A note on deterministic mean field games of controls with state constraints: existence of mild solutions

Jameson Graber∗  Sergio Mayorga†

January 24, 2023

Abstract

We show the existence of mild solutions for a first-order mean field game of controls under the state constraint that trajectories be confined in a close $d$ and bounded set in euclidean space. This extends the results of [CC18] to the case of a mean field game of controls. Our controls are velocities and we find that the existence of an equilibrium is complicated by the requirement that they should have enough regularity. We solve this by imposing a small Lipschitz constant on the dependence of the Lagrangian on the joint measure of states and controls, and showing that regular paths can be approximated within the same class of functions despite the constraint.

MSC: 28B20, 49J15, 49N70, 35Q91, 91A13, 91A23.

Keywords: mean field games, differential games, optimal control, calculus of variations, fixed point, Nash equilibrium.

1 Introduction

Mean field games have been introduced to study large-scale interactions between rational agents [LL07; HMC06]. In a mean field game, one assumes a population of anonymous identical players. A representative player solves an optimal control problem depending on a flow of measures $(\nu_t)_{0 \leq t \leq T}$, e.g.

$$\inf_{x(\cdot)} \left\{ \int_0^T l(t, x(t), \dot{x}(t), \nu_t) \, dt + l_T(x(T), \nu_T) \right\}$$

for a given running cost $l$ and final cost $l_T$. Here and throughout this article the game is deterministic, i.e. the players’ dynamics are not subject to noise. In most of the literature on mean field games, $\nu_t = m_t$ is a measure only on the state space. In this case a Nash equilibrium occurs when $m_t$ is precisely the distribution of states given that every player follows optimal trajectories of (1).

One then derives the following system of partial differential equations for the equilibrium:

$$\begin{cases}
-\partial_t u + h(t, x, \nabla u, m_t) = 0, & u(x, T) = l_T(x, m_T), \\
\partial_t m_t - \nabla \cdot (m_t \nabla h(t, x, \nabla u, m_t)) = 0, & m_t|_{t=0} = m_0,
\end{cases}$$

(2)

where $h(t, x, p, m)$ is the Legendre transform of $l(t, x, v, m)$ with respect to $v$. The first equation in System (2) is a backward-in-time Hamilton-Jacobi equation with terminal condition $l_T$ that depends on $\mu_T$, and the second is a forward-in-time continuity equation with given initial condition $m_0$. System (2) is now commonplace in mean field game theory, especially in the case where $h$ can be written in the form $h(t, x, p, m) = H(x, p) - F(x, m)$. See [Car12, CG15, Car13, GM18] for results on the well-posedness of such systems.

∗Department of Mathematics, Baylor University. Jameson_Graber@baylor.edu
†Euler International Mathematics Institute, Saint Petersburg. mayorga@eulerinstitute.ru
A natural extension is to consider \( \nu_t \) a measure on the Cartesian product of the state space with the space of controls. In this article we will take controls to be velocities. Then a Nash equilibrium occurs when \( \nu_t \) is precisely the distribution of all possible values \( (x(t), \dot{x}(t)) \) for optimal trajectories of \( I \). In this case we have a mean field game of controls [CL18] (also called extended mean field game [GPV14, GV16]). The corresponding system of partial differential equations now reads

\[
\begin{align*}
-\partial_t u + h(t, x, \nabla u, \nu_t) &= 0, \\
\partial_t m_t - \nabla \cdot (m_t \nabla p h(t, x, \nabla u, \nu_t)) &= 0, \\
(I, -\nabla_p h(t, \cdot, \nabla u, \nu_t)) &= m_t = \nu_t,
\end{align*}
\]

where \( I \) is the identity mapping and \( \# \) is the push-forward operator; thus the third equation in \( I \) means that \( \nu_t \) is the push-forward of \( m_t \) through the map \( x \mapsto (x, -\nabla_p h(t, x, \nabla u(x, t), \nu_t)) \). Cardaliaguet and Lehalle give an abstract result on the well-posedness of System \( I \) over \( \mathbb{R}^d \times [0, T] \) in [CL18] (see also [GPV14, GV16, GMP21]).

In this paper we consider mean field games of controls with state constraints, i.e., where the trajectory \( x(t) \) must remain within some given closed subset of \( \mathbb{R}^d \). For the more common case where \( \nu_t = m_t \) and \( h(t, x, p, m) = H(x, p) - F(x, m) \), a full analysis of first-order mean field games with state constraints has been completed by Cannarsa, Capuani, and Cardaliaguet [CC18, CCC18, CCC21]. Their program has three steps. The first is to prove the existence of “mild solutions” to the mean field game with state constraints, which is motivated in the following way. For sufficiently smooth Hamiltonians and terminal conditions, viscosity solutions of the Hamilton-Jacobi equation without state constraints automatically possess \( C^{1,1} \) regularity. This means that the optimal feedback control \( -\nabla_p H(x, \nabla u) \) is defined on a sufficiently “large” set so that the continuity equation is well-posed. In the presence of state constraints, it is known that such regularity can fail for Hamilton-Jacobi equations, e.g., [CM08, Section 1.1]. Thus, a different approach to the existence of Nash equilibria is required. Start with the Lagrangian formulation of the problem, which is taken from [BCS17], rewriting the optimal control problem \( I \) as depending on a measure \( \eta \) on the space of possible paths. By definition, a mild solution occurs when \( \eta \)-a.e. trajectory is a minimizing trajectory; such an \( \eta \) is called a constrained equilibrium. A multivalued mapping \( \eta \mapsto E(\eta) \) is associated with this definition, to which Cannarsa and Capuani apply the Kakutani fixed point theorem [CC18], thereby showing the existence of mild solutions. Once this first step is completed, the next step is to study regularity properties of optimizers to control problems with state constraints [CCC18] and apply the results to the coupled system [CCC21].

Our goal is to follow this same program for mean field games of controls with state constraints. For the present study we will address only the first step, namely the existence of appropriately defined mild solutions. In doing so, we encounter some difficulties that did not appear in [CC18]. We need to come up with an appropriate definition of constrained equilibrium that involves both states and controls, and show that the analogue of the mapping \( E \) can be defined from a compact set into itself, and that it is closed. For compactness, since we are considering measures on both states and velocities, we require a second-order estimate on optimal trajectories; this is in contrast to [CC18], in which only first-order estimates were needed in the first step. Fortunately, we are able to apply results from [CCC18] to obtain the desired compactness. To show that \( E \) is closed, we need a result on the approximation of constrained trajectories, which has a completely different flavor from the corresponding result in [CC18, Section 3.1]. Our argument, which may have independent interest, has a more geometric flavor; see Section 4.

In [BGP21], the authors have also studied a class of mean field games of controls with mixed control-state constraints, using a Lagrangian approach as we do. In their model, the distributions of controls enters into the cost to players exclusively via a price function, which depends on the average of controls. They assume a qualification assumption (Hypothesis (H5)) on the controls that allows them to prove the regularity of minimizers to the control problem, followed by well-posedness of the price function as a fixed point. By contrast, we will assume no restriction on the controls, but only on the states. Regularity of optimizers will come from the results of [CCC18] on Euler-Lagrange equations for constrained optimal control. Moreover, we assume a more abstract dependence on the distribution of controls, where the Lagrangian is Lipschitz in the variable \( \nu \) with respect to a.
Wasserstein metric; the trade-off is that we make a smallness assumption on the Lipschitz constant (see Section 3).

The motivation for our study is far from purely mathematical. Indeed, many applications of mean field game theory in economics require mean field games of controls with state constraints [Ach+14; Ach+17; CS17]. We consider the present article an important step toward rigorously establishing the well-posedness of these models. Let us now present a brief overview of the main points in this article.

In section 2 we fix some terminology and borrow a result of CCC18 on the regularity of the minimizers of the constrained problem for a cost functional with time-dependent Lagrangian \( l \) (Theorem 2.2). All the trajectories are to remain in a closed, bounded smooth domain \( \Omega \), the space of states. Let \( \Gamma = \{(\gamma(\cdot) = (x(\cdot), u(\cdot)) \mid x(\cdot) \in \Gamma_1, u(\cdot) \in \Gamma_2\} \) be a set of “pairs” of trajectories and controls, on the time interval \([0,T]\); see 4 or 5 below for full definiteness. Suppose \( \eta \) is a probability measure on the set \( \Gamma \), and for \( \gamma \in \Gamma \), \( c_\gamma(t) = c(t,x(t),u(t)) \), with the additional property that the first marginal of \( (c_0)\#\eta \) equals a measure \( m_0 \) on \( \Omega \) fixed beforehand. For any given \( x_0 \in \Omega \), consider the functional

\[
I^{x_0}[u(\cdot) ; \eta] := \int_0^T l(t, x^n_n(t), u(t), (e_t)\#\eta) dt + l_T(x(T), (e_T)\#\eta)
\]

where \( x^n_n(\cdot) \) is the trajectory starting at \( x_0 \) with velocity \( u(\cdot) \). This functional can be viewed as acting on \( \Gamma_2 \), the set of second components of \( \Gamma \). Roughly speaking, for a hopeful definition of equilibrium, we would like the sets of minimizers of this functional in \( \Gamma \) to be in the same class as the (second components of the) paths \( \gamma \) that “make up” the input measure \( \eta \) on \( \Gamma \). This is achieved by Corollary 5.3 below. It comes at the expense of requiring the Lipschitz constant of \( l \) and \( D_u l \) with respect to its last variable (the measure variable) sufficiently small, so at this moment we warn the reader that the result in this paper is valid under this “small data” condition in regards of the measure variable; this restriction does not appear in the case with no state constraints [CL18].

Having guaranteed that data \( \eta \) in the space of probability measures on the suitable set \( \Gamma^K \) produces minimizers with second components in \( \Gamma^K_2 \), we proceed to declare \( \eta \) a constrained equilibrium of mean field game of controls (‘cemfgcs’, Definition 5.1 below, inspired directly by CCC18), if \( \eta \)- a.e. \( \gamma = (x(\cdot), u(\cdot)) \in \Gamma^K \) has the property that \( u(\cdot) \) is an optimal control for \( I^{x_0}[\cdot ; \eta] \). A set-valued mapping \( E : P_{\mathcal{F}_0}(\Gamma^K) \to P_{\mathcal{F}_0}(\Gamma^K) \) is associated with this definition (see Remark 5.2), and the main task becomes to show that \( E \) has a fixed point. Now, to prove that \( E \) does not take on empty values requires some work: the multi-set mapping \( (\eta, x_0) \mapsto \Gamma^K \) is closed, which is shown in Lemma 5.3, and it is there where we use Proposition 4.3 for the following reason, in somewhat vague terms: one wishes to show that a control \( u(\cdot) \) with initial state \( x_0 \) is optimal, provided there is a sequence \( x_n \to x_0 \) with a corresponding sequence of optimal \( u_n(\cdot) \) with initial states \( x_n \), and such that the controls \( u_n(\cdot) \) converge to \( u(\cdot) \). To compare the cost of the control \( u(\cdot) \) with that of some other, arbitrary control \( \tilde{u}(\cdot) \) (corresponding to the same initial state \( x_0 \)), one would like to approximate \( \tilde{u}(\cdot) \) by a sequence \( \tilde{u}^n(\cdot) \) with initial states \( x_n \), use the fact that the cost of \( u_n(\cdot) \) is less than the cost of \( \tilde{u}^n(\cdot) \), and then apply the continuity of the functional to pass to the limit and conclude that the cost of \( u(\cdot) \) is less than the cost of \( \tilde{u}(\cdot) \). The construction of the approximating sequence \( \{u_n(\cdot)\}^\infty_1 \) is considerably more involved in our case than in CCC18, where the velocities paths are allowed to have discontinuities. Thus, we have given Proposition 4.3 its own section.

The existence of a cemfgs then follows by the Kakutani-Fan-Glicksberg theorem; see Lemmas 5.4 and 5.5. This justifies the notion of mild solution (Definition 5.6 and Theorem 5.7), essentially just saying that we have a value function \( V(t,x) \) given as the solution of an optimal control problem that is “in equilibrium” with its data \( \eta \); the distribution of the set of its optimal pairs \( (x(\cdot), u(\cdot)) \) is precisely the support of \( \eta \), the distribution of all trajectories and control histories.

2 Preliminaries

In this section we will fix our notation and the main hypotheses of our result and import a regularity result in calculus of variations CCC18.
2.1 Notation and hypotheses

Whenever \( S \) is a complete, separable space, with metric \( d_S \), if \( \mathcal{B}(S) \) denotes the set of Borel probability measures on \( S \) and \( s_0 \in S \) is an arbitrary element, the subset

\[
\mathcal{P}(S) = \{ \eta \in \mathcal{B}(S) \mid \int_S d_S(s, s_0)\eta(s) < \infty \}
\]

is endowed with the 1-Wasserstein distance

\[
d_S(\eta^1, \eta^2) = \inf_{\Gamma(\eta^1, \eta^2)} \int_{S \times S} d_S(s^1, s^2)\gamma(ds^1, ds^2),
\]

where

\[
\Gamma(\eta^1, \eta^2) = \{ \gamma \in \mathcal{B}(S \times S) \mid \pi^1\gamma = \eta^1, \pi^2\gamma = \eta^2 \}.
\]

It is known that, if \( S \) happens to be compact, then \( \mathcal{P}(S) \) is a compact metric space with the distance \( d_S \); see, for instance, Remark 5.1.5 and Proposition 7.1.5 in [AGS08].

If \( \eta \) is a measure, \( \text{spt}(\eta) \) will denote the support of the measure; every Borel measure on a separable metric space has support [Bog07].

For the full extent of this paper we fix a \( C^3 \) bounded domain in \( \mathbb{R}^d \), i.e. a bounded, connected open set \( \Omega \) in \( \mathbb{R}^d \) whose boundary \( \partial\Omega \) is a surface of class \( C^3 \), and \( X \subset \mathbb{R}^d \) will be a fixed open bounded set including \( \Omega \). Let

\[
d_\Omega(x) = \inf\{|x - y| \mid y \in \Omega\}, \quad d_{\Omega^c}(x) = \inf\{|x - y| \mid y \notin \Omega\}
\]

and

\[
b_\Omega(x) = d_\Omega(x) - d_{\Omega^c}(x),
\]

\( x \in X; b_\Omega \) is called the oriented boundary distance. It can be shown that there exists a positive number \( \rho_0 \), such that, if we denote the neighbourhood of \( \partial\Omega \) of radius \( \rho_0 \) by \( \Sigma_{\rho_0} \), then \( b_\Omega \) is of class \( C^2 \) on \( \Sigma_{\rho_0} \) is of class \( C^3 \), i.e. a bounded, connected, twice continuously differentiable with all derivatives up to second order bounded.

We fix a number \( T > 0 \), and a continuous function

\[
l : [0, T] \times X \times \mathbb{R}^d \times \mathcal{P}(M) \rightarrow \mathbb{R},
\]

\[
(t, x, v, \nu) \mapsto l(t, x, v, \nu),
\]

where

\[
M := \bar{\Omega} \times \mathbb{R}^d,
\]

with the following properties:

(L-i) (bound on \( l \) and \( Dl \) at \( v = 0 \)) the function \( l \) is \( C^1 \) in \((x, v)\) and there is a constant \( n_1 > 0 \) such that

\[
|l(t, x, 0, \nu)| + |D_v l(t, x, 0, \nu)| + |D_{vv} l(t, x, 0, \nu)| \leq n_1
\]

for all \((t, x, \nu) \in [0, T] \times X \times \mathcal{P}(M)\).

(L-ii) (a) (structural bounds on the gradient of \( D_v l \)) the function \( D_v l \) is \( C^1 \) in \((t, x, v)\), and there is a constant \( c > 1 \) such that

\[
\frac{1}{c} l_d \leq D_{vv} l(t, x, v, \nu) \leq c l_d,
\]

where \( I_d \) is the \( d \times d \) identity matrix, and a constant \( k_1 \) such that

\[
|D_{vv}^2 l(t, x, v, \nu)| \leq k_1 (1 + |v|),
\]
Lemma 2.1. (The analogous properties of $l$ of $P_l$) We also fix a function $l: [0, T] \times X \times \mathbb{R}^d \times \mathcal{P}(M)$; for all $(t, x, v, \nu) \in [0, T] \times X \times \mathbb{R}^d \times \mathcal{P}(M)$:

(b) (Lipschitz continuity of $l$ and $D_v l$ in the measure) the function $D_v l$ is Lipschitz continuous with respect to the variable $\nu \in \mathcal{P}(M)$: there is a constant $c_1$ such that

$$|l(t, x, v, \nu_1) - l(t, x, v, \nu_2)| \leq c_1 d_M(\nu_1, \nu_2), \quad \nu_1, \nu_2 \in \mathcal{P}(M)$$

$$|D_v l(t, x, v, \nu) - D_v l(t, x, v, \nu_2)| \leq c_1 d_M(\nu_1, \nu_2), \quad \nu_1, \nu_2 \in \mathcal{P}(M)$$

for all $(t, x, v) \in [0, T] \times X \times \mathbb{R}^d$;

(L-iii) ($|v|$-dependent Lipschitz continuity in time of $l$ and $D_v l$) there is a constant $\kappa_1$ such that

$$|l(t, x, v, \nu) - l(s, x, v, \nu)| \leq \kappa_1 (1 + |v|^2)|t - s|,$$

$$|D_v l(t, x, v, \nu) - D_v l(s, x, v, \nu)| \leq \kappa_1 (1 + |v|)|t - s|$$

for all $t, s \in [0, T], (x, v, \nu) \in X \times \mathbb{R}^d \times \mathcal{P}(M)$.

We also fix a function

$$l_T: X \times \mathcal{P}(M) \longrightarrow \mathbb{R}$$

$$(x, \nu) \longmapsto l_T(x, \nu),$$

with the following properties:

(T-i) Both $l_T$ and its derivative in $x$, $Dl_T$, are bounded on $X \times \mathcal{P}(M)$.

(T-ii) Both $l_T$ and $Dl_T$ are Lipschitz continuous in $\nu$ with Lipschitz constant $c_1$ (with respect to the 1-Wasserstein distance $d_M$ on the space $\mathcal{P}(M)$).

This implies, in particular, that $l_T$ is bounded over subsets $X \times \mathcal{K}$, where $\mathcal{K}$ is any compact subset of $\mathcal{P}(M)$.

