

A BLOB METHOD FOR MEAN FIELD CONTROL WITH TERMINAL CONSTRAINTS*

KATY CRAIG[†], KARTHIK ELAMVAZHUTHI[‡], AND HARLIN LEE[§]

Abstract. In the present work, we develop a novel particle method for a general class of mean field control problems, with source and terminal constraints. Specific examples of the problems we consider include the dynamic formulation of the p -Wasserstein metric, optimal transport around an obstacle, and measure transport subject to acceleration controls. Unlike existing numerical approaches, our particle method is meshfree and does not require global knowledge of an underlying cost function or of the terminal constraint. A key feature of our approach is a novel way of enforcing the terminal constraint via a soft, nonlocal approximation, inspired by recent work on blob methods for diffusion equations. We prove convergence of our particle approximation to solutions of the continuum mean-field control problem in the sense of Γ -convergence. A byproduct of our result is an extension of existing discrete-to-continuum convergence results for mean field control problems to more general state and measure costs, as arise when modeling transport around obstacles, and more general constraint sets, including controllable linear time invariant systems. Finally, we conclude by implementing our method numerically and using it to compute solutions the example problems discussed above. We conduct a detailed numerical investigation of the convergence properties of our method, as well as its behavior in sampling applications and for approximation of optimal transport maps.

Key words. optimal transport, mean field optimal control, particle methods, measure transport

MSC codes. 35Q35, 35Q62, 35Q82, 65M12, 82C22, 93A16, 49M41, 49N80

1. Introduction. The goal of the present paper is to develop a particle method for solving the following mean-field control problem:

$$\begin{aligned}
 (\star) \quad & \min_{(\mu, \mathbf{u}) \in \mathcal{A}(m_0, m_1)} \int_0^1 \int_{\mathbb{R}^d} \psi(\mathbf{u}(y, t)) d\mu_t(y) dt + \int_0^1 \int_{\mathbb{R}^d} L(y, \mu_t) d\mu_t(y) dt \\
 & \mathcal{A}(m_0, m_1) := \left\{ (\mu, \mathbf{u}) : \mu \in AC([0, 1]; \mathcal{P}_1(\mathbb{R}^d)), \mathbf{u} \in L^1_{d\mu_t \otimes dt}(\mathbb{R}^d \times [0, 1]; U) \right. \\
 & \quad \left. \partial_t \mu_t + \nabla \cdot (\mathbf{F}(\cdot, \mu_t) \mu_t + \mathbf{u}(\cdot, t) \mu_t) = 0, \mu_0 = m_0, \mu_1 = m_1 \right\}.
 \end{aligned}$$

In particular, we seek an evolving probability measure μ and a control \mathbf{u} , constrained to take values in a vector space U , so that (i) μ evolves from an initial measure m_0 toward a terminal measure m_1 , (ii) (μ, \mathbf{u}) satisfy the continuity equation constraint, subject to additional affects from a measure dependent vector field \mathbf{F} , and (iii) (μ, \mathbf{u}) accomplish this while minimizing the integrals of the control cost ψ and the state and measure cost L . See section 2.1 for precise definitions of the relevant function spaces and notion of solution.

Problems of the form (\star) arise in a variety of contexts, including statistical mechanics, biology, economics, and control theory; see [32, 16, 17, 20, 31, 39, 41, 42, 15, 12, 21, 35, 13] and the references therein. More recently, problems of this form have also attracted interest in the machine learning community, in the context of sampling and generative modeling [3, 36]. The main contribution of the present work is a new method for computing optimal solutions of (\star) via a particle approximation: see equation $(\text{MFC}_{\varepsilon, \delta, N})$ and its numerical implementation in section 4. We prove convergence of this approximation, in the sense of Γ -convergence, and

*Submitted to the editors DATE.

Funding: K. Craig's work was supported by NSF DMS grant 2145900. K. Elamvazhuthi's work was supported by AFOSR MURI FA9550-18-1-0502. H. Lee's work was partially supported by grants NSF DMS-1952339 and TWCF0333 from the Templeton World Charity Foundation.

[†]Department of Mathematics, University of California, Santa Barbara (kcraig@ucsb.edu). Corresponding author.

[‡]Department of Mechanical Engineering, University of California, Riverside (karthike@ucr.edu).

[§]School of Data Science and Society, University of North Carolina at Chapel Hill (harlin@unc.edu). This work was partially completed while the author was at Department of Mathematics, University of California, Los Angeles.

then implement the method numerically to compute several fundamental examples of (\star) , including the dynamic formulation of the p -Wasserstein distance, optimal transport around obstacles, and measure transport subject to acceleration controls.

Traditionally, numerical methods for solving mean field control problems with terminal constraints have relied on an *Eulerian* grid-based approach, such as that proposed in Benamou and Brenier’s original work on the dynamic formulation of the 2-Wasserstein distance [11]. While this approach has been valuable, especially in lower-dimensional contexts [1, 26, 27], it presents a notable limitation in high dimensions. In contrast, our particle method is inherently meshfree, thus feasible in high dimensions.

A second difference between our method and classical methods is that our approach computes the optimal trajectory from m_0 to m_1 by updating the location of each individual particle using only *local information* from m_1 near that particle. This is in contrast to classical grid-based approaches, for which the terminal constraint m_1 must be known globally at all grid points. As a consequence of this, we believe our approach has promise as an optimal transport based sampling method, interpreting the final particle locations at time $t = 1$ as the samples of m_1 . Furthermore, as is evident in our numerical experiments (see section 4.4), the final particle locations obtained by our method exhibit more structure than traditional iid samples, which can be advantageous when using the samples to approximate integrals of smooth functions against m_1 .

Our approach is also different from existing methods for solving (\star) based on a *Lagrangian* transport map approach, in which one reformulates the problem as a Monge problem and the solution is given by an optimal transport map [2, 34, 22]. Classical methods for computing such a map demand an explicit, closed-form expression for the cost function, which is unavailable in many scenarios, such as optimal transport around obstacles. Furthermore, even when an optimal transport map is found, further computation is required to infer the optimal trajectories. In contrast, our particle approach does not require global knowledge of an underlying cost function, and it directly outputs approximate optimal trajectories. Moreover, in cases where the exact optimal transport map possesses strong monotonicity properties, such as the dynamic formulation of the p -Wasserstein distance, interpolating between these particle trajectories can then give a numerical approximation of the optimal transport map. (See section 4.8 for an error analysis of this approximation in the case $p = 2$.)

Our approach is strongly inspired by recent work by Fornasier, Lisini, Orrieri, and Savaré [30], which studied the convergence of finite agent controls to (\star) , in the mean field limit. In this work, the authors analyze when solutions of a mean field control problem *without* a terminal constraint, $\mu_1 = m_1$, may be approximated by particle solutions of a spatially discrete optimization problem. The heuristic idea behind this approach can be seen as follows: given an approximation of m_0 by an empirical measure, $m_0 \approx (1/N) \sum_{i=1}^N \delta_{y_i}$, $\{y_i\}_{i=1}^N \subseteq \mathbb{R}^d$, if the velocity field $\mathbf{v} = \mathbf{F} + \mathbf{u}$ in the continuity equation constraint is sufficiently regular, the continuity equation constraint reduces to a system of ordinary differential equations for the trajectories of the particles, $\dot{y}_i(t) = \mathbf{v}(y_i(t), t)$, where $\mu_t = (1/N) \sum_{i=1}^N \delta_{y_i(t)}$, and the spatial integrals in the objective function reduce to finite sums.

The main strategy of the present paper is to extend the approach developed by Fornasier et. al. [30] to the setting of mean field problems with terminal constraints (\star) , while at the same time generalizing the hypotheses on the state and measure cost L and the constraint set U . Indeed, a byproduct of our result is that we prove convergence of the discrete to the continuum problems under these weaker assumptions, which is new even in the absence of a terminal constraint. On the other hand, our extension to control problems with a terminal constraint requires a novel approach, via a soft, nonlocal approximation of the constraint. This is inspired by classical vortex blob methods for the Euler and Navier Stokes equations [9, 10]

and the aggregation equation [23], which have more recently been extended to the case of diffusion equations [18, 24, 25, 19, 37, 38, 14].

With this strategy in mind, we consider solutions of (\star) under the following assumptions on the source and target distributions m_0 and m_1 , the range of the control map U , the state and measure cost L , the control cost ψ , and the measure dependent vector field \mathbf{F} :

ASSUMPTION 1.1. *We suppose that the following hold:*

- (i) $m_0, m_1 \in \mathcal{P}_1(\mathbb{R}^d)$ with $dm_1(x) = m_1(x)dx$ for $m_1 \in L^2(\mathbb{R}^d)$.
- (ii) $U \subseteq \mathbb{R}^d$ is a subspace.
- (iii) $L : \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \rightarrow [0, +\infty]$ satisfies one of the following:
 - (a) L is jointly uniformly continuous and $\{L = +\infty\} = \emptyset$;
 - (b) L is independent of the second variable, i.e., $L(y, \mu) = L(y, \nu)$, for all $y \in \mathbb{R}^d$, $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$, $\{L = +\infty\}$ is open, and L is continuous on $\{L < +\infty\}$.
- (iv) $\psi : U \rightarrow [0, +\infty) : u \mapsto \frac{1}{p}|u|^p$, for $p > 1$,
or, more generally, ψ may be any function satisfying Assumption 2.1 below;
- (v) $\mathbf{F} : \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{R}^d$ is jointly uniformly continuous and there exist constants $C_F, C'_F \geq 0$ so that

$$|\mathbf{F}(y, \mu)| \leq C_F + C'_F \left(|y| + \int_{\mathbb{R}^d} |z| d\mu(z) \right), \quad \forall y \in \mathbb{R}^d, \mu \in \mathcal{P}_1(\mathbb{R}^d).$$

The main distinction between the above hypotheses and previous work by Fornasier, et. al. [30] is item (iiib), which is of interest when the state and measure cost L is used to enforce an obstacle; see section 1.2 below. We also note that we commit a mild notational abuse in part (i) above: if a measure m is absolutely continuous with respect to Lebesgue measure, we use m to denote both the measure and its density with respect to Lebesgue. See also Remark 4.1 below, for an approach to relaxing the assumption $m_1 \in L^2(\mathbb{R}^d)$ in the context of numerical simulation.

Of particular interest in the present paper, and the cases for which we obtain the strongest convergence results, are the cases of an unconstrained control $U = \mathbb{R}^d$ and the case of a controllable linear time invariant system, $L = 0$, $F(y, \mu) = Ay$, for which the system is controllable between any two states and $\text{supp } m_0 \subseteq N(A)$; see Theorem 1.9. While the former was considered by Fornasier, et. al. [30], to our knowledge the latter is new in context of particle approximation of mean field control, with or without a terminal constraint. A key motivation underlying both of these hypotheses is that they ensure the discrete optimization problem at the particle level is *feasible*, in the sense that there exists an element of the constraint set for which the objective function is finite.

The remainder of the introduction is organized as follows. In Section 1.1, we describe our particle approximation and state our main results. In Section 1.2, we describe three key examples of mean field control problems that motivate our Assumption 1.1 above. In Section 1.3, we outline the strategy of our approach.

1.1. Main results. In order to state our main results, we begin by introducing an equivalent formulation of our mean-field control problem (\star) in momentum coordinates. This is motivated by the fact that, in the original formulation, \mathbf{u} belongs to a function space that depends on μ , and we seek to remove this dependence.

Define the control cost functional

$$\Psi : \mathcal{M}(\mathbb{R}^d \times [0, 1]; U) \times C([0, 1]; \mathcal{P}_1(\mathbb{R}^d)) \rightarrow [0, +\infty],$$

$$\Psi(\nu | \mu) := \begin{cases} \int_0^1 \int_{\mathbb{R}^d} \psi(\mathbf{u}(y, t)) d\mu_t(y) dt & \text{if } d\nu(y, t) = \mathbf{u}(y, t) d\mu_t(y) dt, \\ +\infty & \text{otherwise.} \end{cases}$$

Likewise, define the penalization for the terminal constraint, $\mathcal{F} : \mathcal{P}_1(\mathbb{R}^d) \rightarrow [0, +\infty]$,

$$(1.1) \quad \mathcal{F}(\mu) := \begin{cases} 0 & \text{if } \mu = m_1, \\ +\infty & \text{otherwise.} \end{cases}$$

The original problem (\star) is equivalent to the following mean field control problem in momentum coordinates

$$(MFC) \quad \min_{(\mu, \mathbf{v}) \in C(m_0)} \mathcal{E}(\mu, \mathbf{v}), \quad \mathcal{E}(\mu, \mathbf{v}) := \mathcal{G}(\mu, \mathbf{v}) + \mathcal{F}(\mu_1),$$

$$\mathcal{G}(\mu, \mathbf{v}) := \Psi(\mathbf{v}|\mu) + \int_0^1 \int_{\mathbb{R}^d} L(y, \mu_t) d\mu_t(y) dt,$$

$$C(m_0) := \{(\mu, \mathbf{v}) \in AC([0, 1]; \mathcal{P}_1(\mathbb{R}^d)) \times \mathcal{M}(\mathbb{R}^d \times [0, 1]; U) : \\ \partial_t \mu_t + \nabla \cdot (\mathbf{F}(\cdot, \mu_t) \mu_t + \mathbf{v}_t) = 0, \mu_0 = m_0\}.$$

See Lemma 2.3 below for a proof of this equivalence.

With this formulation of the problem in hand, we now describe our approach to softening the terminal constraint and constructing the particle approximation. The first step is to replace the characteristic function \mathcal{F} with the soft L^2 penalty:

$$(1.2) \quad \mathcal{F}_\varepsilon(\mu) := \begin{cases} \frac{1}{\varepsilon} \int_{\mathbb{R}^d} |\mu(y) - m_1(y)|^2 dy & \text{if } d\mu(y) = \mu(y)dy \text{ and } \mu \in L^2(\mathbb{R}^d), \\ +\infty, & \text{otherwise.} \end{cases}$$

This leads to the optimization problem

$$(MFC_\varepsilon) \quad \min_{(\mu, \mathbf{v}) \in C(m_0)} \mathcal{E}_\varepsilon(\mu, \mathbf{v}), \quad \mathcal{E}_\varepsilon(\mu, \mathbf{v}) := \mathcal{G}(\mu, \mathbf{v}) + \mathcal{F}_\varepsilon(\mu_1).$$

We have the following convergence result for solutions of (MFC_ε) to (MFC) as $\varepsilon \rightarrow 0$.

PROPOSITION 1.2 (Convergence as $\varepsilon \rightarrow 0$). *Suppose Assumption 1.1 holds and (MFC) is feasible. Suppose that $(\mu_\varepsilon, \mathbf{v}_\varepsilon)$ is a sequence of solutions to (MFC_ε) . Then, there exists $(\mu, \mathbf{v}) \in C(m_0)$ so that, up to a subsequence,*

$$(1.3) \quad (\mu_\varepsilon, \mathbf{v}_\varepsilon) \rightarrow (\mu, \mathbf{v}) \text{ in } C([0, 1]; \mathcal{P}_1(\mathbb{R}^d)) \times \mathcal{M}(\mathbb{R}^d \times [0, 1]; U)$$

and (μ, \mathbf{v}) solves (MFC) .

REMARK 1.3 (Convergence to unique minimizer). *The fact that the preceding convergence result only holds “up to a subsequence” is due the generality of our assumptions, which do not necessarily ensure that solutions of (MFC) are unique. In particular, it is possible that different subsequences of $(\mu_\varepsilon, \mathbf{v}_\varepsilon)$ approximate different minimizers of (MFC) .*

On the other hand, if one imposes additional hypotheses to ensure that the minimizer (μ, \mathbf{v}) of (MFC) is unique, as is true in many important examples (see section 1.2), then Proposition 1.2 implies that every sequence $(\mu_\varepsilon, \mathbf{v}_\varepsilon)$ has a further subsequence that converges to (μ, \mathbf{v}) . Therefore, $(\mu_\varepsilon, \mathbf{v}_\varepsilon)$ itself must converge to (μ, \mathbf{v}) .

While smaller values of ε lead to minimizers μ_ε that are closer to the desired target measure m_1 at time $t = 1$, it is clear from the definition of \mathcal{F}_ε that minimizers are forced to satisfy $\mu_1 \in L^2$, which will always fail in the case of particle measures, $\mu_t = (1/N) \sum_{i=1}^N \delta_{y_i(t)}$. We navigate this issue by incorporating an additional regularization into the terminal constraint.

Consider a mollifier k_δ satisfying the following assumption:

ASSUMPTION 1.4. *Suppose $k \in C_b(\mathbb{R}^d)$ is nonnegative, even, and $\int |x|k(x)dx < +\infty$. For any $\delta > 0$, let $k_\delta(y) = k(y/\delta)/\delta^d$.*

For a probability measure ν , the convolution of ν with k_δ is a bounded, continuous function, given by

$$k_\delta * \nu(y) = \int_{\mathbb{R}^d} k_\delta(y-x)d\nu(x).$$

In this way, we can define the regularized functional

$$(1.4) \quad \mathcal{F}_{\varepsilon,\delta}(\mu) := \frac{1}{\varepsilon} \int_{\mathbb{R}^d} |k_\delta * \mu(y) - k_\delta * m_1(y)|^2 dy.$$

When $\int k = 1$ and $\mu \in L^2(\mathbb{R}^d)$, $\mathcal{F}_{\varepsilon,\delta}(\mu)$ converges to $\mathcal{F}_\varepsilon(\mu)$ as $\delta \rightarrow 0$. However, unlike $\mathcal{F}_\varepsilon(\mu)$, $\mathcal{F}_{\varepsilon,\delta}(\mu)$ is always finite — in fact, bounded by $2\|k\|_\infty/(\varepsilon\delta^d)$ — for all $\mu \in \mathcal{P}(\mathbb{R}^d)$. This leads to the optimization problem

$$(\text{MFC}_{\varepsilon,\delta}) \quad \min_{(\mu,\nu) \in \mathcal{C}(m_0)} \mathcal{E}_{\varepsilon,\delta}(\mu, \nu), \quad \mathcal{E}_{\varepsilon,\delta}(\mu, \nu) := \mathcal{G}(\mu, \nu) + \mathcal{F}_{\varepsilon,\delta}(\mu_1).$$

We will often use that, expanding the square, using associativity of convolution, and abbreviating $K_\delta := k_\delta * k_\delta$, we may express $\mathcal{F}_{\varepsilon,\delta}$ as follows:

$$(1.5) \quad \begin{aligned} \mathcal{F}_{\varepsilon,\delta}(\mu) &= \frac{1}{\varepsilon} \left[\int (k_\delta * k_\delta * \mu) d\mu - 2 \int (k_\delta * k_\delta * m_1) d\mu + \int (k_\delta * k_\delta * m_1) dm_1 \right] \\ &= \frac{1}{\varepsilon} \left[\int ((K_\delta * \mu) - 2(K_\delta * m_1)) d\mu + C_{\delta,m_1} \right], \end{aligned}$$

where $C_{\delta,m_1} := \int (K_\delta * m_1) dm_1$ is a constant independent of μ .

Under mild hypotheses, we show that, as long as (MFC_ε) is feasible, solutions of $(\text{MFC}_{\varepsilon,\delta})$ converge to a solution of (MFC_ε) as $\delta \rightarrow 0$, in the sense of the following theorem. (Note that, if (MFC) is feasible, then both (MFC_ε) and $(\text{MFC}_{\varepsilon,\delta})$ are feasible for all $\varepsilon, \delta > 0$.)

THEOREM 1.5 (Convergence as $\delta \rightarrow 0$). *Suppose Assumptions 1.1 and 1.4 hold and $\int k = 1$. Fix $\varepsilon > 0$, and suppose that (MFC_ε) is feasible. Then, for any sequence (μ_δ, ν_δ) of solutions to $(\text{MFC}_{\varepsilon,\delta})$, there exists (μ, ν) so that, up to a subsequence,*

$$(1.6) \quad (\mu_\delta, \nu_\delta) \rightarrow (\mu, \nu) \text{ in } C([0, 1]; \mathcal{P}_1(\mathbb{R}^d)) \times \mathcal{M}(\mathbb{R}^d \times [0, 1]; U)$$

and (μ, ν) solves (MFC_ε) .

As described in Remark 1.3, if one additionally assumes that the solution of (MFC_ε) is unique, then any sequence of minimizers of $(\text{MFC}_{\varepsilon,\delta})$ converges to this unique solution, without passing to a subsequence.

As indicated above, the key advantage of the regularized problem $(\text{MFC}_{\varepsilon,\delta})$, compared to (MFC) , is that the regularized problem admits a natural particle discretization. Replacing μ_0 with its empirical approximation,

$$\mu_0 \approx \frac{1}{N} \sum_{i=1}^N \delta_{y_{i,0}}, \quad \mathbf{y}_0 := [y_{i,0}]_{i=1}^N \in (\mathbb{R}^d)^N,$$

we obtain the following finite dimensional, ODE constrained optimization problem:

$$(\text{MFC}_{\varepsilon,\delta,N}) \quad \min_{(\mathbf{y}, \mathbf{u}) \in \mathcal{A}_N(\mathbf{y}_0)} \mathcal{E}_{\varepsilon,\delta,N}(\mathbf{y}, \mathbf{u})$$

where

$$(1.7) \quad \mathcal{E}_{\varepsilon, \delta, N}(\mathbf{y}, \mathbf{u}) := \frac{1}{N} \sum_{i=1}^N \int_0^1 \psi(u_i(t)) dt + \frac{1}{N} \sum_{i=1}^N \int_0^1 L_N(y_i(t), \mathbf{y}(t)) dt \\ + \frac{1}{\varepsilon} \left[\frac{1}{N^2} \sum_{i,j=1}^N K_\delta(y_i(1) - y_j(1)) - \frac{2}{N} \sum_{i=1}^N (K_\delta * m_1)(y_i(1)) + C_{\delta, m_1} \right] \\ \mathcal{A}_N(\mathbf{y}_0) := \left\{ (\mathbf{y}, \mathbf{u}) : \mathbf{y} \in AC([0, 1]; (\mathbb{R}^d)^N), \mathbf{u} \in L^1([0, 1]; U^N) \right. \\ \left. \dot{\mathbf{y}}_i(t) = \mathbf{F}_N(y_i(t), \mathbf{y}(t)) + u_i(t), \forall i = 1, \dots, N, \mathbf{y}(0) = \mathbf{y}_0 \right\}.$$

where C_{δ, m_1} is the constant from equation (1.5), U^N denotes the cartesian product of U , and the system of ordinary differential equations holds in the Carathéodory sense.

In order to describe our hypotheses on L_N and \mathbf{F}_N , which follow Fornasier et. al., we begin by recalling the notions of symmetry and convergence introduced in their work [30]. A map $G_N : \mathbb{R}^d \times (\mathbb{R}^d)^N \rightarrow \mathbb{R}^k$ is *symmetric* if

$$G_N(x, \mathbf{y}) = G_N(x, \sigma(\mathbf{y})) \text{ for all permutations } \sigma : (\mathbb{R}^d)^N \rightarrow (\mathbb{R}^d)^N.$$

Given a symmetric and continuous map G_N , we will consider the following notion of convergence to $G : \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{R}^k$:

(1.8) for any sequence of natural numbers $N \rightarrow +\infty$

$$\text{and } \mathbf{y}_N := [y_{i,N}]_{i=1}^N \subseteq (\mathbb{R}^d)^N \text{ satisfying } \mu_N := \frac{1}{N} \sum_{i=1}^N \delta_{y_{i,N}} \xrightarrow{N \rightarrow +\infty} \mu \text{ in } W_1,$$

$$\text{we have } \lim_{N \rightarrow +\infty} \sup_{z \in C} |G_N(z, \mathbf{y}_N) - G(z, \mu)| = 0, \quad \text{for all } C \subset \subset \mathbb{R}^d.$$

With this in hand, we turn to the hypotheses on L_N and \mathbf{F}_N :

ASSUMPTION 1.6. *We assume L_N and F_N satisfy either of the following hypotheses:*

- (i) (a) *If L satisfies Assumption 1.1(iii), we assume $L_N : \mathbb{R}^d \times (\mathbb{R}^d)^N \rightarrow [0, +\infty)$ is symmetric and continuous and L_N converges to L as in equation (1.8).*
- (b) *If L satisfies Assumption 1.1(iiib), we assume $L_N : \mathbb{R}^d \times (\mathbb{R}^d)^N \rightarrow [0, +\infty)$ is symmetric, lower semicontinuous, and independent of the second variable, i.e. $L_N(x, \mathbf{y}) = L_N(x, \mathbf{z})$ for all $x \in \mathbb{R}^d, \mathbf{y}, \mathbf{z} \in (\mathbb{R}^d)^N$. Furthermore, we assume $L_N \nearrow L$ pointwise.*
- (ii) $\mathbf{F}_N : \mathbb{R}^d \times (\mathbb{R}^d)^N \rightarrow \mathbb{R}^d$ *is continuous, symmetric, and there exist $C_{F_N}, C'_{F_N} \geq 0$ so that*

$$|\mathbf{F}_N(z, \mathbf{y})| \leq C_{F_N} + C'_{F_N} (|z| + \|\mathbf{y}\|), \quad \forall z \in \mathbb{R}^d, \mathbf{y} \in (\mathbb{R}^d)^N.$$

Furthermore, we assume the compatibility condition

$$\mathbf{F}_N(z, \mathbf{y}) - \mathbf{F}(z, \mu) \in U \text{ for all } z \in \mathbb{R}^d, \mathbf{y} \in (\mathbb{R}^d)^N, \mu \in \mathcal{P}_1(\mathbb{R}^d).$$

Lastly, we assume \mathbf{F}_N converges to \mathbf{F} as in equation (1.8).

