Available online at www.sciencedirect.com
ScienceDirect
J. Differential Equations ••• (••••) •••–•••
Journal of Differential Equations
www.elsevier.com/locate/jde

Mayer control problem with probabilistic uncertainty on initial positions

Antonio Marigonda [a], Marc Quincampoix [b]

[a] *Department of Computer Science, University of Verona, Strada Le Grazie 15, I-37134 Verona, Italy*
[b] *Laboratoire de Mathématiques de Bretagne Atlantique, CNRS-UMR 6205, Université de Brest, 6, avenue Victor Le Gorgeu, CS 93837, 29238 Brest cedex 3, France*

Received 26 September 2017; revised 10 November 2017

Abstract

In this paper we introduce and study an optimal control problem in the Mayer's form in the space of probability measures on $\mathbb{R}^n$ endowed with the Wasserstein distance. Our aim is to study optimality conditions when the knowledge of the initial state and velocity is subject to some uncertainty, which are modeled by a probability measure on $\mathbb{R}^d$ and by a vector-valued measure on $\mathbb{R}^d$, respectively. We provide a characterization of the value function of such a problem as unique solution of an Hamilton–Jacobi–Bellman equation in the space of measures in a suitable viscosity sense. Some applications to a pursuit-evasion game with uncertainty in the state space is also discussed, proving the existence of a value for the game.

MSC: 49N70; 49L15

Keywords: Differential games; Optimal transport; Hamilton–Jacobi–Bellman equation

1. Introduction

We consider the following controlled differential equation

$$\dot{x}(t) = f(x(t), u(t)),\ u(t) \in U,\ t \in [0, T], \tag{1}$$

E-mail addresses: antonio.marigonda@univr.it (A. Marigonda), marc.quincampoix@univ-brest.fr (M. Quincampoix).

https://doi.org/10.1016/j.jde.2017.11.014
0022-0396/© 2017 Published by Elsevier Inc.

where $f : \mathbb{R}^d \times U \to \mathbb{R}^d$ is a Lipschitz continuous function, the control set U is a compact subset of some finite dimensional vector space, and the control function $u(\cdot)$ is a Borel measurable function $u : [0, T] \mapsto U$.

The main features of the optimal control system we will investigate in the paper are the following:

- The initial position x_0 is not exactly known by the controller, but only a probabilistic description is available. More precisely, the initial state is described by a measure μ_0 with the following property: given any Borel set $A \subseteq \mathbb{R}^d$, the quantity $\mu_0(A)$ gives the probability that the initial position lies in the set A.
- Because of the uncertain initial position, to every point of the support of μ_0 there may correspond a possibly different choice of the control – hence a different possible velocity. Moreover we allow the "division of mass", i.e., even if the initial condition x_0 is known (namely $\mu_0 = \delta_{x_0}$), it can be split into different trajectories by several possible velocities in $f(x_0, U)$ but of course the total weight of these trajectories must remain equal to one.

So the natural state space of our control problem is the space $\mathscr{P}(\mathbb{R}^d)$ of Borel probability measures on $\mathbb{R}^d$. The conservation of mass along the corresponding trajectory $\boldsymbol{\mu} = \{\mu_t\}_{t\in[0,T]}$ (seen as a time-dependent probability measure), and the controlled dynamics, can be summarized in the following dynamical system

$$\begin{cases} \partial_t \mu_t + \operatorname{div}(v_t \mu_t) = 0, \\ \mu|_{t=0} = \mu_0, \\ v_t(x) \in F(x) := f(x, U), \quad \text{for } \mu_t\text{-almost every } x \in \mathbb{R}^d, \text{ a.e. } t \in [0, T], \end{cases} \tag{2}$$

where the first equation of the above system should be understood in the sense of distributions in $[0, T] \times \mathbb{R}^d$.

Observe that when $v_t(\cdot)$ is sufficiently regular (i.e., Lipschitz continuous), then the unique solution μ_t of (2) is the pushforward of the measure μ_0 by the flow at time t of the differential equation $\dot{x}(t) = v_t(x(t))$. We also note that the trajectories depend only on F and not on the specific parametrization $F(x) := f(x, U)$ and, consequently, we will mainly consider the differential inclusion $\dot{x}(t) \in F(x(t))$ whose trajectories are the same as those of (1).

We stress the fact that the measures μ_T that can be reached at time T from an initial measure μ by mean of an admissible trajectory in the sense of (2) are *not* simply the ones which are pushforward of the initial measure μ_0 by any Borel selection ϕ of the reachable set for the finite-dimensional underlying system. An example of this situation is provided by Example 2.10.

The controller aims to minimize the cost function depending on the value of trajectory at the terminal time T

$$\mathscr{J}(\boldsymbol{\mu}) := \mathscr{G}(\mu_T) \tag{3}$$

where, $\mathscr{G} : \mathscr{P}(\mathbb{R}^d) \to \mathbb{R}$ is Lipschitz continuous with respect to the Wasserstein distance. A particular case of such a cost function is $\mathscr{G}_g(\mu) := \displaystyle\int_{\mathbb{R}^d} g(x)\, d\mu(x)$ where $g : \mathbb{R}^d \to \mathbb{R}$ is Lipschitz continuous and bounded. In this case, $\mathscr{G}_g$ turns out to be a Lipschitz map with respect to the Wasserstein distance on probability measures, and the final cost $\mathscr{G}_g(\mu_T)$ represents the expectation of the final cost g with respect to the probability measure μ_T. But such a cost is of moderate interest because the optimal value is nothing else that $\int_{\mathbb{R}^d} \boldsymbol{V}(0,x)\, d\mu_0(x)$ where $\boldsymbol{V}$ is the value function of the Mayer problem associated with (1) and the terminal cost $g(x(T))$.

In this paper we will consider a general Lipschitz function $\mathscr{G}$ from the set $\mathscr{P}(\mathbb{R}^d)$ of probability measures on $\mathbb{R}^d$ to $\mathbb{R}$, thus allowing terminal costs more general than those defined by the expectation of a function g. For instance, we can minimize the variance of the terminal probability distribution μ_T, or minimize the Wasserstein distance between μ_T and a given measure $\bar{\mu}$. Thinking of μ_T is a distribution of individuals, this means that we want to arrange these individuals into a preset formation.

The conservation law (2) has been extensively studied in the literature, we refer to [2] for a general overview, and [15,16,25] for the controlled case. The described framework has been studied in many papers (see [14], [16], [17]), mainly concerning time-optimal control problems, where, for instance, a large population of agents is macroscopically described by an equation such as (2), and the aim is to steer them to a sort of *safe* target region in the smallest possible amount of time, under different assumptions on the target region, on the way to compute the running cost, and on the possibility or not to remove the agents from the system as soon as they arrive. All these papers provide a Hamilton–Jacobi–Bellman equation which is solved by the value function. Nevertheless, the lack of a regularity theory for it (this aspect is partially addressed in [14]), does not allow one to prove a suitable comparison principle for the equation, thus preventing a full characterization of the value function. We refer the reader to [12] for a notion of viscosity solution on the Wasserstein space with a comparison principle, but in [12], the dynamics is much less general than 2 (cf. also [1,24]).

Strictly related to this class of problem, there is the so-called *confinement problem*, where it is studied the evolution of a time-depending *set* whose points follows the trajectory of a suitable differential inclusion in order to minimize a certain cost functional. For the applications, the initial set may be a flock of animals, crowd of pedestrians, or the frontline of a fire. We refer the reader to [8–10,19–21] to have a survey on the most recent results about this widely studied problem. We can link the set-dependent point of view with our approach by thinking to the evolving set to be the *support* of a measure which describes the initial state. Of course, a measure-evolving approach offers much more information on the initial state (since in general we are allowed to take measure that are not uniformly distributed), and this could provide on a more accurate description of the evolution in the model cases (see, e.g., [11]).

The first main goal of the present paper is to study the regularity of the value function associated with the dynamics (2) and the cost (3) and to provide a characterization of this value in terms of the unique solution of a suitable Hamilton–Jacobi–Bellman equation in the space of probability measures. This will require the introduction and investigation of a suitable notion of solution for such kinds of equations.

The second main goal consists in investigating the following game theory problem, where the first player acts on the system

$$\partial_t \mu_t + \operatorname{div}(v_t \mu_t) = 0,\ v_t(x) \in F(x),\ \text{for } \mu_t \text{ a.e. } x \in \mathbb{R}^d\ t \in [0,T],$$

while the second player controls the system

$$\partial_t \nu_t + \operatorname{div}(\theta_t \nu_t) = 0,\ \theta_t(x) \in G(x),\ \text{for } \nu_t \text{ a.e. } x \in \mathbb{R}^d\ t \in [0,T].$$

Associated to both the above dynamics, the following cost function is defined

$$\mathscr{J} := \mathscr{G}(\mu_T, \nu_T),$$

that the first and the second player wish to minimize and maximize, respectively.

This problem could be viewed as a zero-sum differential game in the space of probability measures. Our aim is to obtain the existence of a value for this game: namely to show that the upper value and the lower value coincide. We accomplish this task by proving that the lower value and the upper value are both solutions of a Hamilton–Jacobi–Bellman equation and by showing that the Hamilton–Jacobi–Bellman equation has a unique solution. We refer the reader to [12,13,24] for other differentials games problems in the space of measures.

The paper is structured as follows: in Section 2 we formulate the Mayer problem in the space of probability measures, proving also some relevant properties of the set of admissible trajectories (see Subsection 2.1) and of the value function (see Subsection 2.2). In Section 3 we introduce a notion of viscosity solution for general first-order Hamilton–Jacobi–Bellman equations in Wasserstein space, proving a uniqueness result in this setting. In Section 4 we use the results of the previous sections to characterize the value function of the Mayer problem. Section 5 introduces and discusses a pursuit-evasion game in the Wasserstein space, proving the existence of a value for the game. In Appendix A, we provide the basic definitions and notations used in the paper, while Appendix B is devoted to the comparison between the notion of generalized gradients that we used and other notions available in the literature.

We will use the following notation, referring to Appendix A for the definitions.

$C^0_b(X;Y)$	the set of continuous bounded function from a Banach space X to Y, endowed with $\|f\|_\infty = \sup_{x\in X} |f(x)|$ (if $Y=\mathbb{R}$, Y will be omitted);
$C^0_c(X;Y)$	the set of compactly supported functions of $C^0_b(X;Y)$, with the topology induced by $C^0_b(X;Y)$;
$\mathscr{P}(X)$	the set of Borel probability measures on a Banach space X, endowed with the weak* topology induced by $C^0_b(X)$;
$\mathscr{M}(\mathbb{R}^d;\mathbb{R}^d)$	the set of vector-valued Borel measures on $\mathbb{R}^d$ with values in $\mathbb{R}^d$, endowed with the weak* topology induced by $C^0_c(\mathbb{R}^d;\mathbb{R}^d)$;
$\mathrm{m}_2(\mu)$	the second moment of a probability measure $\mu\in\mathscr{P}(x)$;
$r\sharp\mu$	the push-forward of the measure μ by the Borel map r;
$\mu\otimes\eta_x$	the product measure of $\mu\in\mathscr{P}(X)$ with the Borel family of measures $\{\eta_x\}_{x\in X}$;
π_i	the i-th projection map from $\mathbb{R}^d$ to $\mathbb{R}$ yielding the i-th component;
π_{ij}	the projection map from $\mathbb{R}^d$ to $\mathbb{R}^2$ yielding the i-th and j-th components;
$\Pi(\mu,\nu)$	the set of admissible transport plans from μ to ν;
$\Pi_o(\mu,\nu)$	the set of optimal transport plans from μ to ν;
$W_2(\mu,\nu)$	the 2-Wasserstein distance between μ and ν;
$\mathscr{P}_2(X)$	the subset of the elements $\mathscr{P}(X)$ with finite p-moment, endowed with the 2-Wasserstein distance;
$\mathrm{Bar}_i(\gamma)$	the i-th barycentric projection of γ;
$\dfrac{\nu}{\mu}$	the Radon–Nikodym derivative of the measure ν w.r.t. the measure μ;
$\mathrm{Lip}(f)$	the Lipschitz constant of a function f.

2. A Mayer problem in the Wasserstein space

To maintain the flow of the paper we postpone to Appendix A several known results and definitions.

In this section we will introduce the Mayer problem in finite horizon. Given a cost function $\mathscr{G} : \mathscr{P}(\mathbb{R}^d) \to \mathbb{R}$ and a time horizon $T > 0$, we will consider the problem of minimizing the cost over all the endpoints of the trajectories in the space of measures that can be represented as a superposition of trajectories defined in $[0, T]$ of a given differential inclusions $\dot{x}(t) \in F(x(t))$, weighted by a probability measure μ on the initial state.

Throughout this section, we will made the following standing assumptions, referring the reader to [3] for an introduction to differential inclusions in finite-dimensional spaces:

(F) $F : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ is a Lipschitz continuous set-valued map with nonempty compact convex values;

($\mathscr{G}$) $\mathscr{G} : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ is bounded and Lipschitz continuous w.r.t. W_2 metric.

Given $a, b \in \mathbb{R}$, $a < b$, we will denote by $\Gamma_{[a,b]} = C^0([a, b]; \mathbb{R}^d)$ the space of continuous curves from $[a, b]$ to $\mathbb{R}^d$ endowed with the uniform convergence norm, and for all $t \in [a, b]$ we define the *evaluation operator* $e_t : \mathbb{R}^d \times \Gamma_{[a,b]} \to \mathbb{R}^d$ by setting $e_t(x, \gamma) = \gamma(t)$. When $[a, b] = [0, T]$ we will write Γ_T in place of $\Gamma_{[0,T]}$. We have that e_t is continuous.

2.1. Admissible trajectories and their properties

Given $N \in \mathbb{N}$, consider a smooth function φ, and n_N agents initially at the points $x_i^N \in \mathbb{R}^d$ and moving along the corresponding trajectory of the control system $\dot{\gamma}_i^N(t) = f(\gamma_i^N(t), u_i^N(t))$ satisfying $\gamma_i^N(0) = x_i^N$. If the i-th agent has weight $\lambda_i^N \in [0, 1]$ with $\sum \lambda_i^N = 1$, we have

$$\frac{d}{dt} \sum_{i=1}^{n_N} \lambda_i^N \varphi(\gamma_i^N(t)) = \sum_{i=1}^{n_N} \lambda_i^N \nabla\varphi(\gamma_i^N(t)) f(\gamma_i^N(t), u_i^N(t)).$$

By defining $\mu_t^N = \displaystyle\sum_{i=1}^{n_N} \lambda_i^N \delta_{\gamma_i^N(t)}$, $\vec{\nu}_t^N = \displaystyle\sum_{i=1}^{n_N} \lambda_i^N f(\cdot, u_i^N(t))\delta_{\gamma_i^N(t)}$, we have

$$\frac{d}{dt} \int_{\mathbb{R}^d} \varphi(x)\, d\mu_t^N(x) = \int_{\mathbb{R}^d} \nabla\varphi(x)\, d\vec{\nu}_t^N(x).$$

Clearly we have that $\vec{\nu}_t^N$ is absolutely continuous w.r.t. μ_t^N, and so we can write $\vec{\nu}_t^N = v_t^N \mu_t^N$ for a map $v_t^N \in L^1_{\mu_t^N}(\mathbb{R}^d; \mathbb{R}^d)$. Recalling the convexity of $F(x)$ we have also $v_t^N(x) \in F(x)$ for μ_t^N-a.e. $x \in \mathbb{R}^d$. Nevertheless, we must be careful in considering a notion of convergence for the family $\{v_t^N\}_{N\in\mathbb{N}}$, since even for fixed t they belong to L^p spaces of different measures. One of the natural solution to this problem, see e.g. the discussion on [2] before Definition 5.4.3 p. 127, is to consider directly the (time-depending) sequence of vector-valued measures $\{\nu_t^N\}_{N\in\mathbb{N}}$ and study its weak* limit points. With this notion of convergence, by taking a limit as $N \to +\infty$ of the above equation, we are naturally led to the following definition.

Please cite this article in press as: A. Marigonda, M. Quincampoix, Mayer control problem with probabilistic uncertainty on initial positions, J. Differential Equations (2017), https://doi.org/10.1016/j.jde.2017.11.014

Definition 2.1 *(Admissible trajectories).* Let $a, b \in \mathbb{R}$, $a < b$, $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, $F : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ be satisfying $(\boldsymbol{F})$. We say that $\boldsymbol{\mu} = \{\mu_t\}_{t\in[a,b]} \subseteq \mathscr{P}_2(\mathbb{R}^d)$ is an *admissible trajectory* starting from μ defined on $[a, b]$ if there exists a family of time-dependent Borel vector-valued measures $\vec{\boldsymbol{\nu}} = \{\vec{\nu}_t\}_{t\in[a,b]} \subseteq \mathscr{M}(\mathbb{R}^d; \mathbb{R}^d)$ such that

- $\partial_t \mu_t + \operatorname{div} \vec{\nu}_t = 0$ in the sense of distributions, and $\mu_a = \mu$,
- $|\vec{\nu}_t| \ll \mu_t$ for a.e. $t \in [a, b]$, i.e., the total variation $|\vec{\nu}_t|$ of the vector-valued measure $\vec{\nu}_t$ is absolutely continuous w.r.t. μ_t for a.e. $t \in [a, b]$;
- $\dfrac{\vec{\nu}_t}{\mu_t}(x) \in F(x)$ for a.e. $t \in [a, b]$ and μ_t-a.e. $x \in \mathbb{R}^d$.

In this case, we will say that $\boldsymbol{\mu}$ is *driven* by $\vec{\boldsymbol{\nu}}$. We will denote by $\mathscr{A}^F_{[a,b]}(\mu)$ the set of all admissible trajectories starting from μ and defined on $[a, b]$, and we set $\mathscr{A}^F_{[a,b]} = \displaystyle\bigcup_{\mu\in\mathscr{P}_2(\mathbb{R}^d)} \mathscr{A}^F_{[a,b]}(\mu)$. When $a = 0$, we will denote $\mathscr{A}^F_{[0,b]}(\mu)$ by $\mathscr{A}^F_b(\mu)$ and $\mathscr{A}^F_{[a,b]}$ by $\mathscr{A}^F_b$.

Remark 2.2. If we take $F(x) = -\partial Z(x)$, where $Z : \mathbb{R}^d \to]-\infty, +\infty]$ is a λ-convex l.s.c. (in the sense of (10.4.10) of [2]) function whose domain has nonempty interior, the admissible curves according to the previous definition reduces to the *gradient flow* of the potential energy functional $\mathscr{F}(\mu) = \displaystyle\int_{\mathbb{R}^d} Z(x)\, d\mu(x)$. We refer the reader to Chapter 10 and 11 in [2] for a complete treatment of gradient flow equations in Wasserstein spaces.

The following result provides some basic properties of the admissible trajectories.

Proposition 2.3 *(Properties of the admissible trajectories). Let $a, b, c \in \mathbb{R}$, $a < b < c$, $F : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ be satisfying $(\boldsymbol{F})$. Recalling that the space $X := C^0([a, b]; \mathscr{P}_2(\mathbb{R}^d))$ equipped with the metric*

$$d_X(\boldsymbol{\mu}, \boldsymbol{\nu}) = \sup_{t\in[a,b]} W_2(\mu_t, \nu_t), \textit{ for all } \boldsymbol{\mu} = \{\mu_t\}_{t\in[a,b]},\ \boldsymbol{\nu} = \{\nu_t\}_{t\in[a,b]},$$

is a complete metric space, we have that

(1) ***(closedness)*** *the set of admissible trajectories is closed in (X, d_X);*
(2) ***(compactness)*** *if $\{\boldsymbol{\mu}^N\}_{N\in\mathbb{N}}$ is a sequence of admissible trajectories satisfying $\sup_{N\in\mathbb{N}} \{\mathrm{m}_2(\mu^N_0)\} < \infty$, then it admits a convergent subsequence in (X, d_X).*
(3) ***(concatenation)*** *given $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, $\boldsymbol{\mu} = \{\mu_t\}_{t\in[a,b]} \in \mathscr{A}^F_{[a,b]}(\mu)$, $\boldsymbol{\nu} = \{\nu_t\}_{t\in[b,c]} \in \mathscr{A}^F_{[b,c]}(\mu_b)$ then, set*

$$\boldsymbol{\mu} \odot \boldsymbol{\nu} := \{\zeta_t\}_{t\in[a,c]}, \textit{ with } \zeta_t = \begin{cases} \mu_t, & \textit{if } a \le t \le b, \\ \nu_t, & \textit{if } b < t \le c, \end{cases}$$

we have $\boldsymbol{\mu} \odot \boldsymbol{\nu} \in \mathscr{A}^F_{[a,c]}(\mu)$.

Please cite this article in press as: A. Marigonda, M. Quincampoix, Mayer control problem with probabilistic uncertainty on initial positions, J. Differential Equations (2017), https://doi.org/10.1016/j.jde.2017.11.014

(4) ***(estimate)*** *if* $\boldsymbol{\mu} = \{\mu_t\}_{t\in[a,b]}$ *is an admissible trajectory, and* $\boldsymbol{\eta} \in \mathscr{P}(\mathbb{R}^d \times \Gamma_{[a,b]})$ *satisfies* $\mu_t = e_t \sharp \boldsymbol{\eta}$ *for all* $t \in [a,b]$*, then for* $s_1, s_2 \in [a,b]$ *we have*

$$\|e_{s_1} - e_{s_2}\|_{L^2_{\boldsymbol{\eta}}} \le Ce^{2(b-a)C}\left(1 + \min_{i=1,2} \mathrm{m}_2^{1/2}(\mu_{s_i})\right)|s_1 - s_2|,$$

where $C = \max_{y\in F(0)} |y| + \mathrm{Lip}(F)$.

(5) ***(convergence)*** *if* $\boldsymbol{\mu} = \{\mu_t\}_{t\in[a,b]}$ *is an admissible trajectory, and* $\boldsymbol{\eta} \in \mathscr{P}(\mathbb{R}^d \times \Gamma_{[a,b]})$ *satisfies* $\mu_t = e_t \sharp \boldsymbol{\eta}$ *for all* $t \in [a,b]$*, given* $\bar{t} \in [a,b]$ *and a sequence* $\{t_i\}_{i\in\mathbb{N}} \subseteq [a,b]$ *with* $t_i \to \bar{t}$*, every limit for* $i \to +\infty$ *of a* $L^2_{\boldsymbol{\eta}}$*-weak converging sequence* $\dfrac{e_{t_i} - e_{\bar{t}}}{t_i - \bar{t}}$ *belongs to the set*

$$\{v \circ e_{\bar{t}} : v \in L^2_{\mu_{\bar{t}}},\ v(x) \in F(x) \text{ for } \mu_{\bar{t}}\text{-a.e. } x \in \mathbb{R}^d\}.$$

Proof. Items (1) and (2) were proved in Proposition 3 and Theorem 3 of [16], respectively. Item (3) follows from Lemma 4.4 in [23] and from the definition of admissible trajectory. We prove now (4). By the Lipschitz continuity of F, for any $x \in \mathbb{R}^d$ and $v \in F(x)$, we have $|v| \le C(|x| + 1)$. If γ is a trajectory of the differential inclusion $\dot{x}(t) \in F(x(t))$ defined on $[a,b]$, for all $s_1, s_2 \in [a,b]$ we have

$$\begin{aligned}|\gamma(s_1) - \gamma(s_2)| &\le \left|\int_{s_1}^{s_2} |\dot{\gamma}(w)|\,dw\right| \le C|s_1 - s_2| + C\left|\int_{s_1}^{s_2} |\gamma(w)|\,dw\right| \\ &\le C(1 + |\gamma(s_i)|)|s_1 - s_2| + C\left|\int_{s_1}^{s_2} |\gamma(w) - \gamma(s_i)|\,dw\right|\end{aligned}$$

Grönwall inequality yields

$$|\gamma(s_1) - \gamma(s_2)| \le C(1 + |\gamma(s_i)|)|s_1 - s_2|e^{C|s_1 - s_2|} \le Ce^{2(b-a)C}(1 + |\gamma(s_i)|)|s_1 - s_2|$$

Integrating this relation w.r.t. $\boldsymbol{\eta}$, yields for $i = 1, 2$,

$$\begin{aligned}\|e_{s_1} - e_{s_2}\|_{L^2_{\boldsymbol{\eta}}} &\le Ce^{2(b-a)C}|s_1 - s_2|\left(\int_{\mathbb{R}^d \times \Gamma_{[a,b]}} (1 + |\gamma(s_i)|)^2\,d\boldsymbol{\eta}\right)^{1/2} \\ &= Ce^{2(b-a)C}|s_1 - s_2|\left(\int_{\mathbb{R}^d} (1 + |x|)^2\,d\mu_{s_i}\right)^{1/2} = Ce^{2(b-a)C}(1 + \mathrm{m}_2^{1/2}(\mu_{s_i}))|s_1 - s_2|,\end{aligned}$$

and we conclude by taking the minimum on $i = 1, 2$.

