The Annals of Applied Probability
2019, Vol. 29, No. 3, 1640–1684
https://doi.org/10.1214/18-AAP1429
© Institute of Mathematical Statistics, 2019

NUMERICAL METHOD FOR FBSDES OF MCKEAN–VLASOV TYPE

BY JEAN-FRANÇOIS CHASSAGNEUX, DAN CRISAN AND FRANÇOIS DELARUE

Université Paris Diderot, Imperial College London and Université de Nice Sophia-Antipolis

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 nonlocal PDE:

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

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 ν is a notation for the image of the probability measure μ 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, that is,

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

Received June 2017; revised March 2018.
MSC2010 subject classifications. 60H10, 65C30, 91A13.
Key words and phrases. McKean–Vlasov, master equation, forward–backward SDE, mean field games, numerical approximation.

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 π on $\mathbb{R}^d \times \mathbb{R}^d$ whose marginals on $\mathbb{R}^d$ are respectively μ and μ'.

While the first two lines in (1) form a classical nonlinear parabolic equations, the last two terms are nonstandard. Not only are they nonlocal, in the sense that the solution or its derivatives are computed at points υ different from x, but also they involve derivatives in the argument μ, which lives in a space of probability measures. In this regard, the notation $\partial_\mu \mathcal{U}(t, x, \mu)(\upsilon)$ denotes the so-called *Wasserstein derivative* of the function $\mathcal{U}$ in the direction of the measure, computed at point (t, x, μ) and taken at the *continuous coordinate* υ. We provide below a short reminder of the construction of this derivative, as introduced by Lions; see [14] or [18], Chapter 5.

These PDEs arise in the study of large population stochastic control problems, either of mean field game type (see, for instance, [14, 15, 21, 27] or [19], Chapter 12, and the references therein) or of mean field control type; see, for instance, [11, 12, 21, 29]. 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 monographes 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 [21], 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 [19], Chapter 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; see [13] for linear equations and [21] for nonlinear equations. 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:

$$X_s^{t,x,\mu} = x + \int_t^s b\big(X_r^{t,x,\mu}, Y_r^{t,x,\mu}, [X_r^{t,\xi}, Y_r^{t,\xi}]\big)\,\mathrm{d}r + \int_t^s \sigma\big(X_r^{t,x,\mu}, [X_r^{t,\xi}]\big)\,\mathrm{d}W_r, \tag{2}$$

$$\begin{aligned} Y_s^{t,x,\mu} = {} & g\big(X_T^{t,x,\mu}, [X_T^{t,\xi}]\big) \\ & + \int_s^T f\big(X_r^{t,x,\mu}, Y_r^{t,x,\mu}, Z_r^{t,x,\mu}, [X_r^{t,\xi}, Y_r^{t,\xi}]\big)\,\mathrm{d}r \\ & - \int_s^T Z_r^{t,x,\mu} \cdot \mathrm{d}W_r, \end{aligned} \tag{3}$$

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

$$X_s^{t,\xi} = \xi + \int_t^s b\big(X_r^{t,\xi}, Y_r^{t,\xi}, [X_r^{t,\xi}, Y_r^{t,\xi}]\big)\,\mathrm{d}r + \int_t^s \sigma\big(X_r^{t,\xi}, [X_r^{t,\xi}]\big)\,\mathrm{d}W_r, \tag{4}$$

$$Y_s^{t,\xi} = g\big(X_T^{t,\xi}, [X_T^{t,\xi}]\big) + \int_s^T f\big(X_r^{t,\xi}, Y_r^{t,\xi}, Z_r^{t,\xi}, [X_r^{t,\xi}, Y_r^{t,\xi}]\big)\,\mathrm{d}r - \int_s^T Z_r^{t,\xi}\cdot \mathrm{d}W_r, \tag{5}$$

where $(W_t)_{0\le t\le T}$ is a Brownian motion and ξ has μ as distribution. In the previous equations and in the sequel, we use the notation $[\theta]$ for the law of a random variable θ. 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 [6] 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 [6], 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. In 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 the 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 μ, 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 [6].

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 [7] 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 [7] 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 [22, 23] and [9], 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 quite 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, 4, 5, 24] and [3, 8, 26]. In [1, 2, 5], the authors provide a discretization scheme in the form of a *coupled* forward–backward system of two difference equations; convergence is proved. Still, in order to make such a discretization scheme fully implementable, it is necessary to decouple the two difference equations therein—which is similar to the issue we face here with the probabilistic approach: In [4], a numerical method of Newton type is proposed to decouple the two discrete equations; numerical examples are given but the convergence is not investigated from the theoretical point of view; Newton's method is also used in [24] to handle mean field games with quadratic Hamiltonians; In [30], Picard's iterations (of a very similar form to ours) are used to decouple the discretized forward–backward system, but again convergence is just addressed through numerical examples. In [3, 8, 26], numerical schemes are investigated for so-called potential mean field games, namely for mean field games whose equilibria coincide with the minimizers of a control problem of mean field type; schemes are obtained by discretizing the corresponding optimization problem and convergence is checked on a series of examples. As none of the aforementioned references explicitly address the rate of convergence of the implemented version of the underlying scheme, it is pretty difficult to make a sharp comparison with our own results. However, we are convinced that all these results offer relevant prospects for further probabilistic investigations: Potential mean field games should deserve a special care and Newton's (instead of Picard's) method should be also considered within our framework. Conversely, it could be worth testing the continuation method we describe below to decouple the forward–backward difference equations that arise from the PDE approach. In a similar spirit, we also draw the following two works to the reader's attention. In [16], a decoupling procedure for the mean field game system (comprising two PDEs: a forward Fokker–Planck equation and a backward Hamilton–Jacobi–Bellman equation) is proven to be convergent (without any rate) in the potential case again; this procedure is based upon an iteration rule that consists in taking as proxy for the mean field interaction the mean of the distributions over all the previous iterations. This is in contrast with Picard's method that we use below: In Picard's method, the sole return of the last iteration serves as a proxy. We feel that it should be worth studying this updating rule in our framework. This is all the more true that, in the recent preprint [25] that was published on arXiv at the same period as our work, the author succeeded to extend the result obtained in [16] to first order (i.e., without any diffusive term) mean field games with monotone coefficients (but possibly nonpotential).

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:

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

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 $(W_t)_{0\le t\le T}$. To simplify, we assume that the state process $(X_t)_{0\le t\le T}$ is of the same dimension. The process $(Y_t)_{0\le t\le T}$ is 1-dimensional. As a result, $(Z_t)_{0\le t\le T}$ is d-dimensional.

In (6), the three processes $(X_t)_{0\le t\le T}$, $(Y_t)_{0\le t\le T}$ and $(Z_t)_{0\le t\le T}$ are required to be $\mathbb{F}$-progressively measurable. Both $(X_t)_{0\le t\le T}$ and $(Y_t)_{0\le t\le T}$ 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, $(X_t)_{0\le t\le T}$, $(Y_t)_{0\le t\le T}$ and $(Z_t)_{0\le t\le T}$ must satisfy

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

The domains and codomains of the coefficients are defined accordingly. The assumption that σ is 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 [21] for a complete overview on the subject, together with the references [10, 14, 18–20]. In light of the examples tackled in [21], 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 [22, 23] 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, for example, Proposition 2.2 in [21],

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

where $\mathcal{U} : [0,T]\times\mathbb{R}^d\times\mathcal{P}_2(\mathbb{R}^d)\to\mathbb{R}$ is assumed to be the classical solution, in the sense of [21], 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

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

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, ξ) 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, μ, v) such that v is in the support of μ.

ASSUMPTIONS. Our analysis requires some minimal regularity assumptions on the coefficients b, σ, 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, σ, f and g are Λ-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 σ is bounded by Λ.

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 [21], Definition 2.6, to (1); and $\mathcal{U}$ and its derivatives satisfy

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

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

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

and

$$|\partial^2_{xx} \mathcal{U}(t, x, \mu) - \partial^2_{xx} \mathcal{U}(t, x', \mu)| \leq \Lambda |x - x'|, \tag{11}$$

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

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

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 $\upsilon, \upsilon' \in \mathbb{R}^d$,

$$
\begin{aligned}
&\big|\partial_\upsilon \partial_\mu \mathcal{U}\big(t, x, [\xi]\big)(\upsilon) - \partial_\upsilon \partial_\mu \mathcal{U}\big(t, x, [\xi]\big)\big(\upsilon'\big)\big| \\
&\qquad \leq \Lambda\big\{1 + |\upsilon|^{2\alpha} + \big|\upsilon'\big|^{2\alpha} + \|\xi\|_2^{2\alpha}\big\}^{\frac{1}{2}}\big|\upsilon - \upsilon'\big|,
\end{aligned} \tag{13}
$$

for some $\alpha > 0$.

REMARK 1. In [21], see also [15] for the periodic case and the forthcoming [19], Chapter 12, for a new point of view on [21], it is shown that, under some conditions on the coefficients b, f and σ, the PDE (1) has indeed a unique classical solution which satisfies assumption (H1). Generally speaking, those results apply under the following two main assumptions: Coefficients are to be smooth enough, which includes smoothness in the direction of the measure argument, and either time is small enough or some form of monotonicity holds true in the direction μ. Basically, monotonicity precludes the emergence of singularities over a time interval of arbitrary length: When the time interval $[0, T]$ is small enough, smoothness is indeed transmitted from the coefficients to the solution; however, singularities may develop as time grows up, as it happens in inviscid conservation laws (think for instance of Burgers equation), except if the coefficients are monotone in a suitable sense. Here, the measure argument does not feel the noise (since the dynamics of the law of the solution in (6) is purely deterministic): Similar to inviscid conservation laws, an additional monotonicity condition is thus needed to prevent any singularities. For simplicity, we feel better not to expand any of the sets of assumptions introduced in [15, 19, 21], as it would require an additional lengthy list of notation for the underlying set-up. Still, we make clear how to prove each of the above assumptions from existing references:

(1) Estimate (13) is obtained by combining Definition 2.6 and Proposition 3.9 in [21]. A major difficulty in the analysis provided below is the fact that α 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 [22, 23]. Below, we tackle two cases: the case when $\alpha \leq 1$, which has been investigated in [15] and [19], Chapter 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 [21]. 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 μ, it suffices to prove

$$\begin{aligned}
&\big|\mathcal{U}(t+h,x,[X^{t,\xi}_{t+h}]) - \mathcal{U}(t,x,[X^{t,\xi}_t])\big| \\
&\quad + \big|\partial_x\mathcal{U}(t+h,x,[X^{t,\xi}_{t+h}]) - \partial_x\mathcal{U}(t,x,[X^{t,\xi}_t])\big| \\
&\quad \le \Lambda h^{\frac{1}{2}}\big(1+|x|+\|\xi\|_2\big).
\end{aligned}$$

As stated in Proposition 2.2 in [21], for $\xi \sim \mu$, $\mathcal{U}(s,x,[X^{t,\xi}_s]) = u_{t,\mu}(s,x)$ where $u_{t,\mu}$ is solution to a quasi-linear PDE. Then the estimate (12) follows from standard results on nonlinear PDEs; see, for example, Theorem 2.1 in [22].

In comparison with the assumption used in [22], the condition (H1) is more demanding. In [22], 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, while in [22], 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 [23].

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 [22, 23], 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 nontractable 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 $\Re = \{r_0 = 0, \dots, r_N = T\}$ 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 δ 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 $L^2(\Omega, \mathcal{F}_{r_k}, \mathbb{P}; \mathbb{R}^d)$):

$$X_t = \xi + \int_{r_k}^t b\bigl(X_s, Y_s, [X_s, Y_s]\bigr)\,\mathrm{d}s + \int_{r_k}^t \sigma\bigl(X_s, [X_s]\bigr)\,\mathrm{d}W_s, \tag{14}$$

$$Y_t = g\bigl(X_T, [X_T]\bigr) + \int_t^T f\bigl(X_s, Y_s, Z_s, [X_s, Y_s]\bigr)\,\mathrm{d}s - \int_t^T Z_s \cdot \mathrm{d}W_s. \tag{15}$$

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

$$\begin{cases} X_t = \xi + \displaystyle\int_{r_k}^t b\bigl(X_s, Y_s, [X_s, Y_s]\bigr)\,\mathrm{d}s + \int_{r_k}^t \sigma\bigl(X_s, [X_s]\bigr)\,\mathrm{d}W_s, \\ Y_t = \mathcal{U}\bigl(r_{k+1}, X_{r_{k+1}}, [X_{r_{k+1}}]\bigr) \\ \qquad + \displaystyle\int_t^{r_{k+1}} f\bigl(X_s, Y_s, Z_s, [X_s, Y_s]\bigr)\,\mathrm{d}s - \int_t^{r_{k+1}} Z_s \cdot \mathrm{d}W_s, \end{cases}$$

for $t \in [r_k, r_{k+1}]$, recall (7).

