A CUBATURE BASED ALGORITHM TO SOLVE DECOUPLED MCKEAN-VLASOV FORWARD BACKWARD STOCHASTIC DIFFERENTIAL EQUATIONS

P.E. CHAUDRU DE RAYNAL AND C.A. GARCIA TRILLOS

Abstract. We propose a new algorithm to approach weakly the solution of a McKean-Vlasov SDE. Based on the cubature method of Lyons and Victoir [LV04], the algorithm is deterministic differing from the the usual methods based on interacting particles. It can be parametrized in order to obtain a given order of convergence.

Then, we construct implementable algorithms to solve decoupled Forward Backward Stochastic Differential equations (FBSDE) of McKean-Vlasov type, which appear in some stochastic control problems in a mean field environment. We give two algorithms and show that they have convergence of order one and two under appropriate regularity conditions.

This is a reprint version of an article published in Stochastic Processes and their Applications 125 (2015) 2206–2255

1 Introduction

We call decoupled McKean-Vlasov forward backward stochastic differential equation (MKV-FBSDE) the following FBSDE system:

$$\begin{cases} dX_t^x = \sum_{i=0}^d V_i(t, X_t^x, \mathbb{E}\varphi_i(X_t^x))dB_t^i \\ dY_t^x = -f(t, X_t^x, Y_t^x, Z_t^x, \mathbb{E}\varphi_f(X_t^x, Y_t^x))dt + Z_t^x dB_t^{1:d} \\ X_0^x = x, \quad Y_T^x = \phi(X_T^x) \end{cases} \tag{1.1}$$

for any t in $[0,T]$, $T>0$ be given. We place ourselves in a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0})$, with $B_t^{1:d}$ a d-dimensional adapted Brownian motion and $B_t^0 = t$. We take $V_i : (t,y,w) \in [0,T]\times\mathbb{R}^d\times\mathbb{R} \mapsto V_i(t,y,w)$; functions $\varphi_i : y \in \mathbb{R}^d \mapsto \varphi_i(y) \in \mathbb{R}$, $i = 0, \cdots, d$ and $\varphi_f : (y,y') \in \mathbb{R}^d\times\mathbb{R} \mapsto \varphi_f(y,y')$ and the mapping $f : t,y,y',z,w \in [0,T]\times\mathbb{R}^d\times\mathbb{R}\times\mathbb{R}^d\times\mathbb{R} \mapsto f(t,y,y',z,w) \in \mathbb{R}$ to be infinitely differentiable with bounded derivatives. The mapping ϕ is an at least Lipschitz function from $\mathbb{R}^d$ to $\mathbb{R}$ whose precise regularity is given below.

McKean Vlasov processes may be regarded as a limit approximation for interacting systems with large number of particles. They appeared initially in statistical mechanics, but are now used in many fields because of the wide range of applications requiring large populations interactions. For example, they are used in finance, as factor stochastic volatility models [Ber09] or uncertain volatility models [GHL11]; in economics, in the theory of "mean field games" recently developed by J.M. Lasry and P.L. Lions in a series of papers [LL06a, LL06b, LL07a, LL07b] (see also [CDL12, CD12a, CD12b] for the probabilistic counterpart) and also in physics, neuroscience, biology, etc. In section 5, we present a class of control problems in which equation (1.1) explicitly appears.

The note of Sznitman [Szn91] gives a complete overview on the topic of systems with a large number of particles. A proof of the existence and uniqueness of the solution of a MKV-FBSDE system related but different to the one of our setup is found in [BLP09]. These existence and uniqueness results are

Key words and phrases. Cubature; McKean-Vlasov processes; BSDE; mean field games; non-local PDE.

easily extended to (1.1).

A cubature algorithm for MKV-FBSDE processes. Cubature on Wiener space was introduced in 2004 by T.Lyons and N.Victoir [LV04], following the earlier work of S. Kusuoka [Kus01]. Ever since, the cubature method has been used to solve the problem of calculating Greeks in finance [Tei06], non-linear filtering problems [CG07], stochastic partial differential equations [BT08], [DTV12] and backward stochastic differential equations [CM10b, CM10a].
The main idea of the cubature method consists in replacing the Brownian motion by choosing randomly a path among an a priori (finite) set[1] of continuous functions from $[0,T]$ to $\mathbb{R}^d$ with bounded variations such that the expectation of the iterated integrals against both the Brownian and such paths are the same, up to a given order m. Hence, the SDE is replaced by a system of weighted ODEs.

We give the main idea to construct a cubature based approximation scheme for (1.1). The main issue in the case of a MKV-FBSDE is the McKean-Vlasov dependence that appears in the coefficients. This dependence breaks the Markov property (considered only on $\mathbb{R}^d$) of the process so that it is not possible to apply, a priori, many classical analysis tools. In order to handle this problem, the idea consists in taking benefit on the following observation: *given the law of the solution of the system,* (1.1) *is a classical time inhomogeneous FBSDE* (the law just acts as a time dependent parameter).

Let $(\eta_t)_{0\leq t\leq T}$ be a family of probability measures on $\mathbb{R}^d$, and let us fix the law in the McKean-Vlasov terms of (1.1) to be $(\eta_t)_{0\leq t\leq T}$. For this modified system, we may apply a classical cubature FBSDE scheme for the forward component (the time dependence of the coefficients being handled as an additional dimension). The trick consists in taking advantage of the decoupled setting: we first build a cubature tree (depending on the order of the cubature) and then go back along the nodes of the tree by computing the current value of the backward process as a conditional expectation at each node. We refer to [CM10a] or [CM10b] for a detailed description of such algorithm.

Obviously, at each step of the scheme, we pay the price of using an arbitrary probability measure as parameter for the coefficients instead of the law of the process. Therefore this law has to be chosen carefully in order to keep a good control on the error and achieve convergence. An example of a "good choice" is to take at each step of the cubature tree the discrete marginal law given by the solution of the ODEs along the cubature paths and corresponding weights. We show that for a cubature of order m and a number N of discretization steps, this choice of approximation law leads to a $N^{-(m-1)/2}$ order approximation of the expectation of any $m+2$ times continuously differentiable functional of the forward component, when all the derivatives are bounded[2], and to a first order approximation scheme of the backward component, where the given orders stand for the supremum norm error. Higher orders of approximation are also obtained by correcting some terms in the algorithm.

As it is pointed out in [LV04] and [CM10b], the regularity of the terminal condition ϕ in (1.1) may be relaxed to Lipschitz and the approximation convergence rate preserved, provided that the vector fields are uniformly non-degenerate (in fact, the condition in the given references is weaker, since the vector fields are supposed to satisfy an UFG condition, see [KS87]). This relies on the regularization properties of parabolic and semi-linear parabolic PDEs (see [Fri08] for an overview in the elliptic case and respectively [KS87] and [CD12c] for the UFG case). We show that this remains true in the McKean-Vlasov case and that the convergence rate still holds when the function ϕ is Lipschitz only and when the vector fields are uniformly elliptic.

[1] Explicit examples of such functions are given in [LV04]. We put back this discussion to a next subsection.

[2] this is the special case when ϕ is $m+2$ times continuously differentiable with bounded derivatives and $f=0$ in (1.1).

Usually, forward MKV-SDEs are solved by using particle algorithms (see for example [AKH02, TV03] or [Bos05] and references therein) in which the McKean term is approached with the empirical measure of a large number of interacting particles with independent noise. Adapting such algorithms to the forward-backward problem is not obvious as the high dimension of the involved Brownian motion (given by the number of particles) induces, a priori, a high dimension backward problem with the obvious consequences for the numerical implementation. In comparison, our proposed algorithm gives a deterministic approximation of the McKean term, and since it does not induce any additional noise, it does not increase the dimension of the backward problem.

Although our algorithm works for decoupled MKV-FBSDEs, we believe this solver may also be considered as a building block if one is interested in approaching the fully coupled case (when the forward coefficients depend on the backward variable), for example via fixed point procedures. Nevertheless, a lot of work is required to define the precise conditions and setup in which such algorithm would converge to the (or at least a) solution of the fully coupled problem.

The conditional system. Let us shortly develop what we mean with the sentence "*given the law of the solution of the system,* (1.1) *is a classical time inhomogeneous FBSDE*".
Working with a non-linear problem, such as MKV-FBSDE, could be tricky. In our case, the main object to work with is the *conditional system*. This is the formulation that allows to get rid of the dependence on the law and to replace it by a time dependent parameter.

Following the same line of arguments presented in Buckdahn et al. [BLP09], it is well seen that there exists a unique solution $\{X_t^x, Y_t^x\}_{t\geq 0}$ to the system (1.1). In a Markovian setting, the law of this couple is entirely determined by the law $\mu = (\mu_t)_{0\leq t\leq T}$ of the forward process $(X_t)_{0\leq t\leq T}$ and a given deterministic function $u : [0,T]\times\mathbb{R}^d \to \mathbb{R}$. In our case, one can show that this remains true (see Section 8 below for a proof) so that there exists a deterministic $u(t,y)$ such that for all t,

$$Y_t = u(t, X_t). \tag{1.2}$$

We prove that under appropriate assumptions u is regular and satisfies the parametrized non-local semi linear PDE:

$$\begin{cases} \partial_t u(t,y) + \mathcal{L}^\mu u(t,y) = f\left(t, y, u(t,y), (\mathcal{V}^\mu u(t,y))^T, \langle \mu_t, \varphi_f(\cdot, u(t,\cdot))\rangle\right) \\ u(T,y) = \phi(y) \end{cases}, \tag{1.3}$$

where $\mathcal{V}^\mu u$ stands for the row vector $(\nabla u \cdot V_1, \dots, \nabla u \cdot V_d)$, $(\mathcal{V}^\mu u)^T$ is the transpose of $\mathcal{V}^\mu u$ and $\mathcal{L}^\mu$ is the generator of the forward component in (1.5) below and given by:

$$\mathcal{L}^\mu := V_0(\cdot,\cdot,\langle\mu_\cdot,\varphi_0\rangle)\cdot D_y + \frac{1}{2}\mathrm{Tr}[VV^T(\cdot,\cdot,\langle\mu_\cdot,\varphi_i\rangle)D_y^2]. \tag{1.4}$$

Here, we used the duality notation $\langle\mu,\varphi_i\rangle$ for $\int\varphi_i d\mu$. Likewise, the superscript μ means that the vector fields are taken at the point $(\cdot,\cdot,\langle\mu_\cdot,\varphi_i\rangle)$ (where the $i\in\{0,\cdots,d\}$ is taken with respect to the corresponding vector field), V is the matrix $[V_1,\cdots,V_d]$, "$\cdot$" stands for the euclidean scalar product on $\mathbb{R}^d$ and "Tr" for the trace.

The conditional MKV-FBSDE system is then defined as

$$\begin{cases} dX_s^{t,y,\mu} = \sum_{i=0}^d V_i(s, X_s^{t,y,\mu}, \langle\mu_s,\varphi_i\rangle)dB_s^i \\ dY_s^{t,y,\mu} = -f(s, X_s^{t,y,\mu}, Y_s^{t,y,\mu}, Z_s^{t,y,\mu}, \langle\mu_s, \varphi_f(\cdot, u(s,\cdot))\rangle)ds + Z_s^{t,y}dB_s^{1:d} \\ X_t^{t,y,\mu} = y, \quad Y_T^{t,y,\mu} = \phi(X_T^{t,y,\mu}). \end{cases} \tag{1.5}$$

Let us remark that in this setting, we can also define the deterministic mapping $v : [0,T]\times\mathbb{R}^d \to \mathbb{R}$ such that

$$Z_t = v(t, X_t). \tag{1.6}$$

As in the classical BSDE theory, under appropriate regularity conditions,

$$v(t,x) = (\mathcal{V}^{\mu} u(t,x))^T.$$

Assumptions. As the reader might guess from the previous discussion, the error analysis of the proposed algorithm uses extensively the regularity of $(t,x) \in [0,T] \times \mathbb{R}^d \mapsto \mathbb{E}\phi(X_T^{t,x,\mu})$ for the forward part and of the solution u of (1.3) for the backward part. Therefore, we present two types of hypotheses that guarantee that the required regularity is attained.

The first option we present is to require smoothness on the boundary condition and all the coefficient functions, from where we will deduce the necessary regularity. However, it is also interesting to consider boundary conditions with less regularity. In this case, we need to compensate the regularity loss by imposing stronger diffusion conditions on the forward variable, namely asking for uniform ellipticity of the diffusion matrix V.

(SB): We say that assumption **(SB)** holds if the mapping ϕ in (1.1) is C_b^∞.

(LB): We say that assumption **(LB)** holds if the mapping ϕ in (1.1) is uniformly Lipschitz continuous and if the matrix VV^T is uniformly elliptic i.e., there exists $c > 0$ such that

$$\forall (t,y,w) \in [0,T] \times \mathbb{R}^d \times \mathbb{R}, \forall \varsigma \in \mathbb{R}^d, c^{-1}|\varsigma|^2 \leq VV^T(t,y,w)\varsigma \cdot \varsigma \leq c|\varsigma|^2.$$

Remark. *A reader familiarized with the cubature method might wonder why we assume uniform ellipticity instead of the weaker UFG condition usually needed for applying the method with Lipschitz boundary conditions. The reason is that the smoothing results of Kusuoka and Stroock* [KS87] *hold for space dependent vector fields only, and therefore do not apply directly to our framework with a time dependence coming from the McKean term. There is some extension in the time inhomogeneous case that do not include derivatives in the V_0 direction (see for example* [CM02] *and references therein), but, to the best of our knowledge, there is no result that could be applied to our framework.*

Towards a more general class of coefficients. We have chosen to work with the assumed explicit dependence of the coefficients with respect to the law, as it is very natural in practice. In fact as a consequence of our analysis, our algorithm works for a more general class of MKV-FBSDE, i.e. for a system written as

$$\begin{cases} dX_t^x = \sum_{i=0}^d V_i(t, X_t^x, \mu_t) dB_t^i \\ dY_t^x = -f(t, X_t^x, Y_t^x, Z_t^x, \mu_t^{X,Y}) dt + Z_t^x dB_t^{1:d} \\ X_0^x = x, \quad Y_T^x = \phi(X_T^x), \end{cases} \tag{1.7}$$

where $\mu^{X,Y} = (\mu_t^{X,Y})_{0 \leq t \leq T}$ denotes the joint law of $(X_t, Y_t)_{0 \leq t \leq T}$, the coefficients V_i, $0 \leq i \leq d$ (and f) are Lipschitz continuous with respect to an appropriately defined distance in the space of probability measures on $\mathbb{R}^d$ (respectively $\mathbb{R}^d \times \mathbb{R}$) which integrates the usual Euclidean norm. The distance we consider is defined by duality: let $\mathcal{F}$ be a (sufficiently rich[3]) class of functions (that will be detailed in the following). Then we define the distance $d_{\mathcal{F}}$ between two probability measures on $\mathbb{R}^n$ by

$$d_{\mathcal{F}}(\mu, \nu) = \sup_{\varphi \in \mathcal{F}} |\langle \varphi, \mu - \nu \rangle|. \tag{1.8}$$

In this decoupled case the Lipschitz property of the coefficients with respect to $d_{\mathcal{F}}$ ensures the existence of a unique solution of (1.7) [4]. Then, we are able to analyze the convergence of our procedure in two different cases. When $\mathcal{F}$ is the class of 1-Lipschitz functions, i.e. when the distance is the so-called Wasserstein-1 distance, and when the vector fields are uniformly elliptic our algorithm leads to an $N^{-1/2}$ order approximation[5]. When $\mathcal{F}$ is the class of C_b^∞ functions we obtain an N^{-1} order approximation without any ellipticity assumption on the diffusion matrix.

[3]It should contain the space of 1-Lipschitz functions.

[4]by definition of $\mathcal{F}$, the distance $d_{\mathcal{F}}$ is less than or equal to the Wasserstein 2 distance. Then, one uses the same kind of arguments as in [Szn91].

[5]Recall that N denotes the number of discretization steps.

Objectives and organization of this paper. As a corollary of the discussion on the conditional system we can resume our objective as the approximation of $\mathbb{E}\phi(X_T^x)$, where $(X_t^x)_{0\leq t\leq T}$ is the solution of (1.1) and of u satisfying (1.2).

This paper is organized as follows: Section 2 states the algorithm, while the convergence rate of the forward and backward approximations is stated in Section 3. Then, we give a numerical example for each set of hypotheses **(SB)** and **(LB)** in Section 4. A class of control problems is introduced in Section 5. The remainder of the paper is dedicated to the proof of the convergence. For the sake of simplicity, we first recall some definitions, basic facts and notations in Section 6. The forward and backward convergence rates for regular boundary conditions are successively proved in Section 7 and the common mathematical tools are given in Section 8. Section 9 presents the extension to the Lipschitz boundary condition case and the announced generalization of the law dependence of the McKean terms.

Notations. As we are treating with objects exhibiting different dependences, the notation can become a bit heavy. For the sake of simplicity, we adopt the following conventions. We denote by φ the function $\varphi = [\varphi_1, \cdots, \varphi_d]$. For two positive integers $i < j$, the notation "$i : j$" means "from index i to j". For all $\varsigma \in \mathbb{R}^n$, $n \in \mathbb{N}$ the partial derivative $[\partial/\partial\varsigma]$ is denoted by ∂_ς. Let $g : \ y \in \mathbb{R}^d \mapsto g(y) \in \mathbb{R}$ be a p-continuously differentiable function. We set $||g||_{\infty,p} := \max_{j\leq p} ||\partial_y^j g||_\infty$. We say that a function g from $[0,T] \times \mathbb{R}^d \times \mathbb{R}^d$ is C_b^p, $p \in \mathbb{N}^*$ if it is bounded and p-times continuously differentiable with bounded derivatives. We usually denote by η (eventually with an exponent) a family of probability measures $(\eta_t)_{0\leq t\leq T}$ on $\mathbb{R}^d$. For such a family, we set $\mathcal{L}^\eta$ to be the second order operator of the form (1.4) with η instead of μ. In general, we will work with the vector fields taken at the point $(\cdot, \cdot, \langle\mu_\cdot, \varphi_i\rangle)$ (where the $i \in \{0, \cdots, d\}$ signals the corresponding vector field) and in general we will omit the explicit dependence on μ in the notation. In any case, we will mark the law dependence explicitly when needed, in particular when a dependence with respect to a different law appears.

2 Algorithm

Multi-index. Multi-indices allow to easily manage differentiation and integration in several dimensions. Let

$$\mathcal{M} = \{\emptyset\} \cup \bigcup_{l\in\mathbb{N}^*} \{0, 1, \ldots, d\}^l, \tag{2.1}$$

denotes the set of multi-indices where $\emptyset$ refers, for the sake of completeness, to the zero-length multi-index. We define "$*$" to be the concatenation operator such that if $\beta^1 = (\beta_1^1, \ldots, \beta_l^1)$ and $\beta^2 = (\beta_1^2, \ldots, \beta_n^2)$ then $\beta^1 * \beta^2 = (\beta_1^1, \ldots, \beta_l^1, \beta_1^2, \ldots, \beta_n^2)$.

Cubature on Wiener Space. In the introduction, we mentioned that the cubature method consists in replacing the Brownian path by choosing randomly a path ω among a finite subset $\{\omega_1, \cdots, \omega_\kappa\}$, $\kappa \in \mathbb{N}^*$, of $C_{\text{bv}}^0([0,T], \mathbb{R}^d)$ (the set of continuous functions from $[0,T]$ to $\mathbb{R}^d$ with bounded variations) with probability λ in $\{\lambda_1, \cdots, \lambda_\kappa\} \subset \mathbb{R}^+$. We precise this notion with the definition given by Lyons and Victoir [LV04]:

Definition 2.1. *Let m be a natural number and $t \in \mathbb{R}^+$. A m-cubature formula on the Wiener space $C^0([0,t], \mathbb{R}^d)$ is a discrete probability measure $\mathbb{Q}_t$ with finite support on $C_{\text{bv}}^0([0,t], \mathbb{R}^d)$ such that the expectation of the iterated Stratonovitch integrals of degree m under the Wiener measure and under the cubature measure $\mathbb{Q}_t$ are the same, i.e., for all multi-index $(i_1, \cdots, i_l) \in \{1, \cdots, d\}^l$, $l \leq m$*

$$\mathbb{E}\int_{0<t_1<\cdots<t_l<t} \circ dB_{t_1}^{i_1} \cdots \circ dB_{t_l}^{i_l} = \mathbb{E}_{\mathbb{Q}_t}\int_{0<t_1<\cdots<t_l<t} \circ dB_{t_1}^{i_1} \cdots \circ dB_{t_l}^{i_l} = \sum_{j=1}^{l} \lambda_j \int_{0<t_1<\cdots<t_l<t} d\omega_j^{i_1}(t_1) \cdots d\omega_j^{i_l}(t_l),$$

where "$\circ$" stands for the Stratonovich operator and ω_j^i for the i^{th} coordinate of the j^{th} path.

As a direct consequence of the Taylor-Stratonovitch expansion, a cubature formula of degree m is such that:

$$|(\mathbb{E} - \mathbb{E}_{\mathbb{Q}_t})F(B_t^{1:d})| \leq Ct^{(m+1)/2}||F||_{m+2,\infty}, \tag{2.2}$$

for all bounded and $m+2$ times continuously differentiable function F with bounded derivatives.