We prove (5). Given $x \in \mathbb{R}^d$, let $\gamma_x(\cdot)$ be any trajectory of the differential inclusion satisfying $\gamma_x(\bar{t}) = x$. Fix now $\varepsilon > 0$, $\bar{x} \in \mathbb{R}^d$. Then there exists $\delta_{\bar{x}} > 0$ such that $F(y) \subseteq F_\varepsilon(\bar{x}) := F(\bar{x}) + \varepsilon\overline{B(0,1)}$ for all $y \in B(\bar{x}, \delta_{\bar{x}})$. As in the proof of (4), we have

$$|\gamma_x(s) - x| \le Ce^{2(b-a)C}(1+|x|)|s-\bar{t}|.$$

In particular, there exists $\tau_{\bar{x}} > 0$ such that if $|s-\bar{t}| \le \tau_{\bar{x}}$, then $|\gamma_x(s) - x| \le \delta_{\bar{x}}$ for all $x \in B(\bar{x}, \delta_{\bar{x}}/2)$, and so $\dot{\gamma}_x(s) \in F(\bar{x}) + \varepsilon\overline{B(0,1)}$ for a.e. $s \in [a,b]$ with $|s-\bar{t}| \le \tau_{\bar{x}}$ and for all $x \in B(\bar{x}, \delta_{\bar{x}}/2)$. Consider now a sequence $\{t_i\}_{i\in\mathbb{N}}$ such that $\dfrac{e_{t_i}(x,\gamma) - e_{\bar{t}}(x,\gamma)}{t_i - \bar{t}} \rightharpoonup w(x,\gamma)$ in $L^2_{\boldsymbol{\eta}}$. The vector-valued measure $w\boldsymbol{\eta}$ is absolutely continuous w.r.t. $\mu_{\bar{t}}$ by the disintegration theorem (w.r.t. $e_{\bar{t}}$), and so we have $w\boldsymbol{\eta} = v\mu_{\bar{t}}$ for a certain $v \in L^1_{\mu_{\bar{t}}}(\mathbb{R}^d)$. For all $\xi \in \mathbb{R}^d$, $\bar{x} \in \mathbb{R}^d$ density point of $\mu_{\bar{t}}$, and $0 < \delta < \delta_{\bar{x}}/2$ we have

$$\begin{aligned}\int_{B(\bar{x},\delta)} \langle \xi, v(x)\rangle \, d\mu_{\bar{t}} &= \lim_{i\to+\infty} \int_{B(\bar{x},\delta)\times\Gamma_{[a,b]}} \langle \xi, \frac{e_{t_i}(x,\gamma) - e_{\bar{t}}(x,\gamma)}{t_i - \bar{t}}\rangle \, d\boldsymbol{\eta}(x,\gamma)\\ &\le \sup_{y\in F_\varepsilon(\bar{x})} \langle \xi, y\rangle \, \mu_{\bar{t}}(B(\bar{x},\delta)).\end{aligned}$$

Dividing by $\mu_{\bar{t}}(B(\bar{x},\delta))$ and letting $\delta \to 0^+$, this implies that $\langle \xi, v(\bar{x})\rangle \le \sup_{y\in F_\varepsilon(\bar{x})} \langle \xi, y\rangle$ for all $\xi \in \mathbb{R}^d$, and density point $\bar{x} \in \mathbb{R}^d$ of $\mu_{\bar{t}}$. By convexity of $F_\varepsilon(\bar{x})$, we have $v(\bar{x}) \in F_\varepsilon(\bar{x})$ for $\mu_{\bar{t}}$-a.e. $\bar{x} \in \mathbb{R}^d$. We conclude by letting $\varepsilon \to 0^+$ and noticing that, since $v(x) \in F(x)$ for $\mu_{\bar{t}}$-a.e. $x \in \mathbb{R}^d$, we have $v \in L^2_{\mu_{\bar{t}}}$ since F has linear growth. □

We will prove now a result allowing to use some Gronwall-like estimates on the admissible trajectories.

Proposition 2.4 *(Gronwall-like estimate in W_2). Assume that F satisfies $(\boldsymbol{F})$. Let $a, b \in \mathbb{R}$ with $a < b$. Then there exists $K > 0$ such that given $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$, $\boldsymbol{\mu} = \{\mu_t\}_{t\in[a,b]} \in \mathscr{A}^F_{[a,b]}(\mu)$ it is possible to find $\boldsymbol{\nu} = \{\nu_t\}_{t\in[a,b]} \in \mathscr{A}^F_{[a,b]}(\nu)$ satisfying*

$$W_2(\mu_t, \nu_t) \le K \cdot W_2(\mu, \nu), \text{ for all } t \in [a,b].$$

Proof. By the assumption on $F(\cdot)$, there exists a compact set U and a continuous function f, Lipschitz in the first variable uniformly w.r.t. the second, such that $F(x) = \{f(x,u) : u \in U\}$. By the Superposition Principle (Theorem A.8), let $\boldsymbol{\eta} = \mu \otimes \eta_x \in \mathscr{P}(\mathbb{R}^d \times \Gamma_{[a,b]})$ be such that $\mu_t = e_t \sharp \boldsymbol{\eta}$, for a suitable Borel family $\{\eta_x\}_{x\in\mathbb{R}^d} \subseteq \mathscr{P}(\Gamma_{[a,b]})$ uniquely defined for μ-a.e. $x \in \mathbb{R}^d$. Given an optimal transport plan $\pi \in \Pi_o(\mu,\nu)$, we define $\boldsymbol{\pi} = \pi \otimes \eta_x \in \mathscr{P}(\mathbb{R}^d \times \mathbb{R}^d \times \Gamma_{[a,b]})$.

For any $(x,\gamma) \in \mathbb{R}^d \times \Gamma_{[a,b]}$ we consider

$$H(x,\gamma) = \left\{ u \in L^1([a,b];U) : \gamma(t) - x - \int_0^t f(\gamma(s), u(s))\, ds = 0 \text{ for all } t \in [a,b] \right\}.$$

By Theorem 8.2.9 p.315 in [5], we can find a Borel map $(x,\gamma) \mapsto u_{x,\gamma}$ such that

$$\gamma(t) = x + \int_a^t f(\gamma(s), u_{x,\gamma}(s))\, ds, \text{ for all } t \in [a,b] \text{ and } \boldsymbol{\eta}-\text{a.e. } (x,\gamma) \in \mathbb{R}^d \times \Gamma_{[a,b]}.$$

Define now a Borel map $\tau : \mathbb{R}^d \times \Gamma_{[a,b]} \to \Gamma_{[a,b]}$ as follows:

- if $\dot{\hat{\gamma}} \in F(\hat{\gamma}(t))$ for a.e. $t \in [a,b]$, we set $\tau(\hat{y}, \hat{\gamma}) \in \Gamma_{[a,b]}$ to be the unique solution of

$$\begin{cases} \dot{\gamma}(t) = f(\gamma(t), u_{\hat{\gamma}(a),\hat{\gamma}}(t)), \text{ for a.e. } t \in [a,b], \\ \gamma(a) = \hat{y}. \end{cases}$$

- if $\dot{\hat{\gamma}} \notin F(\hat{\gamma}(t))$ for a.e. $t \in [a,b]$, we set $\tau(\hat{y}, \hat{\gamma}) = \hat{\gamma}$.

Clearly, for all $y \in \mathbb{R}^d$, a.e. $t \in [a,b]$, and $\boldsymbol{\eta}$-a.e. $(\gamma(a), \gamma) \in \mathbb{R}^d \times \Gamma_{[a,b]}$ we have

- $\tau(y,\gamma)(a) = y$;
- $\dfrac{d}{dt}\tau(y,\gamma)(t) = f(\tau(y,\gamma)(t), u_{\gamma(a),\gamma}(t)) \in F(\tau(y,\gamma)(t))$,
- $|\tau(y,\gamma)(t) - \gamma(t)| \le |y - \gamma(a)| \cdot e^{\mathrm{Lip}(f)(t-a)}$,

where the last assertion follows from Grönwall inequality since

$$|\tau(y,\gamma)(t) - \gamma(t)| \le |y - \gamma(a)| + \int_a^t |f(\tau(y,\gamma)(s), u_{\gamma(a),\gamma}(s)) - f(\gamma(s), u_{\gamma(a),\gamma}(s))|\, ds$$

$$\le |y - \gamma(a)| + \mathrm{Lip}(f) \int_a^t |\tau(y,\gamma)(s) - \gamma(s)|\, ds.$$

Define now $\boldsymbol{\nu} = \{\nu_t\}_{t\in[a,b]}$ by setting

$$\int_{\mathbb{R}^d} \varphi(y)\, d\nu_t(y) = \iiint_{\mathbb{R}^d\times\mathbb{R}^d\times\Gamma_{[a,b]}} \varphi \circ e_t(y, \tau(y,\gamma))\, d\boldsymbol{\pi}(x,y,\gamma),$$

for every $\varphi \in C_c^1(\mathbb{R}^d)$.

Evaluating the above expression for $t = a$ we have

$$\int_{\mathbb{R}^d} \varphi(y)\, d\nu_a(y) = \iiint_{\mathbb{R}^d\times\mathbb{R}^d\times\Gamma_{[a,b]}} \varphi \circ e_a(y, \tau(y,\gamma))\, d\boldsymbol{\pi}(x,y,\gamma)$$

$$= \iiint_{\mathbb{R}^d\times\mathbb{R}^d\times\Gamma_{[a,b]}} \varphi(y)\, d\boldsymbol{\pi}(x,y,\gamma) = \int_{\mathbb{R}^d} \varphi(y)\, d\nu(y),$$

and so $\nu_{|t=a} = \nu$. By deriving w.r.t. t, we obtain

$$\frac{d}{dt}\int_{\mathbb{R}^d} \varphi(y)\,d\nu_t(y) =$$

$$= \iiint_{\mathbb{R}^d\times\mathbb{R}^d\times\Gamma_{[a,b]}} \langle \nabla\varphi\circ e_t(y,\tau(y,\gamma)),\, f(\tau(y,\gamma)(t), u_{\gamma(a),\gamma}(t))\rangle\, d\boldsymbol{\pi}(x,y,\gamma).$$

Disintegrating $\boldsymbol{\pi}$ w.r.t. the map $g_t(x,y,\gamma) := e_t(y,\tau(y,\gamma))$, and recalling that $g_t\sharp\boldsymbol{\pi} = \nu_t$, we have $\boldsymbol{\pi} = \nu_t \otimes \boldsymbol{\pi}_{x,y,\gamma}$ and

$$\frac{d}{dt}\int_{\mathbb{R}^d} \varphi(y)\,d\nu_t(y) = \int_{\mathbb{R}^d} \langle \nabla\varphi(\xi), \iiint_{g_t^{-1}(\xi)} f(\xi, u_{\gamma(a),\gamma}(t))\,d\boldsymbol{\pi}_{x,y,\gamma}(x,y,\gamma)\rangle\, d\nu_t(\xi).$$

Recalling the convexity of F, we have

$$v_t(\xi) := \iiint_{g_t^{-1}(\xi)} f(\xi, u_{\gamma(a),\gamma}(t))\,d\boldsymbol{\pi}_{x,y,\gamma}(x,y,\gamma) \in F(\xi), \text{ for a.e. } t\in[a,b],$$

and so $\boldsymbol{\nu} \in \mathscr{A}^F_{[a,b]}(\nu)$ is an admissible trajectory.

Finally, set $\pi_{13}(x,y,\gamma) = (x,\gamma)$, we have $(e_t\circ\pi_{13}, g_t)\sharp\boldsymbol{\pi} \in \Pi(\mu_t,\nu_t)$, and so

$$\begin{aligned} W_2^2(\mu_t,\nu_t) &\le \int_{\mathbb{R}^d\times\mathbb{R}^d\times\Gamma_{[a,b]}} |\gamma(t) - \tau(y,\gamma)(t)|^2\, d\boldsymbol{\pi}(x,y,\gamma) \\ &\le e^{2\mathrm{Lip}(f)(t-a)} \int_{\mathbb{R}^d\times\mathbb{R}^d\times\Gamma_{[a,b]}} |y-\gamma(a)|^2\, d\boldsymbol{\pi}(x,y,\gamma) \\ &= e^{2\mathrm{Lip}(f)(t-a)} \int_{\mathbb{R}^d\times\mathbb{R}^d} |y-x|^2\, d\pi(x,y) = e^{2\mathrm{Lip}(f)(t-a)} W_2^2(\mu,\nu), \end{aligned}$$

and so we can choose $K = e^{2\mathrm{Lip}(f)(b-a)}$. The proof is complete. □

The following proposition illustrate the fact that for an initial condition $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ any selection $v(\cdot) \in F(\cdot)$ with $v \in L^2_\mu(\mathbb{R}^d)$ can be the initial velocity of an admissible trajectory (this is a well-known consequence of Filippov Theorem in the context of differential inclusions).

Proposition 2.5 *(Initial velocity of smooth trajectories). Let $a,b\in\mathbb{R}$, $a<b$, $\mu\in\mathscr{P}_2(\mathbb{R}^d)$, $F:\mathbb{R}^d \rightrightarrows \mathbb{R}^d$ be satisfying* $(\boldsymbol{F})$*, $\mu\in\mathscr{P}_2(\mathbb{R}^d)$. Then for every $v_a \in L^2_\mu(\mathbb{R}^d)$ such that $v_a(x)\in F(x)$ for μ-a.e. $x\in\mathbb{R}^d$ there exist $\boldsymbol{\eta}\in\mathscr{P}(\mathbb{R}^d\times\Gamma_{[a,b]})$ such that $\boldsymbol{\mu} = \{e_t\sharp\boldsymbol{\eta}\}_{t\in[a,b]} \in \mathscr{A}^F_{[a,b]}(\mu)$ and*

$$\lim_{t\to a^+} \int_{\mathbb{R}^d\times\Gamma_{[a,b]}} \langle \varphi\circ e_a(x,\gamma), \frac{e_t(x,\gamma)-e_a(x,\gamma)}{t-a}\rangle\, d\boldsymbol{\eta}(x,\gamma) = \int_{\mathbb{R}^d} \langle\varphi(x), v_a(x)\rangle\, d\mu(x).$$

Please cite this article in press as: A. Marigonda, M. Quincampoix, Mayer control problem with probabilistic uncertainty on initial positions, J. Differential Equations (2017), https://doi.org/10.1016/j.jde.2017.11.014

Proof. Without loss of generality, we assume that $[a,b]=[0,T]$. According to the assumptions on F, by Theorem 9.7.1 and Theorem 9.7.2 in [5], there exists $f:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}^d$ such that [1]

$$\begin{cases} x\mapsto f(x,u) \text{ is Lipschitz continuous with constant } 5d\,\mathrm{Lip}(F),\\ |f(x,u)-f(x,v)|\le 5d\cdot\max\{|y|:\ y\in F(x)\}\cdot|u-v|,\\ F(x)=\{f(x,u):\ u\in\overline{B(0,1)}\}.\end{cases}\tag{4}$$

By Filippov's Implicit Function Theorem (see e.g. Theorem 8.2.10 in [5]), there exists a measurable selection u_x of $F(\cdot)$ such that $v_0(x)=f(x,u_x)$ for μ-a.e. $x\in\mathbb{R}^d$. For every $x\in\mathbb{R}^d$ define γ_x to be the unique solution of $\dot\gamma_x(t)=f(\gamma_x(t),u_x)$, $\gamma_x(0)=x$. The map $x\mapsto\gamma_x$ is Borel, thus we can define the product measure $\boldsymbol{\eta}=\mu\otimes\delta_{\gamma_x}$. We notice that

$$\begin{aligned}|\gamma_x(t)-\gamma_x(0)-tf(\gamma_x(0),u_x)|&\le\int_0^t|f(\gamma_x(s),u_x)-f(\gamma_x(0),u_x)|\,ds\\ &\le 5d\,\mathrm{Lip}(F)\int_0^t|\gamma_x(s)-\gamma_x(0)-sf(\gamma_x(0),u_x)|\,ds+\frac{C}{2}t^2(|x|+1),\end{aligned}$$

where $C>0$ is a constant satisfying $|y|\le C(|x|+1)$ for all $y\in F(x)$, $x\in\mathbb{R}^d$. Using Grönwall inequality and dividing by t, we have

$$\left|\frac{\gamma_x(t)-\gamma_x(0)}{t}-f(\gamma_x(0),u_x)\right|\le\frac{CT}{2}e^{5dT\,\mathrm{Lip}(F)}(|x|+1),$$

which, squaring and integrating in x w.r.t. the measure μ, yields that the map $g_t(x):=\dfrac{\gamma_x(t)-\gamma_x(0)}{t}$ has L^2_μ norm bounded by $e^{5dT\,\mathrm{Lip}(F)}(CT+1)(\mathrm{m}_2^{1/2}(\mu)+1)$. For every $\varphi\in L^2_\mu(\mathbb{R}^d)$ we have

$$\lim_{t\to0^+}\int_{\mathbb{R}^d\times\Gamma_{[a,b]}}\langle\varphi\circ e_0(x,\gamma),\frac{e_t(x,\gamma)-e_a(x,\gamma)}{t-a}\rangle\,d\boldsymbol{\eta}(x,\gamma)=\lim_{t\to0^+}\int_{\mathbb{R}^d}\langle\varphi(x),g_t(x)\rangle\,d\mu(x)$$

Since by Hölder inequality we have

$$\|\langle\varphi,g_t\rangle\|_{L^1_\mu}\le\|\varphi\|_{L^2_\mu}\cdot e^{5dT\,\mathrm{Lip}(F)}(CT+1)(\mathrm{m}_2^{1/2}(\mu)+1),$$

we can apply the Dominated Convergence Theorem to pass to the limit under the integral sign, obtaining

$$\lim_{t\to0^+}\int_{\mathbb{R}^d}\langle\varphi(x),g_t(x)\rangle\,d\mu(x)=\int_{\mathbb{R}^d}\langle\varphi(x),\lim_{t\to0^+}g_t(x)\rangle\,d\mu(x)=\int_{\mathbb{R}^d}\langle\varphi(x),v_a(x)\rangle\,d\mu(x).\quad\square$$

[1] Of course when F comes from the control system (1) we do not change the parametrization of the map F.

2.2. The value function for the Mayer problem

Given $s \in [0,T]$, $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, we define the *value function* $V : [0,T] \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ by setting

$$V(s,\mu) = \inf\left\{\mathscr{G}(\mu_T) : \{\mu_t\}_{t\in[s,T]} \in \mathscr{A}^F_{[s,T]}(\mu)\right\}.$$

We say that $\{\mu_t\}_{t\in[s,T]} \in \mathscr{A}^F_{[s,T]}(\mu)$ is an *optimal trajectory* for $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ if $V(s,\mu) = \mathscr{G}(\mu_T)$.

Remark 2.6. From Proposition 2.3, since $\mathscr{G}(\cdot)$ is l.s.c., we deduce immediately the existence of optimal trajectories for every $\mu \in \mathscr{P}_2(\mathbb{R}^d)$.

Proposition 2.7 *(Dynamic Programming Principle for the Mayer problem). For all $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ and $\tau \in [0,T]$ we have*

$$V(\tau,\mu) = \inf\left\{V(s,\mu_s) : \{\mu_t\}_{t\in[\tau,T]} \in \mathscr{A}^F_{[\tau,T]}(\mu),\ s \in [\tau,T]\right\},$$

i.e., $V(\tau,\mu_\tau) \le V(s,\mu_s)$ for all $\tau \le s \le T$ and $\{\mu_t\}_{t\in[\tau,T]} \in \mathscr{A}^F_{[\tau,T]}(\mu)$, and $V(\tau,\mu_\tau) = V(s,\mu_s)$ for all $\tau \le s \le T$ if and only if $\{\mu_t\}_{t\in[\tau,T]}$ is an optimal trajectory for μ.

Proof. By contradiction, assume that there exist $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, $\boldsymbol{\mu} = \{\mu_t\}_{t\in[\tau,T]} \in \mathscr{A}^F_{[\tau,T]}(\mu)$, $\tau < s \le T$ and $\varepsilon > 0$ with $V(\tau,\mu_\tau) = V(s,\mu_s) + \varepsilon$. In particular, there exists $\hat{\boldsymbol{\mu}} = \{\hat\mu_t\}_{t\in[s,T]} \in \mathscr{A}^F_{[s,T]}(\mu_s)$ such that $V(s,\mu_s) \ge \mathscr{G}(\hat\mu_T) - \varepsilon/2$, and so $V(\tau,\mu_\tau) \ge \mathscr{G}(\hat\mu_T) + \varepsilon/2$. We consider the new trajectory $\bar{\boldsymbol{\mu}} = \{\bar\mu_t\}_{t\in[0,T]} \in \mathscr{A}^F_{[0,T]}$ defined as $\bar{\boldsymbol{\mu}} = \boldsymbol{\mu} \odot \hat{\boldsymbol{\mu}}$ (i.e. $\bar\mu_t = \mu_t$ for $t \in [0,s]$ and $\bar\mu_t = \hat\mu_t$ for $t \in [s,T]$). Clearly we have $\bar\mu_\tau = \mu_\tau$, $\bar\mu_s = \hat\mu_s = \mu_s$, $\bar\mu_T = \hat\mu_T$ and $\{\bar\mu_t\}_{t\in[\tau,T]} \in \mathscr{A}^F_{[\tau,T]}(\mu_\tau)$. By definition we must then have $V(\tau,\mu_\tau) \le \mathscr{G}(\bar\mu_T)$, leading to a contradiction with $V(\tau,\mu_\tau) - \varepsilon/2 \ge \mathscr{G}(\hat\mu_T) = \mathscr{G}(\bar\mu_T)$.

Assume now to have the equality $V(\tau,\mu_\tau) = V(s,\mu_s)$ for all $\tau \le s \le T$. In particular, we have $V(\tau,\mu_\tau) = V(T,\mu_T) = \mathscr{G}(\mu_T) =$, so $\{\mu_t\}_{t\in[\tau,T]}$ is an optimal trajectory for $\mu_\tau = \mu$. Conversely, assume that $\{\mu_t\}_{t\in[\tau,T]}$ is an optimal trajectory for μ, and take $s \in [\tau,T]$. By definition, we have $V(s,\mu_s) \le \mathscr{G}(\mu_T)$ since the restriction $\{\mu_t\}_{t\in[s,T]} \in \mathscr{A}^F_{[s,T]}(\mu_s)$, furthermore, by the monotonicity property we have $V(\tau,\mu) \le V(s,\mu_s) \le \mathscr{G}(\mu_T)$ but the first and the last term of the inequality coincides by the optimality assumption, so we have $V(\tau,\mu) = V(s,\mu_s) = \mathscr{G}(\mu_T)$ for all $s \in [\tau,T]$. □

Proposition 2.8 *(Regularity of the value function). Let $T > 0$, $F, \mathscr{G}$ be satisfying* ***(F)*** *and* $(\mathscr{G})$*, respectively. Then $V : [0,T] \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ is bounded and for every $K \ge 0$, it is Lipschitz continuous on the set $\{(t,\mu) \in [0,T] \times \mathscr{K},\ \mathrm{m}_2(\mu) \le K\}$.*

Proof. The boundedness follows directly from the definition.

