HAL
open science

Metric gradient flows with state dependent functionals: the Nash-MFG equilibrium flows and their numerical schemes

Gabriel Turinici

To cite this version:

Gabriel Turinici. Metric gradient flows with state dependent functionals: the Nash-MFG equilibrium flows and their numerical schemes. Nonlinear Analysis: Theory, Methods and Applications, 2017, 165 (December 2017), pp.163-181. 10.1016/j.na.2017.10.002 . hal-01528480v2

HAL Id: hal-01528480

https://hal.science/hal-01528480v2

Submitted on 7 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Metric gradient flows with state dependent functionals: the Nash-MFG equilibrium flows and their numerical schemes

Gabriel Turinici*

October 7, 2017

Abstract

We investigate the convergence of a relaxed version of the *best reply* numerical schemes (also known as *best response* or *fictitious play*) used to find Nash-mean field games equilibriums. This leads us to consider evolution equations in metric spaces similar to gradient flows except that the functional to be differentiated depends on the current point; these are called equilibrium flows. We give two definitions of solutions and prove, through the introduction of a specific index Υ depending on the trajectory, that, as the time step tends to zero, the interpolated (*à la* de Giorgi) numerical curves converge to equilibrium flows. As a by-product we obtain a sufficient condition for the uniqueness of a mean field games equilibrium. We close with applications to congestion and vaccination mean field games.
Keywords: gradient flows; mean field games; vaccination games;

1 Introduction

Let $\mathcal{X}$ be a a Polish geodesic metric space (see [11] for an introduction to metric spaces) and $\mathcal{C}(\cdot,\cdot) : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ a functional. We investigate in this work the equation:

$$\partial_t x_t + \nabla_1 \mathcal{C}(x_t, x_t) = 0, \ x_0 = \bar{x}. \tag{1}$$

*Université Paris-Dauphine, PSL Research University, CNRS UMR 7534, CEREMADE, 75016 PARIS, FRANCE & Institut Universitaire de France, gabriel.turinici@dauphine.fr

Such an equation is called an *equilibrium flow* or *partial flow* for reasons that will be made clear in the sequel.

Discrete (numerically computable) versions of this evolution equation are the numerical schemes defined by the recurrence:

$$x_0^\tau = \bar{x},\ x_{k+1}^\tau \in argmin_{y\in\mathcal{X}} \frac{d(y, x_k^\tau)^2}{2\tau} + \mathcal{C}(y, x_k^\tau), \quad k \geq 0. \tag{2}$$

These numerical schemes are relaxed versions of the *best reply / best response / fictitious play* algorithms (see [7]); the original schemes take sometimes $\tau = \infty$ i.e., omit the first term in (2).

Our goal is to give a rigorous definition of the concept of solution of (1) and show that the numerical schemes (2) converges, when $\tau \to 0$, to a solution of (1). Finally we give examples that show that the equilibrium flows can be successfully used in the study of mean field games equilibriums.

The present work provides thus a rigorous treatment of a two-argument (partial) gradient flow distinct from previous, time-dependent approaches; this allows to obtain the convergence of the "best reply" numerical schemes (2) but also novel uniqueness results for MFGs.

These results are not available with previous techniques from [1, 17, 29, 17, 42, 37, 38]), see remark 5 after the proof of theorem 1; in order to succeed, we introduce a new index Υ dependent of the trajectory (see definition 16) and formalize its expected properties in assumptions $(\mathbf{A}_7)$ and $(\mathbf{A}_8)$, which we prove to be compatible with many different MFG applications (see sections 3.1 to 3.4 and their application-dependent metric spaces that fulfill assumptions $(\mathbf{A}_7)$ and $(\mathbf{A}_8)$); the manipulation of the index Υ requires to obtain some upper bounds (see the proof of theorem 1, one of our main results); finally we are able to obtain estimates of the partial flow divergence by making use of the hypothesis $(\mathbf{A}_8)$.

1.1 Motivation and literature review

The equilibrium in non-cooperative multi-player games are often formulated as mixed strategy Nash equilibriums (see [39]). The computation of such equilibriums and the procedure for players to reach them has been the object of many contributions and give rise to several proposals e.g., replicator dynamics and fictitious play / best reply / best response dynamics, see [19] for details.

The relatively recent introduction of the mean field games (abbreviated from now on as MFG) by Lasry and Lions [34, 33, 35] and conjointly by Huang, Malhamé and Caines [26, 25] (see also [36, 12, 6, 24, 14, 13, 22, 23, 41, 40, 15] for entry points to the literature) allow to extend this concept

to games with an infinite number of players. In this context the players are considered similar (or decomposed in several classes, each with an infinite number of individuals) and an equilibrium is attained when all agents in a class use same mixed strategy, which is optimal in the Nash sense. Mixed strategies are probability measures over the state of pure strategies and thus form a metric space (we will come back later to the topological description of that space); in order to gain in generality we will suppose from now on that the space of all mixed strategies is metrizable and will be denoted $\mathcal{X}$. The cost of the individual strategy x depends on the choice of everybody else's strategy $y \in \mathcal{X}$ and is encoded through the cost function $\mathcal{C}(x, y)$. A MFG equilibrium is thus a point $x \in \mathcal{X}$ such that

$$\mathcal{C}(x, x) \leq \mathcal{C}(z, x), \forall z \in \mathcal{X}. \tag{3}$$

In this context, the relaxed best reply algorithm, which corresponds to (2) has been proposed and tested (see e.g., [7, 8]) with successful results. However only very few works concern the behavior of solutions for $\tau \to 0$ in the general framework of metric spaces or the meaning to be given to the limit equation (1).

Note that when $\mathcal{C}$ is independent of the second argument, i.e.,

$$\mathcal{C}(x, y) = E(x), \tag{4}$$

the relation (2) becomes the celebrated implicit Euler-type scheme of Jordan, Kinderlehrer and Otto [28] for the definition of gradient flows in metric spaces

$$\partial_t y_t + \nabla E(y_t) = 0, \ y_0 = \bar{y}, \tag{5}$$

and received considerable attention (see [44, 43, 2] for instance). However, the situation when E has dependence on other variables has not been treated to the same extent and the related contributions involve gradient flows of time dependent functionals $\mathcal{E}(t, u)$ with a known dependence on time (see [29, 17, 42, 37, 38]). Of course, formally one can set

$$\mathcal{E}(t, u) = \mathcal{C}(u, x_t). \tag{6}$$

However the cited papers use implicit (in time) optimization

$$y_0^\tau = \bar{x}, \ y_{k+1}^\tau \in argmin_{y \in \mathcal{X}} \frac{d(y, y_k^\tau)^2}{2\tau} + \mathcal{E}((k+1)\tau, y_k^\tau), \quad k \geq 0, \tag{7}$$

which are not equivalent to the numerical scheme (2) because (7) is fully implicit while the (relaxed) best-reply scheme is semi-explicit. Furthermore, technical assumptions invoked in previous works assume that the time-dependence of $\mathcal{E}(t, u)$ is smooth (e.g., in [42, assumption (2.19c) page 109])

or that its differentiability points have additional properties (for instance independent of x in [17, assumption E3.1], [29, assumption A4]) while here the mere absolute continuity of $t \mapsto x_t$ does not allow to fulfill *a priori* such assumptions. Nevertheless our contribution ows much to all these previous works which it translates to this specific setting.

2 Theoretical results

2.1 Basic reminders and motivation

The absence of a vector operations in a metric space does no allow to develop fully a differential calculus and requires adaptation of notions of derivative. Accordingly the definition of evolution equations have to use alternative properties.

We recall below the main ideas of such an alternative formulation (see [1]) for the particular case (4)-(5); suppose for a moment than $\mathcal{X}$ is an Euclidian space and E a smooth (C^1 or above) function; then:

$$\frac{d}{dt}E(x_t) = \langle \nabla E(x_t), x_t' \rangle \geq -\left|\nabla E(x_t)\right| \cdot \left|x_t'\right| \geq -\frac{1}{2}\left|x_t'\right|^2 - \frac{1}{2}\left|\nabla E\right|^2(x_t),$$

or equivalently,

$$\frac{d}{dt}E(x_t) + \frac{1}{2}\left|x_t'\right|^2 + \frac{1}{2}\left|\nabla E\right|^2(x_t) \geq 0 \ \forall t, \tag{8}$$

with equality only if x is solution of (5). Therefore asking that

$$\frac{d}{dt}E(x_t) + \frac{1}{2}\left|x_t'\right|^2 + \frac{1}{2}\left|\nabla E\right|^2(x_t) \leq 0 \ \forall t, \tag{9}$$

is an equivalent characterization of (5) (more preciselly called the EDI formulation). The integral form can also be used:

$$E(x_b) - E(x_a) + \int_a^b \left(\frac{1}{2}\left|x_t'\right|^2 + \frac{1}{2}\left|\nabla E\right|^2(x_t)\right) dt \leq 0 \ \forall \ 0 \leq a \leq b. \tag{10}$$

The advantage of formulation (10) is that it only uses quantities that can be defined in a metric space (see below for definition of $|x_t'|$ and $|\nabla E|$). The corresponding computation for a bi-variate functional $\mathcal{C}$ is:

$$\int_a^b \left(\frac{d}{dt}\mathcal{C}(x_t,\nu)\Big|_{\nu=x_t}\right) dt + \int_a^b \left(\frac{1}{2}\left|x_t'\right|^2 + \frac{1}{2}\left|\nabla_1\mathcal{C}\right|^2(x_t,x_t)\right) dt \leq 0 \, \forall \, 0 \leq a \leq b. \tag{11}$$

However this formulation poses specific problems (see also section 3.1) as in general the solution $(x_t)_{t\geq 0}$ is only absolutely continuous (with respect to time) while the manipulation of the term $\frac{d}{dt}\mathcal{C}(x_t,\nu)\Big|_{\nu=x_t}$ requires additional assumptions. This is the object of the next section.

Before that, let us recall the following definition:

Definition 1 *A curve $x:[0,T]\to(\mathcal{X},d)$ is called absolutely continuous if there exists $f\in L^1(0,T)$ such that*

$$d(x_{t_1},x_{t_2})\leq\int_{t_1}^{t_2}f(t)dt,\ \forall t_1<t_2,\ t_1,t_2\in[0,T]. \tag{12}$$

For an absolutely continuous curve $(x_t)_{t\in[0,T]}$ the metric derivative of x at r defined by

$$|x_r'|=\lim_{h\to 0}\frac{d(x_{r+h},x_r)}{|h|}, \tag{13}$$

exists a.e., belongs to $L^1(0,T)$ and is the smallest L^1 function that verifies (12).