If δ is small enough, a natural approach is to introduce a Picard iteration scheme to approximate the solution of the above equation. Indeed, we can regard the above forward–backward system as a system of type (6), set on a small time interval of length δ and driven by the terminal boundary condition $\mathcal{U}(r_{k+1}, \cdot, \cdot)$. So, in the ideal case when $\mathcal{U}(r_{k+1}, \cdot, \cdot)$ is perfectly known, one can implement the following Picard recursion (with respect to the index j):

$$\begin{cases} X_t^j = \xi + \displaystyle\int_{r_k}^t b\bigl(X_s^j, Y_s^j, [X_s^j, Y_s^j]\bigr)\,\mathrm{d}s + \int_{r_k}^t \sigma\bigl(X_s^j, [X_s^j]\bigr)\,\mathrm{d}W_s, \\ Y_t^j = \mathcal{U}\bigl(r_{k+1}, X_{r_{k+1}}^{j-1}, [X_{r_{k+1}}^{j-1}]\bigr) \\ \qquad + \displaystyle\int_t^{r_{k+1}} f\bigl(X_s^{j-1}, Y_s^j, Z_s^j, [X_s^{j-1}, Y_s^j]\bigr)\,\mathrm{d}s - \int_t^{r_{k+1}} Z_s^j \cdot \mathrm{d}W_s, \end{cases} \tag{16}$$

with the initialization $(X_s^0 = \xi + \int_{r_k}^t b(X_s^0, 0, [X_s^0, 0])\,\mathrm{d}s + \int_{r_k}^t \sigma(X_s^0, [X_s^0])\,\mathrm{d}W_s)_{r_k \leq s \leq r_{k+1}}$ and $(Y_s^0 = 0)_{r_k \leq s \leq r_{k+1}}$. It is known that, for δ small enough, $(X^j, Y^j, Z^j) \to_{j \to \infty} (X, Y, Z)$, in the sense that $\|\!|(X^j - X, Y^j - Y, Z^j - Z)\|\!|_{[r_k, r_{k+1}]} \to_{j \to \infty} 0$.

As we just said, this works in an ideal case only and, 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 how to overcome (1) and (2) above. The main idea is to use a backward recursive algorithm (with a new recursion, but on the time parameter) in order to compute at each index k of the time mesh the value of the terminal boundary at the jth step of the Picard iteration (16).

For sure, to solve the jth step of (16), we need an approximation, which we call below a *solver* and which we denote by $\texttt{solver}[k+1]$, for $\mathcal{U}(r_{k+1},\cdot,\cdot)$. The point is thus to define backwards these solvers as $k+1$ runs from N to 1. While it is obvious that $\texttt{solver}[N]$ should be given by the terminal condition of (6), $\texttt{solver}[N-1]$, $\texttt{solver}[N-2],\dots,$ must be defined recursively. To make it clear, we assume that, for $k \le N-1$, we are given a solver which computes an approximation of $\mathcal{U}(r_{k+1},\cdot,\cdot)$ in the form

$$\texttt{solver}[k+1](\xi) = \mathcal{U}\big(r_{k+1},\xi,[\xi]\big) + \varepsilon^{k+1}(\xi), \tag{17}$$

where ε is the error made by the solver. The above decomposition should hold for any $\xi \in L^2(\mathcal{F}_{r_{k+1}})$. We shall sometimes refer to $\texttt{solver}[k+1](\cdot)$ as "the solver at level $k+1$".

Taking these observations into account, the next step is to define the solver at level k, namely $\texttt{solver}[k](\cdot)$. To do so, we first define the solver at level k in the form of an ideal (and thus not implementable) solver. Basically, this assumes that each Picard iteration in the approximation (16) of the solution of the forward–backward system can be perfectly computed. To distinguish this *ideal* solver at level k from the *implementable* solver at level k that we need in the end, we denote the ideal one by $\texttt{picard}[k]()$. Accordingly, we do (for the time being) as if, to compute the terminal boundary condition in (16), we could directly use $\texttt{picard}[k+1]()$ instead of the implementable version $\texttt{solver}[k+1]()$. Given $\texttt{picard}[k+1]()$, $\texttt{picard}[k]()$ is defined as follows:

$$\begin{cases} \tilde{X}_t^{k,j} = \xi + \displaystyle\int_{r_k}^t b\big(\tilde{X}_s^{k,j},\tilde{Y}_s^{k,j},[\tilde{X}_s^{k,j},\tilde{Y}_s^{k,j}]\big)\,\mathrm{d}s \\ \qquad\quad + \displaystyle\int_{r_k}^t \sigma\big(\tilde{X}_s^{k,j},[\tilde{X}_s^{k,j}]\big)\,\mathrm{d}W_s, \\ \tilde{Y}_t^{k,j} = \texttt{picard}[k+1]\big(\tilde{X}_{r_{k+1}}^{k,j-1}\big) - \displaystyle\int_t^{r_{k+1}} \tilde{Z}_s^{k,j}\cdot \mathrm{d}W_s \\ \qquad\quad + \displaystyle\int_t^{r_{k+1}} f\big(\tilde{X}_s^{k,j-1},\tilde{Y}_s^{k,j},\tilde{Z}_s^{k,j},[\tilde{X}_s^{k,j-1},\tilde{Y}_s^{k,j}]\big)\,\mathrm{d}s, \end{cases} \tag{18}$$

for $j \geq 1$ and with

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

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

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

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

$$(19) \quad \begin{cases} \tilde{X}_t^{N-1,j} = \xi + \displaystyle\int_{r_{N-1}}^t b(\tilde{X}_s^{N-1,j}, \tilde{Y}_s^{N-1,j}, [\tilde{X}_s^{N-1,j}, \tilde{Y}_s^{N-1,j}])\,\mathrm{d}s \\ \qquad\qquad + \displaystyle\int_{r_{N-1}}^t \sigma(\tilde{X}_s^{N-1,j}, [\tilde{X}_s^{N-1,j}])\,\mathrm{d}W_s, \\ \tilde{Y}_t^{N-1,j} = g(\tilde{X}_T^{N-1,j-1}, [\tilde{X}_T^{N-1,j-1}]) - \displaystyle\int_t^T \tilde{Z}_s^{N-1,j} \cdot \mathrm{d}W_s \\ \qquad\qquad + \displaystyle\int_t^T f(\tilde{X}_s^{N-1,j-1}, \tilde{Y}_s^{N-1,j}, \tilde{Z}_s^{N-1,j}, \\ \qquad\qquad [\tilde{X}_s^{N-1,j-1}, \tilde{Y}_s^{N-1,j}])\,\mathrm{d}s. \end{cases}$$

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

$$(20) \qquad \texttt{picard[}N\texttt{](}\xi\texttt{)} = g(\xi, [\xi]) \quad \text{and} \quad \varepsilon^N(\xi) = 0.$$

Practical implementation. As already noticed in item (3) of the list of difficulties spelled out right after (16), it is not possible to solve the backward and forward equations in (18) perfectly, even though the system is decoupled. Hence, we need to introduce an approximation that can be implemented in practice and that will give rise to the true `solver[]()` (instead of the fake one `picard[]()`). Given a continuous adapted input process $\mathfrak{X} = (\mathfrak{X}_s)_{r_k \leq s \leq r_{k+1}}$ such that $\mathbb{E}[\sup_{r_k \leq s \leq r_{k+1}} |\mathfrak{X}_s|^2] < \infty$ and $\eta \in L^2(\Omega, \mathcal{F}_{r_{k+1}}, \mathbb{P}; \mathbb{R})$, we thus would like to solve

$$(21) \quad \begin{cases} \tilde{X}_t = \mathfrak{X}_{r_k} + \displaystyle\int_{r_k}^t b(\tilde{X}_s, \tilde{Y}_s, [\tilde{X}_s, \tilde{Y}_s])\,\mathrm{d}s + \int_{r_k}^t \sigma(\tilde{X}_s, [\tilde{X}_s])\,\mathrm{d}W_s, \\ \tilde{Y}_t = \eta + \displaystyle\int_t^{r_{k+1}} f(\mathfrak{X}_s, \tilde{Y}_s, \tilde{Z}_s, [\mathfrak{X}_s, \tilde{Y}_s])\,\mathrm{d}s - \int_t^{r_{k+1}} \tilde{Z}_s \cdot \mathrm{d}W_s, \end{cases}$$

for $t \in [r_k, r_{k+1}]$. To make it clear, the reader may think of (21) as a variant of (16), in which $\mathfrak{X}$ is given by the forward component X^{j-1} of the $(j-1)$th Picard iteration and η is obtained by calling $\texttt{solver[}k+1\texttt{](}X_{r_{k+1}}^{j-1}\texttt{)}$. Observing that all these data should be indeed available before running the jth Picard iteration, we call them *inputs*. Continuing the parallel with (16), the output $(\tilde{X}, \tilde{Y}, \tilde{Z})$ should be identified with the result (X^j, Y^j, Z^j) of the jth Picard iteration in (16).

The numerical computation of the output $(\tilde{X}, \tilde{Y}, \tilde{Z})$ in terms of the input should go along the following general principle. First, it requires a refinement of the original time mesh $\Re$ in order to discretize (21) between r_k and r_{k+1}. This prompts us to let π be a discrete time grid of $[0, T]$ such that $\Re \subset \pi$,

$$\pi := \{t_0 := 0 < \cdots < t_n := T\} \quad \text{and} \quad |\pi| := \max_{i<n}(t_{i+1} - t_i). \tag{22}$$

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

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

for all $k < N$. So, instead of a perfect solver for an iteration of the Picard scheme (18), we assume that we are given a numerical solver, denoted by $\overline{\texttt{solver}}[k](\bar{\mathfrak{X}}, \eta, f)$, which computes an approximation of the process $(\tilde{X}_s, \tilde{Y}_s, \tilde{Z}_s)_{r_k\le s\le r_{k+1}}$ on π^k for a discretization $(\bar{\mathfrak{X}}_t)_{t\in\pi^k}$ of the time continuous process $(\mathfrak{X}_s)_{r_k\le s\le r_{k+1}}$. 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 η, the discrete-time process $(\bar{\mathfrak{X}}_t)_{t\in\pi^k}$ and the driver f of the backward equation (in other words, we include for later purpose the driver f in the list of inputs). In short, the output is what the numerical solver returns after one iteration in the Picard scheme when the discrete input is $(\eta, \bar{\mathfrak{X}}, f)$. Pay attention that, in contrast with b and σ, 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 $(\eta, (\bar{\mathfrak{X}}_t)_{t\in\pi^k})$.

An example for $\overline{\texttt{solver}}[\,](,,)$ is given in Example 2 below.

Full algorithm for `solver[]()`. Using $\overline{\texttt{solver}}[\,](,,)$ for each level, we can now give a completely implementable algorithm for `solver[]()`. Intuitively, it suffices to duplicate the construction of `picard[`k`]()` in terms of `picard[`$k+1$`]()` in order to define `solver[`k`]()` in terms of `solver[`$k+1$`]()`. To do so, we must go back to (18), replace therein `picard[`$k+1$`]()` by `solver[`$k+1$`]()` and then regard the numerical solution of the forward–backward system (18) as the result of $\overline{\texttt{solver}}[k](\bar{\mathfrak{X}}, \eta, f)$, with $\bar{\mathfrak{X}}$ given by (a discrete version of) the forward component $\tilde{X}^{k,j-1}$ and η obtained by calling `solver[`$k+1$`](`$X^{k,j-1}_{r_{k+1}}$`)`.