We now consider existence of solutions to $(\text{MFC}_{\varepsilon, \delta, N})$. We show that if either (i) the control is unconstrained and the initial locations lie in the domain of the state and measure cost L_N or if (ii) the state and measure cost L_N is continuous, then $(\text{MFC}_{\varepsilon, \delta, N})$ is feasible. It then follows quickly that, whenever $(\text{MFC}_{\varepsilon, \delta, N})$ is feasible, a solution exists. Our proof is a mild adaptation of [30, Proposition 4.2], extending this result to the case when L_N satisfies Assumption 1.6(ib).

PROPOSITION 1.7. *Suppose Assumptions 1.1, 1.4 and 1.6 hold. Fix $\varepsilon > 0$, $\delta > 0$, $N \in \mathbb{N}$ and $\mathbf{y}_0 \in (\mathbb{R}^d)^N$. Suppose at least one of the following structural assumptions holds:*

- (a) *the control is unconstrained, $U = \mathbb{R}^d$ and $y_{i,0} \in \{L_N < +\infty\}$ for all $i = 1, \dots, N$;*
- (b) *the state and measure cost L_N satisfies Assumption 1.6(ia).*

Then $(\text{MFC}_{\varepsilon,\delta,N})$ is feasible.

Furthermore, whenever $(\text{MFC}_{\varepsilon,\delta,N})$ is feasible, then a solution of $(\text{MFC}_{\varepsilon,\delta,N})$ exists.

As a corollary of this and the preceding convergence results, we obtain sufficient conditions for existence of minimizers to the continuum optimization problems we consider.

COROLLARY 1.8. *Suppose Assumptions 1.1, 1.4, and 1.6 hold and $\int \phi(|x|)dm_0 < +\infty$, where ϕ is as in Assumption 2.1 below. Furthermore, suppose that (MFC) is feasible and that at least one of the following holds:*

- (a) *the control is unconstrained, $U = \mathbb{R}^d$, and $\text{supp } m_0 \subseteq \{L < +\infty\}$;*
- (b) *the state and measure costs L and L_N satisfy Assumptions 1.1(iia) and 1.6(ia).*

Then, for all $\varepsilon, \delta > 0$, solutions of $(\text{MFC}_{\varepsilon,\delta})$, (MFC_ε) , and (MFC) exist.

Finally, we consider the convergence of the discrete particle approximation $(\text{MFC}_{\varepsilon,\delta,N})$ to the continuum. In particular, we show that, if the problem either has an unconstrained control or is a controllable linear time invariant system, then, for fixed $\varepsilon, \delta > 0$, solutions of $(\text{MFC}_{\varepsilon,\delta,N})$ converge to a solution of $(\text{MFC}_{\varepsilon,\delta})$ as $N \rightarrow +\infty$.

THEOREM 1.9 (Convergence of minimizers as $N \rightarrow +\infty$). *Suppose Assumptions 1.1, 1.4, and 1.6 hold, and ϕ is as in Assumption 2.1 below. Suppose there exists a convex set Ω so that*

$$\text{supp } m_0 \subset\subset \Omega \subseteq \{L < +\infty\}.$$

Fix $\varepsilon, \delta > 0$, and suppose $(\text{MFC}_{\varepsilon,\delta})$ is feasible. Furthermore, assume at least one of the following structural assumptions holds:

- (a) *Unconstrained control: $U = \mathbb{R}^d$*
- (b) *Controllable linear time invariant system:*
 - (i) $L = L_N = 0$;
 - (ii) *There exists $A \in M_{d \times d}(\mathbb{R}^d)$ so that $\mathbf{F}_N(y, \cdot) = \mathbf{F}(y, \cdot) = Ay$;*
 - (iii) *There exists a full rank matrix $B \in M_{d \times k}(\mathbb{R}^d)$ with $R(B) = U$ for which the system is controllable: that is, the controllability gramian, $\Gamma(T) = \int_0^T e^{-\tau A} B B^T e^{-\tau A^T} d\tau$ is nonsingular for all $T > 0$.*
 - (iv) *For any sequence $\eta \rightarrow 0$,*

$$\eta \phi(|\Gamma^{-1}(\eta)(I - e^{-\eta A})z|) \rightarrow 0 \text{ uniformly for } z \in \text{supp } m_0.$$

In particular, it is sufficient that $\text{supp } m_0 \subseteq N(A)$.

Fix $\mathbf{y}_{N,0} \in \Omega^N$ with $\sup_{i,N} |y_{i,N,0}| < +\infty$, satisfying

$$(1.9) \quad \frac{1}{N} \sum_{i=1}^N \delta_{y_{i,N,0}} \xrightarrow{N \rightarrow +\infty} m_0 \text{ in } \mathcal{P}_1(\mathbb{R}^d).$$

Then, for any sequence $(\mathbf{y}_N, \mathbf{u}_N) \in \mathcal{A}(\mathbf{y}_{N,0})$ of minimizers of $(\text{MFC}_{\varepsilon,\delta,N})$, there exists $(\mu, \mathbf{v}) \in C(m_0)$ so that, up to a subsequence,

$$(1.10) \quad \frac{1}{N} \sum_{i=1}^N \delta_{y_{i,N}} \rightarrow \mu \text{ in } C([0, 1]; \mathcal{P}_1(\mathbb{R}^d)),$$

$$(1.11) \quad \frac{1}{N} \sum_{i=1}^N \int_0^1 u_{i,N}(t) \delta_{y_{i,N}(t)} dt \rightarrow \mathbf{v} \text{ in } \mathcal{M}(\mathbb{R}^d \times [0, 1]; U).$$

Furthermore, any such limit point (μ, \mathbf{v}) is a minimizer of $(\text{MFC}_{\varepsilon, \delta})$.

REMARK 1.10 (Discrete to continuum in the absence of a terminal constraint). *While the main focus on the present paper is mean field control in the presence of a terminal constraint, we note that the previous theorem includes the case of no terminal constraint, which arises when $k = k_\delta = 0$, so $\mathcal{F}_{\varepsilon, \delta} = 0$. In this way, the previous theorem extends the gamma convergence result of Fornasier, Lisini, Orrieri, and Savaré [30, Theorem 3.3] to the following two cases:*

- *unbounded state and measure costs L and L_N , as arise when modeling an obstacle,*
- *controllable linear time invariant systems, with constrained controls, $U \neq \mathbb{R}^d$.*

We now state our main result, which shows that solutions of $(\text{MFC}_{\varepsilon, \delta, N})$ converge to a solution of (MFC) in the limit $\varepsilon, \delta \rightarrow 0$ and $N \rightarrow +\infty$.

THEOREM 1.11 (Convergence of Minimizers as $\varepsilon, \delta \rightarrow 0$ and $N \rightarrow \infty$). *Suppose Assumptions 1.1, 1.4, and 1.6 hold and $\int k = 1$. Suppose (MFC) is feasible and there exists a convex set Ω so that*

$$\text{supp } m_0 \subset\subset \Omega \subseteq \{L < +\infty\}.$$

Furthermore, assume either the unconstrained control hypothesis (a) or the controllable linear time invariant system hypothesis (b) from Theorem 1.9.

Fix $\mathbf{y}_{N,0} \in \Omega^N$ with $\sup_{i,N} |y_{i,N,0}| < +\infty$ satisfying

$$\frac{1}{N} \sum_{i=1}^N \delta_{y_{i,N,0}} \xrightarrow{N \rightarrow +\infty} m_0 \text{ in } \mathcal{P}_1(\mathbb{R}^d).$$

Then, for any sequence $(\mathbf{y}_{\varepsilon, \delta, N}, \mathbf{u}_{\varepsilon, \delta, N}) \in \mathcal{A}(\mathbf{y}_{N,0})$ of minimizers of $(\text{MFC}_{\varepsilon, \delta, N})$, there exists subsequences ε_m, δ_m , and N_m so that, defining

$$(1.12) \quad \mathbf{y}_m := \mathbf{y}_{\varepsilon_m, \delta_m, N_m}, \quad \mathbf{u}_m = \mathbf{u}_{\varepsilon_m, \delta_m, N_m},$$

$$(1.13) \quad \mu_m := \frac{1}{N} \sum_{i=1}^N \delta_{y_{i,m}}, \quad \mathbf{v}_m := \frac{1}{N} \sum_{i=1}^N u_{i,m}(t) \delta_{y_{i,m}},$$

we have

$$(1.14) \quad \mu_m \rightarrow \mu \text{ in } C([0, 1]; \mathcal{P}_1(\mathbb{R}^d)),$$

$$(1.15) \quad \mathbf{v}_m \rightarrow \mathbf{v} \text{ in } \mathcal{M}(\mathbb{R}^d \times [0, 1]; U),$$

where (μ, \mathbf{v}) is a minimizer of (MFC) .

Note that, in the preceding theorem, even if (MFC) has a unique minimizer, our convergence result will only hold up to subsequences. This is due to the facts that the discretization parameter N must grow sufficiently quickly with respect to ε and δ and that δ must decay sufficiently quickly with respect to ε . We leave the question of quantifying the relationship between ε, δ , and N for which convergence holds, without passing to a subsequence, to future work.

1.2. Motivating Examples. We now describe three important examples of mean field control problems of the form (\star) , which motivate our numerical study.

A first special case is the dynamic formulation of the p -Wasserstein distance on $\mathcal{P}_p(\mathbb{R}^d)$,

the space of probability measures with finite p th moment, $M_p(\mu) = \int |x|^p d\mu(x)$, $p > 1$,

$$(W_p^p) \quad \min_{(\mu, \mathbf{u}) \in \mathcal{A}_W(m_0, m_1)} \int_0^1 \int_{\mathbb{R}^d} |\mathbf{u}(y, t)|^p d\mu_t(y) dt$$

$$\mathcal{A}_W(m_0, m_1) := \left\{ (\mu, \mathbf{u}) : \mu \in AC([0, 1], \mathcal{P}_1(\mathbb{R}^d)), \mathbf{u} \in L^1_{d\mu_t \otimes dt}(\mathbb{R}^d \times [0, 1]; \mathbb{R}^d) \right. \\ \left. \partial_t \mu_t + \nabla \cdot (\mathbf{u}_t \mu_t) = 0, \mu_0 = m_0, \mu_1 = m_1 \right\},$$

which arises from (\star) by taking the choices $\psi = |\cdot|^p$, $L = 0$, $U = \mathbb{R}^d$, and $\mathbf{F} = 0$, as introduced in the $p = 2$ case by Benamou and Brenier [11] and generalized to $p > 1$ by Ambrosio, Gigli, and Savaré [7]. To discretize this problem via a particle approximation, $(\text{MFC}_{\varepsilon, \delta, N})$, we take $L_N = 0$ and $\mathbf{F}_N = 0$. When $m_0, m_1 \in \mathcal{P}_p(\mathbb{R}^d)$ and m_1 is absolutely continuous with respect to Lebesgue measure, there exists a unique minimizer of (W_p^p) , and the optimal value of the objective function coincides with the p th power of the p -Wasserstein distance

$$W_p^p(m_0, m_1) = (W_p^p).$$

A second special case of the mean field control problem is the dynamic formulation of the p -Wasserstein distance around an obstacle Ω , represented by a open subset of \mathbb{R}^d on which the interpolating measure μ_t is forbidden from placing mass,

$$(W_p^p(\Omega^c)) \quad \min_{(\mu, \mathbf{u}) \in \mathcal{A}_{\Omega^c}(m_0, m_1)} \int_0^1 \int_{\mathbb{R}^d} |\mathbf{u}(y, t)|^p d\mu_t(y) dt$$

$$\mathcal{A}_{\Omega^c}(m_0, m_1) := \left\{ (\mu, \mathbf{u}) : \mu \in AC([0, 1], \mathcal{P}_1(\mathbb{R}^d)), \mathbf{u} \in L^1_{d\mu_t \otimes dt}(\mathbb{R}^d \times [0, 1]; \mathbb{R}^d) \right. \\ \left. \partial_t \mu_t + \nabla \cdot (\mathbf{u}_t \mu_t) = 0, \mu_0 = m_0, \mu_1 = m_1, \text{supp } \mu_t \subseteq \Omega^c \text{ for a.e. } t \in [0, 1] \right\}.$$

This arises from (\star) under the same choices as the preceding example, with the exception that

$$L(y, \mu) = \chi_{\Omega^c}(y), \quad \chi_{\Omega^c}(y) := \begin{cases} 0 & \text{if } y \in \Omega^c, \\ +\infty & \text{if } y \in \Omega. \end{cases}$$

To discretize this problem with the nonlocal terminal constraint, $(\text{MFC}_{\varepsilon, \delta, N})$, we may take $\mathbf{F}_N = 0$ and L_N to be any lower semicontinuous function $L_N(y, \mu_t) = L_N(y)$ that vanishes on Ω^c and converges up to $+\infty$ on Ω . This variant of the optimal transport problem can also be interpreted, more geometrically, as optimal transport on the manifold with holes, $\mathbb{R}^d \setminus \Omega$. Unlike in the case without obstacles, minimizers in general are not unique, unless Ω^c is convex.

A third important example of the mean field control problem arises when an acceleration control is imposed on the evolution of the measure μ_t ,

$$(\text{MFC}_a) \quad \min_{(\mu, \mathbf{a}) \in \mathcal{A}_a(m_0, m_1)} \int_0^1 \int_{\mathbb{R}^{2d}} |\mathbf{a}(x, v, t)|^2 d\mu_t(x, v) dt$$

$$\mathcal{A}_a(m_0, m_1) = \left\{ (\mu, \mathbf{a}) : \mu \in AC([0, 1], \mathcal{P}_1(\mathbb{R}^{2d})), \mathbf{a} \in L^1_{d\mu_t \otimes dt}(\mathbb{R}^{2d} \times [0, 1]; \mathbb{R}^d) \right. \\ \left. \partial_t \mu_t + \nabla \cdot ([v \ 0]^t \mu_t) + \nabla \cdot ([0 \ \mathbf{a}]^t \mu_t) = 0, \mu_0 = m_0, \mu_1 = m_1 \right\},$$

which arises from (\star) by the choices $y = (x, v) \in \mathbb{R}^{2d}$, $\psi = |\cdot|^2$, $L = 0$, $\mathbf{u} = (0, \mathbf{a}) \in U = \{0\} \times \mathbb{R}^d \subseteq \mathbb{R}^{2d}$, and $\mathbf{F}((x, v), \mu) = \mathbf{F}((x, v)) = [v \ 0]^t$. To discretize this problem with the nonlocal terminal constraint, $(\text{MFC}_{\varepsilon, \delta, N})$, we take $L_N = 0$ and $\mathbf{F}_N((x, v), \mathbf{z}) = [v \ 0]^t$. In this case, uniqueness of solutions is known [34, 22], when the probability measures are absolutely continuous and have compact support.

1.3. Strategy of approach. The remainder of the paper is organized as follows. In section 2, we specify the precise function spaces and notions of solution that we consider and prove the equivalence of the original formulation of the mean field control problem and the formulation in momentum coordinates. In section 3, we prove our main convergence results, relating minimizers of the optimization problems $(\text{MFC}_{\varepsilon,\delta,N})$, $(\text{MFC}_{\varepsilon,\delta})$, (MFC_ε) , and (MFC) in the limits as $\varepsilon, \delta \rightarrow 0$ and $N \rightarrow +\infty$. Finally, in section 4, we implement our particle method numerically and use it to compute solutions of the dynamical formulation of the p -Wasserstein metric, optimal transport around an obstacle, and measure transport with acceleration constraints, as well as numerically analyzing the convergence properties of the method.

2. Preliminaries.

2.1. Notation. Given $X \subseteq \mathbb{R}^d$, let $\mathcal{M}(X)$ denote the space of finite signed Borel measures on X , let $\mathcal{P}(X)$ denote the space of Borel probability measures, and, for any $p \geq 1$, let $\mathcal{P}_p(X)$ denote the subset of $\mathcal{P}(X)$ with finite p th moments, $M_p(\mu) := \int_X |x|^p d\mu(x) < +\infty$. We consider the space $\mathcal{P}_p(\mathbb{R}^d)$ to be endowed with the p -Wasserstein metric, W_p . We will often use the fact that convergence in the W_p metric for any $p \geq 1$ implies narrow convergence, which is to say, convergence in the duality with bounded continuous functions $f : \mathbb{R}^d \rightarrow \mathbb{R}$. For further details on the Wasserstein metrics and optimal transport, we refer the reader to one of the many excellent textbooks on the subject [7, 44, 43, 28, 4].

Given U , a subspace of \mathbb{R}^d , let $\mathcal{M}(X; U)$ denote the space of vector-valued finite signed Borel measures on X with range in U . We will consider $\mathcal{M}(X; U)$ to be endowed with the narrow topology, which is to say, convergence in $\mathcal{M}(X; U)$ is given by convergence in the duality with bounded continuous functions $f : X \rightarrow U$. For $X \subseteq \mathbb{R}^d$, $Y \subseteq \mathbb{R}^d$, and a Borel map $\mathbf{t} : X \rightarrow Y$, define $\mathbf{t}_\# : \mathcal{P}(X) \rightarrow \mathcal{P}(Y)$ by $(\mathbf{t}_\#\mu)(B) = \mu(\mathbf{t}^{-1}(B))$ for all Borel subsets $B \subseteq Y$. For any Borel set B , we let 1_B denote the indicator function on B , that is, $1_B(x) = 1$ if $x \in B$ and $1_B(x) = 0$ if $x \notin B$.

For any product space $\prod_{i=1}^N X_i$, $X_i \subseteq \mathbb{R}^d$, let π^i denote the projection onto the i th coordinate. When elements of the product space are denoted by a distinguished variable, e.g. $(x, y, z) \in X \times Y \times Z$, we write π^x, π^y, π^z for projections onto the corresponding components.

The set of continuous curves in $\mathcal{P}_1(\mathbb{R}^d)$ will be denoted by $C([0, 1]; \mathcal{P}_1(\mathbb{R}^d))$ and the set of absolutely continuous curves will be denoted by $AC([0, 1]; \mathcal{P}_1(\mathbb{R}^d))$. We will often identify curves $\mu \in C([0, 1]; \mathcal{P}_1(\mathbb{R}^d))$ with the element $\tilde{\mu} \in \mathcal{P}_1(\mathbb{R}^d \times [0, 1])$ satisfying

$$(2.1) \quad \int_{\mathbb{R}^d \times [0, 1]} f(y, t) d\tilde{\mu}(y, t) = \int_0^1 \int_{\mathbb{R}^d} f(y, t) d\mu_t(y) dt$$

for all $f \in C_b(\mathbb{R}^d \times [0, 1])$. We will abbreviate $\tilde{\mu}$ by $d\mu_t \otimes dt$. Under the hypotheses of Assumption 1.1, $\mu \in AC([0, 1]; \mathcal{P}_1(\mathbb{R}^d))$ and $\mathbf{u} \in L^1_{d\mu_t \otimes dt}(\mathbb{R}^d \times [0, 1]; U)$ is a distributional solution of the equation $\partial_t \mu_t + \nabla \cdot ((\mathbf{F}(\cdot, \mu_t)\mu_t + \mathbf{u}(\cdot, t)\mu_t)) = 0$ if, for all $\varphi \in C_c^\infty(\mathbb{R}^d \times [0, 1])$,

$$(2.2) \quad \int_0^1 \int_{\mathbb{R}^d} \partial_t \varphi(y, t) + \nabla \varphi(y, t) \cdot ((\mathbf{F}(y, \mu_t) + \mathbf{u}(y, t)) d\mu_t(y) dt = 0.$$

Given $\mathbf{v} \in \mathcal{M}(\mathbb{R}^d \times [0, 1]; U)$ and $\mu \in C([0, 1]; \mathcal{P}_1(\mathbb{R}^d))$, if \mathbf{v} is absolutely continuous with respect to the measure $\tilde{\mu} = d\mu_t \otimes dt$, we let $\mathbf{u} : \mathbb{R}^d \times [0, 1] \rightarrow U$ denote the Radon-Nikodym derivative: $\mathbf{u}(y, t) d\tilde{\mu}(y, t) = \mathbf{u}(y, t) d\mu_t(y) dt = d\mathbf{v}(y, t)$. Furthermore, for any $\mathbf{v} \in \mathcal{M}(\mathbb{R}^d \times [0, 1]; U)$, we let \mathbf{v}_t denote the disintegration of \mathbf{v} with respect to $\pi^t_\# \mathbf{v} \in \mathcal{M}([0, 1])$, where π^t denotes the projection onto the second, temporal component of $\mathbb{R}^d \times [0, 1]$. In this

way, for all $f \in C_b(\mathbb{R}^d \times [0, 1])$,

$$\int_{\mathbb{R}^d \times [0, 1]} f(y, t) d\mathbf{v}(y, t) = \int_0^1 \left(\int_{\mathbb{R}^d} f(y, t) d\mathbf{v}_t(y) \right) d(\pi_{\#}^t \mathbf{v})(t).$$

We say that $(\mu, \mathbf{v}) \in AC([0, 1]; \mathcal{P}_1(\mathbb{R}^d)) \times \mathcal{M}(\mathbb{R}^d \times [0, 1]; U)$ is a distributional solution of $\partial_t \mu_t + \nabla \cdot ((\mathbf{F}(\cdot, \mu_t) \mu_t + \mathbf{v}_t)) = 0$ if, for all $\varphi \in C_c^\infty(\mathbb{R}^d \times [0, 1])$,

$$(2.3) \quad \int_0^1 \int_{\mathbb{R}^d} (\partial_t \varphi(y, t) + \nabla \varphi(y, t) \cdot \mathbf{F}(y, \mu_t)) d\mu_t(y) dt + \int_0^1 \int_{\mathbb{R}^d} \nabla \varphi(y, t) \cdot d\mathbf{v}(y, t) = 0.$$

2.2. Control cost and momentum coordinates. We now state the general hypothesis on our control cost ψ , following previous work by Fornasier, et. al. [30, Section 2.2].

ASSUMPTION 2.1. *Assume the control cost $\psi : U \rightarrow [0, +\infty)$ is convex, lower semi-continuous, and satisfies $\psi(0) = 0$. Furthermore, assume there exists a moderating function $\phi : [0, +\infty) \rightarrow [0, +\infty)$ that is strictly convex, continuously differentiable, satisfies $\phi(0) = 0$ and $\phi'(0) = 0$, is superlinear at $+\infty$, and for which there exists $K > 0$ so that*

$$(2.4) \quad \phi(2r) \leq K(1 + \phi(r)) \quad \text{for all } r \in [0, +\infty).$$

Finally, assume that the control cost and the moderating function are related as follows: there exists $C > 0$ so that

$$(2.5) \quad \phi(|x|) - 1 \leq \psi(x) \leq C(1 + \phi(|x|)), \quad \forall x \in U.$$

Examples of control costs satisfying this assumption include

- $\psi(x) = \frac{1}{p}|x|^p$ for $p > 1$;
- $\psi(x) = \begin{cases} \frac{1}{p}|x| & \text{for } |x| \leq 1 \\ \frac{1}{p}|x|^p & \text{for } |x| > 1 \end{cases}$ for $p > 1$.

In particular, the role of the moderating function ϕ is that it allows one to decouple the smoothness and monotonicity properties from the local behavior of ψ . In the following lemma, we gather two elementary observations about the relationship between ψ and ϕ that we will use in what follows.

LEMMA 2.2. *Suppose ϕ and ψ satisfy Assumption 2.1.*

(i) *There exists $R_\phi > 0$ so that $|x| \leq \phi(|x|) + R_\phi \leq \psi(x) + R_\phi + 1$, $\forall x \in U$.*

(ii) *There exists $K' > 0$ so that, for all $D \geq 1$, $\phi(Dr) + Dr \leq K' D^{K'} (\phi(r) + r)$, $\forall r \in [0, +\infty)$.*

We now make precise the sense in which (\star) and its the momentum formulation (MFC) are equivalent. See Appendix A for the proof.

