Selection by vanishing common noise for potential finite state mean field games

Alekos Cecchina and François Delarue

ABSTRACT
The goal of this paper is to provide a selection principle for potential mean field games on a finite state space and, in this respect, to show that equilibria that do not minimize the corresponding mean field control problem should be ruled out. Our strategy is a tailor-made version of the vanishing viscosity method for partial differential equations. Here, the viscosity has to be understood as the square intensity of a common noise that is inserted in the mean field game or, equivalently, as the diffusivity parameter in the related parabolic version of the master equation. As established in the recent contribution (Bayraktar et al., 2021, J. Math. Pures Appl. 147:98–162), the randomly forced mean field game becomes indeed uniquely solvable for a relevant choice of a Wright-Fisher common noise, the counterpart of which in the master equation is a Kimura operator on the simplex. We here elaborate on (Bayraktar et al., 2021, J. Math. Pures Appl. 147:98–162) to make the mean field game with common noise both uniquely solvable and potential, meaning that its unique solution is in fact equal to the unique minimizer of a suitable stochastic mean field control problem. Taking the limit as the intensity of the common noise vanishes, we obtain a rigorous proof of the aforementioned selection principle. As a byproduct, we get that the classical solution to the viscous master equation associated with the mean field game with common noise converges to the gradient of the value function of the mean field control problem without common noise. We hence select a particular weak solution of the master equation of the original mean field game. Lastly, we establish an intrinsic uniqueness criterion for this solution within a suitable class of weak solutions to the master equation satisfying a weak one-sided Lipschitz inequality.

1. Introduction
The theory of mean field games (MFGs) addresses Nash equilibria within infinite population of rational players subjected to mean field interactions. It has received a lot of attention since the pioneering works of Lasry and Lions [1–3] and of Huang, Caines and Malhamé [4–6]. Earlier works in the field were mostly dedicated to proving the existence of such equilibria in various types of settings, including deterministic or...
stochastic dynamics, stationary or time-inhomogeneous models, continuous or finite state spaces, local or nonlocal couplings. Many of the proofs in this direction go through the analysis of the so-called MFG system, which is a system of two forward and backward partial differential equations (PDEs) (PDEs reducing to mere ODEs for finite state spaces) describing both the dynamics of an equilibrium and the evolution of the cost to a typical player along this equilibrium; see for instance [3, 7, 8] and [9, Chapter 3] for a tiny example, together with the notes and complements in [10, Chapter 3] for more references. Another and slightly more recent object in the field is the master equation which describes the evolution of the value of the game in the form of a PDE set on the space of probability measures. This master equation can be viewed as the analogue of the so-called Nash system for games with finitely many players, which is a system of Hamilton-Jacobi-Bellman (HJB) equations for the value functions of all the players in the game. The passage from a collection of value functions for the players in the finite game to a single value function in the corresponding MFG follows from the symmetry structure of the model. Informally, the connection between the MFG system and the master equation is pretty simple: The MFG system is nothing but the system of characteristics of the master equation. This picture may be made rigorous when the MFG has a unique equilibrium. Provided that the coefficients of the game are smooth enough, the master equation might then be expected to have itself a classical solution, but this turns out to be a challenging question. In all the instances where this guess can be indeed demonstrated (see for instance [9, 11–14] in the continuous setting and [15–17] in the discrete case, together with [7, 18–21] and [10, Chapter 7] for related works), the standard assumption that is used for ensuring uniqueness is the so-called Lasry-Lions monotonicity condition; see [1–3] and [10, Chapter 3] for monotonicity on continuous state spaces and [8, 17] on finite state spaces. Up to some extensions that include another form of monotonicity (e.g., see [10, Chapter 3] or [22]), it is in fact the only main type of uniqueness criterion that exists. Actually, monotonicity has the great advantage of being very robust, meaning that it not only forces uniqueness but also stability of the equilibria; at the same time, it has the drawback of being rather restrictive from a practical point of view. Unfortunately, the master equation becomes poorly understood beyond the monotonous case. In particular, the connection between the MFG system and the master equation takes a dramatic turn whenever equilibria are no longer unique: In the latter case, there may be several possible values for the game; accordingly, classical solutions to the master equation cease to exist and almost nothing is then known on the master equation, except maybe in few examples in which the master equation can be reduced to a one-dimensional PDE.

The goal of our paper is hence to make one new step forward and to address in a more systematic way the following two questions for a suitable class of MFGs without uniqueness:

1. Is it possible to select some of the equilibria of the MFG?
2. Is it possible to select one specific solution of the master equation?

For sure, these two questions are very challenging in full generality. Subsequently, we cannot hope for a class of MFGs that is too big. So far, the typical examples for which
these two questions have been addressed rigorously in the literature are cases where equilibria can be described through a one-dimensional parameter only. In [23–25], the state space can be embedded in $\mathbb{R}$ and the one-dimensional parameter is taken as the mean of the population: [25] addresses a linear-quadratic 1d MFG with Gaussian equilibria and [23, 24] address two examples of MFGs on $\{0, 1\}$ and $\{-1, 1\}$ respectively. We here intend to study a generalization of [23, 24] and to consider MFGs on a finite state space of any cardinality. However, even the latter would remain too much. We thus restrict ourselves to so-called potential games, namely to games whose cost coefficients derive from potentials. As explained in Subsection 2.6 below, this still covers the framework of [23, 24] and, interestingly, this provides an example where equilibria cannot be described by a single parameter.

The great interest of potential games is that they are intrinsically associated with a variational problem, usually referred to as a mean field control problem (MFCP): The MFG indeed reads as a first order condition for the MFCP, meaning that any optimal trajectory of the corresponding MFCP solves the MFG; see for instance [3, 7] and [8, 26] for earlier references on the continuous and discrete settings respectively. In short, the MFCP is here a deterministic control problem with trajectories taking values within the space of probability measures (over the state space supporting the MFG) and the cost functional of which is driven by the potentials of the cost coefficients of the original MFG. Noticeably, this variational interpretation of MFGs has been widely used in the analysis of the MFG system; see for example [27–29]. Here, we want to use it as a way to rule out some of the equilibria of the MFG, namely those that are not minimizers of the MFCP; we provide examples of such equilibria in Subsection 2.6. For sure, we could decide to impose this selection principle arbitrarily but, in the end, this would lack a convincing explanation. Using terminology from game theory, equilibria would be hence selected according to a mere cooperative criterion. Instead, our purpose is to identify a framework in which no other solution can emerge from a true competition between the players. We think that it provides in the end a more satisfactory justification.

Before we say more on different approaches to this selection principle, it might be worth recalling that, intuitively, MFG are to be thought of as asymptotic versions of games with finitely many players; see for instance [3] for an earlier discussion on this question together with [6, 30, 31] for a generic manner to reconstruct approximate equilibria to the finite game from solutions of the MFG. In this respect, the most convincing strategy for justifying a selection principle would certainly consist in proving that the Nash equilibria of the corresponding finite player version of the game converge (in some way) to minimizers of the corresponding MFCP. Actually, this direct approach is precisely what is done in [23, 24] in a specific case where the state space has exactly two elements. However, this turns out to be a difficult strategy since the passage from games with finitely many players to MFGs remains in general rather subtle, regardless of any question of selection; see for instance [9, 32–34] and [15, 16] for several contributions on this matter under the additional Lasry-Lions monotonicity conditions in continuous and discrete settings respectively.

Differently from their mean field counterpart, finite player equilibria are in fact unique in several important examples of non-monotone stochastic differential games
that are studied in the MFG literature. For instance, so is the case when the global noise driving the whole particle system is non-degenerate. Then, uniqueness follows from the well-posedness of the Nash system, which is in turn a consequence of solvability results for uniformly parabolic systems of nonlinear PDEs; see for instance [11, Chapter 6]. In line with this observation, another conceivable strategy for the selection principle is thus to reintroduce some macroscopic diffusion in the mean field setting in order to impose uniqueness and then to take the limit as the corresponding viscosity tends to 0. This macroscopic diffusion should be thought of as a restoration of the global noise that forces the Nash system to be uniformly parabolic in the corresponding finite player stochastic game. To clarify this picture, we may think of $N$ players forced by independent copies of the same noise in the finite game. In that case, the corresponding empirical distribution is indeed subjected to fluctuations of size $1/\sqrt{N}$ within a suitable metric space; see [35–37]. This factor $1/\sqrt{N}$ should be regarded as the intensity of the global noise in the finite player game; obviously, it tends to 0 as $N$ tends to $\infty$. Restoring a small but macroscopic diffusion in the mean field setting is thus a way to reproduce some of the features of the finite player game, but it turns out to be more advantageous (than taking directly the limit $N \to \infty$) since the two mean field and vanishing viscosity limits are then no longer taken simultaneously but successively. Also, working with the mean field limit is easier since any deviation of the reference player has then no influence on the statistical state of the population.

Hence, the approach we implement here may be summarized as follows: We pass to the limit in a randomly forced version of the MFG, the limit being taken as the intensity of the random forcing (called macroscopic diffusion in the previous paragraph) tends to 0. Noticeably, this strategy was used in [25], but in a linear quadratic setting only. In the MFG folklore, this random forcing is usually referred to as a common or systemic noise. Rigorously, it must be understood in the finite version of the MFG as a noise that is indeed common to all the players, in contrast to idiosyncratic noises that are independent and specific to each given player. Differently from the framework with idiosyncratic noises only, the fluctuations of the empirical distribution then become truly stochastic in the limit because of the correlations between the noises. We refer for instance to [9, 38] for two distinct approaches to continuous state MFGs with a common noise; as for the finite state case, we refer, among others, to [39], the key idea of which is to force the finite-player system to have many simultaneous jumps at some random times prescribed by the common noise. The reader may also have a look at [40] which offers a discrete point of view on [9].

The key fact for the selection argument is that the common noise may force uniqueness provided that it satisfies a suitable form of non-degeneracy. Finding such a kind of common noise is however a difficult problem, since it requires to construct a random forcing that has strong enough smoothing properties and that preserves at the same time the geometry of the space of probability measures (no matter the cardinality of the state space). For a complete review, we refer to [41, 42] for continuous state MFGs and to the recent article [43] for finite state MFGs (noticeably, the latter also involves a repulsive forcing at the boundary), bearing in mind that we are not aware of any other result of the same type that we could use for our approach. In brief, the cornerstone in [43] is to design a form of common noise, which we call Wright-Fisher, so that the
The corresponding master equation becomes a system of nonlinear PDEs driven by a so-called Kimura operator and hence enjoys the related Schauder like smoothing estimates established in [44] (we refer to Definition 2.2.1 in the latter reference for a complete definition of a Kimura operator and to [43, Subsection 3.2] for an overview). This paves the way for the following sketch: If we succeed to associate a variational structure to the MFG with a common noise, meaning that the unique equilibrium of the MFG with common noise is also the unique minimizer of some MFCP with common noise, and if we then manage to show that the minimizer of the MFCP with common noise converges in some suitable sense to solutions of the original MFCP without common noise, then we are done! In this prospect, requiring the MFG with common noise to derive itself from a potential structure turns out to be very useful since it permits to verify more easily that the limiting equilibria are themselves minimizers of the MFCP without common noise.

Although it is quite clear, this idea is not so simple to implement: In short, the procedure used in [43] to restore uniqueness in finite state MFGs does not preserve the potential structure. Part of our job here is thus to elaborate on [43] in order to cook up a randomly forced version of the MFG that is uniquely solvable and that derives from a potential; equivalently, the corresponding master equation is required to coincide with the derivative of a suitable parabolic HJB equation on the simplex, the analysis of which is here carried out explicitly by means of the properties of the Kimura operator associated with the common noise. Another task is then to take the limit as the intensity of the common noise tends to zero and to show that the solutions that are selected in this way are indeed minimizers of the MFCP without common noise, hence justifying the selection principle that we figured out. The last step in our program is to make the connection between the selection principle and the master equation. As for the potential MFG with a common noise, we show that the master equation has indeed a unique classical solution and that the latter converges almost everywhere to the gradient of the value function of the MFCP without common noise. Following an earlier work of Kružkov [45], we are then able to prove that this limit is in fact a weak solution to a conservative form of the master equation and that it is the unique one that satisfies in addition a weak one-sided Lipschitz inequality; this conservative form of the master equation was already pointed out in [17]. Remarkably, all our results are consistent with the two-state examples addressed in [23, 24]; see Subsection 2.6 for more details. We also provide more details about the connection with the N-player game in Subsection 2.7.

In the end, even though we manage to carry out our program, it may be objected that our selection principle relies on some a priori choices and that other options might have been conceivable. For example, our preference for a Wright-Fisher common noise may look arbitrary. This is true, but beyond the obvious fact that this solution is somehow dictated by the uniqueness results that exist in the literature, it is also fair to say that it looks a quite natural and reasonable choice. Firstly, one of the main features of the Kimura operator associated to our Wright-Fisher noise is that it keeps the simplex invariant, without boundary conditions. This is an important property that makes the choice of a Wright-Fisher noise quite natural. Secondly, we stress that our MFG with common noise has itself a particle interpretation; see for instance [46]. This
Throughout the text, the state space is taken as $[d] := \{1, \ldots, d\}$, for an integer $d \geq 2$. We use the generic notation $p = (p_i)_{i \in [d]}$ (with $i$ in subscript) for elements of $\mathbb{R}^d$, while processes and functions are usually denoted by $\mathbf{p} = ((p_i)_{i=1, \ldots, d})_{t \in I}$ (for some time interval $I$) and $f = (f^i)_{i=1, \ldots, d}$ respectively (with $i$ in superscript). Also, we let $\mathcal{S}_d := \{(p_1, \ldots, p_d) \in (\mathbb{R}_+)^d : \sum_{i \in [d]} p_i = 1\}$ be the $(d-1)$-dimensional simplex. The Euclidean norm of $p \in \mathbb{R}^d$ is denoted by $|p|$. We can identify $\mathcal{S}_d$ with the convex polyhedron of $\mathbb{R}^{d-1}$ $\mathcal{S}_d := \{(x_1, \ldots, x_{d-1}) \in (\mathbb{R}_+)^{d-1} : \sum_{i \in [d-1]} x_i \leq 1\}$. In particular, we sometimes write “the interior” of $\mathcal{S}_d$: in such a case, we implicitly define the interior of $\mathcal{S}_d$ as the $(d-1)$-dimensional interior of $\mathcal{S}_d$. To make it clear, for some $p \in \mathcal{S}_d$, we write $p \in \text{Int}(\mathcal{S}_d)$ to say that $p_i > 0$ for any $i \in [d]$. We also write $x \in \text{Int}(\mathcal{S}_d)$ to say that $x \in \mathcal{S}_d$, $x_i > 0$ for each $i \in [d-1]$ and $\sum_{i \in [d-1]} x_i < 1$.

We use the same convention when speaking about the boundary of $\mathcal{S}_d$: For some $p \in \mathcal{S}_d$, we may write $p \in \partial \mathcal{S}_d$ to say that $p_i = 0$ for some $i \in [d]$. Finally, $\delta_{i,j}$ is the Kronecker symbol, $r_+$ denotes the positive part of $r \in \mathbb{R}$ and, for two elements $(v_i)_{i \in [d]}$ and $(w_i)_{i \in [d]}$ of $\mathbb{R}^d$, we sometimes denote the inner product $\sum_{i \in [d]} v_i w_i$ by $\langle v_*, w_* \rangle$.

2. Main results

In order to state our main results, we first introduce step by step the two forms of MFGs that we handle in the paper. We start with the game without common noise, which is assumed to be potential. This game could be called *inviscid*, a terminology that is commonly used in the PDE literature in order to emphasize the absence of diffusivity. As we already explained in the introduction, this inviscid MFG might not be uniquely solvable, which fact is the basic rationale for inserting next a common noise in the dynamics. In [43], we indeed cooked up a form of forcing, comprising a noise and a
repulsive drift at the boundary, under which the simplex is kept invariant by the equilibrium dynamics and that forces the MFG to become uniquely solvable. Accordingly, the game with common noise should be called viscous. Unfortunately, a striking point in our study is that the common noise, at least in the form postulated in [43], destroys the potential structure of the game. This prompts us to address in the end a new and tailor-made form of MFG that is driven by both a common noise and a potential structure.

2.1. A first form of MFG

The general form of inviscid MFGs that we here consider is given by the following fixed point problem: For some time horizon $T > 0$, find an $S_d$-valued continuous trajectory $\mathbf{p} = (p_t)_{0 \leq t \leq T}$ that is an optimal trajectory to the $\mathbf{p}$-dependent control problem

$$
\inf_{\mathbf{x} = (x_t)_{0 \leq t \leq T}} J(\mathbf{x}; \mathbf{p}) = \int_0^T \sum_{i \in [d]} q^i_t \left( L^i(x_t) + f^i(p_t) \right) dt + \sum_{i \in [d]} q^i_t g^i(p_T),
$$

(2.1)

where $\mathbf{q} = (q_t)_{0 \leq t \leq T}$ solves the Fokker-Planck (FP) equation

$$
\dot{q}^i_t = \sum_{j \in [d]} q^j_t x^{i,j}_t = \sum_{j \neq i} \left( q^j_t x^{i,j}_t - q^i_t x^{j,i}_t \right), \quad t \in [0, T], \quad i \in [d],
$$

(2.2)

subjected to the initial condition $q_0 = p_0$ and to the control $\mathbf{x} = (x^{i,j}_t)_{i,j \in [d]}_{0 \leq t \leq T}$ satisfying the constraint

$$
x^{i,j}_t \geq 0, \quad i,j \in [d], i \neq j; \quad x^{i,i}_t = -\sum_{j \neq i} x^{i,j}_t, \quad i \in [d], \quad t \in [0, T].
$$

(2.3)

Obviously, the latter constraint says that the trajectory $\mathbf{q}$ may be interpreted as the collection of marginal distributions of a Markov process with transition jump rates $((x^{i,j}_t)_{i,j \in [d]}_{0 \leq t \leq T})$. In the definition of the cost functional (2.1), $(f^i)_i \in [d]$ and $(g^i)_i \in [d]$ are tuples of real valued functions, the regularity of which is specified in the next subsection. As for the cost $L^i_{i \in [d]}$, we take for convenience

$$
L^i(\mathbf{x}) = \frac{1}{2} \sum_{j \neq i} |x_{i,j}|^2.
$$

(2.4)

The MFG associated with (2.1) and (2.2) has been widely studied. In this respect, it is worth recalling that uniqueness is known to hold true in a few settings only and may actually fail in many cases. The typical condition that is used in practice to ensure uniqueness is a form of monotonicity of the cost coefficients $f$ and $g$, but as recalled in the introduction and as shown in the recent paper [43], uniqueness can be also forced without any further need of monotonicity by adding to the dynamics of $\mathbf{q}$ a convenient kind of common noise together with a repulsive forcing at the boundary. In the presence of common noise, equilibria become random. In [43], candidates $\mathbf{p}$ for solving the equilibria are then sought as $S_d$-valued continuous stochastic processes on $[0, T]$ that are adapted to the (complete) filtration $\mathcal{F}$ generated by a collection of independent Brownian motions $((\mathcal{B}^{i,j})_{0 \leq t \leq T})_{i,j \in [d], i \neq j}$. Those Brownian motions form the common noise and are constructed on a given (complete) probability space $(\Omega, \mathcal{A}, \mathbb{P})$. 

COMMUNICATIONS IN PARTIAL DIFFERENTIAL EQUATIONS
Accordingly, the FP equation (2.2) for \( q = (q_t)_{0 \leq t \leq T} \) becomes a Stochastic Differential Equation (SDE) driven by both the common noise \((B^{i,j}_t)_{0 \leq t \leq T}, i, j \in [d], i \neq j\) and the environment \( p \), the general form of which is

\[
dq^i_t = \sum_{j \neq i} \left( q^j_t \left( \varphi (p^j_t) + x^j_t \right) - q^i_t \left( \varphi (p^i_t) + x^i_t \right) \right) dt + \frac{\epsilon}{\sqrt{2}} \sum_{j \neq i} \sqrt{p^j_t} \left( dB^i_t - dB^j_t \right),
\]

for \( i \in [d] \) and \( t \in [0, T] \), with \( q_0 = p_0 \) as initial condition. A peculiar point with (2.5) is that, generally speaking, the components \((q^i_t)_{0 \leq t \leq T}, i \in [d]\) are positive but the mass process \((\sum_{i=1}^d q^i_t)_{0 \leq t \leq T}\) is just equal to 1 in the mean under the expectation \( \mathbb{E} \) carrying the common noise. We refer to [43, Prop 2.4] for more details on this subtlety, but also on the solvability of (2.5). Basically, (2.5) is uniquely solvable if \( \alpha \) is a bounded process and \( \epsilon^2 \int_0^T (1/p^i_t) dt \) has finite exponential moments of sufficiently high order for any \( i \in [d] \). In the sequel we use the fact that unique solvability is actually true for any initial condition \( q_0 \in S_d \), even if different from \( p_0 \). Consistently with the fact that both \( p \) and \( q \) are random, the control process \( \alpha \) is also assumed to be progressively-measurable with respect to \( \mathbb{F} \) and, in the resulting MFG with a common noise, the cost (2.1) is averaged out with respect to the expectation \( \mathbb{E} \), namely the cost functional becomes

\[
\inf_{\alpha = (\alpha_t)_{0 \leq t \leq T}} J^{\varphi, \psi}_{e\varphi} (\alpha; p), J^{\varphi, \psi}_{e\varphi} (\alpha; p) = \mathbb{E} \left[ \int_0^T \sum_{i \in [d]} q^{i, \varphi}_{t} \left( L^i(\alpha_t) + f^i(p_t) \right) dt + \sum_{i \in [d]} q^{i, \varphi}_{T} g^i(p_T) \right],
\]

where \( q^{i, \varphi} = (q^{i, \varphi}_{t})_{0 \leq t \leq T}, i \in [d] \) solves (2.5).

The reader must pay attention to the superscript \( \varphi \) right above. Indeed, in addition to the common noise, the intensity of which is denoted by the positive parameter \( \epsilon \) in (2.5) (which we take in \((0, 1]\) in the sequel), the other main feature of (2.5) is the additional \( \varphi \) therein. As we alluded to, \( \varphi \) is actually intended to induce a repulsive drift that forces equilibria (dynamics of equilibria are obtained by taking \( p = q \) in (2.5)) to stay away from the boundary of the simplex whenever \( p_0 \in \text{Int}(S_d) \), and this despite the presence of the noise. The reader may notice that, instead, the deterministic dynamics (2.2) stay away from the boundary when \( \varphi \equiv 0 \). For our choice of \( \varphi \), we have in particular that \( \epsilon^2 \int_0^T (1/p^i_t) dt \) is exponentially integrable, as we required above to ensure that (2.5) is uniquely solvable; see [43, Prop. 2.4] for the details of the computations.

To achieve this goal and apply the results of [43], it suffices for the moment to assume that \( \varphi \) is a non-increasing smooth function on \([0, + \infty)\) such that

\[
\varphi (r) = \begin{cases} \kappa, & \text{if } r \in [0, \theta], \\ 0, & \text{if } r > \theta. \end{cases}
\]

Here, \( \kappa \) and \( \theta \) are two additional positive parameters that permit to tune the intensity of the drift induced by \( \varphi \). In this framework, the main result [43, Theorem 3.2] says that, for given positive values of \( \epsilon \) and \( \theta \), the MFG associated with the dynamics (2.5) and the cost functional (2.6) is uniquely solvable if \( \kappa \) is large enough. The form of the condition on \( \kappa \) together with the assumption on \( f \) and \( g \) are clarified in the statement of Theorems 2.1 and 2.2 even though the latter two address a slightly different problem. Indeed, we stress that the MFG (2.5)-(2.6) is not the system that we effectively use in
the sequel for addressing the selection principle. The main reason is that this MFG is not potential, whilst the potential structure is needed in our argument. Instead we will focus on a variant of the cost (2.6) that is introduced in the statement of Theorem 2.2.

2.2. Potential structure

As announced in the introduction, we here restrict ourselves to the so-called potential case. Following [3, 7, 26], we hence assume that the coefficients \( f \) and \( g \) derive from smooth potentials \( F \) and \( G \). Roughly speaking, this means that

\[
f^i(p) = \frac{\partial F}{\partial p_i}(p), \quad g^i(p) = \frac{\partial G}{\partial p_i}(p), \quad p \in S_d,
\]

(2.8)

but this writing is not completely satisfactory: In order to give a meaning to the two derivatives in the right-hand side above, both \( F \) and \( G \) must be in fact defined on an open subset of \( \mathbb{R}^d \) containing \( S_d \) (recall that the latter is a \((d - 1)\)-dimensional manifold). In case when \( F \) and \( G \) are just defined on the simplex, we may use instead the intrinsic derivative on the simplex, which identifies with a \((d - 1)\)-dimensional instead of \(d\)-dimensional vector. We refer to [43, Subsection 3.1.2] for the definition of intrinsic derivatives, but say to clarify that, whenever \( F \) is differentiable on a neighborhood of the simplex in \( \mathbb{R}^d \), the intrinsic gradient \( D F = (d_1 F, \ldots, d_d F) \in \mathbb{R}^d \) of \( F \) is simply given by the orthogonal projection of the \( d\)-dimensional gradient \( \nabla F \) onto the tangent space to the simplex, which is the orthogonal space to the \(d\)-dimensional vector \( 1 = (1, \ldots, 1) \). Hence we define \( D F = \nabla F - \frac{1}{d} (\nabla F, 1)1 \), and, when \( F \) is just defined on the simplex, the intrinsic derivative is defined by the same formula, but rewritten as

\[
d_i F(p) = \partial_i [F(p + \varepsilon(e_i - \bar{e}))]|_{\varepsilon = 0}, \quad p \in \text{Int}(S_d), \quad i \in [d].
\]

In the above definition, \( e_i \) is defined as the \(i\)th vector of the canonical basis of \( \mathbb{R}^d \) and \( \bar{e} \) as \( \bar{e} := (e_1 + \cdots + e_d)/d \); in particular, \( e_i - \bar{e} \) is a tangent vector to the simplex. From the construction, we have \( \sum_{i=1}^d d_i F = 0 \). Therefore, from now on, we consider the potentials to be just defined on the simplex and accordingly the derivatives to be taken in the intrinsic sense. In clear, we assume throughout the paper that there exist \( F, G : S_d \to \mathbb{R} \) such that, for any \( i \in [d] \) and \( p \in \text{Int}(S_d) \),

\[
d_i F(p) = f^i(p) - \frac{1}{d} \sum_{j=1}^d f^j(p), \quad d_i G(p) = g^i(p) - \frac{1}{d} \sum_{j=1}^d g^j(p).
\]

(2.9)

Note that this is obviously the case if \( F \) and \( G \) are extended into functions that are differentiable on a \(d\)-dimensional neighborhood and that the corresponding Euclidean derivatives satisfy (2.8). However, it is worth observing that (2.9) is slightly more natural than (2.8) because it involves \(d - 1\) entries instead of \(d\). In other words, \( F \) and \( G \) are required to be defined only on \( S_d \). In particular, we will see in Subsection 2.6 that any two state MFG is potential, in the sense that we can always find \( F \) and \( G \) satisfying (2.9). Whilst the proof of the latter is almost straightforward, the proof of the existence of an extension of both \( F \) and \( G \) to \( \mathbb{R}^d \) that satisfy (2.8) leads to heavier formulas. Anyway, beyond its interest for formulating the potential structure, the intrinsic derivatives will be also useful to state the PDEs associated with both the MFG and MFCP in
hand. Last, we recall that a variant of Whitney’s Theorem permits to extend smooth functions from the simplex to the whole Euclidean space; see for instance [47, Chapter 1]. If needed, the reader can use the latter as a quite systematic tool to express the various derivatives on the simplex in the form of Euclidean derivatives.

A very appealing fact with potential games without common noise is that they are naturally associated with a control problem. Actually, this connection is a general fact in game theory and it goes far beyond the single scope of MFGs. In the specific framework of MFGs, the underlying control problem is an MFCP, as we pointed out in the introduction. In our setting (and once again without common noise), the MFCP takes the form:

$$\inf_{\mathbf{a}=(\mathbf{a}_i)_{0 \leq t \leq T}} J(\mathbf{a}), \quad J(\mathbf{a}) = \int_0^T \left( \sum_{i \in [d]} q^i_t L^i(\mathbf{a}_t) + F(q_t) \right) dt + G(q_T), \quad (2.10)$$

where, as in (2.10), $\mathbf{q} = (q_t)_{0 \leq t \leq T}$ is a deterministic trajectory solving (2.2) subjected to the initial condition $q_0 = p_0$, for some given $p_0 \in \text{Int}(S_d)$, and to the deterministic control $\mathbf{a} = ((a_i^j)_{i,j \in [d]})_{0 \leq t \leq T}$ satisfying the constraint (2.3). For convenience, we also assume that admissible controls are bounded, meaning that $(a_i^j)_{0 \leq t \leq T} \in L^\infty(0, T)$ (see Remark 5.4 for more details on the rationale for this assumption). The connection between the MFCP (2.10)--(2.2) and the MFG (2.1)--(2.2) has been widely addressed in the literature; see for instance [3, 7] for continuous state MFGs and [8, 26] for finite state MFGs. Generally speaking, it says that any optimal trajectory $\mathbf{p} = (p_t)_{0 \leq t \leq T}$ to (2.10)--(2.2) that stays away from the boundary of the simplex solves the MFG associated with (2.1)--(2.2). Noticeably, the fact that the potential structure is expressed in intrinsic derivatives does not change anything, see the introduction of Subsection 5.2. However, there are known instances of MFG equilibria that are not minimizers of the corresponding MFCP; see Subsection 2.6 below for a benchmark example. This is by the way the starting point of our paper. Our main result here is indeed to construct a selection procedure that it rules out these non-minimal equilibria, meaning that it rules out solutions $\mathbf{p}$ to the MFG (2.1)--(2.2) that are not optimal trajectories of (2.10)--(2.2). Using the same terminology as in the previous subsection, we succeed to associate with the inviscid MFG, which is hence an inviscid potential game, a viscous potential game with the following four features:

1. The viscous potential game derives from a viscous MFCP, that is an MFCP with a common noise of intensity $\varepsilon$. In words, the viscous MFG writes as the first order condition of some viscous MFCP and, accordingly, any minimizer of the viscous MFCP is an equilibrium of the viscous potential game.
2. The viscous potential game is uniquely solvable, hence implying that its unique solution, say $\mathbf{p}^\varepsilon$, is also the unique optimal trajectory of the viscous MFCP.
3. The laws of the optimal trajectories $(\mathbf{p}^\varepsilon)_{\varepsilon, \phi}$ are tight and, accordingly, the optimal trajectory $\mathbf{p}^\varepsilon$ converges in the weak sense up to a subsequence, as the viscosity $\varepsilon^2$ and the forcing drift $\phi$ both vanish. Any limit point $\mathbb{M}$ is a probability distribution on $C([0, T]; S_d)$ that is supported by the set of optimal trajectories of the inviscid MFCP. These limit points are in particular uniquely determined.
when the inviscid MFCP has a unique optimizer, which happens to be the case for almost every initial condition when $F$ and $G$ are smooth enough.

4. For a convenient choice of $\phi$ in terms of $\varepsilon$, the cost functional of the viscous potential game, which is in the end a variant of $J^{\varepsilon,\phi}$ in (2.6), converges in a suitable sense to the cost functional $J$ in (2.1). In particular, the equilibrium cost of the viscous potential game converges, along the same subsequences as in item (3), to the mean of the equilibrium costs of the inviscid potential game under the limiting distribution $M$. Again, this equilibrium cost is uniquely determined when the inviscid MFCP has a unique optimal trajectory.

The combination of the first three items reads as a selection principle since it rules out equilibria of (2.1)–(2.2) that are not optimizers of (2.10)–(2.2), whilst the last item guarantees some consistency in our approach as it says that the cost functional underpinning the approximating viscous potential game is itself a good approximation of the original cost function $J$ in (2.1). Although this strategy looks quite natural, it is in fact rather subtle. The major obstacle is that, as we already said, the pair (2.5)–(2.6) is not a potential game, hence advocating for the search of a version that derives from a potential.

Our first step in the construction of a suitable viscous potential game is the construction of the viscous MFCP itself. To do so, we elaborate on [43]. Following (2.10), we can indeed associate with the dynamics (2.5) a stochastic control problem, which we precisely call viscous MFCP. It has the following form:

$$\inf_{\alpha} J^{\alpha,\phi}(\alpha), \quad J^{\alpha,\phi}(\alpha) = \mathbb{E}\left[ \int_0^T \left( \sum_{i \in [d]} p_i^t L_i(x_i) + F(p_i^t) \right) dt + G(p_T) \right],$$

(2.11)

where $p = (p_t)_{0 \leq t \leq T}$ solves the $\alpha$-driven SDE

$$dp_i^t = \sum_{j \neq i} \left( p_i^t (\phi(p_i^t) + x_i^t) - p_i^t (\phi(p_i^t) + x_i^t) \right) dt + \frac{\varepsilon}{\sqrt{2}} \sum_{j \neq i} \sqrt{p_i^t p_j^t (dB_i^{t,i} - dB_j^{t,i})},$$

(2.12)

for $i \in [d]$ and $t \in [0, T]$, with $q_0 = p_0$ as initial condition, and, as before, $\alpha$ is an $\mathcal{F}$-progressively measurable process satisfying (2.3) except for the fact that, for purely technical reasons, we will restrict ourselves to processes whose off-diagonal coordinates are bounded by a constant $M$ that is explicitly given in terms of $f, g$ and $T$ (even though (2.3) just implies that the diagonal coordinates are bounded by $(d-1)M$, we will say abusively that such processes are bounded by $M$). The function $\phi$ is chosen as in (2.7) and the initial condition $p_0$ belongs to the interior of the simplex. Such a condition on $p_0$ will be always assumed in the rest of the paper. As we already explained, it permits to apply [43, Propositions 2.2, 2.3, 2.4]: If $\kappa$ in (2.7) is large enough, the SDE (2.12) (which is usually called a Wright-Fisher SDE) is uniquely solvable in the strong sense and the solution remains in the interior of the simplex; further, $\int_0^T (1/p_t^i) dt$ has exponential moments of sufficiently high order, which ensures that (2.5) is well-posed.

In this framework, our first main result has some interest in its own, independently of the aforementioned selection principle, but the statement requires new material. The functional spaces to which $F$ and $G$ are assumed to belong, and to which the value function is proved to belong, are defined in detail in the appendix. These are called
Wright-Fisher, as introduced in [44], and are used in [43] to prove well posedness of the MFG master equation. We just say here that:

1. The space $C^{0, \gamma}_{WF}(S_d)$ consists of continuous functions on $S_d$ that are $\gamma$-Hölder continuous up to the boundary with respect to the metric associated with the Wright-Fisher noise. In short, the distance between two points $p = (p_i)_{1 \leq i \leq d}$ and $q = (q_i)_{1 \leq i \leq d}$ in $S_d$ is then given by the Euclidean norm of the difference $(\sqrt{p_i} - \sqrt{q_i})_{1 \leq i \leq d}$; see (A.3) in the appendix together with [43, Remark 3.1] in (2.12). Similarly, $C^{k, \gamma}_{WF}(S_d)$ consists of continuous functions on $S_d$ that are continuously differentiable in $\text{Int}(S_d)$, with Hölder continuous derivatives up to the boundary. Both spaces are equipped with norms $\| \cdot \|_{WF, 0, \gamma}$ and $\| \cdot \|_{WF, 1, \gamma}$.

2. For $k = 0, 1$, $C^{k, \gamma}_{WF}(S_d)$ consists of continuous functions on $S_d$ that are $2 + k$ times continuously differentiable in $\text{Int}(S_d)$, with derivatives satisfying a suitable behavior at the boundary and a suitable Hölder regularity that depend on the order of the derivative. In particular, the derivatives of order $1$ (if $k = 0$) and of order $1$ and $2$ (if $k = 1$) are Hölder continuous up to the boundary, but the derivative of order $2 + k$ (i.e. $2$ if $k = 0$ and $3$ if $k = 1$) may blow up at the boundary and be only locally Hölder continuous in the interior. Both spaces are equipped with norms $\| \cdot \|_{WF, 0, 2 + \gamma}$ and $\| \cdot \|_{WF, 1, 2 + \gamma}$.

3. The spaces $C^{0, \gamma}_{WF}([0, T] \times S_d)$ and $C^{k, 2 + \gamma}_{WF}([0, T] \times S_d)$ are the parabolic versions of $C^{0, \gamma}_{WF}(S_d)$, and $C^{k, 2 + \gamma}_{WF}(S_d)$. While the former consists of functions on $[0, T] \times S_d$ that are Hölder continuous for the time-space variant of the Wright-Fisher metric (given as the Wright-Fisher metric between space locations plus the root of the Euclidean distance between time instants; see (A.5) in the appendix), the latter consists of continuous functions on $[0, T] \times S_d$ that are continuously differentiable in time $t \in [0, T]$ and that are $2 + k$ times continuously differentiable in space in $\text{Int}(S_d)$, with derivatives satisfying a suitable behavior at the boundary and a suitable Hölder regularity. In particular, the time derivative and the space derivatives up to order $1 + k$ are Hölder continuous up to the boundary but the derivative of order $2 + k$ may blow up at the boundary. The norms are also denoted by $\| \cdot \|_{WF, 0, \gamma}$ and $\| \cdot \|_{WF, k, 2 + \gamma}$ (below, the norm is understood as being for the parabolic space if the function in argument of the norm is time-space dependent).

Throughout the sequel, the parameter $\gamma \in (0, 1)$ is fixed. Moreover, we recall that we are always assuming that $F$ and $G$ satisfy (2.9).

**Theorem 2.1.** Recall $(\theta, \kappa)$ from (2.7). If $F \in C^{1, \gamma}_{WF}(S_d)$ and $G \in C^{1, 2 + \gamma}_{WF}(S_d)$, then there exists a constant $k_1 > 0$ only depending on $\|f\|_\infty, \|g\|_\infty$, $T$ and $d$, such that for any $\varepsilon \in (0, 1], \theta > 0$ and $\kappa \geq k_1 / \varepsilon^2$, and any initial state $p_0$ in $\text{Int}(S_d)$, the MFCP (2.11)–(2.12) set over $\mathcal{F}$-progressively measurable processes $\alpha$ that are bounded by $M = 2(\|g\|_\infty + T\|f\|_\infty)$ has a unique solution. Moreover, there exists $\gamma' \in (0, \gamma]$, possibly depending on $\varepsilon$ and $\kappa$, such that the corresponding HJB (2.13) equation has a unique solution $V^{\varepsilon, \phi}$ in $C^{2 + \gamma'}_{WF}([0, T] \times S_d)$.

The proof of this result is given in Section 3; see Theorem 3.3. We use the notations $D^2 = (d_{j,k})_{j,k \in [d]}$ for the first and second order derivatives on the simplex, the second
derivative being defined similarly to the first (see Subsection 3.1 for a short account and [43, Subsection 3.1.2] for more details). The HJB equation has the following form:

\[
\begin{aligned}
\partial_t \mathcal{V} + \mathcal{H}_M^\phi(p, D\mathcal{V}) + F(p) + \frac{\varepsilon^2}{2} \sum_{j, k} (p_j \delta_{j,k} - p_j p_k) \mathcal{d}_{j,k}^2 \mathcal{V} = 0,
\end{aligned}
\]

(2.13)

for \((t, p) \in [0, T] \times \mathcal{S}_d\), where \(\mathcal{H}_M^\phi\) is a Hamiltonian term depending explicitly on \(\phi\) and \(M\), the precise form of which is not so relevant at this early stage of the paper and will be just given in the sequel of the text; see (3.3). In fact, we feel more useful for the reader to be aware of the key fact that, here, this HJB equation is shown to have a unique classical solution, denoted by \(\mathcal{V}_{\nu, \phi}\). Obviously, this is a strong result that is true because of the presence of the common noise and, in particular, that bypasses any use of convexity on \(F\) and \(G\) (and hence of monotonicity on \(f\) and \(g\)). The proof makes use of the smoothing properties obtained in [44] and [43] for so-called Kimura operators, which are here the generators associated with the simplex-valued diffusion processes (2.12), when the latter are driven by controls in Markov feedback form \((x^j_t = x^j(t, p_t))_{0 \leq t \leq T}\) for \(x^j: [0, T] \times \mathcal{S}_d \rightarrow \mathbb{R}_+\) if \(i \neq j\). For such an \(x\), the operator may be expanded in the form

\[
\sum_{j, k} (p_k (\phi(p_j) + \mathcal{x}^j(t, p)) - p_j (\phi(p_k) + \mathcal{x}^k(t, p))) \mathcal{d}_j + \frac{\varepsilon^2}{2} \sum_{j, k} (p_j \delta_{j,k} - p_j p_k) \mathcal{d}_{j,k}^2;
\]

see [44, Definition 2.2.1] for a complete review and [43, Subsection 3.2] for an overview. The main difficulty in the analysis is that this operator degenerates at the boundary of the simplex. The latter explains why we need the forcing \(\phi\) to be sufficiently strong in order to guarantee the existence of a classical solution, whence the condition \(\kappa \geq \kappa_1/\varepsilon^2\) in the statement.

Our second main result is to prove that there is a uniquely solvable MFG that derives from the viscous MFCP. Noticeably, this is a non-trivial fact. The reason is that, because of the presence of stochastic terms in (2.12), the standard computations that permit to pass from inviscid MFCPs to inviscid potential games are no longer true. To wit, the result below says that the shape of the cost of the viscous potential game is not the same as the shape of the cost of the original inviscid one.