We prove first the Lipschitz continuity w.r.t. the second variable, so let $s \in [0,T]$ be fixed. Fix $\varepsilon > 0$. Given $\mu^{(1)}, \mu^{(2)} \in \mathscr{P}_2(\mathbb{R}^d)$, let $\{\mu^{(2)}_t\}_{t\in[s,T]} \in \mathscr{A}^F(\mu^{(2)})$ be such that $V(s,\mu^{(2)}) \ge \mathscr{G}(\mu^{(2)}_T) - \varepsilon$. By Proposition 2.4, there exists $\{\mu^{(1)}_t\}_{t\in[s,T]} \in \mathscr{A}^F(\mu^{(1)})$ satisfying $W_2(\mu^{(1)}_T, \mu^{(2)}_T) \le K W_2(\mu^{(1)}, \mu^{(2)})$, where $K = e^{5d\mathrm{Lip}(F)\cdot(T-s)} \le e^{5d\mathrm{Lip}(F)\cdot T}$. We have

$$\begin{aligned} V(s,\mu^{(1)})-V(s,\mu^{(2)}) &\le \mathscr{G}(\mu_T^{(1)})-\mathscr{G}(\mu_T^{(2)})+\varepsilon \le \varepsilon+\mathrm{Lip}(\mathscr{G})\cdot W_2(\mu_T^{(1)},\mu_T^{(2)}) \\ &\le \varepsilon+\mathrm{Lip}(\mathscr{G})e^{5d\mathrm{Lip}(F)\cdot T}\cdot W_2(\mu^{(1)},\mu^{(2)}). \end{aligned}$$

By letting $\varepsilon\to 0^+$ and interchanging the roles of $\mu^{(1)}$ and $\mu^{(2)}$, we obtain

$$\left|V(s,\mu^{(1)})-V(s,\mu^{(2)})\right|\le \mathrm{Lip}(\mathscr{G})e^{5d\mathrm{Lip}(F)\cdot T}W_2(\mu^{(1)},\mu^{(2)}).$$

We prove now the Lipschitz continuity w.r.t. the first variable, so let $\mu\in\mathscr{P}_2(\mathbb{R}^d)$ be fixed, $s_1,s_2\in[0,T]$. Fix $\varepsilon>0$ and let $\{\mu_t^{(2)}\}_{t\in[s_2,T]}\in\mathscr{A}^F_{[s_2,T]}(\mu)$, be such that $V(s_2,\mu)\ge\mathscr{G}(\mu_T^{(2)})-\varepsilon$. This yields for all $t\in[s_2,T]$

$$V(t,\mu_t^{(2)})-V(s_2,\mu_{s_2}^{(2)})\le V(t,\mu_t^{(2)})-\mathscr{G}(\mu_T^{(2)})+\varepsilon\le\varepsilon.$$

We distinguish now two cases:

- Assume that $s_1\le s_2$. In this case, given any $\{\mu_t^{(1)}\}_{t\in[s_1,T]}\in\mathscr{A}^F_{[s_1,T]}(\mu)$, and recalling the monotonicity property provided by Proposition 2.7 and the fact that we have $\mu=\mu_{s_i}^{(i)}$, $i=1,2$, we have

$$\begin{aligned} V(s_1,\mu)-V(s_2,\mu) &\le V(s_2,\mu_{s_2}^{(1)})-V(s_2,\mu)=V(s_2,\mu_{s_2}^{(1)})-V(s_2,\mu_{s_1}^{(1)}) \\ &\le \mathrm{Lip}(\mathscr{G})e^{5d\mathrm{Lip}(F)\cdot T}W_2(\mu_{s_2}^{(1)},\mu_{s_1}^{(1)}) \\ &\le \mathrm{Lip}(\mathscr{G})e^{5d\mathrm{Lip}(F)\cdot T}W_2(e_{s_1}\sharp\boldsymbol{\eta_1},e_{s_2}\sharp\boldsymbol{\eta_1}) \\ &\le \mathrm{Lip}(\mathscr{G})e^{5d\mathrm{Lip}(F)\cdot T}\|e_{s_1}-e_{s_2}\|_{L^2_{\boldsymbol{\eta_1}}}, \end{aligned}$$

  where we used the Lipschitz continuity of $V(s_2,\cdot)$.
- Assume that $s_2\le s_1$, since $\mu=\mu_{s_i}^{(i)}$, $i=1,2$, we have

$$\begin{aligned} V(s_1,\mu)-V(s_2,\mu) &\le V(s_1,\mu_{s_2}^{(2)})-V(s_1,\mu_{s_1}^{(2)})+V(s_1,\mu_{s_1}^{(2)})-V(s_2,\mu_{s_2}^{(2)}) \\ &\le V(s_1,\mu_{s_2}^{(2)})-V(s_1,\mu_{s_1}^{(2)})+\varepsilon \\ &\le \varepsilon+\mathrm{Lip}(\mathscr{G})e^{5d\mathrm{Lip}(F)\cdot T}W_2(\mu_{s_2}^{(2)},\mu_{s_1}^{(2)}) \\ &= \varepsilon+\mathrm{Lip}(\mathscr{G})e^{5d\mathrm{Lip}(F)\cdot T}\|e_{s_1}-e_{s_2}\|_{L^2_{\boldsymbol{\eta_2}}}. \end{aligned}$$

By Proposition 2.3 applied to $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$, we have

$$\|e_{s_1}-e_{s_2}\|_{L^2_{\boldsymbol{\eta_i}}}\le(CTe^{2TC}+1)Ce^{2TC}(1+\mathrm{m}_2^{1/2}(\mu))|s_1-s_2|,$$

where $C=\max_{y\in F(0)}|y|+\mathrm{Lip}(F)$. Having defined

$$K''=2K'(CTe^{2TC}+1)Ce^{2TC}\sup_{\mu\in\mathscr{K}}(1+\mathrm{m}_2^{1/2}(\mu)),$$

we have

$$V(s_1,\mu) - V(s_2,\mu) \le \varepsilon + K''|s_1 - s_2|,$$

and we conclude by letting $\varepsilon \to 0^+$ and interchanging the roles of s_1 and s_2. □

We will show now a feature that marks a significative difference between the classical case and our framework.

Lemma 2.9. *For each initial point $x(0) = x_0$, consider the reachable set at time T.*

$$R(T; x_0) := \{x(T) : \dot{x}(t) \in F(x(t)) \text{ for a.e. } t \in [0,T],\ x(0) = x_0\}.$$

Then, for all $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ we have

$$V(0,\mu) \le \inf\{\mathscr{G}(\phi\sharp\mu) : \phi \text{ is a Borel selection of } R(T;\cdot)\},$$

and the inequality may be strict.

Proof. To prove the statement it is enough to show that given a Borel selection ϕ of $R(T;\cdot)$, we can represent $\phi\sharp\mu$ as terminal point of an admissible trajectory. For every $x \in \mathbb{R}^d$, we consider the set

$$\tilde{R}_\phi(T; x) := \{\gamma \in \Gamma : \dot{\gamma}(t) \in F(\gamma(t)) \text{ for a.e. } t \in [0,T],\ \gamma(0) = x,\ \gamma(T) = \phi(x)\}.$$

The set-valued map $\tilde{R}_\phi(T;\cdot) : \mathbb{R}^d \rightrightarrows \Gamma$ is Borel according to the properties of $F(\cdot)$, and so we can find a Borel map $\tilde{\phi} : \mathbb{R}^d \to \Gamma$ such that $\tilde{\phi}(x)(\cdot)$ is an admissible trajectory of the finite-dimensional differential inclusion joining x and $\phi(x)$. Set now $\boldsymbol{\eta}_\phi = \mu \otimes \delta_{\tilde{\phi}(x)} \in \mathscr{P}(\mathbb{R}^d \times \Gamma)$. Since for $\boldsymbol{\eta}_\phi$-a.e. $(x,\gamma) \in \mathbb{R}^d \times \Gamma$ we have that $\gamma(0) = x$ and γ is an admissible trajectory for F, we have that $\boldsymbol{\mu} := \{\mu_t\}_{t\in[0,T]}$ defined by $\mu_t = e_t\sharp\boldsymbol{\eta}_\phi$ is an admissible trajectory satisfying $\mu_0 = \mu$ and $\mu_T = \phi\sharp\mu$, as desired. This trajectory is indeed driven by $\boldsymbol{\nu} := \{\nu_t\}_{t\in[0,T]}$, where $\nu_t = v_t\mu_t$ and for μ_t-a.e. $x \in \mathbb{R}^d$

$$v_t(x) := \int_{e_t^{-1}(y)} \dot{\gamma}(t)\, d\eta_{t,x}(y,\gamma) \in F(x),$$

where $\eta_{t,x}$ is the disintegration of $\boldsymbol{\eta}_\phi$ w.r.t. e_t, i.e. $\boldsymbol{\eta}_\phi = e_t \otimes \eta_{t,x}$. We refer the reader to e.g. [18] for the details. □

The example below shows that the inequality may be strict in general.

Example 2.10. In $\mathbb{R}$, set $F(x) = [-1,1]$ for all $x \in \mathbb{R}$, $T = 1$. For every $x_0 \in \mathbb{R}$, we have $R(T; x_0) = [x_0 - 1, x_0 + 1]$. Denoted by $\delta_a \in \mathscr{P}(\mathbb{R})$ the Dirac delta with mass concentrated in $a \in \mathbb{R}$, we define the terminal cost $\mathscr{G}(\mu) := \min\{2, W_2(\mu,\theta)\}$ where $\theta = \dfrac{1}{2}(\delta_{-1} + \delta_1)$. The map

Please cite this article in press as: A. Marigonda, M. Quincampoix, Mayer control problem with probabilistic uncertainty on initial positions, J. Differential Equations (2017), https://doi.org/10.1016/j.jde.2017.11.014

$\mathscr{G}:\mathscr{P}_2(\mathbb{R})\to\mathbb{R}$ is Lipschitz continuous w.r.t. W_2. We notice that there are no maps $\phi:\mathbb{R}\to\mathbb{R}$ satisfying $\phi(x_0)\in R(T;x_0)=[x_0-1,x_0+1]$ and $\theta=\phi\sharp\delta_0=\delta_{\phi(0)}$, in fact

$$\theta([-1,1]\setminus\{\phi(0)\})\geq 1/2>\phi\sharp\delta_0([-1,1]\setminus\{\phi(0)\})=0.$$

We compute now

$$W_2^2(\phi\sharp\delta_0,\theta)=\inf_{\pi\in\Pi(\phi\sharp\delta_0,\theta)}\int_{\mathbb{R}\times\mathbb{R}}|x-y|^2\,d\pi(x,y).$$

Since $\phi\sharp\delta_0=\delta_{\phi(0)}$, the set of admissible transport plans $\Pi(\phi\sharp\delta_0,\theta)$ reduces to the unique element $\pi=\delta_{\phi(0)}\otimes\theta$, and so

$$W_2^2(\phi\sharp\delta_0,\theta)=\int_{\mathbb{R}\times\mathbb{R}}|x-y|^2\,d(\delta_{\phi(0)}\otimes\theta)=\frac{1}{2}|\phi(0)-1|^2+\frac{1}{2}|\phi(0)+1|^2.$$

Since $\phi(0)\in[-1,1]$, we obtain $W_2^2(\phi\sharp\delta_0,\theta)\geq 1$. In particular, $\mathscr{G}(\phi\sharp\delta_0)\geq 1$ for *every* map $\phi:\mathbb{R}\to\mathbb{R}$ such that $\phi(y)\in R(T;y)$.

Set now $v:\mathbb{R}\to\mathbb{R}$ to be $v(x)=\text{sign}(x)$ for $x\neq 0$, $v(0)=0$, and $\mu_t=\frac{1}{2}(\delta_{-t}+\delta_t)$, we have that $\boldsymbol{\mu}=\{\mu_t\}_{t\in[0,1]}$ solves $\partial_t\mu_t+\text{div}(v\mu_t)=0$ according to the Superposition Principle (Theorem 8.2.1 in [2]), moreover $\mu_0=\delta_0$ and $\mu_1=\theta$ and $v(x)\in F(x)$ for all $x\in\mathbb{R}^d$. Thus in this case θ can be reached from δ_0 at time 1, and then $V(0,\delta_0)=\mathscr{G}(\theta)=0$.

3. Hamilton–Jacobi–Bellman equations in Wasserstein space

The aim of this section is to introduce the essential differential structure on $\mathscr{P}_2(\mathbb{R}^d)$ in order to define suitable notions of sub/super-differentials and viscosity solution (cf. [22] for viscosity solution of Hamilton Jacobi equations not stated in a finite dimensional space).

Lemma 3.1 *(Representation of optimal plans). Let $\mu,\nu\in\mathscr{P}_2(\mathbb{R}^d)$, $\gamma\in\Pi_o(\mu,\nu)$. Then*

- *there exist unique functions $p_\gamma^\mu\in L^2_\mu(\mathbb{R}^d)$ and $q_\gamma^\nu\in L^2_\nu(\mathbb{R}^d)$ such that for all Borel map $\phi:\mathbb{R}^d\to\mathbb{R}^d$ satisfying $\phi\in L^2_\mu(\mathbb{R}^d)\cap L^2_\nu(\mathbb{R}^d)$ we have*

$$\int_{\mathbb{R}^d\times\mathbb{R}^d}\langle\phi(x),x-y\rangle\,d\gamma(x,y)=\int_{\mathbb{R}^d}\langle\phi(x),p_\gamma^\mu(x)\rangle\,d\mu(x)=\int_{\mathbb{R}^d}\langle\phi(y),q_\gamma^\mu(y)\rangle\,d\nu(y),$$

- *we have $p_\gamma^\mu=\text{Id}_{\mathbb{R}^d}-\text{Bar}_1(\gamma)$, $q_\gamma^\nu=\text{Id}_{\mathbb{R}^d}-\text{Bar}_1(\gamma^{-1})$ where the barycenter Bar_1 is defined in Definition A.5.*

Proof. The first statement has been proved in Lemma 4 of [12]. Indeed, to prove the existence of p_γ^μ is enough to notice that

$$\phi \mapsto \int_{\mathbb{R}^d\times\mathbb{R}^d} \langle \phi(x), x-y\rangle\, d\gamma(x,y),$$

is a linear and continuous operator from $L^2_\mu(\mathbb{R}^d)$ to $\mathbb{R}$, and then to use Riesz representation theorem. Similarly, we prove the existence of q^ν_γ, by noticing that

$$\phi \mapsto \int_{\mathbb{R}^d\times\mathbb{R}^d} \langle \phi(x), x-y\rangle\, d\gamma(x,y) = \int_{\mathbb{R}^d\times\mathbb{R}^d} \langle \phi(y), y-x\rangle\, d\gamma^{-1}(x,y),$$

is a linear and continuous operator from $L^2_\nu(\mathbb{R}^d)$ to $\mathbb{R}$.

The second statement follows from the disintegration theorem w.r.t. the first marginals of γ and γ^{-1}, respectively. Indeed, if $\gamma = \mu \otimes \gamma_x$ and $\gamma^{-1} = \nu \otimes \gamma^{-1}_y$, we can identify $\{\gamma_x\}_{x\in\mathbb{R}^d}$ and $\{\gamma_y\}_{y\in\mathbb{R}^d}$ with subsets of $\mathscr{P}_2(\mathbb{R}^d)$, obtaining for all $\phi \in L^2_\mu(\mathbb{R}^d)\cap L^2_\nu(\mathbb{R}^d)$

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \langle \phi(x), x-y\rangle\, d\gamma(x,y) = \begin{cases} \displaystyle\int_{\mathbb{R}^d} \langle \phi(x), x - \int_{\mathbb{R}^d} y\, d\gamma_x(y)\rangle\, d\mu(x), \\ \displaystyle\int_{\mathbb{R}^d} \langle \phi(y), y - \int_{\mathbb{R}^d} x\, d\gamma^{-1}_y(x)\rangle\, d\nu(y). \end{cases} \qquad \square$$

We introduce now a notion of viscosity sub/super-differentials that will be used in the rest of the paper. The comparison between this notion of sub/super-differential and other notions available in literature is discussed in Appendix B.

Definition 3.2 *(Viscosity sub/super-differentials).* Let $w : [0,T]\times\mathscr{P}_2(\mathbb{R}^d)\to\mathbb{R}$ be a map, $(\bar t,\bar\mu)\in]0,T[\times\mathscr{P}_2(\mathbb{R}^d)$, $\delta>0$. We say that $(p_{\bar t}, p_{\bar\mu})\in\mathbb{R}\times L^2_{\bar\mu}(\mathbb{R}^d)$ belongs to the viscosity δ-superdifferential of w at $(\bar t,\bar\mu)$ if

i.) there exists $\bar\nu$ and $\gamma\in\Pi_o(\bar\mu,\bar\nu)$ such that for all Borel map $\phi:\mathbb{R}^d\to\mathbb{R}^d$ satisfying $\phi\in L^2_\mu(\mathbb{R}^d)\cap L^2_\nu(\mathbb{R}^d)$ we have

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \langle \phi(x), x-y\rangle\, d\gamma(x,y) = \int_{\mathbb{R}^d} \langle \phi(x), p^\mu_\gamma(x)\rangle\, d\mu(x),$$

i.e., $p_{\bar\mu} = p^{\bar\mu}_\gamma$ where $p^{\bar\mu}_\gamma$ is defined as in Lemma 3.1.

ii.) for all $\mu\in\mathscr{P}_2(\mathbb{R}^d)$ we have

$$\begin{aligned} w(t,\mu) - w(\bar t,\bar\mu) \le p_t(t-\bar t) + &\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} \langle x_2, x_3 - x_1\rangle\, d\tilde\mu(x_1,x_2,x_3) + \\ &+ \delta\sqrt{(t-\bar t)^2 + W^2_{2,\tilde\mu}(\bar\mu,\mu)} + o(|t-\bar t| + W_{2,\tilde\mu}(\bar\mu,\mu)), \end{aligned}$$

for all $\tilde\mu\in\mathscr{P}(\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d)$ satisfying $\pi_{12}\sharp\tilde\mu = (\mathrm{Id}_{\mathbb{R}^d}, p_{\bar\mu})\sharp\bar\mu$ and $\pi_{13}\sharp\tilde\mu\in\Pi(\bar\mu,\mu)$.

Please cite this article in press as: A. Marigonda, M. Quincampoix, Mayer control problem with probabilistic uncertainty on initial positions, J. Differential Equations (2017), https://doi.org/10.1016/j.jde.2017.11.014

We denote the set of the viscosity δ-superdifferentials of w at $(\bar{t},\bar{\mu})$ by $D^+_\delta w(\bar{t},\bar{\mu})$. Similarly, we define the set of the viscosity δ-subdifferentials $D^-_\delta w(\bar{t},\bar{\mu})$ of w at $(\bar{t},\bar{\mu})$ by setting $-D^-_\delta w(\bar{t},\bar{\mu}) = D^+_\delta(-w)(\bar{t},\bar{\mu})$.

We will use the following concept of viscosity solution (see [12]).

Definition 3.3 *(Hamilton–Jacobi–Bellman Equation).* We consider an equation in the form

$$\partial_t w(t,\mu) + \mathscr{H}(\mu, Dw(t,\mu)) = 0, \tag{5}$$

where $\mathscr{H}(\mu,p)$ is defined for any $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ and $p \in L^2_\mu(\mathbb{R}^d)$. We say that a function $w : [0,T] \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ is

- a *subsolution* of (5) if w is u.s.c. and there exists a constant $C > 0$ such that
$$p_t + \mathscr{H}(\mu, p_\mu) \geq -C\delta,$$
for all $(t,\mu) \in]0,T[\times \mathscr{P}_2(\mathbb{R}^d)$, $(p_t,p_\mu) \in D^+_\delta w(t_0,\mu_0)$, and $\delta > 0$.
- a *supersolution* of (5) if w is l.s.c. and there exists a constant $C > 0$ such that
$$p_t + \mathscr{H}(\mu, p_\mu) \leq C\delta,$$
for all $(t,\mu) \in]0,T[\times \mathscr{P}_2(\mathbb{R}^d)$, $(p_t,p_\mu) \in D^-_\delta w(t_0,\mu_0)$, and $\delta > 0$.
- a *solution* of (5) if w is both a supersolution and a subsolution.

We will prove now a comparison principle between sub- and supersolutions by using the doubling of variable method.

Theorem 3.4 *(Comparison principle). Consider the equation* (5) *for an Hamiltonian function $\mathscr{H}$ satisfying the following properties*

- *positive homogeneity: for every $\lambda \geq 0$, $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, $p \in L^2_\mu(\mathbb{R}^d)$ we have $\mathscr{H}(\mu,\lambda p) = \lambda\mathscr{H}(\mu,p)$;*
- *dissipativity: there exists $k \geq 0$ such that for all $\mu,\nu \in \mathscr{P}_2(\mathbb{R}^d)$, $\gamma \in \Pi_o(\mu,\nu)$, defined $p^\mu_\gamma = \mathrm{Id}_{\mathbb{R}^d} - \mathrm{Bar}_1(\gamma)$, $q^\nu_\gamma = \mathrm{Id}_{\mathbb{R}^d} - \mathrm{Bar}_1(\gamma^{-1})$, we have*
$$\mathscr{H}_F(\mu, p_\mu) - \mathscr{H}_F(\nu, q_\nu) \leq k W_2^2(\mu,\nu).$$

Let w_1 be a bounded and Lipschitz continuous subsolution and w_2 be a bounded and Lipschitz continuous supersolution to (5). *Then*

$$\inf_{(s,\mu)\in[0,T]\times\mathscr{P}_2(\mathbb{R}^d)} w_2(s,\mu) - w_1(s,\mu) = \inf_{\mu\in\mathscr{P}_2(\mathbb{R}^d)} w_2(T,\mu) - w_1(T,\mu).$$

In particular, equation (5) *admits at most one Lipschitz continuous bounded solution.*

Proof. The proof will be in the same spirit of Theorem 1 of [12].

Without loss of generality, we may assume $k \geq \mathrm{Lip}(w_i)$, $i = 1, 2$. Set

$$A := \inf_{\mu \in \mathscr{P}_2(\mathbb{R}^d)} w_2(T, \mu) - w_1(T, \mu),$$

and notice that since $\mathscr{H}$ does not involve w, $w_1 - A$ is still a subsolution. Thus, without loss of generality, we can assume $A = 0$. We will prove the result by contradiction, so assume that

$$-\xi := \inf_{(s,\mu) \in [0,T] \times \mathscr{P}_2(\mathbb{R}^d)} w_2(s, \mu) - w_1(s, \mu) < 0,$$

and choose $(t_0, \mu_0) \in [0, T] \times \mathscr{P}_2(\mathbb{R}^d)$ such that $w_2(t_0, \mu_0) - w_1(t_0, \mu_0) < -\xi/2$.

Consider now the space $X = [0, T] \times \mathscr{P}_2(\mathbb{R}^d)$ endowed with the metric

$$d_X(\xi_1, \xi_2) = \sqrt{(s_1 - s_2)^2 + W_2^2(\mu_1, \mu_2)} \text{ where } \xi_i = (s_i, \mu_i),\ i = 1, 2.$$

Clearly, (X, d_X) is a complete metric space. We endow $X \times X$ with the metric $d_{X \times X}$ defined by

$$d_{X \times X}(z_1, z_2) = d_X((s_1, \mu_1), (s_2, \mu_2)) + d_X((t_1, \nu_1), (t_2, \nu_2)),$$

for all $z_i = (s_i, \mu_i, t_i, \nu_i) \in X \times X$, $i = 1, 2$. Again, we have that $(X \times X, d_{X \times X})$ is a complete metric space.