Of course this error control is not in general small, but the Markovian and scaling properties of the Brownian motion can be used to apply the cubature method iteratively in small subdivisions of the interval $[0,t]$ for which we have a good error control.

Indeed, consider a cubature formula $\mathbb{Q}_1$ of order $m \in \mathbb{N}^*$ with support $\{\omega_1, \dots, \omega_\kappa\}$ and corresponding weights $\{\lambda_1, \dots, \lambda_\kappa\}$. For all $h > 0$ and any $t \in [0, T-h]$, one can deduce a cubature measure $\mathbb{Q}_{t,t+h}$ of order m with finite support on $C^0_{\text{bv}}([t,t+h], \mathbb{R}^d)$ equal to $\{\tilde{\omega}_1, \cdots, \tilde{\omega}_\kappa\}$ with the same weights $\{\lambda_1, \dots, \lambda_\kappa\}$ and where the paths are defined as $\tilde{\omega}_j : s \in [t, t+h] \mapsto \tilde{\omega}(s) = \sqrt{h}\omega_j((s-t)/h)$ for all $1 \leq j \leq \kappa$.

Then, by virtue of the Markovian property, this subdivision leads to the construction of a tree which has κ^k nodes (corresponding to the number of paths) at the k^{th} subdivision. Each path $\tilde{\omega}_{(i_1,\dots,i_k)}$, where $(i_1, \dots, i_k)$ stands for the trajectory of the path, has then a cumulate weight of the form $\Lambda_{(i_1,\dots,i_k)} = \prod_{j=1}^k \lambda_{i_j}$, see the example and figure 1 below.

2.1 Main idea

Take a subdivision of the time interval $[0,T]$ into $N \in \mathbb{N}^*$ steps $0 = T_0 < \cdots < T_N = T$. The procedure can be decomposed in two parts:

(1) **Building the tree $\mathcal{T}$.** This part of our algorithm can be resumed as a combined Euler-cubature approach and can be divided in four steps.
   (a) First, freeze in the space of probability measures the law that appears in the coefficients of (1.1) (the choice of this measure is explained below). At step 0, this measure is the Dirac mass in the starting point.
   (b) Freeze, in time, the given deterministic measure: this is an Euler step.
   (c) Apply the cubature method. This will produce a cloud of deterministic particles given by the solution of the resulting ODEs
   (d) At each step, construct a discrete measure coming from the obtained cloud of particles and their associated cumulative weights. This is the law to be used to approximate the law in the coefficients of (1.1).

   The reader might guess that the order of approximation of such an algorithm is one, due to the Euler step. Hence, in order to obtain higher order, we expand the function that appears in the McKean-Vlasov part up to a certain order, which is denoted by "q" in the sequel.

(2) **The backward component**. The backward component of the algorithm runs by assigning the value of the function on the boundary (known from the definition of the equation), and then back-propagating its value thanks to
   (a) A discretization scheme for the backward approximation
   (b) The cubature measure for finding conditional expectations.

   As mentioned before, we present two versions of the algorithm, with convergence of order one and two. As will be clear in the definition, the change in convergence order requires the use of a different backward scheme and cubature order.

Example: one dimensional cubature of order $m = 3$. In this case, we may use a cubature formula with $\kappa = 2$ paths given by $\{+t, -t\}$, and associated weights: $\{\lambda_1 = 1/2, \lambda_2 = 1/2\}$. Let us explain the idea behind the algorithm with an example for 2 steps, as shown in Figure 1. We initialize the tree at a given point x, and the law at T_0 as δ_x. Then, we find two descendants given as the solution of an ODE that uses the position X, the two cubature paths, and an approximated law using

the information at time 0. Each descendant will have a weight equal to the product of the weight of its parent times the weight given to the corresponding cubature path. Once all nodes at time T_1 are calculated, we obtain the discrete measure $\hat{\mu}_{T_1}$, the law approximation at time T_1. The process is then repeated for each node at time T_1 to reach the final time $T = T_2$.

Figure 1 right illustrates the idea behind the backward approximation: the approximated function $\hat{u}$ is defined first at the leaves of the constructed tree, and then back-propagates using the approximated law to obtain $\hat{u}$ at previous times. The back-propagation is made by conditional expectation: average with respect to the weight of each cubature path.

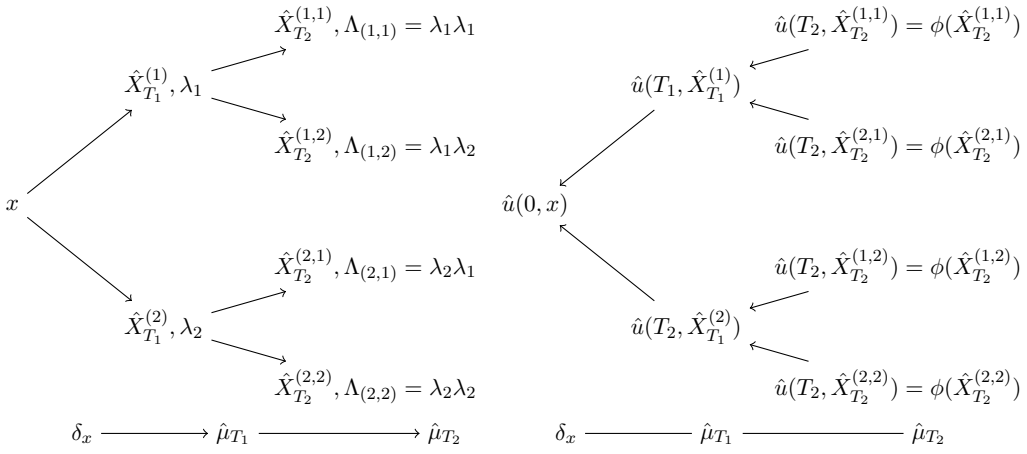

FIGURE 1. Left: Cubature tree. Right: Backward scheme.

2.2 Algorithms

Having the general idea in mind, we can give a precise description of each of the two main parts of our proposed algorithm. Since the cubature involves Stratanovitch integrals, we set:

$$\bar{V}_0^k = V_0^k - \frac{1}{2}\sum_{i,j=1}^{d} V_j^i \frac{\partial}{\partial x_j} V_k^j, \tag{2.3}$$

for all $k \in \{1, \cdots, d\}$ and rewrite the system (1.1) as:

$$\begin{cases} dX_t^x = \bar{V}_0(t, X_t^x, \mathbb{E}\varphi_0(X_t^x))dt + \sum_{i=1}^{d} V_i(t, X_t^x, \mathbb{E}\varphi_i(X_t^x)) \circ dB_t^i \\ dY_t^x = -f(t, X_t^x, Y_t^x, Z_t^x, \mathbb{E}\varphi_f(X_t^x, Y_t^x))dt + Z_t^x dB_t^{1:d} \\ X_0^x = x, \quad Y_T^x = \phi(X_T^x). \end{cases} \tag{2.4}$$

In order to make the description of the algorithm as clear as possible, for any k, κ in $\mathbb{N}$, we set $\mathcal{S}_\kappa(k) = \{\text{multi-index } (j_1, \cdots, j_k) \in \{1, \cdots, \kappa\}^k\}$, i.e., $\mathcal{S}_\kappa(k)$ is the set of multi-indices with entries between $1, \ldots, \kappa$ of length (exactly) k.

2.2.1 Building the tree $\mathcal{T}(\gamma, q, m)$

The subdivision. Let $\gamma > 0$, $N \in \mathbb{N}^*$, let $0 = T_0 < \ldots < T_N = T$ be a discretization of the time interval $[0, T]$ given as

$$T_k = T\left[1 - \left(1 - \frac{k}{N}\right)^\gamma\right] \tag{2.5}$$

and let $\Delta_{T_k} = T_k - T_{k-1}$.

Remark. *When the boundary condition is not smooth, we take a non-uniform subdivision in order to refine the discretization step close to the boundary as proposed by Kusuoka in* [Kus01]. *If, on the contrary, the boundary condition is smooth, we may use a classical uniform discretization. For this reason, in the following we assume that* $\gamma = 1$ *if* ***(SB)*** *holds, and that* $\gamma > m - 1$ *under* ***(LB)***.

Let γ be given as explained above, q and m be two given integers, and $\big\{\{\omega_1, \cdots, \omega_\kappa\}, \{\lambda_1, \cdots, \lambda_\kappa\}\big\}$ be a m order cubature (the number κ of paths and weights depends on m). Recall that $\omega_j : t \in [0,1] \mapsto (\omega_j^1(t), \ldots, \omega_j^d(t)) \in \mathbb{R}^d$ is some continuous function with bounded variation and for all t in $[0,T]$, we set $\omega_0(t) = t$. Examples of cubature formulas of order $3, 5, 7, 9, 11$ can be found in [LV04] or [GL11].

Algorithm 1 Cubature Tree $\mathcal{T}(\gamma, q, m)$

Set $(X^{\emptyset}, \hat{\mu}_{T_0}, \Lambda_0) = (x, \delta_x, 1)$
for $0 \le i \le d$ **do**
    Set $F_i(t, \hat{\mu}_{T_0}) = \sum_{p=0}^{q-1} \frac{1}{p!}(t - T_0)^p \langle \delta_x, (\mathcal{L}^{\delta_x})^p \varphi_i \rangle$
end for
for $1 \le k \le N-1$ **do**
    for $\pi \in \mathcal{S}_\kappa(k)$ **do**
        for $1 \le j \le \kappa$ **do**
            Define $\hat{X}^{\pi * j}_{T_{k+1}}$ as the solution of the ODE:

$$d\hat{X}_t^{\pi*j} = \textstyle\sum_{i=0}^d V_i(t, \hat{X}_t^{\pi*j}, F_i(t, \hat{\mu}_{T_k}))\sqrt{\Delta_{T_{k+1}}} d\omega_j^i((t - T_k)/\Delta_{T_{k+1}}),$$
$$\hat{X}_{T_k}^{\pi*j} = \hat{X}_{T_k}^{\pi}$$

            Set the associated weight: $\Lambda_{\pi * j} = \Lambda_\pi \lambda_j$
        end for
    end for
    Set $\hat{\mu}_{T_{k+1}} = \sum_{\pi \in \mathcal{S}_\kappa(k+1)} \Lambda_\pi \delta_{\hat{X}^{\pi}_{T_{k+1}}}$
    for $0 \le i \le d$ **do**
        Set $F_i(t, \hat{\mu}_{T_{k+1}}) = \sum_{p=0}^{q-1} \frac{1}{p!}(t - T_k)^p \langle \hat{\mu}_{T_{k+1}}, (\mathcal{L}^{\hat{\mu}})^p \varphi_i \rangle$
    end for
end for

2.2.2 Backward scheme

Algorithm 2 First order backward scheme

for $\pi \in \mathcal{S}_\kappa(N)$ **do**
  Set $\hat{u}^1(T_N, \hat{X}^\pi_{T_N}) = \phi(\hat{X}^\pi_{T_N})$
  Set $\hat{v}^1(T_N, \hat{X}^\pi_{T_N}) = 0$
end for
for $N-1 \geq k \geq 1$ **do**
  for $\pi \in \mathcal{S}_\kappa(k)$ **do**
    $\hat{v}^1(T_k, \hat{X}^\pi_{T_k}) = \dfrac{1}{\Delta_{T_{k+1}}} \displaystyle\sum_{j=1}^{\kappa} \lambda_j \hat{u}^1(T_{k+1}, X^{\pi * j}_{T_{k+1}}) \sqrt{\Delta_{T_{k+1}}} \omega_j(1)$
    for $1 \leq j \leq \kappa$ **do**
      $\hat{\Theta}^{\pi,1}_{k+1,k}(j) = \left(T_{k+1}, \hat{X}^{\pi * j}_{T_{k+1}}, \hat{u}^1(T_{k+1}, \hat{X}^{\pi * j}_{T_{k+1}}), \hat{v}^1(T_k, \hat{X}^\pi_{T_k}), F^1(T_{k+1}, \hat{\mu}_{T_{k+1}})\right)$
    end for
    $\hat{u}^1(T_k, \hat{X}^\pi_{T_k}) = \displaystyle\sum_{j=1}^{\kappa} \lambda_j \left(\hat{u}^1(T_{k+1}, X^{\pi * j}_{T_{k+1}}) + \Delta_{T_{k+1}} f(\hat{\Theta}_{k+1,k}(j))\right)$
  end for
  Set $F^1(T_{k+1}, \hat{\mu}_{T_{k+1}}) = \langle \hat{\mu}_{T_{k+1}}, \varphi_f(\cdot, \hat{u}^1(T_{k+1}, \cdot)) \rangle$
end for

Algorithm 3 Second order backward scheme

for $\pi \in \mathcal{S}_\kappa(N)$ **do**
  Set $\hat{u}^2(T_N, \hat{X}^\pi_{T_N}) = \phi(\hat{X}^\pi_{T_N})$
  Set $\hat{v}^2(T_N, \hat{X}^\pi_{T_N}) = 0$
end for
for $\pi \in \mathcal{S}_\kappa(N-1)$ **do**
  Set $\hat{u}^2(T_{N-1}, \hat{X}^\pi_{T_{N-1}}) = \hat{u}^1(T_{N-1}, \hat{X}^\pi_{T_{N-1}})$ and $\hat{v}^2(T_{N-1}, \hat{X}^\pi_{T_{N-1}}) = \hat{v}^1(T_{N-1}, \hat{X}^\pi_{T_{N-1}})$
  Set $F^2(T_{N-1}, \hat{\mu}_{T_{N-1}}) = \langle \hat{\mu}_{T_{N-1}}, \varphi_f(\cdot, \hat{u}^2(T_{N-1}, \cdot)) \rangle$
end for
for $N-2 \geq k \geq 1$ **do**
  for $\pi \in \mathcal{S}_\kappa(k)$ **do**
    for $1 \leq j \leq \kappa$ **do**
      $\hat{\Theta}^{\pi,2}_{k+1}(j) = \left(T_{k+1}, \hat{X}^{\pi * j}_{T_{k+1}}, \hat{u}^2(T_{k+1}, \hat{X}^{\pi * j}_{T_{k+1}}), \hat{v}^2(T_{k+1}, \hat{X}^{\pi * j}_{T_{k+1}}), F^2(T_{k+1}, \hat{\mu}_{T_{k+1}})\right)$
      $\hat{\zeta}^{\pi * j}_{k+1} := 4 \dfrac{1}{\Delta_{T_{k+1}}} \sqrt{\Delta_{T_{k+1}}} \omega_j(1) - 6 \dfrac{1}{\Delta^2_{T_{k+1}}} \displaystyle\int_{T_k}^{T_{k+1}} (s - T_k) \sqrt{\Delta_{T_{k+1}}} d\omega_j((s - T_k)/\Delta_{T_{k+1}})$
    end for
    Set $\hat{v}^2(T_k, \hat{X}^\pi_{T_k}) = \displaystyle\sum_{j=1}^{\kappa} \lambda_j \left(\hat{u}^2(T_{k+1}, X^{\pi * j}_{T_{k+1}}) + \Delta_{T_{k+1}} f(\hat{\Theta}^{\pi * j, 2}_{k+1})\right) \hat{\zeta}^{\pi * j}_{k+1}$
    [*Predictor*] $\tilde{u}(T_k, \hat{X}^\pi_{T_k}) = \displaystyle\sum_{j=1}^{\kappa} \lambda_j \left(\hat{u}^2(T_{k+1}, X^{\pi * j}_{T_{k+1}}) + \Delta_{T_{k+1}} f(\hat{\Theta}^{\pi * j, 2}_{k+1})\right)$
    Set $\tilde{F}(T_k, \hat{\mu}_{T_k}) = \mathbb{E}_{\hat{\mu}_{T_k}} \varphi_f(\cdot, \tilde{u}(T_k, \cdot))$
    [*Corrector*] $\tilde{\Theta}^\pi_k = \left(T_k, \hat{X}^\pi_{T_k}, \tilde{u}(T_k, \hat{X}^\pi_{T_k}), \hat{v}^2(T_k, \hat{X}^\pi_{T_k}), \tilde{F}(T_k, \hat{\mu}_{T_k})\right)$
    $\hat{u}^2(T_k, \hat{X}^\pi_{T_k}) = \displaystyle\sum_{j=1}^{\kappa} \lambda_j \left(\hat{u}^2(T_{k+1}, X^{\pi * j}_{T_{k+1}}) + \frac{1}{2} \Delta_{T_{k+1}} \left(f(\hat{\Theta}^{\pi * j, 2}_{k+1}) + f(\tilde{\Theta}^\pi_k)\right)\right)$
    Set $F^2(T_k, \hat{\mu}_{T_k}) = \langle \hat{\mu}_{T_k}, \varphi_f(\cdot, \hat{u}^2(T_k, \cdot)) \rangle$
  end for
end for

Remark. *The initialization value for v at the boundary, that we have fixed in 0, is arbitrary, given that the first steps in both algorithms does not use this value.*

*However, if the algorithm is used under **(SB)** and the values of $D_x u$ can be easily calculated on the boundary, we have a natural initialization value for v. In this case, we may initialize the backward algorithms of order one and two to reflect this additional information, by setting $v(t,x) = (\mathcal{V}^\mu u(t,x))^T$.*

This modification will have no effect at all for the first order scheme, and is interesting only from the point of view of consistence. On the other hand, on the second order algorithm, the natural initialization of v allows to skip the first order step. This change does not affect the overall rate of convergence of the algorithm but will induce a reduction in the error constant whence of the total approximation error.

It is worth noticing that the given algorithm is particularly effective for treating the McKean dependence of the backward component. Indeed, note that the expectation of any regular enough function of $\hat{u}$ is readily available given that the support of the approximating measure $\hat{\mu}$ coincides with the points where $\hat{u}$ is available. Of course the situation is quite different when a different approach, like a particle method, is used.

3 Main Results

In this section, we first give the rate of convergence of our algorithms 1, 2 and 3 when both the coefficients and terminal condition in (1.1) are smooth. This is given in Theorem 3.1 below. Then, we give the rate when the boundary condition is Lipschitz and when the diffusion part of (1.1) is uniformly non-degenerate. This does not really affect the convergence order, provided the subdivision is taken appropriately. The result is summarized in Corollary 3.2 below. Finally, we give the convergence of a version of our algorithm applied to equation (1.7): when the dependence of the coefficients with respect to the law is general. This is given in Corollary 3.3.

In order to make the exposition of our results clear, let us define, for $i = 1, 2$:

$$\mathcal{E}_u^i(k) := \max_{\pi \in \mathcal{S}_\kappa(k)} |u(T_k, \hat{X}^\pi_{T_k}) - \hat{u}^i(T_k, \hat{X}^\pi_{T_k})|; \qquad \mathcal{E}_v^i(k) := \max_{\pi \in \mathcal{S}_\kappa(k)} |v(T_k, \hat{X}^\pi_{T_k}) - \hat{v}^i(T_k, \hat{X}^\pi_{T_k})|, \tag{3.1}$$

with $\hat{u}^1, \hat{u}^2, \hat{v}^1$ and $\hat{v}^2$ as defined by the algorithms 2 and 3 and where u, v are defined in (1.2), (1.6).

Main result in a smooth setting. We have that

Theorem 3.1. *Suppose that assumption **(SB)** holds. Let m be a given cubature order, q a given non-negative integer and $\mathcal{T}(1, q, m)$ the cubature tree defined by Algorithm 1. Then, there exists a positive constant C, depending only on T, q, d, $||\varphi_{0:d}||_{2q+m+2,\infty}$, $||\phi||_{m+2,\infty}$, such that:*

$$\max_{k \in \{0,\dots,N\}} |\langle \mu_{T_k} - \hat{\mu}_{T_k}, \phi \rangle| \leq C \left(\frac{1}{N}\right)^{[(m-1)/2] \wedge q}, \tag{3.2}$$

with $\hat{\mu}$ as defined in Algorithm 1.
Suppose in addition that $q \geq 1$ and $m \geq 3$. Then, there exists a positive constants C_1, depending only on T, q, d, $||\varphi_{0:d}||_{2q+m+2,\infty}$, $||\varphi_f||_{m+2,\infty}$,$||\phi||_{m+3,\infty}$, such that for all $k = 0, \dots, N$:

$$\mathcal{E}_u^1(k) + \Delta_{T_k}^{1/2} \mathcal{E}_v^1(k) \leq C_1 \left(\frac{1}{N}\right), \tag{3.3}$$

Moreover, suppose in addition that $q \geq 2$ and $m \geq 7$. Then, there exists a positive constant C_2, depending only on T, q, d, $||\varphi_{0:d}||_{2q+m+2,\infty}$, $||\varphi_f||_{m+2,\infty}$,$||\phi||_{m+4,\infty}$, such that for all $k = 0, \dots, N$:

$$\mathcal{E}_u^2(k) + \Delta_{T_k}^{1/2} \mathcal{E}_v^2(k) \leq C_2 \left(\frac{1}{N}\right)^2. \tag{3.4}$$

Convergence order for a Lipschitz boundary condition.

