Finite state Mean Field Games with Wright-Fisher common noise

Erhan Bayraktar, Alekos Cecchin, Asaf Cohen, François Delarue

Version of Record: https://www.sciencedirect.com/science/article/pii/S002178242100012X
Manuscript_2f010936b5f2c3e9b672f971f84d80ff

Finite state Mean Field Games with Wright–Fisher common noise

Erhan Bayraktar[a,1], Alekos Cecchin[b,2], Asaf Cohen[a,3], François Delarue [b,4,*]

[a] *Department of Mathematics, University of Michigan, Ann Arbor, Michigan 48109, United States*
[b] *Université Côte d'Azur, CNRS, Laboratoire J.A. Dieudonné, 06108 Nice France*

Abstract

We force uniqueness in finite state mean field games by adding a Wright–Fisher common noise. We achieve this by analyzing the master equation of this game, which is a degenerate parabolic second-order partial differential equation set on the simplex whose characteristics solve the stochastic forward-backward system associated with the mean field game; see Cardaliaguet et al. [10]. We show that this equation, which is a non-linear version of the Kimura type equation studied in Epstein and Mazzeo [28], has a unique smooth solution whenever the normal component of the drift at the boundary is strong enough. Among others, this requires a priori estimates of Hölder type for the corresponding Kimura operator when the drift therein is merely continuous.

Résumé

Nous imposons l'unicité à des jeux à champ moyen sur des espaces d'état finis en forçant la dynamique sous-jacente par un bruit commun de type Wright–Fisher. Notre approche s'appuie sur l'analyse de l'équation maîtresse correspondante ; celle-ci s'écrit comme une équation aux dérivées partielles dégénérée sur le simplexe, dont les caractéristiques sont précisément les solutions du système progressif-rétrograde associé au jeu à champ moyen, voir par exemple Cardaliaguet et al. [10]. Nous démontrons que cette équation, qui s'apparente ici à une version non-linéaire des équations de Kimura étudiées dans Epstein and Mazzeo [28], a une unique solution classique lorsque la composante normale du terme de transport est suffisamment forte au bord. Un point essentiel dans notre preuve est l'obtention, pour de tels opérateurs de Kimura, d'estimations Hölder valables lorsque le coefficient de transport n'est que continu.

Keywords: Mean-field games, master equation, Kimura operator, non-linear PDEs, forcing uniqueness, common noise, Wright–Fisher diffusion.
2010 MSC: 91A13, 91A15, 35K65.

*Corresponding author
Email addresses: erhan@umich.edu (Erhan Bayraktar), alekos.Cecchin@unice.fr (Alekos Cecchin), asafc@umich.edu (Asaf Cohen), francois.delarue@unice.fr (François Delarue)

[1] E. Bayraktar is partially supported by the National Science Foundation and by the Susan M. Smith chair.
[2] A. Cecchin acknowledges the financial support of French ANR project ANR-16-CE40-0015-01 on "Mean Field Games".
[3] A. Cohen acknowledges the financial support of Research supported by the National Science Foundation (DMS-2006305).
[4] F. Delarue thanks Institut Universitaire de France, French ANR project ANR-16-CE40-0015-01 on "Mean Field

1. Introduction

Aiming at forcing uniqueness in the theory of mean field games (MFGs), a more complete account of which we provide below, we analyze here a system of parabolic partial differential equations (PDEs) with the main feature of being set on the space of probability measures on $[\![d]\!] := \{1, \cdots, d\}$, the latter being referred to as the $(d-1)$-dimensional simplex, for a fixed integer $d \geq 1$. This system is indexed by the elements i of $[\![d]\!]$ itself and has the following generic form:

$$\begin{aligned} &\partial_t U^i - \frac{1}{2}\sum_{j=1}^d (U^i - U^j)_+^2 + f^i(t,p) + \sum_{j=1}^d \varphi(p_j)[U^j - U^i] \\ &\quad + \sum_{1\leq j,k\leq d} p_k[\varphi(p_j) + (U^k - U^j)_+]\left(\partial_{p_j} U^i - \partial_{p_k} U^i\right) \\ &\quad + \varepsilon^2 \sum_{j=1}^d p_j\left(\partial_{p_i} U^i - \partial_{p_j} U^i\right) + \frac{\varepsilon^2}{2}\sum_{1\leq j,k\leq d}(p_j\delta_{jk} - p_j p_k)\partial^2_{p_j p_k} U^i = 0, \\ &U^i(T,p) = g^i(p), \end{aligned} \tag{1.1}$$

for $i \in [\![d]\!]$, where $(t,p) \in [0,T]\times\mathcal{S}_{d-1}$ and $\mathcal{S}_{d-1}$ is the $(d-1)$-dimensional simplex. Whenever ε is equal to zero and φ is also identically equal to zero, this system is the so-called *master equation* that describes the values of the equilibria in a (finite state) MFG driven by a simple continuous time Markov decision process on $[\![d]\!]$ and by the functions $(f^i)_{i\in[\![d]\!]}$ and $(g^i)_{i\in[\![d]\!]}$ as respective running and terminal costs. To wit, the first line in (1.1), which has a form similar to a Hamilton-Jacobi equation on $[\![d]\!]$, accounts for the optimization problem in the underyling MFG, whilst the second line accounts for the dynamics of the equilibria. Although we provide a longer review on MFG later in the text, we feel useful to quote, at this early stage of the introduction, [34, 35, 48] and [11, Chapter 7] as references on the master equation for finite state mean field games, and [9, 7, 12, 10, 18, 49, 50] as references for continuous state mean field games. The main novelty here is the third line in (1.1). Therein, ε is a (strictly) positive viscosity parameter which we call the *intensity of the common noise*. This terminology comes from the fact that equation (1.1) is associated with a new form of MFG, which we are going to describe later, in which equilibria are no longer deterministic but are subjected to a so-called *common noise* and are hence randomized. Under the action of the common noise, the master equation becomes a system of second order PDEs, the principal part of which is the second-order operator in the third line of (1.1) and is called a Kimura operator on the simplex (see [28, 43]). Accordingly, the master equation here reads as a system of non-linear parabolic PDEs of Kimura type. As for the additional function φ in the first two lines of (1.1), it should be understood as a forcing term in the dynamics of the equilibria that allow the latter to escape for free from the boundary of the simplex. In this context, one of our contributions (see Meta-Theorem 2.9 and Theorem 3.4) is to show that, when ε is strictly positive and the functions $(f^i)_{i\in[\![d]\!]}$ and $(g^i)_{i\in[\![d]\!]}$ satisfy some smoothness conditions, we can choose φ large enough in the neighborhood of the boundary of $\mathcal{S}_{d-1}$ and null everywhere else in such a way that (1.1) has a unique smooth solution (in a so-called Wright–Fischer Hölder space of functions that are once differentiable in time and twice in space with a suitable behavior

Games" and French ANR projet ANR-19-P3IA-0002 "3IA Côte d'Azur - Nice - Interdisciplinary Institute for Artificial Intelligence".

at the boundary of the simplex). Accordingly, our main result is that the corresponding MFG is uniquely solvable for a prescribed initial condition (see Meta-Theorem 2.7 and Theorem 3.2). Importantly, there are many examples for which the latter is false when $\varepsilon = 0$, which explains why we refer quite often to the concept of "forcing uniqueness".

A general framework to analyze linear *Kimura PDEs* was introduced by Epstein and Mazzeo in [28] and this framework was extended subsequently in [29, 30, 53, 54]. Generally speaking, the analysis of Kimura PDEs suffers from two main difficulties: (1) the simplex boundary is not smooth, and (2) the PDE degenerates at the boundary. Despite these difficulties, the authors of [28] were able to prove the existence and uniqueness of smooth solutions to linear Kimura PDEs under enough regularity of the coefficients. However, these results do not apply to (1.1) because the coefficients therein are time-dependent (Kimura operators are assumed to be time-homogeneous in [28]) and, most of all, because the equation is non-linear. While the additional time dependence can be handled with relative ease (see Lemma 4.4), the non-linearity requires a sophisticated analysis, which, in fact, is the main technical part of this paper. In this respect, the main step in our study is Theorem 2.10, which provides an a priori Hölder estimate to solutions of linear Kimura PDEs when driven by merely continuous drift terms that point inward the simplex in a sufficiently strong manner, whence our need for the additional φ in equation (1.1). The proof of this a priori estimate uses a tailor-made *coupling by reflection* argument inspired by earlier works on couplings for multidimensional processes (see e.g., [20]). However, the coupling by itself, as usually implemented in the literature for proving various types of smoothing effects for diffusion processes, is in fact not enough for our purpose. We indeed pay a price for the degeneracy of the equation at the boundary and, similar to other works on Kimura operators (see for instance [2]), we need to perform an induction over the dimension to handle the degeneracy properly; see Proposition 5.6 for the details of the induction property. Once we reach this point, the proof of existence of a solution to (1.1) is straightforward, provided that φ therein is chosen in a relevant way, and uses Schauder's fixed point theorem on the proper Wright–Fisher[5] space, as well as Schauder's estimates derived in [28] for the linear equation and Lemma 4.4 mentioned earlier (see Theorem 3.4).

Let us now clarify our technical contribution into the context of MFGs. MFGs were introduced in the seminal works of Lasry and Lions [44, 45, 46], and Huang, Malhamé, and Caines [41, 40]. Merging intuition from statistical physics and classical game theory, this paradigm provides the asymptotic behavior of many weakly interacting strategic players who are in a Nash equilibrium. Formally, this asymptotic equilibrium is described as the fixed point of a best response map, which sends a given flow of measures to the distribution of a controlled state-dynamics. For recent theoretical developments and applications of this theory, we refer the reader to [1, 6, 9, 11, 12] and the references therein. MFGs with (a fixed and) finite number of states were analyzed by [36, 37, 39, 48]; for a probabilistic approach to finite state MFGs we refer to [13, 16].

Typically, MFGs do not admit unique solutions. Two known instances of uniqueness are the small T case and the so-called monotonous case due to Lasry and Lions, see [44, Section 4] for the latter. The thrust of our paper is to establish uniqueness by adding a common noise[6] that emerges from the limiting behavior of Wright–Fisher population-genetics models. The special structure of the common noise we use leads to stochastic dynamics evolving inside the multidi-

[5]Most of the time, we just say Wright–Fisher space instead of Wright–Fisher Hölder space.

[6]Recently, Bertucci, Lasry, and Lions [8] mentioned that "The addition of a common noise in the MFG setting remains one of the most important questions in the MFG theory." We feel that our paper may be one step forward in this direction.

mensional simplex $\mathcal{S}_{d-1}$ and eventually to the second order form of (1.1). Besides the forcing uniqueness result, this is another interest of our work to incorporate population-genetics models into MFGs; to the best of our knowledge, this is a new feature in the field. In this regard, it is worth pointing out that, even though we do not speak about it in this paper, it is in fact possible to explain the common noise at the level of a particle system by a diffusion approximation. We refer to our companion work [3] for a complete overview.

In fact, we must stress that the recent work [8] (to which we already alluded in the footnote 6) also addresses a form of common noise for finite state MFGs. As explained therein, the key point in this direction is to force the finite-player system to have many simultaneous jumps at some random times prescribed by the common noise. Although we share a similar idea in our construction, our common noise structure is in the end different from [8]: While the simultaneous jumps in [8] are governed by a deterministic transformation of the state space, they here obey a resampling procedure that is typical, as we have just said, of population-genetics models. Moreover, one of the questions in [8] is to decide whether the solution preserves monotonicity of the coefficients; in this regard, forcing uniqueness (outside the monotonous setting) is not discussed in [8]. In fact, forcing uniqueness for MFGs was addressed in other works but in different settings. Recently, Delarue [23] established a forcing uniqueness result for a continuous state MFG obtained by forcing a deterministic (meaning that the players follow ordinary differential equations) MFG by means of a common noise. In this case, the common noise is infinite-dimensional and henceforth differs from the most frequent instance of common noise used in the literature, since the latter has very often a finite dimension, see e.g., [12]. At this point, it is worth mentioning that forcing uniqueness is studied in [57] under the action of a standard finite-dimensional common noise, but for a linear quadratic MFG. This is due to the fact that the equilibrium distribution in that paper is Gaussian and is parametrized by its mean and variance, which reduces the dimension of the problem. On a more prospective level, forcing uniqueness by common noise might enable a selection criterion by taking small noise limit for cases where the limiting problem does not have a unique equilibrium. This question was addressed in some specific cases in [24] for a continuous state model and in [14] for a finite state model (see also [5]) and is the purpose of the forthcoming work [15] in a more general setting (with a finite state space).

Once the master equation (1.1) has been solved, the equilibrium distribution of the mean field game, which becomes random under the action of the common noise, is provided by the solution of the forward component of the following *stochastic mean field game system*, given by the forward-backward stochastic differential equation

$$\begin{aligned} dP_t^i &= \sum_{j=1}^{d}\Big(P_t^j(\varphi(P_t^i) + (u_t^j - u_t^i)_+) - P_t^i(\varphi(P_t^j) + (u_t^i - u_t^j)_+)\Big)\mathrm{d}t \\ &\quad + \frac{\varepsilon}{\sqrt{2}}\sum_{j=1}^{d}\sqrt{P_t^iP_t^j}d[W_t^{i,j} - W_t^{j,i}], \\ du_t^i &= -\Big(\sum_{j=1}^{d}\varphi(P_t^j)[u_t^j - u_t^i] - \frac{1}{2}\sum_{j=1}^{d}(u_t^i - u_t^j)_+^2 + f^i(P_t)\Big)\mathrm{d}t \\ &\quad - \frac{\varepsilon}{\sqrt{2}}\sum_{j=1}^{d}\sqrt{\frac{P_t^j}{P_t^i}}(\nu_t^{i,i,j} - \nu_t^{i,j,i})\mathrm{d}t + \sum_{1\le j\neq k\le d}\nu_t^{i,j,k}\mathrm{d}W_t^{j,k}, \end{aligned} \tag{1.2}$$

with a given initial deterministic probability vector $(P_0^i = p_{0,i})_{i\in[\![d]\!]}$ for the forward equation and

the terminal condition $(u_T^i = g^i(P_T))_{i\in[\![d]\!]}$ for the backward equation. The process $((W_t^{i,j})_{i,j\in[\![d]\!]:i\neq j})_{0\le t\le T}$ is a Wiener process and, as usual in the theory of backward stochastic differential equations, the role of the processes $(((\nu_t^{i,j,k}))_{0\le t\le T})_{i,j,k\in[\![d]\!]:j\neq k}$ is to force the solution $((u_t^i)_{0\le t\le T})_{i\in[\![d]\!]}$ to be non-anticipating. The process $((P_t^i)_{i\in[\![d]\!]})_{0\le t\le T}$ is a Wright–Fisher diffusion process (taking values in the $(d-1)$-dimensional simplex), see [32, 31, 55]; accordingly, the forward equation in (1.2) must be interpreted as a stochastic Fokker-Planck equation on $\mathcal{S}_{d-1}$. The process $((u_t^i)_{i\in[\![d]\!]})_{0\le t\le T}$ stands for the game-value for the representative player and the system of equations that it solves in (1.2) must be read as a (backward) stochastic Hamilton–Jacobi–Bellman equation on $\mathcal{S}_{d-1}$. The connection between (1.1) and (1.2) is given by the following relationship.

$$u_t^i = U^i(t, P_t), \qquad \text{and} \qquad \nu_t^{i,j,k} = V^{i,j,k}(t, P_t),$$

where

$$V^{i,j,k}(t,p) = \frac{\varepsilon}{\sqrt{2}}\left(\partial_{p_j} U^i(t,p) - \partial_{p_k} U^i(t,p)\right)\sqrt{p_j p_k}.$$

In fact, this relationship is the cornerstone to prove uniqueness of the solution of (1.1) through a verification argument, see Theorem 4.3. This argument is inspired from the original four-step-scheme in [52]; in the framework of continuous state MFGs, it has already been used in [12, 10, 18]. On a more elaborated level, we should point out that the master equation has been a key tool (e.g., see [10] for continuous state mean field games and [4, 17] for finite state mean field games) to show convergence of the closed loop Nash-equilibrium of the N-player system to the MFG equilibrium. In both [4, 17], there is no common noise and, in all the latter three cases, the Lasry–Lions monotonicity condition is assumed to be force, which is obviously in stark contrast to the setting of the current paper. The convergence problem in our setup is thus an interesting question, which we resolve in our aforementioned work [3].

The rest of the paper is organized as follows. In Section 2 we introduce the MFG model and provide preliminary versions of the main MFG and PDE results of the paper. The first one (Meta-Theorem 2.7) states that the MFG with common noise admits a unique solution; the second one (Meta-Theorem 2.9) states that the related master equation (1.1) has a unique smooth solution; the last one (Theorem 2.10) is an a priori Hölder regularity estimate for linear PDEs driven by a merely continuous drift and a Kimura operator. In Section 3 we provide more material for the analysis of PDEs set on the simplex. This allows us to formulate more rigorous versions of Meta-Theorems 2.7 and 2.9, see Theorems 3.2 and 3.4. Section 4 is dedicated to the proofs of Theorems 3.2 and 3.4, taken for granted the a priori estimate from Theorem 2.10. A key point therein is to make the connection between the master equation (1.1) and the MFG forward-backward system (1.2). The proof of Theorem 2.10 is the most demanding one of the paper and is given in Section 5. The main two ingredients in the proof are the coupling construction provided in Proposition 5.3 and the induction step in Proposition 5.6.

We now provide frequently used notation.

Notation. For $a, b \in \mathbb{R}$, we let $a \wedge b := \min\{a, b\}$. We use the notation $M^\dagger$ to denote the transpose of a matrix M. Moreover, we use the generic notation $p = (p_i)_{i\in[\![d]\!]}$ (with p in lower case and i in subscript) for elements of $\mathbb{R}^d$, while processes are usually denoted by $\boldsymbol{P} = ((P_t^i)_{i=1,\dots,d})_{0\le t\le T}$ (with P in upper case and i in superscript). For a subset A of a Euclidean space, we denote by $\mathrm{Int}(A)$ the interior of A. Also, we recall the notation $\mathcal{S}_{d-1} := \{(p_1, \cdots, p_d) \in (\mathbb{R}_+)^d : \sum_{i\in[\![d]\!]} p_i = 1\}$, where $[\![d]\!] := \{1, \dots, d\}$. We can identify $\mathcal{S}_{d-1}$ with the convex polyhedron of $\mathbb{R}^{d-1}$ $\hat{\mathcal{S}}_{d-1} := \{(x_1, \cdots, x_{d-1}) \in (\mathbb{R}_+)^{d-1} : \sum_{i\in[\![d-1]\!]} x_i \le 1\}$. In particular, we sometimes write "the interior" of $\mathcal{S}_{d-1}$; in such a case, we implicitly consider the interior of $\mathcal{S}_{d-1}$ as the $(d-1)$-dimensional interior

of $\hat{\mathcal{S}}_{d-1}$. Obviously, the interior of $\mathcal{S}_{d-1}$, when regarded as a subset of $\mathbb{R}^d$, is empty, which makes it of a little interest. To make it clear, for some $p \in \mathcal{S}_{d-1}$, we sometimes write $p \in \mathrm{Int}(\hat{\mathcal{S}}_{d-1})$, meaning that $p_i > 0$ for any $i \in [\![d]\!]$. We use the same convention when speaking about the boundary of $\mathcal{S}_{d-1}$: For some $p \in \mathcal{S}_{d-1}$, we may write $p \in \partial\hat{\mathcal{S}}_{d-1}$ to say that $p_i = 0$ for some $i \in [\![d]\!]$. For $p \in \mathcal{S}_{d-1}$, we write $\sqrt{p}$ for the vector $(\sqrt{p_1}, \cdots, \sqrt{p_d})$. Finally, $\delta_{i,j}$ is the Kronecker symbol and r_+ denotes the positive part of $r \in \mathbb{R}$.

2. Model and preliminary versions of the main results

The purpose of this section is to introduce step by step the model and then to provide preliminary versions of our main results. Although not definitive, those versions should help the reader to have a quick overview of the content of the paper. More complete statements are given in the next section. As we already accounted for in the introduction, our general objective is to prove that a relevant form of common noise may force uniqueness of equilibria to mean field games on a finite state space, see Meta-Theorem 2.7. Our approach relies on the analysis of the master equation (1.1), whose solvability is addressed by means of a suitable smoothing property for so-called Kimura operators. We refer to Subsection 2.3 for an outlook on our PDE results. The complete description of Kimura operators is postponed to Section 3.

2.1. A preliminary version of the mean field game

The first point that we need to clarify is the form of the mean field game itself. Whilst it is absolutely standard when there is no common noise, the mean field game addressed below takes indeed a more intricate and less obvious form in the presence of common noise. In fact, the somewhat non-classical structure that we use throughout the paper is specifically designed in order to be in correspondence with the class of second order differential operators on the simplex, referred to as Kimura operators in the text, for which we can indeed prove the smoothing results announced in introduction, see Subsections 2.3 for a first account and 3.2 for more details.

Clearly, the sharpest way to derive the form of mean field games that is used below would consist in going back to a game with a large but finite number N of players and in justifying that, under the limit $N \to \infty$, this finite game converges in some sense to our form of mean field games. Instead, we directly write down our version of mean field game with a common noise on a finite state space. Our rationale for doing so is that it allows the reader to jump quickly into the article. If she or he is interested, she or he may have a look at our companion paper [3], in which the discrete model is described in depth and the convergence problem is entirely resolved.

2.1.1. Mean field game without a common noise

When there is no common noise, our form of mean field game is directly taken from the earlier works [36, 37]. In short, a given tagged player evolves according to a Markov process with values in a finite state space E, which we will take for convenience as $E = [\![d]\!] := \{1, \cdots, d\}$. At any time $t \in [0, T]$, for a finite time horizon $T > 0$, she chooses her transition rates in the form of a time-measurable $d \times d$–matrix $(\beta_t^{i,j})_{i,j \in [\![d]\!]}$ satisfying

$$\beta_t^{i,j} \geq 0, \quad i \neq j, \quad \beta_t^{i,i} = -\sum_{j \neq i} \beta_t^{i,j}, \quad t \in [0, T]. \tag{2.1}$$

Given the rates $((\beta_t^{i,j})_{i,j\in[\![d]\!]})_{0\le t\le T}$, the marginal distribution $((Q_t^i)_{i\in[\![d]\!]})_{0\le t\le T}$ of the states of the tagged player evolves according to the discrete Fokker–Planck (or Kolmogorov) equation:

$$\frac{\mathrm{d}}{\mathrm{d}t}Q_t^i = \sum_{j\in[\![d]\!]} Q_t^j \beta_t^{j,i}, \quad t\in[0,T], \tag{2.2}$$

the initial statistical state $(Q_0^i)_{i\in[\![d]\!]}$ being prescribed as an element of $\mathcal{S}_{d-1}$. In words, Q_t^i is the probability that the tagged player be in state i at time t.

With the tagged player, we assign a cost functional depending on a deterministic time-measurable $\mathcal{S}_{d-1}$-valued path $(P_t)_{0\le t\le T}$, referred to as an environment and starting from the same initial state as $(Q_t)_{0\le t\le T}$, namely $Q_0^i = P_0^i$ for $i\in[\![d]\!]$. Intuitively, P_t is understood as the statistical state at time t of all the *other* players in the continuum, which are basically assumed to be independent and identically distributed. Given $(P_t)_{0\le t\le T}$, the cost to the tagged player is written in the form

$$\mathcal{J}((\beta_t)_{0\le t\le T},(P_t)_{0\le t\le T}) := \sum_{i\in[\![d]\!]}\Big[Q_T^i g(i,P_T) + \int_0^T Q_t^i\Big(f(t,i,P_t) + \frac{1}{2}\sum_{j\neq i}\big|\beta_t^{i,j}\big|^2\Big)\mathrm{d}t\Big], \tag{2.3}$$

where g is a function from $[\![d]\!]\times\mathcal{S}_{d-1}$ into $\mathbb{R}$ and f is a function from $[0,T]\times[\![d]\!]\times\mathcal{S}_{d-1}$ into $\mathbb{R}$. To simplify the notations, we will sometimes write $\boldsymbol{\beta}$ for $(\beta_t)_{0\le t\le T}$ and $\boldsymbol{P}$ for $(P_t)_{0\le t\le T}$. Accordingly, we will write $\mathcal{J}(\boldsymbol{\beta},\boldsymbol{P})$ for the cost to the tagged player.

In this setting, a mean field equilibrium is a path $\boldsymbol{P} = (P_t)_{0\le t\le T}$ as before for which we can find an optimal control $(\beta_t^\star)_{0\le t\le T}$ to $\mathcal{J}(\cdot,\boldsymbol{P})$ such that the corresponding solution to (2.2) is $(P_t)_{0\le t\le T}$ itself. We stress the fact that here $\boldsymbol{P}$ and $\boldsymbol{Q}$ are deterministic paths.

2.1.2. Stochastic Fokker–Planck equation

We now introduce a special form of common noise in order to force the equilibria to satisfy a relevant form of diffusion processes with values in the simplex $\mathcal{S}_{d-1}$. To make it clear, our aim is to force equilibria to satisfy the following stochastic variant of equation (2.2):

$$\mathrm{d}P_t^i = \sum_{j\in[\![d]\!]} P_t^j \alpha_t^{j,i}\mathrm{d}t + \frac{\varepsilon}{\sqrt{2}}\sum_{j\in[\![d]\!]}\sqrt{P_t^iP_t^j}\,\mathrm{d}[W_t^{i,j}-W_t^{j,i}], \tag{2.4}$$

for $t\in[0,T]$, where $((W_t^{i,j})_{0\le t\le T})_{i,j\in[\![d]\!]:i\neq j}$ is a collection of independent $1d$ Brownian motions, referred to as the common noise, and $\boldsymbol{\alpha} = (\alpha_t)_{0\le t\le T}$ is a progressively measurable process (with respect to the augmented filtration $\mathbb{F}^{\boldsymbol{W}} = (\mathcal{F}_t^{\boldsymbol{W}})_{0\le t\le T}$ generated by $\boldsymbol{W} = ((W_t^{i,j})_{i,j\in[\![d]\!]:i\neq j})_{0\le t\le T}$) satisfying (2.1). All these processes are constructed on some probability space $(\Omega,\mathcal{A},\mathbb{P})$. Throughout, we use the convention $\boldsymbol{W}^{i,i} = (W_t^{i,i})_{0\le t\le T}\equiv 0$, for any $i\in[\![d]\!]$. Above, the parameter ε reads as the intensity of the common noise. Accordingly, the collection $((\overline{W}_t^{i,j} := (W_t^{i,j} - W_t^{j,i})/\sqrt{2})_{0\le t\le T})_{i,j\in[\![d]\!]:i\neq j}$ forms an antisymmetric Brownian motion.

Although it looks rather unusual, the form of the stochastic integration in (2.4) is in fact directly inspired by stochastic models of population genetics. To wit, for $i,j\in[\![d]\!]$, the (i,j)-bracket writes (with a somewhat abusive but quite useful notation in the first term in the right-hand side below)

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\langle P^i,P^j\rangle_t &= \Big\langle\frac{\varepsilon}{\sqrt{2}}\sum_{k\in[\![d]\!]}\sqrt{P_t^iP_t^k}\,\mathrm{d}[W_t^{i,k}-W_t^{k,i}],\frac{\varepsilon}{\sqrt{2}}\sum_{l\in[\![d]\!]}\sqrt{P_t^jP_t^l}\,\mathrm{d}[W_t^{j,l}-W_t^{l,j}]\Big\rangle\\ &= \varepsilon^2\sum_{k,l\in[\![d]\!]}\sqrt{P_t^iP_t^jP_t^kP_t^l}\big(\delta_{i,j}\delta_{k,l}-\delta_{i,l}\delta_{k,j}\big) = \varepsilon^2\big(P_t^i\delta_{i,j}-P_t^iP_t^j\big).\end{aligned} \tag{2.5}$$

The last term on the right-hand side is known as being the diffusion matrix of the Wright–Fisher model, see for instance [32, 31, 55]. It is also the leading part of so-called Kimura operators, see Subsection 3.2.

Below, we will be specifically interested in cases when the equilibrium strategies are in feedback form, meaning that $\alpha_t^{i,j} = \alpha(t, i, P_t)(j)$ for a function $\alpha : [0,T] \times [\![d]\!] \times \mathcal{S}_{d-1} \times [\![d]\!] \ni (t, i, p, j) \mapsto \alpha(t, i, p)(j) \in \mathbb{R}$ such that, for any $(t, p) \in [0,T] \times \mathcal{S}_{d-1}$ and any $i \in [\![d]\!]$,

$$\alpha(t, i, p)(j) \geq 0, \quad j \in [\![d]\!] \setminus \{i\}, \quad \alpha(t, i, p)(i) = -\sum_{j \neq i} \alpha(t, i, p)(j), \tag{2.6}$$

in which case (2.4) becomes a stochastic differential equation, the well-posedness of which is addressed in the next section, at least in a setting that is relevant to us, see Proposition 2.2. The function α is said to be a *feedback strategy*. One of the key point in the latter statement is that the solution takes values in $\mathcal{S}_{d-1}$ itself. Another key point is that, whenever each p_0^i is in $(0, +\infty)$ and each $\alpha(t, j, p)(i)$ remains away from zero for p_i is in the right neighborhood of 0, the coordinates of the solution are shown to remain almost surely (strictly) positive, which plays a crucial role in the definition of our mean field game with common noise. Below, we ensure strict positivity of the rate transition from j to i for p_i small enough by forcing accordingly the dynamics at the boundary of the simplex[7] of the $(d-1)$-dimensional simplex (where d is the cardinality of the state space); we make this point clear in §2.2.1. Importantly, if this additional forcing at the boundary is strong enough, it also ensures that $\int_0^T 1/P_t^i dt$ has exponential moments of sufficiently high order, see Proposition 2.3 For the time being, we observe that the strict positivity of the solution (provided that we take it for granted) permits to rewrite the equation (2.7) in the form:

$$\mathrm{d}P_t^i = \sum_{j \in [\![d]\!]} P_t^j \alpha(t, j, P_t)(i)\mathrm{d}t + \varepsilon P_t^i \sum_{j \in [\![d]\!]} \sqrt{\frac{P_t^j}{P_t^i}} \mathrm{d}\overline{W}_t^{i,j}, \quad t \in [0,T], \tag{2.7}$$

where, for consistency, we have replaced $\alpha_t^{j,i}$ by $\alpha(t, j, P_t)(i)$.

Now that we have equation (2.7), we can formulate our mean field game. As we already accounted for, the first observation is that the Brownian motions $((W^{i,j})_{0 \leq t \leq T})_{i,j \in [\![d]\!] : i \neq j}$ in (2.7) should be regarded as common noises (or the whole collection should be regarded as a common noise). The second key point is that equation (2.7) should be understood as the equation for an environment $(P_t)_{0 \leq t \leq T}$, candidate for being a solution of the mean field game. It thus remains to introduce the equation for a tagged player evolving within the environment $(P_t)_{0 \leq t \leq T}$. Our key idea in this respect is to linearize (2.7) in order to describe the statistical marginal states of the tagged player, provided $\int_0^T 1/P_t^i dt$ is enough exponentially integrable, namely

$$\mathrm{d}Q_t^i = \sum_{j \in [\![d]\!]} Q_t^j \beta(t, j, P_t)(i)\mathrm{d}t + \varepsilon Q_t^i \sum_{j \in [\![d]\!]} \sqrt{\frac{P_t^j}{P_t^i}} \mathrm{d}\overline{W}_t^{i,j}, \quad t \in [0,T], \tag{2.8}$$

where β stands for the feedback function (hence satisfying (2.6)) used by the tagged player to implement her own strategy in the form of a progressively-measurable (with respect to the

[7] We recall the convention introduced in the very beginning of the paper according to which the boundary is here understood as the boundary of $\hat{\mathcal{S}}_{d-1}$ under the identification of $\mathcal{S}_{d-1}$ and $\hat{\mathcal{S}}_{d-1}$ (and similarly for the interior). We take this convention for granted in the rest of the paper.

filtration $\mathbb{F}^{\boldsymbol{W}}$) process $\boldsymbol{\beta} = ((\beta_t^{i,j} = \beta(t,i,P_t)(j))_{i,j\in[\![d]\!]})_{0\le t\le T}$. The main difficulty here is to interpret (2.8) in a convenient manner. Notice in this regard that our choice to take here $(\beta_t)_{0\le t\le T}$ in a closed feedback form (or semi-closed since it depends on the environment $(P_t)_{0\le t\le T}$) is only for consistency with (2.7) and just plays a little role in our interpretation. Actually, one important fact in this respect is that the variable Q_t^i in (2.8) should read as a *conditional expected mass* when the tagged player is in state i at time t. Here, the reader must be aware of the terminology that we use: We say conditional expected mass instead of conditional probability because, although Q_t is shown below to have non-negative entries and to satisfy $\mathbb{E}[\sum_{i\in[\![d]\!]} Q_t^i] = 1$, it may not be a probability measure, meaning that $\sum_{i\in[\![d]\!]} Q_t^i$ may differ from 1 with a positive probability, which is the whole subtlety of our model.

In order to clarify the equation (2.8), we may indeed associate a Lagrangian or particle representation with it. In the mean field game, we hence assume that the representative agent, at time t, has not only a position $X_t \in [\![d]\!]$, but also another feature $Y_t \in \mathbb{R}_+$, which we call a *mass*. To state the dynamics more precisely, it is convenient to enlarge the current probability space in the form of a product space $(\Omega\times\Xi, \mathcal{A}\otimes\mathcal{G}, \mathbb{P}\otimes\mathbf{P})$, where $(\Xi,\mathcal{G},\mathbf{P})$ denotes another probability space that is just used here (and nowhere else in the text). Whilst $(\Omega,\mathcal{A},\mathbb{P})$ is still equipped with the process $\boldsymbol{W}$, $(\Xi,\mathcal{G},\mathbf{P})$ is now intended to carry the additional idiosyncratic noise to which the representative player is subjected. Accordingly, both processes $\boldsymbol{X} = (X_t)_{0\le t\le T}$ and $\boldsymbol{Y} = (Y_t)_{0\le t\le T}$ are constructed on the product space $\Omega\times\Xi$. Our aim is then to show (at least informally) that, when the environment $\boldsymbol{P}$ is given on the original space $(\Omega,\mathcal{A},\mathbb{P})$, the solution $\boldsymbol{Q}$ of (2.8) satisfies

$$Q_t^i := \mathbf{E}[Y_t \mathbf{1}_{\{X_t=i\}}], \qquad i\in[\![d]\!], \tag{2.9}$$

which means that $\boldsymbol{Q}$ is in fact the conditional expected mass of the reference player, *conditional* here being understood as *conditional on the common noise*. For a given realization of the common noise, the state process $\boldsymbol{X}$ is then required to obey standard Markovian dynamics of the form

$$\mathbf{P}(X_{t+h} = j | X_t = i) = \beta_t^{i,j} h + o(h), \quad t\in[0,T),\ h>0, \quad i,j\in[\![d]\!], \quad i\neq j, \tag{2.10}$$

the transitions $\boldsymbol{\beta}$ being as in (2.8), namely $\boldsymbol{\beta} = ((\beta_t^{i,j} = \beta(t,i,P_t)(j))_{i,j\in[\![d]\!]})_{0\le t\le T}$. Provided $\int_0^T 1/P_t^i \mathrm{d}t$ is sufficiently exponentially integrable, the dynamics of $\boldsymbol{Y}$ are instead given by an $\boldsymbol{X}$-dependent equation:

$$\mathrm{d}Y_t = \frac{\varepsilon}{\sqrt{2}} Y_t \sum_{j\in[\![d]\!]} \sqrt{\frac{P_t^j}{P_t^{X_t}}} \mathrm{d}(W^{X_t,j} - W^{j,X_t}) = \frac{\varepsilon}{\sqrt{2}} Y_t \sum_{i\in[\![d]\!]} \mathbf{1}_{\{X_t=i\}} \sum_{j\in[\![d]\!]} \sqrt{\frac{P_t^j}{P_t^i}} \mathrm{d}(W^{i,j} - W^{j,i}). \tag{2.11}$$

Writing informally $\mathrm{d}Q_t^i$ as

$$\mathrm{d}Q_t^i = \mathbf{E}[\mathbf{1}_{\{X_t=i\}} \mathrm{d}Y_t] + \sum_{j\in[\![d]\!]} Q_t^j \beta_t^{j,i} \mathrm{d}t, \quad t\in[0,T], \quad i\in[\![d]\!],$$

we then recover (2.8). For sure, this argument may be made rigorous. In particular, it may be very convenient to represent the Markov dynamics of $\boldsymbol{X}$ in (2.10) by means of an extra Poisson random measure constructed on $(\Xi,\mathcal{G},\mathbf{P})$, very much in the spirit of [16].

Remark 2.1. *The very mechanism underpinning* (2.10) *and* (2.11) *is in fact different from the true Wright–Fisher model even though it shares some similarities. Instead, when working with*

a finite version of the game, we can force the players with a pure Wright–Fisher noise by resampling their states (in $[\![d]\!]$) independently, according to their instantaneous empirical distribution at the times of a Poisson process. Such an approach is simpler as it bypasses any additional mass process $\boldsymbol{Y}$, but it has a serious drawback for the game: The empirical measure is then strongly attractive since the system is resampled recurrently. This feature precludes any interesting deviating phenomenon, whilst, in our model, the tagged player may really deviate (in law) from the population. In the end, the macroscopic equation (2.7) *for the population is a Wright–Fisher diffusion, but the microscopic behavior of one particle in* (2.10) *and* (2.11) *is different. Noticeably, the state $(\boldsymbol{X}, \boldsymbol{Y})$ of the tagged player is hence encoded, at any time $t \in [0, T]$, in the form $(Y_t \mathbf{1}_{\{X_t = i\}})_{i \in [\![d]\!]}$. Since Q_t in* (2.9) *is just the expectation of the latter quantity, the state variable of the tagged player thus takes values within the same space as the feature Q_t that is used to describe its statistical behavior.*

2.1.3. Cost functional and first formulation of the game

It now remains to associate a cost functional with the tagged player. Consistently with (2.3) we here let

$$\mathcal{J}(\beta, (P_t)_{0 \le t \le T}) := \sum_{i \in [\![d]\!]} \mathbb{E}\Big[Q_T^i g(i, P_T) + \int_0^T Q_t^i \Big(f(t, i, P_t) + \frac{1}{2} \sum_{j \neq i} \big| [\beta(t, i, P_t)](j) \big|^2 \Big) \mathrm{d}t \Big]. \quad (2.12)$$

In the above left-hand side, β stands for the strategy used in the equation for $\boldsymbol{Q} = (Q_t)_{0 \le t \le T}$ in (2.8); also, $\boldsymbol{P} = (P_t)_{0 \le t \le T}$ denotes the environment (as the cost functional does depend upon the environment), defined as the solution of (2.7).

Hence, for an initial condition $p_0 = (p_{0,i})_{i \in [\![d]\!]} \in \mathcal{S}_{d-1}$ with positive entries (that is $p_{0,i} > 0$ for each $i \in [\![d]\!]$), our definition of a mean field game solution comes in the following three steps:

1. Consider a feedback function α as in (2.6) such that (2.7), with p_0 as initial condition, has a unique solution $(P_t)_{0 \le t \le T}$ (say on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ equipped with a collection of $1d$ Brownian motions $((W_t^{i,j})_{0 \le t \le T})_{i,j \in [\![d]\!] : i \neq j}$, with the same convention as before that $\boldsymbol{W}^{i,i} = (W_t^{i,i})_{0 \le t \le T} \equiv 0$ for $i \in [\![d]\!]$), which remains positive with probability 1; the process $(P_t)_{0 \le t \le T}$ is then called an environment;

2. On the same space $(\Omega, \mathcal{A}, \mathbb{P})$, solve, for any bounded and measurable feedback function β, equation (2.8) for $(Q_t)_{0 \le t \le T}$ with p_0 as initial condition, and then find the optimal trajectories (if they exist) of the minimization problem

$$\inf_{\beta} \mathcal{J}(\beta, \boldsymbol{P}). \quad (2.13)$$

3. Find an environment $(P_t)_{0 \le t \le T}$ such that $(P_t)_{0 \le t \le T}$ is an optimal trajectory of (2.13). Such a $(P_t)_{0 \le t \le T}$ is called an MFG equilibrium or a solution to the MFG.

The precise definition is given in the next section (Definition 2.6). Let us remark that an equivalent definition could be also given in terms of the particle representation described above. Indeed, for a given environment $\boldsymbol{P}$ and control β, if $\boldsymbol{X}$ denotes the corresponding position of the reference player and $\boldsymbol{Y}$ its mass, the cost (2.12) rewrites, thanks to (2.9), as

$$\mathcal{J}(\beta, (P_t)_{0 \le t \le T}) := \mathbb{E}\Big[Y_T g(X_T, P_T) + \int_0^T Y_t \Big(f(t, X_t, P_t) + \frac{1}{2} \sum_{j \neq X_t} \big| [\beta(t, X_t, P_t)](j) \big|^2 \Big) \mathrm{d}t \Big].$$

Then, a mean field game solution can be defined as a couple $(\alpha, \boldsymbol{P})$ such that α is optimal for (2.13) and $P^i_t = \mathbb{E}[Y_t \mathbf{1}_{\{X_t=i\}}|\mathcal{F}^{\boldsymbol{W}}_t]$, where $\boldsymbol{X}$ and $\boldsymbol{Y}$ satisfy (2.10) and (2.11) for the given α and $\boldsymbol{P}$.