For every $(t, x, \nu) \in [0, T] \times X \times \mathcal{P}(\bar{\Omega} \times \mathbb{R}^d)$, let $h(t, x, \cdot, \nu)$ be the Legendre transform of $l(t, x, \cdot, \nu)$, i.e.

$$h(t, x, p, \nu) = l^*(t, x, p, \nu) := \sup_{v \in \mathbb{R}^d} \{p \cdot v - l(t, x, v, \nu)\}.$$

Lemma 2.1. (The analogous properties of $h$) It follows that:

(H-i) (bound on $h$ and $Dh$ at $p = 0$) the function $h$ is $C^1$ in $(x, p)$ and there is a constant $n_2 > 0$ which depends only on $n_1$, such that

$$|h(t, x, 0, \nu)| + |D_x h(t, x, 0, \nu)| + |D_p h(t, x, 0, \nu)| \leq n_2$$

for all $(t, x, \nu) \in [0, T] \times X \times \mathcal{P}(M)$.

(H-ii) (a) (structural bounds on the gradient of $D_p h$) the function $D_p h$ is $C^1$ in $(t, x, p)$, and

$$\frac{1}{c} I_d \leq D_{pp}^2 h(t, x, p, \nu) \leq c I_d,$$

and there is a constant $k_2 = k_2(n_2, c)$ (i.e. $k_2$ depends only on $n_2$ and $c$) such that

$$|D_{xp}^2 h(t, x, p, \nu)| \leq k_2(1 + |p|),$$

for all $(t, x, p, \nu) \in [0, T] \times X \times \mathbb{R}^d \times \mathcal{P}(M)$;

(b) (Lipschitz continuity of $h$ and $D_p h$ in the measure) the function $D_p h$ is Lipschitz continuous with respect to the variable $\nu \in \mathcal{P}(M)$ and

$$|D_p h(t, x, p, \nu_1) - D_p h(t, x, p, \nu_2)| \leq c_1 d_M(\nu_1, \nu_2), \quad \nu_1, \nu_2 \in \mathcal{P}(M)$$

for all $(t, x, p) \in [0, T] \times X \times \mathbb{R}^d$;
(H-iii) \(|p|\)-dependent Lipschitz continuity in time of \(D_p h\) there is a constant \(\kappa_2 = \kappa_2(n_2, c, \kappa_1)\) such that

\[
|h(t, x, v, \nu) - h(s, x, v, \nu)| \leq \kappa_2(1 + |p|^2)|t - s|,
\]

\[
|D_p h(t, x, v, \nu) - D_p h(s, x, v, \nu)| \leq \kappa_2(1 + |p|)|t - s|
\]

for all \(t, s \in [0, T]\), \((x, p, \nu) \in X \times \mathbb{R}^d \times \mathcal{P}(M)\).

Proof. All these properties are commonly found in the literature; we only show the proof of (H-iii), since it involves the measure variable, though this does not make the problem harder. Fix \(p \in \mathbb{R}^d\), let \(v_1\) and \(v_2\) be the (unique, in this case) maximizers of the functions \(v \mapsto p \cdot v - l(t, x, v, \nu_1)\) and \(v \mapsto p \cdot v - l(t, x, v, \nu_2)\) respectively. This implies that

\[
p = D_p l(t, x, v_1, \nu_1) = D_p l(t, x, v_2, \nu_2).
\]

It can be shown (e.g. see [Dac08, p. 138]) that

\[
v_1 = D_p h(t, x, p, \nu_1), \quad v_2 = D_p h(t, x, p, \nu_2).
\]

By the lower bound on second derivative \(D^2_{\nu\nu} l\) above, we have

\[
\frac{1}{c}|v_2 - v_1|^2 \leq (D_p l(t, x, v_2, \nu) - D_p l(t, x, v_1, \nu)) \cdot (v_2 - v_1),
\]

for any \(t, x, \nu\). Therefore

\[
\frac{1}{c}|v_2 - v_1|^2 \leq (D_p l(t, x, v_2, \nu_1) - D_p l(t, x, v_2, \nu_2)) \cdot (v_2 - v_1)
\]

\[
\quad + (D_p l(t, x, v_2, \nu_2) - D_p l(t, x, v_1, \nu_1)) \cdot (v_2 - v_1)
\]

\[
\leq c_1 d_M(v_1, v_2)|v_2 - v_1| + (p - p) \cdot (v_2 - v_1) = c_1 d_M(v_1, v_2)|v_2 - v_1|,
\]

from which the desired inequality follows. \(\square\)

Whenever \(K_1, K_2\) are positive numbers, we denote \(K = (K_1, K_2)\). Let

\[
\Gamma^K_1 := \{x : [0, T] \rightarrow \bar{\Omega} \mid x(\cdot) \text{ is continuously differentiable}, |\dot{x}(t)| \leq K_1, 0 \leq t \leq T\}.
\]

To define the set of regular controls, for every \(x_0 \in \bar{\Omega}\) we let

\[
\Gamma^K_2 :\{ u : [0, T] \rightarrow \bar{B}_{K_1}(0) \mid u(\cdot) \text{ is Lipschitz,} \quad |\dot{u}(t)| \leq K_2 \text{ a.e., } x_0 + \int_0^t u(\tau)d\tau \in \bar{\Omega} \text{ for } t \in [0, T]\}
\]

for \(x_0 \in \bar{\Omega}\), and

\[
\Gamma^K_2 := \bigcup_{x_0 \in \bar{\Omega}} \Gamma^K_2, x_0.
\]

Both \(\Gamma^K_1\) and \(\Gamma^K_2\) are compact with respect to the uniform metrics \(d_1\) and \(d_2\):

\[
d_1(x_1, x_2) := \max_{0 \leq t \leq T} |x_1(t) - x_2(t)|, \quad d_2(u_1, u_2) := \max_{0 \leq t \leq T} |u_1(t) - u_2(t)|.
\]

Let

\[
\Gamma^K = \{ \gamma = (x, u) : [0, T] \rightarrow \mathbb{R}^d \times \mathbb{R}^d \mid x(\cdot) \in \Gamma^K_1, u(\cdot) \in \Gamma^K_2\}.
\]

By definition, the paths \(\gamma \in \Gamma^K\) are bounded. On \(\Gamma^K\) we put the metric

\[
d_1((x_1, u_1), (x_2, u_2)) := d_1(x_1, x_2) + d_2(u_1, u_2),
\]

so that \(\Gamma^K\) is compact with respect to this metric. We choose not to specify \(K\) in the notation \(d_1\) because the formula is the same regardless of the values of \(K_1, K_2\).
2.2 Regularity of constrained minimizers

In order to clarify the discussion that follows, we introduce the following sets:

\[ A\Gamma_1 := AC(0,T;\Omega), \]

\[ A\Gamma_2 := \{ u : [0,T] \to \mathbb{R}^d \mid \exists x(\cdot) \in A\Gamma_1 \mid u(t) = \dot{x}(t) \text{ a.e.} \} \subset L^1([0,T];\mathbb{R}^d), \]

\[ A\Gamma = \{ \gamma = (x,u) : [0,T] \to \Omega \times \mathbb{R}^d \mid x(\cdot) \in A\Gamma_1, u(\cdot) \in A\Gamma_2 \}. \tag{4} \]

We note that \((x,u) \in A\Gamma\) does not necessarily mean that \(\dot{x} = u\); \(A\Gamma_2\) are merely functions in \(L^1([0,T]:\mathbb{R}^d)\) that are derivatives of paths in \(AC(0,T;\Omega)\). Clearly, \(\Gamma^K_1 \subset A\Gamma_1\) and \(\Gamma^K_2 \subset A\Gamma_2\) for every \(K\). The set \(\Gamma^K_2\) consists of Lipschitz paths (with Lipschitz constant \(K_2\)) that can be used as controls for the dynamics (the corresponding integrated paths remain in \(\bar{\Omega}\)). Regardless of how we turn \(A\Gamma\) into a measurable space, to each probability measure \(\eta\) on \(A\Gamma\) and \(x(\cdot) \in A\Gamma_1\) we may associate the cost

\[ \int_0^T l(t,x(t),\dot{x}(t),(e_t)_\#\eta)dt + l_T(x(T),(e_T)_\#\eta), \]

where \(e_t : \Gamma^K \to \mathbb{R}^d \times \mathbb{R}^d\), \(0 \leq t \leq T\) are the evaluation mappings:

\[ e_t(\gamma) = e_t((x(\cdot),u(\cdot))) := (x(t),u(t)). \]

This cost can be considered as a functional acting on the set of paths \(A\Gamma_1\) or its subset \(\Gamma^K_1\); alternatively, it may be considered as a functional on the set of controls \(A\Gamma_2\) or its subset \(\Gamma^K_2\).

Given a path \(\nu\) in \(\mathcal{P}(M)\), it is known that hypotheses \(L(\cdot)\) through \(L(\cdot\cdot)\) above are more than enough to ensure that if one considers the functional

\[ J[x(\cdot);\nu] := \int_0^T l(t,x(t),\dot{x}(t),\nu_t)dt + l_T(x(T),\nu_T) \tag{5} \]

to be defined on the set \(A\Gamma_1 = AC(0,T;\bar{\Omega})\), then there exists a minimizer of this functional in \(AC^2(0,T;\bar{\Omega})\) where the minimization is taken with respect to the paths \(x(\cdot)\) with a common initial point \(x(0) = x_0\); see, e.g., [CS04].

In the present manuscript we wish to have minimizers with better regularity, which is why we introduced the sets \(\Gamma^K\). For our purposes we will heavily rely on [CCC18, Theorem 3.1] in section 3 below.

**Theorem 2.2.** [Borrowed \& adapted from [CCC18]] Let \(l\) be subject to the conditions above, \(h = h^*, \)
let \(\nu_t\) be a Lipschitz path in \(\mathcal{P}(M)\), and consider the functional \(J[ \cdot ;\nu]\) defined above, on the set of \(AC(0,T;\bar{\Omega})\) paths \(x(\cdot)\) with a common initial point \(x(0) = x_0\), \(x_0 \in \bar{\Omega}\). Let \(x(\cdot)\) be a minimizer. Then \(x(\cdot)\) has the following properties:

1. \(x(\cdot) \in C^{1,1}([0,T];\bar{\Omega})\), i.e. it is continuously differentiable with Lipschitz derivative,
2. a Lipschitz function \(p : [0,T] \to \mathbb{R}^d\) exists such that

\[ \dot{x}(t) = -D_p h(t,x(t),p(t),\nu_t), \quad t \in [0,T], \]

\[ \dot{p}(t) = D_x h(t,x(t),p(t),\nu_t) - L(t,x(t),p(t),\nu_t)1_{\partial \Omega}(x(t)) Db_{\Omega}(x(t)), \quad a.e. \ t \in [0,T], \]

\[ p(T) = Dl_T(x(T),\nu_T) + \beta Db_{\Omega}(x(T))1_{\partial \Omega}(x(T)). \]

In this theorem,

- \(\beta\) is a constant bounded by either 1 or the \(2c\sup_{x \in X}|D_p h(T,x,Dl_T(x,\nu_T),\nu_T)|\)

\(^1\)The dot as subindex is meant to distinguish the path \(t \mapsto \nu_t\) from any single measure value of this path.
• $A$ is a bounded continuous function defined on $[0, T] \times \Sigma_{\rho_0} \times \mathbb{R}^d \times \mathcal{P}(\Omega)$.

• $1_{\partial \Omega}(\cdot)$ is the characteristic function of the boundary $\partial \Omega$.

**Remark 2.3.** The result [CCC18, Theorem 3.1] deals with a time-dependent Lagrangian $f(t, x, v)$ and no dependence on the measure. To adapt it to our needs, we will consider the path $t \mapsto \nu_t$ as given, put

$$f(t, x, v) := l(t, x, v, \nu_t), \quad 0 \leq t \leq T,$$

and then use [CCC18, Theorem 3.1], which is supported in several facts that are established in that paper. We will carry out now a careful analysis of these, suited to our needs.

*In what follows, $m_0$ denotes a fixed element of $\mathcal{P}(\bar{\Omega})$.*

### 3 Estimates and a small data condition

As explained in [CCC18, p. 179, 181], we have:

**Lemma 3.1.** As a consequence of hypotheses (LIII), there are constants $C = C(c, n_1)$ and $C(k_1, n_1)$ such that

$$|D_v l(t, x, v, \nu_t)| \leq C(c, n_1)(1 + |v|) \quad \text{and} \quad |D_x l(t, x, v, \nu_t)| \leq C(k_1, n_1)(1 + |v|^2),$$

independently of the path $\nu$. and likewise constants $C(c, n_2)$ and $C(k_2, n_2)$ such that

$$|D_p h(t, x, p, \nu_t)| \leq C(c, n_2)(1 + |p|) \quad \text{and} \quad |D_x h(t, x, p, \nu_t)| \leq C(k_2, n_2)(1 + |p|^2)$$

independently of the path $\nu$.

In this section, we fix a Lipschitz path $\nu$ with values in $\mathcal{P}(M)$.

**Sketch of proof of Theorem (2.2).** First, it is assumed that $X = \mathbb{R}^d$. That is, assume the hypotheses (LIII) and (TBB) are stated for functions $l : [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}(M) \rightarrow \mathbb{R}$ and $l_T : \mathbb{R}^d \times \mathcal{P}(M) \rightarrow \mathbb{R}$.

Let us denote $g(x) = l_T(x, \nu_T)$, $x \in X$. Let

$$\theta := T|C(c, n_1) + n_1 + 2 \sup_{x \in X}|g(x)|. \quad (6)$$

We refer to Remark 2.3 where $f(t, x, v) := l(t, x, v, \nu_t)$. For fixed $x_0 \in \Omega$, $\epsilon > 0$, $\delta > 0$, the penalized problem (compare with (LIII)) is to minimize

$$J^{\epsilon, \delta}[x(\cdot)] := \int_0^T \left[ f(t, x(t), \dot{x}(t)) + \frac{d_\Omega(x(t))}{\epsilon} \right] dt + \frac{d_\Omega(x(T))}{\delta} + g(x(T))$$

over the set $AC(0, T; \mathbb{R}^d)$ with common initial points $x(0) = x_0$. It is known [Ces83] that there exists at least one minimizer for each $x_0 \in \Omega$; the set of such is denoted by $\mathcal{X}_{\epsilon, \delta}[x_0]$; the set of minimizers of the unpenalized problem (i.e. (LIII)) with initial state at $x_0 \in \Omega$ is denoted by $\mathcal{X}[x_0]$. We will now describe, by steps, how CCC18, Theorem 3.1 is established.

**Step 1.** For any $\rho \in (0, \rho_0)$ there exists $\epsilon(\rho) > 0$ such that, picking $\epsilon \in (0, \epsilon(\rho))$ and $\delta > 0$, it follows that $\sup_{t \in [0, T]} d_\Omega(x(t)) \leq \rho$ whenever $x_0 \in \Omega$ and $x(\cdot) \in \mathcal{X}_{\epsilon, \delta}(x_0)$. That is, the problem can be sufficiently penalized so that minimizers remain as close to $\Omega$ as desired. We must determine the function $\epsilon(\rho)$ precisely. It is readily shown that if $x(\cdot) \in \mathcal{X}_{\epsilon, \delta}(x_0)$ then

$$\frac{1}{4c} \int_0^T ||\dot{x}(t)||^2 + \frac{1}{\epsilon} d_\Omega(x(t)) dt \leq \theta.$$
This inequality implies that
\[ |d_{\Omega}(x(t)) - d_{\Omega}(x(s))| \leq \sqrt{4c\theta}|t - s|^{1/2} \]
for all \( s, t \in [0, T] \) and \( x(\cdot) \in \mathcal{X}_{\varepsilon, \delta}[x_0], x_0 \in \bar{\Omega} \). Denote, for the moment, \( H_0 = \sqrt{4c\theta} \); thus, \( H_0 \) is the 1/2-Hölder constant of \( d_{\Omega}(x(\cdot)) \). Let
\[ \epsilon_0 := \frac{\rho_0^3}{64c^2\theta^2} = \frac{\rho_0^3}{16c\theta H_0^2}. \]
Let \( J \) be any subinterval of \([0, T]\) of length \( |J| = \frac{1}{2}\rho_0^3/H_0^2 \). Then
\[ \frac{2}{3}H_0|J|^{1/2} + \frac{4\theta c\epsilon}{|J|} < \rho_0. \] (7)
Now,
\[ \int_J d_{\Omega}(x(t))dt \leq \int_0^T d_{\Omega}(x(t))dt \leq 4\theta c\epsilon \]
and, since \( d_{\Omega}(x(\cdot)) \) is 1/2-Hölder with constant \( H_0 \), a calculation (e.g. arguing by contradiction) shows that this implies that \( d_{\Omega}(x(t)) \) is no larger than the left-hand side of (7). Therefore
\[ d_{\Omega}(x(t)) < \rho_0 \]
for all \( t \in J \). Since the subinterval \( J \) of that length is arbitrary, this argument shows the following: if \( 0 < \rho \leq \rho_0 \), then there is a function \( \epsilon(\rho) \), namely, \( \epsilon(\rho) = \frac{\rho^3}{16c\theta H_0^2} \), such that, for any \( x_0 \in \bar{\Omega}, \delta > 0 \), \( x(\cdot) \in \mathcal{X}_{\varepsilon, \delta}[x_0] \), the distance between \( x(t) \) and \( \Omega \) is never larger than \( \rho \). We see that the function \( \epsilon(\cdot) \) is increasing, and \( \epsilon_0 = \epsilon(\rho_0) \). From now on \( \delta \) will have a specific value, namely:
\[ \delta := \min\left\{ \frac{1}{2mc}, 1 \right\}, \quad \text{where } m := \sup_{x \in \mathcal{X}} |D_{h}h(T, x, D_T(x, \nu_T), \nu_T)|. \] (8)
Because of Lemma 3.4 \( m \) is a constant whose bound involves only \( c, n_2, \|D_T\|_{\infty} \), and, hence, the same is true of \( \delta \). Lastly, let
\[ C_1 := 8c + 8c \sup_{x \in \mathcal{X}} |D_T(x, \nu_T)|^2 + 2k_2 + \tilde{\kappa}_1(c_1, \text{Lip}(\nu))(T + 4c\theta), \] (9)
where
\[ \tilde{\kappa}_1(c_1, \text{Lip}(\nu)) := \kappa_1 + c_1 \text{Lip}(\nu) \] (10)
and we will sometimes omit the brackets that indicate the dependence on the other constants. We set \( f(t, x, v) = l(t, x, v, \nu_t) \), as explained in Remark 2.3 where \( f \) is the running cost in [CCC18]. It is important to identify precisely the time regularity for \( f \), i.e. the analogue of (L.31) for \( f \). We have
\begin{align*}
|f(t, x, v) - f(s, x, v)| &= |l(t, x, v, \nu_t) - l(s, x, v, \nu_s)| \\
&\leq \kappa_1(1 + |v|^2)|t - s| + c_1 d_M(\nu_t, \nu_s) \\
&\leq (\kappa_1(1 + |v|^2) + c_1 \text{Lip}(\nu))|t - s| \\
&\leq (\kappa_1 + c_1 \text{Lip}(\nu))(1 + |v|^2)|t - s| = \tilde{\kappa}_1(c_1, \text{Lip}(\nu))(1 + |v|^2)|t - s| \quad \text{(11)}
\end{align*}
Similarly,
\[ |D_v f(t, x, v) - D_v f(s, x, v)| \leq \tilde{\kappa}_1(c_1, \text{Lip}(\nu))(1 + |v|)|t - s|. \] (12)
Step 2. Let \( \rho \in (0, \rho_0), \epsilon \in (0, \epsilon(\rho)), x_0 \in \bar{\Omega} \) and \( x(\cdot) \in \mathcal{X}_{\varepsilon, \delta}[x_0] \). Then \( x(\cdot) \) is of class \( C^{1,1} \) and there is a Lipschitz path \( p(\cdot) \) in \( \mathbb{R}^d \) such that, for a.e. \( t \in [0, T] \),
\[
\begin{aligned}
\dot{x}(t) &= D_p f^*(t, x(t), p(t)) \\
\dot{p}(t) &= D_x f^*(t, x(t), p(t)) - \frac{\lambda(t)}{\epsilon} D_{b\Omega}(x(t)) \\
p(T) &= D_g(x(T)) + \frac{\beta}{\delta} D_{b\Omega}(x(T))
\end{aligned}
\] (13)