Corollary 3.2. *Suppose that assumption **(LB)** holds. Let m be a given cubature order, q a given non-negative integer, γ a non negative real and $\mathcal{T}(\gamma, q, m)$ the cubature tree defined by the algorithm 1. Then, there exists a positive constant C depending only on T, $||\varphi||_{2q+m+2,\infty}$, $||\phi||_{1,\infty}$, such that:*

$$|\langle \mu_T - \hat{\mu}_T, \phi \rangle| \quad \leq \quad C \left(\left(\frac{1}{N} \right)^{(\gamma/2)\wedge q} \vee L(\gamma, m) \right). \tag{3.5}$$

where

$$L(\gamma, m) = \begin{cases} N^{-\gamma/2} & \text{if} \quad \gamma \in (0, m-1) \\ N^{-(m-1)/2} \ln(N) & \text{if} \quad \gamma = m-1 \\ N^{-(m-1)/2} & \text{if} \quad \gamma \in (m-1, +\infty) \end{cases} \tag{3.6}$$

Moreover, if $\gamma > m-1$, the results on the error control of $\hat{u}^1, \hat{v}^1; \hat{u}^2$, and $\hat{v}^2$ respectively given by (3.3) and (3.4) remain valid (with a constant C_2' depending only on T, q, d, $||\varphi_{0:d}||_{2q+m+2,\infty}$, $||\varphi_f||_{m+2,\infty}$,$||\phi||_{1,\infty}$).

Note that, in (3.5), the control holds only at time T although it holds at each step in (3.2): this is because the boundary condition is Lipschitz only so that we have to wait for the smoothing effect to take place.

We emphasize that the result still applies if we let the boundary condition $\phi : (y, w) \in \mathbb{R}^d \times \mathbb{R} \mapsto \phi(y, w)$ depend also on the law of the process $(X_t^x,\ 0 \leq t \leq T)$. For example one can consider

$$\mathbb{E}\left[\phi(X_t^x, \mathbb{E}[\varphi_\phi(X_T^x)])\right]$$

for a given $\varphi_\phi \in C_b^{m+2}$ and where ϕ is Lipschitz in w uniformly in y.

The algorithm can be easily adapted to the case of the particular dependence explored in [BLP09]:

$$V_i(t, y, \mu) = \langle \mu_t, V_i(t, y, \cdot) \rangle,\ i = 0, \cdots, d,$$

and the result of Theorem 3.1 and Corollary 3.2 remain valid. Note that in that case the uniform ellipticity **(LB)** has to be understood for the matrix $[\langle \eta_t, V(t, y, \cdot) \rangle] [\langle \eta_t, V(t, y, \cdot) \rangle]^*$ uniformly in y, t in $\mathbb{R}^d \times \mathbb{R}^+$ and η family of probability measures on $\mathbb{R}^d$.

Results under a more general law dependence.

As mentioned in the introduction, the algorithm may be modified to solve problems in a naturally extended framework. Let us precise the framework of this extension.

Let $\mathcal{F}$ and $\mathcal{F}'$ be two classes of functions, dense in the space of continuous functions that are zero at infinity. Let $d_{\mathcal{F}}$ $d_{\mathcal{F}'}$ be two distances as defined in (1.8). Recall that we suppose the vector fields V_i $0 \leq i \leq d$ that appear in (1.7) to be Lipschitz continuous with respect to $d_{\mathcal{F}}$ and the driver f to be Lipschitz continuous with respect to $d_{\mathcal{F}'}$. Furthermore, let us suppose that there exists a unique solution X_t^x, Y_t^x, Z_t^x to such a system and, as before, denote by u the decoupling function defined as (1.2) given its existence.

Clearly, we need to modify Algorithm 1 in the natural way to be used in this framework, that is, at each discretization time, we plug directly in the coefficients the cubature based law.

In order to retrieve higher orders of convergence, we need to expand the McKean term that appears in the coefficients. In this extended case, we have to be careful when considering the forward algorithm with $q > 1$: indeed, we must give sense to the expansion proposed at the definition of the functions F_i $0 \leq i \leq d$ in Algorithm 1. A good notion may be the one proposed in Section 7 of [Car10]. To avoid further technicalities, we will consider here only the case with $q = 1$, i.e., when no expansion is performed.

With this definitions and observations in mind, we give the main result under a more general law dependence.

Corollary 3.3. *Let μ_T be the marginal law of the forward process in (1.7) at time T. Let $m \geq 3$ be a given cubature order and $\hat{\mu}_T$ be the discrete measure given by the cubature tree $\mathcal{T}(1,1,m)$ defined by the algorithm 1. Then, there exist two positive constants C_1 and C_2, depending only on T, d such that:*

- *If **(SB)** holds and $\mathcal{F}$ (resp. $\mathcal{F}'$) is the class of functions φ in $C_b^\infty(\mathbb{R}^d,\mathbb{R})$ (resp. $C_b^\infty(\mathbb{R}^d\times\mathbb{R},\mathbb{R})$) such that $||\varphi||_{\infty,\infty} \leq 1$, then*

$$d_{\mathcal{F}}(\mu_T,\hat{\mu}_T) + \mathcal{E}_u^1(k) + \Delta_{T_k}^{1/2}\mathcal{E}_v^1(k) \leq C_1 N^{-1}. \tag{3.7}$$

- *If **(LB)** holds and $\mathcal{F}$ (resp. $\mathcal{F}'$) is the class of functions φ in $C_b^1(\mathbb{R}^d,\mathbb{R})$ (resp. $C_b^1(\mathbb{R}^d\times\mathbb{R},\mathbb{R})$) such that $||\varphi||_{1,\infty} \leq 1$, then*

$$d_{\mathcal{F}}(\mu_T,\hat{\mu}_T) + \mathcal{E}_u^1(k) + \Delta_{T_k}^{1/2}\mathcal{E}_v^1(k) \leq C_2 N^{-1/2}. \tag{3.8}$$

We emphasize that, when $\mathcal{F} = \{\varphi \in C_b^1(\mathbb{R}^d,\mathbb{R}),\ s.t.\ ||\varphi||_{1,\infty} \leq 1\}$, thanks to the Monge-Kantorovitch duality theorem, the distance $d_{\mathcal{F}}$ is the so-called Wasserstein-1 distance.

4 Numerical examples

In this section, we illustrate the algorithm behavior by applying it to a toy model for which the exact solution is available.

Consider the $d-$dimensional MKV-FBSDE on the interval $[0,1]$ with dynamics given by

$$\begin{aligned} dX_t &= \mathbb{E}\left[\sin(X_t)\right]dt + dB_t \\ -dY_t &= \left(\frac{\mathbf{1}\cdot\cos(X_t)}{2} + \mathbb{E}\left[(\mathbf{1}\cdot\sin(X_t))\exp(-Y_t^2)\right]\right)dt - Z_t\cdot dB_t, \end{aligned}$$

where $(B_t)_{0\leq t\leq 1}$ is a $d-$dimensional Brownian motion, $\mathbf{1}$ is a $d-$dimensional vector having each entry equal to one and the sin and cos functions are applied entry-wise. Moreover, suppose that $X_0 = \mathbf{0}$. It is easily verified that a solution for the forward variable is $X = B$, and thanks to the uniqueness result this is the unique solution for the forward variable.

With respect to the backward part, take two different boundary conditions corresponding to the two considered set of assumptions **(SB)** and **(LB)**.

(SB): For $x \in \mathbb{R}^d$, we fix $\phi(x) = \mathbf{1}\cdot\cos(x)$. In this case, the solution to the backward part of the problem is

$$u(t,x) = \mathbf{1}\cdot\cos(x);\ \text{and}\ v(t,x) = -\sin(x),$$

which clearly implies $Y_t = \mathbf{1}\cdot\cos(X_t)$ and $Z_t = -\sin(X_t)$.

(LB): We fix the boundary condition to be $\phi(x) := \phi'(d^{-1/2}(\mathbf{1}\cdot x))$ where ϕ' is the triangular function defined for all $y \in \mathbb{R}$ as

$$\phi'(y) = \begin{cases} y+K & \text{if } y \in (-K,0] \\ -y+K & \text{if } y \in (0,K] \\ 0 & \text{otherwise.} \end{cases}$$

In this case, the solution is given by

$$u(t,x) = \mathbb{E}\left[\phi(X_T^x) + \int_t^T \left(\frac{\mathbf{1}\cdot\cos(X_t)}{2} + \mathbb{E}\left[(\mathbf{1}\cdot\sin(X_t))\exp(-Y_t^2)\right]\right)ds\right].$$

Basic properties of the Brownian motion imply that

$$u(t,x) = U(t,d^{-1/2}(\mathbf{1}\cdot x)) + (\mathbf{1}\cdot\cos(x))\left[\exp\left(\frac{t-1}{2}\right) - 1\right]$$

where

$$U(t,y) = \sqrt{\frac{1-t}{2\pi}}\left[\exp\left(\frac{-(K+y)^2}{2(1-t)}\right) + \exp\left(\frac{-(K-y)^2}{2(1-t)}\right) - 2\exp\left(\frac{-y^2}{2(1-t)}\right)\right]$$
$$+ (K+y)\left[F\left(\frac{-y}{\sqrt{1-t}}\right) - F\left(\frac{-K-y}{\sqrt{1-t}}\right)\right] + (K-y)\left[F\left(\frac{K-y}{\sqrt{1-t}}\right) - F\left(\frac{-y}{\sqrt{1-t}}\right)\right]$$

and F is the cumulative distribution function of the standard normal distribution. Evidently $v(t,x) = D_x u(t,x)$, and is defined for $t < 1$.

4.1 Tests in dimension one

Given that the law dependence already increases the dimension of the problem, we start by presenting some results when we fix $d = 1$.

4.1.1 Forward component

To implement the forward variable, we use the cubature formulae of order 3 and 5 presented in [LV04], which have paths support of size $\kappa = 2$ and $\kappa = 3$ respectively. Given the simple structure of the forward variable dynamics and the piecewise linear definition of the cubatures, we are able to solve explicitly the ODEs appearing during the tree construction. Hence there is no need to use any ODE solver.

In our first test we evaluate the weak approximation error of X using the function ϕ as test function. Indeed we plot as error for the **(SB)** case

$$\max_{k=1,\dots,N} |\langle \hat{\mu}_{T_k} - \mu_{T_k}, \cos\rangle|,$$

while for the **(LB)** case, we plot

$$|\langle \hat{\mu}_{T_k} - \mu_{T_k}, \phi\rangle|,$$

where ϕ is the defined triangular function with $K = 0.6$. As was pointed out before, the difference between the kind of error we are observing for each case is justified as a smoothing effect is needed for the approximation to be valid under **(LB)** .

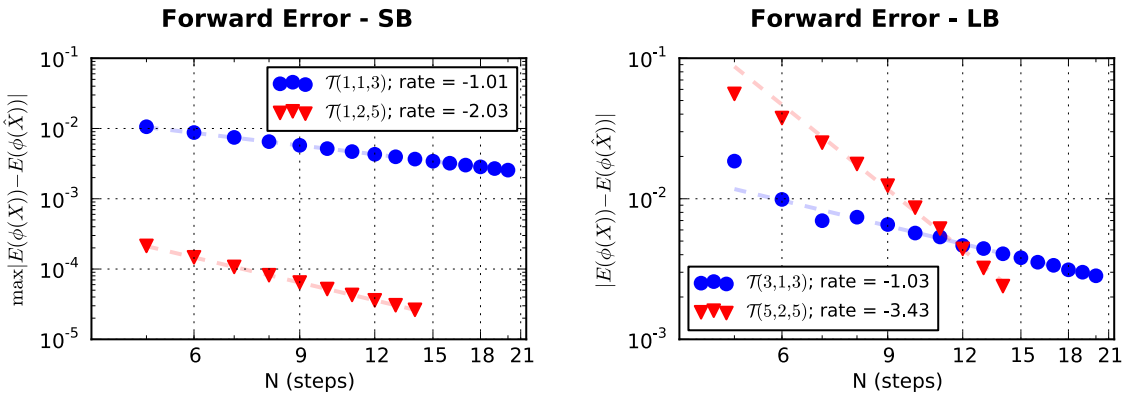

FIGURE 2. Weak approximation of the forward variable: The calculated rates are the slope of a linear regression on the last 8 points.

Figure 2 shows the obtained rate of convergence where we have used the uniform discretization in the **(SB)** case and the discretization with $\gamma = 2$ for the **(LB)** case. With the exception of the rate of convergence for the second order algorithm under **(LB)** (which is actually better than the predicted one), the expected rates of convergence are verified in both cases.

Moreover, under the smooth case, the benefit of using the higher order scheme is not only evident from a quickest convergence, but the error constant itself is smaller. This is an effect that depends on the particular example, but we remark it as it is interesting to notice that a higher order of convergence does not imply necessarily a higher initial constant.

4.1.2 Backward component

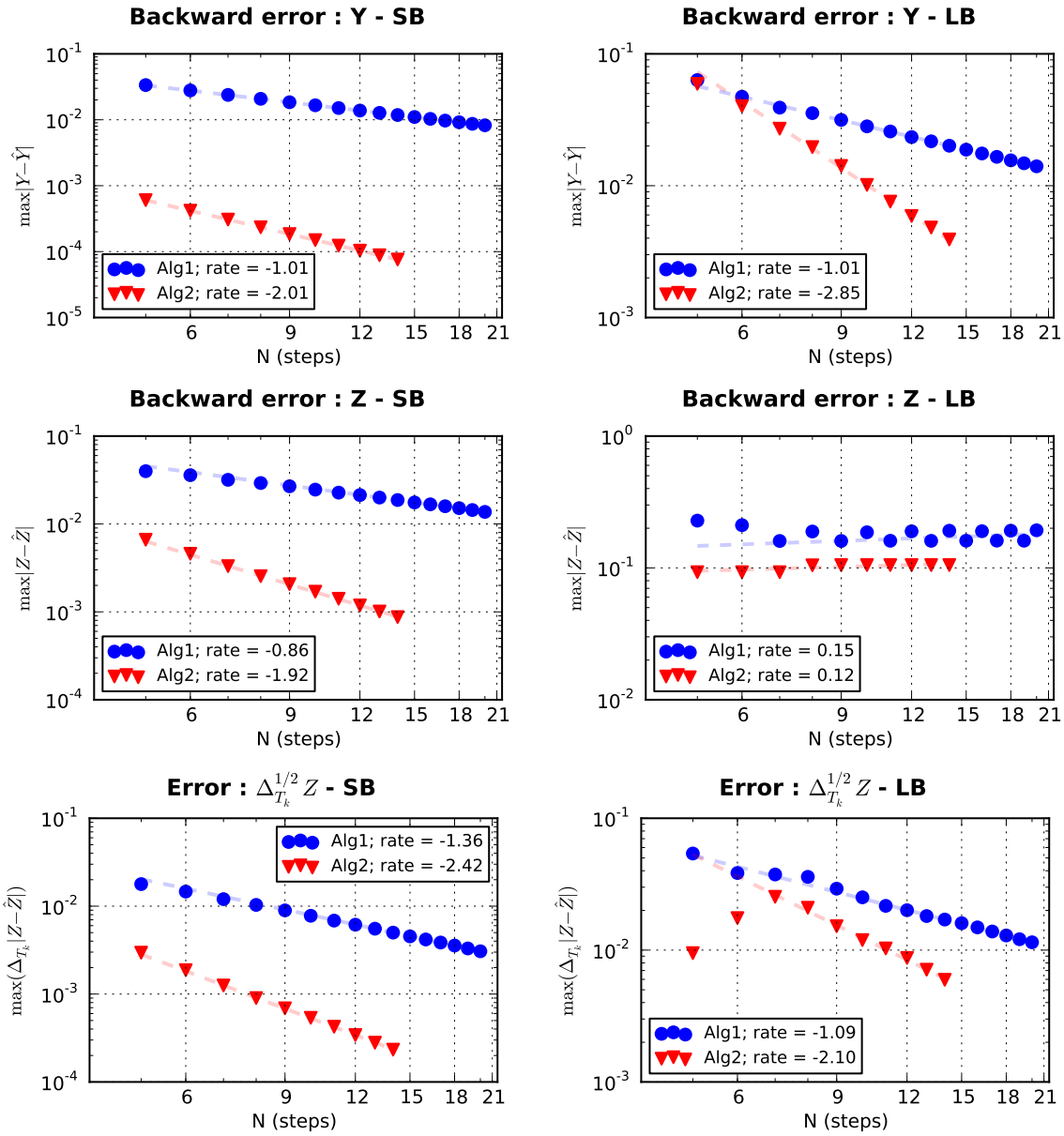

FIGURE 3. Weak approximation of the backward variable: The calculated rates are the slope of a linear regression on the last 8 points.

Let us check now the approximation of the backward variable. We evaluate numerically

$$\max_{0\le k\le N-2;\ \pi\in\mathcal{S}_\kappa(k)} \left|\hat{u}^1(T_k,\hat{X}^\pi_{T_k}) - u(T_k\hat{X}^\pi_{T_k})\right| \quad \text{and} \quad \max_{0\le k\le N-2;\ \pi\in\mathcal{S}_\kappa(k)} \left|\hat{u}^2(T_k,\hat{X}^\pi_{T_k}) - u(T_k\hat{X}^\pi_{T_k})\right|;$$

for both the (SB) and (LB) cases, where we fix $K = 0.6$ for the latter.

The specific structure of our examples allows us to obtain a second order convergence scheme with a cubature of order only 5. Indeed, in such a case, the terms in front of the leading rate of convergence on the cubature error estimate (cf Claim 7.8) are identically 0. Given that the order 5 cubature induces a lower complexity, it is simpler to carry out simulations for a larger number of steps.

As can be appreciated from the two uppermost plots in Figure 3, the expected rates of convergence for both algorithms are verified under the smooth and Lipschitz conditions. Just as we remarked in the forward approximation, solving the backward variable in the smooth case with the higher order scheme has the double benefit of better rate of convergence and smaller constant. As one would expect, due to the use of higher order derivatives, this is no longer true for the Lipschitz case.

It is interesting to look at the behavior of the other backward variable, Z. We look first at an error of the type

$$\max_{0\leq k\leq N;\ \pi\in\mathcal{S}_\kappa(k)} \left|\hat{v}^1(T_k,\hat{X}^\pi_{T_k}) - v(T_k\hat{X}^\pi_{T_k})\right| \quad \text{and} \quad \max_{0\leq k\leq N;\ \pi\in\mathcal{S}_\kappa(k)} \left|\hat{v}^2(T_k,\hat{X}^\pi_{T_k}) - v(T_k\hat{X}^\pi_{T_k})\right|.$$

The two plots in the middle of Figure 3 are concerned with these errors. Although nice convergence is obtained in the smooth case, this is no longer true for the **(LB)** case, where the error stagnates. As will be clear from the analysis, this is a consequence of the singularity appearing at the boundary on the control of derivatives in this case. Hence, a more adequate error analysis considers errors given by

$$\max_{0\leq k\leq N;\ \pi\in\mathcal{S}_\kappa(k)} \Delta_{T_k}^{1/2}\left|\hat{v}^1(T_k,\hat{X}^\pi_{T_k}) - v(T_k\hat{X}^\pi_{T_k})\right| \quad \text{and} \quad \max_{0\leq k\leq N;\ \pi\in\mathcal{S}_\kappa(k)} \Delta_{T_k}^{1/2}\left|\hat{v}^2(T_k,\hat{X}^\pi_{T_k}) - v(T_k\hat{X}^\pi_{T_k})\right|.$$

The expected rate of convergence of this type of error is, respectively for $\hat{v}^1$ and $\hat{v}^2$, of the same order of the order of the error of $\hat{u}^1, \hat{u}^2$ with respect to u. As shown in the bottommost plots in Figure 3, the numerical tests for the **(SB)** and **(LB)** cases reflect the expected rates.

4.2 Tests in higher dimensions

We evaluate as well the algorithm using our test models **(SB)** , **(LB)** with dimensions $d = 2$ and $d = 4$. For these tests, we evaluate only the first order schemes and use the 3-cubature formulae presented in [GL11] which have supports of size $\kappa = 4$ and $\kappa = 6$ respectively.

Figure 4 shows that, just as is in the one-dimensional case, the announced rates of convergence for the forward and backward variables are verified. Note that for the particular chosen examples, the error value changes just slightly with dimension.