It is worth noticing that, whenever $\varepsilon = 0$ in (2.7) and (2.8), the system (2.7)–(2.8) becomes a simpler system of two decoupled Fokker–Planck (or Kolmogorov) equations that are similar to (2.2). While existence of a mean field game solution (in the case $\varepsilon = 0$) is by now well-understood, uniqueness remains a difficult issue. In fact, there are few generic conditions that ensure uniqueness. Generally speaking, the two known instances of uniqueness are (besides some specific examples that can be treated case by case) the short time horizon case (namely T is small enough in comparison with the regularity properties of the underlying cost coefficients) and the so-called monotonous case due to Lasry and Lions [46, 45] (which does not require T to be small enough). In short, the cost coefficients f and g are said to be monotonous (in the sense of Lasry and Lions) if, for any $p, p' \in \mathcal{S}_{d-1}$ and for any $t \in [0, T]$,

$$\sum_{i\in[\![d]\!]} (g(i,p) - g(i,p'))(p_i - p'_i) \geq 0, \quad \sum_{i\in[\![d]\!]} (f(t,i,p) - f(t,i,p'))(p_i - p'_i) \geq 0. \tag{2.14}$$

The main goal of the rest of the paper is precisely to prove that, whenever ε in (2.7)–(2.8) is strictly positive, uniqueness may hold true for our MFG under quite mild regularity conditions on the coefficients and in particular without requiring any monotonicity properties; in fact, the main constraint that we ask is that the coordinates of the solutions of (2.7) stay sufficiently far away from zero (provided that the coordinates of the initial condition themselves are not zero). We address this requirement in the next subsection: Basically, it will prompt us to introduce a new term in the dynamics of both $(P_t)_{0\leq t\leq T}$ and $(Q_t)_{0\leq t\leq T}$ to force the coordinates to stay positive.

2.2. New MFG and first meta-statement

As we already alluded to, an important observation is that, for any solution $((P^i_t)_{i\in[\![d]\!]})_{0\leq t\leq T}$ to (2.7), it holds

$$\mathrm{d}\Big(\sum_{i\in[\![d]\!]} P^i_t\Big) = 0, \tag{2.15}$$

which can be easily proved by summing over the coordinates in (2.7). In particular, since the initial condition is taken in $\mathcal{S}_{d-1}$, the mass remains constant, equal to 1. Subsequently, if the coordinates of $(P_t)_{0\leq t\leq T}$ remain non-negative (which we discuss right below), the process $(P_t)_{0\leq t\leq T}$ lives in $\mathcal{S}_{d-1}$, which is of a special interest for us. In fact, non-negativity of the coordinates may be easily seen by rewriting (2.7) in the form

$$\mathrm{d}P^i_t = a_i(t, P_t)\mathrm{d}t + \varepsilon\sqrt{P^i_t(1 - P^i_t)}\mathrm{d}\widetilde{W}^i_t, \tag{2.16}$$

for a new Brownian motion $(\widetilde{W}^i_t)_{0\leq t\leq T}$, where $a_i(t,p) := \sum_{j\in[\![d]\!]}[p_j\alpha(t,j,p)(i) - p_i\alpha(t,i,p)(j)]$, for $i \in [\![d]\!]$, and where the form of the stochastic integral follows from (2.5) with $i = j$ therein. (Notice that the form of a_i differs from the writing used in (2.7), but both are obviously equivalent since $\sum_{j\in[\![d]\!]} \alpha(t,i,p)(j) = 0$.) We have that $a_i(t,p) \geq -Cp_i$, for a constant $C > 0$, since $\alpha(t,j,p)(i) \geq 0$ for $j \neq i$ and we assume α bounded. By stochastic comparison with Feller's $(1d)$ branching diffusion [25, Exercise 5.1], we easily deduce that the coordinates of $(P_t)_{0\leq t\leq T}$ should remain non-negative (the details are left to the reader and a rigorous statement, tailored to our framework, is given below).

2.2.1. Equations that are repelled from the boundary

In the sequel, we are interested in solutions to (2.7) that stay sufficiently far away from the boundary. As we already explained several times, the reason is that our uniqueness result is based upon the smoothing properties of the operator generated by (2.7). Since the latter degenerates at the boundary of the simplex, we want to keep the solutions to (2.7) as long as possible within the relative interior of $\mathcal{S}_{d-1}$. In this regard, it is worth observing from [25, Exercise 5.1] that the sole condition (2.6) is not enough to prevent solutions to (2.7) to touch the boundary of the simplex. To guarantee that no coordinate vanishes, more is needed. For instance, in Feller's branching diffusions, the solution does not vanish if the drift is sufficiently positive in the neighborhood of 0. This prompts us to revisit the two equations (2.7) and (2.8) and to consider instead (notice that, in the two formulas below, the value of $\alpha(t, i, P_t)(i)$ is in fact useless)

$$\mathrm{d}P_t^i = \sum_{j\in[\![d]\!]} \Big(P_t^j(\varphi(P_t^i) + \alpha(t, j, P_t)(i)) - P_t^i(\varphi(P_t^j) + \alpha(t, i, P_t)(j))\Big)\mathrm{d}t + \varepsilon \sum_{j\in[\![d]\!]} \sqrt{P_t^i P_t^j}\mathrm{d}\overline{W}_t^{i,j}, \quad (2.17)$$

and

$$\mathrm{d}Q_t^i = \sum_{j\in[\![d]\!]} \Big(Q_t^j(\varphi(P_t^i) + \beta(t, j, P_t)(i)) - Q_t^i(\varphi(P_t^j) + \beta(t, i, P_t)(j))\Big)\mathrm{d}t + \varepsilon Q_t^i \sum_{j\in[\![d]\!]} \sqrt{\frac{P_t^j}{P_t^i}}\mathrm{d}\overline{W}_t^{i,j}, \quad (2.18)$$

for $t \in [0, T]$, with the same deterministic initial condition $P_0 = (P_0^i = p_{0,i})_{i\in[\![d]\!]}$. Here the function φ is a non-increasing Lipschitz function from $[0, \infty)$ into itself such that

$$\varphi(r) := \begin{cases} \kappa & r \le \delta, \\ 0 & r > 2\delta, \end{cases} \quad (2.19)$$

δ being a positive parameter whose value next is somewhat arbitrary. As for κ, we clarify its main role in the statements of Meta-Theorems 2.7 and 2.9 and of Theorems 2.10, 3.2 and 3.4; the reader may also find a taste of it in the three propositions right below. In the two equations (2.17) and (2.18), α and β are the same as in (2.7) and (2.8). Hence, the drift in the first equation now reads

$$a_i(t, p) := \sum_{j\in[\![d]\!]} \Big(p_j[\varphi(p_i) + \alpha(t, j, p)(i)] - p_i[\varphi(p_j) + \alpha(t, i, p)(j)]\Big). \quad (2.20)$$

It still satisfies $\sum_{i\in[\![d]\!]} a_i(t, p) = 0$. And, importantly, whenever $p_i = 0$ (with $p = (p_1, \cdots, p_d) \in \mathcal{S}_{d-1}$), it satisfies $a_i(t, p) \ge \kappa$. In this framework, we have the following three statements, the proofs of which are postponed to Subsection 3.4.

Proposition 2.2. *Consider φ as in* (2.19) *with $\delta \in (0, 1)$ and $\kappa \ge \varepsilon^2/2$, for $\varepsilon > 0$. Then, for a bounded (measurable) feedback function α as in* (2.6)*, the stochastic differential equation* (2.17) *has a unique (strong) solution whenever the initial condition is prescribed and satisfies $p_{0,i} > 0$ for each $i \in [\![d]\!]$ and $\sum_{i\in[\![d]\!]} p_{0,i} = 1$. Moreover, the coordinates of the solution remain almost surely (strictly) positive and satisfy $\sum_{i\in[\![d]\!]} P_t^i = 1$, for any time.*

The following statement provides a stronger version.

Proposition 2.3. *Under the assumptions and notation of Proposition 2.2, for κ as in* (2.19) *and for $\lambda > 0$, let $\gamma := \kappa - (1+\lambda)\varepsilon^2/2$. Then, the solution to* (2.17) *satisfies*

$$\mathbb{E}\left[\exp\left(\lambda\gamma \int_0^T \frac{1}{P_s^i}\mathrm{d}s\right)\right] \leq C p_{0,i}^{-\lambda}, \quad \text{for each } i \in [\![d]\!], \tag{2.21}$$

together with

$$\sup_{0\leq t\leq T} \mathbb{E}\left[(P_t^i)^{-\lambda}\right] \leq C p_{0,i}^{-\lambda}, \quad \text{for each } i \in [\![d]\!], \tag{2.22}$$

for a constant C that only depends on δ, κ, λ, T and on the supremum norm of α.

Proposition 2.4. *Under the assumption and notation of Proposition 2.2, assume that β is bounded and measurable and that κ in* (2.19) *satisfies $\kappa \geq (61+d)\varepsilon^2$. Assume also that the initial condition $p_0 = (p_{0,i})_{i\in[\![d]\!]} \in \mathcal{S}_{d-1}$ of* (2.17) *has positive entries (that is $p_{0,i} > 0$ for each $i \in [\![d]\!]$). Then, for any initial condition $q_0 = (q_{0,i})_{i\in[\![d]\!]} \in \mathcal{S}_{d-1}$ (possibly different from p_0),* (2.18) *has a unique pathwise solution $((Q_t^i)_{0\leq t\leq T})_{i\in[\![d]\!]}$. It satisfies $\mathbb{E}[\sup_{0\leq t\leq T}|Q_t^i|^4] < \infty$ for any $i \in [\![d]\!]$. The coordinates of the solution are (strictly) positive and satisfy*

$$\mathbb{E}\Big[\sum_{i\in[\![d]\!]} Q_t^i\Big] = 1. \tag{2.23}$$

2.2.2. Reformulation of the game

We now have most of the needed ingredients to formulate the setting to which our main result applies. Roughly speaking, our result addresses the mean field game associated with the pair (2.17)–(2.18) instead of (2.7)–(2.8) and with the cost functional (2.12); so this is the same MFG as the one described in Subsection 2.1 except for the fact that we included the forcing φ in the state equations and that, in the equation (2.18), we will allow for a more general (random) rate function instead of the feedback function β. So, in lieu of (2.18), we will consider (in the mean field game)

$$\mathrm{d}Q_t^i = \sum_{j\in[\![d]\!]} \Big(Q_t^j(\varphi(P_t^i) + \beta_t^{j,i}) - Q_t^i(\varphi(P_t^j) + \beta_t^{i,j})\Big)\mathrm{d}t + \varepsilon Q_t^i \sum_{j\in[\![d]\!]} \sqrt{\frac{P_t^j}{P_t^i}}\mathrm{d}\overline{W}_t^{i,j}, \quad t \in [0,T]. \tag{2.24}$$

Here, $((\beta_t^{i,j})_{0\leq t\leq T})_{i,j\in[\![d]\!]}$ is a collection of bounded $\mathbb{F}^{\boldsymbol{W}}$-progressively-measurable processes that are required to satisfy

$$\beta_t^{i,j} \geq 0 \quad i,j \in [\![d]\!],\ i \neq j. \tag{2.25}$$

(Notice that the diagonal terms $((\beta_t^{i,i})_{0\leq t\leq T})_{i\in[\![d]\!]}$ play no role.) Such processes are called *admissible open-loop strategies*. Somehow, this is to say that we can work (at least for (2.18)) with strategies that may depend upon the whole past of the environment $\boldsymbol{P} = (P_t)_{0\leq t\leq T}$, which is in contrast to strategies of the form $(\beta(t,i,P_t)(j))_{0\leq t\leq T}$ in (2.18) which depend, at time t, on the environment through its current state only. Latter strategies are said to be semi-closed. We explain below how such semi-closed strategies manifest through the master equation.

Remark 2.5. *We let the reader check that Propositions 2.2 and 2.3 given above, see Subsection 3.4 for their proofs, can be also extended to the case where the process $((\alpha(t,i,P_t)(j))_{i,j\in[\![d]\!]})_{0\leq t\leq T}$ in* (2.17) *is replaced by a more general bounded progressively measurable process $((\alpha_t^{i,j})_{i,j\in[\![d]\!]})_{0\leq t\leq T}$ satisfying the analogue of* (2.25)*. The proof of the solvability of* (2.17) *in the statement of Proposition 2.2 is even simpler since $((\alpha_t^{i,j})_{i,j\in[\![d]\!]})_{0\leq t\leq T}$ is then in open-loop form. Similarly, we can replace* (2.18) *by* (2.24) *in the statement of* (2.4)*.*

Importantly, we regard the two cost coefficients f and g in (2.12) as being defined on $[0,T]\times[\![d]\!]\times\mathcal{S}_{d-1}$ and $[\![d]\!]\times\mathcal{S}_{d-1}$ respectively. To make it simpler, we write $g^i(p)$ for $g(i,p)$ and $f^i(t,p)$ for $f(t,i,p)$. Accordingly, for a progressively-measurable strategy $\boldsymbol{\beta}=((\beta_t^{i,j})_{i,j\in[\![d]\!]})_{0\le t\le T}$, the cost functional becomes

$$\mathcal{J}(\boldsymbol{\beta},\boldsymbol{P}) := \sum_{i\in[\![d]\!]}\mathbb{E}\Big[Q_T^i g^i(P_T)+\int_0^T Q_t^i\Big(f^i(t,P_t)+\frac{1}{2}\sum_{j\neq i}\big|\beta_t^{i,j}\big|^2\Big)\mathrm{d}t\Big], \tag{2.26}$$

where $(Q_t)_{0\le t\le T}$ solves (2.24), with $Q_0=P_0$ (the latter being equal to some deterministic $p_0\in\mathcal{S}_{d-1}$).

Definition 2.6. *Given a deterministic initial condition p_0, a* solution of the mean field game (with common noise) *is a pair $(\boldsymbol{P},\alpha)$ such that*

(i) $\boldsymbol{P}=(P_t)_{0\le t\le T}$ is an $\mathcal{S}_{d-1}$-valued process, progressively measurable with respect to $\mathbb{F}^{\boldsymbol{W}}$, with p_0 as initial condition, and $\alpha:[0,T]\times[\![d]\!]\times\mathcal{S}_{d-1}\times[\![d]\!]\to\mathbb{R}$ is a bounded feedback strategy;

(ii) $\boldsymbol{P}$ and α satisfy Equation (2.17) *in the strong sense;*

(iii) $\mathcal{J}(\alpha,\boldsymbol{P})\le\mathcal{J}(\boldsymbol{\beta},\boldsymbol{P})$ for any admissible open-loop strategy $\boldsymbol{\beta}$.

We say that the solution $(\boldsymbol{P},\alpha)$ is unique if given another solution $(\widetilde{\boldsymbol{P}},\widetilde{\alpha})$, we have $P_t=\widetilde{P}_t$ for any $t\in[0,T]$, $\mathbb{P}$-a.s., and $\alpha(t,i,P_t)(j)=\widetilde{\alpha}(t,i,P_t)(j)$ $dt\otimes\mathbb{P}$-a.e., for each $i,j\in[\![d]\!]$.

We recall that the probability space and the Brownian motion $\boldsymbol{W}$ are fixed and then Equations (2.17) and (2.24) have unique strong solutions. The above hence defines *strong* mean field game solutions, in the sence that $\boldsymbol{P}$ is adapted to $\mathbb{F}^{\boldsymbol{W}}$. For a comparison between strong and weak MFG solutions, in the diffusion case, we refer to [12, Chapter 2].

Here is now a meta form of our main statement.

Meta-Theorem 2.7. *Assume that the coefficients f and g are* sufficiently regular. *Then, for any $\varepsilon\in(0,1)$, there exists a threshold $\kappa_0>0$, only depending on ε, $\|f\|_\infty$, $\|g\|_\infty$ and T, such that, for any $\kappa\ge\kappa_0$ and $\delta\in(0,1/(4\sqrt{d}))$, and for any (deterministic) initial condition $(p_{0,i})_{i\in[\![d]\!]}\in\mathcal{S}_{d-1}$ with positive entries, the mean field game has a unique solution as defined by Definition 2.6.*

The statement is said to be in meta-form since the assumptions on f and g are not clear. The definitive version is given in Theorem 3.2 below and the proof is completed in Section 4.2.

Remark 2.8. *At this stage, it is worth mentioning that our notion of solution, as defined in Definition 2.6, could be relaxed: Instead of requiring the strategies to be in feedback form (namely, in the form $((\alpha(t,i,P_t)(j))_{i,j\in[\![d]\!]})_{0\le t\le T}$), we could allow them to be in open-loop form (namely, to be given by more general bounded progressively-measurable processes $((\alpha_t^{i,j})_{i,j\in[\![d]\!]})_{0\le t\le T}$). Our claim is that Meta-Theorem 2.7, and in fact Theorem 3.2 as well, extend to this case: The solution given by Meta-Theorem 2.7 and Theorem 3.2 remains unique within the larger class of open-loop solutions. The proof is exactly the same. Actually, our choice to use feedback strategies is for convenience only since we feel better to keep, in our main statements, the same framework as the one used in the exposition of the problem.*

Moreover, we are confident that our result also extends to random initial conditions, but the proof would certainly require an additional effort since the initial conditions should then satisfy suitable integrability properties. To wit, an expectation must be added to the right-hand side of both (2.21) *and* (2.22) *when $(p_{0,i})_{i\in[\![d]\!]}$ becomes random: The resulting expectations might be infinite unless some integrability properties are indeed satisfied.*

2.3. Overview of the PDE results

The proof of Meta-Theorem 2.7 (or equivalently of the definitive version Theorem 3.2) is highly based upon the so-called master equation associated with the mean field game. We refer to [9, 10, 12, 49, 50] for foundations of the topics for mean field games set on $\mathbb{R}^d$ and to [4, 17, 8, 48] for related issues for mean field games with a finite state space.

2.3.1. Well-posedness of the master equation

Generally speaking, the master equation here takes the form of a system of nonlinear parabolic equations driven by a so-called Kimura operator, the latter being carefully described in Section 3. This system is nothing but the system of second order PDEs (1.1) stated on the $(d-1)$-dimensional simplex. Although the formulation of (1.1) looks fine, it is in fact rather abusive since $\mathcal{S}_{d-1}$ has an empty interior in $\mathbb{R}^d$: In other words, except if the unknown U^i therein is defined on a neighborhood of the simplex, the derivatives that appear in the equations are not properly defined. Although we just clarify this point in the next section, we feel useful to state a preliminary version of our main result on the well-posedness of (1.1). Very much in the spirit of Meta-Theorem 2.7, we have it in the form of a meta-statement.

Meta-Theorem 2.9. *Assume that the coefficients f and g are* sufficiently regular. *Then, for any $\varepsilon \in (0,1)$, there exists a threshold $\kappa_0 > 0$, only depending on ε, $\|f\|_\infty$, $\|g\|_\infty$ and T, such that, for any $\kappa \geq \kappa_0$ and $\delta \in (0, 1/(4\sqrt{d}))$, the master equation* (1.1) *has a unique smooth solution $U = (U^1, \cdots, U^d)$.*

Similar to Meta-Theorem 2.7, Meta-Theorem 2.9 leaves unclear the assumptions on f and g as well as the smoothness of the solution of the master equation. The definitive version is given in Theorem 3.4 below; this includes a clear definition of the various derivatives that appear in (1.1). We refer to Section 4.3 for the proof.

2.3.2. Smoothing estimates for PDEs set on the simplex

Solvability of the master equation (1.1) is clearly the main issue in our analysis. In fact, the key step in our approach is to prove a priori estimates for a so-called *linear* version of the master equation, the latter being obtained by freezing the nonlinear component U therein. To make it clear, the linear analogue of each equation in the system (1.1) may be written in the generic form

$$\begin{aligned} &\partial_t u(t,p) + \sum_{j \in [\![d]\!]} \Big(\varphi(p_j) + b_j(t,p) + p_j b_j^\circ(t,p) \Big) \partial_{p_j} u(t,p) \\ &\quad + \frac{\varepsilon^2}{2} \sum_{j,k \in [\![d]\!]} (p_j \delta_{jk} - p_j p_k) \partial^2_{p_j p_k} u(t,p) + h(t,p) = 0, \\ &u(T,p) = \ell(p), \end{aligned} \tag{2.27}$$

for $(t,p) \in [0,T] \times \mathrm{Int}(\hat{\mathcal{S}}_{d-1})$ ($\mathrm{Int}(\hat{\mathcal{S}}_{d-1})$ being here regarded as a subset of $\mathcal{S}_{d-1}$), $b = (b_j)_{j \in [\![d]\!]} : [0,T] \times \mathcal{S}_{d-1} \to (\mathbb{R}_+)^d$, $b^\circ = (b_j^\circ)_{j \in [\![d]\!]} : [0,T] \times \mathcal{S}_{d-1} \to \mathbb{R}^d$, $h : [0,T] \times \mathcal{S}_{d-1} \to \mathbb{R}$ and $\ell : \mathcal{S}_{d-1} \to \mathbb{R}$ are bounded and satisfy

$$\sum_{j \in [\![d]\!]} \Big(\varphi(p_j) + b_j(t,p) + p_j b_j^\circ(t,p) \Big) = 0, \quad t \in [0,T], \quad p \in \mathcal{S}_{d-1}, \tag{2.28}$$

the unknown u in (2.27) being here real-valued (in words, it is an equation and not a system of equations). Obviously, the function φ is the same as in (2.19) and the constraint (2.28) is

reminiscent of (2.6) as it guarantees that the diffusion process associated with (2.27) leaves the simplex invariant.

The technical result below is the core of our paper. It provides an a priori Hölder estimate to solutions of (2.27) under the rather weak assumption that b and b° are merely continuous. This is made possible under the constraint that φ points inward the simplex at the boundary in a sufficiently strong manner, whence our need to have κ large enough in Meta-Theorems 2.7 and 2.9. Differently from the latter two, this new statement must be regarded as being complete except for the fact that the notion of derivatives in the PDE (2.27) must be clarified.

Theorem 2.10. *Assume that $(b_j)_{j\in[\![d]\!]}$, $(b_j^\circ)_{j\in[\![d]\!]}$ and h are time-space continuous and that ℓ is Lipschitz continuous (we let $\|\ell\|_{1,\infty} = \|\ell\|_\infty + \sup_{p\neq q}|\ell(p)-\ell(q)|/|p-q|$). Then, there exists an exponent $\eta \in (0,1)$ such that, for any given $\delta \in (0, 1/(4\sqrt{d}))$ and $\varepsilon \in (0,1)$, we can find a threshold $\kappa_0 > 0$, depending on ε and $(\|b_j\|_\infty)_{j\in[\![d]\!]}$, such that, for any $\kappa \geq \kappa_0$, we can find another constant C, only depending on δ, ε, κ, $(\|b_j\|_\infty)_{j\in[\![d]\!]}$, $(\|b_j^\circ\|_\infty)_{j\in[\![d]\!]}$, $\|h\|_\infty$, $\|\ell\|_{1,\infty}$ and T, such that any solution $u \in C^{1,2}([0,T]\times \mathrm{Int}(\hat{\mathcal{S}}_{d-1}),\mathbb{R}) \cap C^0([0,T]\times\mathcal{S}_{d-1},\mathbb{R})$ of (2.27) satisfies*

$$\big|u(t,p)-u(s,q)\big| \leq C(|t-s|^{\eta/2}+|p-q|^\eta), \quad (s,t)\in[0,T],\ (p,q)\in\mathcal{S}_{d-1}.$$

Moreover, $\|u\|_\infty$ is less than $\|\ell\|_\infty + T\|h\|_\infty$.

The proof of the theorem is provided in Section 5.

Remark 2.11.

1. *The notation $C^{1,2}$ in the statement is here understood in the usual sense: u is required to be once continuously differentiable in t and twice continuously differentiable in p on the interior of the simplex, the notion of derivative being fully clarified in Subsections 3.1 and 3.2. As for the notation C^0, it refers to functions that are continuous in (t,p).*

2. *We stress that we are not aware of any similar a priori Hölder estimate in the literature. There are some papers about the Hölder regularity of elliptic equations with degeneracies near the boundary, but they do not fit our framework (besides the obvious fact that the underlying equations are elliptic whist ours is parabolic): We refer for instance to [38] for a case with a specific instance of drift that does not cover our needs. We also emphasize that the Hölder estimate in Theorem 2.10 does not depend on the modulus of continuity of the coefficients $(b_j)_{j\in[\![d]\!]}$, $(b_j^\circ)_{j\in[\![d]\!]}$ and h. In fact, we here assume the latter to be continuous for convenience only as it suffices for our own purposes. We believe that the result would remain true if $(b_j)_{j\in[\![d]\!]}$, $(b_j^\circ)_{j\in[\![d]\!]}$ and h were merely bounded and measurable; this would require to adapt accordingly the proof, which consists in mollifying the coefficients, as explained in the introduction of Section 5, see in particular footnote* (16).

3. *On another matter, it is worth noticing that we may trace back explicitly the dependence of κ_0 over ε. The key point in the proof is inequality* (5.35)*, which shows that κ_0 may be taken of the form $\kappa_0 = \varepsilon^{-2}\kappa_{00}$, for κ_{00} only depending on $(\|b_j\|_\infty)_{j\in[\![d]\!]}$. The parameter η therein is a free parameter that is eventually chosen as $1/2$, see the discussion after Proposition 5.6.*

3. From Wright-Fischer spaces to complete statements

The purpose of this section is twofold. First, we introduce useful material for studying the PDEs addressed in the article. By the way, we provide a more rigorous formulation of the two equations (1.1) and (2.27), using appropriate systems of derivatives. This allows us to obtain complete and definitive versions of the two Meta-Theorems 2.7 and 2.9. In a second step of the section, we provide the proofs of Propositions 2.2, 2.3 and 2.4 that we used in the previous section to formulate our MFG.

3.1. Reformulation of the master equation

On the road towards a complete version of Meta-Theorem 2.9, we first provide a more rigorous formulation of the master equation (1.1) since the latter should be regarded as a PDE stated on a $(d-1)$-dimensional manifold.

3.1.1. Local coordinates

The first way is to reformulate the master equation in so-called local coordinates, using the identification of $\mathcal{S}_{d-1}$ with $\hat{\mathcal{S}}_{d-1} = \{(x_1, \cdots, x_{d-1}) \in (\mathbb{R}_+)^{d-1} : \sum_{i\in[\![d-1]\!]} x_i \le 1\}$ (see the introduction for the notation). For a real-valued function h defined on $\mathcal{S}_{d-1}$, we may indeed define the function

$$\begin{aligned} \hat{h}^i(p^{-i}) := h(p) = h\Big(p_1, \cdots, p_{i-1}, 1 - \sum_{k\neq i} p_k, p_{i+1}, \cdots, p_d\Big), \\ \text{with } p^{-i} = \Big(p_1, \cdots, p_{i-1}, p_{i+1}, \cdots, p_d\Big), \end{aligned} \tag{3.1}$$

for $p \in \mathcal{S}_{d-1}$ and hence $p^{-i} \in \hat{\mathcal{S}}_{d-1}$, and for $i \in [\![d]\!]$. We then say that h is differentiable on $\mathcal{S}_{d-1}$ if $\hat{h}^i$ is differentiable on $\hat{\mathcal{S}}_{d-1}$ for some (and hence for any[8]) $i \in [\![d]\!]$. In case when h is defined on a (d-dimensional) neighborhood of $\mathcal{S}_{d-1}$, we then have $\partial_{p_j}\hat{h}^i(p^{-i}) = \partial_{p_j}h(p) - \partial_{p_i}h(p)$, for $j \neq i$. We then end up with

$$\partial_{p_j}\hat{h}^i(p^{-i}) - \partial_{p_k}\hat{h}^i(p^{-i}) = \partial_{p_j}h(p) - \partial_{p_k}h(p), \tag{3.2}$$

for any $j, k \in [\![d]\!] \setminus \{i\}$ and $p \in \mathcal{S}_{d-1}$. Similarly, the second order derivative may be written as

$$\partial^2_{p_jp_k}\hat{h}^i(p^{-i}) = \partial^2_{p_jp_k}h(p) - \partial^2_{p_ip_j}h(p) - \partial^2_{p_ip_k}h(p) + \partial^2_{p_ip_i}h(p), \tag{3.3}$$

and the second order term in (1.1) with $h = U^i$ (the reader should not make any confusion between U^i and $\hat{h}^i$: U^i is the ith coordinate of the solution to the master equation whilst $\hat{h}^i$ is the projection of h, whenever the latter is real-valued, onto a real-valued function on $\hat{\mathcal{S}}_{d-1}$) becomes

$$\frac{\varepsilon^2}{2}\sum_{j,k\in[\![d]\!]}(p_j\delta_{jk} - p_jp_k)\partial^2_{p_jp_k}h(p) = \frac{\varepsilon^2}{2}\sum_{j,k\neq i}(p_j\delta_{jk} - p_jp_k)\partial^2_{p_jp_k}\hat{h}^i(p^{-i}), \tag{3.4}$$

for $p \in \mathcal{S}_{d-1}$. In fact, according to the definition of the Wright–Fisher spaces in the forthcoming §3.2.2, we will consider functions that are just twice-differentiable on the interior $\mathrm{Int}(\hat{\mathcal{S}}_{d-1})$ of the simplex and for which the second-order derivatives may not extend to the boundary of the simplex.

[8] In order to prove the differentiability of $\hat{h}^j$ for any $j \in [\![d]\!] \setminus \{i\}$, it suffices to see that (say that $j > i$ to simplify) $\hat{h}^j(p^{-j}) = \hat{h}^i(p_1, \cdots, p_{i-1}, p_{i+1}, \cdots, p_{j-1}, 1 - \sum_{k\neq j} p_k, p_{j+1}, \cdots)$.

3.1.2. Intrinsic derivatives

Equivalently, we may formulate the derivatives in (1.1) in terms of the intrinsic gradient on $\mathcal{S}_{d-1}$, regarded as a $(d-1)$-dimensional manifold. Indeed, whenever h is defined on a neighborhood of $\mathcal{S}_{d-1}$, we may denote by $\nabla h = (\partial_{p_1} h, \dots, \partial_{p_d} h)$ the standard gradient in $\mathbb{R}^d$. Identifying the tangent space $\mathcal{T}_{d-1}$ to the simplex at a given point $p \in \mathcal{S}_{d-1}$ with the orthogonal space to the d-dimensional vector $\mathbf{1} = (1, \cdots, 1)$, the intrinsic gradient of h, seen as a function defined on the simplex, at p identifies with the orthogonal projection of ∇h on $\mathcal{T}_{d-1}$. We denote it by $\mathfrak{D}h := (\mathfrak{d}_1 h, \dots, \mathfrak{d}_d h)$, that is

$$\mathfrak{D}h := \nabla h - \frac{1}{d}(\mathbf{1} \cdot \nabla h)\mathbf{1}\,; \quad \mathfrak{d}_{p_i} h := \partial_{p_i} h - \frac{1}{d} \sum_{j \in [\![d]\!]} \partial_{p_j} h, \quad i \in [\![d]\!],$$

or, equivalently (which allows to define $\mathfrak{D}h$ when h is just defined on $\mathcal{S}_{d-1}$),

$$\mathfrak{d}_{p_i} h(p) = -\frac{1}{d} \sum_{j \neq i} \partial_{p_j} \hat{h}^i(p^{-i}), \quad p \in \mathcal{S}_{d-1}.$$

Of course we have $\sum_j \mathfrak{d}_{p_j} h = \mathbf{1} \cdot \mathfrak{D}h = 0$ by construction[9]. And the following holds true

$$\mathfrak{d}_{p_i} h(p) - \mathfrak{d}_{p_j} h(p) = \partial_{p_i} h(p) - \partial_{p_j} h(p),$$

for $i, j \in [\![d]\!]$ and $p \in \mathcal{S}_{d-1}$. As for the second order derivatives, we have[10]

$$\begin{aligned} \mathfrak{d}^2_{p_i p_j} h(p) &= \partial^2_{p_i p_j} h(p) - \frac{1}{d} \sum_{k \in [\![d]\!]} (\partial^2_{p_i p_k} h(p) + \partial^2_{p_j p_k} h(p)) + \frac{1}{d^2} \sum_{k,l \in [\![d]\!]} \partial^2_{p_k p_l} h(p) \\ &= \frac{1}{d^2} \sum_{k,l \neq i} \partial^2_{p_k p_l} \hat{h}^i(p^{-i}) - \frac{1}{d} \sum_{k \neq i} \partial^2_{p_k p_j} \hat{h}^i(p^{-i}), \end{aligned} \tag{3.5}$$

[9] When h is just defined on the simplex, the proof is slightly less obvious, but it may be achieved by checking that $\partial_{p_j} \hat{h}^i(p^{-i}) = -\partial_{p_i} \hat{h}^j(p^{-j})$, for $i \neq j$. And then, $\sum_{i \in [\![d]\!]} \sum_{j \neq i} \partial_{p_j} \hat{h}^i(p^{-i}) = -\sum_{i \in [\![d]\!]} \sum_{j \neq i} \partial_{p_i} \hat{h}^j(p^{-j})$. By Fubini's theorem, the latter is also equal to $-\sum_{j \in [\![d]\!]} \sum_{i \neq j} \partial_{p_j} \hat{h}^i(p^{-i}) = -\sum_{i \in [\![d]\!]} \sum_{j \neq i} \partial_{p_i} \hat{h}^j(p^{-j})$, from which we deduce that it is indeed equal to 0.

[10] The second line follows from the identity $\partial_{p_l} \hat{h}^j(p^{-j}) = \partial_{p_l} \hat{h}^i(p^{-i}) - \partial_{p_j} \hat{h}^i(p^{-i})$ if $l \neq i$ and $\partial_{p_l} \hat{h}^j(p^{-j}) = -\partial_{p_j} \hat{h}^i(p^{-i})$ if $l = i$, which, in turn, implies

$$\begin{aligned} \mathfrak{d}_{p_i}(\mathfrak{d}_{p_j} h) &= -\frac{1}{d} \sum_{k \neq i} \partial_{p_k} \widehat{\mathfrak{d}_{p_j} h}^i(p^{-i}) = -\frac{1}{d} \sum_{k \neq i} \partial_{p_k} \Big(-\frac{1}{d} \sum_{l \neq i,j} \big[\partial_{p_l} \hat{h}^i(p^{-i}) - \partial_{p_j} \hat{h}^i(p^{-i}) \big] + \frac{1}{d} \partial_{p_j} \hat{h}^i(p^{-i}) \Big) \\ &= -\frac{1}{d} \sum_{k \neq i} \partial_{p_k} \Big(-\frac{1}{d} \sum_{l \neq i} \partial_{p_l} \hat{h}^i(p^{-i}) + \partial_{p_j} \hat{h}^i(p^{-i}) \Big). \end{aligned}$$

and then the second order term in (1.1) (with $h = U^i$, see (3.4)) becomes[11]

$$\sum_{j,k\in[\![d]\!]} (p_j\delta_{jk} - p_jp_k)\partial^2_{p_jp_k}h(p) = \sum_{j,k\in[\![d]\!]} (p_j\delta_{jk} - p_jp_k)\mathfrak{d}^2_{p_jp_k}h,$$

since $\sum_{k\in[\![d]\!]}(p_j\delta_{jk} - p_jp_k) = 0$ for any $j \in [\![d]\!]$.

3.1.3. Two equivalent formulations

In the end, the master equation (1.1) may be written in two equivalent forms. The first one may be written in terms of the derivatives in (3.2)–(3.3):

$$\begin{aligned}
&\partial_t U^i(t,p) - \frac{1}{2}\sum_{j=1}^d (U^i(t,p) - U^j(t,p))^2_+ + f^i(t,p) + \sum_{j\in[\![d]\!]} \varphi(p_j)[U^j(t,p) - U^i(t,p)] \\
&\quad + \sum_{j\neq i}\sum_{k\in[\![d]\!]} \Big(p_k[\varphi(p_j) + (U^k(t,p) - U^j(t,p))_+] \\
&\qquad\qquad - p_j[\varphi(p_k) + (U^j(t,p) - U^k(t,p))_+]\Big)\partial_{p_j}\widehat{\{U^i\}}^i(t,p^{-i}) \\
&\quad - \varepsilon^2\sum_{j\neq i} p_j\partial_{p_j}\widehat{\{U^i\}}^i(t,p^{-i}) + \frac{\varepsilon^2}{2}\sum_{j,k\neq i}(p_j\delta_{jk} - p_jp_k)\partial^2_{p_jp_k}\widehat{\{U^i\}}^i(t,p^{-i}) = 0, \\
&U^i(T,p) = g^i(p),
\end{aligned} \tag{3.6}$$

for $(t,p) \in [0,T]\times \mathrm{Int}(\hat{\mathcal{S}}_{d-1})$. Above, the function $\widehat{\{U^i\}}^i$ is defined on $\hat{\mathcal{S}}_{d-1}$ in the same way as before, namely $\widehat{\{U^i\}}^i(t,p^{-i}) = U^i(t,p_1,\cdots,p_{i-1},1-\sum_{j\neq i}p_j,p_{i+1},\cdots,p_d)$, for $p \in \mathcal{S}_{d-1}$. For sure, we could rewrite the equation for U^i in terms of the variable p^{-j} (instead of p^{-i}) for another index $j \neq i$, but this would be of little interest for us. In fact, we will make greater use of a second form of (1.1) that may be written in terms of the intrinsic derivative:

$$\begin{aligned}
&\partial_t U^i(t,p) - \frac{1}{2}\sum_{j=1}^d (U^i(t,p) - U^j(t,p))^2_+ + f^i(t,p) + \sum_{j\in[\![d]\!]} \varphi(p_j)[U^j(t,p) - U^i(t,p)] \\
&\quad + \sum_{j,k\in[\![d]\!]} p_k[\varphi(p_j) + (U^k(t,p - U^j(t,p))_+]\Big(\mathfrak{d}_{p_j}U^i(t,p) - \mathfrak{d}_{p_k}U^i(t,p)\Big) \\
&\quad + \varepsilon^2\sum_{j\neq i} p_j\Big(\mathfrak{d}_{p_i}U^i(t,p) - \mathfrak{d}_{p_j}U^i(t,p)\Big) + \frac{\varepsilon^2}{2}\sum_{j,k\in[\![d]\!]}(p_j\delta_{jk} - p_jp_k)\mathfrak{d}^2_{p_jp_k}U^i(t,p) = 0, \\
&U^i(T,p) = g^i(p),
\end{aligned} \tag{3.7}$$

[11] The result may be also proved by combining (3.4) and (3.5), hence bypassing the derivatives of h themselves. Indeed,

$$\begin{aligned}
\sum_{j,k\in[\![d]\!]} (p_j\delta_{jk} - p_jp_k)\mathfrak{d}^2_{p_jp_k}h &= -\frac{1}{d}\sum_{j,k\in[\![d]\!]} (p_j\delta_{jk} - p_jp_k)\sum_{l\neq k}\partial^2_{p_lp_j}\hat{h}^k(p^{-k}) \\
&= -\frac{1}{d}\sum_{j,k\in[\![d]\!]} (p_j\delta_{jk} - p_jp_k)\Big(\sum_{l\neq k,i}[\partial^2_{p_lp_j}\hat{h}^i(p^{-i}) - \partial^2_{p_kp_j}\hat{h}^i(p^{-i})] - \partial^2_{p_kp_j}\hat{h}^i(p^{-i})\Big) \\
&= -\frac{1}{d}\sum_{j,k\in[\![d]\!]} (p_j\delta_{jk} - p_jp_k)\Big(\sum_{l\neq i}\partial^2_{p_lp_j}\hat{h}^i(p^{-i}) - d\cdot\partial^2_{p_kp_j}\hat{h}^i(p^{-i})\Big) \\
&= \sum_{j,k\in[\![d]\!]} (p_j\delta_{jk} - p_jp_k)\partial^2_{p_kp_j}\hat{h}^i(p^{-i}).
\end{aligned}$$

for $(t, p) \in [0, T] \times \mathrm{Int}(\hat{\mathcal{S}}_{d-1})$.

3.2. Kimura operators

We now clarify the choice of the functional spaces for f and g in Meta-Theorems 2.7 and 2.9 and for U in Meta-Theorem 2.9. Basically, we take those spaces from a recent Schauder like theory due to Epstein and Mazzeo [28] for what we called Kimura operators, the latter being operators of the very same structure as the second order generator of (2.7), which we will denote by $(\mathcal{L}_t)_{0\le t\le T}$.

3.2.1. Normal form

Following (2.5), we get (at least informally) that, for any twice differentiable real-valued function h on $\mathbb{R}^d$,

$$\mathcal{L}_t h(p) = \sum_{i\in[\![d]\!]} a_i(t,p)\partial_{p_i} h(p) + \frac{\varepsilon^2}{2}\sum_{i,j\in[\![d]\!]} (p_i\delta_{i,j} - p_i p_j)\partial^2_{p_i p_j} h(p). \tag{3.8}$$

Obviously, we here recover the same second-order operator as in the master equation (1.1), but as we have just explained, this writing is rather abusive. For sure, we have introduced in the previous subsection convenient systems of coordinates and of derivatives that permit to get a more rigorous form. However we need here a more refined formulation of (3.8) that brings out the degeneracies of the operator and, most of all, that fits the framework of Definition 2.2.1 for Kimura operators in the aforementioned monograph [28]. In this regard, it is worth noticing that the second-order term in $\mathcal{L}_t$ is obviously degenerate when it is acting on functions defined on $\mathbb{R}^d$. This follows from the fact that the matrix $(p_i\delta_{i,j} - p_i p_j)_{i,j\in[\![d]\!]}$ has $(1, \cdots, 1)$ in its kernel whenever $(p_1, \cdots, p_d)$ is in $\mathcal{S}_{d-1}$; this is somehow the price to pay for forcing the solutions to (2.7) to stay within $\mathcal{S}_{d-1}$. More subtly, we are concerned in the rest of this section with the degeneracies of $\mathcal{L}_t$ when it is truly acting on functions defined on the simplex and hence when it is expressed in local coordinates.