2.2 Definition of EDI / EVI equilibrium flows

Let us denote

$$\mathcal{D}^s(\mathcal{C})=\{x\in\mathcal{X}|\mathcal{C}(x,x)<\infty\}. \tag{14}$$

We suppose from now on that $\mathcal{C}$ satisfies the assumption:

($\mathbf{A}_1$) There exists $C_1<\infty$ such that $\mathcal{C}(y,x)\geq -C_1$, $\forall x,y\in\mathcal{X}$.

For any $\alpha,\beta\in\mathbb{R}$, $\alpha\leq\beta$, we denote by $\mathcal{S}(\alpha,\beta)$ the set of divisions of the interval $[\alpha,\beta]$. Let $x=(x_t)_{t\in[0,T]}\subset\mathcal{D}^s(\mathcal{C})$ be an absolutely continuous curve in $\mathcal{X}$; define for $0\leq a\leq b\leq T$ and a division $\Delta=\{a=t_0<t_1<...t_{N_\Delta}=b\}\in\mathcal{S}(a,b)$:

$$\Upsilon(\Delta;x,a,b)=\sum_k\mathcal{C}(x_{k+1},x_k)-\mathcal{C}(x_k,x_k). \tag{15}$$

$$\Upsilon(x,a,b)=\liminf_{\Delta\in\mathcal{S}(a,b),\ |\Delta|\to 0}\Upsilon(\Delta;x,a,b). \tag{16}$$

To ease notations, when there is no ambiguity about the set $\mathcal{S}(\cdot,\cdot)$ we will omit it and write for instance $\Upsilon(x,a,b)=\liminf_{|\Delta|\to 0}\Upsilon(\Delta;x,a,b)$ instead of (16).

Remark 1 *When $\mathcal{X}$ is e.g., Euclidian and under regularity assumptions on $\mathcal{C}$ it is easy to check that $\Upsilon(x,a,b)=\int_a^b\frac{d}{dt}\mathcal{C}(x_t,\nu)\Big|_{\nu=x_t}dt$.*

Definition 2 (EDI equilibrium flow) *An absolutely continuous curve* $(x_t)_{t\in[0,T]}$ *is called an EDI-equlibrium flow starting from* $\bar{x}$ *if* $\lim_{t\to 0} x_t = \bar{x}$ *and:*

$$\forall s \geq 0,\ \Upsilon(x,0,s) + \frac{1}{2}\int_0^s |x_r'|^2 \,\mathrm{d}r + \frac{1}{2}\int_0^s |\nabla_1 \mathcal{C}|^2 (x_r, x_r)\,\mathrm{d}r \leq 0, \tag{17}$$

$$a.e.\ t > 0,\ \forall s \geq t,\ \Upsilon(x,t,s) + \frac{1}{2}\int_t^s |x_r'|^2 \,\mathrm{d}r + \frac{1}{2}\int_t^s |\nabla_1 \mathcal{C}|^2 (x_r, x_r)\,\mathrm{d}r \leq 0, \tag{18}$$

where, for any $y \in \mathcal{D}^s(\mathcal{C})$ *and any point* (x,y) *with* $\mathcal{C}(x,y) < \infty$*, the slope* $|\nabla_1 \mathcal{C}|(x,y)$ *of* $\mathcal{C}(\cdot,\cdot)$ *with respect to the first argument evaluated at* (x,y) *is:*

$$|\nabla_1 \mathcal{C}|(x,y) = \limsup_{z\to x} \frac{(\mathcal{C}(x,y) - \mathcal{C}(z,y))^+}{d(x,z)} = \max\left\{\limsup_{z\to x} \frac{\mathcal{C}(x,y) - \mathcal{C}(z,y)}{d(x,z)}, 0\right\}.$$

Remark 2 *For the particular case* (4) *the definition above coincides with the definition of a EDI gradient flow, see [1]. Moreover an equilibrium flow for* $\mathcal{C}(x,y)$ *is also an equilibrium flow for* $\mathcal{C}(x,y) + G(y)$ *for any function* G*.*

Remark 3 *A natural question is if there exist functions* $\mathcal{C}$ *that satisfy the above assumptions and that cannot be treated with developments in previous works [29, 17]; the answer is positive, see for instance [29, Example 3 page 11]) for* $\mathcal{X} = R$*,* $d(x,y) = |x-y|$*,* $\mathcal{C}(x,y) = d(x,y)$ *(and also* $d(x,y) + \mathcal{F}(x) + G(y)$ *with* $\mathcal{F}(x)$ *smooth). One cannot use previous theories because setting* $\mathcal{E}(t,u) = \mathcal{C}(u,x_t) = |u - x_t|$ *the derivative with respect to* t *exists except when* $u = x_t$*, or it is exactly there that it should be used (see [17, equation 2.3] and the curve* $x_t = t$*). See also section 3.1 below.*

As the previous works concerned a different setting we do not claim that this contribution is a generalization but rather an extension of these results for our situation.

When $\mathcal{C}$ has further convexity properties one can adapt the EVI formulation as in [29, 30]. We introduce the following assumption (corresponding to [1, Assumption 4.24 page 77]):

($\mathbf{A}_2$) Suppose $\mathcal{C}$ is a lower semicontinuous functional with respect to the first argument and there exists $\lambda \in \mathbb{R}$ such that for any $x_0, x_1, z, v \in \mathcal{X}$ there exists a curve γ connecting x_0 and x_1 such that for all $s \in [0,1]$:

$$\mathcal{C}(\gamma(s), z) \leq (1-s)\mathcal{C}(x_0, z) + s\mathcal{C}(x_1, z) - \lambda\frac{(1-s)s}{2} d^2(x_0, x_1). \tag{19}$$

$$d^2(\gamma(s), v) \leq (1-s)d^2(x_0, v) + sd^2(x_1, v) - (1-s)sd^2(x_0, x_1). \tag{20}$$

In general a function (of one or several variables) that satisfies (19)-(20) with respect to one of its variables is called λ-convex in that variable.

Definition 3 (EVI equilibrium flow) *An absolutely continuous curve* $(x_t)_{t\in[0,T]}$ *is called an EVI-equlibrium flow starting from* $\bar{x}$ *if* $\lim_{t\to 0} x_t = \bar{x}$ *and :*

$$\mathcal{C}(x_t, x_t) + \frac{1}{2}\frac{d}{dt}d^2(x_t, y) + \frac{\lambda}{2}d^2(x_t, y) \leq \mathcal{C}(y, x_t), \forall y, \ a.e. \ t \geq 0. \tag{21}$$

Note that the definition is valid because if $(x_t)_{t\in[0,T]}$ is absolutely continuous then $t \mapsto d^2(x_t, y)$ is also absolutely continuous thus differentiable a.e. with respect to t.

2.3 Convergence of numerical schemes: general situation

Let us denote

$$\mathcal{M}(x, \tau) = argmin_{y\in\mathcal{X}} \frac{d(y, x)^2}{2\tau} + \mathcal{C}(y, x). \tag{22}$$

With this definition the relaxed *best reply* (or *fictitious play*) (see e.g., [7, 8]) numerical scheme in equation (2) can be written as $x^\tau_{k+1} \in \mathcal{M}(x^\tau_k, \tau)$.

We investigate in this section whether when $\tau \to 0$ the set $\{x^\tau_k, k \geq 1\}$ converges to a solution of (1) as defined in (17)-(18) or (21).

In order to work with meaningful objects, we introduce the following assumption which is the analogue of [1, Assumption 4.8 page 67]:

($\mathbf{A}_3$) There exists $\bar{\tau} > 0$ such that for any $\tau \leq \bar{\tau}$ and $\bar{x} \in \mathcal{D}^s(\mathcal{C})$:

$$\mathcal{M}(\bar{x}, \tau) \neq \emptyset. \tag{23}$$

We will make use of the de Giorgi interpolation introduced in [16] under the name of "minimizing movement" which was originally designed to give a unified formulation of many problems such as time-dependent PDE, steepest descent methods, heat equation. The method starts from a set of points obtained by parameter-indexed minimizations and constructs a continuous curve of minimizers ; the construction keeps many important properties of the original set of points such as compactness and equicontinuity and allows to find limit curves (see also [1] and [2, Definition 2.0.6 page 42]).

Assuming that assumption $(\mathbf{A}_3)$ is satisfied, we can define the interpolation *à la* de Giorgi which is a curve $t \in [0,T] \mapsto x_t^\tau$ such that $x_0^\tau = \bar{x}$ and

$$x_t^\tau \in \mathcal{M}(x_{k\tau}^\tau, t-k\tau), \forall t \in]k\tau,(k+1)\tau],\ k \geq 1. \tag{24}$$

For such a map, we define the discrete speed $\mathrm{Dsp}^\tau : [0,+\infty) \to [0,+\infty)$ by

$$\mathrm{Dsp}_t^\tau = \frac{d(x_{n\tau}^\tau, x_{(n+1)\tau}^\tau)}{\tau} \text{ for } t \text{ in } (n\tau,(n+1)\tau),$$

and the discrete slope $\mathrm{Dsl}^\tau : [0,+\infty) \to [0,+\infty)$ by

$$\mathrm{Dsl}_t^\tau = \frac{d(x_{n\tau}^\tau, x_t^\tau)}{t-n\tau} \text{ for } t \text{ in } (n\tau,(n+1)\tau).$$

We will need some additional hypothesis:

$(\mathbf{A}_4)$ For any $c \in \mathbb{R}$, $r > 0$ and $x \in \mathcal{X}$ the set $\{y \in \mathcal{X} | \mathcal{C}(y,y) \leq c, d(y,x) \leq r\}$ is compact.

$(\mathbf{A}_5)$ The slope $|\nabla_1 \mathcal{C}|$ is lower semicontinuous.

$(\mathbf{A}_6)$ Lipschitz property with respect to the second argument: there exists $L > 0$ such that:

$$\mathcal{C}(x,y) \leq Ld(y,z) + \mathcal{C}(x,z),\ \forall x,y,z \in \mathcal{X}. \tag{25}$$

$(\mathbf{A}_7)$ For any absolutely continuous curve $(x_t)_{t\in[a,b]}$:

$$\Upsilon(x,a,b) \leq \liminf_{|\Delta_n|\to 0, y_n \to x, y_n \subset \mathcal{D}^s(\mathcal{C}), \sup_n \int_a^b |\dot{y}_n(t)| < \infty} \Upsilon(\Delta_n; y_n, a, b), \tag{26}$$

where the convergence of the curve y_n to x is in the uniform (on compacts) norm.