The case of dimensions 2 and 4 show one of the current limitations of the method: its complexity grows, in general, exponentially both in terms of the number of iterations and the dimension of the problem. Indeed, considering once again the 3-cubature formula, we have that in general the number of nodes in the tree is

$$\sharp(nodes) = \frac{(2d)^{n-1}-1}{2d-1},$$

with the obvious effects on memory management and execution time. We remark that for some particular cases, the complexity can be radically lower. For instance, under the case of constant drift and diffusion coefficients and smooth boundary conditions, using symmetric cubature formulae (as we did here) leads to a kind of "auto-pruning" of the cubature tree leading to complexity grow of the form

$$\sharp(nodes) = \sum_{i=1}^{n} i^d,$$

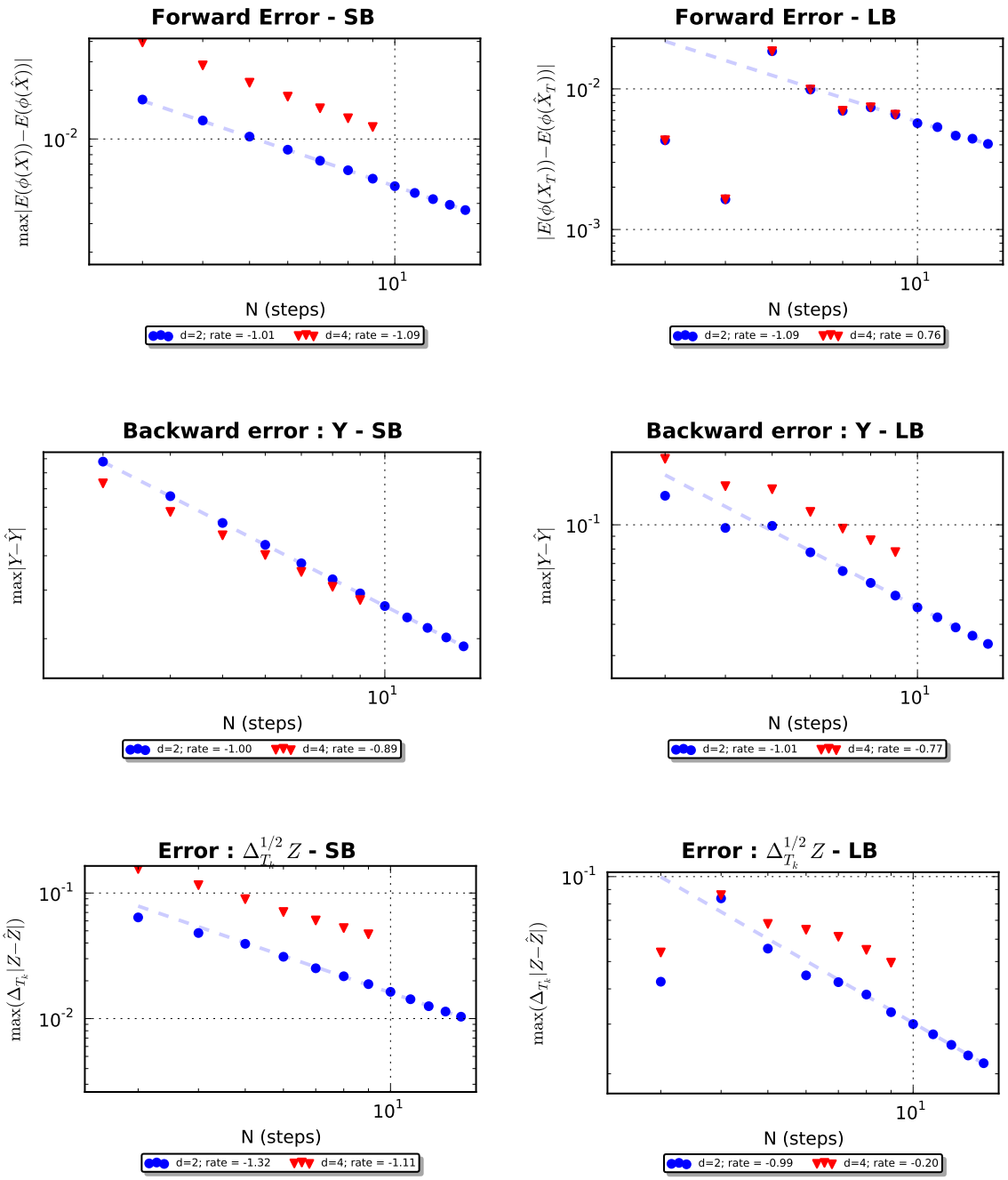

FIGURE 4. Results in dimension 2 and 4.

i.e. polynomial in n with the order of the polynomial depending on the dimension d.

5 A class of control problems in a mean field environment

In this section, we show that equation (1.1) appears when solving a class of control problems inspired from the theory of mean field games but designed in such a way that the dynamics of the controlled process have no influence on the mean field environment.

For the sake of illustration, consider for instance the problem of optimization of an issuer having a large portfolio of credit assets inspired in the framework presented in [BHH+11]. One of the methods used to model credit asset dynamics is the so called structural model (see [BJPR09] for a review on credit risk models). Under this model, we assume that a credit default is triggered when the value of the corresponding credit asset is below a certain threshold. In the original Merton setup, the default may only be triggered at a certain fixed maturity time T. In a rather more realistic view, the default is triggered the first time the credit asset is below the threshold.

We assume that the credit assets in the basket are small and homogeneous (for example, we suppose they belong to the same economic sector) so that their value is modeled by SDEs with the same volatility and drift function terms. To simplify, we will consider the simpler Merton model. Moreover, in order to account for sector-wise contagion effects, we suppose there is a mean field type dependence in the dynamics. In addition to the credit assets, we suppose the issuer has a market portfolio used by the issuer to backup the credit risk, for example to comply with credit risk regulations, or to provide liquidity to its credit branch. Then, the value of the position of the issuer position is modeled by an SDE with coefficients depending on the contribution of all credit assets. The objective of the control problem is to maximize the value of the issuer position.

We will formalize mathematically a generalized version of the presented example. For this, we introduce a system in which a *marked* particle (the issuer in our example) with a controlled state variable $\Xi(\alpha, M)$ is immersed in an *environment* of M interacting particles (the credit assets in our example) with state variables $\xi^1, \dots, \xi^M$, and which dynamics are given by

$$\begin{cases} d\xi_t^1 = b(t, \xi_t^1, \frac{1}{M}\sum_{i=1}^M \delta_{\xi_t^i})dt + \sigma(t, \xi_t^1, \frac{1}{M}\sum_{i=1}^M \delta_{\xi_t^i})dB_t^1 \\ \quad \vdots \qquad\qquad\qquad \vdots \\ d\xi_t^M = b(t, \xi_t^M, \frac{1}{M}\sum_{i=1}^M \delta_{\xi_t^i})dt + \sigma(t, \xi_t^M, \frac{1}{M}\sum_{i=1}^M \delta_{\xi_t^i})dB_t^M \\ d\Xi_t(\alpha, M) = b^0(t, \xi_t^1, \Xi_t(\alpha, M), \frac{1}{M}\sum_{i=1}^M \delta_{\xi_t^i}, \mu_t, \alpha_t)dt + \sigma^0(t, \xi_t^1, \Xi_t(\alpha, M), \frac{1}{M}\sum_{i=1}^M \delta_{\xi_t^i}, \mu_t)dW_t \\ \xi_0^1 = \dots \xi_0^M = \xi_0, \quad \Xi_0(\alpha, M) = \Xi_0 \end{cases} \tag{5.1}$$

where ξ_0, Ξ_0 belong to $\mathbb{R}^d$, $(\alpha_t,\ t \geq 0)$ is a progressively measurable process with image in $A \subset \mathbb{R}$, $B_t^1, \dots, B_t^M$ are M independent Brownian motions and $(W_t,\ t \geq 0)$ is a d-dimensional Brownian motion possibly correlated with $(B_t^1, \dots, B_t^M)^T$. Note that, in this framework, the marked player does not influence the dynamics of the other players. For a large number of environment players, the system is approximated by the McKean-Vlasov system

$$\begin{cases} d\xi_t = b(t, \xi_t, \mu_t)dt + \sigma(t, \xi_t, \mu_t)dB_t \\ d\bar{\Xi}_t(\alpha) = b^0(t, \xi_t, \bar{\Xi}_t(\alpha), \mu_t, \alpha_t)dt + \sigma^0(t, \xi_t, \bar{\Xi}_t(\alpha), \mu_t)dW_t \\ \xi_0 = \xi, \quad \bar{\Xi}_0(\alpha) = \Xi_0 \end{cases} \tag{5.2}$$

where μ_t is the law of ξ_t that we will assume in the following to be fixed. Assume that the marked player is interested in minimizing the cost functional

$$J(t, \xi_0, \Xi_0, \alpha) = \mathbb{E}\left[g\left(\xi_T^{t,\xi_0}, \bar{\Xi}_T^{t,\Xi_0}(\alpha), \mu_T\right) + \int_0^T f(s, \xi_s^{t,\xi_0}, \bar{\Xi}_s^{t,\Xi_0}(\alpha), \mu_s)ds\right],$$

for $\alpha \in \mathcal{A}$, the set of all progressively measurable process $\alpha = (\alpha_t,\ t \geq 0)$ valued in A (the maximization case is available up to a change of sign). Suppose that we want to compute the optimal value function $u(t, \xi_0, \Xi_0) = \inf\{J(t, \xi_0, \Xi_0, \alpha),\ \alpha \in \mathcal{A}\}$. Then, we know that under appropriate assumptions, u can be obtained as the solution of the following Hamilton Jacobi Bellman equation on $[0, T] \times \mathbb{R} \times \mathbb{R}^d$

$$0 = \partial_t u(t, x, \bar{x}) + \frac{1}{2}\mathrm{Tr}(\bar{a} D_{x,\bar{x}}^2 u(t, x, \bar{x})) + b(t, x, \mu_t) D_x u(t, x, \bar{x}) + H(t, x, \bar{x}, D_x u, \mu_t). \tag{5.3}$$

Here H is the Hamiltonian

$$H(t, x, \bar{x}, z, \mu_t) = \inf_{\alpha \in \mathcal{A}} \left[b^0(t, x, \bar{x}, \mu_t, \alpha)z + f(t, x, \bar{x}, \mu_t)\right]$$

and

$$\bar{a} = \begin{bmatrix} \sigma\sigma^T & \sigma\rho(\sigma^0)^T \\ \sigma^0\rho^T\sigma^T & \sigma^0(\sigma^0)^T \end{bmatrix}, \qquad \rho = [B, W],$$

where $[\cdot,\cdot]$ stands for the quadratic variation.

We will not discuss the resolvability of the HJB equation (see e.g. [FS06] or [Pha09] for a partial review). Given the existence of an optimal control, we can interpret (5.3) from a probabilistic point of view: we have that $u(t,x,\bar{x}) = Y_t^{t,x,\bar{x}}$ where $Y^{t,x,\bar{x}}$ is given by the MKV-FBSDE

$$\begin{cases} dX_s^{t,x,\bar{x}} = b(s, X_s^{t,x,\bar{x}}, \mu_s)ds + \sigma(s, X_s^{t,x,\bar{x}}, \mu_s)dB_s \\ d\bar{X}_s^{t,x,\bar{x}} = \sigma^0(s, X_s^{t,x,\bar{x}}, \bar{X}_s^{t,x,\bar{x}}, \mu_s)dW_s \\ -dY_s^{t,x,\bar{x}} = H(s, X_s^{t,x,\bar{x}}, \bar{X}_s^{t,x,\bar{x}}, \bar{Z}_s^{t,x,\bar{x}}, \mu_s) - \bar{Z}_s^{t,x,\bar{x}}dW_s + Z^{t,x,\bar{x}}dB_s \\ X_t^{t,x,\bar{x}} = x, \quad \bar{X}_t^{t,x,\bar{x}} = \bar{x}, \quad Y_0^{t,x,\bar{x}} = g(X_T^{t,x,\bar{x}}, \bar{X}_T^{t,x,\bar{x}}, \mu_T). \end{cases} \tag{5.4}$$

The reader may object that the Hamiltonian H does not satisfy the boundedness condition we have assumed for the analysis of the algorithm (bounded with bounded derivatives w.r.t. the variable z). However, some relatively mild assumptions guarantee that the first derivative term $\bar{Z}$ will be bounded. This is almost direct when the boundary condition g is bounded and smooth and proved in [CD12c] when g is Lipschitz and the diffusion matrix uniformly elliptic. Hence, given an estimate on this quantity, one may introduce a modified system in which we replace in the function $(Z, \bar{Z})$ by $(\psi(Z), \bar{\psi}(\bar{Z}))$, where $\psi, \bar{\psi}$ are truncation functions used to make the value of $Z, \bar{Z}$ satisfy its known estimates, as in [Ric11] (if the estimate is not explicitly known, a sequence of functions approximating the identity may be used as in [IDR10], but some additional work would be needed to account for the truncation error). In both cases, the truncated problem will then satisfy the needed assumptions and may be solved with the presented Algorithm 1, 2.

Remark. Let us just comment some structural assumptions on our class of exemples. We first emphasize that we do not allow the marked player in (5.1) to influence the dynamic of the other player : this allows to make the forward component in (5.4) independant on the law of the backward component. Secondly, the diffusion part of the marked player is not allowed to be controlled : this is the reason why the position of the backward variable does not appear in the forward component in (5.4). Then, the second component in (5.4) has no drift : this is because the drift part of the second component in (5.2) is included in the Hamiltonian in (5.4). Finally, we emphasize that here we do not solve the stochastic control problem but we only compute the value of the optimal cost.

6 Preliminaries

In the following we set a subdivision $T_0 = 0 < \cdots < T_N = T$ of $[0,T]$.

Artificial dynamics. We denote by $\underline{s}$ the mapping $s \mapsto \underline{s} = T_k$ if $s \in [T_k, T_{k+1})$, $k \in \{0, \cdots, N-1\}$.

For any family of probability measures η^1 and η^2, one denotes by P^{η^1} and $\tilde{P}^{\eta^2}$ the operators such that, for all $t < s$ in $[0,T]$, for all measurable function g from $\mathbb{R}^d$ to $\mathbb{R}$ and for all y in $\mathbb{R}^d$:

$$P_{t,s}^{\eta^1} g(y) = \mathbb{E}[g(X_s^{t,y,\eta^1})] \text{ and } \tilde{P}_{t,s}^{\eta^2} g(y) = \mathbb{E}[g(\tilde{X}_s^{t,y,\eta^2})]$$

and $(\mathcal{L}_s^{\eta^1})_{t\le s\le T}$ and $(\tilde{\mathcal{L}}_s^{\eta^2})_{t\le s\le T}$ their infinitesimal generator, where for all g in $C^2(\mathbb{R}^d, \mathbb{R})$

$$\mathcal{L}_s^{\eta^1} g(y) := V_0(s, y, \langle \eta_s^1, \varphi_0 \rangle) \cdot D_y g(y) + \frac{1}{2} \mathrm{Tr}[VV^T(s, y, \langle \eta_s^1, \varphi \rangle) D_y^2 g(y)] \tag{6.1}$$

and by definition $\tilde{\mathcal{L}}_s^{\eta^2} = \mathcal{L}_{\underline{s}}^{\eta^2}$. Here X^{t,y,η^1} and $\tilde{X}^{t,y,\eta^2}$ are the respective solutions of

$$dX_s^{t,y,\eta^1} = \sum_{i=0}^{d} V_i\left(s, X_s^{t,y,\eta^1}, \langle \eta_s^1, \varphi_i\rangle\right) dB_s^i, \quad X_t^{t,y,\eta^1} = y, \tag{6.2}$$

$$d\tilde{X}_s^{t,y,\eta^2} = \sum_{i=0}^{d} V_i\left(s, \tilde{X}_s^{t,y,\eta^2}, \sum_{p=0}^{q-1}[(t-\underline{t})^p/p!]\langle \eta_{\underline{s}}^2, (\tilde{\mathcal{L}}^{\eta^2})^p\varphi_i\rangle\right) dB_t^i, \quad \tilde{X}_t^{t,y,\eta^2} = y. \tag{6.3}$$

Finally, let us define the operator associated to the cubature measure, $Q^{\hat{\mu}}$, as

$$Q_{t,s}^{\hat{\mu}} g(y) = \mathbb{E}_{\mathbb{Q}_{t,s}}[g(\tilde{X}_s^{t,y,\hat{\mu}})] \tag{6.4}$$

for all $t < s$ in $[0,T]$ for all y in $\mathbb{R}^d$ and for all measurable function g from $\mathbb{R}^d$ to $\mathbb{R}$. Note that for all k in $\{1,\dots,N\}$:

$$Q_{0,T_k}^{\hat{\mu}} g(x) = \langle \hat{\mu}_{T_k}, g\rangle.$$

Multi-index (2). Let $\mathcal{M}$ be defined by (2.1). Let $\beta \in \mathcal{M}$. We define $|\beta| = l$ if $\beta = (\beta_1,\dots,\beta_l)$, $|\beta|_0 := \text{card}\{i : \beta_i = 0\}$ and $\|\beta\| := |\beta| + |\beta|_0$. Naturally $|\emptyset| = |\emptyset|_0 = \|\emptyset\| = 0$. For every $\beta \neq \emptyset$, we set $-\beta := (\beta_2,\dots,\beta_l)$ and $\beta- := (\beta_1,\dots,\beta_{l-1})$. We set β^+ the multi-index obtained by deleting the zero components of β.

We will frequently refer to the set of multi-indices of degree at most l denoted by $\mathcal{A}_l := \{\beta \in \mathcal{M} : \|\beta\| \leq l\}$. We define as well its *frontier set* $\partial\mathcal{A} := \{\beta \in \mathcal{M} \setminus \mathcal{A} : -\beta \in \mathcal{A}\}$. We can easily check that $\partial\mathcal{A}_l \subset \mathcal{A}_{l+2} \setminus \mathcal{A}_l$.

Directional derivatives. For notational convenience, let us define the second order operator

$$\mathcal{V}_{(0)} := \partial_t + \mathcal{L},$$

and for $j = \{1,\dots,d\}$ the operator

$$\mathcal{V}_{(j)} := V_j \frac{\partial}{\partial x_j}.$$

where, as announced in the notation section, we do not mark explicitly the time, space and law dependence. For every $\|\beta\| \leq l$ let us define recursively

$$\mathcal{V}_\beta g := \begin{cases} g & \text{if } |\beta| = 0 \\ V_{\beta_1}\mathcal{V}_{-\beta} g & \text{if } |\beta| > 0, \end{cases} \tag{6.5}$$

provided that $g : [0,T]\times\mathbb{R}^d \mapsto \mathbb{R}$ is smooth enough. Hence, for $n \in \mathbb{N}$ we denote by $\mathcal{D}_b^n$, the space of such functions g for which $\mathcal{V}_\beta g$ exists and is bounded for every $\beta \in \mathcal{A}_n$. For any function g in $\mathcal{D}_b^n$, we set for all $\beta \in \mathcal{A}_n$,

$$D_\beta g := \frac{\partial}{\partial y_{\beta_1}} \cdots \frac{\partial}{\partial y_{\beta_{|\beta|}}} g,$$

where $[\partial/\partial y_0]$ must be understood as $[\partial/\partial t]$.

Iterated integrals. For any multi-index β and adapted process g we define for all $t < s \in [0,T]$ the *multiple Itô integral* $I_\beta^{t,s}[g]$ recursively by

$$I_\beta^{t,s}(g) = \begin{cases} g(\tau) & \text{if } |\beta| = 0 \\ \int_t^s I_{\beta-}^{\rho,r}(g)dr & \text{if } |\beta| > 0 \text{ and } \beta_l = 0 \\ \int_t^s I_{\beta-}^{t,r}(g)dB_r^{\beta_l} & \text{if } |\beta| > 0 \text{ and } \beta_l > 0 \end{cases}$$

We will write $I_\beta^{t,s} := I_\beta^{t,s}(1)$.

The previous notation is very convenient to introduce an Itô-Taylor expansion, that is an analogue of Taylor formula when dealing with Itô processes. The proof follows simply by repeated iteration of Itô's lemma, and may be found (without the law dependence) in [KP92].

Lemma 6.1. *Let $t < s \in [0,T]$ and $y \in \mathbb{R}^d$. Let $n \in \mathbb{N}^*$ and let g in D_b^n Then, for each family of probability measures η on $\mathbb{R}^d$, we have have an* Itô-Taylor expansion of order n, *that is*

$$g(s, X_s^{t,y,\eta}) = g(t,y) + \sum_{\beta \in \mathcal{A}_n} \mathcal{V}_\beta g(t,y) I_\beta^{t,s} + \sum_{\beta \in \partial \mathcal{A}_n} I_\beta^{t,s}[\mathcal{V}_\beta g(.,X_.^{t,y,\eta})]$$

where $(X_s^{t,y,\eta},\ t \le s \le T)$ is the solution of (6.3).

The following lemma is a particular case of a result in [KP92]. It follows from integration by parts formula and expectation properties.