In case when the drift a in (3.8) is zero, $\mathcal{L}_t$ becomes time independent and coincides with the generator of the d-dimensional Wright–Fisher model. In case when a is non-zero but is time-independent and satisfies

$$a_i(p) \ge 0 \quad \text{if } p_i = 0, \tag{3.9}$$

(which means that a points inward at points p that belong to the boundary of $\hat{\mathcal{S}}_{d-1}$), the operator $\mathcal{L}_t$ itself becomes a time-homogeneous Kimura diffusion operator. Below, we make an intense use of the recent monograph of Epstein and Mazzeo [28] on those types of operators, see also [19, 43, 56] for earlier results. The key feature is that, under the identification of $\mathcal{S}_{d-1}$ with $\hat{\mathcal{S}}_{d-1}$, we may regard the simplex as a $d-1$ dimensional *manifold with corners*, the corners being obtained by intersecting at most $d-1$ 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 $\hat{\mathcal{S}}_{d-1}$ (we then call the codimension of the corner the number of hyperplanes showing up in the intersection). Accordingly and consistently with the analysis performed in Subsection 3.1, we can rewrite (3.8) as an operator acting on functions from $\hat{\mathcal{S}}_{d-1}$ to $\mathbb{R}$, the resulting operator being then a Kimura diffusion operator on $\hat{\mathcal{S}}_{d-1}$. It suffices to reformulate (3.8) in terms of the sole $d-1$ first coordinates $(p_1, \cdots, p_{d-1})$ or, more generally, in terms of $(p_i)_{i\in[\![d]\!]\setminus\{l\}}$ for any given coordinate $l \in [\![d]\!]$. Somehow, choosing the coordinate l amounts to choosing a system of local coordinates and, as we explain below, the choice of l is mostly dictated by the position of $(p_1, \cdots, p_d)$ inside the simplex. For instance, whenever all the

entries of $p = (p_1, \cdots, p_d)$ are positive, meaning that $(p_1, \cdots, p_{d-1})$ belongs to the interior of $\hat{\mathcal{S}}_{d-1}$, the choice of l does not really matter and we may work, for convenience, with $l = d$. We then rewrite the generator $\mathcal{L}_t$, as given by (3.8), in the form

$$\hat{\mathcal{L}}_t \hat{h}(x) = \sum_{i=1}^{d-1} \hat{a}_i(t,x) \partial_{x_i} \hat{h}(x) + \frac{\varepsilon^2}{2} \sum_{i,j=1}^{d-1} (x_i \delta_{i,j} - x_i x_j) \partial^2_{x_i x_j} \hat{h}(x), \tag{3.10}$$

where now $x \in \hat{\mathcal{S}}_{d-1}$, $\hat{h}$ is a smooth function on $\mathbb{R}^{d-1}$ and $\hat{a}_i(t,x) = a_i(t, \check{x})$, with $\check{x} = (x_1, \cdots, x_{d-1}, 1 - x_1 - \cdots - x_{d-1})$. Following the computations of Subsection 3.1, the connection between (3.8) and (3.10) is that $\hat{\mathcal{L}}_t \hat{h}(x) = \mathcal{L}_t\{h(\check{x})\}$, whenever $\hat{h}$ is defined from h through the identity $\hat{h}(x) = h(\check{x})$.

Importantly, we then notice that, in the interior of $\hat{\mathcal{S}}_{d-1}$, $\hat{\mathcal{L}}_t$ is elliptic. Indeed, for any $(x_1, \cdots, x_{d-1}) \in \hat{\mathcal{S}}_{d-1}$ and $(\xi_1, \cdots, \xi_{d-1}) \in \mathbb{R}^{d-1}$,

$$\begin{aligned} \sum_{i,j=1}^{d-1} \xi_i (x_i \delta_{i,j} - x_i x_j) \xi_j &= \sum_{i=1}^{d-1} \xi_i^2 x_i - \Big(\sum_{i=1}^{d-1} \xi_i x_i \Big)^2 \geq \sum_{i=1}^{d-1} \xi_i^2 x_i - \sum_{j=1}^{d-1} x_j \sum_{i=1}^{d-1} \xi_i^2 x_i \\ &= \sum_{i=1}^{d-1} \xi_i^2 x_i \Big(1 - \sum_{j=1}^{d-1} x_j \Big), \end{aligned} \tag{3.11}$$

which suffices to prove ellipticity whenever $x_1, \cdots, x_{d-1} > 0$ and $\sum_{j=1}^{d-1} x_j < 1$.

Take now a corner of codimension $l \in \{1, \cdots, d-1\}$. If the l hyperplanes entering the definition of the corner are of the form $\mathcal{H}_{i_1} = \{x \in \mathbb{R}^{d-1} : x_{i_1} = 0\}, \ldots, \mathcal{H}_{i_l} = \{x \in \mathbb{R}^{d-1} : x_{i_l} = 0\}$, for $1 \leq i_1 < i_2 < \cdots < i_l \leq d-1$, then we can rewrite (3.10) in the form

$$\begin{aligned} \hat{\mathcal{L}}_t \hat{h}(x) = \frac{\varepsilon^2}{2} \sum_{i \in I} (1 - x_i) x_i \partial^2_{x_i} \hat{h}(x) + \frac{\varepsilon^2}{2} \sum_{j,k \notin I} x_j (\delta_{j,k} - x_k) \partial^2_{x_j x_k} \hat{h}(x) + \sum_{i \in I} \hat{a}_i(t,x) \partial_{x_i} \hat{h}(x) \\ - \frac{\varepsilon^2}{2} \sum_{i,j \in I : i \neq j} x_i x_j \partial^2_{x_i x_j} \hat{h}(x) - \varepsilon^2 \sum_{i \in I, j \notin I} x_i x_j \partial^2_{x_i x_j} \hat{h}(x) + \sum_{i \notin I} \hat{a}_i(t,x) \partial_{x_i} \hat{h}(x), \end{aligned} \tag{3.12}$$

with $I = \{i_1, \cdots, i_l\}$. Here, we observe from (3.9) that $\hat{a}_i(t,x) \geq 0$ if $i \in I$ and $x \in \cap_{j \in I} \mathcal{H}_j$. Also, as long as x belongs to $\cap_{j \in I} \mathcal{H}_j$ but $(x_j)_{j \notin I}$ and $1 - \sum_{j \notin I} x_j$ remain positive (which is necessarily true in the relative interior of $\cap_{j \in I} \mathcal{H}_j \cap \hat{\mathcal{S}}_{d-1}$), then, by the same argument as in (3.11), the matrix $(x_j(\delta_{j,k} - x_k))_{j,k \notin I}$ is non degenerate. Hence, up to the intensity factor ε, the above decomposition fits (2.4) in [28, Definition 2.2.1], which is of crucial interest for us[12].

Assume now that the corner of codimension l is given by the intersection of the hyperplanes $\{x \in \mathbb{R}^{d-1} : x_{i_1} = 0\}, \ldots, \{x \in \mathbb{R}^{d-1} : x_{i_{l-1}} = 0\}$ and $\{x \in \mathbb{R}^{d-1} : x_1 + \cdots + x_{d-1} = 1\}$. In order to recover the same form as in [28, (2.4)] (or equivalently in (3.12)), we perform the following change of variables: consider $(y_1, \cdots, y_{d-1}) := (p_1, \cdots, p_{i_l - 1}, p_{i_l + 1}, \cdots, p_d)$ as a new system of local coordinates, for some given index $i_l \in \{i_{l-1} + 1, \cdots, d-1\}$ (which is indeed possible if $i_{l-1} < d-1$; if not, i_l must be chosen as the largest index that is different from $i_1, \cdots, i_{l-1}, d$ and

[12] The reader may also notice that, in [28], the operator $\mathcal{L}_t$ is passed in a normal form, see Proposition 2.2.3 therein which guarantees that such a normal form indeed exists. Here the change of variable to get the normal form may be easily clarified by adapting the 1d case accordingly, see [27]: It suffices to change x_i into the new coordinate $\arcsin^2(\sqrt{x_i})$. In fact, the normal form in [28] plays a key role in the definition of the Wright-Fischer Hölder spaces that we recall in the next paragraph.

hence is lower than i_{l-1}, which asks for reordering the coordinates in the change of variables). For a test function h as before, we then expand $\mathcal{L}_l[h(p_1, \cdots, p_{i_l-1}, 1-\sum_{j\neq i_l} p_j, p_{i_{l-1}+1}, \cdots, p_d)]$ as before and check that we recover the same structure as in [28, (2.4)], but in the new coordinates. This demonstrates how to check the setting of [28, Definition 2.2.1].

3.2.2. Wright–Fisher spaces

The rationale for double-checking [28, Definition 2.2.1] is that we want to use next the Schauder theory developed in [28] for Kimura operators. This prompts us to introduce here the functional spaces that underpin the corresponding Schauder estimates.

For a point $x^0 \in \hat{\mathcal{S}}_{d-1}$ in the relative interior of a corner C of $\hat{\mathcal{S}}_{d-1}$ of codimension $l \in \{0, \cdots, d-1\}$ (if $l = 0$, then x^0 is in the interior of $\hat{\mathcal{S}}_{d-1}$), we may consider a new system of coordinates $(y_1, \cdots, y_{d-1})$ (of the same type as in the previous paragraph) such that $C = \{y \in \hat{\mathcal{S}}_{d-1} : y_{i_1} = \cdots = y_{i_l} = 0\}$, for $1 \leq i_1 < \cdots < i_l$. Letting $I := \{i_1, \cdots, i_l\}$ and denoting by $(y^0_1, \cdots, y^0_{d-1})$ the coordinates of x^0 in the new system (for sure $y^0_{i_j} = 0$ for $j = 1, \cdots, l$), we may find a $\delta^0 > 0$ such that:

1. the closure $\overline{\mathcal{U}}(\delta^0, x^0)$ of $\mathcal{U}(\delta^0, x^0) := \{y \in (\mathbb{R}_+)^{d-1} : \sup_{i=1,\cdots,d-1} |y_i - y^0_i| < \delta^0\}$ is included in $\hat{\mathcal{S}}_{d-1}$,
2. for y in $\overline{\mathcal{U}}(\delta^0, x^0)$, for $j \notin I$, $y_j > 0$,
3. for y in $\overline{\mathcal{U}}(\delta^0, x^0)$, for $y_1 + \cdots + y_{d-1} < 1 - \delta^0$.

A function $\hat{h}$ defined on $\overline{\mathcal{U}}(\delta^0, x^0)$ is then said to belong to $\mathscr{C}^\gamma_{\mathrm{WF}}(\overline{\mathcal{U}}(\delta^0, x^0))$, for some $\gamma \in (0,1)$, if, in the new system of coordinates, $\hat{h}$ is Hölder continuous on $\overline{\mathcal{U}}(\delta^0, x^0)$ with respect to the distance[13]

$$d(y, y') := \sum_{i=1}^{d-1} \big| \sqrt{y_i} - \sqrt{y'_i} \big|. \tag{3.13}$$

We then let

$$\big\|\hat{h}\big\|_{\gamma;\mathcal{U}(\delta^0,x^0)} := \sup_{y\in\overline{\mathcal{U}}(\delta^0,x^0)} \big|\hat{h}(y)\big| + \sup_{y,y'\in\overline{\mathcal{U}}(\delta^0,x^0)} \frac{|\hat{h}(y) - \hat{h}(y')|}{d(y,y')^\gamma}.$$

Following [28, Lemma 5.2.5 and Definition 10.1.1], we say that a function $\hat{h}$ defined on $\mathcal{U}(\delta^0, x^0)$ belongs to $\mathscr{C}^{2+\gamma}_{\mathrm{WF}}(\mathcal{U}(\delta^0, x^0))$ if, in the new system of coordinates,

1. $\hat{h}$ is continuously differentiable on $\mathcal{U}(\delta^0, x^0)$ and $\hat{h}$ and its derivatives extend continuously to $\overline{\mathcal{U}}(\delta^0, x^0)$ and the resulting extensions belong to $\mathscr{C}^\gamma_{\mathrm{WF}}(\overline{\mathcal{U}}(\delta^0, x^0))$;

[13] There is a subtlety here: In fact, the distance used in [28, (1.32), (5.42)] is defined in terms of the coordinates that are used in the normal form of the operator, see footnote 12. So rigorously, we should not use the variables $(y_1, \cdots, y_{d-1})$ but the variables $(\arcsin^2(\sqrt{y_1}), \cdots, \arcsin^2(\sqrt{y_{d-1}}))$ in the definition of the distance. Fortunately, since we have the condition $y_1 + \cdots + y_{d-1} < 1 - \delta^0$, the change of variable $y_i \mapsto \arcsin^2(\sqrt{y_i})$ is a smooth diffeomorphism, from which we deduce that the distance (3.13) is equivalent to the same distance but with the new variables. And, in fact, once we have made the change of variables, there is another subtlety: The reader may indeed notice that the distance defined in (3.13) does not match the distance defined in [28, (1.32), (5.42)] since, for $j \notin I$, we should consider $|y_j - y'_j|$ instead of $|\sqrt{y_j} - \sqrt{y'_j}|$. Anyhow, since y_j and y'_j are here required to be away from 0, the distance used in [28] is equivalent to ours.

2. The function $\hat{h}$ is twice continuously differentiable on $\mathcal{U}_+(\delta^0, x^0) = \mathcal{U}(\delta^0, x^0) \cap \{(y_1, \cdots, y_{d-1}) \in (\mathbb{R}_+)^{d-1} : \forall i \in I, y_i > 0\}$. Moreover[14], for any $i, j \in I$ and any $k, l \notin I$,

$$\lim_{\min(y_i, y_j) \to 0_+} \sqrt{y_i y_j} \partial^2_{y_i y_j} \hat{h}(y) = 0, \quad \lim_{y_i \to 0_+} \sqrt{y_i} \partial^2_{y_i y_k} \hat{h}(y) = 0, \tag{3.14}$$

and the functions $y \mapsto \sqrt{y_i y_j} \partial^2_{y_i y_j} \hat{h}(y)$, $y \mapsto \sqrt{y_i} \partial^2_{y_i y_k} \hat{h}(y)$ and $y \mapsto \partial^2_{y_k y_l} \hat{h}(y)$ belong to $\mathscr{C}^{\gamma}_{\mathrm{WF}}(\overline{\mathcal{U}}(\delta^0, x^0))$ (meaning in particular that they can be extended by continuity to $\overline{\mathcal{U}}(\delta^0, x^0)$).

We then call

$$\begin{aligned} \|\hat{h}\|_{2+\gamma; \mathcal{U}(\delta^0, x^0)} := \|\hat{h}\|_{\gamma; \mathcal{U}(\delta^0, x^0)} + \sum_{i=1}^{d-1} \|\partial_{y_i} \hat{h}\|_{\gamma; \mathcal{U}(\delta^0, x^0)} + \sum_{i,j \in I} \| \sqrt{y_i y_j} \partial^2_{y_i y_j} \hat{h}\|_{\gamma; \mathcal{U}(\delta^0, x^0)} \\ + \sum_{k,l \notin I} \|\partial^2_{y_k y_l} \hat{h}\|_{\gamma; \mathcal{U}(\delta^0, x^0)} + \sum_{i \in I} \sum_{k \notin I} \| \sqrt{y_i} \partial^2_{y_i y_k} \hat{h}\|_{\gamma; \mathcal{U}(\delta^0, x^0)}, \end{aligned}$$

where $\sqrt{y_i y_j} \partial^2_{y_i y_j} \hat{h}$ is a shorten notation for $y \mapsto \sqrt{y_i y_j} \partial^2_{y_i y_j} \hat{h}(y)$ (and similarly for the others). For a given finite covering $\cup_{i=1}^K \mathcal{U}(\delta^0, x^{0,i})$ of $\hat{\mathcal{S}}_{d-1}$, a function $\hat{h}$ (or equivalently the associated function h defined on $\mathcal{S}_{d-1}$) is said to be in $\mathscr{C}^{2+\gamma}_{\mathrm{WF}}(\mathcal{S}_{d-1})$ if $\hat{h}$ belongs to each $\mathscr{C}^{2+\gamma}_{\mathrm{WF}}(\mathcal{U}(\delta^0, x^{0,i}))$. We then let

$$\|\hat{h}\|_{2+\gamma} := \sum_{i=1}^K \|\hat{h}\|_{2+\gamma; \mathcal{U}(\delta^0, x^{0,i})}.$$

We refer to [28, Chapter 10] for more details.

A similar definition holds for the space $\mathscr{C}^{1+\gamma/2, 2+\gamma}_{\mathrm{WF}}([0,T] \times \mathcal{S}_{d-1})$ of functions that are once continuously differentiable in time and twice continuously differentiable in space, with derivatives that are locally γ-Hölder continuous with respect to the time-space distance (in the local system of coordinates)

$$D((t,y),(t',y')) := |t - t'|^{1/2} + d(y, y'). \tag{3.15}$$

To make it clear, a function $\hat{h}$ defined on $[0,T] \times \overline{\mathcal{U}}(\delta^0, x^0)$ is then said to belong to $\mathscr{C}^{\gamma/2, \gamma}_{\mathrm{WF}}([0,T] \times \overline{\mathcal{U}}(\delta^0, x^0))$, for some $\gamma \in (0,1)$, if, in the new system of coordinates, $\hat{h}$ is Hölder continuous on $[0,T] \times \overline{\mathcal{U}}(\delta^0, x^0)$ with respect to the distance D. We then let

$$\left\|\hat{h}\right\|_{\gamma/2, \gamma; [0,T] \times \mathcal{U}(\delta^0, x^0)} := \sup_{(t,y) \in [0,T] \times \overline{\mathcal{U}}(\delta^0, x^0)} \left|\hat{h}(t,y)\right| + \sup_{t,t' \in [0,T],\ y, y' \in \overline{\mathcal{U}}(\delta^0, x^0)} \frac{|\hat{h}(t,y) - \hat{h}(t',y')|}{D((t,y),(t',y'))^{\gamma}}.$$

Following [28, Lemma 5.2.7], we say that a function $\hat{h}$ defined on $\mathcal{U}(\delta^0, x^0)$ belongs to the space $\mathscr{C}^{1+\gamma/2, 2+\gamma}_{\mathrm{WF}}([0,T] \times \mathcal{U}(\delta^0, x^0))$ if, in the new system of coordinates,

1. $\hat{h}$ is continuously differentiable on $[0,T] \times \mathcal{U}(\delta^0, x^0)$; $\hat{h}$ and its time and space derivatives extend continuously to $[0,T] \times \overline{\mathcal{U}}(\delta^0, x^0)$ and the resulting extensions belong to $\mathscr{C}^{\gamma/2, \gamma}_{\mathrm{WF}}([0,T] \times \overline{\mathcal{U}}(\delta^0, x^0))$;

[14] Similar to footnote 13, the reader should observe that, in [28, Lemma 5.2.5 and Definition 10.1.1], the two limits in (3.14) are in fact regarded in terms of the coordinates used in the normal form of the operator. To make it clear, we should here require $\lim_{\min(z_i, z_j) \to 0_+} \sqrt{z_i z_j} \partial^2_{z_i z_j} \hat{\ell}(z) = 0$ with $\hat{\ell}(z_1, \cdots, z_{d-1}) = \hat{h}(\sin^2(\sqrt{z_1}), \cdots, \sin^2(\sqrt{z_{d-1}}))$ (and similarly for the second limit). It is an easy exercise to check that (3.14) would then follow.

2. The function $\hat{h}$ 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 $k,l\notin I$,

$$\lim_{\min(y_i,y_j)\to 0_+}\sqrt{y_iy_j}\partial^2_{y_iy_j}\hat{h}(t,y)=0,\quad \lim_{y_i\to 0_+}\sqrt{y_i}\partial^2_{y_iy_k}\hat{h}(t,y)=0, \tag{3.16}$$

and the functions $(t,y)\mapsto\sqrt{y_iy_j}\partial^2_{y_iy_j}\hat{h}(t,y)$, $(t,y)\mapsto\sqrt{y_i}\partial^2_{y_iy_k}\hat{h}(y)$ and $(t,y)\mapsto\partial^2_{y_ky_l}\hat{h}(t,y)$ belong to $\mathscr{C}^{\gamma/2,\gamma}_{\mathrm{WF}}([0,T]\times\overline{\mathcal{U}}(\delta^0,x^0))$.

We then call

$$\begin{aligned}\|\hat{h}\|_{1+\gamma/2,2+\gamma;\mathcal{U}(\delta^0,x^0)}:=\|\hat{h}\|_{\gamma/2,\gamma;\mathcal{U}(\delta^0,x^0)}+\|\partial_t\hat{h}\|_{\gamma/2,\gamma;\mathcal{U}(\delta^0,x^0)}+\sum_{i=1}^{d-1}\|\partial_{y_i}\hat{h}\|_{\gamma/2,\gamma;\mathcal{U}(\delta^0,x^0)}\\+\sum_{i,j\in I}\|\sqrt{y_iy_j}\partial^2_{y_iy_j}\hat{h}\|_{\gamma/2,\gamma;\mathcal{U}(\delta^0,x^0)}+\sum_{k,l\notin I}\|\partial^2_{y_ky_l}\hat{h}\|_{\gamma/2,\gamma;\mathcal{U}(\delta^0,x^0)}\\+\sum_{i\in I}\sum_{k\notin I}\|\sqrt{y_i}\partial^2_{y_iy_k}\hat{h}\|_{\gamma/2,\gamma;\mathcal{U}(\delta^0,x^0)}.\end{aligned}$$

For a given finite covering $\cup_{i=1}^K\mathcal{U}(\delta^0,x^{0,i})$ of $\hat{\mathcal{S}}_{d-1}$, a function $\hat{h}$ (or equivalently the associated function h defined on $[0,T]\times\mathcal{S}_{d-1}$) is said to be in $\mathscr{C}^{1+\gamma/2,2+\gamma}_{\mathrm{WF}}([0,T]\times\mathcal{S}_{d-1})$ if $\hat{h}$ belongs to each $\mathscr{C}^{2+\gamma}_{\mathrm{WF}}([0,T]\times\mathcal{U}(\delta^0,x^{0,i}))$. We then let

$$\|\hat{h}\|_{1+\gamma/2,2+\gamma}:=\sum_{i=1}^K\|\hat{h}\|_{1+\gamma/2,2+\gamma;\mathcal{U}(\delta^0,x^{0,i})}.$$

We stress the fact that the finite covering that we use in the sequel is fixed once for all. There is no need to change it.

Remark 3.1. *Importantly, for $y,y'\in\bar{\mathcal{U}}(\delta^0,x^0)$, and for a constant $c\geq 1$, c depending on δ_0,*

$$\Big|\Big(1-\sum_{i=1}^{d-1}y_i\Big)^{1/2}-\Big(1-\sum_{i=1}^{d-1}y'_i\Big)^{1/2}\Big|\leq\frac{c-1}{2}\Big|\sum_{i=1}^{d-1}(y_i-y'_i)\Big|\leq(c-1)\sum_{i=1}^{d-1}\Big|\sqrt{y_i}-\sqrt{y'_i}\Big|.$$

Recalling that the vector $y=(y_1,\cdots,y_{d-1})$ (resp. $y'=(y'_1,\cdots,y'_{d-1})$) stand for the new coordinates of an element $x\in\hat{\mathcal{S}}_{d-1}$ (resp. x') and that x (resp. x') itself is canonically associated with $\check{x}=(x_1,\cdots,x_{d-1},1-x_1-\cdots-x_{d-1})\in\mathcal{S}_{d-1}$ (resp. $\check{x}'$), we deduce that

$$c^{-1}\sum_{i=1}^d\Big|\sqrt{\check{x}_i}-\sqrt{\check{x}'_i}\Big|\leq d(y,y')\leq c\sum_{i=1}^d\Big|\sqrt{\check{x}_i}-\sqrt{\check{x}'_i}\Big|,$$

which permits to reformulate the modulus of continuity showing up in the Hölder condition of the Wright–Fisher space in an intrinsic manner. Since the number of neighborhoods of the form $\mathcal{U}(\delta^0,x^0)$ used to cover $\hat{\mathcal{S}}_{d-1}$ is finite, we can choose the same c for all those neighborhoods.

3.2.3. Back to the master equation

Actually, we must point out that, in order to apply Theorem 10.0.2 in [28] (existence of classical solutions to Kimura type PDEs together with related Schauder's estimates), as we do later, the master equation should be satisfied also in boundary points, under the appropriate local chart; see (10.1) therein. This would require to formulate (3.6) for each local chart used in the construction of the Wright–Fischer spaces, that is, for any projection p^{-i} with $i \in [\![d]\!]$, since we verified in (3.12) that those changes of variables make the second order operator fit the setup of [28]. In this respect, we observe that there is no hindrance for us to restate (3.6) in the right system of coordinates. Also, we already noticed that, as long as we look for solutions $U^1, \cdots, U^d$ in the space $\mathscr{C}^{1+\gamma/2,2+\gamma}_{\mathrm{WF}}([0,T] \times \mathcal{S}_{d-1})$ for some $\gamma \in (0,1)$, the first order derivatives always extend by continuity up to the boundary. Moreover, even though the second order derivatives are defined only in the interior of the simplex and are allowed to blow-up at the boundary, the rate of explosion of those second order derivatives, as prescribed by (3.16), combine well with the degeneracy property of the operator on the boundary. Hence, by a standard continuity argument, it is enough to require that (3.7), which is written in terms of the intrinsic derivative, holds in $\mathrm{Int}(\hat{\mathcal{S}}_{d-1})$.

3.3. Complete versions of the main statements

We now provide refined versions of Meta-Theorems 2.7 and 2.9.

3.3.1. Existence and uniqueness of an MFG equilibrium

We have indeed all the ingredients to clarify Meta-Theorem 2.7 and to formulate the statement in a rigorous manner. We recall that the proof is given in Section 4.2.

Theorem 3.2. *Assume that, for some $\gamma > 0$, each f^i, for $i \in [\![d]\!]$, belongs to $\mathscr{C}^{\gamma/2,\gamma}_{\mathrm{WF}}([0,T]\times\mathcal{S}_{d-1})$, and each g^i, for $i \in [\![d]\!]$, belongs to $\mathscr{C}^{2+\gamma}_{\mathrm{WF}}(\mathcal{S}_{d-1})$. Then, for any $\varepsilon \in (0,1)$, there exists a threshold $\kappa_0 > 0$, only depending on ε, $\|f\|_\infty$, $\|g\|_\infty$ and T, such that, for any $\kappa \geq \kappa_0$ and $\delta \in (0, 1/(4\sqrt{d}))$, and for any (deterministic) initial condition $(p_{0,i})_{i\in[\![d]\!]} \in \mathcal{S}_{d-1}$ with positive entries, the mean field game has a unique solution, in the sense of Definition 2.6.*

Remark 3.3. *As already emphasized in Remark 2.8, we could certainly extend the uniqueness result to the larger class of open-loop strategies and also to the case when the initial condition is random.*

As for the assumptions on the coefficients, the key fact is that there is no need for any monotonicity condition in the statement. Still, it would be interesting to see whether the result remains true under lower regularity conditions on the function g. Assuming g to have two Hölder continuous derivatives (as we do here) is quite convenient since it allows to find a solution to the master equation (see the next section) that remains smooth up to the boundary at time T. More effort would be needed to allow for more general (and hence less regular) terminal costs; accordingly, it would require to address with care the rate at which the derivatives of the solution to the master equation would blow up at terminal time. We leave this problem for future work.

3.3.2. Existence and uniqueness of a solution to the master equation

Here is now the rigorous version of Meta-Theorem 2.9, the proof of which is given in Section 4.3.

Theorem 3.4. *Assume that, for some $\gamma \in (0, 1/2]$, each f_i, for $i \in [\![d]\!]$, belongs to $\mathscr{C}^{\gamma/2,\gamma}_{\mathrm{WF}}([0,T] \times \mathcal{S}_{d-1})$, and each g_i, for $i \in [\![d]\!]$, belongs to $\mathscr{C}^{2+\gamma}_{\mathrm{WF}}(\mathcal{S}_{d-1})$. Then, for any $\varepsilon \in (0,1)$, there exist a universal exponent $\eta \in (0,1)$ (hence independent of ε) and a threshold $\kappa_0 > 0$, only depending on ε, $\|f\|_\infty$, $\|g\|_\infty$ and T, such that, for any $\kappa \geq \kappa_0$ and $\delta \in (0, 1/(4\sqrt{d}))$, the master equation (3.6)–(3.7) has a solution in $[\mathscr{C}^{1+\gamma'/2,2+\gamma'}_{\mathrm{WF}}([0,T] \times \mathcal{S}_{d-1})]^d$, for $\gamma' = \min(\gamma, \eta)/2$.*

Moreover, the solution is unique in the sense that $(U^1, \cdots, U^d)$ is the unique tuple such that, for any $i \in [\![d]\!]$, U^i belongs to the Wright–Fisher space $\mathscr{C}^{1+\gamma'/2,2+\gamma'}_{\mathrm{WF}}([0,T] \times \mathcal{S}_{d-1})$ for some $\gamma' > 0$, and (3.6)–(3.7) hold at any $(t,p) \in [0,T] \times \mathrm{Int}(\hat{\mathcal{S}}_{d-1})$.

Remark 3.5. *We do not provide any new version of Theorem 2.10. The reader will find in (4.12) a more suitable version of equation (2.27), when expressed in intrinsic derivatives. We let her/him adapt accordingly the definition of the space $C^{1,2}([0,T] \times \mathrm{Int}(\hat{\mathcal{S}}_{d-1}), \mathbb{R})$ that is used in the statement.*

3.4. Proofs of auxiliary results from Section 2

We now prove the auxiliary results stated in §2.2.1.

3.4.1. Proof of Proposition 2.2

The proof holds in two steps. We give a sketch of it only.

First Step. By (2.15), it suffices to solve the equation for $(P^1_t, \cdots, P^{d-1}_t)_{0 \leq t \leq T}$. As long as the latter stays in the interior of $\hat{\mathcal{S}}_{d-1}$, the equation satisfied by the process is non-degenerate, see (3.11). Moreover, the diffusion matrix is Lipschitz away from the boundary of the simplex and the drift is bounded. Therefore, by a standard localization argument, we can easily adapt the strong existence and uniqueness result of Veretennikov [58] (see Remark 3 therein for the case when the initial condition is random and Remark 4 therein for the case when the state variable and the underpinning Brownian motion do not have the same dimension) and then deduce that, up until one coordinate (including the dth coordinate, as given by $(P^d_t = 1 - (P^1_t + \cdots + P^{d-1}_t))_{0 \leq t \leq T}$) reaches a given positive threshold ϵ, a (strong) solution exists and is pathwise unique. By letting ϵ tend to 0, we deduce that there exists a solution up to the first time it reaches the boundary of the simplex (or, equivalently, one of the coordinates vanishes) and this solution is pathwise unique (once again, up to the first time it reaches the boundary).

Second Step. The goal in this step is to prove that, for $\kappa \geq \varepsilon^2/2$, the solution of (2.17) until the first time it reaches the boundary of $\hat{\mathcal{S}}_{d-1}$ stays in fact away from the boundary of $\hat{\mathcal{S}}_{d-1}$. As a consequence, the solution of (2.17) is well defined on the entire $[0,T]$. To do so, we come back to (2.16), namely, we write the dynamics of the ith coordinate (for $i = 1, \cdots, d-1$) in the form

$$\mathrm{d}P^i_t = a_i(t, (P^1_t, \cdots, P^d_t))\mathrm{d}t + \varepsilon \sqrt{P^i_t(1 - P^i_t)}\mathrm{d}\widetilde{W}^i_t,$$

with a_i as in (2.20). The above holds true up until the first time $\tau := \inf\{t \in [0,T] : (P^1_t, \cdots, P^{d-1}_t) \in \partial\hat{\mathcal{S}}_{d-1}\} \wedge T$. Then, using the fact that α is bounded and $\alpha(t,j,x)(i) \geq 0$ for $i \neq j$ in (2.20), we can easily compare the process $(P^i_t)_{0 \leq t \leq \tau}$ with the solution of the equation

$$\mathrm{d}\overline{P}^i_t = \big(\varphi(\overline{P}^i_t) - C\overline{P}^i_t\big)\mathrm{d}t + \varepsilon \sqrt{(\overline{P}^i_t)_+(1 - \overline{P}^i_t)_+}\mathrm{d}\widetilde{W}^i_t,$$

with $\bar{p}^i_0 = p^i_0$ as initial condition, for a constant $C \geq 0$. Above, $(\cdot)_+$ stands for the positive part. By [42, Chapter 5, Proposition 2.13], the above equation has a unique strong solution. Letting

$\bar\tau^i := \inf\{t \in [0,T] : \overline{P}^i_t \in \{0,1\}\} \wedge T$ and choosing C large enough, we have $\overline{P}^i_t \le P^i_t$, for all $t \le \tau \wedge \bar\tau^i$. We then apply Feller's test (see [42, Chapter 5, Proposition 5.22]) to $(\overline{P}^i_t)_{0\le t\le T}$ (the reader may notice that the fact that the initial condition is random is not a hindrance since it belongs to $(0,1)$ with probability 1). The natural scale (see [42, Chapter 5, (5.42)]) is here given by

$$\Phi(r) := \int_\delta^r \exp\Big(-2\int_\delta^s \frac{\varphi(u)-Cu}{\varepsilon^2 u(1-u)}\mathrm{d}u\Big)\mathrm{d}s, \quad r \in (0,1).$$

For $r \in (0,\delta]$,

$$-\Phi(r) \ge \int_r^\delta \exp\Big(2\frac{\kappa}{\varepsilon^2}\ln(\frac{\delta}{s}) - \frac{2C\delta}{\varepsilon(1-\delta)}\Big)\mathrm{d}s,$$

from which we get that $\Phi(0+) = -\infty$ if $2\kappa/\varepsilon^2 \ge 1$. We deduce that, if the latter is true, $(\overline{P}^i_t)_{0\le t\le T}$ does not touch 0. By comparison, we deduce that $(P^i_t)_{0\le t\le\tau}$ cannot touch 0 before it touches 1, that is $P^i_\tau = 1$ if the set $\{t \in [0,\tau] : P^i_t \in \{0,1\}\}$ is not empty. This holds true for $i = 1,\cdots,d-1$, but by choosing another system of coordinates, we get the same result for the coordinate $i = d$. Assume now that we can find some coordinate $i \in [\![d]\!]$ such that $P^i_\tau \in \{0,1\}$, in which case our analysis says that $P^i_\tau = 1$. Since $\sum_{j=1}^d P^j_\tau = 1$, we deduce that $P^j_\tau = 0$ for all $j \in [\![d]\!] \setminus \{i\}$, which is a contradiction with our analysis. So, the conclusion is that, at time τ, we must have $P^i_\tau \in (0,1)$ for all $i \in [\![d]\!]$. That is, $\tau = T$ and the process $(P_t = (P^1_t,\cdots,P^d_t))_{0\le t\le T}$ remains in the $((d-1)$-dimensional) interior of $\mathcal{S}_{d-1}$. □

3.4.2. Proof of Proposition 2.3

As in the proof of Proposition 2.2, we write the equation for $(P^i_t)_{0\le t\le T}$ (for a given $i \in [\![d]\!]$, i being possibly equal to d) in the form

$$\mathrm{d}P^i_t = a_i(t,P_t)\mathrm{d}t + \varepsilon\sqrt{P^i_t(1-P^i_t)}\mathrm{d}\widetilde{W}^i_t,$$

with a_i as in (2.20). Then, we get, by Itô's formula (recall that the left-hand side below is well-defined since $(P^i_t)_{0\le t\le T}$ does not vanish),

$$\begin{aligned}\mathrm{d}[\ln P^i_t] = &\sum_{j\in[\![d]\!]}\Big[\frac{P^j_t}{P^i_t}\Big(\varphi(P^i_t)+\alpha(t,j,P_t)(i)\Big) - \Big(\varphi(P^j_t)+\alpha(t,i,P_t)(j)\Big)\Big]\mathrm{d}t - \frac{\varepsilon^2}{2}\frac{1-P^i_t}{P^i_t}\mathrm{d}t \\ &+ \varepsilon\sqrt{\frac{1-P^i_t}{P^i_t}}\mathrm{d}\widetilde{W}^i_t.\end{aligned} \tag{3.17}$$

For a constant C depending on the same parameters as those quoted in the statement, we can lower bound the drift in (3.17) as follows (using the definition of φ in (2.19) together with the fact that $\alpha(t,j,P_t)(i)$ is non-negative if $j \ne i$)

$$\begin{aligned}&\sum_{j\in[\![d]\!]}\Big[\frac{P^j_t}{P^i_t}\Big(\varphi(P^i_t)+\alpha(t,j,P_t)(i)\Big) - \Big(\varphi(P^j_t)+\alpha(t,i,P_t)(j)\Big)\Big] - \frac{\varepsilon^2}{2}\frac{1-P^i_t}{P^i_t} \\ &\ge \frac{1-P^i_t}{P^i_t}\kappa\mathbf{1}_{\{P^i_t\le\delta\}} - \frac{\varepsilon^2}{2}\frac{1-P^i_t}{P^i_t} - C.\end{aligned}$$

Allowing the value of C to change from line to line and recalling that $\kappa \geq \varepsilon^2/2$, we get

$$\sum_{j\in[\![d]\!]}\Big[\frac{P_t^j}{P_t^i}\Big(\varphi(P_t^i)+\alpha(t,j,P_t)(i)\Big)-\Big(\varphi(P_t^j)+\alpha(t,i,P_t)(j)\Big)\Big]-\frac{\varepsilon^2}{2}\frac{1-P_t^i}{P_t^i}$$
$$\geq \frac{1-P_t^i}{P_t^i}\Big(\kappa-\frac{\varepsilon^2}{2}\Big)-C.$$

Hence, integrating (3.17) from 0 to some stopping time τ (with values in $[0,T]$), adding and subtracting the compensator $(\lambda\varepsilon^2/2)\int_0^\tau(1-P_t^i)/P_t^i\mathrm{d}t$, multiplying by λ and then taking the exponential, we get

$$\begin{aligned}&(P_\tau^i)^\lambda\exp\Big(-\lambda\varepsilon\int_0^\tau\sqrt{\frac{1-P_t^i}{P_t^i}}\mathrm{d}\widetilde{W}_t^i-\frac{\lambda^2\varepsilon^2}{2}\int_0^\tau\frac{1-P_t^i}{P_t^i}\mathrm{d}t\Big)\\&\geq(p_{0,i})^\lambda\exp\Big(\lambda\Big[\kappa-\frac{\varepsilon^2(1+\lambda)}{2}\Big]\int_0^\tau\frac{1}{P_t^i}\mathrm{d}t-C\Big).\end{aligned}\tag{3.18}$$

Choosing $\tau=\inf\{t\in[0,T]:P_t^i\leq\epsilon\}\wedge T$, for $\epsilon>0$ as small as needed, the left-hand side has conditional expectation less than 1. So, taking expectation and letting ϵ tend to 0, we deduce that

$$\mathbb{E}\left[\exp\Big(\lambda\Big[\kappa-\frac{\varepsilon^2(1+\lambda)}{2}\Big]\int_0^T\frac{1}{P_t^i}\mathrm{d}t\Big)\right]\leq C(p_{0,i})^{-\lambda}.$$

The bound (2.21) easily follows.

It then remains to prove (2.22). To do so, we come back to (3.18). Using the fact that γ is positive and choosing $\tau=t\in[0,T]$, we rewrite it in the form

$$(P_t^i)^{-\lambda}\leq(p_{0,i})^{-\lambda}\exp\Big(C-\lambda\varepsilon\int_0^t\sqrt{\frac{1-P_s^i}{P_s^i}}\mathrm{d}\widetilde{W}_t^i-\frac{\lambda^2\varepsilon^2}{2}\int_0^t\frac{1-P_s^i}{P_s^i}\mathrm{d}s\Big).$$

Taking expectations on both sides, we easily complete the proof of (2.22). □

3.4.3. Proof of Proposition 2.4

For each $i\in[\![d]\!]$, we call $(\mathcal{E}_t^i)_{0\leq t\leq T}$ the Doléans–Dade exponential

$$\mathcal{E}_t^i:=\exp\Big(\varepsilon\sum_{j\in[\![d]\!]}\int_0^t\sqrt{\frac{P_s^j}{P_s^i}}\mathrm{d}\overline{W}_s^{i,j}-\frac{\varepsilon^2}{2}\int_0^t\frac{1-P_s^i}{P_s^i}\mathrm{d}s\Big),\quad t\in[0,T].$$

Then, $(Q_t^i)_{0\leq t\leq T}$ is a solution to (2.18) if and only if

$$\mathrm{d}[(\mathcal{E}_t^i)^{-1}Q_t^i]=\sum_{j\in[\![d]\!]}(\mathcal{E}_t^i)^{-1}\Big(Q_t^j(\varphi(P_t^i)+\beta(t,j,P_t)(i))-Q_t^i(\varphi(P_t^j)+\beta(t,i,P_t)(j))\Big)\mathrm{d}t,$$

which may be rewritten in the form

$$\mathrm{d}\widetilde{Q}_t^i=\sum_{j\in[\![d]\!]}\Big((\mathcal{E}_t^i)^{-1}\mathcal{E}_t^j\widetilde{Q}_t^j(\varphi(P_t^i)+\beta(t,j,P_t)(i))-\widetilde{Q}_t^i(\varphi(P_t^j)+\beta(t,i,P_t)(j))\Big)\mathrm{d}t,\tag{3.19}$$

for $t \in [0,T]$, with $(\widetilde{Q}^i_0 = q_{0,i})_{i\in[\![d]\!]} \in \mathcal{S}_{d-1}$ as initial condition and under the change of variable

$$\widetilde{Q}^i_t := (\mathcal{E}^i_t)^{-1} Q^i_t, \quad t \in [0,T]. \tag{3.20}$$

Obviously, (3.19) has a unique pathwise solution. It is continuous and adapted to the filtration $\mathbb{F}^{\boldsymbol{W}}$. Since $\beta(t,j,P_t)(i) \geq 0$ for $j \neq i$, it is pretty easy to check that all the coordinates remain (strictly) positive.