$(\mathbf{A}_8)$ There exists $C_L < \infty$ such that for any $x,y,z \in \mathcal{D}^s(\mathcal{C})$:

$$|\mathcal{C}(x,y) + \mathcal{C}(y,z) - \mathcal{C}(x,z) - \mathcal{C}(y,y)| \leq C_L d(x,y) d(y,z). \tag{27}$$

Remark 4 *The assumption* $(\mathbf{A}_8)$ *implies (see Lemma 2 in the appendix):*

$$\Upsilon(x,a,b) = \lim_{|\Delta|\to 0} \Upsilon(\Delta; x, a, b) = \lim_{|\Delta_n|\to 0, y_n \to x, y_n \subset \mathcal{D}^s(\mathcal{C}), \sup_n \int_a^b |\dot{y}_n(t)| < \infty} \Upsilon(\Delta_n; y_n, a, b), \tag{28}$$

which implies $(\mathbf{A}_7)$.

In particular note that $\mathcal{C}(x,y) = d(x,y)$ satisfies $(\mathbf{A}_7)$ *but does* **not** *satisfy* (28) *(thus neither* $(\mathbf{A}_8)$*) while $\mathcal{C}(x,y) = d^2(x,y)$ satisfies* $(\mathbf{A}_8)$ *with $C_L = 2$. On the other hand for $\mathcal{X} = \mathbb{R}$ the mapping $(x,y) \mapsto \mathcal{C}(x,y) = \int_0^x \int_0^y f(s,t)dsdt$, with f a bounded function will also satisfy assumption* $(\mathbf{A}_8)$*.*

Assumptions $(\mathbf{A}_1)$,$(\mathbf{A}_4)$, $(\mathbf{A}_5)$ are classical (see for instance [1, Assumption 4.13 page 69]) and $(\mathbf{A}_7)$ belongs to the same class; on the other hand, the assumption $(\mathbf{A}_6)$ is the analogue of [29, assumption A4 page 11] through the correspondence in equation (6).

The properties of the curves obtained by the numerical scheme (2) are detailed in the following result.

Theorem 1 *Let $\mathcal{C}$ satisfying assumptions* $(\mathbf{A}_1)$*,* $(\mathbf{A}_3)$*,* $(\mathbf{A}_4)$*,* $(\mathbf{A}_5)$*,* $(\mathbf{A}_6)$ *and* $(\mathbf{A}_7)$*. Take $x_0 \in \mathcal{D}^s(\mathcal{C})$. Then the set of curves $\{(x_t^\tau)_{t\in[0,T]}; \tau \leq \bar{\tau}\}$ defined in* (24) *is relatively compact in the set of curves in $\mathcal{X}$ with local uniform convergence and any limit curve is an EDI equilibrium flow in the sense of Definition 2.*

Proof We follow, when possible, the proof of [1, Theorem 4.14 page 69]; first we proceed as in [1, Theorem 4.9 page 67] and note that for fixed $x \in \mathcal{X}$, $\tau \leq \bar{\tau}$ and $x_\tau \in \mathcal{M}(x,\tau)$, the map $\tau \mapsto \frac{d(x_\tau,x)^2}{2\tau} + \mathcal{C}(x_\tau,x)$, is locally Lipschitz and

$$\frac{d}{d\tau}\left(\frac{d(x_\tau,x)^2}{2\tau} + \mathcal{C}(x_\tau,x)\right) = -\frac{d^2(x,x_\tau)}{2\tau^2}. \tag{29}$$

The proof of this identity follow precisely the reference and the details are left to the reader.

The minimal property of $x^\tau_{(k+1)\tau}$ implies that for any k: $\mathcal{C}(x^\tau_{(k+1)\tau}, x^\tau_{k\tau}) - \mathcal{C}(x^\tau_{k\tau}, x^\tau_{k\tau}) \leq 0$; combined with (25) this shows that $\mathcal{C}(x^\tau_{(k+1)\tau}, x^\tau_{(k+1)\tau}) < \infty$.

On the other hand, as in [1, Lemma 4.10 page 67] we obtain (again we skip the details) that $\tau \mapsto d(x,x_\tau)$ is non decreasing and

$$|\nabla_1 \mathcal{C}|\,(x_\tau,x) \leq \frac{d(x,x_\tau)}{2\tau}. \tag{30}$$

Integrating (29) we obtain the following identity:

$$\sum_{k=n}^{m-1} \mathcal{C}(x^\tau_{(k+1)\tau}, x^\tau_{k\tau}) - \mathcal{C}(x^\tau_{k\tau}, x^\tau_{k\tau}) + \frac{1}{2}\int_{n\tau}^{m\tau} (\mathrm{Dsp}^\tau_r)^2\, dr + \frac{1}{2}\int_{n\tau}^{m\tau} (\mathrm{Dsl}^\tau_r)^2\, dr = 0, \tag{31}$$

or equivalently, denoting Δ^τ the division of $[0,T]$ containing the points $k\tau$ $(1 \le k \le T/\tau)$:

$$\Upsilon(\Delta^\tau, x^\tau, n\tau, m\tau) + \frac{1}{2}\int_{n\tau}^{m\tau} (\mathrm{Dsp}_r^\tau)^2\, dr + \frac{1}{2}\int_{n\tau}^{m\tau} (\mathrm{Dsl}_r^\tau)^2\, dr = 0. \tag{32}$$

In particular for any $k \in \mathbb{N}^*$:

$$\begin{aligned}
&\left(\sum_{\ell=0}^{k} d(x_{\ell\tau}^\tau, x_{\ell+1\tau}^\tau)\right)^2 \le \left(\int_0^{(k+1)\tau} |\mathrm{Dsp}_r^\tau|\, dr\right)^2 \le (k+1)\tau \int_0^{(k+1)\tau} |\mathrm{Dsp}_r^\tau|^2\, dr \\
&\le -2(k+1)\tau\Upsilon(\Delta^\tau; x^\tau, 0, (k+1)\tau) \\
&\le -2(k+1)\tau\left(\mathcal{C}(\bar{x},\bar{x}) - \mathcal{C}(x_{(k+1)\tau}^\tau, x_{(k+1)\tau}^\tau) + L\sum_{\ell=0}^{k} d(x_{\ell\tau}^\tau, x_{\ell+1\tau}^\tau)\right).
\end{aligned} \tag{33}$$

This shows that in particular $\sum_{\ell=0}^{k} d(x_{\ell\tau}^\tau, x_{\ell+1\tau}^\tau)$ is bounded by a constant depending on T, L, $\mathcal{C}(\bar{x},\bar{x})$ and C_1 but independent on τ. On the other hand for $t \in]k\tau, (k+1)\tau]$, $t \le T$:

$$d^2(x_t^\tau, \bar{x}) \le \left(d(x_t^\tau, x_{k\tau}^\tau) + \sum_{\ell=0}^{k-1} d(x_{\ell\tau}^\tau, x_{\ell+1\tau}^\tau)\right)^2 \le \left(\sum_{\ell=0}^{k} d(x_{\ell\tau}^\tau, x_{\ell+1\tau}^\tau)\right)^2,$$

where we used that $\tau \mapsto d(x_t^\tau, x_{k\tau}^\tau)$ is non decreasing. Thus the set of curves $\{(x_t^\tau)_{t\in[0,T]}; \tau \le \bar{\tau}\}$ is uniformly bounded with respect to τ. As a by-product $\Upsilon(\Delta^\tau; x^\tau, n\tau, m\tau)$ is bounded uniformly with respect to τ, and $n, m \le T/\tau$.

A similar estimation starting from:

$$d^2(x_{n\tau}^\tau, x_{n\tau}^\tau) \le \left(\int_{n\tau}^{m\tau} |\mathrm{Dsp}_r^\tau|\, dr\right)^2 \le (n\tau - m\tau)\int_{n\tau}^{m\tau} |\mathrm{Dsp}_r^\tau|^2\, dr, \tag{34}$$

allows to see that the set of curves is also equicontinuous. By Arzela-Ascoli one obtains the relative compactness.

Let now $\tau_n \downarrow 0$ and $(x_t)_{t\in[0,T]}$ a limit curve of $\{(x_t^{\tau_n})_{t\in[0,T]}; n \ge 1\}$. From (34) one obtains that $(x_t)_{t\in[0,T]}$ is absolutely continuous and for $t \le s \le T$:

$$\int_t^s |\dot{x}_r|^2\, dr \le \liminf_{n\to\infty} \int_t^s |\mathrm{Dsp}_r^{\tau_n}|^2\, dr. \tag{35}$$

On the other hand, from the lower semicontinuity of $|\nabla_1\mathcal{C}|$ and (30) we obtain

$$|\nabla_1\mathcal{C}|\,(x_t, x_t) \le \liminf_{n\to\infty} |\nabla_1\mathcal{C}|\,(x_t^{\tau_n}, x_{k_n\tau_n}^{\tau_n}) \le \liminf_{n\to\infty} \mathrm{Dsl}_t^{\tau_n}. \tag{36}$$

where k_n is such that $t \in]k_n\tau_n, (k_n+1)\tau_n]$.

Using $(\mathbf{A}_7)$ and similar arguments as in the end of the proof of [1, Theorem 4.14 page 69] (details are left to the reader) one can pass to the limit in (32) and obtain relations (17)-(18). ■

Remark 5 *The previous works (see e.g. [1]) take advantage of a simplification occurring in the treatment of the term* $\sum_{k=n}^{m-1} \mathcal{C}(x^\tau_{(k+1)\tau}, x^\tau_{k\tau}) - \mathcal{C}(x^\tau_{k\tau}, x^\tau_{k\tau})$ *in* (31) *because, when* (4) *is valid, this term reduces to* $E(x^\tau_m) - E(x^\tau_n)$*; the same is true to some extent for time-dependent approaches in [17] (recall equation* (6)*) which exploit the fully-implicit scheme* (7)*, not available in our setting. For us, the absence of such a simplification does not allow to obtain the uniform boundedness of* x^τ *directly and requires care when passing to the limit* $\tau \to 0$*. These points are addressed by a specific treatment of the index* Υ*, some upper bounds on* x^τ *from the assumption* $(\mathbf{A}_6)$ *and, latter in theorem 2, by making use of the hypothesis* $(\mathbf{A}_2)$ *and* $(\mathbf{A}_8)$*, see both proofs for details.*

2.4 Convergence of numerical schemes: convex case

For the convex case more precise information can be obtained and is gathered in the following theorem.

Theorem 2 *Let* $\mathcal{C}$ *satisfying assumptions* $(\mathbf{A}_1)$*,* $(\mathbf{A}_2)$*,* $(\mathbf{A}_4)$ *and* $(\mathbf{A}_6)$*.*

1. *For every* $x \in \mathcal{D}^s(\mathcal{C})$*, the set* $\mathcal{M}(x, \tau)$ *contains exactly one element for any* $\tau \leq 1/\lambda^-$*. Moreover, for* $x_0 \in \mathcal{D}^s(\mathcal{C})$ *the set of curves* $\{(x^\tau_t)_{t\in[0,T]}; \tau \leq \bar{\tau}\}$ *defined in* (24) *is relatively compact and any limit curve* $(x_t)_{t\in[0,T]}$ *is an EVI equilibrium flow in the sense of Definition 3.*

2. *Suppose in addition that for any* $x \in \mathcal{D}^s(\mathcal{C})$ *and* $u, v \in \mathcal{D}^s(\mathcal{C})$ *in a neighborhood of* x*:*

$$|C(u, v) + C(v, u) - C(u, u) - C(v, v)| = O(d(u, v)^2). \tag{37}$$

   Then the (EVI) equilibrium flow is unique.