Given $\varepsilon, \eta > 0$, we define the functional $\Phi_{\varepsilon\eta} : X \times X \to \mathbb{R}$ by setting

$$\Phi_{\varepsilon\eta}(s, \mu, t, \nu) = -w_1(s, \mu) + w_2(t, \nu) + \frac{1}{\varepsilon} d_X^2((s, \mu), (t, \nu)) - \eta s.$$

Define $z_0 = (t_0, \mu_0, t_0, \mu_0) \in X \times X$. Since $\Phi_{\varepsilon\eta}$ is continuous and bounded from below, and $(X \times X, d_{X \times X})$ is complete, by Ekeland Variational Principle (see e.g. Theorem 1 p.255 in [4]), for any $\delta > 0$ there exists $z_{\varepsilon\eta\delta} = (s_{\varepsilon\eta\delta}, \mu_{\varepsilon\eta\delta}, t_{\varepsilon\eta\delta}, \nu_{\varepsilon\eta\delta}) \in X \times X$ such that for any $z = (s, \mu, t, \nu) \in X \times X$ we have

$$\begin{cases} \Phi_{\varepsilon\eta}(z_{\varepsilon\eta\delta}) + \delta d_{X \times X}(z_0, z_{\varepsilon\eta\delta}) \leq \Phi_{\varepsilon\eta}(z_0), \\ \Phi_{\varepsilon\eta}(z_{\varepsilon\eta\delta}) \leq \Phi_{\varepsilon\eta}(z) + \delta d_{X \times X}(z, z_{\varepsilon\eta\delta}). \end{cases} \tag{6}$$

Furthermore, we set $\rho_{\varepsilon\eta\delta} = d_X((s_{\varepsilon\eta\delta}, \mu_{\varepsilon\eta\delta}), (t_{\varepsilon\eta\delta}, \nu_{\varepsilon\eta\delta}))$.

By taking $z = (s_{\varepsilon\eta\delta}, \mu_{\varepsilon\eta\delta}, s_{\varepsilon\eta\delta}, \mu_{\varepsilon\eta\delta})$ in (6), we have

$$\Phi_{\varepsilon\eta}(z_{\varepsilon\eta\delta}) \leq \Phi_{\varepsilon\eta}(s_{\varepsilon\eta\delta}, \mu_{\varepsilon\eta\delta}, s_{\varepsilon\eta\delta}, \mu_{\varepsilon\eta\delta}) + \delta \rho_{\varepsilon\eta\delta}.$$

Recalling the definition of Φ, this implies

$$\begin{aligned} &- w_1(s_{\varepsilon\eta\delta}, \mu_{\varepsilon\eta\delta}) + w_2(t_{\varepsilon\eta\delta}, \nu_{\varepsilon\eta\delta}) + \frac{1}{\varepsilon} \rho_{\varepsilon\eta\delta}^2 - \eta s_{\varepsilon\eta\delta} \leq \\ &\qquad \leq -w_1(s_{\varepsilon\eta\delta}, \mu_{\varepsilon\eta\delta}) + w_2(s_{\varepsilon\eta\delta}, \mu_{\varepsilon\eta\delta}) - \eta s_{\varepsilon\eta\delta} + \delta \rho_{\varepsilon\eta\delta}, \end{aligned}$$

thus

$$w_2(t_{\varepsilon\eta\delta}, \nu_{\varepsilon\eta\delta}) - w_2(s_{\varepsilon\eta\delta}, \mu_{\varepsilon\eta\delta}) + \frac{1}{\varepsilon}\rho^2_{\varepsilon\eta\delta} \le \delta\rho_{\varepsilon\eta\delta},$$

which implies $\rho_{\varepsilon\eta\delta} \le \varepsilon(k+\delta)$ recalling the smoothness assumptions on w_2.

Claim 1: If $s_{\varepsilon\eta\delta}, t_{\varepsilon\eta\delta} \in]0,T[$ *then*

$$\left(\frac{2}{\varepsilon}(s_{\varepsilon\eta\delta} - t_{\varepsilon\eta\delta}) - \eta, \frac{2}{\varepsilon}p_{\varepsilon\eta\delta}\right) \in D^+_\delta w_1(s_{\varepsilon\eta\delta}, \mu_{\varepsilon\eta\delta}), \tag{7}$$

$$\left(\frac{2}{\varepsilon}(s_{\varepsilon\eta\delta} - t_{\varepsilon\eta\delta}), \frac{2}{\varepsilon}q_{\varepsilon\eta\delta}\right) \in D^-_\delta w_1(s_{\varepsilon\eta\delta}, \mu_{\varepsilon\eta\delta}), \tag{8}$$

where $p_{\varepsilon\eta\delta} = \mathrm{Id}_{\mathbb{R}^d} - \mathrm{Bar}_1(\gamma) \in L^2_{\mu_{\varepsilon\eta\delta}}(\mathbb{R}^d)$, $q_{\varepsilon\eta\delta} = \mathrm{Id}_{\mathbb{R}^d} - \mathrm{Bar}_1(\gamma^{-1}) \in L^2_{\nu_{\varepsilon\eta\delta}}(\mathbb{R}^d)$, and $\gamma \in \Pi_o(\mu_{\varepsilon\eta\delta}, \nu_{\varepsilon\eta\delta})$ is the unique solution of the minimization problem

$$\min\{\|\mathrm{Id}_{\mathbb{R}^d} - \mathrm{Bar}_1(\gamma')\|_{L^2_{\mu_{\varepsilon\eta\delta}}} : \gamma' \in \Pi_o(\mu_{\varepsilon\eta\delta}, \nu_{\varepsilon\eta\delta})\}.$$

Proof (of Claim 1). By taking $t = t_{\varepsilon\eta\delta}$ and $\nu = \nu_{\varepsilon\eta\delta}$ in (6), we have

$$\Phi_{\varepsilon\eta}(z_{\varepsilon\eta\delta}) \le \Phi_{\varepsilon\eta}(s, \mu, t_{\varepsilon\eta\delta}, \nu_{\varepsilon\eta\delta}) + \delta d_X((s,\mu), (t_{\varepsilon\eta\delta}, \nu_{\varepsilon\eta\delta})), \text{ for all } (s,\mu) \in X,$$

which, recalling the definition of Φ, yields

$$\begin{aligned} w_1(s,\mu) - w_1(s_{\varepsilon\eta\delta}, \mu_{\varepsilon\eta\delta}) \le \\ \le \frac{1}{\varepsilon}\left[W_2^2(\mu, \nu_{\varepsilon\eta\delta}) - W_2^2(\mu_{\varepsilon\eta\delta}, \nu_{\varepsilon\eta\delta}) + (s - t_{\varepsilon\eta\delta})^2 - (s_{\varepsilon\eta\delta} - t_{\varepsilon\eta\delta})^2\right] + \\ + \delta\sqrt{W_2^2(\mu, \mu_{\varepsilon\eta\delta}) + |s - s_{\varepsilon\eta\delta}|^2} + \eta(s_{\varepsilon\eta\delta} - s). \end{aligned} \tag{9}$$

Recalling the choice of γ, the definition of $p_{\varepsilon\eta\delta}$, and Theorem B.5(3), for every $\tilde{\mu} \in \mathscr{P}(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$ satisfying $\pi_3 \sharp \tilde{\mu} = \mu$ and $\pi_{12} \sharp \tilde{\mu} = (\mathrm{Id}_{\mathbb{R}^d}, p_{\varepsilon\eta\delta}) \sharp \mu_{\varepsilon\eta\delta}$ we have

$$\begin{aligned} \frac{1}{2}W_2^2(\mu, \nu_{\varepsilon\eta\delta}) - \frac{1}{2}W_2^2(\mu_{\varepsilon\eta\delta}, \nu_{\varepsilon\eta\delta}) \le \\ \le \int_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} \langle x_2, x_3 - x_1 \rangle \, d\tilde{\mu}(x_1, x_2, x_3) + o(W^2_{2,\tilde{\mu}}(\mu_{\varepsilon\eta\delta}, \mu)), \\ = \int_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} \langle p_{\varepsilon\eta\delta}(x_1), x_3 - x_1 \rangle \, d(\pi_{13} \sharp \tilde{\mu})(x_1, x_3) + o(W^2_{2,\tilde{\mu}}(\mu_{\varepsilon\eta\delta}, \mu)), \end{aligned} \tag{10}$$

In particular, the conditions on $\tilde{\mu}$ imply also $\pi_{13} \sharp \tilde{\mu} \in \Pi(\mu_{\varepsilon\eta\delta}, \mu)$. By combining (10) and (9) we obtain (7) recalling the definition of viscosity superdifferential. The proof of (7) is symmetric, and this ends the proof of Claim 1. ◇

Claim 2: Assume that $-2k\varepsilon(k+\delta)^2 \geq C\delta - \eta$, then $s_{\varepsilon\eta\delta}, t_{\varepsilon\eta\delta} \notin]0,T[$.

Proof (of Claim 2). We argue by contradiction, assuming that $s_{\varepsilon\eta\delta}, t_{\varepsilon\eta\delta} \notin]0,T[$. By Claim 1, since w_1 and w_2 are a sub- and super-solution, respectively, and recalling the positive homogeneity of the Hamiltonian, we have

$$-C\delta \leq \frac{1}{\varepsilon}(s_{\varepsilon\eta\delta} - t_{\varepsilon\eta\delta}) - \eta + \mathscr{H}\left(\mu_{\varepsilon\eta\delta}, \frac{2}{\varepsilon}p_{\varepsilon\eta\delta}\right) = \frac{1}{\varepsilon}(s_{\varepsilon\eta\delta} - t_{\varepsilon\eta\delta}) - \eta + \frac{2}{\varepsilon}\mathscr{H}\left(\mu_{\varepsilon\eta\delta}, p_{\varepsilon\eta\delta}\right)$$

$$C\delta \geq \frac{1}{\varepsilon}(s_{\varepsilon\eta\delta} - t_{\varepsilon\eta\delta}) + \mathscr{H}\left(\nu_{\varepsilon\eta\delta}, \frac{2}{\varepsilon}q_{\varepsilon\eta\delta}\right) = \frac{1}{\varepsilon}(s_{\varepsilon\eta\delta} - t_{\varepsilon\eta\delta}) + \frac{2}{\varepsilon}\mathscr{H}\left(\nu_{\varepsilon\eta\delta}, q_{\varepsilon\eta\delta}\right),$$

where C is the constant appearing in Definition 3.3. By combining the above relations, we have

$$\mathscr{H}\left(\nu_{\varepsilon\eta\delta}, q_{\varepsilon\eta\delta}\right) - \mathscr{H}\left(\mu_{\varepsilon\eta\delta}, p_{\varepsilon\eta\delta}\right) \leq \frac{\varepsilon}{2}(C\delta - \eta).$$

By assumption, we have

$$\mathscr{H}\left(\nu_{\varepsilon\eta\delta}, q_{\varepsilon\eta\delta}\right) - \mathscr{H}\left(\mu_{\varepsilon\eta\delta}, p_{\varepsilon\eta\delta}\right) \geq -k\rho^2_{\varepsilon\eta\delta},$$

and so, recalling that $\rho_{\varepsilon\eta\delta} \leq \varepsilon(k+\delta)$,

$$-k\left(\varepsilon(k+\delta)\right)^2 \leq \frac{\varepsilon}{2}(C\delta - \eta),$$

leading to a contradiction with the choice of $\varepsilon, \delta, \eta$. ◇

Claim 3: Assume $\xi > 2\eta T - 2\varepsilon(k+\delta)\delta$, then $s_{\varepsilon\eta\delta} \neq T$ and $t_{\varepsilon\eta\delta} \neq T$.

Proof (of Claim 3). We notice that, by definition of ξ and recalling (6),

$$-\frac{\xi}{2} \geq w_2(t_0, \mu_0) - w_1(t_0, \mu_0) - \eta t_0 = \Phi_{\varepsilon\eta}(z_0) \geq \Phi_{\varepsilon\eta}(z_{\varepsilon\eta\delta}).$$

We prove the assertion by contradiction, assuming first $s_{\varepsilon\eta\delta} = T$.

$$\begin{aligned}-\frac{\xi}{2} &\geq \Phi_{\varepsilon\eta}(T, \mu_{\varepsilon\eta\delta}, t_{\varepsilon\eta\delta}, \nu_{\varepsilon\eta\delta}) = -w_1(T, \mu_{\varepsilon\eta\delta}) + w_2(t_{\varepsilon\eta\delta}, \nu_{\varepsilon\eta\delta}) + \frac{1}{\varepsilon}\rho^2_{\varepsilon\eta\delta} - \eta T \\ &\geq - w_1(T, \mu_{\varepsilon\eta\delta}) + w_2(T, \mu_{\varepsilon\eta\delta}) + \frac{1}{\varepsilon}\rho^2_{\varepsilon\eta\delta} - k\rho_{\varepsilon\eta\delta} - \eta T\end{aligned}$$

Since we have assumed $A = 0$, we have $0 \leq -w_1(T, \mu_{\varepsilon\eta\delta}) + w_2(T, \mu_{\varepsilon\eta\delta})$, thus

$$-\frac{\xi}{2} \geq \varepsilon(k+\delta)\delta - \eta T,$$

which leads to a contradiction with the choice of $\varepsilon, \delta, \eta$. Thus $s_{\varepsilon\eta\delta} \neq T$ and the proof showing $t_{\varepsilon\eta\delta} \neq T$ can be done in the same way.

We show now that $s_{\varepsilon\eta\delta} \neq 0$. Since $\Phi_{\varepsilon\eta}$ is continuous, there exists $h_{\varepsilon\eta\delta} > 0$ such that

$$\Phi_{\varepsilon\eta}(0, \mu_{\varepsilon\eta\delta}, t_{\varepsilon\eta\delta}, \nu_{\varepsilon\eta\delta}) \geq \Phi_{\varepsilon\eta}(h, \mu_{\varepsilon\eta\delta}, t_{\varepsilon\eta\delta}, \nu_{\varepsilon\eta\delta}) - \eta T,$$

Please cite this article in press as: A. Marigonda, M. Quincampoix, Mayer control problem with probabilistic uncertainty on initial positions, J. Differential Equations (2017), https://doi.org/10.1016/j.jde.2017.11.014

for all $0 \leq h \leq h_{\varepsilon\eta\delta}$, so we have

$$\begin{aligned}
-\frac{\xi}{2} &\geq \Phi_{\varepsilon\eta}(0, \mu_{\varepsilon\eta\delta}, t_{\varepsilon\eta\delta}, \nu_{\varepsilon\eta\delta}) \geq \Phi_{\varepsilon\eta}(h, \mu_{\varepsilon\eta\delta}, t_{\varepsilon\eta\delta}, \nu_{\varepsilon\eta\delta}) - \eta T \\
&= - w_1(h, \mu_{\varepsilon\eta\delta}) + \big(w_2(t_{\varepsilon\eta\delta}, \nu_{\varepsilon\eta\delta}) - w_2(0, \mu_{\varepsilon\eta\delta})\big) + w_2(0, \mu_{\varepsilon\eta\delta}) + \frac{1}{\varepsilon}\rho^2_{\varepsilon\eta\delta} - \eta T \\
&\geq - w_1(h, \mu_{\varepsilon\eta\delta}) + \big(w_2(0, \mu_{\varepsilon\eta\delta}) - w_2(h, \mu_{\varepsilon\eta\delta})\big) + w_2(h, \mu_{\varepsilon\eta\delta}) + \frac{1}{\varepsilon}\rho^2_{\varepsilon\eta\delta} - k\rho_{\varepsilon\eta\delta} - \eta T \\
&\geq - w_1(h, \mu_{\varepsilon\eta\delta}) + w_2(h, \mu_{\varepsilon\eta\delta}) + \frac{1}{\varepsilon}\rho^2_{\varepsilon\eta\delta} - k(h + \rho_{\varepsilon\eta\delta}) - \eta T.
\end{aligned}$$

Since we have assumed $A = 0$, we have $0 \leq -w_1(h, \mu_{\varepsilon\eta\delta}) + w_2(h, \nu_{\varepsilon\eta\delta})$, thus

$$-\frac{\xi}{2} \geq \varepsilon(k + \delta)\delta - \eta T - kh,$$

which, by letting $h \to 0^+$, leads again to a contradiction with the choice of $\varepsilon, \delta, \eta$. Thus $s_{\varepsilon\eta\delta} \neq 0$ and the proof showing $t_{\varepsilon\eta\delta} \neq 0$ can be done in the same way. $\diamond$

By Claim 2 and Claim 3, if we choose $\varepsilon, \delta, \eta > 0$ such that

$$\xi > 2\eta T - 2\varepsilon(k + \delta)^2, \qquad -2k\varepsilon(k + \delta)^2 \geq C\delta - \eta,$$

we have $s_{\varepsilon\eta\delta}, t_{\varepsilon\eta\delta} \notin [0, T]$, against the definition of ξ. Thus we have $\xi = 0$ and the proof is completed. $\square$

4. Hamilton–Jacobi–Bellman equation for the Mayer's problem

We will now characterize the value function of the Mayer's problem as the unique Lipschitz continuous viscosity solution of a suitable Hamilton–Jacobi–Bellman equation in the space of probability measures.

Definition 4.1 *(HJB equation for the Mayer's problem).* Given $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, $p_\mu \in L^2_\mu(\mathbb{R}^d; \mathbb{R}^d)$, we set

$$\mathscr{H}_F(\mu, p_\mu) := \inf \left\{ \int_{\mathbb{R}^d} \langle p_\mu(x), v_\mu(x)\rangle \, d\mu(x) : \begin{array}{l} v_\mu : \mathbb{R}^d \to \mathbb{R}^d \text{ Borel map} \\ v_\mu(x) \in F(x) \text{ for } \mu\text{-a.e. } x \in \mathbb{R}^d \end{array} \right\}.$$

Remark 4.2. By Theorem 8.2.12 in [5], we have that the map

$$x \mapsto h(x) := \inf_{v \in F(x)} \langle p_\mu(x), v\rangle$$

is Borel, thus for every Borel selection $v : \mathbb{R}^d \to \mathbb{R}^d$ of F we have:

$$\int_{\mathbb{R}^d} \langle v(x), p_\mu(x)\rangle \, d\mu(x) \geq \int_{\mathbb{R}^d} \inf_{v\in F(x)} \langle p_\mu(x), v\rangle \, d\mu(x),$$

and then by taking the infimum on $v(\cdot)$ we obtain

$$\mathscr{H}_F(\mu, p_\mu) \geq \int_{\mathbb{R}^d} \inf_{v\in F(x)} \langle p_\mu(x), v\rangle \, d\mu(x).$$

Thanks to assumptions (F), there exists $C > 0$ such that for all $\mu \in \mathscr{P}_2(\mathbb{R}^d)$

$$\int_{\mathbb{R}^d} \inf_{v\in F(x)} \langle p_\mu(x), v\rangle \, d\mu(x)$$
$$\geq -C \int |p_\mu(x)| \cdot (|x| + 1) \, d\mu(x) \geq -C \|p_\mu\|_{L^2_\mu} (\mathrm{m}_2^{1/2}(\mu) + 1) > -\infty.$$

For every $\varepsilon > 0$, define the Borel set-valued map $G_\varepsilon : \mathbb{R}^d \rightrightarrows \mathbb{R}$ by setting $G_\varepsilon(x) = [h(x) - \varepsilon, h(x) + \varepsilon]$. This map is Borel with closed images. Define the map $g : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ by $g(x, v) = \langle p_\mu(x), v\rangle$. The map g is Carathéodory, i.e., for every $v \in \mathbb{R}^d$ we have that $x \mapsto g(x, v)$ is Borel, and for every $x \in \mathbb{R}^d$ we have that $v \mapsto g(x, v)$ is continuous. Thus by Theorem 8.2.9 in [5] we have that for every $\varepsilon > 0$ there exists a measurable selection v_ε satisfying $\langle p_\mu(x), v_\varepsilon(x)\rangle \leq \inf_{v\in F(x)} \langle p_\mu(x), v\rangle + \varepsilon$, and so

$$\mathscr{H}(\mu, p_\mu) \leq \int_{\mathbb{R}^d} \langle v_\varepsilon(x), p_{\mu(x)} \, d\mu(x) \leq \varepsilon + \int_{\mathbb{R}^d} \inf_{v\in F(x)} \langle p_\mu(x), v\rangle \, d\mu(x) + \varepsilon.$$

By letting $\varepsilon \to 0$ we have the equality

$$\mathscr{H}(\mu, p_\mu) = \int_{\mathbb{R}^d} \inf_{v\in F(x)} \langle p_\mu(x), v\rangle \, d\mu(x).$$

Proposition 4.3 *(Smoothness of the Hamiltonian). Let $F : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ be satisfying* **(*F*)**. *Then the Hamiltonian $\mathscr{H}_F$ defined in* (4.1) *satisfies*

- *for all $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, $\lambda \geq 0$, $p_\mu \in L^2_\mu(\mathbb{R}^d)$ we have $\mathscr{H}_F(\mu, \lambda p_\mu) = \lambda \mathscr{H}_F(\mu, p_\mu)$;*
- *there exists $k \geq 0$ such that for all $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$, $\gamma \in \Pi_o(\mu, \nu)$, defined $p_\mu = \mathrm{Id}_{\mathbb{R}^d} - \mathrm{Bar}_1(\gamma)$, $q_\nu = \mathrm{Id}_{\mathbb{R}^d} - \mathrm{Bar}_1(\gamma^{-1})$, we have*

$$\mathscr{H}_F(\mu, p_\mu) - \mathscr{H}_F(\nu, q_\nu) \leq k W_2^2(\mu, \nu).$$

Proof. The first property is trivial. Let $\varepsilon > 0$ and $w_\nu^\varepsilon : \mathbb{R}^d \to \mathbb{R}^d$ be a Borel map belonging to the set of L^2_ν-selections of F, and such that

$$\mathscr{H}_F(\nu, q_\nu) + \varepsilon \geq \int_{\mathbb{R}^d} \langle w_\nu^\varepsilon(y), q_\nu(y)\rangle \, d\nu.$$

For all $v_\mu \in L^2_\mu$ such that $v_\mu(x) \in F(x)$ for μ-a.e. $x \in \mathbb{R}^d$ we have

$$\mathscr{H}_F(\mu, p_\mu) - \mathscr{H}_F(\nu, q_\nu) - \varepsilon \leq \int_{\mathbb{R}^d\times\mathbb{R}^d} \langle v_\mu(x) - w_\nu^\varepsilon(y), x-y\rangle \, d\gamma(x,y).$$

Recalling that F is Lipschitz continuous, we have that there exists $f : \mathbb{R}^d \to \mathbb{R}^d$ such that (4) holds true. By Filippov's Implicit Function Theorem (see e.g. Theorem 8.2.10 in [5]), there exists a Borel map $y \mapsto u_y^\varepsilon \in \overline{B(0,1)}$ satisfying $w_{\nu^\varepsilon}(y) = f(y, u_y^\varepsilon)$. Since $\mathscr{H}(\mu, p_\mu) = \int_{\mathbb{R}^d} \inf_{v\in F(x)} \langle p_\mu(x), v\rangle \, d\mu(x)$ by Remark 4.2, we have

$$\begin{aligned}\mathscr{H}_F(\mu, p_\mu) - \mathscr{H}_F(\nu, q_\nu) - \varepsilon &\leq \int_{\mathbb{R}^d} \inf_{v\in F(x)} \langle v, p_\mu(x)\rangle \, d\mu - \int_{\mathbb{R}^d} \langle w_\nu^\varepsilon(y), q_\nu(y)\rangle \, d\nu \\ &\leq \int_{\mathbb{R}^d} \langle f(x,u_y) - f(y,u_y), x-y\rangle \, d\gamma(x,y) \\ &\leq 5d\mathrm{Lip}(F) \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 \, d\gamma(x,y) = 5d\mathrm{Lip}(F) W_2^2(\mu,\nu),\end{aligned}$$

recalling the optimality of γ. We conclude by letting $\varepsilon \to 0^+$. □

Theorem 4.4 *(Characterization of the value function). Let $T > 0$, $F : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ be a Lipschitz continuous set-valued map with nonempty compact convex values, $\mathscr{G} : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ be a bounded and Lipschitz continuous map. Then for any $K \geq 0$, the value function $V(\cdot)$ is the unique Lipschitz continuous solution of the equation*

$$\begin{cases} \partial_t w(t,\mu) + \mathscr{H}_F(\mu, Dw(t,\mu)) = 0, \\ w(T,\mu) = \mathscr{G}(\mu), \end{cases} \tag{11}$$

stated on the set $\{(t,\mu) \in [0,T] \times \mathscr{K},\ \mathrm{m}_2(\mu) \leq K\}$.