LEMMA 2.3. *If (μ, \mathbf{u}) solves (\star) , then $(\mu, \mathbf{u} d\mu_t \otimes dt)$ solves (MFC). On the other hand, if (μ, \mathbf{v}) solves (MFC), then $d\mathbf{v} \ll d\mu_t \otimes dt$ and $(\mu, d\mathbf{v}/(d\mu_t \otimes dt))$ solves (\star) .*

3. Convergence of minimizers as $\varepsilon \rightarrow 0$, $\delta \rightarrow 0$, and $N \rightarrow +\infty$. We now turn to the proofs of our major results, relating minimizers of the problems (MFC $_{\varepsilon, \delta, N}$), (MFC $_{\varepsilon, \delta}$), (MFC $_{\varepsilon}$), and (MFC) in the limits as $\varepsilon \rightarrow 0$, $\delta \rightarrow 0$, and $N \rightarrow +\infty$. We begin, in section 3.1, by establishing some fundamental lower semicontinuity and compactness properties of the continuum mean field control problem. In section 3.2, we consider the $\varepsilon \rightarrow 0$ limit. In section 3.3, we consider the behavior as $\delta \rightarrow 0$, focusing on the proof of Γ -convergence of the objective functionals, after which convergence of minimizers follows quickly. Finally, in section 3.4, we consider the $N \rightarrow +\infty$ limit. Our arguments in section 3.4 are strongly inspired by previous work by Fornasier, et. al. [30]. Furthermore, as explained in Remark 1.10, we succeed in extending these arguments to our more general hypotheses on the state and measure cost L and to controllable linear time invariant systems, with $U \neq \mathbb{R}^d$.

3.1. Lower semicontinuity and compactness. In this section, we collect some fundamental lower semicontinuity and compactness properties for the continuum mean field control problem. See appendix A for the proofs.

First, we note that, under the hypotheses of Assumption 1.1, the functional \mathcal{G} is lower semicontinuous.

LEMMA 3.1. *Suppose Assumption 1.1 holds. Then the functional \mathcal{G} , defined in (MFC), is lower semicontinuous on $C([0, 1], \mathcal{P}_1(\mathbb{R}^d)) \times \mathcal{M}(\mathbb{R}^d \times [0, 1]; U)$.*

Next, we observe that sublevels of \mathcal{G} in the constraint set $C(m_0)$ are sequentially compact.

LEMMA 3.2. *Suppose Assumption 1.1 holds. If $(\mu_n, \nu_n) \in C(m_0)$ and $\sup_n \mathcal{G}(\mu_n, \nu_n) < +\infty$, then there exists $(\mu, \nu) \in C(m_0)$ so that, up to a subsequence,*

$$(3.1) \quad (\mu_n, \nu_n) \text{ converges to } (\mu, \nu) \text{ in } C([0, 1]; \mathcal{P}_1(\mathbb{R}^d)) \times \mathcal{M}(\mathbb{R}^d \times [0, 1]; U)$$

and $\mathcal{G}(\mu, \nu) \leq \sup_n \mathcal{G}(\mu_n, \nu_n)$.

3.2. Convergence as $\varepsilon \rightarrow 0$: soft to hard terminal constraint. We now collect our results on the convergence of (MFC $_\varepsilon$) to (MFC) as $\varepsilon \rightarrow 0$. The proofs follow standard Γ -convergence arguments, which we defer to Appendix B.

We first observe that the sequence of functionals \mathcal{E}_ε , defined in (MFC $_\varepsilon$), Γ -converges to the functional \mathcal{E} , defined in (MFC), as $\varepsilon \rightarrow 0$.

PROPOSITION 3.3 (Γ -convergence as $\varepsilon \rightarrow 0$). *Suppose Assumption 1.1 holds.*

- (i) *For every sequence $(\mu_\varepsilon, \nu_\varepsilon)$ converging in $C([0, 1]; \mathcal{P}_1(\mathbb{R}^d)) \times \mathcal{M}(\mathbb{R}^d \times [0, 1]; U)$ to (μ, ν) , we have $\mathcal{E}(\mu, \nu) \leq \liminf_{\varepsilon \rightarrow 0} \mathcal{E}_\varepsilon(\mu_\varepsilon, \nu_\varepsilon)$.*
- (ii) *For every $(\mu, \nu) \in C([0, 1]; \mathcal{P}_1(\mathbb{R}^d)) \times \mathcal{M}(\mathbb{R}^d \times [0, 1]; U)$, $\mathcal{E}(\mu, \nu) \geq \limsup_{\varepsilon \rightarrow 0} \mathcal{E}_\varepsilon(\mu, \nu)$.*

It is then an immediate consequence of this result that minimizers of (MFC $_\varepsilon$) converge to a minimizer of (MFC), up to a subsequence, as stated in Proposition 1.2; see appendix B.

3.3. Convergence as $\delta \rightarrow 0$: nonlocal to local penalization on terminal measure. In the present section, we show that, for fixed $\varepsilon > 0$, the sequence of functionals $\mathcal{E}_{\varepsilon, \delta}$, defined in (MFC $_{\varepsilon, \delta}$), Γ -converges to the functional \mathcal{E}_ε , defined in equation MFC $_\varepsilon$, as $\delta \rightarrow 0$.

PROPOSITION 3.4 (Γ -convergence as $\delta \rightarrow 0$). *Suppose Assumptions 1.1 and 1.4 hold and $\int k = 1$. Fix $\varepsilon > 0$.*

- (i) *For every sequence (μ_δ, ν_δ) converging in $C([0, 1]; \mathcal{P}_1(\mathbb{R}^d)) \times \mathcal{M}(\mathbb{R}^d \times [0, 1]; U)$ to (μ, ν) , we have $\mathcal{E}_\varepsilon(\mu, \nu) \leq \liminf_{\delta \rightarrow 0} \mathcal{E}_{\varepsilon, \delta}(\mu_\delta, \nu_\delta)$.*
- (ii) *For all $(\mu, \nu) \in C([0, 1]; \mathcal{P}_1(\mathbb{R}^d)) \times \mathcal{M}(\mathbb{R}^d \times [0, 1]; U)$, $\mathcal{E}_\varepsilon(\mu, \nu) \geq \limsup_{\delta \rightarrow 0} \mathcal{E}_{\varepsilon, \delta}(\mu, \nu)$.*

Proof. First, we consider part (i). Without loss of generality, we may pass to a subsequence so that

$$(3.2) \quad \lim_{\delta \rightarrow 0} \mathcal{E}_{\varepsilon, \delta}(\mu_\delta, \nu_\delta) = \liminf_{\delta \rightarrow 0} \mathcal{E}_{\varepsilon, \delta}(\mu_\delta, \nu_\delta) < +\infty.$$

In particular, this ensures that

$$\sup_{\delta} \varepsilon^{-1} \|k_\delta * \mu_{\delta, 1} - k_\delta * m_1\|_{L^2(\mathbb{R}^d)}^2 = \sup_{\delta} \mathcal{F}_{\varepsilon, \delta}(\mu_{\delta, 1}) < +\infty.$$

Since $m_1 \in L^2(\mathbb{R}^d)$, $k_\delta * m_1 \rightarrow m_1$ in $L^2(\mathbb{R}^d)$, and in particular, $k_\delta * m_1$ is uniformly bounded in $L^2(\mathbb{R}^d)$. Thus $k_\delta * \mu_{\delta, 1}$ must be uniformly bounded in $L^2(\mathbb{R}^d)$, and, up to another subsequence, $k_\delta * \mu_{\delta, 1}$ converges weakly in $L^2(\mathbb{R}^d)$. By the convergence of μ_δ to μ , we have

that $\mu_{\delta,1}$ converges to μ_1 in $\mathcal{P}_1(\mathbb{R}^d)$. Furthermore, [18, Lemma 2.3] ensures $k_\delta * \mu_{\delta,1} \rightarrow \mu_1$ in distribution. Thus, by uniqueness of limits, $k_\delta * \mu_{\delta,1}$ converges to μ_1 weakly in $L^2(\mathbb{R}^d)$ and, in particular, $\mu_1 \in L^2(\mathbb{R}^d)$.

Now we consider the limits of the functionals along these sequences. Since $\mu_{\delta,1}$ converges to μ_1 in $\mathcal{P}_1(\mathbb{R}^d)$, by [18, Theorem 4.1],

$$(3.3) \quad \liminf_{\delta \rightarrow 0} \int K_\delta * \mu_{\delta,1} d\mu_{\delta,1} \geq \int \mu_1^2.$$

Since $k_\delta * \mu_{\delta,1} \rightarrow \mu_1$ weakly in $L^2(\mathbb{R}^d)$ and $k_\delta * m_1 \rightarrow m_1$ strongly in $L^2(\mathbb{R}^d)$, for C_{δ,m_1} as in equation (1.5), we obtain

$$(3.4) \quad \begin{aligned} \liminf_{\delta \rightarrow 0} -2 \int (K_\delta * m_1) d\mu_{\delta,1} + C_{\delta,m_1} \\ = \liminf_{\delta \rightarrow 0} -2 \int (k_\delta * \mu_{\delta,1})(k_\delta * m_1) + \int (K_\delta * m_1) dm_1 \geq -2 \int m_1 \mu_1 + \int m_1^2. \end{aligned}$$

Thus, combining equation (3.3) and (3.4) with the expression for $\mathcal{F}_{\varepsilon,\delta}$ in equation (1.5), we obtain

$$\mathcal{F}_\varepsilon(\mu_1) = \frac{1}{\varepsilon} \|\mu_1 - m_1\|_{L^2(\mathbb{R}^d)}^2 \leq \liminf_{\delta \rightarrow 0} \mathcal{F}_{\varepsilon,\delta}(\mu_{\delta,1})$$

Due to the lower semicontinuity of the functional \mathcal{G} , shown in Lemma 3.1, we can therefore conclude that

$$\mathcal{E}_\varepsilon(\mu, \mathbf{v}) \leq \liminf_{\delta \rightarrow 0} \mathcal{E}_{\varepsilon,\delta}(\mu_\delta, \mathbf{v}_\delta).$$

Now, we turn to part (ii). We may assume that $\mathcal{E}_\varepsilon(\mu, \mathbf{v}) < \infty$, otherwise the inequality is trivial. Thus, $\mathcal{F}_\varepsilon(\mu_1) < +\infty$, so $\mu_1 \in L^2(\mathbb{R}^d)$ and $k_\delta * \mu_1 \rightarrow \mu_1$ in $L^2(\mathbb{R}^d)$. Thus, $\lim_{\delta \rightarrow 0} \mathcal{F}_{\varepsilon,\delta}(\mu_1) = \mathcal{F}_\varepsilon(\mu_1)$. Therefore, we conclude that,

$$\mathcal{E}_\varepsilon(\mu, \mathbf{v}) \geq \limsup_{\delta \rightarrow 0} \mathcal{E}_{\varepsilon,\delta}(\mu^\delta, \mathbf{v}^\delta). \quad \square$$

Theorem 1.5, on the convergence of minimizers of $(\text{MFC}_{\varepsilon,\delta})$ to a minimizer of (MFC_ε) now follows from this Γ -convergence result via a standard argument; see Appendix B.

3.4. Convergence as $N \rightarrow +\infty$: discrete to continuum. We now turn to the convergence of the spatially discrete problem, in the continuum limit $N \rightarrow +\infty$. Note that, by definition of $\mathcal{E}_{\varepsilon,\delta,N}$ in equation (1.7), for any $(\mathbf{y}, \mathbf{u}) \in (\mathbb{R}^d)^N \times U^N$,

$$\begin{aligned} \mathcal{E}_{\varepsilon,\delta,N}(\mathbf{y}, \mathbf{u}) &= \mathcal{G}_N(\mathbf{y}, \mathbf{u}) + \mathcal{F}_{\varepsilon,\delta} \left(\frac{1}{N} \sum_{i=1}^N \delta_{y_i(1)} \right) \\ \mathcal{G}_N(\mathbf{y}, \mathbf{u}) &:= \frac{1}{N} \sum_{i=1}^N \int_0^1 \psi(u_i(t)) dt + \frac{1}{N} \sum_{i=1}^N \int_0^1 L_N(y_i(t), \mathbf{y}(t)) dt. \end{aligned}$$

We begin with the following lemma.

LEMMA 3.5. *Suppose Assumptions 1.1 and 1.4 hold, and suppose $\mathbf{y}_N \in (\mathbb{R}^d)^N$ satisfies $\frac{1}{N} \sum_{i=1}^N \delta_{y_i} \rightarrow \mu$ narrowly for some $\mu \in \mathcal{P}(\mathbb{R}^d)$. Then, for all $\varepsilon, \delta > 0$,*

$$\lim_{N \rightarrow +\infty} \mathcal{F}_{\varepsilon,\delta} \left(\frac{1}{N} \sum_{i=1}^N \delta_{y_i} \right) = \mathcal{F}_{\varepsilon,\delta}(\mu).$$

Proof. Let $\mu_N = \frac{1}{N} \sum_{i=1}^N \delta_{y_i}$. Note that, using the expression (1.5) and neglecting the constant C_{δ, m_1} , we may rewrite the left hand side of the expression in the following way:

$$\int [K_\delta * \mu_N - 2(K_\delta * m_1)] d\mu_N = \iint K_\delta(x-y) d\mu_N(x) d\mu_N(y) - 2 \int (K_\delta * m_1) d\mu_N$$

By Assumption 1.4, $(x, y) \mapsto K_\delta(x-y)$ and $x \mapsto k_\delta * m_1(x)$ are bounded and continuous, so the result is an immediate consequence of the definition of narrow convergence. \square

Now, combining the preceding lemma with the Γ -convergence result of Fornasier et al. [30, Theorem 3.2], making appropriate adaptations when the state and measure costs satisfy our alternative assumptions 1.1(iiib) and 1.6(ib), we obtain the following result.

PROPOSITION 3.6 (Γ -convergence as $N \rightarrow +\infty$). *Suppose Assumptions 1.1, 1.4, and 1.6 hold. Fix $\varepsilon > 0$ and $\delta > 0$.*

(i) *Consider $(\mathbf{y}_N, \mathbf{u}_N) \in C([0, 1]; (\mathbb{R}^d)^N \times L^1([0, 1]; U^N))$, and suppose there exists (μ, ν) for which*

$$(3.5) \quad \frac{1}{N} \sum_{i=1}^N \delta_{y_{i,N}} \rightarrow \mu \text{ in } C([0, 1]; \mathcal{P}_1(\mathbb{R}^d)),$$

$$(3.6) \quad \frac{1}{N} \sum_{i=1}^N \int_0^1 u_{i,N}(t) \delta_{y_{i,N}(t)} dt \rightarrow \nu \text{ in } \mathcal{M}(\mathbb{R}^d \times [0, 1]; U).$$

Then, $\liminf_{N \rightarrow +\infty} \mathcal{E}_{\varepsilon, \delta, N}(\mathbf{y}_N, \mathbf{u}_N) \geq \mathcal{E}_{\varepsilon, \delta}(\mu, \nu)$.

(ii) *Suppose $\int \phi(|x|) dm_0(x) < +\infty$. For every $(\mu, \nu) \in C(m_0)$ such that $\mathcal{G}(\mu, \nu) < +\infty$ and $N \in \mathbb{N}$, there exists $\mathbf{y}_{N,0} \in (\mathbb{R}^d)^N$ and $(\mathbf{y}_N, \mathbf{u}_N) \in \mathcal{A}_N(\mathbf{y}_{N,0})$ with $y_{i,N,t} \subseteq \text{supp } \mu_t$ for every $i = 1, \dots, N$ and $t \in [0, 1]$, so that*

$$(3.7) \quad \frac{1}{N} \sum_{i=1}^N \delta_{y_{i,N}} \rightarrow \mu \text{ in } C([0, 1]; \mathcal{P}_1(\mathbb{R}^d)),$$

$$(3.8) \quad \frac{1}{N} \sum_{i=1}^N \int_0^1 u_{i,N}(t) \delta_{y_{i,N}(t)} dt \rightarrow \nu \text{ in } \mathcal{M}(\mathbb{R}^d \times [0, 1]; U),$$

$$(3.9) \quad \limsup_{N \rightarrow +\infty} \mathcal{E}_{\varepsilon, \delta, N}(\mathbf{y}_N, \mathbf{u}_N) \leq \mathcal{E}_{\varepsilon, \delta}(\mu, \nu).$$

Proof. First, we show part (i). Up to passing to a subsequence, we may assume without loss of generality that

$$(3.10) \quad \lim_{N \rightarrow +\infty} \mathcal{E}_{\varepsilon, \delta, N}(\mathbf{y}_N, \mathbf{u}_N) = \liminf_{N \rightarrow +\infty} \mathcal{E}_{\varepsilon, \delta, N}(\mathbf{y}_N, \mathbf{u}_N) < +\infty.$$

By Lemma 3.5,

$$(3.11) \quad \mathcal{F}_{\varepsilon, \delta} \left(\frac{1}{N} \sum_{i=1}^N \delta_{y_{i,N}(1)} \right) \rightarrow \mathcal{F}_{\varepsilon, \delta}(\mu_1).$$

Suppose that L and L_N satisfy Assumption 1.1(iia) and Assumption 1.6(ia). By [30, Theorem 3.2(i)], we have

$$\liminf_{N \rightarrow +\infty} \mathcal{G}_N(\mathbf{y}_N, \mathbf{u}_N) \geq \mathcal{G}(\mu, \nu).$$

Since $\mathcal{E}_{\varepsilon, \delta}(\mu, \nu) = \mathcal{G}(\mu, \nu) + \mathcal{F}_{\varepsilon, \delta}(\mu_1)$, combining the preceding inequality with (3.11) gives the result.

On the other hand, suppose that L and L_N satisfy Assumption 1.1(iii b) and Assumption 1.6(ib). By [30, Theorem 3.2(i)], we have

$$(3.12) \quad \liminf_{N \rightarrow +\infty} \frac{1}{N} \sum_{i=1}^N \int_0^1 \psi(u_i(t)) dt \geq \Psi(\mathbf{v}|\mu).$$

Furthermore, for $\mu_N := \frac{1}{N} \sum_{i=1}^N \delta_{y_{i,N}}$, Fatou's lemma ensures

$$\begin{aligned} \liminf_{N \rightarrow +\infty} \frac{1}{N} \sum_{i=1}^N \int_0^1 L_N(y_{i,N}(t)) dt &= \liminf_{N \rightarrow +\infty} \int_0^1 \int_{\mathbb{R}^d} L_N d\mu_{N,t} \\ &\geq \int_0^1 \liminf_{N \rightarrow +\infty} \int_{\mathbb{R}^d} L_N d\mu_{N,t} dt. \end{aligned}$$

For any $M \in \mathbb{N}$, the fact that $L_N \nearrow L$ and L_M is lower semicontinuous and nonnegative ensures

$$\liminf_{N \rightarrow +\infty} \int_{\mathbb{R}^d} L_N d\mu_{N,t} \geq \liminf_{N \rightarrow +\infty} \int_{\mathbb{R}^d} L_M d\mu_{N,t} \geq \int_{\mathbb{R}^d} L_M d\mu_t.$$

Sending $M \rightarrow +\infty$ on the right hand side, the monotone convergence theorem ensures

$$(3.13) \quad \liminf_{N \rightarrow +\infty} \int_{\mathbb{R}^d} L_N d\mu_{N,t} \geq \int_{\mathbb{R}^d} L d\mu_t.$$

Combining this with (3.12), we obtain $\liminf_{N \rightarrow +\infty} \mathcal{G}_N(\mathbf{y}_N, \mathbf{u}_N) \geq \mathcal{G}(\mu, \mathbf{v})$. Finally, combining this with (3.11) gives the result.

We now show part (ii). It is an immediate consequence of [30, Theorem 3.2(i)] that (3.7) and (3.8) hold and

$$(3.14) \quad \limsup_{N \rightarrow +\infty} \frac{1}{N} \sum_{i=1}^N \int_0^1 \psi(u_i(t)) dt \leq \Psi(\mathbf{v}|\mu).$$

(The fact that we may choose $y_{i,N}(t) \in \text{supp}(\mu_t)$ for every $i = 1, \dots, N$ and $t \in [0, 1]$ can be seen by inspection of the proof: in the equation following [30, equation (6.21)], we may assume $\gamma_{i,k,m} \in \text{supp} \tilde{\pi}^k \subseteq \text{supp} \pi$.) As before, by Lemma 3.5,

$$(3.15) \quad \mathcal{F}_{\varepsilon,\delta} \left(\frac{1}{N} \sum_{i=1}^N \delta_{y_{i,N}(1)} \right) \rightarrow \mathcal{F}_{\varepsilon,\delta}(\mu_1).$$

If L and L_N satisfy Assumption 1.1(iii a) and Assumption 1.6(ia), [30, Theorem 3.2(i)] also gives $\limsup_{N \rightarrow +\infty} \mathcal{G}_N(\mathbf{y}_N, \mathbf{u}_N) \leq \mathcal{G}(\mu, \mathbf{v})$. Since $\mathcal{E}_{\varepsilon,\delta}(\mu, \mathbf{v}) = \mathcal{G}(\mu, \mathbf{v}) + \mathcal{F}_{\varepsilon,\delta}(\mu_1)$, combining the preceding inequality with (3.15) gives the result in this case.

On the other hand, suppose L and L_N satisfy Assumption 1.1(iii b) and Assumption 1.6(ib). Without loss of generality, we may assume $\mathcal{E}(\mu, \mathbf{v}) < +\infty$, so that

$$\begin{aligned} \int_0^1 \int_{\mathbb{R}^d} L d\mu_t dt < +\infty &\implies \int_{\mathbb{R}^d} L d\mu_t < +\infty \text{ for a.e. } t \in [0, 1] \\ &\implies \{y_{i,N}(t)\}_{i=1}^N \subseteq \text{supp} \mu_t \subseteq \{L < +\infty\} \text{ for a.e. } t \in [0, 1]. \end{aligned}$$

Thus, the fact that $L_N \nearrow L$ implies

$$\begin{aligned} \limsup_{N \rightarrow +\infty} \int_0^1 \int_{\mathbb{R}^d} L_N d\mu_{N,t} dt &\leq \limsup_{N \rightarrow +\infty} \int_0^1 \int_{\mathbb{R}^d} L d\mu_{N,t} dt \\ &= \limsup_{N \rightarrow +\infty} \int_0^1 \int_{\mathbb{R}^d} L1_{\{L < +\infty\}} d\mu_{N,t} dt \leq \int_0^1 \int_{\mathbb{R}^d} L1_{\{L < +\infty\}} d\mu_t dt \leq \int_0^1 \int_{\mathbb{R}^d} L d\mu_t dt, \square \end{aligned}$$

where the second to last inequality follows since L is continuous on the closed set $\{L < +\infty\}$, so $L1_{\{L < +\infty\}}$ is upper semicontinuous. Combining this with (3.14) and (3.15) gives the result.

The preceding convergence result can now be used to show that, for fixed $\varepsilon, \delta > 0$, as $N \rightarrow +\infty$, minimizers of the spatially discrete problem $(\text{MFC}_{\varepsilon, \delta, N})$ converge to a solution of $(\text{MFC}_{\varepsilon, \delta})$, up to a subsequence. The result follows from a standard argument, which we defer to Appendix B.

THEOREM 3.7 (Convergence of minimizers as $N \rightarrow +\infty$). *Suppose Assumptions 1.1, 1.4, and 1.6 hold, and suppose $\int \phi(|x|) dm_0(x) < +\infty$. Fix $\varepsilon, \delta > 0$, and suppose $(\text{MFC}_{\varepsilon, \delta})$ is feasible.*

Then, there exists $\mathbf{y}_{N,0} \in (\mathbb{R}^d)^N$ satisfying

$$(3.16) \quad y_{i,N,0} \in \text{supp } m_0 \text{ for all } i = 1, \dots, N,$$

$$(3.17) \quad \frac{1}{N} \sum_{i=1}^N \delta_{y_{i,N,0}} \rightarrow m_0 \text{ in } \mathcal{P}_1(\mathbb{R}^d),$$

so that, if $(\mathbf{y}_N, \mathbf{u}_N) \in \mathcal{A}(\mathbf{y}_{N,0})$ is a minimizer of $(\text{MFC}_{\varepsilon, \delta, N})$, then

$$(3.18) \quad \lim_{N \rightarrow +\infty} \mathcal{E}_{\varepsilon, \delta, N}(\mathbf{y}_N, \mathbf{u}_N) = \inf_{(\mu, \nu) \in \mathcal{C}(m_0)} \mathcal{E}_{\varepsilon, \delta}(\mu, \nu).$$

and there exists $(\mu, \nu) \in \mathcal{C}(m_0)$ so that, up to a subsequence,

$$(3.19) \quad \frac{1}{N} \sum_{i=1}^N \delta_{y_{i,N}} \rightarrow \mu \text{ in } C([0, 1]; \mathcal{P}_1(\mathbb{R}^d)),$$

$$(3.20) \quad \frac{1}{N} \sum_{i=1}^N \int_0^1 u_{i,N}(t) \delta_{y_{i,N}(t)} dt \rightarrow \nu \text{ in } \mathcal{M}(\mathbb{R}^d \times [0, 1]; U).$$

Furthermore, any such limit point (μ, ν) is a minimizer of $(\text{MFC}_{\varepsilon, \delta})$.