Lemma 6.2. *Let $\beta \in \mathcal{M}$, and let $t_1 < t_2 \in [0,T]$. Then for any bounded and measurable functions g_1 and g_2 in $[t_1,t_2]$ there exists a constant depending only on β and i such that*

$$\mathbb{E}\left[I_\beta^{t_1,t_2}(g_1) I_{(i)}^{t_1,t_2}(g_2) \middle| \mathcal{F}_{t_1}\right] \le \mathbf{1}_{\{(\beta)^+ = i\}} C(\beta,i)(t_2-t_1)^{(||\beta||+1)/2} \sup_{t_1 \le s \le t_2} |g_1(s)| \sup_{t_1 \le s \le t_2} |g_2(s)|,$$

$$\mathbb{E}\left[I_\beta^{t_1,t_2}(g_1) I_{(0,i)}^{t_1,t_2}(g_2) \middle| \mathcal{F}_{t_1}\right] \le \mathbf{1}_{\{(\beta)^+ = i\}} C'(\beta,i)(t_2-t_1)^{(||\beta||+3)/2} \sup_{t_1 \le s \le t_2} |g_1(s)| \sup_{t_1 \le s \le t_2} |g_2(s)|.$$

7 Proof of Theorem 3.1 under (SB)

7.1 Rate of convergence of the forward approximation: proof of (3.2)

Here we prove the approximation order of the forward component. Let η be a given family of probability measures on $\mathbb{R}^d$. We have the following decomposition of the error:

$$\begin{aligned}
(P_{T_0,T_N} - Q^{\hat\mu}_{T_0,T_N})\phi(x) &= (P_{T_0,T_N} - P^\eta_{T_0,T_N})\phi(x) + Q^{\hat\mu}_{T_0,T_{N-1}}(P^\eta_{T_{N-1},T_N} - Q^{\hat\mu}_{T_{N-1},T_N})\phi(x) \\
&\quad + Q^{\hat\mu}_{T_0,T_{N-2}} P^\eta_{T_{N-2},T_N}\phi(x) - Q^{\hat\mu}_{T_0,T_{N-1}} P^\eta_{T_{N-1},T_N}\phi(x) \\
&\quad + Q^{\hat\mu}_{T_0,T_{N-3}} P^\eta_{T_{N-3},T_N}\phi(x) - Q^{\hat\mu}_{T_0,T_{N-2}} P^\eta_{T_{N-2},T_N}\phi(x) \\
&\quad \vdots \\
&\quad + P^\eta_{T_0,T_N}\phi(x) - Q^{\hat\mu}_{T_0,T_1} P^\eta_{T_1,T_N}\phi(x) \\
&= (P_{T_0,T_N} - P^\eta_{T_0,T_N})\phi(x) + (P^\eta_{T_0,T_1} - Q^{\hat\mu}_{T_0,T_1}) P^\eta_{T_1,T_N}\phi(x) \\
&\quad + \sum_{j=1}^{N-1} Q^{\hat\mu}_{T_0,T_j}\left[(P^\eta_{T_j,T_{j+1}} - Q^{\hat\mu}_{T_j,T_{j+1}}) P^\eta_{T_{j+1},T_N}\phi(x)\right],
\end{aligned}$$

so that:

$$\begin{aligned}
(P_{T_0,T_N} - Q^{\hat\mu}_{T_0,T_N})\phi(x) &= (P_{T_0,T_N} - P^\eta_{T_0,T_N})\phi(x) + (P^\eta_{T_0,T_1} - Q^{\hat\mu}_{T_0,T_1}) P^\eta_{T_1,T_N}\phi(x) \\
&\quad + \sum_{j=1}^{N-2} Q^{\hat\mu}_{T_0,T_j}\left[(P^\eta_{T_j,T_{j+1}} - Q^{\hat\mu}_{T_j,T_{j+1}}) P^\eta_{T_{j+1},T_N}\phi(x)\right] \\
&\quad + Q^{\hat\mu}_{T_0,T_{N-1}}\left[(P^\eta_{T_{N-1},T_N} - Q^{\hat\mu}_{T_{N-1},T_N})\phi(x)\right]
\end{aligned}$$

That is, the global error is decomposed as a sum of local errors (i.e., as the sum of errors on each interval). These local errors can be also split. Let us define the function

$$\psi(T_k,x) := P^\eta_{T_k,T_N}\phi(x). \tag{7.1}$$

We emphasize that for all t in $[0,T)$, $y \mapsto \psi(t,y)$ is C_b^∞. Indeed, this function can be seen as the solution of the PDE

$$\begin{cases} \partial_t \psi(t,y) + \mathcal{L}^\eta \psi(t,y) = 0, \text{ on } [0,T]\times\mathbb{R}^d \\ \psi(T,y) = \phi(y) \end{cases} \tag{7.2}$$

taken at time T_k, where $\mathcal{L}^\eta$ is defined by (6.1). The claim follows from Lemma 8.1. On each interval Δ_{T_k}, $k = 1,\cdots,N-1$, the local error $(P^\eta_{T_k,T_{k+1}} - Q^{\hat\mu}_{T_k,T_{k+1}})\psi(T_{k+1},x)$ is expressed as

$$\begin{aligned} (P^\eta_{T_k,T_{k+1}} - Q^{\hat\mu}_{T_k,T_{k+1}})\psi(T_{k+1},x) &= (P^\eta_{T_k,T_{k+1}} - \tilde P^{\hat\mu}_{T_k,T_{k+1}})\psi(T_{k+1},x) &(7.3)\\ &\quad + (\tilde P^{\hat\mu}_{T_k,T_{k+1}} - Q^{\hat\mu}_{T_k,T_{k+1}})\psi(T_{k+1},x). &(7.4)\end{aligned}$$

Error (7.3) can be identified as a frozen (in time) error (and so, a sort of weak Euler error) plus an approximation error, in the sense that in step k, the measure μ_{T_k} is approximated by the discrete law $\hat\mu_{T_k}$. Then, (7.4) is a (purely) cubature error on one step, and we have:

$$\begin{aligned} &(P_{T_0,T_N} - Q^{\hat\mu}_{T_0,T_N})\phi(x)\\ &= (P_{T_0,T_N} - P^\eta_{T_0,T_N})\phi(x)\\ &\quad + (P^\eta_{T_0,T_1} - \tilde P^{\hat\mu}_{T_0,T_1})\psi(T_1,x) + \sum_{j=1}^{N-2} Q^{\hat\mu}_{T_0,T_j}\left[(P^\eta_{T_j,T_{j+1}} - \tilde P^{\hat\mu}_{T_j,T_{j+1}})\psi(T_{j+1},x)\right]\\ &\quad + (\tilde P^{\hat\mu}_{T_0,T_1} - Q^{\hat\mu}_{T_0,T_1})\psi(T_1,x) + \sum_{j=1}^{N-2} Q^{\hat\mu}_{T_0,T_j}\left[(\tilde P^{\hat\mu}_{T_j,T_{j+1}} - Q^{\hat\mu}_{T_j,T_{j+1}})\psi(T_{j+1},x)\right]\\ &\quad + Q^{\hat\mu}_{T_0,T_{N-1}}\left[(P^\eta_{T_{N-1},T_N} - Q^{\hat\mu}_{T_{N-1},T_N})\phi(x)\right] \end{aligned}$$

We have the two following Claims:

Claim 7.1. *There exists a positive constant $C(T,V_{0:d})$ depending on the regularity of the $V_{0:d}$ and on T such that, for all y in $\mathbb{R}^d$, for all family of probability measures η, for all $k \in \{1,\cdots,N-1\}$, one has:*

$$\begin{aligned} &\left|(P^\eta_{T_k,T_{k+1}} - \tilde P^{\hat\mu}_{T_k,T_{k+1}})\psi(T_{k+1},y)\right|\\ &\le C(T,V_{0:d})||\psi(T_{k+1},.)||_{2,\infty}\sum_{i=0}^{d}\int_{T_k}^{T_{k+1}}\left|\langle\eta_t,\varphi_i\rangle - \sum_{p=0}^{q-1}[(t-T_k)^p/p!]\langle\eta_{T_k},(\mathcal{L}^\eta)^p\varphi_i\rangle\right| dt\\ &\quad + C(T,V_{0:d})||\psi(T_{k+1},.)||_{2,\infty}\sum_{i=0}^{d}\sum_{p=0}^{q-1}[\Delta^p_{T_{k+1}}/p!]\left|\langle\eta_{T_k},(\mathcal{L}^\eta)^p\varphi_i\rangle - \langle\hat\mu_{T_k},(\mathcal{L}^{\hat\mu})^p\varphi_i\rangle\right|. \end{aligned}$$

Proof. We deduce the claim by applying Lemma 8.3 to $y \mapsto \psi(T_k,y)$ for each $k \in \{0,\cdots,N-2\}$. □

Claim 7.2. *There exists a positive constant $C(T,V_{0:d},d,m)$ depending on the regularity of the $V_{0:d}$ on the dimension d and the cubature order m such that: for all y in $\mathbb{R}^d$, for all family of probability measure η, for all $k \in \{0,\cdots,N-1\}$, one has:*

$$\left|(\tilde P^{\hat\mu}_{T_k,T_{k+1}} - Q^{\hat\mu}_{T_k,T_{k+1}})\psi(T_{k+1},y)\right| \le C(T,V_{0:d},d,m)\sum_{l=m+1}^{m+2}||\psi(T_{k+1},\cdot)||_{l,\infty}\Delta^{\frac{l}{2}}_{T_{k+1}}.$$

Proof. The claim follows, by applying Lemma 8.4 with $\hat\mu$ to the function $y \in \mathbb{R}^d \mapsto \psi(T_k,y)$ for each $k \in \{1,\cdots,N-1\}$. □

Then, by plugging estimates of Claims 7.1 and 7.2 in the error expansion we deduce that:

$$\begin{aligned}
&\left|(P_{T_0,T_N} - Q^{\hat{\mu}}_{T_0,T_N})\phi(x)\right| \qquad (7.5)\\
&\leq C(T,V_{0:d})\sum_{j=0}^{N-1}||\psi(T_{j+1},\cdot)||_{2,\infty}\int_{T_j}^{T_{j+1}}\sum_{i=0}^{d}\left|\langle\eta_t,\varphi_i\rangle - \sum_{p=0}^{q-1}[(t-T_k)^p/p!]\langle\eta_{T_j},(\mathcal{L}^\eta)^p\varphi_i\rangle\right|dt\\
&\quad +C(T,V_{0:d})\sum_{j=0}^{N-1}||\psi(T_{j+1},\cdot)||_{2,\infty}\sum_{p=0}^{q-1}\Delta_{T_{j+1}}^{p+1}\sum_{i=0}^{d}\left|\langle\eta_{T_j},(\mathcal{L}^\eta)^p\varphi_i\rangle - \langle\hat{\mu}_{T_j},(\mathcal{L}^{\hat{\mu}})^p\varphi_i\rangle\right|\\
&\quad +C(T,V_{0:d},d,m)\sum_{j=0}^{N-1}\sum_{l=m+1}^{m+2}||\psi(T_{j+1},\cdot)||_{l,\infty}\Delta_{T_{j+1}}^{\frac{l}{2}} + \left|(P_{T_0,T_N} - P^{\eta}_{T_0,T_N})\phi(x)\right|
\end{aligned}$$

Up to now, the analysis holds for any family of probability measures η. *The key point in the proof is to note that we can actually choose* $\eta = \mu$, that is, the law of the solution of the forward component in (1.1). In that case for all measurable function g:

$$\langle\eta_\cdot, g\rangle = \langle\mu_\cdot, g\rangle = \mathbb{E}[g(X^x_\cdot)].$$

Then, (7.5) becomes:

$$\begin{aligned}
&\left|(P_{T_0,T_N} - Q^{\hat{\mu}}_{T_0,T_N})\phi(x)\right| \qquad (7.6)\\
&\leq C(T,V_{0:d})\sum_{j=0}^{N-1}||\psi(T_{j+1},\cdot)||_{2,\infty}\int_{T_j}^{T_{j+1}}\sum_{i=0}^{d}\left|\mathbb{E}\left[\varphi_i(X_t^{x,\mu})\right] - \sum_{p=0}^{q-1}[(t-T_k)^p/p!]\mathbb{E}\left[(\mathcal{L}^\mu)^p\varphi_i(X_{T_j}^{x,\mu})\right]\right|dt\\
&\quad +C(T,V_{0:d})\sum_{j=0}^{N-1}||\psi(T_{j+1},\cdot)||_{2,\infty}\sum_{p=0}^{q-1}\Delta_{T_{j+1}}^{p+1}\sum_{i=0}^{d}\left|\mathbb{E}\left[(\mathcal{L}^\mu)^p\varphi_i(X_{T_j}^{x,\mu})\right] - \mathbb{E}\left[(\mathcal{L}^{\hat{\mu}})^p\varphi_i(\hat{X}_{T_j}^{x,\hat{\mu}})\right]\right|\\
&\quad +C(T,V_{0:d},d,m)\sum_{j=0}^{N-1}\sum_{l=m+1}^{m+2}||\psi(T_{j+1},\cdot)||_{l,\infty}\Delta_{T_{j+1}}^{\frac{l}{2}}
\end{aligned}$$

since $P_{s,t} = P^{\mu}_{s,t}$ for all $s < t \in [0,T]$, by definition. Now we have:

Claim 7.3. *For any $k \in \{0,\cdots,N-1\}$, and for all t in $[T_k;T_{k+1})$ there exists a positive constant $C(d,V_{0:d})$ such that:*

$$\int_{T_k}^{T_{k+1}}\sum_{i=0}^{d}\left|\mathbb{E}\left[\varphi_i(X_t^{x,\mu})\right] - \sum_{p=0}^{q-1}[(t-T_k)^p/p!]\mathbb{E}\left[(\mathcal{L}^\mu)^p\varphi_i(X_{T_k}^{x,\mu})\right]\right|dt \leq C(d,V_{0:d})\|\varphi\|_{2q,\infty}\Delta_{T_{k+1}}^{q+1}$$

Proof. This follows by Itô-Taylor expansion of order q of $\varphi_i(X^{X_{T_k},\mu})$ for each $i = 0,\cdots,d$ and for any k in $\{0,\cdots,N-1\}$. □

Therefore,

$$\begin{aligned}
&\left|(P_{T_0,T_N} - Q^{\hat{\mu}}_{T_0,T_N})\phi(x)\right| \qquad (7.7)\\
&\leq C(T, V_{0:d}) \sum_{j=0}^{N-1} ||\psi(T_{j+1},\cdot)||_{2,\infty} \sum_{p=0}^{q-1} \Delta_{T_{j+1}}^{p+1} \sum_{i=0}^{d} \left|(P_{T_0,T_N} - Q^{\hat{\mu}}_{T_0,T_N})\mathcal{L}^p\varphi_i(x)\right|\\
&\quad + C(T, V_{0:d}, d)\|\varphi\|_{2q,\infty} \sum_{j=0}^{N-1} ||\psi(T_{j+1},\cdot)||_{2,\infty} \Delta_{T_{j+1}}^{q+1}\\
&\quad + C(T, V_{0:d}, d, m) \sum_{j=0}^{N-1} \sum_{l=m+1}^{m+2} ||\psi(T_{j+1},\cdot)||_{l,\infty} \Delta_{T_{j+1}}^{\frac{l}{2}}.
\end{aligned}$$

Thanks to estimate (8.2) in Lemma 8.1, for all n in $\mathbb{N}$, we have the following bound on the supremum norm of the derivatives of ψ up to order n:

$$||\nabla_y^n \psi(t,\cdot)||_\infty \leq C(T, V_{0:d})||\phi||_{n,\infty}. \qquad (7.8)$$

By plugging this bound in (7.7) we get

$$\begin{aligned}
&\left|(P_{T_0,T_N} - Q^{\hat{\mu}}_{T_0,T_N})\phi(x)\right|\\
&\leq C(T, V_{0:d})||\phi||_{2,\infty} \sum_{j=0}^{N-1} \sum_{p=0}^{q-1} \Delta_{T_{j+1}}^{p+1} \sum_{i=0}^{d} \left|(P_{T_0,T_j} - Q^{\hat{\mu}}_{T_0,T_j})\mathcal{L}^p\varphi_i(x)\right|\\
&\quad + C(T, V_{0:d}, d, m)(||\phi||_{2m+2,\infty} + ||\varphi||_{2q,\infty}) \left(\frac{1}{N}\right)^{q\wedge(m-1)/2}.
\end{aligned}$$

It should be remarked that the term on the right hand side is controlled in terms of the approximation error itself, acting on the functions $\mathcal{L}^p\varphi_i,\ i = 1,\cdots,d,\ p = 0,\cdots,q-1$, from step 0 to j for any j in $\{1,\cdots,N-1\}$. To proceed, the argument is the following one: since these bounds hold for all (at least) ϕ smooth enough, one can let $\phi = \varphi_0$, and use the discrete Gronwall Lemma to get the following bound on φ_0:

$$\begin{aligned}
\left|(P_{T_0,T_N} - Q^{\hat{\mu}}_{T_0,T_N})\varphi_0(x)\right| \ &\leq\ C(T, V_{0:d}, m, ||\phi||_{m+2,\infty}, ||\varphi||_{2q,\infty}, ||\varphi_0||_{2q+m,\infty}) \left(\frac{1}{N}\right)^{q\wedge[(m-1)/2]}\\
&\quad \times \Bigg\{ \sum_{j=0}^{N-1} \sum_{p=0}^{q-1} \Delta_{T_{j+1}}^{p+1} \sum_{i=1}^{d} \left|(P_{T_0,T_j} - Q^{\hat{\mu}}_{T_0,T_j})\mathcal{L}^p\varphi_i(x)\right|\\
&\quad + \sum_{j=0}^{N-1} \sum_{p=1}^{q-1} \Delta_{T_{j+1}}^{p+1} \left|(P_{T_0,T_j} - Q^{\hat{\mu}}_{T_0,T_j})\mathcal{L}^p\varphi_0(x)\right| \Bigg\}
\end{aligned}$$

It is clear that, by iterating this argument (i.e., by letting $\phi = \varphi_1$ and then $\phi = \varphi_2$,..., $\phi = \mathcal{L}\varphi_0$, etc...) we obtain:

$$\left|(P_{T_0,T_N} - Q^{\hat{\mu}}_{T_0,T_N})\phi(x)\right| \ \leq\ C(T, V, d, q, m, ||\phi||_{m+2,\infty}, ||\varphi||_{2q+m,\infty}) \left(\frac{1}{N}\right)^{q\wedge[(m-1)/2]}.$$

This concludes the proof (3.2) at time T. From these arguments, we easily deduce that the estimate holds for any T_k, $k = 1,\ldots,N$. □

7.2 Rate of convergence for the backward approximation: proof of (3.3) and (3.4)

Here we prove the approximation order of the backward component. Before presenting the proof, we introduce some notations. Let us define the Brownian counterparts of $\hat{\Theta}_{k+1,k}, \hat{\Theta}_k$ and $\hat{\zeta}_k$ given in

step 9 in Algorithm 2 and steps 12, 18 and 13 in Algorithm 3. For all family of probability measures η we set

$$\Theta_k^{\eta}(y) := (T_k, y, u(T_k,y), v(T_k,y), \langle \eta_{T_k}, \varphi_f[\cdot, u(T_k,\cdot)]\rangle)\,, \tag{7.9}$$
$$\bar{\Theta}_{k+1,k}^{\eta^1,\eta^2}(y) := \left(T_{k+1}, X_{T_{k+1}}^{T_k,y,\eta^1}, u(T_{k+1}, X_{T_{k+1}}^{T_k,y,\eta^1}), v(T_k,y), \langle \eta^2_{T_k}, \varphi_f[\cdot, u(T_k,\cdot)]\rangle\right),$$

and

$$\zeta_k = 4\frac{B_{T_{k+1}} - B_{T_k}}{\Delta_{T_{k+1}}} - 6\frac{\int_{T_k}^{T_{k+1}}(s-T_k)dB_s}{\Delta^2_{T_{k+1}}}.$$

The proof uses extensively the regularity of the function u. From Lemma 8.2, for all $t \in [0,T)$, the function $y \in \mathbb{R}^d \mapsto u(t,y)$ is C_b^∞ with uniform bounds in time. In the elliptic case the same situation holds, although the bounds depend on time and blow up in the boundary. Hence, we keep track of the explicit dependence of each error term on u and its derivatives in such a way that the proof is simplified for the elliptic case.

Moreover, we will expand and bound terms of the form $y \mapsto f(\cdot, y, u(\cdot,y), \mathcal{V}u(\cdot,y), \cdot)$. When differentiating such a term, the bounds involve the product of the derivatives of u with respect to the space variable. Namely, the r^{th} differentiation of f involves a product of at most $r+1$ derivatives of u. To keep track of the order of the derivatives that appear in the bound, we introduce the set of positive integers for which their sum is less than or equal to r: $\mathcal{I}(l,r) = \{I = (I_1,\dots,I_l) \in \{1,\dots,r\}^l : \sum_j I_j \le r\}$, and define the quantity:

$$M_u(r,s) := \sum_{l=1}^{r}\sum_{I\in\mathcal{I}(l,r)}\prod_{j=1}^{l}||u(s,.)||_{I_j,\infty}. \tag{7.10}$$

(1) Proof of the order of convergence for the first order algorithm (Algorithm 2).