Given the solution to (3.19), we may reconstruct $((Q^i_t)_{0\leq t\leq T})_{i\in[\![d]\!]}$ from the change of variable (3.20). Then, taking the power l in (2.18), for an exponent $l \geq 1$, we get

$$\begin{aligned} \mathrm{d}(Q^i_t)^l = {} & l \sum_{j\in[\![d]\!]} (Q^i_t)^{l-1}\Big(Q^j_t(\varphi(P^i_t) + \beta(t,j,P_t)(i)) - Q^i_t(\varphi(P^j_t) + \beta(t,i,P_t)(j))\Big)\mathrm{d}t \\ & + \frac{l(l-1)}{2}\varepsilon^2 (Q^i_t)^l \frac{1-P^i_t}{P^i_t}\mathrm{d}t + \varepsilon l (Q^i_t)^l \sum_{j\in[\![d]\!]} \sqrt{\frac{P^j_t}{P^i_t}}\mathrm{d}\overline{W}^{i,j}_t, \end{aligned} \tag{3.21}$$

for $t \in [0,T]$. As a result, we can find a constant C, only depending on l, κ and on the supremum norm of β, such that

$$\mathrm{d}\Big[\sum_{i\in[\![d]\!]} (Q^i_t)^l\Big] \leq \Big[C + \varepsilon^2 \frac{l(l-1)}{2} \sum_{j\in[\![d]\!]} \frac{1-P^j_t}{P^j_t}\Big] \cdot \Big[\sum_{i\in[\![d]\!]} (Q^i_t)^l\Big]\mathrm{d}t + \mathrm{d}m_t,$$

where $(m_t)_{0\leq t\leq T}$ is a local martingale[15]. By a standard localization argument, we end up with

$$\sup_{0\leq t\leq T} \mathbb{E}\Big[\Big(\sum_{i\in[\![d]\!]} (Q^i_t)^l\Big)\exp\Big(-\varepsilon^2 \frac{l(l-1)}{2}\int_0^T \sum_{j\in[\![d]\!]} \frac{1-P^j_s}{P^j_s}\mathrm{d}s\Big)\Big] \leq C,$$

for a new value of C. And then, applying Cauchy-Schwarz inequality and invoking the above inequality with $2l$ instead of l, we get

$$\sup_{0\leq t\leq T} \mathbb{E}\Big[\sum_{i\in[\![d]\!]} (Q^i_t)^l\Big] \leq C \sup_{0\leq t\leq T} \mathbb{E}\Big[\exp\Big(\varepsilon^2 l(2l-1)\int_0^T \sum_{j\in[\![d]\!]} \frac{1-P^j_t}{P^j_t}\mathrm{d}t\Big)\Big]^{1/2}.$$

Take $l = 8$ and choose $\gamma \geq 60\varepsilon^2$ and $\lambda = 2d$ in the statement of Proposition 2.3 (which is indeed possible since $\kappa - (1+\lambda)\varepsilon^2/2 \geq (61+d)\varepsilon^2 - (1/2+d)\varepsilon^2 \geq 60\varepsilon^2$). Then, by Hölder's inequality (in order to handle the sum inside the exponential), the right-hand side is upper bounded. Returning to (3.21), invoking Burkholder–Davis–Gundy inequalities, (2.22) and the bound $\sup_{0\leq t\leq T}\mathbb{E}[(Q^i_t)^4/P^i_t] \leq \sup_{0\leq t\leq T}\mathbb{E}[(Q^i_t)^8]^{1/2}\sup_{0\leq t\leq T}\mathbb{E}[(P^i_t)^{-2}]^{1/2}$, for $i \in [\![d]\!]$, we deduce that $\sup_{0\leq t\leq T}|Q^i_t|$ has a finite fourth moment for each $i \in [\![d]\!]$. Equality (2.23) is easily proved by summing over $i \in [\![d]\!]$ in (2.18). □

4. From the MFG system to the master equation

This section is dedicated to the proofs of Theorems 3.2 and 3.4, taken for granted the statement of Theorem 2.10. Throughout the section, we assume that the condition $\kappa \geq (61+d)\varepsilon^2$ is in force.

[15] Here and throughout, the notation $\mathrm{d}X^1_t \geq \mathrm{d}X^2_t$, $t \in [0,T]$, for two stochastic processes $((X^i_t)_{0\leq t\leq T})_{i=1,2}$, is understood as $(X^2_t - X^1_t)_{0\leq t\leq T}$ is a non-decreasing process.

4.1. MFG system

With the optimization problem driven by the cost functional (2.12) and the state equation (2.24) (within the environment (2.17)), we may associate a value function. Obviously, we may expect this value function to solve a stochastic (because of the common noise) variant of the usual Hamilton–Jacobi–Bellman equation for a stochastic optimal control problem on a discrete state space. The combination of this Stochastic Hamilton–Jacobi–Bellman (SHJB) equation with the equation (2.17) for the environment will lead us to a relevant version of the so-called MFG system (which is a key tool in the standard theory of mean field games, see for instance the references [9, 10, 12, 48, 49, 50]).

4.1.1. Formulation of the system

In order to proceed, we recall (2.26). Importantly, $(P_t)_{0\leq t\leq T}$ therein is regarded as a stochastic environment. Typically, it is the solution of an equation of the form (2.17). In any case, it is a continuous $\mathcal{S}_{d-1}$-valued process that is progressively-measurable with respect to the filtration $\mathbb{F}^{\boldsymbol{W}}$ and that satisfies the conclusion of Proposition 2.3, see (2.21) and (2.22). In particular, it remains away from the boundary of the simplex.

The related value function at time $t \in [0,T]$ is defined as

$$\begin{aligned} &u^l\Big(t,(P_s)_{t\leq s\leq T}\Big) := \operatorname*{ess\,inf}_{(\beta_s)_{t\leq s\leq T}} \mathcal{J}^l\Big(t,(\beta_s)_{t\leq s\leq T},(P_s)_{t\leq s\leq T}\Big), \quad l\in[\![d]\!],\\ &\mathcal{J}^l\Big(t,(\beta_s)_{t\leq s\leq T},(P_s)_{t\leq s\leq T}\Big)\\ &\quad := \sum_{i\in[\![d]\!]}\mathbb{E}\Big[Q^i_T[t,l]g^i(P_T)+\int_t^T\Big(Q^i_s[t,l][f^i(s,P_s)+\frac{1}{2}\sum_{j\neq i}|\beta^{i,j}_t|^2]\Big)\mathrm{d}s\,\Big|\,\mathcal{F}^{\boldsymbol{W}}_t\Big], \quad l\in[\![d]\!], \end{aligned} \tag{4.1}$$

whereas $(Q^i_s[t,l])_{t\leq s\leq T}$ is the solution to (2.24) when the initial time is $t\in[0,T)$ and the initial distribution is $Q^i_t[t,l]=\delta_{l,i}$, for $i\in[\![d]\!]$. Importantly, the value function is random: Stochasticity accounts for the fact that the cost functionals f and g in the optimal control problem depend upon the environment $(P_s)_{0\leq s\leq T}$, which is random itself. Hence the corresponding HJB equation is a backward stochastic HJB equation (SHJB) that takes the form of a system of backward SDEs indexed by $i\in[\![d]\!]$:

$$\begin{aligned} \mathrm{d}u^i_t &= -\Big(\sum_{j\in[\![d]\!]}\varphi(P^j_t)[u^j_t-u^i_t]+H^i(u_t)+f^i(t,P_t)\Big)\mathrm{d}t-\frac{\varepsilon}{\sqrt{2}}\sum_{j\in[\![d]\!]:j\neq i}\sqrt{\frac{P^j_t}{P^i_t}}(\nu^{i,i,j}_t-\nu^{i,j,i}_t)\mathrm{d}t\\ &\quad+\sum_{j,k\in[\![d]\!]:j\neq k}\nu^{i,j,k}_t\mathrm{d}W^{j,k}_t,\\ u^i_T &= g^i(P_T), \end{aligned} \tag{4.2}$$

where H^i is the Hamiltonian

$$H^i(y) := -\frac{1}{2}\sum_{j\in[\![d]\!]}(y_i-y_j)^2_+, \quad y=(y_j)_{j\in[\![d]\!]}. \tag{4.3}$$

It is worth emphasizing that, in the equation (4.2), the unknown is the larger family of processes $((u^i_t)_{i\in[\![d]\!]},(\nu^{i,j,k}_t)_{i,j,k\in[\![d]\!]:j\neq k})_{0\leq t\leq T}$, which are required to be progressively measurable with respect to $\mathbb{F}^{\boldsymbol{W}}$. This is a standard fact in the theory of backward SDEs and the role of the processes $((\nu^{i,j,k}_t)_{i,j,k\in[\![d]\!]:j\neq k})_{0\leq t\leq T}$ is precisely to force the solution of the stochastic HJB equation to be non-anticipative. The reason why we here choose indices (i,j,k) with $j\neq k$ is quite clear: there are no noises of the form $((W^{j,j}_t)_{0\leq t\leq T})_{j\in[\![d]\!]}$ in the forward equation.

4.1.2. Verification argument

The following verification argument clarifies the connection between (4.1) and (4.2).

Lemma 4.1. *For an environment $\boldsymbol{P} = (P_t)_{0\le t\le T}$ as before (satisfying in particular* (2.21) *and* (2.22)*), assume that there exists a solution $((u_t^i)_{i\in[\![d]\!]}, (\nu_t^{i,j,k})_{i,j,k\in[\![d]\!]:\, j\neq k})_{0\le t\le T}$ to* (4.2) *such that $((u_t^i)_{i\in[\![d]\!]})_{0\le t\le T}$ is bounded (by a deterministic constant) and*

$$\sum_{i,j,k\in[\![d]\!]:\, j\neq k} \mathbb{E}\int_0^T |\nu_t^{i,j,k}|^2 \mathrm{d}t < \infty.$$

For $t\in[0,T]$, let $(\beta_s^\star)^{i,j} := (u_s^i - u_s^j)_+$ for $i\neq j$ and $s\in[t,T]$. Then, $\mathcal{J}^l(t,(\beta_s^\star)_{t\le s\le T},(P_s)_{t\le s\le T}) = u_t^l$ and, for any other (bounded) strategy, say $(\beta_s)_{t\le s\le T}$, such that

$$\sum_{i,j\in[\![d]\!]:i\neq j}\int_t^T \mathbb{P}(\beta_s^{i,j} \neq (\beta_s^\star)^{i,j})\mathrm{d}s > 0,$$

the cost $\mathcal{J}^l(t,(\beta_s)_{t\le s\le T},(P_s)_{t\le s\le T})$ is strictly higher than u_t^l.

In words, the above says that $(((\beta_s^\star)^{i,j} = (u_s^i - u_s^j)_+)_{i,j\in[\![d]\!]:i\neq j})_{t\le s\le T}$ is the unique optimal control. In fact, the solvability of the equation (4.2) is guaranteed by the following lemma.

Lemma 4.2. *For $\boldsymbol{P} = (P_t)_{0\le t\le T}$ as before (satisfying* (2.21) *and* (2.22)*),* (4.2) *has a unique (progressively-measurable) solution $((u_t^i)_{i\in[\![d]\!]}, (\nu_t^{i,j,k})_{i,j,k\in[\![d]\!]:\, j\neq k})_{0\le t\le T}$ such that $((u_t^i)_{i\in[\![d]\!]})_{0\le t\le T}$ is almost surely bounded by a deterministic constant and $((\nu_t^{i,j,k})_{i,j,k\in[\![d]\!]:\, j\neq k})_{0\le t\le T}$ satisfies*

$$\sum_{i,j,k\in[\![d]\!]:\, j\neq k}\mathbb{E}\Big[\int_0^T \exp\Big(\varepsilon^2\sum_{l\in[\![d]\!]}\int_0^t \frac{1}{P_s^l}\mathrm{d}s\Big)|\nu_t^{i,j,k}|^2\mathrm{d}t\Big] < \infty.$$

Abusively, such a solution is said to be bounded.

The two lemmas will be proved in Subsection 4.1.3 below. For the time being, we observe, by combining the two of them, that, for a given $\boldsymbol{P} = (P_t)_{0\le t\le T}$ satisfying (2.21) and (2.22), the solution to the optimal control problem (2.26) is entirely described by the SHJB equation (4.2), as it suffices to solve the forward equation (2.24) with $((\beta_t^{i,j} = (u_t^i - u_t^j)_+)_{i,j\in[\![d]\!]:i\neq j})_{0\le t\le T}$ therein. Now, Definition 2.6 implies that an environment $\boldsymbol{P}$ is a solution to the MFG in hand if and only if it solves the forward equation in the forward-backward system:

$$\begin{aligned}
\mathrm{d}P_t^i &= \sum_{j\in[\![d]\!]}\Big(P_t^j(\varphi(P_t^i) + (u_t^j - u_t^i)_+) - P_t^i(\varphi(P_t^j) + (u_t^i - u_t^j)_+)\Big)\mathrm{d}t + \varepsilon\sum_{j\in[\![d]\!]}\sqrt{P_t^iP_t^j}\mathrm{d}\overline{W}_t^{i,j},\\
\mathrm{d}u_t^i &= -\Big(\sum_{j\in[\![d]\!]}\varphi(P_t^j)[u_t^j - u_t^i] + H^i(u_t) + f^i(t,P_t)\Big)\mathrm{d}t \qquad (4.4)\\
&\quad - \frac{\varepsilon}{\sqrt{2}}\sum_{j\in[\![d]\!]:\, j\neq i}\sqrt{\frac{P_t^j}{P_t^i}}(\nu_t^{i,i,j} - \nu_t^{i,j,i})\mathrm{d}t + \sum_{j,k\in[\![d]\!]:\, j\neq k}\nu_t^{i,j,k}\mathrm{d}W_t^{j,k},
\end{aligned}$$

with $(P_0^i = p_{0,i})_{i\in[\![d]\!]} \in \mathcal{S}_{d-1}$ as deterministic initial condition for the forward equation and $(u_T^i = g^i(P_T))_{i\in[\![d]\!]}$ as terminal boundary condition for the backward equation. System (4.4) is the (stochastic) MFG system that characterizes the solutions of the MFG described in Definition 2.6. Hence, proving Theorem 3.2 is here the same as proving that (4.4) is uniquely solvable (within the space of processes that satisfy the conditions described in Proposition 2.4 and Lemma 4.2).

4.1.3. Proofs of Lemmas 4.1 and 4.2

Proof of Lemma 4.1. Call $((Q^i_s)_{i\in[\![d]\!]})_{t\le s\le T}$ the solution to (2.24) with $Q^i_t = \delta_{i,l}$ for some $l \in [\![d]\!]$ and expand

$$\begin{aligned}
&\mathrm{d}\Big(\sum_{i\in[\![d]\!]} Q^i_s u^i_s + \int_t^s \sum_{i\in[\![d]\!]} Q^i_r(f^i(r,P_r) + \frac{1}{2}\sum_{j\neq i}|\beta^{i,j}_r|^2)\mathrm{d}r\Big)\\
&= -\sum_{i\in[\![d]\!]} Q^i_s\Big(\sum_{j\in[\![d]\!]}\varphi(P^j_s)[u^j_s - u^i_s] + H^i(u_s)\Big)\mathrm{d}s - \frac{\varepsilon}{\sqrt{2}}\sum_{i\in[\![d]\!]} Q^i_s \sum_{j\neq i}\sqrt{\frac{P^j_s}{P^i_s}}(\nu^{i,i,j}_s - \nu^{i,j,i}_s)\mathrm{d}s\\
&\quad + \sum_{i\in[\![d]\!]} u^i_s \sum_{j\in[\![d]\!]}\Big(Q^j_s(\varphi(P^i_s) + \beta^{j,i}_s) - Q^i_s(\varphi(P^j_s) + \beta^{i,j}_s)\Big)\mathrm{d}s + \frac{1}{2}\sum_{i\in[\![d]\!]} Q^i_s\sum_{j\neq i}|\beta^{i,j}_s|^2\mathrm{d}s\\
&\quad + \frac{\varepsilon}{\sqrt{2}}\sum_{i\in[\![d]\!]} Q^i_s\sum_{j\in[\![d]\!]}\sqrt{\frac{P^j_s}{P^i_s}}d[W^{i,j}_s - W^{j,i}_s]\cdot\sum_{j,k\in[\![d]\!]:\,j\neq k}\nu^{i,j,k}_s\mathrm{d}W^{j,k}_s + \mathrm{d}m_s,
\end{aligned} \tag{4.5}$$

where $(m_s)_{t\le s\le T}$ is a uniformly integrable martingale. On the last line, the dot in the first term is used to compute the underlying bracket. On the second line,

$$\sum_{i\in[\![d]\!]} u^i_s\sum_{j\in[\![d]\!]}\Big(Q^j_s\varphi(P^i_s) - Q^i_s\varphi(P^j_s)\Big) = \sum_{i\in[\![d]\!]}\sum_{j\in[\![d]\!]} Q^j_s\varphi(P^i_s)(u^i_s - u^j_s) = \sum_{i\in[\![d]\!]}\sum_{j\in[\![d]\!]} Q^i_s\varphi(P^j_s)(u^j_s - u^i_s),$$

which cancels out with the first term on the first line. Moreover,

$$\begin{aligned}
&-\sum_{i\in[\![d]\!]} Q^i_s H^i(u_s) + \sum_{i,j\in[\![d]\!]} u^i_s(Q^j_s\beta^{j,i}_s - Q^i_s\beta^{i,j}_s) + \frac{1}{2}\sum_{i\in[\![d]\!]} Q^i_s\sum_{j\neq i}|\beta^{i,j}_s|^2\\
&= \frac{1}{2}\sum_{i\in[\![d]\!]} Q^i_s\sum_{j\neq i}(u^i_s - u^j_s)^2_+ - \sum_{i\in[\![d]\!]} Q^i_s\sum_{j\neq i}\beta^{i,j}_s(u^i_s - u^j_s) + \frac{1}{2}\sum_{i\in[\![d]\!]} Q^i_s\sum_{j\neq i}|\beta^{i,j}_s|^2\\
&\ge \frac{1}{2}\sum_{i\in[\![d]\!]} Q^i_s\sum_{j\neq i}(u^i_s - u^j_s)^2_+ - \sum_{i\in[\![d]\!]} Q^i_s\sum_{j\neq i}\beta^{i,j}_s(u^i_s - u^j_s)_+ + \frac{1}{2}\sum_{i\in[\![d]\!]} Q^i_s\sum_{j\neq i}|\beta^{i,j}_s|^2\\
&= \frac{1}{2}\sum_{i\in[\![d]\!]} Q^i_s\sum_{j\neq i}\big|\beta^{i,j}_s - (u^i_s - u^j_s)_+\big|^2,
\end{aligned}$$

the inequality being in fact an equality if $\beta \equiv \beta^\star$.

It remains to compute the bracket on the last line of (4.5). We get

$$\frac{\varepsilon}{\sqrt{2}}Q^i_s\sum_{j\in[\![d]\!]}\sqrt{\frac{P^j_s}{P^i_s}}\mathrm{d}[W^{i,j}_s - W^{j,i}_s]\cdot\sum_{j,k\in[\![d]\!]:\,j\neq k}\nu^{i,j,k}_s\mathrm{d}W^{j,k}_s = \frac{\varepsilon}{\sqrt{2}}Q^i_s\sum_{j\in[\![d]\!]:\,j\neq i}\sqrt{\frac{P^j_s}{P^i_s}}\Big(\nu^{i,i,j}_s - \nu^{i,j,i}_s\Big)\mathrm{d}s,$$

which cancels out with the last term on the first line of (4.5).

Integrating from t to T and taking conditional expectation in (4.5), we then deduce that

$$\begin{aligned}
&\sum_{i\in[\![d]\!]} Q^i_t u^i_t + \frac{1}{2}\mathbb{E}\Big[\sum_{i\in[\![d]\!]}\int_t^T Q^i_s\sum_{j\neq i}\big|\beta^{i,j}_s - (u^i_s - u^j_s)_+\big|^2\mathrm{d}s\,|\,\mathcal{F}^{\boldsymbol{W}}_t\Big]\\
&\le \mathbb{E}\Big[\sum_{i\in[\![d]\!]} Q^i_T u^i_T + \int_t^T\sum_{i\in[\![d]\!]} Q^i_s(f^i(s,P_s) + \frac{1}{2}\sum_{j\neq i}|\beta^{i,j}_s|^2)\mathrm{d}s\,|\,\mathcal{F}^{\boldsymbol{W}}_t\Big],
\end{aligned} \tag{4.6}$$

the inequality being an equality if $\beta \equiv \beta^\star$. Recalling that $Q_t^i = \delta_{i,l}$, this is what we want. □

Proof of Lemma 4.2. First Step. The first step of the proof is to consider a truncated version of (4.2). Hence, for a given constant $c > 0$, we consider the equation

$$\begin{aligned} \mathrm{d}u_t^i &= -\Big(\sum_{j\in[\![d]\!]} \varphi(P_t^j)[u_t^j - u_t^i] + H_c^i(u_t) + f^i(t,P_t)\Big)\mathrm{d}t - \frac{\varepsilon}{\sqrt{2}} \sum_{j\neq i} \sqrt{\frac{P_t^j}{P_t^i}}(\nu_t^{i,i,j} - \nu_t^{i,j,i})\mathrm{d}t \\ &\quad + \sum_{j\neq k} \nu_t^{i,j,k}\mathrm{d}W_t^{j,k}, \\ u_T^i &= g^i(P_T), \end{aligned} \tag{4.7}$$

where H_c^i stands for the truncated Hamiltonian

$$H_c^i(y) := -\frac{1}{2} \sum_{j\in[\![d]\!]} [(y_i - y_j)_+^2 \mathbf{1}_{\{y_i - y_j \le c\}} + (2c(y_i - y_j) - c^2)\mathbf{1}_{\{y_i - y_j > c\}}], \quad y = (y_j)_{j\in[\![d]\!]}.$$

Then, (4.7) is a backward equation with a time dependent driver that is Lipschitz continuous with respect to the entries $(u_t^i)_{i\in[\![d]\!]}$ and $(\nu_t^{i,j,k})_{i,j,k\in[\![d]\!]:j\neq k}$, the Lipschitz constant with respect to the entries $(u_t^i)_{i\in[\![d]\!]}$ being bounded by a deterministic constant C (possibly depending on c) and the Lipschitz constant with respect to $(\nu_t^{i,j,k})_{i,j,k\in[\![d]\!]:j\neq k}$ being bounded by

$$c_t := \frac{\varepsilon}{\sqrt{2}} \Big[\sum_{i\in[\![d]\!]} \frac{1}{P_t^i} \Big]^{1/2},$$

in the sense that (using the fact that the driver is linear in $(\nu_t^{i,j,k})_{i,j,k\in[\![d]\!]:j\neq k}$)

$$\begin{aligned} \frac{\varepsilon}{\sqrt{2}} \Big(\sum_{i\in[\![d]\!]} \Big[\sum_{j\neq i} \sqrt{\frac{P_t^j}{P_t^i}}(\nu_t^{i,i,j} - \nu_t^{i,j,i}) \Big]^2 \Big)^{1/2} &\le \frac{\varepsilon}{\sqrt{2}} \Big(\sum_{i\in[\![d]\!]} \Big[\frac{1}{P_t^i} \sum_{j\neq i} (\nu_t^{i,i,j} - \nu_t^{i,j,i})^2 \Big] \Big)^{1/2} \\ &\le \frac{\varepsilon}{\sqrt{2}} \Big[\sum_{i\in[\![d]\!]} \frac{1}{P_t^i} \Big]^{1/2} \Big[\sum_{i,j\in[\![d]\!]:j\neq i} (\nu_t^{i,i,j} - \nu_t^{i,j,i})^2 \Big]^{1/2} \\ &= c_t \Big[\sum_{i,j\in[\![d]\!]:j\neq i} (\nu_t^{i,i,j} - \nu_t^{i,j,i})^2 \Big]^{1/2}. \end{aligned}$$

By Proposition 2.3 (with $\lambda = 2d - 1$ and $\gamma \ge 60\varepsilon^2$) and from the condition $\kappa \ge (61 + d)\varepsilon^2$ together with Hölder's inequality, we notice that $\mathbb{E}[\exp(2\varepsilon^2 \sum_{l\in[\![d]\!]} \int_0^T (1/P_s^l)ds)] < \infty$. Then, by [33, Theorem 2.1 (i)] (with, using the notations therein, $\gamma \equiv 1$, β_1 a positive constant, c_1 a non-negative constant, $\beta_2 = 2$ and $c_2(t) = c_t$), there exists a unique solution to (4.7) satisfying

$$\begin{aligned} &\sum_{i\in[\![d]\!]} \mathbb{E}\Big[\sup_{0\le t\le T} \Big(\exp\Big(\varepsilon^2 \sum_{l\in[\![d]\!]} \int_0^t \frac{1}{P_s^l}\mathrm{d}s \Big) |u_t^i|^2 \Big) \Big] < \infty, \\ &\sum_{i,j,k\in[\![d]\!]:j\neq k} \mathbb{E}\Big[\int_0^T \exp\Big(\varepsilon^2 \sum_{l\in[\![d]\!]} \int_0^t \frac{1}{P_s^l}\mathrm{d}s \Big) \Big[|\nu_t^{i,j,k}|^2 + \Big(1 + \sum_{l\in[\![d]\!]} \int_0^t \frac{1}{P_s^l}\mathrm{d}s \Big) |u_t^i|^2 \Big] \mathrm{d}t \Big] < \infty. \end{aligned}$$

Second Step. We now prove that we can find a bound for the solution that is independent of c. To do so, we follow the proof of Lemma 4.1, noticing that the Hamiltonian H_c introduced in the first step is associated with the same cost functional $\mathcal{J}$ as in (2.26) except that the processes $((\beta^{i,j}_t)_{0\le t\le T})_{i,j\in[\![d]\!]:i\neq j}$ therein are required to be bounded by c, and similarly for $\mathcal{J}^l$ in (4.1). In particular, u^l_t defined in the first step satisfies $u^l_t = \operatorname{ess\,inf}_{\boldsymbol{\beta};\boldsymbol{\beta}^{i,j}\le c}\mathcal{J}^l(t,(\beta_s)_{t\le s\le T},(P_s)_{t\le s\le T})$. Call now $((Q^i_s)_{i\in[\![d]\!]})_{t\le s\le T}$ the solution to (2.24) with $Q^i_t = \delta_{i,l}$ for some $l\in[\![d]\!]$ and $\beta\equiv 0$. Then, by (4.6) (but with the solution $((u^i_t)_{i\in[\![d]\!]})_{0\le t\le T}$ to (4.7) and so with the new Hamiltonian)

$$u^l_t \le \mathbb{E}\Big[\sum_{i\in[\![d]\!]} Q^i_T u^i_T + \int_t^T \sum_{i\in[\![d]\!]} Q^i_s(f^i(s,P_s) + \frac{1}{2}\sum_{j\neq i}|\beta^{i,j}_s|^2)\mathrm{d}s\,|\,\mathcal{F}^{\boldsymbol{W}}_t\Big].$$

Here, $\beta\equiv 0$ and $u^i_T = g^i(P_T)$, which provides an upper bound for $((u^l_t)_{l\in[\![d]\!]})_{0\le t\le T}$, by using (2.23) and the L^∞ bounds on f and g. Importantly, the upper bound is independent of c. In order to obtain a lower bound, we call $((Q^i_s)_{i\in[\![d]\!]})_{t\le s\le T}$ the solution to (2.24) with $Q^i_t = \delta_{i,l}$ for some $l\in[\![d]\!]$, given an open-loop strategy $\boldsymbol{\beta}$ whose coordinates are bounded by c. Using again the bounds on f and g, we get

$$\begin{aligned}\mathcal{J}^l\big(t,(\beta_s)_{t\le s\le T},(P_s)_{t\le s\le T}\big) &= \mathbb{E}\Big[\sum_{i\in[\![d]\!]} Q^i_T u^i_T + \int_t^T \sum_{i\in[\![d]\!]} Q^i_s(f^i(s,P_s) + \frac{1}{2}\sum_{j\neq i}|\beta^{i,j}_s|^2)\mathrm{d}s\,|\,\mathcal{F}^{\boldsymbol{W}}_t\Big]\\ &\ge \mathbb{E}\Big[-\sum_{i\in[\![d]\!]} Q^i_T\|g^i\|_\infty - \int_t^T\sum_{i\in[\![d]\!]} Q^i_s\|f^i\|_\infty\mathrm{d}s\,|\,\mathcal{F}^{\boldsymbol{W}}_t\Big] \ge -C_0,\end{aligned}$$

by (2.23), for a constant C_0 independent of c and $\boldsymbol{\beta}$; so $u^l_t\ge -C_0$.

In the end, we may find a constant C_0 such that, whatever the value of c in (4.7), the solution is bounded by C_0. We deduce that, whenever $c\ge 2C_0$, the solution of (4.7) is also a solution of the backward equation (4.2). This proves the existence of a bounded solution to (4.2). As for uniqueness, it suffices to notice that a bounded solution to (4.2) is also a solution to (4.7), but for a large enough c inside. Hence, we get that any bounded solution to (4.2) is bounded by C_0, which shows uniqueness. □

4.2. Proof of Theorem 3.2: Existence and uniqueness of an MFG equilibrium

The system (4.4) is what we call a forward-backward stochastic differential equation. But, differently from most of the cases that have been addressed so far in the literature (see for instance [11, Chapter 3]), solutions to the forward equation are here regarded as processes with values in $\mathcal{S}_{d-1}$. This requires a special treatment, which is the main rationale of our paper. Whatever the setting, a standard strategy for solving forward-backward stochastic differential equations (at least in the so-called Markovian case, see for instance [52, 22]) is to regard the system formed by the two forward and backward equations as the characteristics of a system of parabolic second order PDEs. In our framework, this system of PDEs is precisely the master equation (3.6)–(3.7). Therein, the unknown is the tuple of functions $(U^i)_{i\in[\![d]\!]}$, each U^i standing for a real valued function defined on $[0,T]\times\mathcal{S}_{d-1}$. Accordingly, the master equation is formally obtained by imposing $u^i_t = U^i(t,P_t)$ and expanding U using Itô formula; this is basically what we do later for proving Theorem 4.3.

The precise connection between the master equation and the MFG system (4.4) is given by the following statement, which is in fact a general feature of MFGs: Once the master equation is known to have a unique smooth solution, the MFG should be uniquely solvable.

Theorem 4.3. *Assume that there exists a d-tuple $(U^1, \cdots, U^d)$ of real valued functions defined on $[0,T] \times \mathcal{S}_{d-1}$ such that, for any $i \in [\![d]\!]$, U^i belongs to the Wright–Fisher space $\mathscr{C}_{\mathrm{WF}}^{1+\gamma'/2,2+\gamma'}([0,T] \times \mathcal{S}_{d-1})$ for some $\gamma' > 0$ (see §3.2.2 for the definition), and for any $(t,p) \in [0,T] \times \mathrm{Int}(\hat{\mathcal{S}}_{d-1})$, equation (3.6) holds at (t,p). Then, for any (deterministic) initial condition $(p_{0,i})_{i\in[\![d]\!]} \in \mathcal{S}_{d-1}$, with $p_{0,i} > 0$ for any $i \in [\![d]\!]$, the MFG system (4.4) has a unique solution $(P_t = (P_t^i)_{i\in[\![d]\!]}, u_t = (u_t^i)_{i\in[\![d]\!]}, \nu_t = (\nu_t^{i,j,k})_{i,j,k\in[\![d]\!]:j\neq k})_{0\le t\le T}$ in the class of $\mathbb{F}^{\boldsymbol{W}}$-progressively-measurable processes $(\widetilde{P}_t = (\widetilde{P}_t^i)_{i\in[\![d]\!]}, \widetilde{u}_t = (\widetilde{u}_t^i)_{i\in[\![d]\!]}, \widetilde{\nu}_t = (\widetilde{\nu}_t^{i,j,k})_{i,j,k\in[\![d]\!]:j\neq k})_{0\le t\le T}$ such that $(\widetilde{P}_t)_{0\le t\le T}$ is continuous and takes values in $\mathcal{S}_{d-1}$, $(\widetilde{u}_t)_{0\le t\le T}$ is continuous and is bounded by a deterministic constant and $(\widetilde{\nu}_t)_{0\le t\le T}$ satisfies $\mathbb{E}[\int_0^T \exp(\varepsilon^2 \sum_{l\in[\![d]\!]} \int_0^t (1/P_s^l)\mathrm{d}s)|\widetilde{\nu}_t|^2 \mathrm{d}t] < \infty$. The solution satisfies, $\mathrm{d}\mathbb{P}$ almost surely, for all $t \in [0,T]$ and all $i \in [\![d]\!]$, $u_t^i = U^i(t,P_t)$, and, $\mathrm{d}\mathbb{P} \otimes \mathrm{d}t$ almost everywhere, for all $i,j,k \in [\![d]\!]$, with $j \neq k$, $\nu_t^{i,j,k} = V^{i,j,k}(t,P_t)$, where*

$$V^{i,j,k}(t,p) := \frac{\varepsilon}{\sqrt{2}} \Big(\mathfrak{d}_{p_j} U^i(t,p) - \mathfrak{d}_{p_k} U^i(t,p) \Big) \sqrt{p_j p_k}.$$

We notice that Theorem 3.2 from the above theorem, thanks to what we discussed in §4.1.2.

Proof. The proof is inspired by [52], but the fact that the master equation is set on the simplex makes it more difficult. Also, we recall that $\kappa \ge (61+d)\varepsilon^2$.

First Step. In order to prove the existence of a solution, one may first solve the SDE

$$\begin{aligned} \mathrm{d}P_t^i = & \sum_{j\in[\![d]\!]} \Big(P_t^j[\varphi(P_t^i) + (U^j(t,P_t) - U^i(t,P_t))_+] - P_t^i[\varphi(P_t^j) + (U^i(t,P_t) - U^j(t,P_t))_+] \Big) \mathrm{d}t \\ & + \varepsilon \sum_{j\in[\![d]\!]:j\neq i} \sqrt{P_t^i P_t^j} \mathrm{d}\overline{W}_t^{i,j}, \quad t \in [0,T], \end{aligned} \tag{4.8}$$

with $p_0 = (p_{0,i})_{i\in[\![d]\!]}$ as initial condition. Solvability is a mere consequence of Proposition 2.2.

Then, it suffices to let $u_t^i := U^i(t,P_t)$ and $\nu_t^{i,j,k} := V^{i,j,k}(t,P_t)$, for $t \in [0,T]$ and $i,j,k \in [\![d]\!]$, with $j \neq k$. By Itô's formula (see the next step if needed for the details), we may easily expand $(u_t^i)_{0\le t\le T}$ and check that it solves the backward equation in (4.4). (The fact that the second-order derivatives are just defined on the interior of simplex is not a hindrance since $(P_t)_{0\le t\le T}$ does not touch the boundary, see Proposition 2.2.) Obviously, the processes $((u_t^i)_{0\le t\le T})_{i\in[\![d]\!]}$ and $((\nu_t^{i,j,k})_{0\le t\le T})_{i,j,k\in[\![d]\!]:j\neq k}$ are bounded and hence satisfy the required growth conditions.

