



**HAL**  
open science

# NUMERICAL METHOD FOR FBSDES OF MCKEAN-VLASOV TYPE

Jean-François Chassagneux, Dan Crisan, Francois Delarue

► **To cite this version:**

Jean-François Chassagneux, Dan Crisan, Francois Delarue. NUMERICAL METHOD FOR FBSDES OF MCKEAN-VLASOV TYPE. The Annals of Applied Probability, 2019. hal-01711187

**HAL Id: hal-01711187**

**<https://hal.science/hal-01711187>**

Submitted on 16 Feb 2018

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

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

# NUMERICAL METHOD FOR FBSDES OF MCKEAN-VLASOV TYPE

JEAN-FRANÇOIS CHASSAGNEUX\*, DAN CRISAN† AND FRANÇOIS DELARUE‡

ABSTRACT. This paper is dedicated to the presentation and the analysis of a numerical scheme for forward-backward SDEs of the McKean-Vlasov type, or equivalently for solutions to PDEs on the Wasserstein space. Because of the mean field structure of the equation, earlier methods for classical forward-backward systems fail. The scheme is based on a variation of the method of continuation. The principle is to implement recursively local Picard iterations on small time intervals.

We establish a bound for the rate of convergence under the assumption that the decoupling field of the forward-backward SDE (or equivalently the solution of the PDE) satisfies mild regularity conditions. We also provide numerical illustrations.

## 1. INTRODUCTION

In this paper, we investigate a probabilistic numerical method to approximate the solution of the following non-local PDE

$$\begin{aligned}
 & \partial_t \mathcal{U}(t, x, \mu) + b(x, \mathcal{U}(t, x, \mu), \nu) \cdot \partial_x \mathcal{U}(t, x, \mu) \\
 & + \frac{1}{2} \text{Tr}[\partial_{xx}^2 \mathcal{U}(t, x, \mu) a(x, \mu)] + f(x, \mathcal{U}(t, x, \mu), \partial_x \mathcal{U}(t, x, \mu) \sigma(x, \mu), \nu) \\
 & + \int_{\mathbb{R}^d} \partial_\mu \mathcal{U}(t, x, \mu)(v) \cdot b(v, \mathcal{U}(t, v, \nu), \nu) d\mu(v) \\
 & + \int_{\mathbb{R}^d} \frac{1}{2} \text{Tr}[\partial_x \partial_\mu \mathcal{U}(t, x, \mu)(v) a(v, \mu)] d\mu(v) = 0,
 \end{aligned} \tag{1}$$

for  $(t, x, \mu) \in [0, T) \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$  with the terminal condition  $\mathcal{U}(T, \cdot) = g(\cdot)$ , where  $\nu$  is a notation for the image of the probability measure  $\mu$  by the mapping  $\mathbb{R}^d \ni x \mapsto (x, \mathcal{U}(t, x, \mu)) \in \mathbb{R}^{2d}$ . Above,  $a(x, \mu) = [\sigma \sigma^\dagger](x, \mu)$ . The set  $\mathcal{P}_2(\mathbb{R}^d)$  is the set of probability measures with a finite second-order moment, endowed with the Wasserstein distance i.e.

$$\mathcal{W}_2(\mu, \mu') := \inf_{\pi} \left( \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - x'|^2 d\pi(x, x') \right)^{\frac{1}{2}},$$

for  $(\mu, \mu') \in \mathcal{P}_2(\mathbb{R}^d) \times \mathcal{P}_2(\mathbb{R}^d)$ , the infimum being taken over the probability distributions  $\pi$  on  $\mathbb{R}^d \times \mathbb{R}^d$  whose marginals on  $\mathbb{R}^d$  are respectively  $\mu$  and  $\mu'$ .

Whilst the first two lines in (1) form a classical non-linear parabolic equations, the last two terms are non-standard. Not only are they non-local, in the sense that the solution or its derivatives are computed at points  $v$  different from  $x$ , but also they involve derivatives in the argument  $\mu$ , which lives in a space of probability measures. In this regard, the notation  $\partial_\mu \mathcal{U}(t, x, \mu)(v)$  denotes the so-called *Wasserstein derivative* of the function  $\mathcal{U}$

---

\*Laboratoire de Probabilités et Modèles aléatoires, Université Paris Diderot. [chassagneux@math.univ-paris-diderot.fr](mailto:chassagneux@math.univ-paris-diderot.fr)

†Department of Mathematics, Imperial College London. [d.crisan@imperial.ac.uk](mailto:d.crisan@imperial.ac.uk)

‡Laboratoire Jean-Alexandre Dieudonné, Université de Nice Sophia-Antipolis. [delarue@unice.fr](mailto:delarue@unice.fr)

in the direction of the measure, computed at point  $(t, x, \mu)$  and taken at the *continuous coordinate*  $v$ . We provide below a short reminder of the construction of this derivative, as introduced by Lions, see [12] or [17, Chap. 5].

These PDEs arise in the study of large population stochastic control problems, either of mean field game type, see for instance [12, 13, 20, 30] or [18, Chap. 12] and the references therein, or of mean field control type, see for instance [9, 10, 20, 33]. In both cases,  $\mathcal{U}$  plays the role of a value function or, when the above equation is replaced by a system of equations of the same form, the gradient of the value function. Generally speaking, these types of equations are known as “master equations”. We refer to the aforementioned papers and monographs for a complete overview of the subject, in which existence and uniqueness of classical or viscosity solutions have been studied. In particular, in our previous paper [20], we tackled classical solutions by connecting  $\mathcal{U}$  with a system of fully coupled Forward-Backward Stochastic Differential Equations of the McKean-Vlasov type (MKV FBSDE), for which  $\mathcal{U}$  plays the role of a decoupling field. We also refer to [18, Chap. 12] for a similar approach.

In the current paper, we build on this link to design our numerical method.

The connection between  $\mathcal{U}$  and FBSDEs may be stated as follows. Basically,  $\mathcal{U}$  may be written as  $\mathcal{U}(t, x, \mu) = Y_t^{t,x,\mu}$  for all  $(t, x, \mu) \in [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$ , where  $Y^{t,x,\mu}$  together with  $(X^{t,x,\mu}, Z^{t,x,\mu})$  solves the following standard FBSDE:

$$X_s^{t,x,\mu} = x + \int_t^s b(X_r^{t,x,\mu}, Y_r^{t,x,\mu}, [X_r^{t,\xi}, Y_r^{t,\xi}])dr + \int_t^s \sigma(X_r^{t,x,\mu}, [X_r^{t,x,\mu}])dW_r \quad (2)$$

$$Y_s^{t,x,\mu} = g(X_T^{t,x,\mu}, [X_T^{t,\xi}]) + \int_s^T f(X_r^{t,x,\mu}, Y_r^{t,x,\mu}, Z_r^{t,x,\mu}, [X_r^{t,\xi}, Y_r^{t,\xi}])dr - \int_s^T Z_r^{t,x,\mu} \cdot dW_r, \quad (3)$$

which is parametrized by the law of the following MKV FBSDE:

$$X_s^{t,\xi} = \xi + \int_t^s b(X_r^{t,\xi}, Y_r^{t,\xi}, [X_r^{t,\xi}, Y_r^{t,\xi}])dr + \int_t^s \sigma(X_r^{t,\xi}, [X_r^{t,\xi}])dW_r \quad (4)$$

$$Y_s^{t,\xi} = g(X_T^{t,\xi}, [X_T^{t,\xi}]) + \int_s^T f(X_r^{t,\xi}, Y_r^{t,\xi}, Z_r^{t,\xi}, [X_r^{t,\xi}, Y_r^{t,\xi}])dr - \int_s^T Z_r^{t,\xi} \cdot dW_r, \quad (5)$$

where  $(W_t)_{0 \leq t \leq T}$  is a Brownian motion and  $\xi$  has  $\mu$  as distribution. In the previous equations and in the sequel, we use the notation  $[\theta]$  for the law of a random variable  $\theta$ . In particular, in the above, we have that  $[\xi] = \mu$ . So, to obtain an approximation of  $\mathcal{U}(t, x, \mu)$  given by the initial value of (3), our strategy is to approximate the system (4)-(5) as its solution appears in the coefficients of (2)-(3). In this regard, our approach is probabilistic.

Actually, our paper is not the first one to address the numerical approximation of equations of the type (1) by means of a probabilistic approach. In its PhD dissertation, Alanko [4] develops a numerical method for mean field games based upon a Picard iteration: Given the proxy for the equilibrium distribution of the population (which is represented by the mean field component in the above FBSDE), one solves for the value function by approximating the solution of the (standard) BSDE associated with the control problem; given the solution of the BSDE, we then get a new proxy for the equilibrium distribution and so on... Up to a Girsanov transformation, the BSDE associated with the control problem coincides with the backward equation in the above FBSDEs. In [4], the Girsanov transformation is indeed used to decouple the forward and backward equations and it is the keystone of the paper to address the numerical impact of the change of

measure onto the mean field component. Under our setting, this method would more or less consist in solving for the backward equation given a proxy for the forward equation and then in iterating, which is what we call the *Picard method* for the FBSDE system. Unfortunately, convergence of the Picard iterations is a difficult issue, as the convergence is known in small time only, see the numerical examples in Section 4 below. It is indeed well-known that Picard theorem only applies in small time for fully coupled problems. In this regard, it must be stressed that our system (4)-(5) is somehow doubly coupled, once in the variable  $x$  and once in the variable  $\mu$ , which explains why a change of measure does not permit to decouple it entirely. As a matter of fact, the convergence of the numerical method is not explicitly addressed in [4].

In fact, a similar limitation on the length of the time horizon has been pointed out in other works on the numerical analysis of a mean field game. For instance, in a slightly different setting from ours, which does not explicitly appeal to a forward-backward system of the type (4)-(5), Bayraktar, Budhiraja and Cohen [5] provide a probabilistic numerical approach for a mean field game with state constraints set over a queuing system. The scheme is constructed in two steps. The authors first consider a discrete form of the original mean field game based upon a Markov chain approximation method *à la* Kushner-Dupuis of the underlying continuous-time control problem. The solution to the discrete-time mean field game is then approximated by means of a Picard scheme: Given a proxy for the law of the optimal trajectories, the discrete Markov decision problem is solved first; the law of the solution then serves as a new proxy for the next step in the Picard sequence. The authors are then successful in proving the convergence of their approximation but again for a small time interval only, see Section 5.2 in [5] for details.

The goal of our paper is precisely to go further and to propose an algorithm whose convergence with a rate is known on any interval of a given length. In the classical case, this question has been addressed by several authors, among which [21, 22] and [7], but all these methods rely on the Markov structure of the problem. Here, the Markov property is true but at the price of regarding the entire  $\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$  as state space: The fact that the second component is infinite dimensional makes intractable the complexity of these approaches. To avoid any similar problem, we use a pathwise approach for the forward component; it consists in iterating successively the *Picard method* on small intervals, all the Picard iterations being implemented with a tree approximation of the Brownian motion. This strategy is inspired from the method of continuation, the parameter in the continuation argument being the time length  $T$  itself. The advantage for working on a tree is twofold: as we said, we completely bypass any Markov argument; also, we get, not only, an approximation of the system (4)-(5) but also, for free, an approximation of the system (2)-(3), which “lives” on a subtree obtained by conditioning on the initial root. We prove that the method is convergent and provide a rate of convergence for it. Numerical examples are given in Section 4. Of course, the complexity remains pretty high in comparison with the methods developed in the classical non McKean-Vlasov case. This should not come as a surprise since, as we already emphasized, the problem is somehow infinite dimensional.

We refer the interested reader to the following papers for various numerical methods, based upon finite differences or variational approaches, for mean field games: [1, 2, 3] and [6, 26, 25].

The paper is organized as follows. The method for the system (4)-(5) is exposed in Section 2. The convergence is addressed in Section 3. In Section 4, we explain how to compute in practice  $\mathcal{U}(t, x, \mu)$  (and thus approximate (2)-(3)) from the approximation of

the sole (4)-(5) and we present some numerical results validating empirically the convergence results obtained in Section 3. We collect in the appendix some key results for the convergence analysis.

## 2. A NEW ALGORITHM FOR COUPLED FORWARD BACKWARD SYSTEMS

As announced right above, we will focus on the approximation of the following type of McKean-Vlasov forward-backward stochastic differential equation:

$$\begin{aligned} dX_t &= b(X_t, Y_t, [X_t, Y_t])dt + \sigma(X_t, [X_t])dW_t, \\ dY_t &= -f(X_t, Y_t, Z_t, [X_t, Y_t])dt + Z_t \cdot dW_t, \quad t \in [0, T], \\ Y_T &= g(X_T, [X_T]) \quad \text{and} \quad X_0 = \xi, \end{aligned} \tag{6}$$

for some time horizon  $T > 0$ . Throughout the analysis, the equation is regarded on a complete filtered probability space  $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ , equipped with a  $d$ -dimensional  $\mathbb{F}$ -Brownian motion  $(W_t)_{0 \leq t \leq T}$ . To simplify, we assume that the state process  $(X_t)_{0 \leq t \leq T}$  is of the same dimension. The process  $(Y_t)_{0 \leq t \leq T}$  is 1-dimensional. As a result,  $(Z_t)_{0 \leq t \leq T}$  is  $d$ -dimensional.

In (6), the three processes  $(X_t)_{0 \leq t \leq T}$ ,  $(Y_t)_{0 \leq t \leq T}$  and  $(Z_t)_{0 \leq t \leq T}$  are required to be  $\mathbb{F}$ -progressively measurable. Both  $(X_t)_{0 \leq t \leq T}$  and  $(Y_t)_{0 \leq t \leq T}$  have continuous trajectories. Generally speaking, the initial condition  $X_0$  is assumed to be square-integrable, but at some point, we will assume that  $X_0$  belongs to  $L^p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d)$ , for some  $p > 2$ . Accordingly,  $(X_t)_{0 \leq t \leq T}$ ,  $(Y_t)_{0 \leq t \leq T}$  and  $(Z_t)_{0 \leq t \leq T}$  must satisfy:

$$\| \| (X, Y, Z) \| \|_{[0, T]} := \mathbb{E} \left[ \sup_{0 \leq t \leq T} (|X_t|^2 + |Y_t|^2) + \int_0^T |Z_t|^2 dt \right]^{1/2} < \infty.$$

The domains and codomains of the coefficients are defined accordingly. The assumption that  $\sigma$  is assumed to be independent of the variable  $y$  is consistent with the global solvability results that exist in the literature for equations like (6). For instance, it covers cases coming from optimization theory for large mean field interacting particle systems. We refer to our previous paper [20] for a complete overview on the subject, together with the references [8, 12, 17, 18, 19]. In light of the examples tackled in [20], the fact that  $b$  is independent of  $z$  may actually seem more restrictive, as it excludes cases when the forward-backward system of the McKean-Vlasov type is used to represent the value function of the underlying optimization problem. It is indeed a well-known fact that, with or without McKean-Vlasov interaction, the value function of a standard optimization problem may be represented as the backward component of a standard FBSDE with a drift term depending upon the  $z$  variable. This says that, in order to tackle the aforementioned optimization problems of the mean field type by means of the numerical method investigated in this paper, one must apply the algorithm exposed below to the Pontryagin system. The latter one is indeed of the form (6), provided that  $Y$  is allowed to be multi-dimensional. (Below, we just focus on the one-dimensional case, but the adaptation is straightforward.)

In fact, our choice for assuming  $b$  to be independent of  $z$  should not come as a surprise. The same assumption appears in the papers [21, 22] dedicated to the numerical analysis of standard FBSDEs, which will serve us as a benchmark throughout the text. See however Remark 4.

Finally, the fact that the coefficients are time-homogeneous is for convenience only.

As a key ingredient in our analysis, we use the following representation result given in e.g. Proposition 2.2 in [20],

$$Y_t^\xi := \mathcal{U}(t, X_t^\xi, [X_t^\xi]) , \quad (7)$$

where  $\mathcal{U} : [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$  is assumed to be the classical solution, in the sense of [20, Definition 2.6], to (1). In this regard, the derivative with respect to the measure argument is defined according to Lions' approach to the Wasserstein derivative. In short, the *lifting*  $\hat{\mathcal{U}}$  of  $\mathcal{U}$  to  $L^2(\Omega, \mathcal{F}_0, \mathbb{P}; \mathbb{R}^d)$ , which we define by

$$\hat{\mathcal{U}}(t, x, \xi) = \mathcal{U}(t, x, [\xi]), \quad t \in [0, T], \quad x \in \mathbb{R}^d, \quad \xi \in L^2(\Omega, \mathcal{F}_0, \mathbb{P}; \mathbb{R}^d),$$

is assumed to be Fréchet differentiable. Of course, this makes sense as long as the space  $(\Omega, \mathcal{F}_0, \mathbb{P})$  is rich enough so that, for any  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ , there exists a random variable  $\xi \in L^2(\Omega, \mathcal{F}_0, \mathbb{P}; \mathbb{R}^d)$  such that  $\xi \sim \mu$ . So, in the sequel,  $(\Omega, \mathcal{F}_0, \mathbb{P})$  is assumed to be atomless, which makes it rich enough. A crucial point with Lions' approach to Wasserstein differential calculus is that the Fréchet derivative of  $\hat{\mathcal{U}}$  in the third variable, which can be identified with a square-integrable random variable, may be represented at point  $(t, x, \xi)$  as  $\partial_\mu \mathcal{U}(t, x, [\xi])(\xi)$  for a mapping  $\partial_\mu \mathcal{U}(t, x, \mu)(\cdot) : \mathbb{R}^d \ni v \mapsto \partial_\mu \mathcal{U}(t, x, \mu)(v) \in \mathbb{R}^d$ . This latter function plays the role of Wasserstein derivative of  $\mathcal{U}$  in the measure argument. To define a classical solution, it is then required that  $\mathbb{R}^d \ni v \mapsto \partial_\mu \mathcal{U}(t, x, \mu)(v)$  is differentiable, both  $\partial_\mu \mathcal{U}$  and  $\partial_v \partial_\mu \mathcal{U}$  being required to be continuous at any point  $(t, x, \mu, v)$  such that  $v$  is in the support of  $\mu$ .

**Assumptions.** Our analysis requires some minimal regularity assumptions on the coefficients  $b, \sigma, f$  and the function  $\mathcal{U}$ . As for the coefficients functions, we assume that there exists a constant  $\Lambda \geq 0$  such that:

- (H0): The functions  $b, \sigma, f$  and  $g$  are  $\Lambda$ -Lipschitz continuous in all the variables, the space  $\mathcal{P}_2(\mathbb{R}^d)$  being equipped with the Wasserstein distance  $\mathcal{W}_2$ . Moreover, the function  $\sigma$  is bounded by  $\Lambda$ .

We now state the main assumptions on  $\mathcal{U}$ , see Remark 1 for comments.

- (H1): for any  $t \in [0, T]$  and  $\xi \in L^2(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^d)$ , the McKean-Vlasov forward-backward system (6) set on  $[t, T]$  instead of  $[0, T]$  with  $X_t = \xi$  as initial condition at time  $t$  has a unique solution  $(X_s^{t, \xi}, Y_s^{t, \xi}, Z_s^{t, \xi})_{t \leq s \leq T}$ ; in parallel,  $\mathcal{U}$  is the classical solution, in the sense of [20, Definition 2.6], to (1); and  $\mathcal{U}$  and its derivatives satisfy

$$|\mathcal{U}(t, x, \mu) - \mathcal{U}(t, x, \mu')| + |\partial_x \mathcal{U}(t, x, \mu) - \partial_x \mathcal{U}(t, x, \mu')| \leq \Lambda \mathcal{W}_2(\mu, \mu') , \quad (8)$$

$$|\partial_x \mathcal{U}(t, x, \mu)| + \|\partial_\mu \mathcal{U}(t, x, [\xi])(\xi)\|_2 \leq \Lambda , \quad (9)$$

$$|\partial_{xx}^2 \mathcal{U}(t, x, \mu)| + \|\partial_v \partial_\mu \mathcal{U}(t, x, [\xi])(\xi)\|_2 \leq \Lambda , \quad (10)$$

$$\text{and } |\partial_{xx}^2 \mathcal{U}(t, x, \mu) - \partial_{xx}^2 \mathcal{U}(t, x', \mu)| \leq \Lambda |x - x'| , \quad (11)$$

for  $(t, x, x', \xi) \in [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \times L^2(\Omega, \mathcal{F}_0, \mathbb{P}; \mathbb{R}^d)$  and  $\mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d)$ . Also, we require that

$$\begin{aligned} & |\mathcal{U}(t+h, x, [\xi]) - \mathcal{U}(t, x, [\xi])| + |\partial_x \mathcal{U}(t+h, x, [\xi]) - \partial_x \mathcal{U}(t, x, [\xi])| \\ & \leq \Lambda h^{\frac{1}{2}} (1 + |x| + \|\xi\|_2) , \end{aligned} \quad (12)$$

and for all  $h \in [0, T)$ ,  $(t, x) \in [0, T - h] \times \mathbb{R}^d$ ,  $\xi \in L^2(\Omega, \mathcal{F}_0, \mathbb{P}; \mathbb{R}^d)$  and  $v, v' \in \mathbb{R}^d$ ,

$$|\partial_v \partial_\mu \mathcal{U}(t, x, [\xi])(v) - \partial_v \partial_\mu \mathcal{U}(t, x, [\xi])(v')| \leq \Lambda \{1 + |v|^{2\alpha} + |v'|^{2\alpha} + \|\xi\|_2^{2\alpha}\}^{\frac{1}{2}} |v - v'|, \quad (13)$$

for some  $\alpha > 0$ .