Let $k \in \{1,\dots,N-1\}$. We first break the error between u and $\hat{u}^1$ as follows:

$$\begin{aligned}
u(T_k,\hat{X}^\pi_{T_k}) - \hat{u}^1(T_k,\hat{X}^\pi_{T_k}) &= u(T_k,\hat{X}^\pi_{T_k}) - \mathbb{E}\left[u(T_{k+1}, X_{T_{k+1}}^{T_k,\hat{X}^\pi_{T_k},\mu}) + \Delta_{T_{k+1}} f(\bar{\Theta}^{\mu,\mu}_{k+1,k}(\hat{X}^\pi_{T_k}))\right] && (7.11)\\
&\quad + \mathbb{E}\left[u(T_{k+1}, X_{T_{k+1}}^{T_k,\hat{X}^\pi_{T_k},\mu}) + \Delta_{T_{k+1}} f(\bar{\Theta}^{\mu,\mu}_{k+1,k}(\hat{X}^\pi_{T_k}))\right] && (7.12)\\
&\quad - \mathbb{E}_{\mathbb{Q}_{T_k,T_{k+1}}}\left[u(T_{k+1}, X_{T_{k+1}}^{T_k,\hat{X}^\pi_{T_k},\hat{\mu}}) + \Delta_{T_{k+1}} f(\bar{\Theta}^{\hat{\mu},\mu}_{k+1,k}(\hat{X}^\pi_{T_k}))\right]\\
&\quad + \mathbb{E}_{\mathbb{Q}_{T_k,T_{k+1}}}\Big[u(T_{k+1}, X_{T_{k+1}}^{T_k,\hat{X}^\pi_{T_k},\hat{\mu}}) - \hat{u}^1(T_{k+1}, X_{T_{k+1}}^{T_k,\hat{X}^\pi_{T_k},\hat{\mu}}) && (7.13)\\
&\quad + \Delta_{T_{k+1}}\left(f(\bar{\Theta}^{\hat{\mu},\mu}_{k+1,k}(\hat{X}^\pi_{T_k})) - f(\hat{\Theta}^{\pi,1}_{k+1,k})\right)\Big].
\end{aligned}$$

Similarly, we can expand the error between v and $\hat{v}^1$ as:

$$\begin{aligned}
&\Delta_{T_{k+1}}\left[v(T_k,\hat{X}^\pi_{T_k}) - \hat{v}^1(T_k,\hat{X}^\pi_{T_k})\right]\\
&= \Delta_{T_{k+1}} v(T_k,\hat{X}^\pi_{T_k}) - \mathbb{E}\left[u\left(T_{k+1}, X_{T_{k+1}}^{T_k,\hat{X}^\pi_{T_k},\mu}\right)\Delta B_{T_{k+1}}\right] && (7.14)\\
&\quad + \mathbb{E}\left[u\left(T_{k+1}, X_{T_{k+1}}^{T_k,\hat{X}^\pi_{T_k},\mu}\right)\Delta B_{T_{k+1}}\right] - \mathbb{E}_{\mathbb{Q}_{T_k,T_{k+1}}}\left[u\left(T_{k+1}, X_{T_{k+1}}^{T_k,\hat{X}^\pi_{T_k},\hat{\mu}}\right)\Delta B_{T_{k+1}}\right] && (7.15)\\
&\quad + \mathbb{E}_{\mathbb{Q}_{T_k,T_{k+1}}}\left(\left[u\left(T_{k+1}, X_{T_{k+1}}^{T_k,\hat{X}^\pi_{T_k},\hat{\mu}}\right) - \hat{u}^1\left(T_{k+1}, X_{T_{k+1}}^{T_k,\hat{X}^\pi_{T_k},\hat{\mu}}\right)\right]\Delta B_{T_{k+1}}\right). && (7.16)
\end{aligned}$$

Then, at each step, the approximation error on the backward variables can be expanded as: a first term (7.11) and (7.14), corresponding to scheme errors; a second term, (7.12) and (7.15), corresponding

to generalized cubature errors and can be viewed as one step versions of the forward error (3.2) in Theorem 3.1 ; and a third term, (7.13) and (7.16), which are propagation errors.

Let us explain how the proof works. We will bound separately each error: the scheme, cubature and propagation errors. Each bound is summarized in a Claim (respectively Claims 7.4, 7.5 and 7.5 below). Then, we will deduce the dynamics of the error at step k, $\mathcal{E}_u^1(k)$ defined as (3.1) and conclude with a Gronwall argument.

The first claim below gives the bounds on the scheme errors.

Claim 7.4. *There exists a constant C depending on the regularity of $V_{0:d}$ and f (and not on k) such that the scheme errors (7.11) and (7.14) are bounded by:*

$$\left|u(T_k,\hat{X}^\pi_{T_k}) - \mathbb{E}\left[u(T_{k+1},X^{T_k,\hat{X}^\pi_{T_k},\mu}_{T_{k+1}}) + \Delta_{T_{k+1}} f(\bar{\Theta}^{\mu,\mu}_{k+1,k}(\hat{X}^\pi_{T_k}))\right]\right| \leq C \sup_{s\in[T_k,T_{k+1}]} ||u(s,\cdot)||_{4,\infty}\Delta^2_{T_{k+1}}$$

$$\left|\Delta_{T_{k+1}} v(T_k,\hat{X}^\pi_{T_k}) - \mathbb{E}\left[u\left(T_{k+1},X^{T_k,\hat{X}^\pi_{T_k},\mu}_{T_{k+1}}\right)\Delta B_{T_{k+1}}\right]\right| \leq C \sup_{s\in[T_k,T_{k+1}]} ||u(s,\cdot)||_{3,\infty}\Delta^2_{T_{k+1}}$$

Proof. The proof of the first estimate follows from a second order Itô-Taylor expansion. Applying Lemma 6.1 with $n=2$ to u and taking the expectation leads to:

$$\mathbb{E}\left[u(T_{k+1},X^{T_k,y,\mu}_{T_{k+1}})\right] = u(T_k,y) + \Delta_{T_k}\mathcal{V}_{(0)}u(T_k,y) + \sum_{\beta\in\partial\mathcal{A}_2}\mathbb{E}\left(I^{T_k,T_{k+1}}_\beta[\mathcal{V}_\beta u(\cdot,X^{T_k,y,\mu}_\cdot)]\right) \tag{7.17}$$

and applying again Lemma 6.1 with $n=1$ to $\mathcal{V}_{(0)}u$ and taking the expectation gives:

$$\mathbb{E}\left[\mathcal{V}_{(0)}u(T_{k+1},X^{T_k,y,\mu}_{T_{k+1}})\right] = \mathcal{V}_{(0)}u(T_k,y) + \sum_{\beta\in\partial\mathcal{A}_1}\mathbb{E}\left(I^{T_k,T_{k+1}}_\beta[\mathcal{V}_\beta\mathcal{V}_{(0)}u(\cdot,X^{T_k,y,\mu}_\cdot)]\right). \tag{7.18}$$

Now, note that since u is the solution of PDE (1.3) we have $f=\mathcal{V}_{(0)}u$. So that, by combining (7.17), (7.18) and estimate of Lemma 6.2, we obtain

$$\begin{aligned}
&\left|u(T_k,\hat{X}^\pi_{T_k}) - \mathbb{E}\left[u(T_{k+1},X^{T_k,\hat{X}^\pi_{T_k},\mu}_{T_{k+1}}) + \Delta_{T_{k+1}} f(\bar{\Theta}^{\mu,\mu}_{k+1,k}(\hat{X}^\pi_{T_k}))\right]\right| \\
&\leq 2\sum_{\beta\in\partial\mathcal{A}_2}\mathbb{E}\left(I^{T_k,T_{k+1}}_\beta[\mathcal{V}_\beta u(\cdot,X^{T_k,y,\mu}_\cdot)]\right) \\
&\leq C(T,V_{0:d},f)\sup_{s\in[T_k,T_{k+1}]}||u(s,\cdot)||_{4,\infty}\Delta^2_{T_{k+1}}.
\end{aligned}$$

This concludes the proof of the first estimate. The proof of the second estimate is similar. We first apply an Itô-Taylor expansion of Lemma 6.1 with $n=1$ on u. Then, by noticing that $\Delta B_{T_{k+1}} = (I^{T_k,T_{k+1}}_{(1)},\dots,I^{T_k,T_{k+1}}_{(d)})^T$ and by multiplying by $I^{T_k,T_{k+1}}_{(j)}$ the previous expansion of u, and taking the expectation gives, thanks to Itô's Formula

$$\mathbb{E}\left[u(T_{k+1},X^{T_k,y,\mu}_{T_{k+1}})I^{T_k,T_{k+1}}_{(j)}\right] = \mathcal{V}_{(j)}u(T_k,y)\Delta_{T_{k+1}} + \sum_{\beta\in\partial\mathcal{A}_1}\mathbb{E}\left(I^{T_k,T_{k+1}}_\beta[\mathcal{V}_\beta u(\cdot,X^{T_k,y,\mu}_\cdot)]I^{T_k,T_{k+1}}_{(j)}\right),$$

for $j=1,\dots,d$ and where the first term in the right hand side is the bracket between the stochastic integrals. The last term is controlled by using Lemma 6.2. Recalling that the j-th component of the function v is given by $\mathcal{V}_{(j)}u$ and reordering the terms we obtain the second inequality. □

We now turn to bound the cubature like error terms (7.12) and (7.15). This is summarized by:

Claim 7.5. *There exist two constants C, depending on d, q, T, m, and the regularity of $V_{0:d}$ and $\varphi_{0:d}$ (and not on k), and C', depending in addition on the regularity of f, such that:*

$$\begin{aligned}
&\left|\mathbb{E}\left[u(T_{k+1}, X_{T_{k+1}}^{T_k,\hat{X}^\pi_{T_k},\mu}) + \Delta_{T_{k+1}} f(\bar{\Theta}^{\mu,\mu}_{k+1,k}(\hat{X}^\pi_{T_k}))\right] - \mathbb{E}_{\mathbb{Q}_{T_k,T_{k+1}}}\left[u(T_{k+1}, X_{T_{k+1}}^{T_k,\hat{X}^\pi_{T_k},\hat{\mu}}) + \Delta_{T_{k+1}} f(\bar{\Theta}^{\hat{\mu},\mu}_{k+1,k}(\hat{X}^\pi_{T_k}))\right]\right| \\
&\quad \le C\Bigg(||u(T_{k+1},\cdot)||_{2,\infty}\left[\Delta^{q+1}_{T_{k+1}} + \Delta_{T_{k+1}} N^{-[(m-1)\wedge 2q]/2}\right] + ||u(T_{k+1},\cdot)||_{m+1,\infty}\Delta^{(m+1)/2}_{T_{k+1}} \\
&\qquad + ||u(T_{k+1},\cdot)||_{m+2,\infty}\Delta^{(m+2)/2}_{T_{k+1}}\Bigg) + C'\Bigg(M_u(2,T_{k+1})\left[\Delta^{q+1}_{T_{k+1}} + \Delta_{T_{k+1}} N^{-[(m-1)\wedge 2q]/2}\right] \\
&\qquad + M_u(m+1,T_{k+1})\Delta^{(m+1)/2}_{T_{k+1}} + M_u(m+2,T_{k+1})\Delta^{(m+2)/2}_{T_{k+1}}\Bigg) \\
&\left|\mathbb{E}\left[u\left(T_{k+1}, X_{T_{k+1}}^{T_k,\hat{X}^\pi_{T_k},\mu}\right)\Delta B_{T_{k+1}}\right] - \mathbb{E}_{\mathbb{Q}_{T_k,T_{k+1}}}\left[u\left(T_{k+1}, X_{T_{k+1}}^{T_k,\hat{X}^\pi_{T_k},\hat{\mu}}\right)\Delta B_{T_{k+1}}\right]\right| \\
&\quad \le C\Bigg(||u(T_{k+1},\cdot)||_{3,\infty}\left[\Delta^{q+1}_{T_{k+1}} + \Delta^2_{T_{k+1}} N^{-[(m-1)\wedge 2q]/2}\right] + ||u(T_{k+1},\cdot)||_{m,\infty}\Delta^{(m+1)/2}_{T_{k+1}} \\
&\qquad + ||u(T_{k+1},\cdot)||_{m+1,\infty}\Delta^{(m+2)/2}_{T_{k+1}}\Bigg)
\end{aligned}$$

Proof. Note that the r^{th} derivative of the function $y \mapsto f(\cdot, y, u(\cdot, y), \cdot, \cdot)$ is bounded by $C' M_u(r,\cdot)$ defined by (7.10). Then, the proof of the first assertion follows from (8.12) in Lemma 8.5 applied to u and f and the second assertion from (8.13) in Lemma 8.5 applied to u. □

Finally, an estimate on the propagation error (7.13) is given by:

Claim 7.6. *There exists a constant C depending on d, q, T, m, and the regularity of $V_{0:d}$ and $\varphi_{0:d}$ (and not on k) such that:*

$$\begin{aligned}
&\left|\mathbb{E}_{\mathbb{Q}_{T_k,T_{k+1}}}\left[u(T_{k+1}, X_{T_{k+1}}^{T_k,\hat{X}^\pi_{T_k},\hat{\mu}}) - \hat{u}^1(T_{k+1}, X_{T_{k+1}}^{T_k,\hat{X}^\pi_{T_k},\hat{\mu}}) + \Delta_{T_{k+1}}\left(f(\bar{\Theta}^{\hat{\mu},\mu}_{k+1,k}(\hat{X}^\pi_{T_k})) - f(\hat{\Theta}^{\pi,1}_{k+1,k})\right)\right]\right| \\
&\quad \le (1 + C\Delta_{T_{k+1}})\mathcal{E}^1_u(k+1) + C\Bigg(||u(T_{k+1},\cdot)||_{m+2,\infty}\Delta_{T_{k+1}} N^{-[(m-1)/2]\wedge q} \\
&\qquad + ||u(T_{k+1},\cdot)||_{3,\infty}\left[\Delta^2_{T_{k+1}} + \Delta^{q+1}_{T_{k+1}} + \Delta^2_{T_{k+1}} N^{-[(m-1)\wedge 2q]/2}\right] \\
&\qquad + ||u(T_{k+1},\cdot)||_{m+1,\infty}\Delta^{(m+1)/2}_{T_{k+1}} + ||u(T_{k+1},\cdot)||_{m+2,\infty}\Delta^{(m+2)/2}_{T_{k+1}}\Bigg)
\end{aligned}$$

Proof. Let us start by expanding the f term. We get from the mean value theorem that there exist three random variable Ψ_1, Ψ_2, Ψ_3, respectively bounded by $\|\partial_{y'} f\|_\infty$, $\|\partial_z f\|_\infty$ and $\|\partial_w f\|_\infty$ almost surely, depending on each argument of $\hat{\Theta}^{\pi,1}_{k+1,k}$ and $\bar{\Theta}^{\hat{\mu},\mu}_{k+1,k}$ such that:

$$\begin{aligned}
&\Delta_{T_{k+1}}\left(f(\bar{\Theta}^{\hat{\mu},\mu}_{k+1,k}(\hat{X}^\pi_{T_k})) - f(\hat{\Theta}^{\pi,1}_{k+1,k})\right) \\
&= \Delta_{T_{k+1}}\Psi_1\left(u(T_{k+1}, X_{T_{k+1}}^{T_k,\hat{X}^\pi_{T_k},\hat{\mu}}) - \hat{u}^1(T_{k+1}, X_{T_{k+1}}^{T_k,\hat{X}^\pi_{T_k},\hat{\mu}})\right) + \Delta_{T_{k+1}}\Psi_2\left(v(T_k,\hat{X}^\pi_{T_k}) - \hat{v}^1(T_k,\hat{X}^\pi_{T_k})\right) \\
&\quad + \Delta_{T_{k+1}}\Psi_3\left(\langle\mu_{T_{k+1}}, \varphi_f[\cdot, u(T_{k+1},\cdot)]\rangle - \langle\hat{\mu}_{T_{k+1}}, \varphi_f[\cdot, \hat{u}^1(T_{k+1},\cdot)]\rangle\right).
\end{aligned}$$

Now, we can use the error expansion (7.14), (7.15) and (7.16) of $\Delta_{T_{k+1}}\left(v(T_k,\hat{X}^\pi_{T_k})-\hat{v}^1(T_k,\hat{X}^\pi_{T_k})\right)$ together with the second assertion of Claims 7.4 and 7.5 to get

$$\begin{aligned}
&\left|\mathbb{E}_{\mathbb{Q}_{T_k,T_{k+1}}}\left[u(T_{k+1},X^{T_k,\hat{X}^\pi_{T_k},\hat{\mu}}_{T_{k+1}})-\hat{u}^1(T_{k+1},X^{T_k,\hat{X}^\pi_{T_k},\hat{\mu}}_{T_{k+1}})+\Delta_{T_{k+1}}\left(f(\bar{\Theta}^{\hat{\mu},\mu}_{k+1,k}(\hat{X}^\pi_{T_k}))-f(\hat{\Theta}^{\pi,1}_{k+1,k})\right)\right]\right| \\
&\leq\left|\mathbb{E}_{\mathbb{Q}_{T_k,T_{k+1}}}\left[\left(u(T_{k+1},X^{T_k,\hat{X}^\pi_{T_k},\hat{\mu}}_{T_{k+1}})-\hat{u}^1(T_{k+1},X^{T_k,\hat{X}^\pi_{T_k},\hat{\mu}}_{T_{k+1}})\right)(1+\Psi_1\Delta_{T_{k+1}}+\Psi_2\Delta B_{T_{k+1}})\right]\right| \\
&\quad+C||u(T_{k+1},\cdot)||_{3,\infty}\left[\Delta^2_{T_{k+1}}+\Delta^{q+1}_{T_{k+1}}+\Delta^2_{T_{k+1}}N^{-[(m-1)\wedge 2q]/2}\right]+C||u(T_{k+1},\cdot)||_{m+1,\infty}\Delta^{(m+1)/2}_{T_{k+1}} \\
&\quad+C||u(T_{k+1},\cdot)||_{m+2,\infty}\Delta^{(m+2)/2}_{T_{k+1}}+\Delta_{T_{k+1}}\|\partial_w f\|_\infty\left|\langle\mu-\hat{\mu}_{T_{k+1}},\varphi_f[\cdot,u(T_{k+1},\cdot)]\rangle\right| \\
&\quad+\Delta_{T_{k+1}}\|\partial_w f\|_\infty\left|\langle\hat{\mu}_{T_{k+1}},\varphi_f[\cdot,u(T_{k+1},\cdot)]-\varphi_f[\cdot,\hat{u}^1(T_{k+1},\cdot)]\rangle\right|. \qquad (7.19)
\end{aligned}$$

Note that from the forward result (3.2) in Theorem 3.1, we have

$$\Delta_{T_{k+1}}\|\partial_w f\|_\infty\left|\langle\mu-\hat{\mu}_{T_{k+1}},\varphi_f[\cdot,u(T_{k+1},\cdot)]\rangle\right|\leq C(||u(T_{k+1},\cdot)||_{m+2,\infty})\Delta_{T_{k+1}}N^{-((m-1)\wedge 2q)/2}, \qquad (7.20)$$

while using the regularity of φ_f and the definition of $\hat{\mu}$ gives,

$$\Delta_{T_{k+1}}\|\partial_w f\|_\infty\left|\langle\hat{\mu}_{T_{k+1}},\varphi_f[\cdot,u(T_{k+1},\cdot)]-\varphi_f[\cdot,\hat{u}^1(T_{k+1},\cdot)]\rangle\right|\leq C'\Delta_{T_{k+1}}\mathcal{E}^1_u(k+1). \qquad (7.21)$$

The Claim follows by applying the Cauchy-Schwartz inequality on the first term in the right hand side of (7.19) and plugging (7.20) and (7.21) in (7.19). □

We can now analyze the local error at step k. By plugging the estimates from Claims 7.4, 7.5 and 7.6 in the expansion (7.11), (7.12) and (7.13) of $u-\hat{u}$ we obtain that:

$$\mathcal{E}^1_u(k)\leq\left(1+C\Delta_{T_{k+1}}\right)\mathcal{E}^1_u(k+1)+\bar{\epsilon}(k+1), \qquad (7.22)$$

with

$$\begin{aligned}
\bar{\epsilon}(k+1) \;=\; & C\Bigg(\sup_{s\in[T_k;T_{k+1}]}||u(s,\cdot)||_{2,\infty}\Delta^4_{T_{k+1}}+||u(T_{k+1},\cdot)||_{2,\infty}\Delta_{T_{k+1}}N^{-[(m-1)\wedge 2q]/2} \\
&+||u(T_{k+1},\cdot)||_{m+1,\infty}\Delta^{(m+1)/2}_{T_{k+1}}+||u(T_{k+1},\cdot)||_{m+2,\infty}\Delta^{(m+2)/2}_{T_{k+1}} \\
&+||u(T_{k+1},\cdot)||_{3,\infty}\left[\Delta^2_{T_{k+1}}+\Delta^{q+1}_{T_{k+1}}+\Delta^2_{T_{k+1}}N^{-[(m-1)\wedge 2q]/2}\right]\Bigg) \qquad (7.23) \\
&+C'\Bigg(M_u(m+1,T_{k+1})\Delta^{(m+1)/2}_{T_{k+1}}+M_u(m+2,T_{k+1})\Delta^{(m+2)/2}_{T_{k+1}} \\
&+M_u(2,T_{k+1})\left[\Delta^{q+1}_{T_{k+1}}+\Delta_{T_{k+1}}N^{-[(m-1)\wedge 2q]/2}\right]\Bigg).
\end{aligned}$$

Under **(SB)** , we have from Lemma 8.2 that for all $n\in\mathbb{N}^*$ there exists a constant K, depending on the regularity of $V_{0:d}$ and ϕ, such that for all $k\in\{0,\cdots,N-1\}$,

$$M_u(n,T_{k+1})+\sup_{s\in[T_k,T_{k+1}]}||u(s,\cdot)||_{n,\infty}\leq K.$$

Therefore, Gronwall's Lemma applied to (7.22) and the definition of Δ_{T_k} implies

$$\mathcal{E}^1_u\leq CN^{-1}. \qquad (7.24)$$

Moreover, Claims 7.4 and 7.5 and expansion (7.14), (7.15) and (7.16) show that:

$$\begin{aligned}
\Delta_{T_{k+1}}\mathcal{E}^1_v(k+1)\leq C\Bigg(&\sup_{s\in[T_k,T_{k+1}]}||u(s,\cdot)||\Delta^2_{T_{k+1}}+||u(T_{k+1},\cdot)||_{3,\infty}\left[\Delta^{q+1}_{T_{k+1}}+\Delta^2_{T_{k+1}}N^{-[(m-1)\wedge 2q]/2}\right] \\
&+||u(T_{k+1},\cdot)||_{m+1,\infty}\Delta^{(m+1)/2}_{T_{k+1}}+C||u(T_{k+1},\cdot)||_{m+2,\infty}\Delta^{(m+2)/2}_{T_{k+1}}\Bigg)+\Delta^{1/2}_{T_{k+1}}\mathcal{E}^1_u(k+1),
\end{aligned}$$

which together with (7.24) imply

$$\Delta_{T_k}^{1/2}\mathcal{E}_v^1 \leq CN^{-1},$$

and the result holds.