Second Step. Consider now another solution, say

$$\Big(((\widetilde{P}_t^i)_{0\le t\le T})_{i\in[\![d]\!]}, ((\widetilde{u}_t^i)_{0\le t\le T})_{i\in[\![d]\!]}, ((\widetilde{\nu}_t^{i,j,k})_{0\le t\le T})_{i,j,k\in[\![d]\!]:j\neq k} \Big)$$

to (4.4). We denote $\widetilde{U}_t^i := U^i(t,\widetilde{P}_t)$, $\partial_{p_j}\widetilde{U}_t^i := \mathfrak{d}_{p_j} U^i(t,\widetilde{P}_t)$, $\partial^2_{p_j p_k}\widetilde{U}_t^i := \mathfrak{d}^2_{p_j p_k} U^i(t,\widetilde{P}_t)$, for $t \in [0,T]$ and $i,j,k \in [\![d]\!]$, $j \neq k$. Thanks to the fact that $U^i \in \mathcal{C}^{1,2}([0,T] \times \mathrm{Int}(\hat{\mathcal{S}}_{d-1}))$, we can apply Itô's formula to $(\widetilde{\{U^i\}}^i(t,\widetilde{P}_t))_{0\le t\le T}$ (which obviously coincides with $(\widetilde{U}_t^i)_{0\le t\le T}$). We get (the computations of the various intrinsic derivatives that appear in the expansion are similar to those

in (3.7))

$$
\begin{aligned}
\mathrm{d}\widetilde{U}_t^i &= \Big\{\partial_t \widetilde{U}_t^i + \frac{\varepsilon^2}{2}\sum_{j,k\in[\![d]\!]}(\widetilde{P}_t^j\delta_{jk} - \widetilde{P}_t^j\widetilde{P}_t^k)\partial^2_{p_jp_k}\widetilde{U}_t^i \\
&\quad + \sum_{j,k\in[\![d]\!]}\widetilde{P}_t^k\Big[\varphi(\widetilde{P}_t^j) + (\widetilde{U}_t^k - \widetilde{U}_t^j)_+\Big]\Big(\partial_{p_j}\widetilde{U}_t^i - \partial_{p_k}\widetilde{U}_t^i\Big)\Big\}\mathrm{d}t \\
&\quad + \frac{\varepsilon}{\sqrt{2}}\sum_{j,k\in[\![d]\!]:j\neq k}\sqrt{\widetilde{P}_t^j\widetilde{P}_t^k}\Big(\partial_{p_j}\widetilde{U}_t^i - \partial_{p_k}\widetilde{U}_t^i\Big)\mathrm{d}W_t^{j,k} \\
&= -\Big\{H^i(\widetilde{U}_t) + f^i(t,\widetilde{P}_t) + \sum_{j\in[\![d]\!]}\varphi(\widetilde{P}_t^j)[\widetilde{U}_t^j - \widetilde{U}_t^i] \\
&\quad + \varepsilon^2\sum_{j\in[\![d]\!]}\widetilde{P}_t^j\Big(\partial_{p_i}\widetilde{U}_t^i - \partial_{p_j}\widetilde{U}_t^i\Big)\Big\}\mathrm{d}t + \frac{\varepsilon}{\sqrt{2}}\sum_{j,k\in[\![d]\!]:j\neq k}\sqrt{\widetilde{P}_t^j\widetilde{P}_t^k}\Big(\partial_{p_j}\widetilde{U}_t^i - \partial_{p_k}\widetilde{U}_t^i\Big)\mathrm{d}W_t^{j,k},
\end{aligned}
$$

where in the last equality we used the equation (3.7) satisfied by U. This prompts us to let

$$
\widetilde{V}_t^{i,j,k} = \frac{\varepsilon}{\sqrt{2}}\sqrt{\widetilde{P}_t^j\widetilde{P}_t^k}\Big(\partial_{p_j}\widetilde{U}_t^i - \partial_{p_k}\widetilde{U}_t^i\Big), \quad t\in[0,T], \quad i,j,k\in[\![d]\!], \quad j\neq k.
$$

Subtracting the equation satisfied by $((\widetilde{u}_t^i)_{i\in[\![d]\!]})_{0\le t\le T}$, we get

$$
\begin{aligned}
\mathrm{d}(\widetilde{U}_t^i - \widetilde{u}_t^i) &= -\Big\{H^i(\widetilde{U}_t) - H^i(\widetilde{u}_t) + \sum_{j\in[\![d]\!]}\varphi(\widetilde{P}_t^j)[\widetilde{U}_t^j - \widetilde{u}_t^j - (\widetilde{U}_t^i - \widetilde{u}_t^i)] \\
&\quad + \frac{\varepsilon}{\sqrt{2}}\sum_{j\in[\![d]\!]}\sqrt{\frac{\widetilde{P}_t^j}{\widetilde{P}_t^i}}\big(\widetilde{V}_t^{i,i,j} - \widetilde{\nu}_t^{i,i,j} - (\widetilde{V}_t^{i,j,i} - \widetilde{\nu}_t^{i,j,i})\big)\Big\}\mathrm{d}t \\
&\quad + \sum_{j,k\in[\![d]\!]:j\neq k}(\widetilde{V}_t^{i,j,k} - \widetilde{\nu}_t^{i,j,k})\mathrm{d}W_t^{j,k}, \quad t\in[0,T], \quad i\in[\![d]\!].
\end{aligned}
$$

Consider now

$$
e_t := \exp\Big(\varepsilon^2\int_0^t\sum_{j\in[\![d]\!]}\frac{1}{\widetilde{P}_s^j}\mathrm{d}s\Big), \quad t\in[0,T].
$$

Then, by Itô's formula, we obtain, for any $t\in[0,T]$,

$$
\begin{aligned}
&e_t|\widetilde{U}_t^i - \widetilde{u}_t^i|^2 + \int_t^T \varepsilon^2 e_s|\widetilde{U}_s^i - \widetilde{u}_s^i|^2\Big(\sum_{j\in[\![d]\!]}\frac{1}{\widetilde{P}_s^j}\Big)\mathrm{d}s + \int_t^T e_s\sum_{j,k\in[\![d]\!]:j\neq k}\Big|\widetilde{V}_s^{i,j,k} - \widetilde{\nu}_s^{i,j,k}\Big|^2\mathrm{d}s \\
&= 2\int_t^T e_s(\widetilde{U}_s^i - \widetilde{u}_s^i)\Big\{H^i(\widetilde{U}_s) - H^i(\widetilde{u}_s) + \sum_{j\in[\![d]\!]}\varphi(\widetilde{P}_s^j)[\widetilde{U}_s^j - \widetilde{u}_s^j - (\widetilde{U}_s^i - \widetilde{u}_s^i)] \\
&\quad + \frac{\varepsilon}{\sqrt{2}}\sum_{j\in[\![d]\!]}\sqrt{\frac{\widetilde{P}_s^j}{\widetilde{P}_s^i}}\big(\widetilde{V}_s^{i,i,j} - \widetilde{\nu}_s^{i,i,j} - (\widetilde{V}_s^{i,j,i} - \widetilde{\nu}_s^{i,j,i})\big)\Big\}\mathrm{d}s \qquad (4.9) \\
&\quad + 2\int_t^T e_s(\widetilde{U}_t^i - \widetilde{u}_t^i)\sum_{j,k\in[\![d]\!]:j\neq k}(\widetilde{V}_s^{i,j,k} - \widetilde{\nu}_s^{i,j,k})\mathrm{d}W_s^{j,k}.
\end{aligned}
$$

By Proposition 2.3 (together with Remark 2.5), with $\lambda = 2d - 1$, $\gamma \geq 60\varepsilon^2$ and $\kappa \geq (61 + d)\varepsilon^2$ therein, and by Hölder's inequality, $\mathbb{E}[e_T^4]$ is finite. Since $((\widetilde{U}_t^i)_{i\in[\![d]\!]})_{0\leq t\leq T}$ and $((\widetilde{u}_t^i)_{i\in[\![d]\!]})_{0\leq t\leq T}$ are bounded (by deterministic constants) and $((\widetilde{V}_t^{i,j,k})_{i,j,k\in[\![d]\!]:j\neq k})_{0\leq t\leq T}$ and $((\widetilde{u}_t^{i,j,k})_{i\in[\![d]\!]:j\neq k})_{0\leq t\leq T}$ are square-integrable, all the terms in the right-hand side have integrable sup norm (over $t \in [0, T]$); as for the last term in the right-hand side, the latter follows from Burkholder–Davis–Gundy inequalities. Also, we can treat the difference $H^i(\widetilde{U}_s) - H^i(\widetilde{u}_s)$ as a Lipschitz difference, since U and $\tilde{u}$ are bounded. Hence, taking expectations and applying Young's inequality, we can find a constant C such that

$$\begin{aligned}
&\mathbb{E}\Big[e_t|\widetilde{U}_t^i - \widetilde{u}_t^i|^2 + \int_t^T \varepsilon^2 e_s|\widetilde{U}_s^i - \widetilde{u}_s^i|^2\Big(\sum_{j\in[\![d]\!]}\frac{1}{\widetilde{P}_s^j}\Big)\mathrm{d}s + \int_t^T e_s \sum_{j,k\in[\![d]\!]:j\neq k}\Big|\widetilde{V}_s^{i,j,k} - \widetilde{v}_s^{i,j,k}\Big|^2\,\mathrm{d}s\Big] \\
&\leq C\sum_{j\in[\![d]\!]}\mathbb{E}\Big[\int_t^T e_s|\widetilde{U}_s^j - \widetilde{u}_s^j|^2\mathrm{d}s\Big] + \varepsilon^2\mathbb{E}\Big[\int_t^T \frac{e_s}{\widetilde{P}_s^i}|\widetilde{U}_s^i - \widetilde{u}_s^i|^2\mathrm{d}s\Big] \\
&\quad + \mathbb{E}\Big[\int_t^T e_s \sum_{j,k\in[\![d]\!]:j\neq k}\Big|\widetilde{V}_s^{i,j,k} - \widetilde{v}_s^{i,j,k}\Big|^2\,\mathrm{d}s\Big].
\end{aligned}$$

We obtain,

$$\sum_{j\in[\![d]\!]}\mathbb{E}\left[e_t|\widetilde{U}_t^j - \widetilde{u}_t^j|^2\right] \leq C\sum_{j\in[\![d]\!]}\int_t^T \mathbb{E}\left[e_s|\widetilde{U}_s^j - \widetilde{u}_s^j|^2\right]\mathrm{d}s,$$

and thus Gronwall's lemma yields, for any $i \in [\![d]\!]$ and any $t \in [0, T]$,

$$\mathbb{P}\left(\widetilde{u}_t^i = \widetilde{U}_t^i = U^i(t, \widetilde{P}_t)\right) = 1.$$

This permits to identify $(\widetilde{P}_t)_{0\leq t\leq T}$ with the solution of (4.8). It is then pretty straightforward to show that $(\widetilde{u}_t^i)_{0\leq t\leq T}$ coincides with $(u_t^i)_{0\leq t\leq T}$, for each $i \in [\![d]\!]$, and then that $(\widetilde{v}_t^{i,j,k})_{0\leq t\leq T}$ coincides with $(v_t^{i,j,k})_{0\leq t\leq T}$, for each $i, j, k \in [\![d]\!]$, $j \neq k$. □

4.3. Proof of Theorem 3.4: Unique solvability of the master equation

In order to establish the unique solvability of the master equation (3.6)–(3.7), the first point is to observe that it may be rewritten in a somewhat generic form. Indeed, for a given coordinate $i \in [\![d]\!]$, we may let

$$\begin{aligned}
B_j^i(t, p, y) &:= \varphi(p_j) + \sum_{k\in[\![d]\!]} p_k(y_k - y_j)_+ - p_j\sum_{k\in[\![d]\!]}[\varphi(p_k) + (y_j - y_k)_+] + \varepsilon^2(\delta_{i,j} - p_j), \\
F^i(t, p, y) &:= H^i(y) + f^i(t, p) + \sum_{k\in[\![d]\!]}\varphi(p_j)(y_j - y_i),
\end{aligned} \tag{4.10}$$

where $t \in [0, T]$, $p \in \mathcal{S}_{d-1}$ and $y = (y_k)_{k\in[\![d]\!]} \in \mathbb{R}^d$. Recalling (4.3), we may rewrite (3.7) as

$$\begin{aligned}
&\partial_t U^i(t, p) + F^i(t, p, U(t, p)) + \sum_{j\in[\![d]\!]} B_j^i(t, p, U(t, p))\mathfrak{d}_{p_j}U^i(t, p) \\
&\qquad + \frac{\varepsilon^2}{2}\sum_{j,k\in[\![d]\!]}(p_j\delta_{jk} - p_jp_k)\mathfrak{d}^2_{p_jp_k}U^i(t, p) = 0, \\
&U^i(T, p) = g^i(p),
\end{aligned} \tag{4.11}$$

for $t \in [0,T]$ and $p \in \mathrm{Int}(\mathcal{S}_{d-1})$, with the shorten notation $U(t,p) = (U^i(t,p))_{i\in[\![d]\!]}$. For sure, we could write (3.6) in a similar form. In fact, what really matters is that

$$\sum_{j\in[\![d]\!]} B^i_j(t,p,y) = 0,$$

for any $i \in [\![d]\!]$, $t \in [0,T]$, $p \in \mathrm{Int}(\hat{\mathcal{S}}_{d-1})$ and $y = (y_i)_{i\in[\![d]\!]} \in \mathbb{R}^d$, and that $B^i_j(t,p,y) > 0$ whenever $p_j = 0$.

In the sequel, solvability of (4.11) is addressed in several steps. The first one is to address the solvability of the linear version of (4.11) obtained by freezing the nonlinear component U in B^i and F^i; as we make it clear below, this mostly follows from the earlier results of [28]. The second one is to prove a priori estimates for the solutions to the latter linear version independently of the nonlinear component U that is frozen in the coefficients B^i and F^i; this is where we invoke Theorem 2.10. The last step is to deduce the existence of a classical solution to the master equation by means of Schauder's fixed point theorem.

4.3.1. Linear version

We first establish a preliminary solvability result for the linear version (2.27), keeping in mind that it should be reformulated in intrinsic derivatives as follows:

$$\begin{aligned}
&\partial_t u(t,p) + \sum_{j\in[\![d]\!]} \Big(\varphi(p_j) + b_j(t,p) + p_j b^\circ_j(t,p)\Big)\mathfrak{d}_{p_j} u(t,p) \\
&\qquad + \frac{\varepsilon^2}{2} \sum_{j,k\in[\![d]\!]} (p_j\delta_{jk} - p_j p_k)\mathfrak{d}^2_{p_j p_k} u(t,p) + h(t,p) = 0, \\
&u(T,p) = \ell(p),
\end{aligned} \tag{4.12}$$

for $(t,p) \in [0,T]\times \mathrm{Int}(\hat{\mathcal{S}}_{d-1})$ ($\mathrm{Int}(\hat{\mathcal{S}}_{d-1})$ being here regarded as a subset of $\mathcal{S}_{d-1}$), where we recall that $b = (b_j)_{j\in[\![d]\!]} : [0,T]\times\mathcal{S}_{d-1} \to (\mathbb{R}_+)^d$, $b^\circ = (b^\circ_j)_{j\in[\![d]\!]} : [0,T]\times\mathcal{S}_{d-1} \to \mathbb{R}^d$, $h : [0,T]\times\mathcal{S}_{d-1} \to \mathbb{R}$ and $\ell : \mathcal{S}_{d-1} \to \mathbb{R}$ are bounded and satisfy

$$\sum_{j\in[\![d]\!]} \Big(\varphi(p_j) + b_j(t,p) + p_j b^\circ_j(t,p)\Big) = 0, \quad t \in [0,T], \quad p \in \mathcal{S}_{d-1}. \tag{4.13}$$

Our solvability result is

Lemma 4.4. *Assume that the functions $(b_j)_{j\in[\![d]\!]}$ and $(b^\circ_j)_{j\in[\![d]\!]}$ are in $\mathscr{C}^{\eta/2,\eta}_{\mathrm{WF}}([0,T]\times\mathcal{S}_{d-1})$, that h is in $\mathscr{C}^{\eta/4,\eta/2}_{\mathrm{WF}}([0,T]\times\mathcal{S}_{d-1})$ and ℓ is in $\mathscr{C}^{2+\eta/2}_{\mathrm{WF}}(\mathcal{S}_{d-1})$, for some $\eta \in (0,1)$. Then, equation* (4.12) *has a unique classical solution in the space $\mathscr{C}^{1+\eta/4,2+\eta/2}_{\mathrm{WF}}([0,T]\times\mathcal{S}_{d-1})$.*

Proof. For a fixed $t_0 \in [0,T]$, we rewrite (4.12) in the form

$$\begin{aligned}
&\partial_t u(t,p) + \sum_{j\in[\![d]\!]} \Big(\varphi(p_j) + b_j(t_0,p) + p_j b^\circ_j(t_0,p)\Big)\mathfrak{d}_{p_j} u(t,p) + \frac{\varepsilon^2}{2} \sum_{j,k\in[\![d]\!]} (p_j\delta_{jk} - p_j p_k)\mathfrak{d}^2_{p_j p_k} u(t,p) \\
&\qquad + \sum_{j\in[\![d]\!]} \Big((b_j(t,p) + p_j b^\circ_j(t,p)) - (b_j(t_0,p) + p_j b^\circ_j(t_0,p))\Big)\mathfrak{d}_{p_j} u(t,p) + h(t,p) = 0, \\
&u(T,p) = \ell(p),
\end{aligned} \tag{4.14}$$

for $(t,p) \in [0,T]\times \mathrm{Int}(\hat{\mathcal{S}}_{d-1})$. Our first goal is to solve the equation on $[t_0,T]\times\mathrm{Int}(\hat{\mathcal{S}}_{d-1})$ provided t_0 is chosen close enough to T.

In order to solve the above equation, we define the following mapping. For a vector-valued function $w = (w_j)_{j\in[\![d]\!]} : [0,T]\times\mathcal{S}_{d-1} \to \mathbb{R}^d$ whose components are in $\mathscr{C}^{\eta/4,\eta/2}_{\mathrm{WF}}([0,T]\times\mathcal{S}_{d-1})$, we call v the solution of the equation

$$\begin{aligned}
&\partial_t v(t,p) + \sum_{j\in[\![d]\!]} \Big(\varphi(p_j) + b_j(t_0,p) + p_j b_j^\circ(t_0,p)\Big)\mathfrak{d}_{p_j} v(t,p) + \frac{\varepsilon^2}{2}\sum_{j,k\in[\![d]\!]} (p_j\delta_{jk} - p_j p_k)\mathfrak{d}^2_{p_j p_k} v \\
&\quad + \sum_{j\in[\![d]\!]} \Big((b_j(t,p) + p_j b_j^\circ(t,p)) - (b_j(t_0,p) + p_j b_j^\circ(t_0,p))\Big) w_j(t,p) + h(t,p) = 0, \\
&v(T,p) = \ell(p),
\end{aligned}$$

for $(t,p) \in [t_0,T]\times\mathrm{Int}(\hat{\mathcal{S}}_{d-1})$, the solution being known, by [28, Theorem 10.0.2], to exist and to satisfy

$$\|v\|_{1+\eta/4,2+\eta/2;[t_0,T])} \le C\Big(\|\ell\|_{2+\eta/2} + \|W\|_{\eta/4,\eta/2;[t_0,T]} + \|h\|_{\eta/4,\eta/2;[t_0,T]}\Big), \tag{4.15}$$

where we added the notation $[t_0,T]$ in the Wright–Fisher norm in order to emphasize the fact that the underlying domain is $[t_0,T]\times\mathcal{S}_{d-1}$ and not $[0,T]\times\mathcal{S}_{d-1}$, and with

$$W(t,p) := \sum_{j\in[\![d]\!]} \Big((b_j(t,p) + p_j b_j^\circ(t,p)) - (b_j(t_0,p) + p_j b_j^\circ(t_0,p))\Big) w_j(t,p).$$

Clearly, we can find a universal constant $c > 0$ such that

$$\big\|W\big\|_{\eta/4,\eta/2;[t_0,T]} \le c\sum_{j\in[\![d]\!]} \big\|B_j^\circ\big\|_{\eta/4,\eta/2;[t_0,T]} \|w_j\|_{\eta/4,\eta/2;[t_0,T]},$$

with

$$B_j^\circ(t,p) := b_j(t,p) + p_j b_j^\circ(t,p) - \Big(b_j(t_0,p) + p_j b_j^\circ(t_0,p)\Big), \quad (t,p)\in[t_0,T]\times\mathcal{S}_{d-1}.$$

Now, we can find a constant C, only depending on the Wright–Fisher norm of $b = (b_j)_{j\in[\![d]\!]}$ such that, for any $s,t\in[t_0,T]$ and any $p,q\in\mathcal{S}_{d-1}$,

$$\begin{aligned}
&\big|b(t,p) - b(t_0,p) - (b(s,q) - b(t_0,q))\big| \\
&= \big|b(t,p) - b(t_0,p) - (b(s,q) - b(t_0,q))\big|^{1/2}\big|b(t,p) - b(t_0,p) - (b(s,q) - b(t_0,q))\big|^{1/2} \\
&\le C\Big(\big|b(t,p) - b(t_0,p)\big| + \big|b(s,q) - b(t_0,q)\big|\Big)^{1/2}\Big(\big|b(t,p) - b(s,q)\big| + \big|b(t_0,p) - b(t_0,q)\big|\Big)^{1/2} \\
&\le C(T-t_0)^{\eta/4}\Big(|t-s|^{\eta/4} + |\sqrt{p} - \sqrt{q}|^{\eta/2}\Big),
\end{aligned}$$

which shows that the Wright–Fisher Hölder norm (of exponents $(\eta/4,\eta/2)$) of $b - b(t_0,\cdot)$ is small with $T - t_0$ (it is easy to see that sup norm is small with $T - t_0$). Proceeding in a similar way with the other functions entering the definition of B°, we deduce that the Wright–Fisher Hölder norm (of exponent $\eta/2$) of B° is small with $T - t_0$.

Therefore, Schauder's estimates (4.15) imply that, for $T - t_0$ small enough

$$\|\mathfrak{d}_p v\|_{\eta/4,\eta/2;[t_0,T]} \le C\Big(\|\ell\|_{2+\eta/2;[t_0,T]} + \|h\|_{\eta/4,\eta/2;[t_0,T]}\Big) + \frac{1}{2}\big\|w\big\|_{\eta/4,\eta/2;[t_0,T]},$$

for a constant C which is independent of w and t_0. This shows in particular that $\|\mathfrak{d}_p v\|_{\eta/4,\eta/2;[t_0,T]} \leq 2C(\|\ell\|_{2+\eta,[t_0,T]} + \|h\|_{\eta/4,\eta/2;[t_0,T]})$ whenever $\|w\|_{\eta/4,\eta/2;[t_0,T]} \leq 2C(\|\ell\|_{2+\eta,[t_0,T]} + \|h\|_{\eta/4,\eta/2;[t_0,T]})$. In particular, the map $w \mapsto v$ preserves a closed ball of $[\mathscr{C}_{\mathrm{WF}}^{\eta/4,\eta/2}([t_0,T]\times\mathcal{S}_{d-1})]^d$. By linearity, the map $w \mapsto v$ is obviously continuous from $[\mathscr{C}_{\mathrm{WF}}^{\eta'/4,\eta'/2}([t_0,T]\times\mathcal{S}_{d-1})]^d$ into itself, for any $\eta' \in (0,\eta]$. By Schauder's theorem (regarding any closed ball of $[\mathscr{C}_{\mathrm{WF}}^{\eta/4,\eta/2}([t_0,T]\times\mathcal{S}_{d-1})]^d$ as a compact subset of $[\mathscr{C}_{\mathrm{WF}}^{\eta'/4,\eta'/2}([t_0,T]\times\mathcal{S}_{d-1})]^d$, for $\eta' \in (0,\eta)$), we deduce that there exists a solution v to (4.14) (and hence to (4.12)) in $\mathscr{C}_{\mathrm{WF}}^{1+\eta/4,2+\eta/2}([t_0,T]\times\mathcal{S}_{d-1})$ (so on $[t_0,T]\times\mathcal{S}_{d-1}$). By iterating in time, we deduce that there exists a solution to (4.12) on the entire $[0,T]\times\mathcal{S}_{d-1}$ in the space $\mathscr{C}_{\mathrm{WF}}^{1+\eta/4,2+\eta/2}([0,T]\times\mathcal{S}_{d-1})$.

Uniqueness follows from a straightforward application of Kolmogorov representation formula, see Proposition 5.1 if needed. □

4.3.2. Fixed point argument via Schauder's theorem

Here is now the last step of our proof of Theorem 3.4, which strongly relies on Theorem 2.10.

Proof of Theorem 3.4.

Existence. The proof of existence holds in two steps.

First Step. We first consider the following nonlinear variant of (4.12):

$$\begin{aligned}
&\partial_t U^i(t,p) + \sum_{j\in[\![d]\!]} \Big(\varphi(p_j) + b^i_j(t,p,U(t,p)) + p_j b^\circ_j(t,p,U(t,p))\Big)\mathfrak{d}_{p_j} U^i(t,p)\\
&\qquad + \frac{\varepsilon^2}{2}\sum_{j,k\in[\![d]\!]} (p_j\delta_{jk} - p_j p_k)\mathfrak{d}^2_{p_j p_k} U^i(t,p) + h^i(t,p,U(t,p)) = 0,\\
&U^i(T,p) = g^i(p),
\end{aligned} \tag{4.16}$$

where for any $i \in [\![d]\!]$, $b^i = (b^i_j)_{j\in[\![d]\!]} : [0,T]\times\mathcal{S}_{d-1}\times\mathbb{R}^d \to (\mathbb{R}_+)^d$, $b^\circ = (b^\circ_j)_{j\in[\![d]\!]} : [0,T]\times\mathcal{S}_{d-1}\times\mathbb{R}^d \to \mathbb{R}^d$ and $h^i : [0,T]\times\mathcal{S}_{d-1}\times\mathbb{R}^d \to \mathbb{R}$. We are going to prove the existence of a solution $U = (U^1,\cdots,U^d)$ to (4.16) whenever, for some constant $C_0 \geq 0$, the functions $(b^i)_{i\in[\![d]\!]}$, b° and $(h^i)_{i\in[\![d]\!]}$ are bounded by C_0 and satisfy the following regularity properties

$$\begin{aligned}
&\big|b^i(t,p,y) - b^i(s,q,z)\big| + \big|b^\circ(t,p,y) - b^\circ(s,q,z)\big| + \big|h^i(t,p,y) - h^i(s,q,z)\big|\\
&\qquad \leq C_0(|t-s|^{\gamma/2} + |\sqrt{p} - \sqrt{q}|^\gamma + |y-z|),
\end{aligned} \tag{4.17}$$

for $i \in [\![d]\!]$, $s,t \in [0,T]$, $p,q \in \mathcal{S}_{d-1}$ and $y,z \in \mathbb{R}^d$. Without any loss of generality, we can assume that $\sup_{i\in[\![d]\!]} \|g^i\|_{1,\infty} \leq C_0$. Existence of a classical solution to (4.16) is then proved by a new application of Schauder's fixed point theorem. To do so, we call η and C the exponent and the constant from Theorem 2.10 when $\|b\|_\infty$, $\|b^\circ\|_\infty$, $\|h\|_\infty$ and $\|\ell\|_{1,\infty}$ are less than C_0. We then take an input function $V = (V^1,\cdots,V^d) \in [\mathscr{C}_{\mathrm{WF}}^{\eta/2,\eta}([0,T]\times\mathcal{S}_{d-1})]^d$ such that, for each $i \in [\![d]\!]$, $\|V^i\|_{\eta/2,\eta} \leq C$. By Lemma 4.4 with η therein being replaced by $\min(\eta,\gamma)$, we can solve (4.16) for each $i \in [\![d]\!]$ when, in the nonlinear terms, U is replaced by V. We call $U = (U^1,\cdots,U^d)$ the solution. It belongs to $[\mathscr{C}_{\mathrm{WF}}^{1+\gamma'/2,2+\gamma'}([0,T]\times\mathcal{S}_{d-1})]^d$. By Theorem 2.10, it also satisfies $\|U^i\|_{\eta/2,\eta} \leq C$, for each $i \in [\![d]\!]$. Revisiting if needed the proof of Lemma 4.4, there is no difficulty in proving that the resulting map $V \mapsto U$ is continuous from $[\mathscr{C}_{\mathrm{WF}}^{\eta'/2,\eta'}([0,T]\times\mathcal{S}_{d-1})]^d$ into itself, for any $\eta' \in (0,\eta)$. This permits to apply Schauder's theorem.

Second Step. The goal now is to choose $(b^i)_{i\in[\![d]\!]}$, b° and $(h^i)_{i\in[\![d]\!]}$ (and hence C_0 as well) accordingly so that the solution to (4.16) is in fact a solution to the master equation (4.11). In order to proceed, we follow the same idea as in the proof of Lemma 4.2 and recall the truncated Hamiltonian

$$H_c^i(y) = -\frac{1}{2}\sum_{j\in[\![d]\!]}\big[(y_i-y_j)_+^2\mathbf{1}_{\{y_i-y_j\le c\}} + (2c(y_i-y_j)-c^2)\mathbf{1}_{\{y_i-y_j>c\}}\big], \quad y=(y_j)_{j\in[\![d]\!]},$$

for a constant c to be fixed later. Also, for another constant Γ, the value of which will be also fixed later on, we call ψ_Γ the function

$$\psi_\Gamma(r) := \begin{cases} r, & \text{if } |r|\le\Gamma,\\ \Gamma\mathrm{sign}(r), & \text{if } |r|\ge\Gamma,\end{cases} \qquad r\in\mathbb{R}.$$

Given these notations, we let (compare with (4.10))

$$\begin{aligned} b_j^i(t,p,y) &:= \sum_{k\in[\![d]\!]} p_k \min\Big(c,(y_k-y_j)_+\Big) + \varepsilon^2\delta_{i,j}, \quad i,j\in[\![d]\!],\\ b_j^\circ(t,p,y) &:= -\sum_{k\in[\![d]\!]}\Big[\varphi(p_k)+\min\Big(c,(y_j-y_k)_+\Big)\Big] - \varepsilon^2, \quad j\in[\![d]\!],\\ h^i(t,p,y) &:= \psi_\Gamma\Big(H_c^i(y)+f^i(t,p)+\sum_{j\in[\![d]\!]}\varphi(p_j)\psi_\Gamma(y_j-y_i)\Big), \quad i\in[\![d]\!], \end{aligned} \tag{4.18}$$

for $(t,p,y)\in[0,T]\times\mathcal{S}_{d-1}\times\mathbb{R}^d$. For a given value of c, we can choose Γ (hence depending on c) such that all the above coefficients are bounded by Γ. Moreover, the coefficients satisfy (4.17) for a suitable choice of C_0 therein (notice in this regard that this is the specific interest of the second occurence of ψ_Γ to force the whole term to be jointly Lipschitz in (p,y)). By the first step, there exists a classical solution, say $U=(U^1,\cdots,U^d)$, to (4.16). As in the proof of Theorem 4.3, we can represent U through a forward-backward stochastic differential equation. Following (4.4), the backward equation writes

$$\mathrm{d}u_t^i = -h^i(t,P_t,u_t)\mathrm{d}t - \frac{\varepsilon}{\sqrt{2}}\sum_{j\in[\![d]\!]}\sqrt{\frac{P_t^j}{P_t^i}}(\nu_t^{i,i,j}-\nu_t^{i,j,i})\mathrm{d}t + \sum_{j,k\in[\![d]\!]}\nu_t^{i,j,k}\mathrm{d}W_t^{j,k}, \tag{4.19}$$

with $u_T^i = g^i(P_T)$ as terminal boundary condition, where $(P_t)_{0\le t\le T}$ is the solution to the corresponding forward equation (but there is no need to write it down). The key point here is to observe that h^i is at most of linear growth in u, uniformly in (t,p), the constant in the linear growth depending on c but not on Γ. By considering the drifted Brownian motions $((W_t'^{,i,j} = W_t^{i,j} - \int_0^t(\varepsilon\sqrt{P_s^j})/(\sqrt{2P_s^i})\mathrm{d}s)_{0\le t\le T})_{j\in[\![d]\!]:j\ne i}$ and $((W_t'^{,j,i} = W_t^{j,i} + \int_0^t(\varepsilon\sqrt{P_s^j})/(\sqrt{2P_s^i})\mathrm{d}s)_{0\le t\le T})_{j\in[\![d]\!]:j\ne i}$, for a given value of $i\in[\![d]\!]$, we can apply Girsanov theorem to get rid of the second term in the equation for $(u_t^i)_{0\le t\le T}$ in (4.19), the application of Girsanov theorem being here made licit by Proposition 2.3. We easily deduce that, with U as in (4.19), $\|U(t,\cdot)\|_\infty \le C(1+\int_t^T\|U(s,\cdot)\|_\infty\mathrm{d}s)$ and then deduce that U and hence $(u_t)_{0\le t\le T}$ are bounded by a constant C that depends on c but not on Γ. In particular, it makes sense to choose Γ large enough such that, for $y=(y_j)_{j\in[\![d]\!]}$ with $|y|\le C$, $h^i(t,p,y)$ in (4.18) is also equal to

$$h^i(t,p,y) = H_c^i(y) + f^i(t,p) + \sum_{j\in[\![d]\!]}\varphi(p_j)(y_j-y_i).$$

It says that the backward equation (4.19) identifies with the backward equation (4.7) in the proof of Lemma 4.2. But the point in the second step of the proof of Lemma 4.2 is precisely to show that the solution to (4.19) can be bounded independently of c. In words, we can find a constant C_1, independent of c (and of course of Γ) such that $U = (U^i)_{i\in[\![d]\!]}$ is bounded by C_1. Then, choosing $c \geq 2C_1$ (and Γ large enough as before), we have that, for any $y = (y_j)_{j\in[\![d]\!]}$ with $|y| \leq C_1$, and any $(t, p) \in [0, T] \times \mathcal{S}_{d-1}$ and $i, j \in [\![d]\!]$,

$$B^i_j(t, p, y) = \varphi(p_j) + b^i_j(t, p, y) + b^\circ_j(t, p, y)p_j, \quad F^i(t, p, y) = h^i(t, p, y),$$

with B^i and F^i as in (4.10). This shows that $U = (U^1, \cdots, U^d)$ solves the master equation (4.11).

Uniqueness. By Theorem 4.3, we know that, for any initial condition $p_0 = (p_{0,i})_{i\in[\![d]\!]} \in \mathrm{Int}(\mathcal{S}_{d-1})$, the system (4.4) has a unique solution. Hence, for any two solutions U and U' to the master equation, one has $U(0, p_0) = U'(0, p_0)$. Since p_0 is arbitrary, we get that $U(0, \cdot)$ and $U'(0, \cdot)$ coincide on $\mathrm{Int}(\hat{\mathcal{S}}_{d-1})$. By continuity, they coincide up to the boundary. Here, the initial time is arbitrary and we can replace the initial time 0 by any other initial time $t \in (0, T)$. □

5. Proof of the a priori Hölder estimate

The proof of Theorem 2.10 is the core of the paper. The main ingredient is a coupling estimate for the diffusion process associated with the linear equation (4.12), see the statement of Proposition 5.3. Whilst this approach is mostly inspired by earlier coupling arguments used to prove regularity of various classes of harmonic functions (see for instance [20, 21, 26, 51]), we here need a tailored version that fits the specificities of Kimura operators. In short, the coupling estimate we obtain below does not suffice to conclude directly in full generality. In fact, it just permits to derive the required Hölder estimate in the case $d = 2$. In the higher dimensional setting, we need an additional argument that uses induction on the dimension of the state space to pass from the coupling estimate to the Hölder bound; see Remark 5.4 for a first account and Subsection 5.2 for more details. In short, the rationale for this additional induction argument is that the coupling estimate obtained in Proposition 5.3 blows up near the boundary, except when $d = 2$. As for the induction argument itself, it is based on a conditioning property that is proper to Kimura type operators: Roughly speaking, the last $d - m$ coordinates of the diffusion process associated with the linear equation (4.12) (see also (2.27)) behave, conditional on the first m coordinates, as a diffusion associated with a linear equation of the same type as (4.12) but in dimension $d - m$ instead of d, see Proposition 5.2 for the complete statement. It is worth mentioning that our induction argument is inspired by the work [2]. Therein, the authors prove a gradient estimate for simpler and more regular forms of drifts by iterating on the dimension of the state space. Differently from ours, their approach is purely deterministic: As a result, the conditioning principle exposed in Proposition 5.2 manifests implicitly in [2] through the form of the underlying Kimura operators.

Throughout the section, we are given coefficients $b = (b_i)_{i\in[\![d]\!]}$, $b^\circ = (b^\circ_i)_{i\in[\![d]\!]}$, h and ℓ as in the statement of Theorem 2.10. Then, the aforementioned diffusion process associated with (4.12) is given by the following statement.

Proposition 5.1. *Consider φ as in* (2.19) *with $\delta \in (0, 1)$ and $\kappa \geq \varepsilon^2/2$, for $\varepsilon > 0$. Then, the stochastic differential equation*

$$\mathrm{d}P^i_s = \Big(\varphi(P^i_s) + b_i(s, P_s) + P^i_s b^\circ_i(s, P_s)\Big)\mathrm{d}t + \varepsilon\sqrt{P^i_s}\sum_{j\in[\![d]\!]}\sqrt{P^j_s}\mathrm{d}\overline{W}^{i,j}_t, \quad s \in [t, T],\ i \in [\![d]\!], \quad (5.1)$$

is uniquely solvable for any initial time in $t \in [0,T]$ and any (possibly random) initial condition in $\mathrm{Int}(\mathcal{S}_{d-1})$*. Moreover, the coordinates of the solution remain almost surely strictly positive. In particular, for any* $(t,p) \in [0,T]\times\mathrm{Int}(\hat{\mathcal{S}}_{d-1})$*, for any* $[t,T]$*-valued stopping time* τ *(with respect to the filtration* $\mathbb{F}^{\boldsymbol{W}}$*), any function* u *as in the statement of Theorem 2.10 (with* (2.27) *therein being replaced by the more rigorous version* (4.12)*) admits the representation*

$$u(t,p) = \mathbb{E}\Big[u(\tau, P_\tau^{t,p}) + \int_t^\tau h(s, P_s^{t,p})\mathrm{d}s\Big], \tag{5.2}$$

where $P^{t,p}$ *is the* d*-dimensional process whose dynamics are given by* (5.1) *and starts from* p *at time* t*. In particular, the* L^∞ *bound in Theorem 2.10 holds true.*

Proof. Strong existence and uniqueness may be proven in Proposition 2.2 (Equation (5.1) is slightly more general than the equation handled in Proposition 2.2, but the proof works in the same way). Representation of u is a straightforward consequence of Itô's formula (as in the proof of Theorem 4.3). □

It is worth observing that, taking $\tau = T$ in (5.2), u has the (standard) representation:

$$u(t,p) = \mathbb{E}\Big[\ell(P_T^{t,p}) + \int_t^T h(s, P_s^{t,p})\mathrm{d}s\Big], \quad (t,p) \in [0,T]\times\mathrm{Int}(\hat{\mathcal{S}}_{d-1}), \tag{5.3}$$

which is of course very useful to us. Indeed, using a standard mollification argument (taking benefit of the fact that the coefficients $(b_i)_{i\in[\![d]\!]}$ and $(b_i^\circ)_{i\in[\![d]\!]}$ are continuous), we can easily approximate the coefficients $(b_i)_{i\in[\![d]\!]}$ and $(b_i^\circ)_{i\in[\![d]\!]}$ for the sup norm by sequences of coefficients $((b_i^n)_{i\in[\![d]\!]})_{n\geq 1}$ and $((b_i^{\circ,n})_{i\in[\![d]\!]})_{n\geq 1}$ that are time-space continuous and Lipschitz continuous in the space variable (uniformly in the time variable). Hence, if we prove that u in (5.3) satisfies the Hölder estimate stated in Theorem 2.10 for coefficients $(b_i)_{i\in[\![d]\!]}$ and $(b_i^\circ)_{i\in[\![d]\!]}$ that are Lipschitz continuous in space (uniformly in time), we can deduce that the same holds when $(b_i)_{i\in[\![d]\!]}$ and $(b_i^\circ)_{i\in[\![d]\!]}$ are merely continuous by passing to the limit along the aforementioned mollification[16].

In other words, we may assume for our purpose that $(b_i)_{i\in[\![d]\!]}$ and $(b_i^\circ)_{i\in[\![d]\!]}$ are Lipschitz continuous in space, uniformly in time, provided that we prove that the resulting Hölder estimate does not depend on the Lipschitz constants of $(b_i)_{i\in[\![d]\!]}$ and $(b_i^\circ)_{i\in[\![d]\!]}$.

Throughout the section, we assume that, as in the statement of Theorem 2.10, ε is in $(0,1)$.

5.1. Preliminary results on coupling and conditioning

5.1.1. Conditioning on the m first coordinates

The core of the analysis is based upon the probabilistic representation (5.2) and in turn on the properties of the process $\boldsymbol{P} = (P_t^1, \cdots, P_t^d)_{0\leq t\leq T}$ solving equation (5.1).

As we already alluded to a few lines before, our general strategy relies on an induction argument based upon the dimension of the state variable. This is precisely the goal of this paragraph to clarify the way we may reduce dimension inductively. General speaking, the arguments is based on a conditioning argument.

[16] As noticed in Remark 2.11, we may think of adapting the argument when $(b_i)_{i\in[\![d]\!]}$ and $(b_i^\circ)_{i\in[\![d]\!]}$ are just measurable. In such a case, we could no longer approximate them in sup norm and we should work instead with some L^p norm. This should require to control the mean occupation measure of the process $\boldsymbol{P}^{t,p}$ in terms of this L^p norm. We believe that this is possible thanks to the ellipticity property of the operator inside the domain.

In order to make it clear, we rewrite (5.1), but using (at least for the sole purpose of the statement of Proposition 5.2 right below) the letter X_t instead of P_t for the unknown:

$$\mathrm{d}X_t^i = \Big(\varphi(X_t^i) + b_i(t, X_t) + X_t^i b_i^\circ(t, X_t)\Big)\mathrm{d}t + \varepsilon\sqrt{X_t^i}\sum_{j\in[\![d]\!]}\sqrt{X_t^j}\mathrm{d}\overline{W}_t^{i,j}, \quad t \in [t_0, T],\ i \in [\![d]\!], \tag{5.4}$$

for a given initial time t_0. Our rationale to change P_t into X_t is motivated by the fact that we feel better to keep the letter P_t for the new state variable once the dimension has been reduced. The objective is then to write the law of $(X_t)_{t_0\le t\le T}$ in the form

$$(X_t)_{t_0\le t\le T} \overset{\text{law}}{=} \Big(P_t^{\circ,1}, \cdots, P_t^{\circ,m}, \varsigma^2(P_t^\circ)P_t^1, \cdots, \varsigma^2(P_t^\circ)P_t^{d-m}\Big)_{t_0\le t\le T}, \tag{5.5}$$

where $\boldsymbol{P}^\circ = (P_t^\circ = (P_t^{\circ,1}, \cdots, P_t^{\circ,m}))_{t_0\le t\le T}$ and $\boldsymbol{P} = (P_t = (P_t^1, \cdots, P_t^{d-m}))_{t_0\le t\le T}$ are new stochastic processes taking respectively values within the set $\hat{\mathcal{S}}_m = \{(p_1^\circ, \cdots, p_m^\circ) \in (\mathbb{R}_+)^m : \sum_{i=1}^m p_i^\circ \le 1\}$ and $\mathcal{S}_{d-m-1} = \{(p_1, \cdots, p_{d-m}) \in (\mathbb{R}_+)^{d-m} : \sum_{i=1}^{d-m} p_i = 1\}$. Above, ς is given by

$$\varsigma(p^\circ) := \sqrt{1 - (p_1^\circ + \cdots + p_m^\circ)}, \quad (p_1^\circ, \cdots, p_m^\circ) \in \hat{\mathcal{S}}_m.$$

Proposition 5.2. *Given coefficients $(b_i)_{i\in[\![d]\!]}$ and $(b_i^\circ)_{i\in[\![d]\!]}$ as in the statement of Theorem 2.10, there exist new coefficients*

- *$(\widetilde{b_i})_{i\in[\![d-m]\!]}$, with values in $(\mathbb{R}_+)^{d-m}$, that are bounded by a constant that only depends on $(\|b_i\|_\infty)_{i\in[\![d]\!]}$,*
- *$(\widetilde{b_i^\circ})_{i\in[\![d-m]\!]}$, with values in $\mathbb{R}^{d-m}$, that are bounded by a constant that only depends on $(\|b_i\|_\infty)_{i\in[\![d]\!]}$ and $(\|b_i^\circ\|_\infty)_{i\in[\![d]\!]}$,*

and that are Lipschitz continuous in space uniformly in time such that

- *whenever $\delta \in (0,1)$ and $\kappa \ge \varepsilon^2/2$,*
- *for any family of antisymmetric Brownian motions $\overline{\boldsymbol{W}}^\circ = (\overline{W}_t^\circ = (\overline{W}_t^{\circ,i,j})_{i,j\in[\![d]\!]:i\ne j})_{0\le t\le T}$ of dimension $d(d-1)/2$ that is independent of $\overline{\boldsymbol{W}} = (\overline{W}_t = (\overline{W}_t^{i,j})_{i,j\in[\![d]\!]:i\ne j})_{0\le t\le T}$,*
- *for any given initial condition $(t_0, p^\circ, p) \in [0,T]\times\hat{\mathcal{S}}_m\times\mathcal{S}_{d-m-1}$,*

the system

$$\begin{aligned}
\mathrm{d}P_t^i &= \varsigma^{-2}(P_t^\circ)\Big(\varphi(\varsigma^2(P_t^\circ)P_t^i) + \widetilde{b_i}(t, P_t^\circ, P_t) + P_t^i\widetilde{b_i^\circ}(t, P_t^\circ, P_t)\Big)\mathrm{d}t \\
&\quad + \varepsilon\varsigma^{-1}(P_t^\circ)\sum_{j\in[\![d-m]\!]:j\ne i}\sqrt{P_t^iP_t^j}\mathrm{d}\overline{W}_t^{i,j}, \quad i \in [\![d-m]\!], \\
\mathrm{d}P_t^{\circ,i} &= \Big(\varphi(P_t^{\circ,i}) + b_i(t, (P_t^\circ, \varsigma^2(P_t^\circ)P_t)) + P_t^ib_i^\circ(t, P_t^\circ, \varsigma^2(P_t^\circ)P_t)\Big)\mathrm{d}t \\
&\quad + \varepsilon\sum_{j\in[\![m]\!]:j\ne i}\sqrt{P_t^{\circ,i}P_t^{\circ,j}}\mathrm{d}\overline{W}_t^{\circ,i,j} + \varepsilon\varsigma(P_t^\circ)\sqrt{P_t^{\circ,i}}\sum_{j\in[\![d-m]\!]}\sqrt{P_t^j}\mathrm{d}\overline{W}_t^{\circ,i,m+j}, \quad i \in [\![m]\!],
\end{aligned} \tag{5.6}$$

for $t \in [t_0, T]$, with $(P_{t_0}^\circ, P_{t_0}) = (p^\circ, p)$, has a unique strong solution, which satisfies the identity in law (5.5) *whenever* (5.4) *is initialized from $(p^\circ, \varsigma^2(p^\circ)p)$ at time t_0.*

The proof of Proposition 5.2 is deferred to Subsection 5.4. Throughout, we denote by $\mathbb{F}^{\boldsymbol{W}^\circ,\boldsymbol{W}} = (\mathcal{F}_t^{\boldsymbol{W}^\circ,\boldsymbol{W}})_{0\le t\le T}$ the augmented filtration generated by $\boldsymbol{W}^\circ = ((W_t^{\circ,i,j})_{0\le t\le T})_{i,j\in[\![d]\!]:i\neq j}$ and $\boldsymbol{W} = ((W_t^{i,j})_{0\le t\le T})_{i,j\in[\![d]\!]:i\neq j}$, the latter two being implicitly understood as two independent collections of Brownian motions such that $\overline{\boldsymbol{W}}^{\circ,i,j} = (\boldsymbol{W}^{\circ,i,j} - \boldsymbol{W}^{\circ,j,i})/\sqrt{2}$ and $\overline{\boldsymbol{W}}^{i,j} = (\boldsymbol{W}^{i,j} - \boldsymbol{W}^{j,i})/\sqrt{2}$, when $\overline{\boldsymbol{W}}^\circ$ and $\overline{\boldsymbol{W}}$ are as in the statement of Proposition 5.2.