9
and
\[ |p(t)|^2 \leq 4c \left[ \frac{1}{\epsilon} d_\Omega(x(t)) + \frac{C_1(c, n_1, n_2, \theta, \kappa_1)}{\delta^2} \right] \] (14)
where \( C_1 \) was introduced in (10). Note that \( C_1 \) depends on \( c, n_1, n_2, \theta \) and on \( \kappa_1 \); \( \kappa_1 \) is the function (10) and \( \theta \) is given in (4), \( f^* \) is the Legendre transform of \( f \), i.e. \( f^*(t, x, p) = h(t, x, p, \nu_t) \) and \( \lambda : \quad [0, T] \to [0, 1] \) is a measurable map. The proof is based on Pontryagin's maximum principle \[ Vin00 \]. The vector \( p \) is the adjoint vector in that context; see formula (3.32) in \[ CCC18 \], and the paragraph that follows it. Said formula means that the value of \( p(t) \) depends on the limiting subdifferential in \((t, x)\) of the penalized Lagrangian \( f(t, x, v) + d_\Omega(x)/\epsilon \) at the point \((t, x(t), \dot{x}(t))\): if \( x(t) \in \Omega \), then it is \( D_x f(t, x(t), \dot{x}(t)) \); if \( 0 < b_\Omega(x(t)) < \rho \), then it is \( D_x f(t, x(t), \dot{x}(t)) + \frac{1}{\epsilon} D_\Omega b_\Omega(x(t)) \), while if \( x(t) \notin \partial \Omega \), then Lemma 2.1 in \[ CCC18 \] provides a formula for the limiting subdifferential of the function \( d_\Omega \) in terms of the \( C^{1,1} \) function \( b_\Omega \), gives that \( p(t) \) is a point in the segment \( D_x f(t, x(t), \dot{x}(t)) + \frac{1}{\epsilon} D_\Omega b_\Omega(x(t)) \) belongs to \([0, 1]\) for a.e. \( t \in [0, T] \).

Step 3. Suppose \( \epsilon \leq \epsilon_0 \), \( x(\cdot) \in X_{\epsilon, \delta}[x_0], |a, b| \subset [0, \Omega], d_\Omega(x(a)) = 0, d_\Omega(x(t)) > 0 \) for \( t \in (a, b) \), and \( d_\Omega(x(b)) = 0 \) or \( b = T \). Suppose \( d_\Omega(x(\cdot)) \) attains its maximum on \([a, b]\) at \( t' \). Then, it can be shown \[ CCC18 \] Lemma 3.6 and 3.7 that
\[ \frac{d^2}{dt^2} d_\Omega(x(t)) \big|_{t = t'} \leq 0. \] (15)
We mention this is where \( \delta \) has to be as in (5). Using the formulas \[ ICC18 \], inequality (15) gives:
\[ 0 \geq [D^2 d_\Omega(x(t')))D_p f^*(t', x(t'), p(t'))] \cdot D_p f^*(t', x(t'), p(t')) - D d_\Omega(x(t')) \cdot D_{pp} f^*(t', x(t'), p(t')) + D d_\Omega(x(t')) \cdot D_{pp} f^*(t', x(t'), p(t')) D_p f^*(t', x(t'), p(t')) - D d_\Omega(x(t')) \cdot D_{pp} f^*(t', x(t'), p(t')) D_p f^*(t', x(t'), p(t')) + \frac{1}{\epsilon} D d_\Omega(x(t')) \cdot [D_{pp} f^*(t', x(t'), p(t'))] D d_\Omega(x(t')))].
\]
At this moment we must remark that, letting
\[ \tilde{\kappa}_2 = \kappa_1 C(c, n_2), \]
one can verify that inequalities (12) and (13) hold for \( f^* \) with \( \tilde{\kappa}_2 \) in place of \( \kappa_1 \), that is, in particular:
\[ |D_p f^*(t, x, p) - D_p f^*(s, x, p)| \leq \tilde{\kappa}_2 (1 + |p|)|t - s|. \] (16)
To continue with the argument, recall the bounds on \( D^2_{pp} f^* \) from the hypotheses (i.e. Lemma 2.1), due to which the term in the last line is bounded below by \( \frac{1}{2} \| b_\Omega \|_{C^2} \). Using the formula for the bounds of the terms in the other lines (namely: \( 10, (H.iii) \) and Lemma 3.1), after dividing throughout by \( \| b_\Omega \|_{C^2} \) we get:\n\[ \frac{1}{\epsilon c} \leq C(c, n_2)(1 + |p|)^2 + \kappa_2(c,1, \text{Lip}(\nu_\cdot))(1 + |p|) + k_2 C(c, n_2)(1 + |p|)^2 + \frac{1}{2} dc C(k_2, n_2) \] \[ \leq [2C(c, n_2) + 2\kappa_2(c,1, \text{Lip}(\nu_\cdot)) + 2k_2 C(c, n_2) + \frac{1}{2} dc C(k_2, n_2)](1 + |p|^2) \leq [C(c, n_2)(2 + 2k_2) + 2\kappa_2(c,1, \text{Lip}(\nu_\cdot)) + \frac{1}{2} dc C(k_2, n_2)](1 + 4c \left[ \frac{1}{\epsilon} d_\Omega(x(t')) + \frac{C_1 + \tilde{\kappa}_1 (T + 4c \theta)}{\delta^2} \right]), \]
that is,
\[ \frac{1}{\epsilon c} \leq (\bar{C}_2 + 2\tilde{\kappa}_2)(1 + 4c \left[ \frac{1}{\epsilon} d_\Omega(x(t')) + \frac{C_1 + \tilde{\kappa}_1 (T + 4c \theta)}{\delta^2} \right]), \] (17)
where we have set —recall (14)—
\[
\tilde{C}_1 = 8c + 8c \sup_{x \in \mathcal{X}} |Dt_x(\nu, \nu_t)|^2 + 2k_2, \quad (\implies C_1 = \tilde{C}_1 + \tilde{\kappa}(T + 4c\theta))
\]
\[
\tilde{C}_2 = C(c, n_2)(2 + 2k_2) + \frac{1}{2} dcC(k_2, n_2)
\]
to distinguish constants that do not depend on Lip(\nu). We want to set the value of \( \epsilon \) precisely, so that the obtained inequality (17) is impossible. This would imply that for this value of \( \epsilon \), the path \( x(\cdot) \) remains in \( \Omega \). By what was discussed above, if \( \epsilon \leq \epsilon_0 \) and \( x(\cdot) \in \mathcal{X}^{\epsilon, \delta}[x_0] \), then \( d_\Omega(x(t)) \leq \rho(\epsilon) \) for all \( t \). Let \( \epsilon_1 \) be so that \( \rho_1 := \epsilon^{-1}(\epsilon_1) \) satisfies
\[\gamma \frac{1}{\epsilon_1} = (\tilde{C}_2 + 2\tilde{\kappa}_2)4c\rho_1,\]
where \( \gamma \in (0, 1) \) is to be chosen later. We can find \( \epsilon_1 \) explicitly by the formula for \( \rho(\epsilon) \), indeed:
\[
\epsilon_1 := \frac{\gamma^3}{64\epsilon^2\theta^2} \cdot \frac{1}{4^3\epsilon^6(\tilde{C}_2 + 2\tilde{\kappa}_2)^3} = \frac{\gamma^3}{4^6\epsilon^2\theta^3(\tilde{C}_2 + 2\tilde{\kappa}_2)^3}
\]
In this case from (17) it follows that
\[
\frac{1}{\epsilon_1} (1 - \gamma) \leq (\tilde{C}_2 + 2\tilde{\kappa}_2)[1 + \frac{\tilde{C}_1 + \tilde{\kappa}_1(T + 4c\theta)}{\delta^2}],
\]
i.e.
\[
4^6\theta^2\epsilon^7(\tilde{C}_2 + 2\tilde{\kappa}_2)^2 \left( \frac{1 - \gamma}{\gamma^3} \right) \leq 1 + \frac{\tilde{C}_1 + \tilde{\kappa}_1(T + 4c\theta)}{\delta^2},
\]
giving
\[
\frac{1 - \gamma}{\gamma^3} \leq \frac{1}{4^6\theta^2\epsilon^7(\tilde{C}_2 + 2\tilde{\kappa}_2)^2} \left( 1 + \frac{\tilde{C}_1 + \tilde{\kappa}_1(T + 4c\theta)}{\delta^2} \right)
\]
Since \( (1 - \gamma)/\gamma^3 \rightarrow +\infty \) as \( \gamma \rightarrow 0^+ \), it is clear we can set \( \gamma \) to a value that will make the latter inequality absurd. To make it easier to deal with, note that \( (1 - \gamma)/\gamma < (1 - \gamma)/\gamma^3 \). So we can set \( \gamma \) so that \( (1 - \gamma)/\gamma \) is larger than the right-hand side of the inequality. This is achieved by
\[
\gamma := \left[ 1 + \frac{1}{4^6\theta^2\epsilon^7(\tilde{C}_2 + 2\tilde{\kappa}_2)^2} \left( 1 + \frac{\tilde{C}_1 + \tilde{\kappa}_1(T + 4c\theta)}{\delta^2} \right) \right]^{-1},
\]
giving, for \( \epsilon_1 \),
\[
\epsilon_1 := \frac{1}{4^6\theta^2\epsilon^7(\tilde{C}_2 + 2\tilde{\kappa}_2)^2} \left[ 1 + \frac{1}{4^6\theta^2\epsilon^7(\tilde{C}_2 + 2\tilde{\kappa}_2)^2} \left( 1 + \frac{\tilde{C}_1 + \tilde{\kappa}_1(T + 4c\theta)}{\delta^2} \right) \right]^{-3}
\]
With \( \epsilon = \epsilon_1 \), as long as \( \epsilon_1 \leq \epsilon_0 \), we, therefore, have a contradiction, because (17) does not hold. We ignore a priori, however, whether \( \epsilon_1 \) of formula (19) is not larger than \( \epsilon_0 \): if it is larger, then the point \( x(t') \) (where \( x(\cdot) \in \mathcal{X}^{\epsilon_1, \delta}[x_0] \)) is not within the distance \( \rho_0 \) that makes the calculations above valid. The value \( \epsilon_0 \) depends directly on \( \rho_0 \) and thus the geometry of \( \Omega \). If \( \epsilon_1 \geq \epsilon_0 \), we can choose \( \gamma \) so that \( \gamma < 1 \) and (see (18)) \( \gamma^3/(4^6\theta^2\epsilon^7(\tilde{C}_2 + 2\tilde{\kappa}_2)^3) \) is less than \( \epsilon_0 \). This is achieved by the extra requirement \( \gamma < \min\{1/2, (\epsilon_0 4^6\theta^2\epsilon^7\tilde{C}_2)^{1/3}\} \). Correspondingly, by letting
\[
\epsilon_1 := \min\left\{ \frac{1}{8}, \frac{\epsilon_0 \tilde{C}_2}{(\tilde{C}_2 + 2\tilde{\kappa}_2)^3}, \frac{\left[ 1 + \frac{1}{4^6\theta^2\epsilon^7(\tilde{C}_2 + 2\tilde{\kappa}_2)^2} \left( 1 + \frac{\tilde{C}_1 + \tilde{\kappa}_1(T + 4c\theta)}{\delta^2} \right) \right]^{-3}}{(\tilde{C}_2 + 2\tilde{\kappa}_2)^3} \right\},
\]
we ensure that the calculations above are valid and the contradiction is obtained. Thus, if \( \epsilon \leq \epsilon_1 \), where \( \epsilon_1 \) is given by (20), the minimizer \( x(\cdot) \in \mathcal{X}^{\epsilon, \delta}[x_0] \) is sure to remain in \( \Omega \) in the penalized
problem. For such a value of ε, \( x(\cdot) \) is a minimizer of the penalized problem if and only if it is a minimizer of the unpenalized problem. Thus, we can say the following about \( \dot{p}(t) \):

\[
|\dot{p}(t)| \leq C(k_2, n_2)(1 + 4\frac{cC_1}{\delta^2}) + P_3(c_1\text{Lip}(\nu)),
\]

where \( P_3 = P_3(y) \) is a polynomial of third degree in \( y \) whose coefficients depend on all data constants other than \( c_1 \) and Lip \((\nu)\). We can denote the right-hand side of (21) by Lip \((p(\cdot))\).

Last step. Bounds are established and an explicit formula is obtained for \( \lambda(t) \), proving that it’s continuous. Then \( \Lambda(t, x(t), p(t), \nu_t) := \frac{\mathcal{N}(t)}{c} \) is a definition and the formula they obtain for \( \lambda(t) \) proves that \( \Lambda \) is continuous. Of importance to us is to note the following: for \( \epsilon \leq \epsilon_1 \) with \( \epsilon_1 \) as above (i.e. (20)), since \( d_\Omega(x(t)) = 0 \) for \( x(\cdot) \in \mathcal{X}'[x_0] \) for every \( t \in [0, T] \) (and therefore in \( \mathcal{X}[x_0] \) also), we have, from (14), that for every \( x(\cdot) \in \mathcal{X}[x_0] \),

\[
|p(t)|^2 \leq 4c\frac{C_1}{\delta^2},
\]

where \( C_1 \), we remind again, is

\[
C_1 = C_1(c, n_1, n_2, \theta, \tilde{\kappa}_1) = 8c + 8c \sup_{x \in \mathcal{X}} |Dl_T(x, \nu_T)|^2 + 2k_2 + \tilde{\kappa}_1(T + 4c\theta)
= \tilde{C}_1 + \tilde{\kappa}_1(T + 4c\theta);
\]

see the definition following formula (17), and (19).

The case when \( \mathcal{X} \) is a bounded open set including \( \bar{\Omega} \) is treated by extending the running and terminal costs to \( \mathbb{R}^d \) in a way that preserves the hypotheses (L-iii) and (T-i,ii). Thus Theorem 2.2 is established.

We will now record the Lipschitz constant of the optimal trajectory of \( J[\cdot, \nu], \) and of its velocity. For the former, the bound on \( D_pJ \) from Lemma 6.11 and the fact that for the optimal trajectory \( x(\cdot) \) we have \( \dot{x}(t) = -D_pJ(t, x(t), p(t), \nu_t) = -D_pJ^*(t, x(t), p(t)) \) from Theorem 2.2. That is,

\[
|\dot{x}(t)| \leq C(c, n_2)(1 + 2\sqrt{cC_1}),
\]

so

\[
0 \leq t \leq T, \text{ so the right-hand side is a Lipschitz constant for } x(\cdot); \text{ we can denote it by Lip}(x(\cdot)).
\]

For the latter, using the formula for \( \dot{x}(t) \) from Theorem 2.2 and the hypotheses on the data, we estimate:

\[
|\dot{x}(s) - \dot{x}(t)| = |D_pJ^*(s, x(s), p(s), \nu_s) - D_pJ^*(t, x(t), p(t), \nu_t)|
\leq |D_pJ^*(s, x(s), p(s)) - D_pJ^*(t, x(t), p(s))| + |D_pJ^*(t, x(t), p(s)) - D_pJ^*(t, x(t), p(t))| + |D_pJ^*(t, x(t), p(t)) - D_pJ^*(s, x(s), p(s))|
\]

so

\[
|\dot{x}(s) - \dot{x}(t)| \leq \tilde{\kappa}_2 \times (1 + |p(t)|)|t - s| \quad \text{(by (10))}
+ k_2 \times (1 + |p(t)|)|\text{Lip}(x(\cdot))||t - s| \quad \text{(by H.-iii)}
+ c\sqrt{d} \text{Lip}(p(\cdot))|t - s| \quad \text{(by H.-iii)}
\]

so

\[
0 \leq t, s \leq T. \text{ Therefore, relying on the formula for the Lipschitz constant of } p(\cdot), \text{ i.e. (21) and the bound on } |p(t)|, \text{ i.e. (22), we get}
\]

\[
|\dot{x}(s) - \dot{x}(t)| \leq |s - t| \left( 1 + 2\frac{\sqrt{cC_1}}{\delta} \right) \left[ \tilde{\kappa}_2 + k_2 C(c, n_2)(1 + 2\sqrt{cC_1}/\delta) \right] + c\sqrt{d}\text{Lip}(p(\cdot)).
\]

12
We have assumed that \( l, l_T \) satisfy the hypotheses of section 2.1. Let \( \nu \) be a Lipschitz path in \( \mathcal{P}(M) \), \( x_0 \in \bar{\Omega} \), and \( l, l_T \) satisfy the hypotheses of section 2.1. Then there exist numbers \( b_1, b_2, \ldots, b_{10} \), that depend on all the hypotheses constants of section 2.1, with the exception of \( c_1 \) and \( \text{Lip}(\nu) \), for which, if

\[
K_1 > b_1 + b_2 \sqrt{b_3 + b_4 c_1 \text{Lip}(\nu)},
\]

\[
K_2 > b_5 + b_6 \sqrt{b_7 + b_8 c_1 \text{Lip}(\nu)} + b_9 c_1 \text{Lip}(\nu) + b_{10} (c_1 \text{Lip}(\nu))^3,
\]

then any minimizer \( x(\cdot) \) of

\[
\min \{ J[x(\cdot); \nu] \mid x(\cdot) \in A\Gamma_1 \}
\]

is such that \((x(\cdot), \dot{x}(\cdot))\) lies in \( \Gamma^K \), where \( K = (K_1, K_2) \).

