the functional $F(\mu) = C_f(\mu, \mu^*)$ at $\mu \in \mathcal{P}^r(\Omega)$ be denoted as ξ_μ. The functional $F(\mu) = C_f(\mu, \mu^*)$ satisfies:*

$$\left|\int_\Omega \langle \xi_{\mu_1} - \xi_{\mu_2}, T_{\mu_2\to\mu_1} - \mathrm{id}\rangle \, d\mu_2\right| \leq l \int_\Omega |T_{\mu_2\to\mu_1} - \mathrm{id}|^2 \, d\mu_2,$$

where $T_{\mu_2\to\mu_1}$ is the optimal transport map from μ_2 to μ_1 w.r.t. C_f.

Proof. Let $\phi_\mu = \frac{\delta C_f(\mu,\mu^*)}{\delta\mu}$ be the Kantorovich potential for the optimal transport from μ to μ^*. We now have the following relation [42]:

$$T_{\mu\to\mu^*} = \mathrm{id} - (\nabla h)^{-1}(\nabla\phi_\mu),$$

where the function $h : \mathbb{R}^d \to \mathbb{R}$ is such that $h(\mathbf{v}) = f(|\mathbf{v}|)$. It follows from the l-smoothness of f that the function h is also l-smooth. From the above and l-smoothness of h, we get:

$$\begin{aligned}
&\left|\int_\Omega \langle \xi_{\mu_1} - \xi_{\mu_2}, T_{\mu_2\to\mu_1} - \mathrm{id}\rangle \, d\mu_2\right| \\
&= \left|\int_\Omega \langle \nabla h\left(\mathrm{id} - T_{\mu_1\to\mu^*}\right) - \nabla h\left(\mathrm{id} - T_{\mu_2\to\mu^*}\right), T_{\mu_2\to\mu_1} - \mathrm{id}\rangle \, d\mu_2\right| \\
&\leq \int_\Omega |\langle \nabla h\left(\mathrm{id} - T_{\mu_1\to\mu^*}\right) - \nabla h\left(\mathrm{id} - T_{\mu_2\to\mu^*}\right), T_{\mu_2\to\mu_1} - \mathrm{id}\rangle| \, d\mu_2 \\
&\leq l \int_\Omega |T_{\mu_2\to\mu_1} - \mathrm{id}|^2 \, d\mu_2.
\end{aligned}$$

□