3. *If the stronger assumption* $(\mathbf{A}_8)$ *(instead of* (37) *) is satisfied, then for any two EVI equilibrium flows* $(x_t)_{t\in[0,T]}$ *and* $(y_t)_{t\in[0,T]}$ *starting from* $\bar{x} \in \mathcal{D}^s(\mathcal{C})$ *and* $\bar{y} \in \mathcal{D}^s(\mathcal{C})$ *respectively:*

$$d(x_t, y_t) \leq e^{-(\lambda - C_L)t} d(\bar{x}, \bar{y}), \ \ a.e., t \geq 0. \tag{38}$$

4. *Under assumption* $(\mathbf{A}_8)$ *if* $\lambda > C_L$ *then the mean field game with cost functional* $\mathcal{C}(\cdot, \cdot)$ *has an unique Nash-MFG equilibrium* $x^{\mathcal{C}}$ *and any equilibrium flow* $(x_t)_{t\geq 0}$ *starting from some* $x(0) = \bar{x} \in \mathcal{D}^s(\mathcal{C})$ *converges exponentially fast to* $x^{\mathcal{C}}$*.*

Proof

Proof of item 1: Following the proof of [1, Theorem 4.25 page 77] it is possible to prove the existence and uniqueness of the minimizers in $\mathcal{M}(x,\tau)$ using the same arguments i.e. taking a minimizing sequence $(x_n)_{n\geq 1}$ and using assumption $(\mathbf{A}_2)$ for $s=1/2$, $x_0=x_n$, $x_1=x_m$, $y=x$, $v=x$. Moreover, the following estimation (replacing [1, equation (69) page 78]) is obtained using arguments in the same proof as soon as $x^\tau\in\mathcal{M}(x,\tau)$:

$$\frac{d^2(x^\tau,y)-d^2(x,y)}{2\tau}+\frac{\lambda}{2}d^2(x^\tau,y)\leq\mathcal{C}(y,x)-\mathcal{C}(x^\tau,x),\ \forall y\in\mathcal{X}. \tag{39}$$

Summing up such estimations it follows that for $t=n\tau<m\tau=s$:

$$\begin{aligned}&\frac{d^2(x_t^\tau,y)-d^2(x_s^\tau,y)}{2(s-t)}+\frac{\lambda\tau}{2(s-t)}\sum_{\ell=n}^{m-1}d^2(x^\tau_{(\ell+1)\tau},y)\\&\leq\frac{\tau}{s-t}\left(\sum_{\ell=n}^{m-1}\mathcal{C}(y,x^\tau_{\ell\tau})-\mathcal{C}(x^\tau_{(\ell+1)\tau},x^\tau_{\ell\tau})\right),\ \forall y\in\mathcal{X}.\end{aligned} \tag{40}$$

We can prove as in theorem 1 that the curves $\{x_t^\tau\}_{\tau\leq 1/\lambda^-}$ form a relatively compact set. Taking $\tau_n\downarrow 0$ one obtains for $0\leq t<s\leq T$:

$$\begin{aligned}&\frac{d^2(x_t,y)-d^2(x_s,y)}{2(s-t)}+\frac{\lambda}{2(s-t)}\int_t^s d^2(x_r,y)dr\\&\leq\frac{1}{s-t}\left(\int_t^s\left[\mathcal{C}(y,x_r)-\mathcal{C}(x_r,x_r)\right]dr\right),\ \forall y\in\mathcal{X},\end{aligned} \tag{41}$$

which is the integral form of the definition 3 of a EVI equilibrium flow. In the passage from (40) to (41) the terms $\sum_{\ell=n}^{m-1}d^2(x^\tau_{(\ell+1)\tau},y)$ and $\sum_{\ell=n}^{m-1}\mathcal{C}(y,x^\tau_{\ell\tau})$ are, thanks to the Lipschitz property of $\mathcal{C}(y,\cdot)$ and $d^2(\cdot,y)$, arbitrarily close to Riemann sums of the corresponding integrals. On the other hand for the term $\sum_{\ell=n}^{m-1}\mathcal{C}(x^\tau_{(\ell+1)\tau},x^\tau_{\ell\tau})$ the inequality is obtained from the Fatou lemma applied to $f_n(r)=\sum_{\ell=n}^{m-1}1_{[\ell\tau_n,(\ell+1)\tau_n[}(r)\mathcal{C}(x^{\tau_n}_{(\ell+1)\tau_n},x^{\tau_n}_{\ell\tau_n})$ (recall that we have a lower bound on $\mathcal{C}$ from assumption $(\mathbf{A}_1)$) and the lower semi-continuity of $\mathcal{C}$.

Proof of item 3 To prove contraction (38) we use the EVI inequality for x_t and y_t and write:

$$\begin{aligned}&\frac{1}{2}\frac{d}{ds}d^2(x_s,y_t)\Big|_{s=t}+\frac{\lambda}{2}d^2(x_t,y_t)\leq\mathcal{C}(y_t,x_t)-\mathcal{C}(x_t,x_t)\ a.e.\ t\geq 0.\\&\frac{1}{2}\frac{d}{ds}d^2(y_s,x_t)\Big|_{s=t}+\frac{\lambda}{2}d^2(y_t,x_t)\leq\mathcal{C}(x_t,y_t)-\mathcal{C}(y_t,y_t)\ a.e.\ t\geq 0.\end{aligned} \tag{42}$$

Adding up the two identities and invoking (27) for $x = z = x_t$, $y = y_t$ we obtain (heuristically for the moment):

$$\frac{1}{2}\frac{d}{dt}d^2(x_t, y_t) + \lambda d^2(x_t, y_t) \leq C_L d^2(x_t, y_t) \text{ a.e. } t \geq 0, \tag{43}$$

and (38) follows. In practice, the manipulations above can be made precise by a doubling of variables argument (see [1, page 80], [31], [9, pages 179, 183]).

Proof of item 2: the proof is similar to that of item 3: let $(x_t)_{t\in[0,T]}$ and $(y_t)_{t\in[0,T]}$ be two EVI equilibrium flows starting from the same point $\bar{x} \in \mathcal{D}^s(\mathcal{C})$ and $C_M > 0$ a constant (large enough). Since the flows are continuous in time, for some small time $T_M > 0$ we have $d(x_t, y_t)$ small enough such that the $O(d(x_t, y_t)^2)$ appearing in (42) via (37) is smaller than $C_M d^2(x_t, y_t)$. Then, a computation similar to (43) shows that:

$$d(x_t, y_t) \leq e^{-(\lambda - C_M)t} d(\bar{x}, \bar{x}) = 0, \ \forall t \in [0, T_M]. \tag{44}$$

This shows that the solution is locally unique on $[0, T_M]$ and then globally unique by a standard maximal interval argument (for instance using Zorn's lemma).

Proof of item 4: the proof is a consequence of item 3 as soon as one proves that the curve $y_t = x^{\mathcal{C}}$ is an EVI equilibrium flow. To prove this, note that the definition (3) of a Nash-MFG equilibrium implies $\mathcal{M}(x^{\mathcal{C}}, \tau) = \{x^{\mathcal{C}}\}$ for all $\tau > 0$. Then $x_t^\tau = x^{\mathcal{C}}$ for any $\tau, t \geq 0$ thus the (only) limit curve is $y_t = x^{\mathcal{C}}$ and by item 1 it is an EVI equilibrium flow. ■

Remark 6 *The uniqueness of the mean field games equilibrium is often a by-product of the monotonicity of the payoff function, as in [34, 33, 12] or in [21, Section 2.8 page 109]; however the monotonicity of the payoff is not a generic property for general MFG. On the other hand recall that MFG with multiple equilibria exist, see the example in section 3.3 below. The item 4 gives a novel route to prove the uniqueness of mean field games equilibriums, which is worth exploring because very few such procedures exist in the literature. The requirements here are*

- *a weak form of convexity: note that assumption* $(\mathbf{A}_2)$ *only uses convexity in the first argument, which is a consequence of the linearity of* $\mathcal{C}$ *with respect to the strategy of the player;*
- *a regularity assumption (hypothesis* $(\mathbf{A}_8)$*);*
- *and the important upper bound* $\lambda > C_L$*.*

Among the three, it is the latter property that seems to be the most restrictive in applications.

3 Applications

For the notations used in this section see the appendix A.

3.1 The treatment of the term $\int_a^b \frac{d}{dt}\mathcal{C}(x_t,\nu)\big|_{\nu=x_t}dt$

We consider first an example that shows how our proposed methodology avoids problems with the term $\int_a^b \frac{d}{dt}\mathcal{C}(x_t,\nu)\big|_{\nu=x_t}dt$ which does not always has a well-defined meaning, even when the dynamics is C^∞ with respect to time.

Set $\mathcal{X}=\mathbb{R}$ and $\mathcal{C}(x,y)=\frac{x^2}{2}+\theta|x-y|$ for some fixed $\theta>0$. After elementary computations we obtain from the definition (3) that the set of Nash equilibriums is the entire interval $[-\theta,\theta]$. In particular there is no uniqueness; we check thus the assumptions of the theorem 1:

- $(\mathbf{A}_1)$ is valid with $C_1=0$;
- $(\mathbf{A}_3)$ is valid because we minimize a continuous coercive real function;
- $(\mathbf{A}_4)$ is valid trivially on $\mathbb{R}$;
- $(\mathbf{A}_5)$ results from the computation $|\nabla_1\mathcal{C}|(x,y)=(x-\theta\cdot 1_{x\le y}+\theta\cdot 1_{x>y})_+$
- $(\mathbf{A}_6)$ is valid with $L=\theta$;
- $(\mathbf{A}_7)$ results from the formula $\Upsilon(x,a,b)=\frac{x_b^2-x_a^2}{2}+length(x,a,b)$ (with $length(x,a,b)$ being the length of the curve $(x_t)_{t\ge 0}$ between $t=a$ and $t=b$); note that $(\mathbf{A}_8)$ does not hold in this case.

We can apply the theorem 1 to conclude that at least an equilibrium flow exists. In fact, a more careful inspection reveals that the flow is:
- when $|x_0|\le\theta$: the flow is the constant curve $x_t=x_0$;
- when $x_0>\theta$: $x_t=\theta+e^{-t}(x_0-\theta)$;
- when $x_0<\theta$: $x_t=-\theta+e^{-t}(x_0+\theta)$.