**Remark 1.** In [20], it is shown that, under some conditions on the coefficients  $b$ ,  $f$  and  $\sigma$ , the PDE (1) has indeed a unique classical solution which satisfies the assumption (H1).

- (1) Estimate (13) is obtained by combining Definition 2.6 and Proposition 3.9 in [20]. A major difficulty in the analysis provided below is the fact that  $\alpha$  may be larger than 1, in which case the Lipschitz bound for the second order derivative is super-linear. This problem is proper to the McKean-Vlasov structure of the equation and does not manifest in the classical setting, compare for instance with [21, 22]. Below, we tackle two cases: the case when  $\alpha \leq 1$ , which has been investigated in [13] and [18, Chap. 12] under stronger conditions on the coefficients, and the case when  $\alpha > 1$  but  $\mathcal{U}$  is bounded.
- (2) Estimates (8)-(12) are required to control the convergence error when the coefficients ( $b$  or  $f$ ) depend on  $Z$ .
  - (a) The estimate (8) can be retrieved from the computations made in [20]. See the comments at the bottom of page 60, near equation (4.58).
  - (b) The estimate (12) comes from the theory of FBSDEs (without McKean-Vlasov interaction). Indeed, using the Lipschitz property of  $\mathcal{U}$  and  $\partial_x \mathcal{U}$  in the variable  $\mu$ , it suffices to prove

$$\begin{aligned} & |\mathcal{U}(t+h, x, [X_{t+h}^{t,\xi}]) - \mathcal{U}(t, x, [X_t^{t,\xi}])| + |\partial_x \mathcal{U}(t+h, x, [X_{t+h}^{t,\xi}]) - \partial_x \mathcal{U}(t, x, [X_t^{t,\xi}])| \\ & \leq \Lambda h^{\frac{1}{2}} (1 + |x| + \|\xi\|_2). \end{aligned}$$

As stated in Proposition 2.2 in [20], for  $\xi \sim \mu$ ,  $\mathcal{U}(s, x, [X_s^{t,\xi}]) = u_{t,\mu}(s, x)$  where  $u_{t,\mu}$  is solution to a quasi-linear PDE. Then the estimate (12) follows from standard results on non-linear PDEs, see e.g. Theorem 2.1 in [21].

In comparison with the assumption used in [21], the condition (H1) is more demanding. In [21], there is no need for assuming the second-order derivative to be Lipschitz in space. This follows from the fact that, here, we approximate the Brownian increments by random variables taking a small number of values, whilst in [21], the Brownian increments are approximated by a quantization grid with a larger number of points. In this regard, our approach is closer to the strategy implemented in [22].

**2.1. Description.** The goal of the numerical method exposed in the paper is to approximate  $\mathcal{U}$ . The starting point is the formula (6) and, quite naturally, the strategy is to approximate the process  $(X^\xi, Y^\xi, Z^\xi) := (X^{0,\xi}, Y^{0,\xi}, Z^{0,\xi})$ .

Generally speaking, this approach raises a major difficulty, as it requires to handle the strongly coupled forward-backward structure of (6). Indeed, theoretical solutions to (6) may be constructed by means of basic Picard iterations but in small time only, which comes in contrast with similar results for decoupled forward or backward equations for which Picard iterations converge on any finite time horizon. In the papers [21, 22] –which deal with the non McKean-Vlasov case–, this difficulty is bypassed by approximating the decoupling field  $\mathcal{U}$  at the nodes of a time-space grid. Obviously, this strategy is hopeless in the McKean-Vlasov setting as the state variable is infinite dimensional; discretizing it

on a grid would be of a non-tractable complexity. This observation is the main rationale for the approach exposed below.

Our method is a variation of the so-called *method of continuation*. In full generality, it consists in increasing step by step the coupling parameter between the forward and backward equations. Of course, the intuition is that, for a given time length  $T$ , the Picard scheme should converge for very small values of the coupling parameter. The goal is then to insert the approximation computed for a small coupling parameter into the scheme used to compute a numerical solution for a higher value of the coupling parameter. Below, we adapt this idea, but we directly regard  $T$  itself as a coupling parameter. So we increase  $T$  step and by step and, on each step, we make use of a Picard iteration based on the approximations obtained at the previous steps.

This naturally motivates the introduction of an equidistant grid  $\mathfrak{R} = \{r_0 = 0, \dots, r_N = T\}$  of the time interval  $[0, T]$ , with  $r_k = k\delta$  and  $\delta = \frac{T}{N}$  for  $N \geq 2$ . In the following we shall consider that  $\delta$  is ‘‘small enough’’ and state more precisely what it means in the main results, see Theorem 5 and Theorem 7.

For  $0 \leq k \leq N - 1$ , we consider intervals  $I_k = [r_k, T]$  and on each interval, the following FBSDE, for  $\xi \in L^2(\mathcal{F}_{r_k})$  (which is a shorter notation for  $L^2(\Omega, \mathcal{F}_{r_k}, \mathbb{P}; \mathbb{R}^d)$ ):

$$X_t = \xi + \int_{r_k}^t b(X_s, Y_s, [X_s, Y_s]) ds + \int_{r_k}^t \sigma(X_s, [X_s]) dW_s, \quad (14)$$

$$Y_t = g(X_T, [X_T]) + \int_t^T f(X_s, Y_s, Z_s, [X_s, Y_s]) ds - \int_t^T Z_s \cdot dW_s. \quad (15)$$

Picard iterations. We need to compute backwards the value of  $\mathcal{U}(r_k, \xi, [\xi])$  for some  $\xi \in L^2(\mathcal{F}_{r_k})$ ,  $0 \leq k \leq N - 2$ . We are then going to solve the FBSDE (14)-(15) on the interval  $I_k$ . As explained above, the difficulty is the arbitrariness of  $T$ : When  $k$  is large,  $I_k$  is of a small length, but this becomes false as  $k$  decreases. Fortunately, we can rewrite the forward-backward system on a smaller interval at the price of changing the terminal boundary condition. Indeed, from (H1), we know that  $(X_s^{r_k, \xi}, Y_s^{r_k, \xi}, Z_s^{r_k, \xi})_{r_k \leq s \leq r_{k+1}}$  solves:

$$\begin{cases} X_t &= \xi + \int_{r_k}^t b(X_s, Y_s, [X_s, Y_s]) ds + \int_{r_k}^t \sigma(X_s, [X_s]) dW_s, \\ Y_t &= \mathcal{U}(r_{k+1}, X_{r_{k+1}}, [X_{r_{k+1}}]) + \int_t^{r_{k+1}} f(X_s, Y_s, Z_s, [X_s, Y_s]) ds - \int_t^{r_{k+1}} Z_s \cdot dW_s, \end{cases}$$

for  $t \in [r_k, r_{k+1}]$ .

If  $\delta$  is small enough, a natural approach is to introduce a Picard iteration scheme to approximate the solution of the above equation. To do so, one can implement the following recursion (with respect to the index  $j$ ):

$$\begin{cases} X_t^j &= \xi + \int_{r_k}^t b(X_s^j, Y_s^j, [X_s^j, Y_s^j]) ds + \int_{r_k}^t \sigma(X_s^j, [X_s^j]) dW_s, \\ Y_t^j &= \mathcal{U}(r_{k+1}, X_{r_{k+1}}^{j-1}, [X_{r_{k+1}}^{j-1}]) \\ &\quad + \int_t^{r_{k+1}} f(X_s^{j-1}, Y_s^j, Z_s^j, [X_s^{j-1}, Y_s^j]) ds - \int_t^{r_{k+1}} Z_s^j \cdot dW_s \end{cases}.$$

with  $(X_s^0 = \xi + \int_{r_k}^t b(X_s^0, 0, [X_s^0, 0]) ds + \int_{r_k}^t \sigma(X_s^0, [X_s^0]) dW_s)_{r_k \leq s \leq r_{k+1}}$  and  $(Y_s^0 = 0)_{r_k \leq s \leq r_{k+1}}$ .

It is known that, for  $\delta$  small enough,  $(X^j, Y^j, Z^j) \rightarrow_{j \rightarrow \infty} (X, Y, Z)$ , in the sense that  $\| (X^j - X, Y^j - Y, Z^j - Z) \|_{[r_k, r_{k+1}]} \rightarrow_{j \rightarrow \infty} 0$ .

But in practice we will encounter three main difficulties.

- (1) The procedure has to be stopped after a given number of iterations  $J$ .
- (2) The above Picard iteration assumes the perfect knowledge of the map  $\mathcal{U}$  at time  $r_k$ , but  $\mathcal{U}$  is exactly what we want to compute...

(3) The solution has to be discretized in time and space.

**Ideal recursion.** We first discuss 1) and 2) above. The main idea is to use a recursive algorithm (with a new recursion, but on the time parameter).

Namely, for  $k \leq N - 1$ , we assume that we are given a solver which computes

$$\text{solver}[k+1](\xi) = \mathcal{U}(r_{k+1}, \xi, [\xi]) + \epsilon^{k+1}(\xi), \quad (16)$$

where  $\epsilon$  is an error made, for any  $\xi \in L^2(\mathcal{F}_{r_{k+1}})$ . We shall sometimes refer to  $\text{solver}[k+1](\cdot)$  as “the solver at level  $k+1$ ”.

Taking these observations into account, we first define an ideal solver, which assumes that each Picard iteration in the approximation of the solution of the forward-backward system can be perfectly computed. We denote it by  $\text{picard}[\cdot](\cdot)$ . Accordingly, we identify (for the time being)  $\text{solver}[k+1](\cdot)$  with  $\text{picard}[k+1](\cdot)$ . Given  $\text{picard}[k+1](\cdot)$ ,  $\text{picard}[k](\cdot)$  is defined as follows.

$$\begin{cases} \tilde{X}_t^{k,j} &= \xi + \int_{r_k}^t b(\tilde{X}_s^{k,j}, \tilde{Y}_s^{k,j}, [\tilde{X}_s^{k,j}, \tilde{Y}_s^{k,j}]) ds + \int_{r_k}^t \sigma(\tilde{X}_s^{k,j}, [\tilde{X}_s^{k,j}]) dW_s, \\ \tilde{Y}_t^{k,j} &= \text{picard}[k+1](\tilde{X}_{r_{k+1}}^{k,j-1}) - \int_t^{r_{k+1}} \tilde{Z}_s^{k,j} \cdot dW_s \\ &\quad + \int_t^{r_{k+1}} f(\tilde{X}_s^{k,j-1}, \tilde{Y}_s^{k,j}, \tilde{Z}_s^{k,j}, [\tilde{X}_s^{k,j-1}, \tilde{Y}_s^{k,j}]) ds, \end{cases} \quad (17)$$

for  $j \geq 1$  and with

$$\left( \tilde{X}_s^{k,0} = \xi + \int_{r_k}^t b(X_s^{k,0}, 0, [X_s^{k,0}, 0]) ds + \int_{r_k}^t \sigma(X_s^{k,0}, [X_s^{k,0}]) dW_s \right)_{r_k \leq s \leq r_{k+1}},$$

and  $(\tilde{Y}_s^{k,0} = 0)_{r_k \leq s \leq r_{k+1}}$ . We then define

$$\text{picard}[k](\xi) := Y_{r_k}^{k,J} \text{ and } \epsilon^k(\xi) := Y_{r_k}^{k,J} - \mathcal{U}(r_k, \xi, [\xi]),$$

where  $J \geq 1$  is the number of Picard iterations.

At level  $N - 1$ , which is the last level for our recursive algorithm, the Picard iteration scheme is given by

$$\begin{cases} \tilde{X}_t^{N-1,j} &= \xi + \int_{r_{N-1}}^t b(\tilde{X}_s^{N-1,j}, \tilde{Y}_s^{N-1,j}, [\tilde{X}_s^{N-1,j}, \tilde{Y}_s^{N-1,j}]) ds \\ &\quad + \int_{r_{N-1}}^t \sigma(\tilde{X}_s^{N-1,j}, [\tilde{X}_s^{N-1,j}]) dW_s, \\ \tilde{Y}_t^{N-1,j} &= g(\tilde{X}_T^{N-1,j-1}, [\tilde{X}_T^{N-1,j-1}]) - \int_t^T \tilde{Z}_s^{N-1,j} \cdot dW_s \\ &\quad + \int_t^T f(\tilde{X}_s^{N-1,j-1}, \tilde{Y}_s^{N-1,j}, \tilde{Z}_s^{N-1,j}, [\tilde{X}_s^{N-1,j-1}, \tilde{Y}_s^{N-1,j}]) ds. \end{cases} \quad (18)$$

Here, the terminal condition  $g$  is known and the error comes from the fact that the Picard iteration is stopped. It is then natural to set, for  $\xi \in L^2(\mathcal{F}_T)$ ,

$$\text{picard}[N](\xi) = g(\xi, [\xi]) \text{ and } \epsilon^N(\xi) = 0. \quad (19)$$

**Practical implementation.** As already noticed in 3) above, it is not possible to solve the backward and forward equations in (17) perfectly, even though the system is decoupled. Hence, we need to introduce an approximation that can be implemented in practice. Given a continuous adapted input process  $\tilde{\mathfrak{X}} = (\tilde{\mathfrak{X}}_s)_{r_k \leq s \leq r_{k+1}}$  such that  $\mathbb{E}[\sup_{r_k \leq s \leq r_{k+1}} |\tilde{\mathfrak{X}}_s|^2] < \infty$  and  $\eta \in L^2(\Omega, \mathcal{F}_{r_{k+1}}, \mathbb{P}; \mathbb{R})$ , we thus would like to solve

$$\begin{cases} \tilde{X}_t &= \tilde{\mathfrak{X}}_{r_k} + \int_{r_k}^t b(\tilde{X}_s, \tilde{Y}_s, [\tilde{X}_s, \tilde{Y}_s]) ds + \int_{r_k}^t \sigma(\tilde{X}_s, [\tilde{X}_s]) dW_s \\ \tilde{Y}_t &= \eta + \int_t^{r_{k+1}} f(\tilde{\mathfrak{X}}_s, \tilde{Y}_s, \tilde{Z}_s, [\tilde{\mathfrak{X}}_s, \tilde{Y}_s]) ds - \int_t^{r_{k+1}} \tilde{Z}_s \cdot dW_s, \end{cases}$$

for  $t \in [r_k, r_{k+1}]$ .

Let  $\pi$  be a discrete time grid of  $[0, T]$  such that  $\mathfrak{R} \subset \pi$ ,

$$\pi := \{t_0 := 0 < \dots < t_n := T\} \text{ and } |\pi| := \max_{i < n} (t_{i+1} - t_i). \quad (20)$$

For  $0 \leq k \leq N - 1$ , we note  $\pi^k := \{t \in \pi \mid r_k \leq t \leq r_{k+1}\}$  and for later use, we define the indices  $(j_k)_{0 \leq k \leq N}$  as follows

$$\pi^k = \{t_{j_k} := r_k < \dots < t_i < \dots < r_{k+1} =: t_{j_{k+1}}\},$$

for all  $k < N$ . So, instead of a perfect solver for an iteration of the Picard scheme (17), we assume that we are given a numerical solver, denoted by  $\overline{\text{solver}}[k](\tilde{\mathfrak{X}}, \eta, f)$ , which computes an approximation of the process  $(\tilde{X}_s, \tilde{Y}_s, \tilde{Z}_s)_{r_k \leq s \leq r_{k+1}}$  on  $\pi^k$  for a discretization  $(\tilde{\mathfrak{X}}_t)_{t \in \pi^k}$  of the time continuous process  $(\mathfrak{X}_s)_{r_k \leq s \leq r_{k+1}}$ . The *output* is denoted by  $(\bar{X}_t, \bar{Y}_t, \bar{Z}_t)_{t \in \pi^k}$ . In parallel, we call *input* the triplet formed by the random variable  $\eta$ , the discrete-time process  $(\tilde{\mathfrak{X}}_t)_{t \in \pi^k}$  and the driver  $f$  of the backward equation. In short, the output is what the numerical solver returns after one iteration in the Picard scheme when the discrete input is  $(\eta, \tilde{\mathfrak{X}}, f)$ . Pay attention that, in contrast with  $b$  and  $\sigma$ , we shall allow  $f$  to vary; this is the rationale for regarding it as an input. However, when the value of  $f$  is clear, we shall just regard the input as the pair  $(\eta, (\tilde{\mathfrak{X}}_t)_{t \in \pi^k})$ .

The full convergence analysis, including the discretization error, will be discussed in the next section in the following two cases: first for a generic (or abstract) solver  $\overline{\text{solver}}[\cdot](\cdot, \cdot)$  and second for an explicit solver, as given in the example below.

**Example 2.** *This example is the prototype of the solver  $\overline{\text{solver}}[\cdot](\cdot, \cdot)$ . We consider an approximation of the Brownian motion obtained by quantization of the Brownian increments. At every time  $t \in \pi$ , we denote by  $\bar{W}_t$  the value at time  $t$  of the discretized Brownian motion. It may expressed as*

$$\bar{W}_{t_i} := \sum_{j=0}^{i-1} \Delta \bar{W}_j,$$

where

$$\Delta \bar{W}_j := h_j^{\frac{1}{2}} \varpi_j, \quad \varpi_j := \Gamma_d \left( h_j^{-\frac{1}{2}} (W_{t_{j+1}} - W_{t_j}) \right),$$

$\Gamma_d$  mapping  $\mathbb{R}^d$  onto a finite grid of  $\mathbb{R}^d$ . Importantly,  $\Gamma_d$  is assumed to be bounded by  $\Lambda$  and each  $\varpi_j$  is assumed to be centered and to have the identity matrix as covariance matrix. Of course, this is true if  $\Gamma_d$  is of the form

$$\Gamma_d(w_1, \dots, w_d) := (\Gamma_1(w_1), \dots, \Gamma_1(w_d)), \quad (w_1, \dots, w_d) \in \mathbb{R}^d,$$

where  $\Gamma_1$  is a bounded odd function from  $\mathbb{R}$  onto a finite subset of  $\mathbb{R}$  with a normalized second order moment under the standard Gaussian measure. In practice,  $\Gamma_d$  is intended to take a small number of values. Of course, the typical example is the so-called binomial approximation, in which case  $\Gamma_1$  is the sign function.

On each interval  $[r_k, r_{k+1}]$ , given a discrete-time input process  $\tilde{\mathfrak{X}}$  and a terminal condition  $\eta$ , we thus implement the following scheme (below,  $\mathbb{E}_{t_i}$  is the conditional expectation given  $\mathcal{F}_{t_i}$ ):

(1) For the backward component:

- (a) Set as terminal condition,  $(\bar{Y}_{t_{j_{k+1}}}, \bar{Z}_{t_{j_{k+1}}}) = (\eta, 0)$ .
- (b) For  $j_k \leq i < j_{k+1}$ , compute recursively

$$\bar{Y}_{t_i} = \mathbb{E}_{t_i} \left[ \bar{Y}_{t_{i+1}} + (t_{i+1} - t_i) f(\tilde{\mathfrak{X}}_{t_i}, \bar{Y}_{t_i}, \bar{Z}_{t_i}, [\tilde{\mathfrak{X}}_{t_i}, \bar{Y}_{t_i}]) \right], \quad \bar{Z}_{t_i} = \mathbb{E}_{t_i} \left[ \frac{\Delta \bar{W}_i}{t_{i+1} - t_i} \bar{Y}_{t_{i+1}} \right].$$

(2) For the forward component:

- (a) Set as initial condition,  $\bar{X}_{t_{j_k}} = \tilde{\mathfrak{X}}_{r_k}$ .
- (b) For  $j_k < i \leq j_{k+1}$ , compute recursively

$$\bar{X}_{t_{i+1}} = \bar{X}_{t_i} + b(\bar{X}_{t_i}, \bar{Y}_{t_i}, [\bar{X}_{t_i}, \bar{Y}_{t_i}])(t_{i+1} - t_i) + \sigma(\bar{X}_{t_i}, [\bar{X}_{t_i}])\Delta\bar{W}_i.$$

Full algorithm for  $\text{solver}[\cdot](\cdot)$ . Using  $\overline{\text{solver}}[\cdot](\cdot, \cdot)$ , for each level, we can now give a completely implementable algorithm for  $\text{solver}[\cdot](\cdot)$ . Its description is as follows.