Suppose now that \( \eta \in \mathcal{P}(\Gamma^K) \), and \( \nu_T = (e_T)\# \eta \), and that \( \nu \) is defined by \( \nu_t = (e_t)\# \eta \), \( 0 \leq t \leq T \). It can be easily checked that \( \text{Lip}(\nu) = K_1 + K_2 \).

**Corollary 3.3.** Let \( x_0 \in \bar{\Omega}, l, l_T \) satisfy the hypotheses of section 2.1. If the constant \( c_1 \) is sufficiently small, then there exists \( K = (K_1, K_2) \) such that for any \( \eta \in \mathcal{P}(\Gamma^K) \), setting

\[
\nu_t := (e_t)\# \eta, \quad 0 \leq t \leq T,
\]

any minimizer \( x(\cdot) \) of

\[
\min \{ J[x(\cdot); \nu] \mid x(\cdot) \in A\Gamma_1 \}
\]

satisfies: \((x(\cdot), \dot{x}(\cdot))\) \in \( \Gamma^K \).

**Proof.** As noted, for any choice of \( K = (K_1, K_2) \), if \( \eta \in \mathcal{P}(\Gamma^K) \) and \( \nu_t := (e_t)\# \eta \), \( 0 \leq t \leq T \), then \( \text{Lip}(\nu) = K_1 + K_2 \). Denote the right-hand sides of Lemma 3.2 for such a choice by \( \tilde{K}_1, \tilde{K}_2 \), i.e.

\[
\tilde{K}_1 := b_1 + b_2 \sqrt{b_3 + b_4 c_1 (K_1 + K_2)}
\]

\[
\tilde{K}_2 := b_5 + b_6 \sqrt{b_7 + b_8 c_1 (K_1 + K_2)} + b_9 c_1 (K_1 + K_2) + b_{10} (c_1 (K_1 + K_2))^3,
\]

where \( b_1, \ldots, b_{10} \) are as in the lemma, which says that \( \tilde{K}_1 \) and \( \tilde{K}_2 \) will be the Lipschitz constant of the minimizer \( x(\cdot) \) and of its velocity \( \dot{x}(\cdot) \), respectively.

Now choose a number \( K_1 \) such that \( K_1 > b_1 + b_2 \sqrt{b_3 + 1} \) and \( K_2 \) so that \( K_2 > b_5 + b_6 \sqrt{b_7 + 1} \). Then, by letting \( c_1 \to 0^+ \) in the expressions for \( \tilde{K}_1 \) and \( \tilde{K}_2 \), we see that there will be some \( c_1 \) such that \( \tilde{K}_1 < K_1 \) and \( \tilde{K}_2 < K_2 \).

Note that there is no conflict between the two latter propositions in the space where \( \nu \) is taken, since any probability measure \( \nu \) on \( \bar{\Omega} \times \bar{B}_{K_1}(0) \) can be regarded as a probability measure \( \tilde{\nu} \) on \( \bar{\Omega} \times \mathbb{R}^d \) by setting \( \tilde{\nu}(E) := \nu(E \cap (\bar{\Omega} \times \bar{B}_{K_1}(0))) \) for every Borel subset \( E \) of \( \bar{\Omega} \times \mathbb{R}^d \).

Henceforth we suppose that \( c_1 \) and \( K \) are as in Corollary 3.3.

## 4 Approximation of constrained paths

We dedicate this section to a proposition which we will use but may also be of independent interest.

Denote \( K^{+1} = (K_1 + 1, K_2 + 1) \).
Proposition 4.1. Let \( \{x_k\}_{k=1}^\infty \subset \Omega \) be a sequence converging to \( x_0 \in \Omega \), and let \( u^0(\cdot) \in \Gamma^{K,x_0}_2 \). There exists a sequence \( \{u^k\}_{k=1}^\infty \subset \Gamma^{K+1}_2 \), with \( u^k(\cdot) \in \Gamma^{K+1}_2 \) for every \( k \), such that \( u^k(t) \to u^0(t) \) uniformly in \( t \in [0,T] \).

Before proceeding with the proof of Proposition 4.1, we mention the following:

Lemma 4.2. Let \( \Omega \) be a domain of class \( C^3 \). Then for each \( \xi \in \partial \Omega \), there exist open sets \( U \supset \xi, V \subset \mathbb{R}^d \) and a diffeomorphism \( \psi : U \to V \subset \mathbb{R}^d \) such that the following conditions hold:

1. \( \psi \) and \( \psi^{-1} \) are both \( C^2 \);
2. \( \psi \) preserves boundary and interior of \( \Omega \) in the following sense:
   \[
   \psi(U \cap \overline{\Omega}) = V \cap (\mathbb{R}^{d-1} \times [0,\infty)) \quad \text{and} \quad \psi(U \cap \partial \Omega) = V \cap (\mathbb{R}^{d-1} \times \{0\});
   \]
3. if \( e_1, \ldots, e_d \) are the standard basis vectors in \( \mathbb{R}^d \) and if \( x \in U \), then \( \psi(x) \cdot e_d = d(x, \partial \Omega) \) if \( x \in \overline{\Omega} \) and \( \psi(x) \cdot e_d = d(x, \partial \Omega) \) if \( x \notin \overline{\Omega} \).

Proof. See the Appendix. \( \square \)

The proof of Proposition 4.1 can be best understood by considering a simplifying case. For the moment let \( \Omega = (0,\infty) \); this is not allowed by our hypotheses because it is unbounded, but nevertheless it will be instructive. Assuming \( x_k(0) \to x(0) \), we want to construct \( x_k(t) \) such that \( \dot{x}_k(t) \) converges uniformly to \( \dot{x}(t) \) while \( x_k(t) \geq 0 \) for all \( t \). If \( x_k(0) \geq x(0) \), we find no difficulty letting \( \dot{x}_k(t) = \dot{x}(t) \), i.e. the paths run in parallel. Suppose, however, that \( x(0) > x_k(0) \geq 0 \). In this case we define

\[
 x_k(t) = \frac{x_k(0)}{x(0)} x(t).
\]

Then \( x_k(\cdot) \) is \( C^{1,1} \), and we have an estimate:

\[
 \sup_{t \in [0,T]} |\dot{x}_k(t) - \dot{x}(t)| \leq \sup_{t \in [0,T]} \left| \frac{x_k(0)}{x(0)} \right| |\dot{x}(t)| \leq K \left| \frac{x_k(0)}{x(0)} \right|.
\]

Since \( x_k(t) \geq 0 \) for all \( t \), we see that \( x_k(\cdot) \) satisfies the desired properties.

For a general (but bounded) domain \( \Omega \), the main idea of the proof is to follow this example whenever \( x(t) \) is close to the boundary. Since the boundary is curved and multi-dimensional, we will choose local coordinates in which to imitate this simpler case. In the simple example above, \( x_k(t) \) is identical to the distance from the boundary; in higher dimensions, we will use Lemma 4.2 to work in coordinates where the last one is the distance from the boundary.

With this basic idea in mind, the steps of the proof are as follows:

1. Cover the boundary \( \partial \Omega \) by neighborhoods with local coordinates as in Lemma 4.2.
2. Divide \( \Omega \) into a “collar region,” denoted \( V \), and an “interior region” \( \Omega \setminus V \). When a trajectory is in the collar region, we represent it using local coordinates from the first step, while if it is in the interior region, we can use standard coordinates. Partition the time horizon \([0,T]\) using evenly spaced intervals \([t_\ell, t_{\ell+1}]\), such that the trajectory can be represented in a single coordinate system on each \([t_\ell, t_{\ell+1}]\).
3. Define \( x_k(t) \) recursively: given \( x_k(t) \) up to \( t = t_\ell \), define it on \([t_\ell, t_{\ell+1}]\) by using the simple example above as a guide when \( x_k(t) \) must stay in the collar region \( V \). Show that \( \max_{t \in [t_\ell, t_{\ell+1}]} |\dot{x}_k(t) - \dot{x}(t)| \) is controlled by \( \max_{t \in [0,T]} |\dot{x}_k(t) - \dot{x}(t)| \). (Actually, it is a bit more complicated than this; see Assumption 4.3 below.)
4. Conclude that \( \dot{x}_k(t) \) converges uniformly to \( \dot{x}(t) \), which implies \( x_k(t) \) converges uniformly to \( x(t) \) as well.

We hope this outline will keep the reader oriented through the technical details, which we now give below.
Proof of Proposition 4.1. Let $x(t) := x^0(t), 0 \leq t \leq T$. Clearly, it will be sufficient to show that there is a sequence $x_k(\cdot) \in C^{1,1}([0, T])$ such that $x_k(0) = x_k$ for all $k$, and $x_k(t) \to x(t), \dot{x}_k(t) \to \dot{x}(t)$ uniformly in $t \in [0, T]$.

Step 1: Since $\Omega$ is a domain of class $C^3$, Lemma 4.2 applies. For each $\xi \in \partial \Omega$ we can find $r(\xi) > 0$, an open set $U_\xi \subset \mathbb{R}^d$, and a diffeomorphism $\psi_\xi : U_\xi \to B(0, 2r(\xi)) \subset \mathbb{R}^d$ such that $\psi_\xi$ and $\psi_\xi^{-1}$ are both $C^2$ and we have $\psi_\xi(\xi) = 0$,

$$\psi_\xi(U_\xi \cap \Omega) = B(0, 2r(\xi)) \cap (\mathbb{R}^{d-1} \times [0, \infty)) \quad \text{and} \quad \psi_\xi(U_\xi \cap \partial \Omega) = B(0, 2r(\xi)) \cap (\mathbb{R}^{d-1} \times \{0\});$$

if $e_1, \ldots, e_d$ are the standard basis vectors in $\mathbb{R}^d$ and if $x \in U_\xi$, then $\psi_\xi(x) \cdot e_d \geq 0$ if and only if $x \in \Omega$, in which case $\psi_\xi(x) \cdot e_d = d(x, \partial \Omega)$. Let $\hat{r}(\xi) > 0$ be such that

$$B(\xi, 4\hat{r}(\xi)) \subset U_\xi \quad \text{and} \quad \psi_\xi(B(\xi, 4\hat{r}(\xi))) \subset B(0, r(\xi)).$$

Since $\overline{\Omega}$ is compact we can find $\xi_1, \ldots, \xi_N$ such that $\partial \Omega \subset \bigcup_{j=1}^N B(\xi_j, \hat{r}(\xi_j)) =: \hat{U}$. We will denote $r_j = r(\xi_j), \hat{r}_j = \hat{r}(\xi_j), U_j = U_{\xi_j}, B_j = B(\xi_j, \hat{r}(\xi_j))$, and $\psi_j = \psi_{\xi_j}$, and we will define $U = \bigcup_{j=1}^N U_j$. Pick some $j_0$ such that $x_0 \in B_{j_0}$. We take

$$C = \max\left\{\|D\psi_j\|_\infty, \|D^2\psi_j\|_\infty, \|D\psi_j^{-1}\|_\infty, \|D^2\psi_j^{-1}\|_\infty : j = 1, \ldots, N\right\},$$

(23) (Note that $C \geq 1$.) Set

$$r = \min\left\{\min\{r(\xi_j) : j = 1, \ldots, N\}, \frac{1}{2}\right\}, \quad \hat{r} = \min\left\{\min\{\hat{r}(\xi_j) : j = 1, \ldots, N\}, \frac{1}{2}\right\}$$

and

$$K_0 = \max\{K_1, K_2, 1\}.$$

(24)

Step 2: Define

$$V = \{x \in \overline{\Omega} : d(x, \partial \Omega) \leq 2\hat{r}\}.$$

If $x \in V$ then there exists $y \in \partial \Omega$ such that $|x - y| \leq 2\hat{r}$; then if $y \in B(\xi_j, \hat{r}(\xi_j))$, we have $x \in B(\xi_j, 3\hat{r}(\xi_j))$. It follows that $V \subset \bigcup_{j=1}^N B(\xi_j, 3\hat{r}(\xi_j))$. Let $L = \left[K_1 T \hat{r}^{-1}\right] + 1$ and set $t_\ell = \frac{\ell}{L}$ for $\ell = 0, 1, \ldots, L$. Let

$$M := \frac{3C^5 K_0}{\hat{r}^2},$$

and let $\varepsilon > 0$ be an arbitrarily small number such that

$$(3M)^L \varepsilon < \min\{\hat{r}, r\}.$$  

(26)

We may assume without loss of generality that

$$|x_k - x_0| < \varepsilon.$$

Fix $\ell \in \{0, \ldots, L\}$. Notice that

$$|x(t) - x(t_\ell)| \leq K_1 (t - t_\ell) \leq \frac{K_1 T}{L} < \hat{r} \quad \forall t \in [t_\ell, t_{\ell+1}].$$

(27)

We know that either $d(x(t), \partial \Omega) \geq 2\hat{r}$ or else $x(t_\ell) \in B(\xi_j, 3\hat{r}(\xi_j))$ for some $j_\ell$. In the first case, (27) implies that $d(x(t), \partial \Omega) \geq \hat{r}$ for all $t \in [t_\ell, t_{\ell+1}]$; in the second case, (27) implies that $x(t) \in B(\xi_j, 4\hat{r}(\xi_j)) \subset U_j$ for all $t \in [t_\ell, t_{\ell+1}]$.

Step 3: To define $x_k(t)$, we will proceed in a recursive, piece-wise fashion. First we will assume that for some given $\ell \in \{0, 1, \ldots, L - 1\}$, $x_k(t)$ has been defined on $[0, t_\ell]$ in such a way that, if $\ell \geq 1$, we have $x_k(\cdot) \in C^{1,1}([0, t_\ell]; \overline{\Omega})$, and so that the following assumptions are satisfied:
Assumption 4.3. (a) If $\ell \geq 1$ and $x(t_s) \notin V$, then $x_k(t_s) \in \Omega$ and $\frac{d}{dt} x_k(t_s) = \dot{x}(t_s)$.

(b) If $\ell \geq 1$ and $x(t_s) \in V$, so that $x(t_s) \in B(\xi_{j_s}, 3\hat{r}(j_s))$ for some $j_s$, then $x_k(t_s) \in U_\ell$, and for $s = 1, \ldots, d - 1$, we have

$$\frac{d}{dt} \psi_j(x_k(t)) \cdot e_s|_{t=t_s} = \frac{d}{dt} \psi_j(x(t)) \cdot e_s|_{t=t_s}. \quad (28)$$

Moreover,

(i) if $\psi_j(x(t_s)) \cdot e_d > 0$, then

$$\frac{d}{dt} \psi_j(x_k(t)) \cdot e_d|_{t=t_s} = \psi_j(x_k(t_s)) \cdot e_d \frac{d}{dt} \psi_j(x(t)|_{t=t_s}; \quad (29)$$

(ii) if $\psi_j(x(t_s)) \cdot e_d = 0$, then

$$\frac{d}{dt} \psi_j(x_k(t)) \cdot e_d|_{t=t_s} = 0.$$

(c) We have the following estimates:

$$|x_k(t) - x(t)| \leq (3M)^\ell \varepsilon \quad \forall t \in [0, t_{\ell+1}], \quad \text{and}$$

$$|\dot{x}_k(t) - \dot{x}(t)| \leq (3M)^\ell \varepsilon \quad \forall t \in [0, t_{\ell}], \quad (30)$$

$$|d(x_k(t), \partial \Omega) - d(x(t), \partial \Omega)| \leq (3M)^\ell \varepsilon d(x(t), \partial \Omega),$$

where $M$ is defined in (20).

We will now define $x_k(t)$ on $[t_\ell, t_{\ell+1}]$ in such a way that $x_k(\cdot) \in C^{1,1}([0, t_{\ell+1}; \Omega])$ and the same assumptions are satisfied with $t_\ell$ replaced by $t_{\ell+1}$. We divide into four cases.

Case 1: First, consider the case in which $x(t_s) \notin V$ and $x(t_{s+1}) \notin V$. Then we define

$$x_k(t) = x_k(t_s) - x(t_s) + x(t) \quad \forall t \in [t_s, t_{s+1}].$$

It follows that

$$|x_k(t) - x(t)| = |x_k(t_s) - x(t_s)|, \quad \dot{x}_k(t) = \dot{x}(t) \quad \forall t \in [t_s, t_{s+1}].$$

Since Assumption 4.3(a) is satisfied, we see that $x_k$ is $C^{1,1}$ on all of $[0, t_{\ell+1})$. Moreover,

$$|x_k(t_s) - x(t_s)| < (3M)^\ell \varepsilon < \hat{r}$$

by Assumption 4.3(c) and equation (20). Since $d(x(t), \partial \Omega) \geq \hat{r}$, it follows that $x_k(t) \in \Omega$ for all $t \in [t_s, t_{s+1})$. It is straightforward to check that Assumption 4.3 is satisfied with $t_\ell$ replace by $t_{\ell+1}$.

Case 2: Next, consider the case in which $x(t_s) \notin V$ and $x(t_{s+1}) \in V$, so we have $x(t_{s+1}) \in B(\xi_{j_{s+1}}, 3\hat{r}(j_{s+1}))$. We will let $\lambda \in (0, 1)$ be small (depending on $\ell$) but fixed. Define $t_{\ell, \lambda} = \frac{(\ell+1-\lambda)T}{\lambda} = t_{\ell+1} - \lambda T$ where $\lambda_T := \frac{\lambda}{\ell}$. Now let

$$x_k(t) = x_k(t_s) - x(t_s) + x(t) \quad \forall t \in [t_s, t_{\ell, \lambda}].$$

Since Assumption 4.3(a) is satisfied, we see that $x_k(\cdot)$ is $C^{1,1}$ on all of $[0, t_{\ell, \lambda}]$. We also have

$$|x_k(t) - x(t)| = |x_k(t_s) - x(t_s)|, \quad \dot{x}_k(t) = \dot{x}(t) \quad \forall t \in [t_s, t_{\ell, \lambda}], \quad (31)$$

We obtain the estimate

$$|x_k(t_{\ell, \lambda}) - x(t_{\ell+1})| \leq |x_k(t_s) - x(t_s)| + |x(t_{\ell, \lambda}) - x(t_{\ell+1})|$$

$$\leq |x_k(t_s) - x(t_s)| + K_1 \lambda T$$

$$\leq (3M)^\ell \varepsilon + K_1 \lambda T, \quad (32)$$

We also have

$$|x_k(t_{\ell, \lambda}) - x(t_{\ell+1})| \leq |x_k(t_{\ell, \lambda}) - x(t_{\ell, \lambda})| + |x(t_{\ell, \lambda}) - x(t_{\ell+1})|$$

$$\leq |x_k(t_{\ell, \lambda}) - x(t_{\ell, \lambda})| + K_1 \lambda T$$

$$\leq (3M)^\ell \varepsilon + K_1 \lambda T,$$
which for \( \lambda \) small enough is less than \( \bar{r} \) by (20). It follows that \( x_k(t_{\ell, \lambda}) \in B(\xi_{j, t_{\ell, \lambda}}, 4\bar{r}(j_{t_{\ell, \lambda}})) \subset \psi_{j_{t_{\ell, \lambda}}^{-1}}(B(0, r(j_{t_{\ell, \lambda}}))) \). We also note for future reference that \( d(x_k(t_{\ell, \lambda}, \partial \Omega)) \leq \bar{r} + d(x(t_{\ell, \lambda}, \partial \Omega)) \leq 3\bar{r} \).