**Theorem 2.2.** Take \(F, G, M\), and \(\kappa_1\) as in the statement of Theorem 2.1. Then, there exists \(\kappa_2 \geq \kappa_1\), only depending on \(||f||_\infty, ||g||_\infty\), \(T\) and \(d\), such that, for any \(\varepsilon \in (0, 1]\), any \(\theta > 0\) and any \(\kappa \geq \kappa_2/\varepsilon^2\), we can find a time-dependent coefficient \(\theta_{\nu, \phi} : [0, T] \times \mathcal{S}_d \rightarrow \mathbb{R}_d\) that is continuous on \([0, T] \times \mathcal{S}_d\) such that, for any initial condition \(p_0 \in \text{Int} (\mathcal{S}_d)\), the optimal trajectory \(p^{\varepsilon, \phi}\) of the MFCP (2.11)–(2.12) is also the unique equilibrium of the MFG with common noise driven by the \(p\)-dependent cost functional

\[
\hat{j}^{\varepsilon, \phi}(x; p) = \mathbb{E} \left[ \int_0^T \sum_{i \in [d]} q_i \left( L_i(x_t) + f^i(p_t) + \theta_{\nu, \phi}^i (t, p_t) \right) dt + \sum_{i \in [d]} q_i g^i(p_T) \right],
\]

(2.14)

defined over pairs \((q, x)\) solving (2.5), for \(\mathbb{F}\)-progressively measurable processes \(x\) that are bounded by \(M\), and over \(\mathbb{F}\)-adapted continuous processes \(p\) that take values in \(\text{Int} (\mathcal{S}_d)\).
Notice that, in the notation $\vartheta_{e,\phi}$, we put the parameters $(e, \phi)$ in subscript as we sometimes write $\vartheta^i_{e,\phi}$ for denoting the coordinate $i$ of $\vartheta_{e,\phi}$. Consistently, we adopt the same convention for all the functions that are defined on a finite dimensional set and that are parameterized by $e$ and $\phi$.

The statement of Theorem 2.2 deserves some explanations. First, we feel useful to specify the definition of an equilibrium in our framework:

**Definition 2.3.** With the same framework as in Theorem 2.2 (in particular $\kappa$ is large enough), an $\mathbb{F}$-adapted continuous process $p$ with values in $S_d$ is said to be an equilibrium if the following two properties are satisfied:

i. There exists an $M$-bounded and $\mathbb{F}$-progressively measurable process $a$ such that $p$ solves the SDE (2.12)—obtained by equalizing $p$ and $q$ in (2.5)—, with $p_0$ as initial condition;
ii. For any other $M$-bounded and $\mathbb{F}$-progressively measurable process $b$ for which (2.5) is uniquely solvable, $\int_0^T \varphi(\mathbb{E},\mathbb{U})(a_t, p_t) dt \leq \int_0^T \varphi(\mathbb{E},\mathbb{U})(b_t, p_t) dt$.

In particular, from item (i) in the above definition, $p$ in (2.14) is implicitly required to solve (2.12) for some $M$-bounded and $\mathbb{F}$-progressively measurable control process $a$.

As recalled above, (2.12) is in fact uniquely solvable under the assumption of Theorem 2.2; its solution is thus $p$ itself. Moreover, $\int_0^T (1/p_t^i) dt$ has exponential moments of sufficiently high order, which guarantees that (2.5) and hence (2.14) make sense. Actually, it is worth mentioning that the same would be true under the weaker condition $\kappa \leq \kappa_0 \epsilon^2$, with $\kappa_0$ only depending on the dimension; see again [43, Proposition 2.3]. The stronger threshold of the form $\kappa_2/\epsilon^2$ in the statement of Theorem 2.2 is needed for the well-posedness of the master equation, as in [43, Remark 2.11 and Theorem 3.4].

The proof of Theorem 2.2 is given in Section 4, together with the precise definition of the additional cost $\vartheta_{e,\phi}$; see Theorem 4.1 and (4.6). At this stage, it is certainly fair to say that the definition of $\vartheta_{e,\phi}$ is implicit, meaning that it depends on $V_{e,\phi}$ itself, which might seem a bit disappointing but looks in the end inevitable. However, we insist on the fact there exists an explicit procedure for constructing $\vartheta_{e,\phi}$ in terms of $V_{e,\phi}$. In particular, $\vartheta_{e,\phi}$ is given once the viscous MFCP has been solved, and this before any further definition of the viscous MFG. This ensures that our construction is free of any vicious circle. As for the proof of Theorem 2.2 itself, it relies on a variant of the verification argument used in [43, Theorem 3.2]. The main point is to take benefit of the smoothness of the value function $V_{e,\phi}$ in order to expand the intrinsic gradient of $\partial_{e,\phi}$ along any possible equilibrium $p$. This allows us to prove that, whatever the equilibrium $p$, the optimal solution to $\tilde{J}_{e,\phi}(\cdot ; p)$ is in the form

$$\left(\tilde{J}_{e,\phi}(\cdot ; p)^i\right)_t = a^*(d_i V_{e,\phi}(t, p_t) - d_j V_{e,\phi}(t, p_t)), \quad t \in [0, T], \quad i \neq j, \tag{2.15}$$

where

$$a^*(r) = \begin{cases} 0 & \text{if } r < 0, \\ r & \text{if } r \in [0, M], \\ M & \text{if } r > M. \end{cases} \tag{2.16}$$
By plugging (2.15) into (2.12), we get that all the equilibria satisfy the same SDE. Thanks to the bounds we have on $D\nabla_e^\nu_{e,\phi}$, the latter SDE is uniquely solvable, which implies uniqueness of the MFG solution. See Section 4 for more details.

### 2.3. Selection

The next step in our program is to address the asymptotic behavior of the equilibria $(p^e, u^e)_{e,\phi}$ as $e$ tends to 0 and the support of $\phi$ shrinks to the boundary of the simplex. In this regard, one difficulty is that, in the statements of both Theorems 2.1 and 2.2, the function $\nu$ is implicitly required to become larger and larger, as $e$ tends to 0, on the interval $[0, \theta]$. Equivalently, the constant $\kappa$ therein blows up as $e$ tends to 0. Obviously, this looks a serious hindrance for passing to the limit. In Section 5 below, we bypass this difficulty by using the fact that, in the limit, the solutions of the Fokker-Planck equation (2.2) without common noise cannot reach the boundary when starting from the interior of the simplex, and in fact the solution stays away from the boundary with an explicit threshold (this advocates once more for taking $p_0$ in $\text{Int}(S_d)$). Also, for the subsequent analysis, we introduce a new parameter $\delta$ and we choose $\nu = u^e_\delta$ satisfying (in addition to the aforementioned monotonicity and regularity properties)

$$
\varphi_{\theta, \delta, e}(r) = \begin{cases} 
\kappa_e & r \leq \theta, \\
\kappa_0 & 2\theta \leq r \leq \delta, \\
0 & r \geq 2\delta,
\end{cases}
$$

for $0 < \delta \leq 1/2$, $0 < 2\theta < \delta$ and $\kappa_0 > 0$, together with

$$
|\varphi'_{\theta, \delta, e}(r)| \leq \frac{2\kappa_e}{\theta} 1_{[0,2\theta]}(r) + \frac{2\kappa_0}{\delta} 1_{[2\theta,\delta]}(r), \quad r \geq 0.
$$

Above, $\kappa_e$ is of the form $\kappa_e := e^{-2}\kappa_2$, for $\kappa_2$ as in the statement of Theorem 2.2, while $\kappa_0$ is a constant to be determined. We implicitly require $\kappa_e \geq \kappa_0$ in order to guarantee that $\varphi_{\theta, \delta, e}$ is non-increasing. The rationale for introducing an additional parameter $\delta$ in the decomposition (2.17) is the following: Whilst $\kappa_e$ blows up with $e$, $\kappa_0$ does not. Here, $\kappa_e$ is used to force uniqueness of the MFG equilibrium when the intensity $e$ of the common noise is given, hence the need to have it large when $e$ is small, see Theorem 2.2. Differently, $\kappa_0$ is used below to force the equilibrium to stay sufficiently far away from the boundary, uniformly with respect to the intensity $e$; the reader may have a look at Proposition 5.5 for a precise statement.

In order to state our main result about convergence of the equilibria, we also let $E := L^2([0,T]; A)$, where $A = \{(a_{i,j})_{i,j=1,...,d} \in \mathbb{R}^{d \times d} : a_{i,j} \in [0,M], i \neq j; a_{i,i} = -\sum_{j \neq i} a_{i,j}\}$. We equip $E$ with the weak topology; moreover, we endow $C([0,T]; S^d)$ with the topology of uniform convergence. The dependence of the solution $p^e, \phi$ on $\delta$, $\theta$ and $\kappa$ is implicitly written as a dependence upon $\phi$. We then get the following result:

**Theorem 2.4.** Let the assumptions of both Theorems 2.1 and 2.2 be in force and, with the same notation as in (2.15) and for $\kappa_0 > 0$ and $\phi$ as in (2.17)-(2.18), let

$$
a^e, \phi_t := \varphi_t^e, \phi, \quad t \in [0,T].
$$
Then, for any initial condition \( p_0 \in \text{Int}(S_d) \), there is a constant \( \delta_0 > 0 \) such that the family of laws \( (\mathbb{P}^\varepsilon(p^{e,\varepsilon}, x^{e,\varepsilon})^{-1})_{\varepsilon \in (0, \sqrt{\kappa_2/k_0}], \delta \in (0, \delta_0), 2\theta \leq \delta} \) is tight in \( \mathcal{P}(C([0, T]; S_d) \times \mathcal{E}) \). Moreover, any weak limit \( \mathbb{M}_e \), as \( \varepsilon \) and \( \delta \) tend to 0, is supported by pairs \( (p, x) \) that minimize the cost functional \( J \) in (2.10), with \( p = q \) and \( p_0 \) as initial condition therein.

The proof is given in Subsection 5.1; see Theorem 5.1. We emphasize that the statement holds true in the general case when the inviscid MFCP has multiple optimizers. It then provides a form of selection: Up to a subsequence, the solutions of the viscous MFG converge to a probability measure whose support is included in the set of minimizers; recalling that the minimizers of the inviscid MFCP are equilibria of the inviscid MFG, this suffices to rule out the equilibria of the inviscid MFG that are not minimizers; see Corollary 5.3. Actually, uniqueness is by far not limited to the example of Subsection 2.6. As explained in the next Subsection, it happens quite often that the inviscid MFCP starting at \((t, p)\) (in the sense that \( q \) in (2.2) starts from \( p \) at time \( t \)) has a unique solution. Indeed, under the additional assumption that \( F \) and \( G \) are in \( C^{1,1}(S_d) \), the set of points \((t, p) \in [0, T] \times \text{Int}(S_d)\) for which the inviscid MFCP starting at \((t, p)\) is uniquely solvable consists of points of differentiability of the value function of the inviscid MFCP. This set has full measure since the value function can be shown to be Lipschitz in time and space; see Proposition 5.2.

Now that we have obtained from Theorem 2.4 the announced limiting behavior for the equilibria of the viscous MFG, we may also wonder about the behavior of the cost functional \( \tilde{J}^e,\varphi \) in (2.14) as \( \varepsilon \) tends to 0 and the support of \( \varphi \) shrinks to the boundary (and hence \( \delta \) vanishes). In fact, this asks us to revisit the shape of the coefficient \( \vartheta_{e,\varepsilon,\varphi} \), which is certainly the most intriguing term therein, see again Section 4. Importantly, we learn from our construction that, in order to control the impact of \( \vartheta_{e,\varepsilon,\varphi} \) accurately in the cost functional \( \tilde{J}^e,\varphi \), we cannot play for free with \( e, \delta \) and \( \theta \) at the same time—the three of them popping up in the definition of \( \varphi \) in (2.17). The reason is that, even though this may only happen with small probability, the process \( p^{e,\varepsilon,\varphi} \) may visit the neighborhood of the boundary of the simplex where the function \( \varphi \) is non-zero. Even more, \( \varphi \) may become very large when \( \varepsilon \) tends to 0. Since the geometry of this neighborhood of the boundary of the simplex is determined by \( \delta \) and \( \theta \), this explains why some tradeoff between \( e, \delta \) and \( \theta \) is necessary when averaging out the cost functional \( \vartheta_{e,\varepsilon,\varphi} \) with respect to all the possible trajectories of \( p^{e,\varepsilon,\varphi} \). In this context, the following result says that we can choose \( \kappa_0 \) in function of \( \kappa_2 \) and then tune \( \delta \) and \( \theta \) in terms of \( \varepsilon \) such that, along the equilibrium, the influence of \( \vartheta_{e,\varepsilon,\varphi} \) vanishes as \( \varepsilon \) tends to 0:

**Proposition 2.5.** Let the assumptions of both Theorems 2.1 and 2.2 be in force. Then, there exists \( \bar{\kappa}_0 \geq 0 \), only depending on \( \kappa_2 \) and \( d \), such that, for any \( \kappa_0 \geq \bar{\kappa}_0 \) and \( \varepsilon \in (0, \sqrt{\kappa_2/k_0}] \), we can choose \( \delta = \hat{\delta}(\varepsilon) \) and \( \theta = \theta(\varepsilon) \leq \delta(\varepsilon)/2 \), for some (strictly) positive-valued functions \( \hat{\delta} \) and \( \theta \), converging to 0 with \( \varepsilon \), and then \( \varphi = \hat{\varphi}(\varepsilon) \) in
(2.17)–(2.18), such that all the assumptions required in the statements of Theorems 2.1 and 2.2 are satisfied together with the following limit:

\[
\lim_{\varepsilon \to 0} \mathbb{E}^M \left[ \int_0^T \sum_{i \in [d]} q_i^e \phi(\varepsilon) v_i^e \phi(\varepsilon) (p^e \phi(\varepsilon)) dt \right] = 0,
\]

for any \( p_0 \in \text{Int}(S_d) \) and \( q_0 \in S_d \), where \( q^e \phi(\varepsilon) \) solves (2.5) with initial condition \( q_0 \) and \( p = p^e \phi(\varepsilon) \) therein.

In the statement, it is implicitly understood that \( \phi(\varepsilon) \) is parameterized by \( \phi(\varepsilon) \) and \( \phi(\varepsilon) \). As for the proof, it is given in Subsection 5.3, see Proposition 5.10 and Theorem 5.12. For sure, the above result says that sup \( p \in \mathbb{M} \) \( J(e, p) \) is a minimizer of the inviscid MFCP and hence an MFG equilibrium (it is worth recalling again that there might be MFG equilibria that are not minimizers). In particular, \( J(e, p) \) is nothing but \( J(e, p^e) \), the infimum being here taken over all the deterministic processes \( \beta \), see (2.1). At the end of the day, we may interpret the right-hand side of (2.19) as a mean over the values of some of the equilibria of the inviscid MFG. Obviously, the argument inside the limit symbol in the left-hand side is also the value of the unique equilibrium of the viscous MFG, hence proving that the limit points of the values of the viscous MFGs are means over the values of the inviscid MFG. Importantly, the probability \( \mathbb{M} \) here just charges the minimizers of the inviscid MFCP: In case when the inviscid MFCP has a unique solution, the expectation \( \mathbb{E}^M[J(e, p)] \) then reduces to \( \inf_{\mathbb{M}} J(e, p) \), where \( p \) is the unique minimal path of the inviscid MFCP.

Notice that (2.19) could be recasted in a fashion closer to \( \Gamma \)-convergence, but this would ask for more materials in the text and we would make little use of it in the end. Instead, our formulation suffices to address the convergence of the solution to the master equation, which is a key point in our paper.

### 2.4. Master equation

It is worth recalling that the value of a uniquely solvable MFG has a nice interpretation in terms of a PDE set on the space of probability measures and known as the master equation. The unknown of the latter reads in the form of a \( d \)-tuple of functions \( [0, T] \times \mathcal{S} \rightarrow U(t, p) = (U^1(t, p), \ldots, U^d(t, p)) \). For each \( i \in [d] \), \( U^i(t, p) \) should be regarded as the optimal cost of the optimal control problem solved by a reference player starting from state \( i \) at time \( t \) whenever the population follows the equilibrium initialized from state \( p \) a time \( t \). We often refer to this tuple \( (U^1, \ldots, U^d) \) as the conditional value of the game, the word “conditional” being here used to indicate that the initial state of the
reference player is forced to start from a deterministic state. In comparison, the (exact) value of the game (to which we referred in the previous paragraph) is the optimal cost to the reference player when the latter starts from any random initial condition with $p$ as distribution; this exact value is $\sum_{i \in [d]} p_i U^i(t, p)$. Our first main result in this direction concerns the master equation of the viscous MFG: It is here a system of second-order PDEs on the simplex, driven by the same Kimura operator as the HJB equation (2.13). It has the following general form:

$$
\begin{align*}
\partial_t U^i + H_M((U^i - U^j)_{j \in [d]}) + \sum_{j \in [d]} \phi(p_j)(U^j - U^i) + (j^i + \vartheta^i_{w, \varphi})(t, p) \\
+ \sum_{j, k \in [d]} p_j \left( \phi(p_j) + (U^k - U^j)_+ \right) (d_i U^i - d_k U^k) \\
+ \varepsilon^2 \sum_{j \in [d]} p_j (d_i U^j - d_j U^i) + \frac{1}{2} \varepsilon^2 \sum_{j, k \in [d]} (p_j \delta_{j,k} - p_j p_k) d^2_{j,k} U^i = 0,
\end{align*}
$$

(2.20)

for $(t, p) \in [0, T] \times \mathcal{S}_d$, where $H_M$ is the Hamiltonian:

$$
H_M(w) = \sum_{j \in [d]} \left\{ - a^*(w_j) w_j + \frac{1}{2} |a^*(w_j)|^2 \right\}, \quad w \in \mathbb{R}^d,
$$

(2.21)

with $a^*$ as in (2.16). A key fact is that, under the assumption of Theorem 2.2, this master equation has a unique classical solution (with a suitable behavior close to the boundary, see the definition of the so-called Wright-Fisher spaces in the appendix): This result is mostly a consequence of [43, Theorem 3.4], provided $\vartheta$ defined by (4.6) belongs to the right Wright-Fisher space, which will be shown in Theorem 4.1. Given a classical solution $U_{w, \varphi} = (U^i_{w, \varphi})_{i \in [d]}$ to (2.20), the value of the viscous MFG, when initialized from a state $p \in \text{Int}(\mathcal{S}_d)$ at some time $t \in [0, T)$, is given by $\sum_{i \in [d]} p_i U^i_{w, \varphi}(t, p)$. In other words, $U^i_{w, \varphi}(t, p)$ is nothing but $\inf_{x} \tilde{J}_{w, \varphi}^i(x; p_{w, \varphi})$ whenever $p_{w, \varphi}$ is initialized from $p$ at time $t$ and $\varphi$ in (2.5) is initialized at time $t$ from $(q^i = \delta_{i,j})_{j \in [d]}$.

Due to the potential structure of the game, there is in fact a strong connection between the HJB equation (2.13) and the master equation (2.20). We cannot have directly $U^i_{w, \varphi}(t, p) = d_i V_{w, \varphi}(t, p)$ for any $i \in [d]$, because the intrinsic gradient sum to zero, while the functions $U^i_{w, \varphi}$ do not. This is by the way part of the difficulty in proving Theorem 2.2; see Section 4. What we can show is that

$$
U^i_{w, \varphi}(t, p) - U^j_{w, \varphi}(t, p) = d_i V_{w, \varphi}(t, p) - d_j V_{w, \varphi}(t, p),
$$

(2.22)

for $t \in [0, T], p \in \text{Int}(\mathcal{S}_d)$ and $i, j \in [d]$, which is reminiscent of [9, Theorem 3.7.1] (in the sense that, heuristically, space derivatives in continuous state space are replaced here by differences). Notably, (2.22) suffices to prove that the MFG and the MFCP have the same solution, because the optimal control is given by (2.15). Let us point out that the construction of $\vartheta_{w, \varphi}$ in the proof of Theorem 2.2 plays a key role in obtaining (2.22). In short, $\vartheta_{w, \varphi}$ is precisely chosen in such a way that (2.22) is eventually satisfied: In the latter identity, the value function $V_{w, \varphi}$ is constructed first; then, $\vartheta_{w, \varphi}$ is defined in terms of $V_{w, \varphi}$; finally, the cost functional and in turn the value of the viscous MFG rely on the choice of $\vartheta_{w, \varphi}$. Interestingly, Proposition 2.5 provides a way to pass to the limit in the left-hand side of (2.22). When the inviscid MFCP (2.10)–(2.2) has a unique minimizer initialized from $p$ at time $t$, Proposition 2.5 implies that the limit of $U^i_{w, \varphi}$
(choosing, as in Proposition 2.5, δ as δ = ̂δ(ε), iversal parameters to (2.1) is U*(t, p), where now U*(t, p) stands for inf p J(x, p) with p denoting the unique minimizer of the inviscid MFCP initialized from p at time t and q in (2.2) being initialized at time t from (q, δ).

Obviously, a natural question is to relate such a limit U with the value function V of the inviscid MFCP (2.10), where, for (t, p) ∈ [0, T] × Sd, V(t, p) is defined as inf p J(x) whenever q in (2.2) starts from p at time t. We manage to prove (see Theorem 6.2) that V is the unique Lipschitz viscosity solution of the following HJB equation:

\[
\begin{align*}
&\begin{cases}
\partial_t V + \sum_{k,j \in [d]} p_k H(d_k V - d_j V) + F(p) = 0, \\
V(T, p) = G(p),
\end{cases} \\
&\text{where } H \text{ is the Hamiltonian associated to } L \text{ in (2.4), namely}
\end{align*}
\]

\[
H(u) = -\frac{1}{2} \sum_{j \in [d]} (u_j)^2, \quad u \in \mathbb{R}^d.
\] (2.23)

Pay attention that there is no condition on the boundary of the simplex; see Definition 6.1 for the details together with the proof of Theorem 6.2 (comparison principle), which instead exploits the fact that the interior of the simplex is invariant for the dynamics (2.2). Obviously, (2.23) should be regarded as the inviscid version of the equation (2.13) (up to the fact that controls are truncated by M in the latter, but this is in no loss of generality since optimal controls to (2.1) are bounded by M, see Proposition 5.2). Importantly, V is shown to be Lipschitz continuous in time and space, see if needed Proposition 5.2 in the core of the text. Hence, it is almost everywhere differentiable in [0, T] × Sd, which plays a crucial role in our analysis: We also prove in Proposition 5.2 that the inviscid MFCP has a unique solution when it is initialized from p ∈ Int(Sd) at time t such that V is differentiable in (t, p)—and hence for almost every (t, p) ∈ [0, T] × Int(Sd)—, which permits to pass to the limit (as ε tends to 0) in DV,φ almost everywhere in time and space in (2.22) (the simplex being equipped with the (d − 1) Lebesgue measure). To this end, we need to make the slightly stronger assumption that F ∈ C^{1,1}(Sd), meaning that f is Lipschitz continuous (on Sd and hence up to the boundary), in order to ensure that V is semiconcave; see again Proposition 5.2.

We build on this idea to obtain the following:

**Theorem 2.6** (Part I). Under the same assumption and notation as in the statement of Proposition 2.5, we have

\[
\lim_{\varepsilon \to 0} V_{\varepsilon, \hat{\phi}(\varepsilon)} = V \quad \text{locally uniformly in } [0, T] \times \text{Int}(Sd),
\] (2.25)

and, moreover, if in addition F ∈ C^{1,1}(Sd),

\[
\lim_{\varepsilon \to 0} D V_{\varepsilon, \hat{\phi}(\varepsilon)} = D V \quad \text{a.e. in } [0, T] \times \text{Int}(Sd) \quad \text{and} \quad \text{in } [L^1_{\text{loc}}([0, T] \times \text{Int}(Sd))]^d,
\] (2.26)

where, here and throughout, loc means on “any compact subsets.”

This is the most technical and demanding result of the paper, and is proved is Subsection 5.3, see Theorem 5.12. Notice that, in the end, we cannot prove convergence at any point of differentiability of V, as it would be expected from the fact that the optimal trajectories converge at any of those points, but just almost everywhere. Passing
to the limit in (2.22), equation (2.26) provides a strong form of selection for the (conditional) value of the inviscid MFG. In the above notations, it says that, for almost every \((t, p) \in [0, T] \times S_d\), the (conditional) value \(U(t, p) = (U^i(t, p))_{i \in [d]}\) of the game that is selected is given by the derivative of \(V\), namely

\[
U^i(t, p) - U^j(t, p) = d_i V(t, p) - d_j V(t, p),
\]

(2.27)

for any \(i, j \in [d]\). At first sight, it looks like that only finite differences of the vector \((U^1(t, p), \ldots, U^d(t, p))\) are hence selected. In fact, we can reconstruct \emph{a posteriori} the full-fledged collection \((U^i(t, p))_{i \in [d]}\) by observing that each \(U^i(t, p)\) should coincide with the optimal cost to (2.1) when \(p\) is the minimizer of the inviscid MFCP (which is unique for almost every initial point \((t, p)\)) and when \(q\) in (2.2) starts from the Dirac point mass in \(i\) at time \(t\). Hence we may complement Theorem 2.6 in the following way, which is also proved in Theorem 5.12:

**Theorem 2.6 (Part II).** Under the same assumption and notation as in the statement of Proposition 2.5 and provided that \(F \in C^{1,1}(S_d)\), we have

\[
\lim_{\varepsilon \to 0} U_{\varepsilon, \varphi}(x) = U \quad \text{a.e. in} \ [0, T] \times \text{Int}(S_d) \quad \text{and} \quad \text{in} \ [L^1_{\text{loc}}([0, T] \times \text{Int}(S_d))]^d,
\]

(2.28)

where, for any initial condition \((t, p) \in [0, T] \times \text{Int}(S_d)\) from which the inviscid MFCP has a unique optimal trajectory \(p, U^i(t, p) = \inf_{x} J(x; p)\) in (2.1), the problem being set over the time interval \([t, T]\) and \(q\) in (2.2) starting from \(q_t = (q^i_t = \delta_{i,j})_{i \in [d]}\).

We stress again that, in the end, we are not able to address the limit of the MFG (2.5)–(2.6), as \(\varepsilon \to 0\). In particular, we cannot prove that the master equation of the viscous potential MFG (2.20) is close to the master equation of the viscous non-potential MFG studied in [43], even though we manage to show that the additional cost \(\vartheta_{\varepsilon, \varphi}\) (which makes the difference between the two cost functionals) vanishes when computed along the equilibrium (see Proposition 2.5). In order to prove the proximity of the two master equations (with and without \(\vartheta_{\varepsilon, \varphi}\)), we would need a fine analysis of the master equation associated with the viscous potential MFG when the parameter \(\varepsilon\) is small, since singularities are expected to emerge in the neighborhood of initial conditions at which the inviscid MFCP has multiple optimizers. Such an analysis is not available in our quite general situation; it is however carried out in the simpler framework of [25], where convergence by vanishing common noise is established for linear-quadratic 1-dimensional MFGs.

### 2.5. Weak solution to the master equation

The last step in our program is hence to provide an intrinsic approach to the relationship (2.27) by addressing directly the master equation of the inviscid MFG. The latter writes (see for instance [8, 48] and [10, Chapter 7]):

\[
\begin{cases}
\partial_t U^i + H((U^i - U^j)_{j \in [d]}) + f^i(p) + \sum_{j, k \in [d]} p_k (U^k - U^j)^+ (d_j U^i - d_k U^j) = 0, \\
U^i(T, p) = g^i(p),
\end{cases}
\]

(2.29)

for \(i \in [d]\), which is informally obtained by taking \(\varphi \equiv 0\) and \(\varepsilon = 0\) in (2.20). Recast in terms of the centered (conditional) value function \((\hat{U}^i := U^i - \bar{U})_{i \in [d]}\), with \(\bar{U} = \frac{1}{d} \sum_{i \in [d]} U^i\),
\[ \frac{1}{d} \sum_{j \in [d]} U_j, \quad (2.29) \]

becomes:

\[
\begin{aligned}
\partial_t \hat{U}^i + & \quad \hat{H}^i((\hat{U}^j - \hat{U}^k)_{j,k \in [d]}) + \hat{f}^i(p) \\
& + \sum_{j,k \in [d]} p_k(\hat{U}^k - \hat{U}^j)_{+}(d_j \hat{U}^i - d_k \hat{U}^i) = 0, \\
\hat{U}^i(T, p) = & \quad \hat{g}^i(p),
\end{aligned}
\]

for \( i \in [d] \), where we have let

\[ \hat{H}^i((u^{h,k})_{j,k \in [d]}) := H((u^i - u^j)_{j \in [d]}) - \frac{1}{d} \sum_{j \in [d]} H((u^i - u^j)_{k \in [d]}), \]

and similarly \( \hat{f}^i(p) = f^i(p) - \frac{1}{d} \sum_{j \in [d]} f^j(p) \) and \( \hat{g}^i(p) = g^i(p) - \frac{1}{d} \sum_{j \in [d]} g^j(p) \). As we have already explained, the master equation is typically non-uniquely solvable (see the next subsection for a benchmark example). The question for us is thus to rephrase (2.27) as a uniqueness result for the master equation (or at least for its centered version (2.30)) within a well-chosen class of functions. Loosely speaking, we succeed to obtain such a result in Section 6 below but for the conservative form of (2.30), namely

\[
\begin{aligned}
\partial_t \hat{U}^i + & \quad \hat{H}^i((\hat{U}^j - \hat{U}^k)_{j,k \in [d]}) + \hat{f}^i(p) - \frac{1}{2} \sum_{j,k \in [d]} p_k d_i((\hat{U}^k - \hat{U}^j)_+^2) = 0, \\
\hat{U}^i(T, p) = & \quad \hat{g}^i(p),
\end{aligned}
\]

(2.31)

for \( i \in [d] \). Clearly, the two equations (2.30) and (2.31) may be identified within the class of differentiable functions \( \hat{U} \) that satisfy, for any \( i, j, k \in [d] \),

\[ -\frac{1}{2} d_i((\hat{U}^k - \hat{U}^j)_+^2) = (\hat{U}^k - \hat{U}^j)_+(d_j \hat{U}^i - d_k \hat{U}^i), \]

which indeed holds true if the Schwarz identity is satisfied, i.e. for any \( i, j \in [d] \)

\[ d_i \hat{U}^j = d_j \hat{U}^i. \]

(2.32)

As we clarify in Section 6, identity (2.32) guarantees that \( \hat{U} \) derives in space from a potential, meaning that \( \hat{U}^i(t, p) = d_i \hat{V}^\circ(t, p) \) for some real-valued function \( \hat{V}^\circ \) defined on \([0, T] \times S_d\); obviously, the fact that the coordinates of \( \hat{U} \) sum to 0 is a necessary condition to find such a potential. As a byproduct, it prompts us to regard the conservative formulation (2.31) of the master equation as the derivative system obtained by applying the operator \( d_i \), for each \( i \in [d] \), to the HJ equation (2.23). In words, (2.31) may be rewritten as

\[
\begin{aligned}
\partial_t \hat{U}^i + & \quad d_i((\sum_{k,j \in [d]} p_k H(\hat{U}^k - \hat{U}^j)) + d_i F(p) = 0, \\
\hat{U}^i(T, p) = & \quad d_i G(p),
\end{aligned}
\]

(2.33)

for \( i \in [d] \). Interestingly enough, the formulation (2.33) makes clear the link between the HJ equation (2.23) and the master equation, at least when the latter is understood in its conservative form. For scalar conservation laws, the usual notion of admissibility which is used to restore uniqueness of weak solutions is the one of entropy solution. In space
dimension 1, which is the case when \( d = 2 \) (see the next subsection), the entropy solution to a scalar conservation law is also shown to be the space derivative of the viscosity solution of the corresponding HJB equation; see e.g. [49–51]. However, for hyperbolic systems of PDEs with multiple space dimension, which is the case here when \( d \geq 3 \), there might be non-uniqueness of entropy solutions and there are very few results in the literature about such systems. In particular, system (2.33) is hyperbolic in the wide sense, but not strictly hyperbolic. Nevertheless, exploiting the connection with the HJB equation (2.23) and borrowing the idea from the paper of Krúžkov [45], it is possible to establish uniqueness in a suitable set of admissible weak solutions. This is captured by the following statement, which holds true under a weaker assumption on \( G \) than what we required in the previous statements; compare for instance with Theorems 2.1 and 2.2.

**Theorem 2.7.** Assume that \( F \) and \( G \) are in \( C^{1,1}(S_d) \). The conservative form (2.33) of the master equation has a unique weak solution that is bounded and satisfies a weak one-sided Lipschitz inequality in space. This solution is the almost everywhere space derivative of the unique viscosity solution \( V \) of the HJ equation (2.23).

The proof is given in Section 6, see Theorem 6.6. The notions of weak solution and weak one-sided Lipschitz inequality in space are clarified in Definition 6.4 below. In a nutshell, the proof of the above statement holds in three steps: The first one is to show that any weak solution to the conservative form of the master equation derives from a potential; The second one is to prove that the potential must be an almost everywhere and semiconcave solution of the HJ equation (2.23); The last step is to identify almost everywhere and semiconcave solutions with viscosity solutions of (2.23). Importantly, viscosity solutions are shown to be unique, despite the lack of boundary conditions, see Corollary 6.3 below. To put it differently, the striking facts that we use here to restore a form of uniqueness to the master equation are, on the one hand, the existence of a potential and, on the other hand, the semiconcavity assumption. In this regard, it must be fair pointing out that the existence of a potential is somewhat enclosed in the conservative form of (2.33). In other words, the conservative form not only permits to address solutions in a weak sense, but it also permits to reduce the space of solutions to gradient functions. As for the semiconcavity assumption, it plays a crucial role in the selection. To wit, the connection between semiconcave solutions of HJ equations and entropy (or one-sided Lipschitz) solutions of scalar conservation laws has been widely discussed; see for instance the first chapter in the monograph of Cannarsa and Sinestrari [49] together with the bibliography therein. In the case of hyperbolic systems with a potential structure—like (2.33)—, the role of semiconcavity is exemplified in the earlier paper of Krúžkov [45] from which we borrow part of the proof of Theorem 2.7. For sure, it is also important to say that, in the end, we are not able to define weak solutions for the non-conservative versions (2.29) and (2.30) of the master equation. However, we prove in Proposition 6.5 below that classical solutions to (2.30) are indeed weak solutions to (2.33), the key point being that Schwarz identity (2.32) holds true for classical solutions to (2.30).

### 2.6. Example

To illustrate our results, we feel useful to revisit the \( d = 2 \) example addressed in [24] (the reader may also have a look at [23] which shares many features with [24]).
Therein, a selection result is proven by addressing directly the large $N$ behavior of the $N$-player game, both in terms of the value functions for the feedback Nash equilibria and of the optimal trajectories. Although this is certainly a much more satisfactory approach than ours, at least from a modeling point view, making a detour via the finite case remains however much more challenging and difficult; see our preliminary discussion in the introduction together with the prospects addressed in the next subsection. By the way, it must be stressed that the selection result established in [24] is partial only, as it leaves open the case when the initial point of the equilibrium is precisely a singular point of the master equation. Noticeably, this argument in support of the vanishing viscosity approach is exemplified in the paper [25]: Therein, the authors prove a selection result for linear quadratic games with a continuous state space by both methods; In this setting, the vanishing viscosity method is clearly the easiest one. The reader will find further prospects regarding the connection with the $N$-player game in the next subsection.

The case $d=2$ is very special because any MFG becomes potential. Below, we first provide a general description of two state MFGs and then we specialize our results to the example analyzed in [24]. For the same two sets of coefficients $(f^i)_{i=1,2}$ and $(g^i)_{i=1,2}$ as in (2.1), we can easily reconstruct two potentials $F$ and $G$ such that

$$d_1 F(p) = \frac{f^1(p) - f^2(p)}{2} = -d_2 F(p), \quad p \in S_2,$$

by letting

$$F(p_1, 1-p_1) := \int_0^{p_1} (f^1(q, 1-q) - f^2(q, 1-q))dq, \quad p_1 \in [0,1],$$

and similarly for $G$. Interestingly the centered master equation (2.30), which is a system in the general case $d \geq 3$, becomes a mere equation when $d=2$. Indeed, we then have $\hat{U}^1 = -\hat{U}^2$, which implies that (2.30) can be rewritten in terms of the sole $\hat{U}^1 = (U^1 - U^2)/2$. Accordingly, the conservative version of the master equation takes the form

$$\begin{cases}
\partial_t \hat{U}^1 + d_1 (\mathcal{H}(p, \hat{U}^1)) + \frac{1}{2} (f^1(p_1, 1-p_1) - f^2(p_1, 1-p_1)) = 0, \\
\hat{U}^1(T, p) = \frac{1}{2} (g^1(p_1, 1-p_1) - g^2(p_1, 1-p_1)),
\end{cases}$$

(2.36)

where

$$-\mathcal{H}(p, u) = 2p_1 u^2_+ + 2(1-p_1)(-u^2_+) = 2p_1 \left(\frac{|u| + u}{2}\right)^2 + 2(1-p_1) \left(\frac{|u| - u}{2}\right)^2 = u^2 + (2p_1 - 1)u|u|.$$ 

The latter expression prompts us to change the variable $p_1$ into $m = 2p_1 - 1$ (which should be thought of as the mean of $(p_1, p_2)$ if the state space was $\{1, -1\}$ instead of $\{1, 2\}$). Letting $Z(m) := -2\hat{U}^1(\frac{1+m}{2}, \frac{1-m}{2}) = (U^2 - U^1)(\frac{1+m}{2}, \frac{1-m}{2}),$ for $m \in [-1,1]$, we can rewrite (2.36) in the form
\[ \begin{align*}
-\partial_t Z + \partial_m \left( \frac{mZ|Z|}{2} - \frac{Z^2}{2} \right) &= f^2 \left( \frac{1 + m}{2}, \frac{1 - m}{2} \right) - f^1 \left( \frac{1 + m}{2}, \frac{1 - m}{2} \right), \\
Z(T, p) &= g^2 \left( \frac{1 + m}{2}, \frac{1 - m}{2} \right) - g^1 \left( \frac{1 + m}{2}, \frac{1 - m}{2} \right),
\end{align*} \tag{2.37} \]

for \((t, m) \in [0, T] \times [-1, 1] \).

In [24], the cost coefficients are chosen as \(F \equiv 0\) and \(g^1(p) = -(2p_1 - 1)\), \(g^2(p) = 2p_1 - 1\), so that

\[
G(p_1, 1 - p_1) = \int_0^{p_1} -2(2q - 1) dq = -2p_1^2 + 2p_1 = 2p_1 p_2.
\]

The reduced master equation in [24], see (3.11) therein, is exactly Equation (2.37) for \(Z\) (up to a time reversal). Note also that the potential \(G(p_1, p_2) = -(2p_1 - 1)^2/2\) therein differs from \(G\) above by a constant (which is 1/2), but obviously this does not matter. Importantly, the master equation (2.37) may have multiple weak solutions when \(T\) is large enough, hence the need for a selection argument. The solution selected in (2.37), following the theory for scalar conservation laws, is the entropy solution, which can be shown to be unique in this case despite the lackness of boundary conditions. As explained in the previous subsection, the entropy solution is the space derivative of the viscosity solution to the HJB equation, making this selection consistent with Theorem 2.7. Moreover, in [24], the value functions for the feedback Nash equilibrium of the \(N\)-player game are shown to converge to this entropy solution [24, Theorem 8]; this says that the solutions to the master equation that are selected by taking the limit over \(\epsilon\) or over \(N\) are the same. So, in a nutshell, our result is fully consistent with [24].

As far as convergence of the optimal trajectories is concerned, the equilibria are shown to be non-unique, provided that the time horizon \(T\) is chosen large enough: Whatever the initial condition at time 0, there are three solutions to the MFG if \(T > 2\); see [24, Proposition 2]. In this regard, the main result in [24] states that, whenever the initial condition \(p_0 = (p_0^1, p_0^2)\) of the population at time 0 satisfies \(m_0 := 2p_0^1 - 1 \neq 0\) (i.e., the mean parameter is non-zero), there is a unique equilibrium \((p_t)_{0 \leq t \leq T}\) that is selected by letting \(N\) tend to \(\infty\) in the corresponding \(N\)-player game. It satisfies the equation

\[
\frac{d}{dt} m_t = -2m_t Z(t, m_t) + 2Z(t, m_t), \quad t \in [0, T], \tag{2.38}
\]

with \((m_t = p_t^1 - p_t^2)_{0 \leq t \leq T}\) and \(Z\) being the unique entropy solution to (2.37); see [24, (23)]. Notably, this equation is shown to admit a unique solution, when \(m_0 \neq 0\); see [24, Proposition 6]. Again, this is consistent with our results: [24, Theorem 15] asserts that this equilibrium is also the unique minimizer of the corresponding inviscid MFCP initialized from the mean \(m_0\) at initial time 0 (recalling that choosing \(m_0\) is the same as choosing \(p_0^1\)), see (2.10) plugging \(F \equiv 0\) and \(G(p) = 2p_1 p_2\) therein. While the proof of Theorem [24, Theorem 15] is carried out by explicit computations, our Theorem 2.4 applies directly. Interestingly, we may recover (2.38) explicitly. Indeed, in [24], the function \(m \rightarrow Z(0, m)\) is shown to be discontinuous at \(m = 0\) only (provided that \(T\) is large enough; if \(T\) is small, \(m \rightarrow Z(0, m)\) is continuous). Also, the function \(m \rightarrow V(0, \frac{1 - m}{2})\) in (2.23) is continuously differentiable at \(m_0\) (since \(m_0\) is assumed to be non-zero) which,
as we already explained (see also Proposition 5.2), implies that there is indeed a unique minimizer to the MFCP initialized from \( \left( 0, \frac{1-m_0}{2} \right) \). Also, our discussion (see (viii) in Proposition 5.2) says that this unique minimizer, say \((p_t^*)_{0 \leq t \leq T}\), solves the equation

\[
\frac{d}{dt} p_t^{*,1} = (1-p_t^{*,1})(\hat{U}^2(t,p_t^*) - \hat{U}^1(t,p_t^*))_+ - p_t^{1,*}(\hat{U}^1(t,p_t^*) - \hat{U}^2(t,p_t^*))_+,
\]

for \( t \in [0,T] \). Letting \((m_t^* := 2p_t^{1,*} - 1)_{0 \leq t \leq T}\), we easily derive that \(m_t^*\) solves Equation (2.38), whence we get \(m_t = m_t^*\).