The value  $\text{solver}[k](\xi)$ , i.e. the value of the solver at level  $k$  with initial condition  $\xi \in L^2(\mathcal{F}_{r_k})$ , is obtained through:

- (1) Initialize the backward component at  $\bar{Y}_t^{k,0} = 0$  for  $t \in \pi_k$  and regard  $(\bar{X}_t^{k,0})_{t \in \pi_k}$  as the forward component of  $\overline{\text{solver}}[k](\xi, 0, 0)$
- (2) for  $1 \leq j \leq J$ 
  - (a) compute  $\bar{Y}_{r_{k+1}}^{k,j} = \text{solver}[k+1](\bar{X}_{r_{k+1}}^{k,j-1})$ .
  - (b) compute  $(\bar{X}^{k,j}, \bar{Y}^{k,j}, \bar{Z}^{k,j}) = \overline{\text{solver}}[k](\bar{X}^{k,j-1}, \bar{Y}_{r_{k+1}}^{k,j}, f)$
- (3) return  $\bar{Y}_{r_{k+1}}^{k,J}$ .

Following (19), we let

$$\text{solver}[N](\xi) = g(\xi, [\xi]). \quad (21)$$

We first explain the initialization step. The basic idea is to set the backward component to 0 and then to solve the forward component as an approximation of the autonomous McKean-Vlasov diffusion process in which the backward entry is null. Of course, this may be solved by means of any standard method, but to make the notation shorten, we felt better to regard the underlying solver as a specific case of a forward-backward solver with null coefficients in the backward equation. We specify in the analysis below the conditions that this initial solver  $\overline{\text{solver}}[\cdot](\cdot, 0, 0)$  must satisfy.

It is also worth noting that each Picard iteration used to define the solver at level  $k$  calls the solver at level  $k+1$ . This is a typical feature of the way the continuation method manifests from the algorithmic point of view. In particular, the total complexity is of order  $O(J^N \mathfrak{K})$ , where  $\mathfrak{K}$  is the complexity of the solver  $\overline{\text{solver}}[\cdot](\cdot, \cdot)$ . In this regard, it must be stressed that, for a given length  $T$ ,  $N$  is fixed, regardless of the time step  $|\pi|$ . Also,  $J$  is intended to be rather small as the Picard iterations are expected to converge geometrically fast, see the numerical examples in Section 4 in which we choose  $J = 5$ . However, it must be noticed that the complexity increases exponentially fast when  $T$  tends to  $\infty$ , which is obviously the main drawback of this method. Again, we refer to Section 4 for numerical illustrations.

Useful notations. Throughout the paper,  $\|\cdot\|_p$  denotes the  $L^p$  norm on  $(\Omega, \mathcal{F}, \mathbb{P})$ . Also,  $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{P}})$  stands for a copy of  $(\Omega, \mathcal{F}, \mathbb{P})$ . It is especially useful to represent the Lions' derivative of a function of a probability measure and to distinguish the (somewhat artificial) space used for representing these derivatives from the (physical) space carrying the Wiener process. For a random variable  $X$  defined on  $(\Omega, \mathcal{F}, \mathbb{P})$ , we shall denote by  $\langle X \rangle$  its copy on  $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{P}})$ .

We shall use the notations  $C_\Lambda, c_\Lambda$  for constants only depending on  $\Lambda$  (and possibly on the dimension as well). They are allowed to increase from line to line. We shall use the notation  $C$  for constants not depending upon the discretization parameters. Again, they are allowed to increase from line to line. In most of the proofs, we shall just write  $C$  for  $C_\Lambda$ , even if we use the more precise notation  $C_\Lambda$  in the corresponding statement.

**2.2. A first analysis with no discretization error.** To conclude this section, we want to understand how the error propagates through the solvers used at different levels in the ideal case where the Picard iteration in (17) can be perfectly computed or equivalently

when the solver is given by  $\text{solver}[k](\cdot) = \text{picard}[k](\cdot)$ . For  $j \geq 1$ , we then denote by  $(\tilde{X}^{k,j}, \tilde{Y}^{k,j}, \tilde{Z}^{k,j})$ , the solution on  $[r_k, r_{k+1}]$  of (17).

The main result of the section, see Theorem 5, is an upper bound for the error when we use  $\text{picard}[\cdot](\cdot)$  to approximate  $\mathcal{U}$ . The proof of this theorem requires the following proposition, which gives a local error estimate for each level.

**Proposition 3.** *Let us define, for  $j \in \{1, \dots, J\}$ ,  $k \in \{1, \dots, N-1\}$ ,*

$$\Delta_k^j := \left\| \sup_{t \in [r_k, r_{k+1}]} (\tilde{Y}_t^{k,j} - \mathcal{U}(t, \tilde{X}_t^{k,j}, [\tilde{X}_t^{k,j}])) \right\|_2$$

then, there exist constants  $C_\Lambda, c_\Lambda$  such that, for  $\bar{\delta} := C_\Lambda \delta < c_\Lambda$ ,

$$\Delta_k^j \leq \bar{\delta}^j \Delta_k^0 + \sum_{\ell=1}^j \bar{\delta}^{\ell-1} e^{\bar{\delta}} \|\epsilon^{k+1}(\tilde{X}_{r_{k+1}}^{k,j-\ell})\|_2. \quad (22)$$

We recall that  $\epsilon^k(\xi)$  stands for the error term:

$$\epsilon^k(\xi) = \text{picard}[k](\xi) - \mathcal{U}(r_k, \xi, [\xi]), \quad \text{with } \epsilon^N(\xi) = 0.$$

**Remark 4.** *A careful inspection of the proof shows that, whenever  $\sigma$  depends on  $Y$  or  $b$  depends on  $Z$ , the same result holds true but with a constant  $C_\Lambda$  depending on  $N$ . As  $N$  is fixed in practice, this might still suffice to complete the analysis of the discretization scheme in that more general setting.*

**Proof.** We suppose that the full algorithm is initialized at some level  $k \in \{0, \dots, N-1\}$ , with an initial condition  $\xi \in L^2(\mathcal{F}_{r_k})$ . As the value of the index  $k$  is fixed throughout the proof, we will drop it in the notations  $(\tilde{X}^{k,j}, \tilde{Y}^{k,j}, \tilde{Z}^{k,j})$  and  $\Delta_k^j$ .

Applying Ito's formula for functions of a measure argument, see [11, 20], we have

$$\begin{aligned} d\mathcal{U}(t, \tilde{X}_t^j, [\tilde{X}_t^j]) &= \left( b(\tilde{X}_t^j, \tilde{Y}_t^j, [\tilde{X}_t^j, \tilde{Y}_t^j]) \cdot \partial_x \mathcal{U}(t, \tilde{X}_t^j, [\tilde{X}_t^j]) \right. \\ &\quad + \frac{1}{2} \text{Tr}[a(\tilde{X}_t^j, [\tilde{X}_t^j]) \partial_{xx}^2 \mathcal{U}(t, \tilde{X}_t^j, [\tilde{X}_t^j])] \\ &\quad + \hat{\mathbb{E}} \left[ b(\langle \tilde{X}_t^j \rangle, \langle \tilde{Y}_t^j \rangle, [\tilde{X}_t^j, \tilde{Y}_t^j]) \cdot \partial_\mu \mathcal{U}(t, \tilde{X}_t^j, [\tilde{X}_t^j]) \right] \\ &\quad + \hat{\mathbb{E}} \left[ \frac{1}{2} \text{Tr}[a(\langle \tilde{X}_t^j \rangle, [\tilde{X}_t^j]) \partial_v \partial_\mu \mathcal{U}(t, \tilde{X}_t^j, [\tilde{X}_t^j])] \right] + \partial_t \mathcal{U}(t, \tilde{X}_t^j, [\tilde{X}_t^j]) \Big) dt \\ &\quad + \partial_x \mathcal{U}(t, \tilde{X}_t^j, [\tilde{X}_t^j]) \cdot (\sigma(\tilde{X}_t^j, [\tilde{X}_t^j]) dW_t). \end{aligned}$$

Expressing the integral in (1) as expectations on  $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{P}})$  and combining with (1) and (17), we obtain

$$\begin{aligned} d[\tilde{Y}_t^j - \tilde{Y}_t^j] &= \left( \{b(\tilde{X}_t^j, \tilde{Y}_t^j, [\tilde{X}_t^j, \tilde{Y}_t^j]) - b(\tilde{X}_t^j, \tilde{Y}_t^j, [\tilde{X}_t^j, \tilde{Y}_t^j])\} \cdot \partial_x \mathcal{U}(t, \tilde{X}_t^j, [\tilde{X}_t^j]) \right. \\ &\quad + \hat{\mathbb{E}} \left[ \{b(\langle \tilde{X}_t^j \rangle, \langle \tilde{Y}_t^j \rangle, [\tilde{X}_t^j, \tilde{Y}_t^j]) - b(\langle \tilde{X}_t^j \rangle, \langle \tilde{Y}_t^j \rangle, [\tilde{X}_t^j, \tilde{Y}_t^j])\} \cdot \partial_\mu \mathcal{U}(t, \tilde{X}_t^j, [\tilde{X}_t^j]) \right] \\ &\quad + f(\tilde{X}_t^j, \tilde{Y}_t^j, \tilde{Z}_t^j, [\tilde{X}_t^j, \tilde{Y}_t^j]) - f(\tilde{X}_t^j, \tilde{Y}_t^j, \tilde{Z}_t^j, [\tilde{X}_t^j, \tilde{Y}_t^j]) \Big) dt + [\tilde{Z}_t^j - \tilde{Z}_t^j] \cdot dW_t, \end{aligned}$$

where  $\tilde{Y}_t^j := \mathcal{U}(t, \tilde{X}_t^j, [\tilde{X}_t^j])$  and  $\tilde{Z}_t^j := \partial_x u(t, \tilde{X}_t^j, [\tilde{X}_t^j]) \sigma(\tilde{X}_t^j, [\tilde{X}_t^j])$ . Observe that this argument is reminiscent of the four-step scheme, see [32].

Using standard arguments from BSDE theory and (H0)–(H1), we then compute

$$\begin{aligned} \Delta^j &\leq e^{C\delta} \|\mathcal{U}(r_{k+1}, \tilde{X}_{r_{k+1}}^j, [\tilde{X}_{r_{k+1}}^j]) - \tilde{Y}_{r_{k+1}}^j\|_2 \\ &\leq e^{C\delta} \left( \|\epsilon^{k+1}(\tilde{X}_{r_{k+1}}^{j-1})\|_2 + \|\mathcal{U}(r_{k+1}, \tilde{X}_{r_{k+1}}^j, [\tilde{X}_{r_{k+1}}^j]) - \mathcal{U}(r_{k+1}, \tilde{X}_{r_{k+1}}^{j-1}, [\tilde{X}_{r_{k+1}}^{j-1}])\|_2 \right), \end{aligned}$$

recalling  $\tilde{Y}_{r_{k+1}}^j = \text{picard}[k+1](\tilde{X}_{r_{k+1}}^{j-1})$  and (16). Since  $\mathcal{U}$  is Lipschitz, we have

$$\Delta^j \leq e^{C\delta} \left( \|\epsilon^{k+1}(\tilde{X}_{r_{k+1}}^{j-1})\|_2 + 2L \|\tilde{X}_{r_{k+1}}^j - \tilde{X}_{r_{k+1}}^{j-1}\|_2 \right). \quad (23)$$

We also have that

$$\begin{aligned} \tilde{X}_t^j - \tilde{X}_t^{j-1} &= \int_{r_k}^t \{b(\tilde{X}_s^j, \tilde{Y}_s^j, [\tilde{X}_s^j, \tilde{Y}_s^j]) - b(\tilde{X}_s^{j-1}, \tilde{Y}_s^{j-1}, [\tilde{X}_s^{j-1}, \tilde{Y}_s^{j-1}])\} ds \\ &\quad + \int_{r_k}^t \{\sigma(\tilde{X}_s^j, [\tilde{X}_s^j]) - \sigma(\tilde{X}_s^{j-1}, [\tilde{X}_s^{j-1}])\} dW_s. \end{aligned}$$

Using usual arguments (squaring, taking the sup, using B urkholder-Davis-Gundy inequality), we get, since  $b$  and  $\sigma$  are Lipschitz continuous,

$$\left\| \sup_{t \in [r_k, r_{k+1}]} |\tilde{X}_t^j - \tilde{X}_t^{j-1}| \right\|_2 \leq C \left( \delta \left\| \sup_{t \in [r_k, r_{k+1}]} |\tilde{Y}_t^j - \tilde{Y}_t^{j-1}| \right\|_2 + \delta^{\frac{1}{2}} \left\| \sup_{t \in [r_k, r_{k+1}]} |\tilde{X}_t^j - \tilde{X}_t^{j-1}| \right\|_2 \right).$$

Observing that

$$\begin{aligned} |\tilde{Y}_s^j - \tilde{Y}_s^{j-1}| &\leq |\tilde{Y}_s^j - \mathcal{U}(s, \tilde{X}_s^j, [\tilde{X}_s^j])| + |\tilde{Y}_s^{j-1} - \mathcal{U}(s, \tilde{X}_s^{j-1}, [\tilde{X}_s^{j-1}])| \\ &\quad + \Lambda(|\tilde{X}_s^{j-1} - \tilde{X}_s^j| + \|\tilde{X}_s^{j-1} - \tilde{X}_s^j\|_2), \end{aligned}$$

we obtain, for  $\delta$  small enough,

$$\left\| \sup_{t \in [r_k, r_{k+1}]} |\tilde{X}_t^j - \tilde{X}_t^{j-1}| \right\|_2 \leq C\delta(\Delta^j + \Delta^{j-1}). \quad (24)$$

Combining the previous inequality with (23), we obtain, for  $\delta$  small enough,

$$\Delta^j \leq e^{C\delta} \|\epsilon^{k+1}(\tilde{X}_{r_{k+1}}^{j-1})\|_2 + C\delta\Delta^{j-1},$$

which by induction leads to

$$\Delta^j \leq (C\delta)^j \Delta^0 + \sum_{\ell=1}^j (C\delta)^{\ell-1} e^{C\delta} \|\epsilon^{k+1}(\tilde{X}_{r_{k+1}}^{j-\ell})\|_2,$$

and concludes the proof.  $\square$

We now state the main result of this section, which explains how the local error induced by the fact that the Picard iteration is stopped at rank  $J$  propagates through the various levels  $k = N - 1, \dots, 0$ .

**Theorem 5.** *We can find two constants  $C_\Lambda, c_\Lambda > 0$  and a continuous non-decreasing function  $\mathfrak{B} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  matching 0 in 0, only depending on  $\Lambda$ , such that, for  $\bar{\delta} := C_\Lambda \delta < \min(c_\Lambda, 1)$  and  $\beta \geq \mathfrak{B}(\bar{\delta})$  satisfying*

$$(J-1)\Lambda\bar{\delta}^J \frac{e^{\beta C_\Lambda T}}{e^{\beta\bar{\delta}} - 1} \leq 1 \quad (25)$$

where  $J$  is the number of Picard iterations in a period, it holds, for any period  $k \in \{0, \dots, N\}$  and  $\xi \in L^2(\mathcal{F}_{r_k})$ ,

$$\|\text{solver}[k](\xi) - \mathcal{U}(r_k, \xi, [\xi])\|_2 \leq \Lambda \frac{e^{\beta C_\Lambda T}}{\beta} \bar{\delta}^{J-1} (1 + \|P_{r_k, T}^*(\xi)\|_2), \quad (26)$$

where  $P_{r_k, t}(\xi)$  is the solution at time  $t$  of the stochastic differential equation

$$dX_s^0 = b(X_s^0, 0, [X_s^0, 0])ds + \sigma(X_s^0, [X_s^0])dW_s,$$

with  $X_{r_k}^0 = \xi$  as initial condition, and  $P_{r_k, t}^*(\xi) = \sup_{s \in [r_k, t]} |P_{r_k, s}(\xi)|$ .

Of course, it is absolutely straightforward to bound  $\|P_{r_k, T}^*(\xi)\|_2$  by  $C(1 + \|\xi\|_2)$  in (26). Theorem 5 may be restated accordingly, but the form used in the statement is more faithful to the spirit of the proof.

**Proof.** We prove the claim by an induction argument. We show below that for all  $k \in \{0, \dots, N\}$ ,

$$\|\epsilon^k(\xi)\|_2 = \|\text{solver}[k](\xi) - \mathcal{U}(r_k, \xi, [\xi])\|_2 \leq \theta_k \left(1 + \|P_{r_k, T}^*(\xi)\|_2\right), \quad (27)$$

where  $(\theta_k)_{k=0, \dots, N-1}$  is defined by the following backward induction:  $\theta_N := 0$ , recall (19), and for  $k \in \{0, \dots, N-1\}$ ,

$$\theta_k := \Lambda \bar{\delta}^J + e^{\beta \bar{\delta}} \theta_{k+1}, \quad (28)$$

where  $\beta$  is such that

$$\left(\gamma + \gamma \bar{\delta} e^{\gamma \bar{\delta}} \left(\gamma + \frac{\Lambda}{1 - \bar{\delta}}\right)\right) \leq e^{\beta \bar{\delta}}, \quad \text{with } \gamma := \frac{e^{\bar{\delta}}}{1 - \bar{\delta}}. \quad (29)$$

With this definition, we have, for all  $k \in \{0, \dots, N\}$ ,

$$\theta_k = \Lambda \bar{\delta}^J \sum_{j=0}^{N-k-1} e^{j\beta \bar{\delta}} \leq \Lambda \bar{\delta}^J \frac{e^{\beta C_\Lambda T}}{e^{\beta \bar{\delta}} - 1}, \quad (30)$$

which gives the expected result.

We now prove (27). Observe that it is obviously true for the last step  $N$ . Assume now that it holds true at step  $k+1$ , for  $k < N$ , and that (30) holds true for  $\theta_{k+1}$ . Then, using (25), we have

$$\theta_{k+1}^j \leq 1, \quad \text{for all } j \leq J-1. \quad (31)$$

From Proposition 3, we have

$$\Delta_k^j \leq \bar{\delta}^j \Delta_k^0 + \sum_{\ell=1}^j \bar{\delta}^{\ell-1} e^{\bar{\delta}} \|\epsilon^{k+1}(\tilde{X}_{r_{k+1}}^{k, j-\ell})\|_2. \quad (32)$$

Using the induction hypothesis (27), we compute

$$\Delta_k^j \leq \bar{\delta}^j \Delta_k^0 + \frac{e^{\bar{\delta}}}{1 - \bar{\delta}} \theta_{k+1} + e^{\bar{\delta}} \theta_{k+1} \sum_{\ell=0}^{j-1} \bar{\delta}^{j-1-\ell} \|P_{r_{k+1}, T}^*(\tilde{X}_{r_{k+1}}^{k, \ell})\|_2. \quad (33)$$

We study the last sum. Observe that for  $\ell \in \{1, \dots, j-1\}$ ,

$$\|P_{r_{k+1}, T}^*(\tilde{X}_{r_{k+1}}^{k, \ell})\|_2 \leq \|P_{r_{k+1}, T}^*(\tilde{X}_{r_{k+1}}^{k, 0})\|_2 + \sum_{i=1}^{\ell} \|P_{r_{k+1}, T}^*(\tilde{X}_{r_{k+1}}^{k, i}) - P_{r_{k+1}, T}^*(\tilde{X}_{r_{k+1}}^{k, i-1})\|_2.$$

We observe that  $P_{r_{k+1},t}(\tilde{X}_{r_{k+1}}^{k,0}) = P_{r_k,t}(\tilde{X}_{r_k}^{k,0}) = P_{r_k,t}(\xi)$ , for  $t \in [r_{k+1}, T]$ . Hence,  $P_{r_{k+1},T}^*(\tilde{X}_{r_{k+1}}^{k,0}) \leq P_{r_k,T}^*(\xi)$ . Also, it is well-checked that there exists a constant  $C_\Lambda$  such that each  $P_{t,T}^*$  is  $C_\Lambda$ -Lipschitz continuous from  $L^2(\mathcal{F}_t)$  into  $L^2(\mathcal{F}_T)$ . Then,

$$\sum_{\ell=0}^{j-1} \bar{\delta}^{j-1-\ell} \|P_{r_k,T}^*(\tilde{X}_{r_{k+1}}^{k,\ell})\|_2 \leq C_\Lambda \sum_{\ell=1}^{j-1} \bar{\delta}^{j-1-\ell} \sum_{i=1}^{\ell} \|\tilde{X}_{r_{k+1}}^{k,i} - \tilde{X}_{r_{k+1}}^{k,i-1}\|_2 + \sum_{\ell=0}^{j-1} \bar{\delta}^\ell \|P_{r_k,T}^*(\xi)\|_2.$$

Using (24) in the proof of Proposition 3 and changing the definition of  $\bar{\delta}$ , we obtain

$$\sum_{\ell=0}^{j-1} \bar{\delta}^{j-1-\ell} \|P_{r_k,T}^*(\tilde{X}_{r_{k+1}}^{k,\ell})\|_2 \leq \bar{\delta} \sum_{i=1}^{j-1} (\Delta_k^i + \Delta_k^{i-1}) \sum_{\ell=i}^{j-1} \bar{\delta}^{j-1-\ell} + \sum_{\ell=0}^{j-1} \bar{\delta}^\ell \|P_{r_k,T}^*(\xi)\|_2. \quad (34)$$