Proof. Recalling Proposition 4.3 and Theorem 3.4, it is enough to show that $V(\cdot)$ is a viscosity solution of (11).

Claim 1: V is a subsolution of (11).

Proof (of Claim 1). Take $(\bar{t}, \bar{\mu}) \in]0, T[\times \mathscr{P}_2(\mathbb{R}^d)$, $\delta > 0$, $(p_t, p_\mu) \in D_\delta^+ V(\bar{t}, \bar{\mu})$. Let $v_{\bar{t}} : \mathbb{R}^d \to \mathbb{R}^d$ be a Borel map such that $v_{\bar{t}}(x) \in F(x)$ for μ-a.e. $x \in \mathbb{R}^d$ By Proposition 2.5, it is possible to find an admissible curve $\boldsymbol{\mu} = \{\mu_t\}_{t\in[\bar{t},T]} \in \mathscr{A}^F_{[\bar{t},T]}(\bar{\mu})$ and $\boldsymbol{\eta} \in \mathscr{P}(\mathbb{R}^d \times \Gamma_{[\bar{t},T]})$ such that $\mu_t = e_t \sharp \boldsymbol{\eta}$ for all $t \in [\bar{t}, T]$ and

$$\lim_{t\to\bar t}\int_{\mathbb{R}^d\times\Gamma_{[\bar t,T]}}\langle p_\mu\circ e_{\bar t}(x,\gamma),\frac{e_t(x,\gamma)-e_{\bar t}(x,\gamma)}{t-\bar t}\rangle\,d\boldsymbol{\eta}(x,\gamma)=\int_{\mathbb{R}^d}\langle p_\mu(x),v_{\bar t}(x)\rangle\,d\bar\mu(x).$$

According to the Dynamic Programming Principle in Proposition 2.7, we have $V(t,\mu_t)-V(\bar t,\bar\mu)\le 0$, moreover, if we define $\tilde\mu=(e_{\bar t},p_\mu\circ e_{\bar t},e_t)\sharp\boldsymbol{\eta}$, we have $\pi_{12}\sharp\tilde\mu=(\mathrm{Id}_{\mathbb{R}^d},p_\mu)\sharp\bar\mu$ and $\pi_{13}\sharp\tilde\mu=(e_{\bar t},e_t)\sharp\boldsymbol{\eta}\in\Pi(\bar\mu,\mu_t)$. Moreover, we have $W_{2,\tilde\mu}(\bar\mu,\mu_t)=\|e_t-e_{\bar t}\|_{L^2_{\boldsymbol{\eta}}}$, which tends to 0 as $t\to\bar t^+$ due to the continuity of $t\mapsto e_t$ (see Proposition 2.3). By applying the definition of viscosity superdifferential with $\tilde\mu=(e_{\bar t},p_\mu\circ e_{\bar t},e_t)\sharp\boldsymbol{\eta}$, we have that

$$\begin{aligned}
0&\le V(t,\mu_t)-V(\bar t,\bar\mu)\\
&\le p_t(t-\bar t)+\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d}\langle x_2,x_3-x_1\rangle\,d\tilde\mu(x_1,x_2,x_3)+\\
&\quad+\delta\sqrt{(t-\bar t)^2+W^2_{2,\tilde\mu}(\bar\mu,\mu)}+o(|t-\bar t|+W_{2,\tilde\mu}(\bar\mu,\mu_t))\\
&=p_t(t-t_0)+\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d}\langle p_\mu\circ e_{\bar t}(x,\gamma),e_t(x,\gamma)-e_{\bar t}(x,\gamma)\rangle\,d\boldsymbol{\eta}(x,\gamma)+\\
&\quad+\delta\sqrt{(t-\bar t)^2+W^2_{2,\tilde\mu}(\bar\mu,\mu)}+o(|t-\bar t|+W_{2,\tilde\mu}(\bar\mu,\mu_t)).
\end{aligned}$$

Dividing by $\sqrt{(t-\bar t)^2+W^2_{2,\tilde\mu}(\bar\mu,\mu)}$ and letting $t\to\bar t^+$ yields

$$\begin{aligned}
-\delta&\le\lim_{t\to\bar t^+}p_t\cdot\frac{t-\bar t}{\sqrt{(t-\bar t)^2+W^2_{2,\tilde\mu}(\bar\mu,\mu)}}+\\
&\quad+\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d}\langle p_\mu\circ e_{\bar t}(x,\gamma),\frac{e_t(x,\gamma)-e_{\bar t}(x,\gamma)}{t-\bar t+W_{2,\tilde\mu}(\bar\mu,\mu_t)}\rangle\,d\boldsymbol{\eta}(x,\gamma)\\
&\le p_t+\lim_{t\to\bar t^+}\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d}\langle p_\mu\circ e_{\bar t}(x,\gamma),\frac{e_t(x,\gamma)-e_{\bar t}(x,\gamma)}{t-\bar t}\rangle\,d\boldsymbol{\eta}(x,\gamma)\\
&=p_t+\int_{\mathbb{R}^d}\langle p_\mu(x),v_{\bar t}(x)\rangle\,d\bar\mu(x).
\end{aligned}$$

By the arbitrariness of $v_{\bar t}$ among the $L^2_{\bar\mu}$-selections of F, taking the infimum on $v_{\bar t}$ we have

$$p_t+\mathscr{H}(\bar\mu,p_\mu)\ge-\delta,$$

which ends the proof of Claim 1. $\diamond$

Claim 2: V is a supersolution of (11).

Proof (of Claim 2). Take $(\bar t,\bar\mu)\in]0,T[\times\mathscr{P}_2(\mathbb{R}^d)$, $\delta>0$, $(p_t,p_\mu)\in D^-_\delta V(\bar t,\bar\mu)$. By Proposition 2.3, there exists it is possible to find an optimal curve $\boldsymbol{\mu}=\{\mu_t\}_{t\in[\bar t,T]}\in\mathscr{A}^F_{[\bar t,T]}(\bar\mu)$ and

$\boldsymbol{\eta} \in \mathscr{P}(\mathbb{R}^d \times \Gamma_{[\bar{t},T]})$ such that $\mu_t = e_t \sharp \boldsymbol{\eta}$ for all $t \in [\bar{t}, T]$ and $V(t, \mu_t) = V(\bar{t}, \bar{\mu})$ for all $t \in [\bar{t}, T]$. By choosing as before $\tilde{\mu} = (e_{\bar{t}}, p_\mu \circ e_{\bar{t}}, e_t)\sharp \boldsymbol{\eta}$, we have $W_{2,\tilde{\mu}}(\mu_{\bar{t}}, \mu_t) = \|e_{\bar{t}} - e_t\|_{L^2_{\boldsymbol{\eta}}}$, and we obtain

$$\begin{aligned} 0 &= V(t, \mu_t) - V(\bar{t}, \bar{\mu}) \\ &\geq p_t(t - \bar{t}) + \int_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} \langle p_\mu \circ e_{\bar{t}}(x, \gamma), e_t(x, \gamma) - e_{\bar{t}}(x, \gamma) \rangle \, d\boldsymbol{\eta}(x, \gamma) + \\ &\quad - \delta \sqrt{(t - \bar{t})^2 + W^2_{2,\tilde{\mu}}(\bar{\mu}, \mu)} + o(|t - \bar{t}| + W_{2,\tilde{\mu}}(\bar{\mu}, \mu_t)). \end{aligned}$$

Dividing by $\sqrt{(t - \bar{t})^2 + W^2_{2,\tilde{\mu}}(\bar{\mu}, \mu)}$ yields

$$\begin{aligned} 0 \geq & \frac{(t - \bar{t})}{\sqrt{(t - \bar{t})^2 + W^2_{2,\tilde{\mu}}(\bar{\mu}, \mu)}} \left[p_t + \int_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} \langle p_\mu \circ e_{\bar{t}}(x, \gamma), \frac{e_t(x, \gamma) - e_{\bar{t}}(x, \gamma)}{t - \bar{t}} \rangle \, d\boldsymbol{\eta}(x, \gamma) \right] + \\ & - \delta + \frac{o(|t - \bar{t}| + W_{2,\tilde{\mu}}(\bar{\mu}, \mu_t))}{\sqrt{(t - \bar{t})^2 + W^2_{2,\tilde{\mu}}(\bar{\mu}, \mu)}}. \end{aligned}$$

We conclude by applying Proposition 2.3 to take the limit along a sequence $t_i \to \bar{t}$ such that $\dfrac{e_t - e_{\bar{t}}}{t - \bar{t}}$ weakly converges to $v_{\bar{t}} \circ e_0$ in $L^2_{\boldsymbol{\eta}}$, for a suitable $L^2_{\bar{\mu}}$-selection $v_{\bar{t}}$ of F. Indeed, we have

$$K'\delta \geq p_t + \int_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} \langle p_\mu, v_{\bar{t}}(x) \rangle \, d\bar{\mu}(x) \geq p_t + \mathscr{H}(\bar{\mu}, p_\mu),$$

where $K' = 1 + Ce^{2(b-a)C}\left(1 + \sup_{\mu \in \mathscr{K}} \mathrm{m}_2^{1/2}(\mu_{\bar{t}})\right)$ and $C = \max_{y \in F(0)} |y| + \mathrm{Lip}(F)$. □

5. A pursuit-evasion game

In this section we apply the result obtained to the study of a pursuit-evasion game in Wasserstein space. Our goal will be to show that this game admits a value, proving that the upper and the lower values are sub- and supersolution of the same Hamilton–Jacobi equations. The comparison principle will be used to conclude the existence of a value for this game. For an introduction to differential games, we refer the reader to [7], and to [6] for a survey on the most recent developments.

5.1. Dynamics and strategies

We consider two set-valued map $F, G : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ satisfying $(\boldsymbol{F})$. Given $\mu_a \in \mathscr{P}_2(\mathbb{R}^d)$, the set of admissible trajectories starting from μ_a at time $t = a$ defined on $[a, b]$ for the first player will be $\mathscr{A}^F_{[a,b]}(\mu_a)$, and, similarly, given $\nu_a \in \mathscr{P}_2(\mathbb{R}^d)$, the set of admissible trajectories starting from ν_a at time $t = a$ defined on $[a, b]$ for the second player will be $\mathscr{A}^G_{[a,b]}(\nu_a)$.

Definition 5.1 *(Nonanticipative strategies).* A *strategy* for the first player defined on $[t_0, T]$ will be a map $\alpha : \mathscr{A}^G_{[t_0,T]} \to \mathscr{A}^F_{[t_0,T]}$. A strategy for the first player α defined on $[t_0, T]$ will be called *nonanticipative with delay* τ if there exists $\tau > 0$ such that given $t_0 \le s \le T$, $\boldsymbol{\nu}^i = \{\nu^i_t\}_{t\in[t_0,T]} \in \mathscr{A}^G_{[t_0,T]}$, $i = 1, 2$, satisfying $\nu^1_t = \nu^2_t$ for all $t \in]t_0, s[$, and set $\alpha(\boldsymbol{\nu}^i) = \{\mu^i_t\}_{t\in[t_0,T]}$, $i = 1, 2$, we have $\mu^1_t = \mu^2_t$ for all $t_0 \le t \le \min\{s + \tau, T\}$.

Given $\mu_0 \in \mathscr{P}_2(\mathbb{R}^d)$, we define

$$\mathcal{A}_\tau(t_0) := \left\{\alpha : \mathscr{A}^G_{[t_0,T]} \to \mathscr{A}^F_{[t_0,T]} : \alpha \text{ is a nonanticipative strategy with delay } \tau\right\},$$

$$\mathcal{A}_\tau(t_0, \mu_0) := \left\{\alpha \in \mathcal{A}_\tau(t_0) : \alpha(\mathscr{A}^G_{[t_0,T]}) \subseteq \mathscr{A}^F_{[t_0,T]}(\mu_0)\right\},$$

$$\mathcal{A}(t_0) := \bigcup_{\tau>0} \mathcal{A}_\tau(t_0),$$

$$\mathcal{A}(t_0, \mu_0) := \left\{\alpha \in \mathcal{A}(t_0) : \alpha(\mathscr{A}^G_{[t_0,T]}) \subseteq \mathscr{A}^F_{[t_0,T]}(\mu_0)\right\}.$$

By switching the roles of F and G in the previous definitions, we obtain the corresponding definition of strategy and nonanticipative strategy defined on $[t_0, T]$ with delay τ for the second player. The corresponding defined sets are named by $\mathcal{B}_\tau(t_0)$, $\mathcal{B}_\tau(t_0, \nu_0)$, $\mathcal{B}(t_0)$, $\mathcal{B}(t_0, \nu_0)$, respectively, for any given $\nu_0 \in \mathscr{P}_2(\mathbb{R}^d)$.

Lemma 5.2 *(Normal form). Let $t_0 < \tau < T$. For any $(\alpha, \beta) \in \mathcal{A}_\tau(t_0) \times \mathcal{B}_\tau(t_0)$ there is a unique pair $(\boldsymbol{\mu}, \boldsymbol{\nu}) \in \mathscr{A}^F_{[t_0,b]} \times \mathscr{A}^G_{[t_0,b]}$ such that $\alpha(\boldsymbol{\nu}) = \boldsymbol{\mu}$ and $\beta(\boldsymbol{\mu}) = \boldsymbol{\nu}$.*

Proof. The proof will follow the line of Lemma 1 in [12]. Let $(\alpha, \beta) \in (\mathcal{A}(t_0) \times \mathcal{B}(t_0))$. Clearly we have $\mathcal{A}_{\tau_1}(t_0) \subseteq \mathcal{A}_{\tau_2}(t_0)$ and $\mathcal{B}_{\tau_1}(t_0) \subseteq \mathcal{B}_{\tau_2}(t_0)$ if $\tau_1 \ge \tau_2$, thus, without loss of generality, we may assume that there exists $\tau > 0$ such that $(\alpha, \beta) \in (\mathcal{A}_\tau(t_0) \times \mathcal{B}_\tau(t_0))$. We consider a partition of the interval $[t_0, T]$ by defining $N_\tau = \min\{k \in \mathbb{N} : t_0 + k\tau < T\}$, and set $t_k = t_0 + k\tau$ for $k = 0, \dots, N_\tau$ and $t_{N_\tau+1} = T$. We will proceed by induction, defining $(\boldsymbol{\mu}, \boldsymbol{\nu})$ on $[t_0, t_k[$.

Recalling Definition 5.1, the restriction of $\alpha(\boldsymbol{\nu}')$ to $[t_0, t_1[$ does not depend on the particular choice of $\boldsymbol{\nu}' \in \mathscr{A}^G_{[t_0,T]}$: indeed, if we have $\boldsymbol{\nu_1}, \boldsymbol{\nu_2} \in \mathscr{A}^G_{[t_0,T]}$, by taking $s = t_0$, we have $\alpha(\boldsymbol{\nu_1}) = \alpha(\boldsymbol{\nu_2})$ in $[t_0, t_0 + \tau[$. We set then $\boldsymbol{\mu}$ to be equal to $\alpha(\boldsymbol{\nu}')$ in $[t_0, t_1[$ for any choice of $\boldsymbol{\nu}' \in \mathscr{A}_{[t_0,T]}$, and moreover $\boldsymbol{\nu} = \beta(\boldsymbol{\mu})$ is uniquely defined in $[t_0, t_1[$ since β is nonanticipative. Suppose to have defined $(\boldsymbol{\mu}, \boldsymbol{\nu})$ on $[t_0, t_k[$, where $0 \le k \le N_\tau$. For every $\boldsymbol{\nu} \in \mathscr{A}^G_{[t_0,T]}$, the restriction of $\alpha(\boldsymbol{\nu})$ to $[t_0, t_{k+1}[$ depends uniquely to the restriction of $\boldsymbol{\nu}$ on $[t_0, t_k[$, in particular $\boldsymbol{\mu} = \alpha(\boldsymbol{\nu})$ is uniquely defined on $[t_0, t_{k+1}[$, and so we can define $\boldsymbol{\nu}$ in $[t_k, t_{k+1}[$ by taking the restriction of $\beta(\boldsymbol{\mu})$ to $[t_k, t_{k+1}[$. By induction we conclude that $\boldsymbol{\mu}, \boldsymbol{\nu}$ are well-defined in $[0, T[$, and we conclude by noticing that indeed, $\alpha(\boldsymbol{\nu})$ at time T is fully determined by the restriction of $\boldsymbol{\nu}$ on $[t_0, T - \varepsilon]$ for all $0 \le \varepsilon \le \min\{\tau, T - t_0\}$, thus $\boldsymbol{\mu} = \alpha(\boldsymbol{\nu})$ is uniquely determined also at $t = T$, and the same for $\boldsymbol{\nu} = \alpha(\boldsymbol{\mu})$. □

5.2. Value functions and Dynamic Programming Principle

Definition 5.3 *(Upper and lower value functions).* We consider a *payoff function* $\mathcal{G} : \mathscr{P}(\mathbb{R}^d) \times \mathscr{P}(\mathbb{R}^d) \to \mathbb{R}$ bounded and locally Lipschitz continuous, and we assume that F and G satisfy $(\boldsymbol{F})$. Given $t_0 \in [0, T]$, $\mu_0, \nu_0 \in \mathscr{P}_2(\mathbb{R}^d)$, $(\alpha, \beta) \in \mathcal{A}(\mu_0, t_0) \times \mathcal{B}(\nu_0, t_0)$ we define

Please cite this article in press as: A. Marigonda, M. Quincampoix, Mayer control problem with probabilistic uncertainty on initial positions, J. Differential Equations (2017), https://doi.org/10.1016/j.jde.2017.11.014

$$J(t_0,\mu_0,\nu_0,\alpha,\beta)=\mathcal{G}(\mu_T,\nu_T),$$

where $\boldsymbol{\mu}=\{\mu_t\}_{t\in[0,T]}\in\mathscr{A}^F_{[t_0,T]}(\mu_0)$, $\boldsymbol{\nu}=\{\nu_t\}_{t\in[0,T]}\in\mathscr{A}^G_{[t_0,T]}(\nu_0)$, and $(\boldsymbol{\mu},\boldsymbol{\nu})\in\mathscr{A}^F_{[t_0,T]}(\mu_0)\times\mathscr{A}^G_{[t_0,T]}(\nu_0)$ is the unique element of $\mathscr{A}^F_{[t_0,T]}(\mu_0)\times\mathscr{A}^G_{[t_0,T]}(\nu_0)$, given by Lemma 5.2, satisfying $\alpha(\boldsymbol{\nu})=\boldsymbol{\mu}$ and $\beta(\boldsymbol{\nu})=\boldsymbol{\mu}$.

The *upper* and *lower* value function $V^\pm:[0,T]\times\mathscr{P}_2(\mathbb{R}^d)\times\mathscr{P}_2(\mathbb{R}^d)\to\mathbb{R}$ are defined by setting

$$V^+(t_0,\mu_0,\nu_0)=\inf_{\alpha\in\mathcal{A}(t_0,\mu_0)}\sup_{\beta\in\mathcal{B}(t_0,\nu_0)}J(t_0,\mu_0,\nu_0,\alpha,\beta),$$

$$V^-(t_0,\mu_0,\nu_0)=\sup_{\beta\in\mathcal{B}(t_0,\nu_0)}\inf_{\alpha\in\mathcal{A}(t_0,\mu_0)}J(t_0,\mu_0,\nu_0,\alpha,\beta).$$

Remark 5.4. For the pursuit-evasion game, a relevant example of payoff in Definition 5.3 is given by $\mathcal{G}(\mu,\nu)=g(W_2^2(\mu,\nu))$, where $g:[0,+\infty[\,\to[0,+\infty[$ is strictly increasing, bounded, Lipschitz continuous and $g(0)=0$ (e.g. $g(r)=\arctan(r)$).

Definition 5.5 *(Shifting strategies).* Let $T>0$, $t_0\in[0,T]$, $\bar\mu\in\mathscr{P}_2(\mathbb{R}^d)$. A map $\xi^{F,\bar\mu}_{t_0}:\mathscr{A}^F_{[t_0,T]}\to\mathscr{A}^F_{[t_0,T]}(\bar\mu)$ will be called a *shifting strategy* in $[t_0,T]$ for F if there exists $K>0$ such that given $\boldsymbol{\mu}^{(i)}=\{\mu^{(i)}_t\}_{t\in[0,T]}\in\mathscr{A}^F_{[t_0,T]}$, $i=1,2$, and set $\boldsymbol{\mu}^{(3)}=\{\mu^{(3)}_t\}_{t\in[0,T]}=\xi^{F,\mu}_{t_0}(\boldsymbol{\mu}^{(1)})$, $\boldsymbol{\mu}^{(4)}=\{\mu^{(4)}_t\}_{t\in[0,T]}=\xi^{F,\bar\mu}_{t_0}(\boldsymbol{\mu}^{(2)})$, the following hold

i.) $W_2(\mu^{(1)}_t,\mu^{(3)}_t)\le K\,W_2(\mu^{(1)}_{t_0},\mu^{(3)}_{t_0})$ for all $t\in[t_0,T]$;
ii.) if there exists $t_0<s<T$ such that $\mu^{(2)}_t=\mu^{(1)}_t$ for all $t\in[t_0,s]$ then $\mu^{(4)}_t=\mu^{(3)}_t$ for all $t\in[t_0,s]$.

The same definition with F replaced by G will give the definition of shifting strategy for G. We notice that, from the definition, we have $\mu^{(3)}_{t_0}=\bar\mu$; moreover, given any strategy $\alpha\in\mathcal{A}_\tau(t_0)$, the composition $\xi^{F,\bar\mu}_{t_0}\circ\alpha:\mathscr{A}^G_{[t_0,T]}\to\mathscr{A}^F_{[t_0,T]}(\bar\mu)$ is a nonanticipative strategy with delay τ, thus $\xi^{F,\bar\mu}_{t_0}\circ\alpha\in\mathcal{A}_\tau(t_0,\bar\mu)$.