For \( \lambda \) small enough we have \( x(t) \in B(\xi_{j_{t_{\ell, \lambda}}}, 4\bar{r}(j_{t_{\ell, \lambda}})) \) for all \( t \in [t_{\ell, \lambda}, t_{\ell+1}] \). We define

\[
\tilde{x}(t) = \psi_{j_{t_{\ell, \lambda}}}(x(t)) \quad \forall t \in [t_{\ell, \lambda}, t_{\ell+1}], \quad \tilde{x}^j(t) = \tilde{x}(t) \cdot e_j, \quad j = 1, \ldots, d.
\]

Next, we will define \( \tilde{x}_k(t) \) on \([t_{\ell, \lambda}, t_{\ell+1}]\), show that it is also in \( B(0, 2\bar{r}(j_{t_{\ell, \lambda}})) \), and then define \( x_k(t) = \psi_{j_{t_{\ell, \lambda}}^{-1}}(\tilde{x}_k(t)) \) on \([t_{\ell, \lambda}, t_{\ell+1}]\).

Set

\[
\tilde{x}_k(t_{\ell, \lambda}) = \psi_{j_{t_{\ell, \lambda}}}(x_k(t_{\ell, \lambda})),
\]

\[
v_0 := \left. \frac{d}{dt} \right|_{t=t_{\ell, \lambda}} \psi_{j_{t_{\ell, \lambda}}}(x_k(t)) = D_\lambda \psi_{j_{t_{\ell, \lambda}}}(x_k(t_{\ell, \lambda})) \tilde{x}(t_{\ell, \lambda}),
\]

\[
v_1 := \tilde{x}(t_{\ell+1}) = \left. \frac{d}{dt} \right|_{t=t_{\ell+1}} \psi_{j_{t_{\ell, \lambda}}}(x(t)) = D_\lambda \psi_{j_{t_{\ell, \lambda}}}(x(t_{\ell+1})) \tilde{x}(t_{\ell+1}).
\]

Here we denote by \( \left. \frac{d}{dt} \right| \) the left-hand derivative, i.e.

\[
\left. \frac{d}{dt} \right| f(t) = \lim_{s \to t} \frac{f(s) - f(t)}{s - t}.
\]

Observe that \( |v_0|, |v_1| \leq C K_1 \), and

\[
|v_0 - v_1| \leq |D_\lambda \psi_{j_{t_{\ell, \lambda}}}(x_k(t_{\ell, \lambda})) \tilde{x}(t_{\ell, \lambda}) - D_\lambda \psi_{j_{t_{\ell, \lambda}}}(x(t_{\ell+1})) \tilde{x}(t_{\ell+1})| + |D_\lambda \psi_{j_{t_{\ell, \lambda}}}(x(t_{\ell+1})) \tilde{x}(t_{\ell, \lambda}) - D_\lambda \psi_{j_{t_{\ell, \lambda}}}(x(t_{\ell+1})) \tilde{x}(t_{\ell+1})|
\]

\[
\leq C K_1 |x_k(t_{\ell, \lambda}) - x(t_{\ell+1})| + C |\tilde{x}(t_{\ell, \lambda}) - \tilde{x}(t_{\ell+1})|
\]

\[
\leq C K_1 |x_k(t_{\ell, \lambda}) - x(t_{\ell+1})| + C K_1^2 \lambda_T + C K_1^2 \lambda_T
\]

\[
\leq C K_0(3M)^\ell \varepsilon + 2 C K_0^2 \lambda_T,
\]

where we have used (32). We will define

\[
\tilde{x}_k(t) = \tilde{x}_k(t_{\ell, \lambda}) + (t - t_{\ell, \lambda}) v_0 + \frac{1}{2}(t - t_{\ell, \lambda})^2 v
\]

with \( v = v(\lambda) \) chosen in order that

\[
\tilde{x}^j_k(t_{\ell+1}) = \left. \frac{\tilde{x}^j_k(t_{\ell+1})}{\tilde{x}^d_k(t_{\ell+1})} \right|_{t=t_{\ell+1}}^{v_j^0} = v_j^0 \quad \forall j = 1, \ldots, d - 1.
\]

Here for any vector \( v \in \mathbb{R}^d \), we denote \( v^j = v \cdot e_j \), the \( j \)th standard coordinate. We take \( \lambda \) small enough so that

\[
|\lambda_T| \leq \frac{\bar{r}}{C K_1} \implies \left| \frac{\lambda_T v_0^d}{2 \tilde{x}^d_k(t_{\ell+1})} \right| \leq \frac{1}{2}.
\]

The second inequality follows from the first after recalling that \( \tilde{x}^d_k(t_{\ell+1}) = d(x(t_{\ell+1}), \partial \Omega) \geq \bar{r} \) and \( |v_0^d| \leq C K_1 \). With this assumption, we can uniquely solve for \( v \) to get (33) and (34) with the formula

\[
v^d = \frac{1}{\lambda_T} \left( 1 - \frac{\lambda_T v_0^d}{2 \tilde{x}^d_k(t_{\ell+1})} \right)^{-1} \left( \frac{\tilde{x}^d_k(t_{\ell, \lambda}) + \lambda_T v_0^d}{\tilde{x}^d_k(t_{\ell+1})} v^d_0 - v^d_0 \right), \quad v^j = \frac{v^j_0 - v^d_0}{\lambda_T}.
\]

We estimate, using (35),

\[
|\lambda_T v^d| \leq 2 \left| \frac{\tilde{x}^d_k(t_{\ell, \lambda}) + \lambda_T v_0^d}{\tilde{x}^d_k(t_{\ell+1})} v^d_0 - v^d_0 \right| \leq 2 C K_1 \left( \frac{\tilde{\bar{r}} + C K_1 \lambda_T}{\tilde{r}} + 1 \right) \leq 2 C K_1 \left( \frac{\tilde{\bar{r}} + \bar{r}}{\tilde{r}} + 1 \right) = 10 C K_1.
\]
Now returning to the identity (34), we get
\[
|\dot{x}_k(t_{\ell+1}) - \dot{x}(t_{\ell+1})| = |\dot{x}_k(t_{\ell+1}) - \dot{x}(t_{\ell+1})| \\
= |\dot{x}_k(t_{\ell,\lambda}) - \dot{x}(t_{\ell+1}) + \lambda_T v_0^d + \frac{1}{2} \lambda T^2 \zeta d^2| \\
= \frac{C K_1}{r} (|x_k(t_{\ell,\lambda}) - x(t_{\ell+1})| + 6 C K_1 \lambda_T) \\
\leq \frac{C^2 K_1}{r} (|x_k(t) - x(t)| + 7 K_1 \lambda_T),
\]
where we have used (32) and (34). Next, we note that \( \dot{x}_k(\cdot) \) is linear. On the other hand, by the definitions (23) and (24), recalling \( t \) where we have recalled (32), (31) and (36). Coupled with (31) and (37) we obtain that, for all \( \dot{\tilde{\gamma}}(\cdot) \), we have
\[
|\dot{x}(t)| = |D^2 \psi_{j_{\ell+1}}(x(t)) |\dot{x}(t), \ddot{x}(t)| + D^2 \psi_{j_{\ell+1}}(x(t)) |\ddot{x}(t)| \leq C (K^2 + K_2) \leq 2 C K_0^2,
\]
and so we deduce that, for all \( t \in [t_{\ell,\lambda}, t_{\ell+1}], \)
\[
|\dot{x}_k(t) - \dot{x}(t)| \leq \frac{t_{\ell+1} - t}{t_{\ell+1} - t_{\ell,\lambda}} |x_0 - \dot{x}(t_{\ell,\lambda})| + \frac{t - t_{\ell,\lambda}}{t_{\ell+1} - t_{\ell,\lambda}} |\dot{x}_k(t_{\ell+1}) - \dot{x}(t_{\ell+1})| + 2 C K_0^2 \lambda_T \\
\leq |(D^2 \psi_{j_{\ell+1}}(x_k(t_{\ell,\lambda}))) \dot{x}(t_{\ell,\lambda})| + \frac{C^2 K_1}{r} (|x_k(t) - x(t)| + 7 K_1 \lambda_T) + 2 C K_0^2 \lambda_T \\
\leq \frac{2 C^2 K_0}{r} (|x(t) - x(t)| + 9 C K_0^2 \lambda_T),
\]
where we have recalled (32), (31) and (30). Coupled with (31) and (37) we obtain that, for all \( t \in [t_{\ell,\lambda}, t_{\ell+1}], \)
\[
|\ddot{x}_k(t) - \ddot{x}(t)| \leq |\ddot{x}_k(t_{\ell,\lambda}) - \ddot{x}(t_{\ell,\lambda})| + \int_{t_{\ell,\lambda}}^t |\ddot{x}_k(s) - \ddot{x}(s)| \, ds \\
\leq C |x_k(t_{\ell,\lambda}) - x(t_{\ell,\lambda})| + \int_{t_{\ell,\lambda}}^t |\ddot{x}_k(s) - \ddot{x}(s)| \, ds \\
\leq \frac{3 C^2 K_0}{r} |x(t) - x(t)| + \frac{9 C K_0^2}{r} \lambda_T \\
\leq M (3 M)^{\ell} \varepsilon + 3 M \lambda_T \leq (3 M)^{\ell+1} \varepsilon < \min \{r, \bar{r} \},
\]
where we have taken \( \lambda_T \leq 1 \) small enough. Therefore, \( x_k(t) \in B(0, 2 r(j_{\ell+1})) \). Also, recall from Step 2 that for this case, \( x^d(t) = d(x(t), \partial \Omega) \geq \bar{r} \) for all \( t \in [t_{\ell}, t_{\ell+1}] \); thus \( x^d(t) > 0 \) for \( t \in [t_{\ell,\lambda}, t_{\ell+1}] \) by (33). Thus we can define \( x_k(t) = \psi^{-1}_{j_{\ell+1}}(\tilde{x}_k(t)) \in \Omega \) on \( [t_{\ell,\lambda}, t_{\ell+1}] \), as desired. By construction, \( x_k(\cdot) \) is \( C^{1,1} \) on the interval \( [0, t_{\ell+1}] \); it suffices to observe that the left- and right-hand derivatives agree at \( t = t_{\ell,\lambda} \). Assumption (1.1.b) is satisfied with \( t_{\ell+1} \) in place of \( t_{\ell} \); in particular, (28) and (29) follow directly from (41). Moreover, we have the following estimates on the interval \( [t_{\ell,\lambda}, t_{\ell+1}] \). First, noting that
\[
|x_k(t) - x(t)| \leq C |\ddot{x}_k(t) - \ddot{x}(t)|,
\]
we deduce
\[
|x_k(t) - x(t)| \leq \frac{3 C^3 K_0}{r} |x_k(t) - x(t)| + \frac{9 C^3 K_0^2}{r} \lambda_T \leq M (3 M)^{\ell} \varepsilon + 3 M \lambda_T.
\]
Then we compute
\[
\dot{x}(t) - \dot{x}(t) = D^2 \psi_{j_{\ell+1}}(\tilde{x}_k(t)) (\dot{x}_k(t) - \dot{x}(t)) + (D^2 \psi_{j_{\ell+1}}(\dot{x}_k(t)) - D^2 \psi_{j_{\ell+1}}(\dot{x}(t))) \dot{x}(t) = \dot{x}(t).
\]
Applying estimates (37) and (38), and recalling (25) and (24), we get

\[
|\dot{x}_k(t) - \dot{x}(t)| \leq \frac{2C^3K_0}{\ell} |x_k(t) - x(t)| + \frac{9C^3K_0^2}{\ell} \lambda_T + C |\ddot{x}_k(t) - \ddot{x}(t)| K_1
\]

\[
\leq \frac{2C^3K_0}{\ell} |x_k(t) - x(t)| + \frac{9C^3K_0^2}{\ell} \lambda_T + \frac{3C^3K_0^2}{\ell} |x_k(t) - x(t)| + \frac{9C^3K_0^3}{\ell} \lambda_T,
\]

which, after recalling (25), using Assumption (13 c), and doing some straightforward estimates, becomes

\[
|\dot{x}_k(t) - \dot{x}(t)| \leq 2M(3M)^{\ell} \varepsilon + 6MK_0^2 \lambda_T.
\]

Finally, applying (38), we get

\[
\frac{|x_k(t_{\ell+1}) - x(t_{\ell+1})|}{d(x(t_{\ell+1}), \partial \Omega)} \leq \frac{1}{\ell} |x_k(t_{\ell+1}) - x(t_{\ell+1})| \leq \frac{3C^3K_0}{\ell^2} |x_k(t) - x(t)| + \frac{9C^3K_0^2}{\ell^2} \lambda_T,
\]

which implies

\[
|d(x_k(t_{\ell+1}), \partial \Omega) - d(x(t_{\ell+1}), \partial \Omega)| \leq (M(3M)^{\ell} \varepsilon + 3M \lambda_T) d(x(t_{\ell+1}), \partial \Omega).
\]

By taking \( \lambda \) small enough, and recalling (31), we see that equation (30) holds with \( \ell + 1 \) in place of \( \ell \).

**Case 3:** Suppose now that \( x(t_{\ell}) \in V \) and \( x(t_{\ell+1}) \notin V \). We then have \( x(t_{\ell}) \in B(\xi_j, 3\hat{r}(j)) \).

From Step 2, we see that \( x(t) \in B(\xi_j, 4\hat{r}(j)) \) for all \( t \in [t_{\ell}, t_{\ell+1}] \), hence also for all \( t \in [t_{\ell} - \delta, t_{\ell+1}] \) for \( \delta > 0 \) small enough (or for \( \delta = 0 \) if \( \ell = 0 \)). In addition, if \( \ell \geq 1 \), then by Assumption (13 b), we also have, possibly after making \( \delta > 0 \) smaller, \( x_k(t) \in U_\ell \) for \( t \in [t_{\ell} - \delta, t_{\ell}] \); if \( \ell = 0 \) then we have \( x_k(t_{\ell}) = x_k \in U_\ell \) by the setup in Step 2.

Define

\[
\tilde{x}(t) = \psi_{j_1}(x(t)) \quad \forall t \in [t_{\ell} - \delta, t_{\ell+1}], \quad \tilde{x}_k(t) = \psi_{j_1}(x_k(t)) \quad \forall t \in [t_{\ell} - \delta, t_{\ell}].
\]

As in the previous step, suppose \( 0 < \lambda < 1 \) and \( \lambda \) is a small but fixed number, and let \( t_{\ell,\lambda} = \frac{(\ell+1-\lambda)T}{\lambda T} = t_{\ell+1} - \lambda T \) where \( \lambda_T := \frac{\lambda T}{\ell} \). By taking \( \lambda \) small enough, we have that \( x(t_{\ell,\lambda}) \notin V \). For \( t \in [t_{\ell}, t_{\ell,\lambda}] \) define

\[
\tilde{x}_k^a(t) = \tilde{x}_k(t) - \tilde{x}_k(t_{\ell}) + \tilde{x}_k(t) \quad \forall s = 1, \ldots, d - 1, \quad \tilde{x}_k^a(t) = \begin{cases} \frac{x_k(t)}{\tilde{x}_k(t)} \tilde{x}_k^d(t), & \text{if } \tilde{x}_k^d(t) \neq 0, \\ \tilde{x}_k(t), & \text{if } \tilde{x}_k^d(t) = 0. \end{cases}
\]

Then for all \( t \in [t_{\ell}, t_{\ell,\lambda}] \), we have

\[
\tilde{\dot{x}}_k^a(t) = \dot{x}_k^a(t) \quad \forall s = 1, \ldots, d - 1, \quad \tilde{\dot{x}}_k^d(t) = \begin{cases} \frac{\dot{x}_k^a(t)}{2\tilde{x}_k(t)} \tilde{x}_k^d(t), & \text{if } \tilde{x}_k^d(t) \neq 0, \\ \tilde{x}_k^d(t), & \text{if } \tilde{x}_k^d(t) = 0. \end{cases}
\]

If \( \ell \geq 1 \) we must check that (41) holds even at \( t = t_{\ell} \); in particular we need to show the formula holds as \( t \) approaches \( t_{\ell} \) from the right or the left. Since Assumption (13 b) is satisfied, we see that

\[
\tilde{x}_k^a(t_{\ell}+) = \tilde{x}_k(t_{\ell}+) \quad \forall s = 1, \ldots, d - 1, \quad \tilde{x}_k^a(t_{\ell}+) = \begin{cases} \frac{x_k(t^+)}{\tilde{x}_k(t)} \tilde{x}_k^d(t), & \text{if } \tilde{x}_k^d(t) \neq 0, \\ 0, & \text{if } \tilde{x}_k^d(t) = 0, \end{cases}
\]

where \( f(t^+) := \lim_{\varepsilon \to 0^+} f(s) \). Equation (42) accords with (41) as \( t \to t_{\ell} \). Indeed, we can check that \( \tilde{x}_k^d(t_{\ell}) = 0 \) when \( \tilde{x}_k^d(t_{\ell}) = 0 \) by observing that \( \tilde{x}_k(t) \) is necessarily at a minimum whenever it is zero; the other cases are straightforward.