Observing that, for all  $i \leq j-1$ ,  $\sum_{\ell=i}^{j-1} \bar{\delta}^{j-1-\ell} \leq \frac{1}{1-\bar{\delta}}$ , we get

$$\sum_{\ell=0}^{j-1} \bar{\delta}^{j-\ell} \|P_{r_k,T}^*(\tilde{X}_{r_{k+1}}^{k,\ell})\|_2 \leq \frac{2\bar{\delta}}{1-\bar{\delta}} \mathcal{S}_k^{j-1} + \frac{1}{1-\bar{\delta}} \|P_{r_k,T}^*(\xi)\|_2, \quad (35)$$

where  $\mathcal{S}_k^n := \sum_{i=0}^n \Delta_k^i$ . Inserting the previous estimate into (33) and changing  $\bar{\delta}$  into  $2\bar{\delta}$ , we obtain

$$\Delta_k^j \leq \bar{\delta}^j \Delta_k^0 + \frac{e^{\bar{\delta}}}{1-\bar{\delta}} \theta_{k+1} (1 + \|P_{r_k,T}^*(\xi)\|_2) + \theta_{k+1} \frac{\bar{\delta} e^{\bar{\delta}}}{1-\bar{\delta}} \mathcal{S}_k^{j-1}. \quad (36)$$

We note that  $\Delta_k^0 \leq \Lambda(1 + \|P_{r_k,T}^*(\xi)\|_2)$ . Recalling  $\gamma$  in (29), equation (36) leads to

$$\Delta_k^j \leq a_j + \gamma \theta_{k+1} \bar{\delta} \mathcal{S}_k^{j-1}. \quad (37)$$

where we set  $a_j := (\Lambda \bar{\delta}^j + \gamma \theta_{k+1})(1 + \|P_{r_k,T}^*(\xi)\|_2)$ . We have

$$\mathcal{S}_k^j - \mathcal{S}_k^{j-1} = \Delta_k^j \leq a_j + \gamma \theta_{k+1} \bar{\delta} \mathcal{S}_k^{j-1},$$

and then

$$\mathcal{S}_k^j \leq e^{\gamma \theta_{k+1} \bar{\delta} j} \mathcal{S}_k^0 + \sum_{\ell=1}^j e^{\gamma \theta_{k+1} \bar{\delta} (j-\ell)} a_\ell. \quad (38)$$

We compute

$$\sum_{\ell=1}^j a_\ell \leq \left( j \gamma \theta_{k+1} + \frac{\Lambda \bar{\delta}}{1-\bar{\delta}} \right) (1 + \|P_{r_k,T}^*(\xi)\|_2),$$

which combined with the properties (31) and (38) leads to, for all  $j \leq J-1$ ,

$$\mathcal{S}_k^j \leq e^{\gamma \bar{\delta}} \left( \gamma + \frac{\Lambda}{1-\bar{\delta}} \right) (1 + \|P_{r_k,T}^*(\xi)\|_2),$$

where we recall that  $\mathcal{S}_k^0 = \Delta_k^0 \leq \Lambda(1 + \|P_{r_k,T}^*(\xi)\|_2)$ . We insert the previous inequality into (37) for  $j = J$  and get

$$\Delta_k^J \leq \left( \Lambda \bar{\delta}^J + \left( \gamma + \gamma \bar{\delta} e^{\gamma \bar{\delta}} \left( \gamma + \frac{\Lambda}{1-\bar{\delta}} \right) \right) \theta_{k+1} \right) (1 + \|P_{r_k,T}^*(\xi)\|_2).$$

Using (29), this rewrites

$$\Delta_k^J \leq \left( \Lambda \bar{\delta}^J + e^{\beta \bar{\delta}} \theta_{k+1} \right) (1 + \|P_{r_k,T}^*(\xi)\|_2),$$

and validates (28) and thus (30). We then obviously have that (27) holds true.  $\square$

## 3. CONVERGENCE ANALYSIS

**3.1. Error analysis in the generic case.** We now study the convergence of a generic implementable solver  $\text{solver}[\cdot](\cdot)$ , based upon the local solver  $\overline{\text{solver}}[\cdot](\cdot, \cdot)$  as described above, as long as the output of the local solver  $\overline{\text{solver}}[k](\cdot, \cdot)$  satisfies some conditions, which are shown to be true for Example 2.

In order to define the required assumption, we use the same letters  $\Lambda$  and  $\alpha$  as in (H0) and (H1), except that, without any loss of generality, we assume that  $\alpha$  is greater than 1. For the same coefficients as in the equation (6), and in particular for the same driver  $f$ , we then ask  $\overline{\text{solver}}[k](\cdot, \cdot)$  to satisfy the following three conditions.

$$(A1) \quad \sup_{t \in \pi^k} \|\mathcal{U}(t, \bar{X}_t, [\bar{X}_t]) - \bar{Y}_t\|_{2\alpha} \leq e^{\Lambda\delta} \|\mathcal{U}(r_{k+1}, \bar{X}_{r_{k+1}}, [\bar{X}_{r_{k+1}}]) - \bar{Y}_{r_{k+1}}\|_{2\alpha} \\ + \Lambda \max_{j_k \leq i < j_{k+1}} \|\bar{X}_{t_i} - \bar{\mathfrak{X}}_{t_i}\|_{2\alpha} + \mathcal{D}^1(|\pi|) + \mathcal{D}^2(|\pi|)(1 + \|\xi\|_{2\alpha}^\alpha),$$

$$(A2) \quad \sup_{t \in \pi^k} \|\bar{X}_t - \bar{X}'_t\|_{2\alpha} \leq \Lambda\delta \sup_{t \in \pi^k} \|\bar{Y}_t - \bar{Y}'_t\|_{2\alpha},$$

$$(A3) \quad \|\mathcal{U}(r_{k+1}, \bar{X}_{r_{k+1}}, [\bar{X}_{r_{k+1}}]) - \bar{Y}_{r_{k+1}}\|_{2\alpha}^\alpha \leq \Lambda \|\mathcal{U}(r_{k+1}, \bar{X}_{r_{k+1}}, [\bar{X}_{r_{k+1}}]) - \bar{Y}_{r_{k+1}}\|_{2\alpha},$$

where  $(\bar{X}, \bar{Y}, \bar{Z}) := \overline{\text{solver}}[k](\bar{\mathfrak{X}}, \eta, f)$ , for  $f$  as before, and  $(\bar{X}', \bar{Y}', \bar{Z}') := \overline{\text{solver}}[k](\bar{\mathfrak{X}}', \eta', f')$ , for another  $f'$  either equal to  $f$  or 0, are two output values of  $\overline{\text{solver}}[\cdot](\cdot, \cdot)$  associated to two input processes  $\bar{\mathfrak{X}}, \bar{\mathfrak{X}}'$ , with the same initial condition  $\bar{\mathfrak{X}}_{r_k} = \bar{\mathfrak{X}}'_{r_k} = \xi$ , and to two different terminal conditions  $\eta$  and  $\eta'$ . For  $i \in \{1, 2\}$ , the function  $\mathcal{D}^i : [0, \infty) \rightarrow [0, \infty)$  is a discretization error associated to the use of the grid  $\pi$ , which satisfies  $\lim_{h \downarrow 0} \mathcal{D}^i(h) = 0$ . Importantly, both  $\mathcal{D}^1$  and  $\mathcal{D}^2$  are independent of  $\bar{\mathfrak{X}}, \bar{\eta}, J$  and  $N$ .

In full analogy with the discussion right below Theorem 5, we shall also need some conditions on the solver  $\overline{\text{solver}}[k](\cdot, 0, 0)$  used to initialize the algorithm at each step. Following the definition of  $(P_{r_k, t})_{0 \leq t \leq T}$  introduced in the statement of Theorem 5, we let by induction, for a given  $k \in \{0, \dots, N-1\}$ :

$$P_{r_k, t}(\xi) = (\overline{\text{solver}}[k](\xi, 0, 0))_t^1, \quad t \in \pi^k, \quad \xi \in L^2(\mathcal{F}_{r_k}),$$

where we recall that  $(\overline{\text{solver}}[k](\xi, 0, 0))_t^1$  is the forward component of the algorithm's output, and, for  $k \leq N-2$ ,

$$P_{r_k, t}(\xi) = P_{r_\ell, t}(P_{r_k, r_\ell}(\xi)), \quad t \in \pi^\ell, \quad k < \ell \leq N-1,$$

and then  $P_{r_k, T}^*(\xi) = \max_{s \in \pi, s \in [r_k, T]} |P_{r_k, s}(\xi)|$ , for  $\xi \in L^2(\mathcal{F}_{r_k})$ . It then makes sense to assume

$$(A4) \quad \|P_{r_k, T}^*(\xi) - P_{r_k, T}^*(\xi')\|_{2\alpha} \leq \Lambda \|\xi - \xi'\|_{2\alpha}$$

$$(A5) \quad \|P_{r_k, T}^*(\xi)\|_{2\alpha} \leq \Lambda(1 + \|\xi\|_{2\alpha})$$

where  $\xi, \xi' \in L^{2\alpha}(\mathcal{F}_{r_k})$  and  $k \in \{0, \dots, N-1\}$ .

**Remark 6.** *The main challenging assumption (and maybe the most surprising one) is (A3). It is obviously satisfied when  $\alpha = 1$  as long as  $\Lambda$  is assumed to be greater than 1. We refer to [13] and [17, Chap. 12] for sets of conditions under which this is indeed true. When  $\alpha > 1$ , Assumption (A3) is checked provided we have an a priori bound on  $\|\mathcal{U}(r_{k+1}, \bar{X}_{r_{k+1}}, [\bar{X}_{r_{k+1}}]) - \bar{Y}_{r_{k+1}}\|_{2\alpha}$ , see Lemma 10. This permits to invoke the result proven in our previous paper [20], which holds true in a weaker setting than the solvability results obtained in [13] and [17, Chap. 12].*

**Theorem 7.** *We can find two constants  $C_\Lambda, c_\Lambda > 0$  and a continuous non-decreasing function  $\mathfrak{B} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  matching 0 in 0, only depending on  $\Lambda$ , such that, for  $\bar{\delta} := C_\Lambda \delta < \min(c_\Lambda, 1)$  and  $\beta \geq \mathfrak{B}(\bar{\delta})$  satisfying*

$$(J-1) \left( \Lambda \bar{\delta}^J + e^{\beta \bar{\delta}} \mathcal{D}^2(|\pi|) \right) \frac{e^{\beta C_\Lambda T}}{e^{\beta \bar{\delta}} - 1} \leq 1, \quad (39)$$

where  $J$  is the number of Picard iterations in a period, it holds, for any period  $k \in \{0, \dots, N\}$  and  $\xi \in L^2(\mathcal{F}_{r_k})$ ,

$$\begin{aligned} & \|\text{solver}[k](\xi) - \mathcal{U}(r_k, \xi, [\xi])\|_{2\alpha} \\ & \leq C (\bar{\delta}^{J-1} + (N-k) \mathcal{D}^2(|\pi|)) (1 + \|\xi\|_{2\alpha}^\alpha) + C(N-k) \mathcal{D}^1(|\pi|), \end{aligned}$$

for a constant  $C$  independent of the discretization parameters.

**Proof.** The proof will follow closely the proof of Theorem 5 but we now have to take into account the discretization error. We will first show that for all  $k = \{0, \dots, N\}$ ,

$$\|\epsilon^k(\xi)\|_{2\alpha} \leq \theta_k (1 + \|\mathbf{P}_{r_k, T}^*(\xi)\|_{2\alpha}^\alpha) + \vartheta_k \mathcal{D}^1(|\pi|), \quad (40)$$

where

$$\epsilon^k(\xi) = \text{solver}[k](\xi) - \mathcal{U}(r_k, \xi, [\xi]),$$

and  $(\theta_k, \vartheta_k)_{k=0, \dots, N}$  is defined by the following backward induction:  $(\theta_N, \vartheta_N) := (0, 0)$ , recall (21), and for  $k \in \{0, \dots, N-1\}$ ,

$$\theta_k := \Lambda \bar{\delta}^J + e^{\beta \bar{\delta}} \{\theta_{k+1} + \mathcal{D}^2(|\pi|)\} \quad \text{and} \quad \vartheta_k := e^{\beta \bar{\delta}} (\vartheta_{k+1} + 1), \quad (41)$$

$\beta$  being defined as in equation (48).

Assume for a while that this holds true. Then, we have, for all  $k = \{0, \dots, N-1\}$ ,

$$\theta_k \leq (\Lambda \bar{\delta}^J + e^{\beta \bar{\delta}} \mathcal{D}^2(|\pi|)) \frac{e^{\beta \bar{\delta}(N-k)} - 1}{e^{\beta \bar{\delta}} - 1} \quad \text{and} \quad \vartheta_k \leq e^{\beta \bar{\delta}} \frac{e^{\beta(N-k)\bar{\delta}} - 1}{e^{\beta \bar{\delta}} - 1}. \quad (42)$$

Recalling that  $\bar{\delta}N = C_\Lambda T$ , we get the announced inequality.

We now prove (40). Obviously, it holds true for the last step  $N$ . Assume now that it is true at step  $k+1$ , for  $k < N$  and that (42) holds for  $\theta_{k+1}$  and  $\vartheta_{k+1}$ .

In particular, using (39), we observe that

$$\theta_{k+1} j \leq 1, \quad \text{for all } j \leq J-1. \quad (43)$$

*First Step.* For  $j \in \{0, \dots, J\}$ , let

$$\bar{\Delta}_k^j := \sup_{t \in \pi^k} \|\mathcal{U}(t, \bar{X}_t^{k,j}, [\bar{X}_t^{k,j}]) - \bar{Y}_t^{k,j}\|_{2\alpha}.$$

Under (A1)–(A2), we will prove in this first step an upper bound for  $\bar{\Delta}_k^j$ , for  $j = 1, \dots, J$ , similar to the one obtained in Proposition 3.

Using (A1) and (H1) and the fact that

$$\bar{Y}_{r_{k+1}}^{k,j} = \mathcal{U}(r_{k+1}, \bar{X}_{r_{k+1}}^{k,j-1}, [\bar{X}_{r_{k+1}}^{k,j-1}]) + \epsilon^{k+1}(\bar{X}_{r_{k+1}}^{k,j-1}),$$

we observe that

$$\begin{aligned}
 \bar{\Delta}_k^j &\leq e^{\Lambda\delta} \left[ \|\mathcal{U}(r_{k+1}, \bar{X}_{r_{k+1}}^{k,j}, [\bar{X}_{r_{k+1}}^{k,j}]) - \mathcal{U}(r_{k+1}, \bar{X}_{r_{k+1}}^{k,j-1}, [\bar{X}_{r_{k+1}}^{k,j-1}])\|_{2\alpha} \right. \\
 &\quad \left. + \|\epsilon^{k+1}(\bar{X}_{r_{k+1}}^{k,j-1})\|_{2\alpha} \right] \\
 &\quad + \Lambda \max_{j_k \leq i < j_{k+1}} \|\bar{X}_{t_i}^{k,j} - \bar{X}_{t_i}^{k,j-1}\|_{2\alpha} + \mathcal{D}^1(|\pi|) + \mathcal{D}^2(|\pi|)(1 + \|\xi\|_{2\alpha}^\alpha) \\
 &\leq C_\Lambda \max_{t \in \pi^k} \|\bar{X}_t^{k,j} - \bar{X}_t^{k,j-1}\|_{2\alpha} + e^{\Lambda\delta} \|\epsilon^{k+1}(\bar{X}_{r_{k+1}}^{k,j-1})\|_{2\alpha} \\
 &\quad + \mathcal{D}^1(|\pi|) + \mathcal{D}^2(|\pi|)(1 + \|\mathbf{P}_{r_k, T}^*(\xi)\|_{2\alpha}^\alpha).
 \end{aligned} \tag{44}$$

Using (A2), we also have

$$\begin{aligned}
 \sup_{t \in \pi^k} \|\bar{X}_t^{k,j} - \bar{X}_t^{k,j-1}\|_{2\alpha} &\leq \Lambda\delta \sup_{t \in \pi^k} \left[ \|\bar{Y}_t^{k,j} - \mathcal{U}(t, \bar{X}_t^{k,j}, [\bar{X}_t^{k,j}])\|_{2\alpha} + \Lambda \|\bar{X}_t^{k,j} - \bar{X}_t^{k,j-1}\|_{2\alpha} \right. \\
 &\quad \left. + \|\mathcal{U}(t, \bar{X}_t^{k,j-1}, [\bar{X}_t^{k,j-1}]) - \bar{Y}_t^{k,j-1}\|_{2\alpha} \right] \\
 &\leq C_\Lambda \delta \left( \bar{\Delta}_k^j + \bar{\Delta}_k^{j-1} \right),
 \end{aligned}$$

for  $\delta$  small enough. Inserting the previous inequality in (44), we get

$$\begin{aligned}
 \bar{\Delta}_k^j &\leq C_\Lambda \delta \bar{\Delta}_k^{j-1} + e^{C_\Lambda \delta} \|\epsilon^{k+1}(\bar{X}_{r_{k+1}}^{k,j-1})\|_{2\alpha} + \mathcal{D}^1(|\pi|) + \mathcal{D}^2(|\pi|)(1 + \|\mathbf{P}_{r_k, T}^*(\xi)\|_{2\alpha}^\alpha), \\
 &\leq \bar{\delta}^j \bar{\Delta}_k^0 + e^{\bar{\delta}} \sum_{\ell=0}^{j-1} \bar{\delta}^\ell \|\epsilon^{k+1}(\bar{X}_{r_{k+1}}^{k,j-1-\ell})\|_{2\alpha} + \frac{\mathcal{D}^1(|\pi|)}{1-\bar{\delta}} + \frac{\mathcal{D}^2(|\pi|)}{1-\bar{\delta}} (1 + \|\mathbf{P}_{r_k, T}^*(\xi)\|_{2\alpha}^\alpha),
 \end{aligned}$$

with  $\bar{\delta} := C_\Lambda \delta$ . We note that compared to (22), there is a new term, namely  $(\mathcal{D}^1(|\pi|) + \mathcal{D}^2(|\pi|)(1 + \|\mathbf{P}_{r_k, T}^*(\xi)\|_{2\alpha}^\alpha))/(1 - \bar{\delta})$ , which is due to the discretization.

*Second Step.* Using (40) at the previous step  $k+1$  and noting that  $\bar{\Delta}_k^0 \leq \Lambda(1 + \|\mathbf{P}_{r_k, T}^*(\xi)\|_{2\alpha}) \leq 2\Lambda(1 + \|\mathbf{P}_{r_k, T}^*(\xi)\|_{2\alpha}^\alpha)$ , we claim that

$$\begin{aligned}
 \bar{\Delta}_k^j &\leq (2\Lambda\bar{\delta}^j + \gamma\mathcal{D}^2(|\pi|)) (1 + \|\mathbf{P}_{r_k, T}^*(\xi)\|_{2\alpha}^\alpha) + \gamma(\vartheta_{k+1} + 1)\mathcal{D}^1(|\pi|) \\
 &\quad + e^{\bar{\delta}} \theta_{k+1} \sum_{\ell=0}^{j-1} \bar{\delta}^{j-1-\ell} \left( 1 + \|\mathbf{P}_{r_{k+1}, T}^*(\bar{X}_{r_{k+1}}^{k,\ell})\|_{2\alpha}^\alpha \right),
 \end{aligned} \tag{45}$$

where  $\gamma := e^{\bar{\delta}}/(1 - \bar{\delta})$ .