Last but not least, the case \(m_0 = 0\) is left open in [24]. In that case, the three MFG solutions consist of a constant equilibrium, equal to 0, and two non-trivial symmetric equilibria. The latter two are also the minimizers of the inviscid MFCP; see again [24, Theorem 15]. It is also claimed in [24], see Section 4 therein, that, numerically, equilibria of the \(N\)-player game are tending to converge in law to those two minimizers, with weight 1/2 each. In other words, there is a numerical evidence for ruling out the zero equilibrium. Obviously, our Theorem 2.4 sounds as a confirmation of this latter intuition, as it precisely says that the zero equilibrium is indeed excluded by the vanishing viscosity method. The fact that the two remaining ones should be charged with probability 1/2 each comes from an additional symmetry argument, which is similar to the one used in [25]: If \(p^{*,1,e,\varphi} = (p^{*,1,e,\varphi}, p^{*,2,e,\varphi})_{0 \leq t \leq T}\) is an optimal trajectory of the viscous MFCP, then, thanks to the symmetric form of \(G\), \((p^{*,1,e,\varphi}, p^{*,2,e,\varphi})_{0 \leq t \leq T}\) is an admissible path with the same cost and hence is also an optimal trajectory but for the common noise \((B^{1,2}_t, B^{2,1}_t)_{0 \leq t \leq T}\) (instead of \((B^{1,2}_t, B^{2,1}_t)_{0 \leq t \leq T}\)). By uniqueness in law of the equation characterizing the optimal trajectory, this shows that \((p^{*,1,e,\varphi}, p^{*,2,e,\varphi})_{0 \leq t \leq T}\) and \((p^{*,2,e,\varphi}, p^{*,1,e,\varphi})_{0 \leq t \leq T}\) have the same distribution. Consequently, under any weak limit \(\mathbb{M}\) as in the statement of Theorem 2.4, the marginal law of the first variable—which must be understood as the law of \(p\)—has to be symmetric. Here, we know that the support of \(\mathbb{M}\) is necessarily included in a set of two non-trivial trajectories. Hence, each of them should be charged with probability 1/2.

Obviously, the thrust of our approach is that it applies to more general coefficients \(F\) and \(G\) and to any number of states \(d \geq 2\). Of course, the symmetry argument we have just alluded to only applies under appropriate assumptions.

### 2.7. Prospects and extensions

We now provide some additional observations that may be useful for the reader.

In line with the discussion about the case \(d = 2\), we notice that, in [24], the entropy characterization of the admissible solution to the master equation is not used to prove the convergence of the \(N\)-Nash system. This point is emphasized in the concluding remarks therein. Instead, the authors use the particular structure of the example under study. However, this characterization based on entropy solutions might be useful to address the same convergence problem but for more general two state MFGs. Similarly, the characterization of the unique solution given by Theorem 2.7 (see also Definition 6.4) might help to establish the convergence of the \(N\)-player value functions for potential MFGs with \(d \geq 3\) states. This remains however a difficult open question, even in the case \(d = 2\), and we leave it to future works.
We now outline another possible way to clarify the connection between our selection result and the \(N\)-player game associated with the potential MFG. As we already alluded to, we indeed cooked up in a recent preprint [46] a relevant version of an \(N\)-player game with a Wright-Fisher common noise that converges, as \(N\) grows up, to the unique equilibrium of the non-potential viscous MFG, it being understood that both \(\varepsilon\) and \(\phi\) therein are fixed. Quite clearly, a similar \(N\)-player game can be associated with the potential viscous MFG, taking into account the additional \(\partial_{\varepsilon, \phi}\) in the cost functional (2.14). By combining those results with the results obtained here, we deduce that, if we start from the \(N\)-player game associated with the potential viscous MFG (with the appropriate choice of \(\partial_{\varepsilon, \phi}\)), let \(N\) tend to \(\infty\) first and then take the same limit over \(\varepsilon\) and \(\phi\) as in Theorem 2.4, then the solutions of the inviscid MFG that are selected in this way must be minimizers of the inviscid MFCP. For sure, if ever we could prove that the distance between the equilibria of the \(N\)-player games with and without common noise becomes small, uniformly in \(N\), as \(\varepsilon\) tends to 0 and the support of \(\phi\) shrinks, then we would get a commutative diagram and hence prove that the limit points of the equilibria of the \(N\)-player game are solutions of the inviscid MFCP. This looks however a very challenging strategy and there is no clear evidence that this should be indeed feasible. A simpler problem would consist in allowing \(N\) to depend on \(\varepsilon\) and \(\phi\) and then to find some \(N^{\varepsilon, \phi}\) (with \(N \to \infty\) as \(\varepsilon \to 0\) and the support of \(\phi\) shrinks) in such a way that the limit points of the equilibria (along subsequences) of the \(N^{\varepsilon, \phi}\)-player game are indeed minimizers of the inviscid MFCP.

On another matter, we emphasize that all the results of the paper can still be proved if, instead of being purely quadratic, the cost functional of the inviscid MFG in (2.2)–(2.4) satisfies the following more general conditions:

i. The analogue of the Lagrangian (2.4) writes in the form 
\[
(L^i(x) = L^i((x_{i,j})_{j \neq i}))_{i \in [d]},
\]
where, for each \(i \in [d]\), \(L^i\) is uniformly convex on the \((d - 1)\)-dimensional orthant;

ii. The transition rates underpinning the inviscid MFG are of the form \(\alpha_{i,j} + b^{i,j}(p)\), for \(j \neq i\), for an extra Lipschitz-continuous drift \(b : S_d \to \mathbb{R}^{d \times d}\), with the same sign constraints as before that \(\alpha_{i,j} \geq 0\) and \(b^{i,j}(p) \geq 0\);

iii. For each \(i \in [d]\), the Hamiltonian (\(\text{Tex translation failed}\)) is \(C^{1,1}\) in the adjoint variable \(u = (u_j)_{j \in [d]} \in \mathbb{R}^d\).

Moreover, in order to guarantee that the two Theorems 2.6 and 2.7 remain true, the following is also needed:

i. The reduced Hamiltonian associated with the inviscid MFCP, namely
\[
\widehat{S}_d(x_1, \ldots, x_{d-1}) \mapsto \sum_{i,j \in [d-1]} (x_j(x_{j,i} + b^{i,j}(\bar{x})) - x_i(x_{i,j} + b^{i,j}(\bar{x})))
+ \sum_{i \in [d-1]} x^{-d}(x_{d,i} + b^{d,i}(\bar{x})) - x_i(x_{i,d} + b^{i,d}(\bar{x}))
+ \sum_{i \in [d-1]} x_iL^i((x_{i,j})_{j \neq i}) + x^{-d}L^d((x_{d,j})_{j \in [d-1]}),
\]
with \(x^{-d} := 1 - (x_1 + \cdots + x_{d-1})\) and \(\bar{x} := (x_1, \ldots, x_{d-1}, x^{-d})\), is strictly convex in the adjoint variable on the interior \(\text{Int}(\hat{S}_d)\) of the simplex when the corresponding state
trajectories are defined in $\hat{S}_d$; see (5.9) for the point where this reduced Hamiltonian is used in the quadratic case.

Of course, we prefer to work with the purely quadratic Hamiltonian since it simplifies the notations.

Lastly, we point out that a related work ([52]) appeared after the submission of the first version of the paper. Therein, a priori Hölder estimates are proven for Kimura equations set on the orthant and driven by merely bounded drifts that are pointing inwards at the boundary of the domain. Although this pointing condition is consistent with our requirement to include the forcing term $\varphi$ in the dynamics (2.5), it is less demanding from a purely quantitative point of view. In words, provided that the results from [52] can be applied to the simplex (and not only to the orthant) and then that the analysis carried out in [43] for solving the master equation of the viscous MFG can be repeated, this would allow us to work with a lower threshold $\kappa_\varepsilon$ in (2.17). Our guess is that $\kappa_\varepsilon$ could be in fact chosen independently of $\varepsilon$. For sure, this would not change our selection results, but this would allow for slightly simpler computations in Section 5 (anyway, estimates for the reciprocal of $\varphi$, which are the most demanding, would be still needed since our definition of the MFG indeed involves the reciprocal of $\varphi$).

3. Mean field control problem

The main goal of this section is to prove Theorem 2.1. We feel useful to recall that, for a function $\varphi$ as in (2.7), we aim at minimizing $J^{\varepsilon, \varphi}(x)$ in (2.11) where $p = (p_t)_{0 \leq t \leq T}$ therein solves the $x$-driven SDE (2.12). Importantly, the pair $(\varepsilon, \varphi)$ is kept fixed throughout the section, which prompts us to drop out the superscript $(\varepsilon, \varphi)$ in the subsequent notations. As explained in the previous section, we restrict ourselves to processes $x$ that are bounded by $M = 2(\|g\|_\infty + T\|f\|_\infty)$, in the sense that $|x_i^{(t)}| \leq M, dt \otimes P$ almost everywhere, for any $i, j \in [d]$ with $i \neq j$. The bound $M$ has the following interpretation in terms of the inviscid MFG (2.1)-(2.2): For a given (deterministic) path $p = (p_t)_{0 \leq t \leq T}$ with values in $S_d$, optimizers of (2.1) are given in terms of the value function $((u_i^t)_{0 \leq t \leq T})_{i \in [d]}$, namely $x_i^{(t)} = (u_i^t - u_j^t)_+$, for $t \in [0, T]$ and $i, j \in [d]$ with $i \neq j$; see [10, Chapter 7]. Here, $u_i^t$ is defined as the optimal cost when $q$ starts at time $t$ from the initial condition $q_0^t = \delta_{i,j}$ and hence satisfies $\|u_i^t\| \leq T\|f\|_\infty + \|g\|_\infty$ : the upper bound holds by choosing the zero control, while the lower bound follows from the sign of $L$. With the same meaning for $a*$ as in (2.16), this allows us to express the corresponding Hamiltonian in the form

$$\tilde{H}_M(p, w) := \inf_{(x_i)_{i \in [d]}} \sum_{i \in [d]} p_i \tilde{H}_M^i(w),$$

(3.1)

for $p \in S_d$ and $w = (w_i)_{i \in [d]} \in \mathbb{R}^d$, with

$$\tilde{H}_M(p, x, w) := \sum_{i \in [d]} p_i \sum_{j \in [d]; j \neq i} \left(x_i(w_j - w_i) + \frac{1}{2} |x_i|^2\right)$$

$$\tilde{H}_M^i(w) := \inf_{(x_j)_{j \in [d]; j \neq i}} \sum_{j \in [d]} \left(x_j(w_j - w_i) + \frac{1}{2} |x_j|^2\right)$$

(3.2)

$$= \sum_{j \neq i} \left\{ a^*(w_i - w_j)(w_j - w_i) + \frac{1}{2} |a^*(w_i - w_j)|^2 \right\}.$$
By boundedness of \( a^* \) (which in turn follows from our choice to restrict ourselves to controls that are bounded by \( M \)), \( \check{H}_i^j \) is Lipschitz continuous and continuously differentiable with Lipschitz and bounded derivatives (pay attention that it is not \( C^2 \)). Indeed, a simple computation shows that 
\[
\check{H}_i^j(w) = \sum_{j \neq i} \check{S}(w_i - w_j),
\]
where, for \( r \in \mathbb{R} \),
\[
\check{S}(r) = \begin{cases} 
0 & r \leq 0, \\
- \frac{1}{2} r^2 & 0 \leq r \leq M, \\
- rM + \frac{1}{2} M^2 & r \geq M,
\end{cases}
\]
and \( \check{S}'(r) = -a^*(r), \quad r \in \mathbb{R} \).

The Hamiltonian \( \check{H}_i^j \) is used in the rest of the paper. The HJB equation for the value function is nothing but (2.13), with \( \mathcal{H}^0_M \) therein given by
\[
\mathcal{H}^0_M(p, w) := \mathcal{H}_M(p, w) + \sum_{i \in [d]} \sum_{j \neq i} p_i \varphi(p_j)(w_j - w_i). \tag{3.3}
\]
The following is straightforward but useful for us:
\[
\partial_w \mathcal{H}^0_M(p, w) = \sum_{j \in [d]} \sum_{j \neq i} p_j (\varphi(p_i) + a^*(w_j - w_i)) - p_i \sum_{j \in [d]} (\varphi(p_j) + a^*(w_i - w_j)). \tag{3.4}
\]

**Remark 3.1.** Note that, in Section 2, see (2.21), we preferred to use the slightly different Hamiltonian \( H_M \) instead of \( \check{H}_i^j \), but the two are clearly related by the identity 
\[
\check{H}_i^j(w) = H_M((w_i - w_j)_{i \in [d]}), \quad w \in \mathbb{R}^d.
\]

### 3.1. Classical solutions

The well-known verification argument may be easily adapted to the simplex: If there exists a classical solution \( \mathcal{V} \) to the HJB equation, then the optimal control is unique (clearly bounded), if the initial condition is in the interior of the simplex, and given in feedback form through the feedback function \( \check{z}^*: \mathcal{V} \to D \mathcal{V} \). The proof proceeds in the same way, by expanding \( \mathcal{V} \) along the trajectories by means of Itô formula, and by using the fact that solutions to (2.12) remain in \( \text{Int}(\mathcal{S}_d) \) (which makes it possible to use interior smoothness of \( \mathcal{V} \) and uniform convexity of the Hamiltonian on \( \text{Int}(\mathcal{S}_d) \)).

Although intrinsic derivatives are the most canonical ones, and will hence be used in the next sections, a key tool to prove the well-posedness of the HJB equation (2.13) is to work with local charts. In this respect, it is worth recalling that any function \( h \) defined in the simplex \( \mathcal{S}_d \) may be easily regarded as a function defined on the set \( \check{S}_d \). It suffices to identify \( h \) with \( \check{h} \) defined by
\[
\check{h}(x) := h(\check{x}), \quad \check{x} := (x_1, \ldots, x_{d-1}, 1 - (x_1 + \cdots + x_{d-1})), \quad \check{x} \in \check{S}_d. \tag{3.5}
\]

As explained in [43], \( h \) is then once or twice differentiable on the (interior of) the simplex if \( \check{h} \) is once or twice differentiable in the usual sense as a function defined on an open subset of \( \mathbb{R}^{d-1} \), in which case we have a dictionary to pass from \( Dh \) and \( D^2h \) to \( D\check{x} \check{h} \) and \( D^2\check{x} \check{h} \) and conversely, whereas \( D_x \) is the gradient in the variable \( x \in \mathbb{R}^{d-1} \) and \( D^2_x \) is the Hessian. In short, 
\[
\partial_{\check{x}} \check{h}(x) = D_x h(\check{x}) - D_d h(\check{x}) = D_x h(\check{x}) + \sum_{j \in [d-1]} D_j h(\check{x}), \quad \text{for } i \in [d-1] \text{ and } x \in \text{Int}(\check{S}_d), \text{ and conversely } D_i h(p) = \left( \partial_{\check{x}} \check{h} - \frac{1}{d} \sum_{j \in [d-1]} \partial_{\check{x}} \check{h} \right)(p_1, \ldots, p_{d-1}), \quad \text{for } i \in [d-1],
\]
and $d_i h(p) = -\frac{1}{d} \sum_{j \in [d-1]} \partial_i \hat{h}(p_1, \ldots, p_{d-1})$, for $p \in \text{Int}(S_d)$. As for the second order derivatives, $\partial_{x_i x_j} \hat{h}(x) = d_{ij} h(\tilde{x}) - d_{j} d_{i} h(\tilde{x}) + d_{ii} h(\tilde{x})$, for $i, j = 1, \ldots, d - 1$. As a byproduct, the HJB has the following writing in local charts (sums being taken over $[d-1]$):

$$
\begin{aligned}
\partial_t \hat{V} + \hat{H}_M(x, D_x \hat{V}) + \hat{F}(x) + \frac{1}{2} \varepsilon^2 \sum_{j,k} (x_j \partial_{j,k} - x_k x_j) \partial_{x_j x_k} \hat{V} \\
+ \sum_i x_i \left[ \sum_j \phi(x_j) (\partial_x \hat{V} - \partial_{x_j} \hat{V}) - \phi(x^d) \partial_{x_j} \hat{V} \right] + x^{-d} \sum_j \phi(x_j) \partial_{x_j} \hat{V} = 0,
\end{aligned}
$$

(3.6)

for $t \in [0, T]$ and $x \in \text{Int}(\hat{S}_d)$, where $\hat{H}_M(x,z) = H_M(\tilde{x}, z) = \sum_{j \in [d-1]} x_j \hat{H}_M^j(z) + x^{-d} \hat{H}_M^d(z)$. Here, $\hat{H}_M^i$, for $i = 1, \ldots, d$, is the Hamiltonian

$$
\hat{H}_M^i(z) := \hat{H}_M^i(\Theta(z)), \quad \Theta(z) = \left( z_1 - \frac{1}{d} \sum_{j=1}^{d-1} z_j, \ldots, z_{d-1} - \frac{1}{d} \sum_{j=1}^{d-1} z_j, - \frac{1}{d} \sum_{j=1}^{d-1} z_j \right),
$$

(3.7)

for $z = (z_1, \ldots, z_{d-1}) \in \mathbb{R}^{d-1}$ and we denote $x^{-d} = 1 - \sum_{j=1}^{d-1} x^j$; we refer to [43] for the derivation of the second order term, see Equation (3.3) therein. Interestingly enough, the optimal feedback then writes (in local chart) in the form (provided that the HJB equation has a classical solution) $\left( \hat{a}^\ast_{i,j}(D_x \hat{V}) \right)_{i,j \in [d], i \neq j}$ with (recall the definition of $a^\ast$ in (2.16))

$$
\hat{a}^\ast_{i,j}(z) = \begin{cases} 
    a^\ast(z_i - z_j), & i, j \in [d-1], \\
    a^\ast(z_i), & j = d, \\
    a^\ast(-z_j), & i = d.
\end{cases}
$$

(3.8)

We remark that, if the value function is in the Wright-Fisher space $C^{1,2+\varepsilon}_{WF}([0, T] \times S_d)$ (to which we already alluded and which is defined in more details in the appendix), then $V$ solves (2.13) if and only in $\hat{V}$ solves (3.6).

We choose to express the last coordinate in terms of the first $d-1$ for convenience only, and in fact the choice of the local chart is arbitrary. This is one reason why we expressed the main results in terms of intrinsic derivatives. Anyhow, the local chart is more adapted to the proof of Theorem 3.2 below. Indeed, it is worth emphasizing that, in order to prove the well-posedness of (2.13), it is enough to check that, provided that it belongs to the right space, $\hat{V}$ solves (3.6) in the interior of the simplex for the fixed chart we have chosen. In this regard, the precise choice of the local chart is not of a great importance and expressing any other coordinate than $x_i$ in terms of the other ones would work as well; to wit, by the same arguments as in [43, Subsection 3.2.3], Equation (3.6) can be equivalently written in terms of another local chart. This claim holds also for the derivative systems (3.9) and (3.10) that we introduce below.

### 3.2. Derivative system

In order to address the HJB equation (2.13), we first study the derivative system. The rationale to do so is that, obviously, the nonlinear term in the derivative system is of
order zero only while it is of order one in the HJB equation. As a byproduct, it makes it possible to apply \textit{a priori} estimates proved in [43, Theorem 2.10]. As explained above, we can use both intrinsic derivatives and local charts. Before we state the corresponding forms of the derivative system, we caution that the computations in the derivation of (3.9) and (3.10) below are rather tedious; anyhow, there is nothing difficult and we feel it is sufficient to just provide the final results. Taking the derivative in (2.13) (by means of (3.4)) and applying the Schwarz identity \( d_i V^i = d_j V^j \), we formally get the following expression for \( V = D V \) (the indices in the sums below belonging to \([d]\)):

\[
\begin{align*}
\partial_t V^i + \tilde{H}_M^i(V) - \frac{1}{d} \sum_j \tilde{H}_M^j(V) + \sum_j (\varphi(p_j) - p_j \varphi'(p_j))(V^j - V^i) \\
- \frac{1}{d} \sum_j \sum_l (\varphi(p_j) - p_j \varphi'(p_j))(V^j - V^l) + f^i(p) - \frac{1}{d} \sum_l f^l(p) \\
+ \sum_{j,k} p_k(\varphi(p_j) + a^*(V^k - V^i))(d_j V^i - d_k V^i) + \frac{1}{2} \varepsilon^2 \sum_{j,k} (p_j \delta_{j,k} - p^j p^k) d_{jk}^2 V^i \\
+ \frac{1}{2} \varepsilon^2 \left( d_i V^i - 2 \sum_j p_j d_j V^i - \frac{1}{d} \sum_j d_j V^i \right) = 0, \\
V^i(T, p) = g^i(p) - \frac{1}{d} \sum_j g^j(p),
\end{align*}
\]

(3.9)

where \( \tilde{H}_M^i(V) \) is defined by (3.7). Instead, differentiating (3.6) with respect to \( x \) (using in the sequel the generic notation \( Z \) for \( D_x \tilde{V} \)) and applying the Schwarz identity \( \partial_x Z^i(t,x) = \partial_{x,i} Z^i(t,x) \), for \( i,j \in [d-1] \), we then get, at least formally, the following system of equations (all the sums below are taken over \([d-1]\)):

\[
\begin{align*}
\partial_t Z^i + \tilde{H}_M^i(Z) - \tilde{H}_M^d(Z) + f^i(x) - f^d(x) + \sum_j (\tilde{b}^j(x,Z) + \frac{1}{2} \varepsilon^2 \delta_{ij} - \varepsilon^2 x_j) \partial_{x,i} Z^j \\
+ \sum_{j,k} \tilde{c}^l(x) Z^j + \frac{1}{2} \varepsilon^2 \sum_{j,k} (x_j \delta_{j,k} - x_j x_k) \partial_{x,j} Z^j = 0, \\
Z^i(T, x) = \tilde{g}^i(x) - \tilde{g}^d(x),
\end{align*}
\]

(3.10)
on \([0,T] \times \text{Int}(\hat{S}_d)\), for \( i \in [d-1] \), where, for \( j \in [d-1] \) and \( z = (z_k)_{k \in [d-1]} \in \mathbb{R}^d \),

\[
\begin{align*}
\tilde{c}^l(x) &= (\varphi'(x_l) - \varphi(x^d) - \sum_{k \in [d-1]} \varphi(x_k)) \delta_{l,j} + (\varphi'(x^d) - \varphi'(x_l)) x_l \\
\tilde{b}^l(x,z) &= \sum_{k \in [d-1]} \left[ x_k [\varphi(x_j) + a^*(z_k - z_j)] - x_j [\varphi(x_k) + a^*(z_j - z_k)] \right] \\
&+ x^d \left[ \varphi(x_j) + a^*(z_j) \right] - x_j \left[ \varphi(x^d) + a^*(z_j) \right].
\end{align*}
\]

(3.11)

The two equations are equivalent, by using the identities \( Z^i = \hat{V}^i - \hat{V}^d \), \( \hat{V}^i = Z^i - \frac{1}{d} \sum_{j=1}^{d-1} Z^j \) and \( \hat{V}^d = -\frac{1}{d} \sum_{j=1}^{d-1} Z^j \), given by the aforementioned dictionary to pass from one derivative to another.

Here, we prove well-posedness of (3.10), because it is needed for solving the HJB equation (3.6). Recalling the shape of \( \varphi \) from (2.7), our main solvability result is:

\textbf{Theorem 3.2.} If \( f \in [C_{WF}^0(S_d)]^d \) and \( g \in [C_{WF}^{0,2\gamma}(S_d)]^d \) for a given \( \gamma \in (0,1) \), then there exists a constant \( K_1 > 0 \) only depending on \( M, T \) and \( d \), such that for any \( \varepsilon \in (0,1], \theta > 0 \) and \( \kappa \geq \kappa_1 / \varepsilon^2 \), there exists \( \gamma' \in (0,\gamma] \), possibly depending on \( \varepsilon \) and \( \kappa \), such that \textbf{Equation (3.10) admits a unique solution in} \([C_{WF}^{0,2\gamma}(\mathbb{R} \times S_d)]^d\).
Proof. The proof of existence is done via Leray-Schauder fixed point theorem. Throughout, the value of $\varepsilon \in (0,1]$ is given; the value of $\theta > 0$ in (2.7) is also given and the value of $\kappa$ in the same formula is fixed next. Also, we let $\gamma' \in (0,\gamma]$ to be chosen later. Letting\(^1\) $\mathcal{X} = \mathcal{C}^{0,\gamma'}_{WF}([0,T] \times \hat{S}_d)$, we consider the map $\Phi : \mathcal{X}^{d-1} \to \mathcal{X}^{d-1}$, defined by $\Phi(Z) = Y^i$, where $Y^i$ is the solution to the linear equation obtained by freezing the zero order terms in (3.10) (all the sums being taken over $[d-1]$):

$$
\begin{align*}
\partial_t Y^i + \sum_j (\hat{b}^i(x,Z) + \frac{1}{2} e^2 \delta_{i,j} - e^2 x_j) \partial_{x_j} Y^i + \frac{1}{2} e^2 \sum_{j,k} (x_j \delta_{i,k} - x_j x_k) \partial^2_{x_jx_k} Y^i &= -\left[\hat{H}^i_M(Z) - \hat{H}^d_M(Z) + \hat{f}^i(x) - \hat{f}^d(x) + \sum_j \hat{c}^i_j(x)Z_j\right] \\
Y^i(T,x) &= \hat{g}^i(x) - \hat{g}^d(x).
\end{align*}
$$

(3.12)

The key remark is that, once $Z$ is given, this is a scalar equation for each $Y^i$, in the sense that there is no $Y^j$, $j \neq i$, in the equation. Therefore we are allowed to invoke Theorem 10.0.2 of [44], which states that there exists a unique solution $Y^i \in \mathcal{C}^{0,2+\gamma'}_{WF}([0,T] \times \hat{S}_d)$ to (3.12), for any $i$, if the right hand side and the drift belong to $\mathcal{C}^{0,\gamma'}_{WF}([0,T] \times \hat{S}_d)$ and the terminal condition is in $\mathcal{C}^{0,2+\gamma'}_{WF}((0,T) \times \hat{S}_d)$. Such assumptions are satisfied in the present situation because $\hat{H}^i_M - \hat{H}^d_M$ and $a^*$ (which shows up in $\hat{b}$, see (3.11)) are Lipschitz continuous and $\varphi$ and $\varphi'$ are bounded and Lipschitz; thus the map $\Phi$ is well-defined. The claim hence follows if $\Phi$ admits a fixed point. In order to apply Leray-Schauder fixed point theorem we must show that $\Phi$ is continuous and compact and that the set

$$
\mathcal{X} = \{Z \in \mathcal{X}^{d-1} : Z = \lambda \Phi(Z) \text{ for some } \lambda \in (0,1]\}
$$

is bounded in $\mathcal{X}^{d-1}$, when the product space $\mathcal{X}^{d-1}$ is equipped with the norm $||Z||_{\mathcal{X}^{d-1}} = \max_{j \in [d-1]} ||Z^j||_{WF,0,\gamma'}$, for $Z \in \mathcal{X}^{d-1}$.

**Step 1.** We first show that $\Phi$ is continuous and compact. To do so, we may restrict ourselves to inputs $Z$ such that $\max_{j \in [d-1]} ||Z^j||_{WF,0,\gamma'}$ is less than some arbitrarily fixed real $R > 0$. Then, Theorem 10.0.2 of [44] gives, for any $i \in [d-1]$,

$$
||Y^i||_{WF,0,2+\gamma'} \leq C_R(\max_{j \in [d-1]} ||Z^j||_{WF,0,\gamma'} + ||f^i - f^d||_{WF,0,\gamma'} + ||g^i - g^d||_{WF,0,2+\gamma'}),
$$

(3.13)

for some constant $C_R \geq 0$ depending on $R$ through the drift $\hat{b}(x,Z)$ in (3.12), which yields (up to a new value of $C_R$ that may also depend also on the other inputs of the PDE)

$$
\max_{i \in [d-1]} ||Y^i||_{WF,0,2+\gamma'} \leq C_R.
$$

(3.14)

The above inequality implies that the map $\Phi$ is compact, as $\mathcal{C}^{0,2+\gamma'}_{WF}([0,T] \times \hat{S}_d)$ is compactly embedded in $\mathcal{C}^{0,\gamma'}_{WF}([0,T] \times \hat{S}_d)$, see the appendix. To prove continuity, we consider the analogue of (3.13), but applied to $Y - Y'$ with $(Y,Y') = (\Phi(Z),\Phi(Z'))$, for $(Z,Z') \in (\mathcal{X}^{d-1})^2$. Again, we assume that $\max_{j \in [d-1]} ||Z^j||_{WF,0,\gamma'}$ and

\(^1\)Our notation for the Wright-Fisher space here is a bit abusive since it is regarded as a space of functions on $[0,T] \times \hat{S}_d$, but, as we already explained, there is no difficulty in passing from functions defined on $[0,T] \times \hat{S}_d$ to functions defined on $[0,T] \times \hat{S}_d$, and conversely.
max_{j \in [d-1]} \| (Z_i')' \|_{WF, 0, \gamma'} are less than \( R \). So, using (3.14) together with the fact that the derivatives of \( \tilde{H}_M - \tilde{H}_M^d \) are Lipschitz, we have
\[
\| \Phi(Z') - \Phi(Z) \|_{\mathcal{X}^{d-1}} \leq \max_{j \in [d-1]} \| (Y')' - Y' \|_{WF, 0, 2+\gamma'} \leq C_R \| X' - Z' \|_{\mathcal{X}^{d-1}},
\]
which proves continuity.

**Step 2.** We now prove an \( L^\infty \) bound of \( \xi \). For \( Z \in \mathcal{X} \), we have, for some \( \lambda \in (0, 1] \),
\[
\begin{aligned}
\partial_t Z_i^j + \sum_{i,j}(b_i^j(x, Z) + \frac{1}{2} \varepsilon^2 \delta_i - \varepsilon^2 x_j) \partial_x Z_i^j + \frac{1}{2} \varepsilon^2 \sum_{i,k}(\delta_{i,k} - x_j x_k) \partial_{x} Z_i^j \\
= -\lambda \left[ \tilde{H}_M^d(Z) - \tilde{H}_M^d(Z) + f_i^j(x) - \tilde{f}_i^j(x) + \sum_j \tilde{c}_{i,j}^j(x)Z_i \right] \\
Z_i^j(T, x) = \hat{\lambda}(\hat{g}^j(x) - \hat{g}^j(d(x))).
\end{aligned}
\]
(3.15)
The proof follows from a standard representation of \( Z \) along the solution of the SDE that is driven by the second-order differential operator appearing in (3.15). To make it clear, we have, for any \( i \in [d-1] \) and \((t, x) \in [0, T] \times \text{Int}(\hat{S}_d),
\[
Z_i^j(t, x) = \lambda \mathbb{E} \left[ \int_t^T f_i^j(X_i^j, Z(s, X_i^j)) \, ds + \hat{g}(X_i^j) \right],
\]
(3.16)
where, for convenience, we have let \( Z(s, X_i^j) := (Z_i^j(s, X_i^j))_{j \in [d-1]} \) together with
\[
\begin{aligned}
\hat{g}^j(x) &= \hat{g}^j(x) - \hat{g}^j(d), \\
\hat{f}_i^j(x, z) &= \hat{H}_M^d(z) - \hat{H}_M^d(z) + f_i^j(x) - \tilde{f}_i^j(x) + \sum_{j,i} \tilde{c}_{i,j}^j(x)z_j,
\end{aligned}
\]
(3.17)
for \( x \in \hat{S}_d \) and \( z \) in \( \mathbb{R}^{d-1} \). In (3.16), \( X^i \cdot = (X_i^j, (X_i^j)_{t \leq s \leq T})_{j \in [d-1]} \) denotes a \((d-1)\)-dimensional process solving the SDE
\[
dX_i^j = (b_i^j(X_i^j, Z(s, X_i^j)) + \frac{\varepsilon^2}{2} \delta_i - \varepsilon^2 X_i^j) \, ds \\
+ \frac{e}{\sqrt{2}} \left\{ \sum_{k \in [d-1]} X_i^j \sqrt{X_i^j} \, dW_i^k - W_i^k + \sqrt{X_i^j} X_i^j \, dW_i^d - W_i^d \right\}.
\]
(3.18)
for \( t \leq s \leq T \), with initial condition \( X_i^j = x \), where we have denoted \( X_i^{d-1} = 1 - \sum_{j=1}^{d-1} X_i^j \).

Representation (3.16) follows from the fact that \( Z_i \in C^{1,2}([0, T] \times \text{Int}(\hat{S}_d)) \) (which is here the usual space of functions that are once continuously differentiable in time and twice in space) and hence from Itô’s formula applied to \( (Z_i^j(s, X_i^j))_{t \leq s \leq T} \), provided that the solution to (3.18) remains in \( \text{Int}(\hat{S}_d) \). Assume for a while that the latter holds true. Then, having (3.16) (together with the notations (3.11) and (3.17)), we exploit the Lipschitz continuity of \( (\tilde{H}_M^i)_{i \in [d]} \), the boundedness of \( \varphi \) and \( \varphi' \), the fact \( \lambda \leq 1 \), and the uniform bounds on \( f \) and \( g \) to obtain
\[
|Z_i^j(t, x)| \leq \| \hat{g}^j - \hat{g}^d \|_\infty + T \| f_i^j - \hat{f}_i^j \|_\infty + C \int_t^T \max_{i \in [d-1]} \sup_{x' \in \text{Int}(\hat{S}_d)} \| Z_i^j(s, x') \| \, ds.
\]
(3.19)
Taking the supremum over \( x \in \text{Int}(\hat{S}_d) \) and the maximum over \( i \in [d-1] \) in the left-hand side and applying Gronwall’s lemma, we get a bound for
The key fact is then to observe that, whenever $X_i^{j,-d}$ solves (noticing that the sum over $j$ in the first line in the definition (3.11) of $\bar{b}$ is null and similarly for the first term in the second line of (3.18))

$$dX_i^{j,-d} = \left\{ \sum_{j \in [d-1]} X_i^{j,0} \left[ \varphi(X_i^{j,-d}) + a^*(Z_j(s, X_i^{j,-d})) \right] - X_i^{j,-d} \sum_{j \in [d-1]} \left[ \varphi(X_i^{j,0}) + a^*(-Z_j(s, X_i^{j,0})) \right] + \varepsilon^2 \left( \frac{1}{2} - X_i^{j,-d} \right) \right\} ds$$

(3.20)

The key fact is then to observe that, whenever $X_i^{j,0}$ is close to zero, $\varphi(X_i^{j,0})$ (which shows up in the definition of the drift, compare (3.11) with (3.18)) is greater than $\kappa$, and thus helps for pushing the particle toward the interior of the simplex. This guarantees that, provided that $\kappa \geq \varepsilon^2/2$, the equation is well-posed and that the unique solution stays in $\text{Int}(\mathcal{S}_d)$, see [43, Proposition 2.2] for the details.

**Step 3.** We now provide a (uniform) H"older estimate for the elements of $\mathcal{X}$. Again we borrow the result from [43]. Indeed, (3.15) can be rewritten as a system of $d-1$ equations on $[0,T] \times \mathcal{S}_d$, using the dictionary to pass from intrinsic derivatives to derivatives in the local chart. Thus we can apply\footnote{In fact, this requires a modicum of care, since the function $\varphi$ in [2] is assumed to vanish outside $[0,2\theta]$ and $\theta$ itself is required to be small enough, see (2.16) therein (our $\theta$ being denoted by $\delta$ in [2]). Obviously, our own $\varphi$ in (2.7) does not satisfy this requirement (paying attention that our $\delta$ in (2.7) is different from $\delta$ in [2], since the latter should be understood as our $\theta$). However, we can recover the setting of [2]. Indeed, the key point is that we can always modify our function $\varphi$ in (2.7) so that it fits the assumption of [2, Theorem 2.10]: Going back to [2, (2.28)], it is indeed easy to see that the values of $\varphi$ taken outside $[0,2\theta]$ can be inserted in the function $\bar{b}$ therein. Since $\bar{b}$ does not enter the definition of the threshold $\kappa_0$ in [2, Theorem 2.10], this leaves the conclusion of [2, Theorem 2.10] unchanged.} Theorem 2.10 of [43], which states that there exists $\kappa_1$ as in the statement and that, for any $\kappa \geq \kappa_1/\varepsilon^2$, there exists $\gamma'' \in (0,\gamma]$, possibly depending on $\varepsilon$ and $\kappa$, such that $||Z||_{WF,\gamma''} \leq C'$, for a constant $C'$ depending on $\varepsilon$, $\kappa$, $\theta$, $M$, $d$, $T$ and on the $L^\infty$ norm of the right-hand side of (3.12) (hence on $f$, $g$, $\varphi$, $\varphi'$, and $||Z||_{\infty}$, the latter being uniformly bounded by Step 2). Therefore, by choosing $\gamma'$ in the definition $\mathcal{X}$ as $\gamma''$, we deduce that $\mathcal{X}$ is bounded.

**Step 4.** Uniqueness of classical solutions can be proved by using the so-called four stepscheme; see [53, 54]. Any classical solution $Z$ can be indeed represented in the form of a multi-dimensional forward-backward SDE (which is nothing but a system of stochastic characteristics). In turn, the fact that (3.10) has a classical solution forces the former forward-backward SDE to be uniquely solvable, and hence (3.10) itself to be also uniquely solvable. This argument is in fact explained in detail in [43, Theorem 4.3]. The specific subtlety (which is common to [43] and to our case) is that, due to the fact that the Kimura operator driving (3.10) degenerates near the boundary, some exponential integrability is needed for the reciprocal of the forward component in the forward-backward system of characteristics. In fact, this integrability property is very similar to
the integrability property discussed after Definition 2.3. In short, it holds true provided that $\kappa$ is bigger than (up to a multiplicative constant) $\varepsilon^2$, which is obviously the case in our setting since $\kappa$ scales here (at least) like $\varepsilon^{-2}$. This point is discussed with care in the paper [43], see the proof of Theorem 4.3. □

3.3. Solving for the HJB equation

We now turn to the well posedness of (2.13), or equivalently of (3.6), and prove the following refined version of Theorem 2.1.

**Theorem 3.3.** If $F \in C_{WF}^{1,\gamma}(S_d)$ and $G \in C_{WF}^{1,2+\gamma}(S_d)$, for a given $\gamma \in (0, 1)$, then there exists a constant $\kappa_1 > 0$ only depending on $M$, $T$ and $d$, such that for any $\varepsilon \in (0, 1], \theta > 0$ and $\kappa \geq \kappa_1/\varepsilon^2$, Equation (2.13) admits a unique solution $V \in C_{WF}^{1,2+\gamma}([0, T] \times S_d)$. The solution $V$ is the value function of the viscous MFCP and the optimal feedback function is given by

$$\tilde{z}^{*,i,j}(t, p) = a^*(dV(t, p) - d_jV(t, p)).$$

(3.21)

The latter gives the unique optimal control in the sense that, for any initial state $p_0 \in \operatorname{Int}(S_d)$ and any pair of optimal trajectory $p$ and optimal control $\tilde{z}$ (which is an $\mathbb{F}$-progressively measurable process bounded by $M$), it holds $z_t = \tilde{z}^*(t, p_t)$ for $dt \otimes \mathbb{P}$ a.e. $(t, \omega)$. Moreover, the derivative $D\tilde{V}$ is the unique solution to (3.9) in $C_{WF}^{0,2+\gamma}([0, T] \times S_d)$.

Equivalently, with the same assumptions and in the same space (up to a change of coordinate), Equation (3.6) admits a unique solution $\hat{V}$ and its derivative $D_x\hat{V}$ is the unique solution to (3.10) (denoted by $Z$ in the statement of Theorem 3.2).

**Proof.** As announced before, we prove well posedness of (3.6). The candidate for being the optimal feedback is (see (3.8)) $z_{i,j}^*(t, x) = \tilde{a}_{i,j}^*(Z(t, x), i \neq j$, for $Z$ given by Theorem 3.2. Using the same notation as in (3.2) and (3.8), we thus consider, on $[0, T] \times \hat{S}_d$, the PDE (sums being taken over $[d - 1]$):

$$\begin{cases}
\partial_t Z + \mathcal{H}_M(x, Z) + \hat{F}(x) + \frac{1}{2} \varepsilon^2 \sum_{j,k} (x_j \partial_{j,k} - x_j x_k) \partial^2_{x_j x_k} Z \\
\quad + \sum_k x_k \left[ \sum_i \phi(x_j)(\partial_{x_j} Z - \partial_{x_k} Z) - \phi(x^{-d}) \partial_{x_j} Z \right] + x^{-d} \sum_i \phi(x_j) \partial_{x_j} Z = 0,
\end{cases}$$

$$Z(T, x) = G(x),$$

(3.22)

In particular, since the argument $Z$ of the Hamiltonian is frozen (i.e. it is an input of the equation), we can regard (3.22) as a linear Kimura PDE with $Z$ as unknown (the drift coefficient driving the first order term is nothing but $\hat{b}_j(x, 0)$ in (3.11) and hence points inward the simplex). Since $Z \in C_{WF}^{0,2+\gamma}([0, T] \times \hat{S}_d)$, we know from Theorem 10.0.2 of [44] that (3.22) admits a unique solution $Z \in C_{WF}^{0,2+\gamma}([0, T] \times \hat{S}_d)$.