5.1.2. Main coupling estimate

Proposition 5.3. *Assume that δ in (2.19) belongs to $(0, 1/(4\sqrt{d}))$ Moreover, take two initial conditions (t_0, p°, p) and (t_0, q°, q) in $[0,T]\times\hat{\mathcal{S}}_m\times\mathcal{S}_{d-m-1}$, with $m\in\{1,\cdots,d-1\}$, such that $|p^\circ|_1 := p_1^\circ + \cdots + p_m^\circ \le 1/2$, $|q^\circ|_1 \le 1/2$, and $|p-q| < \delta^2/(64\sqrt{d})$. On the (filtered) probability space carrying $\overline{\boldsymbol{W}}^\circ$ and $\overline{\boldsymbol{W}}$, call $(P_t^\circ = (P_t^{\circ,1},\cdots,P_t^{\circ,m}), P_t = (P_t^1,\cdots,P_t^{d-m}))_{t_0\le t\le T}$ the solution to (5.6) with (t_0,p°,p) as initial condition.*

Then, for any $\eta\in(0,1)$, there exists a threshold $\kappa_0\ge 2$, only depending on η, ε and $(\|b_i\|_\infty)_{i\in[\![d]\!]}$ (but not on δ) such that, for any $\kappa\ge\kappa_0$, we can find a constant C, depending on δ, ε, κ, η, $(\|b_i\|_\infty)_{i\in[\![d]\!]}$, $(\|b_i^\circ\|_\infty)_{i\in[\![d]\!]}$ and T such that, provided that $|p-q|^{1/3}\le T-t_0$, there exists an adapted process $(Q_t^\circ = (Q_t^{\circ,1},\cdots,Q_t^{\circ,m}), Q_t = (Q_t^1,\cdots,Q_t^{d-m}))_{t_0\le t\le T}$ that has the same law as the solution to (5.6) with (t_0,q°,q) as initial condition and for which the following property holds true.

If we call

$$\widetilde{P}_t := \big(\sqrt{P_t^1},\cdots,\sqrt{P_t^{d-m}}\big),\quad \widetilde{Q}_t := \big(\sqrt{Q_t^1},\cdots,\sqrt{Q_t^{d-m}}\big),\quad t\in[t_0,T],$$

and

$$\begin{aligned}&\varrho := \inf\{s\ge t_0 : \big|P_s^\circ - Q_s^\circ\big| > \big|\widetilde{P}_s - \widetilde{Q}_s\big|\},\quad \rho := \inf\{s\ge t_0 : \big|P_s^\circ\big|_1 \ge 3/4\},\\ &\sigma := \inf\{s\ge t_0 : \big|\widetilde{P}_s - \widetilde{Q}_s\big| \ge \delta/4\},\quad \tau := \inf\{s\ge t_0 : \big|\widetilde{P}_s - \widetilde{Q}_s\big| = 0\},\end{aligned} \tag{5.7}$$

then

$$\mathbb{P}(\{\varpi_S < \tau\wedge\varrho\wedge\rho\}) \le C\frac{|p-q|^{1/12}}{\min_{i\in[\![d-m]\!]}(\max(p_i,q_i))^\eta}, \tag{5.8}$$

where $\varpi_S := \varpi\wedge S$, with $\varpi := \rho\wedge\varrho\wedge\sigma\wedge\tau$ and $S := t_0 + |p-q|^{1/3}$.

The proof of Proposition 5.3 is deferred to Subsection 5.3.

Remark 5.4. *We now explain the difficulty when the dimension d is greater than or equal to 3 and the reason why we need an induction argument to derive the required Hölder estimate.*

A naive way to proceed is indeed to choose $m = 0$ in the above statement. In such a case, the process $(P^1,\cdots,P^{d-m})$ in (5.6) coincides with the solution $(P^1,\cdots,P^d)$ in (5.1). In other words, Proposition 5.3 with $m = 0$ reads as a coupling estimate for the solution to (5.1).

Let us now see what the right hand side of (5.8) becomes when $m = 0$. Up to an obvious change of coordinates, we then may assume that $\min_{i\in[\![d]\!]}(\max(p_i,q_i)) = \max(p_1,q_1)$, in which case the right-hand side of (5.8) writes $|p-q|^{1/12}/\max(p_1,q_1)^\eta$, both $p = (p_1,\cdots,p_d)$ and $q = (q_1,\cdots,q_d)$ being now regarded as d-dimensional vectors. The point is then to upper bound $|p-q|^{1/12}/\max(p_1,q_1)^\eta$. When $d = 2$, this is pretty easy because

$$|p-q| = \sqrt{|p_1-q_1|^2+|p_2-q_2|^2} = \sqrt{|p_1-q_1|^2+|(1-p_1)-(1-q_1)|^2} \le \sqrt{2}|p_1-q_1|,$$

and then we get

$$\frac{|p-q|^{1/12}}{\max(p_1,q_1)^\eta} \le \sqrt{2}|p-q|^{1/12-\eta}.$$

Unfortunately, this argument no longer works when $d \ge 3$ since, in that case, one of the entries $(|p_i - q_i|)_{i=2,\cdots,d}$ may be much larger than $|p_1 - q_1|$.

5.2. Derivation of the Hölder estimate and proof of Theorem 2.10

We now explain how to derive the Hölder estimate in Theorem 2.10 from Proposition 5.3. As we already alluded to, it relies on an additional iteration on the dimension, which is in turn inspired by earlier PDE results on Kimura diffusions, see for instance [2]. The induction assumption takes the following form.

Take h, ℓ and u as in (4.12). For a given $m \in [\![d]\!]$, call $\mathscr{P}_m$ the following property: For any $\varepsilon, \eta \in (0,1)$ and $\delta \in (0, 1/(4\sqrt{d}))$, there exist a threshold κ_0, only depending on ε, η, m and $(\|b_i\|_\infty)_{i\in[\![d]\!]}$, and an exponent $\alpha \in (0,1)$, only depending on η and m, such that, for any $\kappa \ge \kappa_0$, we can find a constant C, depending on δ, ε, κ, η, $(\|b_i\|_\infty)_{i\in[\![d]\!]}$, $(\|b_i^\circ\|_\infty)_{i\in[\![d]\!]}$, $\|h\|_\infty$, $\|\ell\|_{1,\infty}$ and T such that, for any $p = (p_1, \cdots, p_d)$ and $q = (q_1, \cdots, q_d)$ in $\mathcal{S}_{d-1}$,

$$|u(t,p_1,\cdots,p_d) - u(t,q_1,\cdots,q_d)| \le C\frac{|p-q|^\alpha}{\max(p,q)_{(m)}^{\eta\alpha}}, \tag{5.9}$$

where $\max(p,q)_{(m)}$ denotes the mth element in the increasing reordering of

$$\max(p,q) = (\max(p_1,q_1), \cdots, \max(p_d,q_d)).$$

We then have the following two propositions:

Proposition 5.5. *Within the framework of Theorem 2.10, $\mathscr{P}_1$ holds true.*

Proposition 5.6. *Within the framework of Theorem 2.10, assume that there exists an integer $m \in \{1, \cdots, d-1\}$ such that $\mathscr{P}_m$ holds true. Then, $\mathscr{P}_{m+1}$ holds true.*

Notice that Proposition 5.6 implies Theorem 2.10: It suffices to choose $\eta = 1/2$ in $\mathscr{P}_d$, noticing that $\max(p,q)_{(d)}$ is necessarily greater than $1/d$. Below, we directly prove Proposition 5.6. The proof of Proposition 5.5 is completely similar (somehow, everything works as if we had a property $\mathscr{P}_0$).

Proof. For some $m \in \{1, \cdots, d-1\}$ and some $\eta \in (0, 1/4)$, we consider κ_0 as being the maximum of κ_0 given by Proposition 5.3 with η replaced by $\eta/12$ therein and of κ_0 given by $\mathscr{P}_m$ with η therein. Also, we consider α given by $\mathscr{P}_m$ and $\kappa \ge \kappa_0$. We then assume that $\mathscr{P}_m$ holds true and we take $p, q \in \mathcal{S}_{d-1}$ together with $t_0 \in [0,T]$. Without any loss of generality, we may assume that

$$\max(p_1,q_1) \le \max(p_2,q_2) \le \cdots \le \max(p_m,q_m) \le \frac{1}{2md}. \tag{5.10}$$

Observe that if the last inequality is not satisfied, the bound (5.9) at rank $m+1$ (at t_0 instead of t) is a straightforward consequence of the bound at rank m with $(2md)^{\eta\alpha}C$ instead of C as constant. For sure, we may also assume that

$$\max(p_1,q_1) \le \max(p_2,q_2) \le \cdots \le \max(p_d,q_d), \tag{5.11}$$

in which case $\max(p_d, q_d)$ is the largest element in the sequence $\max(p_i, q_i)$. In particular, at least p_d or q_d is above $1/d$ (since one of the two elements dominates all the other elements in the family $(p_1, \cdots, p_d, q_1, \cdots, q_d)$). Hence, we may assume that $\min(p_d, q_d) \geq 1/(2d)$. Again, the proof is over if not since $|p - q|$ is then necessarily larger than $1/(2d)$: Tuning C accordingly, (5.9) follows from the fact that u is bounded, see Proposition 5.1. By the same argument, we may assume that $|p - q| < \delta^2/(128d^{3/2})$.

First Step. Clearly,

$$\begin{aligned}
&|u(t_0, p_1, \cdots, p_d) - u(t_0, q_1, \cdots, q_d)| \\
&\leq \sum_{i \in [\![m]\!]} \Big(\big| u(t_0, q_1, \cdots, q_{i-1}, p_i, \cdots, p_{d-1}, 1 - q_1 - \cdots q_{i-1} - p_i - \cdots - p_{d-1}) \\
&\qquad - u(t_0, q_1, \cdots, q_i, p_{i+1}, \cdots, 1 - q_1 - \cdots q_i - p_{i+1} - \cdots - p_{d-1}) \big| \Big) \\
&\quad + \big| u(t_0, q_1, \cdots, q_m, p_{m+1}, \cdots, p_{d-1}, 1 - q_1 - \cdots q_m - p_{m+1} - \cdots - p_{d-1}) \\
&\qquad - u(t_0, q_1, \cdots, q_m, q_{m+1}, \cdots) \big|,
\end{aligned} \tag{5.12}$$

with the obvious convention that $(q_1, \cdots, q_{i-1}, p_i, \cdots, p_{d-1}, 1 - q_1 - \cdots q_{i-1} - p_i - \cdots - p_{d-1}) = (p_1, \cdots, p_d)$ when $i = 1$. Notice from (5.10) and from the bound $\min(p_d, q_d) \geq 1/(2d)$ that, for $i \in [\![m]\!]$, $q_1 + \cdots + q_i \leq 1/(2d)$ and $p_i + \cdots + p_{d-1} \leq 1 - 1/(2d)$, which fully justifies the fact that all the entries above are non-negative. Obviously, by the induction assumption, for any $i \in [\![m]\!]$,

$$\begin{aligned}
&\sum_{i \in [\![m]\!]} \Big(\big| u(t_0, q_1, \cdots, q_{i-1}, p_i, \cdots, p_{d-1}, 1 - q_1 - \cdots q_{i-1} - p_i - \cdots - p_{d-1}) \\
&\qquad - u(t_0, q_1, \cdots, q_i, p_{i+1}, \cdots, 1 - q_1 - \cdots q_i - p_{i+1} - \cdots - p_{d-1}) \big| \Big) \\
&\leq C \sum_{i \in [\![m]\!]} \frac{|p_i - q_i|^\alpha}{\max(p, q)_{(m)}^{\eta\alpha}} \leq C |p - q|^{(1-\eta)\alpha},
\end{aligned}$$

where we used the fact that $\max(p, q)_{(m)} = \max(p_m, q_m)$ and where we modified the value of C in the last term. The conclusion is that, in (5.12), we can focus on the last term. Equivalently, we can assume that $p_i = q_i$, for $i = 1, \cdots, m$, provided we replace (5.11) by (which is weaker, but which is the right assumption here since there is no way to compare properly the last coordinates in the last term of (5.12))

$$\max_{i \in [\![m]\!]} p_i = \max_{i \in [\![m]\!]} q_i \leq \max(p_{m+1}, q_{m+1}) \leq \cdots \leq \max(p_{d-1}, q_{d-1}), \tag{5.13}$$

with $q_d \geq 1/(2d)$. We now invoke Proposition 5.1 to represent $u(t, p_1, \cdots, p_d)$ and $u(t, q_1, \cdots, q_d)$ through the respective solutions to (5.1) together with Proposition 5.2 above which provides another representation for the process used in the Kolmogorov formula (5.2). In particular, we can find $(P^{\circ,1}, \cdots, P^{\circ,m}, P^1, \cdots, P^{d-m})$ as in the statement of Proposition 5.3 such that the tuple $(P^{\circ,1}, \cdots, P^{\circ,m}, \varsigma^2(P^\circ)P^1, \cdots, \varsigma^2(P^\circ)P^{d-m})$ has the same law as the solution to (5.1) when starting from p at time t_0, and, in a similar manner, $(Q^{\circ,1}, \cdots, Q^{\circ,m}, Q^1, \cdots, Q^{d-m})$ such that the tuple $(Q^{\circ,1}, \cdots, Q^{\circ,m}, \varsigma^2(Q^\circ)Q^1, \cdots, \varsigma^2(Q^\circ)Q^{d-m})$ has the same law as the solution to (5.1) when starting from q at time t_0. In particular, we have $(P^{\circ,1}_{t_0}, \cdots, P^{\circ,m}_{t_0}) = (Q^{\circ,1}_{t_0}, \cdots, Q^{\circ,m}_{t_0}) = p^\circ$, with $p^\circ = (p_1, \cdots, p_m)$, $(P^1_{t_0}, \cdots, P^{d-m}_{t_0}) = \varsigma^{-2}(p^\circ)(p_{m+1}, \cdots, p_d)$ and $(Q^1_{t_0}, \cdots, Q^{d-m}_{t_0}) = \varsigma^{-2}(p^\circ)(q_{m+1}, \cdots, q_d)$.

Then, for any deterministic time $S \in [t_0, T]$, using the same notation as in the statement of Proposition 5.3,

$$\begin{aligned} u(t_0, p_1, \cdots, p_d) &= \mathbb{E}\Big[u\Big(\varpi_S, P^{\circ,1}_{\varpi_S}, \cdots, P^{\circ,m}_{\varpi_S}, \varsigma^2(P^\circ_{\varpi_S})P^1_{\varpi_S}, \cdots, \varsigma^2(P^\circ_{\varpi_S})P^{d-m}_{\varpi_S}\Big)\Big] + O(S - t_0), \\ u(t_0, q_1, \cdots, q_d) &= \mathbb{E}\Big[u\Big(\varpi_S, Q^{\circ,1}_{\varpi_S}, \cdots, Q^{\circ,m}_{\varpi_S}, \varsigma^2(Q^\circ_{\varpi_S})Q^1_{\varpi_S}, \cdots, \varsigma^2(Q^\circ_{\varpi_S})Q^{d-m}_{\varpi_S}\Big)\Big] + O(S - t_0), \end{aligned} \tag{5.14}$$

where $|O(r)| \le \|h\|_\infty r$. To make it simpler, we also let (the notation $(X_t)_{t_0 \le t \le T}$ below is rather abusive since $(X_t)_{t_0 \le t \le T}$ also denotes the solution to (5.4), but, in fact, Proposition 5.2 says both $(X_t)_{t_0 \le t \le T}$'s have the same law)

$$\begin{aligned} X_t &= \Big(P^{\circ,1}_t, \cdots, P^{\circ,m}_t, \varsigma^2(P^\circ_t)P^1_t, \cdots, \varsigma^2(P^\circ_t)P^{d-m}_t\Big), \\ Y_t &= \Big(Q^{\circ,1}_t, \cdots, Q^{\circ,m}_t, \varsigma^2(Q^\circ_t)Q^1_t, \cdots, \varsigma^2(Q^\circ_t)Q^{d-m}_t\Big), \quad t \in [t_0, T]. \end{aligned}$$

We then denote by $(\max(X_t, Y_t)_{(1)}, \cdots, \max(X_t, Y_t)_{(d)})$ the order statistic of the d-dimensional tuple $(\max(X_t, Y_t)_1, \cdots, \max(X_t, Y_t)_d)$.

Second Step. We first assume that $S := t_0 + |p - q|^{1/3} \le T$. The strategy is to split into four events the set Ω over which the expectations appearing in (5.14) are computed.

1st event. On the event $E_1 := \{\varpi_S = \tau\} \subset \{(P^1_{\varpi_S}, \cdots, P^{d-m}_{\varpi_S}) = (Q^1_{\varpi_S}, \cdots, Q^{d-m}_{\varpi_S})\}$, we have, by the induction assumption and from the Lipschitz property of ς^2,

$$\begin{aligned} \big|u(\varpi_S, X_{\varpi_S}) - u(\varpi_S, Y_{\varpi_S})\big| &= \Big|u\Big(\varpi_S, P^{\circ,1}_{\varpi_S}, \cdots, P^{\circ,m}_{\varpi_S}, \varsigma^2(P^\circ_{\varpi_S})P^1_{\varpi_S}, \cdots, \varsigma^2(P^\circ_{\varpi_S})P^{d-m}_{\varpi_S}\Big) \\ &\qquad - u\Big(\varpi_S, Q^{\circ,1}_{\varpi_S}, \cdots, Q^{\circ,m}_{\varpi_S}, \varsigma^2(Q^\circ_{\varpi_S})Q^1_{\varpi_S}, \cdots, \varsigma^2(Q^\circ_{\varpi_S})Q^{d-m}_{\varpi_S}\Big)\Big| \\ &\le \frac{C}{\max(X_{\varpi_S}, Y_{\varpi_S})^{\eta\alpha}_{(m)}}\big|P^\circ_{\varpi_S} - Q^\circ_{\varpi_S}\big|^\alpha, \end{aligned} \tag{5.15}$$

the constant C being allowed to change from line to line provided that it only depends on the parameters listed in the induction assumption.

Assume that $\max(X_{\varpi_S}, Y_{\varpi_S})_{(m)} < \max(X^l_{\varpi_S}, Y^l_{\varpi_S})$ for any $l = m+1, \cdots, d$, then necessarily $\max(X_{\varpi_S}, Y_{\varpi_S})_{(m)} \ge \max(X^i_{\varpi_S}, Y^i_{\varpi_S})$ for any $i = 1, \cdots, m$. We then obtain $C|P^\circ_{\varpi_S} - Q^\circ_{\varpi_S}|^{(1-\eta)\alpha}$ as upper bound for the right-hand side of (5.15). Therefore, we can focus on the complementary event when $\max(X_{\varpi_S}, Y_{\varpi_S})_{(m)} \ge \max(X^l_{\varpi_S}, Y^l_{\varpi_S})$ for some $l = m+1, \cdots, d$. We obtain

$$\big|u(\varpi_S, X_{\varpi_S}) - u(\varpi_S, Y_{\varpi_S})\big| \le C|P^\circ_{\varpi_S} - Q^\circ_{\varpi_S}|^{(1-\eta)\alpha} + \sum_{l=m+1}^{d} \frac{C}{\max(X^l_{\varpi_S}, Y^l_{\varpi_S})^{\eta\alpha}}\big|P^\circ_{\varpi_S} - Q^\circ_{\varpi_S}\big|^\alpha,$$

which we rewrite into (recalling that $p_1 = q_1, \cdots, p_m = q_m$)

$$\begin{aligned} \big|u(\varpi_S, X_{\varpi_S}) - u(\varpi_S, Y_{\varpi_S})\big| &\le C\big|P^\circ_{\varpi_S} - P^\circ_{t_0} - (Q^\circ_{\varpi_S} - Q^\circ_{t_0})\big|^{(1-\eta)\alpha} \\ &\quad + \sum_{l=m+1}^{d} \frac{C}{\max(X^l_{\varpi_S}, Y^l_{\varpi_S})^{\eta\alpha}}\big|P^\circ_{\varpi_S} - P^\circ_{t_0} - (Q^\circ_{\varpi_S} - Q^\circ_{t_0})\big|^\alpha. \end{aligned} \tag{5.16}$$

In order to upper bound $C/(\max(X^l_{\varpi_S}, Y^l_{\varpi_S}))^{\eta\alpha}$, we expand $((X^l_t)^{-2\eta\alpha})_{t_0 \le t \le \varpi_S}$ by Itô's formula, for $l = m+1, \cdots, d$. To do so, we recall that $(X^1, \cdots, X^d)$ has the same law as the solution of

(5.4). It is an Itô process with bounded coefficients (in terms of $(\|b_i\|_\infty)_{i=1,\cdots,d}$ and $(\|b_i^\circ\|_\infty)_{i=1,\cdots,d}$). Importantly, the drift of the lth coordinate is lower bounded by $\kappa - \|b_l^\circ\|_\infty p_l$ when $p_l \le \delta$. Recalling that $\kappa_0 \ge 2\varepsilon^2$, we easily deduce that, for a new value of C (say $C \ge 1$) whose value is allowed to change from line to line (see also footnote 15 for the meaning of the inequality right below),

$$\begin{aligned}
&\mathrm{d}\Big((X_t^l)^{-2\eta\alpha}(t-t_0)^{\alpha/2}\Big)\\
&\le \Big(C(t-t_0)^{\alpha/2-1}(X_t^l)^{-2\eta\alpha} - \eta\alpha(2\kappa - \varepsilon^2(1+2\eta\alpha))(t-t_0)^{\alpha/2}(X_t^l)^{-(1+2\eta\alpha)}\Big)\mathrm{d}t + \mathrm{d}m_t\\
&\le \Big(C(t-t_0)^{\alpha/2-1}(X_t^l)^{-2\eta\alpha} - C^{-1}(t-t_0)^{\alpha/2}(X_t^l)^{-(1+2\eta\alpha)}\Big)\mathrm{d}t + \mathrm{d}m_t\\
&= (t-t_0)^{\alpha/2}(X_t^l)^{-2\eta\alpha}\Big(C(t-t_0)^{-1} - C^{-1}(X_t^l)^{-1}\Big)\mathrm{d}t + \mathrm{d}m_t\\
&\le (t-t_0)^{\alpha/2}(X_t^l)^{-2\eta\alpha}\Big(C(t-t_0)^{-1} - C^{-1}(X_t^l)^{-1}\Big)\mathbf{1}_{\{X_t^l \ge (t-t_0)/C^2\}}\mathrm{d}t + \mathrm{d}m_t\\
&\le C(t-t_0)^{\alpha/2-1-2\eta\alpha}\mathrm{d}t + \mathrm{d}m_t,
\end{aligned}$$

where $(m_t)_{t\ge 0}$ is a local martingale. Recall that $\eta < 1/4$ (see the remark at the very beginning of the proof), we get by a standard localization argument:

$$\mathbb{E}[(X_{\varpi_S}^l)^{-2\eta\alpha}(\varpi_S - t_0)^{\alpha/2}] \le C.$$

Using the fact that the coefficients in the second equation of (5.6) are bounded, we deduce that, from Kolmogorov–Centsov theorem, $P^\circ - Q^\circ$ has a version that is $1/3$-Hölder continuous and the Hölder constant Λ has a finite fourth moment, which we may assume to be bounded by C. Hence,

$$\mathbb{E}\Big[\frac{|P_{\varpi_S}^\circ - Q_{\varpi_S}^\circ - (P_{t_0}^\circ - Q_{t_0}^\circ)|^{2\alpha}}{(\varpi_S - t_0)^{\alpha/2}}\Big] \le \mathbb{E}[\Lambda^{2\alpha}(\varpi_S - t_0)^{\alpha/6}] \le C\mathbb{E}[(\varpi_S - t_0)^{\alpha/3}]^{1/2}.$$

Similarly,

$$\begin{aligned}
\mathbb{E}\Big[|P_{\varpi_S}^\circ - Q_{\varpi_S}^\circ - (P_{t_0}^\circ - Q_{t_0}^\circ)|^{(1-\eta)\alpha}\Big] &\le \mathbb{E}[\Lambda^{(1-\eta)\alpha}(\varpi_S - t_0)^{(1-\eta)\alpha/3}]\\
&\le C\mathbb{E}[(\varpi_S - t_0)^{2(1-\eta)\alpha/3}]^{1/2}.
\end{aligned}$$

From (5.16), we get (use the condition $\eta < 1/4$ to get the last bound)

$$\begin{aligned}
\mathbb{E}\Big[\big|u(\varpi_S, X_{\varpi_S}) - u(\varpi_S, Y_{\varpi_S})\big|\mathbf{1}_{E_1}\Big] &\le C\mathbb{E}[(\varpi_S - t_0)^{2(1-\eta)\alpha/3}]^{1/2} + C\mathbb{E}[(\varpi_S - t_0)^{\alpha/3}]^{1/4}\\
&\le C|p-q|^{\alpha/36} + C|p-q|^{(1-\eta)\alpha/9} \le C|p-q|^{\alpha/36}. \qquad (5.17)
\end{aligned}$$

2nd event. On the event $E_2 := \{|P_{\varpi_S}^\circ - Q_{\varpi_S}^\circ| \ge |\widetilde{P}_{\varpi_S} - \widetilde{Q}_{\varpi_S}|\} \supset \{\varpi_S = \varrho\}$ (with the same notation as in the statement of Proposition 5.3), we have

$$\begin{aligned}
|X_{\varpi_S} - Y_{\varpi_S}|^2 &= \sum_{i\in[\![m]\!]}\big|P_{\varpi_S}^{\circ,i} - Q_{\varpi_S}^{\circ,i}\big|^2 + \sum_{i\in[\![d-m]\!]}\big|P_{\varpi_S}^i - Q_{\varpi_S}^i\big|^2\\
&= \sum_{i\in[\![m]\!]}\big|P_{\varpi_S}^{\circ,i} - Q_{\varpi_S}^{\circ,i}\big|^2 + \sum_{i\in[\![d-m]\!]}\big|\widetilde{P}_{\varpi_S}^i + \widetilde{Q}_{\varpi_S}^i\big|^2\big|\widetilde{P}_{\varpi_S}^i - \widetilde{Q}_{\varpi_S}^i\big|^2 \le 5|P_{\varpi_S}^\circ - Q_{\varpi_S}^\circ|^2.
\end{aligned}$$

Modifying the value of the constant C in (5.9), we deduce that

$$\begin{aligned}\big|u(\varpi_S, X_{\varpi_S}) - u(\varpi_S, Y_{\varpi_S})\big| &= \Big|u\Big(\varpi_S, P^{\circ,1}_{\varpi_S}, \cdots, P^{\circ,m}_{\varpi_S}, \varsigma^2(P^\circ_{\varpi_S})P^1_{\varpi_S}, \cdots, \varsigma^2(P^\circ_{\varpi_S})P^{d-m}_{\varpi_S}\Big) \\ &\qquad - u\Big(\varpi_S, Q^{\circ,1}_{\varpi_S}, \cdots, Q^{\circ,m}_{\varpi_S}, \varsigma^2(Q^\circ_{\varpi_S})Q^1_{\varpi_S}, \cdots, \varsigma^2(Q^\circ_{\varpi_S})Q^{d-m}_{\varpi_S}\Big)\Big| \\ &\le \frac{C}{\max(X_{\varpi_S}, Y_{\varpi_S})^{\eta\alpha}_{(m)}}\big|P^\circ_{\varpi_S} - Q^\circ_{\varpi_S}\big|^\alpha,\end{aligned}$$

which is the same as (5.15). Therefore, we get the same conclusion as in the first step, see (5.17):

$$\mathbb{E}\Big[\big|u(\varpi_S, X_{\varpi_S}) - u(\varpi_S, Y_{\varpi_S})\big|\mathbf{1}_{E_2}\Big] \le C|p-q|^{\alpha/36}. \tag{5.18}$$

3rd event. We now consider the event $E_3 := \{\sup_{t\in[t_0,S]}|P^\circ_t|_1 \ge \frac{3}{4}\} \supset \{\varpi_S = \rho\}$, where we recall from (5.10) that $|P^\circ_{t_0}|_1 = \sum_{i=1}^m p_i \le \frac{1}{2d} \le \frac{1}{2}$. Obviously (since the SDE for P° has bounded coefficients), there exists a constant c such that $\mathbb{E}[\sup_{t\in[t_0,S]}|P^\circ_t - P^\circ_{t_0}|^2] \le c(S-t_0) = c|p-q|^{1/3}$. Therefore, using the fact that u is bounded together with Markov's inequality, we deduce that

$$\mathbb{E}\Big[\big|u(\varpi\wedge S, X_{\varpi\wedge S}) - u(\varpi\wedge S, Y_{\varpi\wedge S})\big|\mathbf{1}_{E_3}\Big] \le C|p-q|^{1/3}. \tag{5.19}$$

4th event. Lastly, we let $E_4 := \{\varpi_S < \tau\wedge\varrho\wedge\rho\}$. Since $\varsigma^{-2}(p^\circ) = \varsigma^{-2}(q^\circ) \le 2d$ and $|p-q| < \delta^2/(128d^{3/2})$, we have $|\varsigma^{-2}(p^\circ)(p_{m+1},\cdots,p_d) - \varsigma^{-2}(q^\circ)(q_{m+1},\cdots,q_d)| < \delta^2/(64\sqrt{d})$. Hence, by Proposition 5.3 (with p therein being given by $\varsigma^{-2}(p^\circ)(p_{m+1},\cdots,p_d)$ and similarly for q) and with η therein being replaced by $\eta/12$, we know that, for a new value of C,

$$\mathbb{P}(E_4) \le C\frac{|p-q|^{1/12}}{\min_{i=m+1,\cdots,d}(\max(p_i,q_i))^{\eta/12}} \le C\frac{|p-q|^{1/12}}{\max(p_{m+1},q_{m+1})^{\eta/12}} = C\frac{|p-q|^{1/12}}{\max(p,q)^{\eta/12}_{(m+1)}},$$

where the derivation of the last two terms follows from (5.13) and from the condition $\max(p_d,q_d) \ge 1/(2d)$. Since the left-hand side is less than 1, we deduce that, for any exponent $\beta\in(0,1]$,

$$\mathbb{P}(E_4) \le \mathbb{P}(E_4)^\beta \le C^\beta\frac{|p-q|^{\beta/12}}{\max(p,q)^{\beta\eta/12}_{(m+1)}},$$

and then, for a new value of C possibly depending on β,

$$\mathbb{E}\Big[\big|u(\varpi\wedge S, X_{\varpi\wedge S}) - u(\varpi\wedge S, Y_{\varpi\wedge S})\big|\mathbf{1}_{E_4}\Big] \le C\frac{|p-q|^{\beta/12}}{\max(p,q)^{\beta\eta/12}_{(m+1)}}. \tag{5.20}$$

Conclusion. Here is now the conclusion of the second step. For the same η and α as before, choose the largest $\beta\in(0,1]$ such that $\beta/12 \le \alpha/36$. Finally, let $\alpha' = \beta/12$. Deduce that, for a possibly new value of the constant C therein, all the terms in (5.17), (5.18), (5.19) and (5.20) are bounded by $C|p-q|^{\alpha'}/\max(p,q)^{\eta\alpha'}_{(m+1)}$. Since $E_1\cup E_2\cup E_3\cup E_4 = \Omega$, we deduce that

$$\mathbb{E}\Big[\big|u(\varpi\wedge S, X_{\varpi\wedge S}) - u(\varpi\wedge S, Y_{\varpi\wedge S})\big|\Big] \le C\frac{|p-q|^{\alpha'}}{\max(p,q)^{\alpha'\eta}_{(m+1)}},$$

which together with (5.14), is $\mathscr{P}_{m+1}$, at least for initial conditions (t_0, p) and (t_0, q) such that $T - t_0 \geq |p - q|^{1/3}$, and $\eta < 1/4$. As for the requirement $\eta < 1/4$, this is not a hindrance: Since the denominator in (5.9) is less than 1, the exponent can be increased for free. As for the case $T - t_0 < |p - q|^{1/3}$, it is discussed in the next step.

Third Step. It remains to handle the case that $t_0 + |p - q|^{1/3} \geq T$. We then rewrite (5.14) in the form

$$u(t_0, p_1, \cdots, p_d) = \mathbb{E}\Big[\ell\Big(P_T^{\circ,1}, \cdots, P_T^{\circ,m}, \varsigma^2(P_T^\circ)P_T^1, \cdots, \varsigma^2(P_T^\circ)P_T^{d-m}\Big)\Big] + O(T - t_0),$$
$$u(t_0, q_1, \cdots, q_d) = \mathbb{E}\Big[\ell\Big(Q_T^{\circ,1}, \cdots, Q_T^{\circ,m}, \varsigma^2(Q_T^\circ)Q_T^1, \cdots, \varsigma^2(Q_T^\circ)Q_T^{d-m}\Big)\Big] + O(T - t_0).$$

By (5.5), each expectation may be (directly) rewritten by means of the solution to the stochastic differential equation (5.1). Since the latter has bounded coefficients and since ℓ is Lipschitz continuous, we deduce that

$$|u(t_0, p) - u(t_0, q)| \leq |\ell(p) - \ell(q)| + |u(t_0, p) - \ell(p)| + |u(t_0, q) - \ell(q)| \leq C|p - q| + C(T - t_0)^{1/2},$$

for a constant C that only depends on κ, $\|\ell\|_{1,\infty}$, $(\|b_i\|_\infty)_{i\in[\![d]\!]}$ and $(\|b_i^\circ\|_\infty)_{i\in[\![d]\!]}$. Since $T - t_0 \leq |p - q|^{1/3}$, this completes the proof of $\mathscr{P}_{m+1}$ in the remaining case when $T - t_0 \leq |p - q|^{1/3}$. □

5.3. Proof of the coupling property

We now prove Proposition 5.3 by means of a reflection coupling, inspired by [20, 47]. Throughout, we use the notation $Z_t := \widetilde{P}_t - \widetilde{Q}_t$, for $t \in [0, T]$.

5.3.1. Preliminary result

Proposition 5.7. *Under the same assumption and notation as in the statement of Proposition 5.3, there exists a constant C, only depending on δ, κ, $(\|b_i\|_\infty)_{i\in[\![d]\!]}$, $(\|b_i^\circ\|_\infty)_{i\in[\![d]\!]}$ and T such that (see footnote 15 for the meaning of the inequality right below),*

$$\mathrm{d}|Z_t| \leq C\mathrm{d}t + \sum_{i\in[\![d-m]\!]} \frac{Z_t^i}{|Z_t|} \frac{\widetilde{\beta}_t^i}{\max(\widetilde{P}_t^i, \widetilde{Q}_t^i)}\mathrm{d}t + \varepsilon\frac{\varsigma^{-1}(P_t^\circ) + \varsigma^{-1}(Q_t^\circ)}{2} \sum_{i,j\in[\![d-m]\!]:i\neq j} \frac{Z_t^i}{|Z_t|}\mathrm{d}\overline{W}_t^{i,j}\widetilde{P}_t^j, \quad (5.21)$$

for $t \in [t_0, \varpi \wedge T)$, where for each $i \in [\![d - m]\!]$, $(\beta_t^i)_{0\leq t\leq T}$ is a progressively measurable process that is dominated by $4\|\widetilde{b}_i\|_\infty$.

Proof. Throughout the proof, we use the convention $\overline{W}_t^{i,i} := 0$, for $t \in [0, T]$ and $i \in [\![d - m]\!]$.

First Step. The first step is to perform a change of variable in equation (5.6), letting therein

$$\widetilde{P}_t^i := \sqrt{P_t^i}, \quad t \in [t_0, T],\ i \in [\![d - m]\!].$$

By Itô's formula, we get

$$\begin{aligned}\mathrm{d}\widetilde{P}_t^i = \varsigma^{-2}(P_t^\circ)\Big(&\frac{\varphi(\varsigma^2(P_t^\circ)P_t^i) + \widetilde{b}_i(t, P_t^\circ, P_t) + (\widetilde{P}_t^i)^2\widetilde{b}_i^\circ(t, P_t^\circ, P_t)}{2\widetilde{P}_t^i} - \frac{\varepsilon^2}{8\widetilde{P}_t^i} + \frac{\varepsilon^2}{8}\widetilde{P}_t^i\Big)\mathrm{d}t \\ &+ \varepsilon\frac{\varsigma^{-1}(P_t^\circ)}{2} \sum_{j\in[\![d-m]\!]} \widetilde{P}_t^j\mathrm{d}\overline{W}_t^{i,j},\end{aligned} \quad (5.22)$$

for $t \in [t_0, T]$, which prompts us to let

$$\widetilde{B}_i(t, r^\circ, r) := \frac{\varphi(\varsigma^2(r^\circ)r_i) + \widetilde{b}_i(t, r^\circ, r) + r_i\widetilde{b}_i^\circ(t, r^\circ, r)}{2\sqrt{r_i}} - \frac{\varepsilon^2}{8\sqrt{r_i}} + \frac{\varepsilon^2}{8}\sqrt{r_i},$$

for $r^\circ \in \hat{\mathcal{S}}_m$ and $r \in \mathcal{S}_{d-m-1}$. Denoting by B° and Σ° the drift and diffusion coefficients in the dynamics of P° in equation (5.6), we then look at the solutions of the coupled SDEs:

$$\begin{aligned}
\mathrm{d}\widetilde{P}_t^i &= \varsigma^{-2}(P_t^\circ)\widetilde{B}_i(t, P_t^\circ, P_t)\mathrm{d}t + \varepsilon\frac{\varsigma^{-1}(P_t^\circ)}{2}\sum_{j\in[\![d-m]\!]}\widetilde{P}_t^j\mathrm{d}\overline{W}_t^{i,j} \\
\mathrm{d}P_t^\circ &= B^\circ(t, P_t^\circ, P_t)\mathrm{d}t + \Sigma^\circ(t, P_t^\circ, P_t)\mathrm{d}W_t^\circ \\
\mathrm{d}\widetilde{Q}_t^i &= \varsigma^{-2}(Q_t^\circ)\widetilde{B}_i(t, Q_t^\circ, Q_t)\mathrm{d}t + \varepsilon\frac{\varsigma^{-1}(Q_t^\circ)}{2}\sum_{j\in[\![d-m]\!]}\widetilde{Q}_t^j(R_t\mathrm{d}\overline{W}_tR_t)^{i,j}, \\
\mathrm{d}Q_t^\circ &= B^\circ(t, Q_t^\circ, Q_t)\mathrm{d}t + \Sigma^\circ(t, Q_t^\circ, Q_t)\mathrm{d}W_t^\circ,
\end{aligned} \tag{5.23}$$

where $Q_t^i := (\widetilde{Q}_t^i)^2$, for $t \in [t_0, T]$ and $i \in [\![d-m]\!]$, and where R_t denotes the reflection matrix:

$$R_t := I_{d-m} - 2\frac{(\widetilde{P}_t - \widetilde{Q}_t)(\widetilde{P}_t - \widetilde{Q}_t)^\dagger}{|\widetilde{P}_t - \widetilde{Q}_t|^2}\mathbf{1}_{\{t<\tau\}}, \quad t \in [t_0, T], \tag{5.24}$$

with $\tau := \inf\{t \geq t_0 : \widetilde{P}_t = \widetilde{Q}_t\}$, I_{d-m} standing for the identity matrix of dimension $d-m$. The initial conditions are $\widetilde{P}_{t_0} = \widetilde{p} := \sqrt{p}$, $P_{t_0}^\circ = p^\circ$, $\widetilde{Q}_{t_0} = \widetilde{q} := \sqrt{q}$ and $Q_{t_0}^\circ = q^\circ$, for the same p, p°, q and q° as in the statement of Proposition 5.3.

We claim that (5.23) is uniquely solvable in the strong sense, see Lemma 5.9. Importantly, we prove in Lemma 5.8 below that $(\int_{t_0}^t R_s\mathrm{d}\overline{W}_sR_s)_{t\geq t_0}$ is an antisymmetric Brownian motion of dimension $(d-m)(d-m-1)/2$. Since (5.6) is uniquely solvable, this proves that the law of $(Q_t^\circ, \widetilde{Q}_t)_{t_0\leq t\leq T}$ coincides with the law of the solution to the first two equations in (5.23) when the latter are initiated from $(q^\circ, \sqrt{q})$ at time t_0. In other words, the law of $(Q_t^\circ, Q_t)_{t_0\leq t\leq T}$ coincides with the law of the solution to (5.6) when the latter is initiated from (q°, q) at time t_0. We then define the stopping times ρ, ϱ and σ as in the statement of Proposition 5.3.