This corresponds to equation (33) adapted to our context. By (A2), we have, for  $\ell \leq J-1$ ,

$$\|\mathbf{P}_{r_{k+1}, T}^*(\bar{X}_{r_{k+1}}^{k,\ell}) - \mathbf{P}_{r_{k+1}, T}^*(\bar{X}_{r_{k+1}}^{k,0})\|_{2\alpha} \leq C_\Lambda \sup_{t \in \pi^k} \|\bar{X}_t^{k,\ell} - \bar{X}_t^{k,0}\|_{2\alpha}. \tag{46}$$

Using (A4), we then compute, recalling that  $\bar{Y}^{k,0} = 0$ ,

$$\begin{aligned}
 \sup_{t \in \pi^k} \|\bar{X}_t^{k,\ell} - \bar{X}_t^{k,0}\|_{2\alpha} &\leq \Lambda\delta \sup_{t \in \pi^k} \|\bar{Y}_t^{k,\ell}\|_{2\alpha} \\
 &\leq \Lambda\delta \left( \bar{\Delta}_k^\ell + \Lambda \sup_{t \in \pi^k} \|\bar{X}_t^{k,\ell} - \bar{X}_t^{k,0}\|_{2\alpha} + \Lambda(1 + \|\xi\|_{2\alpha}^\alpha) \right) \\
 &\leq C_\Lambda \delta \bar{\Delta}_k^\ell + C_\Lambda \delta (1 + \|\xi\|_{2\alpha}^\alpha),
 \end{aligned}$$

where for the last inequality we used the fact that  $\delta$  is small enough. Observing that  $\|\xi\|_{2\alpha} \leq \|\mathbf{P}_{r_k, T}^*(\xi)\|_{2\alpha}$  and combining the previous inequality with (46), we obtain

$$\|\mathbf{P}_{r_{k+1}, T}^*(\bar{X}_{r_{k+1}}^{k, \ell}) - \mathbf{P}_{r_{k+1}, T}^*(\bar{X}_{r_{k+1}}^{k, 0})\|_{2\alpha} \leq C_\Lambda \delta \bar{\Delta}_k^\ell + C_\Lambda \delta (1 + \|\mathbf{P}_{r_k, T}^*(\xi)\|_{2\alpha}).$$

So that, by using the fact that  $\mathbf{P}_{r_{k+1}, T}^*(\bar{X}_{r_{k+1}}^{k, 0}) \leq \mathbf{P}_{r_k, T}^*(\xi)$  together with a convexity argument,

$$\begin{aligned} \|\mathbf{P}_{r_{k+1}, T}^*(\bar{X}_{r_{k+1}}^{k, \ell})\|_{2\alpha}^\alpha &\leq \left( C_\Lambda \delta \bar{\Delta}_k^\ell + (1 + C_\Lambda \delta) (1 + \|\mathbf{P}_{r_k, T}^*(\xi)\|_{2\alpha}) \right)^\alpha, \\ &\leq (1 + 2C_\Lambda \delta)^{\alpha-1} \left( C_\Lambda \delta (\bar{\Delta}_k^\ell)^\alpha + (1 + C_\Lambda \delta) \|\mathbf{P}_{r_k, T}^*(\xi)\|_{2\alpha}^\alpha \right), \end{aligned}$$

Appealing to (A3) and redefining  $\bar{\delta}$ , we get

$$\|\mathbf{P}_{r_{k+1}, T}^*(\bar{X}_{r_{k+1}}^{k, \ell})\|_{2\alpha}^\alpha \leq \bar{\delta} \bar{\Delta}_k^\ell + e^{\bar{\delta}} (1 + \|\mathbf{P}_{r_k, T}^*(\xi)\|_{2\alpha}^\alpha),$$

which may be rewritten as

$$\sum_{\ell=0}^{j-1} \bar{\delta}^{j-1-\ell} \|\mathbf{P}_{r_{k+1}, T}^*(\bar{X}_{r_{k+1}}^{k, \ell})\|_{2\alpha}^\alpha \leq \bar{\delta} \sum_{\ell=1}^{j-1} \bar{\delta}^{j-1-\ell} \bar{\Delta}_k^\ell + \frac{e^{\bar{\delta}}}{1 - \bar{\delta}} (1 + \|\mathbf{P}_{r_k, T}^*(\xi)\|_{2\alpha}^\alpha).$$

Recalling the notation  $\gamma = e^{\bar{\delta}}/(1 - \bar{\delta})$  and letting  $\bar{\mathcal{S}}_k^n := \sum_{i=0}^n \bar{\delta}^{n-i} \bar{\Delta}_k^i$ , we obtain a new version of (37), namely

$$\bar{\Delta}_k^j \leq \Lambda \bar{\delta}^j \left( \frac{1}{2} + \|\mathbf{P}_{r_k, T}^*(\xi)\|_{2\alpha}^\alpha \right) + \bar{a} + \theta_{k+1} \gamma \bar{\delta} \bar{\mathcal{S}}_k^{j-1}, \quad (47)$$

where we changed the constant  $2\Lambda$  in (45) into  $\frac{1}{2}\Lambda$  as we changed the value of  $\bar{\delta}$ , and where we put

$$\bar{a} = (\gamma^2 \theta_{k+1} + \gamma \mathcal{D}^2(|\pi|)) (1 + \|\mathbf{P}_{r_k, T}^*(\xi)\|_{2\alpha}^\alpha) + \gamma (\vartheta_{k+1} + 1) \mathcal{D}^1(|\pi|).$$

We straightforwardly deduce that

$$\begin{aligned} \bar{\mathcal{S}}_k^j &= \bar{\Delta}_k^j + \bar{\delta} \bar{\mathcal{S}}_k^{j-1} \leq \Lambda \bar{\delta}^j (1 + \|\mathbf{P}_{r_k, T}^*(\xi)\|_{2\alpha}^\alpha) + \bar{a} + (1 + \gamma \theta_{k+1}) \bar{\delta} \bar{\mathcal{S}}_k^{j-1} \\ &\leq e^{\gamma \theta_{k+1} j} \bar{\delta}^j \bar{\mathcal{S}}_k^0 + \sum_{\ell=0}^{j-1} e^{\gamma \theta_{k+1} \ell} \bar{\delta}^\ell \left( \Lambda \bar{\delta}^{j-\ell} (1 + \|\mathbf{P}_{r_k, T}^*(\xi)\|_{2\alpha}^\alpha) + \bar{a} \right), \end{aligned}$$

which yields

$$\bar{\mathcal{S}}_k^j \leq \Lambda (j+2) \bar{\delta}^j e^{\gamma \theta_{k+1} (j-1)} (1 + \|\mathbf{P}_{r_k, T}^*(\xi)\|_{2\alpha}^\alpha) + \frac{\bar{a}}{1 - e^{\gamma \theta_{k+1} \bar{\delta}}},$$

where we used  $\bar{\mathcal{S}}_k^0 \leq 2\Lambda (1 + \|\mathbf{P}_{r_k, T}^*(\xi)\|_{2\alpha}^\alpha)$ . Thanks to (47), we get

$$\bar{\Delta}_k^J \leq \Lambda \bar{\delta}^J \left( \frac{1}{2} + \bar{\delta} \gamma (J+2) \theta_{k+1} e^{\gamma \theta_{k+1} (J-1)} \right) (1 + \|\mathbf{P}_{r_k, T}^*(\xi)\|_{2\alpha}^\alpha) + \frac{\bar{a}}{1 - e^{\gamma \theta_{k+1} \bar{\delta}}}.$$

Recalling that  $(J-1)\theta_{k+1} \leq 1$ , we deduce that, for  $\bar{\delta}$  small enough,

$$\bar{\Delta}_k^J \leq (\Lambda \bar{\delta}^J + e^{\beta \bar{\delta}} \{\theta_{k+1} + \mathcal{D}^2(|\pi|)\}) (1 + \|\mathbf{P}_{r_k, T}^*(\xi)\|_{2\alpha}^\alpha) + e^{\beta \bar{\delta}} (\vartheta_{k+1} + 1) \mathcal{D}^1(|\pi|),$$

provided that  $\beta$  satisfies

$$\frac{\gamma^2}{1 - e^{\gamma \theta_{k+1} \bar{\delta}}} \leq e^{\beta \bar{\delta}}. \quad (48)$$

This validates (41) and concludes the proof.  $\square$

**3.2. Convergence error for the implemented scheme.** We now analyse the global error of our method when the numerical algorithm is given by our benchmark Example 2, see Section 4.1.

**Lemma 8.** *(Scheme stability) Condition (A2) holds true for the scheme given in Example 2.*

**Proof.** For  $k \leq N-1$ , we consider  $(\bar{X}, \bar{Y}, \bar{Z}) := \overline{\text{solver}}[k](\bar{\mathcal{X}}, \eta, f)$  and  $(\bar{X}', \bar{Y}', \bar{Z}') := \overline{\text{solver}}[k](\bar{\mathcal{X}}', \eta', f')$  with  $\bar{\mathcal{X}}_{r_k} = \bar{\mathcal{X}}'_{r_k} = \xi$ . Letting  $\Delta X_i = \bar{X}_{t_i} - \bar{X}'_{t_i}$  and  $\Delta Y_i = \bar{Y}_{t_i} - \bar{Y}'_{t_i}$ , we observe

$$|\Delta X_{i+1}| \leq \left| \sum_{\ell=j_k}^i (t_{\ell+1} - t_\ell) \Delta b_\ell \right| + \left| \sum_{\ell=j_k}^i \Delta \sigma_\ell \Delta \bar{W}_\ell \right|,$$

for  $i \in \{j_k, \dots, j_{k+1}\}$ , where  $\Delta b_\ell := b(\bar{X}_{t_\ell}, \bar{Y}_{t_\ell}, [\bar{X}_{t_\ell}, \bar{Y}_{t_\ell}]) - b(\bar{X}'_{t_\ell}, \bar{Y}'_{t_\ell}, [\bar{X}'_{t_\ell}, \bar{Y}'_{t_\ell}])$  and, similarly,  $\Delta \sigma_\ell := \sigma(\bar{X}_{t_\ell}, [\bar{X}_{t_\ell}]) - \sigma(\bar{X}'_{t_\ell}, [\bar{X}'_{t_\ell}])$ .

Invoking Cauchy-Schwartz inequality for the first term and the Burkholder-Davis-Gundy inequality for discrete martingales for the second term and appealing to the Lipschitz property of  $b$  and  $\sigma$ , we get

$$\begin{aligned} \|\Delta X_{i+1}\|_{2\alpha} &\leq C\delta \max_{\ell=j_k, \dots, i} (\|\Delta Y_\ell\|_{2\alpha} + \|\Delta X_\ell\|_{2\alpha}) + C \left\| \sum_{\ell=j_k}^i |\Delta \sigma_\ell|^2 \cdot |\Delta \hat{W}_\ell|^2 \right\|_{\alpha}^{\frac{1}{2}} \\ &\leq C\delta \max_{\ell=j_k, \dots, i} (\|\Delta Y_\ell\|_{2\alpha} + \|\Delta X_\ell\|_{2\alpha}) + C \left( \sum_{\ell=j_k}^i (t_{\ell+1} - t_\ell) \|\Delta X_\ell\|_{2\alpha}^2 \right)^{\frac{1}{2}} \\ &\leq C\delta \max_{\ell=j_k, \dots, i} (\|\Delta Y_\ell\|_{2\alpha} + \|\Delta X_\ell\|_{2\alpha}) + C\delta^{1/2} \max_{\ell=j_k, \dots, i} (\|\Delta X_\ell\|_{2\alpha}), \end{aligned}$$

where we used the identity  $t_{\ell+1} - t_\ell = \delta/(j_{k+1} - j_k)$ . For  $\delta$  small enough (taking the sup in the sum), we then obtain

$$\max_{j_k \leq i \leq j_{k+1}} \|\Delta X_i\|_{2\alpha} \leq C\delta \max_{j_k \leq i \leq j_{k+1}} \|\Delta Y_i\|_{2\alpha}, \quad (49)$$

which concludes the proof.  $\square$

We now turn to the study of the approximation error.

**Lemma 9.** *Assume that (H0)-(H1) are in force. Then, condition (A1) holds true for the scheme given in Example 2 with*

$$\mathcal{D}^1(|\pi|) \leq C\sqrt{|\pi|} \text{ and } \mathcal{D}^2(|\pi|) \leq C\sqrt{|\pi|}.$$

**Proof.** *First Step.* Given the scheme defined in Example 2, we introduce its piecewise continuous version, which we denote by  $(\bar{X}_s)_{0 \leq s \leq T}$ . For  $i < n$ ,  $t_i < s < t_{i+1}$ ,

$$\bar{X}_s := \bar{X}_{t_i} + b_i(s - t_i) + \sigma_i \sqrt{s - t_i} \varpi_i, \quad \varpi_i := \frac{1}{\sqrt{t_{i+1} - t_i}} \Delta \bar{W}_i,$$

with  $(b_i, \sigma_i) := (b(\bar{X}_{t_i}, \bar{Y}_{t_i}, [\bar{X}_{t_i}, \bar{Y}_{t_i}]), \sigma(\bar{X}_{t_i}, [\bar{X}_{t_i}]))$ . In preparation for the proof, we also introduce a piecewise càd-làg version, denoted by  $(\bar{X}_s^{(\lambda)})_{0 \leq s \leq T}$ , where  $\lambda$  is a parameter in  $[0, 1)$ . For  $i < n$ ,  $t_i < s < t_{i+1}$ ,

$$\bar{X}_s^{(\lambda)} := \bar{X}_{t_i} + b_i(s - t_i) + \lambda \sigma_i \sqrt{s - t_i} \varpi_i.$$

For the reader's convenience, we also set

$$\begin{aligned}\bar{U}_s &:= \mathcal{U}(s, \bar{X}_s, [\bar{X}_s]), \\ \bar{V}_s^x &:= \partial_x \mathcal{U}(s, \bar{X}_s, [\bar{X}_s]), \quad \bar{V}_s^\mu := \partial_\mu \mathcal{U}(s, \bar{X}_s, [\bar{X}_s]) \langle \langle \bar{X}_s \rangle \rangle, \\ \bar{V}_s^{x,0} &:= \partial_x \mathcal{U}(s, \bar{X}_s^{(0)}, [\bar{X}_s]).\end{aligned}$$

Applying the discrete Itô formula given in Proposition 14, and using the PDE solved by  $\mathcal{U}$ , recall (1), we compute

$$\begin{aligned}\bar{U}_{t_{i+1}} &= \bar{U}_{t_i} + \int_{t_i}^{t_{i+1}} \bar{V}_s^x \cdot \{b(\bar{X}_{t_i}, \bar{Y}_{t_i}, [\bar{X}_{t_i}, \bar{Y}_{t_i}]) - b(\bar{X}_{t_i}, \bar{U}_{t_i}, [\bar{X}_{t_i}, \bar{U}_{t_i}])\} ds \\ &\quad + \int_{t_i}^{t_{i+1}} \hat{\mathbb{E}}[\bar{V}_s^\mu \cdot \{\langle b(\bar{X}_{t_i}, \bar{Y}_{t_i}, [\bar{X}_{t_i}, \bar{Y}_{t_i}]) - b(\bar{X}_{t_i}, \bar{U}_{t_i}, [\bar{X}_{t_i}, \bar{U}_{t_i}]) \rangle\}] ds \\ &\quad - (t_{i+1} - t_i) f(\bar{\mathfrak{X}}_{t_i}, \bar{U}_{t_i}, \sigma^\dagger(\bar{X}_{t_i}, [\bar{X}_{t_i}]) \bar{V}_{t_i}^x, [\bar{\mathfrak{X}}_{t_i}, \bar{U}_{t_i}]) \\ &\quad + \bar{V}_{t_i}^x \cdot \left( \sqrt{t_{i+1} - t_i} \sigma(\bar{X}_{t_i}, [\bar{X}_{t_i}]) \varpi_i \right) \\ &\quad + \mathcal{R}_i^w + \mathcal{R}_i^f + \mathcal{R}_i^{bx} + \mathcal{R}_i^{b\mu} + \mathcal{R}_i^{\sigma x} + \mathcal{R}_i^{\sigma\mu} + \delta\mathcal{M}(t_i, t_{i+1}) + \delta\mathcal{T}(t_i, t_{i+1}),\end{aligned}$$

with

$$\begin{aligned}\mathcal{R}_i^w &:= \int_{t_i}^{t_{i+1}} (\bar{V}_s^{x,0} - \bar{V}_{t_i}^{x,0}) \cdot \frac{\sigma(\bar{X}_{t_i}^{(0)}, [\bar{X}_{t_i}]) \varpi_i}{2\sqrt{s - t_i}} ds, \\ \mathcal{R}_i^f &:= \int_{t_i}^{t_{i+1}} \left\{ f(\bar{X}_s, \bar{U}_s, \sigma^\dagger(\bar{X}_s, [\bar{X}_s]) \bar{V}_s^x, [\bar{X}_s, \bar{U}_s]) \right. \\ &\quad \left. - f(\bar{\mathfrak{X}}_{t_i}, \bar{U}_{t_i}, \sigma^\dagger(\bar{X}_{t_i}, [\bar{X}_{t_i}]) \bar{V}_{t_i}^x, [\bar{\mathfrak{X}}_{t_i}, \bar{U}_{t_i}]) \right\} ds, \\ \mathcal{R}_i^{bx} &:= \int_{t_i}^{t_{i+1}} \bar{V}_s^x \cdot \{b(\bar{X}_{t_i}, \bar{U}_{t_i}, [\bar{X}_{t_i}, \bar{U}_{t_i}]) - b(\bar{X}_s, \bar{U}_s, [\bar{X}_s, \bar{U}_s])\} ds, \\ \mathcal{R}_i^{b\mu} &:= \int_{t_i}^{t_{i+1}} \hat{\mathbb{E}}[\bar{V}_s^\mu \cdot \{\langle b(\bar{X}_{t_i}, \bar{U}_{t_i}, [\bar{X}_{t_i}, \bar{U}_{t_i}]) - b(\bar{X}_s, \bar{U}_s, [\bar{X}_s, \bar{U}_s]) \rangle\}] ds,\end{aligned}$$

and

$$\mathcal{R}_i^{\sigma x} = \frac{1}{2} \int_{t_i}^{t_{i+1}} \int_0^1 \Delta^x(s, \lambda) d\lambda ds, \quad \mathcal{R}_i^{\sigma\mu} = \frac{1}{2} \int_{t_i}^{t_{i+1}} \int_0^1 \Delta^\mu(s, \lambda) d\lambda ds, \quad (50)$$

where

$$\begin{aligned}\Delta^x(s, \lambda) &:= \text{Tr} \left\{ \partial_{xx}^2 \mathcal{U}(s, \bar{X}_s^{(\lambda)}, [\bar{X}_s]) a(\bar{X}_{t_i}, [\bar{X}_{t_i}]) - \partial_{xx}^2 \mathcal{U}(s, \bar{X}_s, [\bar{X}_s]) a(\bar{X}_s, [\bar{X}_s]) \right\} \\ \Delta^\mu(s, \lambda) &:= \hat{\mathbb{E}} \left[ \text{Tr} \left\{ \partial_v \partial_\mu \mathcal{U}(s, \bar{X}_s, [\bar{X}_s]) \langle \langle \bar{X}_s^{(\lambda)} \rangle \rangle \langle a(\bar{X}_{t_i}, [\bar{X}_{t_i}]) \rangle \right. \right. \\ &\quad \left. \left. - \partial_v \partial_\mu \mathcal{U}(s, \bar{X}_s, [\bar{X}_s]) \langle \langle \bar{X}_s \rangle \rangle \langle a(\bar{X}_s, [\bar{X}_s]) \rangle \right\} \right].\end{aligned}$$

Also,  $\delta\mathcal{M}(t_i, t_{i+1})$  is a martingale increment satisfying  $\mathbb{E}[|\delta\mathcal{M}(t_i, t_{i+1})|^{2\alpha} | \mathcal{F}_{t_i}]^{1/(2\alpha)} \leq Ch_i$  and  $\|\delta\mathcal{T}(t_i, t_{i+1})\|_{2\alpha} \leq C_\Lambda h_i^{\frac{3}{2}}$ , recall Proposition 14.