Now if \( \tilde{x}_k^d(t_{\ell}) > 0 \) we have (by algebraic rearrangement)

\[
\tilde{x}_k^a(t) - \tilde{x}_k^a(t_{\ell}) = \tilde{x}_k^a(t_{\ell}) - \tilde{x}_k^a(t) + \left( \tilde{x}_k^a(t_{\ell}) - \tilde{x}_k^a(t_{\ell}) \right) \frac{\tilde{x}_k^d(t) - \tilde{x}_k^d(t_{\ell})}{\tilde{x}_k^d(t_{\ell})}
\]

19
which, combined with \( \hat{x}_k(t) - \hat{x}(t) = \tilde{x}_k(t) - \tilde{x}(t) \) for all \( s = 1, \ldots, d - 1 \), implies
\[
|\tilde{x}_k(t) - \tilde{x}(t)| \leq |\tilde{x}_k(t) - \tilde{x}(t)| + |\hat{x}_k(t) - \hat{x}(t)| - |\hat{x}_d(t) - \hat{x}(t)|
\]
\[
\leq C|x_k(t) - x(t)| + \frac{|d(x_k(t), \partial \Omega) - d(x(t), \partial \Omega)|}{d(x(t), \partial \Omega)}
\]
\[
\leq (C + \tilde{r})(3M)^{\ell} \varepsilon \leq 2C(3M)^{\ell} \varepsilon \quad \forall t \in [t_{\ell}, t_{\ell+1}]
\]
using (39) from Assumption 4.3(c) and (27) from Step 2. If \( \hat{x}_d(t_{\ell}) = 0 \) we have
\[
|x_k(t) - x(t)| \leq |\tilde{x}_k(t) - \tilde{x}(t)| \leq C|x_k(t) - x(t)| \leq C(3M)^{\ell} \varepsilon \quad \forall t \in [t_{\ell}, t_{\ell+1}].
\]
In either case,
\[
|x_k(t) - x(t)| \leq 2C(3M)^{\ell} \varepsilon \leq (3M)^{\ell+1} \varepsilon < r \quad \forall t \in [t_{\ell}, t_{\ell+1}]
\]
using (20). It follows that \( \tilde{x}_k(t) \in B(0, 2r(j_{\ell})) \) for all \( t \in [t_{\ell}, t_{\ell+1}] \). Thus we can define
\[
x_k(t) = \psi_{j_{\ell}}^{-1}(\tilde{x}_k(t)) \quad \forall t \in [t_{\ell}, t_{\ell+1}].
\]
As for the derivative, we have \( \hat{x}_k(t) = \hat{x}(t) \) for all \( t \in [t_{\ell}, t_{\ell+1}] \) if \( \hat{x}_d(t_{\ell}) = 0 \); if \( \hat{x}_d(t_{\ell}) > 0 \), we have
\[
|\hat{x}_k(t) - \hat{x}(t)| = \frac{\hat{x}_d(t_{\ell}) - \hat{x}_d(t)}{\hat{x}_d(t_{\ell})} |\hat{x}_d(t)| = \frac{d(x_k(t), \partial \Omega) - d(x(t), \partial \Omega)}{d(x(t), \partial \Omega)} |\hat{x}_d(t)| \leq C K_1(3M)^{\ell} \varepsilon.
\]
We obtain the estimates
\[
|x_k(t) - x(t)| \leq C|\tilde{x}_k(t) - \tilde{x}(t)| \leq 2C^2(3M)^{\ell} \varepsilon \quad (43)
\]
and, using the formula (39) with \( j_{\ell} \) in place of \( j_{\ell+1} \),
\[
|\hat{x}_k(t) - \hat{x}(t)| \leq C|\tilde{x}_k(t) - \tilde{x}(t)| + CK_1|\tilde{x}_k(t) - \tilde{x}(t)| \leq 3C^2K_1(3M)^{\ell} \varepsilon \quad (44)
\]
for all \( t \in [t_{\ell}, t_{\ell+1}] \).

If \( \hat{x}_d(t_{\ell}) > 0 \) we also have
\[
|\hat{x}_d(t_{\ell,\lambda}) - \hat{x}_d(t_{\ell})| \leq \frac{|\hat{x}_d(t_{\ell}) - \hat{x}_d(t_{\ell})|}{\hat{x}_d(t_{\ell})} \hat{x}_d(t_{\ell})
\]
and, recalling that \( \hat{x}_d(t_{\ell}) \leq 2\tilde{r} \leq 1 \), this means
\[
|d(x_k(t_{\ell,\lambda}), \partial \Omega) - d(x(t_{\ell,\lambda}), \partial \Omega)| \leq \frac{|d(x_k(t_{\ell}), \partial \Omega) - d(x(t_{\ell}), \partial \Omega)|}{d(x(t_{\ell}), \partial \Omega)} \leq (3M)^{\ell} \varepsilon < \tilde{r}.
\]
If \( \hat{x}_d(t_{\ell}) = 0 \), then \( \hat{x}_d(t_{\ell,\lambda}) = \hat{x}_d(t_{\ell}) \), i.e.
\[
d(x_k(t_{\ell,\lambda}), \partial \Omega) = d(x(t_{\ell,\lambda}), \partial \Omega).
\]
Recalling that \( d(x(t_{\ell,\lambda}), \partial \Omega) \geq 2\tilde{r} \), it follows that \( d(x_k(t_{\ell,\lambda}), \partial \Omega) \geq \tilde{r} \).

Now, for \( t \in [t_{\ell,\lambda}, t_{\ell+1}] \) we let
\[
x_k(t) = x_k(t_{\ell,\lambda}) + (t - t_{\ell,\lambda}) \hat{x}_k(t_{\ell,\lambda}) + \frac{1}{2\lambda_{t_{\ell,\lambda}}}(t - t_{\ell,\lambda})^2(\hat{x}(t_{\ell+1}) - \hat{x}_k(t_{\ell,\lambda})).
\]
Notice that, after taking \( \lambda \) small enough,
\[
|x_k(t) - x_k(t_{\ell,\lambda})| \leq \frac{3}{2} |\hat{x}_k(t_{\ell,\lambda})| + \frac{1}{2} |\hat{x}(t_{\ell+1})| < \tilde{r} \quad \forall t \in [t_{\ell,\lambda}, t_{\ell+1}],
\]
so that \( x_k(t) \in \Omega \) for all \( t \in [t_{\ell,\lambda}, t_{\ell+1}] \). Since \( \dot{x}_k(t) \) is linear on \([t_{\ell,\lambda}, t_{\ell+1}]\), \( \dot{x}_k(t_{\ell+1}) = \dot{x}(t_{\ell+1}) \), and \( |\dot{x}(t)| \leq K_2 \), we have
\[
|\dot{x}_k(t) - \dot{x}(t)| = |\dot{x}_k(t_{\ell,\lambda}) - \dot{x}(t_{\ell,\lambda})| + K_2 \lambda T \leq 3C^2 K_1 (3M)^{\ell} \varepsilon + K_2 \lambda T \quad \forall t \in [t_{\ell,\lambda}, t_{\ell+1}],
\]
from which it follows that
\[
|x_k(t) - x(t)| \leq |x_k(t_{\ell,\lambda}) - x(t_{\ell,\lambda})| + (3C^2 K_1 (3M)^{\ell} \varepsilon + K_2 \lambda T) \lambda T \quad \forall t \in [t_{\ell,\lambda}, t_{\ell+1}].
\]
Putting together (43), (44), (46), and (45), then taking \( \lambda \) small enough, we deduce
\[
|x_k(t) - x(t)|, |\dot{x}_k(t) - \dot{x}(t)| \leq 4C^2 K_0 (3M)^{\ell} \varepsilon \leq (3M)^{\ell+1} \varepsilon \quad \forall t \in [t_{\ell}, t_{\ell+1}].
\]
It follows that
\[
|d(x_k(t_{\ell+1}), \partial \Omega) - d(x(t_{\ell+1}), \partial \Omega)| \leq \frac{4C^2 K_0}{2\ell} (3M)^{\ell} \varepsilon \quad \forall t \in [t_{\ell,\lambda}, t_{\ell+1}].
\]
Putting all these facts together, it follows that \( x_k(\cdot) \in C^{1,1}([0, t_{\ell+1}]; \Omega) \) and that Assumption 13 is satisfied with \( \ell \) replaced by \( \ell + 1 \).

**Case 4:** Finally, suppose \( x(t) \) and \( x(t_{\ell+1}) \) are both in \( V \). We have \( x(t_{\ell+1}) \in B(\xi_{j_{\ell+1}}, 3\hat{r}(j_{\ell+1})) \) and \( x(t_{\ell+1}) \in B(\xi_{j_{\ell+1}}, 3\hat{r}(j_{\ell+1})) \). Again let \( \lambda \in (0, 1) \) be small but fixed, define \( t_{\ell,\lambda} = \frac{(t_{\ell+1} - \lambda T)}{L} = t_{\ell+1} - \lambda T \) where \( \lambda T := \frac{\varepsilon T}{L} \), so that, in particular, \( x(t) \in B(\xi_{j_{\ell+1}}, 3\hat{r}(j_{\ell+1})) \) for \( t \in [t_{\ell,\lambda} - \delta, t_{\ell+1}] \) and for sufficiently small \( \delta > 0 \) for \( t \in [t_{\ell,\lambda}], \) we define \( \tilde{x}_k(t), \tilde{x}(t), \) and \( x_k(t) \) exactly as in the previous case, so that \( x_k(t) = \psi_{j_{\ell}}(\tilde{x_k}(t)) \) and \( x(t) = \psi_{j_{\ell}}(\tilde{x}(t)) \). In particular, equations (40), (41), (43), and (44) hold. Recalling (43), we see that \( x_k(t) \in B(\xi_{j_{\ell+1}}, 4\hat{r}(j_{\ell+1})) \) for \( t \in [t_{\ell,\lambda} - \delta, t_{\ell+1}] \) for \( \delta > 0 \) small enough, and on this interval we define \( \tilde{y}_k(t) = \psi_{j_{\ell+1}}(x_k(t)) \). On the other hand, we may define \( \tilde{y}(t) = \psi_{j_{\ell+1}}(x(t)) \) for all \( t \in [t_{\ell,\lambda} - \delta, t_{\ell+1}] \). Observe that
\[
\begin{align*}
\tilde{y}^d_k(t) &= \tilde{x}^d(t) = d(x(t), \partial \Omega) \quad \forall t \in [t_{\ell,\lambda} - \delta, t_{\ell+1}], \\
\tilde{y}^d_k(t) &= \tilde{x}^d_k(t) = d(x_k(t), \partial \Omega) \quad \forall t \in [t_{\ell,\lambda} - \delta, t_{\ell+1}].
\end{align*}
\]
Thus we may continue the pattern established by equation (40) to define \( \tilde{g}_k(t) \) by setting
\[
\begin{align*}
\tilde{g}_k(t) &= \begin{cases} 
\tilde{x}^d_k(t) & \text{if } \tilde{x}^d_k(t) \neq 0, \\
\tilde{x}^d(t) & \text{if } \tilde{x}^d(t) = 0
\end{cases} = \begin{cases} 
\tilde{x}^d_k(t) & \text{if } \tilde{x}^d_k(t) \neq 0, \\
\tilde{x}^d(t) & \text{if } \tilde{x}^d(t) = 0
\end{cases} \quad \forall t \in [t_{\ell,\lambda}, t_{\ell+1}].
\end{align*}
\]
For the remaining coordinates, we set
\[
\tilde{y}_k^s(t) = \tilde{y}_k^s(t_{\ell,\lambda}) + (t - t_{\ell,\lambda}) \tilde{y}_k^s(t_{\ell,\lambda}) + \frac{1}{2\lambda T}(t - t_{\ell,\lambda})^2 (\tilde{y}^s(t_{\ell+1}) - \tilde{y}_k^s(t_{\ell,\lambda})), \quad s = 1, \ldots, d - 1.
\]
We will need the following estimates. First,
\[
|\tilde{y}_k(t_{\ell,\lambda}) - \tilde{y}(t_{\ell,\lambda})| \leq C|x_k(t_{\ell,\lambda}) - x(t_{\ell,\lambda})| \leq 2C^3 (3M)^{\ell} \varepsilon
\]
by using (43). Then, arguing as in the previous cases and taking \( \lambda \) small enough, we get
\[
|\tilde{y}_k(t) - \tilde{y}(t)| \leq |\tilde{y}_k(t_{\ell,\lambda}) - \tilde{y}(t_{\ell,\lambda})| + 2C^2 K_0 \lambda T \leq (3M)^{\ell} \varepsilon + 2C^2 K_0 \lambda T \leq (3M)^{\ell} \varepsilon \quad \forall t \in [t_{\ell,\lambda}, t_{\ell+1}],
\]
which yields
\[
|\tilde{y}_k(t) - \tilde{y}(t)| \leq 2C^3 (3M)^{\ell} \varepsilon + 2(3M)^{\ell} \varepsilon \lambda T \leq 3C^3 (3M)^{\ell} \varepsilon \quad \forall t \in [t_{\ell,\lambda}, t_{\ell+1}],
\]
and from which we can derive:
\[
\begin{align*}
|x_k(t) - x(t)| &\leq 3C^4 (3M)^{\ell} \varepsilon, \\
|\dot{x}_k(t) - \dot{x}(t)| &\leq 2C(3M)^{\ell} \varepsilon + K_0 C^5 (3M)^{\ell} \varepsilon \leq 3K_0 C^5 (3M)^{\ell} \varepsilon \quad \forall t \in [t_{\ell,\lambda}, t_{\ell+1}].
\end{align*}
\]
Taking into account (45), and (44), we see that (47) holds for all \( t \in [t_\ell, t_{\ell+1}] \). Finally, note that
\[
\tilde{y}^d(t_{\ell+1}) - \tilde{y}^d(t_{\ell+1}) = \begin{cases} \frac{\dot{x}^d(t_{\ell}) - \dot{x}^d(t_\ell)}{\dot{x}^d(t_\ell)} \tilde{y}^d(t_{\ell+1}) & \text{if } \dot{x}^d(t_\ell) \neq 0, \\ 0 & \text{if } \dot{x}(t_\ell) = 0 \end{cases} \quad \forall t \in [t_\ell, t_{\ell+1}].
\]

From this formula and equation (30), it follows that
\[
|d(x_k(t_{\ell+1}), \partial\Omega) - d(x(t_{\ell+1}), \partial\Omega)| \leq (3M)^t \varepsilon d(x(t_{\ell+1}), \partial\Omega).
\]

From (17) and (30), we deduce (31) with \( t_{\ell+1} \) in place of \( t_\ell \).

**Step 4:** Applying Step 3 recursively for each \( \ell = 0, 1, \ldots, L - 1 \), we see that \( x_k(t) \) is defined for all \( t \in [0, T] \), that \( x_k(t) \in C^{1.1}([0, t_{\ell+1}]; \Omega) \), and that equation (31) holds. In particular, this means
\[
|\dot{x}_k(t) - \dot{x}(t)| \leq (3M)^t \varepsilon.
\]

Since \( \varepsilon > 0 \) is arbitrary, the proof is complete.

\[
\square
\]

## 5 Notion of mild solution and existence

Recall that \( m_0 \in \mathcal{P}(\Omega) \) is fixed. Suppose that \( \eta \) is a probability measure on \( A \Gamma \) —recall the definitions (1)— such that \( (\pi^1 \circ c_0)_\# \eta = m_0 \), or that \( \eta \in \mathcal{P}(\Gamma^K) \), where the latter is the subset of those \( \eta \in \mathcal{P}(\Gamma^K) \) such that \( (\pi^1 \circ c_0)_\# \eta = m_0 \). Let \( x_0 \in \bar{\Omega} \), and consider the functional
\[
I^{\alpha}[u(\cdot); \eta] := \int_0^T l(t, x^u_x(t), u(t), (e_t)_\# \eta) dt + l_T(x(T), (e_T)_\# \eta),
\]
defined for \( u(\cdot) \in A\Gamma^2_0 \), where
\[
A\Gamma^2_0 = \{ u(\cdot) \in A\Gamma_2 \mid x_0 + \int_0^t u(\tau) d\tau \in \bar{\Omega} \quad \forall t \in [0, T] \}
\]
and
\[
x^u_x(t) := x_0 + \int_0^t u(\tau) d\tau.
\]

However, if \( K \) is large enough (see Corollary 3.3 and its proof), we know that if \( \eta \in \mathcal{P}(\Gamma^K) \), any optimal control \( u(\cdot) \) for \( I^{\alpha}[\cdot \mid \eta] \) lies in \( \Gamma_2^{x, x_0} \). Inspired by [CC18], we make the following definition.

**Definition 5.1.** We say that \( \eta \in \mathcal{P}(\Gamma^K) \) is a cemfgcs (a «constrained equilibrium of mean field game of controls») if:

for \( \eta \)-a.e. \((x(\cdot), u(\cdot)) \in \Gamma^K : I^{x(\cdot)}[u(\cdot); \eta] \leq I^{x(\cdot)}[\tilde{u}(\cdot); \eta] \) for all \( \tilde{u}(\cdot) \in \Gamma_2^{x, x(\cdot)} \).

Similarly to [CC18], we set
\[
\Gamma^0[x_0] := \{ u(\cdot) \in \Gamma_2^{x, x_0} \mid I^{\alpha}[u(\cdot); \eta] \leq I^{x(\cdot)}[\tilde{u}(\cdot); \eta] \forall \tilde{u} \in \Gamma_2^{x, x_0} \}, \quad x_0 \in \bar{\Omega}, \quad \eta \in \mathcal{P}(\Gamma^K),
\]
i.e., \( \Gamma^0[x_0] \) is the set of optimal controls in \( \Gamma_2^{x, x_0} \) for the functional \( I^{\alpha}[\cdot; \eta] \).

Recall that \( c_1 : A \Gamma \rightarrow \bar{\Omega} \times \mathbb{R}^d \) denotes the evaluation mappings, i.e. \( c_1(\gamma(\cdot)) = c_1(x(\cdot), u(\cdot)) = (x(t), u(t)) \). In particular, \( \pi^1 \circ c_0 \) is the \( \bar{\Omega} \)-valued mapping that assigns to each pair \((x(\cdot), u(\cdot))\) in \( A \Gamma \) the initial point of the trajectory, \( x(0) \). Given an arbitrary measure \( \eta \in \mathcal{P}(\Gamma^K) \), the disintegration theorem (see, e.g., [Bog07]) applied to the map \( \pi^1 \circ c_0 : \Gamma^K \rightarrow \Omega \) yields an \( m_0 \)-unique Borel family of probability measures \( \{\eta_{x_0}\}_{x_0 \in \bar{\Omega}} \) on \( \Gamma^K \) such that

- for \( m_0 \)-a.e. \( x_0 \in \bar{\Omega} \), \( \eta_{x_0}(\Gamma^K \setminus (\pi^1 \circ c_0)^{-1}(x_0)) = 0 \),
- \( \int_{\Gamma^K} \phi(\gamma) d\gamma = \int_{\bar{\Omega}} \int_{\Gamma^K} \phi(\gamma) \eta_{x_0}(d\gamma)m_0(dx_0) \) for every Borel function \( \phi : \Gamma^K \rightarrow [0, \infty) \).
Given \( \eta \in \mathcal{P}_{m_0}(\Gamma^K) \), similarly to [CC18], we define

\[
E : \mathcal{P}_{m_0}(\Gamma^K) \to 2\mathcal{P}_{m_0}(\Gamma^K)
\]

\[
\eta \mapsto E(\eta) := \{ \hat{\eta} \in \mathcal{P}_{m_0}(\Gamma^K) \mid \text{for } m_0\text{-a.e. } x_0 \in \hat{\Omega} : \pi^2 \circ \text{spt}(\hat{\eta}_{x_0}) \subset \Gamma^0[x_0] \}
\]

(50)

where the measures \( \hat{\eta}_{x_0}, x_0 \in \hat{\Omega} \) are the ones into which \( \hat{\eta} \) disintegrates, furnished by the disintegration theorem, and by \( \pi^2 \circ \text{spt}(\hat{\eta}_{x_0}) \) we mean:

\[
\pi^2 \circ \text{spt}(\hat{\eta}_{x_0}) = \{ u(\cdot) \in \Gamma^2 \mid \exists x(\cdot) \in \Gamma^1 \text{ s.t. } (x(\cdot), u(\cdot)) \in \text{spt}(\hat{\eta}_{x_0}) \}.
\]

**Remark 5.2.** Given \( m_0 \in \mathcal{P}(\hat{\Omega}), \eta \in \mathcal{P}_{m_0}(\Gamma^K) \), if \( \eta \in E(\eta) \), then \( \eta \) is a cemfgcs.