As described in Remark 1.3, if one additionally knows that the solution of $(\text{MFC}_{\varepsilon, \delta})$ is unique, then any sequence of minimizers $(\mathbf{y}_N, \mathbf{u}_N)$ converges to the unique solution, without passing to a subsequence. However, note that while the above theorem ensures *there exists* a sequence of initial conditions $\mathbf{y}_{N,0}$ for which minimizers of $(\text{MFC}_{\varepsilon, \delta, N})$ converge to a minimizer of $(\text{MFC}_{\varepsilon, \delta})$, at this level of generality, it is not known if the result is true for *all* choices of initial conditions.

In the special cases of unconstrained controls and controllable linear time invariant systems, Theorem 1.9 ensures that this result indeed continues to hold for all well-prepared initial conditions $\mathbf{y}_{N,0}$. Our argument is strongly inspired by the proof of [30, Theorem 3.3(iii)], which we adapt to our more general hypotheses on the state and measure cost L and the setting of controllable linear time invariant systems.

Proof of Theorem 1.9. Let $\mathbf{y}_{N,0} \in \Omega^N$ be an arbitrary sequence s.t. $\sup_{i,N} |y_{i,N,0}| < +\infty$ and (1.9) holds. Suppose we can show that, for any sequence of minimizers $(\mathbf{y}_N, \mathbf{u}_N) \in \mathcal{A}_N(\mathbf{y}_{N,0})$ of $(\text{MFC}_{\varepsilon,\delta,N})$ we have

$$(3.21) \quad \limsup_{N \rightarrow +\infty} \mathcal{E}_{\varepsilon,\delta,N}(\mathbf{y}_N, \mathbf{u}_N) \leq e_{m_0} < +\infty,$$

with e_{m_0} as in equation (B.1). Then, up to a subsequence, we must have

$$\sup_N \frac{1}{N} \sum_{i=1}^N \int_0^1 \psi(u_{i,N}(t)) dt < +\infty.$$

Thus [30, Theorem 3.1] ensures there exists $(\mu, \nu) \in C(m_0)$ so that, up to a subsequence, (1.10)-(1.11) hold. Furthermore, for any such limit point (μ, ν) , Proposition 3.6(i) ensures

$$(3.22) \quad e_{m_0} \leq \mathcal{E}_{\varepsilon,\delta}(\mu, \nu) \leq \liminf_{N \rightarrow +\infty} \mathcal{E}_{\varepsilon,\delta,N}(\mathbf{y}_N, \mathbf{u}_N) \leq e_{m_0}.$$

Thus, (μ, ν) is a minimizer of $(\text{MFC}_{\varepsilon,\delta})$.

In this way, it suffices to show (3.21). Let $\tilde{\mathbf{y}}_{N,0}$ be the sequence of initial conditions from Theorem 3.7. Since m_0 is compactly supported, we have $\sup_{i,N} |\tilde{y}_{i,N,0}| < +\infty$. Let

$$R = \max\{\sup_{i,N} |y_{i,N,0}|, \sup_{i,N} |\tilde{y}_{i,N,0}|\}.$$

Fix $(\tilde{\mathbf{y}}_N, \tilde{\mathbf{u}}_N) \in \mathcal{A}(\tilde{\mathbf{y}}_{N,0})$ that minimize $(\text{MFC}_{\varepsilon,\delta,N})$. Theorem 3.7 ensures that

$$(3.23) \quad \lim_{N \rightarrow +\infty} \mathcal{E}_{\varepsilon,\delta,N}(\tilde{\mathbf{y}}_N, \tilde{\mathbf{u}}_N) = e_{m_0},$$

and there exists $(\tilde{\mu}, \tilde{\nu}) \in C(m_0)$ so that, up to a subsequence, (3.19-3.20) hold. For the remainder of the proof, we work with this subsequence of N .

Define

$$(3.24) \quad \zeta(r) := \phi(r) + r.$$

By Lemma 2.2, $\zeta(r) \leq 2\phi(r) + R_\phi$. Thus, by [30, Lemma 2.5],

$$\lim_{n \rightarrow +\infty} C_\zeta \left(\frac{1}{N} \sum_{i=1}^N \delta_{y_{i,N,0}}, \frac{1}{N} \sum_{i=1}^N \delta_{\tilde{y}_{i,N,0}} \right) = 0,$$

where $C_\zeta(\mu, \nu)$ denotes the optimal value of the optimal transport problem between μ and ν with cost matrix $\zeta(x - y)$. By Birkhoff's theorem, up to a permutation of the indices $i \mapsto y_{i,N}$, we may assume that

$$c_N := C_\zeta \left(\frac{1}{N} \sum_{i=1}^N \delta_{y_{i,N,0}}, \frac{1}{N} \sum_{i=1}^N \delta_{\tilde{y}_{i,N,0}} \right) = \frac{1}{N} \sum_{i=1}^N \zeta(|y_{i,N,0} - \tilde{y}_{i,N,0}|) \text{ for all } N.$$

We now seek to make a small modification to $(\tilde{\mathbf{y}}_N, \tilde{\mathbf{u}}_N)$ so that the initial condition agrees with $\mathbf{y}_{N,0}$ and, as $N \rightarrow +\infty$, the value of the discrete energy along the modified sequence converges to e_{m_0} . Our modification takes different forms, depending on the structural assumptions we consider. Fix $\eta \in (0, 1)$, and recall that ϕ is increasing, convex, and satisfies the doubling condition (2.4). Note that, for any $r, C \geq 0$,

$$\phi(r + C) \leq \begin{cases} \phi(2r) & \text{if } r \geq C \\ \phi(2C) & \text{if } r \leq C \end{cases} \leq \phi(2r) + \phi(2C).$$

In case (a), in which $U = \mathbb{R}^d$, define

$$\begin{aligned} y_{i,N,\eta}^{(a)}(t) &= \begin{cases} (1-t/\eta)y_{i,N,0} + (t/\eta)\tilde{y}_{i,N,0} & \text{if } t \in [0, \eta), \\ \tilde{y}_{i,N}(t-\eta) & \text{if } t \in [\eta, 1], \end{cases} \\ u_{i,N,\eta}^{(a)}(t) &= \begin{cases} (y_{i,N,0} - \tilde{y}_{i,N,0})/\eta - \mathbf{F}_N(y_{i,N,\eta}^{(a)}(t), \mathbf{y}_{N,\eta}^{(a)}(t)) & \text{if } t \in [0, \eta), \\ \tilde{u}_{i,N}(t-\eta) & \text{if } t \in [\eta, 1]. \end{cases} \end{aligned}$$

By construction, we have $(\mathbf{y}_{N,\eta}^{(a)}, \mathbf{u}_{N,\eta}^{(a)}) \in A(\mathbf{y}_{N,0})$. Furthermore, there exists $R > 0$ so that, for all $t \in [0, \eta)$, $\eta \in (0, 1)$,

$$(3.25) \quad \sup_{i,N} |y_{i,N,\eta}^{(a)}(t)| \leq R < +\infty.$$

Thus, the convergence of $\mathbf{F}_N(x, \mathbf{y}_{N,\eta})$ to $\mathbf{F}(x, m_0)$ as $N \rightarrow +\infty$ locally uniformly in x and t , implies there exists $C_F > 0$ so that

$$\sup_{t \in [0, \eta), i, N} |\mathbf{F}_N(y_{i,N,\eta}^{(a)}(t), \mathbf{y}_{N,\eta}^{(a)}(t))| < C_F.$$

Now, we estimate the state and measure cost in case (a) for $t \in [0, \eta)$. If L and L_N satisfy Assumptions 1.1(iia) and 1.6 (ia), then the convergence of $L_N(x, \mathbf{y}_{N,\eta})$ to $L(x, m_0)$ as $N \rightarrow +\infty$ locally uniformly in x and t , and the estimate (3.25) implies

$$\sup_{t \in [0, \eta), i, N} |L_N(y_{i,N,\eta}^{(a)}(t), \mathbf{y}_{N,\eta}^{(a)}(t))| < +\infty.$$

On the other hand, if L and L_N satisfy Assumptions 1.1(iiib) and 1.6(ib),

$$\sup_{t \in [0, \eta), i, N} L_N(y_{i,N,\eta}^{(a)}(t), \mathbf{y}_{N,\eta}^{(a)}(t)) \leq \sup_{t \in [0, \eta), i, N} L(y_{i,N,\eta}^{(a)}(t)) < +\infty,$$

where we use that $\{y_{i,N,\eta}^{(a)}(t)\}_{t \in [0, \eta), i, N}$ is a bounded subset of $\Omega \subseteq \{L < +\infty\}$, so the image of this set under L is also bounded. Thus, under either assumption on L , there exists $C_L > 0$ so that

$$\frac{1}{N} \sum_{i=1}^N \int_0^\eta L_N(y_{i,N,\eta}^{(a)}(t), \mathbf{y}_{N,\eta}^{(a)}(t)) dt \leq \eta C_L.$$

Finally, we estimate the control cost in case (a), estimating first in terms of the admissible function ϕ . For $t \in [0, \eta)$,

$$\begin{aligned} \phi\left(\left|u_{i,N,\eta}^{(a)}(t)\right|\right) &= \phi\left(\left|(y_{i,N,0} - \tilde{y}_{i,N,0})/\eta - \mathbf{F}_N(y_{i,N,\eta}^{(a)}(t), \mathbf{y}_{N,\eta}^{(a)}(t))\right|\right) \\ &\leq \phi\left(\left|(y_{i,N,0} - \tilde{y}_{i,N,0})/\eta\right| + C_F\right) \\ &\leq \phi\left(2\left|(y_{i,N,0} - \tilde{y}_{i,N,0})/\eta\right|\right) + \phi(2C_F) \end{aligned}$$

Now, we turn to case (b). In this case, define

$$\begin{aligned} y_{i,N,\eta}^{(b)}(t) &= \begin{cases} e^{tA}y_{i,N,0} + \int_0^t e^{-\tau A}u_{i,N,\eta}^{(b)}(\tau)d\tau & \text{if } t \in [0, \eta), \\ \tilde{y}_{i,N}(t-\eta) & \text{if } t \in [\eta, 1], \end{cases} \\ u_{i,N,\eta}^{(b)}(t) &= \begin{cases} -BB^T e^{-tA^T} \Gamma^{-1}(\eta)[y_{i,N,0} - e^{-\eta A}\tilde{y}_{i,N,0}] & \text{if } t \in [0, \eta), \\ \tilde{u}_{i,N}(t-\eta) & \text{if } t \in [\eta, 1]. \end{cases} \end{aligned}$$

Again, by construction, we have $(\mathbf{y}_{N,\eta}^{(b)}, \mathbf{u}_{N,\eta}^{(b)}) \in A(\mathbf{y}_{N,0})$. (See, for example, [8, Theorem 5.2].) Likewise, in case (b), for $t \in [0, \eta]$,

$$\begin{aligned} & \phi\left(\left|u_{i,N,\eta}^{(b)}(t)\right|\right) \\ &= \phi\left(\left|-BB^T e^{-tA^T} \Gamma^{-1}(\eta)[y_{i,N,0} - e^{-\eta A} \tilde{y}_{i,N,0}]\right|\right) \\ &\leq \phi\left(\left|BB^T e^{-tA^T} \Gamma^{-1}(\eta)[y_{i,N,0} - \tilde{y}_{i,N,0}]\right| + \left|BB^T e^{-tA^T} \Gamma^{-1}(\eta)[I - e^{-\eta A}]\tilde{y}_{i,N,0}\right|\right) \\ &\leq \phi\left(2\left|BB^T e^{-tA^T} \Gamma^{-1}(\eta)[y_{i,N,0} - \tilde{y}_{i,N,0}]\right|\right) + \phi\left(2\left|BB^T e^{-tA^T} \Gamma^{-1}(\eta)[I - e^{-\eta A}]\tilde{y}_{i,N,0}\right|\right) \\ &\leq \phi(C_{A,B}\|\Gamma^{-1}(\eta)\| |y_{i,N,0} - \tilde{y}_{i,N,0}|) + \phi(C_{A,B}|\Gamma^{-1}(\eta)(I - e^{-\eta A})\tilde{y}_{i,N,0}|) \end{aligned}$$

for some $C_{A,B} > 0$, depending on A and B .

Since $\|\Gamma^{-1}(\eta)\|$ is a continuous function of $\eta > 0$ that diverges to $+\infty$ as $\eta \rightarrow 0$, there exists a decreasing function $\sigma : (0, 1) \rightarrow \mathbb{R}$ so that $\sigma(\eta) \geq \frac{1}{\eta}$ and $\sigma(\eta) \geq \|\Gamma^{-1}(\eta)\|$ for all $\eta \in (0, 1)$. Define

$$\omega(\eta, \tilde{y}_{i,N,0}) = \begin{cases} 0 & \text{in case (a),} \\ C_{A,B}|\Gamma^{-1}(\eta)(I - e^{-\eta A})\tilde{y}_{i,N,0}| & \text{in case (b).} \end{cases}$$

Therefore, considering both case (a) and case (b) simultaneously, there exists $C' \geq 1$ so that, for $K' > 0$ as in Lemma 2.2 and $\zeta(|x|) = \phi(|x|) + |x|$,

$$\begin{aligned} \phi\left(\left|u_{i,N,\eta}(t)\right|\right) &\leq \phi(C'\sigma(\eta) |y_{i,N,0} - \tilde{y}_{i,N,0}|) + \phi(\omega(\eta, \tilde{y}_{i,N,0})) + C' \\ &\leq K'(C'\sigma(\eta))^{K'} \zeta(|y_{i,N,0} - \tilde{y}_{i,N,0}|) + \phi(\omega(\eta, \tilde{y}_{i,N,0})) + C' \end{aligned}$$

Combining this estimate with inequality (2.5), there exists $C'' > 0$ depending on K' , C' , and C so that

$$\begin{aligned} & \frac{1}{N} \sum_{i=1}^N \int_0^\eta \psi(u_{i,N,\eta}(t)) dt \\ &\leq \eta C + \frac{C}{N} \sum_{i=1}^N \int_0^\eta \phi(|u_{i,N,\eta}(t)|) dt \\ &\leq \eta C'' + C'' \eta (\sigma(\eta))^{K'} \left(\frac{1}{N} \sum_{i=1}^N \zeta(|y_{i,N,0} - \tilde{y}_{i,N,0}|) \right) + \eta C'' \frac{1}{N} \sum_{i=1}^N \phi(\omega(\eta, \tilde{y}_{i,N,0})) \\ &= C'' \eta (\sigma(\eta))^{K'} c_N + o(1), \text{ as } \eta \rightarrow 0, \end{aligned}$$

where we use the fact that, in case (a), $\omega \equiv 0$, and in case (b), hypothesis (biv) ensures that

$$\begin{aligned} \eta \phi(\omega(\eta, \tilde{y}_{i,N,0})) &\leq \eta K'(C_{A,B})^{K'} \zeta(|\Gamma^{-1}(\eta)(I - e^{-\eta A})\tilde{y}_{i,N,0}|) \\ &\leq \eta K'(C_{A,B})^{K'} (2\phi(|\Gamma^{-1}(\eta)(I - e^{-\eta A})\tilde{y}_{i,N,0}|) + R_\phi) \rightarrow 0, \end{aligned}$$

uniformly in i, N .

Finally note that, by definition of $y_{i,N,\eta}$ in both case (a) and case (b),

$$\begin{aligned} & \mathcal{F}_{\varepsilon,\delta} \left(\frac{1}{N} \sum_{i=1}^N \delta_{y_{i,N,\eta}(1)} \right) - \mathcal{F}_{\varepsilon,\delta} \left(\frac{1}{N} \sum_{i=1}^N \delta_{\tilde{y}_{i,N}(1)} \right) \\ &= \mathcal{F}_{\varepsilon,\delta} \left(\frac{1}{N} \sum_{i=1}^N \delta_{\tilde{y}_{i,N}(1-\eta)} \right) - \mathcal{F}_{\varepsilon,\delta} \left(\frac{1}{N} \sum_{i=1}^N \delta_{\tilde{y}_{i,N}(1)} \right). \end{aligned}$$

By Lemma 3.5, this vanishes as $N \rightarrow +\infty$ and $\eta \rightarrow 0$, since the empirical measures in both arguments converge to $\tilde{\mu}_1$.

Combining the above estimates, we have shown that, in both case (a) and case (b),

$$\begin{aligned} & \mathcal{E}_{\varepsilon,\delta,N}(\mathbf{y}_{N,\eta}, \mathbf{u}_{N,\eta}) \\ & \leq \frac{1}{N} \sum_{i=1}^N \int_0^\eta \psi(\mathbf{u}_{N,\eta}(t)) dt + \frac{1}{N} \sum_{i=1}^N \int_0^\eta L_N(y_{i,N,\eta}(t), \mathbf{y}_{N,\eta}(t)) dt + \mathcal{E}_{\varepsilon,\delta,N}(\tilde{\mathbf{y}}_N, \tilde{\mathbf{u}}_N) \\ & \quad + \mathcal{F}_{\varepsilon,\delta} \left(\frac{1}{N} \sum_{i=1}^N \delta_{y_{i,N,\eta}(1)} \right) - \mathcal{F}_{\varepsilon,\delta} \left(\frac{1}{N} \sum_{i=1}^N \delta_{\tilde{y}_{i,N}(1)} \right) \\ & \leq C'' \eta (\sigma(\eta))^{K'} c_N + \mathcal{E}_{\varepsilon,\delta,N}(\tilde{\mathbf{y}}_N, \tilde{\mathbf{u}}_N) + o(1), \text{ as } \eta \rightarrow 0, N \rightarrow +\infty. \end{aligned}$$

Taking $\eta = \sigma^{-1}(c_N^{-1/K'})$ implies $\sigma(\eta) = c_N^{-1/K'}$ and $\eta(\sigma(\eta))^{K'} c_N = \eta \xrightarrow{N \rightarrow +\infty} 0$. Thus,

$$\limsup_{N \rightarrow +\infty} \mathcal{E}_{\varepsilon,\delta,N}(\mathbf{y}_{N,\eta}, \mathbf{u}_{N,\eta}) \leq \limsup_{N \rightarrow +\infty} \mathcal{E}_{\varepsilon,\delta,N}(\tilde{\mathbf{y}}_N, \tilde{\mathbf{u}}_N) = e_{m_0}$$

Finally, since $(\mathbf{y}_{N,\eta}, \mathbf{u}_{N,\eta}) \in \mathcal{A}(\mathbf{y}_{N,0})$ for all $\eta \in (0, 1)$, for any sequence of minimizers $(\mathbf{y}_N, \mathbf{u}_N) \in \mathcal{A}(\mathbf{y}_{N,0})$ of $(\text{MFC}_{\varepsilon,\delta,N})$, we have

$$\limsup_{N \rightarrow +\infty} \mathcal{E}_{\varepsilon,\delta,N}(\mathbf{y}_N, \mathbf{u}_N) \leq \limsup_{N \rightarrow +\infty} \mathcal{E}_{\varepsilon,\delta,N}(\mathbf{y}_{N,\eta}, \mathbf{u}_{N,\eta}) \leq e_{m_0}.$$

This shows (3.21), which completes the proof. \square

Our main theorem, Theorem 1.11, is now an immediate consequence of Theorem 1.2, Theorem 1.5, and Theorem 1.9, via a standard diagonal argument. See appendix B.

4. Numerics.

4.1. Numerical Implementation. We now apply the particle approximation developed in the previous sections to compute three key examples of the mean field optimal control problem (\star) : the dynamic formulation of the 2-Wasserstein distance, 2-Wasserstein optimal transport around obstacles, and measure transport with acceleration constraints; see section 1.2 for more details. For simplicity of exposition, we describe our approach in the context of the first two 2-Wasserstein examples. See section 4.6 for the case of acceleration controls.

For classical $p = 2$ optimal transport, with and without obstacles, the particle discretization of the mean field optimal control problem $(\text{MFC}_{\varepsilon,\delta,N})$ becomes

$$(4.1) \quad \inf_{y_i, u_i} \left\{ \frac{1}{N} \sum_{i=1}^N \int_0^1 |u_i(t)|^2 dt + \frac{1}{N} \sum_{i=1}^N \int_0^1 L_N(y_i(t)) dt + \frac{1}{\varepsilon} \left[\frac{1}{N^2} \sum_{i,k=1}^N K_\delta(y_i(1) - y_k(1)) - \frac{2}{N} \sum_{i=1}^N (K_\delta * m_1)(y_i(1)) \right] \right\}$$

subject to the differential equations

$$(4.2) \quad \dot{y}_i(t) = u_i(t), \quad y_i(0) = y_{i,0}.$$

(Note that we neglect the constant C_{δ, m_1} in the objective functional, since it does not affect minimizers.) In order to numerically approximate solutions of this constrained optimization problem, we must discretize time, incorporate the ODE constraints, and develop a method for approximating the minimizer.

First, we discretize time $t \in [0, 1]$ on a uniform grid with M grid points and time step $h = 1/(M - 1)$, approximating $y_i(t)$ by the vector

$$(4.3) \quad [y_{i,j}]_{j=1}^M \approx [y_{i,0}, y_i(h), y_i(2h), \dots, y_i(1-h), y_i(1)],$$

which, by definition, incorporates initial condition constraint (4.2). Next, we approximate the velocity by a first order finite difference,

$$(4.4) \quad u_i(t) \approx \frac{y_i(t+h) - y_i(t)}{h}.$$

Substituting into the objective function (4.1), we arrive at a fully discrete problem, which is to minimize the sum of the kinetic energy, the potential energy, and the nonlocal energy,

$$(4.5) \quad \min_{y_{i,j}} \{ \text{KE}(y_{i,j}) + \text{PE}(y_{i,j}) + \text{NE}(y_{i,j}) : y_{i,1} = y_{i,0} \forall i = 1, \dots, N \},$$

for

$$(4.6) \quad \text{KE}(y_{i,j}) := \frac{M-1}{N} \sum_{i=1}^N \left| \text{diff}(y_{i,j}) \right|^2$$

$$(4.7) \quad \text{PE}(y_{i,j}) := \frac{1}{N(M-1)} \sum_{i=1}^N \sum_{j=1}^M L_N(y_{i,j})$$

$$(4.8) \quad \text{NE}(y_{i,j}) := \frac{1}{\varepsilon} \left[\frac{1}{N^2} \sum_{i,k=1}^N K_\delta(y_{i,M} - y_{k,M}) - \frac{2}{N} \sum_{i=1}^N (K_\delta * m_1)(y_{i,M}) \right],$$

where $i = 1, \dots, N$ and $j = 2, \dots, M$, with

$$\text{diff}(y_{i,j}) := [y_{i,2} - y_{i,1}, y_{i,3} - y_{i,2}, \dots, y_{i,M} - y_{i,M-1}, 0].$$

The resulting optimization problem is unconstrained. In what follows, we commit a mild abuse of notation and let $y_i(t)$ denote the linear interpolation of $\{y_{i,j}\}_{j=1}^M$ in time. In the majority of the simulations that follow, we consider the case without obstacles $L_N \equiv 0$; section 4.5 considers the optimal trajectories in the presence of obstacles.

In the present simulations, we choose $k_\delta, K_\delta : \mathbb{R}^d \rightarrow \mathbb{R}$ to be Gaussian mollifiers,

$$(4.9) \quad K_\delta(\theta) = (2\pi\delta^2)^{-d/2} \exp(-\|\theta\|^2/2\delta^2), \quad k_\delta = K_\delta/\sqrt{2}.$$

REMARK 4.1 (target measures $m_1 \notin L^2(\mathbb{R}^d)$). *While our theoretical convergence results require $m_1 \in L^2(\mathbb{R}^d)$ (see Assumption 1.1), our numerical approach extends naturally to empirical target measures, $m_1 = \frac{1}{N} \sum_{i=1}^N \delta_{w_i}$. In this case, NE becomes*

$$(4.10) \quad \text{NE}(y_{i,j}) = \frac{1}{\varepsilon N^2} \sum_{i,k=1}^N [K_\delta(y_{i,M} - y_{k,M}) - 2K_\delta(y_{i,M} - w_k)].$$

Finally, once we have arrived at the fully discrete minimization problem (4.5), we compute an approximate minimizer by gradient descent. With this approach, updating the trajectory of the i th particle only requires local information at the terminal time regarding the proximity to other particles and the value of the target m_1 . Derivatives are calculated automatically by PyTorch [40], with a prescribed learning rate α and maximum number of descent steps n . In addition, we use standard learning rate reduction and early stopping mechanisms. If the objective function value does not decrease for two steps, the learning rate α is reduced to 0.2α , as long as the reduction is larger than 10^{-8} , as implemented in PyTorch. If the objective function value does not decrease for 5 steps, the algorithm is terminated early.