*Second Step.* Denoting  $\underline{\varpi}_i := \varpi_i / \sqrt{t_{i+1} - t_i}$  and

$$\begin{aligned} \delta b_i &:= \frac{1}{h_i} \int_{t_i}^{t_{i+1}} \bar{V}_s^x \cdot \{b(\bar{X}_{t_i}, \bar{Y}_{t_i}, [\bar{X}_{t_i}, \bar{Y}_{t_i}]) - b(\bar{X}_{t_i}, \bar{U}_{t_i}, [\bar{X}_{t_i}, \bar{U}_{t_i}])\} ds \\ &\quad + \frac{1}{h_i} \int_{t_i}^{t_{i+1}} \hat{\mathbb{E}}[\bar{V}_s^\mu \cdot \{\langle b(\bar{X}_{t_i}, \bar{Y}_{t_i}, [\bar{X}_{t_i}, \bar{Y}_{t_i}]) - b(\bar{X}_{t_i}, \bar{U}_{t_i}, [\bar{X}_{t_i}, \bar{U}_{t_i}]) \rangle\}] ds, \end{aligned}$$

the previous equation reads

$$\begin{aligned} \bar{U}_{t_{i+1}} &= \bar{U}_{t_i} + \zeta_i \\ &\quad + h_i \left[ \delta b_i - f(\bar{\mathfrak{X}}_{t_i}, \bar{U}_{t_i}, \sigma^\dagger(\bar{X}_{t_i}, [\bar{X}_{t_i}]) \bar{V}_{t_i}^x, [\bar{\mathfrak{X}}_{t_i}, \bar{U}_{t_i}]) + \bar{V}_{t_i}^x \cdot (\sigma(\bar{X}_{t_i}, [\bar{X}_{t_i}]) \underline{\varpi}_i) \right], \end{aligned} \quad (51)$$

where

$$\zeta_i := \mathcal{R}_i^w + \mathcal{R}_i^f + \mathcal{R}_i^{bx} + \mathcal{R}_i^{b\mu} + \mathcal{R}_i^{\sigma x} + \mathcal{R}_i^{\sigma\mu} + \delta\mathcal{M}(t_i, t_{i+1}) + \delta\mathcal{T}(t_i, t_{i+1}).$$

On the other hand, the scheme can be rewritten as

$$\bar{Y}_{t_i} = \bar{Y}_{t_{i+1}} + h_i f(\bar{\mathfrak{X}}_{t_i}, \bar{Y}_{t_i}, \bar{Z}_{t_i}, [\bar{\mathfrak{X}}_{t_i}, \bar{Y}_{t_i}]) - h_i \bar{Z}_{t_i} \cdot \underline{\varpi}_i - \Delta M_i, \quad (52)$$

where  $\Delta M_i$  satisfies

$$\mathbb{E}_{t_i}[\Delta M_i] = 0, \quad \mathbb{E}_{t_i}[\underline{\varpi}_i \cdot \Delta M_i] = 0 \quad \text{and} \quad \mathbb{E}[|\Delta M_i|^2] < \infty. \quad (53)$$

Denoting  $\Delta \bar{Y}_i = \bar{Y}_{t_i} - \bar{U}_{t_i}$ ,  $\Delta \bar{Z}_i = \bar{Z}_{t_i} - \sigma^\dagger(\bar{X}_{t_i}, [\bar{X}_{t_i}]) \bar{V}_{t_i}^x$ , and adding (51) and (52), we get

$$\Delta \bar{Y}_i = \Delta \bar{Y}_{i+1} + h_i (\delta b_i + \delta f_i) + \zeta_i - h_i \Delta \bar{Z}_i \cdot \underline{\varpi}_i - \Delta M_i, \quad (54)$$

where

$$\delta f_i = f(\bar{\mathfrak{X}}_{t_i}, \bar{Y}_{t_i}, \bar{Z}_{t_i}, [\bar{\mathfrak{X}}_{t_i}, \bar{Y}_{t_i}]) - f(\bar{\mathfrak{X}}_{t_i}, \bar{U}_{t_i}, \sigma^\dagger(\bar{X}_{t_i}, [\bar{X}_{t_i}]) \bar{V}_{t_i}^x, [\bar{\mathfrak{X}}_{t_i}, \bar{U}_{t_i}]).$$

For later use, we observe that

$$|\delta b_i| + |\delta f_i| \leq C_\Lambda (|\Delta \bar{Y}_i| + \|\Delta \bar{Y}_i\|_2 + |\Delta \bar{Z}_i|). \quad (55)$$

Summing the equation (54) from  $i$  to  $j_{k+1} - 1$ , we obtain

$$\Delta \bar{Y}_i + \sum_{\ell=i}^{j_{k+1}-1} \{h_\ell \Delta \bar{Z}_\ell \cdot \underline{\varpi}_\ell + \Delta M_\ell\} = \Delta \bar{Y}_{j_{k+1}} + \sum_{\ell=i}^{j_{k+1}-1} h_\ell (\delta b_\ell + \delta f_\ell) - \sum_{\ell=i}^{j_{k+1}-1} \zeta_\ell.$$

Squaring both sides and taking expectation, we compute, using (53) for the left side and Young's and conditional Cauchy-Schwarz inequality for the right side,

$$\begin{aligned} &\mathbb{E}_{t_q}[|\Delta \bar{Y}_i|^2] + \sum_{\ell=i}^{j_{k+1}-1} h_\ell \mathbb{E}_{t_q}[|\Delta \bar{Z}_\ell|^2] \\ &\leq \mathbb{E}_{t_q} \left[ (1 + C\delta) |\Delta \bar{Y}_{j_{k+1}}|^2 + C \sum_{\ell=i}^{j_{k+1}-1} h_\ell |\delta b_\ell + \delta f_\ell|^2 + \frac{C}{\delta} \left( \sum_{\ell=i}^{j_{k+1}-1} \zeta_\ell \right)^2 \right], \end{aligned}$$

for  $i \geq q \geq j_k$ . Combining (55) and Young's inequality, this leads to

$$\begin{aligned} & \mathbb{E}_{t_q} [|\Delta \bar{Y}_i|^2] + \frac{1}{2} \sum_{\ell=i}^{j_{k+1}-1} h_\ell \mathbb{E}_{t_q} [|\bar{Z}_\ell|^2] \\ & \leq \mathbb{E}_{t_q} \left[ e^{C\delta} |\Delta \bar{Y}_{j_{k+1}}|^2 + C \sum_{\ell=i}^{j_{k+1}-1} h_\ell |\Delta \bar{Y}_\ell|^2 + \frac{C}{\delta} \left( \sum_{\ell=i}^{j_{k+1}-1} \zeta_\ell \right)^2 \right]. \end{aligned}$$

Using the discrete version of Gronwall's lemma and recalling that  $\sum_{\ell=j_k}^{j_{k+1}-1} h_\ell = \delta$ , we obtain, for  $i = q$ ,

$$|\Delta \bar{Y}_i|^2 \leq \mathbb{E}_{t_i} \left[ e^{C\delta} |\Delta \bar{Y}_{j_{k+1}}|^2 + \frac{C}{\delta} \max_{j_k \leq i \leq j_{k+1}-1} \left( \sum_{\ell=i}^{j_{k+1}-1} \zeta_\ell \right)^2 \right],$$

and then,

$$\Delta_Y^2 := \max_{j_k \leq i \leq j_{k+1}} \|\Delta \bar{Y}_i\|_{2\alpha}^2 \leq e^{C\delta} \|\Delta \bar{Y}_{j_{k+1}}\|_{2\alpha}^2 + \frac{C}{\delta} \left\| \max_{j_k \leq i \leq j_{k+1}-1} \left( \sum_{\ell=i}^{j_{k+1}-1} \zeta_\ell \right) \right\|_{2\alpha}^2. \quad (56)$$

*Third Step.* To conclude, we need an upper bound for the error  $\|\max_{j_k \leq i \leq j_{k+1}-1} (\sum_{\ell=i}^{j_{k+1}-1} \zeta_\ell)\|_{2\alpha}^2$  where  $\zeta_\ell$  is defined in (52). To do so, we study each term in (52) separately. We also define  $\Delta_X := \max_{t \in \pi^k} \|\bar{X}_t - \tilde{\mathfrak{X}}_t\|_{2\alpha}$  and we recall that  $\bar{X}_{r_k} = \xi$ .

*Third Step a.* We first study the contribution of  $\mathcal{R}_i^f$  to the global error term and note that

$$\left\| \max_{j_k \leq i \leq j_{k+1}} \left( \sum_{\ell=i}^{j_{k+1}-1} \mathcal{R}_\ell^f \right) \right\|_{2\alpha}^2 \leq C \frac{\delta}{|\pi|} \sum_{\ell=j_k}^{j_{k+1}-1} \|\mathcal{R}_\ell^f\|_{2\alpha}^2. \quad (57)$$

We will upper bound this last term.

Let us first observe, that, for  $t_i \leq s \leq t_{i+1}$ ,

$$\begin{aligned} |\bar{V}_s^x - \bar{V}_{t_i}^x| & \leq |\partial_x \mathcal{U}(s, \bar{X}_s, [\bar{X}_s]) - \partial_x \mathcal{U}(t_i, \bar{X}_{t_i}, [\bar{X}_{t_i}])| \\ & \leq C \left( |\bar{X}_s - \bar{X}_{t_i}| + \mathcal{W}_2([\bar{X}_s], [\bar{X}_{t_i}]) + h_i^{\frac{1}{2}} (1 + |\bar{X}_{t_i}| + \|\bar{X}_{t_i}\|_2) \right), \end{aligned}$$

where we used the Lipschitz property of  $\partial_x \mathcal{U}$  given in (H1), together with (8) and (12). Hence,

$$\|\bar{V}_s^x - \bar{V}_{t_i}^x\|_{2\alpha}^2 \leq C \left( \|\bar{X}_s - \bar{X}_{t_i}\|_{2\alpha}^2 + h_i (1 + \|\bar{X}_{t_i}\|_{2\alpha}^2) \right). \quad (58)$$

From the boundedness of  $\sigma$  and the Lipschitz property of  $b$  and  $\mathcal{U}$ , we compute

$$\|\bar{X}_s - \bar{X}_{t_i}\|_{2\alpha}^2 \leq C_\Lambda \left( h_i + h_i^2 \|\bar{U}_{t_i} - Y_{t_i}\|_{2\alpha}^2 + h_i^2 \|\bar{X}_{t_i}\|_{2\alpha}^2 \right). \quad (59)$$

Using Lemma 15 from the appendix below, we obtain

$$\|\bar{V}_s^x - \bar{V}_{t_i}^x\|_{2\alpha}^2 \leq C \left( h_i (1 + \|\xi\|_{2\alpha}^2) + h_i^2 \Delta_Y^2 \right).$$

From the boundedness of  $\partial_x \mathcal{U}$ ,  $\sigma$  and the Lipschitz property of  $\sigma$ , we obtain

$$\|\sigma^\dagger(\bar{X}_s, [\bar{X}_s]) \bar{V}_s^x - \sigma^\dagger(\bar{X}_{t_i}, [\bar{X}_{t_i}]) \bar{V}_{t_i}^x\|_{2\alpha}^2 \leq C \left( h_i (1 + \|\xi\|_{2\alpha}^2) + h_i^2 \Delta_Y^2 \right),$$

where we used the same argument as above to handle the difference between the two  $\sigma$  terms. Combining the previous inequality with the Lipschitz property of  $f$  and replicating the analysis to handle the difference between the  $\bar{U}$  terms, we deduce

$$\|\mathcal{R}_i^f\|_{2\alpha}^2 \leq Ch_i^2 \left( \Delta_X^2 + h_i(1 + \|\xi\|_{2\alpha}^2) + h_i^2 \Delta_Y^2 \right). \quad (60)$$

*Third Step b.* Combining the Lipschitz property of  $b$ , the fact that  $|\bar{V}_s^x|^2 + \hat{\mathbb{E}}[|\bar{V}_s^\mu|^2] \leq C$  and Cauchy-Schwarz inequality, we get

$$\|\mathcal{R}_i^{bx}\|_{2\alpha}^2 + \|\mathcal{R}_i^{b\mu}\|_{2\alpha}^2 \leq Ch_i^2 \|\bar{U}_s - \bar{U}_{t_i}\|_{2\alpha}^2 + \|\bar{X}_s - \bar{X}_{t_i}\|_{2\alpha}^2. \quad (61)$$

Arguing as in the previous step, we easily get

$$\|\mathcal{R}_i^b\|_{2\alpha}^2 \leq Ch_i^2 \left( h_i(1 + \|\xi\|_{2\alpha}^2) + h_i^2 \Delta_Y^2 \right). \quad (62)$$

*Third Step c.* We now study the contribution of the terms  $\mathcal{R}_i^w$  to the global error. From the independence property of  $(\varpi_i)_{i=0, \dots, n-1}$ , we may regard each  $\mathcal{R}_\ell^w$  as a martingale increment. By Burkholder-Davies-Gundy inequalities for discrete martingales, we first compute, using the fact that each  $\varpi_i$  is uniformly bounded,

$$\begin{aligned} \left\| \max_{j_k \leq i \leq j_{k-1}} \left( \sum_{\ell=i}^{j_{k+1}-1} \mathcal{R}_\ell^w \right) \right\|_{2\alpha}^2 &\leq C \left\| \sum_{\ell=j_k}^{j_{k+1}-1} \left| \int_{t_i}^{t_{i+1}} \sigma^\dagger(\bar{X}_{t_i}^{(0)}, [\bar{X}_{t_i}]) \frac{\bar{V}_s^{0,x} - \bar{V}_{t_i}^{0,x}}{\sqrt{s-t_i}} ds \right|^2 \right\|_{\alpha} \\ &\leq C \sum_{\ell=j_k}^{j_{k+1}-1} h_i \left( h_i(1 + \|\xi\|_{2\alpha}^2) + \left\| \sup_{s \in [t_i, t_{i+1}]} |\bar{X}_s^{(0)} - \bar{X}_{t_i}|^2 \right\|_{\alpha} \right). \end{aligned}$$

Since  $|\bar{X}_s^{(0)} - \bar{X}_{t_i}| \leq h_i |b(\bar{X}_{t_i}, \bar{Y}_{t_i}, [\bar{X}_{t_i}, \bar{Y}_{t_i}])|$ , for  $s \in [t_i, t_{i+1}]$ , so that  $\|\bar{X}_s^{(0)} - \bar{X}_{t_i}\|_{2\alpha} \leq C_\Lambda h_i(1 + \|\bar{X}_{t_i}\|_{2\alpha} + \|\bar{Y}_{t_i}\|_{2\alpha}) \leq C_\Lambda h_i(1 + \|\bar{X}_{t_i}\|_{2\alpha} + \Delta_Y^2)$ , the previous inequality, together with Lemma 15, leads to

$$\left\| \max_{j_k \leq i \leq j_{k-1}} \left( \sum_{\ell=i}^{j_{k+1}-1} \mathcal{R}_\ell^w \right) \right\|_{2\alpha}^2 \leq C \delta |\pi| \left( 1 + \|\xi\|_{2\alpha}^2 + |\pi| \Delta_Y^2 \right).$$

Similarly,

$$\begin{aligned} \left\| \max_{j_k \leq i \leq j_{k-1}} \left( \sum_{\ell=i}^{j_{k+1}-1} \delta \mathcal{M}(t_\ell, t_{\ell+1}) \right) \right\|_{2\alpha}^2 &\leq C \sum_{\ell=j_k}^{j_{k+1}-1} \left\| |\delta \mathcal{M}(t_\ell, t_{\ell+1})|^2 \right\|_{\alpha} \\ &\leq C \delta |\pi|. \end{aligned}$$

Hence,

$$\begin{aligned} &\left\| \max_{j_k \leq i \leq j_{k-1}} \left( \sum_{\ell=i}^{j_{k+1}-1} \mathcal{R}_\ell^w \right) \right\|_{2\alpha}^2 + \left\| \max_{j_k \leq i \leq j_{k-1}} \left( \sum_{\ell=i}^{j_{k+1}-1} \delta \mathcal{M}(t_\ell, t_{\ell+1}) \right) \right\|_{2\alpha}^2 \\ &\leq C \delta |\pi| (1 + \|\xi\|_{2\alpha}^2) + C \delta |\pi|^2 \Delta_Y^2. \end{aligned} \quad (63)$$

*Third Step d.* (i) We study the contribution of  $\mathcal{R}_i^{\sigma x}$ . We observe that

$$\begin{aligned} |\Delta^x(s, \lambda)| &\leq \left| \partial_{xx}^2 \mathcal{U}(s, \bar{X}_s^{(\lambda)}, [\bar{X}_s]) - \partial_{xx}^2 \mathcal{U}(s, \bar{X}_s, [\bar{X}_s]) \right| \cdot |a(\bar{X}_{t_i}, [\bar{X}_{t_i}])| \\ &\quad + \left| \partial_{xx}^2 \mathcal{U}(s, \bar{X}_s, [\bar{X}_s]) \right| \cdot |a(\bar{X}_{t_i}, [\bar{X}_{t_i}]) - a(\bar{X}_s, [\bar{X}_s])|, \end{aligned}$$

for  $s \in [t_i, t_{i+1}]$ . Using the boundedness and Lipschitz continuity of  $\partial_{xx}^2 \mathcal{U}$  and  $\sigma$ , we get, from the previous expression,

$$\|\Delta^x(s, \lambda)\|_{2\alpha}^2 \leq C \left( \|\bar{X}_s^{(\lambda)} - \bar{X}_s\|_{2\alpha}^2 + \|\bar{X}_s - \bar{X}_{t_i}\|_{2\alpha}^2 \right). \quad (64)$$

Observing that  $\|\bar{X}_s^{(\lambda)} - \bar{X}_s\|_{2\alpha} \leq C\sqrt{h_i}$ , we obtain using (59), for  $t_i \leq s \leq t_{i+1}$

$$\|\Delta^x(s, \lambda)\|_{2\alpha}^2 \leq Ch_i(1 + h_i\|\bar{U}_{t_i} - Y_{t_i}\|_{2\alpha}^2 + h_i\|\bar{X}_{t_i}\|_{2\alpha}^2).$$

which leads, using Lemma 15 again, to

$$\|\mathcal{R}_i^{\sigma x}\|_{2\alpha}^2 \leq Ch_i^2 \left( h_i + h_i^2(\Delta_Y^2 + \|\xi\|_{2\alpha}^2) \right). \quad (65)$$

(ii) To study  $\mathcal{R}_i^{\sigma\mu}$ , we first observe that

$$\begin{aligned} |\Delta^\mu(s, \lambda)| &\leq C \hat{\mathbb{E}} \left[ \left| \partial_v \partial_\mu \mathcal{U}(s, \bar{X}_s, [\bar{X}_s])(\langle \bar{X}_s^{(\lambda)} \rangle) - \partial_v \partial_\mu \mathcal{U}(s, \bar{X}_s, [\bar{X}_s])(\langle \bar{X}_s \rangle) \right| \right] \\ &\quad + \hat{\mathbb{E}} \left[ \left| \partial_v \partial_\mu \mathcal{U}(s, \bar{X}_s, [\bar{X}_s])(\langle \bar{X}_s \rangle) \right| \cdot \left| \langle a(\bar{X}_{t_i}, [\bar{X}_{t_i}]) - a(\bar{X}_s, [\bar{X}_s]) \rangle \right| \right]. \end{aligned} \quad (66)$$

For the last term, we combine Cauchy-Schwarz inequality (10) and boundedness and Lipschitz continuity of  $\sigma$  to get

$$\begin{aligned} &\hat{\mathbb{E}} \left[ \left| \partial_v \partial_\mu \mathcal{U}(s, \bar{X}_s, [\bar{X}_s])(\langle \bar{X}_s \rangle) \right| \cdot \left| \langle a(\bar{X}_{t_i}, [\bar{X}_{t_i}]) - a(\bar{X}_s, [\bar{X}_s]) \rangle \right| \right] \\ &\leq C \|\bar{X}_{t_i} - \bar{X}_s\|_2 \leq C \|\bar{X}_{t_i} - \bar{X}_s\|_{2\alpha}. \end{aligned}$$

Recalling from (59) that  $\|\bar{X}_s - \bar{X}_{t_i}\|_{2\alpha}^2 \leq C_\Lambda(h_i + h_i^2(\Delta_Y^2 + \|\bar{X}_{t_i}\|_{2\alpha}^2))$ , we obtain, using Lemma 15, that

$$\begin{aligned} &\hat{\mathbb{E}} \left[ \left| \partial_v \partial_\mu \mathcal{U}(s, \bar{X}_s, [\bar{X}_s])(\langle \bar{X}_s \rangle) \right| \cdot \left| \langle a(\bar{X}_{t_i}, [\bar{X}_{t_i}]) - a(\bar{X}_s, [\bar{X}_s]) \rangle \right| \right] \\ &\leq C_\Lambda h_i^{\frac{1}{2}} \left( 1 + h_i^{\frac{1}{2}} \{ \Delta_Y + \|\xi\|_{2\alpha} \} \right). \end{aligned} \quad (67)$$

For the first term in (66), we use (H1) equation (13) to get

$$\begin{aligned} &\left| \partial_v \partial_\mu \mathcal{U}(s, \bar{X}_s, [\bar{X}_s])(\langle \bar{X}_s^{(\lambda)} \rangle) - \partial_v \partial_\mu \mathcal{U}(s, \bar{X}_s, [\bar{X}_s])(\langle \bar{X}_s \rangle) \right| \\ &\leq C \left\{ 1 + |\langle \bar{X}_s^{(\lambda)} \rangle|^{2\alpha} + |\langle \bar{X}_s \rangle|^{2\alpha} + \|\bar{X}_s\|_2^{2\alpha} \right\}^{\frac{1}{2}} |\langle \bar{X}_s^{(\lambda)} \rangle - \langle \bar{X}_s \rangle|. \end{aligned}$$

By Cauchy Schwarz inequality, we obtain

$$\begin{aligned} &\hat{\mathbb{E}} \left[ \left| \{ \partial_v \partial_\mu \mathcal{U}(s, \bar{X}_s, [\bar{X}_s])(\langle \bar{X}_s^{(\lambda)} \rangle) - \partial_v \partial_\mu \mathcal{U}(s, \bar{X}_s, [\bar{X}_s])(\langle \bar{X}_s \rangle) \} \right| \right] \\ &\leq C\sqrt{h_i} \left( 1 + \|\bar{X}_s^{(\lambda)}\|_{2\alpha}^\alpha + \|\bar{X}_s\|_{2\alpha}^\alpha \right). \end{aligned} \quad (68)$$

We then observe that

$$\begin{aligned} \|\bar{X}_s^{(\lambda)}\|_{2\alpha} + \|\bar{X}_s\|_{2\alpha} &\leq C \left( \|\bar{X}_{t_i}\|_{2\alpha} + h_i\|\bar{U}_{t_i} - \bar{Y}_{t_i}\|_{2\alpha} + \sqrt{h_i} \right) \\ &\leq C(1 + \|\xi\|_{2\alpha} + \delta\Delta_Y), \end{aligned}$$

where we used lemma 15 for the last inequality. Combining the last inequality with (68) and using also (67), we compute

$$|\mathcal{R}_i^{\sigma\mu}| \leq Ch_i^{\frac{3}{2}} (1 + \|\xi\|_{2\alpha} + \delta\Delta_Y),$$

and then

$$\left\| \sum_{\ell=j_k}^{j_{k+1}-1} |\mathcal{R}_\ell^{\sigma\mu}| \right\|_{2\alpha}^2 \leq C|\pi|\delta^2 (1 + \delta^2 \Delta_Y^2 + \|\xi\|_{2\alpha}^2) . \quad (69)$$

4. Collecting the estimates (60), (62) and (65), we compute

$$\left( \sum_{\ell=j_k}^{j_{k+1}-1} \|\mathcal{R}_\ell^f + \mathcal{R}_\ell^b + \mathcal{R}_\ell^{\sigma x}\|_{2\alpha} \right)^2 \leq C\delta^2 (\Delta_X^2 + |\pi|\{1 + \|\xi\|_{2\alpha}^2\} + |\pi|^2 \Delta_Y^2) .$$

Observing that

$$\left( \sum_{\ell=j_k}^{j_{k+1}-1} \|\delta\mathcal{T}(t_i, t_{i+1})\|_{2\alpha} \right)^2 \leq C\delta^2 |\pi| ,$$

and combining the previous inequality with (69), (63) and (56), we obtain

$$\Delta_Y^2 \leq e^{C\delta} \|\bar{U}_{r_{k+1}} - \bar{Y}_{r_{k+1}}\|_{2\alpha}^2 + C \left( \delta \Delta_X^2 + |\pi|(1 + \|\xi\|_{2\alpha}^2) + |\pi|\delta \Delta_Y^2 \right) ,$$

which concludes the proof for  $\delta$  small enough.  $\square$

**Lemma 10.** *Assume that  $g$  and  $f(\cdot, 0, 0, [\cdot, 0])$  are bounded. Then (A3) is satisfied whatever the value of  $\alpha$ .*

**Proof.** It suffices to prove that  $\mathcal{U}$  is bounded on the whole space and that  $\bar{Y}$  is bounded independently of the discretization parameters.