The dynamics is thus smooth; however the term $\int_a^b \frac{d}{dt}\mathcal{C}(x_t,\nu)\big|_{\nu=x_t}dt$ cannot be given any meaning because of the non-differentiability with respect to t of $|x_t-x_s|$ in $s=t$.

Remark 7 *In this case the assumption* $(\mathbf{A}_2)$ *is also valid thus one can apply the first point of the theorem 2 to conclude to the existence of a EVI - equilibrium flow. However one can propose examples of* $\mathcal{C}$ *being the sum of a smooth, coercive but non-convex function of* x *(as* $(x^2-1)^2$*) and the term* $\theta|x-y|$ *that can still check the hypothesis of theorem 1 but not those of the theorem 2. Here we took the first term to be quadratic in order to have explicit analytic solutions.*

3.2 Multiple agent types games with potential and interaction terms

We follow here the framework in [7] and in particular show how their Corrolary 5.2 (corresponding to their example in equation (5.9) Section 5) can be treated with tools provided in this work. In this situation, the players are assigned to one of several types $x \in X$; once an agent is given a type x it cannot change it. As a consequence, the probability measure on the set X is fixed, let us denote it by $\mu \in \mathcal{P}(X)$. Each agent can take actions y in a set Y. Both X and Y are supposed compact and moreover Y is a convex, closed subset of an Euclidian space. A strategy is a joint probability measure $\gamma \in \mathcal{P}(X \times Y)$ with $\pi^1_{\#}\gamma = \mu$. We assign:

$$\tilde{\mathcal{X}} = \left\{\gamma \in \mathcal{P}(X \times Y) | \pi^1_{\#}\gamma = \mu\right\}. \tag{45}$$

The cost of the action y for the individual of type $x \in X$ when everybody else is doing γ is $C(x, y, \gamma)$. Denote $\mathcal{C}(\gamma_1, \gamma_2) = \int_{X\times Y} C(x, y, \gamma_2)\gamma_1(dx, dy)$. The Nash equilibrium (termed Cournot-Nash in [7]) is defined as in equation (3).

Note that by the disintegration theorem [2, Theorem 5.3.1 page 121] for any $\gamma \in \tilde{\mathcal{X}}$ there exist a a unique (μ- a.e.) set $(\gamma_x)_{x\in X} \subset \mathcal{P}(Y)$ such that for any Borel function f:

$$\int_{X\times Y} f(x, y)\gamma(dx, dy) = \int_X \left(\int_Y f(x, y)\gamma_x(dy)\right) \mu(dx). \tag{46}$$

When there is no ambiguity, we will use the notation $(\gamma_x)_{x\in X} \subset \mathcal{P}(Y)$ to designate the disintegration of the measure γ along the elements $x \in X$.

The measure γ_x can be interpreted as the strategy of the agents of type x.

Remark 8 *In [7] a specific situation is considered when the cost does not depend on the type of other agents choosing a given action in Y but only on how many they are i.e., on the measure $\pi^2_{\#}\gamma$. In this case $C(x, y, \gamma) = C(x, y, \pi^2_{\#}\gamma)$.*

For $p \geq 1$ take $\nu \in \mathcal{P}_p(Y)$ (arbitrary but fixed) and define

$$\mathcal{X}_p = \left\{\gamma \in \tilde{\mathcal{X}} \,\middle|\, \int_X \mathbb{W}_p(\gamma_x, \nu)^p \mu(dx) < \infty\right\}. \tag{47}$$

Note that the definition of $\mathcal{X}_p$ does not depend on ν. When we do not mention the p index we mean $p = 2$, i.e, $\mathcal{X} = \mathcal{X}_2$. We introduce on $\mathcal{X}_p$ the distance:

$$\forall\ \gamma_0, \gamma_1 \in \mathcal{X}_p :\ d_{\mathcal{X},p}(\gamma_0, \gamma_1) = \left(\int_X \mathbb{W}_p(\gamma_{0,x}, \gamma_{1,x})^p \mu(dx)\right)^{1/p}. \tag{48}$$

Note that from (70) using the Holder inequality we obtain a similar relation for $d_{\mathcal{X},p}(\gamma_0,\gamma_1)$:

$$\forall 1 \le p_1 \le p_2,\ \ \gamma_0,\gamma_1 \in \mathcal{X}_{p_1} : d_{\mathcal{X},p_1}(\gamma_0,\gamma_1) \le d_{\mathcal{X},p_2}(\gamma_0,\gamma_1). \tag{49}$$

With these provisions we can state the following result (compare with [7, Corrolary 5.2 page 425]):

Proposition 1 *Suppose Y is the closure of some open set in $\mathbb{R}^d$. Let V_0 be a smooth function satisfying the hypothesis of Lemma 1 (in the appendix) and $\phi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ a function with continuous second derivatives everywhere. Then, if $\epsilon > 0$ is small enough, the game with payoff*

$$C(x,y,\gamma) = \frac{|x-y|^2}{2} + V_0(y) + \epsilon \int_{X\times Y} \phi(y,y_2)\gamma(dx,dy_2), \tag{50}$$

admits a unique Nash-MFG equilibrium and any EVI-equilibrium flow (as in theorem 2) converges exponentially fast to it.

Proof We derive first a duality formula for the distance $d_{\mathcal{X},1}$; for any $g : X \times Y \to \mathbb{R}$, $\gamma_0,\gamma_1 \in \mathcal{X}_1$, we obtain from (71):

$$\begin{aligned}\int_{X\times Y} g(x,y)(\gamma_0(dx,dy) - \gamma_1(dx,dy)) &= \int_X \left(\int_Y g(x,y)\left[\gamma_{0,x}(dy) - \gamma_{1,x}(dy)\right]\right)\mu(dx)\\ \le \int_X \|g(x,\cdot)\|_{Lip} \cdot \mathbb{W}_1(\gamma_{0,x},\gamma_{1,x})\mu(dx) &\le \left[\sup_{x\in X}\|g(x,\cdot)\|_{Lip}\right]\cdot d_{\mathcal{X},1}(\gamma_0,\gamma_1).\end{aligned} \tag{51}$$

We use theorem 2 on $(\mathcal{X}, d_{\mathcal{X},2})$ and check all assumptions.
• First, since V_0 and ϕ are smooth and X, Y are compact one obtains assumption $(\mathbf{A}_1)$.
• The assumption $(\mathbf{A}_2)$ is also verified because the function $C(x,y,\gamma)$ is $1+\lambda_V - \epsilon C_2$-convex ($C_2$ being a constant that only depends on ϕ and Y). In this case the geodesics between γ_0 and γ_1 are the curves γ_t such that $t \mapsto \gamma_{t,x}$ is the generalized geodesic (with some fixed base ν, $\nu \in \mathcal{P}(Y)$, ν being absolutely continuous with respect to the Lebesgue measure dx) having end points $\gamma_{0,x}$ and $\gamma_{1,x}$. Here $(\gamma_{t,x})_{x\in X}$ is the disintegration of γ_t for any $t \in [0,1]$.
• Assumption $(\mathbf{A}_4)$ is obvious since $\mathcal{X}$ is compact.
• Remain only assumptions $(\mathbf{A}_6)$ and $(\mathbf{A}_8)$. We only prove $(\mathbf{A}_8)$, the other being easier to obtain (and uses essentially the same arguments). The idea is to prove the upper bound for the distance $d_{\mathcal{X},1}$ invoking (51) then use the

ordering in (49). Let $\gamma_0, \gamma_1, \gamma_2 \in \mathcal{X}$. Then

$$
\begin{aligned}
&|\mathcal{C}(\gamma_0,\gamma_1) + \mathcal{C}(\gamma_1,\gamma_2) - \mathcal{C}(\gamma_0,\gamma_2) - \mathcal{C}(\gamma_1,\gamma_1)| = \\
&\epsilon \left| \int_{X\times Y} \int_{X\times Y} \phi(y_1,y_2) \left[\gamma_1(dx,dy_2) - \gamma_2(dx,dy_2)\right] \left[\gamma_0(dx,dy_1) - \gamma_1(dx,dy_1)\right] \right| \\
&\leq \epsilon \left\| \int_{X\times Y} \phi(\cdot,y_2) \left[\gamma_1(dx,dy_2) - \gamma_2(dx,dy_2)\right] \right\|_{Lip} d_{\mathcal{X},1}(\gamma_0,\gamma_1), \qquad (52)
\end{aligned}
$$

where (51) has been used. It remains to see that the Lipschitz norm of $\int_{X\times Y} \phi(\cdot,y_2) \left[\gamma_1(dx,dy_2) - \gamma_2(dx,dy_2)\right]$ can be upper bounded by $d_{\mathcal{X},1}(\gamma_1,\gamma_2)$. But for any $y, \tilde{y} \in Y$ with $y \neq \tilde{y}$:

$$
\begin{aligned}
&\left| \int_{X\times Y} \phi(y,y_2) \left[\gamma_1(dx,dy_2) - \gamma_2(dx,dy_2)\right] - \int_{X\times Y} \phi(\tilde{y},y_2) \left[\gamma_1(dx,dy_2) - \gamma_2(dx,dy_2)\right] \right| \\
&= \left| \int_{X\times Y} \left[\phi(y,y_2) - \phi(\tilde{y},y_2)\right] \left[\gamma_1(dx,dy_2) - \gamma_2(dx,dy_2)\right] \right| \\
&\leq |y-\tilde{y}| \cdot d_{\mathcal{X},1}(\gamma_1,\gamma_2) \left[\sup_{y,\tilde{y}\in Y} \left\| \frac{\phi(y,\cdot) - \phi(\tilde{y},\cdot)}{|y-\tilde{y}|} \right\|_{Lip} \right]. \qquad (53)
\end{aligned}
$$

Since ϕ is twice differentiable, assumption $(\mathbf{A}_8)$ follows with $C_L = \epsilon C_3$ where C_3 is a constant depending only on ϕ and the space Y. ■

3.2.1 Congestion Mean Field Games

We consider next a congestion mean field game. Consider a population that has to choose a place on a real line (an adaptation of the "beach example" in P.L. Lions' lectures at Collège de France). The strategy of any individual is a probability law ξ on $\mathbb{R}$. The individuals prefer to be at the origin and the cost increases with the distance from the origin; this is modeled by a term $\int_{\mathbb{R}} F_0(x)\xi(dx)$ with, for instance, $F_0(x) = c_0 x^2/2$, $c_0 > 0$. The individuals dislike congestion which means that there is a penalty to choose a location where the local density of others is large. If the mean field density of others is $m(dx)$ the term that models congestion is $\int_{\mathbb{R}} \int_{\mathbb{R}} \rho_\sigma(y-x) m(dx)\xi(dy)$ where ρ_σ is a given kernel, for instance $\rho_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$, for some $\sigma \geq 0$ (σ models the range of interaction). Thus the cost functional is:

$$
\mathcal{C}(\xi, m) = \int_{\mathbb{R}} F_0(y)\xi(dy) + \int_{\mathbb{R}} \int_{\mathbb{R}} \rho_\sigma(y-x) m(dx)\xi(dy). \qquad (54)
$$

This can be put into the framework of the previous section taking X to be a singleton (all individuals are alike) but here Y is potentially the whole

real line (thus in particular not compact). However, since the individual minimizes his / her cost, any Nash-MFG equilibrium strategy ξ, if it exists, must satisfy

$$supp(\xi) \subset argmin_{y\in\mathbb{R}} \left\{ F_0(y) + \int_{\mathbb{R}} \rho_\sigma(y-x)\xi(dx) \right\}. \tag{55}$$

In particular any y in the support of ξ will satisfy $c_0 y^2/2 \leq 1/(\sigma\sqrt{2\pi})$, thus the support of ξ is included in $Y = \left[-\frac{2}{c_0\sigma\sqrt{2\pi}}, \frac{2}{c_0\sigma\sqrt{2\pi}}\right]$.