The key fact is to show that $\zeta = Z$ where $\zeta = D_x Z$. Since the second order operator driving (3.22) is elliptic in the interior of the simplex (and non-degenerate in any ball, see for instance [43, (3.11)]) and the source term is differentiable in space, with time-space H"older continuous derivatives, we know from interior estimates for parabolic PDEs (see Theorem 8.12.1 in [55]) that $\zeta$ is once continuously differentiable in time
and twice in space on $[0, T] \times \text{Int}(\hat{S}_d)$—even though we have no guarantee on the behavior at the boundary. This suffices to differentiate (3.22). We then get the following variant of (3.10) at any point $(t, x)$ of $[0, T] \times \text{Int}(\hat{S}_d)$ (the sums below being taken over $j \in [d - 1]$):

$$
\begin{align*}
\frac{\partial \xi^i}{\partial t} + \sum_j (\hat{b}_2^j(x) + \frac{1}{2} \varepsilon^2 \delta_{i,j} - \varepsilon^2 x_j) \partial_{x_j} \xi^i + \frac{1}{2} \varepsilon^2 \sum_{j,k} (x_j \delta_{j,k} - x_j x_k) \partial^2_{x_j x_k} \xi^i + \sum_j \hat{c}^{i,j}(x) \xi^j
= - \left[ \hat{H}^d_M(Z) - \hat{H}^d_{\text{WF}}(Z) + \hat{f}^d(x) - \hat{f}^d(x) + \sum_j (\hat{b}_1^j(x, Z) + \frac{1}{2} \varepsilon^2 \delta_{i,j} - \varepsilon^2 x_j) \partial_{x_j} Z^j \right],
\end{align*}
$$

where, for $j \in [d - 1]$, $x$ and $z$ (as usual sums below are over $k \in [d - 1]$),

$$
\begin{align*}
\hat{b}_1^j(x, z) &= \sum_k \{ x_k a^*(z_k - z_j) - x_j a^*(z_j - z_k) \} + x^{-d} a^*(-z_j) - x_j a^*(z_j), \\
\hat{b}_2^j(x) &= \sum_k \{ x_k \varphi(x_k) - x_j \varphi(x_k) \} + x^{-d} \varphi(x) - x_j \varphi(x^{-d}).
\end{align*}
$$

Obviously, $\hat{b}_1^j$ and $\hat{b}_2^j$ should be compared with $\hat{b}_1$ in (3.11). In particular, $\hat{b}_1(x, z)$ is nothing but $\hat{b}_1^i(x, z) + \hat{b}_2^i(x)$. This prompts us to make the difference with (3.10), from which we get

$$
\begin{align*}
\frac{\partial (\xi^i - Z^i)}{\partial t} + \sum_j (\hat{b}_2^j(x) + \frac{1}{2} \varepsilon^2 \delta_{i,j} - \varepsilon^2 x_j) \partial_{x_j} (\xi^i - Z^i)
+ \frac{1}{2} \varepsilon^2 \sum_{j,k} (x_j \delta_{j,k} - x_j x_k) \partial^2_{x_j x_k} (\xi^i - Z^i) + \sum_j \hat{c}^{i,j}(x) (\xi^j - Z^j) &= 0,
\end{align*}
$$

$$
(\xi^i - Z^i)(T, x) = 0.
$$

In order to prove that $\xi = Z$, we can use Itô’s formula as we did in the proof of Theorem 3.2. Indeed, the interior smoothness of $\xi^i$, for each $i = 1, \ldots, d - 1$, suffices to apply Itô’s formula to $\left( \sum_{i \in [d - 1]} R_{\ell,i}^j(\xi^i - Z^i)(s, X^\ell_i) \right)_{t \leq s \leq T}$ for any given $t$ where $X^\ell_i$ solves (3.18), but for $\hat{b}_1(x, z)$ therein replaced by $\hat{b}_2$, with some $x \in \text{Int}(\hat{S}_d)$ as initial condition at time $t$, and $((R_{\ell,i}^1)_{i \in [d - 1]})_{t \leq s \leq T}$ solves the SDE $dR_{\ell,i}^j = \sum_{s \in [d - 1]} R_{\ell,i}^j \varepsilon^{i,j}(X^\ell_i) ds$, for $s \in [t, T]$ with $(R_{\ell,i}^1) = \delta_{i,j} i_{i \in [d - 1]}$. Following the standard proof of Feynman-Kac formula, we get that $\xi^i(t, x) = Z^i(t, x)$. Hence, $\xi^i(t, \cdot)$ and $Z^i(t, \cdot)$ coincide on $\text{Int}(\hat{S}_d)$ and then, by continuity, on the entire $\hat{S}_d$. In particular, this implies that $Z \in C_{\text{WF}}^{1, 2+\gamma}(0, T) \times \hat{S}_d$, see the definition of the hybrid spaces in the appendix.

By replacing $Z$ by $D_x Z$ in (3.22), we deduce that $Z$ solves (3.6). By a straightforward adaptation of the verification theorem, we deduce that $Z$ must be the value function of the MFCP and, as by-product, it must be the unique solution of (3.6) in the space $C_{\text{WF}}^{1, 2+\gamma}(0, T) \times \hat{S}_d$. Also, since $\mathcal{H}_M(p, z, w)$ (see (3.2)) is strictly convex with respect to $x$ as long as $p$ is in $\text{Int}(\hat{S}_d)$ and since any controlled trajectory $\tilde{p}$ in (2.12) stays in $\text{Int}(\hat{S}_d)$ (see [43, Proposition 2.2]), we deduce that the optimal control is unique and is in a feedback form. In local coordinates, the optimal feedback function writes

$$
\begin{align*}
\dot{x}^i(t, x) &= a^*(\partial_{x_i} Z(t, x) - \partial_{x_i} Z(t, x)) \quad \text{if } i, j \in [d - 1], \\
\dot{x}^d &= a^*(\partial_{x_j} Z(t, x)), \quad \dot{x}^d = a^*(-\partial_{x_j} Z(t, x)),
\end{align*}
$$

and this is equivalent to (3.21) in intrinsic derivatives. Relabeling $Z$ into $\tilde{V}$, this completes the proof. \(\square\)
4. Potential game with a common noise

The main purpose of this section is to prove Theorem 2.2.

4.1. New MFG

Our first step is to introduce a viscous MFG that derives from the viscous MFCP studied in the previous section. Equivalently, we would like the corresponding MFG system to represent the necessary condition for optimality of the MFCP. As we already explained in Section 2, the problem is that, if we use the same dynamics as in (2.5) (which are the basis of the results of [43], on which our paper is built), we can no longer use the cost functional \( J^{\infty, \phi} \) (see (2.6)) to get a potential structure. To wit, the master equation associated with (2.6) (which may be computed along the same lines as in [43], see (3.7) therein) does not identify with the derivative system (3.10). In particular, the master equation with a common noise

\[ dv_t = -\left( \frac{1}{d} \sum_{i} H_{M}^i (v_t) + f^i (p_t^*) \right) dt + \frac{1}{\sqrt{d}} \sum_{i,k} \sqrt{p_t^{*,i} (p_t^{*,i})^{-1}} \left( w_t^{i,k} + w_t^{i,k,i} \right) dB_t^{i,k}, \]

for some initial condition \( p_0 \in \text{Int}(S_d) \); see [43, Proposition 2.2] for the unique solvability, the unique solution remaining inside the interior of \( S_d \). By Itô’s formula (the fact that we can apply Itô’s formula with intrinsic derivatives can be justified by using the local chart, at least in the interior of the simplex), we get (sums below being over indices in \([d]\])

\[ \frac{dp_t^{*,i}}{dt} = \left( \sum_{j} p_t^{*,j} (\varphi (p_t^{*,j}) + \alpha (v_t^{j} - v_t^{i})) - \sum_{k} \nu_t^{*,i} (\varphi (p_t^{*,k}) + \alpha (v_t^{k} - v_t^{i})) \right) dt + \frac{1}{\sqrt{d}} \sum_{k} \sqrt{p_t^{*,i} (p_t^{*,i})^{-1}} (dB_t^{i,k} - dB_t^{k,i}), \]

where

\[ dp_t^* = (\sum_{i} p_t^{*,i}) \frac{dS_t}{dS_0}, \]

\[ dv_t = -\left( \frac{1}{d} \sum_{i} H_{M}^i (v_t) + f^i (p_t^*) \right) dt + \frac{1}{\sqrt{d}} \sum_{i,k} \sqrt{p_t^{*,i} (p_t^{*,i})^{-1}} \left( w_t^{i,k} + w_t^{i,k,i} \right) dB_t^{i,k}, \]

for some initial condition \( p_0 \in \text{Int}(S_d) \); see [43, Proposition 2.2] for the unique solvability, the unique solution remaining inside the interior of \( S_d \). By Itô’s formula (the fact that we can apply Itô’s formula with intrinsic derivatives can be justified by using the local chart, at least in the interior of the simplex), we get (sums below being over indices in \([d]\])

\[ \frac{dp_t^*}{dt} = \left( \sum_{i} p_t^{*,i} (\varphi (p_t^{*,i}) + \alpha (v_t^{i} - v_t^*)) \right) dt + \frac{1}{\sqrt{d}} \sum_{k} \sqrt{p_t^{*,i} (p_t^{*,i})^{-1}} (dB_t^{i,k} - dB_t^{k,i}), \]
\[ w^{i,j,k}_t = W^{i,j,k}(t, p^*_t), \quad \text{with} \quad W^{i,j,k}(t, p) = \frac{1}{\sqrt{2}} \varepsilon \sqrt{p_i p_k (d_j V^i - d_k V^i)}(t, p). \] (4.3)

Notice in particular that
\[
\frac{1}{\sqrt{2}} \varepsilon \sum_j \sqrt{p_i^{\ast,j} (p_i^{\ast,j})^{-1}} w^{i,j,k}_t = \frac{1}{2} \varepsilon^2 \sum_j p_i^{\ast,j} (d_j V^i - d_i V^j)(t, p^*_t), \]
\[
\frac{1}{\sqrt{2}} \varepsilon \sum_j \sqrt{p_i^{\ast,j} (p_i^{\ast,j})^{-1}} w^{i,j,k}_t = \frac{1}{2} \varepsilon^2 \sum_j p_i^{\ast,j} (d_j V^j - d_i V^j)(t, p^*_t). \] (4.4)

We will use several times Schwarz' identity for intrinsic derivatives, which we already invoked in (2.32). Together with the fact that \( V \) is twice differentiable on \( \text{Int}(S_d) \), it says that, for any \( i, j \in [d], d_i V^j = d_j V^i \). Hence (4.4) above implies that the term in the penultimate line of (4.2) is equal to
\[
\frac{1}{2} \varepsilon^2 (d_i V^i(t, p^*_t)) - 2 \sum_j p_i^{\ast,j} d_j V^i(t, p^*_t) - \frac{1}{d} \sum_j d_j V^j(t, p^*_t)),
\]
which is in fact the term in the penultimate line of the PDE in (3.9).

Ideally, we would like to see (4.2) as the stochastic HJB equation associated with our new MFG with common noise (see [43, Lemma 4.1] for the derivation of this HJB equation). However, we cannot do so directly because the pair \((p^*, \nu)\) in (4.2) takes values in the tangent bundle to the simplex, namely \( \sum_{i \in [d]} v_i = 0 \) for any \( t \in [0, T] \). Obviously, the latter is not consistent with our original MFG, whether there is a common noise or not. Indeed, if this were consistent, then, discarding for a while the common noise, we would have to think of \( v_0^i \) as the minimum of \( J(\cdot; p^*) \) in (2.1) whenever \( q \) therein starts from the Dirac mass at point \( i \), but, then, there would be no reason why the sum of all these costs over \( i \in [d] \) should be null. In fact, we here recover the point raised in (2.22): Therein, we can identify the two vectors \((U^i_{e, \nu}(t, p))_{i \in [d]} \) and \((d_j V^j_{e, \nu}(t, p))_{i \in [d]} \) up to a constant only. The idea below is thus to reconstruct from scratch the sum of the coordinates of the conditional value function. To do so, we notice from [44, Theorem 10.0.2] again that we can solve the PDE in the simplex (sums below are over indices in \([d]\))

\[
\begin{align*}
\partial_t \mathcal{V} + \sum_{j, k} p_j (p_k + a^*(V^j - V^k))(d_j \mathcal{V} - d_k \mathcal{V}) + \frac{\varepsilon^2}{2} \sum_{j, k} (p_j \delta_{j, k} - p_j p_k) d_{j, k} \mathcal{V} \\
+ \frac{1}{2} \sum_{j, k} p_j a^*(V^j - V^k))^2 + \sum_{j, k} p_j p_k \phi'(p_j)(V^j - V^k) + \langle p, f^*(p) \rangle = 0,
\end{align*}
\] (4.5)

where we recall that \( \langle p, f^*(p) \rangle \) (and similarly with \( f \) replaced by \( g \)) here denotes the inner product \( \sum p_j f^j(p) \). The rationale for such a definition is that \( \mathcal{V}(t, p) \) is eventually equal to \( \langle p, U_{e, \nu}(t, p) \rangle \) for \( U_{e, \nu} \) the solution of the master equation of the viscous potential MFG that we introduce next; obviously, this is the same \( U_{e, \phi} \) as in (2.22). Intuitively, if the potentials were given by (2.8) and the HJB equation (2.13) were true on a \( d \)-dimensional neighborhood of the simplex, then \( \mathcal{V}(t, p) \) would be equal to the inner product \( \langle p, D\mathcal{V}(t, p) \rangle \), for \( D\mathcal{V} \) the Euclidean gradient of \( \mathcal{V} \).

We are now in the position to elucidate the shape of \( \vartheta^i_{e, \nu} \) in (2.14), by letting (we remove the superscripts \( \varepsilon \) and \( \phi \) for simplicity)
\[
\vartheta^j(t,p) := \sum_i \left[ p_j \varphi'(p_i)(V^i - V^j)(t,p) + \frac{1}{\sqrt{2}} \varepsilon \sqrt{p_j p_i^{-1} \left( \tilde{W}^{i,j,i} - \tilde{W}^{i,j,j} - 2 \gamma^i(t)(t,p) \right)} \right],
\]

where
\[
\tilde{W}^{i,j,k}(t,p) = W^{i,j,k}(t,p) - \langle p, W^{\bullet,j,k}(t,p) \rangle
\]
\[
\gamma^i(t,p) = \frac{1}{\sqrt{2}} \varepsilon \sqrt{p_j p_i^{-1} (d_i \gamma(t,p) - d_j \gamma(t,p) - (V^i - V^j))(t,p)).}
\]

Observe in particular that, despite the factor \(\sqrt{p_i^{-1}}\) in (4.6), the function \(\vartheta\) is bounded and continuous on the entire \([0,T] \times S_u\). Using (4.4), we indeed have
\[
\vartheta^j(t,p) = \sum_i p_j \varphi'(p_i)(V^i - V^j)(t,p)
+ \frac{1}{2} \varepsilon^2 \sum_i p_i (d_i V^i - d_i V^j)(t,p) + \varepsilon^2 \sum_i p_i (d_i \gamma - d_j \gamma - (V^i - V^j))(t,p)
+ \varepsilon^2 \sum_i p_i p_k (d_i V^k - d_j V^k)(t,p).
\]

Now, we recall (2.14) together with Definition 2.3: For an adapted continuous process \(p\) with values in \(S_u\), such that \(\int_0^T \frac{1}{p_i} dt\) has exponential moments of sufficiently high order (which we recall holds true if \(p\) solves an equation of the same type as (2.12) and \(\kappa\) is large enough independently of \(p\)) for a progressively-measurable process \(\alpha = (\alpha^{i,j}_{t})_{i,j \in [a], i \neq j} \) such that \(0 \leq \alpha^{i,j}_t \leq M\) and for \(q\) solving (2.5), we let
\[
\tilde{J}^{i,q}(\alpha;p) := \mathbb{E} \left[ \int_0^T \sum_i \frac{1}{p_i} q_i (\alpha^{i}(t) + f_i(t;p_t) + \vartheta^i(t,p_t)) dt + \sum_i q_i g_i(p_T) \right].
\]

By following [43, Subsection 4.1.1], the Stochastic HJB (SHJB) equation associated with this minimization problem here writes down (sums being taken over \([d]\))
\[
du_t^i = - (\hat{H}_M(u_t) + \sum_j \varphi'(p_t)(u_t^j - u_t^i) + f^i(p_t) + \vartheta^i(t,p_t)) dt
- \frac{1}{\sqrt{2}} \varepsilon \sum_{j \neq i} \sqrt{p_j^{i}} (p_i^{-1})(\nu^{i,j,j}_t - \nu^{i,j,i}_t) dt + \sum_{j \neq k} \nu^{i,j,k}_t dB_t^{i,k},
\]
\[
u_t^i = g^i(p_T).
\]

Equation (4.10) is a backward SDE. The additional process \(v\) is part of the solution and ensures that the process \(u\) is adapted to the noise. Hence, our new MFG (in the sense of Definition 2.3) is characterized by the forward-backward system made of the SHJB equation (4.10) and of the Stochastic FP (SFP) equation (2.12); see again [43, Subsection 4.1.1] for the proof.

Of course, the core of our construction is to show that the optimal trajectory \(p^*\) of the MFCP is the unique possible equilibrium of this new MFG. In this regard, our choice for \(\vartheta\) is especially designed so that \((v_t^i - \langle p_t^*, v_t^* \rangle + \gamma(t,p_t^*))_{0 \leq t \leq T}\) solves (4.10) whenever \(p\) is taken as \(p^*\). In such a case, by setting the martingale terms equal to each other in the expansions of \((u_t^i)_{0 \leq t \leq T}\) and \((v_t^i - \langle p_t^*, v_t^* \rangle + \gamma(t,p_t^*))_{0 \leq t \leq T}\,\), we get from (4.2) and (4.3)
\[
\nu_t^{i,j,k} = w_t^{i,j,k} - \langle p_t^*, w_t^{i,j,k} \rangle + \frac{1}{\sqrt{2}} \varepsilon \sqrt{p_t^{j,k}(d_j \gamma(t,p_t^*) - d_k \gamma(t,p_t^*) - (V_t^i - V_t^k))}
= \tilde{w}_t^{i,j,k} + \gamma^{i,k}(t,p_t^*),
\]
\[
\nu_t^i = g^i(p_T).
\]
where $\tilde{w}_t^{i,j,k} = w_t^{i,j,k} - \langle p_t, w_t^{i,j,k} \rangle$, which explains why $Y$ appears in (4.6). The details are given in the proof of Theorem 4.1 below.

### 4.2. Solvability

This is the refined version of Theorem 2.2. Recall that $\phi$ satisfies (2.7).

**Theorem 4.1.** If $F \in C^{1,\gamma}_{WF}(\mathcal{S}_d)$ and $G \in C^{2+\gamma}_{WF}(\mathcal{S}_d)$ for a given $\gamma \in (0, 1)$, then there exists a constant $\kappa_2 \geq \kappa_1$ ($\kappa_1$ and $\gamma^\prime$ being given by Theorem 3.2) only depending on $M$, $T$ and $\delta$, such that for any $\varepsilon \in (0, 1]$, $\theta > 0$ and $\kappa \geq \kappa_2/\varepsilon^2$, there exists $\gamma'' \in (0, \gamma']$, possibly depending on $\varepsilon$ and $\kappa$, such that the new MFG, associated with the dynamics (2.5) and with the cost (4.9), admits a unique solution $(p, \mathbf{x})$ for any $p_0 \in \text{Int}(\mathcal{S}_d)$. It is equal to the unique optimizer of the MFCP (2.11)-(2.12). Moreover, the master equation (2.20) associated with the modified MFG admits a unique solution $U \in [C^{0,2+\gamma''}_{WF}([0, T] \times \mathcal{S}_d)]^d$ and (2.22) holds.

**Proof.** We first prove existence of a MFG solution, by using the solution of the MFCP, and then show uniqueness by invoking the results from [43].

Existence. As announced in the previous subsection, we choose $p = p^*$ with $p^*$ as in (4.1) (here, we drop the superscript $*$ to alleviate the notation) and then let, as a candidate for solving the SHJB equation (4.10):

$$u_t^i := v_t^i - \langle p_t, v_t^i \rangle + Y(t, p_t), \quad t \in [0, T], \quad i \in [d].$$

(4.12)

Importantly, we notice that $u_t^i - u_t^j = v_t^i - v_t^j$ for any $i, j \in [d]$ with $i \neq j$. With the same notation as in (4.11), we then get (some explanations are given after the formula; moreover, the sums below are over $[d]$)

\[
\begin{align*}
    du_t^i &= -\langle \tilde{H}_M^i(u_t) - \langle p_t, \tilde{H}_M^i(u_t) \rangle, f^i(p_t) - \langle p_t, f^*(p_t) \rangle \rangle dt \\
    &\quad - \sum_{j,k} \phi^i_j(u_t^j - u_t^i) - \sum_{j,k} p_t^i \phi^*(p_t^i)(u_t^j - u_t^i) dt \\
    &\quad + \sum_{j,k} p_t^i \phi^*(p_t^i)(u_t^j - u_t^i) - \sum_{j,k} p_t^i \phi^*(p_t^i)(u_t^j - u_t^i) dt \\
    &\quad - \frac{1}{\sqrt{2}} \varepsilon \left( \sum_{j} \sqrt{p_t^j(p_t^j)^{-1}(w_t^j, w_t^j, w_t^j, w_t^j)} - \sum_{j,k} \sqrt{p_t^j p_t^k (w_t^{i,j,k} + w_t^{i,j,k})} \right) dt \\
    &\quad - \left( \sum_{j,k} p_t^j \phi^*(p_t^j) + a^*(u_t^j - u_t^i) \right) (u_t^j - u_t^i) dt - \frac{1}{\sqrt{2}} \varepsilon \sum_{j,k} p_t^j p_t^k (w_t^{i,j,k} + w_t^{i,j,k}) dt \\
    &\quad - \left( \frac{1}{2} \sum_{j,k} p_t^j \phi^*(p_t^j) (u_t^j - u_t^i)^2 + \sum_{j,k} p_t^j p_t^k \phi^*(p_t^j)(u_t^j - u_t^i) + \langle p_t, f^*(p_t) \rangle \right) dt + \sum_{j,k} v_t^{i,j,k} dB_t^{i,j,k}.
\end{align*}
\]

In short, the term on the first line comes from the expansion of $dv_t^i - d\langle p_t, v_t^i \rangle$, see the first line in (4.2). Similarly, the terms on the second and third lines come from the second line in (4.2). And the fourth line derives from the third line in (4.2). The first term on the fifth line comes from $\langle v_t^i, dp_t \rangle$ and the second term on the same line is the bracket in the expansion of the inner product $d\langle v_t^*, p_t \rangle$. The first term on the last line comes from the expansion of $\langle Y(t, p_t) \rangle_{0 \leq t \leq T}$ by means of Itô’s formula. The last term is given by (4.11).

We first treat terms that cancel in the above expansion. Obviously, the inner products $\langle p_t, f^*(p_t) \rangle$ on the top and bottom lines cancel. Similarly, the second term on the second
line cancel out with half of the first term on the penultimate line, and the second term on the third line cancel out with the second term on the last line. As for the inner product \( \langle p_t, H_M(u_t) \rangle \) on the first line, it cancels with the second half of the first term on the fifth line and with the first on term on the last line. Now, using the fact that \( w_{i,j,k}^t = -w_{i,k,j}^t \), we have

\[
\frac{1}{\sqrt{2}} \varepsilon \sum_{j,k} \sqrt{p_t^i p_t^j} (w_{i,j,k}^t + w_{i,k,j}^t) - \frac{1}{\sqrt{2}} \varepsilon \sum_{j,k} \sqrt{p_t^i p_t^j} (w_{i,j,k}^t - w_{i,k,j}^t) = 0,
\]

so that the last terms on the fourth and fifth lines also cancel out. Moreover, adding \( \partial^t (t, p_t) \) (using (4.6)) to the first term on the third line and the first term on the fourth line, we get

\[
\partial^t (t, p_t) + \sum_i \partial^t \varepsilon(p_t) (u_t^i - u_t^i) - \frac{1}{\sqrt{2}} \varepsilon \sum_i \sqrt{p_t^i (p_t^i)^{-1}} (w_{i,i,j}^t + w_{j,i}^t) - \frac{1}{\sqrt{2}} \varepsilon \sum_i \sqrt{p_t^i (p_t^i)^{-1}} (w_{i,i,j}^t - w_{i,j,i}^t) = - \langle p_t, w_t^{*,i,j} \rangle - 2 \langle p_t, w_t^*, t, p_t \rangle
\]

where, in the second and third lines, we used the two equalities \( \langle p_t, w_t^{*,i,j} \rangle = - \langle p_t, w_t^{*,j,i} \rangle \) and \( \gamma^{i,j}(t, p_t) = - \gamma^{j,i}(t, p_t) \).

It remains to see from (4.11) that

\[
\partial^t (t, p_t) + \sum_i \partial^t \varepsilon(p_t) (u_t^i - u_t^i) - \frac{1}{\sqrt{2}} \varepsilon \sum_i \sqrt{p_t^i (p_t^i)^{-1}} (w_{i,i,j}^t + w_{j,i}^t) - \frac{1}{\sqrt{2}} \varepsilon \sum_i \sqrt{p_t^i (p_t^i)^{-1}} (w_{i,i,j}^t - w_{i,j,i}^t) = - \langle p_t, w_t^{*,i,j} \rangle - 2 \langle p_t, w_t^*, t, p_t \rangle + \gamma^{i,j}(t, p_t) - \gamma^{j,i}(t, p_t)
\]

We hence get that the pair \( (p_t, u_t)_{0 \leq t \leq T} \) solves (4.1) (with \( v_t \) replaced by \( u_t \) therein) and (4.10).

**Uniqueness.** For \( \kappa_2 \) as in the statement, uniqueness follows from \(^3\) [43, Theorem 3.2], using the fact that \( \partial \) in (4.8) is Hölder continuous, which in turn follows from the fact that the solution to the linear equation (4.5) belongs to \( C_0^{T,2+\gamma}([0, T] \times S_d) \), by [44, Theorem 10.0.2]. The new exponent \( \gamma'' \), as well as existence and uniqueness of a classical solution to the master equation (2.20), then follow from [43, Theorem 3.4]. Finally, (2.22) follows from (4.12). □

\(^3\) As we already explained in footnote 2, some care is needed to apply the results of [2], since the framework therein is not exactly the same. In footnote 2, we already commented on the shape of the function \( \phi \). This observation is still relevant here. Also, it must be stressed that, in [2], the constant \( \kappa_2 \) (with the same notation as therein, \( \kappa_2 \) denoting the threshold for \( \kappa \) in [2] and hence being the analogue of \( \kappa_2/c \) here) is allowed to depend on \( ||f||_\infty \) and \( ||g||_\infty \). In our analysis, \( \kappa_2 \) is allowed to depend on \( M \), which in turn depends on \( ||f||_\infty \) and \( ||g||_\infty \). So, the latter is consistent with the results of [2]. In fact, our framework is easier since the controls are already required to be bounded by \( M \), which is not the case in [2]. This explains why \( M \) directly shows up in our statement; in short, it provides an upper bound for the drift in the dynamics of \( p \).
5. Selection by vanishing viscosity

5.1. Selection of equilibria

The purpose of this subsection is to prove Theorem 2.4. We recall that $\phi_{\theta, \delta, \epsilon}$ is defined by (2.17)--(2.18), whereas $\kappa_i := \epsilon^{-2} \kappa_2$, for $\kappa_2$ as in the statement of Theorem 2.2. In particular, $\kappa_2$ is fixed once for all and only depends on $||f||_\infty, ||g||_\infty, T$ and $d$, and is thus independent of the four remaining parameters $\theta, \delta, \epsilon$ and $\kappa_0$ in (2.17). As for $\kappa_0$, it is a constant whose value is fixed later on; say that, in the end, it must be above some threshold $\kappa_0$, only depending on $d$ and $\kappa_2$, see for instance Theorem 5.12. In order for $\phi_{\theta, \delta, \epsilon}$ to be non-increasing, we will impose a smallness condition on $\epsilon$, namely $\epsilon^2 \leq \epsilon_0^2 = \min(\kappa_2/\kappa_0, 1)$. Again, we stress the fact that, in this condition, the constant $\kappa_0$ will be chosen later on, while $\kappa_2$ is fixed.

Accordingly, for any initial condition $(t_0, p_0) \in [0, T] \times \text{Int}(S_d)$, we write $(p^0_{[t_0, p_0]}, x^0_{[t_0, p_0], t}) := (p^0_{[t_0, p_0]}, x^0_{[t_0, p_0], t})_{t_0 \leq t \leq T}$ for the minimizer of $\mathcal{J}^\epsilon, \phi$ defined by (2.11) with $\phi$ being given by (2.17)--(2.18). When there is no ambiguity on the choice of the initial condition, we merely write $(p, x)$. Importantly, in this notation, the $\mathbb{F}$-progressively measurable process $x^0_{[t_0, p_0], t}$ is given by (2.15) through a feedback function, whose off-diagonal entries are bounded by $M$. We recall that $\mathcal{E}$ (defined in before Theorem 2.4) is endowed with the weak topology, which makes it a compact metric (and hence Polish) space and for which the convergence is denoted by $\rightarrow_{\mathbb{P}}$. We also denote by $\mathcal{V}_{\theta, \delta, \epsilon}$ the value function of the viscous MFCP and by $\mathcal{V}$ the value function of the inviscid MFCP. The following result then subsumes Theorem 2.4:

**Theorem 5.1.** Assume that $F \in C^{1,\gamma}_{WF}(S_d) \quad \text{and} \quad G \in C^{1,2+\gamma}_{WF}(S_d)$ for a given $\gamma \in (0, 1)$. Moreover, fix the value of $\kappa_0$ in (2.17) and let $p_0 \in \text{Int}(S_d)$ stand for the initial condition of (2.12) at time 0. Then there exists $\delta_0$ depending only on $p_0, M, T$ and $d$, such that the family $(p^0_{t_0, \delta, \epsilon}, x^0_{t_0, \delta, \epsilon})_{0 < \theta < \delta \leq \delta_0, 0 \leq \epsilon \leq \epsilon_0}$ is tight in $C([0, T]; S_d) \times \mathcal{E}$. The limit in law of any converging subsequence $(p^0_{\theta_n, \delta_n, \epsilon_n}, x^0_{\theta_n, \delta_n, \epsilon_n})$, with $\lim_{n \to \infty} (\theta_n, \delta_n, \epsilon_n) = (0, 0, 0)$, is a probability $\mathbb{P}$ that satisfies the conclusion of Theorem 2.4. Moreover, for any such converging subsequence,

$$\lim_{n \to \infty} J^{\theta_n, \delta_n, \epsilon_n, \min} (x^0_{\theta_n, \delta_n, \epsilon_n}) = \min_{\beta \in \mathcal{E}} J(\beta).$$

(5.1)

In particular, for any $t_0 \in [0, T]$ and any $p_0 \in \text{Int}(S_d), \quad \lim_{(\theta_n, \delta_n, \epsilon_n) \rightarrow (0, 0, 0)} \mathcal{V}_{\theta_n, \delta_n, \epsilon_n}(t_0, p_0) = \mathcal{V}(t_0, p_0).$ (5.2)

**Proof.** Throughout the proof, we use the following notation. For $\mathbf{x} = (x_i)_{0 \leq i \leq T}$ a bounded deterministic path in $\mathcal{E}$, we call $\mathbf{p} (\mathbf{x})$ the solution of the equation (obviously, the solution exists and is unique as the equation is linear)

$$p^i_t = p^i_0 + \int_0^t \sum_{j \in [d]} (p^j_s x^i_j - p^i_s x^j_i) \, ds, \quad t \in [0, T], \quad i \in [d].$$

(5.3)
Step 1. The distributions of the random variables \((x^{0,\delta,\epsilon})_{0,\delta,\epsilon}\) (regarded as taking values within \(\mathcal{E}\)) is tight as \(\mathcal{E}\) is compact. For any \(x\) in \(\mathcal{E}\), the corresponding solution \(p^o := p^o(x)\) to (5.3) is such that \(p_t^{i,i} \geq p_0^i - M(d-1) \int_0^t p_i^{s,i}ds\) and thus, by Gronwall’s lemma,

\[
p_t^{s,i} \geq p_0^i e^{-TM(d-1)}.
\]  

(5.4)

This prompts us to define

\[\delta_0 := \frac{1}{4} e^{-TM(d-1)} \min_{i \in [d]} p_0^i.\]

and let

\[\tau_{e,0,\delta} := \inf \{0 \leq t \leq T : \min_{i \in [d]} p_t^{i,0,\delta,\epsilon} \leq 3\delta_0\} \wedge T,\]

with the convention \(\inf \emptyset = +\infty\), and \(\hat{p}^{0,\delta,\epsilon} := (p^{0,\delta,\epsilon}_{t_0,0,\epsilon} + t)_{0 \leq t \leq T}\).

Notice in particular that, for \(\delta < \delta_0\), \(\hat{p}^{0,\delta,\epsilon}\) does not see the function \(\varphi\) in its drift since the support of latter is restricted to \([0, 2\delta]\). By Kolomogorov’s criterion, since \(x\) is bounded by \(M\) and \(p\) is bounded by 1, it is then standard to show the tightness of the distributions of the processes \((\hat{p}^{0,\delta,\epsilon}_{t_0,0,\epsilon} + t)_{0 \leq 2\delta \leq \delta_0, 0 < t < \epsilon_0}\) in \(C([0, T]; \mathcal{S}_d)\).

Step 2. We hence consider a weakly convergent subsequence \((p^{0,\delta,\epsilon}_{t_0,0,\epsilon} + x^{0,\delta,\epsilon}_{0,\delta,\epsilon})_{n \geq 0}\), with some \((p, x)\) as weak limit, where \(\lim_{n \to \infty} (\theta^{0,\delta,\epsilon}_{n,\delta,\epsilon} + \frac{1}{n}) = 0\). To simplify the notations, we let \((\hat{p}^{n}, x^{(n)}) := (\hat{p}^{0,\delta,\epsilon}_{i_0,0,\epsilon} + x^{0,\delta,\epsilon}_{0,\delta,\epsilon})\). Applying Skorohod’s representation Theorem, we can assume without any loss generality that the convergence holds almost surely, provided that we allow the Brownian motions \((B^{i,j})_{i,j \in [d], i \neq j}\) to depend on \(n\). We hence write the latter in the form \(B^{(n)} = (B^{(n),i,j})_{i,j \in [d], i \neq j}\). So, we can assume that there exists a full event \(\Omega_0\) on which \(\sup_{0 \leq t \leq T} |\hat{p}^{(n)}_t - \hat{p}_t| \to 0\) and \(x^{(n)} \to x\).

We then write (pay attention that, although we don’t mention it explicitly, the last four terms in the right-hand side below depend on \(i\))

\[\hat{p}^{(n),i} := p_0^i + R^{(n),1}_t + R^{(n),2}_t + R^{(n),3}_t + R^{(n),4}_t, \quad t \in [0, T],\]

where (sums being over indices in \([d]\))

\[
R^{(n),1}_t = \int_0^t \sum_{i,j} (p^{n,i}_s x^{(n),i,j}_s - p^{n,i}_s x^{(n),i,j}_s) ds - \int_0^t \sum_{i,j} (p^{i,j}_s x^{n,i,j}_s - p^{i,j}_s x^{n,i,j}_s) ds,
\]

\[
R^{(n),2}_t = \int_0^t \sum_{i,j} (p^{i,j}_s x^{n,i,j}_s - p^{i,j}_s x^{n,i,j}_s) ds,
\]

\[
R^{(n),3}_t = \int_0^t \left[ (1 - \hat{p}^{(n),i}_s) \phi_{0,\delta,\epsilon}(\hat{p}^{(n),i}_s) - \hat{p}^{(n),i}_s \sum_{j \neq i} \phi_{0,\delta,\epsilon}(\hat{p}^{(n),j}_s) \right] ds,
\]

\[
R^{(n),4}_t = \frac{1}{\sqrt{2}} \frac{1}{\epsilon_n} \int_0^t \sum_{j \neq i} \sqrt{\hat{p}^{(n),i}_s \hat{p}^{(n),j}_s} d(B^{(n),i,j}_s - B^{(n),j,i}_s).
\]

We then work on \(\Omega_0\) in order to handle the almost sure convergence of the first three terms. By uniform convergence of \(\hat{p}^{(n)}\) to \(p\) and by uniform boundedness of \(x^{(n)}\), \((R^{(n),1}_t)_{0 \leq t \leq T}\) tends to 0, uniformly in \(t \in [0, T]\). By weak convergence of \(x^{(n)}\) to \(x\), we deduce that, for any \(t \in [0, T]\), \(R^{(n),2}_t\) tends to \(\int_0^t \sum_{i,j} (p^{i,j}_s x^{i,j}_s - p^{i,j}_s x^{i,j}_s) ds\); by Arzelà-Ascoli Theorem, the convergence is uniform in \(t \in [0, T]\). Since it is implicitly required that \(\delta_n < \delta_0\) and hence \(\phi_{0,\delta,\epsilon}(\hat{p}^{(n),j}_t) = 0\) for all \(j \in [d]\) and \(t \in [0, T]\) (recall that \(\hat{p}^{(n),j}_t\)
is stopped before entering the support of \( \phi \), the term \((R_t^{(n)}, 3)_{0 \leq t \leq T}\) is constantly 0. We hence derive the almost sure limit of the first three terms (the initial condition being excluded) in the expansion of \( \bar{p}^{(n)} \). As for \((R_t^{(n)}, 4)_{0 \leq t \leq T}\), we observe by Doob’s inequality that, since the second moment of the stochastic integral is uniformly bounded with respect to \( n \), \( \sup_{t \in [0, T]} R_t^{(n), 4} \) tends to 0 in probability.

Thus we can conclude that, with probability 1, the limit process \((\bar{p}, \bar{\alpha})\) solves equation (5.3).

Step 3. We keep the same notation as in the second step (working in particular with the same Skorokhod representation sequence). Since

\[
\mathbb{P}(\inf_{0 \leq t \leq T} \min_{i \in [d]} \hat{P}^{(n)}_{t} \leq 3\delta_0) \geq \mathbb{P}(\tau_{\theta_n, \delta_n, \epsilon_n}, T) \geq 3\delta_0)
\]

and the limit process satisfies

\[
\mathbb{P}(\inf_{0 \leq t \leq T} \min_{i \in [d]} \hat{P}^{(n)}_{t} \leq 3\delta_0) = 0,
\]

Portmanteau Theorem gives

\[
\lim_{n \to \infty} \mathbb{P}(\tau_{\theta_n, \delta_n, \epsilon_n} < T) = 0. \tag{5.5}
\]

Now,

\[
\mathbb{E}\left[ \sup_{t \in [0, T]} |\hat{p}^{(n)}_{t} - \bar{p}_{t}| \right] \leq \mathbb{E}\left[ \sup_{t \in [0, T]} \left| \hat{p}^{(n)}_{t} - \bar{p}_{t} \right| 1_{\{\tau_{\theta_n, \delta_n, \epsilon_n}, t < T\}} \right] + \mathbb{E}\left[ \sup_{t \in [0, T]} |\hat{p}^{(n)}_{t} - \bar{p}_{t}| \right]
\]

\[
\leq 2\mathbb{P}(\tau_{\theta_n, \delta_n, \epsilon_n}, T) + \mathbb{E}\left[ \sup_{t \in [0, T]} |\hat{p}^{(n)}_{t} - \bar{p}_{t}| \right].
\]

The first term in the right-hand side tends to 0 by (5.5) and the second term by the convergence result proved in the second step (the almost sure convergence also holding true in \( L^1 \) since the underlying processes take values in the simplex). Therefore we obtain \( \lim_{n \to \infty} \mathbb{E}[\sup_{t \in [0, T]} |\hat{p}^{(n)}_{t} - \bar{p}_{t}|] = 0 \). This in particular implies that the collection of the laws of the random variables \((\hat{p}^{(n)}, \hat{\alpha}^{(n)}, \alpha^{(n)}, \hat{\alpha}^{(n)}, \tau_{\theta_n, \delta_n, \epsilon_n})_{0 < 2\theta < \epsilon_0, 0 < \epsilon_0, \alpha^{(n)}, \hat{\alpha}^{(n)}, \tau_{\theta_n, \delta_n, \epsilon_n}}\) is tight (on the same space as before).

Step 4. We pass to the limit in the cost. To this end, we use the convenient notations

\[
\bar{p}^{(n)} := \hat{p}^{(n)}, \delta_n, \epsilon_n \text{ and } \bar{J}^{(n)}(\cdot) := J^{\infty}(\theta_n, \delta_n, \epsilon_n)(\cdot), \text{ see (2.11).}
\]

By convexity (splitting \( \hat{\alpha}^{(n)} \) into \( \hat{\alpha}^{(n)} \cdot \alpha^{(n)} + (\hat{\alpha}^{(n)} - \alpha^{(n)})), \) we have

\[
\bar{J}^{(n)}(\alpha^{(n)}) - \mathbb{E}[\bar{J}(\alpha)]
\]

\[
= \mathbb{E}\left[ \int_0^T \left( \frac{1}{2} \sum_{i,j} \left( \hat{p}^{(n)}_{t[i]} \hat{\alpha}^{(n)}_{t[i,j]} \right)^2 + F(\hat{p}^{(n)}_{t[i]}) - \frac{1}{2} \sum_{i,j} \hat{p}^{(n)}_{t[i]} \hat{\alpha}^{(n)}_{t[i,j]} \hat{\alpha}^{(n)}_{t[i,j]} - G(\hat{p}^{(n)}_{t[i]}) + G(\hat{p}^{(n)}_{T}) - G(\hat{p}^{(n)}_{T}) \right) dt \right]
\]

\[
= \mathbb{E}\left[ \int_0^T \left( \frac{1}{2} \sum_{i,j} (\hat{p}^{(n)}_{t[i]} - \bar{p}^{(n)}_{t[i]}) \sum_{j \neq t[i]} \left( \hat{\alpha}^{(n)}_{t[i,j]} \bar{p}^{(n)}_{t[i,j]} \right) dt \right] + \mathbb{E}\left[ \int_0^T \sum_{i,j} \hat{p}^{(n)}_{t[i]} \sum_{j \neq t[i]} \hat{\alpha}^{(n)}_{t[i,j]} (\hat{\alpha}^{(n)}_{t[i,j]} - \bar{\alpha}^{(n)}_{t[i,j]} \right)
\]

\[
+ \mathbb{E}\left[ \int_0^T (F(\hat{p}^{(n)}_{t[i]}) - F(\bar{p}^{(n)}_{t[i]})) dt + G(\hat{p}^{(n)}_{T}) - G(\bar{p}^{(n)}_{T}) \right].
\]

(5.6)

Since \( \lim_{n \to \infty} \mathbb{E}[\sup_{t \in [0, T]} |\hat{p}^{(n)}_{t[i]} - \bar{p}^{(n)}_{t[i]}|] = 0 \), the first and third term in the lower bound go to 0 (using the boundedness of \( \alpha^{(n)} \) and the regularity of \( F \) and \( G \)). As for the second
term, it can be proved to tend to 0 by combining the convergence \( \mathbf{z}^{(n)} \rightarrow \mathbf{z} \) with Lebesgue dominated convergence theorem. Thus we obtain \( \mathbb{E}[J(\mathbf{z})] \leq \liminf_{n \rightarrow \infty} J^{(n)}(\mathbf{z}^{(n)}) \).