As we illustrate in our simulations (see section 4.7), we do not expect the loss landscape to be convex, so we do not have guarantees that our gradient descent approach will converge to a global minimizer of (4.5). However, as our numerical results attest, even our simple gradient descent approach leads to reasonable results. We leave the important question of developing more accurate methods for computing an approximate minimizer to future work, as the main focus of the present paper is to analyze the effects of discretizing dynamic optimal transport by a regularized particle method.

Due to the nonconvexity of the loss landscape, we expect that the results of gradient descent will depend strongly on the initialization. In the following simulations, we initialize the trajectories to be straight lines terminating at the center of mass of the target distribution. In other words, we choose $y_{i,j}$ so that

$$(4.11) \quad y_{i,M} = \int y m_1(y) dy \text{ for all } i = 1, \dots, N,$$

$$(4.12) \quad [y_{i,j}]_{j=2}^{M-1} \text{ linearly interpolates between } y_{i,0} \text{ and } y_{i,M} \text{ for all } i = 1, \dots, N.$$

Python code for all experiments is available at <https://github.com/HarlinLee/BlobOT>.

4.2. Parameter selection. In practice, we expect the gradient descent dynamics for computing an approximate minimizer of (4.5) to depend on the choice of the parameters $N, \delta, \varepsilon, M, \alpha$ and n . For the continuum formulation of the problem, there is no downside to choosing ε to be arbitrarily small, independent of N and δ , since it simply enforces the terminal constraint more exactly. For this reason, in our simulations, we typically choose ε to be a fixed, very small parameter.

On the other hand, we cannot hope for good results by choosing δ arbitrarily small without regard to N . When δ is very small, the ‘‘sensing radius’’ of the mollifier K_δ becomes very small, preventing particles from detecting the correct target distribution, unless they were coincidentally initialized very close to target. For example, when the target distribution is an empirical measure, particles would not be able to detect desired target locations w_i too far from their current terminal location $y_{i,M}$ because the small radius of concentration of K_δ . In a similar way, if one chooses N arbitrarily large without regard to δ , assuming that the empirical measure $\frac{1}{N} \sum_{i=1}^N \delta_{y_{i,M}}$ narrowly converges to a limiting probability measure μ_1 the nonlocal energy approximates

$$\text{NE}(y_{i,j}) \approx F_{\varepsilon,\delta}(\mu_1) = \frac{1}{\varepsilon} \|k_\delta * \mu_1 - k_\delta * m_1\|_{L^2}^2.$$

It is a classical result that there exist μ_1, m_1 arbitrarily far apart for which the L^2 differences of their regularizations is arbitrarily small, so in this way, when $\delta > 0$ is fixed too large with respect to N , $\text{NE}(y_{i,j})$ is not as accurate for imposing the terminal constraint.

For these reasons, we allow $\delta \rightarrow 0$ and $N \rightarrow +\infty$ simultaneously. Inspired by quantitative error estimates available for classical vortex blob methods for the Euler and Navier-Stokes

equations [9, 10] and blob methods for the aggregation equation [23], we let δ scale with N ,

$$(4.13) \quad \delta = N^{-k/d}, \quad \text{for } 0 \ll k < 1.$$

In the present manuscript, we take $k = 0.99$. As a consequence, two particles distance $O(N^{-1/d})$ apart in d -dimensions will be able to “sense” one another via the mollifier K_δ . In particular, when K_δ is a Gaussian mollifier, equation (4.13) ensures its standard deviation is slightly larger than the interparticle distance on a regular grid.

With regard to the number of time steps M , since we expect good regularity in time, we anticipate higher accuracy when M is large, at the expense of increasing the dimension of the optimization problem (4.5). Due to the fact that the optimal trajectories of the dynamic 2-Wasserstein problem will always follow straight lines, in the case $L_N \equiv 0$, it suffices to take $M = 2$. However, we allow for flexibility in choice of M , to accommodate more general formulations of the problem, including obstacles and acceleration constraints. In the numerical experiments that follow, we choose $M > 2$, even though this is an unnecessarily computational expense in the $L_N \equiv 0$ case, in order to illustrate that the kinetic energy KE on its own is able to effectively straighten the trajectories.

Finally, we consider the choice of parameters α and n for our gradient descent dynamics. As shown in section 4.7, when ε is large, the energy landscape flattens, but we can still achieve good results with a larger learning rate α . For this reason, we typically take α to scale with ε . As a consequence, when ε is small, the learning rate α is small, and it may take more steps n to reach an approximate global minimizer. In this way, while from the perspective of the continuum problem, there is no downside to taking ε arbitrarily small, from the perspective of the fully discrete problem, we do see that when ε is very small, it can require more iterations of gradient descent to reach an approximate global minimizer; see section 4.8.

4.3. Comparison with Python Optimal Transport. In Figure 4.1, we begin by comparing the results of Blob OT to classical approaches for solving the 2-Wasserstein optimal transport problem. In particular, we compare our method to the Python Optimal Transport emd (POT-emd) function, which computes optimal transport between two empirical measures via solving the Kantorovich formulation of the problem using the Hungarian algorithm [29]. As illustrated in the figure, our source and target measures are given by $N = 30$ equally weighted Dirac masses in $d = 2$ spatial dimensions. We choose $M = 3$ time steps, $\delta = N^{-0.99/d}$, $\varepsilon = 0.01$, learning rate $\alpha = 0.01$, and $n = 2000$ gradient descent steps. (See section 4.9 for a more detailed discussion of the relationship between the parameters δ , N , and ε and section 4.8 for a discussion of how these relate to the gradient descent parameters α and n .)

The top left panel of Figure 4.1 shows the trajectories $\{y_{i,j}\}$ computed by our method, linearly interpolating between the $M = 3$ time steps to better illustrate the paths of the particles. The top right panel compares the terminal values of the trajectories $\{y_{i,M}\}$ to the locations of particles in the target $\{w_i\}$. While our method uses a soft constraint to match source to target particles, we observe overall good agreement. The bottom left panel shows the optimal transport matching computed via POT-emd, which is qualitatively similar to our solution, though not identical. However, in the bottom right panel, we see by comparing the Blob OT and POT-emd solutions in terms of the value of the objective function (4.5), they are extremely close in terms of the degree to which they match source to target measure with the smallest possible kinetic energy. While the gradient descent of Blob-OT is initialized far from optimum, we see that, over the course of the gradient descent, it converges to the nearly optimal objective function value achieved by POT-emd. After $n = 2,000$ steps, Blob OT has objective function value of $1.9884 = 1.9779$ (kinetic energy, 4.6) + 0.0105 (nonlocal energy, 4.8), while POT-emd has objective function value of $1.9908 = 1.9908$ (kinetic energy, 4.6) + 0 (nonlocal energy, 4.8), rounded to 4 decimal points. Note that while the kinetic energy for the Blob OT

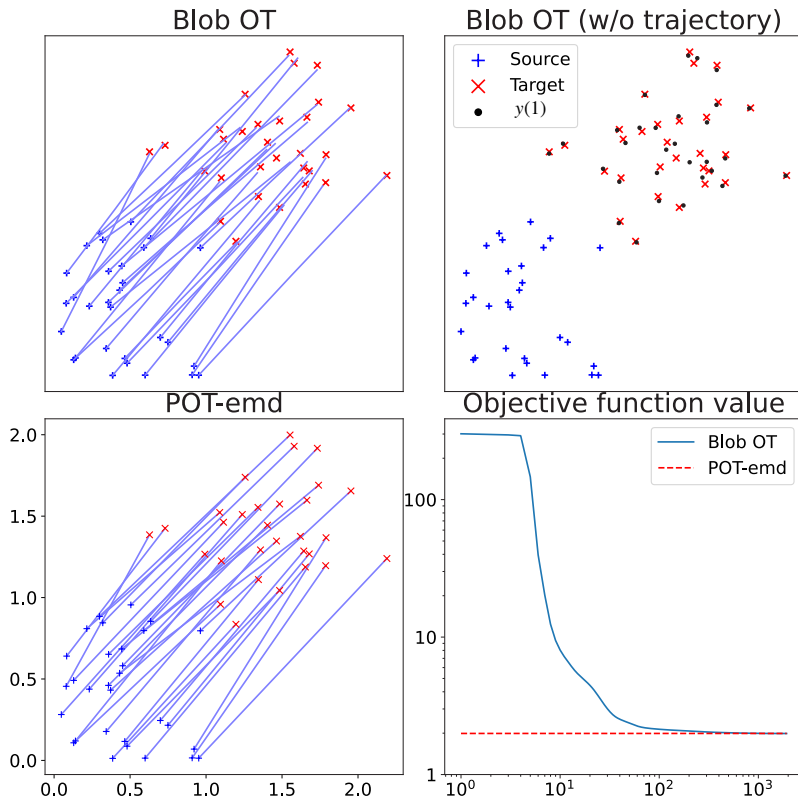


Fig. 4.1: Comparison between our particle method for computing the dynamic formulation of the 2-Wasserstein distance and classical approaches to computing optimal transport, e.g., via the Kantorovich formulation of the problem, as implemented in Python Optimal Transport emd (POT-emd).

is smaller than POT-emd, this is due to the fact that Blob OT has a soft terminal constraint, not because it has indeed found a superior matching between source and target.

4.4. Continuous target Gaussian distribution. In Figure 4.2, we contrast the behavior of our method for computing the 2-Wasserstein optimal transport when the target distribution is given by a continuum Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$, with $\mu = (1.5, 1.5)$, $\sigma = 0.5$, versus when the target is given by iid samples from the same Gaussian, in $d = 2$ dimensions. In both cases, we take our source distribution to be $N = 225$ equally weighted Dirac masses arranged on a grid, as an approximation of the uniform measure on the square $m_0 = 1_{[0,1] \times [0,1]}$. We take $M = 5$ time steps, $\varepsilon = 0.01$, $\delta = N^{-0.99/d}$, $n = 5 \cdot 10^6$ gradient descent steps, and learning rate $\alpha = 0.01\varepsilon$.

In the top row of Figure 4.2, we use our method to compute trajectories from the source measure to iid samples of the Gaussian, drawn using the numpy function `randn`. In the bottom row, we compute trajectories from the source measure to the continuum gaussian, taking $m_1(y) = \varphi_{0.5}(y - (1.5, 1.5))$ and using the approximation $K_\delta * m_1(y) = \varphi_{\sqrt{0.5+\delta^2}}(y - (1.5, 1.5)) \approx m_1(y)$ in the definition of the nonlocal energy, equation (4.8). The left column shows the trajectories $\{y_{i,j}\}$ computed by our method, linearly interpolating between time

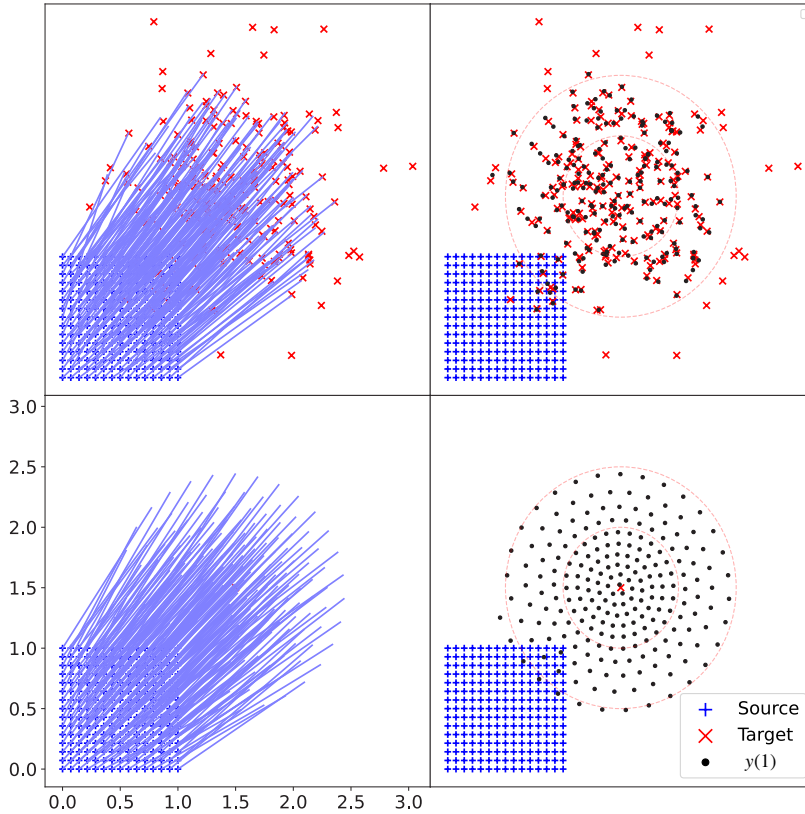


Fig. 4.2: Comparison between our particle method for computing the dynamic formulation of the 2-Wasserstein distance between a uniform grid and iid samples from a Gaussian (top row) and the continuum Gaussian itself (bottom row).

steps to show the path of each particle. The right column shows the terminal locations of the trajectories, which can be interpreted as samples of m_1 , and compares them to the target measure. The dotted circles around the Gaussian target measure illustrate its standard deviations, σ and 2σ .

In both the top and bottom row, we see that the particles primarily end up within two standard deviations of the Gaussian mean, largely ignoring outliers/tails. We believe this is due to the finite particle nature of our approximation, and we anticipate that, as $N \rightarrow +\infty$, $\delta \rightarrow 0$, and $\varepsilon \rightarrow 0$, minimizers of (4.5) will indeed match to more outliers/tails. While our method succeeds at capturing the irregular samples in the top row, we find it especially interesting that, when the source measure is taken to be a continuum Gaussian in the second row, the final particle locations self-organize with much more regular structure. For this reason, we believe that methods based on our approach may show promise within the context of sampling, especially when the samples are used to discretize integrals of smooth functions.

In order to illustrate the importance of the choice of parameters on behavior of our method, we illustrate how the behavior of the previous figure changes when δ is chosen too small. In Figure 4.3, we consider the same numerical experiment as in the previous Figure 4.2, except that instead of choosing $\delta = N^{-0.99/d}$, we choose $\delta = \sigma d^{-0.5} N^{-0.99/d}$. For this smaller value of δ , we observe that the particles are less able to match the outliers/tails of the target Gaussian.

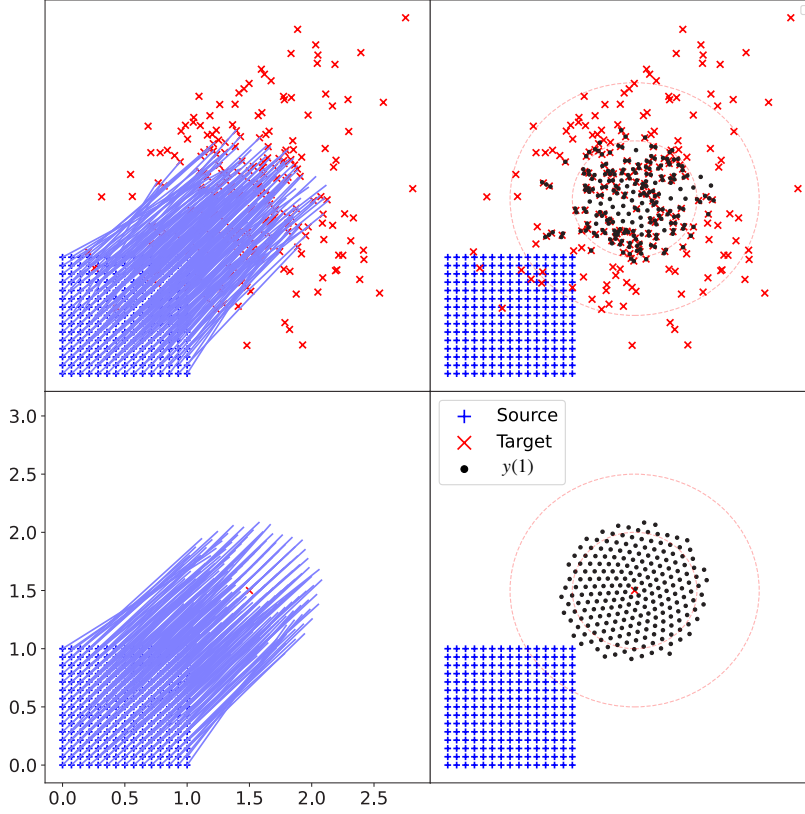


Fig. 4.3: Comparison between computing optimal transport from a uniform grid to iid samples from a Gaussian (top row) and the continuum Gaussian itself (bottom row) when the regularization parameter δ is intentionally chosen too small.

The reason for this is similar in both the iid sample case (top row) and the continuum Gaussian case (bottom row). In the case of iid samples, this is due to the fact that, in our objective function (4.5), the terminal locations of our particles $\{y_{i,M}\}$ are less able to sense the samples $\{w_i\}$ when δ is too small, since $y \mapsto \varphi_\delta(y - w_i)$ decays quickly away from w_i . Interestingly, in this case, when the terminal locations of our particles are unable to sense a sufficiently near sample, they organize themselves to be roughly evenly spaced in the empty regions.

In the case of a continuum Gaussian, $y \mapsto m_1(y)$ becomes very small outside two standard deviations of the mean. In both cases, the particle interactions mediated by the first term in the nonlocal energy $K_\delta(y_{i,M} - y_{j,M})$ become weak when δ is too much smaller than the distance between particles, preventing the particles from repelling each other and exploring more regions of parameter space.

4.5. Obstacles. Building on the preceding section, we now consider optimal transport in the presence of an obstacle Ω , which is given by the union of the interior of two circles of radius $r_k = 0.2$ centered at $c_1 = (1.0, 1.5)$ and $c_2 = (1.25, 1.25)$. We represent the obstacle in our optimization problem via the state and measure cost

$$(4.14) \quad L_N(y) = c_\Omega(N) \sum_{k=1}^2 \max\{r_k^2 - (y - c_k)^2, 0\}.$$

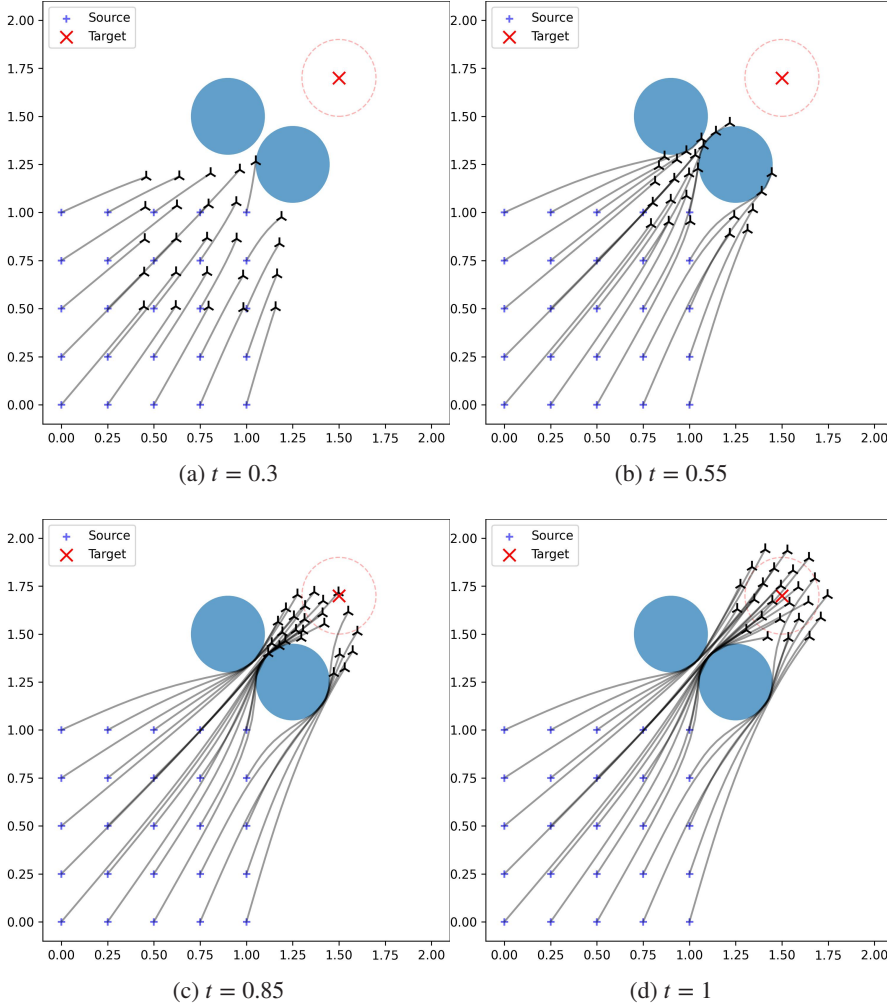


Fig. 4.4: 2-Wasserstein optimal transport from an empirical measure on a uniform grid to a continuum Gaussian distribution, in the presence of two circular obstacles.

for $c_\Omega(N)$ satisfying $\lim_{N \rightarrow +\infty} c_\Omega(N) = +\infty$. By definition, we see that $L_N \nearrow L$, where

$$L(y) = \begin{cases} 0 & \text{if } y \in \Omega^c, \\ +\infty & \text{if } y \in \Omega. \end{cases}$$

The source measure is given by $N = 25$ equally weighted Dirac masses arranged in a uniform grid on the unit square, $m_0 = 1_{[0,1] \times [0,1]}$. For the continuum gaussian target measure, we take $m_1(y) = \varphi_{0,2}(y - (1.5, 1.7))$ and use the approximation $K_\delta * m_1(y) \approx m_1(y)$ in the definition of the nonlocal energy, equation (4.8). We take $M = 21$ time steps, $\varepsilon = 0.1$, $\delta = N^{-0.99/d}$, $n = 2 \cdot 10^5$ gradient descent steps, and learning rate $\alpha = 0.001\varepsilon$. The obstacle constant is $c_\Omega(N) = (h\varepsilon)^{-1}$.

Figure 4.4 shows the evolution of the linear interpolations of the trajectories $y_i(t)$, $i =$

$1, \dots, N$ at times $t = 0.3, 0.55, 0.85$, and 1.0 . For the time continuous problem in the limit $\delta \rightarrow 0, \varepsilon \rightarrow 0$, the particles should follow straight lines, bending only to follow the boundaries of the obstacles and straightening again as they leave the obstacle and approach their terminal points. Overall, we observe good agreement between our numerical approximation and the continuum solution, with only mild bending close to the obstacle.

4.6. Acceleration control. We now consider the performance of our method in the case of measure transport subject to acceleration controls. We begin by describing how to discretize the time continuous approach, in analogy with our approach for the velocity control problem, described at the beginning of section 4.1.

In order to solve (MFC_a) , we apply the nonlocal terminal constraint and particle discretization $(\text{MFC}_{\varepsilon, \delta, N})$, to arrive at the continuous time formulation:

$$(4.15) \quad \inf_{x_i, v_i, a_i} \frac{1}{N} \sum_{i=1}^N \int_0^1 |a_i(t)|^2 dt \\ + \frac{1}{\varepsilon} \left[\frac{1}{N^2} \sum_{i,k=1}^N K_\delta((x_i(1), v_i(1)) - (x_k(1), v_k(1))) - \frac{2}{N} \sum_{i=1}^N (K_\delta * m_1)(x_i(1), v_i(1)) \right]$$

subject to the constraints

$$(4.16) \quad \begin{cases} \dot{x}_i(t) &= v_i(t) \\ x_i(0) &= x_{i,0} \end{cases} \quad \begin{cases} \dot{v}_i(t) &= a_i(t) \\ v_i(0) &= v_{i,0} \end{cases}$$

As before, we discretize time $t \in [0, 1]$ on a uniform grid with M grid points and time step $h = 1/(M-1)$, approximating $x_i(t)$ by the vector

$$(4.17) \quad [x_{i,j}]_{j=1}^M \approx [x_{i,0}, x_{i,0} + hv_{i,0}, x_i(2h), \dots, x_i(1-h), x_i(1)],$$

which, by definition, incorporates initial condition constraints (4.16). As before, we approximate the velocity and the acceleration by first order finite differences,

$$v_{i,j} := \frac{x_{i,j} - x_{i,j-1}}{h}, \quad a_{i,j} := \frac{v_{i,j} - v_{i,j-1}}{h}.$$

Substituting into the objective function (4.15), we arrive at the fully discrete problem, which is to minimize the sum of the control cost and the nonlocal position/velocity energy

$$\min_{x_{i,j}} \{ \text{CC}(x_{i,j}) + \text{NPVE}(x_{i,j}) : x_{i,1} = x_{i,0}, x_{i,2} = x_{i,0} + hv_{i,0} \forall i = 1, \dots, N \}$$

for

$$\text{CC}(x_{i,j}) := \frac{(M-1)^3}{N} \sum_{i=1}^N \left| \text{ddiff}(x_{i,j}) \right|^2$$

$$\text{NPVE}(x_{i,j}) :=$$

$$\frac{1}{\varepsilon} \left[\frac{1}{N^2} \sum_{i,k=1}^N K_\delta((x_{i,M}, v_{i,M}) - (x_{k,M}, v_{k,M})) - \frac{2}{N} \sum_{i=1}^N (K_\delta * m_1)(x_{i,M}, v_{i,M}) \right]$$

where $i = 1, \dots, N$ and $j = 2, \dots, M$, with

$$\text{ddiff}(x_{i,j}) := [x_{i,1} - 2x_{i,2} + x_{i,3}, x_{i,2} - 2x_{i,3} + x_{i,4}, \dots, x_{i,M-2} - 2x_{i,M-1} + x_{i,M}, 0, 0].$$

As in the velocity control case, while the continuum problem motivating the above problem is well-posed only when $m_1 \in L^2(\mathbb{R}^d \times \mathbb{R}^d)$, one particular case of interest is the case when the target measure m_1 is given by a sum of Dirac masses, $\frac{1}{N} \sum_{i=1}^N \delta_{(w_i^x, w_i^v)}$. This can be handled within our framework analogously to the previously described velocity control case.