Corrolary 1 *Define $\mathcal{X} = (\mathcal{P}_2(Y), \mathbb{W}_2)$ and consider the congestion MFG defined in* (54) *for $\xi, m \in \mathcal{X}$. For c_0 large enough, and any $\bar{x} \in \mathcal{X}$ there exists a unique EVI equilibrium flow* (1) *starting in $\bar{x}$. Moreover the congestion MFG* (54) *admits a unique equilibrium and the EVI-equilibrium flow starting from $\bar{x}$ converges exponentially fast to it.*

Proof We apply the Proposition 1 on $\mathcal{X}$ taking into account that any equilibrium is an element of $\mathcal{X}$. ■

Remark 9 *When $\sigma \to 0$ the equilibrium is the semicircle law [3]: $\xi(x) = \frac{2}{\pi r_0^2}(r_0^2 - x^2)_+ dx$ for some constant $r_0 > 0$. Note however that the approach above does not guarantee convexity for any choice of c_0 and in particular when $\sigma \to 0$ no value of c_0 is large enough. On the contrary, a metric similar to that in section 3.4 is* 0*-convex for any c_0 and allows to use theorem 2.*

Remark 10 *Other choices of functions F_0 can also be treated with the above methodology (keeping the coercivity at infinity and λ-convexity). On the other hand, the congestion term $\int_{\mathbb{R}}\int_{\mathbb{R}} \rho_\sigma(y-x)m(dx)\xi(dy)$ may also be modified but it should remain compatible with the assumptions* $(\mathbf{A}_6)$ *and* $(\mathbf{A}_8)$.

3.3 Non-smooth interaction Mean Field Games with multiple equilibria

Consider now the cost

$$\mathcal{C}(\xi,\eta) = \int_{\mathbb{R}} F_0(x)\xi(dx) + \int_{\mathbb{R}}\int_{\mathbb{R}} |x-y|\xi(dx)\eta(dy). \tag{56}$$

This is a MFG model similar to that in sections 3.1 and 3.2.1 where individuals prefer to be at the origin but here they also prefer to be close together.

With the same technique as in (55) we can prove that it is enough to restrict to the domain $[-R, R]$ with $R = \sqrt{\frac{2M_1(\xi)}{c_0}}$, where $M_1(\xi) = \int_{\mathbb{R}} |x|\xi(dx)$. But then

$$M_1(\xi) = \int_{\mathbb{R}} |x|\xi(dx) = \int_{-R}^{R} |x|\xi(dx) \leq R = \sqrt{\frac{2M_1(\xi)}{c_0}}, \tag{57}$$

thus $R \leq 2/c_0$.

Corrolary 2 *Consider $\Omega = [-2/c_0, 2/c_0]$, $(\mathcal{X}, d) = (\mathcal{P}_2(\Omega), \mathbb{W}_2)$ and the cost functional:*

$$\mathcal{C}(\xi, \eta) = \int_{\Omega} F_0(x)\xi(dx) + \int_{\Omega}\int_{\Omega} |x - y|\xi(dx)\eta(dy). \tag{58}$$

Then for any $x_0 \in \mathcal{X}$ the set of curves $\{(x_t^\tau)_{t\in[0,T]}; \tau \leq \bar{\tau}\}$ defined in (24) is relatively compact and any limit curve $(x_t)_{t\in[0,T]}$ is an EVI equilibrium flow in the sense of the Definition 3.

Proof We start checking the assumptions of the theorem 2:
• $(\mathbf{A}_1)$ is verified with $C_1 = 0$;
• $(\mathbf{A}_2)$ is verified with $\lambda = c_0$ taking as interpolating curves the generalized geodesics with some arbitrary, fixed, base; note that there is no restriction on c_0 except strict positivity, in particular it does need to be large;
• $(\mathbf{A}_6)$ is verified for $L = 1$, using arguments similar to those in the proof of Proposition 1 (i.e., bounding first by the $\mathbb{W}_1$ distance);
• $(\mathbf{A}_4)$ is also satisfied. ■

Note that $(\mathbf{A}_8)$ is not valid here. Based on section 3.1 we do not expect to have uniqueness of the Nash-MFG equilibria. Indeed one can check that any δ_{x^*} with $|x^*| \leq 1/c_0$ is a Nash-MFG equilibrium:

$$\begin{aligned}\forall \xi \in \mathcal{X} : \mathcal{C}(\xi, \delta_{x^*}) &= \int_{\Omega} (F_0(x) + |x - x^*|)\xi(dx) \geq \int_{\Omega} F_0(x^*)\xi(dx) \\ &= F_0(x^*) = \int_{\Omega} (F_0(x) + |x - x^*|)\delta_{x^*}(dx) = \mathcal{C}(\delta_{x^*}, \delta_{x^*}).\end{aligned} \tag{59}$$

3.4 Vaccination Mean Field Games

An important class of MFG examples that go beyond potential or interaction terms are vaccination mean field games (see [18, 4, 5, 32] and more generally [10, 20, 46] for vaccination coverage dynamics). In this framework the individual can choose his / her vaccination strategy which is a probability law

set on $\Omega = [0,T] \cup \{\infty\}$ (the point at infinity meaning no vaccination). Thus $\mathcal{X}$ is the space of probability measures on Ω; we can also see it as a subset of $\dot{H}^{-1}(\Omega)$ (the dual of the space of zero-mean $H^1(\Omega)$-Sobolev functions) with metric:

$$d^2(m_1,m_2) = \sum_{\omega\in\{0,T,\infty\}} [m_1(\{\omega\}) - m_2(\{\omega\})]^2 + \sup_{\|\nabla\phi\|_{L^2([0,T];\mathbb{R})}\leq 1,\ \int_0^T \phi = 0} \left\{ \int_0^T \phi(m_1 - m_2) \right\}^2. \tag{60}$$

This metric can also be given another expression (see [43, Section 5.5.2 page 210]: denote by u_{m_1,m_2} the solution on $]0,T[$ of the Laplace problem $-\Delta u_{m_1,m_2} = m_1 - m_2$ with homogeneous Neumann boundary conditions $\frac{\partial u}{\partial n} = 0$ in 0 and T; note that u_{m_1,m_2} depends linearly of $m_1 - m_2$. Then:

$$d^2(m_1,m_2) = \left(\sum_{\omega\in\{0,T,\infty\}} [m_1(\{\omega\}) - m_2(\{\omega\})]^2 \right) + \|\nabla u_{m_1,m_2}\|^2_{L^2([0,T];\mathbb{R})}. \tag{61}$$

The cost can be expressed, for some functional $F : (\mathcal{X}, d) \to C(\Omega;\mathbb{R})$, as

$$\mathcal{C}(\xi, m) = \int_\Omega F(m)(x)\xi(dx). \tag{62}$$

The functional F is highly non-linear and depends on the solution of a system of measure-driven ODEs, see [32, equations (1) and (4)] for details. More precisely, for some constants $\beta, \gamma > 0$, $0 < r_V < r_I < \infty$, $S_0, I_0 \geq 0$, $S_0 + I_0 < 1$, consider the system (see also [9, Theorem 10.2.3 page 246] for the well-posedness):

$$\begin{cases} dS^m(t) = -\beta S^m(t) I^m(t) dt - dm(t), & S^m(0^-) = S_0, \\ dI^m(t) = \Big(\beta S^m(t) - \gamma\Big) I^m(t) dt, & I^m(0^-) = I_0, \\ d\varphi^m(t) = \beta I^m(t)(1 - \varphi^m(t)) dt, & \varphi^m(0^-) = 0, \end{cases} \tag{63}$$

where m is prolonged by 0 on $]T,\infty[$. Note that $S^m(t)$, $S^m(t)$ and $\varphi^m(t)$ take values in $[0,1]$. The cost is:

$$\mathcal{C}(\xi, m) = -\xi(\{\infty\}) r_I \varphi^m(\infty) + \int_0^T \Big(r_V + (r_I - r_V)\varphi^m(t)\Big)\xi(dt), \tag{64}$$

and thus

$$F(m)(t) = r_V + (r_I - r_V)\varphi^m(t) \text{ for } t < \infty,\ F(m)(\infty) = -r_I \varphi^m(\infty). \tag{65}$$

Corrolary 3 *Consider $\mathcal{X}$ as above and MFG with cost functional* (64)*. Then for any $x_0 \in \mathcal{X}$ the set of curves $\{(x_t^\tau)_{t\in[0,T]}; \tau \leq \bar{\tau}\}$ defined in* (24) *converges when $\tau \to 0$ to a limit curve $(x_t)_{t\in[0,T]}$ which is the unique EVI equilibrium flow in the sense of Definition 3 starting from x_0.*

Proof We check the hypothesis of items 1 and 2 in the theorem 2:
• F is bounded from below which implies hypothesis $(\mathbf{A}_1)$.
• we move next to the assumption $(\mathbf{A}_2)$. Note that there is no apparent convexity in the function $x \mapsto F(m)(x)$. But, the cost $\mathcal{C}$ is linear in the first variable with respect to affine combinations of measures. Therefore we choose as curves in the assumption $(\mathbf{A}_2)$ the segments $\gamma(t) = (1-t)\gamma(0) + t\gamma(1)$. With this choice the distance (61) satisfies the equation (20) (in fact we have equality) while $\mathcal{C}$ will satisfy (19) with $\lambda = 0$ thus assumption $(\mathbf{A}_2)$ is valid.
• Assumption $(\mathbf{A}_4)$ is immediate, the space being compact.
• Assumptions $(\mathbf{A}_6)$ follows immediately after noting that $\frac{d}{dt}F(m)(t)$ is bounded independently on m and t because $\frac{d}{dt}\varphi^m(t)$ is.
• Assumption $(\mathbf{A}_8)$ requires longer computations and is left as an exercise for the reader. ■

Remark 11 *Within the vaccination MFG framework one can encounter also different, more exotic, metric spaces, see for instance [27] where:*

$$\mathcal{X} = \{u : [0,\infty[\to \mathbb{R} | u(0) = 0, \forall b \geq a \geq 0 : |u(a) - u(b)| \leq |a-b|, u(b) \geq u(a)\}, \tag{66}$$

the metric $d^2(f,g) = \sum_{n\in\mathbb{N}^*} \frac{\|f-g\|^2_{L^2([0,n])}}{2^n}$, and the cost functional

$$\mathcal{C}(\xi, m) = \int_{\mathbb{R}} F(m)(t)\dot{u}(t)e^{-rt}dt, \tag{67}$$

for some functional $F : \mathcal{X} \to C(\mathbb{R};\mathbb{R})$.