Lemma 5.6 *(Existence and properties of shifting strategies). Assume that F satisfies* **(*F*)**. *Let $T>0$, $t_0\in[0,T]$, $\bar\mu\in\mathscr{P}_2(\mathbb{R}^d)$. Then there exists at least one shifting strategy $\xi^{F,\bar\mu}_{t_0}$ in $[t_0,T]$ for F.*

Proof. We will consider the construction made in Proposition 2.4: given $\boldsymbol{\mu}\in\mathscr{A}_{[t_0,T]}$, let $\xi^{F,\bar\mu}_{t_0}(\boldsymbol{\mu})=\bar{\boldsymbol{\mu}}\in\mathscr{A}_{[t_0,T]}(\bar\mu)$ constructed as in Proposition 2.4. Property (i) in Definition 5.5 is satisfied, we prove now (ii). From the proof of Proposition 2.4, we have that if $\gamma_1,\gamma_2\in\Gamma_{[t_0,T]}$ are trajectories of the differential inclusion $\dot\gamma(t)\in F(\gamma(t))$ with $\gamma_1(t)=\gamma_2(t)$ for all $t\in[t_0,s]$, then $\tau(y,\gamma_1)(t)=\tau(y,\gamma_2)(t)$ for all $t\in[t_0,s]$, in particular, the restriction of the curve $\tau(y,\gamma)$ on $[t_0,s]$ depends only on the values of γ on $[t_0,s]$. Let $\boldsymbol{\mu}^1=\{\mu^1_t\}_{t\in[t_0,T]}$, $\boldsymbol{\mu}^2=\{\mu^2_t\}_{t\in[t_0,T]}\in\mathscr{A}_{[t_0,T]}(\mu)$, $t_0<s\le T$ such that $\mu^1_t=\mu^2_t$ for $t\in[t_0,s]$. In particular, we have that

$$\partial_t\mu^i_t+\operatorname{div}(v^i_t\mu^i_t)=0,\ \text{ for } i=1,2,\ \text{ with } v^t_1=v^t_2 \text{ in } [t_0,s].$$

Please cite this article in press as: A. Marigonda, M. Quincampoix, Mayer control problem with probabilistic uncertainty on initial positions, J. Differential Equations (2017), https://doi.org/10.1016/j.jde.2017.11.014

Recalling the proof of Superposition Principle Theorem 8.2.1 in [2], we take a family of strictly positive convolution kernels $\{\rho_\varepsilon\}_{\varepsilon>0}$, define $v_t^{i,\varepsilon} = \dfrac{(v_t^i \mu_t^i) * \rho_\varepsilon}{\mu_t^i * \rho_\varepsilon}$, and so $v_t^{1,\varepsilon} = v_t^{2,\varepsilon}$ for $t \in [t_0, s]$. Denote by $X_t^{i,\varepsilon}(x)$ the unique solution of $\dfrac{d}{dt} X_t^{i,\varepsilon}(x) = v_t^{i,\varepsilon}(X_t^{i,\varepsilon}(x))$ such that $X_{t_0}^{i,\varepsilon}(x) = x$. We notice that $X_t^{1,\varepsilon}(x) = X_t^{2,\varepsilon}(x)$ for $(t,x) \in [t_0, s] \times \mathbb{R}^d$. We consider the map $X_\cdot^{i,\varepsilon} : \mathbb{R}^d \to \mathbb{R}^d \times \Gamma_{[t_0,T]}$ defined by

$$X_\cdot^{i,\varepsilon}(x) = (x, \gamma), \text{ where } \gamma(t) = X_t^{i,\varepsilon}(x),$$

and set $\boldsymbol{\eta}^{i,\varepsilon} = X_\cdot^{i,\varepsilon} \sharp \mu_0^i * \rho_\varepsilon$. By taking any limit point for $\varepsilon = 0$, if we denote by $R_{[t_0,s]}$ the restriction operator on curves, we have $(\mathrm{Id}_{\mathbb{R}^d}, R_{[t_0,s]})\sharp \boldsymbol{\eta}_1 = (\mathrm{Id}_{\mathbb{R}^d}, R_{[t_0,s]})\sharp \boldsymbol{\eta}_2$, thus the construction of 2.4 yields $\xi_{t_0}^{F,\bar{\mu}}(\boldsymbol{\mu}^1) = \xi_{t_0}^{F,\bar{\mu}}(\boldsymbol{\mu}^2)$ on $[t_0, s]$. □

Lemma 5.7 *(Regularity of upper and lower values). We have $V^-(t_0, \mu_0, \nu_0) \le V^+(t_0, \mu_0, \nu_0)$ for all $(t_0, \mu_0, \nu_0) \in [0, T] \times \mathscr{P}_2(\mathbb{R}^d) \times \mathscr{P}_2(\mathbb{R}^d)$. Moreover, the functions $V^\pm(\cdot)$ are bounded and locally Lipschitz continuous.*

Proof. The first assertion follows directly from the definition of $V^\pm(\cdot)$. To simplify the notation, given $\boldsymbol{\mu} = \{\mu_t\}_{t\in[t_0,T]}$ and $\boldsymbol{\nu} = \{\nu_t\}_{t\in[t_0,T]}$, we will write $\mathcal{G}_T(\boldsymbol{\mu}, \boldsymbol{\nu})$ instead of $\mathcal{G}(\mu_T, \nu_T)$.

We will prove the second statement only for V^+, being the corresponding proof for V^- completely similar. Due to Lemma 5.2, we have

$$V^+(t_0, \mu^i, \nu^i) = \inf_{\alpha \in \mathcal{A}(t_0,\mu^i)} \sup_{\boldsymbol{\nu}^i \in \mathscr{A}_{[t_0,T]}(\nu^i)} \mathcal{G}_T(\alpha(\boldsymbol{\nu}^i), \boldsymbol{\nu}^i), \; i = 0, 1.$$

We prove first the Lipschitz continuity with respect to μ_0. Fix $\varepsilon > 0$, $t_0 \in [0, T]$, $\mu^i, \nu^i \in \mathscr{P}_2(\mathbb{R}^d)$, $i = 0, 1$. There exist $\alpha^{1,\varepsilon} \in \mathcal{A}_\tau(t_0, \mu^1)$ such that

$$\sup_{\boldsymbol{\nu} \in \mathscr{A}_{[t_0,T]}(\nu^1)} \mathcal{G}_T(\alpha^{1,\varepsilon}(\boldsymbol{\nu}^1), \boldsymbol{\nu}^1) \le V^+(t_0, \mu^1, \nu^1) + \varepsilon.$$

We take two shifting strategy $\xi_{t_0}^{F,\mu^0}$, $\xi_{t_0}^{G,\nu^1}$, and define

$$\alpha^{0,\varepsilon} = \xi_{t_0}^{F,\mu^0} \circ \alpha^{1,\varepsilon} \circ \xi_{t_0}^{G,\nu^1} : \mathscr{A}^G_{[t_0,T]}(\nu^0) \to \mathscr{A}^F_{[t_0,T]}(\mu^0).$$

Thus we have

$$\begin{aligned} &V^+(t_0, \mu^0, \nu^0) - V^+(t_0, \mu^1, \nu^1) \le \\ &\le \varepsilon + \sup_{\boldsymbol{\nu}^0 \in \mathscr{A}_{[t_0,T]}(\nu^0)} \mathcal{G}_T(\alpha^{0,\varepsilon}(\boldsymbol{\nu}^0), \boldsymbol{\nu}^0) - \sup_{\boldsymbol{\nu}^1 \in \mathscr{A}_{[t_0,T]}(\nu^1)} \mathcal{G}_T(\alpha^{1,\varepsilon}(\boldsymbol{\nu}^1), \boldsymbol{\nu}^1). \end{aligned}$$

Choose now $\boldsymbol{\nu}^{0,\varepsilon} = \{\nu_t^{0,\varepsilon}\}_{t\in[t_0,T]} \in \mathscr{A}_{[t_0,T]}(\nu^0)$ be such that

$$\sup_{\boldsymbol{\nu}^0 \in \mathscr{A}_{[t_0,T]}(\nu^0)} \mathcal{G}_T(\alpha^{0,\varepsilon}(\boldsymbol{\nu}^0), \boldsymbol{\nu}^0) \le \varepsilon + \mathcal{G}_T(\alpha^{0,\varepsilon}(\boldsymbol{\nu}^{0,\varepsilon})), \boldsymbol{\nu}^{0,\varepsilon}).$$

By choosing $\boldsymbol{\nu}^1 = \xi_{t_0}^{G,\nu^1}(\boldsymbol{\nu}^{0,\varepsilon})$, and recalling the definition of $\alpha^{0,\varepsilon}$, we have

$$\begin{aligned} V^+(t_0,\mu^0,\nu^0) - V^+(t_0,\mu^1,\nu^1) - 2\varepsilon \le \\ &\le \mathcal{G}_T(\alpha^{0,\varepsilon}(\boldsymbol{\nu}^{0,\varepsilon}),\boldsymbol{\nu}^{0,\varepsilon}) - \mathcal{G}_T(\alpha^{1,\varepsilon}(\xi_{t_0}^{G,\nu^1}(\boldsymbol{\nu}^{0,\varepsilon})),\xi_{t_0}^{G,\nu^1}(\boldsymbol{\nu}^{0,\varepsilon})) \\ &= \mathcal{G}_T(\xi_{t_0}^{F,\mu_0}\circ\alpha^{1,\varepsilon}(\xi_{t_0}^{G,\nu^1}(\boldsymbol{\nu}^{0,\varepsilon})),\boldsymbol{\nu}^{0,\varepsilon}) - \mathcal{G}_T(\alpha^{1,\varepsilon}(\xi_{t_0}^{G,\nu^1}(\boldsymbol{\nu}^{0,\varepsilon})),\xi_{t_0}^{G,\nu^1}(\boldsymbol{\nu}^{0,\varepsilon})). \end{aligned}$$

Set now $\xi_{t_0}^{G,\nu^1}(\boldsymbol{\nu}^{0,\varepsilon}) = \{\nu_t^{1,\varepsilon}\}_{t\in[t_0,T]}$,

$$\boldsymbol{\mu}^{1,\varepsilon} = \{\mu_t^{1,\varepsilon}\}_{t\in[t_0,T]} = \alpha^{1,\varepsilon}(\xi_{t_0}^{G,\nu^1}(\boldsymbol{\nu}^{0,\varepsilon})), \qquad \xi_{t_0}^{F,\mu_0}(\boldsymbol{\mu}^{1,\varepsilon}) = \{\mu_t^{0,\varepsilon}\}_{t\in[t_0,T]}.$$

We have

$$\begin{aligned} V^+(t_0,\mu^0,\nu^0) - V^+(t_0,\mu^1,\nu^1) - 2\varepsilon \le \\ &\le \mathcal{G}_T(\xi_{t_0}^{F,\mu_0}(\boldsymbol{\mu}^{1,\varepsilon}),\boldsymbol{\nu}^{0,\varepsilon}) - \mathcal{G}_T(\boldsymbol{\mu}^{1,\varepsilon},\xi_{t_0}^{G,\nu^1}(\boldsymbol{\nu}^{0,\varepsilon})) \\ &\le \mathrm{Lip}(\mathcal{G})\cdot\left[W_2(\mu_T^{0,\varepsilon},\mu_T^{1,\varepsilon}) + W_2(\nu_T^{0,\varepsilon},\nu_T^{0,\varepsilon})\right] \\ &\le \mathrm{Lip}(\mathcal{G})\cdot K\cdot\left[W_2(\mu^0,\mu^1) + W_2(\nu^0,\nu^1)\right], \end{aligned}$$

recalling the properties of shifting strategies. By letting $\varepsilon \to 0^+$, and switching the roles of μ_0, ν_0 and μ_1, ν_1, this proves the Lipschitz continuity w.r.t. second and third variables.

We prove now the Lipschitz continuity with respect to the first variable. Fix $\varepsilon > 0$, $\mu,\nu \in \mathscr{P}_2(\mathbb{R}^d)$, $t_0, t_1 \in [0,T]$, $t_0 > t_1$, $\bar{\boldsymbol{\mu}} = \{\bar{\mu}_t\}_{t\in[t_0,T]} \in \mathscr{A}_{[t_0,T]}^F(\mu)$. There exist $\alpha^{1,\varepsilon} \in \mathcal{A}_\tau(t_1,\mu)$ such that

$$\sup_{\boldsymbol{\nu}^1\in\mathscr{A}_{[t_1,T]}(\nu^1)} \mathcal{G}_T(\alpha^{1,\varepsilon}(\boldsymbol{\nu}^1),\boldsymbol{\nu}^1) \le V^+(t_0,\mu,\nu) + \varepsilon.$$

Define a nonanticipative strategy $\alpha^{0,\varepsilon}$ by setting for all $\boldsymbol{\nu}^0 \in \mathscr{A}_{[t_0,T]}^G(\nu_0)$

$$\alpha^{0,\varepsilon}(\boldsymbol{\nu}) = \begin{cases} \bar{\boldsymbol{\mu}}, \text{ on } [t_0,t_1], \\ \xi_{\tau,t_1}^{F,\bar{\mu}_{t_1}}\circ\alpha^{1,\varepsilon}\circ\xi_{\tau,t_1}^{G,\nu}(\boldsymbol{\nu}^0_{|[t_1,T]}), \text{ on } [t_1,T], \end{cases}$$

where $\boldsymbol{\nu}_{|[t_1,T]}$ denotes the restriction of $\boldsymbol{\nu}$ to $[t_1,T]$. This implies

$$V(t_0,\mu,\nu) - V(t_1,\mu,\nu) \le \varepsilon + \sup_{\boldsymbol{\nu}^0\in\mathscr{A}_{[t_0,T]}(\nu)} \mathcal{G}_T(\alpha^{0,\varepsilon}(\boldsymbol{\nu}^0),\boldsymbol{\nu}^0) + \sup_{\boldsymbol{\nu}^1\in\mathscr{A}_{[t_1,T]}^G(\nu)} \mathcal{G}_T(\alpha^{1,\varepsilon}(\boldsymbol{\nu}^1),\boldsymbol{\nu}^1).$$

Select $\nu^{0,\varepsilon} \in \mathscr{A}_{[t_1,T]}^G(\nu)$ such that

$$\sup_{\boldsymbol{\nu}^0\in\mathscr{A}_{[t_0,T]}(\nu)} \mathcal{G}_T(\alpha^{0,\varepsilon}(\boldsymbol{\nu}^0),\boldsymbol{\nu}^0) \le \mathcal{G}_T(\alpha^{0,\varepsilon}(\boldsymbol{\nu}^{0,\varepsilon}),\boldsymbol{\nu}^{0,\varepsilon}) + \varepsilon.$$

So we have

$$\begin{aligned}&V(t_0,\mu,\nu)-V(t_1,\mu,\nu)-2\varepsilon\leq\\ &\leq \mathcal{G}_T(\alpha^{0,\varepsilon}(\boldsymbol{\nu}^{0,\varepsilon}),\boldsymbol{\nu}^{0,\varepsilon})-\mathcal{G}_T(\alpha^{1,\varepsilon}(\xi^{G,\nu}_{\tau,t_1}(\boldsymbol{\nu}^{0,\varepsilon}_{|[t_1,T]}))),\xi^{G,\nu}_{\tau,t_1}(\boldsymbol{\nu}^{0,\varepsilon}_{|[t_1,T]}))\\ &=\mathcal{G}_T(\xi^{F,\bar{\mu}_{t_1}}_{\tau,t_1}\circ\alpha^{1,\varepsilon}\circ\xi^{G,\nu}_{\tau,t_1}(\boldsymbol{\nu}^{0,\varepsilon}),\boldsymbol{\nu}^{0,\varepsilon})-\mathcal{G}_T(\alpha^{1,\varepsilon}(\xi^{G,\nu}_{\tau,t_1}(\boldsymbol{\nu}^{0,\varepsilon}_{|[t_1,T]}))),\xi^{G,\nu}_{\tau,t_1}(\boldsymbol{\nu}^{0,\varepsilon}_{|[t_1,T]})).\end{aligned}$$

Set $\boldsymbol{\nu}^{1,\varepsilon}=\{\nu^{1,\varepsilon}_t\}_{t\in[t_1,T]}=\xi^{G,\nu}_{\tau,t_1}(\boldsymbol{\nu}^{0,\varepsilon}_{|[t_1,T]})$, $\boldsymbol{\mu}^{1,\varepsilon}=\{\mu^{1,\varepsilon}_t\}_{t\in[t_1,T]}=\alpha^{1,\varepsilon}(\xi^{G,\nu}_{\tau,t_1}(\boldsymbol{\nu}^{0,\varepsilon}_{|[t_1,T]})))$, $\boldsymbol{\mu}^{0,\varepsilon}=\{\mu^{0,\varepsilon}_t\}_{t\in[t_0,T]}=\xi^{F,\bar{\mu}_{t_1}}_{\tau,t_1}(\boldsymbol{\mu}^{1,\varepsilon})$, hence,

$$\begin{aligned}V(t_0,\mu,\nu)-V(t_1,\mu,\nu)-2\varepsilon&\leq\mathcal{G}(\mu^{0,\varepsilon}_T,\nu^{0,\varepsilon}_T)-\mathcal{G}_T(\mu^{1,\varepsilon}_T,\nu^{1,\varepsilon}_T)\\ &\leq \mathrm{Lip}(\mathcal{G})\left[W_2(\mu^{0,\varepsilon}_T,\mu^{1,\varepsilon}_T)+W_2(\nu^{0,\varepsilon}_T,\nu^{1,\varepsilon}_T)\right].\end{aligned}$$

We notice that the endpoints of $\boldsymbol{\nu}^{0,\varepsilon}$ and of $\boldsymbol{\nu}^{0,\varepsilon}_{|[t_1,T]}$ are the same, and so, recalling the properties of the shifting strategies,

$$W_2(\nu^{0,\varepsilon}_T,\nu^{1,\varepsilon}_T)\leq K\,W_2(\nu^{0,\varepsilon}_{t_1},\nu^{0,\varepsilon}_{t_0}).$$

With a similar argument, we have

$$W_2(\mu^{0,\varepsilon}_T,\mu^{1,\varepsilon}_T)\leq K\,W_2(\bar{\mu}_{t_1},\mu).$$

By using (4) in Proposition 2.3, we have

$$V(t_0,\mu,\nu)-V(t_1,\mu,\nu)-2\varepsilon\leq Ce^{2TC}\,(2+\mathrm{m}_2(\mu)+\mathrm{m}_2(\nu))\,|t_0-t_1|.$$

We conclude by letting $\varepsilon\to 0^+$. The proof for the case $t_1\leq t_0$ is similar. □

Proposition 5.8 *(Dynamic Programming Principle for the game). Let $t_0,t_1\in[0,T]$ with $t_0<t_1$, $\mu^0,\nu^0\in\mathscr{P}_2(\mathbb{R}^d)$. Then*

$$V^+(t_0,\mu^0,\nu^0)=\inf_{\alpha\in\mathcal{A}(t_0,\mu^0)}\sup_{\beta\in\mathcal{B}(t_0,\nu^0)}\left\{V^+(t_1,\mu_{t_1},\nu_{t_1}):\begin{array}{c}\boldsymbol{\mu}=\{\mu_t\}_{t\in[t_0,T]}=\alpha(\boldsymbol{\nu})\\ \boldsymbol{\nu}=\{\nu_t\}_{t\in[t_0,T]}=\beta(\boldsymbol{\mu})\end{array}\right\},$$

$$V^-(t_0,\mu^0,\nu^0)=\sup_{\beta\in\mathcal{B}(t_0,\nu^0)}\inf_{\alpha\in\mathcal{A}(t_0,\mu^0)}\left\{V^-(t_1,\mu_{t_1},\nu_{t_1}):\begin{array}{c}\boldsymbol{\mu}=\{\mu_t\}_{t\in[t_0,T]}=\alpha(\boldsymbol{\nu})\\ \boldsymbol{\nu}=\{\nu_t\}_{t\in[t_0,T]}=\beta(\boldsymbol{\mu})\end{array}\right\}.$$

Proof. To simplify the notation, given $\boldsymbol{\mu}=\{\mu_t\}_{t\in[t_0,T]}$ and $\boldsymbol{\nu}=\{\nu_t\}_{t\in[t_0,T]}$, we will write $\mathcal{G}_T(\boldsymbol{\mu},\boldsymbol{\nu})$ instead of $\mathcal{G}(\mu_T,\nu_T)$.

We will prove the result only for V^-, since the proof for V^+ can be performed using a very similar argument. Recalling that, thanks to Lemma 5.2, we have

$$V^-(t_0,\mu^0,\nu^0)=\sup_{\beta\in\mathcal{B}(t_0,\nu^0)}\inf_{\boldsymbol{\mu}\in\mathscr{A}^F_{[t_0,T]}(\mu^0)}\mathcal{G}_T(\boldsymbol{\mu},\beta(\boldsymbol{\nu})),$$

we define

$$W(t_0,t_1,\mu^0,\nu^0):=\sup_{\beta\in\mathcal{B}(t_0,\nu^0)}\inf_{\boldsymbol{\mu}\in\mathscr{A}^F_{[t_0,T]}(\mu^0)}\left\{V^-(t_1,\mu_{t_1},\nu_{t_1}):\begin{array}{l}\boldsymbol{\mu}=\{\mu_t\}_{t\in[t_0,T]}\\ \boldsymbol{\nu}=\{\nu_t\}_{t\in[t_0,T]}=\beta(\boldsymbol{\mu})\end{array}\right\},$$

and we want to prove $V^-(t_0,\mu^0,\nu^0)=W(t_0,t_1,\mu^0,\nu^0)$.