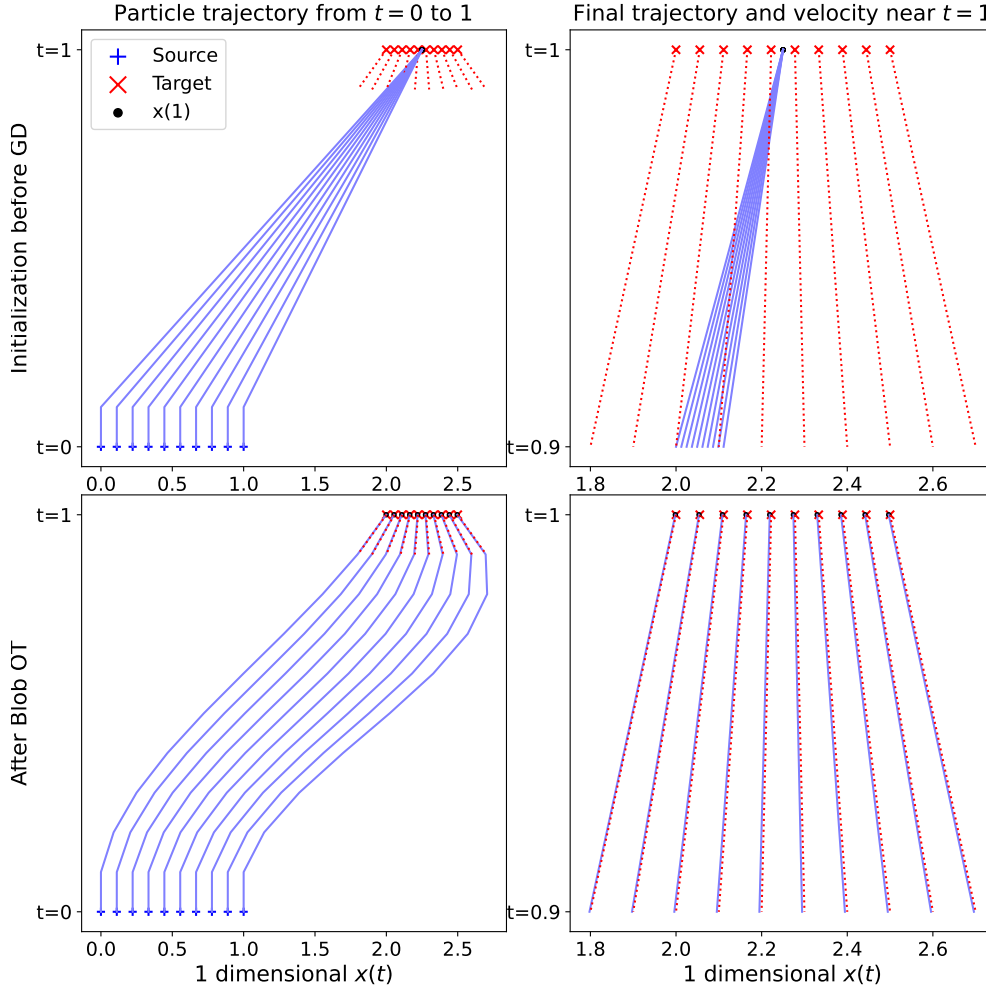


Fig. 4.5: Measure transport subject to acceleration control in one spatial dimension, between two well-ordered empirical distributions. In the top row, we show the initialization of our gradient descent, and in the bottom row, we show the approximate optimizer.

In Figure 4.5, we plot the results of an acceleration control optimization problem in one spatial dimension, in the case that the source distribution is an empirical measure with $N = 10$ particles evenly spaced on $[0, 1]$, with initial velocities zero, and the target distribution is an empirical measure with $N = 10$ particles evenly spaced on $[2, 2.5]$, with target velocities evenly spaced on $[-2, 2]$. Note that this example satisfies the condition of our main theorem that $\text{supp } m_0 \subseteq N(A)$. We consider $M = 11$ time steps, $\epsilon = 10^{-4}$, $\delta = N^{-0.99/2}$, $n = 10^6$ iterations, and $\alpha = 4\epsilon/10^3$ learning rate. (Note that the dimension in our choice of δ is two dimensional, since our mollifier $\varphi_\delta(x, v)$ is a function on $\mathbb{R} \times \mathbb{R}$.)

The top left panel of Figure 4.5 shows our initialization of the gradient descent $x_{i,j}$, linearly interpolating in time. The top right panel zooms in on this initialization and compares it to the desired target distribution. The bottom left panel shows the approximate minimizer $x_{i,j}$ obtained via gradient descent, again linearly interpolating in time, and the bottom right panel compares this to the desired target distribution. The approximate minimizer computed by our method exhibits trajectories with low acceleration and that agree with the desired target distribution in position and velocity.

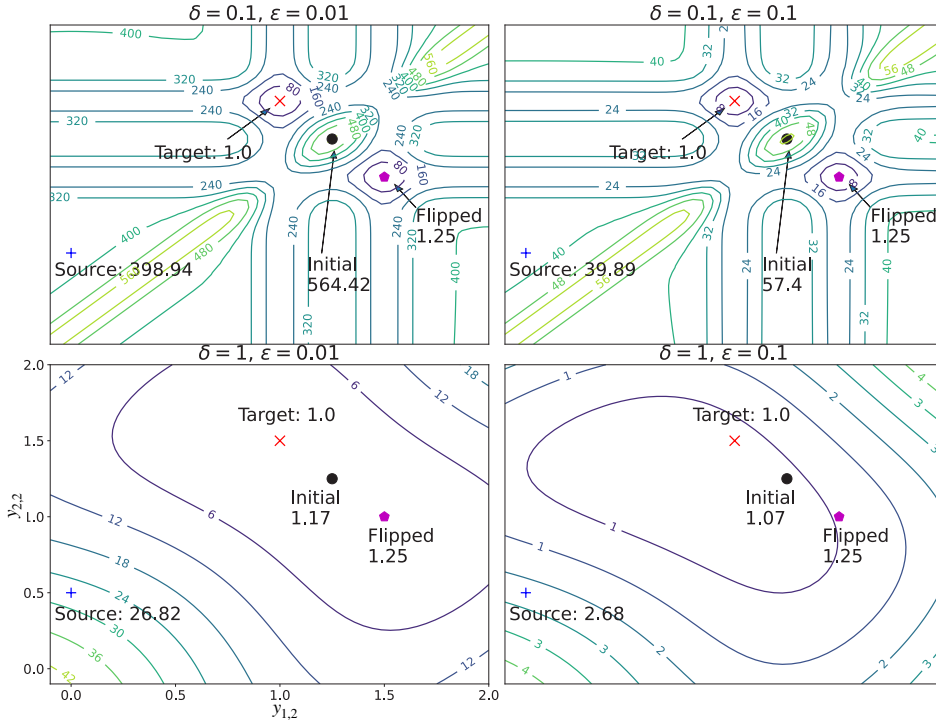


Fig. 4.6: Contour plots of nonconvex loss landscape for our optimization problem (4.5) between two Dirac masses, as a function of the terminal points $y_{1,2}$ and $y_{2,2}$, for different choices of ϵ and δ .

4.7. Illustrating the Loss Landscape. We now turn to an example illustrating basic properties of the loss landscape for our particle discretization of the optimal transport problem (4.5). For simplicity of visualization, we consider $L_N = 0$, $N = 2$ and $M = 2$, with one dimensional source distribution $\frac{1}{2}(\delta_0 + \delta_{0.5})$ and target distribution $\frac{1}{2}(\delta_1 + \delta_{1.5})$. In Figure 4.6, we plot the loss landscape as a function of $y_{1,2}$ and $y_{2,2}$, for different choices of δ and ϵ .

For reference, we have plotted the following values on the loss landscape:

- **Source distribution, +:** the value of the objective function if the particles stay at their initial locations $(y_{1,2}, y_{2,2}) = (y_{1,1}, y_{2,1}) = (0, 0.5)$; we observe that the objective function is large at the source distribution, reflecting the fact that the particles are far from the desired target distribution.
- **Target distribution, x:** the value of the objective function if the particles are optimally transported from source to target, with the leftmost particle in the source get-

ting mapped to the leftmost particle in the target $(y_{1,2}, y_{2,2}) = (1.0, 1.5)$; we observe that the objective function is minimized for this configuration.

- **Flipped distribution**, \blacklozenge : the value of the objective function if the particles in the source are exactly mapped to the particles in the target, but in a nonoptimal order, with the leftmost source particle mapped to the rightmost target particle $(y_{1,2}, y_{2,2}) = (1.5, 1.0)$; we observe that the value of the objective function is small, since the terminal configuration agrees exactly matches the target distribution, but not minimal, since the kinetic energy of achieving this configuration is not as small as possible.
- **Initial distribution**, \bullet : the value of the objective function at our initialization of gradient descent, as described in equations (4.11-4.12).

Note that the value of the objective function is the same at the “Target” and “Flipped” distributions in all four plots, due to the fact that, for both of these configurations, the value of the nonlocal energy term in the objective function (4.5) is zero for any δ and ε .

As anticipated, for all values of δ and ε , the loss landscape is nonconvex. Comparing the first row and second row of Figure 4.6, we observe that smaller values of δ (top row) better distinguish between small changes in $(y_{1,2}, y_{2,2})$, due to the fact that, when δ is small, there is less smoothing in our approximation of the source and target measures.

Comparing the first column and second column of Figure 4.6, we observe that smaller values of ε (left column) place more weight on the nonlocal energy term on the objective function (4.5), rewarding proximity to the target measure. We observe that the plots in the left column are highly symmetric across $y_{1,2} = y_{2,2}$, due to the fact that the objective function primarily considers the final locations of the particles, rather than whether the particles were optimally transported to those locations. On the other hand, larger values of ε (right column) place more weight on the kinetic energy term. In this case, we observe less symmetry across $y_{1,2} = y_{2,2}$, due to the fact that the kinetic energy term prioritizes moving particles a shorter distance from source to target.

Finally, we observe that larger values of δ (bottom row) and ε (right column) flatten the energy landscape. In this way, when using a gradient-based optimization method to compute minimizers of the objective function (4.5), an optimal choice of learning rate will depend on the choices of δ and ε . Since, in the present simulations, we typically choose δ very small, the size of ε is the main factor in the flatness of the energy landscape. For this reason, we choose our learning rate α to scale with ε , taking bigger steps on the flatter energy landscape and smaller steps on the steeper landscape.

4.8. Estimating the error from the optimal transport map. We now analyze the error between the approximate solution of the 2-Wasserstein optimal transport problem computed by our method and the exact optimal transport map. In particular, we examine how this error behaves along the gradient descent which computes the approximate optimizer. On one hand, we expect higher accuracy when ε is small, given that this more strongly imposes the terminal constraint via the nonlocal energy (4.5). On the other hand, due to the fact that we let the learning rate α depend on ε (see sections 4.2 and 4.7), when ε is small, we also have to run gradient descent for more iterations to compute our approximate minimizer. Consequently, in practice, one seeks a value of ε that is small enough to give accurate results but large enough to be computationally efficient.

In the following figures, we consider optimal transport in one dimension, where the source measure is given by $m_0 = 1_{[0,1]}$, discretized as $N = 20$ particles on a uniform grid on $[0, 1]$. The target measure is given by $m_1 = 2 \cdot 1_{[2,2.5]}$. In both cases, we take nonlocal regularization $\delta = N^{-0.99}$, $M = 5$ time steps, and learning rate $\alpha = \max(0.001\varepsilon, 10^{-5})$.

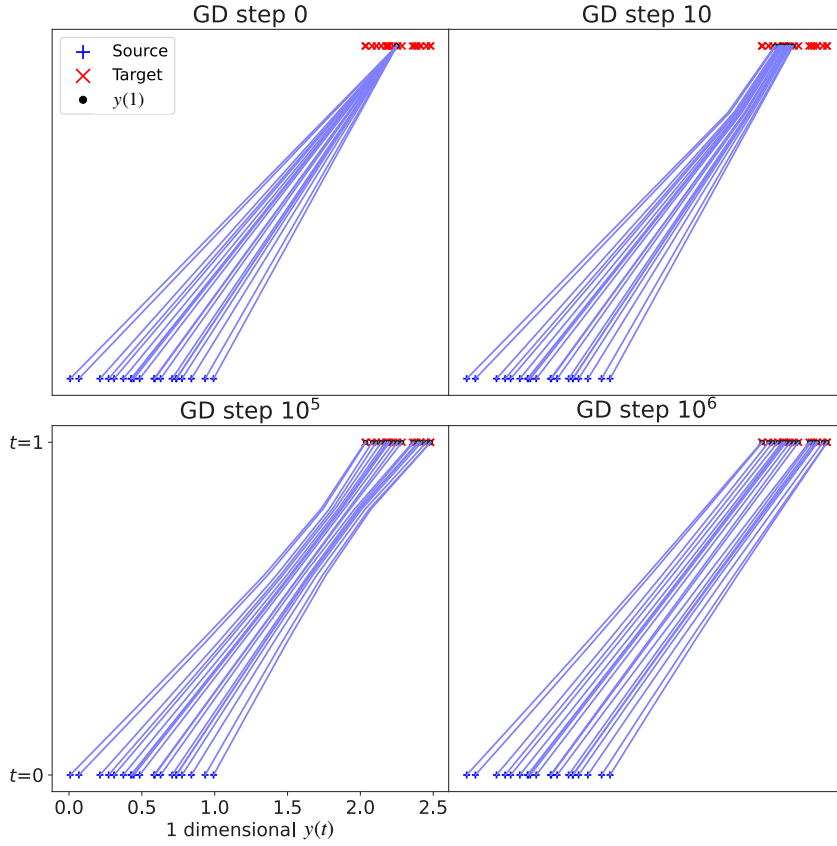


Fig. 4.7: Illustration of how trajectories change along gradient descent to compute an approximate minimizer of (4.5), when computing 2-Wasserstein optimal transport problem between two empirical measures.

At the continuum level, the 2-Wasserstein geodesic from m_0 to m_1 at time t is given by

$$(4.18) \quad T(y, t) := (1 - t)y + t(0.5y + 2).$$

Likewise, $T(y, 1)$ is the optimal transport map from m_0 to m_1 . We consider two measures of error: the average error across all time points in our discretization,

$$(4.19) \quad [\text{Error at all time points}] = \sqrt{\frac{1}{N(M-1)} \sum_{i=1}^N \sum_{j=2}^M \left| y_{i,j} - T\left(y_{i,1}, \frac{j-1}{M-1}\right) \right|^2},$$

and the error at terminal time $t = 1$,

$$(4.20) \quad [\text{Error at } t = 1] = \sqrt{\frac{1}{N} \sum_{i=1}^N |y_{i,M} - T(y_{i,1}, 1)|^2}.$$

We also analyze the decay of both the kinetic energy term (4.6) and the nonlocal energy term (4.10), as well as the total loss, given by their sum (4.5).

Figure 4.7 shows the gradient descent dynamics for a very small choice of $\varepsilon = 0.001$ and $n = 10^6$ gradient descent steps. The top left panel shows the initialization of the gradient descent $y_{i,j}$, the top right panel shows $y_{i,j}$ after 10 gradient descent steps, the bottom left shows the behavior after 10^5 steps, and the bottom right panel shows the behavior after 10^6 steps.

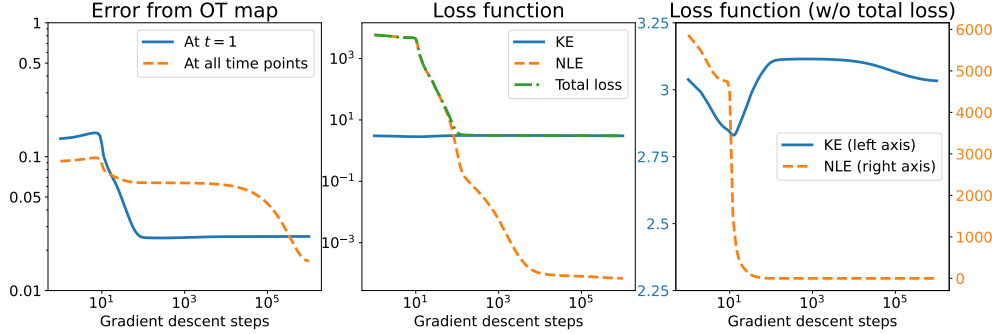


Fig. 4.8: Error analysis between the approximate solution of an optimal transport problem, computed by our method, and the exact solution of the optimal transport problem. Left: While the error at $t = 1$ quickly decays, many more gradient descent steps are required for the error at all time points to decay. Middle: Illustration of how the kinetic energy and nonlocal energy contribute to the total loss. Right: While the nonlocal energy quickly decreases to its optimal value, many more gradient descent steps are required for the kinetic energy to approach its optimal value. Note the dual axis.

A similar phenomena can be observed in Figure 4.8, which shows the behavior of the error and loss function along the gradient descent iterations. In the left plot, we see that both the error at terminal time and the error at all time points decay to zero, though the error at all time points requires a very large number of iterations to decay. This reflects the fact that the very small value of ε forces the trajectories to match the terminal points more strongly than it enforces the kinetic energy, which straightens the trajectories. In the middle plot, we observe the decay of both the kinetic energy and nonlocal energy along iterations. In the right plot, plotting the kinetic energy and nonlocal energy on different axes, we see again that the small value of ε causes the nonlocal energy to decay first, while the kinetic energy takes many more iterations to decay.

4.9. Numerical analysis of rate of convergence to continuum. We conclude by analyzing the rate of convergence to the continuum formulation of the optimal transport problem as $\varepsilon \rightarrow 0$, $\delta \rightarrow 0$, and $N \rightarrow +\infty$. We consider one dimensional source and target distributions given by $m_0 = 1_{[0,1]}$ and $m_1 = 2 \cdot 1_{[2,2.5]}$, discretized on a uniform grid with N particles, where we allow N to vary. We consider $\varepsilon = 0.01$, $M = 5$ time steps, $n = 2 \cdot 10^6$ gradient descent steps, and learning rate $\alpha = 0.003\varepsilon$. As explained in section (4.2), we allow $\delta \rightarrow 0$ and $N \rightarrow +\infty$ simultaneously according to the rate in equation (4.13), with $k = 0.99$.

In order to obtain numerical estimates on the rate of convergence as $N \rightarrow +\infty$ and $\delta \rightarrow 0$, we consider the error at the terminal time $t = 1$, see equation (4.20), using the fact that, in the present example the optimal transport map for the continuum problem is given by $T(y, 1) = y/2 + 2$.

The results of our numerical study are shown in Figure 4.9. Here, we plot how the error from the optimal transport map at time $t = 1$ varies as the number of particles N increases from 5, 10, 20, 40, 80, to 100. Comparing the decay of the error on a log-log scale to the line

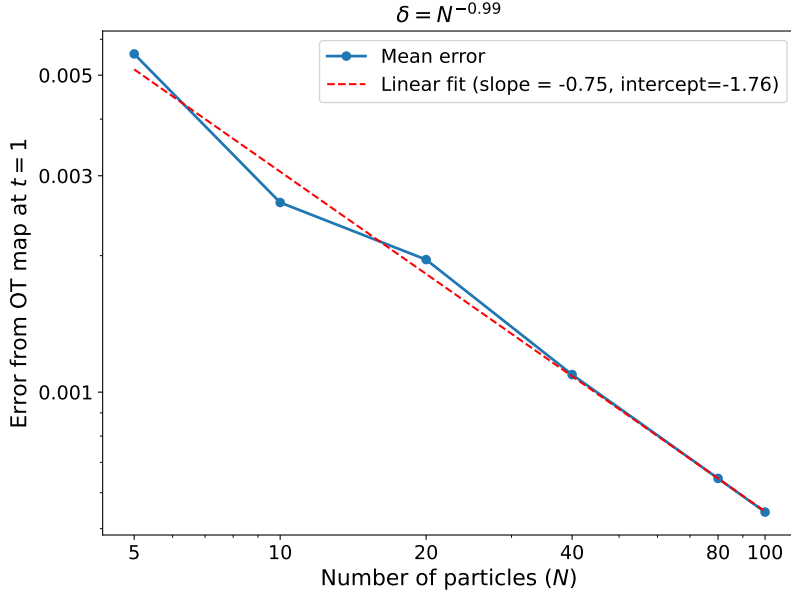


Fig. 4.9: Analysis of the rate of convergence of our particle approximation of 2-Wasserstein optimal transport to the exact solution in one dimension, as $N \rightarrow +\infty$ and $\delta = N^{-0.99} \rightarrow 0$. We observe slightly slower than first order convergence in N .

of best fit (calculated by minimizing least squares deviation as implemented in NumPy [33]), we observe slightly slower than first order convergence as $N \rightarrow +\infty$.

Appendix A. Basic Properties of Mean Field Control Problems. In this section, we collect several results and proofs regarding basic properties of mean field control problems. We begin by proving Lemma 2.2, on elementary properties of our control cost ψ and admissible function ϕ .

Proof of Lemma 2.2. Part (i) is an immediate consequence of the fact that ϕ is superlinear at $+\infty$ and inequality (2.5).

Now, we show part (ii). To begin, we show that there exists $K'' > 1$ so that

$$(A.1) \quad \phi(2r) + 2r \leq K''(\phi(r) + r), \quad \text{for all } r \in [0, +\infty).$$

For any ϕ satisfying Assumption 2.1, there exists r_0 so that $\phi(r_0) = 1$. Then, for all $r \geq r_0$, the fact that ϕ is increasing ensures

$$\phi(2r) \leq K(1 + \phi(r)) \leq 2K\phi(r) \implies \phi(2r) + 2r \leq 2K\phi(r) + 2r.$$

On the other hand, by the convexity of ϕ , $\phi(r) + \phi'(r)(s - r) \leq \phi(s)$ for all $r, s \in [0, +\infty)$. Thus, for all $r \leq r_0$,

$$\phi(2r) \leq \phi(r) + r\phi'(2r) \leq \phi(r) + r\phi'(2r_0) \implies \phi(2r) + 2r \leq \phi(r) + (2 + \phi'(2r_0))r.$$

Therefore, letting $K'' = \max\{2K, 2 + \phi'(2r_0)\}$ gives inequality (A.1).

To prove part (ii), note that, for any $D \geq 1$, choosing $n = \lceil \log_2(D) \rceil$,

$$\begin{aligned} \phi(Dr) + Dr &\leq \phi(2^n r) + 2^n r \leq (K'')^n (\phi(r) + r) \leq (K'')^{\log_2(D)+1} (\phi(r) + r) \\ &= K'' D^{\log_2(K'')} (\phi(r) + r). \end{aligned} \quad \square$$

Now, we prove Lemma 2.3 on the equivalence of the original formulation of the mean field control problem (\star) and the formulation in momentum coordinates (MFC).