The precise description of the algorithm is as follows. The value `solver[`k`](`ξ`)`, that is, 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 $\bar{Y}^{k,0}_t = 0$ for $t \in \pi_k$ and regard $(\bar{X}^{k,0}_t)_{t\in\pi_k}$ as the forward component of $\overline{\texttt{solver}}[k](\xi, 0, 0)$,
(2) for $1 \le j \le J$,
    (a) compute $\bar{Y}^{k,j}_{r_{k+1}} = \texttt{solver}[k+1](\bar{X}^{k,j-1}_{r_{k+1}})$,

(b) compute $(\bar{X}^{k,j}, \bar{Y}^{k,j}, \bar{Z}^{k,j}) = \overline{\texttt{solver}}\texttt{[}k\texttt{]}\texttt{(}\bar{X}^{k,j-1}\texttt{,}\,\bar{Y}^{k,j}_{r_{k+1}}\texttt{,}\,f\texttt{)}$,

(3) return $\bar{Y}^{k,J}_{r_k}$ for $\texttt{solver[}k\texttt{](}\xi\texttt{)}$.

Following (20), we let

$$\texttt{solver[}N\texttt{](}\xi\texttt{)} = g\big(\xi, [\xi]\big). \tag{23}$$

Obviously, (23) reads as the initialization of the backward induction used to compute $\texttt{solver[}k\texttt{]()}$, for k running from N to 0. We now explain the initialization step (1) in the construction of $\texttt{solver[}k\texttt{]()}$, for a given k. 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 $\overline{\texttt{solver}}\texttt{[](,0,0)}$ 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 $\overline{\texttt{solver}}\texttt{[](,,)}$. 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 ∞, which is obviously the main drawback of this method. Again, we refer to Section 4 for numerical illustrations.

The full convergence analysis, including the discretization error, will be discussed in Section 3 in the following two cases: first, for a generic (or abstract) solver $\overline{\texttt{solver}}\texttt{[](,,)}$ and second for an explicit solver, as given in the example below.

EXAMPLE 2. This example is the prototype of the solver $\overline{\texttt{solver}}\texttt{[](,,)}$ used as an elementary block inside the construction of $\texttt{solver[]()}$. It can be used to decouple the system (21) or, equivalently, in Step (2b) of the full algorithm for $\texttt{solver[]()}$.

To start with, 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 be expressed as

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

where

$$\Delta \bar{W}_j := h_j^{\frac{1}{2}} \varpi_j, \qquad \varpi_j := \Gamma_d\big(h_j^{-\frac{1}{2}}(W_{t_{j+1}} - W_{t_j})\big), \tag{24}$$

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

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

where Γ_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, Γ_d is intended to take a small number of values. The typical example is the so-called *binomial approximation*, in which case Γ_1 is the sign function.

On each interval $[r_k, r_{k+1}]$, given a discrete-time input process $\bar{\mathfrak{X}}$ and a terminal condition η, 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 $j_k \le i < j_{k+1}$, compute recursively

$$\bar{Y}_{t_i} = \mathbb{E}_{t_i}\big[\bar{Y}_{t_{i+1}} + (t_{i+1} - t_i) f\big(\bar{\mathfrak{X}}_{t_i}, \bar{Y}_{t_i}, \bar{Z}_{t_i}, [\bar{\mathfrak{X}}_{t_i}, \bar{Y}_{t_i}]\big)\big],$$

$$\bar{Z}_{t_i} = \mathbb{E}_{t_i}\left[\frac{\Delta \bar{W}_i}{t_{i+1} - t_i} \bar{Y}_{t_{i+1}}\right].$$

(2) For the forward component:

(a) Set as initial condition, $\bar{X}_{t_{j_k}} = \bar{\mathfrak{X}}_{r_k}$.

(b) For $j_k < i \le j_{k+1}$, compute recursively

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

Useful notation. 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 pointwise copy on $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{P}})$. To make it clear, $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{P}})$ is equal to $(\Omega, \mathcal{F}, \mathbb{P})$, the hat being just used for notational convenience, and for a random variable X defined on $(\Omega, \mathcal{F}, \mathbb{P})$, $\langle X \rangle$ is defined as $\hat{\Omega} \ni \hat{\omega} \mapsto \langle X \rangle(\hat{\omega}) = X(\hat{\omega})$. The latter makes sense since $\hat{\omega} \in \hat{\Omega} = \Omega$.

We shall use the notation C_Λ, c_Λ for constants only depending on Λ (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_Λ, even if we use the more precise notation C_Λ 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 (18) can be perfectly computed or equivalently when the solver is given by `solver[`k`]()` $=$ `picard[`k`]()`. 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 (18).

The main result of the section (see Theorem 5) is an upper bound for the error when we use `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 $j \in \{1, \dots, J\}$, $k \in \{1, \dots, N-1\}$,*

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

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

$$\Delta_k^j \leq \bar{\delta}^j \Delta_k^0 + \sum_{\ell=1}^{j} \bar{\delta}^{\ell-1} e^{\bar{\delta}} \big\| \varepsilon^{k+1}\big(\tilde{X}_{r_{k+1}}^{k,j-\ell}\big) \big\|_2. \tag{25}$$

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

$$\varepsilon^k(\xi) = \texttt{picard[}k\texttt{](}\xi\texttt{)} - \mathcal{U}\big(r_k, \xi, [\xi]\big) \qquad \text{with } \varepsilon^N(\xi) = 0.$$

REMARK 4. A careful inspection of the proof shows that, whenever σ depends on Y or b depends on Z, the same result holds true but with a constant C_Λ 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 OF PROPOSITION 3. 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 notation $(\tilde{X}^{k,j}, \tilde{Y}^{k,j}, \tilde{Z}^{k,j})$ and Δ_k^j.

Applying Itô's formula for functions of a measure argument (see [13, 21]), we have

$$\begin{aligned} &d\mathcal{U}\big(t, \tilde{X}_t^j, [\tilde{X}_t^j]\big) \\ &\quad = \Big(b\big(\tilde{X}_t^j, \tilde{Y}_t^j, [\tilde{X}_t^j, \tilde{Y}_t^j]\big) \cdot \partial_x \mathcal{U}\big(t, \tilde{X}_t^j, [\tilde{X}_t^j]\big) \end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{2}\mathrm{Tr}\big[a(\tilde X_t^j,[\tilde X_t^j])\partial^2_{xx}\mathcal{U}(t,\tilde X_t^j,[\tilde X_t^j])\big]\\
&+\hat{\mathbb{E}}\big[b(\langle\tilde X_t^j\rangle,\langle\tilde Y_t^j\rangle,[\tilde X_t^j,\tilde Y_t^j])\cdot\partial_\mu\mathcal{U}(t,\tilde X_t^j,[\tilde X_t^j])(\langle\tilde X_t^j\rangle)\big]\\
&+\hat{\mathbb{E}}\bigg[\frac{1}{2}\mathrm{Tr}\big[a(\langle\tilde X_t^j\rangle,[\tilde X_t^j])\partial_\upsilon\partial_\mu\mathcal{U}(t,\tilde X_t^j,[\tilde X_t^j])(\langle\tilde X_t^j\rangle)\big]\bigg]\\
&+\partial_t\mathcal{U}(t,\tilde X_t^j,[\tilde X_t^j])\bigg)\,\mathrm{d}t+\partial_x\mathcal{U}(t,\tilde X_t^j,[\tilde X_t^j])\cdot\big(\sigma(\tilde X_t^j,[\tilde X_t^j])\,\mathrm{d}W_t\big).
\end{aligned}
$$

Expressing the integral in (1) as expectations on $(\hat\Omega,\hat{\mathcal{F}},\hat{\mathbb{P}})$ and combining with (1) and (18), we obtain

$$
\begin{aligned}
&\mathrm{d}\big[\check Y_t^j-\tilde Y_t^j\big]\\
&\quad=\big(\{b(\tilde X_t^j,\tilde Y_t^j,[\tilde X_t^j,\tilde Y_t^j])-b(\tilde X_t^j,\check Y_t^j,[\tilde X_t^j,\check Y_t^j])\}\cdot\partial_x\mathcal{U}(t,\tilde X_t^j,[\tilde X_t^j])\\
&\qquad+\widehat{\mathbb{E}}\big[\{b(\langle\tilde X_t^j\rangle,\langle\tilde Y_t^j\rangle,[\tilde X_t^j,\tilde Y_t^j])-b(\langle\tilde X_t^j\rangle,\langle\check Y_t^j\rangle,[\tilde X_t^j,\check Y_t^j])\}\\
&\qquad\cdot\partial_\mu\mathcal{U}(t,\tilde X_t^j,[\tilde X_t^j])(\langle\tilde X_t^j\rangle)\big]\\
&\qquad+f(\tilde X_t^{j-1},\tilde Y_t^j,\tilde Z_t^j,[\tilde X_t^{j-1},\tilde Y_t^j])-f(\tilde X_t^j,\check Y_t^j,\check Z_t^j,[\tilde X_t^j,\check Y_t^j])\big)\,\mathrm{d}t\\
&\qquad+\big[\check Z_t^j-\tilde Z_t^j\big]\cdot\mathrm{d}W_t,
\end{aligned}
$$

where $\check Y_t^j:=\mathcal{U}(t,\tilde X_t^j,[\tilde X_t^j])$ and $\check Z_t^j:=\partial_x\mathcal{U}(t,\tilde X_t^j,[\tilde X_t^j])\sigma(\tilde X_t^j,[\tilde X_t^j])$. Observe that this argument is reminiscent of the four-step scheme; see [28].

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

$$
\begin{aligned}
\Delta^j&\le e^{C\delta}\Big(\big\|\mathcal{U}(r_{k+1},\tilde X^j_{r_{k+1}},[\tilde X^j_{r_{k+1}}])-\tilde Y^j_{r_{k+1}}\big\|_2\\
&\qquad+C\Big\|\sup_{t\in[r_k,r_{k+1}]}|\tilde X_t^j-\tilde X_t^{j-1}|\Big\|_2\Big)\\
&\le e^{C\delta}\Big(\|\varepsilon^{k+1}(\tilde X^{j-1}_{r_{k+1}})\|_2+\|\mathcal{U}(r_{k+1},\tilde X^j_{r_{k+1}},[\tilde X^j_{r_{k+1}}])\\
&\qquad-\mathcal{U}(r_{k+1},\tilde X^{j-1}_{r_{k+1}},[\tilde X^{j-1}_{r_{k+1}}])\|_2+C\Big\|\sup_{t\in[r_k,r_{k+1}]}|\tilde X_t^j-\tilde X_t^{j-1}|\Big\|_2\Big),
\end{aligned}
$$

recalling $\tilde Y^j_{r_{k+1}}=\texttt{picard[k+1]}(\tilde X^{j-1}_{r_{k+1}})$ and (17). Since $\mathcal{U}$ is Lipschitz, we have

$$
\Delta^j\le e^{C\delta}\Big(\|\varepsilon^{k+1}(\tilde X^{j-1}_{r_{k+1}})\|_2+C\Big\|\sup_{t\in[r_k,r_{k+1}]}|\tilde X_t^j-\tilde X_t^{j-1}|\Big\|_2\Big).\tag{26}
$$

We also have that

$$
\begin{aligned}
\tilde X_t^j-\tilde X_t^{j-1}&=\int_{r_k}^t\{b(\tilde X_s^j,\tilde Y_s^j,[\tilde X_t^j,\tilde Y_t^j])-b(\tilde X_s^{j-1},\tilde Y_s^{j-1},[\tilde X_t^{j-1},\tilde Y_t^{j-1}])\}\,\mathrm{d}s\\
&\qquad+\int_{r_k}^t\{\sigma(\tilde X_s^j,[\tilde X_s^j])-\sigma(\tilde X_s^{j-1},[\tilde X_s^{j-1}])\}\,\mathrm{d}W_s.
\end{aligned}
$$

Using usual arguments (squaring, taking the sup, using Bürkholder–Davis–Gundy inequality), we get, since b and σ are Lipschitz continuous,

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

Observing that

$$|\tilde{Y}_s^j - \tilde{Y}_s^{j-1}| \le |\tilde{Y}_s^j - \mathcal{U}(s, \tilde{X}_s^j, [\tilde{X}_s^j])| + |\tilde{Y}_s^{j-1} - \mathcal{U}(s, \tilde{X}_s^{j-1}, [\tilde{X}_s^{j-1}])| + \Lambda\big(|\tilde{X}_s^{j-1} - \tilde{X}_s^j| + \|\tilde{X}_s^{j-1} - \tilde{X}_s^j\|_2\big),$$

we obtain, for δ small enough,

$$\Big\| \sup_{t\in[r_k,r_{k+1}]} |\tilde{X}_t^j - \tilde{X}_t^{j-1}| \Big\|_2 \le C\delta\big(\Delta^j + \Delta^{j-1}\big). \tag{27}$$

Combining the previous inequality with (26), we obtain, for δ small enough,

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

which by induction leads to

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

and concludes the proof. □

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 nondecreasing function $\mathfrak{B} : \mathbb{R}_+ \to \mathbb{R}_+$ matching 0 in 0, only depending on Λ, such that, for $\bar{\delta} := C_\Lambda \delta < \min(c_\Lambda, 1)$ and $\beta \ge \mathfrak{B}(\bar{\delta})$ satisfying*

$$(J-1)\Lambda \bar{\delta}^J \frac{e^{\beta C_\Lambda T}}{e^{\beta\bar{\delta}} - 1} \le 1, \tag{28}$$

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

$$\begin{aligned} &\| \texttt{solver[}k\texttt{](}\xi\texttt{)} - \mathcal{U}(r_k, \xi, [\xi])\|_2 \\ &\quad \le \Lambda \frac{e^{\beta C_\Lambda T}}{\beta} \bar{\delta}^{J-1}\big(1 + \|P^\star_{r_k,T}(\xi)\|_2\big), \end{aligned} \tag{29}$$

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

$$dX_s^0 = b(X_s^0, 0, [X_s^0, 0])\,\mathrm{d}s + \sigma(X_s^0, [X_s^0])\,\mathrm{d}W_s,$$

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

Of course, it is absolutely straightforward to bound $\|P^\star_{r_k,T}(\xi)\|_2$ by $C(1+\|\xi\|_2)$ in (29). Theorem 5 may be restated accordingly, but the form used in the statement is more faithful to the spirit of the proof.