(2) Proof of the order of convergence for the second order algorithm (Algorithm 3).

Let $k \in \{0, \dots, N-2\}$. We can expand as before the errors on the u and v approximations as:

$$\begin{aligned} u(T_k, \hat{X}_{T_k}^\pi) - \hat{u}^2(T_k, \hat{X}_{T_k}^\pi) = \mathbb{E}_{\mathbb{Q}_{T_k,T_{k+1}}}\left[u(T_{k+1}, X_{T_{k+1}}^{T_k,\hat{X}_{T_k}^\pi,\hat{\mu}}) - \hat{u}^2(T_{k+1}, X_{T_{k+1}}^{T_k,\hat{X}_{T_k}^\pi,\hat{\mu}})\right] \\ + \frac{1}{2}\Delta_{T_{k+1}}\mathbb{E}_{\mathbb{Q}_{T_k,T_{k+1}}}\left[f(\Theta_{k+1}^\mu(X_{T_{k+1}}^{T_k,\hat{X}_{T_k}^\pi,\mu})) - f(\hat{\Theta}_{k+1}^{\pi,2})\right] \quad (7.25) \\ + \frac{1}{2}\Delta_{T_{k+1}}[f(\Theta_k^\mu(\hat{X}_{T_k}^\pi)) - f(\tilde{\Theta}_k^\pi)] \quad (7.26) \\ + \epsilon_{\hat{u}^2,k}^s(\pi) + \epsilon_{\hat{u}^2,k}^c(\pi), \end{aligned}$$

where

$$\epsilon_{\hat{u}^2,k}^s(\pi) = u(T_k, \hat{X}_{T_k}^\pi) - \mathbb{E}\left[u(T_{k+1}, X_{T_{k+1}}^{T_k,\hat{X}_{T_k}^\pi,\mu}) + \frac{\Delta_{T_{k+1}}}{2}\left(f(\Theta_k^\mu(\hat{X}_{T_k}^\pi)) + f(\Theta_{k+1}^\mu(X_{T_{k+1}}^{T_k,\hat{X}_{T_k}^\pi,\mu}))\right)\right] \quad (7.27)$$

$$\begin{aligned} \epsilon_{\hat{u}^2,k}^c(\pi) = \mathbb{E}\left[u(T_{k+1}, X_{T_{k+1}}^{T_k,\hat{X}_{T_k}^\pi,\mu})\right] - \mathbb{E}_{\mathbb{Q}_{T_k,T_{k+1}}}\left[u(T_{k+1}, X_{T_{k+1}}^{T_k,\hat{X}_{T_k}^\pi,\hat{\mu}})\right] \\ + \frac{\Delta_{T_{k+1}}}{2}\left(\mathbb{E}\left[f(\Theta_{k+1}^\mu(X_{T_{k+1}}^{T_k,\hat{X}_{T_k}^\pi,\mu}))\right] - \mathbb{E}_{\mathbb{Q}_{T_k,T_{k+1}}}\left[f(\Theta_{k+1}^\mu(X_{T_{k+1}}^{T_k,\hat{X}_{T_k}^\pi,\mu}))\right]\right). \quad (7.28) \end{aligned}$$

Similarly, we have

$$\begin{aligned} v(T_k, \hat{X}_{T_k}^\pi) - \hat{v}^2(T_k, \hat{X}_{T_k}^\pi) = \mathbb{E}_{\mathbb{Q}_{T_k,T_{k+1}}}\left(\left[u\left(T_{k+1}, X_{T_{k+1}}^{T_k,\hat{X}_{T_k}^\pi,\hat{\mu}}\right) - \hat{u}^2\left(T_{k+1}, X_{T_{k+1}}^{T_k,\hat{X}_{T_k}^\pi,\hat{\mu}}\right)\right]\zeta_{k+1}\right) \\ + \mathbb{E}_{\mathbb{Q}_{T_k,T_{k+1}}}\left[\left(f[\Theta_{k+1}^\mu(X_{T_{k+1}}^{T_k,\hat{X}_{T_k}^\pi,\mu})] - f(\hat{\Theta}_{k+1}^{\pi,2})\right)\Delta_{T_{k+1}}\zeta_{k+1}\right] \quad (7.29) \\ + \epsilon_{\hat{v}^2,k}^s(\pi) + \epsilon_{\hat{v}^2,k}^c(\pi), \end{aligned}$$

with

$$\epsilon_{\hat{v}^2,k}^s(\pi) = v(T_k, \hat{X}_{T_k}^\pi) - \mathbb{E}\left(\left[u(T_{k+1}, X_{T_{k+1}}^{T_k,\hat{X}_{T_k}^\pi,\mu}) + \Delta_{T_{k+1}}f[\Theta_{k+1}^\mu(X_{T_{k+1}}^{T_k,\hat{X}_{T_k}^\pi,\mu})]\right]\zeta_{k+1}\right) \quad (7.30)$$

$$\begin{aligned} \epsilon_{\hat{v}^2,k}^c(\pi) = \mathbb{E}\left[u(T_{k+1}, X_{T_{k+1}}^{T_k,\hat{X}_{T_k}^\pi,\mu})\zeta_{k+1}\right] - \mathbb{E}_{\mathbb{Q}_{T_k,T_{k+1}}}\left[u(T_{k+1}, X_{T_{k+1}}^{T_k,\hat{X}_{T_k}^\pi,\hat{\mu}})\zeta_{k+1}\right] \\ + \mathbb{E}\left[f[\Theta_{k+1}^\mu(X_{T_{k+1}}^{T_k,\hat{X}_{T_k}^\pi,\mu})]\Delta_{T_{k+1}}\zeta_{k+1}\right] - \mathbb{E}_{\mathbb{Q}_{T_k,T_{k+1}}}\left[f[\Theta_{k+1}^\mu(X_{T_{k+1}}^{T_k,\hat{X}_{T_k}^\pi,\mu})]\Delta_{T_{k+1}}\zeta_{k+1}\right]. \quad (7.31) \end{aligned}$$

We identify, as for the first order expansion, some error terms corresponding to the scheme error (7.27) and (7.30), generalized cubature errors (7.28) and (7.31) and propagation errors (7.25), (7.29). Some important changes are clear from the expansion: we have in addition a prediction error term (7.26) reflecting the fact that we perform a new intermediate step, and we have some f term in (7.29), adding to the propagation error.

The proof for the second order approximation is then similar to its first order equivalent, but we will have to consider the mentioned additional terms. In particular, the fact that the second order

approximation $\hat{v}^2(\hat{X}^\pi_{T_k})$ includes the term $f(\hat{\Theta}^{\pi,2}_{T_{k+1}})$, adds an additional coupling effect. With this in mind, and in order to simplify the analysis, we introduce the following quantity

$$\mathcal{E}_f(k) = \max_{\pi\in\mathcal{S}(k)} |f(\Theta^\mu_k(\hat{X}^\pi_{T_k})) - f(\hat{\Theta}^{\pi,2}_k)|$$

and we will analyze the dynamics of the sum of errors at step k, $\mathcal{E}^2_u(k) + \Delta_{T_k}\mathcal{E}_f(k)$.

To this aim, we will bound separately the scheme and cubature errors. Each bound is summarized in a Claim (respectively Claims 7.7, 7.8 below). Then, we will conclude with a Gronwall argument.

The scheme error terms (7.27) and (7.30) and the generalized cubature errors (7.28) and (7.31) are treated similarly as in the first order scheme. We show this in Claims 7.7 and 7.8.

Claim 7.7. *There exists a constant C, depending on the regularity of $V_{0:d}$ and f (and not on k) such that:*

$$\begin{aligned}\left|\epsilon^s_{\hat{u}^2,k}(\pi)\right| &\leq C\sup_{s\in[T_k,T_{k+1}]}||u(s,\cdot)||_{6,\infty}\Delta^3_{T_{k+1}}\\ \left|\epsilon^s_{\hat{v}^2,k}(\pi)\right| &\leq C\sup_{s\in[T_k,T_{k+1}]}||u(s,\cdot)||_{5,\infty}\Delta^2_{T_{k+1}}.\end{aligned}$$

where $\epsilon^s_{\hat{u}^2,k}(\pi), \epsilon^s_{\hat{v}^2,k}(\pi)$ are defined in (7.27) and (7.30).

Proof. The proof follows in the same way as the one of Claim 7.4, by performing a Taylor expansion to one additional order. The choice of ζ_{k+1} is the one needed to match the lower order terms (recall that the i^{th} component of ζ_k is expressed as $\zeta^i_k = 4\Delta^{-1}_{T_k} I^{T_k,T_{k+1}}_{(i)} - 6\Delta^{-2}_{T_k} I^{T_k,T_{k+1}}_{(0,i)}$).

Applying Lemma 6.1 with $n=4$ to u and with $n=2$ to $\mathcal{V}_{(0)}u$ implies, after taking the expectation, that

$$\mathbb{E}\left[u(T_{k+1}, X^{T_k,y,\mu}_{T_{k+1}})\right] = u(T_k,y)+\Delta_{T_k}\mathcal{V}_{(0)}u(T_k,y)+\frac{1}{2}\mathcal{V}_{(0,0)}u(T_k,y)\Delta^2_{T_{k+1}}+\sum_{\beta\in\partial\mathcal{A}_4}\mathbb{E}\left(I^{T_k,T_{k+1}}_\beta[\mathcal{V}_\beta u(.,X^{T_k,y,\mu}_.)]\right)$$

and

$$\mathbb{E}\left[\mathcal{V}_{(0)}u(T_{k+1}, X^{T_k,y,\mu}_{T_{k+1}})\right] = \mathcal{V}_{(0)}u(T_k,y) + \mathcal{V}_{(0,0)}u(T_k,y)\Delta_{T_{k+1}} + \sum_{\beta\in\partial\mathcal{A}_2}\mathbb{E}\left(I^{T_k,T_{k+1}}_\beta[\mathcal{V}_{(\beta*0)}u(.,X^{T_k,y,\mu}_.)]\right).$$

Then, the estimate of Lemma 6.2 gives

$$\begin{aligned}&\mathbb{E}\left[u(T_{k+1}, X^{T_k,y,\mu}_{T_{k+1}}) - \frac{\Delta_{T_k}}{2}\left(\mathcal{V}_{(0)}u(T_k,y) + \mathcal{V}_{(0)}u(T_{k+1}, X^{T_k,y,\mu}_{T_{k+1}})\right)\right] - u(T_k,y)\\ &\quad= \sum_{\beta\in\partial\mathcal{A}_4}\mathbb{E}\left(I^{T_k,T_{k+1}}_\beta[\mathcal{V}_{(\beta)}u(.,X^{T_k,y,\mu}_.)]\right) - \frac{\Delta_{T_k}}{2}\sum_{\beta\in\partial\mathcal{A}_2}\mathbb{E}\left(I^{T_k,T_{k+1}}_\beta[\mathcal{V}_{(\beta*0)}u(.,X^{T_k,y,\mu}_.)]\right)\\ &\quad\leq C(T,V_{0:d},f)\Delta^3_{T_{k+1}}\sup_{s\in[T_k,T_{k+1}]}||u(s,.)||_{6,\infty},\end{aligned}$$

from where we deduce the first inequality.

Similarly, by using Lemma 6.1 with $n=3$ on u and with $n=1$ on $\mathcal{V}_{(0)}u$ and taking the expectation, using the fact that $\zeta^i_k = 4\Delta^{-1}_{T_k} I^{T_k,T_{k+1}}_{(i)} - 6\Delta^{-2}_{T_k} I^{T_k,T_{k+1}}_{(0,i)}$, it follows

$$\begin{aligned}&\mathbb{E}\left[\left[u(T_{k+1}, X^{T_k,y,\mu}_{T_{k+1}}) - \Delta_{T_{k+1}}\mathcal{V}_{(0)}u(T_{k+1}, X^{T_k,y,\mu}_{T_{k+1}})\right]\Delta_{T_{k+1}}\zeta^j_k\right]\\ &= \mathbb{E}\Bigg[\Big[\mathcal{V}_{(j)}u(T_k,y)I^{T_k,T_{k+1}}_{(j)} + \mathcal{V}_{(0,j)}u(T_k,y)I^{T_k,T_{k+1}}_{(0,j)} + \mathcal{V}_{(j,0)}u(T_k,y)I^{T_k,T_{k+1}}_{(j,0)}\\ &\qquad - \Delta_{T_{k+1}}\mathcal{V}_{(j,0)}u(T_k,y)I^{T_k,T_{k+1}}_{(j)}\Big]\left(4I^{T_k,T_{k+1}}_{(j)} - 6\frac{I^{T_k,T_{k+1}}_{(0,j)}}{\Delta_{T_{k+1}}}\right)\Bigg] + \mathcal{R}(k,j)\\ &= \Delta_{T_{k+1}}\mathcal{V}_{(j)}u(T_k,y) + \mathcal{R}(k,j)\end{aligned}$$

where

$$\mathcal{R}(k,j) = \Delta_{T_{k+1}} \sum_{\beta\in\partial\mathcal{A}_3} \mathbb{E}\left(I_\beta^{T_k,T_{k+1}}[\mathcal{V}_{(\beta)}u(.,X_.^{T_k,y,\mu})]\zeta_k^j\right) - \Delta_{T_{k+1}}^2 \sum_{\beta\in\partial\mathcal{A}_1} \mathbb{E}\left(I_\beta^{T_k,T_{k+1}}[\mathcal{V}_{(\beta*0)}u(.,X_.^{T_k,y,\mu})]\zeta_k^j\right).$$

Using Lemma 6.2 we bound the residual term $\mathcal{R}(k,j)$ and obtain

$$\begin{aligned}&\Delta_{T_{k+1}} \sum_{\beta\in\partial\mathcal{A}_3} \mathbb{E}\left(I_\beta^{T_k,T_{k+1}}[\mathcal{V}_{(\beta)}u(.,X_.^{T_k,y,\mu})]\zeta_k^j\right) - \Delta_{T_{k+1}}^2 \sum_{\beta\in\partial\mathcal{A}_1} \mathbb{E}\left(I_\beta^{T_k,T_{k+1}}[\mathcal{V}_{(\beta*0)}u(.,X_.^{T_k,y,\mu})]\zeta_k^j\right)\\ &\le \Delta_{T_k}^3 \sup_{s\in[T_k,T_{k+1}]} ||u(s,.)||_{5,\infty} + \Delta_{T_k}^{7/2} \sup_{s\in[T_k,T_{k+1}]} ||u(s,.)||_{5,\infty},\end{aligned}$$

recalling that $v(T_k,y) = \mathcal{V}_{(j)}u(T_k,y)$ we deduce the second inequality. □

Claim 7.8. *There exist two constants C, depending on d, q, T, m, the regularity of $V_{0:d}$ and $\varphi_{0:d}$ (and not on k), and C', depending in addition on the regularity of f, such that:*

$$\begin{aligned}\left|\epsilon^c_{\hat{u}^2,k}(\pi)\right| &\le C\Big(||u(T_{k+1},\cdot)||_{2,\infty}\left[\Delta_{T_{k+1}}^{q+1} + \Delta_{T_{k+1}}N^{-[(m-1)\wedge 2q]/2}\right] + ||u(T_{k+1},\cdot)||_{m+1,\infty}\Delta_{T_{k+1}}^{(m+1)/2}\\ &\quad + ||u(T_{k+1},\cdot)||_{m+2,\infty}\Delta_{T_{k+1}}^{(m+2)/2}\Big)\\ &\quad + C'\Big(M_u(3,T_{k+1})\left[\Delta_{T_{k+1}}^{q+2} + \Delta_{T_{k+1}}^2 N^{-[(m-1)\wedge 2q]/2}\right] + M_u(m+2,T_{k+1})\Delta_{T_{k+1}}^{(m+3)/2}\\ &\quad + M_u(m+3,T_{k+1})\Delta_{T_{k+1}}^{(m+4)/2}\Big),\\ \left|\epsilon^c_{\hat{v}^2,k}(\pi)\right| &\le C\Big(||u(T_{k+1},\cdot)||_{3,\infty}\left[\Delta_{T_{k+1}}^{q} + \Delta_{T_{k+1}}N^{-[(m-1)\wedge 2q]/2}\right] + \sum_{i=m-2}^{m+1}||u(T_{k+1},\cdot)||_{i,\infty}\Delta_{T_{k+1}}^{(i-1)/2}\Big)\\ &\quad + C'\Big(M_u(4,T_{k+1})\left[\Delta_{T_{k+1}}^{q+1} + \Delta_{T_{k+1}}^2 N^{-[(m-1)\wedge 2q]/2}\right] + \sum_{i=m-1}^{m+2} M_u(i,T_{k+1})\Delta_{T_{k+1}}^{i/2}\Big).\end{aligned}$$

where $\epsilon^c_{\hat{u}^2,k}(\pi), \epsilon^c_{\hat{v}^2,k}(\pi)$ are defined in (7.28) and (7.31).

Remark. *Although the rates of convergence have a leading term of order $\Delta_{T_{k+1}}^{(m-1)}$ that is worst than the one in the first order scheme result, (Claim 7.5), here we assume that m is bigger, and thus they are suitable for a second order scheme.*

Proof. Note first that the r^{th} derivative of the function $y \mapsto f(\cdot,y,u(\cdot,y),\mathcal{V}u(\cdot,y),\cdot)$ is bounded by $M_u(r+1,\cdot)$ defined by (7.10). This estimate goes up to $r+1$ and not just r as in Claim 7.5, because the differentiation of $y\in\mathbb{R}^d \mapsto f(\Theta_{k+1}^\mu(y))$ involves the additional dependence on $\mathcal{V}u$.

Then, the first assertion follows from applying (8.12) in Lemma 8.5 to u and $f(\Theta_{k+1}^\mu)$. For the second assertion, recall that the i^{th} component of ζ_k is expressed as $\zeta_k^i = 4\Delta_{T_k}^{-1}I_{(i)}^{T_k,T_{k+1}} - 6\Delta_{T_k}^{-2}I_{(0,i)}^{T_k,T_{k+1}}$. Then, applying (8.13) and (8.14) with $n=m$ in Lemma 8.5 to u and $f(\Theta_{k+1}^\mu)$, we conclude on the second assertion. □

It will be handy to have an expansion on the prediction error, by recalling that $\tilde{u}$ is essentially an application of the first order scheme, it follows that

$$(u-\tilde{u})(T_k,\hat{X}^\pi_{T_k}) = \mathbb{E}_{\mathbb{Q}_{T_k,T_{k+1}}}\left[(u-\hat{u}^2)(T_{k+1},X^{T_k,\hat{X}^\pi_{T_k},\hat{\mu}}_{T_{k+1}}) + \Delta_{T_{k+1}}\left(f(\Theta^{\hat{\mu}}_{k+1}(X^{T_k,\hat{X}^\pi_{T_k},\hat{\mu}}_{T_{k+1}}) - f(\hat{\Theta}^{\pi,2}_{k+1})\right)\right] + \epsilon^s_{\tilde{u},k}(\pi) + \epsilon^c_{\tilde{u},k}(\pi) \tag{7.32}$$

where

$$\epsilon^c_{\tilde{u},k}(\pi) := \mathbb{E}\left[u(T_{k+1},X^{T_k,\hat{X}^\pi_{T_k},\mu}_{T_{k+1}}) + \Delta_{T_{k+1}}f(\Theta^{\mu}_{k+1}(X^{T_k,\hat{X}^\pi_{T_k},\hat{\mu}}_{T_{k+1}})\right] - \mathbb{E}_{\mathbb{Q}_{T_k,T_{k+1}}}\left[u(T_{k+1},X^{T_k,\hat{X}^\pi_{T_k},\hat{\mu}}_{T_{k+1}}) + \Delta_{T_{k+1}}f(\Theta^{\hat{\mu}}_{k+1}(X^{T_k,\hat{X}^\pi_{T_k},\hat{\mu}}_{T_{k+1}})\right] \tag{7.33}$$

$$\epsilon^s_{\tilde{u},k}(\pi) := u(T_k,\hat{X}^\pi_{T_k}) - \mathbb{E}\left[u(T_{k+1},X^{T_k,\hat{X}^\pi_{T_k},\mu}_{T_{k+1}}) + \Delta_{T_{k+1}}f[\Theta^{\mu}_{k+1}(X^{T_k,\hat{X}^\pi_{T_k},\hat{\mu}}_{T_{k+1}})]\right]. \tag{7.34}$$

Note that $\epsilon^s_{\tilde{u},k}(\pi)$ and $\epsilon^c_{\tilde{u},k}(\pi)$ may be bounded respectively as in Claims (7.4) and (7.5) . The fact that we are using here Θ^μ_{k+1} instead of $\bar{\Theta}^{\mu,\mu}_{k+1,k}$ in those claims, is not really problematic. For the scheme error in Claim (7.4), the difference can be controlled by an additional application of Ito's theorem, but we skip the details. For the cubature error in Claim (7.5) this difference plays no role at all.