Indeed, we have

$$\begin{aligned}W(t_0,t_1,\mu^0,\nu^0)&=\sup_{\beta_0\in\mathcal{B}(t_0,\nu^0)}\inf_{\boldsymbol{\mu}^0\in\mathscr{A}^F_{[t_0,T]}(\mu^0)}\sup_{\substack{\beta_1\in\mathcal{B}(t_0,\nu^0_{t_1})\\ \beta_0(\boldsymbol{\mu}^0)=\boldsymbol{\nu}^0}}\inf_{\boldsymbol{\mu}^1\in\mathscr{A}^F_{[t_1,T]}(\mu^0_{t_1})}\mathcal{G}_T(\boldsymbol{\mu}^1,\beta_1(\boldsymbol{\mu}^1))\\ &\le\sup_{\beta_0\in\mathcal{B}(t_0,\nu^0)}\inf_{\boldsymbol{\mu}^0\in\mathscr{A}^F_{[t_0,T]}(\mu^0)}\sup_{\substack{\beta_1\in\mathcal{B}(t_0,\nu^0_{t_1})\\ \beta_0(\boldsymbol{\mu}^0)=\boldsymbol{\nu}^0}}\mathcal{G}_T(\boldsymbol{\mu}^0_{|[t_1,T]},\beta_1(\boldsymbol{\mu}^0_{|[t_1,T]}))\end{aligned}$$

Fix $\beta_0\in\mathcal{B}(t_0,\nu^0)$, $\boldsymbol{\mu}^0\in\mathscr{A}_{[t_0,T]}(\mu^0)$, set $\boldsymbol{\nu}^0=\{\nu^0_t\}_{t\in[t_0,T]}=\beta_0(\boldsymbol{\mu}^0)$. For every $\beta_1\in\mathcal{B}(t_1,\nu^0_{t_1})$ we set $\beta_{01}\in\mathcal{B}(t_0,\nu^0)$ by $\beta_{01}(\boldsymbol{\mu}^0)=\beta_0(\boldsymbol{\mu}^0)$ in $[t_0,t_1[$, and $\beta_{01}(\boldsymbol{\mu}^0)=\beta_1(\boldsymbol{\mu}^0_{|[t_1,T]})$ in $[t_1,T]$.

$$\begin{aligned}W(t_0,t_1,\mu^0,\nu^0)&\le\sup_{\beta_0\in\mathcal{B}(t_0,\nu^0)}\inf_{\boldsymbol{\mu}^0\in\mathscr{A}^F_{[t_0,T]}(\mu^0)}\sup_{\substack{\beta_1\in\mathcal{B}(t_0,\nu^0_{t_1})\\ \beta_0(\boldsymbol{\mu}^0)=\boldsymbol{\nu}^0}}\mathcal{G}_T(\boldsymbol{\mu}^0,\beta_{01}(\boldsymbol{\mu}^0))\\ &\le\sup_{\beta_0\in\mathcal{B}(t_0,\nu^0)}\inf_{\boldsymbol{\mu}^0\in\mathscr{A}^F_{[t_0,T]}(\mu^0)}\mathcal{G}_T(\boldsymbol{\mu}^0,\beta_0(\boldsymbol{\mu}^0))=V^-(t_0,\mu^0,\nu^0).\end{aligned}$$

We prove now the reverse inequality. Given $\beta_0\in\mathcal{B}(t_0,\nu^0)$ and $\boldsymbol{\mu}^0\in\mathscr{A}^F_{[t_0,T]}(\mu^0)$, we define $\beta_1^{\boldsymbol{\mu}^0}:\mathscr{A}^F_{[t_1,T]}(\mu^0)\to\mathscr{A}^G_{[t_1,T]}(\mu^0_{t_1})$ by setting $\beta_1^{\boldsymbol{\mu}^0}(\boldsymbol{\mu}^1)=(\beta_0(\boldsymbol{\mu}^0))_{|[t_1,T]}$ for all $\boldsymbol{\mu}^1$. Then we have

$$\begin{aligned}W(t_0,t_1,\mu^0,\nu^0)&=\sup_{\beta_0\in\mathcal{B}(t_0,\nu^0)}\inf_{\boldsymbol{\mu}^0\in\mathscr{A}^F_{[t_0,T]}(\mu^0)}\sup_{\substack{\beta_1\in\mathcal{B}(t_0,\nu^0_{t_1})\\ \beta_0(\boldsymbol{\mu}^0)=\boldsymbol{\nu}^0}}\inf_{\boldsymbol{\mu}^1\in\mathscr{A}^F_{[t_1,T]}(\mu^0_{t_1})}\mathcal{G}_T(\boldsymbol{\mu}^1,\beta_1(\boldsymbol{\mu}^1))\\ &\ge\sup_{\beta_0\in\mathcal{B}(t_0,\nu^0)}\inf_{\boldsymbol{\mu}^0\in\mathscr{A}^F_{[t_0,T]}(\mu^0)}\inf_{\boldsymbol{\mu}^1\in\mathscr{A}^F_{[t_1,T]}(\mu^0_{t_1})}\mathcal{G}_T(\boldsymbol{\mu}^1,\beta_1^{\boldsymbol{\mu}^0}(\boldsymbol{\mu}^1))\\ &=\sup_{\beta_0\in\mathcal{B}(t_0,\nu^0)}\inf_{\boldsymbol{\mu}^0\in\mathscr{A}^F_{[t_0,T]}(\mu^0)}\mathcal{G}_T(\boldsymbol{\mu}^0,\beta_0(\boldsymbol{\mu}^0))=V^-(t_0,\mu^0,\nu^0),\end{aligned}$$

which concludes the proof. □

5.3. Existence and characterization of the value

Definition 5.9 *(Hamiltonian function for the pursuit-evasion game).* We consider F, G satisfying **(*F*)**, and define the following Hamiltonian function for all $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$, $p_\mu \in L^2_\mu(\mathbb{R}^d)$, $p_\nu \in L^2_\mu(\mathbb{R}^d)$

$$\mathscr{H}_{PE}(\mu, \nu, p_\mu, p_\nu) = \inf_{\substack{v(\cdot)\in L^2_\mu(\mathbb{R}^d)\\ v(x)\in F(x)\ \mu\text{-a.e.}x}} \int_{\mathbb{R}^d} \langle p_\mu(x), v(x)\rangle\, d\mu(x) + \sup_{\substack{w(\cdot)\in L^2_\nu(\mathbb{R}^d)\\ w(x)\in G(x)\ \nu\text{-a.e.}x}} \int_{\mathbb{R}^d} \langle p_\nu(x), w(x)\rangle\, d\nu(x). \quad (12)$$

Lemma 5.10 *(Smoothness of Hamiltonian function for the pursuit-evasion game). Consider F, G satisfying **(F)**, then the Hamiltonian function $\mathscr{H}_{PE}$ satisfy the following regularity assumptions*

- *for every $\lambda \geq 0$ we have $\mathscr{H}_{PE}(\mu, \nu, \lambda p_\mu, \lambda p_\nu) = \lambda \mathscr{H}_{PE}(\mu, \nu, p_\mu, p_\nu)$;*
- *there exists $k \geq 0$ such that for all $\mu^1, \nu^1, \mu^2, \nu^2 \in \mathscr{P}_2(\mathbb{R}^d)$, $\gamma_\mu \in \Pi_o(\mu^1, \mu^2)$, $\gamma_\nu \in \Pi_o(\nu^1, \nu^2)$, defined $p_{\gamma^\mu} = \mathrm{Id}_{\mathbb{R}^d} - \mathrm{Bar}_1(\gamma_\mu)$, $q_{\gamma^\mu} = \mathrm{Id}_{\mathbb{R}^d} - \mathrm{Bar}_1(\gamma_\mu^{-1})$, $p_{\gamma^\nu} = \mathrm{Id}_{\mathbb{R}^d} - \mathrm{Bar}_1(\gamma_\nu)$, $q_{\gamma^\nu} = \mathrm{Id}_{\mathbb{R}^d} - \mathrm{Bar}_1(\gamma_\nu^{-1})$, we have*

$$\mathscr{H}_{PE}(\mu^1, \nu^1, p_{\gamma^\mu}, p_{\gamma^\mu}) - \mathscr{H}_{PE}(\mu^2, \nu^2, q_{\gamma^\mu}, q_{\gamma^\mu}) \leq k[W_2^2(\mu^1, \mu^2) + W_2^2(\nu^1, \nu^2)].$$

Proof. The first assertion is trivial. For the second one it is sufficient to apply Proposition 4.3 to each term of the sum appearing in (12). □

Proposition 5.11. *The upper and lower value functions $V^\pm(\cdot)$ are viscosity solutions of $\partial_t V + \mathscr{H}_{PE}(\mu, \nu, D_\mu V, D_\nu V) = 0$ on every set with uniformly bounded second moments.*

Proof. The proof will follow the same idea of Theorem 4.4. We will prove only the results for V^-, since the corresponding arguments for V^+ are pretty similar.

Claim 1: V^- is a subsolution of (12).

Proof (of Claim 1). Take $(\bar{t}, \bar{\mu}, \bar{\nu}) \in]0, T[\times \mathscr{P}_2(\mathbb{R}^d) \times \mathscr{P}_2(\mathbb{R}^d)$, $\delta > 0$, $(p_t, p_\mu, p_\nu) \in D^+_\delta V^-(\bar{t}, \bar{\mu}, \bar{\nu})$. Let $v_{\bar{t}} : \mathbb{R}^d \to \mathbb{R}^d$ be a Borel map such that $v_{\bar{t}}(x) \in F(x)$ for μ-a.e. $x \in \mathbb{R}^d$.

According to the Dynamic Programming Principle in Proposition 5.8, for every $\varepsilon > 0$ there exists $\beta_\varepsilon \in \mathcal{B}(\bar{t}, \bar{\nu})$ such that for all $\boldsymbol{\mu} \in \mathscr{A}_{[t_0,T]}(\bar{\mu})$ with $\boldsymbol{\mu} = \{\mu_t\}_{t\in[t_0,T]}$, set $\boldsymbol{\nu}^\varepsilon = \{\nu_t^\varepsilon\}_{t\in[t_0,T]} = \beta_\varepsilon(\boldsymbol{\mu}) \in \mathscr{A}^G_{[t_0,T]}(\bar{\nu})$, we have

$$V^-(t_0, \bar{\mu}, \bar{\nu}) \leq V^-(t, \mu_t, \nu_t^\varepsilon) + \varepsilon.$$

In particular, as in Theorem 4.4, this holds for a $\boldsymbol{\mu} = \{\mu_t\}_{t\in[\bar{t},T]} \in \mathscr{A}^F_{[\bar{t},T]}(\bar{\mu})$ represented by $\boldsymbol{\eta}_\mu \in \mathscr{P}(\mathbb{R}^d \times \Gamma_{[\bar{t},T]})$ such that $\mu_t = e_t \sharp \boldsymbol{\eta}_\mu$ for all $t \in [\bar{t}, T]$ and

$$\lim_{t\to\bar{t}} \int_{\mathbb{R}^d\times\Gamma_{[\bar{t},T]}} \langle p_\mu \circ e_{\bar{t}}(x,\gamma), \frac{e_t(x,\gamma) - e_{\bar{t}}(x,\gamma)}{t - \bar{t}}\rangle\, d\boldsymbol{\eta}_\mu(x,\gamma) = \int_{\mathbb{R}^d} \langle p_\mu(x), v_{\bar{t}}(x)\rangle\, d\bar{\mu}(x).$$

Please cite this article in press as: A. Marigonda, M. Quincampoix, Mayer control problem with probabilistic uncertainty on initial positions, J. Differential Equations (2017), https://doi.org/10.1016/j.jde.2017.11.014

By Proposition 2.3 (2), since the family $\{\boldsymbol{\nu}^\varepsilon\}_{\varepsilon>0}$ is a family of admissible curves satisfying $\nu_0^\varepsilon=\bar\nu$ for all $\varepsilon>0$, every sequence $\{\boldsymbol{\nu}^{\varepsilon_i}\}_{i\in\mathbb{N}}$ with $\varepsilon_i\to0^+$ as $i\to+\infty$, admits a convergent subsequence. In particular, there exists an admissible trajectory $\boldsymbol{\nu}=\{\nu_t\}_{t\in[t_0,T]}\in\mathscr{A}^G_{[t_0,T]}(\bar\nu)$ such that

$$V^-(t,\mu_t,\nu_t)-V^-(t_0,\bar\mu,\bar\nu)\geq 0.$$

As in Theorem 4.4, we define $\tilde\mu=(e_{\bar t},p_\mu\circ e_{\bar t},e_t)\sharp\boldsymbol{\eta}_\mu$ and $\tilde\nu=(e_{\bar t},p_\nu\circ e_{\bar t},e_t)\sharp\boldsymbol{\eta}_\nu$, where $\boldsymbol{\eta}_\nu\in\mathscr{P}(\mathbb{R}^d\times\Gamma_{[\bar t,T]})$ satisfies $\nu_t=e_t\sharp\boldsymbol{\eta}_\nu$ for all $t\in[\bar t,T]$. By applying the definition of viscosity superdifferential, and defined $\tilde\mu$, $\tilde\nu$ as in Claim 1, we have that

$$\begin{aligned}
0\leq & V^-(t,\mu_t,\nu_t)-V^-(\bar t,\bar\mu,\bar\nu)\\
\leq & p_t(t-\bar t)+\int\limits_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d}\langle x_2,x_3-x_1\rangle\,d\tilde\mu(x_1,x_2,x_3)+\\
&+\int\limits_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d}\langle y_2,y_3-y_1\rangle\,d\tilde\nu(y_1,y_2,y_3)+\\
&+\delta\sqrt{(t-\bar t)^2+W^2_{2,\tilde\mu}(\bar\mu,\mu_t)+W^2_{2,\tilde\nu}(\bar\nu,\nu_t)}+o(|t-\bar t|+W_{2,\tilde\mu}(\bar\nu,\nu_t)+W_{2,\tilde\nu}(\bar\nu,\nu_t))\\
= & p_t(t-t_0)+\int\limits_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d}\langle p_\mu\circ e_{\bar t}(x,\gamma),e_t(x,\gamma)-e_{\bar t}(x,\gamma)\rangle\,d\boldsymbol{\eta}_\mu(x,\gamma)+\\
&+\int\limits_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d}\langle p_\nu\circ e_{\bar t}(x,\gamma),e_t(x,\gamma)-e_{\bar t}(x,\gamma)\rangle\,d\boldsymbol{\eta}_\nu(x,\gamma)+\\
&+\delta\sqrt{(t-\bar t)^2+W^2_{2,\tilde\mu}(\bar\mu,\mu)+W^2_{2,\tilde\nu}(\bar\nu,\nu_t)}+o(|t-\bar t|+W^2_{2,\tilde\mu}(\bar\mu,\mu_t)+W_{2,\tilde\nu}(\bar\nu,\nu_t)).
\end{aligned}$$

Dividing by $\sqrt{(t-\bar t)^2+W^2_{2,\tilde\mu}(\bar\mu,\mu_t)+W^2_{2,\tilde\nu}(\bar\nu,\nu_t)}$ and letting $t\to\bar t^+$ along any sequence such that the limit exists yields as in Theorem 4.4

$$-\delta\leq p_t+\int\limits_{\mathbb{R}^d}\langle p_\mu(x),v_{\bar t}(x)\rangle\,d\bar\mu(x)+\lim_{i\to\infty}\int\limits_{\mathbb{R}^d}\langle p_\nu(x),\frac{e_{t_i}(x,\gamma)-e_{\bar t}(x,\gamma)}{t_i-\bar t}\rangle\,d\boldsymbol{\eta}_\nu(x).$$

By (5) in Proposition 2.3, there exists a Borel selection $w_{\bar t}$ of G such that

$$-\delta\leq p_t+\int\limits_{\mathbb{R}^d}\langle p_\mu(x),v_{\bar t}(x)\rangle\,d\bar\mu(x)+\int\limits_{\mathbb{R}^d}\langle p_\nu(x),w_{\bar t}(x)\rangle\,d\bar\nu(x).$$

By the arbitrariness of $v_{\bar t}$ among the $L^2_{\bar\mu}$-selections of F, taking the infimum on $v_{\bar t}$ we have

$$p_t+\mathscr{H}_{PE}(\bar\mu,\bar\nu,p_\mu,p_\nu)\geq-\delta,$$

which ends the proof of Claim 1. ◇

Claim 2: V^- is a supersolution of (12).

Proof (of Claim 2). Take $(\bar t,\bar\mu,\bar\nu)\in]0,T[\times\mathscr{P}_2(\mathbb{R}^d)\times\mathscr{P}_2(\mathbb{R}^d)$, $\delta>0$, $(p_t,p_\mu,p_\nu)\in D^-_\delta V(\bar t,\bar\mu,\bar\nu)$. Let $w_{\bar t}:\mathbb{R}^d\to\mathbb{R}^d$ be a Borel map such that $w_{\bar t}(x)\in F(x)$ for μ-a.e. $x\in\mathbb{R}^d$.

We can find $\boldsymbol{\nu}=\{\nu_t\}_{t\in[\bar t,T]}\in\mathscr{A}^F_{[\bar t,T]}(\bar\nu)$ represented by $\boldsymbol{\eta}_\nu\in\mathscr{P}(\mathbb{R}^d\times\Gamma_{[\bar t,T]})$ such that $\nu_t=e_t\sharp\boldsymbol{\eta}_\nu$ for all $t\in[\bar t,T]$ and

$$\lim_{t\to\bar t}\int_{\mathbb{R}^d\times\Gamma_{[\bar t,T]}}\langle p_\nu\circ e_{\bar t}(x,\gamma),\frac{e_t(x,\gamma)-e_{\bar t}(x,\gamma)}{t-\bar t}\rangle\,d\boldsymbol{\eta}_\nu(x,\gamma)=\int_{\mathbb{R}^d}\langle p_\nu(x),w_{\bar t}(x)\rangle\,d\bar\nu(x).$$

Define the constant strategy $\beta(\boldsymbol{\mu})=\boldsymbol{\nu}$, thus we have from Proposition 5.8

$$V^-(t_0,\bar\mu,\bar\nu)\ge\inf_{\substack{\boldsymbol{\mu}\in\mathscr{A}^F_{[t_0,T]}(\bar\mu)\\ \boldsymbol{\mu}=\{\mu_t\}_{t\in[0,T]}}}V^-(t,\mu_t,\nu_t).$$

As in Claim 1, for any $\varepsilon>0$ we can find $\boldsymbol{\mu}^\varepsilon=\{\mu^\varepsilon_t\}_{t\in[0,T]}\in\mathscr{A}^F_{[t_0,T]}(\bar\mu)$, $\boldsymbol{\mu}=\{\mu_t\}_{t\in[0,T]}\in\mathscr{A}^F_{[t_0,T]}(\bar\mu)$, and $\boldsymbol{\eta}\in\mathscr{P}(\mathbb{R}^d\times\Gamma_{[\bar t,T]})$ such that $\mu_t=e_t\sharp\boldsymbol{\eta}$ for all $t\in[\bar t,T]$ such that $V^-(t_0,\bar\mu,\bar\nu)\ge V^-(t,\mu^\varepsilon_t,\nu_t)-\varepsilon$ and

$$V^-(t,\mu_t,\nu_t)-V^-(t_0,\bar\mu,\bar\nu)\le 0.$$

We proceed now by applying the definition of viscosity subdifferential, dividing by $\sqrt{(t-\bar t)^2+W^2_{2,\tilde\mu}(\bar\mu,\mu_t)+W^2_{2,\tilde\mu}(\bar\nu,\nu_t)}$ and letting $t\to\bar t$ along sequences where the limit exists, as done in Claim 1 and in Theorem 4.4. By the arbitrariness of $w_{\bar t}$, we conclude

$$K'\delta\ge p_t+\mathscr{H}_{PE}(\bar\mu,\bar\nu,p_\mu,p_\nu).\quad\square$$

Theorem 5.12 *(Existence of a value and its characterization). Consider F,G satisfying* **(*F*)***, and a bounded Lipschitz continuous payoff function $\mathcal{G}$. Then the game has a value, i.e., $V^+=V^-=:V$ and V is the unique viscosity solution of the Hamilton–Jacobi–Bellman equation $\partial_t V+\mathscr{H}_{PE}(\mu,\nu,D_\mu V,D_\nu V)=0$, $V(T,\mu,\nu)=\mathcal{G}(\mu,\nu)$.*

Proof. The result follows from the comparison principle proved in Theorem 3.4, and from Proposition 5.11, since both functions solve the same Hamilton–Jacobi–Bellman equation with the same boundary data. □

Finally, we provide an example of possible applications.

Example 5.13 *(Pillage).* Assume that an invader army is sent to plunder a region after having overwhelmed its defending forces. The plundering time is fixed $T>0$. The target of the invaders is to plunder as much as possible food and any other useful supplies, while the target of the civil authorities of the invaded region is to direct the refugees' flow carrying the supplies in order to avoid that they fall in the hands of the enemy. We assume that:

Please cite this article in press as: A. Marigonda, M. Quincampoix, Mayer control problem with probabilistic uncertainty on initial positions, J. Differential Equations (2017), https://doi.org/10.1016/j.jde.2017.11.014

(P_1) The civilian refugees are harmless for the invaders, no matter their concentration compared with the invaders' one.
(P_2) The speed of the invaders are always greater or equal than the speed of the refugees.

We model the situation by defining two time-dependent measures on $\mathbb{R}^2$:

- the *pillage capacity* $\boldsymbol{\mu} = \{\mu_t\}_{t\in[0,T]}$: given a measurable $A \subseteq \mathbb{R}^2$, $\mu_t(A)$ represents the amount of resources that the invaders can plunder from the subregion A at time t: since the spoils of war must be carried back, this quantity can be roughly assumed to be proportional to the number of invader soldiers in the subregion A at time t;
- the *carrying capacity* $\boldsymbol{\nu} = \{\nu_t\}_{t\in[0,T]}$: given a measurable $A \subseteq \mathbb{R}^2$, $\nu_t(A)$ represents the amount of resources that the refugees in the subregion A at time t are carrying with them.

Given $t \in [0,T]$, we write $\nu_t = \dfrac{\nu_t}{\mu_t}\mu_t + \nu_t^s$, where $\nu_t^s \perp \mu_t$ and $\dfrac{\nu_t}{\mu_t}$ is the Radon–Nikodym derivative of ν_t w.r.t. μ_t. Thus the quantity

$$\int_{\mathbb{R}^d} \min\left\{1, \frac{\nu_t}{\mu_t}(x)\right\} d\mu_t(x)$$

represents the spoils of war captured by the invaders at time t, taking into account that the spoils captured cannot exceed the pillage capacity. Given a subregion $C \subseteq \mathbb{R}^2$, we can consider three cases:

- if $\nu_t^s(C) = 0$ and $\dfrac{\nu_t}{\mu_t}(x) = 1$ for μ_t-a.e. $x \in C$, then the invaders have completely plundered the supplies in the subregion C, moreover there are no remaining soldiers in the region C available to be sent to plunder other regions.
- if $\nu_t^s(C) = 0$ and $\dfrac{\nu_t}{\mu_t}(x) < 1$ for μ_t-a.e. $x \in C$, then the invaders have completely plundered the supplies in the subregion C, moreover the spoils carried by the refugees decreased by $\nu_t(C)$ but there are still soldiers available who may be sent to plunder other subregions.
- if none of the above conditions is satisfied, there are still spoils of war in the region C that have not been plundered yet by the invaders.

In the first two cases, the pillage capacity of the invaders and the carrying capacity of the refugees decreases of $\nu_t(C)$.

It is crucial to notice that assumption (P_2) allows us to *postpone* the computation of the spoils captured at the final time $t = T$, since if for $t < T$ we have $0 < \dfrac{\nu_t}{\mu_t} < 1$ in a region C, we can always imagine to split the invaders and the refugees into two populations:

$$\begin{cases} \mu_t = \dfrac{\nu_t}{\mu_t}\mu_{t|C} + \left[\left(1 - \dfrac{\nu_t}{\mu_t}\right)\mu_{t|C} + \mu_{t|\mathbb{R}^2\setminus C}\right], \\ \nu_t = \dfrac{\nu_t}{\mu_t}\mu_{t|C} + \left[\nu^s_{t\,|C} + \nu_{t|\mathbb{R}^2\setminus C}\right]. \end{cases}$$

In this case, any admissible trajectory of the refugees starting from $\frac{\nu_t}{\mu_t}\,\mu_{t|C}$, is also an admissible trajectory for the invaders starting from the same measure, thus these subpopulation will occupy the same position with the same density until the time $t = T$ is reached, while the remaining two subpopulation proceed in the game.

This remark transform the problem with mass loss, in a problem with total mass preserved (both for pillagers and for refugees), and we can assume for both of them that the total mass is normalized to 1. Moreover, the computation of the spoils captured can be made at the final time T only. We model the admissible velocities of the invader soldiers and the refugees by using set-valued maps F and G, respectively, and (P_2) will translate into $F(x) \supseteq G(x)$ for all $x \in \mathbb{R}^d$. The capture functional can be taken to be a variant of the W_2-distance:

$$\mathscr{J}(\mu,\nu) = \min\{W_2(\mu,\nu), C\},$$

where $C > 0$ is a suitable (large) constant, that can be taken to be, for instance, twice the diameter of the invaded region. It is worth of noticing that the problem cannot be reduced to a problem of optimizing the *distance between the supports* of the measures at the final time, because even if μ_T and ν_T have the *same* support, we may have $\mathscr{J}(\mu_T, \nu_T) > 0$. From a model point of view, for the invaders it is not convenient to spread the forces chasing people carrying a low quantity of resources, and, symmetrically, for the refugees is dangerous to convoy supplies in locations where the occupation forces are concentrated.

Acknowledgments

The authors wish to thanks Chloé Jimenez for many valuable discussions during the preparation of the paper, and are indebted with the anonymous referee, whose advises had greatly improved the presentation of the paper.

Appendix A. Preliminaries and notation

In this section we give some preliminaries and fix the notation. Our main references for this part are [2,26,27].

Definition A.1 *(Space of probability measures).* Given Banach spaces X, Y, we denote by $\mathscr{P}(X)$ the set of Borel probability measures on X endowed with the weak* topology induced by the duality with the Banach space $C^0_b(X)$ of the real-valued continuous bounded functions on X with the uniform convergence norm. For any $p \geq 1$, the second moment of $\mu \in \mathscr{P}(X)$ is defined by $\mathrm{m}_2(\mu) = \int_X \|x\|^2_X \, d\mu(x)$, and we set $\mathscr{P}_2(X) = \{\mu \in \mathscr{P}(X) : \mathrm{m}_2(\mathbb{R}^d) < +\infty\}$. For any Borel map $r : X \to Y$ and $\mu \in \mathscr{P}(X)$, we define the *push forward measure* $r\sharp\mu \in \mathscr{P}(Y)$ by setting $r\sharp\mu(B) = \mu(r^{-1}(B))$ for any Borel set B of Y.