PROOF OF THEOREM 5. We prove the claim by an induction argument. We show below that, for all $k \in \{0,\dots,N\}$,

$$\begin{aligned}\|\varepsilon^k(\xi)\|_2 &= \|\texttt{solver[}k\texttt{](}\xi\texttt{)} - \mathcal{U}(r_k,\xi,[\xi])\|_2 \\ &\le \theta_k(1+\|P^\star_{r_k,T}(\xi)\|_2),\end{aligned} \tag{30}$$

where $(\theta_k)_{k=0,\dots,N-1}$ is defined by the following backward induction: $\theta_N := 0$ (recall (20)), and for $k \in \{0,\dots,N-1\}$,

$$\theta_k := \Lambda\bar{\delta}^J + e^{\beta\bar{\delta}}\theta_{k+1}, \tag{31}$$

where $\bar{\delta} := C_\Lambda\delta = C_\Lambda T/N$, for a constant C_Λ that is independent of N, and where β is such that

$$\left(\gamma + \gamma\bar{\delta}e^{\gamma\bar{\delta}}\left(\gamma + \frac{\Lambda}{1-\bar{\delta}}\right)\right) \le e^{\beta\bar{\delta}}, \qquad \text{with } \gamma := \frac{e^{\bar{\delta}}}{1-\bar{\delta}}. \tag{32}$$

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

$$\theta_k = \Lambda\bar{\delta}^J \sum_{j=0}^{N-k-1} e^{j\beta\bar{\delta}} \le \Lambda\bar{\delta}^J \frac{e^{\beta C_\Lambda T}}{e^{\beta\bar{\delta}}-1}, \tag{33}$$

which gives the expected result, recalling that $N\bar{\delta} = C_\Lambda T$.

We now prove (30). 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 (33) holds true for θ_{k+1}. Then, using (28), we have

$$\theta_{k+1} j \le 1, \qquad \text{for all } j \le J-1. \tag{34}$$

From Proposition 3, we have

$$\Delta_k^j \le \bar{\delta}^j\Delta_k^0 + \sum_{\ell=1}^{j} \bar{\delta}^{\ell-1}e^{\bar{\delta}}\|\varepsilon^{k+1}(\tilde{X}_{r_{k+1}}^{k,j-\ell})\|_2. \tag{35}$$

Using the induction hypothesis (30), we compute

$$\Delta_k^j \le \bar{\delta}^j\Delta_k^0 + \frac{e^{\bar{\delta}}}{1-\bar{\delta}}\theta_{k+1} + e^{\bar{\delta}}\theta_{k+1}\sum_{\ell=0}^{j-1}\bar{\delta}^{j-1-\ell}\|P^\star_{r_{k+1},T}(\tilde{X}_{r_{k+1}}^{k,\ell})\|_2. \tag{36}$$

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

$$\begin{aligned}\|P^\star_{r_{k+1},T}(\tilde{X}^{k,\ell}_{r_{k+1}})\|_2 \\ \le \|P^\star_{r_{k+1},T}(\tilde{X}^{k,0}_{r_{k+1}})\|_2 + \sum_{i=1}^{\ell}\|P^\star_{r_{k+1},T}(\tilde{X}^{k,i}_{r_{k+1}}) - P^\star_{r_{k+1},T}(\tilde{X}^{k,i-1}_{r_{k+1}})\|_2.\end{aligned}$$

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

$$\begin{aligned}\sum_{\ell=0}^{j-1}\bar{\delta}^{j-1-\ell}\|P^\star_{r_k,T}(\tilde{X}^{k,\ell}_{r_{k+1}})\|_2 \\ \le C_\Lambda\sum_{\ell=1}^{j-1}\bar{\delta}^{j-1-\ell}\sum_{i=1}^{\ell}\|\tilde{X}^{k,i}_{r_{k+1}} - \tilde{X}^{k,i-1}_{r_{k+1}}\|_2 + \sum_{\ell=0}^{j-1}\bar{\delta}^{\ell}\|P^\star_{r_k,T}(\xi)\|_2.\end{aligned}$$

Using (27) in the proof of Proposition 3 and changing the definition of $\bar{\delta}$, we obtain

$$\begin{aligned}\sum_{\ell=0}^{j-1}\bar{\delta}^{j-1-\ell}\|P^\star_{r_k,T}(\tilde{X}^{k,\ell}_{r_{k+1}})\|_2 \\ \le \bar{\delta}\sum_{i=1}^{j-1}(\Delta^i_k + \Delta^{i-1}_k)\sum_{\ell=i}^{j-1}\bar{\delta}^{j-1-\ell} + \sum_{\ell=0}^{j-1}\bar{\delta}^{\ell}\|P^\star_{r_k,T}(\xi)\|_2.\end{aligned} \tag{37}$$

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

$$\sum_{\ell=0}^{j-1}\bar{\delta}^{j-\ell}\|P^\star_{r_k,T}(\tilde{X}^{k,\ell}_{r_{k+1}})\|_2 \le \frac{2\bar{\delta}}{1-\bar{\delta}}\mathcal{S}^{j-1}_k + \frac{1}{1-\bar{\delta}}\|P^\star_{r_k,T}(\xi)\|_2, \tag{38}$$

where $\mathcal{S}^n_k := \sum_{i=0}^n \Delta^i_k$. Inserting the previous estimate into (36) and changing $\bar{\delta}$ into $2\bar{\delta}$, we obtain

$$\Delta^j_k \le \bar{\delta}^j\Delta^0_k + \frac{e^{\bar{\delta}}}{1-\bar{\delta}}\theta_{k+1}(1 + \|P^\star_{r_k,T}(\xi)\|_2) + \theta_{k+1}\frac{\bar{\delta}e^{\bar{\delta}}}{1-\bar{\delta}}\mathcal{S}^{j-1}_k. \tag{39}$$

We note that $\Delta^0_k \le \Lambda(1 + \|P^\star_{r_k,T}(\xi)\|_2)$. Recalling γ in (32), equation (39) leads to

$$\Delta^j_k \le a_j + \gamma\theta_{k+1}\bar{\delta}\mathcal{S}^{j-1}_k, \tag{40}$$

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

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

and then

$$\mathcal{S}_k^j \leq e^{\gamma\theta_{k+1}\bar{\delta} j}\mathcal{S}_k^0 + \sum_{\ell=1}^{j} e^{\gamma\theta_{k+1}\bar{\delta}(j-\ell)} a_\ell. \tag{41}$$

We compute

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

which combined with the properties (34) and (41) leads to, for all $j \leq J-1$,

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

where we recall that $\mathcal{S}_k^0 = \Delta_k^0 \leq \Lambda(1 + \|P^{\star}_{r_k,T}(\xi)\|_2)$. We insert the previous inequality into (40) for $j = J$ and get

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

Using (32), this rewrites

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

and validates (31) and thus (33). We then obviously have that (30) holds true. □

3. Convergence analysis.

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

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

(A1) $\sup_{t\in\pi^k}\|\mathcal{U}(t,\bar{X}_t,[\bar{X}_t]) - \bar{Y}_t\|_{2\alpha} \leq e^{\Lambda\delta}\|\mathcal{U}(r_{k+1},\bar{X}_{r_{k+1}},[\bar{X}_{r_{k+1}}]) - \bar{Y}_{r_{k+1}}\|_{2\alpha} + \Lambda\max_{j_k\leq i<j_{k+1}}\|\bar{X}_{t_i} - \bar{\mathfrak{X}}_{t_i}\|_{2\alpha} + \mathcal{D}^1(|\pi|) + \mathcal{D}^2(|\pi|)(1+\|\xi\|_{2\alpha}^{\alpha})$,
(A2) $\sup_{t\in\pi^k}\|\bar{X}_t - \bar{X}'_t\|_{2\alpha} \leq \Lambda\delta\sup_{t\in\pi^k}\|\bar{Y}_t - \bar{Y}'_t\|_{2\alpha}$,
(A3) $\|\mathcal{U}(r_{k+1},\bar{X}_{r_{k+1}},[\bar{X}_{r_{k+1}}]) - \bar{Y}_{r_{k+1}}\|_{2\alpha}^{\alpha} \leq \Lambda\|\mathcal{U}(r_{k+1},\bar{X}_{r_{k+1}},[\bar{X}_{r_{k+1}}]) - \bar{Y}_{r_{k+1}}\|_{2\alpha}$,

where $(\bar{X}, \bar{Y}, \bar{Z}) := \overline{\texttt{solver}}\texttt{[}k\texttt{]}\texttt{(}\bar{\mathfrak{X}}\texttt{,}\eta\texttt{,}f\texttt{)}$, for f as before, and $(\bar{X}', \bar{Y}', \bar{Z}') := \overline{\texttt{solver}}\texttt{[}k\texttt{]}\texttt{(}\bar{\mathfrak{X}}'\texttt{,}\eta'\texttt{,}f'\texttt{)}$, for another f' either equal to f or 0, are two output values of $\overline{\texttt{solver}}\texttt{[](,,)}$ 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 η and η'. For $i \in \{1, 2\}$, the function $\mathcal{D}^i : [0, \infty) \to [0, \infty)$ is a discretization error associated to the use of the grid π, 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 $\overline{\texttt{solver}}\texttt{[}k\texttt{](,0,0)}$ used to initialize the algorithm at each step. Following the definition of $(P_{r_k,t})_{0 \le t \le T}$ introduced in the statement of Theorem 5, we let by induction, for a given $k \in \{0, \dots, N-1\}$:

$$\texttt{P}_{r_k,t}(\xi) = \big(\overline{\texttt{solver}}\texttt{[}k\texttt{]}\texttt{(}\xi\texttt{,0,0)}\big)^1_t, \qquad t \in \pi^k, \xi \in L^2(\mathcal{F}_{r_k}),$$

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

$$\texttt{P}_{r_k,t}(\xi) = \texttt{P}_{r_\ell,t}\big(\texttt{P}_{r_k,r_\ell}(\xi)\big), \qquad t \in \pi^\ell, k < \ell \le N-1,$$

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

(A4) $\|\texttt{P}^\star_{r_k,T}(\xi) - \texttt{P}^\star_{r_k,T}(\xi')\|_{2\alpha} \le \Lambda \|\xi - \xi'\|_{2\alpha}$,
(A5) $\|\texttt{P}^\star_{r_k,T}(\xi)\|_{2\alpha} \le \Lambda(1 + \|\xi\|_{2\alpha})$,

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

REMARK 6. Assumptions (A1)–(A5) are mostly related with the choice of the numerical solver $\overline{\texttt{solver}}\texttt{[](,,)}$. We make it clear in Section 3.2: There, we essentially prove that, provided that $\mathcal{U}$ satisfies (H0) and (H1), Example 2 satisfies (A1)–(A5). In fact, the main challenging assumption (and maybe the most surprising one) is (A3), as it explicitly depends on the value of α in assumption (13) for $\mathcal{U}$. Indeed, (A3) is obviously satisfied when $\alpha = 1$ as long as Λ is assumed to be greater than 1. Following Remark 1, we refer to [15] and [18], Chapter 12, for sets of conditions under which the condition $\alpha = 1$ is indeed true. When $\alpha > 1$, Assumption (A3) is checked provided we have an *a priori bound* on $\|\mathcal{U}(r_{k+1}, \bar{X}_{r_{k+1}}, [\bar{X}_{r_{k+1}}]) - \bar{Y}_{r_{k+1}}\|_{2\alpha}$. We provide in Lemma 10 below conditions on the coefficients f and g under which the latter is verified. If true, this permits to work with larger values of α in (13) and so to relax the conditions that are needed on $\mathcal{U}$: For instance, we can invoke the result proven in our previous paper [21], which holds true in a weaker setting than the solvability results obtained in [15] and [18], Chapter 12.