In all cases, the numerical simulations show that the algorithm (2) *converges to a MFG equilibrium.*

Acknowledgements

G.T. acknowledges support from the Agence Nationale de la Recherche (ANR), projects *EMAQS* (ANR-2011-BS01-017-01), CINE-PARA and MFG (*ANR-16-CE40-0015-01*).

A Notations

We recall below some notations and results used in the paper.

For any set Ω we denote by $\mathcal{P}(\Omega)$ the ensemble of probability laws on Ω. For a general application f defined on Ω with values in a measure space, $f_{\#}\nu$ is the push-forward (image) measure of $\nu \in \mathcal{P}(\Omega)$ through f characterized by $f_{\#}\nu(A) = \nu(f^{-1}(A))$ for any measurable set A in the image. When Ω is a metric space, for $p \geq 1$ and some $x_0 \in \Omega$:

$$\mathcal{P}_p(\Omega) = \left\{ \nu \in \mathcal{P}(\Omega) \left| \int_\Omega d(x, x_0)^p \nu(dx) < \infty \right. \right\}, \tag{68}$$

i.e., $\mathcal{P}_p(\Omega)$ is the set of probability laws on Ω with finite p-th moment. Note that the definition does not depend on the choice of x_0.

When Ω is a tensor product $\Omega = \Omega_1 \times \Omega_2 \times ... \times \Omega_M$, and $\gamma \in \mathcal{P}(\Omega)$ we denote $\pi^1_{\#}\gamma \in \mathcal{P}(\Omega_1)$ the first marginal, $\pi^2_{\#}\gamma \in \mathcal{P}(\Omega_2)$ the second marginal, $\pi^{1,2}_{\#}\gamma \in \mathcal{P}(\Omega_1 \times \Omega_2)$ the marginal with respect to the first two coordinates, etc.

The set $\mathcal{P}_p(\Omega)$ can be given the structure of a metric space with the p-Wasserstein metric [43, Section 5] denoted $\mathbb{W}_p$ with

$$\mathbb{W}_p(\nu_1, \nu_2) = \left(\inf \left\{ \int_{\Omega\times\Omega} d(x, y)^p \gamma(x, y) \left| \gamma \in \mathcal{P}(\Omega \times \Omega), \pi^1_{\#}\gamma = \nu_1, \pi^2_{\#}\gamma = \nu_2 \right. \right\} \right)^{1/p}. \tag{69}$$

The set of γ that realize the "inf" in (69) is denoted $\Gamma_o(\nu_1, \nu_2)$.

Proving that $\mathbb{W}_p$ is a distance can be performed with usual techniques [43, Lemmas 5.4 and 5.5 page 182]; also standard is to note that these distances are ordered i.e.

$$\mathbb{W}_{p_1}(\gamma_0, \gamma_1) \leq \mathbb{W}_{p_2}(\gamma_0, \gamma_1), \ \forall\ 1 \leq p_1 \leq p_2, \ \forall \gamma_0, \gamma_1 \in \mathcal{P}_{p_1}(\Omega). \tag{70}$$

Recall that the norm $\|f\|_{Lip}$ of a function f is its smallest Lipschitz constant (and $+\infty$ if no such constant exists). Then (see [45, Remark 6.5 page 95]):

$$\mathbb{W}_1(\gamma_0, \gamma_1) = \sup \left\{ \int_\Omega f(\gamma_0 - \gamma_1) \left| \|f\|_{Lip} \leq 1 \right. \right\}. \tag{71}$$

In order to exploit convexity, we need to define the notion of (generalized) geodesics (see [2, Definition 9.2.2. page 207]): let $\nu_0, \nu_1, \nu \in \mathcal{P}_p(\Omega)$ and $\xi \in \mathcal{P}(\Omega \times \Omega \times \Omega)$ such that $\pi^{1,2}_{\#}\xi \in \Gamma_0(\nu, \nu_0)$, $\pi^{1,3}_{\#}\xi \in \Gamma_0(\nu, \nu_1)$. The generalized geodesic between ν_0 and ν_1 with base ν is the curve $(\nu^\xi(t))_{t\in[0,1]} \subset \mathcal{P}_p(\Omega)$ with

$$\nu^\xi(t) = [(y_1, y_2, y_3) \mapsto ((1 - t)y_2 + ty_3)]_{\#}\, \xi. \tag{72}$$

In particular $\nu^\xi(0) = \nu_0$, $\nu^\xi(1) = \nu_1$. When $\nu = \nu_0$ the generalized geodesic is a (ordinary) geodesic in the space $\mathcal{P}_p(\Omega)$; in this case we do not mention the base any more.

This definition allows to state the following result (for the proof see [2, Lemma 9.2.1 page 206, Proposition 9.3.2 page 210]):

Lemma 1 *Take $p = 2$, $\lambda_V \geq 0$ and $V : \Omega \to \mathbb{R}$ be a λ_V-convex function in the sense that for any $y_1, y_2 \in \Omega$, $t \in [0, 1]$:*

$$V((1-t)y_1 + ty_2) \leq (1-t)V(y_1) + tV(y_2) - \frac{\lambda_V}{2} d^2(y_1, y_2). \tag{73}$$

Then the functional $\mathcal{V} : \mathcal{P}_p(\Omega) \to \mathbb{R}$:

$$\mathcal{V}(\gamma) = \int_\Omega V(y)\gamma(dy), \tag{74}$$

is λ_V-convex on $(\mathcal{P}_p(\Omega), \mathbb{W}_p)$ in the sense of the assumption $(\mathbf{A}_2)$ on the generalized geodesics with base ν for any $\nu \in \mathcal{P}_p(\Omega)$.

A.1 Relationship between assumptions $(\mathbf{A}_8)$ and $(\mathbf{A}_7)$

We prove below the identity (28) in Remark 4.

Lemma 2 *Suppose that $\mathcal{C}$ satisfies assumptions $(\mathbf{A}_1)$, $(\mathbf{A}_6)$ and $(\mathbf{A}_8)$ and $\alpha \mapsto \mathcal{C}(\alpha, \alpha)$ is continuous over $\mathcal{D}^s(\mathcal{C})$. Then (28) holds true.*

Proof . **Step 1** First we prove that $(\alpha, \beta) \mapsto \mathcal{C}(\alpha, \beta)$ is continuous over $\mathcal{D}^s(\mathcal{C}) \times \mathcal{D}^s(\mathcal{C})$. Let $\alpha_n \to \alpha$ and $\beta_n \to \beta$ be two converging sequences. Then

$$\begin{aligned}
&|\mathcal{C}(\alpha_n, \beta_n) - \mathcal{C}(\alpha, \beta)| \leq |\mathcal{C}(\alpha_n, \beta_n) - \mathcal{C}(\alpha_n, \beta)| + |\mathcal{C}(\alpha_n, \beta) - \mathcal{C}(\alpha, \beta)| \\
&\leq Ld(\beta_n, \beta) + |\mathcal{C}(\alpha, \alpha_n) + \mathcal{C}(\alpha_n, \beta) - \mathcal{C}(\alpha, \beta) - \mathcal{C}(\alpha_n, \alpha_n)| + \\
&|\mathcal{C}(\alpha_n, \alpha_n) - \mathcal{C}(\alpha, \alpha_n)| \leq Ld(\beta_n, \beta) + C_L d(\alpha_n, \beta) d(\alpha_n, \alpha) \\
&+|\mathcal{C}(\alpha_n, \alpha_n) - \mathcal{C}(\alpha, \alpha)| + |\mathcal{C}(\alpha, \alpha) - \mathcal{C}(\alpha, \alpha_n)| \\
&\leq (L + C_L d(\alpha_n, \beta))\, Ld(\alpha_n, \alpha) + Ld(\beta_n, \beta) + |\mathcal{C}(\alpha_n, \alpha_n) - \mathcal{C}(\alpha, \alpha)|, \qquad (75)
\end{aligned}$$

which, under present hypothesis, tends to zero when $\alpha_n \to \alpha$ and $\beta_n \to \beta$. In particular the continuity of $\mathcal{C}$ implies that, with the notations in $(\mathbf{A}_8)$, for any fixed $\Delta \in \mathcal{S}(a, b)$: $\lim_{n\to\infty} \Upsilon(\Delta; y_n, a, b) = \Upsilon(\Delta; x, a, b)$.

Step 2 Let now y_n be as in the assumption $(\mathbf{A}_7)$ and $\omega(\cdot)$ a joint continuity modulus for the curves $\{x\} \cup \{y_k, k \geq 1\}$ as functions over $[a, b]$ i.e., for any $z \in \{x\} \cup \{y_k, k \geq 1\}$:

$$d(z_{t_1}, z_{t_2}) \leq \omega(|t_2 - t_1|), \ \forall t_1, t_2 \in [a, b]. \tag{76}$$

Let further M_L be a common bound for the length of all curves $z \in \{x\} \cup \{y_k, k \geq 1\}$, which implies $\sum_{k=0}^{\ell-1} d(z_{t_{k+1}}, z_{t_k}) \leq M_L,\ \forall \{t_0, ..., t_\ell\} \in \mathcal{S}(a,b)$.