We refer to [20] for the proof of the boundedness of  $\mathcal{U}$ .

The bound for  $\bar{Y}$  may be obtained by squaring (52) and then by taking the conditional expectation exactly as done in the second step of the proof of Lemma 9.  $\square$

Assumptions (A4) and (A5) are easily checked. It suffices to observe that  $(\mathbb{P}_{r_k, t}(\xi))_{t \in \pi, t \geq r_k}$  coincides with the solution of the discrete Euler scheme:

$$\bar{X}_{t_{i+1}}^0 = \bar{X}_{t_i}^0 + (t_{i+1} - t_i)b(X_{t_i}^0, 0, [X_{t_i}^0, 0]) + \sqrt{t_{i+1} - t_i}\sigma(X_{t_i}^0, [X_{t_i}^0])\varpi_i ,$$

with  $\bar{X}_{r_k}^0 = \xi$  as initial condition.

Combining Lemma 9, Lemma 8 and Lemma 10 with Theorem 7, we have the following result.

**Corollary 11.** *Under (H1)-(H0), assuming (39), the following holds*

$$\|\text{solver}[k](\xi) - \mathcal{U}(r_k, \xi, [\xi])\|_{2\alpha} \leq C \left( (C\delta)^{J-1} + |\pi|^{\frac{1}{2}}\delta^{-1}(1 + \|\xi\|_{2\alpha}) \right) ,$$

for  $\bar{\delta}$  small enough.

The first term in the right hand side is connected with the local Picard iterations on a step of length  $\delta$ . As expected, it decreases geometrically fast with the number of iterations. The second term is due to the propagation of the error along the mesh. The leading term  $|\pi|^{\frac{1}{2}}$  is consistent with that observed for classical forward-backward systems, see for instance [21, 22]. The normalization by  $\delta$  is due to the propagation of the error through the successive local solvers.

## 4. NUMERICAL APPLICATIONS

In practice, we would like to approximate the value of  $\mathcal{U}(0, \cdot)$  at some point  $(x, \mu) \in \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$ . In the first section below, we explain how to retrieve such approximation using the approximation of  $\mathcal{U}(0, \xi, [\xi])$  given by the algorithm `solver[0]()`, for some  $\xi \sim \mu$ . In a second part, we discuss the numerical results obtained by implementing `solver[0]()` with two levels, i.e.  $N = 2$ . In particular, we show that it is more efficient than an algorithm based simply on Picard Iterations.

**4.1. Approximation of  $\mathcal{U}(0, x, \mu)$ .** The goal of this section is to show how to obtain an approximation of  $\mathcal{U}(0, x, [\xi])$  with  $\xi \sim \mu$  and  $x \in \text{supp}(\mu)$ . We will assume that we thus have at hand a discrete valued random variable  $\xi^{|\pi|} \sim \mu^{|\pi|} = \sum_{\ell=1}^M p_\ell \delta_{x^\ell}$  such that  $\mu^{|\pi|}$  is a good approximation of  $\mu$  for the Wasserstein distance. For instance, such an approximation can be constructed by using quantization techniques. Then, we can use `solver[0]( $\xi^{|\pi|}$ )` to obtain an approximation of  $\mathcal{U}(0, \xi^{|\pi|}, [\xi^{|\pi|}])$ .

Note that `solver[0]( $\xi^{|\pi|}$ )` is a discrete random variable as the algorithm is initialised by a discrete random variable as well. In practice, this means that each point  $x^\ell$  will be the root of a tree and will be associated to an output value  $y^\ell = \mathcal{U}(0, x^\ell, [\xi^{|\pi|}])$  and then `solver[0]( $\xi^{|\pi|}$ )`  $\sim \sum_{\ell=1}^M p_\ell \delta_{y^\ell}$ . It is important to remark that the computations on the trees are connected via the McKean-Vlasov interaction.

Using the Lipschitz continuity of  $\mathcal{U}$ , one easily obtains

$$\begin{aligned} |\mathcal{U}(0, x, \mu) - \mathcal{U}(0, x^{\bar{\ell}}, \mu^{|\pi|})| &\leq C \left( \min_{y \in \text{supp}([\xi^{|\pi|}])} |y - x| + \mathcal{W}_2(\mu^{|\pi|}, \mu) \right) \\ &=: \mathcal{E}_1(|\pi|, \xi), \end{aligned} \quad (70)$$

where  $x^{\bar{\ell}}$  is a point in the support of  $\mu^{|\pi|}$  realising the minimum in the first line.

**Remark 12.** *In many cases, it will be easy to have  $x \in \text{supp}(\mu^{|\pi|})$  and thus reduce the above error to the term  $\mathcal{W}_2(\mu^{|\pi|}, \mu)$ . This is obviously the case if  $\xi$  is deterministic.*

As mentioned above, the approximation of  $\mathcal{U}(0, x^{\bar{\ell}}, \mu^{|\pi|})$  is obtained by running `solver[0]( $\xi^{|\pi|}$ )` and by taking its value on the tree initiated at  $x^{\bar{\ell}}$ , precisely we have  $\mathcal{U}(0, x^{\bar{\ell}}, \mu^{|\pi|}) = y^{\bar{\ell}}$ . The corresponding pointwise error is given by

$$\mathcal{E}_2(|\pi|, \delta, \xi) := |y^{\bar{\ell}} - \mathcal{U}(0, x^{\bar{\ell}}, [\xi^{|\pi|}])|. \quad (71)$$

Of a course, this might be estimated by

$$\mathcal{E}_2(|\pi|, \delta, \xi) \leq \frac{1}{p_{\bar{\ell}}} \left\| \mathcal{U}(0, \xi^{|\pi|}, [\xi^{|\pi|}]) - \text{solver}[0](\xi^{|\pi|}) \right\|_2,$$

but this is very poor when the initial distribution  $\mu$  is diffuse and accordingly when  $\mu^{|\pi|}$  has a large support, in which case  $p_{\bar{\ell}}$  is expected to be small.

To bypass this difficulty, we must regard  $\mathcal{E}_2(|\pi|, \delta, \xi)$  as a conditional error. Somehow, it is the error of the numerical scheme conditional on the initial root of the tree. It requires a new analysis, but it should not be so challenging: Now that we have investigated the error for the McKean-Vlasov component, we can easily revisit the proof of Theorem 7 in order to derive a bound for this conditional error.

Instead of revisiting the whole proof, we can argue by doubling the variables. For  $\xi$  and  $x$  as above, we can regard the four equations (2), (3), (4) and (5) as a single forward-backward system of the McKean-Vlasov type. The forward component of such a doubled system is  $\mathbb{X} = (X^{0,x,\mu}, X^{0,\xi})$  and the backward components are  $\mathbb{Y} = (Y^{0,x,\mu}, Y^{0,\xi})$  and

$\mathbb{Z} = (Z^{0,x,\mu}, Z^{0,\xi})$ . Except for the fact that the dimension of  $\mathbb{X}$  is no longer equal to the dimension of the noise, which we assumed to be true for convenience only, and for the fact that  $\mathbb{Y}$  takes values in  $\mathbb{R}^2$ , the setting is exactly the same as before, namely  $(\mathbb{X}, \mathbb{Y}, \mathbb{Z})$  can be regarded as the solution of a McKean-Vlasov forward-backward SDE in which the mean field component reduces to the marginal law of  $(X^{0,\xi}, Y^{0,\xi})$ . We observe in particular that

$$Y_t^{0,x,\mu} = \mathcal{U}(t, X_t^{0,x,\mu}, [X_t^{0,\xi}]), \quad Y_t^{0,\xi} = \mathcal{U}(t, X_t^{0,\xi}, [X_t^{0,\xi}]), \quad t \in [0, T],$$

with similar relationships for  $Z^{0,x,\mu}$  and  $Z^{0,\xi}$ . Hence,  $\mathbb{Y}_t$  (and  $\mathbb{Z}_t$ ) can be represented as a function of  $\mathbb{X}$ , which was the key assumption in our analysis. For sure, the fact that  $\mathbb{Y}$  takes values in dimension 2 is not a limitation for duplicating the arguments used to prove Theorem 7.

Numerically speaking, the tree initiated at root  $x^{\bar{\ell}}$  under the initial distribution  $\mu^{|\pi|}$  provides an approximation of  $\mathcal{U}(0, x^{\bar{\ell}}, [\xi^{|\pi|}])$ , which is equal to  $Y^{0,x^{\bar{\ell}},[\xi^{|\pi|}]}$ . So our numerical (implemented) scheme is in fact a numerical for the whole process  $(\mathbb{X}, \mathbb{Y}, \mathbb{Z})$ .

This leads us to the following result.

**Theorem 13.** *Let  $y^{\bar{\ell}}$  be the approximation of  $\mathcal{U}(0, x, \mu)$  obtained by calling `solver[0]` ( $\xi^{|\pi|}$ ), where  $\bar{\ell}$  is defined in (70). Then, the following holds*

$$|\mathcal{U}(0, x, \mu) - y^{\bar{\ell}}| \leq \mathcal{E}_1(|\pi|, \xi) + \mathcal{E}_2(|\pi|, \delta, \xi),$$

where  $\mathcal{E}_2(|\pi|, \delta, \xi)$  can be estimated by Corollary 11, with  $(1 + \|\xi\|_{2\alpha})$  replaced by  $(1 + |x^{\bar{\ell}}| + \|\xi\|_{2\alpha})$ .

**4.2. Numerical illustration.** In this section, we will prove empirically the convergence of the approximation obtained by the solver `solver[]()`. In particular, we will compare the output of our algorithm `solver[]()`, when implemented with two levels, i.e.  $N = 2$  (we simply call it *two-level algorithm*), with the output of a basic algorithm based only on Picard iterations, which can be seen as a solver `solver[]()`, but with only one level, i.e.  $N = 1$  (we simply call it *one-level algorithm*). In both cases, we use Example 2 as discretization scheme, with a standard Bernoulli quantization of the normal distribution,  $d$  being equal to 1. In the numerical studies below, we show that the *two-level algorithm* converges in case when the *one-level algorithm* fails.

**4.2.1. The example of a linear model.** In this part, we compare the output of both algorithms for the following linear model where a closed-form solution is available:

$$\begin{aligned} dX_t &= -\rho \mathbb{E}[Y]_t dt + \sigma dW_t, & X_0 &= x, \\ dY_t &= -aY_t dt + Z_t dW_t, & \text{and } Y_T &= X_T, \end{aligned}$$

for  $\rho, a > 0$ , and the true solution for  $\mathbb{E}[X_0] = m_0$  is given by

$$Y_0 = \frac{m_0 e^{aT}}{1 + \frac{\rho}{a}(e^{aT} - 1)}.$$

The errors for various time steps and for both algorithms are shown on the log-log error plot of Figure 1. The parameters are fixed as follows:  $\rho = 0.1$ ,  $a = 0.25$ ,  $\sigma = 1$ ,  $T = 1$  and  $x = 2$ . Moreover, the *two-level algorithm* uses 5 Picard Iterations per level, and the *one-level algorithm* computes 25 Picard Iterations.

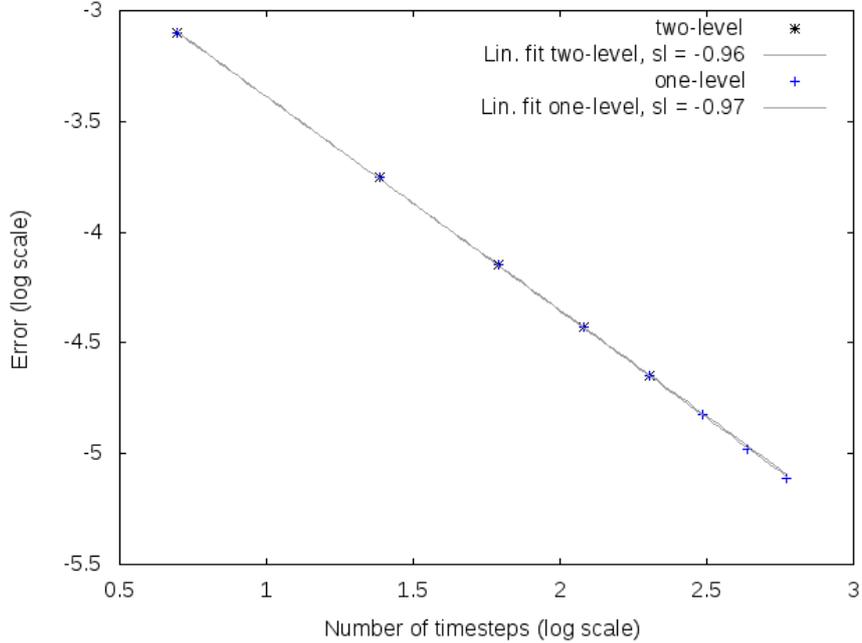


FIGURE 1. Convergence of the algorithms: log-log error plot for the same data as in the text. We can observe that both algorithms return the same value which is close to the true value. This validates the convergence of both methods in this simple linear setting.

4.2.2. *Efficiency of the `solver[]()` algorithm.* In this section, we compare the *two-level algorithm* and the *one-level algorithm* on two models, for which existence and uniqueness to the master equation (or the FBSDE system) hold true for any arbitrary terminal time  $T$  and Lipschitz constant  $L$  of the coefficients function. Nevertheless, as stated in the theorems above, the convergence of the algorithms is guaranteed only for a periods of time which are controlled by  $L$  and  $T$ . Here, we fix the terminal date  $T$  and allow  $L$  to vary with the use of a coupling parameter  $\rho$ , see equations (72) (for a case without McKean-Vlasov interaction) and (73) (for a case with McKean-Vlasov interaction). We will see below that, as expected, the *two-level algorithm* converges for a larger range of coupling parameter than the *one-level algorithm*.

An example with no McKean-Vlasov interaction.

Here, the model is the following

$$\begin{aligned} dX_t &= \rho \cos(Y_t)dt + \sigma dW_t, & X_0 &= x, \\ Y_t &= \mathbb{E}_t[\sin(X_T)]. \end{aligned} \tag{72}$$

On Figure 2, we plot the output of the two-level and one-level algorithm along with a proxy of the true solution computed by usual BSDE approximation method (after a Girsanov transform) and with a very high-level of precision. On the graph, the value  $Y_0$  stands for the approximation of  $\mathcal{U}(0, x)$ : There is no dependence upon the initial measure as there is no MKV interaction in this example. The parameters are fixed as follows:  $\sigma = 1$ ,  $T = 1$  and  $x = 0$ . Moreover, the *two-level algorithm* uses 5 Picard Iterations per level, and the *one-level algorithm* computes 25 Picard Iterations.

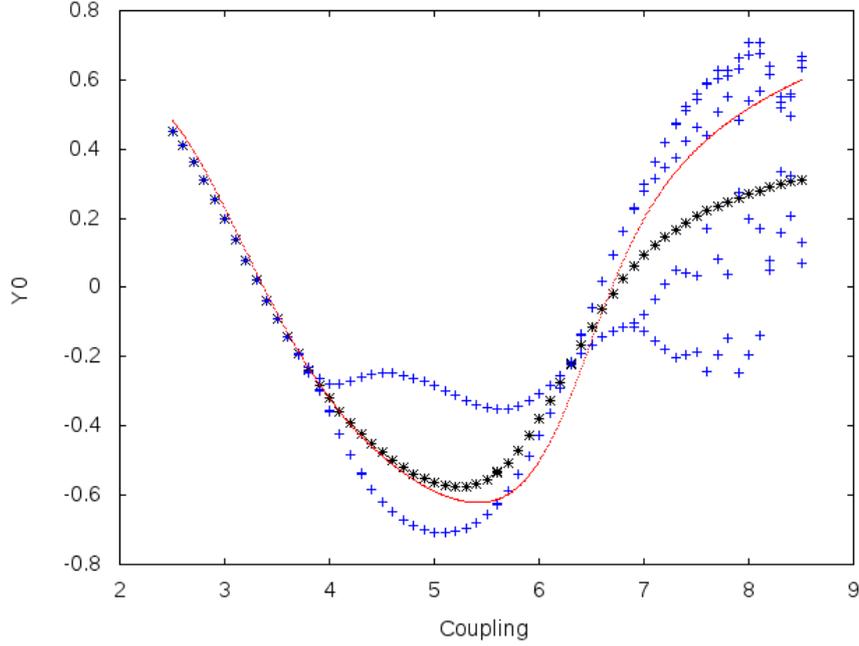


FIGURE 2. Comparison of algorithms' output for different value of the coupling parameter and for the same data as in Example (72): two-level (black star), one-level (blue cross), true value (red line). The *two-level algorithm* converges for larger coupling parameter than the *one-level algorithm*. It is close to the true solution up to parameter  $\rho = 7$ , the discrepancy for large coupling parameter coming most probably from the discrete-time error. Interestingly, the *one-level algorithm* shows bifurcations.

An example from large population stochastic control.

For this part, the model is given by

$$\begin{aligned} dX_t &= -\rho Y_t dt + dW_t, \quad X_0 = x, \\ dY_t &= \text{atan}(\mathbb{E}[X_t]) dt + Z_t dW_t \quad \text{and} \quad Y_T = G'(X_T) := \text{atan}(X_T). \end{aligned} \quad (73)$$

coming from Pontryagin principle applied to MFG

$$\inf_{\alpha} \mathbb{E} \left[ G(X_T^\alpha) + \int_0^T \left( \frac{1}{2\rho} \alpha_t^2 + X_t^\alpha \text{atan}(\mathbb{E}[X_t^\alpha]) \right) dt \right]$$

with  $dX_t^\alpha = \alpha_t dt + dW_t$ , see e.g. [14].

We do not know the exact solution for this model and it is not possible to obtain easily an approximation as in the previous example. We plot on Figure 3, the output value of the *one-level algorithm* and *two-level algorithm*. On the graph, the value  $Y_0$  stands for the approximation of  $\mathcal{U}(0, x, \delta_x)$ . The parameters are fixed as follows:  $\sigma = 1$ ,  $T = 1$  and  $x = 1$ . Moreover, the *two-level algorithm* uses 5 Picard Iterations per level, and the *one-level algorithm* computes 25 Picard Iterations.

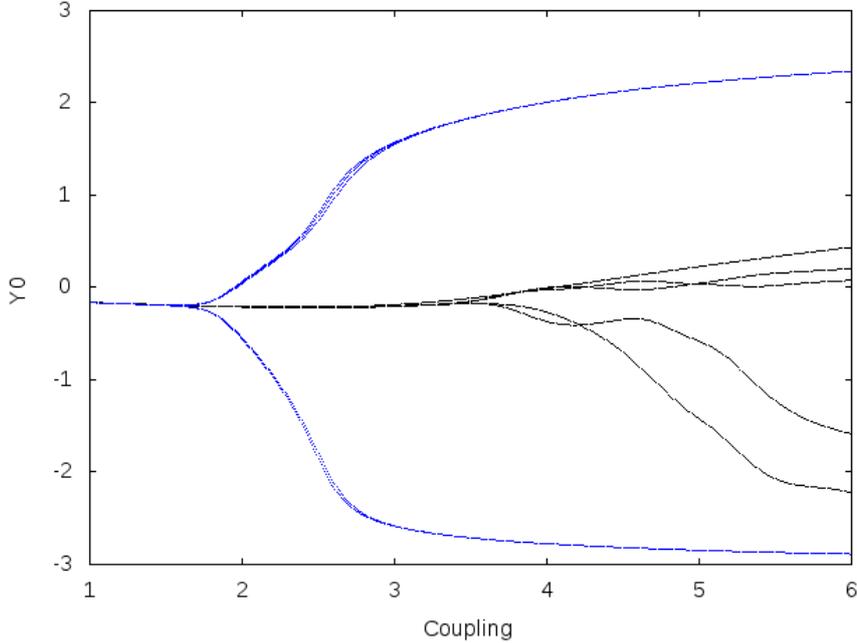


FIGURE 3. Algorithms' output for the same data as in Example (73): *one-level algorithm* (blue line), *two-level algorithm* (black line). We observe the same phenomenon as in the previous model: The *two-level algorithm* converges to a unique value for a larger range of coupling parameter than the *one-level algorithm*, which exhibits a bifurcation. Observe that the *two-level algorithm* fails to converge at some points: One should add a level of computation to shorten the time period  $\delta$ .