Proof of Lemma 2.3. By definition of distributional solutions to the respective PDE constraints, as in equations (2.2) and (2.3), if $(\mu, \mathbf{u}) \in \mathcal{A}(m_0, m_1)$ and the value of the (\star) objective function is finite, then $(\mu, \mathbf{u} d\mu_t \otimes dt) \in C(m_0)$, $\mathcal{F}(\mu_1) = 0$, and $\mathcal{E}(\mu, \mathbf{u} d\mu_t \otimes dt) < +\infty$. Conversely, if $(\mu, \mathbf{v}) \in C(m_0)$ and $\mathcal{E}(\mu, \mathbf{v}) < +\infty$, then $\Psi(\mathbf{v}|\mu) < +\infty$ and $\mathcal{F}(\mu_1) < +\infty$. The fact that $\Psi(\mathbf{v}|\mu) < +\infty$ implies that $d\mathbf{v}(y, t) = \mathbf{u}(y, t)d\mu_t(y)dt$ for some $\mathbf{u} \in L^1_{d\mu_t \otimes dt}(\mathbb{R}^d \times [0, 1]; U)$, and the fact $\mathcal{F}(\mu_1) < +\infty$ implies $\mu_1 = m_1$. Thus $(\mu, \mathbf{u}) \in \mathcal{A}(m_0, m_1)$ and at this point, the value of the objective function in (\star) is finite.

In this way, there is a one to one correspondence between feasible points for (\star) and (MFC). Furthermore, at any feasible point of either problem, the value of the objective functions coincide. This gives the result. \square

Now, we prove Lemma 3.1, which shows the energy \mathcal{G} that appears in (MFC) is lower semicontinuous.

Proof of Lemma 3.1. First, note that convergence in $C([0, 1], \mathcal{P}_1(\mathbb{R}^d))$ implies narrow convergence in $\mathcal{P}(\mathbb{R}^d \times [0, 1])$. The lower semicontinuity of the functional Ψ follows from [6, Lemma 5.4.4].

It remains to consider lower semicontinuity of the second term in \mathcal{G} . If L satisfies assumption (iiib), it is independent of the measure and lower semicontinuous, and this is an immediate consequence of [6, Lemma 5.1.7]. Now suppose L satisfies assumption (iia), so that it depends on the measure but is uniformly continuous and real-valued. Suppose $(\mu^n, \mathbf{v}^n)_{n=1}^\infty$ is a sequence in $C([0, 1], \mathcal{P}_1(\mathbb{R}^d)) \times \mathcal{M}(0, T \times \mathbb{R}^d, U)$ converging to a limit (μ, \mathbf{v}) . Then,

$$\begin{aligned} & \left| \int_{\mathbb{R}^d} L(\cdot, \mu_t) d\mu_t - \int_{\mathbb{R}^d} L(\cdot, \mu_t^n) d\mu_t^n \right| \\ & \leq \left| \int_{\mathbb{R}^d} L(\cdot, \mu_t) d\mu_t - \int_{\mathbb{R}^d} L(\cdot, \mu_t) d\mu_t^n \right| + \int_{\mathbb{R}^d} |L(\cdot, \mu_t) - L(\cdot, \mu_t^n)| d\mu_t^n \end{aligned}$$

As $n \rightarrow +\infty$, both terms tend to zero uniformly in $t \in [0, 1]$, since L is jointly uniformly continuous. Fatou's lemma then ensures that

$$(A.2) \quad \int_0^T \int_{\mathbb{R}^d} L(x, \mu_t) d\mu_t(x) \leq \liminf_{n \rightarrow \infty} \int_0^T \int_{\mathbb{R}^d} L(x, \mu_t^n) d\mu_t(x). \quad \square$$

Next, we prove the following estimate on the time regularity of feasible measures.

PROPOSITION A.1. *Under the hypotheses of Assumption 1.1, suppose*

$$(\mu, \mathbf{v}) \in C([0, 1]; \mathcal{P}_1(\mathbb{R}^d)) \times \mathcal{M}(\mathbb{R}^d \times [0, 1]; U),$$

with $\mu_0 = m_0$ and $\partial_t \mu_t + \nabla \cdot (\mathbf{F}(\cdot, \mu_t) \mu_t + \mathbf{v}_t) = 0$, in the sense of distributions. Furthermore, assume $\Psi(\mathbf{v}|\mu) < +\infty$. Then, there exists $C' > 0$ depending on $C_F, C'_F, \psi, \Psi(\mathbf{v}|\mu)$, and $M_1(m_0)$ so that

$$(A.3) \quad W_1(\mu_s, \mu_t) \leq C'(t-s) + \int_s^t \int_{\mathbb{R}^d} |\mathbf{u}(\cdot, r)| d\mu_r dr.$$

Proof. Since $\Psi(\mathbf{v}|\mu) < +\infty$, we have $d\mathbf{v}(y, t) = \mathbf{u}(y, t)d\mu_t(y)dt$ where

$$\int_0^1 \int_{\mathbb{R}^d} \psi(\mathbf{u}(y, t)) d\mu_t(y) dt < +\infty.$$

Thus, recalling our notions of distributional solution from equations (2.2) and (2.3), we see that μ is a distributional solution of the continuity equation with velocity

$$\mathbf{v}(y, t) := \mathbf{F}(y, \mu_t) + \mathbf{u}(y, t).$$

Furthermore, by Lemma 2.2, we have

$$(A.4) \quad \int_0^1 \int_{\mathbb{R}^d} |\mathbf{u}(y, t)| d\mu_t(y) dt \leq \int_0^1 \int_{\mathbb{R}^d} \psi(\mathbf{u}(y, t)) d\mu_t(y) dt + R_\phi + 1 < +\infty.$$

Combining the above inequalities with our assumptions on \mathbf{F} , we have

$$(A.5) \quad \begin{aligned} \|\mathbf{F}(\cdot, \mu_t) + \mathbf{u}(\cdot, t)\|_{L^1(\mu_t)} &\leq \int_{\mathbb{R}^d} (C_F + C'_F |y| + C'_F M_1(\mu_t)) d\mu_t(y) + \int_{\mathbb{R}^d} |\mathbf{u}(\cdot, t)| d\mu_t \\ &\leq C_F + 2C'_F M_1(\mu_t) + \int_{\mathbb{R}^d} |\mathbf{u}(\cdot, t)| d\mu_t \\ &= C_F + 2C'_F W_1(\mu_t, \delta_0) + \int_{\mathbb{R}^d} |\mathbf{u}(\cdot, t)| d\mu_t. \end{aligned}$$

In particular, since $\mu \in C([0, 1]; \mathcal{P}_1(\mathbb{R}^d))$, we have $t \mapsto W_1(\mu_t, \delta_0)$ is uniformly bounded on $[0, 1]$. Combining this with (A.4), we see that the right hand side of equation (A.5) is integrable in time.

Thus, by [5, Theorem 3.4], there exists $\eta \in \mathcal{M}(\mathbb{R}^d \times C([0, 1]; \mathbb{R}^d))$ so that η is concentrated on sets of pairs so that γ is an absolutely continuous integral solution of

$$\dot{\gamma}(t) = \mathbf{v}(\gamma(t), t), \quad \gamma(0) = x$$

and $\mu_t = e_t \# \eta$, where $e_t : \mathbb{R}^d \times C([0, 1]; \mathbb{R}^d) \rightarrow \mathbb{R}^d : (x, \gamma) \rightarrow \gamma(t)$. Therefore, $(e_t, e_s) \# \eta$ is a transport plan from μ_t to μ_s , so applying the definition of the 1-Wasserstein metric, Jensen's inequality, Tonelli, and inequality (A.5), we obtain

$$(A.6) \quad \begin{aligned} W_1(\mu_s, \mu_t) &\leq \int_{\mathbb{R}^d \times \mathbb{R}^d} |\pi^1 - \pi^2| d(e_s, e_t) \# \eta \\ &= \int_{\mathbb{R}^d \times C([0, 1]; \mathbb{R}^d)} |e_s - e_t| d\eta \\ &= \int_{\mathbb{R}^d \times C([0, 1]; \mathbb{R}^d)} |\gamma(s) - \gamma(t)| d\eta(x, \gamma) \\ &\leq \int_{\mathbb{R}^d \times C([0, 1]; \mathbb{R}^d)} \int_s^t |\dot{\gamma}(r)| dr d\eta(x, \gamma) \\ &= \int_{\mathbb{R}^d \times C([0, 1]; \mathbb{R}^d)} \int_s^t |\mathbf{F}(\gamma(r), \mu_r) + \mathbf{u}(\gamma(r), r)| dr d\eta(x, \gamma) \\ &= \int_s^t \|\mathbf{F}(\cdot, \mu_r) + \mathbf{u}(\cdot, r)\|_{L^1(\mu_r)} dr \\ &\leq \int_s^t \left(C_F + 2C'_F M_1(\mu_r) + \int_{\mathbb{R}^d} |\mathbf{u}(\cdot, r)| d\mu_r \right) dr. \end{aligned}$$

In particular, for all $t \in [0, 1]$, inequality (A.4) ensures that, for all $t \in [0, 1]$,

$$\begin{aligned} M_1(\mu_t) &= W_1(\mu_t, \delta_0) \leq W_1(\mu_0, \mu_t) + W_1(\mu_0, \delta_0) \\ &\leq (C_F + R_\phi + 1) + 2C'_F \int_0^t M_1(\mu_r) dr + \Psi(\mathbf{v}|\mu) + M_1(m_0). \end{aligned}$$

Thus, by Gronwall's inequality, there exists $C > 0$ depending on C_F , R_ϕ , $\Psi(\mathbf{v}|\mu)$, and $M_1(m_0)$ so that

$$(A.7) \quad M_1(\mu_t) \leq C(1 + 2C'_F t e^{2C'_F t}) \leq C(1 + 2C'_F e^{2C'_F}) := C', \quad \forall t \in [0, 1], \forall n \in \mathbb{N}.$$

Substituting this fact into inequality (A.6), we obtain the result. \square

Next, we prove Lemma 3.2, which shows that sublevels of \mathcal{G} in the constraint set $\mathcal{C}(m_0)$ are sequentially compact.

Proof of Lemma 3.2. We begin by observing that, since $\sup_n \mathcal{G}(\mu_n, \mathbf{v}_n) < +\infty$, we also have

$$(A.8) \quad \sup_n \Psi(\mathbf{v}_n | \mu_n) < +\infty$$

and, for all $n \in \mathbb{N}$, there exists \mathbf{u}_n so that $d\mathbf{v}_n(y, t) = \mathbf{u}_n(y, t) d\mu_{n,t}(y) dt$ and

$$(A.9) \quad \Psi(\mathbf{v}_n | \mu_n) = \int_0^1 \int_{\mathbb{R}^d} \psi(\mathbf{u}_n(y, t)) d\mu_{n,t}(y) dt.$$

Furthermore, by inequality (2.5) and Lemma 2.2, we have

$$(A.10) \quad \sup_n \int_0^1 \int_{\mathbb{R}^d} |\mathbf{u}_n(y, t)| d\mu_{n,t}(y) dt \leq \sup_n \int_0^1 \int_{\mathbb{R}^d} \phi(|\mathbf{u}_n(y, t)|) d\mu_{n,t}(y) dt + R_\phi < +\infty.$$

First, we apply Arzelá-Ascoli to show the convergence of μ_n , up to a subsequence. We begin by showing $\{\mu_{n,t}\}_{n \in \mathbb{N}, t \in [0, 1]}$ is relatively sequentially compact with respect to W_1 . Since $(\mu_n, \mathbf{v}_n) \in \mathcal{C}(m_0)$ and $\sup_n \mathcal{G}(\mu_n, \mathbf{v}_n) < +\infty$, by [30, Proposition 5.3], there exists $C > 0$ and a nonnegative, continuously differentiable, convex, and superlinear at $+\infty$ function θ , depending on the choice of ψ , \mathbf{F} and m_0 , as in Assumption 1.1, and the value of $\sup_n \mathcal{G}(\mu_n, \mathbf{v}_n)$ so that

$$\sup_{t \in [0, 1], n \in \mathbb{N}} \int_{\mathbb{R}^d} \theta(|x|) d\mu_{n,t}(x) \leq C \left(1 + \int_{\mathbb{R}^d} \theta(|x|) m_0(x) dx \right) < +\infty.$$

Since θ is superlinear at $+\infty$, for all $\varepsilon > 0$, there exists R so that $r \geq R$ ensures $\theta(r)/r \geq 1/\varepsilon$ and

$$\int_{\mathbb{R}^d \setminus B_0(R)} |x| d\mu_{n,t}(x) \leq \varepsilon \int_{\mathbb{R}^d \setminus B_0(R)} \theta(|x|) d\mu_{n,t}(x) \leq \varepsilon C \left(1 + \int_{\mathbb{R}^d} \theta(|x|) m_0(x) dx \right).$$

Thus, $\{\mu_{n,t}\}_{n \in \mathbb{N}, t \in [0, 1]}$ have uniformly integrable first moments, so by [6, Proposition 7.1.5], $\{\mu_{n,t}\}_{n \in \mathbb{N}, t \in [0, 1]}$ is relatively sequentially compact with respect to W_1 convergence.

Next, we show equicontinuity of $\mu_{n,t}$. By Proposition A.1, there exists $C' > 0$ so that, for all $n \in \mathbb{N}$, $s, t \in [0, 1]$,

$$(A.11) \quad W_1(\mu_{n,s}, \mu_{n,t}) \leq C'(t - s) + \int_s^t \int_{\mathbb{R}^d} |\mathbf{u}_n(\cdot, r)| d\mu_{n,r} dr.$$

Since ϕ is strictly increasing, $\phi^{-1} : [0, +\infty) \rightarrow (0, +\infty)$ is well defined and strictly increasing. Since ϕ is superlinear at $+\infty$, we must have $\lim_{s \rightarrow +\infty} \phi^{-1}(s)/s = 0$. Therefore, we may use Jensen's inequality to estimate the second term above by

$$(A.12) \quad \begin{aligned} \int_s^t \int_{\mathbb{R}^d} |\mathbf{u}_n(\cdot, r)| d\mu_{n,r} dr &= (t - s) \phi^{-1} \circ \phi \left(\frac{1}{t - s} \int_s^t \int_{\mathbb{R}^d} |\mathbf{u}_n(\cdot, r)| d\mu_{n,r} dr \right) \\ &\leq (t - s) \phi^{-1} \left(\frac{1}{t - s} \int_s^t \int_{\mathbb{R}^d} \phi(|\mathbf{u}_n(\cdot, r)|) d\mu_{n,r} dr \right) \\ &\leq (t - s) \phi^{-1} \left(\frac{C''}{t - s} \right) \end{aligned}$$

where the last inequality follow from (A.10), for some $C'' > 0$ independent of $n \in \mathbb{R}^d$ and $s, t \in [0, 1]$. Combining (A.11) and (A.12) shows that $\{\mu_{n,t}\}_{n \in \mathbb{N}}$ are equicontinuous.

Thus, by Arzelá-Ascoli, there exists $\mu \in C([0, 1]; \mathcal{P}_1(\mathbb{R}^d))$ so that, up to a subsequence,

$$(A.13) \quad \mu_n \rightarrow \mu \text{ in } C([0, 1]; \mathcal{P}_1(\mathbb{R}^d)).$$

In particular, this implies $\mu_0 = m_0$ and $d\mu_{n,t} \otimes dt \rightarrow d\mu_t \otimes dt$ narrowly in $\mathcal{P}(\mathbb{R}^d \times [0, 1])$.

Now, we turn to the convergence of \mathbf{v}_n . By inequality (A.8), equation (A.9), and equation (A.13), we may argue in a similar way to [6, Theorem 5.4.4] to obtain that there exists $\mathbf{u} : \mathbb{R}^d \times [0, 1] \rightarrow U$ so that, up to a subsequence, for all $\xi \in C_c^\infty(\mathbb{R}^d \times [0, 1]; \mathbb{R}^d)$,

$$\lim_{n \rightarrow +\infty} \int_0^1 \int_{\mathbb{R}^d} \xi(y, t) \mathbf{u}_n(y, t) d\mu_{n,t}(y) dt = \int_0^1 \int_{\mathbb{R}^d} \xi(y, t) \mathbf{u}(y, t) d\mu_t(y),$$

and

$$(A.14) \quad \int_0^1 \int_{\mathbb{R}^d} |\mathbf{u}(y, t)| d\mu_t(y) dt \leq \int_0^1 \int_{\mathbb{R}^d} \psi(\mathbf{u}(y, t)) d\mu_t(y) dt + R_\phi + 1 < +\infty.$$

Thus, defining $d\mathbf{v}(y, t) := \mathbf{u}(y, t) d\mu_t(y) dt$, we see that, up to a subsequence, $\mathbf{v}_n \rightarrow \mathbf{v}$ narrowly in $\mathcal{M}(\mathbb{R}^d \times [0, 1]; U)$.

It remains to show that $(\mu, \mathbf{v}) \in C(m_0)$. We have already shown that $\mu_0 = m_0$. Since (μ_n, \mathbf{v}_n) is a distributional solution of the continuity equation, in the sense of equation (2.3), for all $n \in \mathbb{N}$ and $\varphi \in C_c^\infty(\mathbb{R}^d \times [0, 1])$, we have

$$\int_0^1 \int_{\mathbb{R}^d} \partial_t \varphi d\mu_{n,t} dt + \int_0^1 \int_{\mathbb{R}^d} \nabla \varphi \cdot \mathbf{F}(\cdot, \mu_{n,t}) d\mu_{n,t} dt + \int_0^1 \int_{\mathbb{R}^d} \nabla \varphi \cdot d\mathbf{v}_n = 0.$$

By the convergence in equation (3.1), it is clear that we may pass to the limit in the first and third terms. For the second term, note that, since \mathbf{F} is uniformly continuous we can conclude that

$$\int_0^1 \int_{\mathbb{R}^d} \nabla \varphi \cdot [\mathbf{F}(\cdot, \mu_{n,t}) - \mathbf{F}(\cdot, \mu_t)] d\mu_{n,t} dt \rightarrow 0.$$

Thus, we may likewise pass to the limit in the second term, since $(y, t) \mapsto \mathbf{F}(y, \mu_t)$ is continuous and $d\mu_{n,t} \otimes dt \rightarrow d\mu_t \otimes dt$ narrowly in $\mathcal{P}(\mathbb{R}^d \times [0, 1])$. This shows that (μ, \mathbf{v}) is a distributional solution of the continuity equation.

It remains to show $\mu \in AC([0, 1]; \mathcal{P}_1(\mathbb{R}^d))$. By Proposition A.1, there exists $C' > 0$ so that, for all $n \in \mathbb{N}$, $s, t \in [0, 1]$,

$$(A.15) \quad W_1(\mu_s, \mu_t) \leq C'(t - s) + \int_s^t \int_{\mathbb{R}^d} |\mathbf{u}(\cdot, r)| d\mu_r dr = \int_s^t \omega(r) dr$$

for

$$\omega(r) := C' + \int_{\mathbb{R}^d} |\mathbf{u}(\cdot, r)| d\mu_r.$$

By inequality (A.14), we have $\omega \in L^1(\mathbb{R}^d)$. Therefore $\mu \in AC([0, 1]; \mathcal{P}_1(\mathbb{R}^d))$.

Finally, the fact that $\mathcal{G}(\mu, \mathbf{v}) \leq \sup_n \mathcal{G}(\mu_n, \mathbf{v}_n)$ is an immediate consequence of the lower semicontinuity proved in Lemma 3.1. \square

Appendix B. Convergence as $\varepsilon \rightarrow 0, \delta \rightarrow 0, N \rightarrow +\infty$. We begin with our proof of Proposition 3.3, which shows Γ -convergence of the objective functionals from (MFC $_\varepsilon$) to (MFC) as $\varepsilon \rightarrow 0$.

Proof of Proposition 3.3. First we prove part (i). Without loss of generality, we may pass to a subsequence so that

$$\lim_{\varepsilon \rightarrow 0} \mathcal{E}_\varepsilon(\mu_\varepsilon, \mathbf{v}_\varepsilon) = \liminf_{\varepsilon \rightarrow 0} \mathcal{E}_\varepsilon(\mu_\varepsilon, \mathbf{v}_\varepsilon) < +\infty.$$

Thus $\mathcal{F}_\varepsilon(\mu_{\varepsilon,1})$, as defined in equation (1.2), must be bounded uniformly in ε , which shows that $\lim_{\varepsilon \rightarrow 0} \|\mu_{\varepsilon,1} - m_1\|_2 = 0$. Since, $\mu_\varepsilon \rightarrow \mu$ in $C([0, 1]; \mathcal{P}_1(\mathbb{R}^d))$, we have $\mu_{\varepsilon,1}$ converges to μ_1 in $\mathcal{P}_1(\mathbb{R}^d)$. Thus, by uniqueness of limits, $\mu_1 = m_1$, so $\mathcal{F}(\mu_1) = 0$. Due to the lower semicontinuity of the functional \mathcal{G} from Lemma 3.1, we can conclude that

$$\mathcal{E}(\mu, \mathbf{v}) = \mathcal{G}(\mu, \mathbf{v}) \leq \liminf_{\varepsilon \rightarrow 0} \mathcal{E}_\varepsilon(\mu_\varepsilon, \mathbf{v}_\varepsilon).$$

Next, we show part (ii). We may assume without loss of generality that $\mathcal{E}(\mu, \mathbf{v}) < +\infty$, so $\mathcal{F}(\mu_1) = 0$ and $\mu_1 = m_1 \in L^2(\mathbb{R}^d)$. The result then follows, since $\mathcal{F}_\varepsilon(\mu_1) \equiv 0$. \square

We now apply this Γ -convergence result to prove Proposition 1.2, which shows that, as $\varepsilon \rightarrow 0$, minimizers of (MFC_ε) converge to a minimizer of (MFC) , up to a subsequence.

Proof of Proposition 1.2. First, note that by the feasibility of (MFC) , there exist $(\mu', \mathbf{u}') \in C(m_0)$ so that $\mathcal{E}(\mu', \mathbf{u}') < +\infty$, so $\mu'_1 = m_1$. Combining this with the fact that $(\mu_\varepsilon, \mathbf{v}_\varepsilon)$ are minimizers of (MFC_ε) , we have

$$\mathcal{G}(\mu_\varepsilon, \mathbf{v}_\varepsilon) \leq \mathcal{E}_\varepsilon(\mu_\varepsilon, \mathbf{v}_\varepsilon) \leq \mathcal{E}_\varepsilon(\mu', \mathbf{u}') = \mathcal{G}(\mu', \mathbf{u}') + \mathcal{F}_\varepsilon(\mu'_1) = \mathcal{G}(\mu', \mathbf{u}') + 0, \quad \forall \varepsilon > 0.$$

Therefore, $\sup_\varepsilon \mathcal{G}(\mu_\varepsilon, \mathbf{v}_\varepsilon) < +\infty$. By Lemma 3.2, there exists $(\mu, \mathbf{v}) \in C(m_0)$ so that, up to a subsequence, (1.3) holds.

Furthermore, for any $(\mu'', \mathbf{v}'') \in C(m_0)$, Proposition 3.3 and the fact that $(\mu_\varepsilon, \mathbf{v}_\varepsilon)$ are minimizers ensure that

$$\mathcal{E}(\mu'', \mathbf{v}'') \geq \limsup_{\varepsilon \rightarrow 0} \mathcal{E}_\varepsilon(\mu'', \mathbf{v}'') \geq \liminf_{\varepsilon \rightarrow 0} \mathcal{E}_\varepsilon(\mu_\varepsilon, \mathbf{v}_\varepsilon) \geq \mathcal{E}(\mu, \mathbf{v})$$

Since $(\mu'', \mathbf{v}'') \in C(m_0)$ was arbitrary, this gives the result. \square

Next we prove Theorem 1.5, on the convergence of minimizers of $(\text{MFC}_{\varepsilon,\delta})$ to minimizers of (MFC_ε) as $\delta \rightarrow 0$.

Proof of Theorem 1.5. First, note that, by the assumption that (MFC_ε) is feasible, there exist $(\mu', \mathbf{v}') \in C(m_0)$ so that $\mathcal{E}_\varepsilon(\mu', \mathbf{v}') < +\infty$. Combining this with the fact that $(\mu_\delta, \mathbf{v}_\delta)$ are minimizers, we have

$$\mathcal{G}(\mu_\delta, \mathbf{v}_\delta) \leq \mathcal{E}_{\varepsilon,\delta}(\mu_\delta, \mathbf{v}_\delta) \leq \mathcal{E}_{\varepsilon,\delta}(\mu', \mathbf{v}') = \mathcal{G}(\mu', \mathbf{v}') + \varepsilon^{-1} \|k_\delta * \mu'_1 - k_\delta * m_1\|_{L^2(\mathbb{R}^d)}^2.$$

Since $\mathcal{E}_\varepsilon(\mu', \mathbf{v}') < +\infty$, $\mu'_1 \in L^2(\mathbb{R}^d)$, and the right hand side is bounded uniformly in δ . Thus, by Lemma 3.2, there exists $(\mu, \mathbf{v}) \in C(m_0)$ so that, up to a subsequence, (1.6) holds.