Second Step. We now have a look at $(Z_t = \widetilde{P}_t - \widetilde{Q}_t)_{t_0\leq t\leq T}$. Using the fact that $R_t\widetilde{Q}_t = \widetilde{P}_t$, for $t \in [t_0, \tau \wedge T)$, we get

$$\mathrm{d}\overline{W}_t\widetilde{P}_t - R_t\mathrm{d}\overline{W}_tR_t\widetilde{Q}_t = \mathrm{d}\overline{W}_t\widetilde{P}_t - R_t\mathrm{d}\overline{W}_t\widetilde{P}_t = 2\frac{Z_tZ_t^\dagger}{|Z_t|^2}\mathrm{d}\overline{W}_t\widetilde{P}_t.$$

We deduce the following expression

$$\frac{Z_t^\dagger}{|Z_t|}(\mathrm{d}\overline{W}_t\widetilde{P}_t - R_t\mathrm{d}\overline{W}_tR_t\widetilde{Q}_t) = 2\frac{Z_t^\dagger}{|Z_t|}\mathrm{d}\overline{W}_t\widetilde{P}_t.$$

Now, we may compute the bracket of the above right-hand side. We get

$$\begin{aligned}
\frac{1}{\mathrm{d}t}\Big\langle\frac{Z_t^\dagger}{|Z_t|}\mathrm{d}\overline{W}_t\widetilde{P}_t\Big\rangle &= \sum_{i,j\in[\![d-m]\!]}\sum_{i',j'\in[\![d-m]\!]}\frac{Z_t^iZ_t^{i'}}{|Z_t|^2}(\delta_{i,i'}\delta_{j,j'} - \delta_{i,j'}\delta_{j,i'})\widetilde{P}_t^j\widetilde{P}_t^{j'} \\
&= 1 - \langle\frac{Z_t}{|Z_t|}, \widetilde{P}_t\rangle^2 = 1 - \frac{(1 - \langle\widetilde{Q}_t, \widetilde{P}_t\rangle)^2}{2 - 2\langle\widetilde{Q}_t, \widetilde{P}_t\rangle} = 1 - \frac{1}{2}(1 - \langle\widetilde{Q}_t, \widetilde{P}_t\rangle) = 1 - \frac{1}{4}|Z_t|^2,
\end{aligned} \tag{5.25}$$

which holds true for $t < T \wedge \tau$. More generally, we need to compute the bracket of $(\int_{t_0}^{t\wedge\tau} \mathrm{d}\overline{W}_s\widetilde{P}_s - R_s d\overline{W}_s R_s\widetilde{Q}_s)_{t_0\le t\le T}$. For $i, j \in [\![d-m]\!]$ and $t \in [t_0, \tau)$, we have

$$\langle(\mathrm{d}\overline{W}_t\widetilde{P}_t - R_t\mathrm{d}\overline{W}_t R_t\widetilde{Q}_t)^i, (\mathrm{d}\overline{W}_t\widetilde{P}_t - R_t\mathrm{d}\overline{W}_t R_t\widetilde{Q}_t)^j\rangle = 4\Big\langle\Big(\frac{Z_tZ_t^\dagger}{|Z_t|^2}\mathrm{d}\overline{W}_t\widetilde{P}_t\Big)^i, \Big(\frac{Z_tZ_t^\dagger}{|Z_t|^2}\mathrm{d}\overline{W}_t\widetilde{P}_t\Big)^j\Big\rangle$$

Here,

$$\Big\langle\Big(\frac{Z_tZ_t^\dagger}{|Z_t|^2}\mathrm{d}\overline{W}_t\widetilde{P}_t\Big)^i, \Big(\frac{Z_tZ_t^\dagger}{|Z_t|^2}\mathrm{d}\overline{W}_t\widetilde{P}_t\Big)^j\Big\rangle = \sum_{k,l\in[\![d-m]\!]}\frac{Z_t^iZ_t^jZ_t^kZ_t^l}{|Z_t|^4}\langle(\mathrm{d}\overline{W}_t\widetilde{P}_t)^k, (\mathrm{d}\overline{W}_t\widetilde{P}_t)^l\rangle.$$

Now,

$$\begin{aligned}\frac{1}{\mathrm{d}t}\langle(\mathrm{d}\overline{W}_t\widetilde{P}_t)^i, (\mathrm{d}\overline{W}_t\widetilde{P}_t)^j\rangle &= \sum_{k,l\in[\![d-m]\!]}\frac{1}{\mathrm{d}t}\langle\mathrm{d}\overline{W}_t^{i,k}\widetilde{P}_t^k, \mathrm{d}\overline{W}_t^{j,l}\widetilde{P}_t^l\rangle\\ &= \sum_{k,l\in[\![d-m]\!]}\widetilde{P}_t^k\widetilde{P}_t^l(\delta_{i,j}\delta_{k,l} - \delta_{i,l}\delta_{j,k}) = \delta_{i,j} - \widetilde{P}_t^i\widetilde{P}_t^j.\end{aligned}$$

So,

$$\begin{aligned}\frac{1}{\mathrm{d}t}\Big\langle\Big(\frac{Z_tZ_t^\dagger}{|Z_t|^2}\mathrm{d}\overline{W}_t\widetilde{P}_t\Big)^i, \Big(\frac{Z_tZ_t^\dagger}{|Z_t|^2}\mathrm{d}\overline{W}_t\widetilde{P}_t\Big)^j\Big\rangle &= \sum_{k,l\in[\![d-m]\!]}\frac{Z_t^iZ_t^jZ_t^kZ_t^l}{|Z_t|^4}\Big(\delta_{k,l} - \widetilde{P}_t^k\widetilde{P}_t^l\Big)\\ &= \frac{Z_t^iZ_t^j}{|Z_t|^2} - \frac{Z_t^iZ_t^j}{|Z_t|^4}\langle Z_t, \widetilde{P}_t\rangle^2.\end{aligned}$$

And then,

$$\frac{1}{\mathrm{d}t}\langle(\mathrm{d}\overline{W}_t\widetilde{P}_t - R_t\mathrm{d}\overline{W}_t R_t\widetilde{Q}_t)^i, (\mathrm{d}\overline{W}_t\widetilde{P}_t - R_t\mathrm{d}\overline{W}_t R_t Q_t)^j\rangle = 4\Big(\frac{Z_t^iZ_t^j}{|Z_t|^2} - \frac{Z_t^iZ_t^j}{|Z_t|^4}\langle Z_t, \widetilde{P}_t\rangle^2\Big). \tag{5.26}$$

Third Step. Now, we return to the equation satisfied by $(Z_t)_{t_0\le t\le T}$:

$$\begin{aligned}\mathrm{d}Z_t = &[\varsigma^{-2}(P_t^\circ)\widetilde{B}(t, P_t^\circ, P_t) - \varsigma^{-2}(Q_t^\circ)\widetilde{B}(t, Q_t^\circ, Q_t)]\mathrm{d}t\\ &+ \varepsilon\frac{\varsigma^{-1}(P_t^\circ)}{2}\mathrm{d}\overline{W}_t\widetilde{P}_t - \varepsilon\frac{\varsigma^{-1}(Q_t^\circ)}{2}R_t\mathrm{d}\overline{W}_t R_t\widetilde{Q}_t,\end{aligned}$$

for $t \in [0, T]$. The point is to apply Itô's formula to $|Z|$. Using the relationship $R_t\widetilde{Q}_t = \widetilde{P}_t$ together with the fact that $Z_t^\dagger R_t = (R_tZ_t)^\dagger = -Z_t^\dagger$, we get, for $t < \varpi \wedge T$,

$$\begin{aligned}\mathrm{d}|Z_t| = &\frac{Z_t^\dagger}{|Z_t|}[\varsigma^{-2}(P_t^\circ)\widetilde{B}(t, P_t^\circ, P_t) - \varsigma^{-2}(Q_t^\circ)\widetilde{B}(t, Q_t^\circ, Q_t)]\mathrm{d}t + \varepsilon\frac{\varsigma^{-1}(P_t^\circ) + \varsigma^{-1}(Q_t^\circ)}{2}\frac{Z_t^\dagger}{|Z_t|}\mathrm{d}\overline{W}_t\widetilde{P}_t\\ &+ \frac{\varepsilon^2}{8}\sum_{i,j\in[\![d-m]\!]}\Big[\Big\langle\Big(\varsigma^{-1}(P_t^\circ)\mathrm{d}\overline{W}_t\widetilde{P}_t - \varsigma^{-1}(Q_t^\circ)R_t\mathrm{d}\overline{W}_t R_t\widetilde{Q}_t\Big)^i,\\ &\qquad\Big(\varsigma^{-1}(P_t^\circ)\mathrm{d}\overline{W}_t\widetilde{P}_t - \varsigma^{-1}(Q_t^\circ)R_t\mathrm{d}\overline{W}_t R_t\widetilde{Q}_t\Big)^j\Big\rangle \times \frac{1}{|Z_t|}\Big(\delta_{i,j} - \frac{Z_t^iZ_t^j}{|Z_t|^2}\Big)\Big].\end{aligned}$$

At this stage of the proof, we have a special look at the brackets in the last two lines in the above expression. By Kunita–Watanabe inequality, we get, for $t < \varpi \wedge T$,

$$\begin{aligned}
&\Big\langle \Big(\varsigma^{-1}(P^\circ_t)\mathrm{d}\overline{W}_t\widetilde{P}_t - \varsigma^{-1}(Q^\circ_t)R_t\mathrm{d}\overline{W}_tR_t\widetilde{Q}_t\Big)^i, \Big(\varsigma^{-1}(P^\circ_t)\mathrm{d}\overline{W}_t\widetilde{P}_t - \varsigma^{-1}(Q^\circ_t)R_t\mathrm{d}\overline{W}_tR_t\widetilde{Q}_t\Big)^j\Big\rangle \\
&= \varsigma^{-2}(P^\circ_t)\Big\langle \Big(\mathrm{d}\overline{W}_t\widetilde{P}_t - R_t\mathrm{d}\overline{W}_tR_t\widetilde{Q}_t\Big)^i, \Big(\mathrm{d}\overline{W}_t\widetilde{P}_t - R_t\mathrm{d}\overline{W}_tR_t\widetilde{Q}_t\Big)^j\Big\rangle + O\Big(\big|\varsigma^{-1}(P^\circ_t) - \varsigma^{-1}(Q^\circ_t)\big|\Big)\mathrm{d}t,
\end{aligned}$$

where $(O(|\varsigma^{-1}(P^\circ_t)-\varsigma^{-1}(Q^\circ_t)|))_{0\le t<\varpi\wedge T}$ stands for a progressively measurable process that is dominated by $C(|\varsigma^{-1}(P^\circ_t) - \varsigma^{-1}(Q^\circ_t)|)_{0\le t<\varpi\wedge T}$ for a universal constant C. Invoking (5.26), we get

$$\begin{aligned}
&\Big\langle \Big(\varsigma^{-1}(P^\circ_t)\mathrm{d}\overline{W}_t\widetilde{P}_t - \varsigma^{-1}(Q^\circ_t)R_t\mathrm{d}\overline{W}_tR_t\widetilde{Q}_t\Big)^i, \Big(\varsigma^{-1}(P^\circ_t)\mathrm{d}\overline{W}_t\widetilde{P}_t - \varsigma^{-1}(Q^\circ_t)R_t\mathrm{d}\overline{W}_tR_t\widetilde{Q}_t\Big)^j\Big\rangle \\
&= 4\varsigma^{-2}(P^\circ_t)\Big(\frac{Z^i_tZ^j_t}{|Z_t|^2} - \frac{Z^i_tZ^j_t}{|Z_t|^4}\langle Z_t, \widetilde{P}_t\rangle^2\Big)\mathrm{d}t + O\Big(\big|P^\circ_t - Q^\circ_t\big|\Big)\mathrm{d}t \\
&= 4\varsigma^{-2}(P^\circ_t)\Big(\frac{Z^i_tZ^j_t}{|Z_t|^2} - \frac{Z^i_tZ^j_t}{|Z_t|^4}\langle Z_t, \widetilde{P}_t\rangle^2\Big)\mathrm{d}t + O(|Z_t|)\mathrm{d}t,
\end{aligned}$$

where we used the fact that $t < \varrho$ to derive the last line of the statement. As before, the process $(O(|Z_t|))_{0\le t<\varpi\wedge T}$ is a progressively measurable process that is dominated by $(C|Z_t|)_{0\le t<\varpi\wedge T}$ for a universal constant C. Returning to the expression for $d|Z_t|$, we then get that

$$\begin{aligned}
&\frac{1}{8}\sum_{i,j\in[\![d-m]\!]}\Big[\Big\langle \Big(\varsigma^{-1}(P^\circ_t)\mathrm{d}\overline{W}_t\widetilde{P}_t - \varsigma^{-1}(Q^\circ_t)R_t\mathrm{d}\overline{W}_tR_t\widetilde{Q}_t\Big)^i, \\
&\qquad\qquad \Big(\varsigma^{-1}(P^\circ_t)\mathrm{d}\overline{W}_t\widetilde{P}_t - \varsigma^{-1}(Q^\circ_t)R_t\mathrm{d}\overline{W}_tR_t\widetilde{Q}_t\Big)^j\Big\rangle \times \frac{1}{|Z_t|}\Big(\delta_{i,j} - \frac{Z^i_tZ^j_t}{|Z_t|^2}\Big)\Big] \qquad (5.27) \\
&= \frac{1}{2|Z_t|}\varsigma^{-2}(P^\circ_t)\sum_{i,j\in[\![d-m]\!]}\Big(\frac{Z^i_tZ^j_t}{|Z_t|^2} - \frac{Z^i_tZ^j_t}{|Z_t|^4}\langle Z_t, \widetilde{P}_t\rangle^2\Big)\Big(\delta_{i,j} - \frac{Z^i_tZ^j_t}{|Z_t|^2}\Big)\mathrm{d}t + O(1)\mathrm{d}t,
\end{aligned}$$

where $(O(1))_{0\le t<\varpi\wedge T}$ stands for a progressively measurable process that is dominated by C for a universal constant C. The key fact here is that

$$\sum_{i,j\in[\![d-m]\!]}\frac{Z^i_tZ^j_t}{|Z_t|^2}\Big(\delta_{i,j} - \frac{Z^i_tZ^j_t}{|Z_t|^2}\Big) = \sum_{i\in[\![d-m]\!]}\frac{(Z^i_t)^2}{|Z_t|^2} - \Big(\sum_{i\in[\![d-m]\!]}\frac{(Z^i_t)^2}{|Z_t|^2}\Big)^2 = 1 - 1 = 0.$$

We deduce that the last line in (5.27) reduces to $O(1)\mathrm{d}t$. We end up with

$$\begin{aligned}
\mathrm{d}|Z_t| = {}& \frac{Z^\dagger_t}{|Z_t|}[\varsigma^{-2}(P^\circ_t)\widetilde{B}(t, P^\circ_t, P_t) - \varsigma^{-2}(Q^\circ_t)\widetilde{B}(t, Q^\circ_t, Q_t)]\mathrm{d}t + \varepsilon\frac{\varsigma^{-1}(P^\circ_t) + \varsigma^{-1}(Q^\circ_t)}{2}\frac{Z^\dagger_t}{|Z_t|}\mathrm{d}\overline{W}_t\widetilde{P}_t \\
&+ O(1)\,\mathrm{d}t, \quad t < \varpi\wedge T.
\end{aligned}$$

Fourth Step. We now have a look at the drift in a more precise way. Recalling that

$$\widetilde{B}_i(t, P^\circ_t, P_t) = \frac{\varphi(\varsigma^2(P^\circ_t)P^i_t) + \widetilde{b}_i(t, P^\circ_t, P_t) + (\widetilde{P}^i_t)^2\widetilde{b}^\circ_i(t, P^\circ_t, P_t)}{2\widetilde{P}^i_t} - \frac{\varepsilon^2}{8\widetilde{P}^i_t} + \frac{\varepsilon^2}{8}\widetilde{P}^i_t,$$

we write

$$\begin{aligned}\mathrm{d}|Z_t| = \sum_{i\in[\![d-m]\!]} \frac{Z_t^i}{|Z_t|}\Big[\varsigma^{-2}(P_t^\circ)\frac{\varphi(\varsigma^2(P_t^\circ)P_t^i)+\widetilde{b}_i(t,P_t^\circ,P_t)-\varepsilon^2/4}{2\widetilde{P}_t^i} \\ - \varsigma^{-2}(Q_t^\circ)\frac{\varphi(\varsigma^2(Q_t^\circ)Q_t^i)+\widetilde{b}_i(t,Q_t^\circ,Q_t)-\varepsilon^2/4}{2\widetilde{Q}_t^i}\Big]\mathrm{d}t + O(1)\,\mathrm{d}t \\ + \varepsilon\frac{\varsigma^{-1}(P_t^\circ)+\varsigma^{-1}(Q_t^\circ)}{2}\frac{Z_t^\dagger}{|Z_t|}\mathrm{d}\overline{W}_t\widetilde{P}_t,\end{aligned}$$

where the constant dominating $O(1)$ is now allowed to depend on $(\|\widetilde{b}_i^\circ\|_\infty)_{i\in[\![d-m]\!]}$. Fix now an index $i \in [\![d-m]\!]$. Then, on the event $P_t^i \le Q_t^i$, we have, for $t < \varpi \wedge T$,

$$\frac{Z_t^i}{|Z_t|}\big|\varsigma^{-2}(P_t^\circ)-\varsigma^{-2}(Q_t^\circ)\big| \le C\frac{|Z_t^i|}{|Z_t|}|P_t^\circ - Q_t^\circ| \le C|Z_t^i| \le C(\widetilde{P}_t^i+\widetilde{Q}_t^i) \le 2C\widetilde{Q}_t^i,$$

for a universal constant C. Proceeding similarly whenever $Q_t^i \le P_t^i$, we can find a collection of non-negative bounded processes $((\zeta_t^i)_{0\le t\le T})_{i\in[\![d-m]\!]}$ (bounded by 4 since $\varsigma^{-2}(P_t^\circ)$ and $\varsigma^{-2}(Q_t^\circ)$ are bounded by 4 for $t < \varpi \wedge T$) such that

$$\begin{aligned}\mathrm{d}|Z_t| = \sum_{i=1}^{d-m}\frac{Z_t^i}{|Z_t|}\zeta_t^i\Big[\frac{\varphi(\varsigma^2(P_t^\circ)P_t^i)+\widetilde{b}_i(t,P_t^\circ,P_t)-\varepsilon^2/4}{2\widetilde{P}_t^i} - \frac{\varphi(\varsigma^2(Q_t^\circ)Q_t^i)+\widetilde{b}_i(t,Q_t^\circ,Q_t)-\varepsilon^2/4}{2\widetilde{Q}_t^i}\Big]\mathrm{d}t \\ + O(1)\,\mathrm{d}t + \varepsilon\frac{\varsigma^{-1}(P_t^\circ)+\varsigma^{-1}(Q_t^\circ)}{2}\frac{Z_t^\dagger}{|Z_t|}\mathrm{d}\overline{W}_t\widetilde{P}_t, \quad t < \varpi\wedge T.\end{aligned}$$

Notice that, on the event $\{\max(P_t^i, Q_t^i) > \delta\}$, we have $\min(P_t^i, Q_t^i) > \delta/2$, for $t < \varpi \wedge T$, since $|P_t^i - Q_t^i| \le 2|\widetilde{P}_t^i - \widetilde{Q}_t^i| \le \delta/2$. Allowing $O(1)$ to depend on δ, κ $(\|\widetilde{b}_i\|_\infty)_{i\in[\![d-m]\!]}$ and $(\|\widetilde{b}_i^\circ\|_\infty)_{i\in[\![d-m]\!]}$, we get

$$\begin{aligned}\mathrm{d}|Z_t| \le \sum_{i=1}^{d-m}\frac{Z_t^i}{|Z_t|}\zeta_t^i\Big[\frac{\varphi(\varsigma^2(P_t^\circ)P_t^i)+\widetilde{b}_i(t,P_t^\circ,P_t)-\varepsilon^2/4}{2\widetilde{P}_t^i} \\ - \frac{\varphi(\varsigma^2(Q_t^\circ)Q_t^i)+\widetilde{b}_i(t,Q_t^\circ,Q_t)-\varepsilon^2/4}{2\widetilde{Q}_t^i}\Big]\mathbf{1}_{\{\max(P_t^i,Q_t^i)\le\delta\}}\mathrm{d}t \\ + O(1)\,\mathrm{d}t + \varepsilon\frac{\varsigma^{-1}(P_t^\circ)+\varsigma^{-1}(Q_t^\circ)}{2}\frac{Z_t^\dagger}{|Z_t|}\mathrm{d}\overline{W}_t\widetilde{P}_t, \quad t < \varpi\wedge T.\end{aligned}$$

Now, it remains to see that

$$\begin{aligned}
&\frac{\varphi(\varsigma^2(P_t^\circ)P_t^i) + \widetilde{b}_i(t, P_t^\circ, P_t) - \varepsilon^2/4}{2\widetilde{P}_t^i} - \frac{\varphi(\varsigma^2(Q_t^\circ)Q_t^i) + \widetilde{b}_i(t, Q_t^\circ, Q_t) - \varepsilon^2/4}{2\widetilde{Q}_t^i} \\
&= -\frac{\varphi(\varsigma^2(Q_t^\circ)Q_t^i) + \widetilde{b}_i(t, Q_t^\circ, Q_t) - \varepsilon^2/4}{2}\Big(\frac{1}{\widetilde{Q}_t^i} - \frac{1}{\widetilde{P}_t^i}\Big) \\
&\quad + \frac{\varphi(\varsigma^2(P_t^\circ)P_t^i) + \widetilde{b}_i(t, P_t^\circ, \widetilde{P}_t) - [\varphi(\varsigma^2(Q_t^\circ)Q_t^i) + \widetilde{b}_i(t, Q_t^\circ, Q_t)]}{2\widetilde{P}_t^i} \\
&= -\frac{\varphi(\varsigma^2(Q_t^\circ)Q_t^i) + \widetilde{b}_i(t, Q_t^\circ, Q_t) - \varepsilon^2/4}{2}\frac{Z_t^i}{\widetilde{P}_t^i\widetilde{Q}_t^i} \\
&\quad + \frac{\varphi(\varsigma^2(P_t^\circ)P_t^i) + \widetilde{b}_i(t, P_t^\circ, P_t) - [\varphi(\varsigma^2(Q_t^\circ)Q_t^i) + \widetilde{b}_i(t, Q_t^\circ, Q_t)]}{2\widetilde{P}_t^i}.
\end{aligned}$$

Recalling that $\widetilde{b}_i$, as given by Proposition 5.2 for each $i \in \{1, \cdots, d-m\}$, has non-negative values, that $\kappa \geq \kappa_0 \geq 2$, see Proposition 5.3, and that $\varepsilon \in (0,1)$, we deduce that, on the event $\{\max(P_t^i, Q_t^i) \leq \delta\}$, $\varphi(\varsigma^2(Q_t^\circ)Q_t^i) + \widetilde{b}_i(t, Q_t^\circ, \widetilde{Q}_t) - \varepsilon^2/4 \geq 0$. Therefore, whenever $P_t^i \geq Q_t^i$ and $P_t^i \leq \delta$,

$$\begin{aligned}
&\frac{\varphi(\varsigma^2(P_t^\circ)P_t^i) + \widetilde{b}_i(t, P_t^\circ, P_t) - \varepsilon^2/4}{2\widetilde{P}_t^i} - \frac{\varphi(\varsigma^2(Q_t^\circ)Q_t^i) + \widetilde{b}_i(t, Q_t^\circ, Q_t) - \varepsilon^2/4}{2\widetilde{Q}_t^i} \\
&\leq \frac{\varphi(\varsigma^2(P_t^\circ)P_t^i) + \widetilde{b}_i(t, P_t^\circ, P_t) - [\varphi(\varsigma^2(Q_t^\circ)Q_t^i) + \widetilde{b}_i(t, Q_t^\circ, Q_t)]}{2\widetilde{P}_t^i} \\
&= \frac{\kappa + b_i(t, P_t^\circ, \widetilde{P}_t) - [\kappa + \widetilde{b}_i(t, Q_t^\circ, Q_t)]}{2\widetilde{P}_t^i} \leq \frac{\|\widetilde{b}_i\|_\infty}{\widetilde{P}_t^i} = \frac{\|\widetilde{b}_i\|_\infty}{\max(\widetilde{P}_t^i, \widetilde{Q}_t^i)}.
\end{aligned}$$

Proceeding similarly when $Q_t^i \geq P_t^i$ and $Q_t^i \leq \delta$ and then letting

$$\widetilde{\beta}_t^i := \zeta_t^i \|\widetilde{b}_i\|_\infty \mathbf{1}_{\{\max(P_t^i, Q_t^i) \leq \delta\}},$$

we get

$$\mathrm{d}|Z_t| \leq \sum_{i=1}^{d-m} \frac{Z_t^i}{|Z_t|} \frac{\widetilde{\beta}_t^i}{\max(\widetilde{P}_t^i, \widetilde{Q}_t^i)} \mathrm{d}t + O(1)\,\mathrm{d}t + \varepsilon \frac{\varsigma^{-1}(P_t^\circ) + \varsigma^{-1}(Q_t^\circ)}{2} \frac{Z_t^\dagger}{|Z_t|} \mathrm{d}\overline{W}_t \widetilde{P}_t, \quad t < \varpi \wedge T,$$

which completes the proof. □

Lemma 5.8. *Take $(\widetilde{P}_t)_{0\leq t\leq T}$ and $(\widetilde{Q}_t)_{0\leq t\leq T}$ two continuous $\mathbb{F}^{\mathbf{W}^\circ, \mathbf{W}}$- adapted processes with values in the intersection of the orthant $(\mathbb{R}_+)^d$ and of the sphere of dimension d. Then, letting*

$$R_t := I_d - 2\frac{(\widetilde{P}_t - \widetilde{Q}_t)(\widetilde{P}_t - \widetilde{Q}_t)^\dagger}{|\widetilde{P}_t - \widetilde{Q}_t|^2} \mathbf{1}_{t<\tau}, \quad t \in [0, T],$$

with $\tau = \inf\{t \geq t_0 : \widetilde{P}_t = \widetilde{Q}_t\}$, I_d standing for the identity matrix of dimension d, the process $(\int_0^t R_s \mathrm{d}\overline{W}_s R_s)_{0\leq t\leq T}$, with the convention $\overline{W}_t^{i,i} := 0$ for $t \in [0,T]$ and $i \in [\![d]\!]$, is an antisymmetric Brownian motion of dimension $d(d-1)/2$ independent of $\mathbf{W}^\circ$.

Proof. We first extend the family $\boldsymbol{W}$ into a new family $\widetilde{\boldsymbol{W}}$ of independent Brownian motions, by letting $\widetilde{\boldsymbol{W}}^{i,j} = \boldsymbol{W}^{i,j}$ for $i, j \in [\![d]\!]$ with $i \neq j$ and by assuming that the family $(\widetilde{\boldsymbol{W}}^{i,i})_{i\in[\![d]\!]}$ is a collection of Brownian motions that is independent of $\boldsymbol{W}$. We then observe that, for any $i, j \in [\![d]\!]$ with $i \neq j$,

$$\Big(\int_0^t R_s \mathrm{d}\overline{W}_s R_s\Big)_{i,j} = \frac{1}{\sqrt{2}}\Big[\Big(\int_0^t R_s \mathrm{d}\widetilde{W}_s R_s\Big)_{i,j} - \Big(\int_0^t R_s \mathrm{d}\widetilde{W}_s R_s\Big)_{j,i}\Big], \quad t \in [0,T].$$

In order to complete the proof, it suffices to show that the family $((\int_0^t R_s \mathrm{d}\overline{W}_s R_s)_{0\le t\le T})_{i,j\in[\![d]\!]:i\neq j}$ forms a collection of independent Brownian motions that is independent of $\boldsymbol{W}^\circ$. Independence between $((\int_0^t R_s \mathrm{d}\widetilde{W}_s R_s)_{0\le t\le T})_{i,j\in[\![d]\!]}$ and $\boldsymbol{W}^\circ$ is obvious. It thus remains to compute the brackets of the family $((\int_0^t R_s \mathrm{d}\widetilde{W}_s R_s)_{0\le t\le T})_{i,j\in[\![d]\!]}$ to conclude. Using the fact that $R_s^\dagger = R_s$ and $R_s R_s = I_d$, we have

$$\begin{aligned}\sum_{k,l\in[\![d]\!]} R_s^{i,k} d\widetilde{W}_s^{k,l} R_s^{l,j} \cdot \sum_{k',l'\in[\![d]\!]} R_s^{i',k'} d\widetilde{W}_s^{k',l'} R_s^{l',j'} &= \sum_{k,l\in[\![d]\!]}\sum_{k',l'\in[\![d]\!]} \Big(R_s^{i,k} R_s^{i',k'} R_s^{l,j} R_s^{l',j'} \delta_{k,k'}\delta_{l,l'}\Big)\\ &= \sum_{k,l\in[\![d]\!]} \Big(R_s^{i,k} R_s^{i',k} R_s^{l,j} R_s^{l,j'}\Big) = \delta_{i,i'}\delta_{j,j'},\end{aligned}$$

which completes the proof. □

Lemma 5.9. *Under the assumption and notations of Propositions 5.3 and 5.7, Equation* (5.23) *is uniquely solvable (in the strong sense).*

Proof. We first observe that, at any time $t \in [0,T]$, the coefficients $\tilde{B}(t,\cdot)$ and $B^\circ(t,\cdot)$ are a priori defined as functions of the space variable $(r^\circ, r) \in \hat{\mathcal{S}}_m \times \mathcal{S}_{d-m-1}$. By projecting $\mathbb{R}^m$ onto $\hat{\mathcal{S}}_m$ (which is convex) and then $\mathbb{R}^{d-m}$ onto $\mathcal{S}_{d-m-1}$ (which is also convex), we may easily extend them to the entire $\mathbb{R}^d$. We then observe from Proposition 5.2 that the full-fledged drift coefficient in the system (5.23) remains Lipschitz continuous in the four entries $(P_t^\circ, P_t, Q_t^\circ, Q_t)$ as long the coordinates of the latter remain away from zero. Similarly, the diffusion coefficient remains Lipschitz continuous in the same four entries as long as the coordinates of the latter remain strictly positive and the distance between P_t and Q_t remains also strictly positive.

Therefore, we deduce that, for any small $a > 0$, the system (5.23) is uniquely solvable up to the first time τ^a when one of the coordinates of the vector $(P^\circ_{\tau^a}, P_{\tau^a}, Q^\circ_{\tau^a}, Q_{\tau^a})$ is less than a or the distance between P_{τ^a} and Q_{τ^a} becomes less than a. Letting a tend to 0, we deduce that (5.23) is uniquely solvable up to $\tau = \lim_{a\searrow 0} \tau^a \wedge T$.

By Lemma 5.8, we know that, up to time τ, we may see $(P_t^\circ, P_t)_{t_0\le t<\tau}$ and $(Q_t^\circ, Q_t)_{t_0\le t<\tau}$ as solutions of an SDE of the same type as (5.6). Hence, by identity in law (5.5) and by Proposition 5.1 (or equivalently by Proposition 2.2, recalling that $\kappa_0 \ge 2$), we deduce that, both processes take values in $\hat{\mathcal{S}}_m \times \mathcal{S}_{d-m}$ and that

$$\mathbb{P}\Big(\inf_{i\in[\![m]\!]}\inf_{t_0\le t<\tau} P_t^{\circ,i} > 0,\ \inf_{i\in[\![m]\!]}\inf_{t_0\le t<\tau} Q_t^{\circ,i} > 0,\ \inf_{i\in[\![d-m]\!]}\inf_{t_0\le t<\tau} P_t^{i} > 0,\ \inf_{i\in[\![d-m]\!]}\inf_{t_0\le t<\tau} Q_t^{i} > 0\Big) = 1.$$

This shows in particular that the drift in (5.23) remains bounded up to time τ and that it makes sense to extend (by continuity) the process (P, P°, Q, Q°) to the closed interval $[0,\tau]$. Moreover, we must have

$$\mathbb{P}(\tau = \tau) = 1,$$

where we recall that τ denotes the first time when the two processes $(P_t)_{0\le t\le\tau}$ and $(Q_t)_{0\le t\le\tau}$ meet. This proves the unique strong solvability on $[0,\tau]$. Unique solvability from τ to T is addressed in a similar manner noting that the diffusion coefficient then becomes simpler, see (5.24). □

5.3.2. Proof of Proposition 5.3

We recall that $|p - q| < \delta^2/(64\sqrt{d})$.

First Step. The proof mostly relies on a Girsanov argument. Using the same notations as in the statement and in the proof of Proposition 5.7, we let (see (5.22) and (5.23))

$$W'^{,i,j}_t := W^{i,j}_t + \frac{1}{\varepsilon}\int_{t_0}^t \Psi^{i,j}_s \mathrm{d}s, \quad \Psi^{i,j}_t := \frac{2\sqrt{2}\widetilde{P}^j_t}{(\varsigma^{-1}(P^\circ_t) + \varsigma^{-1}(Q^\circ_t))} \frac{\widetilde{\beta}^i_t}{\max(\widetilde{P}^i_t, \widetilde{Q}^i_t)} \mathbf{1}_{\{t<\varpi\}}, \tag{5.28}$$

for $t \in [t_0, T]$ and $i, j \in [\![d - m]\!]$ with $i \neq j$, and $W'^{,i,i}_t = 0$ for $i \in [\![d - m]\!]$. Then, for all $t \in [t_0, \varpi \wedge T)$,

$$\varepsilon \frac{Z^\dagger_t}{|Z_t|}\mathrm{d}\overline{W}_t\widetilde{P}_t = \varepsilon \frac{Z^\dagger_t}{|Z_t|}\mathrm{d}\overline{W}'_t\widetilde{P}_t - \frac{Z^\dagger_t}{\sqrt{2}|Z_t|}(\Psi_t - \Psi^\dagger_t)\widetilde{P}_t\mathrm{d}t,$$

with

$$\begin{aligned}
&\frac{Z^\dagger_t}{\sqrt{2}|Z_t|}(\Psi_t - \Psi^\dagger_t)\widetilde{P}_t \\
&= \frac{2}{\varsigma^{-1}(P^\circ_t) + \varsigma^{-1}(Q^\circ_t)} \sum_{i,j\in[\![d-m]\!]} \frac{Z^i_t}{|Z_t|}\Big(\frac{\widetilde{\beta}^i_t\widetilde{P}^j_t}{\max(\widetilde{P}^i_t, \widetilde{Q}^i_t)} - \frac{\widetilde{\beta}^j_t\widetilde{P}^i_t}{\max(\widetilde{P}^j_t, \widetilde{Q}^j_t)}\Big)\widetilde{P}^j_t \\
&= \frac{2}{\varsigma^{-1}(P^\circ_t) + \varsigma^{-1}(Q^\circ_t)}\Big(\sum_{i\in[\![d-m]\!]} \frac{Z^i_t}{|Z_t|}\frac{\widetilde{\beta}^i_t}{\max(\widetilde{P}^i_t, \widetilde{Q}^i_t)} - \Big\langle\frac{Z_t}{|Z_t|}, \widetilde{P}_t\Big\rangle \sum_{j\in[\![d-m]\!]} \frac{\widetilde{\beta}^j_t\widetilde{P}^j_t}{\max(\widetilde{P}^j_t, \widetilde{Q}^j_t)}\Big),
\end{aligned}$$

where we used the identity $\sum_{j\in[\![d-m]\!]}(\widetilde{P}^j_t)^2 = 1$. Plugging the above identity into (5.21), we get

$$\mathrm{d}|Z_t| \le C\mathrm{d}t + \varepsilon\frac{\varsigma^{-1}(P^\circ_t) + \varsigma^{-1}(Q^\circ_t)}{2}\frac{Z^\dagger_t}{|Z_t|}\mathrm{d}\overline{W}'_t\widetilde{P}_t, \quad t \in [t_0, \varpi \wedge T), \tag{5.29}$$

where C is a constant only depending on δ, κ, $(\|b_i\|_\infty)_{i=1,\cdots,d}$, $(\|b^\circ_i\|_\infty)_{i=1,\cdots,d}$ and T.

Second Step. We now introduce the probability measure:

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \exp\Big(-\frac{1}{\varepsilon}\sum_{i,j\in[\![d-m]\!]:i\neq j}\int_0^{\varpi\wedge T}\Psi^{i,j}_t\mathrm{d}W^{i,j}_t - \frac{1}{2\varepsilon^2}\sum_{i,j\in[\![d-m]\!]:i\neq j}\int_0^{\varpi\wedge T}\big|\Psi^{i,j}_t\big|^2\mathrm{d}t\Big). \tag{5.30}$$

Under $\mathbb{Q}$, the processes $((W'^{,i,j}_t)_{t_0\le t\le T})_{i,j\in[\![d-m]\!]:i\neq j}$ are independent Brownian motions (the fact that we can apply Girsanov's theorem is fully justified in the third step of the proof). By (5.25), the bracket of the martingale part in (5.29) is given by (up to the leading multiplicative factor)

$$\frac{1}{\mathrm{d}t}\Big\langle\frac{Z^\dagger_t}{|Z_t|}\mathrm{d}\overline{W}'_t\widetilde{P}_t\Big\rangle = 1 - \frac{1}{4}|Z_t|^2.$$

In particular, there exists a Brownian motion $(B_t)_{t_0\le t\le T}$ under $\mathbb{Q}$ such that

$$d|Z_t| \le C\mathrm{d}t + \varepsilon\frac{\varsigma^{-1}(P^\circ_t) + \varsigma^{-1}(Q^\circ_t)}{2}\sqrt{1 - \frac{1}{4}|Z_t|^2}dB_t, \quad t \in [t_0, \varpi \wedge T). \tag{5.31}$$

Let now

$$\mathrm{d}\Theta_t = C\mathrm{d}t + \varepsilon\theta_t \mathrm{d}B_t, \quad \theta_t := \min\Big(\frac{\varsigma^{-1}(P_t^\circ) + \varsigma^{-1}(Q_t^\circ)}{2}, c'\Big)\sqrt{1 - \frac{1}{4}\min(|Z_t|^2, \frac{\delta^2}{4})},$$

with $|\Theta_{t_0}| = |Z_{t_0}| = |\widetilde{p} - \widetilde{q}|$, with $\widetilde{p} = (\sqrt{p_1}, \cdots, \sqrt{p_{d-m}})$ and similarly for $\widetilde{q}$, and where $c' = \min\{\varsigma^{-1}(p^\circ), |p^\circ|_1 \geq 3/4 + \delta\sqrt{d}/4\} = (1/4 - \delta\sqrt{d}/4)^{-1/2}$. (Note that, for $t \in [t_0, \varpi \wedge T)$, $|P_t^\circ|_1 \leq 3/4$ and $|Q_t^\circ|_1 \leq 3/4 + |P_t^\circ - Q_t^\circ|_1 \leq 3/4 + \sqrt{m}|P_t^\circ - Q_t^\circ| < 3/4 + \delta\sqrt{d}/4 < 3/4 + 1/16 = 13/16$.) Obviously, $|Z_t| \leq \Theta_t$ for all $t \in [t_0, \varpi \wedge T]$ (because, up to time $\varpi \wedge T$, θ_t coincides with the integrand in the stochastic integral appearing in the right-hand side of (5.31)). Since $(\theta_t)_{t_0 \leq t \leq T}$ stays in a (universal) deterministic compact subset of $(0, +\infty)$, we deduce from a new application of Girsanov's theorem that there exists a new probability measure $\mathbb{Q}'$ under which

$$\mathrm{d}\Theta_t = \varepsilon\theta_t \mathrm{d}B'_t, \quad t \in [t_0, T],$$

$(B'_t)_{t_0 \leq t \leq T}$ being a Brownian motion under $\mathbb{Q}'$. By expanding the Girsanov transformation, we can check that $\mathbb{E}^{\mathbb{Q}'}[(\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{Q}')^2] = \mathbb{E}^{\mathbb{Q}}[\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{Q}'] \leq \gamma^2$, that is $\mathbb{Q}(A) = \mathbb{E}^{\mathbb{Q}'}[(\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{Q}')\mathbf{1}_A] \leq \gamma\mathbb{Q}'(A)^{1/2}$ for any event $A \in \mathcal{F}_T^{\boldsymbol{W}^\circ, \boldsymbol{W}}$, for a constant γ that may depend on ε.