In order to complete the proof, consider any deterministic control \( \beta \in \mathcal{E} \) (in particular, the off-diagonal components are bounded by \( M \)). Then, denote by \( \mathbf{p}^{(n),\beta} \) the simplex-valued solution to (2.12) with \( \varepsilon = \varepsilon_n \) therein under the same initial condition \( (0, p_0) \) as before but under the deterministic control \( \beta \) (see [43, Proposition 2.2] for a solvability result). Differently from the analysis performed for \( \mathbf{p}^{(n)} \), the choice of the noise does not really matter here, meaning that we can work with the original Brownian motions \( (\mathcal{B}^{i,j})_{i \neq j} \). Indeed, by the same localization argument as in Steps 2 and 3, it can be proved by a standard stability argument (without any further need of weak compactness) that \( \mathbb{E} \sup_{t \in [0, T]} |J_t^{(n),\beta} - J_t^{(n),\beta}| \) tends to 0 as \( n \) tends to \( \infty \). We deduce \( \lim_{n \rightarrow \infty} J^{(n)}(\beta) = J(\beta) \). Therefore we obtain

\[
\mathbb{E}[J(\mathbf{z})] \leq \liminf_{n \rightarrow \infty} J^{(n)}(\mathbf{z}^{(n)}) \leq \limsup_{n \rightarrow \infty} J^{(n)}(\mathbf{z}^{(n)}) \leq \lim_{n \rightarrow \infty} J^{(n)}(\beta) = J(\beta) \quad (5.7)
\]

for any \( \beta \in \mathcal{E} \). Provided that all the minimizers of \( J \) belong to \( \mathcal{E} \), this implies that \( \mathbf{z} \) belongs with probability 1 to the set of minimizers of \( J \) over \( \mathcal{E} \) and further that (5.1) holds (in particular the limit exists). The fact that optimizers of \( J \)—over \( L^\infty \) controls—belong to \( \mathcal{E} \) is proved in the next Proposition 5.2, together with other properties of the inviscid MFCP.

**Step 5.** As for the proof of (5.2), we can assume without any loss of generality that \( t_0 = 0 \). Observing that the family \( (\mathcal{V}_{\theta, \delta, \varepsilon}(0, p_0))_{\theta, \delta, \varepsilon} \) is bounded (since \( F \) and \( G \) themselves are bounded and the control process in (2.11)–(2.12) is bounded by \( M \)), (5.2) follows from (5.1) together with a standard compactness argument. \( \square \)

### 5.2. Properties of the inviscid MFCP

Before we turn to the proof of Theorem 2.6, we address various properties of the value function of the inviscid MFCP. In this respect, it is useful to work with the system of local coordinates \( (x_1, \ldots, x_{d-1}) \) introduced in Subsection 3.1. The dynamics over which the MFCP (2.10) is defined then have the form (sums are here over \( [d-1] \))

\[
\dot{x}_i^d = \sum_{j \neq i} x_i^j \dot{x}_i^j - x_i^{d,i} + x_i^{d,d} \dot{x}_i^d - x_i^{d,i} \dot{x}_i^d, \quad (5.8)
\]

for \( i \in [d-1] \), with the useful notation that \( x_i^{-d} = 1 - \sum_{i \in [d-1]} x_i^d \). Above, the (deterministic) control \( \mathbf{z} = ((z_i^j)_{i,j \in [d]})_{0 \leq t \leq T} \) is as in (2.3). As we already explained, we assume it to be bounded (but not uniformly bounded by \( M \)), see Remark 5.4 for more detail. Also, the initial condition is taken in the interior of \( \mathcal{S}_d \), which implies in particular that the whole path \( \mathbf{x} \) remains in the interior of the simplex. Lastly, following (3.2) and (3.7) (but paying attention that \( M \) is formally taken as \( +\infty \)), the Hamiltonian of the problem is given, for \( z \in \mathbb{R}^{d-1} \), by (sums are here over \( [d-1] \))
\[ \mathcal{H}(x, z) = \sum_i x_i \dot{H}_i(z) + x^{-d} \dot{H}_d(z), \]
with \[ \dot{H}_i(z) = -\frac{1}{2} \left( \sum_{j \neq i} (z_i - z_j)_+^2 + (z_i)_+^2 \right), \quad \dot{H}_d(z) = -\frac{1}{2} \sum_i (-z_i)_+^2. \] (5.9)

It is important to observe that this Hamiltonian is strictly concave in \( z \), for any \( x \in \text{Int}(\hat{S}_d) \). Indeed, \( \dot{H} \) is the sum of a concave function and of \( -\frac{1}{2} \min_{i \in [d-1]} (x, x^{-d}) \sum_z z_i^2 \). Moreover, we may also write down the corresponding Pontryagin principle:

\[
\begin{aligned}
\dot{x}_i &= \sum_{j \neq i} (x_i(z_i - z_j)_+ - x_i(z_i - z_j^d)_+) + x_i(-z_i)_+ - x_i(z_i^d)_+, \\
\dot{z}_i &= -(H(z_i) - \dot{H}_d(z_i) + \dot{f}_i(x_i) - \dot{f}_i(x_i), \quad \dot{Z}_T = \hat{g}(x_T) - \hat{g}_d(x_T),
\end{aligned}
\] (5.10)

for \( i \in [d - 1] \) and for a given initial condition in \([0, T] \times \text{Int}(\hat{S}_d)\). It is worth noticing that the Pontryagin principle is here stated in local coordinates, or equivalently in dimension \( d - 1 \). For sure, we can also state it in dimension \( d \), in which case the forward-backward system coincides with the standard MFG system (the sum below is over \([d]\))

\[
\begin{aligned}
\dot{p}_i &= \sum_{j \neq i} (p_i(u_i - u_j)_+ - p_i(u_i - u_j^d)_+), \\
\dot{u}_i &= -(H((u_i - u_j)_+)_j) + f_i(p_i), \quad u_T = g_i(p_T),
\end{aligned}
\] (5.11)

with \( H \) as in (2.24). It is pretty easy to see that the two systems (5.10) and (5.11) are equivalent: Given a solution \( u \) to (5.11), it suffices to let \( z = (z_i := (u_i - u_i))_{i \in [d]} \) \( t_0, t \in [d - 1] \), where \( t_0 \) is the initial time. Conversely, given \( z \) a solution to (5.11), it suffices to solve (5.11) where all the occurrences of \( u_i - u_i \) have been replaced by \( z_i - z_i^d \) if \( j \in [d - 1] \) and by \( z_i^d \) if \( j = d \).

The fact that \( \dot{H} \) is strictly concave permits to apply to our situation several results from [49, Chapter 7, Section 4], which we collect in the form of a single proposition, although part of the notions are introduced in detail or explicitly used in Section 6. It is worth mentioning that the results of [49] are stated for a dynamics in \( \mathbb{R}^d \), but is straightforward to see that they apply also to our situation because, when working in local coordinates, any trajectory remains in \( \text{Int}(\hat{S}_d) \), if starting from \( \text{Int}(\hat{S}_d) \).

**Proposition 5.2.** Assume that \( F \) and \( G \) are in \( C^1(\hat{S}_d) \). Recall that \( V : [0, T] \times \hat{S}_d \to \mathbb{R} \) is the value function of the MFCP (2.10), and call \( \hat{V} : [0, T] \times \hat{S}_d \to \mathbb{R} \) its formulation in local coordinates, i.e. \( \hat{V}(t, x) = V(t, \hat{x}) \). If the initial condition \( p_0 \in \text{Int}(\hat{S}_d) \), then

i. An optimal (bounded) control exists and is bounded by \( M = 2(||g||_{\infty} + T||f||_{\infty}) \);

ii. If \( \alpha \) is an optimal control and \( p \) the related optimal trajectory, then there exist \( u \) solving (5.11) and \( z \) solving (5.10), and \( \alpha \) is given by \( (z_i)_{t_0} = (u_i - u_i^d)_{t_0} \leq T \);

iii. \( \hat{V} \) is a viscosity solution of (2.23) on \([0, T] \times \text{Int}(\hat{S}_d)\), at least when (2.23) is formulated in local coordinates, see Definition 6.1;

iv. \( \hat{V} \) is (time-space) Lipschitz-continuous on \([0, T] \times \text{Int}(\hat{S}_d)\) and thus also on \([0, T] \times \hat{S}_d\).

If \( F \) and \( G \) are in \( C^{1,1}(\hat{S}_d) \), then
\begin{enumerate}
  \item \( \dot{V} \) is semiconcave on \([0, T] \times \text{Int}(\hat{S}_d) \) and thus also on \([0, T] \times \hat{S}_d \); 
  \item \( V \) is differentiable at \((t, p_1)\), for any \( t > t_0 \) and any optimal trajectory \( p \) starting from \((t_0, p_0)\), with \( p_0 \in \text{Int}(S_d) \); 
  \item If the optimal control for the MFCP starting from \((t_0, p_0)\) is unique (in particular this holds true when \((5.11) \) is uniquely solvable), then \( V \) is differentiable in \((t_0, p_0)\).
\end{enumerate}

If we assume in addition that \( V \) is differentiable at \((t_0, p_0)\), with \( p_0 \in \text{Int}(S_d) \), then

\begin{enumerate}
  \item There exists a unique optimal control process \( \alpha \) and an optimal trajectory \( p \) for the MFCP starting from \((t_0, p_0)\), and the optimal control is given in feedback form by
    \[ x^i(t) = (d_i V(t, p_1) - d_j V(t, p_1))_+, \quad t \in [t_0, T]; \]
  \item The adjoint equation in \((5.10) \) is such that \( z^i(t, x_t) = \partial_x \dot{V} \) for any \( t \in [t_0, T] \)
    and \( i \in [d - 1] \).
\end{enumerate}

In point (iii), we refer to Definition 6.1 for a reminder on the notion of viscosity solution. In points (v)–(ix), we refer to the assumption of Theorem 2.6 for the meaning of the notation \( C^{1,1} \). In point (v), by time-space semiconcavity, we mean that there exists a constant \( c \) such that, for any \( t \in [0, T], x \in \text{Int}(\hat{S}_d), s \) with \( t \pm s \in [0, T] \) and \( \zeta \) with \( x \pm \zeta \in \text{Int}(\hat{S}_d) \),

\[ v(t + s, x + \zeta) - 2v(t, x) + v(t - s, x - \zeta) \leq c. \tag{5.12} \]

Notice that we also exploit the notion of semiconcavity, but in space only, in the next section, see \((6.11) \). Also, not only in the statement but also throughout the rest of the text, differentiability of \( V \) is understood as time-space differentiability (unless it is stated differently, in which case differentiability is explicitly referred to as space differentiability). Last, we stress that (vii) follows from Theorem 7.4.20 in [49], but the statement therein assumes that the Hamiltonian is strictly convex; in fact, it is clear from the proof that the authors mean strictly convex in \( z \) only.

**Proof.** To prove (i), assume first that controls are bounded by \( R \), for some \( R > M \). Then an optimal control \( \alpha_R \) exists by [49, Theorem 7.4.5] and, by the Pontryagin principle [49, Theorem 7.4.17], point (ii) holds but with the truncated Hamiltonian \( (H_R^i)_{i \in [d]} \) defined as in \((3.2) \). Thus, \( \alpha_R \) induces an equilibrium \( \dot{p}^*(\alpha_R) \) to the MFG \((2.1)–(2.2) \) and (using the coercivity of the Lagrangian on the interior of the simplex) is of the form \((x^i_0, 0)_{i \in [d]} \) is the value process associated with the optimization problem \( J(\cdot; \dot{p}^*(\alpha_R)) \) in \((2.1) \), set over controls that are bounded by \( R \). Choosing 0 as control in \((2.1) \), we observe that \( u \) is upper bounded by \( M/2 \). In order to prove that \(-M/2 \) is a lower bound, it suffices to lower bound the quadratic cost by zero in the cost functional \( J(\cdot; \dot{p}^*(\alpha_R)) \). Hence, \( \alpha_R \) is bounded by \( M \), which is independent of \( R \), implying that an optimal control exists over the set of bounded controls. Therefore, (i) and (ii) are proved and the other points follow now from the results in [49, Section 7.4]. \( \square \)
Since $V$ is almost everywhere differentiable in $[0, T] \times S_d$, the above result, together with Theorem 5.1, implies that the sequence of optimal trajectories $(p^{\theta, \delta, \varepsilon}_{t_0, t})_{0 < t_0 \leq t \leq t_0 + \varepsilon}$ admits a true limit for almost every initial condition $(t_0, p_0)$ (the convergence hence holding true in probability). Moreover, point (vi) above permits to say more about the convergence also when starting from a point of non-differentiability: The randomness of the limit trajectory is enclosed in the initial time only. We summarize in the following:

**Corollary 5.3.** Assume that $F$ is in $C^{1,1}(S_d)$ and $G$ in $C^{1,2+\gamma}_{WF}(S_d)$ for a given $\gamma \in (0, 1)$. Then, if $V$ is differentiable in $(t_0, p_0)$, with $p_0 \in \text{Int}(S_d)$, then, the following holds true in probability (the first one on $C([0, T]; S_d)$ and the second one on $E$),

$$
\lim_{(t_0, \delta, \varepsilon) \to (0, 0, 0)} p_{[t_0, p_0]}^{\theta, \delta, \varepsilon} = p_{[t_0, p_0]} \quad \text{and} \quad \lim_{(t_0, \delta, \varepsilon) \to (0, 0, 0)} \mathbf{x}_{[t_0, p_0]}^{\theta, \delta, \varepsilon} = \mathbf{x}_{[t_0, p_0]},
$$

where $p_{[t_0, p_0]}$ is the unique optimal trajectory and $\mathbf{x}_{[t_0, p_0]}$ the unique optimal control process of the limiting MFCP, see the notation in the introduction of Subsection 5.1.

Moreover, if $V$ is not differentiable in $(t_0, p_0)$, the limit of any converging subsequence is supported on a set of (optimal) trajectories which do not branch strictly after the initial time.

**Remark 5.4.** We comment more on our choice to restrict ourselves to controls that are bounded in (2.3). The unbounded case looks more difficult. One issue is that the Lagrangian $\frac{1}{2} \sum_i p_i \sum_{j \neq i} |x_{ij}|^2$ is not Lipschitz continuous in $p$, uniformly in $x$, if $x$ is not in a compact set. Another issue is that the Lagrangian is not uniformly coercive on the simplex. Due to the latter, we can easily find some unbounded controls that drive the trajectory to the boundary but that remains of a bounded energy. For sure, we could think of using some truncation argument, as done in the proof of Proposition 5.2 above, in order to prove the existence of optimal controls over $L^2([0, T]; \mathbb{R}^{d \times d})$ that are bounded by $M$. However, we prefer not to address this question and assume instead that controls are bounded. We stress that this is in no loss of generality for our own purpose, since the inviscid MFCP is mainly used to select some of the MFG solutions. It is indeed obvious that the latter ones are bounded by $M$, which explains why controls in $L^\infty$ are sufficient in our study.

### 5.3. Selection for the master equation

Although Corollary 5.3 provides an interesting information about the limiting behavior of the equilibrium $p^{\theta, \delta, \varepsilon}$ as the parameters $(\theta, \delta, \varepsilon)$ tend to 0, it says nothing about the asymptotic behavior of the related equilibrium cost. We address this question in this subsection; in particular, we prove here Proposition 2.5 and Theorem 2.6. Throughout, we assume that $F \in C^{1,\gamma}_{WF}(S_d)$ and $G \in C^{1,2+\gamma}_{WF}(S_d)$ for a given $\gamma \in (0, 1)$. At some point, we need to strengthen the condition on $F$ and assume it to belong to $C^{1,1}(S_d)$, see Proposition 5.11 and Theorem 5.12.

Actually, part of the difficulty for passing to the limit in the cost $\tilde{J}^{\varepsilon, \theta}$ defined by (4.9) is to control the distance from the equilibrium to the boundary. Back to the formulas (2.17)–(2.18), it is indeed plain to see that $\varphi$ should become steeper and steeper (and
hence \( |\varphi'| \) larger and larger) in the neighborhood of 0 as \((\theta, \delta, \varepsilon)\) tends to 0, whence the need for some uniform integrability properties on the inverse of the distance from \(p^{0,\theta,\varepsilon}\) to the boundary. We here collect several useful \textit{a priori} bounds in this direction. Proofs of the first three statements are postponed to the end of the section, see Subsection 5.4.

**Proposition 5.5.** For \((\theta, \delta, \varepsilon)\) and \(\varphi = \varphi_{\theta,\delta,\varepsilon}\) as in (2.17)–(2.18), with \(\kappa_0 \geq \varepsilon^2 / 2\) and \(\varepsilon_0 := \kappa_2 / \kappa_0 \geq \varepsilon^2\), and for any initial condition \((t_0, p_0) \in [0, T] \subset \text{Int}(S_d)\) and any \(\lambda > 0\) and \(i \in [d]\),

\[
\mathbb{E} \left[ \exp \left\{ \frac{\lambda}{\varepsilon^2} \left( \kappa_0 - \frac{\varepsilon^2 + \lambda}{2} \right) \int_{t_0}^{T} \frac{1}{p^{i,\theta,\varepsilon}_t} 1_{[0,\delta]}(p^{i,\theta,\varepsilon}_t) dt \right\} \right] \leq \frac{e^{TC_\delta(\varepsilon,\lambda)}}{(p_0^i)^{\lambda/\varepsilon^2}},
\]

(5.14)

\[
\mathbb{E} \left[ \exp \left\{ \lambda \left( \kappa_0 - \frac{\varepsilon^2 (1 + \lambda)}{2} \right) \int_{t_0}^{T} \frac{1}{p^{i,\theta,\varepsilon}_t} 1_{[0,\delta]}(p^{i,\theta,\varepsilon}_t) dt \right\} 1_{\{\inf_{0 \leq t \leq T} p^{i,\theta,\varepsilon}_t > 0\}} \right] \leq \frac{e^{TC(\varepsilon,\lambda)}}{(p_0^i)^{\lambda}},
\]

(5.15)

with \(C(\delta, \varepsilon, \lambda) := e^{-\varepsilon^2}[(1 + \lambda)/(2\delta) + \lambda d(\kappa_\varepsilon + \kappa_0 + M)]\) and \(C(\delta, \lambda) := (1 + \lambda)/(2\delta) + \lambda d(\kappa_0 + M)\), and where \(p^{0,\theta,\varepsilon}\) is here understood as \(p^{0,\theta,\varepsilon}_{[t_0,p_0]}\).

**Proposition 5.6.** For any \(\lambda \geq 1\), there exists a constant \(\kappa_0\) (depending on \(\lambda\) and \(\kappa_2\)) such that, for any \(\kappa_0 \geq \kappa_0\) and any compact subset\(^4\) \(K \subset \text{Int}(S_d)\), we can find (strictly) positive constants \(\tilde{C}, \delta_0, \varepsilon_0\) and (strictly) positive-valued functions \(\tilde{\theta}(\delta, \varepsilon), \tilde{\varepsilon}(\delta)\) and \(\tilde{\varepsilon}(\varepsilon)\) converging to 0 in \((0, 0), 0\) and 0 respectively (all these items only depending on \(\kappa_{0}, \kappa_{2}, K, \lambda, M, T\), and \(d\)), such that

\[\forall \delta \in (0, \delta_0), \forall e \in (0, \varepsilon(\delta)), \forall \theta \in (0, \tilde{\theta}(\delta, e)), \mathbf{P}(\lambda, \theta, \delta, \varepsilon, \mathcal{K}) \leq \tilde{C},\]

(5.16)

\[\forall e \in (0, \varepsilon_0), \forall \theta \in (0, \tilde{\varepsilon}(\delta, e)), \mathbf{P}(\lambda, \theta, \delta, \varepsilon, \mathcal{K}) \leq \tilde{C},\]

(5.17)

where

\[\mathbf{P}(\lambda, \theta, \delta, \varepsilon, \mathcal{K}) = \max_{i \in [d]} \sup_{(t_0, p_0) \in [0, T] \times K} \mathbb{E} \left[ \exp \left\{ \lambda \left( \int_{t_0}^{T} \left[ \varphi_{\theta,\delta,\varepsilon} - \varphi'_{\theta,0,\varepsilon} \right] (p^{i,0,\varepsilon}_{[t_0,p_0],t}) + \frac{1}{p^{i,0,\varepsilon}_{[t_0,p_0],t}} \right) dt \right\} \right].\]

With the same notations, it also holds that

\[\forall e \in (0, \varepsilon_0), \forall \theta \in (0, \tilde{\theta}(\delta, e)),\]

\[\min_{i \in [d]} \inf_{(t_0, p_0) \in [0, T] \times K} \mathbb{P} \left( \inf_{0 \leq t \leq T} p^{i,0,\varepsilon}_{[t_0,p_0],t} \geq \tilde{C} \right) \geq 1 - 2 \exp \left( -e^{-1} \right),\]

(5.18)

In what follows, we prefer to state the convergence result as limits as the viscosity parameter \(\varepsilon^2\) tends to 0, instead of limits as \(\delta\) tends to 0, which explains why, in (5.17) and (5.18), we consider \(\delta\) as a function of \(\varepsilon\), and \(\theta\) as a function of \(\varepsilon\) and \(\delta\). In order to formulate the next statement properly, we need another notation. Similar to \(p^{0,\theta,\varepsilon}_{[t_0,p_0]}\), \(q^{0,\theta,\varepsilon}_{[t_0,p_0],t} := (q^{0,\theta,\varepsilon}_{[t_0,p_0],t})_{0 \leq t \leq T}\) denotes the solution to (2.5) with \(q_0\) as initial condition at time \(t_0\), when \((p, x)\) therein is understood as \((p^{0,\theta,\varepsilon}_{[t_0,p_0]}, x^{0,\theta,\varepsilon}_{[t_0,p_0]})\). In particular, it should be

\(^4\text{Here,} \mathcal{K} \text{ is regarded as a compact subset of} S_d, \text{ but, obviously, we could regard it as a} (d - 1)\text{-dimensional compact subset of} S_d.\)
clear for the reader that $p_{[t_0,p_0]}^{\theta, \delta, \varepsilon}$ and $q_{[t_0,p_0,p_0]}^{\theta, \delta, \varepsilon}$ are the same. When there is no ambiguity on the choice of the initial condition, we merely write $q$.

**Lemma 5.7.** For $\ell \geq 1$, we can find $\lambda := \tilde{\lambda}(\ell)$, only depending on $\ell$ and $d$ and then take $\bar{\kappa}_0$ accordingly in Proposition 5.6 (in terms of $\lambda$ and $\kappa_2$ only) such that, for any $\kappa_0 \geq \bar{\kappa}_0$ and any compact subset $K$ included in $\text{Int}(S_d)$, it holds that, for $\bar{\varepsilon}_0, \bar{\theta}$ and $\bar{\delta}$ as in Proposition 5.6, for any state $i \in [d]$ and any initial point $(t_0, p_0, q_0) \in [0, T] \times K \times S_d$, and for any $\varepsilon \in (0, \bar{\varepsilon}_0]$ and $\theta \in (0, \bar{\theta}(\bar{\delta}(\varepsilon), \varepsilon)]$,

$$
\sup_{t_0 \leq t \leq T} \mathbb{E} \left[ \sum_{i \in [d]} (p_{[t_0,p_0],t}^{i, \theta, \delta(\varepsilon), \varepsilon})^\ell \right] \leq C,
$$

(5.19)

$$
\mathbb{E} \left[ \sup_{t_0 \leq t \leq T} \sum_{i \in [d]} (q_{[t_0,p_0,p_0],t}^{i, \theta, \delta(\varepsilon), \varepsilon})^\ell \right] \leq C,
$$

(5.20)

where $C$ depends only on $\kappa_0, \kappa_2, K, \ell, M, T$ and $d$.

**Proposition 5.8.** We can find $\bar{\kappa}_0 \geq 0$, only depending on $\kappa_2$ and $d$, such that, for any $\kappa_0 \geq \bar{\kappa}_0$ and any compact subset $K$ included in $\text{Int}(S_d)$, there exist constants $\bar{C}$ and $\bar{\varepsilon}_0$, only depending on $\kappa_0, \kappa_2, K, ||f||_{\infty}, ||g||_{\infty}, T$ and $d$ and functions $\bar{\theta}(\bar{\delta}(\varepsilon))$ and $\bar{\hat{\delta}}(\varepsilon)$ as in the statement of Proposition 5.6 (with $\lambda$ therein a fixed constant whose value is made explicit in the proof in terms of $d$ only and is, in particular, required to be greater than $\lambda(12)$ in Lemma 5.7) such that, for $V_{\theta, \delta, \varepsilon} = (V_{\theta, \delta, \varepsilon}^i)_{i \in [d]}$ denoting the solution to (3.9) with $\varphi = \varphi_{\theta, \delta, \varepsilon}$ therein, and for any $\varepsilon \in (0, \bar{\varepsilon}_0]$ and $\theta \in (0, \bar{\theta}(\bar{\delta}(\varepsilon), \varepsilon)]$,

$$
\sup_{t \in [0, T]} \sup_{p \in K} \max_{i \in [d]} |V_{\theta, \delta(\varepsilon), \varepsilon}^i(t, p)| \leq \bar{C}.
$$

(5.21)

Moreover, if $(t, p_0) \in [0, T] \times K$, for the same values of $\kappa_0$, $\varepsilon$ and $\theta$ (indices in the sums below being taken in $[d]$),

$$
\mathbb{E} \left[ \sup_{t_0 \leq t \leq T} \max_{i \in [d]} |V_{\theta, \delta(\varepsilon), \varepsilon}^i(t, p_{[t_0,p_0]}^{\theta, \delta(\varepsilon), \varepsilon})|^2 \right] \leq \bar{C},
$$

(5.22)

$$
\mathbb{E} \left[ \int_{t_0}^T \sum_{i,j,k} |W_{\theta, \delta(\varepsilon), \varepsilon}^{i,j,k}(t, p_{[t_0,p_0]}^{\theta, \delta(\varepsilon), \varepsilon})|^2 dt \right] \leq \bar{C},
$$

(5.23)

$$
\mathbb{E} \left[ \int_{t_0}^T \sum_{i,j,k} |Y_{\theta, \delta(\varepsilon), \varepsilon}^{i,j,k}(t, p_{[t_0,p_0]}^{\theta, \delta(\varepsilon), \varepsilon})|^2 dt \right] \leq \bar{C},
$$

(5.24)

where $W_{\theta, \delta(\varepsilon), \varepsilon}$ is defined by (4.3) and $Y_{\theta, \delta(\varepsilon), \varepsilon}$ by (4.7), with $\varphi = \varphi_{\theta, \delta(\varepsilon), \varepsilon}$.

**Proof of Proposition 5.8.** For a suitable $\lambda > 1$ that will be fixed in (5.25) below in terms of $d$ only, we consider $\bar{\kappa}_0$ as in the statement of Proposition 5.6 but with $\lambda$ therein replaced by $2\lambda d$ (the need for changing $\lambda$ into $2\lambda d$ is made clear in the proof, see again the discussion right after (5.25); in short $\lambda$ in the statement should be understood as $2\lambda d$ in the sequel of the proof). Then, for any $\kappa_0 \geq \bar{\kappa}_0$ and any compact subset $K$ included in $\text{Int}(S_d)$, we consider $\tilde{\delta}_0, \bar{\varepsilon}_0, \tilde{\theta}(\bar{\delta}(\varepsilon))$ and $\tilde{\hat{\delta}}(\varepsilon)$ also as in the statement of Proposition 5.6. We then fix $\varepsilon \in (0, \bar{\varepsilon}_0)$, $\theta \in (0, \tilde{\theta}(\bar{\delta}(\varepsilon), \varepsilon)]$ and $(t_0,p_0) \in [0, T] \times K$ and we write $p$ for the process $p_{[t_0,p_0]}^{\theta, \delta(\varepsilon), \varepsilon} = (p_{[t_0,p_0]}^{i, \theta, \delta(\varepsilon), \varepsilon})_{t_0 \leq t \leq T}$, $\varphi$ for $\varphi_{\theta, \delta(\varepsilon), \varepsilon}$ and $(V_{\theta, \delta(\varepsilon), \varepsilon})_{i \in [d]}$ for the
corresponding solution to (3.9), and similarly for \((W^{i,j,k})_{i,j,k \in [d]}\) and \((\gamma^{i,j})_{i,j \in [d]}\). We then let \((v^i_t := V^i(t,p_t))_{0 \leq t \leq T}\), for \(i \in [d]\), and \((w^{i,j,k}_t := W^{i,j,k}(t,p_t))_{0 \leq t \leq T}\), for \(i,j,k \in [d]\). We know that \((v_t,w_t)_{0 \leq t \leq T}\) satisfy (4.2). We consider then

\[E_t := \exp \left\{ \lambda \int_t^T \sum_{j \in [d]} \left[ (\varphi - \varphi')(p^j_t) + (p^j_t)^{-1} \right] ds \right\}, \quad t \leq T.\]

Itô’s formula and (4.2) give (indices being taken in \([d]\))

\[
d \left( E_t |v^i_t|^2 \right) = 2E_t v^i_t dv^i_t + \lambda E_t |v^i_t|^2 \sum_j (\varphi - \varphi')(p^j_t) + (p^j_t)^{-1} dt + E_t \sum_{j,k} |w^{j,k}_t|^2 dt
\]

\[
= -2E_t v^i_t \left( \frac{1}{d} \sum_j \dot{H}^j(v_t) + f^j(p_t) - \frac{1}{d} \sum_j f^j(p_t) \right) dt
\]

\[
- 2E_t v^i_t \sum_j (\varphi(p^j_t) - p^j_t \varphi'(p^j_t))(v^j_t - v^i_t) - \frac{1}{d} \sum_{j,i} (\varphi(p^j_t) - p^j_t \varphi'(p^j_t))(v^j_t - v^i_t) dt
\]

\[
- \sqrt{2} \lambda E_t v^i_t \left( \sum_j \sqrt{p^j_t^{-1}} (w^{j,j}_t + w^{j,i}_t) - \frac{1}{d} \sum_{j,i} \sqrt{p^j_t^{-1}} (w^{j,j}_t + w^{j,i}_t) \right) dt
\]

\[
+ 2E_t v^i_t \sum_{j,k} w^{j,k}_t dB^{j,k}_t + \lambda E_t |v^i_t|^2 \sum_j (\varphi - \varphi')(p^j_t) + (p^j_t)^{-1} dt + E_t \sum_{j,k} |w^{j,k}_t|^2 dt.
\]

Integrating from \(t \geq t_0\) to \(T\) and using the Lipschitz continuity of the Hamiltonian and the boundedness of \(f\) and \(g\), we deduce that there exists a constant \(C\), which is allowed to vary from line to line as long as it only depends on the same parameters as \(C\) in the statement, such that

\[
E_t |v^i_t|^2 + \lambda \int_t^T E_s |v^i_s|^2 \sum_j (\varphi - \varphi')(p^j_s) + (p^j_s)^{-1} ds + \int_t^T E_s \sum_{j,k} |w^{j,k}_s|^2 ds
\]

\[
\leq E_T |g(p_T)|^2 + 2 \int_t^T E_s v^i_s \sum_{j,k} w^{j,k}_s dB^{j,k}_s + \int_t^T E_s |v^i_s|^2 \left\{ C + C |v_s| + c_d |v_s| \sum_j |\varphi - \varphi'|(p^j_s) \right\} ds
\]

\[
+ c_d \sum_{j,i} \sqrt{p^j_t^{-1}} (|w^{j,j}_t| + |w^{j,i}_t|) \right\} ds,
\]

where \(c_d\) only depends on \(d\). Hence, by Young’s inequality \(ab \leq 2\eta a^2 + b^2/2\eta\), which holds true for any \(\eta > 0\),

\[
E_t |v^i_t|^2 + \lambda \int_t^T E_s |v^i_s|^2 \sum_j (\varphi - \varphi')(p^j_s) + (p^j_s)^{-1} ds + \int_t^T E_s \sum_{j,k} |w^{j,k}_s|^2 ds
\]

\[
\leq C E_T + C \int_t^T E_s (1 + |v_s|^2) ds + c_d \int_t^T E_s v^i_s \sum_j |\varphi - \varphi'|(p^j_s) ds
\]

\[
+ 8\eta \int_t^T E_s \sum_{j,i} |w^{j,j}_s|^2 ds + \frac{c_d^2}{2\eta^2} \int_t^T E_s |v^i_s|^2 \sum_j (p^j_s)^{-1} ds + 2 \int_t^T E_s v^i_s \sum_{j,k} w^{j,k}_s dB^{j,k}_s.
\]

By summing over \(i \in [d]\), we get
\[ E_i|v_i|^2 + \lambda \int_t^T E_i |v_s|^2 \sum_j (|\varphi - \varphi'(p_j)|^2 + (p_j')^{-1}) \, ds + \int_t^T E_i \sum_{i,j,k} |w_s^{i,j,k}|^2 \, ds \]
\[
\leq C E_T + C \int_t^T E_i (1 + |v_s|^2) \, ds + c_d \int_t^T E_i |v_s|^2 \sum_j (|\varphi - \varphi'(p_j)|^2) \, ds \\
+ 8\eta_d \int_t^T E_i \sum_{i,j,k} |w_s^{i,j,k}|^2 \, ds + \frac{c_d^2}{2\eta} \varepsilon^2 \int_t^T E_i |v_s|^2 \sum_j (p_j')^{-1} \, ds + 2 \int_t^T E_i \sum_{i,j,k} v_s^{i,j,k} d B_s^{i,j,k}.
\]

Choosing \( \eta = 1/(16d) \) and \( \lambda = \max(\frac{\lambda(12)}{(2d)}, (32c_d + c_d)d + 1/2) \), we obtain
\[
E_i|v_t|^2 + \frac{1}{2} \int_t^T E_i |v_s|^2 \sum_j (|\varphi - \varphi'(p_j)|^2 + (p_j')^{-1}) \, ds + \frac{1}{2} \int_t^T E_i \sum_{i,j,k} |w_s^{i,j,k}|^2 \, ds \\
\leq C E_T + C \int_t^T E_i (1 + |v_s|^2) \, ds + 2 \int_t^T E_i \sum_{i,j,k} v_s^{i,j,k} d B_s^{i,j,k}.
\] (5.25)

The stochastic integral is a martingale since \( (v_t^i)_{0 \leq t \leq T} \) and \( (w_s^{i,j,k})_{0 \leq t \leq T} \) are bounded (possibly not uniformly in \( \varepsilon \) at this stage of the proof). Also, by (5.17) in Proposition 5.6, replacing therein \( \lambda \) by \( 2\lambda d \) (as we already explained) and then using Hölder’s inequality, we have \( \mathbb{E}[E_T^2] \leq C \) for our choices of \( \kappa_0 \) and \( \kappa_0 \) (the latter being greater than \( \kappa_0 \)). Therefore, taking expectation in the above inequality and applying Gronwall’s lemma, we get
\[
\sup_{0 \leq t \leq T} \mathbb{E}[E_t|v_t|^2] + \mathbb{E}\left[ \int_0^T E_i \sum_{i,j,k} |w_s^{i,j,k}|^2 \, dt \right] \leq C.
\] (5.26)

In order to pass the supremum inside the expectation in the first term of the left-hand side, we return back to (5.25), take the supremum therein and then apply Burkholder-Davis-Gundy’s inequality to handle the martingale, noticing that
\[
\mathbb{E}\left[ \left( \int_0^T E_i \sum_{i,j,k} |v_s^{i,j,k}|^2 \, dt \right)^{1/2} \right] \leq \mathbb{E}\left[ \left( \sup_{0 \leq t \leq T} E_i |v_t|^2 \right)^{1/2} \right] \left( \int_0^T E_i \sum_{i,j,k} |w_s^{i,j,k}|^2 \, dt \right)^{1/2} \\
\leq C \mathbb{E}\left[ \sup_{0 \leq t \leq T} E_i |v_t|^2 \right]^{1/2},
\]
where we used (5.26) together with Cauchy-Schwarz inequality to get the last line. We easily obtain
\[
\mathbb{E}\left[ \sup_{0 \leq t \leq T} E_i |v_t|^2 \right] \leq C + C \mathbb{E}\left[ \sup_{0 \leq t \leq T} E_i |v_t|^2 \right]^{1/2},
\]
which is enough to derive (5.22), recalling that the left-hand side is already known to be finite. Taking \( t = t_0 \) in (5.22) and then letting \( (t_0, \rho_0) \) vary over the entire \( [0, T] \times \mathcal{K} \), we obtain (5.21).

Inequality (5.23) derives from (5.26). Finally, in order to prove (5.24), we return back to (4.5) and then expand \( \mathcal{Y}(t, p_t)_{0 \leq t \leq T} \) by Itô’s formula. We get
\[ \mathcal{V}(t, p_t) = \langle p_T, g^*(p_T) \rangle + \int_t^T \left[ \sum_{j,k} \left( \frac{1}{2} p_j^k x_j^k \right)^2 + p_j^k \phi'(p_j^k)(v_j^k - v_j^k) + \langle p_s, f^*(p_s) \rangle \right] ds \\
+ \frac{1}{\sqrt{2}} \varepsilon \int_t^T \sqrt{p_j^k (d_j \mathcal{V}(s, p_s) - d_k \mathcal{V}(s, p_s))} dB_j^k. \]

By Proposition 5.6 and (5.26) (recall also that \( \varepsilon \) is bounded by \( M \)), we have a bound for the second order moment of the right-hand side in the first line. Taking \( t = t_0 \), passing the stochastic integral to the left, squaring the whole equality and then taking expectation, we get the announced result. \( \square \)

We now address the (local) uniform convergence of the value function \( \mathcal{V}_{\theta, \delta, \varepsilon} \) (of the viscous MFCP) toward the value function \( \mathcal{V} \) of the inviscid MFCP. Recall that the convergence is already known to hold pointwise, see Theorem 5.1. Recall also that \( \mathcal{V}_{\theta, \delta, \varepsilon} = \nabla \mathcal{V}_{\theta, \delta, \varepsilon} \) solves (3.9).

**Proposition 5.9.** We can find \( \bar{\kappa}_0 \geq 0 \), only depending on \( \kappa_2 \) and \( d \), such that, for any \( \kappa_0 \geq \bar{\kappa}_0 \) and any compact subset \( K \) included in \( \text{Int}(S_d) \), for the same two functions \( \tilde{\theta}(\delta, \varepsilon) \) and \( \tilde{\delta}(\varepsilon) \) as in the statement of Proposition 5.8 (which only depend on \( \kappa_0, \kappa_2, K, ||f||_\infty, ||g||_\infty, T \) and \( d \)), it holds that

\[ \lim_{\varepsilon \to 0} \mathcal{V}_{\tilde{\theta}(\delta(\varepsilon), \varepsilon), \tilde{\delta}(\varepsilon), \varepsilon} = \mathcal{V}, \tag{5.27} \]

uniformly on \([0, T] \times K\).

**Proof.** Throughout the proof, we consider \( \bar{\kappa}_0 \) as in the statement of Proposition 5.8 and then, for \( \kappa_0 \geq \bar{\kappa}_0 \) and for two compact subsets \( K \) and \( K' \) included in \( \text{Int}(S_d) \) such that the interior of \( K' \) contains \( K \), we consider \( \tilde{\varepsilon}_0, \tilde{\theta}(\delta, \varepsilon) \) and \( \tilde{\delta}(\varepsilon) \) as in the statement of Proposition 5.8 when the compact subset therein is not \( K \) but \( K' \). For simplicity, we let \( V_{\varepsilon} := \mathcal{V}_{\tilde{\theta}(\delta(\varepsilon), \varepsilon), \tilde{\delta}(\varepsilon), \varepsilon} \) and similarly for \( V_\varepsilon \) for \( \varepsilon \in (0, \tilde{\varepsilon}_0] \).

**Step 1.** The first step is to prove that the functions \( \{V_{\varepsilon}\}_{0 < \varepsilon \leq \tilde{\varepsilon}_0} \) are uniformly continuous on \([0, T] \times K\). In fact, recalling that \( V_{\varepsilon} = \nabla \mathcal{V}_{\varepsilon} \), we already know from (5.21) that the functions \( \{V_{\varepsilon}\}_{0 < \varepsilon \leq \tilde{\varepsilon}_0} \) are uniformly Lipschitz continuous in space on \([0, T] \times K' \).