**Proof.** Let \( m_0 \in \mathcal{P}(\hat{\Omega}), \eta \in \mathcal{P}_{m_0}(\Gamma^K) \) and suppose \( \eta \in E(\eta) \). Let \( \{ \eta_{x_0} \}_{x_0 \in \hat{\Omega}} \) be the disintegration of \( \eta \) with respect to \( \pi^1 \circ e_0 \). Let \( \Omega' \subset \hat{\Omega} \) be a subset such that \( m_0(\Omega' \setminus \hat{\Omega}) = 0 \) and for every \( x \in \Omega' \): \( \pi^2 \circ \text{spt}(\eta_x) \subset \Gamma^0[x] \). Let \( \Gamma' = \bigcup_{x \in \Omega'} \text{spt}(\eta_x) \). Then \( \eta(\Gamma') = 1 \), as can be verified. To show that \( \eta \) is a cemfgcs, pick any \( (x(\cdot), u(\cdot)) \in \Gamma' \). Then \( (x(\cdot), u(\cdot)) \in \text{spt}(\eta_{x_0}) \) for some \( x_0 \in \Omega' \) and since, as can be checked, \( \text{spt}(\eta_x) \subset (\pi^1 \circ e_0)^{-1}(x) \) for every \( x \in \Omega' \), we have \( x(0) = x_0 \). Moreover, since \( x_0 \in \Omega', (x(\cdot), u(\cdot)) \in \text{spt}(\eta_{x_0}) \) means \( u(\cdot) \in \Gamma^0[x_0] \), so \( \Gamma^0[x_0] \) is the set of full measure \( \Gamma' \subset \hat{\Omega} \).

Thus, the definition of cemfgcs is satisfied by \( \eta \) and its corresponding set of full measure \( \Gamma' \subset \hat{\Omega} \).

\[ \square \]

### 5.1 Nonemptiness of the fixed-point mapping

We must check that \( E(\eta) \neq \emptyset \) for each \( \eta \in \mathcal{P}_{m_0}(\Gamma^K) \). The sets \( \Gamma^0[x] \) aren’t empty but may have more than one element. Let \( x_0 \mapsto s(x_0) \in \Gamma^0[x_0] \) be, for the moment, an arbitrary selection of the set-value mapping \( \hat{\Omega} \ni x_0 \mapsto \Gamma^0[x_0] \subset \Gamma^2_+ \). Let \( \eta' \in \mathcal{P}_{m_0}(\Gamma^K) \) be the probability measure defined by:

\[
\int_{\Omega} \phi(\gamma) \eta'(d\gamma) = \int_{\Omega} \int_{\Gamma} \phi(\gamma) \delta_{(x(x_0), s(x_0))}(d\gamma)m_0(dx_0)
\]

for every Borel function \( \phi : \Gamma \to [0, \infty) \) (recall the notation \( x^u \) from [BB]). In order to apply the disintegration theorem to conclude that for \( m_0\text{-a.e. } x_0 \in \hat{\Omega}, \delta_{(x(x_0), s(x_0))} = \eta'_{x_0} \), one would like to know that the family \( \{ \delta_{(x(x_0), s(x_0))} \}_{x_0 \in \hat{\Omega}} \) is a Borel family, i.e., the mapping

\[
\hat{\Omega} \ni x_0 \mapsto \delta_{(x(x_0), s(x_0))}
\]

should be Borel, which means, by definition, that for every open subset \( W \subset \Gamma^K \), the function \( \hat{\Omega} \ni x_0 \mapsto \delta_{(x(x_0), s(x_0))}(W) \) should be Borel. This function takes on values in the set \( \{0, 1\} \), so it suffices to show that the preimage of 1,

\[
\{ x_0 \in \hat{\Omega} \mid \delta_{(x(x_0), s(x_0))}(W) = 1 \} = \{ x_0 \in \hat{\Omega} \mid (x^u(x_0), s(x_0)) \in W \},
\]

is a Borel set whenever \( W \subset \Gamma^K \) is a Borel set. Thus, the selection \( s \) cannot be arbitrary. We provide next the necessary background for the reader’s convenience. The classic reference is [AF90].

### Set-valued maps

**Definition.** A set-valued map \( F : X \rightharpoonup Y \) (i.e. \( F : X \to 2^Y \)), where \( X \) and \( Y \) are metric spaces, is called upper semicontinuous (u.s.c.) at \( x \in \text{Dom}(F) := \{ x \in X \mid F(x) \neq \emptyset \} \) if for every open neighbourhood \( O \) of \( F(x) \) there exists an open ball \( B \subset X \) containing \( x \) such that \( x' \in B \Rightarrow F(x') \subset O \).

It is easily checked that if \( \text{Dom}(F) \) is closed, then \( F : X \rightharpoonup Y \) is u.s.c. if and only if for every closed \( C \subset Y \), \( F^{-1}(C) \) is closed; this criterion will be used a few lines below.
It is also easy to verify the following:

**Fact.** If \( Y \) is compact, and the Graph(\( F := \{(x,y) \in X \times Y \mid y \in F(x)\} \) is closed, then \( F \) is an u.s.c. map with closed domain and closed images.

**Definition.** Let \((\Theta, \mathcal{A})\) be a measurable space, \( Y \) a metric space. A set-valued map \( F : \Theta \rightrightarrows Y \) is called measurable if for every open set \( O \subset Y, F^{-1}(O) \in \mathcal{A} \).

**Fact.** Let \( X \) be a metric space and \( \mathcal{A} \) a \( \sigma \)-algebra on \( X \) that includes all the open sets. Suppose that the set-valued map \( F : X \rightrightarrows Y \) is such that \( F^{-1}(C) \) is closed whenever \( C \subset Y \) is closed. Then \( F \) is measurable.

**Proof.** Fix an open set \( O \subset Y \). For every \( n \in \mathbb{N} \), let \( C_n \) be the closed set \( \{ y \in Y \mid d(y, Y \setminus O) \geq 1/n \} \). Then \( O = \bigcup_{n=1}^{\infty} C_n \) and \( F(x) \cap O \neq \emptyset \) if and only if \( x \in \bigcup_{n=1}^{\infty} F^{-1}(C_n) \), and the latter set belongs to \( \mathcal{A} \). \( \square \)

**Fact ([AF90, Thm 8.1.3])**. Let \( Y \) be a complete separable metric space, \((\Theta, \mathcal{A})\) a measurable space, \( F \) a measurable set-valued map from \( \Theta \) to closed nonempty subsets of \( Y \). Then there exists a measurable selection of \( F \).

By the foregoing, and since \( \Gamma_2^K \) is a compact set, a measurable selection \( x_0 \mapsto s(x_0) \in \Gamma_0^0[x] \subset \Gamma_2^K \) is possible if Graph(\( \Gamma_0^0 \)) is closed. We show a slightly more general statement, which now follows directly from Proposition 4.1.

**Lemma 5.3.** The set-valued mapping

\[
\Gamma^t(\cdot) : \mathcal{P}_{m_0}(\Gamma^K) \times \bar{\Omega} \rightrightarrows \Gamma^K_2 \\
(\eta, x_0) \mapsto \Gamma_0^0[x_0]
\]

is closed.

**Proof.** Let \( \{x_n\}_{n=1}^{\infty} \subset \bar{\Omega} \) be a sequence that converges to \( x_0 \in \bar{\Omega} \), \( \{\eta^n\}_{n=1}^{\infty} \subset \mathcal{P}_{m_0}(\Gamma^K) \) a sequence converging to \( \eta \in \mathcal{P}_{m_0}(\Gamma^K) \), and let \( \{u_n(\cdot)\}_{n=1}^{\infty} \subset \Gamma_2^K \) be a sequence with \( u_n(\cdot) \in \Gamma_0^0[x_n] \) for every \( n \), converging in \( d_2 \) (i.e. uniformly) to \( u(\cdot) \in \Gamma_2^K \). The sequence of paths \( \{x^{u_n(\cdot)}_{x_0(\cdot)}\}_{n=1}^{\infty} \) in the compact metric space \( \Gamma_1^K \) converges to the path \( x^{u(\cdot)}_{x_0(\cdot)} \), so \( u(\cdot) \in \Gamma_2^K(x_0) \). Let \( \tilde{u}(\cdot) \in \Gamma_2^{K,x_0} \) be arbitrary. By Proposition 4.1, we may let \( \{\tilde{u}^n(\cdot)\}_{n=1}^{\infty} \subset \Gamma_2^{K+1} \) be a sequence with \( \tilde{u}^n(\cdot) \in \Gamma_2^{K+1,x_n} \) for every \( n \), that converges to \( \tilde{u}(\cdot) \) uniformly in time. This uniform convergence and the continuity of \( l \) and \( l_T \), along with the fact that

\[
d_{\bar{\Omega} \times B_{K_1}(0)}((e_t) \neq \eta, (e_t) \neq \eta^n) \leq d_{\Gamma^K}(\eta, \eta^n)
\]

for all \( n \) and \( t \in [0, T] \), immediately give that both

\[
I^{x_0}[\tilde{u}; \eta] = \lim_{n \to \infty} I^{x_0}[\tilde{u}^n; \eta^n]. \tag{51}
\]

and

\[
I^{x_0}[u; \eta] = \lim_{n \to \infty} I^{x_0}[u_n; \eta^n]. \tag{52}
\]

Since \( \tilde{u}^n(\cdot) \in \Gamma_2^{x_n,K+1} \subset \mathcal{A}\Gamma_2^{x_n} \) and \( u_n(\cdot) \in \Gamma_0^0[x_n] \), we have

\[
I^{x_0}[u_n; \eta^n] \leq I^{x_n}[\tilde{u}^n; \eta^n], \quad n \in \mathbb{N}.
\]

Passing to the limit as \( n \to \infty \), by (51) and (52), we obtain

\[
I^{x_0}[u(\cdot); \eta] \leq I^{x_0}[\tilde{u}; \eta].
\]

Since \( \tilde{u}(\cdot) \) was an arbitrary element of \( \Gamma_2^{K,x_0} \), this has shown that \( u(\cdot) \in \Gamma_0^0[x_0] \), and, thus, that the mapping \( \Gamma^t(\cdot) \) is closed. \( \square \)
5.2 Existence of equilibrium

We cite:

Kakutani-Fan-Glicksberg theorem. [AB06, p. 17.55] Let $K$ be a nonempty compact convex subset of a locally convex Hausdorff space, and let the multi-valued mapping $\phi : K \to K$ have closed graph and nonempty convex values. Then the set of fixed points of $\phi$ is compact and nonempty.

As mentioned in the section, $\mathcal{P}(\Gamma^K)$ is a compact metric space with the 1-Wasserstein distance, since $\Gamma^K$ is compact (w.r.t. uniform convergence). Because $\Gamma^K$ is compact, it is known that narrow convergence of sequences in $\mathcal{P}(\Gamma^K)$ coincides with convergence in the Wasserstein distance [AGS08, Proposition 7.1.5]. Every element $\mu \in \mathcal{P}(\Gamma^K)$ can be identified with the linear functional $f \mapsto \int_{\Gamma^K} f(\gamma)\mu(\mathcal{D})$ on $C(\Gamma^K)$. Narrow convergence is induced by the weak-* topology of $C(\Gamma^K)'$ — the continuous dual of the space of continuous functions on $\Gamma^K$, and $C(\Gamma^K)'$, with its weak-* topology, is a locally convex, Hausdorff, topological vector space. Thus, $\mathcal{P}(\Gamma^K)$ is a compact, convex subset of $C(\Gamma^K)'$, and it can be easily shown that so is $\mathcal{P}(\tau_0(\Gamma^K))$. Having seen that the multivalued mapping $E : \mathcal{P}(\tau_0(\Gamma^K)) \to \mathcal{P}(\tau_0(\Gamma^K))$ takes on nonempty values, now we will check that they are convex and that the mapping has a closed graph.

**Lemma 5.4.** The set-valued mapping $E : \mathcal{P}(\tau_0(\Gamma^K)) \to \mathcal{P}(\tau_0(\Gamma^K))$, defined above in [50], takes on nonempty, convex values and has a closed graph.

**Proof.** The nonemptiness has been shown. Let $\eta \in \mathcal{P}(\tau_0(\Gamma^K))$, $0 < \lambda < 1$ and $\hat{\eta}^0$ and $\hat{\eta}^1$ be elements of $E(\eta) \subset \mathcal{P}(\tau_0(\Gamma^K))$. It is straightforward to verify that the disintegration of $\lambda \hat{\eta}^0 + (1 - \lambda)\hat{\eta}^1$ coincides $\eta_{0,\text{a.e.}}$ with $\{0\lambda \hat{\eta}^0 + (1 - \lambda)\hat{\eta}^1\}_{\eta_{0,\text{a.e.}}}$ and $\text{spt}(\lambda \hat{\eta}^0 + (1 - \lambda)\hat{\eta}^1) = \text{spt}(\hat{\eta}^0) \cup \text{spt}(\hat{\eta}^1)$ for every $x_0 \in \Omega$. Therefore, for $m_0$: $x_0 \in \Omega$, if $u(\cdot) \in \Gamma_2^0$ and $x(\cdot) \in \Gamma_1^0$ are such that $(\hat{x}(\cdot), u(\cdot)) \in \text{spt}(\lambda \hat{\eta}^0 + (1 - \lambda)\hat{\eta}^1)$, then $u(\cdot) \in \Gamma^0[x_0]$. This shows that $E(\eta)$ is convex.

Let $\{\eta^j\}_{j=1}^{\infty} \subset \mathcal{P}(\tau_0(\Gamma^K))$ be a sequence converging to $\eta \in \mathcal{P}(\tau_0(\Gamma^K))$, take a sequence $\{\hat{\eta}^j\}_{j=1}^{\infty} \subset \mathcal{P}(\tau_0(\Gamma^K))$ with $\hat{\eta}^j \in E(\eta^j)$, and suppose $\lim_{j \to \infty} d_{1,\kappa}^0(\hat{\eta}^j, \hat{\eta}) = 0$ for some $\hat{\eta}$. Let $\{\hat{\eta}_{x_0}\}_{x_0 \in \Omega}$ be the disintegration of $\lambda \hat{\eta}^0 + (1 - \lambda)\hat{\eta}^1$, $\lambda = 1, 2, \ldots$ and take $\Omega^0 \subset \Omega$ to be a subset such that $m_0(\Omega^0 \setminus \Omega^0) = 0$ and $\text{spt}(\hat{\eta}_{x_0}) \subset (\pi^1 \circ e_0)^{-1}(x)$ for all $x \in \Omega^0$. For $j = 1, 2, \ldots$ let $\Omega^j$ be a set such that $m_0(\Omega^j \setminus \Omega^j) = 0$ and for every $x \in \Omega^j$, both $\pi^2 \circ \text{spt}(\hat{\eta}_{x_0}) \subset \Gamma^0 \setminus (\pi^1 \circ e_0)^{-1}(x) = 0$. The subsets $\Omega^0, \Omega^1, \ldots$ have these properties exist by hypothesis. Let

$$\Omega^j := \bigcup_{j=0}^{\infty} \Omega^j, \quad \Omega^0 := (\pi^1 \circ e_0)^{-1}(\Omega^j) \subset \Gamma^K.$$

We have

$$m_0(\Omega^j) = 1, \quad \hat{\eta}(\Omega^j) = 1, \quad \hat{\eta}^j(\Omega^0) = 1, \quad j = 1, 2, \ldots$$

The set $\Gamma^0$ is a Borel subset of $\Gamma^K$, so we can consider all of $(\Gamma^0, \mathcal{B}(\Gamma^0), \hat{\eta}^j |_{\Gamma^0})$ ($j$ ranging from 1 to $\infty$) and $(\Gamma^0, \mathcal{B}(\Gamma^0), \hat{\eta}^j |_{\Omega^0})$ simultaneously as probability spaces on the same set $\Gamma^0$ with the Borel $\sigma$-algebra $\mathcal{B}(\Gamma^0)$. The set $\Gamma^0$ is a separable metric space in its own right (with the metric $d_{1,\kappa}^0 |_{\Omega^0}$). Furthermore, we have

$$\hat{\eta}^j |_{\Omega^0} \to \hat{\eta} |_{\Omega^0} \quad \text{narrowly}. \quad (53)$$

Fix an arbitrary point $x_0 \in \Omega^j$ and let $(\hat{x}, \hat{u}) \in \text{spt}(\hat{\eta}_{x_0})$. Note that $\hat{x}(0) = x_0$, and that $(\hat{x}(\cdot), \hat{u}(\cdot)) \in \text{spt}(\hat{\eta} |_{x_0})$, hence $(\hat{x}(\cdot), \hat{u}(\cdot)) \in \text{spt}(\hat{\eta} |_{x_0})$. The convergence (53) implies the convergence (AGS08, Prop. 5.18) that there exists a sequence $\{(\hat{x}^j, \hat{u}^j)\}_{j=1}^{\infty} \subset \Gamma^0$, with $(\hat{x}^j, \hat{u}^j) \in \text{spt}(\hat{\eta}^j |_{\Omega^0})$ for all $j$, such that $(\hat{x}^j, \hat{u}^j) \to (x, \hat{u})$. Let $x_j = \hat{x}^j(0), j = 1, 2, \ldots$. Clearly, $(\hat{x}^j, \hat{u}^j) \in \text{spt}(\hat{\eta}^j)$ for each $j$, so $(\hat{x}^j, \hat{u}^j) \in \text{spt}(\hat{\eta}_{x_j}^j)$, because, by $x_j$ being in $\Omega^j$, there holds that $\hat{\eta}_{x_j}^j(\Gamma^K \setminus (\pi^1 \circ e_0)^{-1}(x_j)) = 0$. Moreover, since all the $x_j \in \Omega^j$, we have $\pi^2 \circ \text{spt}(\hat{\eta}_{x_j}) \subset \Gamma^0 |_{\Omega^j |_{\Omega^j}}$. In summary, we have sequences $\eta^j \to \eta, x_j \to x_0$ and $\{\hat{u}^j\}_{j=1}^{\infty} \subset \Gamma_2^k$ such that $\hat{u} \in \Gamma^0 |_{\Omega^0 |_{\Omega^0}}$ for every $j$ and the controls $\hat{u}^j$ converge to $\hat{u}$ in $d_2$. We are exactly in the situation of Lemma 5.3, so we conclude that $\hat{u} \in \Gamma^0 |_{\Omega^0 |_{\Omega^0}}$. Thus, we have shown that for an arbitrary point $x_0 \in \Omega^j$,

$$\pi^2 \circ \text{spt}(\hat{\eta}_{x_0}) \subset \Gamma^0 |_{\Omega^0 |_{\Omega^0}}.$$
which means that $\hat{\eta} \in E(\eta)$, because $\Omega'$ has full measure. So we have proved that the set-valued mapping $E : \mathcal{P}_{m_0}(\Gamma^K) \rightsquigarrow \mathcal{P}_{m_0}(\Gamma^K)$ has a fixed point. By Remark 5.2 this means that there exists a constrained equilibrium of mean field game of controls, i.e. we have proved:

**Lemma 5.5.** With the notation of section 2.1 let the hypotheses $(L\Re)$ through $(L_{\omega\Re})$ and $(T\Re)$ hold, and suppose that the constant $c_1$ is sufficiently small and $K = (K_1, K_2)$ as in Corollary 3.3. Then there exists a cemfgcs (constrained equilibrium of mean field game of controls) (see Definition 5.1) $\eta$ in $\mathcal{P}_{m_0}(\Gamma^K)$.