Furthermore, for any $(\mu'', \mathbf{v}'') \in C(m_0)$, Proposition 3.4 and the fact that $(\mu_\delta, \mathbf{v}_\delta)$ are minimizers ensure that

$$\mathcal{E}_\varepsilon(\mu'', \mathbf{v}'') \geq \limsup_{\delta \rightarrow 0} \mathcal{E}_{\varepsilon,\delta}(\mu'', \mathbf{v}'') \geq \liminf_{\delta \rightarrow 0} \mathcal{E}_{\varepsilon,\delta}(\mu_\delta, \mathbf{v}_\delta) \geq \mathcal{E}(\mu, \mathbf{v}) \quad \square$$

Since $(\mu'', \mathbf{v}'') \in C(m_0)$ was arbitrary, this gives the result.

Now, we prove Proposition 1.7, which develops sufficient conditions to ensure that a solution of $(\text{MFC}_{\varepsilon,\delta,N})$ exists. Our proof is a mild adaptation of [30, Proposition 4.2], extending to the case when L_N satisfies Assumption 1.6(ib).

Proof of Proposition 1.7. First, we consider feasibility of $(\text{MFC}_{\varepsilon,\delta,N})$. As observed in the paragraph following [30, equation (3.8)], classical results on existence of solutions to ordinary differential equations ensure that there exists $(\mathbf{y}, 0) \in \mathcal{A}_N(\mathbf{y}_0)$ is nonempty. If L_N satisfies Assumption 1.6(ia), then for $(\mathbf{y}, 0) \in \mathcal{A}_N(\mathbf{y}_0)$, $\mathcal{E}_{\varepsilon,\delta,N}(\mathbf{y}, \mathbf{u}) < +\infty$.

On the other hand, suppose L_N satisfies Assumption 1.6(ib), the control is unconstrained $U = \mathbb{R}^d$, and the initial particle locations are contained in $\{L_N < +\infty\}$. If we consider the curve $y_i(t) \equiv y_{i,0}$ and the velocity $u_i(t) = -\mathbf{F}_N(y_i(t), \mathbf{y}(t)) = -\mathbf{F}_N(y_{i,0}, \mathbf{y}_0)$, then $(\mathbf{y}, \mathbf{u}) \in \mathcal{A}_N(\mathbf{y}_0)$, and since $y_i(t) \in \{L_N < +\infty\}$ for all $t \in [0, 1]$, we have $\mathcal{E}_{\varepsilon,\delta,N}(\mathbf{y}, \mathbf{u}) < +\infty$.

Now, suppose $(\text{MFC}_{\varepsilon,\delta,N})$ is feasible, and we will show that a minimizer exists. Take a minimizing sequence $(\mathbf{y}_k, \mathbf{u}_k) \in \mathcal{A}_N(\mathbf{y}_0)$. Since

$$\sup_k \frac{1}{N} \sum_{i=1}^N \int_0^1 \psi(u_{i,k}(t)) dt < +\infty,$$

as in [30, Proposition 4.2], we obtain that there exist $(\mathbf{y}, \mathbf{u}) \in \mathcal{A}_N(\mathbf{y}_0)$ so that, up to a subsequence, $\mathbf{y}_k \rightarrow \mathbf{y}$ in $C([0, 1]; (\mathbb{R}^d)^N)$ and $\mathbf{u}_k \rightarrow \mathbf{u}$ in $L^1([0, 1]; U^N)$. Furthermore, the proof of [30, Proposition 4.2] shows that

$$\begin{aligned} \liminf_{k \rightarrow +\infty} \frac{1}{N} \sum_{i=1}^N \int_0^1 \psi(u_{i,k}(t)) dt + \frac{1}{N} \sum_{i=1}^N \int_0^1 L_N(y_{i,k}(t), \mathbf{y}_k(t)) dt \\ \geq \frac{1}{N} \sum_{i=1}^N \int_0^1 \psi(u_i(t)) dt + \frac{1}{N} \sum_{i=1}^N \int_0^1 L_N(y_i(t), \mathbf{y}(t)) dt. \end{aligned}$$

By continuity of K_δ and $K_\delta * m_1$, we may likewise pass to the limit in k in the third term in $\mathcal{E}_{\varepsilon,\delta,N}$. This shows $\liminf_{k \rightarrow +\infty} \mathcal{E}_{\varepsilon,\delta,N}(\mathbf{y}_k, \mathbf{u}_k) \geq \mathcal{E}_{\varepsilon,\delta,N}(\mathbf{y}, \mathbf{u})$. Since $(\mathbf{y}_k, \mathbf{u}_k)$ was a minimizing sequence, this shows that (\mathbf{y}, \mathbf{u}) is a minimizer of $(\text{MFC}_{\varepsilon,\delta,N})$. \square

Now, we prove Theorem 3.7, which shows that, for fixed $\varepsilon, \delta > 0$, as $N \rightarrow +\infty$, minimizers of the spatially discrete problem $(\text{MFC}_{\varepsilon,\delta,N})$ converge to a solution of $(\text{MFC}_{\varepsilon,\delta})$, up to a subsequence.

Proof of Theorem 3.7. First, we will show that there exists such a sequence $\mathbf{y}_{N,0}$. Let

$$(B.1) \quad e_{m_0} := \inf_{(\mu, \nu) \in C(m_0)} \mathcal{E}_{\varepsilon,\delta}(\mu, \nu), \quad \mathcal{E}_{\varepsilon,\delta}(\mu, \nu) := \mathcal{G}(\mu, \nu) + \mathcal{F}_{\varepsilon,\delta}(\mu_1).$$

By our hypothesis that $(\text{MFC}_{\varepsilon,\delta})$ is feasible, we have that $e_{m_0} < +\infty$. Thus, for all $N \in \mathbb{N}$, there exists $(\mu_N, \nu_N) \in C(m_0)$ so that $\mathcal{E}_{\varepsilon,\delta}(\mu_N, \nu_N) < e_{m_0} + 1/N$. By Proposition 3.6(ii), there exists $\mathbf{y}_{N,0}$ so that (3.17-3.16) hold and there exist $(\tilde{\mathbf{y}}_N, \tilde{\mathbf{u}}_N) \in \mathcal{A}_N(\mathbf{y}_{N,0})$ so that $\mathcal{E}_{\varepsilon,\delta,N}(\tilde{\mathbf{y}}_N, \tilde{\mathbf{u}}_N) \leq e_{m_0} + 2/N$. Now, suppose $(\mathbf{y}_N, \mathbf{u}_N) \in \mathcal{A}_N(\mathbf{y}_{N,0})$ is an optimizer of $(\text{MFC}_{\varepsilon,\delta,N})$. Then, we have

$$(B.2) \quad \limsup_{N \rightarrow +\infty} \mathcal{E}_{\varepsilon,\delta,N}(\mathbf{y}_N, \mathbf{u}_N) \leq \limsup_{N \rightarrow +\infty} \mathcal{E}_{\varepsilon,\delta,N}(\tilde{\mathbf{y}}_N, \tilde{\mathbf{u}}_N) \leq e_{m_0}$$

and

$$\sup_N \frac{1}{N} \sum_{i=1}^N \int_0^1 \psi(u_{i,N}(t)) dt \leq \sup_N \mathcal{E}_{\varepsilon,\delta,N}(\mathbf{y}_N, \mathbf{u}_N) \leq \sup_N \mathcal{E}_{\varepsilon,\delta,N}(\tilde{\mathbf{y}}_N, \tilde{\mathbf{u}}_N) \leq e_{m_0} + 1 < +\infty.$$

Then [30, Theorem 3.1] ensures that there exists $(\mu, \nu) \in C(m_0)$ so that (3.19)-(3.20) hold. Furthermore, for any such limit point (μ, ν) , by Proposition 3.6(i),

$$(B.3) \quad e_{m_0} \leq \mathcal{E}_{\varepsilon,\delta}(\mu, \nu) \leq \liminf_{N \rightarrow +\infty} \mathcal{E}_{\varepsilon,\delta,N}(\mathbf{y}_N, \mathbf{u}_N).$$

Combining (B.2) and (B.3), we obtain (3.18) and that (μ, ν) is a minimizer of $(\text{MFC}_{\varepsilon,\delta})$. \square

Now, we turn to the proof of Corollary 1.8, which provides sufficient conditions to ensure existence of minimizers to the continuum optimization problems we consider.

Proof of Corollary 1.8. Since (MFC) is feasible, we have that (MFC $_{\varepsilon}$) and (MFC $_{\varepsilon,\delta}$) are feasible for all $\varepsilon, \delta > 0$. Likewise, Proposition 1.7 ensures that, for any $\mathbf{y}_0 \in (\mathbb{R}^d)^N$ with

$$(B.4) \quad y_{i,0} \in \{L_N < +\infty\} \text{ for all } i = 1, \dots, N, \quad \square$$

minimizers of (MFC $_{\varepsilon,\delta,N}$) exist.

By Theorem 3.7, for all $N \in \mathbb{N}$, there exists $\mathbf{y}_{N,0} \in (\mathbb{R}^d)^N$ satisfying (B.4) so that, for any sequence of minimizers $(\mathbf{y}_N, \mathbf{u}_N) \in \mathcal{A}(\mathbf{y}_{N,0})$ of (MFC $_{\varepsilon,\delta,N}$), up to a subsequence, $(\mathbf{y}_N, \mathbf{u}_N)$ converges to a minimizer of (MFC $_{\varepsilon,\delta}$). Thus, minimizers of (MFC $_{\varepsilon,\delta}$) exist.

By Theorem 1.5, any sequence of minimizers of (MFC $_{\varepsilon,\delta}$) converges, up to a subsequence, to a minimizer of (MFC $_{\varepsilon}$), so minimizers of (MFC $_{\varepsilon}$) exist.

Finally, by Proposition 1.2, any sequence of minimizers of (MFC $_{\varepsilon}$) converges, up to a subsequence, to a minimizer of (MFC), so minimizers of (MFC) exist.

We conclude with the proof of our main theorem, Theorem 1.11, on the convergence of minimizers of (MFC $_{\varepsilon,\delta,N}$) to a minimizer of (MFC) as $\varepsilon \rightarrow 0, \delta \rightarrow 0$, and $N \rightarrow +\infty$.

Proof of Theorem 1.11. Recall that, as a consequence of our assumption that (MFC) is feasible, we also have that (MFC $_{\varepsilon}$) and (MFC $_{\varepsilon,\delta}$) are feasible.

By Theorem 1.9, for any $\varepsilon, \delta > 0$, if $(\mathbf{y}_N^{\varepsilon,\delta}, \mathbf{u}_N^{\varepsilon,\delta}) \in \mathcal{A}_N(\mathbf{y}_{N,0})$ minimizes (MFC $_{\varepsilon,\delta,N}$), there is a subsequence $N_l^{\varepsilon,\delta}$ of N and $(\mu_{\varepsilon,\delta}, \mathbf{v}_{\varepsilon,\delta}) \in \mathcal{C}(m_0)$ that minimizes (MFC $_{\varepsilon,\delta}$) for which

$$(B.5) \quad \frac{1}{N_l^{\varepsilon,\delta}} \sum_{i=1}^{N_l^{\varepsilon,\delta}} \delta_{y_{i,N_l^{\varepsilon,\delta}}^{\varepsilon,\delta}} \xrightarrow{l \rightarrow +\infty} \mu_{\varepsilon,\delta} \text{ in } C([0, 1]; \mathcal{P}_1(\mathbb{R}^d))$$

$$(B.6) \quad \frac{1}{N_l^{\varepsilon,\delta}} \sum_{i=1}^{N_l^{\varepsilon,\delta}} u_{i,N_l^{\varepsilon,\delta}}^{\varepsilon,\delta}(t) \delta_{y_{i,N_l^{\varepsilon,\delta}}^{\varepsilon,\delta}(t)} dt \xrightarrow{l \rightarrow +\infty} \mathbf{v}_{\varepsilon,\delta} \text{ in } \mathcal{M}(\mathbb{R}^d \times [0, 1]; U).$$

Likewise, for any ε , Theorem 1.5 ensures that there exists a subsequence δ_k^ε of δ so that $(\mu_{\varepsilon,\delta_k^\varepsilon}, \mathbf{v}_{\varepsilon,\delta_k^\varepsilon})$ converges to some $(\mu_\varepsilon, \mathbf{v}_\varepsilon) \in \mathcal{C}(m_0)$, where $(\mu_\varepsilon, \mathbf{v}_\varepsilon)$ minimizes (MFC $_{\varepsilon}$). Finally, by Theorem 1.2, up to a subsequence, $(\mu_\varepsilon, \mathbf{v}_\varepsilon)$ converges to some (μ, \mathbf{v}) , where (μ, \mathbf{v}) is a minimizer of (MFC).

To obtain the result, we now apply a standard diagonal argument. For simplicity of notation, recall that convergence in the topologies of equations (1.14-1.15) is metrizable, and let d denote such a metric. For $m \in \mathbb{N}$, we may choose a subsequence $(\mu_{\varepsilon_m}, \mathbf{v}_{\varepsilon_m})$ so that

$$d\left((\mu_{\varepsilon_m}, \mathbf{v}_{\varepsilon_m}), (\mu, \mathbf{v})\right) < \frac{1}{3m}.$$

Likewise, we may choose a subsequence $(\mu_{\varepsilon_m, \delta_m^{\varepsilon_m}}, \mathbf{v}_{\varepsilon_m, \delta_m^{\varepsilon_m}})$ so that

$$d\left(\left(\mu_{\varepsilon_m, \delta_m^{\varepsilon_m}}, \mathbf{v}_{\varepsilon_m, \delta_m^{\varepsilon_m}}\right), (\mu_{\varepsilon_m}, \mathbf{v}_{\varepsilon_m})\right) < \frac{1}{3m}.$$

Finally, we may choose a subsequence $\left(\mathbf{y}_{N_m^{j_m, k_m}}^{\varepsilon_m, \delta_m^{\varepsilon_m}}, \mathbf{u}_{N_m^{j_m, k_m}}^{\varepsilon_m, \delta_m^{\varepsilon_m}}\right)$ so that, defining μ_m and \mathbf{v}_m as on the left hand side of (1.12-1.13),

$$d\left((\mu_m, \mathbf{v}_m), \left(\mu_{\varepsilon_m, \delta_m^{\varepsilon_m}}, \mathbf{v}_{\varepsilon_m, \delta_m^{\varepsilon_m}}\right)\right) < \frac{1}{3m}.$$

The result then follows by the triangle inequality. \square

Acknowledgements: K. Craig would like to thank Amir Sagiv for an interesting discussion on the case of optimal transport around obstacles and the connection to manifolds with holes.

REFERENCES

- [1] Y. ACHDOU, F. CAMILLI, AND I. CAPUZZO-DOLCETTA, *Mean field games: numerical methods for the planning problem*, SIAM Journal on Control and Optimization, 50 (2012), pp. 77–109.
- [2] A. AGRACHEV AND P. LEE, *Optimal transportation under nonholonomic constraints*, Transactions of the American Mathematical Society, 361 (2009), pp. 6019–6047.
- [3] M. S. ALBERGO AND E. VANDEN-EIJNDEN, *Building normalizing flows with stochastic interpolants*, arXiv preprint arXiv:2209.15571, (2022).
- [4] L. AMBROSIO, E. BRUÉ, AND D. SEMOLA, *Lectures on optimal transport*, Springer, 2021.
- [5] L. AMBROSIO AND G. CRIPPA, *Continuity equations and ode flows with non-smooth velocity*, Proceedings of the Royal Society of Edinburgh Section A: Mathematics, 144 (2014), pp. 1191–1244.
- [6] L. AMBROSIO, N. GIGLI, AND G. SAVARÉ, *Gradient flows: in metric spaces and in the space of probability measures*, Springer Science & Business Media, 2005.
- [7] L. AMBROSIO, N. GIGLI, AND G. SAVARÉ, *Gradient flows in metric spaces and in the space of probability measures*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, second ed., 2008.
- [8] A. BACCIOTTI, *Stability and control of linear systems*, Springer, 2019.
- [9] J. T. BEALE AND A. MAJDA, *Vortex methods. I. Convergence in three dimensions*, Math. Comp., 39 (1982), pp. 1–27, <https://doi.org/10.2307/2007617>, <http://dx.doi.org/10.2307/2007617>.
- [10] J. T. BEALE AND A. MAJDA, *Vortex methods. II. Higher order accuracy in two and three dimensions*, Math. Comp., 39 (1982), pp. 29–52, <https://doi.org/10.2307/2007618>, <http://dx.doi.org/10.2307/2007618>.
- [11] J.-D. BENAMOU AND Y. BRENIER, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numer. Math., 84 (2000), pp. 375–393, <https://doi.org/10.1007/s002110050002>, <http://dx.doi.org/10.1007/s002110050002>.
- [12] B. BONNET AND H. FRANKOWSKA, *Necessary optimality conditions for optimal control problems in wasserstein spaces*, Applied Mathematics & Optimization, 84 (2021), pp. 1281–1330.
- [13] B. BONNET AND H. FRANKOWSKA, *Semiconcavity and sensitivity analysis in mean-field optimal control and applications*, Journal de Mathématiques Pures et Appliquées, 157 (2022), pp. 282–345.
- [14] M. BURGER AND A. ESPOSITO, *Porous medium equation and cross-diffusion systems as limit of nonlocal interaction*, Nonlinear Analysis, 235 (2023), p. 113347.
- [15] M. BURGER, R. PINNAU, C. TOTZECK, AND O. TSE, *Mean-field optimal control and optimality conditions in the space of probability measures*, SIAM Journal on Control and Optimization, 59 (2021), pp. 977–1006.
- [16] R. CARMONA, F. DELARUE, ET AL., *Probabilistic theory of mean field games with applications I-II*, Springer, 2018.
- [17] J. A. CARRILLO, Y.-P. CHOI, AND M. HAURAY, *The derivation of swarming models: mean-field limit and wasserstein distances*, Collective Dynamics from Bacteria to Crowds: An Excursion Through Modeling, Analysis and Simulation, (2014), pp. 1–46.
- [18] J. A. CARRILLO, K. CRAIG, AND F. S. PATACCINI, *A blob method for diffusion*, Calculus of Variations and Partial Differential Equations, 58 (2019), pp. 1–53.
- [19] J. A. CARRILLO, A. ESPOSITO, AND J. S.-H. WU, *Nonlocal approximation of nonlinear diffusion equations*, arXiv preprint arXiv:2302.08248, (2023).
- [20] J. A. CARRILLO, M. FORNASIER, G. TOSCANI, AND F. VECIL, *Particle, kinetic, and hydrodynamic models of swarming*, Mathematical modeling of collective behavior in socio-economic and life sciences, (2010), pp. 297–336.
- [21] G. CAVAGNARI, S. LISINI, C. ORRIERI, AND G. SAVARÉ, *Lagrangian, eulerian and kantorovich formulations of multi-agent optimal control problems: equivalence and gamma-convergence*, Journal of Differential Equations, 322 (2022), pp. 268–364.
- [22] Y. CHEN, T. T. GEORGIU, AND M. PAVON, *Optimal transport over a linear dynamical system*, IEEE Transactions on Automatic Control, 62 (2016), pp. 2137–2152.
- [23] K. CRAIG AND A. L. BERTOZZI, *A blob method for the aggregation equation*, Math. Comp., 85 (2016), pp. 1681–1717, <https://doi.org/10.1090/mcom3033>, <http://dx.doi.org/10.1090/mcom3033>.
- [24] K. CRAIG, K. ELAMVAZHUTHI, M. HABERLAND, AND O. TURANOVA, *A blob method for inhomogeneous diffusion with applications to multi-agent control and sampling*, Mathematics of Computation, (2023).
- [25] K. CRAIG, M. JACOBS, AND O. TURANOVA, *Nonlocal approximation of slow and fast diffusion*, arXiv preprint arXiv:2312.11438, (2023).
- [26] K. ELAMVAZHUTHI AND P. GROVER, *Optimal transport over nonlinear systems via infinitesimal generators*

- on graphs*, Journal of Computational Dynamics, 5 (2018), pp. 1–32.
- [27] K. ELAMVAZHUTHI, S. LIU, W. LI, AND S. OSHER, *Dynamical optimal transport of nonlinear control-affine systems*, Journal of Computational Dynamics, (2023), pp. 0–0.
- [28] A. FIGALLI AND F. GLAUDO, *An Invitation to Optimal Transport, Wasserstein Distances, and Gradient Flows*, EMS Textbooks in Mathematics, 2021.
- [29] R. FLAMARY, N. COURTY, A. GRAMFORT, M. Z. ALAYA, A. BOISBUNON, S. CHAMBON, L. CHAPEL, A. CORENFLOS, K. FATRAS, N. FOURNIER, L. GAUTHERON, N. T. GAYRAUD, H. JANATI, A. RAKOTOMAMONJY, I. REDKO, A. ROLET, A. SCHUTZ, V. SEGUY, D. J. SUTHERLAND, R. TAVENARD, A. TONG, AND T. VAYER, *Pot: Python optimal transport*, Journal of Machine Learning Research, 22 (2021), pp. 1–8, <http://jmlr.org/papers/v22/20-451.html>.
- [30] M. FORNASIER, S. LISINI, C. ORRIERI, AND G. SAVARÉ, *Mean-field optimal control as gamma-limit of finite agent controls*, European Journal of Applied Mathematics, 30 (2019), pp. 1153–1186.
- [31] M. FORNASIER AND F. SOLOMBRINO, *Mean-field optimal control*, ESAIM: Control, Optimisation and Calculus of Variations, 20 (2014), pp. 1123–1152.
- [32] S.-Y. HA AND E. TADMOR, *From particle to kinetic and hydrodynamic descriptions of flocking*, Kinetic and Related Models, 1 (2008), pp. 415–435, <https://doi.org/10.3934/krm.2008.1.415>, <https://www.aims sciences.org/article/id/ab0b07d7-56dc-4ba7-ab04-85dce5a3c0cc>.
- [33] C. R. HARRIS, K. J. MILLMAN, S. J. VAN DER WALT, R. GOMMERS, P. VIRTANEN, D. COURNAPEAU, E. WIESER, J. TAYLOR, S. BERG, N. J. SMITH, R. KERN, M. PICUS, S. HOYER, M. H. VAN KERKWIJK, M. BRETT, A. HALDANE, J. F. DEL RÍO, M. WIEBE, P. PETERSON, P. GÉRARD-MARCHANT, K. SHEPPARD, T. REDDY, W. WECKESSER, H. ABBASI, C. GOHLKE, AND T. E. OLIPHANT, *Array programming with NumPy*, Nature, 585 (2020), pp. 357–362, <https://doi.org/10.1038/s41586-020-2649-2>, <https://doi.org/10.1038/s41586-020-2649-2>.
- [34] A. HINDAWI, J.-B. POMET, AND L. RIFFORD, *Mass transportation with lq cost functions*, Acta applicandae mathematicae, 113 (2011), pp. 215–229.
- [35] C. JIMENEZ, A. MARIGONDA, AND M. QUINCAMPOIX, *Optimal control of multiagent systems in the wasserstein space*, Calculus of Variations and Partial Differential Equations, 59 (2020), p. 58.
- [36] I. KOBYZEV, S. J. PRINCE, AND M. A. BRUBAKER, *Normalizing flows: An introduction and review of current methods*, IEEE transactions on pattern analysis and machine intelligence, 43 (2020), pp. 3964–3979.
- [37] P.-L. LIONS AND S. MAS-GALLIC, *Une méthode particulière déterministe pour des équations diffusives non linéaires*, Comptes Rendus de l’Académie des Sciences-Series I-Mathematics, 332 (2001), pp. 369–376.
- [38] K. OELSCHLÄGER, *Large systems of interacting particles and the porous medium equation*, Journal of differential equations, 88 (1990), pp. 294–346.
- [39] C. ORRIERI, A. PORRETTA, AND G. SAVARÉ, *A variational approach to the mean field planning problem*, Journal of Functional Analysis, 277 (2019), pp. 1868–1957.
- [40] A. PASZKE, S. GROSS, F. MASSA, A. LERER, J. BRADBURY, G. CHANAN, T. KILLEEN, Z. LIN, N. GIMELSHEIN, L. ANTIGA, ET AL., *Pytorch: An imperative style, high-performance deep learning library*, Advances in neural information processing systems, 32 (2019).
- [41] A. PORRETTA, *On the planning problem for the mean field games system*, Dynamic Games and Applications, 4 (2014), pp. 231–256.
- [42] L. RUTHOTTO, S. J. OSHER, W. LI, L. NURBEKYAN, AND S. W. FUNG, *A machine learning framework for solving high-dimensional mean field game and mean field control problems*, Proceedings of the National Academy of Sciences, 117 (2020), pp. 9183–9193.
- [43] F. SANTAMBROGIO, *Optimal transport for applied mathematicians*, Birkäuser, NY, 55 (2015), p. 94.
- [44] C. VILLANI, *Topics in optimal transportation*, vol. 58, American Mathematical Soc., 2003.