In order to prove uniform continuity in time, we fix some \( \varepsilon \in (0, \tilde{\varepsilon}_0] \) together with an initial condition \((t_0, p_{t_0}) \in [0, T] \times K\). Writing \( p^\varepsilon \) for \( p_{\tilde{\theta}(\delta(\varepsilon), \varepsilon), \tilde{\delta}(\varepsilon), \varepsilon} \) and similarly for \( x^\varepsilon \), we define the stopping time \( \sigma_{\varepsilon} := \inf \{ t \geq t_0 : p_j^\varepsilon \notin K' \} \wedge T \). Since \( \tilde{\delta}(\varepsilon) \) tends to 0 with \( \varepsilon \), we can change the value of \( \tilde{\varepsilon}_0 \) in such a way that \( q_i > \tilde{2}\tilde{\delta}(\varepsilon) \), for any \( \varepsilon \in (0, \tilde{\varepsilon}_0], i \in [d] \) and \( q \in K' \). Since \( \phi_{\varepsilon} := \phi_{\tilde{\theta}(\delta(\varepsilon), \varepsilon), \tilde{\delta}(\varepsilon), \varepsilon} \) is zero outside \([0, \tilde{2}\tilde{\delta}(\varepsilon)]\), we deduce that, up to the stopping time \( \sigma_{\varepsilon} \), \( p^\varepsilon \) does not see the function \( \phi_{\varepsilon} \) in its own dynamics (2.12). Also, since the off-diagonal entries of the control \( x^\varepsilon \) in (2.12) are bounded by \( M \), we easily deduce that there exists a constant \( C \), independent of \( \varepsilon \) and \((t_0, p_{t_0}) \), such that, for any \( t \in [t_0, T] \),

\[ \mathbb{E} \left[ \sup_{t_0 \leq s \leq t \wedge \sigma_{\varepsilon}} |p_j^\varepsilon - p_0^\varepsilon|^2 \right] \leq C(t - t_0). \tag{5.28} \]
In particular, denoting by \( \text{dist}(\mathcal{K}, (\mathcal{K}')^\mathcal{C}) \) the distance from \( \mathcal{K} \) to the complementary of \( \mathcal{K}' \) and then allowing the value of \( C \) to vary from line to line (and to depend on both \( \mathcal{K} \) and \( \mathcal{K}' \) but not on \( \epsilon \)), we have

\[
\mathbb{P}(\sigma_\epsilon < t) \leq \mathbb{P}(\sup_{0 \leq s \leq t \wedge \sigma_\epsilon} |p^\epsilon_s - p_0| \geq \text{dist}(\mathcal{K}, (\mathcal{K}')^\mathcal{C})) \leq C(t - t_0). \tag{5.29}
\]

We now apply Itô’s formula to \( (\mathcal{V}_e(t, p_t))_{t_0 \leq t \leq \sigma_\epsilon} \). By the HJB equation (2.13) (see also (3.6)), we obtain, for any \( t \in [t_0, T] \),

\[
\mathcal{V}_e(t_0, p_0) = \mathbb{E}\left[ \int_{t_0}^{t \wedge \sigma_\epsilon} \left( \frac{1}{2} \sum_{i \in [d]} \sum_{j \in [d], j \neq i} |\mathbf{x}_{s}^{n, i, j}|^2 + F(p^\epsilon_s) \right) ds + \mathcal{V}_e(t \wedge \sigma_\epsilon, p^\epsilon_{t \wedge \sigma_\epsilon}) \right].
\]

Subtracting \( \mathcal{V}_e(t, p_0) \) to both sides and recalling that the integrand in the right-hand side can be bounded independently of \( \epsilon \), we deduce that

\[
|\mathcal{V}_e(t_0, p_0) - \mathcal{V}_e(t, p_0)| \\
\leq C(t - t_0) + \mathbb{E}\left[ |\mathcal{V}_e(t \wedge \sigma_\epsilon, p^\epsilon_{t \wedge \sigma_\epsilon}) - \mathcal{V}_e(t, p^\epsilon_{t \wedge \sigma_\epsilon})| \right] + \mathbb{E}\left[ |\mathcal{V}_e(t, p^\epsilon_{t \wedge \sigma_\epsilon}) - \mathcal{V}_e(t, p_0)| \right],
\]

where we used the Lipschitz property of \( \mathcal{V}_e \) in the space variable (at least whenever the latter belongs to \( \mathcal{K}' \)) to derive the last line. Since the value function \( \mathcal{V}_e \) can be bounded independently of \( \epsilon \) (using for instance the fact that controls themselves are required to be bounded), we deduce from (5.28) and (5.29) that

\[
|\mathcal{V}_e(t_0, p_0) - \mathcal{V}_e(t, p_0)| \leq C(t - t_0)^{1/2},
\]

which shows that the functions \((\mathcal{V}_e)_{0 < \epsilon \leq \epsilon_0}\) are uniformly continuous in time (and hence in time and space) on \([0, T] \times \mathcal{K} \).

**Step 2.** Applying Ascoli-Arzelá theorem, we deduce that there exist a subsequence \((\mathcal{V}_e_n)_{n \geq 0}\) and a function \( \overline{\mathcal{V}}_\mathcal{K} \), a priori depending on \( \mathcal{K} \), such that \( \lim_{n \to \infty} \mathcal{V}_e_n = \overline{\mathcal{V}}_\mathcal{K} \) uniformly in \([0, T] \times \mathcal{K} \). Thanks to (5.1), we have pointwise convergence \( \lim_{\epsilon \to 0} \mathcal{V}_e(t_0, p_0) = \mathcal{V}(t_0, p_0) \) for any \( t_0 \in [0, T] \) and \( p_0 \in \text{Int}(\mathcal{S}_d) \). Hence any subsequence \((\mathcal{V}_{e_n})_{n \geq 0}\) converges uniformly to the same limit which is the value function, and thus we obtain \( \lim_{\epsilon \to 0} \mathcal{V}_e = \mathcal{V} \) uniformly in \([0, T] \times \mathcal{K} \). \( \square \)

We are now in position to prove a preliminary version Proposition 2.5, but restricted to initial conditions in a compact subset of \([0, T] \times \text{Int}(\mathcal{S}_d) \):

**Proposition 5.10.** We can find \( \bar{\kappa}_0 \geq 0 \), only depending on \( \kappa_2 \) and \( d \), such that, for any \( \kappa_0 \geq \bar{\kappa}_0 \) and any compact subset \( \mathcal{K} \) included in \( \text{Int}(\mathcal{S}_d) \), for the same two functions \( \theta(\delta, \epsilon) \) and \( \delta(\epsilon) \) as in the statement of Proposition 5.8 (which only depend on \( \kappa_\theta \), \( \kappa_\delta \), \( \mathcal{K} \), \( ||f||_{\infty}, ||g||_{\infty}, T \) and \( d \)), the additional cost induced by (4.6) tends to 0 with \( \epsilon \):

\[
\lim_{\epsilon \to 0} \Xi_{\theta(\delta(\epsilon), \epsilon), \delta(\epsilon), \epsilon}(t_0, p_0, q_0) = 0, \tag{5.30}
\]

uniformly in \( t_0 \in [0, T], p_0 \in \mathcal{K} \) and \( q_0 \in \mathcal{S}_d \), where
form integrability, namely it is enough to show that with exponents 8, 8 and 4/3, it suffices to prove that grand tends to 0 in probability as and then, for implicitly the same value of 5.6 are applied with 2 lim

...ation for a constant C in independent of, also as in the statement of Proposition 5.8. For simplicity, we let \( \phi_{\varepsilon} := \varphi_{\varepsilon}(\delta(\varepsilon), \varepsilon) \) and \( V_{\varepsilon} := V_{\varepsilon}(\delta(\varepsilon), \varepsilon) \) for \( \varepsilon \in (0, \varepsilon_0] \). Similarly, we use the abbreviated notations \( p^{\varepsilon} \) and \( q^{\varepsilon} \) for the two processes appearing in (5.30), the underlying initial condition \((t_0, p_0, q_0)\) being fixed in \([0, T] \times K \times S_d\) (which is licit provided we prove that the convergences below hold uniformly with respect to \((t_0, p_0, q_0)\)). To prove the claim, we have to show (see (4.6)) that (uniformly with respect to the initial condition)

\[
\lim_{\varepsilon \to 0} \mathbb{E} \left[ \left| \int_{t_0}^{T} \sum_{k} \sum_{j} q_{t}^{k,j} \left( p_{t}^{k,j} \right)^{2} (V_{\varepsilon}(t, p_{t}^{k,j}) - V_{\varepsilon}(t, p_{t}^{k,j})) dt \right|^{3/2} \right] \leq C,
\]

for a constant \( C \) independent of \( \varepsilon \) and of \((t_0, p_0, q_0)\). By (5.22) and by Hölder inequality with exponents 8, 8 and 4/3, it suffices to prove that

\[
\mathbb{E} \left[ \left| \int_{t_0}^{T} \sum_{k} \sum_{j} q_{t}^{k,j} \left( p_{t}^{k,j} \right)^{2} (V_{\varepsilon}(t, p_{t}^{k,j}) - V_{\varepsilon}(t, p_{t}^{k,j})) dt \right|^{12} \right]^{1/8} \leq C.
\]

The first term in the left-hand side is easily bounded by means of Lemma 5.7, recalling that \( \lambda \) in the statement of Proposition 5.8 is required to satisfy \( \lambda \geq \lambda(12) \). As for the second one, it follows from (5.17).

To prove (5.32), we have to show that the expectation is bounded (since there is the additional factor \( \varepsilon \) in front of it), but this easily follows from Holder’s inequality, with \( 1 = 1/3 + 1/6 + 1/2 \), together with (5.19), (5.20), (5.23) and (5.24).

We now address the convergence of the master equation. To do so, we denote by \( U_{0, \delta, \varepsilon} \) the solution to the viscous master equation (2.20) (as provided by Theorem 4.1),

\[
\Xi_{0, \delta, \varepsilon}(t_0, p_0, q_0) = \mathbb{E} \left[ \left| \int_{t_0}^{T} \sum_{k} \sum_{j} q_{t}^{k,j} \left( p_{t}^{k,j} \right)^{2} (V_{\varepsilon}(t, p_{t}^{k,j}) - V_{\varepsilon}(t, p_{t}^{k,j})) dt \right|^{12} \right]^{1/8} \]

Proof. Throughout the proof, we consider \( \bar{\kappa}_0 \) as in the statement of Proposition 5.8 (and implicitly the same value of \( \lambda \) as in its proof, see (5.25) and the discussion below (5.25)) and then, for \( \kappa_0 \geq \bar{\kappa}_0 \) and for a compact subset \( K \) included in Int(\( S_d \)), we consider \( \bar{\theta}_0, \bar{\theta}(\delta, \varepsilon) \) and \( \bar{\delta}(\varepsilon) \), also as in the statement of Proposition 5.8. For simplicity, we let \( \varphi_{\varepsilon} := \varphi_{\theta}(\delta(\varepsilon), \varepsilon) \) and \( V_{\varepsilon} := V_{\theta}(\delta(\varepsilon), \varepsilon) \) for \( \varepsilon \in (0, \varepsilon_0] \). Similarly, we use the abbreviated notations \( p^{\varepsilon} \) and \( q^{\varepsilon} \) for the two processes appearing in (5.30), the underlying initial condition \((t_0, p_0, q_0)\) being fixed in \([0, T] \times K \times S_d\) (which is licit provided we prove that the convergences below hold uniformly with respect to \((t_0, p_0, q_0)\)). To prove the claim, we have to show (see (4.6)) that (uniformly with respect to the initial condition)

\[
\lim_{\varepsilon \to 0} \mathbb{E} \left[ \left| \int_{t_0}^{T} \sum_{k} \sum_{j} q_{t}^{k,j} \left( p_{t}^{k,j} \right)^{2} (V_{\varepsilon}(t, p_{t}^{k,j}) - V_{\varepsilon}(t, p_{t}^{k,j})) dt \right|^{12} \right]^{1/8} \leq C,
\]

where \( W_{\varepsilon} = W_{\theta}(\delta(\varepsilon), \varepsilon) \) and similarly for \( Y_{\varepsilon} \).

We begin by proving (5.31). We know from (5.18) (all the results from Proposition 5.6 are applied with 2\( \lambda d \)), see again the discussion below (5.25) that \( \lim_{\varepsilon \to 0} \mathbb{P}(\inf_{t_0 \leq t \leq T} \min_{k \in [d]} |p_{t}^{k} | \leq \eta) = 0 \) for \( \eta > 0 \) small enough (independently of \( \varepsilon \)), the convergence being uniform with respect to the initial point \((t_0, p_0) \in [0, T] \times K \). Together with the fact that the support of \( \varphi_{\varepsilon} \) shrinks with \( \varepsilon \), we deduce that the integrand tends to 0 in probability as \( \varepsilon \to 0 \). Hence, to obtain (5.31), we have to prove uniform integrability, namely it is enough to show that...
with \( \varphi = \varphi_{\theta, \delta, \epsilon} \) therein. We recall (see Proposition 5.2, (viii)) that there exist a unique optimal control \( \mathbf{a} \) and optimal trajectory \( \mathbf{p} \) for the inviscid MFCP starting at points \( (t_0, p_0) \in [0, T] \times \text{Int}(S_d) \) where the value function \( V \) is differentiable. For such points, let, as in Section 2 (see Theorem 2.6, part II), \( U^j(t_0, p_0) := \inf_q J(x; p) \) with \( q \) being initialized at time \( t_0 \) from \( (q_t^j = \delta_{i,j})_{j \in [d]} \).

**Proposition 5.11.** On top of the assumptions quoted in the beginning of the subsection, assume that \( E \) is in \( C^{1,1}(S_d) \). Then, we can find \( \bar{\kappa}_0 \geq 0 \), only depending on \( \kappa_2 \) and \( d \), such that, for any \( \kappa_0 \geq \bar{\kappa}_0 \) and any compact subset \( K \) included in \( \text{Int}(S_d) \), for the same two functions \( \hat{\theta}(\delta, \epsilon) \) and \( \hat{\delta}(\epsilon) \) as in the statement of Proposition 5.8 (which only depend on \( \kappa_0, \kappa_2, K, ||f||_\infty, ||g||_\infty, T \) and \( d \)), and for any \( (t_0, p_0) \in [0, T] \times K \) at which \( V \) is differentiable,

\[
\lim_{\epsilon \to 0} D\!V_{\hat{\theta}(\delta(\epsilon), \epsilon), \hat{\delta}(\epsilon), \epsilon}(t_0, p_0) = D\!V(t_0, p_0), \tag{5.33}
\]

\[
\lim_{\epsilon \to 0} U_{\hat{\theta}(\delta(\epsilon), \epsilon), \hat{\delta}(\epsilon), \epsilon}(t_0, p_0) = U(t_0, p_0), \tag{5.34}
\]

Moreover, these convergence hold in \([E_{\text{loc}}([0, T] \times \text{Int}(S_d))]^d\), for any \( r \geq 1 \), where \( \text{Int}(S_d) \) is equipped with the \((d - 1)\)-dimensional Lebesgue measure.

**Proof.**

**Step 1.** As in the previous proof, we consider \( \bar{\kappa}_0 \) as in the statement of Proposition 5.8 (and implicitly the same value of \( \lambda \) as in its proof) and then, for \( \kappa_0 \geq \bar{\kappa}_0 \) and for a compact subset \( K \) included in \( \text{Int}(S_d) \), we consider \( \bar{\kappa}_0, \hat{\theta}(\delta, \epsilon) \) and \( \hat{\delta}(\epsilon) \), also as in the statement of Proposition 5.8. We also use the same notations \( \varphi_{\theta, \epsilon}, \psi_{\epsilon}, p^\epsilon \) as in the previous proof, and similarly we write \( x^\epsilon \) for the corresponding optimal control and \( u_\epsilon \) for the solution of the (viscous) master equation. Here, the initial condition of \( p^\epsilon \) is implicitly understood as a point \( (t_0, p_0) \in [0, T] \times K \) at which \( V \) is differentiable. By Corollary 5.3 (writing \((p, a)\) for \((p_{(t_0, p_0)}, a_{(t_0, p_0)})\) therein), the convergence of \( (p^\epsilon, a^\epsilon)_{\epsilon \in (0, \delta_0]} \) to \( (p, a) \) holds in probability (for the same topology as in the statement of Theorem 5.1). In fact, by combining (5.6) and (5.7) (with \( \beta = a \) therein), we have (indices below are in \([d]\))

\[
\lim_{\epsilon \to 0} \mathbb{E} \int_{t_0}^{T} \sum_{i \neq j} p^\epsilon_{t,i} |x^\epsilon_{t,i,j} - x_{t,i,j}|^2 dt = \int_{t_0}^{T} \sum_{i \neq j} p^\epsilon_{t,i} |x^\epsilon_{t,i,j} - x_{t,i,j}|^2 dt = 0. \tag{5.35}
\]

As the limit process \( p \) does not touch the boundary of the simplex, the latter shows that \( x^\epsilon \) converges to \( x \) in probability but for the strong (instead of weak) topology on \( \mathcal{E} \). We make use of this property later on in the proof.

In order to prove (5.34), it is worth recalling that \( U^j(t_0, p_0) \) is the value function of the cost functional \( \mathcal{J}^{p, \epsilon}(\cdot, p^\epsilon) \) when the state trajectory \( q^\epsilon \) in (2.5) is initialized from \( q^\epsilon_{t_0} = \delta_{i,k}, k \in [d] \), and similarly \( U^j(t_0, p_0) \) is the value function of the cost functional \( \mathcal{J}(\cdot, p) \) when the state trajectory in (2.2) is also initialized from \( q^j_{t_0} = \delta_{i,k}, k \in [d] \). Recalling (2.14), we have (indices in the sum belonging to \([d]\))
\[ \tilde{J}^\varepsilon_{t_0}(\alpha^i; p^i) = \mathbb{E}\left[ \int_{t_0}^T \sum_k \tilde{q}^\varepsilon_{t_0}^i (L^k(x^i_t) + f^k(p^i_t)) \, dt + \sum_k \tilde{q}^\varepsilon_T^k g^k(p_T) \right], \]

(5.36)

where \((\{q^\varepsilon_{t_0}^k\}_{k \in [d]})_{t_0 \leq t \leq T}\) solves

\[ \frac{dq^\varepsilon_{t_0}^i}{dt} = \sum_{j \neq k} q^\varepsilon_{t_0}^i (\varphi^i_j(p^i_{t_0}^k) + z^\varepsilon_{t_0}^i k^j) - q^\varepsilon_{t_0}^i (\varphi^i_i(p^i_{t_0}^k) + z^\varepsilon_{t_0}^i k^j) \, dt \]

\[ + \frac{\varepsilon}{\sqrt{2}} \sum_{j \neq k} q^\varepsilon_{t_0}^i \left( \frac{p^i_{t_0}^j}{p^i_{t_0}^k} \right)^q (dB^j_{t_0} - dB^k_{t_0}), \]

with \(q^\varepsilon_{t_0}^i = \delta_{i,k}.\)

By (5.18), there exists \(\eta > 0\) such that \(\lim_{\varepsilon \to 0} \mathbb{P}(\inf_{t_0 \leq t \leq T} \min_{k \in [d]} |q^\varepsilon_{t_0}^k| < \eta) = 0.\) This suffices to kill asymptotically the terms \(\varphi^i_j\) in the drift right above on the model of the proof of Theorem 5.1. As for the martingale part, we may invoke Lemma 5.7 with \(k = 4\) (recall that \(\kappa_0\) is chosen in such a way that Lemma 5.7 applies with \(\ell = 12\)) to show that its supremum norm converges to 0 in probability. Altogether with (5.35), we easily deduce that

\[ q^\varepsilon_{t_0}^i = \delta_{k,i} + \int_{t_0}^T \sum_{j \neq k} (q^\varepsilon_{t_0}^i z^\varepsilon_{t_0}^j - q^\varepsilon_{t_0}^i z^\varepsilon_{t_0}^k) \, ds + r^\varepsilon_{t_0}^i, \quad t \in [t_0, T], \]

where \(\sup_{t_0 \leq t \leq T} |r^\varepsilon_{t_0}^i|\) tends to 0 in probability with \(\varepsilon,\) which prompts us to consider the differential equation

\[ q^\varepsilon_{t_0}^i = \sum_{j \neq k} (q^\varepsilon_{t_0}^i z^\varepsilon_{t_0}^j - q^\varepsilon_{t_0}^i z^\varepsilon_{t_0}^k), \quad q^\varepsilon_{t_0}^i = \delta_{i,k}. \]

Forming the differences \((\{q^\varepsilon_{t_0} - q^k\}_{k \in [d]})_{t_0 \leq t \leq T},\) we easily deduce that \(\sup_{t_0 \leq t \leq T} |q^\varepsilon_t - q_t|\) tends to 0 in probability. By Lemma 5.7 again, the convergence holds in \(L^2\) (recall that \(\lambda \geq \tilde{\lambda}(12)\)). Using (2.14) and (5.35) together with the fact that all the off-diagonal controls are bounded by \(M,\) we get

\[ \lim_{\varepsilon \to 0} \mathbb{E}\left[ \int_{t_0}^T \sum_k \tilde{q}^\varepsilon_{t_0}^i (L^k(x^i_t) + f^k(p^i_t)) \, dt + \sum_k \tilde{q}^\varepsilon_T^k g^k(p_T) \right] \]

\[ = \int_{t_0}^T \sum_k q^\varepsilon_{t_0}^i (L^k(x^i_t) + f^k(p^i_t)) \, dt + \sum_k q^\varepsilon_T^k g^k(p_T) = J(\alpha; p) = U^i(t_0, p_0), \]

which, together with (5.30) that holds for any initial condition, gives (5.34).

**Step 2.** To prove (5.33), we note that, by item (viii) in Proposition 5.2 again \(x^i_{t_0} = (d_0, \nu(t, p_t) - d_0, \nu(t, p_t))\) for any \(t_0 \leq t \leq T\) (recalling that \(\nu\) is differentiable at any \((t, p_t)\)), and that, by item (ix) in Proposition 5.2, the backward equation in (5.10) (in the unknown \(z = (z_t)_{t_0 \leq t \leq T}\)) represents the gradient of \(\nu.\) So in particular, at the initial time, we have \(z^i_{t_0} = \partial_{x^i} \nu(t_0, x_0)\) with \(x_0 = (p^{i-1}_0, ..., p^i_{t_0}).\) But \(z^i_{t_0}\) is also equal to \(U^i(t_0, p_0) - U^d(t_0, p_0),\) which is exactly (5.11). Thus we have \(\partial_{x^i} \nu(t_0, x_0) = U^i(t_0, p_0) - U^d(t_0, p_0).\) Importantly, we have a similar identity when \(\varepsilon \in (0, \varepsilon_0],\) which is provided by (2.22) proved in Theorem 4.1. Therefore (5.33) now follows from (5.34).

**Step 3.** The last claim follows from uniform boundedness of \(DV^\varepsilon\) and \(U^\varepsilon.\) The former is given by (5.21), together with the fact that \(\nu\) is almost everywhere (for the
(d − 1)-dimensional Lebesgue measure) differentiable, while the latter follows easily from the definition (5.36) together with the bounds in Proposition 5.8.

At this stage of the proof, the reader must understand that Propositions 5.10 and 5.11 do not provide complete proofs of Proposition 2.5 and Theorem 2.6. The reason is that the functions θ and δ therein depend on the underlying compact set K. In words, we should write θ_K and δ_K. Now, we would like to choose K = K_ε depending on ε such that, letting

\[
\begin{align*}
\tilde V_\varepsilon &:= \mathcal{V}_{θ_K(δ_{K_ε}(ε), δ_{K_ε}(ε), ε), \varepsilon}, \\
\tilde U_\varepsilon &:= U_{θ_K(δ_{K_ε}(ε), δ_{K_ε}(ε), ε), \varepsilon}, \\
\tilde ξ_\varepsilon &:= \Xi_{θ_K(δ_{K_ε}(ε), δ_{K_ε}(ε), ε), \varepsilon},
\end{align*}
\]

(5.37)

(5.27) and (5.30) hold locally uniformly on [0, T] × Int(S_d), and (5.33) and (5.34) hold almost everywhere. We mostly argue by an inversion argument very similar to the proof of Proposition 5.6 (which is given below).

Theorem 5.12. Under the assumptions quoted in the beginning of the subsection, we can find \( \bar{K}_0 \geq 0 \), only depending on K_2 and d, such that, for any \( \bar{K}_0 \geq K_0 \) and any ε \( \in (0, \varepsilon_0 = \sqrt{K_2/\bar{K}_0}] \), there exist a compact subset \( K_\varepsilon \) included in Int(S_d), with \( K_\varepsilon \supset K_\varepsilon \) if \( \varepsilon < \varepsilon' \) and \( \cup_{\varepsilon \in (0, \varepsilon_0]} K_\varepsilon = \text{Int}(S_d) \), together with functions \( \hat{θ}_{K_\varepsilon}(δ, ε) \) and \( \hat{δ}_{K_\varepsilon}(ε) \) as in the statement of Proposition 5.8 such that, using the same notations as in (5.37) (and in particular letting \( ε = 0 \)),

\[
\begin{align*}
\lim_{\varepsilon \to 0} \tilde V_\varepsilon &= \mathcal{V} \text{ locally uniformly in } [0, T] \times \text{Int}(S_d), \\
\lim_{\varepsilon \to 0} \tilde ξ_\varepsilon &= 0 \text{ locally uniformly in } [0, T] \times \text{Int}(S_d) \times S_d.
\end{align*}
\]

If, in addition \( F \) is in \( C^{1,1}(S_d) \), then

\[
\begin{align*}
\lim_{\varepsilon \to 0} \nabla \tilde V_\varepsilon &= D\mathcal{V} \text{ a.e. on } [0, T] \times \text{Int}(S_d) \text{ and in } L^1_{\text{loc}}([0, T] \times \text{Int}(S_d))^d, \\
\lim_{\varepsilon \to 0} \nabla \tilde U_\varepsilon &= U \text{ a.e. on } [0, T] \times \text{Int}(S_d) \text{ and in } L^1_{\text{loc}}([0, T] \times \text{Int}(S_d))^d,
\end{align*}
\]

where Int(S_d) is equipped with the (d − 1)-dimensional Lebesgue measure.

Proof. Throughout the proof, we consider \( \bar{K}_0 \) as in the statement of Proposition 5.8 and then, for \( \bar{K}_0 \geq K_0 \) and for a compact subset \( K \) included in Int(S_d), we consider \( \hat{θ}_K(δ, ε) \) and \( \hat{δ}_K(ε) \), also as in the statement of Proposition 5.8. Without any loss of generality, the functions \( \hat{θ}_K(δ, ε) \) and \( \hat{δ}_K(ε) \) may be assumed to be defined for all \( δ \in (0, 1/2] \) and \( ε \in (0, \varepsilon_0 = \sqrt{K_2/\bar{K}_0}] \); in fact, only the limits in (0, 0) and 0 matter for our purpose. For any \( n \geq 1 \), let \( K_n \) be the compact set \( \{ x \in S_d : \text{dist}(x, \partial S_d) \geq 1/n \} \). Below, we restrict ourselves to the set \( \mathbb{N}_θ \) of large enough integers \( n \) such that \( K_n \neq \emptyset \). Obviously, \( \mathbb{N}_θ \) is of the form \( \mathbb{N}_θ = \{ n_θ, n_θ + 1, \ldots \} \) for some integer \( n_θ \geq 1 \). For \( n \in \mathbb{N}_θ \), we let \( \mathcal{V}_{n, ε}(t, p) := V_{θ_K(δ_{K_n}(ε), δ_{K_n}(ε), ε)}(t, p) \) for \( t \in [0, T] \) and \( p \in S_d \); similarly, we introduce \( \mathcal{U}_{n, ε} \) and \( \Xi_{n, ε} \) (the latter being defined on \( [0, T] \times S_d \times S_d \)). By Corollary 5.9, for any fixed \( n \in \mathbb{N}_θ \), \( \lim_{ε \to 0} \mathcal{V}_{n, ε} = \mathcal{V} \), uniformly on \( [0, T] \times K_n \), and, by Proposition 5.10, \( \lim_{ε \to 0} \Xi_{n, ε} = 0 \), uniformly on \( [0, T] \times K_n \times S_d \). Further, by Proposition 5.11, \( \lim_{ε \to 0} D\mathcal{V}_{n, ε}(t, p) = D\mathcal{V}(t, p) \) for a.e. \( (t, p) \in [0, T] \times K_n \), and \( \lim_{ε \to 0} D\mathcal{V}_{n, ε} = D\mathcal{V} \) in \( L^1([0, T] \times K_n))^d \), and
similarly with $U_{n,e}$ and $U$. Applying Egoroff’s theorem for any $n \in \mathbb{N}$, there exists $E_n \subset [0,T] \times \mathcal{K}_n$ with $(d$-dimensional—since the simplex is equipped with the $(d - 1)$ Lebesgue measure) Lebesgue measure $|E_n| \leq 2^{-n}$ such that $\lim_{n \to 0} D V_{n,e} = D V$ and $\lim_{n \to 0} U_{n,e} = U$ uniformly on $[0,T] \times \mathcal{K}_n \setminus E_n$. Therefore, for any $n \in \mathbb{N}$, there exists $e_n \in (0,e_0]$ such that, for any $e \leq e_n$,

$$\sup_{(t,p) \in [0,T] \times \mathcal{K}_n} |(V_{n,e} - V)(t,p)| + \sup_{(t,p,q) \in [0,T] \times \mathcal{K}_n \times \mathcal{E}} |\Xi_{n,e}(t,p,q)| + \sup_{(t,p) \in [0,T] \times \mathcal{K}_n \setminus E_n} |(D V_{n,e} - D V)(t,p)| + |(U_{n,e} - U)(t,p)|$$

$$+ \int_0^T \int_{\mathcal{K}_n} |D V_{n,e} - D V|(t,p)|d\varrho(p)dt \leq \frac{1}{n}, \quad (5.42)$$

where $\varrho$ is the image of the $(d - 1)$-dimensional Lebesgue measure by the map $(x_1, ..., x_{d-1}) \mapsto (x_1, ..., x_{d-1}, x^{-d})$. Moreover, we can assume that $e_{n+1} < e_n \leq 1/n$, so that $\lim_{n \to \infty} e_n = 0$.

We now define $n$, and thus $\mathcal{K}_n$, in terms of $e$: for any $e \in (0,e_0)$, let $n_e$ be the unique $n \in \mathbb{N}$ such that $e_{n+1} < e \leq e_n$. Obviously, the function $(0,n_0) \ni e \mapsto n_e$ is decreasing and the supremum, say $N := \sup_{e \in (0,n_0)} n_e$ cannot be finite as otherwise we would have $0 < e_{N+1} \leq e_{n_0} < e$ for any $e \in (0,e_0)$, which is a contradiction. Hence, choosing $n = n_e$ in the right-hand side of (5.42), letting with a slight abuse of notation $\mathcal{K}_e := \mathcal{K}_{n_e}$ (the definition of $\mathcal{K}_e$, for $e \in [e_{n_0}, e_0]$, does not really matter) and then using the same notation as in (5.37), we get, for any compact subset $\mathcal{K}$ included in $\text{Int}(S_d)$,

$$\sup_{(t,p) \in [0,T] \times \mathcal{K}} |\hat{V}_e(t,p) - V(t,p)| \leq \sup_{(t,p) \in [0,T] \times \mathcal{K}_e} |V_{n_e,e}(t,p) - V(t,p)| \leq \frac{1}{n_e}, \quad (5.38)$$

for $e$ small enough, which gives (5.39). Obviously, the proof of (5.39) is similar, and in fact the same argument applies for proving the $L^1$ convergence in (5.40) and (5.41). To prove the a.e. convergence in (5.40) (and similarly in (5.41)), consider again a compact set $\mathcal{K} \subset \text{Int}(S_d)$ and $e$ small enough such that $\mathcal{K} \subset \mathcal{K}_e$. From (5.42) again, we get that

$$|D \hat{V}_e(t,p) - D V(t,p)| \leq \frac{1}{n_e}$$

if $(t,p) \in [0,T] \times \mathcal{K} \setminus E_{n_e}$. Therefore, the set of points $(t,p) \in [0,T] \times \text{Int}(S_d)$ such that $D \hat{V}_e(t,p)$ does not converge to $D V(t,p)$, as $e \to 0$, is included in the set of points $(t,p)$ such that $(t,p) \in E_{n_e}$ for infinitely many $n_e$. The latter is nothing but $\limsup_{n \geq n_0} E_n$, which has Lebesgue measure 0 by Borel-Cantelli lemma, since $\sum_{n=1}^{\infty} |E_n| \leq \sum_{n=1}^{\infty} 2^{-n} < \infty$. Hence $\lim_{e \to 0} D \hat{V}_e(t,p) = D V(t,p)$ for a.e. $(t,p) \in [0,T] \times \text{Int}(S_d)$, from which the a.e. convergence in (5.40) follows. The a.e. convergence in (5.41) is treated in the same way.

\hfill \Box

### 5.4. Proofs of auxiliary exponential integrability properties

**Proof of Proposition 5.5.** We prove (5.14) and (5.15) in a single row, mostly following [43, Proposition 2.3]. Fix $i \in [d]$ and, for simplicity, write $p^i$ for $p^{i,0,\delta,\varepsilon}$ and take $t_0 = 0$. As in the second step of the proof of [43, Proposition 2.2], we write the equation for $p^i$ in the form
\[ dp_i^I = \sum_{j \in [d]} \left[ p_i^I (\varphi_{0, \delta, \varepsilon}(p_i^I) + \chi_{j}^I) - p_i^I (\varphi_{0, \delta, \varepsilon}(p_i^I) + \chi_{j}^I) \right] dt + \varepsilon \sqrt{p_i^I(1 - p_i^I)} d\tilde{W}_i^I, \]  

(5.43)

for \( t \in [0, T] \) and for \( \tilde{W}_i^I = (\tilde{W}_i^I)_{0 \leq t \leq T} \) a 1d-Brownian motion. Then, Itô’s formula yields (the left-hand side below is well-defined since \( p_i^I \) does not vanish)

\[
\begin{align*}
\frac{\lambda}{\varepsilon^2} \ln p_i^I &= \sum_{j \in [d]} \left[ \frac{\lambda}{\varepsilon^2} p_i^I (\varphi_{0, \delta, \varepsilon}(p_i^I) + \chi_{j}^I) - \frac{\lambda}{\varepsilon^2} (\varphi_{0, \delta, \varepsilon}(p_i^I) + \chi_{j}^I) \right] dt - \frac{\lambda}{\varepsilon^2} \frac{1 - p_i^I}{p_i^I} dt \\
&+ \frac{\lambda}{\varepsilon} \sqrt{\frac{1 - p_i^I}{p_i^I}} d\tilde{W}_i^I, \quad t \in [0, T].
\end{align*}
\]

(5.44)

We now subtract the quantity \( \lambda^2 (1 - p_i^I)/(2e^2 p_i^I) \) to the drift of (5.44) and then get the following lower bound (using the definition of \( \varphi \) in (2.17) together with the fact that \( 0 \leq \chi_{j}^I \leq M \) if \( j \neq i \))

\[
\begin{align*}
\sum_{j \in [d]} &\left[ \frac{\lambda^2}{e^2} p_i^I (\varphi_{0, \delta, \varepsilon}(p_i^I) + \chi_{j}^I) - \frac{\lambda}{e^2} (\varphi_{0, \delta, \varepsilon}(p_i^I) + \chi_{j}^I) \right] - \frac{\lambda}{e^2} \frac{1 - p_i^I}{p_i^I} \\
&\geq \frac{\lambda}{e^2} \kappa_0 1_{[0, \delta]}(p_i^I) - \frac{\lambda}{e^2} \sum_{j \in [d]} (\kappa_0 1_{[0, \delta]}(p_i^I) + \kappa_0 1_{[0, \delta]}(p_i^I) + M) - \frac{\lambda}{2e^2} (e^2 + \lambda) \frac{1}{p_i^I} \\
&\geq \frac{\lambda}{e^2} \left( \kappa_0 - \frac{e^2 + \lambda}{2} \right) \frac{1}{p_i^I} 1_{[0, \delta]}(p_i^I) - \frac{\lambda (1 + \lambda)}{2e^2} - \frac{\lambda d}{e^2} (\kappa_0 + \kappa_0 + M).
\end{align*}
\]

(5.45)

Hence, integrating (5.44) from 0 to \( T \), adding and subtracting the compensator \( e^{-2 \lambda^2} \int_0^T (1 - p_i^I)/(2p_i^I) dt \) and then taking the exponential, we get

\[
\begin{align*}
(p_i^I)^{\lambda/2} &\exp \left( -\frac{\lambda}{\varepsilon} \int_0^T \sqrt{\frac{1 - p_i^I}{p_i^I}} d\tilde{W}_i^I - \frac{\lambda^2}{2e^2} \int_0^T \frac{1 - p_i^I}{p_i^I} dt \right) \\
&\geq (p_0^I)^{\lambda/2} \exp \left( \frac{\lambda}{e^2} \left( \kappa_0 - \frac{e^2 + \lambda}{2} \right) \int_0^T \frac{1}{p_i^I} 1_{[0, \delta]}(p_i^I) dt \right) e^{-TC(\delta, \varepsilon, \lambda)}.
\end{align*}
\]

Since the left-hand side has expectation less than 1, claim (5.14) follows. In order to get (5.15), it suffices to replace \( \lambda \) by \( e^2 \lambda \), to observe that the indicator function \( 1_{[0, \delta]} \) in (5.45) has zero value if \( \inf_{0 \leq t \leq T} p_i^I > 2 \), and to integrate from 0 to the first time when \( p_i^I \) becomes lower than 2\( \theta \).

**Proof of Proposition 5.6.** Throughout the proof, the initial condition \((t_0, p_0) \in [0, T] \times \mathcal{K}\) is implicitly understood in the notation \( p_0^{\theta, \delta, \varepsilon} \). Also, we fix the state \( i \in [d] \) and the value of \( \lambda \geq 1 \) and we make explicit the dependence of the various constants upon the two parameters \( \delta \) and \( \varepsilon \). However, we do not indicate the fact that the constants may depend on \( \mathcal{K} \). Below, we use the same notation \( \tilde{W}^I \) as in (5.43).

**Step 1. a.** We first claim that, for any \( \eta > 0 \), there exists \( a_\eta(\delta, \varepsilon) \in (0, 1) \), such that, for all \((t_0, p_0) \in [0, T] \times \mathcal{K} \) and \( \theta \in [0, \delta/2] \),
\[
\mathbb{P}\left( \inf_{t_0 \leq t \leq T} p_t^{i,0,\delta,\varepsilon} > a_\eta(\delta, \varepsilon) \right) \geq 1 - \eta. \tag{5.46} \]

The proof is a consequence of (5.44) and of (5.14) (with \( \lambda = \varepsilon^2 \) and \( \kappa_0 > \varepsilon^2 \)). Indeed, the former, together with Doob’s maximal inequality, yield

\[
\mathbb{P}\left( \sup_{t_0 \leq t \leq T} \left[ - \ln (p_t^{i,0,\delta,\varepsilon}) \right] \right) \geq - \ln (a_\eta) \leq \frac{c}{\ln (a_\eta)} \mathbb{E} \left[ \int_{t_0}^{T} \left( 1 + \frac{1}{p_t^{i,0,\delta,\varepsilon}} \right) dt. \right]
\]

for some \( c \) only depending on \( \varepsilon \) and \( \delta \), \( M \), \( \kappa_0 \) and \( \kappa_2 \). Then, (5.14) gives a bound (depending on \( \delta \) and \( \varepsilon \)) for the above right-hand side.\textsuperscript{b} Our second step is to prove that, provided that \( \kappa_0 \) satisfies

\[
\kappa_0 - 1 - \frac{\lambda}{2} \geq 4\kappa_2, \tag{5.47} \]

there exists \( \theta_1(\delta, \varepsilon, \lambda) > 0 \) such that, for any \( \theta \leq \theta_1(\delta, \varepsilon) \) and any \((t_0, p_0) \in [0, T] \times \mathcal{K}\),

\[
\mathbb{E} \left[ \exp \left\{ - \lambda \int_{t_0}^{T} - \varphi'_\theta(p_t^{i,0,\delta,\varepsilon}) dt \right\} \right] \leq 2, \tag{5.48} \]

where we have let for convenience \( \varphi'_\theta(r) := -(2\kappa_\varepsilon / \theta) 1_{[0,2\theta]}(r) \).

Obviously, \( \int_{t_0}^{T} \varphi'_\theta(p_t^{i,0,\delta,\varepsilon}) dt \) converges to 0 in probability as \( \theta \) tends to 0, uniformly in \((t_0, p_0) \in [0, T] \times \mathcal{K}\) (the other two parameters \( \delta \) and \( \varepsilon \) being kept fixed), as the indicator function appearing in the definition of \( \varphi'_\theta \) vanishes for \( \theta \) small enough (it hence suffices to choose \( 2\theta \leq a_\eta(\delta, \varepsilon) \) for \( a_\eta(\delta, \varepsilon) \) as in (5.46), for a given \( \eta > 0 \) as small as needed). In order to prove (5.48), we then notice that

\[
- \varphi'_\theta(p_t^{i,0,\delta,\varepsilon}) = \frac{2\kappa_\varepsilon}{\theta} 1_{[0,2\theta]}(p_t^{i,0,\delta,\varepsilon}) \leq \frac{2\kappa_\varepsilon}{\theta} \frac{2\theta}{p_t^{i,0,\delta,\varepsilon}} 1_{[0,2\theta]}(p_t^{i,0,\delta,\varepsilon}) \leq \frac{4\kappa_\varepsilon}{p_t^{i,0,\delta,\varepsilon}} 1_{[0,\delta]}(p_t^{i,0,\delta,\varepsilon}). \tag{5.49} \]

Recall now that \( \kappa_\varepsilon = \varepsilon^{-2}\kappa_2 \). Hence, if we choose another real \( \lambda' \) that satisfies \( \kappa_0 \geq 4\kappa_2 + (1 + \lambda')/2 \) (take for instance that \( \lambda' = \lambda + 1 \) and recall \( \kappa_0 \geq 4\kappa_2 + 1 + \lambda/2 \)), then (5.49) and (5.14) yield

\[
\mathbb{E} \left[ \exp \left\{ - \lambda' \int_{t_0}^{T} \varphi'_\theta(p_t^{i,0,\delta,\varepsilon}) dt \right\} \right] \leq \mathbb{E} \left[ \exp \left\{ \frac{4\kappa_2 \lambda'}{\varepsilon^2} \int_{t_0}^{T} \frac{1}{p_t^{i,0,\delta,\varepsilon}} 1_{[0,\delta]}(p_t^{i,0,\delta,\varepsilon}) dt \right\} \right] \leq C_1(\delta, \varepsilon, \lambda'),
\]

where \( C_1(\delta, \varepsilon, \lambda') \) is a constant independent of \( \theta \) and depending on \((t_0, p_0) \) through \( \mathcal{K} \) only.\textsuperscript{5} Combining the above upper bound with the fact that \( \int_{t_0}^{T} \varphi'_\theta(p_t^{i,0,\delta,\varepsilon}) dt \) tends to 0 in probability (uniformly in \((t_0, p_0) \in [0, T] \times \mathcal{K}\)), we easily derive (5.48).