In the same way as [CC18], we define the notion of mild solution for this problem. Recall that $X$ is an open set including $\bar{\Omega}$ and $M = \bar{\Omega} \times \mathbb{R}^d$.

**Definition 5.6.** Let $l : [0, T] \times X \times \mathbb{R}^d \times \mathcal{P}(M) \rightarrow \mathbb{R}$, $l_T : X \times \mathcal{P}(M) \rightarrow \mathbb{R}$ be continuous functions, $m_0 \in \mathcal{P}(\bar{\Omega})$. A pair $(V, \sigma) \in C([0, T] \times \bar{\Omega}) \times C([0, T]; \mathcal{P}(\bar{\Omega}))$ is called a mild solution of the constrained MFGCs problem with initial state distribution $m_0$ if there exists $K = (K_1, K_2)$ and a cemfgcs $\eta \in \mathcal{P}_{m_0}(\Gamma^K)$ such that

(i) $\sigma_0 = m_0$,

(ii) $\sigma_t = (\pi^1 \circ e_t)\# \eta$ for all $t \in [0, T]$,

(iii) $V(t, x) = \inf_{u(.) \in \mathcal{A}_2} \int_0^T l(s, x_{\pi}^u(s), u(s), (e_s)\# \eta)ds + l_T(x(T), (e_T)\# \eta)$ for every $x \in \bar{\Omega}$ and $t \in [0, T]$.

Thus, the main theorem follows:

**Theorem 5.7.** With the notation of section 2.1 let the hypotheses $(L\Re)$ through $(L_{\omega\Re})$ and $(T\Re)$ hold, and suppose that the constant $c_1$ is sufficiently small (see Corollary 3.3). Let $m_0 \in \mathcal{P}(\bar{\Omega})$. Then there exists a mild solution $(V, \sigma) \in C([0, T] \times \bar{\Omega}) \times C([0, T]; \mathcal{P}(\bar{\Omega}))$ of the constrained MFGCs problem with initial state distribution $m_0$.

**Proof.** We apply Lemma 5.5 to obtain $K = (K_1, K_2)$ for a sufficiently small constant $c_1$ and a cemfgcs $\eta \in \mathcal{P}_{m_0}(\Gamma^K)$. Let $\sigma_t = (\pi^1 \circ e_t)\# \eta$ for all $t \in [0, T]$. The continuity of $\sigma$ follows immediately. Then, define $V(t, x)$ as in (iii), $V(t, x) < +\infty$ for all $(t, x) \in [0, T] \times \bar{\Omega}$ because, with the hypotheses, minimizing controls always exist for the problem on any interval $[t, T]$ and initial state $x(t) = x$. The continuity of $V$ follows from [CCC18, Proposition 4.1], by taking into account Remark 2.3 with $v_s = (e_s)\# \eta$, $0 \leq s \leq T$.

**Appendix**

**Proof of Lemma 4.2.** Before proceeding with the proof, we remark that it is only the third condition that needs any proof. Indeed, let $\xi \in \partial \Omega$. Then by definition of a $C^3$ domain, there exist open sets $U \ni \xi, V \subset \mathbb{R}^d$ and a diffeomorphism $\phi : U \rightarrow V \subset \mathbb{R}^d$ such that $\phi$ and $\phi^{-1}$ are both of class $C^3$ and

$$\phi(U \cap \bar{\Omega}) = V \cap (\mathbb{R}^{d-1} \times [0, \infty)) \quad \text{and} \quad \phi(U \cap \partial \Omega) = V \cap (\mathbb{R}^{d-1} \times \{0\}).$$

We will modify $\phi$, $U$, and $V$ to obtain a diffeomorphism $\psi$ satisfying the conditions above, in particular the third “isometry” property. Without loss of generality, we suppose $\phi$ and $\phi^{-1}$ are $C^3$ on the closures $\bar{U}$ and $\bar{V}$, respectively, and that $U$ and $V$ are compact.

**Step 1:** For each $x \in U \cap \partial \Omega$, we can use $\phi$ to define an inward pointing unit normal vector. Indeed, let $\phi^1, \ldots, \phi^d$ denote the vector components of $\phi$ and notice that $\phi^d$ is a $C^3$ function with $\phi^d(U \cap \partial \Omega) = 0$ and $\phi^d(U \cap \Omega) \subset (0, \infty)$, hence $\nabla \phi^d(x)$ is a vector normal to $\partial \Omega$ pointing toward the interior of $\Omega$. We thus define

$$n(x) = \frac{\nabla \phi^d(x)}{|\nabla \phi^d(x)|}.$$
Because $\phi^d$ is $C^3$ on a compact set, it follows that $n$ is of class $C^2$.

**Step 2:** For each $\varepsilon > 0$, we define 

$$U_\varepsilon = \{ x \in U : d(x, \partial \Omega) < \varepsilon \}.$$

If $x \in U \cap \partial \Omega$, then $x + tn(x) \in U_\varepsilon$ for all $t \in (-\varepsilon, \varepsilon)$ and in that case $d(x + tn(x), \partial \Omega) \leq |t|$. In this step we prove that for $\varepsilon > 0$ small enough, we have that for all $x \in U \cap \partial \Omega$, $d(x + tn(x), \partial \Omega) = |t|$ and if $y \in U \cap \partial \Omega$, $y \neq x$, then $|x + tn(x) - y| > |t|$.

Fix $x \in U \cap \partial \Omega$ and let $\tilde{x} = \phi(x) \in \mathbb{R}^{d-1} \times \{ 0 \}$. It suffices to show that for some $\varepsilon > 0$, for all $t \in (-\varepsilon, \varepsilon)$ and all $\tilde{y} \in V \cap (\mathbb{R}^{d-1} \times \{ 0 \})$, $\tilde{y} \neq \tilde{x}$ we have

$$|\phi^{-1}(\tilde{x}) + tn(x) - \phi^{-1}(\tilde{y})| > |t|.$$

We use the fact that $D\phi^{-1}(\tilde{x})(\tilde{y} - \tilde{x})$ is tangent to $\partial \Omega$ at $x$ to derive

$$|\phi^{-1}(\tilde{x}) + tn(x) - \phi^{-1}(\tilde{y})|^2 = |\phi^{-1}(\tilde{x}) - \phi^{-1}(\tilde{y})|^2 + 2(\phi^{-1}(\tilde{x}) - \phi^{-1}(\tilde{y})) \cdot tn(x) + |tn(x)|^2$$

$$= |\phi^{-1}(\tilde{x}) - \phi^{-1}(\tilde{y})|^2$$

$$- 2(\phi^{-1}(\tilde{y}) - \phi^{-1}(\tilde{x}) - D\phi^{-1}(\tilde{x})(\tilde{y} - \tilde{x})) \cdot tn(x) + t^2.$$

We combine the estimates

$$|\tilde{x} - \tilde{y}| \leq \| D\phi \|_\infty |\phi^{-1}(\tilde{x}) - \phi^{-1}(\tilde{y})|$$

and

$$|\phi^{-1}(\tilde{y}) - \phi^{-1}(\tilde{x}) - D\phi^{-1}(\tilde{x})(\tilde{y} - \tilde{x})| \leq \| D^2\phi^{-1} \|_\infty |\tilde{x} - \tilde{y}|^2$$

to deduce

$$|\phi^{-1}(\tilde{x}) + tn(x) - \phi^{-1}(\tilde{y})|^2 \geq \| D\phi \|^{-2}_\infty - 2\| D^2\phi^{-1} \|_\infty |\tilde{x} - \tilde{y}|^2 + t^2.$$

Hence if we pick $\varepsilon > 0$ small enough so that $\| D\phi \|^{-2}_\infty - 2\| D^2\phi^{-1} \|_\infty > 0$, our claim follows.

**Step 3:** Let $\varepsilon > 0$ be as small as in the previous step. For $x \in U_\varepsilon$, define $d(x)$ to be the signed distance from $x$ to $\partial \Omega$, i.e. $d(x) = d(x, \partial \Omega)$ if $x \in \overline{\Omega}$ and $d(x) = -d(x, \partial \Omega)$ if $x \notin \overline{\Omega}$. We will define $P : U_\varepsilon \rightarrow \partial \Omega$ to be a (nonlinear) projection onto $\partial \Omega$ in the following way. For $x \in U_\varepsilon$, there exists $x_0 \in \partial \Omega$ such that $|x - x_0| = d(x, \partial \Omega)$. By standard geometric arguments, $x - x_0$ is normal to $\partial \Omega$ at $x_0$, hence $x - x_0 = d(x)n(x_0)$. By the previous step, it follows that $x_0$ is the unique element of $\partial \Omega$ satisfying $|x - x_0| = d(x, \partial \Omega)$; we define $P(x) = x_0$. Notice that for $x \in U_\varepsilon$, $x_0 \in \partial \Omega \cap U$, we have $x = x_0 + tn(x_0)$ if and only if $x_0 = P(x)$ and $t = d(x)$.

**Step 4:** For $\varepsilon > 0$ as small as in the previous steps, define $\psi : U_\varepsilon \rightarrow \mathbb{R}^d$ by

$$\psi(x) = (\phi(P(x)), d(x)).$$

Notice that $\psi$ has an inverse on its range, namely

$$\psi^{-1}(y) = \psi^{-1}(y^1, \ldots, y^d) = \phi^{-1}(y^1, \ldots, y^{d-1}, 0) + y^d n(\phi^{-1}(y^1, \ldots, y^{d-1}, 0)).$$

We identify $D\psi^{-1}(y)$ with its Jacobian matrix as follows. Let $A(y)$ be the $d \times d$ matrix whose first $d - 1$ rows are given by $\partial_i \phi(y^1, \ldots, y^{d-1}, 0)$ for $i = 1, \ldots, d - 1$ and whose $d$th row is given by $n(\phi^{-1}(y^1, \ldots, y^{d-1}, 0))$. Let $B(y)$ be the $d \times d$ matrix whose first $d - 1$ rows are given by $y^d \partial_n(\phi^{-1}(y^1, \ldots, y^{d-1}, 0))$ for $i = 1, \ldots, d - 1$ and whose $d$th row is zero. Then we have $D\psi^{-1}(y) = A(y) + B(y)$. Both $A(y)$ and $B(y)$ are continuous with respect to $y$. Notice that the first $d - 1$ rows of $A(y)$ form the tangent space to $\partial \Omega$ at $\psi^{-1}(\phi^{-1}(y^1, \ldots, y^{d-1}, 0))$, and thus the $d$th row, $n(\phi^{-1}(y^1, \ldots, y^{d-1}, 0))$, is orthogonal to them. In particular, $A(y)$ is invertible, and its inverse must be continuous with respect to $y$. On the other hand, for $y \in U_\varepsilon$ we see that the matrix norm of $B(y)$ is bounded by $\| Dn \|_\infty \| D\phi \|_\infty \epsilon$. By choosing $\varepsilon$ smaller if necessary, we deduce that $D\psi^{-1}(y)$ is invertible for all $y \in V$ such that $|y^d| < \epsilon$.

By the inverse function theorem, we deduce $\psi$ is also differentiable and $D\psi(x) = [D\psi^{-1}(\psi(x))]^{-1}$. Since $\psi^{-1}$ is in fact of class $C^2$ (because so are $\phi^{-1}$ and $n$), it follows that $\psi$ is as well. Finally, we see that condition 3 of the lemma is satisfied by construction of $\psi$. $\square$
References

[AB06] C. Aliprantis and K. Border. *Infinite Dimensional Analysis*. 3rd. Springer, 2006.

[Ach+14] Y. Achdou et al. “Partial differential equation models in macroeconomics”. In: *Phil. Trans. R. Soc. A* 372.2028 (2014). doi: 10.1098/rsta.2013.0397

[Ach+17] Y. Achdou et al. *Income and wealth distribution in macroeconomics: A continuous-time approach*. Tech. rep. National Bureau of Economic Research, 2017. doi:10.3386/w23732

[AF90] J-P. Aubin and H. Frankowska. *Set-valued analysis*. Birkhäuser, 1990.

[AGS08] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. 2nd. Birkhäuser, 2008.

[BCS17] J-D. Benamou, G. Carlier, and F. Santambrogio. “Variational Mean Field Games”. In: *Modeling and Simulation in Science, Engineering and Technology*. Vol. 1. Springer Basel, 2017, pp. 141–171.

[BGP21] J. F. Bonnans, J. Gianatti, and L. Pfeiffer. “A Lagrangian approach for aggregative mean field games of controls with mixed and final constraints”. arXiv:2103.10743v1. 2021.

[Bog07] V. Bogachev. *Measure Theory*. Vol. 2. Springer, 2007.

[Car12] P. Cardaliaguet. “Notes on Mean Field Games”: https://www.ceremade.dauphine.fr/~cardaliaguet/MFG20130420.pdf. 2012.

[Car15] Pierre Cardaliaguet. “Weak solutions for first order mean field games with local coupling”. In: *Analysis and Geometry in Control Theory and its Applications*. Vol. 11. Springer INdAM Series. Springer, 2015, pp. 111–158. URL: https://arxiv.org/pdf/1305.7015.pdf

[CC18] P. Cannarsa and R. P. Capuani. “Existence and uniqueness for mean field games with state constraints”. In: *PDE models for multi-agent phenomena*. Ed. by P. Cardaliaguet, A. Porretta, and F. Salvarini. Vol. 28. Springer, 2018, pp. 49–71. doi:10.1007/978-3-030-01947-1_3

[CCC18] Y. Cannarsa, R. Capuani, and P. Cardaliaguet. “C^{1,1}-smoothness of constrained solutions in the calculus of variations with applications to mean field games”. In: *Mathematics in Engineering* 1.1 (2018), pp. 174–203. doi:10.3934/Mine.2018.1.174

[CCC21] Piermarco Cannarsa, Rossana Capuani, and Pierre Cardaliaguet. “Mean Field Games with state constraints: from mild to pointwise solutions of the PDE system”. In: *Calc. Var.* 60.3 (2021), pp. 1–33. doi:10.1007/s00526-021-01936-4

[Ces83] L. Cesari. *Optimization — Theory and Applications: problems with ordinary differential equation*. Vol. 17. Springer, 1983.

[CG15] P. Cardaliaguet and P. J. Graber. “Mean field games systems of first order”. In: *ESAIM: COCV* 21.3 (2015), pp. 690–722. doi:10.1051/cocv/2014044

[CL18] P. Cardaliaguet and C. Lehalle. “Mean field game of controls and an application to trade crowding”. In: *Math Finan Econ* 12 (2018), pp. 335–363. doi:10.1007/s11579-017-0206-2

[CM06] P. Cardaliaguet and C. Marchi. “Regularity of the eikonal equation with Neumann boundary conditions in the plane: application to fronts with nonlocal terms”. In: *SIAM J. Control Optim.* 45.3 (2006), pp. 1017–1038. doi:10.1137/050635894

[CS04] P. Cannarsa and C. Sinestrari. *Semiconcave Functions, Hamilton-Jacobi Equations, and Optimal Control*. Birkhäuser, 2004.

[CS17] P. Chan and R. Sirca. “Fracking, Renewables, and Mean Field Games”. In: *SIAM Review* 59.3 (2017), pp. 588–615. doi:10.1137/15M1031424

[Dac08] B. Dacorogna. *Direct Methods in the Calculus of Variations*. 2nd ed. Springer, 2008.

[GM18] P. J. Graber and A. Mészáros. “Sobolev regularity for first order mean field games”. In: *Ann. I. H. Poincaré* 14.1 (2018), pp. 1557–1576. doi:10.1016/j.anihpc.2018.01.002
[GMP21] P. J. Graber, A. Mullenix, and L. Pfeiffer. “Weak solutions for potential mean field games of controls”. In: Nonlinear Differ. Equ. Appl. 28.50 (2021). doi:10.1007/s00030-021-00712-9

[GPV14] Diogo A Gomes, Stefania Patrizi, and Vardan Voskanyan. “On the existence of classical solutions for stationary extended mean field games”. In: Nonlinear Analysis: Theory, Methods & Applications 99 (2014), pp. 49–79. doi:10.1016/j.na.2013.12.016.

[GV16] D. Gomes and V. Voskanyan. “Extended deterministic mean-field games”. In: SIAM J. Control Optim. 54.2 (2016), pp. 1030–1055. doi:10.1137/130944503.

[HMC06] M. Huang, R. Malhamé, and P. Caines. “Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle”. In: Communications in Information & Systems 6.3 (2006), pp. 221–252.

[LL07] J-M. Lasry and P. L. Lions. “Mean field games”. In: Jpn. J. Math. 2.1 (2007), pp. 229–260. doi:10.1007/s11537-007-0657-8

[Vin00] RB. Vinter. Optimal control. Birhäuser, 2000.