The following result is Theorem 5.3.1 in [2].

Theorem A.2 *(Disintegration). Given a measure $\mu \in \mathscr{P}(\mathbb{X})$ and a Borel map $r : \mathbb{X} \to X$, there exists a family of probability measures $\{\mu_x\}_{x\in X} \subseteq \mathscr{P}(\mathbb{X})$, uniquely defined for $r\sharp\mu$-a.e. $x \in X$,* *such that* $\mu_x(\mathbb{X}\setminus r^{-1}(x))=0$ *for* $r\sharp\mu$*-a.e.* $x\in X$*, and for any Borel map* $\varphi : X\times Y\to[0,+\infty]$ *we have*

$$\int_{\mathbb{X}}\varphi(z)\,d\mu(z)=\int_X\left[\int_{r^{-1}(x)}\varphi(z)\,d\mu_x(z)\right]d(r\sharp\mu)(x).$$

We will write $\mu=(r\sharp\mu)\otimes\mu_x$*. If* $\mathbb{X}=X\times Y$ *and* $r^{-1}(x)\subseteq\{x\}\times Y$ *for all* $x\in X$*, we can identify each measure* $\mu_x\in\mathscr{P}(X\times Y)$ *with a measure on* Y*.*

Definition A.3 *(Projections).* Given $N\in\mathbb{N}$, $N>0$ and a finite collection of nonempty sets $X_1,\dots,X_N$, we define the maps $\pi_i : X_1\times\cdots\times X_N\to X_i$ and $\pi_{ij} : X_1\times\cdots\times X_N\to X_i\times X_j$ by setting $\pi_i(x_1,\dots,x_N)=x_i$ and $\pi_{ij}(x_1,\dots,x_N)=(\pi_i\times\pi_j)(x_1,\dots,x_N)=(x_i,x_j)$ for every $x_h\in X_h$, $h=1,\dots,N$. When X_h, $h=1,\dots,N$ are topological spaces, these maps are continuous w.r.t. the product topology.

Definition A.4 *(Transport plans and Wasserstein distance).* Let X be a complete separable Banach space, $\mu_1,\mu_2\in\mathscr{P}(X)$. We define the set of *admissible transport plans* between μ_1 and μ_2 by setting

$$\Pi(\mu_1,\mu_2)=\{\gamma\in\mathscr{P}(X\times X):\pi_i\sharp\gamma=\mu_i,\ i=1,2\}.$$

The *inverse* γ^{-1} of a transport plan $\gamma\in\Pi(\mu,\nu)$ is defined by $\gamma^{-1}=i\sharp\gamma\in\Pi(\nu,\mu)$, where $i(x,y)=(y,x)$ for all $x,y\in X$. The 2-*Wasserstein distance* between μ_1 and μ_2 is

$$W_2^2(\mu_1,\mu_2)=\inf_{\gamma\in\Pi(\mu_1,\mu_2)}\int_{X\times X}|x_1-x_2|^2\,d\gamma(x_1,x_2).$$

If $\mu_1,\mu_2\in\mathscr{P}_2(X)$ then the above infimum is actually a minimum, and we define

$$\Pi_o(\mu_1,\mu_2)=\left\{\gamma\in\Pi(\mu_1,\mu_2):W_2^2(\mu_1,\mu_2)=\int_{X\times X}|x_1-x_2|^p\,d\gamma(x_1,x_2)\right\}.$$

The space $\mathscr{P}_2(X)$ endowed with the W_2-Wasserstein distance is a complete separable metric space, moreover for all $\mu\in\mathscr{P}_2(X)$ there exists a sequence $\{\mu^N\}_{N\in\mathbb{N}}\subseteq\mathrm{co}\{\delta_x : x\in\operatorname{supp}\mu\}$ such that $W_2(\mu^N,\mu)\to 0$ as $N\to+\infty$.

Definition A.5 *(Barycenter).* Given $\gamma\in\Pi(\mu_1,\mu_2)\cap\mathscr{P}_2(\mathbb{R}^d\times\mathbb{R}^d)$ with $\gamma=\mu_i\otimes\gamma_{x_i}$, its *barycentric* i*-th projection* $\mathrm{Bar}_i(\gamma)\in L^2_{\mu_i}(\mathbb{R}^d;\mathbb{R}^d)$, $i=1,2$, is defined by

$$\mathrm{Bar}_i(\gamma)(x_i)=\int_{\mathbb{R}^d}x_j\,d\gamma_{x_i}(x_j),\ \text{for } \mu_i\text{-a.e. } x_i\in\mathbb{R}^d,\ i,j\in\{1,2\},\ i\neq j.$$

Definition A.6 *(Transport multi-plans).* Let $\gamma \in \mathscr{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$ be a transport plan, and let $\mu_3 \in \mathscr{P}_2(\mathbb{R}^d)$. We set $\mu_1 = \pi_1 \sharp \gamma$ and

$$\Pi(\gamma, \mu_3) := \{\tilde{\mu} \in \mathscr{P}_2(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d) : \pi_{12}\sharp\tilde{\mu} = \gamma,\ \pi_3\sharp\tilde{\mu} = \mu_3\},$$
$$\Pi_o(\gamma, \mu_3) := \{\tilde{\mu} \in \mathscr{P}_2(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d) : \pi_{12}\sharp\tilde{\mu} = \gamma,\ \pi_{13}\sharp\tilde{\mu} \in \Pi_o(\mu_1, \mu_3)\}.$$

Given $\tilde{\mu} \in \mathscr{P}_2(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$, $i, j = 1, 2, 3$, we set $\mu_i = \pi_i \sharp \tilde{\mu}$ and

$$W^2_{2,\tilde{\mu}}(\mu_i, \mu_j) = \int_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} |x_i - x_j|^2 \, d\tilde{\mu}(x_1, x_2, x_3).$$

Clearly, $W_{2,\tilde{\mu}}(\mu_i, \mu_j) \geq W_2(\mu_i, \mu_j)$ for all $i, j = 1, 2, 3$.

The following is Lemma 5.3.2 p.122 in [2].

Lemma A.7 *(Composition of transport plans). Let $\gamma_{12}, \gamma_{13} \in \mathscr{P}(\mathbb{R}^d \times \mathbb{R}^d)$ be such that $\pi_1\sharp\gamma_{12} = \pi_1\sharp\gamma_{13} = \mu_1 \in \mathscr{P}(\mathbb{R}^d)$. Then there exists $\tilde{\mu} \in \mathscr{P}(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$ such that $\pi_{12}\sharp\tilde{\mu} = \gamma_{12}$ and $\pi_{13}\sharp\tilde{\mu} = \gamma_{13}$. In particular, if $\gamma_{12} = \mu_1 \otimes \gamma_{12}^{x_1}$, $\gamma_{13} = \mu_1 \otimes \gamma_{13}^{x_1}$, and $\tilde{\mu} = \mu_1 \otimes \tilde{\mu}_{x_1}$, we have $\tilde{\mu}_{x_1} \in \Pi(\gamma_{12}^{x_1}, \gamma_{13}^{x_1})$ for μ_1-a.e. $x_1 \in \mathbb{R}^d$. The measure $\tilde{\mu}$ is unique if γ_{12} or γ_{13} are induced by a transport map.*

Theorem A.8 *(Superposition principle). Let $\boldsymbol{\mu} = \{\mu_t\}_{t\in[0,T]}$ be a solution of the continuity equation $\partial_t \mu_t + \operatorname{div}(v_t \mu_t) = 0$ for a suitable Borel vector field $v : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ satisfying*

$$\int_0^T \int_{\mathbb{R}^d} \frac{|v_t(x)|}{1 + |x|} \, d\mu_t(x) \, dt < +\infty .$$

Then there exists a probability measure $\boldsymbol{\eta} \in \mathscr{P}(\mathbb{R}^d \times \Gamma_T)$, with $\Gamma_T = C^0([0, T]; \mathbb{R}^d)$ endowed with the sup *norm, such that*

(i) *$\boldsymbol{\eta}$ is concentrated on the pairs $(x, \gamma) \in \mathbb{R}^d \times \Gamma_T$ such that γ is an absolutely continuous solution of*

$$\begin{cases} \dot{\gamma}(t) = v_t(\gamma(t)), & \text{for } \mathscr{L}^1\text{-a.e } t \in (0, T) \\ \gamma(0) = x, \end{cases}$$

(ii) *$\mu_t = e_t \sharp \boldsymbol{\eta}$ for all $t \in [0, T]$.*

Conversely, given any $\boldsymbol{\eta}$ satisfying (i) above and defined $\boldsymbol{\mu} = \{\mu_t\}_{t\in[0,T]}$ as in (ii) above, we have that $\partial_t \mu_t + \operatorname{div}(v_t \mu_t) = 0$ and $\mu_{|t=0} = \gamma(0)\sharp\boldsymbol{\eta}$.

Proof. See Theorem 8.2.1 in [2]. □

Please cite this article in press as: A. Marigonda, M. Quincampoix, Mayer control problem with probabilistic uncertainty on initial positions, J. Differential Equations (2017), https://doi.org/10.1016/j.jde.2017.11.014

Appendix B. Comparison with other notion of generalized differentials

Remark B.1. If we require item ii.) of Definition 3.2 to hold only for measures $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ *induced by a transport map* from $\bar{\mu}$, i.e. to measures $\mu = (\mathrm{Id} + \phi)\sharp\bar{\mu}$, we have that there exists only one $\tilde{\mu} \in \mathscr{P}_2(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$ such that $\pi_{12}\sharp\tilde{\mu} = (\mathrm{Id}, p_{\bar{\mu}})\sharp\bar{\mu}$ and $\pi_{13} = (\mathrm{Id}, \mathrm{Id} + \phi)\sharp\bar{\mu}$, due to Lemma A.7, and we have $\tilde{\mu} = (\mathrm{Id}, p_{\bar{\mu}}, \mathrm{Id} + \phi)\sharp\bar{\mu}$. In this case, $W_{2,\tilde{\mu}}(\bar{\mu}, \mu) = \|\phi\|_{L^2_{\bar{\mu}}}$, and we recover essentially the same definition of δ-sub/superdifferential used in [12], in particular the two definitions agrees when $\bar{\mu} \ll \mathscr{L}^d$.

Remark B.2. More generally, in item ii.) of Definition 3.2 we can consider an absolutely continuous $\boldsymbol{\mu} = \{\mu_s\}_{s\in[0,t]}$ curve joining $\bar{\mu}$ to μ, represented by $\boldsymbol{\eta} \in \mathscr{P}(\mathbb{R}^d \times \Gamma_t)$ satisfying $\mu_s = e_s\sharp\boldsymbol{\eta}$, we have that we can choose $\tilde{\mu} \in \mathscr{P}_2(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$ to be $\tilde{\mu} = (e_0, p_{\bar{\mu}} \circ e_0, e_t)\sharp\boldsymbol{\eta}$, thus recovering the same definition of δ-sub/superdifferential used in [16].

We will now compare the definition given in Definition 3.2 with the following one, appeared in Definition 10.3.1 p. 241 of [2].

Definition B.3 *(Fréchet subdifferential).* Let $w : \mathscr{P}_2(\mathbb{R}^d) \to]-\infty, +\infty]$ be proper and l.s.c., $\mu_1 \in \mathscr{P}_2(\mathbb{R}^d)$ such that $w(\mu_1) \in \mathbb{R}$. A plan $\gamma \in \mathscr{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$ belongs to the *Fréchet subdifferential* $\partial w(\mu_1)$ if

- $\pi_1\sharp\gamma = \mu_1$;
- $w(\mu_3) - w(\mu_1) \geq \inf\limits_{\tilde{\mu}\in\Pi_o(\gamma,\mu_3)} \displaystyle\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} \langle x_2, x_3 - x_1\rangle \, d\tilde{\mu} + o(W_2(\mu_1, \mu_3))$

We say that $\gamma \in \mathscr{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$ is a *strong Fréchet subdifferential* if for all $\tilde{\mu} \in \Pi(\gamma, \mu_3)$ we have

$$w(\mu_3) - w(\mu_1) \geq \int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} \langle x_2, x_3 - x_1\rangle \, d\tilde{\mu}(x_1, x_2, x_3) + o(W_{2,\tilde{\mu}}(\mu_1, \mu_3)).$$

Similarly, we say that $\gamma \in \mathscr{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$ is a *strong Fréchet superdifferential* if for all $\tilde{\mu} \in \Pi(\gamma, \mu_3)$ we have

$$w(\mu_3) - w(\mu_1) \leq \int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} \langle x_2, x_3 - x_1\rangle \, d\tilde{\mu}(x_1, x_2, x_3) + o(W_{2,\tilde{\mu}}(\mu_1, \mu_3)).$$

Remark B.4. From the definition, we have that if $(p_t, p_{\bar{\mu}}) \in \bigcap\limits_{\delta>0} D^+_\delta v(\bar{t}, \bar{\mu})$ according to Definition 3.2, then $(\mathrm{Id}_{\mathbb{R}^d}, p_{\bar{\mu}})\sharp\bar{\mu}$ is a strong Fréchet superdifferential of $\mu \mapsto w(\bar{t}, \mu)$ at $\bar{\mu}$. Conversely, given a strong Fréchet superdifferential γ of $\mu \mapsto w(\bar{t}, \mu)$ at $\bar{\mu}$, if there exists $\bar{\nu}$ such that $\gamma \in \Pi_o(\bar{\mu}, \bar{\nu})$, then, set $p_{\bar{\mu}} := \mathrm{Id}_{\mathbb{R}^d} - \mathrm{Bar}_1(\gamma)$, for all $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ and $\delta > 0$ we have

$$w(\bar{t}, \mu) - w(\bar{t}, \bar{\mu}) \leq \int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} \langle x_2, x_3 - x_1\rangle \, d\tilde{\mu}(x_1, x_2, x_3) + \delta \cdot W_{2,\tilde{\mu}}(\bar{\mu}, \mu) + o(W_{2,\tilde{\mu}}(\bar{\mu}, \mu)),$$

for all $\tilde{\mu} \in \mathscr{P}(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$ satisfying $\pi_{12}\sharp\tilde{\mu} = (\mathrm{Id}_{\mathbb{R}^d}, p_{\bar{\mu}})\sharp\bar{\mu}$ and $\pi_{13}\sharp\tilde{\mu} \in \Pi(\bar{\mu}, \mu)$.

We state here the following result, which is contained in Theorem 10.2.2 p. 236 and Theorem 10.4.12 p. 270 in [2],

Theorem B.5. *Let* $\mu_2 \in \mathscr{P}_2(\mathbb{R}^d)$*. Define the map* $\psi : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ *by setting* $\psi(\mu) = -\frac{1}{2}W_2^2(\mu, \mu_2)$ *and set*

$$|\partial\psi|(\mu_1) = \frac{1}{2}\limsup_{\mu_3\to\mu_1} \frac{W_2^2(\mu_3, \mu_2) - W_2^2(\mu_1, \mu_2)}{W_2^2(\mu_3, \mu_1)},$$

i.e., $|\partial\psi|(\mu_1)$ *is the* metric slope *of* ψ *at* μ_1*. Then*

(1) *for every* $\mu_3 \in \mathscr{P}_2(\mathbb{R}^d)$*,* $\gamma \in \Pi_o(\mu_1, \mu_2)$*,* $\tilde{\mu} \in \Pi(\gamma, \mu_3)$ *we have*

$$\frac{1}{2}W_2^2(\mu_3, \mu_2) - \frac{1}{2}W_2^2(\mu_1, \mu_2) \leq$$
$$\leq \int\limits_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} \langle x_1 - x_2, x_3 - x_1\rangle \, d\tilde{\mu}(x_1, x_2, x_3) + o(W_{2,\tilde{\mu}}(\mu_1, \mu_3)),$$

and we can choose $o(W_{2,\tilde{\mu}}(\mu_1, \mu_3)) = W_{2,\tilde{\mu}}^2(\mu_1, \mu_3)$*;*

(2) *for all* $\mu_1 \in \mathscr{P}_2(\mathbb{R}^d)$ *we have*

$$|\partial\psi|(\mu_1) = \inf\left\{ \left\|\mathrm{Bar}_1(\gamma) - \mathrm{Id}_{\mathbb{R}^d}\right\|_{L^2_{\mu_1}} : \gamma \in \Pi_o(\mu_1, \mu_2)\right\}.$$

(3) *the previous infimum is a minimum, and it is attained in a unique point* $\gamma_{12} \in \Pi_o(\mu_1, \mu_2)$*, moreover* $(\mathrm{Id}_{\mathbb{R}^d}, \mathrm{Bar}_1(\gamma_{12}) - \mathrm{Id}_{\mathbb{R}^d})\sharp\mu_1$ *is a strong Fréchet subdifferential of* ψ *at* μ_1*.*

References

[1] S. As Soulaimani, Viability with probabilistic knowledge of initial condition, application to optimal control, Set-Valued Anal. 16 (7–8) (2008) 1037–1060, https://doi.org/10.1007/s11228-008-0097-5.

[2] Luigi Ambrosio, Nicola Gigli, Giuseppe Savaré, Gradient Flows in Metric Spaces and in the Space of Probability Measures, 2nd ed., Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2008, MR2401600 (2009h:49002).

[3] Jean-Pierre Aubin, Arrigo Cellina, Differential Inclusions, Grundlehren der Mathematischen Wissenschaften (Fundamental Principles of Mathematical Sciences), vol. 264, Springer-Verlag, Berlin, 1984, Set-valued maps and viability theory, MR755330.

[4] Jean-Pierre Aubin, Ivar Ekeland, Applied Nonlinear Analysis, Pure and Applied Mathematics (New York), John Wiley & Sons, Inc., New York, 1984, A Wiley-Interscience Publication, MR749753.

[5] Jean-Pierre Aubin, Hélène Frankowska, Set-Valued Analysis, Modern Birkhäuser Classics, Birkhäuser Boston Inc., Boston, MA, 2009, Reprint of the 1990 edition [MR1048347], MR2458436.

[6] R. Buckdahn, P. Cardaliaguet, M. Quincampoix, Some recent aspects of differential game theory, Dyn. Games Appl. 1 (1) (2011) 74–114, https://doi.org/10.1007/s13235-010-0005-0, MR2800786.

[7] Martino Bardi, Italo Capuzzo-Dolcetta, Optimal Control and Viscosity Solutions of Hamilton–Jacobi–Bellman Equations, Systems & Control: Foundations & Applications, Birkhäuser Boston, Inc., Boston, MA, 1997, With appendices by Maurizio Falcone and Pierpaolo Soravia, MR1484411.

[8] Alberto Bressan, Differential inclusions and the control of forest fires, J. Differential Equations 243 (2) (2007) 179–207, https://doi.org/10.1016/j.jde.2007.03.009, MR2371785.

Please cite this article in press as: A. Marigonda, M. Quincampoix, Mayer control problem with probabilistic uncertainty on initial positions, J. Differential Equations (2017), https://doi.org/10.1016/j.jde.2017.11.014

[9] Alberto Bressan, Tao Wang, Equivalent formulation and numerical analysis of a fire confinement problem, ESAIM Control Optim. Calc. Var. 16 (4) (2010) 974–1001, https://doi.org/10.1051/cocv/2009033, MR2744158.
[10] Alberto Bressan, Dongmei Zhang, Control problems for a class of set valued evolutions, Set-Valued Var. Anal. 20 (4) (2012) 581–601, https://doi.org/10.1007/s11228-012-0204-5, MR2990209.
[11] G. Buttazzo, C. Jimenez, E. Oudet, An optimization problem for mass transportation with congested dynamics, SIAM J. Control Optim. 48 (3) (2009) 1961–1976, https://doi.org/10.1137/07070543X, MR2516195.
[12] Pierre Cardaliaguet, Marc Quincampoix, Deterministic differential games under probability knowledge of initial condition, Int. Game Theory Rev. 10 (1) (2008) 1–16.
[13] P. Cardaliaguet, C. Jimenez, M. Quincampoix, Pure and random strategies in differential game with incomplete informations, J. Dyn. Games 1 (3) (2014) 363–375.
[14] Giulia Cavagnari, Regularity results for a time-optimal control problem in the space of probability measures, Math. Control Relat. Fields 7 (2) (2017) 213–233, https://doi.org/10.3934/mcrf.2017007, MR3639457.
[15] Giulia Cavagnari, Antonio Marigonda, Time-optimal control problem in the space of probability measures, in: Large-Scale Scientific Computing, in: Lecture Notes in Comput. Sci., vol. 9374, Springer, Cham, 2015, pp. 109–116, MR3480817.
[16] Giulia Cavagnari, Antonio Marigonda, Khai Tien Nguyen, Fabio S. Priuli, Generalized control systems in the space of probability measures, Set-Valued Var. Anal. (2017), https://doi.org/10.1007/s11228-017-0414-y, in press.
[17] Giulia Cavagnari, Antonio Marigonda, Benedetto Piccoli, Optimal synchronization problem for a multi-agent system, Netw. Heterog. Media 12 (2) (2017) 277–295, https://doi.org/10.3934/nhm.2017012, MR3657115.
[18] Giulia Cavagnari, Antonio Marigonda, Benedetto Piccoli, Superposition principle for differential inclusions, in: Large-Scale Scientific Computing, in: Lecture Notes in Comput. Sci., vol. 10665, Springer, Cham, 2017, in press.
[19] Rinaldo M. Colombo, Thomas Lorenz, Nikolay I. Pogodaev, On the modeling of moving populations through set evolution equations, Discrete Contin. Dyn. Syst. 35 (1) (2015) 73–98, https://doi.org/10.3934/dcds.2015.35.73, MR3286949.
[20] R.M. Colombo, N. Pogodaev, Confinement strategies in a model for the interaction between individuals and a continuum, SIAM J. Appl. Dyn. Syst. 11 (2) (2012) 741–770, https://doi.org/10.1137/110854321, MR2967463.
[21] R.M. Colombo, N. Pogodaev, On the control of moving sets: positive and negative confinement results, SIAM J. Control Optim. 51 (1) (2013) 380–401, https://doi.org/10.1137/12087791X, MR3032880.
[22] Michael G. Crandall, Pierre-Louis Lions, Hamilton–Jacobi equations in infinite dimensions. I. Uniqueness of viscosity solutions, J. Funct. Anal. 62 (3) (1985) 379–396, https://doi.org/10.1016/0022-1236(85)90011-4, MR794776.
[23] Jean Dolbeault, Bruno Nazaret, Giuseppe Savaré, A new class of transport distances between measures, Calc. Var. Partial Differential Equations 34 (2) (2009) 193–231, https://doi.org/10.1007/s00526-008-0182-5, MR2448650 (2009g:49019).
[24] Chloé Jimenez, Marc Quincampoix, Hamilton–Jacobi–Isaacs equations for differential games with asymmetric information on probabilistic initial condition, J. Math. Anal. Appl. 457 (2) (2018) 1422–1451, https://doi.org/10.1016/j.jmaa.2017.08.012.
[25] Nikolay Pogodaev, Optimal control of continuity equations, NoDEA Nonlinear Differential Equations Appl. 23 (2) (2016), https://doi.org/10.1007/s00030-016-0357-2, Art. 21, 24, MR3483165.
[26] Filippo Santambrogio, Optimal Transport for Applied Mathematicians, Progress in Nonlinear Differential Equations and Their Applications, vol. 87, Birkhäuser/Springer, Cham, 2015, Calculus of variations, PDEs, and modeling, MR3409718.
[27] Cédric Villani, Topics in Optimal Transportation, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003, MR1964483 (2004e:90003).