\textit{Step 2.} The goal of this step is to address a similar result to (5.48) but with \( \varphi'_\theta \) replaced by \( \varphi'_\delta \), defined as \( \varphi'_\delta(r) := -(2\kappa_\eta / \delta) 1_{[0,2\delta]}(r) \).

\textsuperscript{5}We do not keep track of the parameters \( \kappa_\eta, \kappa_2, \mathcal{K}, M, T \) and \( d \) in the constants.
a. The first step is to notice that
\[
-\varphi_\delta(p_t^i,\theta,\delta,\epsilon) \leq 2 \frac{\kappa_0 + 2 \delta}{p_t^i} 1_{[0,\Delta]}(p_t^i,\theta,\delta,\epsilon) \leq \frac{4\kappa_0}{p_t^i} 1_{[0,\Delta]}(p_t^i,\theta,\delta,\epsilon) + \frac{4\kappa_0}{p_t^i} 1_{[\Delta,2\Delta]}(p_t^i,\theta,\delta,\epsilon).
\]
(5.50)

b. We address the first term in the right-hand side of (5.50). To do so, we need a finer lower bound on the coordinates of \( p_t^i,\theta,\delta,\epsilon \) and hence we must revisit the proof of Proposition 5.5. As in the first step, we fix some \( \eta > 0 \) and, for \( a_\eta := a_\eta(\delta,\epsilon) \) as therein, we consider the event \( A_{\eta}^{1,i} := \{ \inf_{0 \leq t \leq T} p_t^i,\theta,\delta,\epsilon > a_\eta \} \). Obviously, (5.46) says that \( \mathbb{P}(A_{\eta}^{1,i}) \geq 1 - \eta \).

Next, we define the event
\[
A_{\epsilon}^{2,i} = \left\{ \forall t \in [t_0,T], \quad \epsilon \int_{t_0}^t \left[ 1 - \frac{1 - p_t^i,\theta,\delta,\epsilon}{p_t^i,\theta,\delta,\epsilon} \right] dW_s \geq -1 \right\}.
\]
We observe that the complementary reads
\[
(A_{\epsilon}^{2,i})^c = \left\{ \exists t \in [t_0,T] : \exp \left( - \int_{t_0}^t \left[ 1 - \frac{1 - p_t^i,\theta,\delta,\epsilon}{p_t^i,\theta,\delta,\epsilon} \right] dW_s - \frac{1}{2} \int_{t_0}^t \frac{1 - p_t^i,\theta,\delta,\epsilon}{p_t^i,\theta,\delta,\epsilon} ds \right) \right\},
\]
from which we get by Doob’s inequality that \( \mathbb{P}(A_{\epsilon}^{2,i}) \geq 1 - \exp(-1). \)

We now work on \( (\cap_{\eta \in [d]} A_{\eta}^{1,i}) \cap A_{\epsilon}^{2,i} \) for \( 2\theta \leq a_\eta \). By combining (5.44) and (5.45), we get (choosing \( \lambda = 2^2 \) therein and noticing that, since we work on \( (\cap_{\eta \in [d]} A_{\eta}^{1,i}) \cap A_{\epsilon}^{2,i} \), we can remove the second indicator function in the second line of (5.45)):
\[
\ln \left( p_t^i,\theta,\delta,\epsilon \right) \geq \ln \left( p_0^i,\theta,\delta,\epsilon \right) - 1 + \int_{t_0}^t \left( \kappa_0 + \epsilon \right) \frac{1}{p_t^i,\theta,\delta,\epsilon} 1_{[0,\Delta]}(p_t^i,\theta,\delta,\epsilon) ds - \left( \frac{\epsilon^2}{\delta} + d(\kappa_0 + M) \right) T.
\]

Hence, for \( \kappa_0 \geq \epsilon \), we can find a constant \( C_2 \geq 0 \) (only depending on \( \kappa_0, M, T \) and \( d \)) such that, on \( (\cap_{\eta \in [d]} A_{\eta}^{1,i}) \cap A_{\epsilon}^{2,i} \), for \( 2\theta \leq a_\eta \leq \epsilon \leq \delta \), \( i \in [d] \) and \( t \in [0,T] \), we have \( p_t^i,\theta,\delta,\epsilon \geq \exp(-C_2) \). (Observe indeed that, in (5.46), we can always assume that \( a_\eta(\delta,\epsilon) \leq \min(\epsilon,\delta) \).)

a. Return back to the first term in the right-hand side of (5.50). By (5.15) (applied with \( \lambda \) replaced by \( 8\lambda^2 \) for \( \lambda' > \lambda \)), if \( \kappa_0 \geq \epsilon^2(1 + 8\lambda^2) \), which is for instance true if \( \kappa_0 \geq 2 \) and \( 8\lambda^2 \epsilon^2 \leq 1 \) (in turn the latter is true if \( 8(\lambda + 1)\epsilon^2 \leq 1 \) and \( \lambda' - \lambda = \frac{1}{2} \)), then
\[
\mathbb{E} \left[ \exp \left\{ 4\lambda' \kappa_0 \int_{t_0}^T \frac{1}{p_t^i,\theta,\delta,\epsilon} 1_{[0,\Delta]}(p_t^i,\theta,\delta,\epsilon) dt \right\} 1_{\{ \inf_{0 \leq t \leq T} p_t^i,\theta,\delta,\epsilon > 2\theta \}} \right] \leq C_3(\delta,\lambda'),
\]
(5.51)
where \( C_3(\delta,\lambda') \) is non-increasing with \( \delta \). Under the same condition \( \kappa_0 \geq 2 \) and \( 16\lambda' \epsilon^2 \leq 1 \), (5.14) (applied with \( \lambda \) replaced by \( 8\lambda^2 \lambda') \) yields
at least whenever $2\theta \leq a_\eta$. Choosing $\eta$ such that $C_4(\delta, \varepsilon, \lambda')\eta^{1/2} \leq C_3(\delta, \lambda')$ and allowing for a new value of $C_3(\delta, \lambda')$, we may remove the second indicator function in (5.51). Then, we can easily change the first indicator function in (5.51) into $1_{[0, \omega]\delta}$ by noticing that $r^{-1}1_{[0, \omega]}(r) \leq r^{-1}1_{[0, \delta]}(r) + \delta^{-1}$. For a new value of $C_3(\delta, \lambda')$ (as long as it remains non-increasing with $\delta$), we then have

$$
\mathbb{E}\left[ \exp \left\{ 4\gamma'k_0 \int_{t_0}^T \frac{1}{p_t^{i, \theta, \delta, \varepsilon}} 1_{[0, \omega]}(p_t^{i, \theta, \delta, \varepsilon}) dt \right\} \right] \leq C_3(\delta, \lambda').
$$

(5.52)

Recall $C_2$ from step 2b and deduce by H"older’s inequality that, for $2\theta \leq a_\eta \leq \varepsilon \leq \delta$ and $3\delta \leq \exp(-C_2)$,

$$
\mathbb{E}\left[ \exp \left\{ -\lambda' \int_{t_0}^T \varphi'_\lambda(p_t^{i, \theta, \delta, \varepsilon}) dt \right\} \right] 
\leq \mathbb{E}\left[ \exp \left\{ 4\lambda k_0 \int_{t_0}^T \frac{1}{p_t^{i, \theta, \delta, \varepsilon}} 1_{[0, \omega]}(p_t^{i, \theta, \delta, \varepsilon}) dt \right\} \right] 
= 1 + \mathbb{E}\left[ \exp \left\{ 4\lambda k_0 \int_{t_0}^T \frac{1}{p_t^{i, \theta, \delta, \varepsilon}} 1_{[0, \omega]}(p_t^{i, \theta, \delta, \varepsilon}) dt \right\} 1_{(\gamma|\delta|d_{A_1}; \gamma|A^2\eta)} \right] 
\leq 1 + C_3(\delta, \lambda')^{1/2} \left( \eta d + \exp(-e^{-1}) \right)^{1/\lambda'}.
$$

For $3\delta \leq \exp(-C_2)$, we may choose $\varepsilon \leq \tilde{\varepsilon}(\delta) \wedge \delta$ with $C_3(\delta, \lambda')^{1/\lambda'} \left[ 2 \exp(-\varepsilon^{-1}(\delta)) \right]^{1-1/\lambda'} = 1$ and then $\eta d \leq \exp(-e^{-1})$, with $a_\eta \leq \varepsilon$, and $\theta \leq \min((a_\eta/2, \theta_1(\delta, \varepsilon, \lambda))$ (with $\theta_1(\delta, \varepsilon, \lambda)$ as in Step 1b). We get that the above right-hand side is less than 2, which is the analogue of (5.48).

a. By collecting (2.18), (5.48) and the above conclusion with $\lambda' = \lambda + 1$ and by applying Cauchy-Schwarz inequality, we obtain

$$
\mathbb{E}\left[ \exp \left\{ -\lambda' \int_{t_0}^T \varphi'_\lambda(p_t^{i, \theta, \delta, \varepsilon}) dt \right\} \right] \leq 2,
$$

(5.53)

for the following choices: $\delta \leq \tilde{\delta}_0 := \exp(-C_2)/3$, $\tilde{k}_0 := \max(2, 4k_2 + 1 + \lambda/2)$, $\varepsilon \leq \min(1/\sqrt{16\lambda} + 16, \delta, \sqrt{k_2/\tilde{k}_0}, \tilde{\varepsilon}(\delta))$ and $\tilde{\theta} \leq \min((\theta_1(\delta, \varepsilon, \lambda), a_\eta))$, for $\eta d \leq \exp(-e^{-1})$ such that $a_\eta \leq \varepsilon$. We recall that the condition $\varepsilon^2 \leq \sqrt{k_2/\tilde{k}_0}$ is required to ensure that $\varphi_{\theta, \tilde{\delta}, \varepsilon}$ is non-increasing. This is one part of the inequality in the definition of the term $\Psi$ showing up in (5.16). In fact, the term with $\varphi_{\theta, \tilde{\delta}, \varepsilon}(p_t^{i, \theta, \delta, \varepsilon})$ in $\Psi$ is bounded in the same way since $\varphi_{\theta, \tilde{\delta}, \varepsilon} \leq -(\varphi'_\theta + \varphi'_\varepsilon)$ for $2\theta \leq \tilde{\delta} \leq 1$, yielding (for the same range of parameters)

$$
\mathbb{E}\left[ \exp \left\{ \frac{\lambda'}{2} \int_{t_0}^T \varphi_{\theta, \tilde{\delta}, \varepsilon}(p_t^{i, \theta, \delta, \varepsilon}) dt \right\} \right] \leq 2.
$$

(5.54)

Step 3. We now handle the term with $1/p_t^{i, \theta, \delta, \varepsilon}$ in (5.16).
a. On the one hand, recalling that \( \kappa_0 \geq 2 \), using (5.52) and arguing as above, we obtain
\[
\mathbb{E} \left[ \exp \left\{ \lambda \int_{t_0}^T \frac{1}{p_t^{i,0,\delta,\varepsilon}} 1_{[0,\delta]}(p_t^{i,0,\delta,\varepsilon}) dt \right\} \right] \leq 2,
\]
again for \( \kappa_0, \kappa_1, \theta, \delta, \varepsilon \) as in item \( d \) of the second step.

a. On the other hand, by following the second step, for \( 2\theta \leq a_H \leq \varepsilon \leq \delta \), we get
\[
\mathbb{E} \left[ \exp \left\{ \lambda \int_{t_0}^T \frac{1}{p_t^{i,0,\delta,\varepsilon}} 1_{[0,\delta]}(p_t^{i,0,\delta,\varepsilon}) dt \right\} \right] \leq \exp (\lambda T \mathcal{C}_2) + \mathbb{E} \left[ \exp \left\{ \lambda \int_{t_0}^T \frac{1}{p_t^{i,0,\delta,\varepsilon}} 1_{[0,1]}(p_t^{i,0,\delta,\varepsilon}) dt \right\} 1_{\{(\cap_{i=1}^{3} \mathcal{A}_i') \cap \mathcal{A}_3^c\}^c} \right] \leq \exp (\lambda T \mathcal{C}_2) + \exp \left( \frac{\lambda T}{\delta} \right) (\eta d + \exp (-\varepsilon^{-1})).
\]
Following item \( 2c \), we can render the last term in the right-hand side less than 1.

a. We then combine items \( 3a \) and \( 3b \) by Cauchy-Schwarz inequality. And, then by Hölder inequality, we gather all the three cases addressed in item \( 2d \) and in this third step to get (5.16), provided we replace \( \lambda \) therein by \( \lambda /6 \) and then fix the various parameters as in item \( 2d \) with the additional constraint that
\[
2 \exp (\lambda T /\delta) \exp (-\varepsilon^{-1}) \leq 1 \text{ (which is equivalent to } \varepsilon \leq (\ln (2) + \lambda T /\delta)^{-1} \text{).}
\]

**Step 4.** We now want to prove (5.17) and (5.18). Throughout the step, we fix the value of \( \lambda \).

a. We recall that, for \( 0 < \delta \leq \delta_0, \hat{\delta}(\delta) \) is defined as
\[
\hat{\delta}(\delta) = \min (\frac{1}{4 \sqrt{\lambda}}, \sqrt{\frac{\kappa_2}{\kappa_0}, \delta^2 (\ln (2[C_3(\delta, \lambda + 1)])^{-1}, (\ln (2) + \lambda T /\delta)^{-1}),}
\]
where \( C_3(\delta, \lambda + 1) \) is a non-increasing function of \( \delta \). (We omit to specify the dependence of \( \hat{\delta}(\delta) \) upon \( \lambda \) and \( \kappa_0 \)\.) Clearly, \( \hat{\delta} \) is non-decreasing on \( (0, \delta_0] \), takes positive values and has 0 as limit in 0. We then define
\[
\bar{\delta}(\delta) := \int_0^\delta \hat{\delta}(\delta') d\delta'.
\]
It is straightforward to verify that, for \( \delta \in (0, \delta_0], \ 0 < \bar{\delta}(\delta) \leq \tilde{\delta}(\delta) < \hat{\delta}(\delta) \) (assume without any loss of generality that \( \delta_0 \leq 1 \)). Moreover, \( \bar{\delta} \) extends by continuity to \( [0, \delta_0] \), letting \( \bar{\delta}(0) = 0 \), and the extension, still denoted by \( \bar{\delta} \), is continuous and strictly increasing.

a. We now define \( \tilde{\delta} : [0, \tilde{\delta}_0] \equiv \mathcal{V} \rightarrow \bar{\delta}(\varepsilon) \in [0, \delta_0] \) as the converse of the mapping \( \bar{\delta} : [0, \bar{\delta}_0] \equiv \delta \rightarrow \bar{\delta}(\delta) \in [0, \bar{\delta}_0] \), where \( \bar{\delta}_0 = \bar{\varepsilon}(\delta_0) \). Conclusion (5.17) hence follows from (5.16), noticing that, for any \( \varepsilon \in [0, \bar{\varepsilon}_0], \varepsilon = \bar{\varepsilon}(\delta) < \hat{\delta}(\delta(e)), \) from which we indeed deduce that \( \Psi(\lambda, \theta, \hat{\delta}(\varepsilon), \varepsilon, K) \leq C, \) for \( \theta \leq \hat{\theta}(\hat{\delta}(\varepsilon), \varepsilon) \). As for (5.18), it
follows from Step 2b, recalling that, under our choice for \( g \), the probability of \( (\cap_{j \in [d]} A_{\ell j}^{1,i}) \cap A^{2,i} \) is less than \( 2 \exp(-\varepsilon^{-1}) \). □

**Proof of Lemma 5.7.** We first prove (5.19). For simplicity, we write \( p \) for \( p_{\Theta, \hat{\lambda} (\cdot), \hat{c}} \) and \( \varphi \) for \( \varphi_{\theta, \delta, \varepsilon} \). Using the same notation as in (5.43) and applying Itô’s formula, we expand

\[
\frac{1}{(p_i^t)^{\ell+1}} = \{ -\ell \frac{1}{(p_i^t)^{\ell+1}} \sum_{j \in [d]} \left( p^t_i(\varphi(p_i^t) + x_i^t) - p^t_i(\varphi(p_j^t) + x_i^t) \right) + \ell(\ell + 1) \frac{\varepsilon^2}{2} \frac{1 - p_i^t}{(p_i^t)^{\ell+1}} \} dt
\]

\[
- \ell \varepsilon \frac{\sqrt{p_i^t(1 - p_i^t)}}{(p_i^t)^{\ell+1}} dW_i^t.
\]

Letting

\[
E_t = \exp \left\{ \int_{t_0}^t \left( \ell \sum_{j \in [d]} \varphi(p_j^s) + \ell(\ell + 1) \frac{\varepsilon^2}{2} \frac{1}{p_j^s} - 1 \right) ds \right\}, \quad t \in [t_0, T],
\]

we have

\[
d \left( \frac{E_t^{-1}}{(p_i^t)^{\ell+1}} \right) = -\ell \frac{E_t^{-1}}{(p_i^t)^{\ell+1}} \sum_{j \neq i} p_i^t(\varphi(p_i^t) + x_i^t) dt + \ell \sum_{j \neq i} \frac{E_t^{-1}}{(p_i^t)^{\ell+1}} x_i^t dt + dm_t
\]

\[
\leq (d - 1) \ell M \frac{E_t^{-1}}{(p_i^t)^{\ell+1}} dt + dm_t,
\]

where \( m_t \) is a local martingale. By a standard localization and Gronwall’s lemma, we deduce that there exists a constant \( C \) depending only on the parameters quoted in the statement of the lemma (in particular, it is independent of \( n \)) such that

\[
\sup_{t_0 \leq t \leq T} \mathbb{E} \left[ \frac{E_t^{-1}}{(p_i^t)^{\ell+1}} \right] \leq C.
\]

Applying Cauchy-Schwarz inequality and then the above inequality, with \( \ell \) replaced by \( 2\ell \), we obtain

\[
\sup_{t_0 \leq t \leq T} \mathbb{E} \left[ \frac{1}{(p_i^t)^{\ell+1}} \right] \leq C \left( \mathbb{E} \left[ \exp \left\{ \int_{t_0}^T \left( 2\ell \sum_{j \in [d]} \varphi(p_j^s) + \ell(2\ell + 1) \frac{1}{p_j^s} \right) ds \right\} \right] \right)^{1/2},
\]

which is bounded by a constant thanks to (5.16), choosing \( \lambda \) in terms of \( \ell \) and \( d \).

The proof of (5.20) follows from the same argument and then from Doob’s maximal inequality (to pass the supremum inside the expectation), see for instance [43, Proof of Proposition 2.4]. □

### 6. Uniqueness for the master equation

Here, our aim is to show Theorem 2.7, namely that \( V \), the value function of the inviscid MFCP, is the unique viscosity solution of the HJ equation (2.23) and that its derivative \( V = D V \) is the unique solution, in a suitable class which we will determine, of the conservative form (2.33) of the master equation of the MFG. In this regard, it is worth
emphasizing that we work below with the local coordinates \((x_1, \ldots, x_{d-1}, x^{-d}) = (p_1, \ldots, p_d)\) for \(p \in S_d\) (and thus \(x \in \hat{S}_d\)). We recall that, for any \(x \in \hat{S}_d\), we have \(V(t, x) = V(t, \hat{x})\), where \(\hat{x} = (x_1, \ldots, x_{d-1}, x^{-d})\), and we denote \(x^{-d} = 1 - \sum_{j \in [d-1]} x_j\). Following (3.6), (2.23) may be indeed rewritten

\[
\begin{cases}
\partial_t \hat{V} + \hat{H}(x, D_x \hat{V}) + \hat{F}(x) = 0, \\
V(T, x) = G(x),
\end{cases}
\]

for \(t \in [0, T]\) and \(x \in \text{Int}(\hat{S}_d)\), and with \(\hat{H}\) as in (5.9). Its derivative \(Z = D_x \hat{V}\) should satisfy (2.33), at least when the latter is formulated in local coordinates, namely

\[
\begin{cases}
\partial_t Z^i + \partial_x \left[ \hat{H}(x, Z) + \hat{F}(x) \right] = 0, \\
Z^i(T, x) = \partial_x G(x),
\end{cases}
\]

the latter reading as a multidimensional hyperbolic system of PDEs. Let us point out a common difficulty in the study of the above two equations: Both are set in a bounded domain, but there are no boundary conditions in space, which is due to the fact that the dynamics of the forward characteristics of the MFG system do not see the boundary of the simplex when starting from its interior.

Concerning the HJ equation (6.1), there are no \(C^1\) solutions in general, which prompts us to consider viscosity solutions. Below, we first handle the HJ equation of the MFCP and then turn to the well-posedness of the conservative form of the master equation. The idea for proving uniqueness of the latter is to construct a correspondence between weak solutions in a suitable class and viscosity solutions of the HJ equation.

### 6.1. HJ equation for the MFCP

In this subsection, we assume that \(F\) and \(G\) are just Lipschitz-continuous. As we have just said, the HJ equation (6.1) is set in a bounded domain but without any boundary conditions in space. We hence define viscosity solutions in the interior of the simplex only:

**Definition 6.1.** A function \(v \in C([0, T) \times \text{Int}(\hat{S}_d))\) (hence defined in local coordinates) is said to be:

i. a **viscosity subsolution** of (6.1) on \([0, T) \times \text{Int}(\hat{S}_d)\) if, for any \(\psi \in C^1([0, T) \times \text{Int}(\hat{S}_d))\),

\[
-\partial_t \psi(t, \hat{x}) - \hat{H}(\hat{x}, D_x \psi(t, \hat{x})) - \hat{F}(\hat{x}) \leq 0,
\]

at every \((t, \hat{x}) \in [0, T) \times \text{Int}(\hat{S}_d)\) which is a local maximum of \(v - \psi\) on \([0, T) \times \text{Int}(\hat{S}_d)\);

ii. a **viscosity supersolution** of (6.1) on \([0, T) \times \text{Int}(\hat{S}_d)\) if, for any \(\psi \in C^1([0, T) \times \text{Int}(\hat{S}_d))\),

\[
-\partial_t \psi(t, \hat{x}) - \hat{H}(\hat{x}, D_x \psi(t, \hat{x})) - \hat{F}(\hat{x}) \geq 0,
\]
at every $(t, \bar{x}) \in [0, T) \times \text{Int}(\hat{S}_d)$ which is a local minimum of $v - \psi$ on $[0, T) \times \text{Int}(\hat{S}_d)$;

i. a viscosity solution of (6.1) on $[0, T) \times \text{Int}(\hat{S}_d)$ if it is both a viscosity subsolution and a viscosity supersolution of (6.1) on $[0, T) \times \text{Int}(\hat{S}_d)$.

In order to prove uniqueness of viscosity solutions, in absence of boundary conditions in space, we must use the fact that the forward characteristics, given by an equation of the type (2.2) with $\alpha$ bounded therein, do not leave the interior of the simplex. The result is the following:

**Theorem 6.2** (Comparison Principle). Let $u, v$ be Lipschitz continuous on $[0, T) \times \hat{S}_d$, $u$ be a viscosity subsolution and $v$ be a viscosity supersolution, respectively, of (6.1) on $[0, T) \times \text{Int}(\hat{S}_d)$. If $u(T, x) \leq v(T, x)$ for any $x \in \hat{S}_d$, then $u(t, x) \leq v(t, x)$ for any $t \in [0, T]$ and $x \in \hat{S}_d$.

Before giving the proof, we state an immediate consequence.

**Corollary 6.3.** There exists a unique viscosity solution of (6.1) on $[0, T) \times \text{Int}(\hat{S}_d)$ that is Lipschitz continuous on $[0, T) \times \hat{S}_d$ and satisfies the terminal condition (6.1). It is the value function $\mathcal{V}$ of the MFCP.

**Proof.** Uniqueness holds in $[0, T) \times \hat{S}_d$ by the above theorem. The fact that the value function is a viscosity solution on $\text{Int}(\hat{S}_d)$ is given by Theorem 7.4.14 of [49] (as we already accounted in the statement of Proposition 5.2).

**Proof of Theorem 6.2.** We borrow ideas from the proofs of Theorem 3.8 and Proposition 7.3 in [56]. The idea is to define a supersolution $v_h$ that dominates $u$ at points near the boundary, for any $h$, and then use the comparison principle and pass to the limit in $h$. The parameter $h$ is needed to force $v_h$ to be infinity at the boundary of the simplex. Since the simplex has corners, the distance to the boundary is not a smooth function, so the first step is to construct a nice test function that goes to 0 as $x$ approaches the boundary. Roughly speaking, we consider the product of the distances to the faces of the simplex, and then take its logarithm.

**Step 1.** Let $\rho_i(x)$, for $x \in \text{Int}(\hat{S}_d)$, be the distance from $x$ to the hyperplane $\{y \in \mathbb{R}^{d-1} : y_i = 0\}$, for $i \in [d-1]$, and $\rho_d(x)$ be the distance to $\{y \in \mathbb{R}^{d-1} : \sum_{i=1}^{d-1} y_i = 1\}$. Specifically, for $x \in \text{Int}(\hat{S}_d)$, we have

$$\rho_i(x) = \begin{cases} x_i & i \in [d-1], \\ x^{-d}/\sqrt{d-1} & i = d, \end{cases}$$

where we recall that $x^{-d} = 1 - \sum_{i\in[d-1]} x_i$. Clearly $\rho_i \in C^\infty(\text{Int}(\hat{S}_d))$.

Since $u$ and $v$ are Lipschitz-continuous, we may let $R := \max\{|D_xu|_\infty, |D_xv|_\infty\}$, which is licit since the gradients are defined almost everywhere. Hence it is easy to show\(^6\) that $u$ and $v$ are viscosity subsolution and supersolution, respectively, on $[0, T) \times \text{Int}(\hat{S}_d)$.

\(^6\)In short, the argument is as follows: If $\psi$ is a continuously differentiable function such that $u - \psi$ has a minimum at some point $(t, \bar{x}) \in [0, T) \times \text{Int}(\hat{S}_d)$, then necessarily $|D_x\psi(t, \bar{x})| \leq R$ and similarly when $(t, \bar{x})$ is a maximum of $v - \psi$.\)
Int($\tilde{S}_d$), of the modified HJB equation

$$\partial_t \tilde{V} + \tilde{H}_{2R}(x, D_x \tilde{V}) + \tilde{F}(x) = 0,$$

(6.5)

with $\tilde{H}_{2R}(x, z) = \sum_{k \in [d-1]} x_k \tilde{H}^k_{2R}(z) + x^d \tilde{H}^d_{2R}(z)$, where $\tilde{H}^i_{2R}$, for $i \in [d]$, is given by (3.7), with $M$ replaced by $2R$ therein and also in the definition (2.16) of $a^*$, which we denote here by $a^*_{2R}$ (see also (3.2) for the way the latter shows up in the Hamiltonian). This modified Hamiltonian has the property that $\tilde{H}_{2R}(x, z) = \tilde{H}(x, z)$ for any $z \in \mathbb{R}^{d-1}$ such that $|z| \leq R$, and is further globally Lipschitz continuous in $(x, z)$ and concave in $z \in \mathbb{R}^{d-1}$, even if not strictly. Now, we show that there exists a constant $C_R$, depending on $R$, such that

$$\langle D_z \tilde{H}_{2R}(x, z), D\rho_i(x) \rangle \geq -C_R \rho_i(x),$$

(6.6)

for any $i \in [d], x \in \text{Int}(\tilde{S}_d)$ and $z \in \mathbb{R}^{d-1}$. Indeed, we have

$$\partial_z \tilde{H}_{2R}(x, z) = \begin{cases} \delta_{i, j} & i \in [d - 1], \\ -1/\sqrt{d - 1} & i = d. \end{cases}$$

Similar to (3.4), we also have, for $j \in [d - 1],$

$$\partial_z \tilde{H}_{2R}(x, z) = \sum_{k \in [d-1]} (x_k a^*_2R(z_k - z_j) - x_j a^*_2R(z_j - z_k)) + x^{-d} a^*_2R(-z_j) - x_j a^*_2R(z_j).$$

Hence, for $i \in [d - 1],$

$$\langle D_z \tilde{H}_{2R}(x, z), D\rho_i(x) \rangle = \partial_z \tilde{H}_{2R}(x, z) \geq -2R(d - 1)x_i = -2R(d - 1)\rho_i(x),$$

while (noticing that the contribution of the first sum in the expansion of $D_z \tilde{H}_{2R}$ is null in the computation below)

$$\langle D_z \tilde{H}_{2R}(x, z), D\rho_d(x) \rangle = \frac{1}{\sqrt{d - 1}} \sum_{j \in [d-1]} (x_j a^*_2R(z_j) - x^{-d} a^*_2R(-z_j))$$

$$\geq -2R \frac{d - 1}{\sqrt{d - 1}} x^{-d} = -2R(d - 1)\rho_d(x),$$

and thus (6.6) holds with $C_R = 2R(d - 1)$.

**Step 2.** For any $h > 0$, let

$$v_h(t, x) := v(t, x) - h^2 \sum_{i \in [d]} \ln (\rho_i(x)) + h(T - t), \quad (t, x) \in [0, T] \times \text{Int}(\tilde{S}_d).$$

We claim that $v_h$ is a viscosity supersolution of (6.5) on $[0, T] \times \text{Int}(\tilde{S}_d)$. Let then $\psi \in C^1([0, T] \times \text{Int}(\tilde{S}_d))$, and $(\tilde{t}, \tilde{x}) \in [0, T] \times \text{Int}(\tilde{S}_d)$ be a local minimum of $v_h - \psi$ on $[0, T] \times \text{Int}(\tilde{S}_d)$. Since $v$ is a viscosity supersolution of (6.5) on $[0, T] \times \text{Int}(\tilde{S}_d)$, considering the test function $\psi_h \in C^1([0, T] \times \text{Int}(\tilde{S}_d))$ given by $\psi_h(t, x) = \psi(t, x) + h^2 \sum_{i \in [d]} \ln (\rho_i(x)) - h(T - t)$, we get

$$-\partial_t \psi_h(\tilde{t}, \tilde{x}) - \tilde{H}_{2R}(\tilde{x}, D_x \psi_h(\tilde{t}, \tilde{x})) - \tilde{F}(\tilde{x}) \geq 0.$$
Using the concavity of $\hat{H}_{2R}$ in the second argument, see (3.2), and (6.6), we obtain

$$0 \leq -\partial_t \psi(t, x) - h - \hat{H}_{2R}(\bar{x}, D_x \psi(t, x)) + h^2 \sum_{i \in [d]} \frac{D\rho_i(\bar{x})}{\rho_i(\bar{x})} - \hat{F}(\bar{x})$$

$$\leq -\partial_t \psi(t, x) - h - \hat{H}_{2R}(\bar{x}, D_x \psi(t, x)) - \hat{F}(\bar{x})$$

$$- \left( D_x \hat{H}_{2R}(\bar{x}, D_x \psi(t, x)) + h^2 \sum_{i \in [d]} \frac{D\rho_i(\bar{x})}{\rho_i(\bar{x})} \right)$$

$$\leq -\partial_t \psi(t, x) - h - \hat{H}_{2R}(\bar{x}, D_x \psi(t, x)) - \hat{F}(\bar{x}) + h^2 dC_R,$$

giving

$$-\partial_t \psi(t, x) - \hat{H}_{2R}(\bar{x}, D_x \psi(t, x)) - \hat{F}(\bar{x}) \geq h - h^2 dC_R \geq 0 \quad \text{if } h \leq \frac{1}{dC_R},$$

which implies that $v_h$ is a viscosity supersolution of (6.5) on $[0, T) \times \text{Int}(\hat{S}_d)$.

**Step 3.** As $\rho_i \leq 1$, we have $v_h(t, x) \geq v(t, x)$ for any $(t, x) \in [0, T) \times \text{Int}(\hat{S}_d)$. In particular, $v_h(T, x) \geq v(T, x) \geq u(T, x)$ for any $(t, x) \in [0, T] \times \text{Int}(\hat{S}_d)$. We denote $\rho(x) = \prod_{i=1}^d \rho_i(x)$. Since $u$ and $v$ are bounded, we find that for any $h > 0$ there exists $\eta > 0$ (which may depend on $h$) such that $-h^2 \ln \rho(x) \geq ||u||_{\infty} + ||v||_{\infty}$ if $\rho(x) \leq \eta$. We denote by $\Gamma^u = \{x \in \hat{S}_d : \rho(x) = \eta\}$, $\Omega^u = \{x \in \hat{S}_d : \rho(x) \geq \eta\}$, and $\Omega^\epsilon = \{x \in \hat{S}_d : \rho(x) \leq \eta\}$; note that $\Omega^u$ is a smooth domain. Thus $v_h(t, x) \geq u(t, x)$ for any $t \in [0, T]$ and $x \in \Omega^\epsilon$, in particular for any $x \in \Gamma^u$. Therefore we can apply the comparison principle (Theorem 9.1 page 90 in [57]) in $[0, T] \times \Omega^u$, because $u, v_h \in C([0, T] \times \Omega^u)$; we obtain $u \leq v_h$ on $[0, T] \times \Omega^u$ and hence $u \leq v_h$ on the entire $[0, T] \times \hat{S}_d$, since we already have $u \leq v_h$ on $[0, T] \times \Omega^u$. Finally, the conclusion follows by sending $h$ to 0, as $\lim_{h \to 0} v_h(t, x) = v(t, x)$ for any $(t, x) \in [0, T] \times \text{Int}(\hat{S}_d)$.

---

### 6.2. Uniqueness of the MFG master equation

We now turn to the analysis of (6.2). Clearly, it has to be understood in the sense of distributions. We assume in this subsection that $F$ and $G$ are in $C^{1,1}(S_d)$. The multidimensional hyperbolic system (6.2) is known to be ill-posed in general. Nevertheless, in this specific potential case, it is possible to prove uniqueness of solutions in a suitable class, thanks to a result of Kružkov [45]. We remark that the system is hyperbolic in the wide sense, but not strictly hyperbolic. We denote $Q_T = (0, T) \times \text{Int}(\hat{S}_d)$, $\hat{Q}_T = [0, T] \times \hat{S}_d$, $\hat{f}(x, z) = \hat{H}(x, z) + \hat{F}(x)$ and $\hat{g}(x) = D_x \hat{G}(x)$.

The set of weak solutions in which we prove uniqueness is the following:

**Definition 6.4.** A function $Z \in [C([0, T]; (L^\infty(\hat{S}_d), *))[^d-1]$ (where * denotes the weak star topology $\sigma^*(L^\infty(\hat{S}_d), L^1(\hat{S}_d)))$ is said to be an admissible solution to the Cauchy problem (6.2) if the following three properties hold true:

1. For any $\varphi = (\varphi^1, ..., \varphi^{d-1}) \in C_C(Q_T; \mathbb{R}^{d-1})$ (where the index $C$ means that $\varphi$ is compactly supported),
Proposition 6.5. We have

\[ 0 \leq \int_{Q_T} [Z^j \partial_t \varphi^j + f(x, Z) \partial_x \varphi^j] \, dx \, dt \]  

(6.7)

2. At time \( t = T \), \( Z(T, \cdot) = g \) a.e.; in particular, by time continuity of \( Z \) with respect to the weak star topology,

\[ Z(t, \cdot) \rightharpoonup g \quad \text{as} \quad t \to T; \]  

(6.8)

3. There exists a universal constant \( c \) such that, for any \( \psi \in C^1_c(\operatorname{Int}(\tilde{S}_d); \mathbb{R}^+) \) and any nonnegative matrix \( A = (A_{ij})_{i,j \in [d-1]} \) with \( \operatorname{Trace}(A) \leq 1 \),

\[ \int_{\operatorname{Int}(\tilde{S}_d)} [(D_x \psi, AZ) + c\psi] \, dx \geq 0. \]  

(6.9)

By Banach-Steinhaus theorem, note that \( Z \in [C([0,T]; (L^\infty(\tilde{S}_d), \sigma^*(L^\infty(\tilde{S}_d), L^1(\tilde{S}_d))))]^{d-1} \) implies \( Z \in L^\infty(Q_T; \mathbb{R}^{d-1}) \).

Before we say more about the solvability of (6.2), we feel useful to elucidate the connection between (6.2) and the original form (2.29) of the master equation. For sure, the main difference between the two is that the former is in conservative form while the latter is not, but also the reader must pay attention to the fact that (6.2) is in local coordinates \((x_1, ..., x_{d-1})\) while (2.29) is written in intrinsic coordinates \((p_1, ..., p_d)\). Obviously, (2.29) can be easily written in local coordinates, which makes it easier to compare with (6.2). Similar to (3.10), but with \( \varphi = \varepsilon = 0 \), the version in local coordinates writes (indices in the sums belonging to \([d-1])\):

\[
\begin{cases}
\partial_t \tilde{U}^i + H((\tilde{U}^i - \tilde{U}^j)_j) + \sum_{i,k} x_k (\tilde{U}^k - \tilde{U}^i)_+ - x_i (\tilde{U}^j - \tilde{U}^k)_+ \partial_x \tilde{U}^i \\
+ \sum_j (x^d (\tilde{U}^d - \tilde{U}^j)_+ - x_j (\tilde{U}^j - \tilde{U}^d)_+) \partial_x \tilde{U}^i + \tilde{f}^i(x) = 0,
\end{cases}
\]  

(6.10)

for \((t, x) \in [0,T] \times \operatorname{Int}(\tilde{S}_d)\) and \( i \in [d] \). As we explained in Subsections 2.4 and 3.2, the key step to pass from one formulation to another is Schwarz identity. The following statement clarifies this fact.

**Proposition 6.5.** We have

1. if \( U \in [C^1([0,T] \times S_d)]^d \) is a classical solution of the master equation (2.29) and \( \tilde{U} \) denotes its version in local coordinates, then \( Z \) defined by \( Z^i = \tilde{U}^i - \tilde{U}^d \), for \( i \in [d-1] \), is a weak admissible solution to (6.2); it satisfies \( \partial_x Z^i = \partial_x \tilde{U}^i \) for any \( i,j \in [d-1] \);

2. if \( Z \) is a weak solution to (6.2), in the sense that it satisfies (6.7), and \( Z \) is in \([C^1([0,T] \times S_d)]^{d-1} \), then the master equation (2.29) has a (unique) classical solution \( U \in [C^1([0,T] \times S_d)]^d \); denoting \( \tilde{U} \) its version in local coordinates, the latter satisfies \( \tilde{Z}^i = \tilde{U}^i - \tilde{U}^d \), for \( i \in [d] \).

The proof of Proposition 6.5 is postponed to the end of the section, as we feel better to focus now on the following statement, which is the refined version of Theorem 2.7. Indeed, the next theorem establishes uniqueness of admissible solutions to (6.2), by determining a correspondence with viscosity solutions to (6.1–6.1). The proof is to establish first a connection between admissible solutions to (6.2) and semiconcave
solutions to (6.1–6.1) and then to show that viscosity and semiconcave solutions to (6.1–6.1) are equivalent. We recall that, in our case, a function \( v \in C([0,T] \times \hat{S}_d) \) is called semiconcave (in space) if there exists a constant \( c \) such that, for any \( t \in [0,T], x \in \text{Int}(\hat{S}_d) \) and \( \xi \) with \( x \pm \xi \in \text{Int}(\hat{S}_d) \),

\[
\frac{v(t,x + \xi) - 2v(t,x) + v(t,x - \xi)}{|\xi|^2} \leq c.
\]  
(6.11)

We stress that only semiconcavity in space is needed in the analysis below (for simplicity, we just call it semiconcavity), although the value function \( V \) is shown to be semiconcave in time and space in Proposition 5.2 (v), see (5.12). In this framework, we refer to condition (6.9) as a weak one-sided Lipschitz condition, since it reads as the derivative of the above condition. This terminology has been widely used in Filippov’s theory for differential equations with discontinuous coefficients; see [58]. In a framework even closer to ours, it has been used in the analysis of some scalar conservation laws. One-sided Lipschitz solutions then derive from semiconcave solutions to Hamilton-Jacobi equations. The one-sided Lipschitz condition is then enough to establish a uniqueness criterion that is consistent with the theory of entropy solutions; see e.g. [49, Corollary 1.7.2]. To show in our case that condition (6.9) is satisfied by the derivative of a semiconcave solution (of the HJ equation) if (6.11) holds, we say that \( v \) is a semiconcave solution (of the HJ equation) if (6.11) holds, \( v \) is Lipschitz-continuous in \([0,T] \times \hat{S}_d\), Equation (6.1) holds almost everywhere and the terminal condition (6.1) is satisfied (everywhere).

The proof of the following theorem is mostly due to Kružkov [45], see Theorem 8 therein. For the sake of completeness, we write its adaptation to our framework (as the state variable here belongs to the simplex).