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

$$(J-1)\big(\Lambda \bar{\delta}^J + e^{\beta \bar{\delta}} \mathcal{D}^2(|\pi|)\big) \frac{e^{\beta C_\Lambda T}}{e^{\beta \bar{\delta}} - 1} \leq 1, \tag{42}$$

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

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

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 $k = \{0, \dots, N\}$,

$$\|\varepsilon^k(\xi)\|_{2\alpha} \leq \theta_k\big(1 + \|\mathrm{P}^\star_{r_k,T}(\xi)\|_{2\alpha}^\alpha\big) + \vartheta_k \mathcal{D}^1(|\pi|), \tag{43}$$

where

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

and $(\theta_k, \vartheta_k)_{k=0,\dots,N}$ is defined by the following backward induction: $(\theta_N, \vartheta_N) := (0, 0)$, recall (23), and for $k \in \{0, \dots, N-1\}$,

$$\theta_k := \Lambda \bar{\delta}^J + e^{\beta \bar{\delta}}\{\theta_{k+1} + \mathcal{D}^2(|\pi|)\} \quad \text{and} \quad \vartheta_k := e^{\beta \bar{\delta}}(\vartheta_{k+1} + 1), \tag{44}$$

β being defined as in (51).

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

$$\theta_k \leq \big(\Lambda \bar{\delta}^J + e^{\beta \bar{\delta}} \mathcal{D}^2(|\pi|)\big) \frac{e^{\beta \bar{\delta}(N-k)} - 1}{e^{\beta \bar{\delta}} - 1} \quad \text{and} \quad \vartheta_k \leq e^{\beta \bar{\delta}} \frac{e^{\beta (N-k)\bar{\delta}} - 1}{e^{\beta \bar{\delta}} - 1}. \tag{45}$$

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

We now prove (43). Obviously, it holds true for the last step N. Assume now that it is true at step $k+1$, for $k < N$ and that (45) holds for θ_{k+1} and ϑ_{k+1}.

In particular, using (42), we observe that

$$\theta_{k+1} j \leq 1, \qquad \text{for all } j \leq J-1. \tag{46}$$

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

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

we observe that

$$\begin{aligned}
\bar{\Delta}^j_k &\le e^{\Lambda\delta}\big[\big\|\mathcal{U}\big(r_{k+1}, \bar{X}^{k,j}_{r_{k+1}}, [\bar{X}^{k,j}_{r_{k+1}}]\big) - \mathcal{U}\big(r_{k+1}, \bar{X}^{k,j-1}_{r_{k+1}}, [\bar{X}^{k,j-1}_{r_{k+1}}]\big)\big\|_{2\alpha} \\
&\quad + \big\|\varepsilon^{k+1}\big(\bar{X}^{k,j-1}_{r_{k+1}}\big)\big\|_{2\alpha}\big] \\
&\quad + \Lambda \max_{j_k \le i < j_{k+1}} \big\|\bar{X}^{k,j}_{t_i} - \bar{X}^{k,j-1}_{t_i}\big\|_{2\alpha} + \mathcal{D}^1(|\pi|) + \mathcal{D}^2(|\pi|)\big(1 + \|\xi\|^{\alpha}_{2\alpha}\big) \qquad (47)\\
&\le C_\Lambda \max_{t\in\pi^k} \big\|\bar{X}^{k,j}_t - \bar{X}^{k,j-1}_t\big\|_{2\alpha} + e^{\Lambda\delta}\big\|\varepsilon^{k+1}\big(\bar{X}^{k,j-1}_{r_{k+1}}\big)\big\|_{2\alpha} \\
&\quad + \mathcal{D}^1(|\pi|) + \mathcal{D}^2(|\pi|)\big(1 + \|\mathrm{P}^\star_{r_k,T}(\xi)\|^{\alpha}_{2\alpha}\big).
\end{aligned}$$

Using (A2), we also have

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

for δ small enough. Inserting the previous inequality in (47), we get

$$\begin{aligned}
\bar{\Delta}^j_k &\le C_\Lambda \delta \bar{\Delta}^{j-1}_k + e^{C_\Lambda\delta}\big\|\varepsilon^{k+1}\big(\bar{X}^{k,j-1}_{r_{k+1}}\big)\big\|_{2\alpha} + \mathcal{D}^1(|\pi|) \\
&\quad + \mathcal{D}^2(|\pi|)\big(1 + \|\mathrm{P}^\star_{r_k,T}(\xi)\|^{\alpha}_{2\alpha}\big) \\
&\le \bar{\delta}^j \bar{\Delta}^0_k + e^{\bar{\delta}} \sum_{\ell=0}^{j-1} \bar{\delta}^\ell \big\|\varepsilon^{k+1}\big(\bar{X}^{k,j-1-\ell}_{r_{k+1}}\big)\big\|_{2\alpha} \\
&\quad + \frac{\mathcal{D}^1(|\pi|)}{1-\bar{\delta}} + \frac{\mathcal{D}^2(|\pi|)}{1-\bar{\delta}}\big(1 + \|\mathrm{P}^\star_{r_k,T}(\xi)\|^{\alpha}_{2\alpha}\big),
\end{aligned}$$

with $\bar{\delta} := C_\Lambda\delta$. We note that compared to (25), there is a new term, namely $(\mathcal{D}^1(|\pi|) + \mathcal{D}^2(|\pi|)(1 + \|\mathrm{P}^\star_{r_k,T}(\xi)\|^{\alpha}_{2\alpha})/(1-\bar{\delta})$, which is due to the discretization.

Second Step. Using (43) at the previous step $k+1$ and noting that $\bar{\Delta}^0_k \le \Lambda(1 + \|\mathrm{P}^\star_{r_k,T}(\xi)\|_{2\alpha}) \le 2\Lambda(1 + \|\mathrm{P}^\star_{r_k,T}(\xi)\|^{\alpha}_{2\alpha})$, we claim that

$$\begin{aligned}
\bar{\Delta}^j_k &\le \big(2\Lambda\bar{\delta}^j + \gamma\mathcal{D}^2(|\pi|)\big)\big(1 + \|\mathrm{P}^\star_{r_k,T}(\xi)\|^{\alpha}_{2\alpha}\big) + \gamma(\vartheta_{k+1} + 1)\mathcal{D}^1(|\pi|) \\
&\quad + e^{\bar{\delta}}\theta_{k+1} \sum_{\ell=0}^{j-1} \bar{\delta}^{j-1-\ell}\big(1 + \big\|\mathrm{P}^\star_{r_{k+1},T}\big(\bar{X}^{k,\ell}_{r_{k+1}}\big)\big\|^{\alpha}_{2\alpha}\big), \qquad (48)
\end{aligned}$$

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

This corresponds to equation (36) adapted to our context. By (A2) we have, for $\ell \leq J-1$,

$$\|\mathtt{P}^{\star}_{r_{k+1},T}(\bar{X}^{k,\ell}_{r_{k+1}}) - \mathtt{P}^{\star}_{r_{k+1},T}(\bar{X}^{k,0}_{r_{k+1}})\|_{2\alpha} \leq C_\Lambda \sup_{t\in\pi^k} \|\bar{X}^{k,\ell}_t - \bar{X}^{k,0}_t\|_{2\alpha}. \tag{49}$$

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

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

where for the last inequality we used the fact that δ is small enough. Observing that $\|\xi\|_{2\alpha} \leq \|\mathtt{P}^{\star}_{r_k,T}(\xi)\|_{2\alpha}$ and combining the previous inequality with (49), we obtain

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

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

$$\begin{aligned}
&\|\mathtt{P}^{\star}_{r_{k+1},T}(\bar{X}^{k,\ell}_{r_{k+1}})\|^\alpha_{2\alpha} \\
&\quad\leq \big(C_\Lambda\delta\bar{\Delta}^\ell_k + (1+C_\Lambda\delta)(1+\|\mathtt{P}^{\star}_{r_k,T}(\xi)\|_{2\alpha})\big)^\alpha \\
&\quad\leq (1+2C_\Lambda\delta)^{\alpha-1}\big(C_\Lambda\delta(\bar{\Delta}^\ell_k)^\alpha + (1+C_\Lambda\delta)\|\mathtt{P}^{\star}_{r_k,T}(\xi)\|^\alpha_{2\alpha}\big).
\end{aligned}$$

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

$$\|\mathtt{P}^{\star}_{r_{k+1},T}(\bar{X}^{k,\ell}_{r_{k+1}})\|^\alpha_{2\alpha} \leq \bar{\delta}\bar{\Delta}^\ell_k + e^{\bar{\delta}}(1+\|\mathtt{P}^{\star}_{r_k,T}(\xi)\|^\alpha_{2\alpha}),$$

which may be rewritten as

$$\begin{aligned}
&\sum_{\ell=0}^{j-1}\bar{\delta}^{j-1-\ell}\|\mathtt{P}^{\star}_{r_{k+1},T}(\bar{X}^{k,\ell}_{r_{k+1}})\|^\alpha_{2\alpha} \\
&\qquad\leq \bar{\delta}\sum_{\ell=1}^{j-1}\bar{\delta}^{j-1-\ell}\bar{\Delta}^\ell_k + \frac{e^{\bar{\delta}}}{1-\bar{\delta}}(1+\|\mathtt{P}^{\star}_{r_k,T}(\xi)\|^\alpha_{2\alpha}).
\end{aligned}$$

Recalling the notation $\gamma = e^{\bar{\delta}}/(1-\bar{\delta})$ and letting $\bar{\mathcal{S}}^n_k := \sum_{i=0}^n \bar{\delta}^{n-i}\bar{\Delta}^i_k$, we obtain a new version of (40), namely

$$\bar{\Delta}^j_k \leq \Lambda\bar{\delta}^j\left(\frac{1}{2} + \|\mathtt{P}^{\star}_{r_k,T}(\xi)\|^\alpha_{2\alpha}\right) + \bar{a} + \theta_{k+1}\gamma\bar{\delta}\bar{\mathcal{S}}^{j-1}_k, \tag{50}$$

where we changed the constant 2Λ in (48) into $\frac{1}{2}\Lambda$ as we changed the value of $\bar{\delta}$, and where we put

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

We straightforwardly deduce that

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

which yields

$$\bar{\mathcal{S}}^j_k \le \Lambda(j+2)\bar{\delta}^j e^{\gamma\theta_{k+1}(j-1)}\big(1 + \|\mathrm{P}^{\star}_{r_k,T}(\xi)\|^{\alpha}_{2\alpha}\big) + \frac{\bar{a}}{1-e^{\gamma\theta_{k+1}}\bar{\delta}},$$

where we used $\bar{\mathcal{S}}^0_k \le 2\Lambda(1 + \|\mathrm{P}^{\star}_{r_k,T}(\xi)\|^{\alpha}_{2\alpha})$. Thanks to (50), we get

$$\bar{\Delta}^J_k \le \Lambda\bar{\delta}^J\left(\frac{1}{2} + \bar{\delta}\gamma(J+2)\theta_{k+1}e^{\gamma\theta_{k+1}(J-1)}\right)\big(1 + \|\mathrm{P}^{\star}_{r_k,T}(\xi)\|^{\alpha}_{2\alpha}\big) + \frac{\bar{a}}{1-e^{\gamma\theta_{k+1}}\bar{\delta}}.$$