We invoke the following reformulation of $(\mathbf{A}_8)$: for any curve $z \in \{x\} \cup \{y_k, k \geq 1\}$ and $\{s_0, ..., s_{k-1}, s_k, s_{k+1}, ..., s_\ell\} \in \mathcal{S}(a,b)$:

$$\begin{aligned} &|\Upsilon(\{s_0, ..., s_{k-1}, s_k, s_{k+1}, ..., s_\ell\}; z, a, b) \\ &-\Upsilon(\{s_0, ..., s_{k-1}, s_{k+1}, ..., s_\ell\}; z, a, b)| \leq C_L d(z_{s_{k-1}}, z_{s_k}) d(z_{s_k}, z_{s_{k+1}}). \end{aligned} \tag{77}$$

Repeated application of the inequality (77) for $\Delta_1, \Delta_2 \in \mathcal{S}(a,b)$ allows to obtain for any $z \in \{x\} \cup \{y_k, k \geq 1\}$:

$$|\Upsilon(\Delta_1; z, a, b) - \Upsilon(\Delta_1 \cup \Delta_2; z, a, b)| \leq C_L M_L \omega(|\Delta_1|), \tag{78}$$

and thus

$$|\Upsilon(\Delta_1; z, a, b) - \Upsilon(\Delta_2; z, a, b)| \leq C_L M_L \left(\omega(|\Delta_1|) + \omega(|\Delta_2|)\right). \tag{79}$$

For $z = x$ this means that for $|\Delta| \to 0$ the set $\{\Upsilon(\Delta; x, a, b), \Delta\}$ has a unique limit which has to be $\Upsilon(x, a, b)$ (recall the definition (16)).

Consider now a subsequence of $\{n \geq 1\}$ (which, without loss of generality we can consider to be $\{n \geq 1\}$ itself) such that $\Upsilon(\Delta_n; y_n, a, b)$ is converging to some value J. Then, for fixed $\Delta \in \mathcal{S}(a,b)$:

$$|\Upsilon(\Delta; y_n, a, b) - \Upsilon(\Delta_n; y_n, a, b)| \leq C_L M_L \left(\omega(|\Delta|) + \omega(|\Delta_n|)\right). \tag{80}$$

Passing to the limit when $n \to \infty$ allows to write $|\Upsilon(\Delta; x, a, b) - J| \leq C_L M_L \omega(|\Delta|)$; passing again to the limit when $|\Delta| \to 0$ and taking into account the first, already proven, identity in (28), we obtain $\Upsilon(x, a, b) = J$, hence the conclusion. ■

References

[1] Luigi Ambrosio and Nicola Gigli. *Modelling and Optimisation of Flows on Networks: Cetraro, Italy 2009, Editors: Benedetto Piccoli, Michel Rascle*, chapter A User's Guide to Optimal Transport, pages 1–155. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[2] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures. 2nd ed.* Basel: Birkhäuser, 2nd ed. edition, 2008.

[3] Z. D. Bai and Y. Q. Yin. Convergence to the semicircle law. *Ann. Probab.*, 16(2):863–875, 1988.

[4] Chris T. Bauch and David J. D. Earn. Vaccination and the theory of games. *Proc. Natl. Acad. Sci. USA*, 101(36):13391–13394 (electronic), 2004.

[5] Chris T. Bauch, Alison P. Galvani, and David J. D. Earn. Group interest versus self-interest in smallpox vaccination policy. *Proceedings of the National Academy of Sciences*, 100(18):10564–10567, 2003.

[6] Alain Benoussan, Jens Frehse, and Phillip Yam. *Mean Field Games and Mean Field Type Control Theory*. Springer-Verlag New York, 2013.

[7] A. Blanchet and G. Carlier. Remarks on existence and uniqueness of cournot–nash equilibria in the non-potential case. *Mathematics and Financial Economics*, 8(4):417–433, 2014.

[8] Adrien Blanchet and Guillaume Carlier. From Nash to Cournot-Nash equilibria via the Monge-Kantorovich problem. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 372(2028):20130398, 11, 2014.

[9] Alberto Bressan and Benedetto Piccoli. *Introduction to the mathematical theory of control*, volume 2 of *AIMS Series on Applied Mathematics*. American Institute of Mathematical Sciences (AIMS), Springfield, MO, 2007.

[10] Bruno Buonomo, Alberto d'Onofrio, and Deborah Lacitignola. Global stability of an SIR epidemic model with information dependent vaccination. *Mathematical Biosciences*, 216(1):9 – 16, 2008.

[11] Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A course in metric geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001.

[12] Pierre Cardaliaguet. Notes on Mean Field Games. `https://www.ceremade.dauphine.fr/~cardalia/MFG20130420.pdf`, 2013.

[13] Pierre Cardaliaguet and P. Jameson Graber. Mean field games systems of first order. *ESAIM: COCV*, 21(3):690–722, 2015.

[14] Pierre Cardaliaguet, P. Jameson Graber, Alessio Porretta, and Daniela Tonon. Second order mean field games with degenerate diffusion and local coupling. *Nonlinear Differential Equations and Applications NoDEA*, 22(5):1287–1317, Oct 2015.

[15] Rene Carmona and Francois Delarue. *Probabilistic Theory of Mean Field Games with Applications I-II.* Springer International Publishing, 2017-2018.

[16] Ennio De Giorgi. New problems on minimizing movements. In *Boundary value problems for partial differential equations and applications. Dedicated to Enrico Magenes on the occasion of his 70th birthday*, pages 81–98. Paris: Masson, 1993.

[17] Lucas C. F. Ferreira and Julio C. Valencia-Guevara. Gradient flows of time-dependent functionals in metric spaces and applications to pdes. *Monatshefte für Mathematik*, pages 1–38, 2017.

[18] Paul E. M Fine and Jacqueline A Clarkson. Individual versus public priorities in the determination of optimal vaccination policies. *American Journal of Epidemiology*, 124(6):1012–1020, 1986.

[19] Drew Fudenberg and David K. Levine. *The theory of learning in games*, volume 2 of *MIT Press Series on Economic Learning and Social Evolution*. MIT Press, Cambridge, MA, 1998.

[20] Sebastian Funk, Marcel Salathé, and Vincent A. A. Jansen. Modelling the influence of human behaviour on the spread of infectious diseases: a review. *J. R. Soc. Interface*, 7:1247–1256, 2010.

[21] Diogo A. Gomes, Joana Mohr, and Rafael Rigão Souza. Continuous time finite state mean field games. *Applied Mathematics & Optimization*, 68(1):99–143, 2013.

[22] Diogo A. Gomes, Edgard Pimentel, and Héctor Sánchez-Morgado. Time-dependent mean-field games in the superquadratic case. *ESAIM: COCV*, 22(2):562–580, 2016.

[23] Diogo A. Gomes, Edgard A. Pimentel, and Héctor Sánchez-Morgado. Time-dependent mean-field games in the subquadratic case. *Communications in Partial Differential Equations*, 40(1):40–76, 2015.

[24] Diogo A. Gomes, Edgard A. Pimentel, and Vardan Voskanyan. *Regularity theory for mean-field game systems.* Cham: Springer, 2016.

[25] Minyi Huang, Roland P. Malhamé, and Peter E. Caines. Nash equilibria for large-population linear stochastic systems of weakly coupled agents. In Elkébir Boukas and Roland P. Malhamé, editors, *Analysis, Control and Optimization of Complex Dynamic Systems*, pages 215–252. Springer US, 2005.

[26] Minyi Huang, Roland P. Malhamé, and Peter E. Caines. Large population stochastic dynamic games: closed-loop mckean-vlasov systems and the Nash certainty equivalence principle. *Commun. Inf. Syst.*, 6(3):221–252, 2006.

[27] Emma Hubert and Gabriel Turinici. Nash-MFG equilibrium in a SIR model with time dependent newborn vaccination. https://hal.archives-ouvertes.fr/hal-01389584, October 2016.

[28] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.

[29] E. Kopfer. Gradient flow for the Boltzmann entropy and Cheeger's energy on time-dependent metric measure spaces. *ArXiv e-prints*, November 2016.

[30] E. Kopfer and K.-T. Sturm. Heat Flows on Time-dependent Metric Measure Spaces and Super-Ricci Flows. *ArXiv e-prints*, November 2016.

[31] S. N. Kružkov. First order quasilinear equations with several independent variables. *Mat. Sb. (N.S.)*, 81 (123):228–255, 1970.

[32] Laetitia Laguzet and Gabriel Turinici. Individual vaccination as Nash equilibrium in a SIR model with application to the 2009–2010 influenza A (H1N1) epidemic in France. *Bulletin of Mathematical Biology*, 77(10):1955–1984, 2015.

[33] Jean-Michel Lasry and Pierre-Louis Lions. Jeux à champ moyen. I: Le cas stationnaire. *C. R., Math., Acad. Sci. Paris*, 343(9):619–625, 2006.

[34] Jean-Michel Lasry and Pierre-Louis Lions. Jeux à champ moyen. II: Horizon fini et contrôle optimal. *C. R., Math., Acad. Sci. Paris*, 343(10):679–684, 2006.

[35] Jean-Michel Lasry and Pierre-Louis Lions. Mean field games. *Japanese Journal of Mathematics*, 2(1):229–260, 2007.

[36] Pierre Louis Lions. Cours at Collège de France: Théorie des jeux de champ moyen et applications. `http://www.college-de-france.fr/site/pierre-louis-lions/_audiovideos.htm`, 2007-2011.

[37] Alexander Mielke, Riccarda Rossi, and Giuseppe Savaré. Variational convergence of gradient flows and rate-independent evolutions in metric spaces. *Milan J. Math.*, 80(2):381–410, 2012.

[38] Alexander Mielke, Riccarda Rossi, and Giuseppe Savaré. Nonsmooth analysis of doubly nonlinear evolution equations. *Calc. Var. Partial Differential Equations*, 46(1-2):253–310, 2013.

[39] John Nash. Non-cooperative games. *Ann. of Math. (2)*, 54:286–295, 1951.

[40] Alessio Porretta. On the planning problem for the mean field games system. *Dynamic Games and Applications*, 4(2):231–256, 2014.

[41] Alessio Porretta. Weak solutions to Fokker–Planck equations and mean field games. *Archive for Rational Mechanics and Analysis*, pages 1–62, 2015.

[42] Riccarda Rossi, Alexander Mielke, and Giuseppe Savaré. A metric approach to a class of doubly nonlinear evolution equations and applications. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, 7(1):97–169, 2008.

[43] Filippo Santambrogio. *Optimal transport for applied mathematicians.* Progress in Nonlinear Differential Equations and their Applications, 87. Birkhäuser/Springer, Cham, 2015. Calculus of variations, PDEs, and modeling.

[44] Cédric Villani. *Optimal transport. Old and new*, volume 338 of *Grundlehren der mathematischen Wissenschaften.* Springer, 2009.

[45] Cédric Villani. *Optimal transport, Old and new*, volume 338. Springer-Verlag, Berlin, 2009.

[46] Zhen Wang, Chris T. Bauch, Samit Bhattacharyya, Alberto d'Onofrio, Piero Manfredi, Matjaž Perc, Nicola Perra, Marcel Salathé, and Dawei Zhao. Statistical physics of vaccination. *Physics Reports*, 664:1 – 113, 2016.