**Theorem 6.6.** There exists a unique admissible solution to (6.2). It is given by \( D_x\hat{V} \), where \( V \) is the value function of the inviscid MFCP and \( \hat{V} \) is its version in local chart.

**Proof.** As we have just explained, we first establish a connection between admissible solutions to (6.2) and semiconcave solutions to (6.1–6.1) and, then, we show equivalence between semiconcave and viscosity solutions to (6.1–6.1).

**Step 1.** Let \( Z \in [C([0,T]; (L^\infty(\hat{S}_d), *))])^{d-1} \) be an admissible solution to (6.2). Let \( w \in C^2_c(\text{Int}(\hat{S}_d)) \) and \( \zeta \in C^\infty_c((0,T)) \), and for fixed \( i \neq j \) choose as test functions \( \phi^i(t,x) = \zeta(t)\partial_{x_j} w(x), \phi^j(t,x) = \zeta(t)\partial_{x_i} w(x) \). Then (6.7) provides

\[
\int_0^T \partial_t \zeta(t) \int_{\hat{S}_d} \left[ Z^i(t,x)\partial_{x_j} w(x) - Z^j(t,x)\partial_{x_i} w(x) \right] dx dt = 0,
\]

which, by the fundamental lemma of the calculus of variations, implies that the quantity (Tex translation failed) is a constant for almost every \( t \in [0,T] \). Hence (6.8) and the fact that the final condition is a gradient yield
\[
\int_{\hat{S}_d} \left[ Z'(t,x)\partial_y w(x) - Z'(t,x)\partial_x w(x) \right] dx = 0, \quad \text{for all } w \in C^2_c(\text{Int}(\hat{S}_d)),
\]
which means that \( Z(t,\cdot) \) admits a potential, in the weak sense, for almost every \( t \).

**Step 2.** Fix \( (s,y) \in Q_T \) and choose as test function \( \varphi \) the mollification kernel \( \rho_h(t,x) = h^{-d}\rho((s-t)/h, (y-x)/h) \). Then (6.7) gives

\[
\partial_t Z^h + \partial_x \tilde{f}_h = 0 \quad \text{in } Q^h_T,
\]
where \( Z^h = \rho_h \ast Z, \tilde{f}_h = \rho_h \ast (\tilde{f}(\cdot, Z)) \) and \( Q^h_T \) is the set of \( (s, y) \) in \( Q_T \) with a distance to the (time-space) boundary that is greater than or equal to \( h \). Thanks to (6.12), \( Z^h \) derives from a potential for fixed \( t \), and the equation above implies that \( (-\tilde{f}_h, Z^h_1, \ldots, Z^h_d) \) also derives from a potential (but in time and space) for \( (t, x) \in Q^h_T \). Thus there exists a function \( v_h \) defined in \( Q^h_T \) such that \( \partial_t v_h = -\tilde{f}_h \) and \( \partial_x v_h = Z^h_1 \). Since \( v_h \) is defined up to a constant, we fix \( \nu = (-T-h, x^M) = \tilde{G}(x^M) \), where \( x^M = (1/d, \ldots, 1/d) \in \mathbb{R}^{d-1} \) is the point in the middle of the simplex.

By condition (6.9), substituting again the mollification kernel and integrating over \( t \), we obtain, on \( Q^h_T \),

\[
\sum_{i,j \in [d-1]} A_i,j \partial^2_{x_i x_j} v_h \leq c,
\]
for any nonnegative matrix \( A \) with \( \text{Trace}(A) \leq 1 \), which implies in particular that for any vector \( \nu \) with \( |\nu| = 1 \) we have, also on \( Q^h_T \),

\[
\frac{\partial^2 v_h}{\partial \nu^2} \leq c.
\]

**Step 3.** Let \( h \to 0 \). We have \( \lim_{h \to 0} \partial_t v_h = -\tilde{f}(x, Z) \) and \( \lim_{h \to 0} D_x v_h = Z \) almost everywhere in \( Q_T \). By Ascoli-Arzelà theorem and by boundedness of \( \tilde{f}_h \) and \( Z_h \), uniformly in \( h > 0 \), the sequence \( (v_h)_{h>0} \) is precompact in \( C(\bar{Q}_T) \) endowed with the topology of uniform convergence, it being understood that we extend \( v_h \) outside \( Q^h_T \) as a Lipschitz function. Let \( \nu \) be any limit point. We have necessarily that \( \nu \) is Lipschitz continuous (with a fixed Lipschitz constant) on \( \bar{Q}_T \) and has weak derivatives \( \partial_t \nu = -\tilde{f}(x, Z) \) and \( D_x \nu = Z \) a.e. in \( Q_T \), proving that \( \partial_t \nu + \tilde{f}(x, D_x \nu) = 0 \) a.e. in \( Q_T \).

Since \( v_h(T-h, x^M) = \tilde{G}(x^M) \), we get \( \nu(T, x^M) = \tilde{G}(x^M) \). Moreover, for any test function \( w \in C^2_c(\text{Int}(\hat{S}_d)) \) and any \( h > 0 \) that is less than the distance \( \text{dist}(\text{Supp}(w), \partial \hat{S}_d) \) from the support of \( w \) to the boundary of the simplex, we have

\[
\int_{\hat{S}_d} v_h(T-h, x) D_x w(x) dx = -\int_{\hat{S}_d} Z_h(T-h, x) w(x) dx
\]
\[
= -\int_{\hat{S}_d} (\rho_h \ast Z)(T-h, x) w(x) dx
\]
\[
= -\int_{\mathbb{R}^d} \rho_h(s, y) \left[ \int_{\hat{S}_d} Z(T-h-s, x) w(x+y) dx \right] ds dy.
\]
By (6.8), we know that, for any $|y| \leq \text{dist}(\text{Supp}(w), \partial \hat{S}_d)/2$, $\lim_{h \to 0} [\int_{\hat{S}_d} Z(T - h, x)w(x + y)dx] = \int_{\hat{S}_d} g(x)w(x + y)dx$. Since the function in argument of the limit is uniformly continuous with respect to $y$, the convergence holds uniformly with respect to $y$. Hence, the right-hand side in (6.16) converges to $-\int_{\hat{S}_d} g(x)w(x)dx$. Since the left-hand side in (6.16) converges to $\int_{\hat{S}_d} v(T, x)D_xw(x)dx$, we deduce that $D_xv(T, \cdot) = D_x\hat{G} = g$ a.e. and then $v(T, \cdot) = \hat{G}$ on $\hat{S}_d$ since both are continuous and coincide in $x^d$.

Lastly, by inequality (6.15) (writing first the inequality below for $v_h$ and then taking the limit as $h$ tends to 0)

$$v(t, x + \xi) - 2v(t, x) + v(t, x - \xi) \leq c|\xi|^2,$$

for any $t \in (0, T), x \in \text{Int}(\hat{S}_d)$ and $\xi$ such that $x \pm \xi \in \text{Int}(\hat{S}_d)$, thus (6.11) holds. Hence, we have proved that $v$ is a semiconcave solution to the Cauchy problem (6.1–6.1) and $Z = D_xv$ a.e. in $Q_T$.

**Step 4.** On the converse, if $v$ is a semiconcave solution to (6.1) then, for any $t \in [0, T], v(t, \cdot)$ is a.e. differentiable in $x$. By integration by parts, it is clear that, for any $w \in C^1_c(\hat{S}_d)$, the function $[0, T] \ni t \mapsto \int_{\hat{S}_d} D_xv(t, x)w(x)dx$ is continuous. Since $D_xv \in L^\infty(Q_T)$, the result easily extends to any $w \in L^1(\hat{S}_d)$, hence proving that $D_xv \in C([0, T]; L^\infty(\hat{S}_d), *)$. Also, $v$ is a.e. differentiable in $(t, x)$ and the $x$-derivative clearly satisfies (6.7). Obviously, (6.8) holds true. So we have just to check (6.9). For any $h > 0$, let $v'_h := \rho_h \ast v$ (it being understood that $v$ can be extended in a Lipschitz fashion outside $Q_T$). From (6.11) we derive again inequality (6.15), but with $v_h$ replaced by $v'_h$, and then (6.14) follows. Multiplying (6.14) by $\psi \in C^1_c(\text{Int}(\hat{S}_d); \mathbb{R}_+)$ (provided that $h$ is smaller than dist(Supp($\psi$), $\partial \hat{S}_d$)) and integrating by parts we get (for any nonnegative matrix $A$ with a trace lower than or equal to 1)

$$\int_{\text{Int}(\hat{S}_d)} [\langle D_x\psi, AD_xv'_h \rangle + c\psi] dx \geq 0,$$

and letting $h \to 0$ we obtain (6.9).

**Step 5.** It remains to show that there is a correspondence between semiconcave and viscosity solutions to (6.1–6.1). By Corollary 6.3, any viscosity solution $V$ is in fact the value function of the MFCP. By items (iv) and (v) in Proposition 5.2, the value function is Lipschitz continuous and semiconcave. By Proposition 3.1.7 in [49], $V$ solves (6.1) almost everywhere. On the converse, if $v$ is a semiconcave solution then it is also a viscosity solution on $[0, T] \times \text{Int}(\hat{S}_d)$ by Theorem 10.2 in [51]. By Corollary 6.3, it hence coincides with the value function. □

We now turn:

**Proof of Proposition 6.5.**

**Step 1.** We first assume that $U$ is a classical solution of the master equation (2.29) or equivalently that $\hat{U} \in [C^1([0, T] \times \hat{S}_d)]^d$ is a classical solution to (6.10). Then, $(Z^i = \hat{U}^i - \hat{U}^d)_{i \in [d-1]}$ is a (classical) solution of
\[
\begin{aligned}
&\partial_t Z^i + \partial_x \hat{H}(Z) - \partial_x \tilde{H}(Z) + \sum_{j,k}(x_k (Z^i - Z^j)_+ - x_j (Z^i - Z^k)_+) \partial_x Z^i \\
&+ \sum_j (x^{-d} (Z^i)_+ - x_j (Z^i)_+) \partial_x Z^i + \hat{f}^i(x) - \tilde{f}^d(x) = 0,
\end{aligned}
\]  
(6.17)

for \((t, x) \in [0, T] \times \text{Int}(\hat{S}_d), i \in [d - 1]\). Obviously, the system of characteristics of (6.17) is nothing but the Pontryagin system (in local coordinates) (5.10), see (ii) in Proposition 5.2. Hence, the fact that \(Z\) is a classical solution of (6.17) implies that (5.10) has a unique solution, for any initial condition \((t_0, x_0)\) of the forward equation in (5.10). The argument is pretty standard: By expanding \((Z^i(t, x_t))_{t_0 \leq t \leq T}\) and then comparing with the backward equation, we prove that any solution \((x, z)\) of (5.10) must be of the form \(z^i(t, x_t)\), for \(i \in [d - 1]\) and \(t_0 \leq t \leq T\); Conversely, solving the forward equation with \(z^i(t_0, x_0)\) for \(i \in [d - 1]\) and \(t_0 \leq t \leq T\), we can indeed easily construct a solution. In turn, we deduce that the inviscid MFCP admits a unique optimizer: By (i) in Proposition 5.2, there exists a minimizer; uniqueness follows from the fact the Pontryagin system (5.10) has a unique solution. By Proposition 5.2 (vii), the value function \(\hat{V}\) of the MFCP is differentiable in any \((t_0, x_0)\) and, by point (ix) in the same Proposition, \(z^i(t_0, x_0) = \partial_{x^i} \hat{V}(t_0, x_0)\), whenever the forward equation in (5.10) starts from \(x_0\) at time \(t_0\), but in turn \(z^i(t, x_t) = Z^i(t_0, x_0)\) hence showing that \(Z^i(t, x_t) = \partial_{x^i} \hat{V}(t, x_t)\) for any \(t \in [0, T]\) and \(x \in \text{Int}(\hat{S}_d)\), which implies that, on \([0, T] \times \text{Int}(\hat{S}_d), \hat{V}\) is \(C^2\) and \(\partial_{x^i} \hat{Z}^i = \partial_{x^j} \hat{Z}^j\) for any \(i, j \in [d - 1]\). Recalling (5.9), it is plain to see that \(\partial_{x^i} \hat{H}(x, Z)\) coincides with the nonlinear terms in (6.17), which shows that \(Z\) is a solution to (6.2) on \([0, T] \times \text{Int}(\hat{S}_d)\). It is straightforward to see that it satisfies (6.7) and (6.9), because \(\hat{V}\) is \(C^2\) and (obviously) semiconcave.

**Step 2.** Let \(Z \in [C^1([0, T] \times \hat{S}_d)]^{d-1}\) satisfy (6.7). From (6.12) and the fact that \(Z \in [C^1([0, T] \times \hat{S}_d)]^{d-1}\), we obtain that \(\partial_{\nu^i} \hat{Z}^i = \partial_{\nu^i} \hat{Z}^i\) on \([0, T] \times \text{Int}(\hat{S}_d)\). Thus \(Z\) solves (6.17) on \([0, T] \times \text{Int}(\hat{S}_d)\), but then it solves the equation also at the boundary, because it is differentiable up to the boundary. It remains to construct a classical solution to the master equation (6.10). To do so, it suffices to solve (6.10) with all the occurrences of \(\hat{U}^k - \hat{U}^j\) replaced by \(Z^k - Z^j\) and all the occurrences of \(\hat{U}^j - \hat{U}^d\) replaced by \(Z^j\). By doing so, we hence solve a linear system of transport equations with a vector field that is \(C^2\). Despite the fact that the linear system is set on the simplex, there is no real difficulty for proving that the solution is also \(C^1\). \(\Box\)

**Acknowledgments**

We wish to thank Pierre Cardaliaguet for helpful discussions and suggestions on the paper. We are also very grateful to the two anonymous referees for their useful and insightful comments which helped us improve a lot the paper.

**Funding**

A. Cecchin and F. Delarue acknowledge the financial support of French ANR project ANR-16-CE40-0015-01 on “Mean Field Games.” A. Cecchin also benefited from the support of LABEX Louis Bachelier Finance and Sustainable Growth—project ANR-11-LABX-0019, ECOORES ANR
Project, FDD Chair and Joint Research Initiative FiME in partnership with Europlace Institute of Finance. F. Delarue also thanks French ANR projet ANR-19-P3IA-0002 “3IA Côte d’Azur - Nice - Interdisciplinary Institute for Artificial Intelligence.”

**ORCID**

François Delarue [http://orcid.org/0000-0002-5511-1088](http://orcid.org/0000-0002-5511-1088)

**References**

[1] Lasry, J.-M., Lions, P.-L. (2006). Jeux à champ moyen. I. Le cas stationnaire. *C. R. Math. Acad. Sci. Paris*. 343(9):619–625. DOI: 10.1016/j.crma.2006.09.019.

[2] Lasry, J.-M., Lions, P.-L. (2006). Jeux à champ moyen. II. Horizon fini et contrôle optimal. *C. R. Math. Acad. Sci. Paris*. 343(10):679–684. DOI: 10.1016/j.crma.2006.09.018.

[3] Lasry, J.-M., Lions, P.-L. (2007). Mean field games. *Jpn. J. Math*. 2(1):229–260. DOI: 10.1007/s11537-007-0657-8.

[4] Huang, M., Caines, P., Malhamé, R. (2003). Individual and mass behavior in large population stochastic wireless power control problems: centralized and Nash equilibrium solutions, pp. 98–103.

[5] Huang, M., Caines, P. E., Malhamé, R. P. (2007). The Nash certainty equivalence principle and McKean-Vlasov systems: an invariance principle and entry adaptation. In 2007 46th IEEE Conference on Decision and Control. IEEE, pp. 121–126.

[6] Huang, M., Malhamé, R. P., Caines, P. E. (2006). Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle. *Commun. Inf. Syst*. 6(3):221–251.

[7] Cardaliaguet, P. (2013). Notes on mean field games. [https://www.ceremade.dauphine.fr/cardalia/MFG20130420.pdf/](https://www.ceremade.dauphine.fr/cardalia/MFG20130420.pdf/).

[8] Gomes, D. A., Mohr, J., Souza, R. R. (2013). Continuous time finite state mean field games. *Appl. Math. Optim*. 68(1):99–143. DOI: 10.1007/s00245-013-9202-8.

[9] Cardaliaguet, P., Delarue, F., Lasry, J.-M., Lions, P.-L. (2019). *The Master Equation and the Convergence Problem in Mean Field Games, Volume 201 of Annals of Mathematics Studies*. Princeton, NJ: Princeton University Press.

[10] Carmona, R., Delarue, F. (2018). Mean field FBSDEs, control, and games. In *Probabilistic Theory of Mean Field Games with Applications. I, Volume 83 of Probability Theory and Stochastic Modelling*. Cham, Switzerland: Springer.

[11] Carmona, R., Delarue, F. (2018). Mean field games with common noise and master equations. *Probabilistic Theory of Mean Field Games with Applications. II, Volume 84 of Probability Theory and Stochastic Modelling*. Cham, Switzerland: Springer.

[12] Chassagneux, J.-F., Crisan, D., Delarue, F. A probabilistic approach to classical solutions of the master equation for large population equilibria. *Memoirs of the AMS*. To appear.

[13] Lions, P.-L. Cours au Collège de France, Équations aux dérivées partielles et applications, 2008-09. [https://www.college-de-france.fr/site/pierre-louis-lions/course-2008-2009.htm](https://www.college-de-france.fr/site/pierre-louis-lions/course-2008-2009.htm).

[14] Lions, P.-L. Estimées nouvelles pour les équations quasilinéaires. Seminar in Applied Mathematics at the Collège de France, 2014. [http://www.college-de-france.fr/site/pierre-louis-lions/seminar-2014-11-14-11h15.htm](http://www.college-de-france.fr/site/pierre-louis-lions/seminar-2014-11-14-11h15.htm).

[15] Bayraktar, E., Cohen, A. (2018). Analysis of a finite state many player game using its master equation. *SIAM J. Control Optim*. 56(5):3538–3568. DOI: 10.1137/17M113887X.

[16] Cecchin, A., Pelino, G. (2019). Convergence, fluctuations and large deviations for finite state mean field games via the master equation. *Stoch. Process. Appl*. 129(11):4510–4555. DOI: 10.1016/j.spa.2018.12.002.
[17] Lions, P.-L. Cours au Collège de France, Equations aux dérivées partielles et applications, 2010-11. https://www.college-de-france.fr/site/pierre-louis-lions/course-2010-2011.htm.

[18] Bensoussan, A., Frehse, J., Yam, S. C. P. (2015). The master equation in mean field theory. J. Math. Pures Appl. 103(6):1441–1474. DOI: 10.1016/j.matpur.2014.11.005.

[19] Chassagneux, J.-F., Crisan, D., Delarue, F. (2019). Numerical method for FBSDEs of McKean-Vlasov type. Ann. Appl. Probab. 29(3):1640–1684. DOI: 10.1214/18-AAP1429.

[20] Gomes, D. A., Velho, R. M., Wolfram, M.-T. (2014). Dual two-state mean-field games. In 2014 IEEE 53rd Annual Conference on Decision and Control (CDC). IEEE, pp. 2703–2708.

[21] Gomes, D. A., Velho, R. M., Wolfram, M.-T. (2014). Socio-economic applications of finite state mean field games. Phil. Trans. R Soc. A. 372(2028):20130405. DOI: 10.1098/rsta.2013.0405.

[22] Gangbo, W., Mészáros, A. R. (2020). Global well-posedness of Master equations for deterministic displacement convex potential mean field games. arXiv e-prints.

[23] Bayraktar, E., Zhang, X. (2020). On non-uniqueness in mean field games. Proc. Amer. Math. Soc. 148(9):4091–4106.

[24] Cecchin, A., Dai Pra, P., Fischer, M., Pelino, G. (2019). On the convergence problem in mean field games: a two state model without uniqueness. SIAM J. Control Optim. 57(4):2443–2466. DOI: 10.1137/18M1222454.

[25] Delarue, F., Tchuendom, R. F. (2020). Selection of equilibria in a linear quadratic mean field game. Stoch. Process. Appl. 130(2):1000–1040. DOI: 10.1016/j.spa.2019.04.005.

[26] Guéant, O. (2011). From infinity to one: the reduction of some mean field games to a global control problem. Cahier de la Chaire Finance et Développement Durable, 42.

[27] Cardaliaguet, P. (2015). Weak solutions for first order mean field games with local coupling. In Analysis and Geometry in Control Theory and its Applications. Springer INdAM Series. vol. 11. New York, NY: Springer, pp. 111–158.

[28] Cardaliaguet, P., Graber, J. (2015). Mean field games systems of first order. ESAIM: COCV. 21(3):690–722. DOI: 10.1051/cocv/2014044.

[29] Cardaliaguet, P., Graber, J., Porretta, A., Tonon, D. (2015). Second order mean field games with degenerate diffusion and local coupling. Nonlinear Differ. Equ. Appl. 22(5):1287–1317. DOI: 10.1007/s00030-015-0323-4.

[30] Carmona, R., Delarue, F. (2013). Probabilistic analysis of mean-field games. SIAM J. Control Optim. 51(4):2705–2734. DOI: 10.1137/120883499.

[31] Cecchin, A., Fischer, M. (2020). Probabilistic approach to finite state mean field games. Appl. Math. Optim. 81(2):253–300. DOI: 10.1007/s00245-018-9488-7.

[32] Fischer, M. (2017). On the connection between symmetric N-player games and mean field games. Ann. Appl. Probab. 27(2):757–810. DOI: 10.1214/16-AAP1215.

[33] Lacker, D. (2016). A general characterization of the mean field limit for stochastic differential games. Probab. Theory Relat. Fields. 165(3–4):581–648. DOI: 10.1007/s00440-015-0641-9.

[34] Lacker, D. (2020). On the convergence of closed-loop Nash equilibria to the mean field game limit. Ann. Appl. Probab. 30(4):1693–1761. DOI: 10.1214/19-AAP1541.

[35] Méléard, S. (1996). Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. In Probabilistic Models for Nonlinear Partial Differential Equations (Montecatini Terme, 1995), Volume 1627 of Lecture Notes in Mathematics. New York, NY: Springer, pp. 42–95.

[36] Sznitman, A. (1985). A fluctuation result for nonlinear diffusions. In: Infinite Dimensional Analysis and Stochastic Processes. New York, NY: Springer, 145–160.

[37] Tanaka, H., Hitsuda, M. (1981). Central limit theorem for a simple diffusion model of interacting particles. Hiroshima Math. J. 11(2):415–423. DOI: 10.32917/hmj/1206134109.

[38] Carmona, R., Delarue, F., Lacker, D. (2016). Mean field games with common noise. Ann. Prob. 44:3740–3803.

[39] Bertucci, C., Lasry, J.-M., Lions, P.-L. (2019). Some remarks on mean field games. Commun. Part. Differ. Eq. 44(3):205–227. DOI: 10.1080/03605302.2018.1542438.

[40] Delarue, F. (2021). Master equation for finite state mean field games with additive common noise. In: Cardaliaguet, P., Porretta, A., eds., Mean Field Games, Cetraro, Italy 2019, LNM 2281. New York, NY: Springer, pp. 203–248.
Appendix A.

Wright-Fisher spaces

We describe the so-called Wright-Fisher spaces used in the paper, as recently introduced in the monograph of Epstein and Mazzeo [44]. We here follow the exposition given in [43, Subsection 3.2]. In short, these Wright-Fisher spaces are Hölder spaces, tailor-made to the study of second order operators of the form

\[ \mathcal{L}_i h(p) = \sum_{i \in [d]} a_i(t, p) \partial_i h(p) + \frac{\epsilon^2}{2} \sum_{i, j \in [d]} (p_i \delta_{i,j} - p_i p_j) \partial^2_{ij} h(p), \]  

(A.1)

where \( p \in S_d \) and \( a_i(t, p) \geq 0 \) if \( p_i = 0 \). As we already alluded to, such operators are called Kimura operators; we refer to [59–61] for earlier analyses. Clearly, the second order term in (A.1) is degenerate, which is a price to pay for forcing the corresponding SDE to stay in the simplex. In fact, the latter SDE is nothing but a Wright-Fisher SDE of the same type as (2.12), at least for a relevant choice of the drift \( (a_i)_{i \in [d]} \). The key feature is that, under the identification of \( S_d \) with \( S_d \) (see the introduction for the notation), we may regard the simplex as a \((d - 1)\)-dimensional manifold with corners, the corners being obtained by intersecting at most \( d \) of the hyperplanes \( \{ x \in \mathbb{R}^{d-1} : x_1 = 0 \}, \ldots, \{ x \in \mathbb{R}^{d-1} : x_{d-1} = 0 \}, \{ x \in \mathbb{R}^{d-1} : x_1 + \cdots + x_{d-1} = 1 \} \) with \( S_d \) (we then call the codimension of the corner the number of hyperplanes showing up in the intersection). Accordingly, we can rewrite (A.1) as an operator acting on functions from \( S_d \) to \( \mathbb{R} \), by reformulating (A.1) in terms of the sole \( d - 1 \) first coordinates \((p_1, \ldots, p_{d-1})\) or, more generally, in terms of \((p_i)_{i \in [d] \setminus \ell}\) for any given coordinate \( \ell \in [d] \). Choosing the coordinate \( \ell \) amounts to choosing a system of local coordinates and the choice of \( \ell \) is mostly dictated by the position of \((p_1, \ldots, p_d)\) inside the simplex. Whenever all the entries of \( p = (p_1, \ldots, p_d) \) are positive, meaning that \((p_1, \ldots, p_d)\) belongs to the interior of \( S_d \), the choice of \( \ell \) does not really matter and we work, for convenience, with \( \ell = d \) (which is, in fact, what we have done throughout the paper).

In [43, Subsection 3.2], it is shown that the operator (A.1) fits the decomposition of [44, Definition 2.2.1], which allows to use the Schauder-like theory developed in the latter reference. Roughly speaking, [44, Definition 2.2.1] says that at any point, especially those belonging to the boundary, there exists a system of local coordinates \((x, y)\), \( x = (x_1, \ldots, x_m) \) living in a neighborhood of 0 in the \( n\)-dimensional orthant and \( y = (y_1, \ldots, y_m) \) living in the neighborhood of 0 in \( \mathbb{R}^m \), for which the principal part of the operator (A.1) can be split into the sum of the degenerate operator \( \sum_{j=1}^m x_j \partial^2_{x_j} \) and of a non-degenerate part acting on the variable \( y \). We do not repeat the computations here, but we recall the following two facts (we use below the same notation \( \hat{x} \) for the local coordinates as in (3.5)). Firstly, the operator (A.1) is elliptic non-degenerate in the interior of the simplex, when written in local coordinates in \( S_d \) in the form

\[ \mathcal{L}_i \hat{h}(x) = \sum_{i \in [d-1]} \hat{a}_i(t, x) \partial_i \hat{h}(x) + \frac{\epsilon^2}{2} \sum_{i, j \in [d-1]} (x_i \delta_{i,j} - x_i x_j) \partial^2_{ij} \hat{h}(x), \]

(A.2)

where now \( x \in S_d, \hat{h} \) is a smooth function on \( S_d \) (which must be thought of \( \hat{h}(x) = h(\hat{x}) \)) and \( \hat{a}_i(t, x) = a_i(t, \hat{x}) \). Secondly, for a point in the relative interior of a corner of codimension \( \ell \), there exist local coordinates, of the form \((p_i)_{i \in [d] \setminus \ell}\) for a given \( \ell \) depending on the shape of the corner, such that, in the new coordinates, the operator satisfies the normal form required in [44, Definition 2.2.1]. The details are useless here, but to have some intuition, say that the codimension \( \ell \) is equal to \( m \) in the aforementioned decomposition.

Hence, for a point \( x^0 \in S_d \) in the relative interior of a corner \( C \) of \( S_d \) of codimension \( \ell \in \{0, \ldots, d\} \) (if \( \ell = 0 \), then \( x^0 \) is in the interior of \( S_d \)), we may consider a new system of coordinates \((y_1, \ldots, y_{d-1})\) (obtained as in the second point above) such that \( C = \{ y \in S_d : y_{i_1} = \cdots = y_{i_\ell} = 0 \}, \) for \( 1 \leq i_1 < \cdots < i_\ell \). Letting \( I := \{ i_1, \ldots, i_\ell \} \) and denoting by \((y_j^0, \ldots, y_{d-1}^0)\) the coordinates of \( x^0 \) in the new system (for sure \( y_j^0 = 0 \) for \( j = 1, \ldots, \ell \)), we may find a \( \delta^0 > 0 \) such that:
1. the closure \( \bar{U}(\delta^0, x^0) \) of \( U(\delta^0, x^0) := \{ y \in (\mathbb{R}^d)_{d-1} : \sup_{t \in [d-1]} |y_t - y^0_t| < \delta^0 \} \) is included in \( S_d \).
2. for \( y \) in \( U(\delta^0, x^0) \), for \( j \neq l \), \( y_j > 0 \),
3. for \( y \) in \( U(\delta^0, x^0) \), for \( y_1 + \cdots + y_{d-1} < 1 - \delta^0 \).

A function \( \hat{h} \) defined on \( \bar{U}(\delta^0, x^0) \) is then said to belong to \( C_{WF}^\gamma(\bar{U}(\delta^0, x^0)) \), for some \( \gamma \in (0, 1) \), if, in the new system of coordinates, \( \hat{h} \) is Hölder continuous on \( \bar{U}(\delta^0, x^0) \) with respect to the distance

\[
d(y, y') := \sum_{i \in [d-1]} [\sqrt{y_i} - \sqrt{y'_i}].
\]

We then let

\[
||\hat{h}||_{\gamma, U(\delta^0, x^0)} := \sup_{y \in \bar{U}(\delta^0, x^0)} |\hat{h}(y)| + \sup_{y, y' \in \bar{U}(\delta^0, x^0)} |\hat{h}(y) - \hat{h}(y')|/d(y, y').
\]

Following [44, Lemma 5.2.5 and Definition 10.1.1], we say that a function \( \hat{h} \) defined on \( \bar{U}(\delta^0, x^0) \) belongs to \( C_{WF}^{2+\gamma}(\bar{U}(\delta^0, x^0)) \) if, in the new system of coordinates,

1. \( \hat{h} \) is continuously differentiable on \( U(\delta^0, x^0) \) and \( \hat{h} \) and its derivatives extend continuously to \( \bar{U}(\delta^0, x^0) \) and the resulting extensions belong to \( C_{WF}^\gamma(\bar{U}(\delta^0, x^0)) \);
2. \( \hat{h} \) is twice continuously differentiable on \( U_+(\delta^0, x^0) = U(\delta^0, x^0) \cap \{ (y_1, ..., y_{d-1}) \in (\mathbb{R}^d)_d : \forall i \in I, y_i > 0 \} \). Moreover

\[
\lim_{\min(y_i, y'_{i}) \to 0^+} \sqrt{y_i y'_i} \partial^2_{y_i y'_i} \hat{h}(y) = 0, \quad \lim_{y_i \to 0^+} \sqrt{y_i} \partial^2_{y_i} \hat{h}(y) = 0,
\]

and the functions \( y \to \sqrt{y_i y'_i} \partial^2_{y_i y'_i} \hat{h}(y) \) and \( y \to \sqrt{y_i} \partial^2_{y_i} \hat{h}(y) \) belong to \( C_{WF}^{2+\gamma}(\bar{U}(\delta^0, x^0)) \) (meaning in particular that they can be extended by continuity to \( \bar{U}(\delta^0, x^0) \)).

We then let

\[
||\hat{h}||_{2+\gamma, U(\delta^0, x^0)} := ||\hat{h}||_{\gamma, U(\delta^0, x^0)} + \sum_{i \in [d-1]} ||\partial^2_{y_i} \hat{h}||_{\gamma, U(\delta^0, x^0)} + \sum_{i, j \in I} ||\sqrt{y_i y'_j} \partial^2_{y_i y'_j} \hat{h}||_{\gamma, U(\delta^0, x^0)}
\]

\[
+ \sum_{k, l \in I} ||\partial^2_{y_k y'_l} \hat{h}||_{\gamma, U(\delta^0, x^0)} + \sum_{i \in [d-1]} \sum_{k, l \in I} ||\sqrt{y_i y'_k} \partial^2_{y_i y'_k} \hat{h}||_{\gamma, U(\delta^0, x^0)},
\]

where \( \sqrt{y_i y'_j} \partial^2_{y_i y'_j} \hat{h} \) is a shorten notation for \( y \to \sqrt{y_i y'_j} \partial^2_{y_i y'_j} \hat{h}(y) \) (and similarly for the others). For a given finite covering \( \bigcup_{i=1}^K U(\delta^0, x^0, \cdot, \cdot) \) of \( S_d \), which is then fixed in the rest of the discussion, a function \( h \) is said to be in \( C_{WF}^{2+\gamma}(S_d) \), respectively in \( C_{WF}^{2+\gamma}(S_d) \) if \( h \) belongs to each \( C_{WF}^{2+\gamma}(U(\delta^0, x^0, \cdot, \cdot)) \), respectively each \( C_{WF}^{2+\gamma}(U(\delta^0, x^0, \cdot, \cdot)) \). Equivalently, we write \( h \in C_{WF}^{2+\gamma}(S_d) \) (respectively \( h \in C_{WF}^{2+\gamma}(S_d) \)), for a function \( h \) defined on \( S_d \), if the associated function \( \hat{h} \) defined on \( S_d \) belongs to \( C_{WF}^{2+\gamma}(S_d) \) (respectively \( C_{WF}^{2+\gamma}(S_d) \)). We then let

\[
||\hat{h}||_{WF, \gamma} := \sum_{i=1}^K ||\hat{h}||_{2+\gamma, U(\delta^0, x^0, \cdot, \cdot)} \quad ||\hat{h}||_{WF, 2+\gamma} := \sum_{i=1}^K ||\hat{h}||_{2+\gamma, U(\delta^0, x^0, \cdot, \cdot)}.
\]

We refer to [44, Chapter 10] and to [43, Subsection 3.2] for more details.

**Parabolic Wright-Fisher spaces**

Similar definitions hold for the spaces \( C_{WF}^{\gamma}(0, T \times S_d) \) and \( C_{WF}^{2+\gamma}(0, T \times S_d) \). They are respectively spaces of time-space functions that are \( \gamma \)-Hölder continuous functions and spaces of time-space functions that are continuously differentiable in time and twice continuously differentiable in space, with derivatives that are locally \( \gamma \)-Hölder continuous, Hölder continuity being understood in both cases with respect to the time-space distance (in the local system of coordinates).
We notice that, in Subsection 2.2, the spaces 

\[ C^2_\text{WF} \] 

and the functions 

\[ C^2 \] 

are said to be hybrid since the functions therein have mixed classical and Wright-Fisher regularity. Following [44, Lemma 5.2.7], we say that a function \( \hat{h} \) defined on \([0, T] \times \mathcal{U}(\delta^0, x^0)\) belongs to the space \( C^\gamma_{\text{WF}}([0, T] \times \mathcal{U}(\delta^0, x^0))\) if, in the new system of coordinates, 

1. \( \hat{h} \) is continuously differentiable on \([0, T] \times \mathcal{U}(\delta^0, x^0)\) and \( \hat{h} \) and its time and space derivatives extend continuously to \([0, T] \times \mathcal{U}(\delta^0, x^0)\) and the resulting extensions belong to \( C^\gamma_{\text{WF}}([0, T] \times \mathcal{U}(\delta^0, x^0))\); 

2. \( \hat{h} \) is twice continuously differentiable in space on \([0, T] \times \mathcal{U}(\delta^0, x^0)\). Moreover, for any \( i, j \in I \) and any \( k, l \not\in I \), 

\[
\lim_{\min(y, y) \to 0^+} \sqrt{\gamma_{i \cap j}} \partial^2_{i \cap j \cap k \cap l} \hat{h}(t, y) = 0, \quad \lim_{y \to 0^+} \sqrt{\gamma_{i \cap j}} \partial^2_{i \cap j \cap k \cap l} \hat{h}(t, y) = 0, \tag{A.6}
\]

and the functions 

\[
(t, y) \to \sqrt{\gamma_{i \cap j}} \partial^2_{i \cap j \cap k \cap l} \hat{h}(t, y), \quad (t, y) \to \sqrt{\gamma_{i \cap j}} \partial^2_{i \cap j \cap k \cap l} \hat{h}(y) \quad \text{and} \quad (t, y) \to \partial^2_{i \cap j \cap k \cap l} \hat{h}(t, y)
\]

belong to \( C^\gamma_{\text{WF}}([0, T] \times \mathcal{U}(\delta^0, x^0))\).

We then let 

\[
|\hat{h}|_{2+\gamma;[0, T] \times \mathcal{U}(\delta^0, x^0)} := |\hat{h}|_{[0, T] \times \mathcal{U}(\delta^0, x^0)} + |\partial_t \hat{h}|_{[0, T] \times \mathcal{U}(\delta^0, x^0)} + \sum_{i=1}^d |\partial_i \hat{h}|_{[0, T] \times \mathcal{U}(\delta^0, x^0)} + \sum_{i=1}^d |\partial^2_{i \cap j \cap k \cap l} \hat{h}|_{[0, T] \times \mathcal{U}(\delta^0, x^0)} + \sum_{i=1}^d |\partial^2_{i \cap j \cap k \cap l} \hat{h}|_{[0, T] \times \mathcal{U}(\delta^0, x^0)} + \sum_{i=1}^d |\partial^2_{i \cap j \cap k \cap l} \hat{h}|_{[0, T] \times \mathcal{U}(\delta^0, x^0)}.
\]

For the fixed covering \( \bigcup_{i=1}^K \mathcal{U}(\delta^0, x^0, i) \) of \( \hat{S}_d \), a function \( \hat{h} \) is said to be in \( C^\gamma_{\text{WF}}([0, T] \times \hat{S}_d) \), respectively in \( C^\gamma_{\text{WF}}([0, T] \times \hat{S}_d) \) (as before, the definition extends equivalently to the associated function \( h \) defined on \([0, T] \times S_d \)), if \( \hat{h} \) belongs to each \( C^{\gamma'}_{\text{WF}}([0, T] \times \mathcal{U}(\delta^0, x^0, i)) \), respectively each \( C^{\gamma'}_{\text{WF}}([0, T] \times \mathcal{U}(\delta^0, x^0, i)) \). We then let 

\[
|\hat{h}|_{\text{WF}, \gamma} := \sum_{i=1}^K |\hat{h}|_{[0, T] \times \mathcal{U}(\delta^0, x^0, i)}, \quad |\hat{h}|_{\text{WF}, 2+\gamma} := \sum_{i=1}^K |\hat{h}|_{2+\gamma;[0, T] \times \mathcal{U}(\delta^0, x^0, i)}.
\]

Hybrid spaces

We notice that, in Subsection 2.2, the spaces \( C^\gamma_{\text{WF}}(S_d) \) and \( C^{\gamma'}_{\text{WF}}(S_d) \) are denoted \( C^\gamma_{\text{WF}}(S_d) \) and \( C^{\gamma'}_{\text{WF}}(S_d) \), with a ‘0’ in superscript and without a ‘hat’ on \( S_d \), and similarly for the two norms \( |\hat{h}|_{\text{WF}, \gamma} \) and \( |\hat{h}|_{\text{WF}, 2+\gamma} \), which are written \( |h|_{\text{WF}, 0, \gamma} \) and \( |h|_{\text{WF}, 0, 2+\gamma} \) where \( h : S_d \to \mathbb{R} \) is canonically associated with \( \hat{h} : \hat{S}_d \to \mathbb{R} \). Similar notations, with the additional index ‘0’, are used for the parabolic spaces in the core of the text. This additional ‘0’ permits to distinguish from spaces made of functions whose derivatives belong to either \( C^\gamma_{\text{WF}}(S_d) \) or \( C^{\gamma'}_{\text{WF}}(S_d) \). Those spaces are said to be hybrid since the functions therein have mixed classical and Wright-Fisher regularity. Again, this notion is directly borrowed from [44, Chapter 5]. More precisely, a function \( h \), defined on \( S_d \) belongs to \( C^\gamma_{\text{WF}}(S_d) \) (respectively \( C^{\gamma'}_{\text{WF}}(S_d) \)), for some \( \gamma \in (0, 1) \), if it is continuously differentiable on \( S_d \) (meaning that it is continuously differentiable on the interior and the
derivatives extend by continuity up to the boundary) and each $d_i h$, for $i \in [d]$, belongs to $C^{0,\gamma}_{WF}(S_d)$ (respectively $C^{0,2+\gamma}_{WF}(S_d)$). For $h \in C^{1,\gamma}_{WF}(S_d)$, we then let
$$ ||h||_{WF,1,\gamma} := ||h||_{\infty} + \sum_{i \in [d]} ||d_i h||_{WF,0,\gamma}, $$
and, for $h \in C^{1,2+\gamma}_{WF}(S_d)$, we let
$$ ||h||_{WF,1,2+\gamma} := ||h||_{\infty} + \sum_{i \in [d]} ||d_i h||_{WF,0,2+\gamma}. $$

The parabolic version of $C^{1,2+\gamma}_{WF}(S_d)$ (which is the only one we need in the text) is defined in a similar way. A function $h$, defined on $[0,T] \times S_d$, belongs to $C^{0,2+\gamma}_{WF}([0,T] \times S_d)$, for some $\gamma \in (0,1)$, if it belongs to $C^{0,2+\gamma}_{WF}([0,T] \times S_d)$ (and is hence differentiable in space) and each $d_i h$, for $i \in [d]$, belongs to $C^{0,2+\gamma}_{WF}([0,T] \times S_d)$. For $h \in C^{1,2+\gamma}_{WF}([0,T] \times S_d)$, we then let
$$ ||h||_{WF,1,2+\gamma} := ||h||_{WF,0,\gamma} + ||\partial_t h||_{WF,0,\gamma} + \sum_{i \in [d]} ||d_i h||_{WF,0,2+\gamma}. $$