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

$$\begin{aligned}\bar{\Delta}^J_k \le{}& \big(\Lambda\bar{\delta}^J + e^{\beta\bar{\delta}}\{\theta_{k+1} + \mathcal{D}^2(|\pi|)\}\big)\big(1 + \|\mathrm{P}^{\star}_{r_k,T}(\xi)\|^{\alpha}_{2\alpha}\big) \\ &+ e^{\beta\bar{\delta}}(\vartheta_{k+1}+1)\mathcal{D}^1(|\pi|),\end{aligned}$$

provided that β satisfies

$$\frac{\gamma^2}{1-e^{\gamma\theta_{k+1}}\bar{\delta}} \le e^{\beta\bar{\delta}}. \tag{51}$$

This validates (44) and concludes the proof. □

3.2. *Convergence error for the implemented scheme.* We now analyze 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 $k \le N-1$, we consider $(\bar{X}, \bar{Y}, \bar{Z}) := \overline{\texttt{solver}}[k](\bar{\mathfrak{X}}, \eta, f)$ and $(\bar{X}', \bar{Y}', \bar{Z}') := \overline{\texttt{solver}}[k](\bar{\mathfrak{X}}', \eta', f')$ with $\bar{\mathfrak{X}}_{r_k} = \bar{\mathfrak{X}}'_{r_k} = \xi$. 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

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

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 the Cauchy–Schwarz inequality for the first term and the Bürkholder–Davis–Gundy inequality for discrete martingales for the second term and appealing to the Lipschitz property of b and σ, we get

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

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

$$\max_{j_k \le i \le j_{k+1}} \|\Delta X_i\|_{2\alpha} \le C\delta \max_{j_k \le i \le j_{k+1}} \|\Delta Y_i\|_{2\alpha}, \tag{52}$$

which concludes the proof. □

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*

$$\mathcal{D}^1(|\pi|) \le C\sqrt{|\pi|} \quad \textit{and} \quad \mathcal{D}^2(|\pi|) \le C\sqrt{|\pi|}.$$

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

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

where $(b_i, \sigma_i) := (b(\bar{X}_{t_i}, \bar{Y}_{t_i}, [\bar{X}_{t_i}, \bar{Y}_{t_i}]), \sigma(\bar{X}_{t_i}, [\bar{X}_{t_i}]))$ and ϖ is defined in equation (24). In preparation for the proof, we also introduce a piecewise càdlàg version, denoted by $(\bar{X}_s^{(\lambda)})_{0\le s\le T}$, where λ 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

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

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

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

with

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

and

$$\begin{aligned}\mathcal{R}_i^{\sigma x} &= \frac{1}{2} \int_{t_i}^{t_{i+1}} \int_0^1 \Delta^x(s, \lambda)\, \mathrm{d}\lambda\, \mathrm{d}s,\\ \mathcal{R}_i^{\sigma\mu} &= \frac{1}{2} \int_{t_i}^{t_{i+1}} \int_0^1 \Delta^\mu(s, \lambda)\, \mathrm{d}\lambda\, \mathrm{d}s,\end{aligned} \tag{53}$$

where

$$\Delta^x(s,\lambda) := \mathrm{Tr}\{\partial^2_{xx}\mathcal{U}(s,\bar{X}_s^{(\lambda)},[\bar{X}_s])a(\bar{X}_{t_i},[\bar{X}_{t_i}]) \\ - \partial^2_{xx}\mathcal{U}(s,\bar{X}_s,[\bar{X}_s])a(\bar{X}_s,[\bar{X}_s])\}$$

$$\Delta^\mu(s,\lambda) := \hat{\mathbb{E}}\big[\,\mathrm{Tr}\{\partial_v\partial_\mu\mathcal{U}(s,\bar{X}_s,[\bar{X}_s])(\langle\bar{X}_s^{(\lambda)}\rangle)\langle a(\bar{X}_{t_i},\bar{X}_{t_i}])\rangle \\ - \partial_v\partial_\mu\mathcal{U}(s,\bar{X}_s,[\bar{X}_s])(\langle\bar{X}_s\rangle)\langle a(\bar{X}_s,[\bar{X}_s])\rangle\}\big].$$

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

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

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

the previous equation reads

$$\bar{U}_{t_{i+1}} = \bar{U}_{t_i} + \zeta_i + h_i[\delta b_i - f(\bar{\mathfrak{X}}_{t_i},\bar{U}_{t_i},\sigma^\dagger(\bar{X}_{t_i},[\bar{X}_{t_i}])\bar{V}_{t_i}^x,[\bar{\mathfrak{X}}_{t_i},\bar{U}_{t_i}]) \\ + \bar{V}_{t_i}^x\cdot(\sigma(\bar{X}_{t_i},[\bar{X}_{t_i}])\underline{\varpi}_i)], \tag{54}$$

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

$$\bar{Y}_{t_i} = \bar{Y}_{t_{i+1}} + h_i f(\bar{\mathfrak{X}}_{t_i},\bar{Y}_{t_i},\bar{Z}_{t_i},[\bar{\mathfrak{X}}_{t_i},\bar{Y}_{t_i}]) - h_i\bar{Z}_{t_i}\cdot\underline{\varpi}_i - \Delta M_i, \tag{55}$$

where ΔM_i satisfies

$$\mathbb{E}_{t_i}[\Delta M_i] = 0, \qquad \mathbb{E}_{t_i}[\underline{\varpi}_i\cdot\Delta M_i] = 0 \quad \text{and} \quad \mathbb{E}\big[|\Delta M_i|^2\big] < \infty. \tag{56}$$

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 (54) and (55), we get

$$\Delta\bar{Y}_i = \Delta\bar{Y}_{i+1} + h_i(\delta b_i + \delta f_i) + \zeta_i - h_i\Delta\bar{Z}_i\cdot\underline{\varpi}_i - \Delta M_i, \tag{57}$$

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

$$|\delta b_i| + |\delta f_i| \le C_\Lambda\big(|\Delta\bar{Y}_i| + \|\Delta\bar{Y}_i\|_2 + |\Delta\bar{Z}_i|\big). \tag{58}$$

Summing the equation (57) 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 (56) for the left-hand side and Young's and conditional Cauchy–Schwarz inequality for the right-hand side,

$$\mathbb{E}_{t_q}\big[|\Delta\bar{Y}_i|^2\big] + \sum_{\ell=i}^{j_{k+1}-1} h_\ell \mathbb{E}_{t_q}\big[|\Delta\bar{Z}_\ell|^2\big]$$
$$\le \mathbb{E}_{t_q}\Bigg[(1+C\delta)|\Delta\bar{Y}_{j_{k+1}}|^2$$
$$+ C\sum_{\ell=i}^{j_{k+1}-1} h_\ell |\delta b_\ell + \delta f_\ell|^2 + \frac{C}{\delta}\Bigg(\sum_{\ell=i}^{j_{k+1}-1} \zeta_\ell\Bigg)^2\Bigg],$$

for $i \ge q \ge j_k$. Combining (58) and Young's inequality, this leads to

$$\mathbb{E}_{t_q}\big[|\Delta\bar{Y}_i|^2\big] + \frac{1}{2}\sum_{\ell=i}^{j_{k+1}-1} h_\ell \mathbb{E}_{t_q}\big[|\bar{Z}_\ell|^2\big]$$
$$\le \mathbb{E}_{t_q}\Bigg[e^{C\delta}|\Delta\bar{Y}_{j_{k+1}}|^2 + C\sum_{\ell=i}^{j_{k+1}-1} h_\ell |\Delta\bar{Y}_\ell|^2 + \frac{C}{\delta}\Bigg(\sum_{\ell=i}^{j_{k+1}-1} \zeta_\ell\Bigg)^2\Bigg].$$

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 $i=q$,

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

and then,

$$\Delta_Y^2 := \max_{j_k\le i\le j_{k+1}} \|\Delta\bar{Y}_i\|_{2\alpha}^2$$
$$\le e^{C\delta}\|\Delta\bar{Y}_{j_{k+1}}\|_{2\alpha}^2 + \frac{C}{\delta}\Bigg\|\max_{j_k\le i\le j_{k+1}-1}\Bigg(\sum_{\ell=i}^{j_{k+1}-1} \zeta_\ell\Bigg)\Bigg\|_{2\alpha}^2. \tag{59}$$

Third Step. To conclude, we need an upper bound for the error $\|\max_{j_k\le i\le j_{k+1}-1}(\sum_{\ell=i}^{j_{k+1}-1}\zeta_\ell)\|_{2\alpha}^2$ where ζ_ℓ is defined in (55). To do so, we study each term in (55) separately. We also define $\Delta_X := \max_{t \in \pi^k} \|\bar{X}_t - \bar{\mathfrak{X}}_t\|_{2\alpha}$ 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

$$\left\| \max_{j_k \le i \le j_{k+1}} \left(\sum_{\ell=i}^{j_{k+1}-1} \mathcal{R}_\ell^f \right) \right\|_{2\alpha}^2 \le C \frac{\delta}{|\pi|} \sum_{\ell=j_k}^{j_{k+1}-1} \|\mathcal{R}_\ell^f\|_{2\alpha}^2. \tag{60}$$

We will upper bound this last term.

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

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

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

$$\|\bar{V}_s^x - \bar{V}_{t_i}^x\|_{2\alpha}^2 \le C(\|\bar{X}_s - \bar{X}_{t_i}\|_{2\alpha}^2 + h_i(1 + \|\bar{X}_{t_i}\|_{2\alpha}^2)). \tag{61}$$

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

$$\|\bar{X}_s - \bar{X}_{t_i}\|_{2\alpha}^2 \le C_\Lambda(h_i + h_i^2 \|\bar{U}_{t_i} - Y_{t_i}\|_{2\alpha}^2 + h_i^2 \|\bar{X}_{t_i}\|_{2\alpha}^2). \tag{62}$$

Using Lemma 15 from the Appendix, we obtain

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

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

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

where we used the same argument as above to handle the difference between the two σ 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

$$\|\mathcal{R}_i^f\|_{2\alpha}^2 \le C h_i^2(\Delta_X^2 + h_i(1 + \|\xi\|_{2\alpha}^2) + h_i^2 \Delta_Y^2). \tag{63}$$

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

$$\|\mathcal{R}_i^{bx}\|_{2\alpha}^2 + \|\mathcal{R}_i^{b\mu}\|_{2\alpha}^2 \le C h_i^2 \|\bar{U}_s - \bar{U}_{t_i}\|_{2\alpha}^2 + \|\bar{X}_s - \bar{X}_{t_i}\|_{2\alpha}^2. \tag{64}$$

Arguing as in the previous step, we easily get

$$\|\mathcal{R}_i^b\|_{2\alpha}^2 \le C h_i^2(h_i(1 + \|\xi\|_{2\alpha}^2) + h_i^2 \Delta_Y^2). \tag{65}$$

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 $\underline{\varpi}_i$ is uniformly bounded,

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

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

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

Similarly,

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

Hence,

$$\begin{aligned}
&\Big\| \max_{j_k \le i \le j_{k-1}} \Big(\sum_{\ell=i}^{j_{k+1}-1} \mathcal{R}_\ell^w \Big) \Big\|_{2\alpha}^2 + \Big\| \max_{j_k \le i \le j_{k-1}} \Big(\sum_{\ell=i}^{j_{k+1}-1} \delta\mathcal{M}(t_\ell, t_{\ell+1}) \Big) \Big\|_{2\alpha}^2 \\
&\quad \le C\delta |\pi| \big(1 + \|\xi\|_{2\alpha}^2\big) + C\delta |\pi|^2 \Delta_Y^2.
\end{aligned} \tag{66}$$

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

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

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

$$\|\Delta^x(s, \lambda)\|_{2\alpha}^2 \le C\big(\|\bar{X}_s^{(\lambda)} - \bar{X}_s\|_{2\alpha}^2 + \|\bar{X}_s - \bar{X}_{t_i}\|_{2\alpha}^2\big). \tag{67}$$

Observing that $\|\bar{X}_s^{(\lambda)} - \bar{X}_s\|_{2\alpha} \le C\sqrt{h_i}$, we obtain using (62), for $t_i \le s \le t_{i+1}$

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

which leads, using Lemma 15 again, to

$$\|\mathcal{R}_i^{\sigma x}\|_{2\alpha}^2 \le C h_i^2\big(h_i + h_i^2(\Delta_Y^2 + \|\xi\|_{2\alpha}^2)\big). \tag{68}$$