## 5. APPENDIX

**5.1. A discrete Itô formula.** We consider the following Euler scheme on the discrete time grid  $\pi$  of the interval  $[0, T]$ , recall (20),

$$\bar{X}_{t_{i+1}} = \bar{X}_{t_i} + b_i(t_{i+1} - t_i) + \sigma_i \sqrt{t_{i+1} - t_i} \varpi_i, \quad (74)$$

where  $(\varpi_i)_{i \leq n}$  are i.i.d. centered  $\mathbb{R}^d$ -valued random variables such that the covariance matrix  $\mathbb{E}[\varpi_i \varpi_i^\dagger]$  is the identity matrix and  $\|\varpi_i\|_{2\alpha}^2 \leq \Lambda h_i$ , and  $(b_i, \sigma_i) \in L^2(\mathcal{F}_{t_i})$ , for all  $i \leq n$ .

We also introduce a piecewise continuous version of the previous scheme, for  $i < n$ ,  $t_i \leq s < t_{i+1}$  and  $\lambda \in [0, 1]$ , the process  $(\bar{X}_t^{(\lambda)})_{0 \leq t \leq T}$ ,

$$\bar{X}_s^{(\lambda)} = \bar{X}_{t_i} + b_i(s - t_i) + \sigma_i \lambda \sqrt{s - t_i} \varpi_i \quad (75)$$

and  $\bar{X}_{t_n}^{(\lambda)} = \bar{X}_{t_n}$ . Following the notation used in the proof of Lemma 9, we just write  $(\bar{X}_s)_{0 \leq s \leq T}$  for  $(\bar{X}_s^{(1)})_{0 \leq s \leq T}$ , which defines a continuous version of the Euler scheme given in (74).

**Proposition 14.** *For any  $i \in \{0, \dots, n-1\}$ , the following holds true:*

$$\begin{aligned}
 \mathcal{U}(t_{i+1}, \bar{X}_{t_{i+1}}, [\bar{X}_{t_{i+1}}]) &= \mathcal{U}(t_i, \bar{X}_{t_i}, [\bar{X}_{t_i}]) + \int_{t_i}^{t_{i+1}} \partial_t \mathcal{U}(s, \bar{X}_s, [\bar{X}_s]) ds \\
 &+ \int_{t_i}^{t_{i+1}} \left( \partial_x \mathcal{U}(s, \bar{X}_s, [\bar{X}_s]) \cdot b_i + \frac{1}{2} \int_0^1 \text{Tr}[\partial_{xx}^2 \mathcal{U}(s, \bar{X}_s^{(\lambda)}, [\bar{X}_s]) a_i] d\lambda \right) ds \\
 &+ \int_{t_i}^{t_{i+1}} \hat{\mathbb{E}}[\partial_\mu \mathcal{U}(s, \bar{X}_s, [\bar{X}_s])(\langle \bar{X}_s \rangle) \cdot \langle b_i \rangle] ds \\
 &+ \frac{1}{2} \int_0^1 \hat{\mathbb{E}}\left[\text{Tr}[\partial_v \partial_\mu \mathcal{U}(s, \bar{X}_s, [\bar{X}_s])(\langle \bar{X}_s^{(\lambda)} \rangle) \langle a_i \rangle] d\lambda\right] ds \\
 &+ \int_{t_i}^{t_{i+1}} \partial_x \mathcal{U}(s, \bar{X}_s^{(0)}, [\bar{X}_s]) \frac{\sigma_i \varpi_i}{2\sqrt{s-t_i}} ds + \delta \mathcal{M}(t_i, t_{i+1}) + \delta \mathcal{T}(t_i, t_{i+1}),
 \end{aligned}$$

where  $a_i$  is here equal to  $\sigma_i \sigma_i^\dagger$ , and  $\delta \mathcal{M}(t_i, t_{i+1})$  is a martingale increment satisfying  $\|\delta \mathcal{M}(t_i, t_{i+1})\|_{2\alpha} \leq C_\Lambda h_i^2$  and  $\|\delta \mathcal{T}(t_i, t_{i+1})\|_{2\alpha} \leq C_\Lambda h_i^{\frac{3}{2}}$ .

**Proof.** By writing

$$\bar{X}_{t_{i+1}} = \bar{X}_{t_i} + \int_{t_i}^{t_{i+1}} \left( b_i + \frac{\sigma_i \varpi_i}{2\sqrt{s-t_i}} \right) ds,$$

and by using the standard chain rule for continuously differentiable functions on a Hilbert space, we get

$$\begin{aligned}
 \mathcal{U}(t_{i+1}, \bar{X}_{t_{i+1}}, [\bar{X}_{t_{i+1}}]) &= \mathcal{U}(t_i, \bar{X}_{t_i}, [\bar{X}_{t_i}]) + \int_{t_i}^{t_{i+1}} \partial_t \mathcal{U}(s, \bar{X}_s, [\bar{X}_s]) ds \\
 &+ \int_{t_i}^{t_{i+1}} \left( \partial_x \mathcal{U}(s, \bar{X}_s, [\bar{X}_s]) \cdot \left( b_i + \frac{\sigma_i \varpi_i}{2\sqrt{s-t_i}} \right) ds \right. \\
 &\quad \left. + \hat{\mathbb{E}} \left[ \partial_\mu \mathcal{U}(s, \bar{X}_s, [\bar{X}_s])(\langle \bar{X}_s \rangle) \cdot \left\langle b_i + \frac{\sigma_i \varpi_i}{2\sqrt{s-t_i}} \right\rangle \right] \right) ds.
 \end{aligned}$$

Now we observe that,

$$\begin{aligned}
 \partial_x \mathcal{U}(s, \bar{X}_s, [\bar{X}_s]) &= \partial_x \mathcal{U}(s, \bar{X}_s^{(0)}, [\bar{X}_s]) + \sqrt{s-t_i} \int_0^1 \partial_{xx}^2 \mathcal{U}(s, \bar{X}_s^{(\lambda)}, [\bar{X}_s]) \sigma_i \varpi_i d\lambda \\
 &= \partial_x \mathcal{U}(s, \bar{X}_s^{(0)}, [\bar{X}_s]) + \sqrt{s-t_i} \partial_{xx}^2 \mathcal{U}(s, \bar{X}_s^{(0)}, [\bar{X}_s]) \sigma_i \varpi_i + \sqrt{s-t_i} \mathcal{T}_1(s),
 \end{aligned}$$

where  $\mathcal{T}_1(s)$  is a random variable defined on  $(\Omega, \mathcal{F}, \mathbb{P})$  such that  $\|\mathcal{T}_1(s)\|_{2\alpha} \leq Ch_i^{\frac{1}{2}}$ , and

$$\begin{aligned}
 \partial_\mu \mathcal{U}(s, \bar{X}_s, [\bar{X}_s])(\langle \bar{X}_s \rangle) &= \partial_\mu \mathcal{U}(s, \bar{X}_s, [\bar{X}_s])(\langle \bar{X}_s^{(0)} \rangle) \\
 &+ \sqrt{s-t_i} \int_0^1 \partial_v \partial_\mu \mathcal{U}(s, \bar{X}_s, [\bar{X}_s])(\langle \bar{X}_s^{(\lambda)} \rangle) \langle \sigma_i \varpi_i \rangle d\lambda \\
 &= \partial_\mu \mathcal{U}(s, \bar{X}_s, [\bar{X}_s])(\langle \bar{X}_s^{(0)} \rangle) \\
 &+ \sqrt{s-t_i} \partial_v \partial_\mu \mathcal{U}(s, \bar{X}_s, [\bar{X}_s])(\langle \bar{X}_s^{(0)} \rangle) \langle \sigma_i \varpi_i \rangle + \sqrt{s-t_i} \mathcal{T}_2(s),
 \end{aligned}$$

where  $\mathcal{T}_2(s)$  is a random variable on the enlarged space  $(\Omega \times \hat{\Omega}, \mathcal{F} \otimes \hat{\mathcal{F}}, \mathbb{P} \otimes \hat{\mathbb{P}})$  such that  $\hat{\mathbb{E}}[|\mathcal{T}_2(s)|^{2\alpha}]^{1/(2\alpha)} \leq Ch_i^{\frac{1}{2}}$ .

We insert these expansions back into the identity we obtained for the term  $\mathcal{U}(t_{i+1}, \bar{X}_{t_{i+1}}, [\bar{X}_{t_{i+1}}])$ . We let

$$\begin{aligned} \delta\mathcal{M}(t_i, t_{i+1}) &= \frac{1}{2} \int_{t_i}^{t_{i+1}} \left[ \partial_{xx}^2 \mathcal{U}(s, \bar{X}_s^{(0)}, [\bar{X}_s]) \sigma_i \varpi_i \cdot (\sigma_i \varpi_i) \right. \\ &\quad \left. - \mathbb{E}_{t_i} \left[ \partial_{xx}^2 \mathcal{U}(s, \bar{X}_s^{(0)}, [\bar{X}_s]) \sigma_i \varpi_i \cdot (\sigma_i \varpi_i) \right] \right] ds, \\ \delta\mathcal{T}(t_i, t_{i+1}) &= \frac{1}{2} \int_{t_i}^{t_{i+1}} (\mathcal{T}_1(s) + \mathcal{T}_2(s)) \cdot \sigma_i \varpi_i ds. \end{aligned}$$

It defines a martingale increment satisfying  $\mathbb{E}_{t_i} [|\delta\mathcal{M}(t_i, t_{i+1})|^{2\alpha}]^{1/(2\alpha)} \leq Ch_i$ . Observing that for  $t_i \leq s \leq t_{i+1}$ ,

$$\begin{aligned} \hat{\mathbb{E}} \left[ \partial_\mu \mathcal{U}(s, \bar{X}_s, [\bar{X}_s]) (\langle \bar{X}_s^{(0)} \rangle) \cdot \langle \sigma_i \varpi_i \rangle \right] &= 0, \\ \mathbb{E}_{t_i} \left[ \partial_{xx}^2 \mathcal{U}(s, \bar{X}_s^{(0)}, [\bar{X}_s]) \sigma_i \varpi_i \cdot (\sigma_i \varpi_i) \right] &= \mathbb{E}_{t_i} \left[ \text{Tr}(\partial_{xx}^2 \mathcal{U}(s, \bar{X}_s^{(0)}, [\bar{X}_s]) a_i) \right], \\ \mathbb{E}_{t_i} \left[ \partial_{xx}^2 \mathcal{U}(s, \bar{X}_s^{(0)}, [\bar{X}_s]) \sigma_i \varpi_i \cdot (\sigma_i \varpi_i) \right] &= \mathbb{E}_{t_i} \left[ \text{Tr}(\partial_{xx}^2 \mathcal{U}(s, \bar{X}_s^{(0)}, [\bar{X}_s]) a_i) \right], \\ \hat{\mathbb{E}} \left[ \partial_v \partial_\mu \mathcal{U}(s, \bar{X}_s, [\bar{X}_s]) (\langle \bar{X}_s^{(0)} \rangle) \sqrt{s - t_i} \langle \sigma_i \varpi_i \rangle \cdot \langle \sigma_i \varpi_i \rangle \right] \\ &= \hat{\mathbb{E}} \left[ \text{Tr}(\partial_v \partial_\mu \mathcal{U}(s, \bar{X}_s, [\bar{X}_s]) (\langle \bar{X}_s^{(0)} \rangle) \langle a_i \rangle) \right], \end{aligned}$$

we complete the proof.  $\square$

## 5.2. Estimates for the scheme given in Example 2.

**Lemma 15.** *Under (H0)-(H1), the following holds for the forward component of the scheme given in Example 2 and its continuous version,*

$$\max_{t \in \pi^k} \|\bar{X}_t\|_{2\alpha} \leq C_\Lambda \left( 1 + \|\bar{X}_{r_k}\|_{2\alpha} + \delta \max_{t \in \pi^k} \|\mathcal{U}(t, \bar{X}_t, [\bar{X}_t]) - Y_t\|_{2\alpha} \right), \quad (76)$$

**Proof.** We introduce  $\mathfrak{d}_i := |\mathcal{U}(t_i, \bar{X}_{t_i}, [\bar{X}_{t_i}]) - \bar{Y}_{t_i}|$  and observe from the Lipschitz property of  $b$  and  $\mathcal{U}$  that

$$|b(\bar{X}_{t_i}, \bar{Y}_{t_i}, [\bar{X}_{t_i}, \bar{Y}_{t_i}])| \leq C_\Lambda (1 + |\bar{X}_{t_i}| + \|\bar{X}_{t_i}\|_{2\alpha} + \mathfrak{d}_i + \|\mathfrak{d}_i\|_{2\alpha}). \quad (77)$$

Recall that the scheme for the forward component reads

$$\bar{X}_{t_{i+1}} = \bar{X}_{r_k} + \sum_{\ell=j_k}^i b(\bar{X}_{t_\ell}, \bar{Y}_{t_\ell}, [\bar{X}_{t_\ell}, \bar{Y}_{t_\ell}]) (t_{\ell+1} - t_\ell) + \sum_{\ell=j_k}^i \sigma(\bar{X}_{t_\ell}, [\bar{X}_{t_\ell}]) \Delta \bar{W}_\ell.$$

Squaring the previous inequality, using Cauchy-Schwarz inequality for the first sum and the martingale property for the second sum, we obtain

$$\|\bar{X}_{t_{i+1}}\|_{2\alpha}^2 \leq C \|\bar{X}_{r_k}\|_{2\alpha}^2 + C \sum_{\ell=j_k}^i h_\ell \left( \delta \|b(\bar{X}_{t_\ell}, \bar{Y}_{t_\ell}, [\bar{X}_{t_\ell}, \bar{Y}_{t_\ell}])\|_{2\alpha}^2 + \|\sigma(\bar{X}_{t_\ell}, [\bar{X}_{t_\ell}])\|_{2\alpha}^2 \right),$$

where we used again Burkholder-Davis-Gundy inequality for discrete martingales.

Combining (77) with the boundedness of  $\sigma$ , we then have

$$\|\bar{X}_{t_{i+1}}\|_{2\alpha}^2 \leq C \left( \|\bar{X}_{r_k}\|_{2\alpha}^2 + \delta + \delta^2 \max_{j_k \leq i < j_{k+1}} \|\mathfrak{d}_i\|_{2\alpha}^2 + C\delta \sum_{\ell=j_k}^i h_\ell \|\bar{X}_{t_\ell}\|_{2\alpha}^2 \right).$$

Using the discrete version of Gronwall's lemma, the result easily follows.  $\square$

## REFERENCES

- [1] Y. ACHDOU AND I. CAPUZZO-DOLCETTA (2010) Mean field games: numerical methods *SIAM J. Numer. Anal.*, 48, pp. 1136-1162.
- [2] Y. ACHDOU, F. CAMILLI AND I. CAPUZZO-DOLCETTA (2013) Mean field games: convergence of a finite difference method, *SIAM J. Numer. Anal.*, 51, pp. 2585-2612.
- [3] Y. ACHDOU AND A. PORRETTA (2016) Convergence of a Finite Difference Scheme to Weak Solutions of the System of Partial Differential Equations Arising in Mean Field Games, *SIAM J. Numer. Anal.*, 54, pp. 161-186.
- [4] S. ALANKO *Regression-based Monte Carlo methods for solving nonlinear PDEs.*, PhD dissertation, New York University, 2015.
- [5] BAYRAKTAR, E. AND BUDHIRAJA, A. AND COHEN, A. (2016) *Rate Control under Heavy Traffic with Strategic Servers preprint*, <http://arxiv.org/abs/1605.09010>
- [6] J.D. BENAMOU AND G. CARLIER (2015) Augmented Lagrangian Methods for Transport Optimization, Mean Field Games and Degenerate Elliptic Equations, *Journal of Optimization Theory and Applications*, 167, pp. 1-26.
- [7] C. BENDER AND J. ZHANG (2008) Time discretization and Markovian iteration for coupled FBSDEs, *Ann. Appl. Probab.*, 18 (1), pp. 143-177.
- [8] A. BENSOUSSAN, J. FREHSE AND P. YAM *Mean Field Games and Mean Field Type Control Theory*, Springer Verlag, 2013.
- [9] A. BENSOUSSAN, J. FREHSE AND P. YAM (2015) The Master equation in mean field theory, *Journal de Mathématiques Pures et Appliquées*, 103, pp. 1441-1474.
- [10] A. BENSOUSSAN, J. FREHSE AND P. YAM (2016) On the interpretation of the master equation, *Stochastic Processes and their Applications*, doi = "10.1016/j.spa.2016.10.004"
- [11] R. BUCKDAHN, J. LI, S. PENG AND C. RAINER (2014) Mean-field stochastic differential equations and associated PDEs, forthcoming in *Annals of Probability*.
- [12] P. CARDALIAGUET (2012) Notes from P.L. Lions' lectures at the Collège de France, *notes*, <https://www.ceremade.dauphine.fr/~cardalia/MFG100629.pdf>.
- [13] P. CARDALIAGUET, F. DELARUE, J.-M. LASRY AND P.-L. LIONS (2015) The master equation and the convergence problem in mean field games, *preprint*, <http://arxiv.org/abs/1509.02505>.
- [14] R. CARMONA AND F. DELARUE (2013) Probabilistic Analysis of Mean Field Games, *SIAM Journal on Control and Optimization*, 51 (4), pp. 2705-2734.
- [15] R. CARMONA, AND F. DELARUE (2013) Mean Field Forward-Backward Stochastic Differential Equations, *Electronic Communications in Probability*.
- [16] R. CARMONA AND F. DELARUE The master equation for large population equilibria, In: *Stochastic Analysis and Applications*, Springer Verlag, pp. 77-128, 2014.
- [17] R. CARMONA AND F. DELARUE *Probabilistic Theory of Mean Field Games: vol. I, Mean Field FBSDEs, Control, and Games*, Springer Verlag, 2017.
- [18] R. CARMONA AND F. DELARUE *Probabilistic Theory of Mean Field Games: vol. II, Mean Field Games with Common Noise and Master Equations*, Springer Verlag, 2017.
- [19] R. CARMONA, AND F. DELARUE, AND A. LACHAPELLE (2013) Control of McKean-Vlasov versus Mean Field Games, *Mathematics and Financial Economics*, 7, pp. 131-166.
- [20] J.-F. CHASSAGNEUX, D. CRISAN AND F. DELARUE (2014) A Probabilistic approach to classical solutions of the master equation for large population equilibria, *preprint*.
- [21] F. DELARUE AND S. MENOZZI (2006) A forward-backward stochastic algorithm for quasi-linear PDEs, *Ann. Appl. Probab.*, 16 (1), pp 140-184.
- [22] F. DELARUE AND S. MENOZZI (2008). An Interpolated Stochastic Algorithm for Quasi-Linear PDEs. *Mathematics of Computation*, 77, pp. 125-158.
- [23] D.A. GOMES, L. NURBEKYAN AND E. PIMENTEL, *Economic Models and Mean-field Games Theory*, Publicações Matemáticas, IMPA, Rio, Brazil, 2015.
- [24] D.A. GOMES, E. PIMENTEL AND V. VOSKANYAN *Regularity Theory for Mean-Field Game Systems*, Springer International Publishing Switzerland, 2016.
- [25] O. GUÉANT (2012) New numerical methods for mean field games with quadratic costs, *Networks and Heterogeneous Media*, 2, pp. 315-336.
- [26] A. LACHAPELLE, J. SALOMON AND G. TURINICI (2010) Computation of mean field equilibria in economics, *Mathematical Models and Methods in Applied Sciences*, 20, pp. 567-588.
- [27] J.M. LASRY AND P.L. LIONS (2006) Jeux à champ moyen I. Le cas stationnaire, *Comptes Rendus de l'Académie des Sciences de Paris, ser. A*, 343 (9).

- [28] J.M. LASRY AND P.L. LIONS (2006) Jeux à champ moyen II. Horizon fini et contrôle optimal, *Comptes Rendus de l'Académie des Sciences de Paris, ser. A*, 343 (10).
- [29] J.M. LASRY AND P.L. LIONS (2007) Mean Field Games, *Japanese Journal of Mathematics*, 2 (1), pp. 229-260.
- [30] P.L. LIONS (2014) Estimées nouvelles pour les équations quasiliéaires *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>
- [31] J. MA, H. YIN AND J. ZHANG (2012) On non-Markovian forward-backward SDEs and backward stochastic PDEs. *Stochastic Processes and their Applications*, 122, pp. 3980-4004.
- [32] J. MA, P. PROTTER AND J. YONG (1994) Solving forward-backward stochastic differential equations explicitly – a four step scheme. *Probab. Theory Related Fields*, 98, pp. 339-359.
- [33] H. PHAM, AND X. WEI (2015) Bellman equation and viscosity solutions for mean field stochastic control problem, *preprint*, <http://arxiv.org/abs/1512.07866>