Clearly, $(B'_t)_{t_0 \leq t \leq T}$ can be extended into a Brownian motion (under $\mathbb{Q}'$) on the entire $[t_0, \infty)$ and, similarly, $(\theta_t)_{t_0 \leq t \leq T}$ can be also extended to the entire $[t_0, \infty)$ by letting $\theta_t = c'$ for $t > T$. The process $(\Theta_t)_{t_0 \leq t \leq T}$ can be extended accordingly to the entire $[t_0, \infty)$. Representing $(\Theta_t)_{t \geq t_0}$ in the form a time-changed Brownian motion, there exists a new Brownian motion $(\hat{B}_t)_{t \geq 0}$ under $\mathbb{Q}'$ (with respect to a time-changed filtration) such that

$$\Theta_t = |\widetilde{p} - \widetilde{q}| + \varepsilon\hat{B}_{I_t}, \quad I_t := \int_{t_0}^t \theta_s^2 \mathrm{d}s, \quad t \geq t_0.$$

Obviously, there exists a universal constant $\Gamma \geq 1$ such that, with probability 1 under $\mathbb{Q}'$,

$$\Gamma^{-1}(t - t_0) \leq I_t \leq \Gamma(t - t_0), \quad t \geq t_0.$$

We now call

$$\sigma(\Theta) := \inf\{s \geq t_0 : |\Theta_s| \geq \frac{\delta}{4}\}, \quad \tau(\Theta) := \inf\{s \geq t_0 : \Theta_s = 0\}.$$

Then, recalling that $|p - q| < \delta^2/(64\sqrt{d})$ and observing that

$$|\widetilde{p} - \widetilde{q}|^2 = \sum_{i \in [\![d-m]\!]} \big|\sqrt{p_i} - \sqrt{q_i}\big|^2 \leq \sum_{i \in [\![d-m]\!]} |p_i - q_i| \leq \sqrt{d}|p - q|. \tag{5.32}$$

we get $|\widetilde{p} - \widetilde{q}| < \delta/8$ and then, for $t \in [t_0, T]$,

$$\begin{aligned}
\mathbb{Q}'\Big(\tau(\Theta) < \sigma(\Theta), \tau(\Theta) \leq t\Big) &\geq \mathbb{Q}'\Big(\varepsilon \inf_{0 \leq s \leq I_t} \hat{B}_s \leq -|\widetilde{p} - \widetilde{q}|, \quad \varepsilon \sup_{0 \leq s \leq I_t} \hat{B}_s \leq \frac{\delta}{8}\Big) \\
&\geq 1 - \mathbb{Q}'\Big(\varepsilon \sup_{0 \leq s \leq I_t} \hat{B}_s \leq |\widetilde{p} - \widetilde{q}|\Big) - \mathbb{Q}'\Big(\varepsilon \sup_{0 \leq s \leq I_t} \hat{B}_s \geq \frac{\delta}{8}\Big) \\
&\geq 1 - \mathbb{Q}'\Big(\varepsilon \sup_{0 \leq s \leq (t-t_0)/\Gamma} \hat{B}_s \leq |\widetilde{p} - \widetilde{q}|\Big) - \mathbb{Q}'\Big(\varepsilon \sup_{0 \leq s \leq \Gamma(t-t_0)} \hat{B}_s \geq \frac{\delta}{8}\Big).
\end{aligned}$$

We deduce that

$$\mathbb{Q}'\Big(\tau(\Theta) < \sigma(\Theta), \tau(\Theta) \le t\Big) \ge 1 - C' \frac{|\widetilde{p} - \widetilde{q}|}{\varepsilon \sqrt{t - t_0}} - C' \exp(-\frac{1}{C'\varepsilon^2 (t - t_0)}),$$

for a new constant C' that is independent of ε. Therefore, by (5.32), up to a new value of C',

$$\mathbb{Q}'\Big(\tau(\Theta) < \sigma(\Theta), \tau(\Theta) \le t\Big) \ge 1 - C' \frac{\sqrt{|p - q|}}{\varepsilon \sqrt{t - t_0}} - C' \exp(-\frac{1}{C'\varepsilon^2 (t - t_0)}).$$

In particular, choosing $t - t_0 = |p - q|^{1/3}/2$, which is possible since $|p - q|^{1/3} \le T - t_0$, we deduce that (with $S := t_0 + |p - q|^{1/3}$)

$$\mathbb{Q}'\Big(\tau(\Theta) < \sigma(\Theta), \tau(\Theta) < S\Big) \ge 1 - \frac{C'}{\varepsilon} |p - q|^{1/3},$$

or, equivalently,

$$\mathbb{Q}'\Big(\Big\{\tau(\Theta) < \sigma(\Theta), \tau(\Theta) \le S\Big\}^{\complement}\Big) \le C|p - q|^{1/3},$$

where we recall that C is allowed to depend on ε. Then, returning to $\mathbb{Q}$, we get, for a new value of C,

$$\mathbb{Q}(\tau(\Theta) \ge \sigma(\Theta) \wedge S) = \mathbb{Q}\Big(\Big\{\tau(\Theta) < \sigma(\Theta), \tau(\Theta) \le S\Big\}^{\complement}\Big) \le C|p - q|^{1/6}.$$

We then notice from the inequality $|Z_t| \le \Theta_t$, for $t \in [t_0, \varpi \wedge T]$, that $t \le \varpi \wedge T$ implies $t \le \tau(\Theta)$ and that $\sigma \le \varpi \wedge T$ implies $\sigma(\Theta) \le \sigma$. Therefore, on the event $\{\varpi \ge S\}$, we have $S \le \tau(\Theta)$. Moreover, on the event $\{\tau \wedge \sigma \le \varpi \wedge T\}$,

$$\{\sigma(\Theta) \le \tau(\Theta)\} \supset \{\sigma \le \tau\} = \{\tau < \sigma\}^{\complement}.$$

Hence,

$$\begin{aligned}\mathbb{Q}(\varpi_S < \tau \wedge \varrho \wedge \rho) &\le \mathbb{Q}(\{S \le \varpi\} \cup \{\sigma \le \tau,\ \tau \wedge \sigma \le \varpi \wedge T\}) \\ &\le \mathbb{Q}(\{S \le \tau(\Theta)\} \cup \{\sigma(\Theta) \le \tau(\Theta)\}) = \mathbb{Q}(\tau(\Theta) \ge \sigma(\Theta) \wedge S) \le C|p - q|^{1/6}.\end{aligned}$$

Third Step. In order to complete the proof, it remains to prove that, for a new value of the constant C, for any event $A \in \mathcal{F}_T^{\boldsymbol{W}^\circ, \boldsymbol{W}}$, $\mathbb{P}(A) \le C\mathbb{Q}(A)^{1/2}$, provided that κ is chosen large enough. As for the comparison of $\mathbb{Q}$ and $\mathbb{Q}'$ in the previous step, it suffices to prove that $\mathbb{E}[\mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q}]^{1/2} \le C$ (since $\mathbb{P}(A) = \mathbb{E}^{\mathbb{Q}}[(\mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q})\mathbf{1}_A] \le \mathbb{E}[\mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q}]^{1/2}\mathbb{Q}(A)^{1/2}$). Here (compare with (5.30)),

$$\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}} = \exp\Big(\frac{1}{\varepsilon} \sum_{i,j \in [\![d-m]\!]: i \ne j} \int_{t_0}^{\varpi \wedge T} \Psi_t^{i,j} \mathrm{d}W_t^{i,j} + \frac{1}{2\varepsilon^2} \sum_{i,j \in [\![d-m]\!]: i \ne j} \int_{t_0}^{\varpi \wedge T} \big|\Psi_t^{i,j}\big|^2 \mathrm{d}t\Big).$$

Letting

$$\Big(M_t := \frac{1}{\varepsilon} \sum_{i,j \in [\![d-m]\!]: i \ne j} \int_{t_0}^{t} \Psi_s^{i,j} \mathrm{d}W_s^{i,j}\Big)_{t_0 \le t \le T},$$

this may be rewritten as

$$\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}} = \exp\Big(M_{\varpi \wedge T} + \frac{1}{2}\langle M \rangle_{\varpi \wedge T}\Big),$$

and then

$$\mathbb{E}\Big[\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}\Big] = \mathbb{E}\Big[\exp\Big(M_{\varpi\wedge T} + \frac{1}{2}\langle M\rangle_{\varpi\wedge T}\Big)\Big]$$
$$\leq \mathbb{E}\Big[\exp\Big(M_{\varpi\wedge T} - \langle M\rangle_{\varpi\wedge T} + (1+\frac{1}{2})\langle M\rangle_{\varpi\wedge T}\Big)\Big] \leq \mathbb{E}\Big[\exp\Big(3\langle M\rangle_{\varpi\wedge T}\Big)\Big]^{1/2},$$

where to get the last line, we used the fact that $\mathbb{E}[\exp(2M_{\varpi\wedge T} - 2\langle M\rangle_{\varpi\wedge T})] \leq 1$. Returning to the definition of Ψ in (5.28) and recalling that ς^{-1} is lower bounded by 1, the point is to prove that

$$\mathbb{E}\Big[\exp\Big(\frac{3}{\varepsilon^2}\sum_{i,j\in[\![d-m]\!]:i\neq j}\int_{t_0}^{\varpi\wedge T}\big|\Psi_t^{i,j}\big|^2\mathrm{d}t\Big)\Big] \leq \mathbb{E}\Big[\exp\Big(\frac{6}{\varepsilon^2}\sum_{i\in[\![d-m]\!]}\int_{t_0}^{\varpi\wedge T}\Big|\frac{\widetilde{\beta}_t^i}{\max(\widetilde{P}_t^i,\widetilde{Q}_t^i)}\Big|^2\mathrm{d}t\Big)\Big]$$

is finite, provided that κ is chosen large enough and then to find a tractable bound. The proof is similar to that of Proposition 2.3, but we feel better to expand it as it plays a key role in the determination of the constant κ. Recalling the bound for $(\beta_t^i)_{0\leq t\leq T}$ in the statement of Proposition 5.7, using the fact that $(\widetilde{P}_t^i)^2 = P_t^i$ and invoking Hölder's inequality, it suffices to upper bound

$$\sup_{i\in[\![d-m]\!]}\mathbb{E}\Big[\exp\Big(\frac{6d}{\varepsilon^2}\int_{t_0}^{\varpi\wedge T}\frac{(4\|\widetilde{b}_i\|_\infty)^2}{\max(P_t^i,Q_t^i)}\mathrm{d}t\Big)\Big]^{1/d}. \tag{5.33}$$

Here, we recall from Proposition 5.2 that the coefficients $(\widetilde{b}_j)_{j\in[\![d-m]\!]}$ are bounded by a constant that only depends on $(\|b_j\|_\infty)_{j\in[\![d]\!]}$. Moreover, we recall from (5.6) that

$$\mathrm{d}P_t^i = \varsigma^{-2}(P_t^\circ)\Big(\varphi(\varsigma^2(P_t^\circ)P_t^i) + \widetilde{b}_i(t,P_t^\circ,P_t) + P_t^i\widetilde{b}_i^\circ(t,P_t^\circ,P_t)\Big)\mathrm{d}t + \varepsilon\varsigma^{-1}(P_t^\circ)\sum_{j=1}^{d-m}\sqrt{P_t^iP_t^j}\mathrm{d}\overline{W}_t^{i,j},$$

for $i\in[\![d-m]\!]$ and $t\in[t_0,T]$. Using again the fact that $(1/\mathrm{d}t)\langle\sum_{j\in[\![d-m]\!]}\sqrt{P_t^iP_t^j}\mathrm{d}\overline{W}_t^{i,j}\rangle = P_t^i(1-P_t^i)$, we get, by Itô's formula,

$$\begin{aligned}\mathrm{d}[\ln P_t^i] = {}& \varsigma^{-2}(P_t^\circ)\Big(\frac{\varphi(\varsigma^2(P_t^\circ)P_t^i)+\widetilde{b}_i(t,P_t^\circ,P_t)}{P_t^i} + \widetilde{b}_i^\circ(t,P_t^\circ,P_t)\Big)\mathrm{d}t\\ &+\varepsilon\varsigma^{-1}(P_t^\circ)\frac{1}{\sqrt{P_t^i}}\sum_{j\in[\![d-m]\!]}\sqrt{P_t^j}\mathrm{d}\overline{W}_t^{i,j} - \frac{\varepsilon^2}{2}\varsigma^{-2}(P_t^\circ)\frac{1-P_t^i}{P_t^i}\mathrm{d}t,\quad t\in[t_0,T].\end{aligned} \tag{5.34}$$

Recalling that $\widetilde{b}_i$ takes non-negative values and denoting by $(O(1))_{0\leq t\leq T}$ a progressively measurable process that is dominated by a constant C that may depend on δ, κ, $(\|b_j\|_\infty)_{j\in[\![d]\!]}$, $(\|b_j^\circ\|_\infty)_{j\in[\![d]\!]}$ and T, recalling that $\varphi\equiv\kappa$ on $[0,\delta]$ and $\varsigma^{-2}(P_t^\circ)\leq 4$ for $t\leq\varpi$, and choosing κ as large as needed (in terms of the sole $(\|b_j\|_\infty)_{j\in[\![d]\!]}$), we get, for $t\leq\varpi\wedge T$,

$$\begin{aligned}&\varsigma^{-2}(P_t^\circ)\Big(\frac{\varphi(\varsigma^2(P_t^\circ)P_t^i)+\widetilde{b}_i(t,P_t^\circ,P_t)}{P_t^i} + \widetilde{b}_i^\circ(t,P_t^\circ,P_t) - \frac{\varepsilon^2}{2}\frac{1-P_t^i}{P_t^i}\Big)\\ &\geq \frac{\varsigma^{-2}(P_t^\circ)}{P_t^i}(\kappa-\tfrac{1}{2})\mathbf{1}_{P_t^i\leq\delta} - O(1) \geq \frac{\kappa}{2}\frac{\varsigma^{-2}(P_t^\circ)}{P_t^i} - O(1).\end{aligned}$$

Hence, integrating (5.34), multiplying by some $\eta > 0$ and then taking exponential,

$$(P^i_{\varpi\wedge T})^\eta \exp\Big(-\eta\varepsilon \int_{t_0}^{\varpi\wedge T} \varsigma^{-1}(P^\circ_t)\frac{1}{\sqrt{P^i_t}} \sum_{j\in[\![d-m]\!]} \sqrt{P^j_t}\mathrm{d}\overline{W}^{i,j}_t - \frac{\eta^2\varepsilon^2}{2}\int_{t_0}^{\varpi\wedge T} \varsigma^{-2}(P^\circ_t)\frac{1-P^i_t}{P^i_t}\mathrm{d}t\Big)$$
$$\geq (P^i_0)^\eta \exp\Big((\frac{\eta\kappa-\eta^2\varepsilon^2}{2})\int_{t_0}^{\varpi\wedge T}\frac{\varsigma^{-2}(P^\circ_t)}{P^i_t}\mathrm{d}t - C\Big),$$

where C is a constant as before. For any given $\eta \in (0,1)$, we can choose κ as large as needed (κ now depending on ε, η and $(\|b_j\|_\infty)_{j\in[\![d]\!]}$) such that

$$\frac{\eta\kappa-\eta^2\varepsilon^2}{2} \geq \frac{6d}{\varepsilon^2}(4\max_{j\in[\![d-m]\!]}\|\widetilde{b}_j\|_\infty)^2, \tag{5.35}$$

and then (compare with (5.33))

$$\mathbb{E}\Big[\exp\Big(\frac{6d}{\varepsilon^2}\int_{t_0}^{\varpi\wedge T}\varsigma^{-2}(P^\circ_t)\frac{(4\max_{j\in[\![d-m]\!]}\|\widetilde{b}_j\|_\infty)^2}{P^i_t}\mathrm{d}t\Big)\Big] \leq \frac{C}{p_i^\eta},$$

where C is independent of p_0 but depends on δ, ε, $(\|b_j\|_\infty)_{j\in[\![d]\!]}$ and $(\|b^\circ_j\|_\infty)_{j\in[\![d]\!]}$ and T. Since ς^{-1} is above 1,

$$\mathbb{E}\Big[\exp\Big(\frac{6d}{\varepsilon^2}\int_{t_0}^{\varpi\wedge T}\frac{(4\max_{j\in[\![d-m]\!]}\|\widetilde{b}_j\|_\infty)^2}{P^i_t}\mathrm{d}t\Big)\Big] \leq \frac{C}{p_i^\eta},$$

Similarly, we have the same inequality, but replacing P^i_t by Q^i_t in the left-hand side and p_i by q_i in the right-hand side. Hence, we can upper bound (5.33) by $C/\max(p_i,q_i)^\eta$.

Conclusion. We deduce from the conclusions of the second and third steps that

$$\mathbb{P}(\varpi_S < \tau\wedge\varrho\wedge\rho) \leq C\frac{|p-q|^{1/12}}{\min_{i\in[\![d-m]\!]}(\max(p_i,q_i))^{d\eta/2}},$$

where C depends on δ, ε, κ, η, $(\|b_i\|_\infty)_{i=1,\cdots,d}$, $(\|b^\circ_i\|_\infty)_{i=1,\cdots,d}$ and T. Since the value of η is arbitrary (provided that it belongs to $(0,1)$), we can easily apply the above inequality with $2\eta/d$ instead of η (observe that, whenever $\eta\in(0,1)$, $2\eta/d$ also belongs to $(0,1)$, since $d\geq 2$). □

5.4. Proof of Proposition 5.2

Proof. First Step. We introduce some useful notations. Having in mind the shape of the coefficients in equation (5.1), we let, for $i\in[\![d]\!]$ and for $p\in\mathcal{S}_{d-1}$,

$$\mathfrak{b}_i(t,p) := \varphi(p_i) + b_i(t,p) + p_i b^\circ_i(t,p).$$

Importantly, we recall from (4.13) that, for any $(t,p)\in[0,T]\times\mathcal{S}_{d-1}$, $\sum_{i\in[\![d]\!]}\mathfrak{b}_i(t,p) = 0$. In fact, we can easily extend $\mathfrak{b}_i$, for each $i\in[\![d]\!]$, to the entire $[0,T]\times\mathbb{R}^d$ by composing $\mathfrak{b}_i$ with the orthogonal projection from $\mathbb{R}^d$ into $\mathcal{S}_{d-1}$. This allows us to define the drift B° entering the dynamics of the second equation in (5.6). For a given coordinate $i\in[\![m]\!]$, we indeed let

$$B^\circ_i(t,r^\circ,r) := \mathfrak{b}_i\Big(t,(r^\circ,\varsigma^2(r^\circ)r)\Big), \quad t\in[0,T],$$

for $r^\circ \in \mathbb{R}^m$ and $r \in \mathbb{R}^{d-m}$. Notice that the definition is especially interesting for our purpose whenever $r^\circ \in \hat{\mathcal{S}}_m$ and $r \in \mathcal{S}_{d-m-1}$, but it is well defined in any case (with the obvious convention that $\varsigma^2(r^\circ) = 1 - (r_1^\circ + \cdots + r_m^\circ)$ even if it is negative). Similarly, for $i \in [\![d-m]\!]$, we let

$$B_i(t, r^\circ, r) := \mathfrak{b}_{m+i}\Big(t, (r^\circ, \varsigma^2(r^\circ)r)\Big), \quad t \in [0, T].$$

For a new collection of antisymmetric Brownian motions $\overline{W}^\circ = (\overline{W}^\circ_t = (\overline{W}^{\circ,i,j}_t)_{i,j\in[\![d]\!]:i\neq j})_{0\le t\le T}$ of dimension $d(d-1)/2$ (with the convention that $\overline{W}^{\circ,i,i} \equiv 0$ for $i \in [\![d]\!]$), we consider the system

$$\begin{aligned}
\mathrm{d}P^i_t &= \varsigma^{-2}(P^\circ_t)\Big(B_i(t, P^\circ_t, P_t) - P^i_t \sum_{j\in[\![d-m]\!]} B_j(t, P^\circ_t, P_t)\Big)\mathrm{d}t \\
&\quad + \varepsilon\varsigma^{-1}(P^\circ_t) \sum_{j\in[\![d-m]\!]} \sqrt{P^i_t P^j_t}\,\mathrm{d}\overline{W}^{i+m,j+m}_t, \quad i \in [\![d-m]\!], \\
\mathrm{d}(P^\circ_t)^i &= B^\circ_i(t, P^\circ_t, P_t)\mathrm{d}t + \varepsilon \sum_{j\in[\![m]\!]} \sqrt{(P^\circ_t)^i (P^\circ_t)^j}\,\mathrm{d}\overline{W}^{\circ,i,j}_t \\
&\quad + \varepsilon\varsigma(P^\circ_t) \sum_{j\in[\![d-m]\!]} \sqrt{(P^\circ_t)^i (P_t)^j}\,\mathrm{d}\overline{W}^{\circ,i,m+j}_t, \quad i \in [\![m]\!],
\end{aligned} \tag{5.36}$$

for $t \in [t_0, T]$. The unique solvability of (5.36) is addressed in the next two steps.

Second Step. Observing that $\mathfrak{b}$ is Lipschitz continuous, we deduce that the coefficients of (5.36) are Lipschitz continuous in the entries (P°, P) as long the coordinates of the latter remain bounded and away from zero and as long as the sum of the coordinates of P° remains away below 1, we deduce that, for any small $a > 0$, the system (5.36) is uniquely solvable up to the first τ^a when one of the coordinates of $P^\circ_{\tau^a}$ or of P_{τ^a} becomes lower than a or when the sum of the coordinates of $P^\circ_{\tau^a}$ becomes greater than $1 - a$. Letting a tend to 0, we deduce that (5.36) is uniquely solvable up to time $\tau = \lim_{a\searrow 0} \tau^a \wedge T$.

Hence, unique solvability follows if we can prove that

$$\mathbb{P}\Big(\inf_{i\in[\![m]\!]} \inf_{t_0\le t<\tau} P^{\circ,i}_t > 0, \quad \sup_{t_0\le t<\tau} \sum_{i\in[\![m]\!]} P^{\circ,i}_t < 1, \quad \forall t \in [t_0, \tau), \ \sum_{i\in[\![d-m]\!]} P^i_t = 1\Big) = 1, \tag{5.37}$$

since the latter implies that $\tau = T$.

In order to check (5.37), we first observe that, for $t \in [t_0, \tau)$, $\mathrm{d}(\sum_{i\in[\![d-m]\!]} P^i_t) = 0$. Hence,

$$\sum_{i\in[\![d-m]\!]} P^i_t = 1, \quad t \in [t_0, \tau].$$

(Notice that the time interval is closed: Observing that the coefficients in (5.36) are bounded, we may indeed easily extend the solution in hand at time τ itself.) This prompts us to let

$$\widetilde{X}^i_t = (P^\circ_t)^i, \quad i \in [\![m]\!]\,; \quad \widetilde{X}^i_t = \varsigma^2(P^\circ_t)P^{i-m}_t, \quad i = m+1, \cdots, d,$$

for $t \in [t_0, \tau]$. Observe in particular that $\sum_{i\in[\![d]\!]} \widetilde{X}^i_t = 1$, for all $t \in [t_0, \tau]$. If we prove that $(\widetilde{X}^1_t, \cdots, \widetilde{X}^d_t)_{t_0\le t\le\tau}$ satisfies the SDE (5.1) but for a new choice of the noise, then we are done: Not only we then deduce from Proposition 5.1 (or, equivalently, Proposition 2.2) that (5.37) indeed holds true, but we also obtain the required identity in law, see (5.5).

Third Step. In order to prove that $(\widetilde{X}^1_t, \cdots, \widetilde{X}^d_t)_{t_0 \le t \le \tau}$ satisfies (5.1) (for a new choice of noise), we proceed as follows. First, we notice that, for $i \in [\![m]\!]$,

$$\mathrm{d}\widetilde{X}^i_t = \mathfrak{b}_i(t, \widetilde{X}_t)\mathrm{d}t + \varepsilon \sum_{j \in [\![d]\!]} \sqrt{\widetilde{X}^i_t}\sqrt{\widetilde{X}^j_t}\mathrm{d}\overline{W}^{\circ,i,j}_t, \quad t \in [t_0, \tau].$$

And, for $i = m+1, \cdots, d$,

$$\begin{aligned}
\mathrm{d}\widetilde{X}^i_t = &\Big(B_{i-m}(t, P^\circ_t, P_t) - P^{i-m}_t \sum_{j \in [\![d-m]\!]} B_j(t, P^\circ_t, P_t)\Big)\mathrm{d}t + \varepsilon\varsigma(P^\circ_t) \sum_{j \in [\![d-m]\!]} \sqrt{P^{i-m}_t P^j_t}\mathrm{d}\overline{W}^{i,j+m}_t \\
&- P^{i-m}_t \sum_{j \in [\![m]\!]} B^\circ_j(t, P^\circ_t, P_t)\mathrm{d}t - \sum_{j \in [\![d-m]\!]} \mathrm{d}\langle P^i_t, (P^\circ_t)^j\rangle \\
&- \varepsilon P^{i-m}_t \sum_{j,l \in [\![m]\!]} \sqrt{(P^\circ_t)^j (P^\circ_t)^l}\mathrm{d}\overline{W}^{\circ,j,l}_t - \varepsilon P^{i-m}_t \varsigma(P^\circ_t) \sum_{j \in [\![m]\!]} \sum_{l \in [\![d-m]\!]} \sqrt{(P^\circ_t)^j (P_t)^l}\mathrm{d}\overline{W}^{\circ,j,m+l}_t.
\end{aligned} \tag{5.38}$$

Obviously, the bracket on the second line is zero since the underlying noises are independent. Hence, using the fact that $\sum_{i \in [\![d]\!]} \mathfrak{b}_i(t,p) = 0$ for any $(t,p) \in [0,T] \times \mathbb{R}^d$, the drift reads

$$\begin{aligned}
&B_{i-m}(t, P^\circ_t, P_t) - P^{i-m}_t \sum_{j \in [\![d-m]\!]} B_j(t, P^\circ_t, P_t) - P^{i-m}_t \sum_{j \in [\![m]\!]} B^\circ_j(t, P^\circ_t, P_t) \\
&= \mathfrak{b}_i(t, \widetilde{X}_t) - P^{i-m}_t \sum_{j \in [\![d]\!]} \mathfrak{b}_j(t, P^\circ_t, P_t) = \mathfrak{b}_i(t, \widetilde{X}_t).
\end{aligned}$$

Therefore, in order to prove that $(\widetilde{X}_t)_{t_0 \le t \le \tau}$ satisfies (5.1) (for a new choice of noise), it suffices to identify the martingale structure in (5.38). To do so, we rewrite the three martingale increments in the above expansion for $i = m+1, \cdots, d$ in the form

$$\begin{aligned}
&\varepsilon\varsigma(P^\circ_t) \sum_{j \in [\![d-m]\!]} \sqrt{P^{i-m}_t P^j_t}\mathrm{d}\overline{W}^{i,j+m}_t \\
&\quad - \varepsilon P^{i-m}_t \sum_{j,l \in [\![m]\!]} \sqrt{(P^\circ_t)^j (P^\circ_t)^l}\mathrm{d}\overline{W}^{\circ,j,l}_t - \varepsilon P^{i-m}_t \varsigma(P^\circ_t) \sum_{j \in [\![m]\!]} \sum_{l \in [\![d-m]\!]} \sqrt{(P^\circ_t)^j (P_t)^l}\mathrm{d}\overline{W}^{\circ,j,m+l}_t \\
&= \varepsilon\varsigma^{-1}(P^\circ_t) \sum_{j=m+1}^{d} \sqrt{\widetilde{X}^i_t \widetilde{X}^j_t}\mathrm{d}\overline{W}^{i,j}_t - \varepsilon\varsigma^{-2}(P^\circ_t)\widetilde{X}^i_t \sum_{j \in [\![m]\!]} \sum_{l \in [\![d]\!]} \sqrt{\widetilde{X}^j_t \widetilde{X}^l_t}\mathrm{d}\overline{W}^{\circ,j,l}_t.
\end{aligned}$$

Hence, in order to complete the analysis, it remains to compute the various brackets $\mathrm{d}\langle \widetilde{X}^i_t, \widetilde{X}^j_t\rangle$ for $i, j \in [\![d]\!]$. Obviously, whenever $i, j \in [\![m]\!]$,

$$\mathrm{d}\langle \widetilde{X}^i_t, \widetilde{X}^j_t\rangle = \varepsilon^2\Big(\widetilde{X}^i_t \delta_{i,j} - \widetilde{X}^i_t \widetilde{X}^j_t\Big)\mathrm{d}t.$$

Now, $i, j = m+1, \cdots, d$,

$$
\begin{aligned}
\mathrm{d}\langle \widetilde{X}^i_t, \widetilde{X}^j_t \rangle_t &= \varepsilon^2 \varsigma^{-2}(P^\circ_t) \sqrt{\widetilde{X}^i_t \widetilde{X}^j_t} \sum_{l,l'=m+1}^{d} \sqrt{\widetilde{X}^l_t \widetilde{X}^{l'}_t} (\delta_{i,j}\delta_{l,l'} - \delta_{i,l'}\delta_{j,l})\mathrm{d}t \\
&\quad + \varepsilon^2 \varsigma^{-4}(P^\circ_t) \widetilde{X}^i_t \widetilde{X}^j_t \sum_{k,k' \in [\![m]\!]} \sum_{l,l' \in [\![d]\!]} \sqrt{\widetilde{X}^k_t \widetilde{X}^l_t \widetilde{X}^{k'}_t \widetilde{X}^{l'}_t} (\delta_{k,k'}\delta_{l,l'} - \delta_{k,l'}\delta_{k',l})\mathrm{d}t \\
&= \varepsilon^2 \varsigma^{-2}(P^\circ_t) \Big(\delta_{i,j} \widetilde{X}^i_t \sum_{l=m+1}^{d} \widetilde{X}^l_t - \widetilde{X}^i_t \widetilde{X}^j_t \Big) \mathrm{d}t \\
&\quad + \varepsilon^2 \varsigma^{-4}(P^\circ_t) \widetilde{X}^i_t \widetilde{X}^j_t \Big(\sum_{k \in [\![m]\!]} \sum_{l \in [\![d]\!]} \widetilde{X}^k_t \widetilde{X}^l_t - \sum_{k \in [\![m]\!]} \sum_{l \in [\![m]\!]} \widetilde{X}^k_t \widetilde{X}^l_t \Big) \mathrm{d}t \\
&= \varepsilon^2 \varsigma^{-2}(P^\circ_t) \Big(\delta_{i,j} \widetilde{X}^i_t \sum_{l=m+1}^{d} \widetilde{X}^l_t - \widetilde{X}^i_t \widetilde{X}^j_t \Big) \mathrm{d}t + \varepsilon^2 \varsigma^{-4}(P^\circ_t) \widetilde{X}^i_t \widetilde{X}^j_t \sum_{k=1}^{m} \sum_{l=m+1}^{d} \widetilde{X}^k_t \widetilde{X}^l_t \mathrm{d}t.
\end{aligned}
$$

Now, the key point is to observe that $\sum_{l=m+1}^{d} \widetilde{X}^l_t = 1 - \sum_{l=1}^{m} \widetilde{X}^l_t = \varsigma^2(P^\circ_t)$. Therefore,

$$
\begin{aligned}
\mathrm{d}\langle \widetilde{X}^i_t, \widetilde{X}^j_t \rangle_t &= \Big(\varepsilon^2 \delta_{i,j} \widetilde{X}^i_t - \varepsilon^2 \varsigma^{-2}(P^\circ_t) \widetilde{X}^i_t \widetilde{X}^j_t + \varepsilon^2 \varsigma^{-2}(P^\circ_t) \widetilde{X}^i_t \widetilde{X}^j_t (1 - \varsigma^2(P^\circ_t)) \Big) \mathrm{d}t \\
&= \varepsilon^2 \Big(\delta_{i,j} \widetilde{X}^i_t - \widetilde{X}^i_t \widetilde{X}^j_t \Big) \mathrm{d}t.
\end{aligned}
$$

Now, for $i = 1, \cdots, m$ and for $j = m+1, \cdots, d$,

$$
\begin{aligned}
\mathrm{d}\langle \widetilde{X}^i_t, \widetilde{X}^j_t \rangle_t &= -\varepsilon^2 \varsigma^{-2}(P^\circ_t) \widetilde{X}^j_t \sum_{l,l'=1}^{d} \sum_{k=1}^{m} \sqrt{\widetilde{X}^i_t \widetilde{X}^l_t} \sqrt{\widetilde{X}^k_t \widetilde{X}^{l'}_t} (\delta_{i,k}\delta_{l,l'} - \delta_{i,l'}\delta_{k,l})\mathrm{d}t \\
&= \Big(-\varepsilon^2 \varsigma^{-2}(P^\circ_t) \widetilde{X}^i_t \widetilde{X}^j_t + \varepsilon^2 \varsigma^{-2}(P^\circ_t) \widetilde{X}^i_t \widetilde{X}^j_t \sum_{k=1}^{m} \widetilde{X}^k_t \Big) \mathrm{d}t = -\varepsilon^2 \widetilde{X}^i_t \widetilde{X}^j_t \mathrm{d}t,
\end{aligned}
$$

which completes the proof. □

Acknowledgment. We are thankful to the anonymous referee for his or her suggestions, which helped us improve the paper.

References

[1] Y. Achdou, F. J. Buera, J.-M. Lasry, P.-L. Lions, and B. Moll. Partial differential equation models in macroeconomics. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 372(2028):20130397, 19, 2014.

[2] A. A. Albanese and E. M. Mangino. Analyticity of a class of degenerate evolution equations on the canonical simplex of $\mathbb{R}^d$ arising from Fleming-Viot processes. *J. Math. Anal. Appl.*, 379(1):401–424, 2011.

[3] E. Bayraktar, A. Cecchin, A. Cohen, and F. Delarue. Finite state mean field games with Wright Fisher common noise as limits of N-player weighted games. *arXiv e-prints*, page arXiv:2012.04845, Dec. 2020.

[4] E. Bayraktar and A. Cohen. Analysis of a finite state many player game using its master equation. *SIAM Journal on Control and Optimization*, 56(5):3538–3568, 2018.

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

[6] A. Bensoussan, J. Frehse, and P. Yam. *Mean field games and mean field type control theory*. SpringerBriefs in Mathematics. Springer, New York, 2013.

[7] A. Bensoussan, J. Frehse, and S. C. P. Yam. The master equation in mean field theory. *J. Math. Pures Appl. (9)*, 103(6):1441–1474, 2015.

[8] C. Bertucci, J.-M. Lasry, and P.-L. Lions. Some remarks on mean field games. *Communications in Partial Differential Equations*, 44(3):205–227, 2019.
[9] P. Cardaliaguet. Notes on mean field games. https://www.ceremade.dauphine.fr/ cardalia/MFG20130420.pdf/, 2013.
[10] P. Cardaliaguet, F. Delarue, J.-M. Lasry, and P.-L. Lions. *The master equation and the convergence problem in mean field games*, volume 201 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 2019.
[11] R. Carmona and F. Delarue. *Probabilistic theory of mean field games with applications. I*, volume 83 of *Probability Theory and Stochastic Modelling*. Springer, Cham, 2018. Mean field FBSDEs, control, and games.
[12] R. Carmona and F. Delarue. *Probabilistic theory of mean field games with applications. II*, volume 84 of *Probability Theory and Stochastic Modelling*. Springer, Cham, 2018. Mean field games with common noise and master equations.
[13] R. Carmona and P. Wang. A Probabilistic Approach to Extended Finite State Mean Field Games. *arXiv e-prints*, arXiv:1808.07635, Aug 2018.
[14] A. Cecchin, P. Dai Pra, M. Fischer, and G. Pelino. On the convergence problem in mean field games: a two state model without uniqueness. *SIAM J. Control Optim.*, 57(4):2443–2466, 2019.
[15] A. Cecchin and F. Delarue. Selection by vanishing common noise for potential finite state mean field games. *arXiv e-prints*, page arXiv:2005.12153, May 2020.
[16] A. Cecchin and M. Fischer. Probabilistic approach to finite state mean field games. *Appl. Math. Optim.*, 81(2):253–300, 2020.
[17] A. Cecchin and G. Pelino. Convergence, fluctuations and large deviations for finite state mean field games via the master equation. *Stochastic Processes and their Applications*, 129(11):4510 – 4555, 2019.
[18] J.-F. Chassagneux, D. Crisan, and F. Delarue. Numerical method for FBSDEs of McKean-Vlasov type. *Ann. Appl. Probab.*, 29(3):1640–1684, 2019.
[19] L. Chen and D. W. Stroock. The fundamental solution to the Wright-Fisher equation. *SIAM J. Math. Anal.*, 42(2):539–567, 2010.
[20] M. F. Chen and S. F. Li. Coupling methods for multidimensional diffusion processes. *Ann. Probab.*, 17(1):151–177, 1989.
[21] M. Cranston. A probabilistic approach to gradient estimates. *Canad. Math. Bull.*, 35(1):46–55, 1992.
[22] F. Delarue. On the existence and uniqueness of solutions to fbsdes in a non-degenerate case. *Stochastic Processes and their Applications*, 99(2):209 – 286, 2002.
[23] F. Delarue. Restoring uniqueness to mean-field games by randomizing the equilibria. *Stochastics and Partial Differential Equations: Analysis and Computations*, Mar 2019.
[24] F. Delarue and R. Foguen Tchuendom. Selection of equilibria in a linear quadratic mean-field game. *Stochastic Processes and their Applications*, In press.
[25] R. Durrett. *Stochastic calculus*. Probability and Stochastics Series. CRC Press, Boca Raton, FL, 1996. A practical introduction.
[26] A. Eberle. Reflection couplings and contraction rates for diffusions. *Probab. Theory Related Fields*, 166(3-4):851–886, 2016.
[27] C. L. Epstein and R. Mazzeo. Wright-Fisher diffusion in one dimension. *SIAM J. Math. Anal.*, 42(2):568–608, 2010.
[28] C. L. Epstein and R. Mazzeo. *Degenerate diffusion operators arising in population biology*, volume 185 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 2013.
[29] C. L. Epstein and R. Mazzeo. Harnack inequalities and heat kernel estimates for degenerate diffusion operators arising in population biology. *Appl. Math. Res. Express. AMRX*, (2):217–280, 2016.
[30] C. L. Epstein and C. A. Pop. The Feynman-Kac formula and Harnack inequality for degenerate diffusions. *Ann. Probab.*, 45(5):3336–3384, 2017.
[31] S. N. Ethier. A class of degenerate diffusion processes occurring in population genetics. *Comm. Pure Appl. Math.*, 29(5):483–493, 1976.
[32] W. Feller. Diffusion processes in genetics. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950*, pages 227–246. University of California Press, Berkeley and Los Angeles, 1951.
[33] B. Gashi and J. Li. Backward stochastic differential equations with unbounded generators. *Stoch. Dyn.*, 19(1):1950008, 30, 2019.
[34] D. Gomes, R. M. Velho, and M.-T. Wolfram. Dual two-state mean-field games. In *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*, pages 2703–2708. IEEE, 2014.
[35] D. Gomes, R. M. Velho, and M.-T. Wolfram. Socio-economic applications of finite state mean field games. *Phil. Trans. R. Soc. A*, 372(2028):20130405, 2014.
[36] D. A. Gomes, J. Mohr, and R. R. Souza. Discrete time, finite state space mean field games. *J. Math. Pures Appl. (9)*, 93(3):308–328, 2010.
[37] D. A. Gomes, J. Mohr, and R. R. Souza. Continuous time finite state mean field games. *Appl. Math. Optim.*,

68(1):99–143, 2013.
[38] C. Goulaouic and N. Shimakura. Regularité hölderienne de certains problèmes aux limites elliptiques dégénérés. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, 10(1):79–108, 1983.
[39] O. Guéant. Existence and uniqueness result for mean field games with congestion effect on graphs. *Appl. Math. Optim.*, 72(2):291–303, 2015.
[40] M. Huang, P. E. Caines, and R. P. Malhamé. The Nash certainty equivalence principle and Mckean-Vlasov systems: An invariance principle and entry adaptation. In *Decision and Control, 2007 46th IEEE Conference on*, pages 121–126. IEEE, 2007.
[41] M. Huang, R. P. Malhamé, and P. E. Caines. Large population stochastic dynamic games: Closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle. *Commun. Inf. Syst.*, 6(3):221–251, 2006.
[42] I. Karatzas and S. E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.
[43] M. Kimura. Diffusion models in population genetics. *J. Appl. Probability*, 1:177–232, 1964.
[44] J.-M. Lasry and P.-L. Lions. Jeux à champ moyen. I. Le cas stationnaire. *C. R. Math. Acad. Sci. Paris*, 343(9):619–625, 2006.
[45] J.-M. Lasry and P.-L. Lions. Jeux à champ moyen. II. Horizon fini et contrôle optimal. *C. R. Math. Acad. Sci. Paris*, 343(10):679–684, 2006.
[46] J.-M. Lasry and P.-L. Lions. Mean field games. *Jpn. J. Math.*, 2(1):229–260, 2007.
[47] T. Lindvall and L. C. G. Rogers. Coupling of multidimensional diffusions by reflection. *Ann. Probab.*, 14(3):860–872, 1986.
[48] P.-L. Lions. Cours au collège de france, equations aux dérivées partielles et applications. https://www.college-de-france.fr/site/pierre-louis-lions/course-2010-2011.htm, 2010-11.
[49] P.-L. Lions. Cours au collège de france, equations aux dérivées partielles et applications. https://www.college-de-france.fr/site/pierre-louis-lions/course-2008-2009.htm, 2008-09.
[50] P.-L. Lions. Estimées nouvelles pour les équations quasilinéaires. Seminar in Applied Mathematics at the Collège de France. http://www.college-de-france.fr/site/pierre-louis-lions/seminar-2014-11-14-11h15.htm, 2014.
[51] D. Luo and J. Wang. Coupling by reflection and hölder regularity for non-local operators of variable order. *Trans. Amer. Math. Soc.*, 371(1):431–459, 2019.
[52] J. Ma, P. Protter, and J. Yong. Solving forward-backward stochastic differential equations explicitly — a four step scheme. *Probability Theory and Related Fields*, 98(3):339–359, Sep 1994.
[53] C. A. Pop. C^0-estimates and smoothness of solutions to the parabolic equation defined by Kimura operators. *J. Funct. Anal.*, 272(1):47–82, 2017.
[54] C. A. Pop. Existence, uniqueness and the strong Markov property of solutions to Kimura diffusions with singular drift. *Trans. Amer. Math. Soc.*, 369(8):5543–5579, 2017.
[55] K.-i. Sato. Diffusion processes and a class of Markov chains related to population genetics. *Osaka Math. J.*, 13(3):631–659, 1976.
[56] N. Shimakura. Équations différentielles provenant de la génétique des populations. *Tôhoku Math. J.*, 29(2):287–318, 1977.
[57] R. F. Tchuendom. Uniqueness for linear-quadratic mean field games with common noise. *Dyn. Games Appl.*, 8(1):199–210, 2018.
[58] A. J. Veretennikov. Strong solutions and explicit formulas for solutions of stochastic integral equations. *Mat. Sb. (N.S.)*, 111(153)(3):434–452, 480, 1980.