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

$$\begin{aligned} |\Delta^\mu(s,\lambda)| \le{} & C\hat{\mathbb{E}}\big[\big|\partial_\upsilon\partial_\mu\mathcal{U}(s,\bar{X}_s,[\bar{X}_s])(\langle\bar{X}_s^{(\lambda)}\rangle) \\ & - \partial_\upsilon\partial_\mu\mathcal{U}(s,\bar{X}_s,[\bar{X}_s])(\langle\bar{X}_s\rangle)\big|\big] \\ & + \hat{\mathbb{E}}\big[\big|\partial_\upsilon\partial_\mu\mathcal{U}(s,\bar{X}_s,[\bar{X}_s])(\langle\bar{X}_s\rangle)\big| \\ & \cdot \big|\langle a(\bar{X}_{t_i},[\bar{X}_{t_i}]) - a(\bar{X}_s,[\bar{X}_s])\rangle\big|\big]. \end{aligned} \tag{69}$$

For the last term, we combine the Cauchy–Schwarz inequality (10) and boundedness and Lipschitz continuity of σ to get

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

Recalling from (62) that $\|\bar{X}_s - \bar{X}_{t_i}\|_{2\alpha}^2 \le C_\Lambda(h_i + h_i^2(\Delta_Y^2 + \|\bar{X}_{t_i}\|_{2\alpha}^2))$, we obtain, using Lemma 15, that

$$\begin{aligned} & \hat{\mathbb{E}}\big[\big|\partial_\upsilon\partial_\mu\mathcal{U}(s,\bar{X}_s,[\bar{X}_s])(\langle\bar{X}_s\rangle)\big| \cdot \big|\langle a(\bar{X}_{t_i},[\bar{X}_{t_i}]) - a(\bar{X}_s,[\bar{X}_s])\rangle\big|\big] \\ & \quad \le C_\Lambda h_i^{\frac{1}{2}}\big(1 + h_i^{\frac{1}{2}}\{\Delta_Y + \|\xi\|_{2\alpha}\}\big). \end{aligned} \tag{70}$$

For the first term in (69), we use (H1) equation (13) to get

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

By the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned} & \hat{\mathbb{E}}\big[\big|\{\partial_\upsilon\partial_\mu\mathcal{U}(s,\bar{X}_s,[\bar{X}_s])(\langle\bar{X}_s^{(\lambda)}\rangle) - \partial_\upsilon\partial_\mu\mathcal{U}(s,\bar{X}_s,[\bar{X}_s])(\langle\bar{X}_s\rangle)\}\big|\big] \\ & \quad \le C\sqrt{h_i}\big(1 + \|\bar{X}_s^{(\lambda)}\|_{2\alpha}^\alpha + \|\bar{X}_s\|_{2\alpha}^\alpha\big). \end{aligned} \tag{71}$$

We then observe that

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

where we used Lemma 15 for the last inequality. Combining the last inequality with (71) and using also (70), we compute

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

and then

$$\left\| \sum_{\ell=j_k}^{j_{k+1}-1} |\mathcal{R}_\ell^{\sigma\mu}| \right\|_{2\alpha}^2 \le C|\pi|\delta^2(1+\delta^2\Delta_Y^2+\|\xi\|_{2\alpha}^2). \tag{72}$$

Fourth step. Collecting the estimates (63), (65) and (68), we compute

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

Observing that

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

and combining the previous inequality with (72), (66) and (59), we obtain

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

which concludes the proof for δ 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 α.*

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 [21] for the proof of the boundedness of $\mathcal{U}$.

The bound for $\bar{Y}$ may obtained by squaring (55) 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 $(\mathbb{P}_{r_k,t}(\xi))_{t\in\pi,t\ge r_k}$ 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* (42), *the following holds*:

$$\begin{aligned}&\|\texttt{solver[}k\texttt{](}\xi\texttt{)} - \mathcal{U}(r_k,\xi,[\xi])\|_{2\alpha}\\ &\qquad \le C((C\delta)^{J-1}+|\pi|^{\frac{1}{2}}\delta^{-1}(1+\|\xi\|_{2\alpha})),\end{aligned}$$

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 δ. 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, [22, 23]. The normalization by δ 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, that is, $N = 2$. 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 $x \in \operatorname{supp}(\mu)$. 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 μ 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 initialized by a discrete random variable as well. In practice, this means that each point x^ℓ 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 that `solver[0]`$(\xi^{|\pi|})$ is a random variable with distribution $\sum_{\ell=1}^{M} p_\ell \delta_{\tilde{y}^\ell}$, where $\tilde{y}^\ell$ stands for the approximation of y^ℓ. 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

$$
\begin{aligned}
\big|\mathcal{U}(0,x,\mu) - \mathcal{U}\big(0,x^{\bar{\ell}},\mu^{|\pi|}\big)\big| &\le C\Big(\min_{y\in\operatorname{supp}([\xi^{|\pi|}])} |y-x| + \mathcal{W}_2\big(\mu^{|\pi|},\mu\big)\Big) \\
&=: \mathcal{E}_1\big(|\pi|,\xi\big),
\end{aligned}
\tag{73}
$$

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

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

As mentioned above, the approximation $\tilde{y}^{\bar{\ell}}$ 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 $x^{\bar{\ell}}$. The corresponding pointwise error is given by

$$\mathcal{E}_2(|\pi|, \delta, \xi) := |y^{\bar{\ell}} - \mathcal{U}(0, x^{\bar{\ell}}, [\xi^{|\pi|}])|. \tag{74}$$

Of a course, this might be estimated by

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

but this is very poor when the initial distribution μ 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 ξ 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

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

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 scheme for the whole process $(\mathbb{X}, \mathbb{Y}, \mathbb{Z})$.

This leads us to the following result.

THEOREM 13. *Whenever $\xi^{|\pi|}$ has distribution $\sum_{\ell=1}^M p_\ell \delta_{x^\ell}$, $\texttt{solver[0]}(\xi^{|\pi|})$ has distribution $\sum_{\ell=1}^M p_\ell \delta_{\tilde{y}^\ell}$, where $\tilde{y}^\ell$ is the realization of the random variable $\texttt{solver[0]}(\xi^{|\pi|})$ on the event $\xi^{|\pi|} = x^\ell$. Then, if, for a* *given* $x \in \mathbb{R}^d$, *we call* $\bar{\ell}$ *an index such that* $|x^{\bar{\ell}} - x|$ *achieves the minimum in* $\min_{y \in \mathrm{supp}([\xi^{|\pi|}])} |y - x|$, *then the distance between* $\tilde{y}^{\bar{\ell}}$ *and* $\mathcal{U}(0, x, \mu)$ *is less than*

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

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, that is, $N = 2$ (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, that is, $N = 1$ (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:

$$\begin{aligned} \mathrm{d}X_t &= -\rho \mathbb{E}[Y]_t \,\mathrm{d}t + \sigma \,\mathrm{d}W_t, \qquad X_0 = x, \\ \mathrm{d}Y_t &= -aY_t \,\mathrm{d}t + Z_t \,\mathrm{d}W_t \quad \text{and} \quad Y_T = X_T, \end{aligned}$$

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.

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 period 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 ρ; see equations (75) (for a case without McKean–Vlasov interaction) and (76) (for

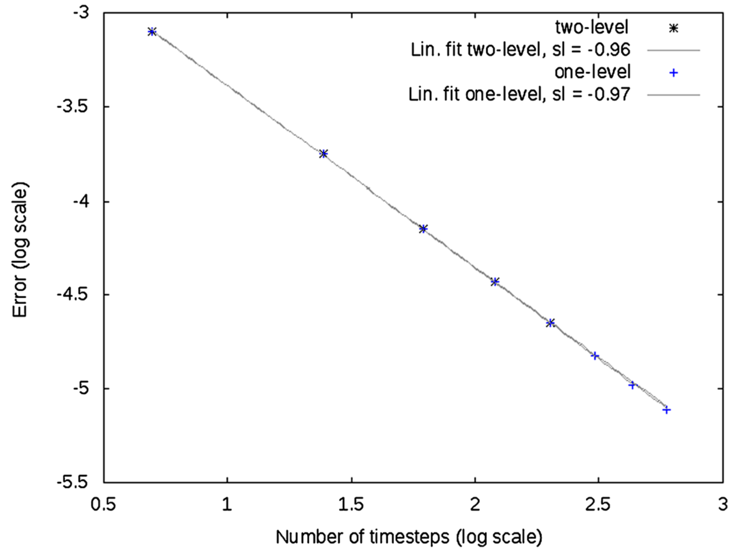

FIG. 1. *Convergence of the algorithms: log-log error plot for the same data as in the text. We can observe that both algorithms return the same value which is close to the true value. This validates the convergence of both methods in this simple linear setting.*

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:

$$\begin{aligned} &\mathrm{d}X_t = \rho\cos(Y_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \qquad X_0 = x, \\ &Y_t = \mathbb{E}_t\big[\sin(X_T)\big]. \end{aligned} \tag{75}$$

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 $Y0$ stands for the approximation of $\mathcal{U}(0,x)$: There is no dependence upon the initial measure as there is no McKean–Vlasov 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.

An example from large population stochastic control. For this part, the model is given by

$$\begin{aligned} &\mathrm{d}X_t = -\rho Y_t\,\mathrm{d}t + \mathrm{d}W_t, \qquad X_0 = x, \\ &\mathrm{d}Y_t = \operatorname{atan}\big(\mathbb{E}[X_t]\big)\,\mathrm{d}t + Z_t\,\mathrm{d}W_t \quad \text{and} \quad Y_T = G'(X_T) := \operatorname{atan}(X_T). \end{aligned} \tag{76}$$

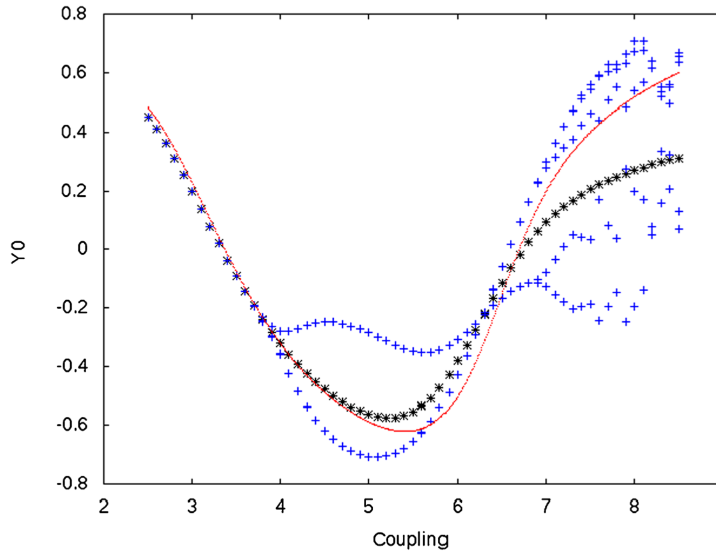

FIG. 2. *Comparison of algorithms' output for different value of the coupling parameter and for the same data as in Example* (75): *two-level* (*black star*), *one-level* (*blue cross*), *true value* (*red line*). *The* two-level algorithm *converges for larger coupling parameter than the* one-level 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*.

It comes from the Pontryagin principle applied to the Mean-Field Game

$$\inf_{\alpha} \mathbb{E}\left[G(X_t^\alpha) + \int_0^T \left(\frac{1}{2\rho}\alpha_t^2 + X_t^\alpha \operatorname{atan}\big(\mathbb{E}[X_t^\alpha]\big)\right) \mathrm{d}t \right]$$

with $\mathrm{d}X_t^\alpha = \alpha_t \,\mathrm{d}t + \mathrm{d}W_t$; see, for example, [17].

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 $Y0$ 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.

APPENDIX

A.1. A discrete Itô formula. We consider the following Euler scheme on the discrete time grid π of the interval $[0, T]$ (recall (22)):

$$\bar{X}_{t_{i+1}} = \bar{X}_{t_i} + b_i(t_{i+1} - t_i) + \sigma_i \sqrt{t_{i+1} - t_i}\,\varpi_i, \tag{77}$$

where $(\varpi_i)_{i \le n}$ 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 $\|\varpi_i\|_{2\alpha}^2 \le \Lambda h_i$, and $(b_i, \sigma_i) \in L^2(\mathcal{F}_{t_i})$, for all $i \le n$.

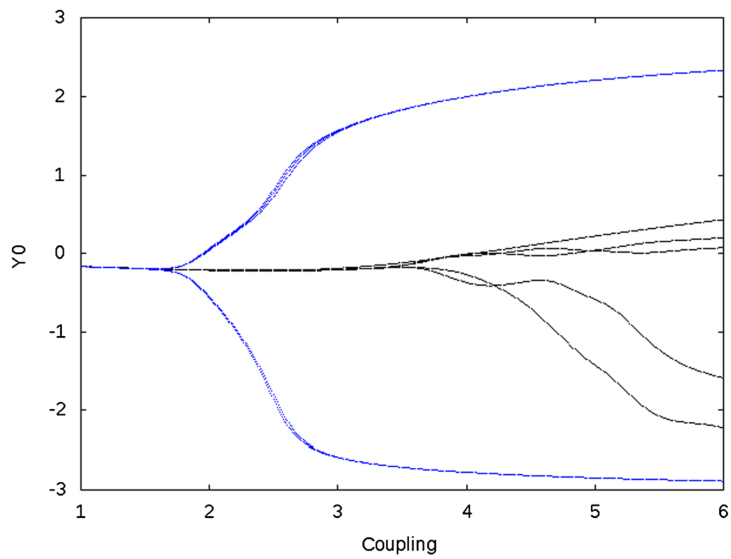

FIG. 3. *Algorithms' output for the same data as in Example* (76): one-level algorithm (*blue line*), two-level algorithm (*black line*). *We observe the same phenomenon as in the previous model*: *The* two-level algorithm *converges to a unique value for a larger range of coupling parameter than the* 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* δ.

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

$$\bar{X}_s^{(\lambda)} = \bar{X}_{t_i} + b_i(s - t_i) + \sigma_i \lambda \sqrt{s - t_i}\, \varpi_i \tag{78}$$

and $\bar{X}_{t_n}^{(\lambda)} = \bar{X}_{t_n}$. Following the notation used in the proof of Lemma 9, we just write $(\bar{X}_s)_{0 \le s \le T}$ for $(\bar{X}_s^{(1)})_{0 \le s \le T}$, which defines a continuous version of the Euler scheme given in (77).

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

$$\begin{aligned}
&\mathcal{U}(t_{i+1}, \bar{X}_{t_{i+1}}, [\bar{X}_{t_{i+1}}]) \\
&\quad = \mathcal{U}(t_i, \bar{X}_{t_i}, [\bar{X}_{t_i}]) + \int_{t_i}^{t_{i+1}} \partial_t \mathcal{U}(s, \bar{X}_s, [\bar{X}_s])\, \mathrm{d}s \\
&\qquad + \int_{t_i}^{t_{i+1}} \left(\partial_x \mathcal{U}(s, \bar{X}_s, [\bar{X}_s]) \cdot b_i + \frac{1}{2} \int_0^1 \mathrm{Tr}\big[\partial_{xx}^2 \mathcal{U}(s, \bar{X}_s^{(\lambda)}, [\bar{X}_s]) a_i\big]\, \mathrm{d}\lambda \right) \mathrm{d}s \\
&\qquad + \int_{t_i}^{t_{i+1}} \hat{\mathbb{E}}\big[\partial_\mu \mathcal{U}(s, \bar{X}_s, [\bar{X}_s])(\langle \bar{X}_s \rangle) \cdot \langle b_i \rangle\big]\, \mathrm{d}s \\
&\qquad + \frac{1}{2} \int_0^1 \hat{\mathbb{E}}\big[\mathrm{Tr}\big[\partial_\upsilon \partial_\mu \mathcal{U}(s, \bar{X}_s, [\bar{X}_s])(\langle \bar{X}_s^{(\lambda)} \rangle) \langle a_i \rangle\big]\, \mathrm{d}\lambda\big]\, \mathrm{d}s
\end{aligned}$$

$$+\int_{t_i}^{t_{i+1}} \partial_x \mathcal{U}\big(s, \bar{X}_s^{(0)}, [\bar{X}_s]\big) \frac{\sigma_i \varpi_i}{2\sqrt{s-t_i}} \, \mathrm{d}s + \delta\mathcal{M}(t_i, t_{i+1}) + \delta\mathcal{T}(t_i, t_{i+1}),$$

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 $\|\delta\mathcal{M}(t_i, t_{i+1})\|_{2\alpha} \le C_\Lambda h_i^2$ and $\|\delta\mathcal{T}(t_i, t_{i+1})\|_{2\alpha} \le C_\Lambda h_i^{\frac{3}{2}}$.

PROOF. By writing

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

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

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

Now we observe that

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

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

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

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 $hat\mathbb{E}[|\mathcal{T}_2(s)|^{2\alpha}]^{1/(2\alpha)} \le Ch_i^{\frac{1}{2}}$.

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

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

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

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

we complete the proof. □

A.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*,

$$\max_{t\in\pi^k}\|\bar{X}_t\|_{2\alpha} \le C_\Lambda\Big(1 + \|\bar{X}_{r_k}\|_{2\alpha} + \delta\max_{t\in\pi^k}\big\|\mathcal{U}\big(t, \bar{X}_t, [\bar{X}_t]\big) - Y_t\big\|_{2\alpha}\Big). \tag{79}$$

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

$$\big|b\big(\bar{X}_{t_i}, \bar{Y}_{t_i}, [\bar{X}_{t_i}, \bar{Y}_{t_i}]\big)\big| \le C_\Lambda\big(1 + |\bar{X}_{t_i}| + \|\bar{X}_{t_i}\|_{2\alpha} + \mathfrak{d}_i + \|\mathfrak{d}_i\|_{2\alpha}\big). \tag{80}$$

Recall that the scheme for the forward component reads

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

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

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

where we used again Bürkholder–Davis–Gundy inequality for discrete martingales.

Combining (80) with the boundedness of σ, we then have

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

Using the discrete version of Gronwall's lemma, the result easily follows. □

REFERENCES

[1] ACHDOU, Y., CAMILLI, F. and CAPUZZO-DOLCETTA, I. (2013). Mean field games: Convergence of a finite difference method. *SIAM J. Numer. Anal.* **51** 2585–2612. MR3097034

[2] ACHDOU, Y. and CAPUZZO-DOLCETTA, I. (2010). Mean field games: Numerical methods. *SIAM J. Numer. Anal.* **48** 1136–1162. MR2679575

[3] ACHDOU, Y. and LAURIÈRE, M. (2016). Mean field type control with congestion (II): An augmented Lagrangian method. *Appl. Math. Optim.* **74** 535–578. MR3575615

[4] ACHDOU, Y. and PEREZ, V. (2012). Iterative strategies for solving linearized discrete mean field games systems. *Netw. Heterog. Media* **7** 197–217. MR2928376

[5] ACHDOU, Y. and PORRETTA, A. (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** 161–186. MR3452251

[6] ALANKO, S. (2015). *Regression-Based Monte Carlo Methods for Solving Nonlinear PDEs*. ProQuest LLC, Ann Arbor, MI. Ph.D. thesis, New York Univ. MR3419247

[7] BAYRAKTAR, E., BUDHIRAJA, A. and COHEN, A. (2018). Rate control under heavy traffic with strategic servers. *Ann. Probab.* To appear.

[8] BENAMOU, J.-D. and CARLIER, G. (2015). Augmented Lagrangian methods for transport optimization, mean field games and degenerate elliptic equations. *J. Optim. Theory Appl.* **167** 1–26. MR3395203

[9] BENDER, C. and ZHANG, J. (2008). Time discretization and Markovian iteration for coupled FBSDEs. *Ann. Appl. Probab.* **18** 143–177. MR2380895

[10] BENSOUSSAN, A., FREHSE, J. and YAM, P. (2013). *Mean Field Games and Mean Field Type Control Theory*. Springer, New York. MR3134900

[11] BENSOUSSAN, A., FREHSE, J. and YAM, S. C. P. (2015). The master equation in mean field theory. *J. Math. Pures Appl.* (9) **103** 1441–1474. MR3343705

[12] BENSOUSSAN, A., FREHSE, J. and YAM, S. C. P. (2017). On the interpretation of the Master Equation. *Stochastic Process. Appl.* **127** 2093–2137. MR3652408

[13] BUCKDAHN, R., LI, J., PENG, S. and RAINER, C. (2017). Mean-field stochastic differential equations and associated PDEs. *Ann. Probab.* **45** 824–878. MR3630288

[14] CARDALIAGUET, P. (2012). Notes from P. L. Lions' lectures at the Collège de France. Notes. Available at https://www.ceremade.dauphine.fr/~cardalia/MFG100629.pdf.

[15] CARDALIAGUET, P., DELARUE, F., LASRY, J.-M. and LIONS, P.-L. (2015). The master equation and the convergence problem in mean field games. Preprint. Available at arXiv:1509.02505.

[16] CARDALIAGUET, P. and HADIKHANLOO, S. (2017). Learning in mean field games: The fictitious play. *ESAIM Control Optim. Calc. Var.* **23** 569–591. MR3608094

[17] CARMONA, R. and DELARUE, F. (2013). Probabilistic analysis of mean-field games. *SIAM J. Control Optim.* **51** 2705–2734. MR3072222

[18] CARMONA, R. and DELARUE, F. (2018). *Probabilistic Theory of Mean Field Games with Applications. I. Mean Field FBSDEs, Control, and Games. Probability Theory and Stochastic Modelling* **83**. Springer, Cham. MR3752669

[19] CARMONA, R. and DELARUE, F. (2018). *Probabilistic Theory of Mean Field Games. Vol. II. Mean Field Games with Common Noise and Master Equations. Probability Theory and Stochastic Modelling* **84**. Springer, Cham.

[20] CARMONA, R., DELARUE, F. and LACHAPELLE, A. (2013). Control of McKean–Vlasov dynamics versus mean field games. *Math. Financ. Econ.* **7** 131–166. MR3045029

[21] CHASSAGNEUX, J.-F., CRISAN, D. and DELARUE, F. (2014). A Probabilistic approach to classical solutions of the master equation for large population equilibria. Preprint. Available at arXiv:1411.3009.

[22] DELARUE, F. and MENOZZI, S. (2006). A forward-backward stochastic algorithm for quasi-linear PDEs. *Ann. Appl. Probab.* **16** 140–184. MR2209339

[23] DELARUE, F. and MENOZZI, S. (2008). An interpolated stochastic algorithm for quasi-linear PDEs. *Math. Comp.* **77** 125–158. MR2353946

[24] GUÉANT, O. (2012). New numerical methods for mean field games with quadratic costs. *Netw. Heterog. Media* **7** 315–336. MR2928382

[25] HADIKHANLOO, S. (2017). Learning in anonymous nonatomic games with applications to first-order mean field games. Preprint. Available at arXiv:1704.00378.

[26] LACHAPELLE, A., SALOMON, J. and TURINICI, G. (2010). Computation of mean field equilibria in economics. *Math. Models Methods Appl. Sci.* **20** 567–588. MR2647032

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

[28] MA, J., PROTTER, P. and YONG, J. M. (1994). Solving forward-backward stochastic differential equations explicitly—A four step scheme. *Probab. Theory Related Fields* **98** 339–359. MR1262970

[29] PHAM, H. and WEI, X. (2018). Bellman equation and viscosity solutions for mean-field stochastic control problem. *ESAIM Control Optim. Calc. Var.* **24** 437–461. MR3843191

[30] POPOV, B. and TOMOV, V. (2015). Central schemes for mean field games. *Commun. Math. Sci.* **13** 2177–2194. MR3399921

J. F. CHASSAGNEUX
LABORATOIRE DE PROBABILITÉS, STATISTIQUE ET MODÉLISATION
UNIVERSITÉ PARIS DIDEROT
PARIS
FRANCE
E-MAIL: chassagneux@lpsm.paris

D. CRISAN
DEPARTMENT OF MATHEMATICS
IMPERIAL COLLEGE LONDON
LONDON SW7 2AZ
UNITED KINGDOM
E-MAIL: d.crisan@imperial.ac.uk

F. DELARUE
LABORATOIRE JEAN-ALEXANDRE DIEUDONNÉ
UNIVERSITÉ DE NICE SOPHIA-ANTIPOLIS
28 AVENUE VALROSE
06100 NICE
FRANCE
E-MAIL: delarue@unice.fr