Let us now focus on the errors when approaching the driver. Using the mean value theorem, we know that there exist Ψ_1,Ψ_2,Ψ_3 respectively bounded by $\|\partial_{y'}f\|_\infty$, $\|\partial_z f\|_\infty$ and $\|\partial_w f\|_\infty$, such that

$$f(\Theta^\mu_k(\hat{X}^\pi_{T_k})) - f(\hat{\Theta}^{\pi,2}_k) = \Psi_1\left[u(T_k,\hat{X}^\pi_{T_k}) - \hat{u}^2(T_k,\hat{X}^\pi_{T_k})\right] + \Psi_2\left[v(T_k,\hat{X}^\pi_{T_k}) - \hat{v}^2(T_k,\hat{X}^\pi_{T_k})\right] \tag{7.35}$$

$$+ \Psi_3\left(\langle\hat{\mu}_{T_k},\varphi_f[\cdot,u(T_k,\cdot)]\rangle - \langle\hat{\mu}_{T_k},\varphi_f[\cdot,\hat{u}^2(T_k,\cdot),\cdot)]\rangle\right) \tag{7.36}$$

$$+ \Psi_3\left(\langle\mu_{T_k},\varphi_f[\cdot,u(T_k,\cdot)]\rangle - \langle\hat{\mu}_{T_k},\varphi_f[\cdot,u(T_k,\cdot),\cdot)]\rangle\right), \tag{7.37}$$

Similarly, there exist random variables and Ψ'_1,Ψ'_2,Ψ'_3 respectively bounded by $\|\partial_{y'}f\|_\infty$, $\|\partial_z f\|_\infty$ and $\|\partial_w f\|_\infty$, such that

$$f(\Theta^\mu_k(\hat{X}^\pi_{T_k})) - f(\tilde{\Theta}^\pi_k) = \Psi'_1\left[u(T_k,\hat{X}^\pi_{T_k}) - \tilde{u}(T_k,\hat{X}^\pi_{T_k})\right] + \Psi'_2\left[v(T_k,\hat{X}^\pi_{T_k}) - \hat{v}^2(T_k,\hat{X}^\pi_{T_k})\right] \tag{7.38}$$

$$+ \Psi'_3\left(\langle\hat{\mu}_{T_k},\varphi_f[\cdot,u(T_k,\cdot)]\rangle - \langle\hat{\mu}_{T_k},\varphi_f[\cdot,\tilde{u}(T_k,\cdot),\cdot)]\rangle\right) \tag{7.39}$$

$$+ \Psi'_3\left(\langle\mu_{T_k},\varphi_f[\cdot,u(T_k,\cdot)]\rangle - \langle\hat{\mu}_{T_k},\varphi_f[\cdot,u(T_k,\cdot),\cdot)]\rangle\right).. \tag{7.40}$$

Then, using the error development for $(u-\hat{u}^2)(T_k,\hat{X}^\pi_{T_k})$ given in (7.27),(7.28), (7.25),(7.26) the error and the ones for $f(\Theta^\mu_k(\hat{X}^\pi_{T_k})) - f(\hat{\Theta}^{\pi,2}_k)$ in (7.35),(7.36) and (7.37) and $f(\Theta^\mu_k(\hat{X}^\pi_{T_k})) - f(\tilde{\Theta}^\pi_k)$ in (7.38), (7.39) and (7.40) , it follows

$$
\begin{aligned}
(u-\hat{u}^2)(T_k,\hat{X}^\pi_{T_k}) &+ \frac{1}{2}\Delta_{T_k}\left(f(\Theta^\mu_k(\hat{X}^\pi_{T_k})) - f(\hat{\Theta}^{\pi,2}_k)\right)\\
&= (1+\Delta_{T_k}\Psi_1)\mathbb{E}_{\mathbb{Q}_{T_k,T_{k+1}}}\Bigg[u(T_{k+1},X^{T_k,\hat{X}^\pi_{T_k},\hat{\mu}}_{T_{k+1}}) - \hat{u}^2(T_{k+1},X^{T_k,\hat{X}^\pi_{T_k},\hat{\mu}}_{T_{k+1}})\\
&\qquad + \frac{1}{2}\Delta_{T_{k+1}}\left(f(\Theta^\mu_{k+1}(X^{T_k,\hat{X}^\pi_{T_k},\mu}_{T_{k+1}})) - f(\hat{\Theta}^{\pi,2}_{k+1})\right)\Bigg]\\
&+ (1+\Delta_{T_k}\Psi_1)(\epsilon^s_{\hat{u}^2,k}(\pi)+\epsilon^c_{\hat{u}^2,k})(\pi)\\
&+ \frac{1}{2}[\Psi_2\Delta_{T_k}+\Psi_2'\Delta_{T_{k+1}}]\big[v(T_k,\hat{X}^\pi_{T_k}) - \hat{v}^2(T_k,\hat{X}^\pi_{T_k})\big]\\
&+ \Delta_{T_{k+1}}\frac{\Psi_1'}{2}\big[u(T_k,\hat{X}^\pi_{T_k}) - \tilde{u}(T_k,\hat{X}^\pi_{T_k})\big]\\
&+ \frac{1}{2}[\Psi_3\Delta_{T_k}+\Psi_3'\Delta_{T_{k+1}}]\left(\langle\mu_{T_k},\varphi_f[\cdot,u(T_k,\cdot)]\rangle - \langle\hat{\mu}_{T_k},\varphi_f[\cdot,u(T_k,\cdot),\cdot)]\rangle\right),\\
&+ \Delta_{T_k}\frac{\Psi_3}{2}\left(\langle\hat{\mu}_{T_k},\varphi_f[\cdot,u(T_k,\cdot)]\rangle - \langle\hat{\mu}_{T_k},\varphi_f[\cdot,\hat{u}^2(T_k,\cdot),\cdot)]\rangle\right)\\
&+ \Delta_{T_{k+1}}\frac{\Psi_3'}{2}\left(\langle\hat{\mu}_{T_k},\varphi_f[\cdot,u(T_k,\cdot)]\rangle - \langle\hat{\mu}_{T_k},\varphi_f[\cdot,\tilde{u}(T_k,\cdot),\cdot)]\rangle\right)
\end{aligned}
$$

Then, replacing the expansion for the error of $\hat{v}^2$ in terms of (7.30),(7.31), (7.29) and the one for $(u-\tilde{u})$ in terms of (7.32), (7.33), (7.34), and up to rescaling some of the Ψ random variables, one obtains

$$
\begin{aligned}
(u-\hat{u}^2)(T_k,\hat{X}^\pi_{T_k}) &+ \frac{1}{2}\Delta_{T_k}\left(f(\Theta^\mu_k(\hat{X}^\pi_{T_k})) - f(\hat{\Theta}^{\pi,2}_k)\right) \qquad (7.41)\\
&= \mathbb{E}_{\mathbb{Q}_{T_k,T_{k+1}}}\Bigg((1+\Delta_{T_{k+1}}\Psi_1''+\Delta_{T_k}\zeta_{k+1}\Psi_2'')\Big[(u-\hat{u}^2)(T_{k+1},X^{T_k,\hat{X}^\pi_{T_k},\hat{\mu}}_{T_{k+1}})\\
&\qquad + \frac{1}{2}\Delta_{T_{k+1}}\left(f(\Theta^\mu_{k+1}(X^{T_k,\hat{X}^\pi_{T_k},\mu}_{T_{k+1}})) - f(\hat{\Theta}^{\pi,2}_{k+1})\right)\Big]\Bigg)\\
&- \mathbb{E}_{\mathbb{Q}_{T_k,T_{k+1}}}\left(\left[(u-\hat{u}^2)(T_{k+1},X^{T_k,\hat{X}^\pi_{T_k},\hat{\mu}}_{T_{k+1}})\right](\Delta_{T_{k+1}}\Psi_1''+\Delta_{T_{k+1}}\zeta_{k+1}\Psi_2'')\right)\\
&+ (1+\Delta_{T_k}\Psi_1)(\epsilon^s_{\hat{u}^2,k}(\pi)+\epsilon^c_{\hat{u}^2,k}(\pi)) + \Delta_{T_{k+1}}\Psi_3''(\epsilon^s_{\tilde{u},k}(\pi)+\epsilon^c_{\tilde{u},k}(\pi))\\
&+ \Delta_{T_{k+1}}\Psi_4''(\epsilon^s_{\hat{v}^2,k}(\pi)+\epsilon^c_{\hat{v}^2,k}(\pi))\\
&+ \frac{1}{2}[\Psi_3\Delta_{T_k}+\Psi_3'\Delta_{T_{k+1}}]\left(\langle\mu_{T_k},\varphi_f[\cdot,u(T_k,\cdot)]\rangle - \langle\hat{\mu}_{T_k},\varphi_f[\cdot,u(T_k,\cdot),\cdot)]\rangle\right),\\
&+ \Delta_{T_k}\frac{\Psi_3}{2}\left(\langle\hat{\mu}_{T_k},\varphi_f[\cdot,u(T_k,\cdot)]\rangle - \langle\hat{\mu}_{T_k},\varphi_f[\cdot,\hat{u}^2(T_k,\cdot),\cdot)]\rangle\right)\\
&+ \Delta_{T_{k+1}}\frac{\Psi_3'}{2}\left(\langle\hat{\mu}_{T_k},\varphi_f[\cdot,u(T_k,\cdot)]\rangle - \langle\hat{\mu}_{T_k},\varphi_f[\cdot,\tilde{u}(T_k,\cdot),\cdot)]\rangle\right).
\end{aligned}
$$

We are now in position to give the dynamics of the sum of the maximal errors $\mathcal{E}^2_u(k)+\Delta_{T_k}\mathcal{E}_f(k)$. We use Cauchy-Schwartz inequality on the first two terms of (7.41), bound the last two terms in (7.41) using the Lipschitz property of φ_f, as in the first order error analysis, we can deduce that for some constants C, C',

$$
\mathcal{E}^2_u(k)+\Delta_{T_{k+1}}\mathcal{E}_f(k) \le (1+C\Delta_{T_{k+1}})\left[\mathcal{E}^2_u(k+1)+\Delta_{T_k}\mathcal{E}_f(k+1)\right] + C'(\bar{\epsilon}_2(k)+\Delta_{T_{k+1}}N^{-[(m-1)\wedge q]/2}) \qquad (7.42)
$$

with

$$\bar{\epsilon}_2(k) = \sup_{\pi\in\mathcal{S}_k} |\epsilon^s_{\hat{u}^2,k}(\pi) + \Delta_{T_{k+1}}\epsilon^s_{\hat{v}^2,k}(\pi)| + \sup_{\pi\in\mathcal{S}_k} |\epsilon^c_{\hat{u}^2,k}(\pi) + \Delta_{T_{k+1}}\epsilon^c_{\hat{v}^2,k}(\pi)| + \Delta_{T_{k+1}}(\sup_{\pi\in\mathcal{S}_k} |\epsilon^s_{\tilde{u},k}(\pi) + \epsilon^c_{\tilde{u},k}(\pi)|). \tag{7.43}$$

We can bound the two first terms of $\bar{\epsilon}_2(k)$ using respectively Claims 7.7, 7.8, while the last term may be bounded using Claims 7.4 and 7.5 as we have discussed before, so that there is a constant C''

$$\begin{aligned}\bar{\epsilon}_2(k) \leq C''\Bigg(&\sup_{s\in[T_k,T_{k+1}]} ||u(s,\cdot)||_{4,\infty}\Delta^3_{T_{k+1}} \\ &+ [||u(T_{k+1},\cdot)||_{3,\infty} + M_u(4,T_{k+1})]\left[\Delta^{q+1}_{T_{k+1}} + \Delta^2_{T_{k+1}}N^{-[(m-1)\wedge 2q]/2}\right] \\ &+ \sum_{i=m-2}^{m+1}[||u(T_{k+1},\cdot)||_{i,\infty} + M_u(i,T_{k+1}))]\Delta^{(i+1)/2}_{T_{k+1}} + \sum_{i=m+2}^{m+3} M_u(i,T_{k+1})\Delta^{(i+1)/2}_{T_{k+1}}\Bigg).\end{aligned} \tag{7.44}$$

Given that the initialization step is the first order scheme, Claims 7.4, 7.5 and 7.6 imply $\mathcal{E}^2_u(N-1) \leq KN^{-2}$ and $\Delta_{T_{N-1}}\mathcal{E}^2_f(N-1) \leq KN^{-2}$. Moreover, under **(SB)** , we have from Lemma 8.2 that for all $n \in \mathbb{N}^*$ there exists a constant K, depending on the regularity of $V_{0:d}$ and ϕ, such that for all $k \in \{0,\cdots,N-1\}$,

$$M_u(n,T_{k+1}) + \sup_{s\in[T_k,T_{k+1}]} ||u(s,\cdot)||_{n,\infty} \leq K.$$

An application of the discrete Gronwall lemma on the sum $\mathcal{E}^2_u(k) + \Delta_{T_{k+1}}\mathcal{E}_f(k)$, gives

$$\sup_{k\leq N-1} \mathcal{E}^2_u(k) + \Delta_{T_k}\mathcal{E}_f(k) \leq C\left(\mathcal{E}^2_u(N-1) + \Delta_{T_N}\mathcal{E}_f(N-1) + N^{-[(m-1)\wedge q]/2} + \sum_{i=0}^{N-2}\bar{\epsilon}_2(k)\right).$$

Using (7.44) we deduce that, if $m \geq 7$,where the bound is a consequence of the second assertion of Claim 7.8, we have

$$\sup_{k\leq N-1} \mathcal{E}^2_u(k) + \Delta_{T_k}\mathcal{E}_f(k) \leq CN^{-2}.$$

As for the first order case, the previous result together with the expansion of $v - \hat{v}^2$ given (7.30), (7.31), (7.29) and Cauchy-Schwartz inequality, implies

$$\Delta^{1/2}_{T_k}\mathcal{E}^2_v(k) \leq CN^{-2}.$$

This concludes the proof of assertions (3.3) and (3.4) in Theorem 3.1. □

8 Mathematical tools

Here we will intensively use the notions defined in section 6.

8.1 The conditional linear PDE

Lemma 8.1. *Let η be a given family of probability measures on $\mathbb{R}^d$ and consider the PDE*

$$\begin{cases} \partial_t\psi(t,y) + \mathcal{L}^\eta\psi(t,y) = 0, \ \text{on } [0,T]\times\mathbb{R}^d \\ \psi(T,y) = \phi(y) \end{cases} \tag{8.1}$$

*where $\mathcal{L}^\eta$ is defined by (6.1). Suppose that assumption **(SB)** holds, then, this PDE admits a unique infinitely differentiable solution ψ and for every multi-index $\beta \in \mathcal{M}$ there exists a positive constant C depending on the regularity of $V_{0:d}$, $\varphi_{0:d}$ and T such that, :*

$$||D_\beta\psi(t,.)||_\infty \leq C||\phi||_{||\beta||,\infty} \tag{8.2}$$

Proof. Since the law that appears in the coefficients of the SDE is fixed, this is an obvious consequence of the regularity of the coefficients and the terminal condition.

□

8.2 The conditional semi-linear PDE

Lemma 8.2. *Under **(SB)** there exists a function u from $[0,T]\times\mathbb{R}^d$ to $\mathbb{R}$ such that*

$$Y_t^y = u(t, X_t^y)$$

where Y_t^y is defined in (1.1). This function is in $C^{1,2}([0,T]\times\mathbb{R}^d,\mathbb{R})$ and is the unique solution of the semi-linear PDE:

$$\begin{cases} \partial_t u(t,y) + \mathcal{L}^\mu u(t,y) = f\left(t, y, u(t,y), (\mathcal{V}^\mu u(t,y))^T, \langle \mu_t, \varphi_f(\cdot, u(t,\cdot))\rangle\right), \ \text{on } [0,T]\times\mathbb{R}^d \\ u(T,y) = \phi(y) \end{cases} \tag{8.3}$$

where $\mathcal{L}^\mu$ is defined as in (1.4).

Moreover, u is infinitely differentiable, and for every multi-index $\beta\in\mathcal{M}$ there exists a positive constant C depending on the regularity of $V_{0:d}, f, \varphi_{0:d}$, ϕ and T such that,

$$||D_\beta u||_\infty \le C \tag{8.4}$$

Proof of Lemma 8.2.

(i) *Existence and PDE solution.* Consider the conditional BSDE:

$$\begin{cases} dX_s^{t,y,x} = \sum_{i=0}^d V_i(s, X_s^{t,y,x}, \mathbb{E}\left[\varphi_i(X_s^x)\right])dB_s^i \\ d\bar{Y}_s^{t,y,x} = -f(s, X_s^{t,y,x}, \bar{Y}_s^{t,y,x}, \bar{Z}_s^{t,y,x}, \mathbb{E}\left[\varphi_f(X_s^x, Y_s^x)\right])ds + \bar{Z}_s^{t,y,x} dB_s^{1:d} \\ X_t^{t,y,x} = y, \quad \bar{Y}_T^{t,y,x} = \phi(X_T^{t,y,x}), \end{cases} \tag{8.5}$$

for s in $[t,T]$. Note that the McKean term in (8.5), $\mathbb{E}\varphi_f(X_s^x, Y_s^x)$, does not depend on y. In fact, if the solution of (1.1) is found, one can consider the term $\mathbb{E}\varphi_f(X_s^x, Y_s^x)$ simply as a term depending on time, so that conditionally to knowing the joint law of $(X_t^x, Y_t^x,\ 0\le t\le T)$, equation (8.5) is classical and Markov. As pointed out before, the existence of a unique solution to (1.1) follows the lines of the results in [BLP09].

It is clear from the previous discussion that we might apply classical results on BSDE to analyze equation (8.5). In particular, we have from the results of Pardoux and Peng in [PP92] given the regularity in space of f and ϕ, that the mapping $(t,y)\mapsto \bar{u}(t,y) = \bar{Y}_t^{t,y,x}$ the solution of (8.5) is differentiable, once in time and twice in space with bounded derivatives and satisfies the PDE:

$$\begin{cases} \partial_t \bar{u}(t,y) + \mathcal{L}^\mu \bar{u}(t,y) = f\left(t, y, \bar{u}(t,y), (\mathcal{V}^\mu \bar{u}(t,y))^T, \mathbb{E}[\varphi_f(X_t^x, Y_t^x)]\right) \\ \bar{u}(T,y) = \phi(y) \end{cases},$$

Finally, one can show that $\bar{u}$ solves (8.3). To see this, notice that due to the uniqueness of the solutions to (1.1) and (8.5),

$$\bar{u}(t, X_t^x) = \bar{u}(t, X_t^{0,x,x}) = \bar{Y}_t^{t, X_t^{0,x}, x} = \bar{Y}_t^{0,x,x} = Y_t^x.$$

This equality implies

$$\mathbb{E}[\varphi_f(X_s^x, \bar{u}(s, X_s^x))] = \mathbb{E}[\varphi_f(X_s^x, Y_s^x)]$$

for all s in $[0,T]$. Therefore we set $u := \bar{u}$. This concludes the proof of the first assertion.

(ii) *Control on the derivatives.*

To prove the regularity of u and the the bound on its derivatives, we consider first the case involving only space derivatives. In this case the whole argument of Pardoux and Peng may be iterated reasoning

on the BSDE for the first derivative, as long as the hypotheses remain valid, to obtain a BSDE for higher order derivatives in space. We turn the reader to the paper of Crisan and Delarue [CD12c] where this is done in detail (taking into account the additional law dependence that must be considered in our framework).

It remains to consider the case of general derivatives including time derivatives. As we have said before, iterative applications of the Pardoux and Peng argument lead to PDEs similar to (8.3). Then, we can argue that we are able to differentiate once in time for every two derivatives in space. It is also clear that the control on the space derivatives plus the regularity properties of the coefficients imply the control for time derivatives.

8.3 One-step errors

Let η be a given family of probability measures and $X^{t,y,\eta}$, $\tilde{X}^{t,y,\eta}$ defined as in (6.2) and (6.3). We recall that $\hat{\mu}$ denotes the discrete probability measure defined by Algorithm 1 and μ the law of the forward part of (1.1).

Lemma 8.3. *Let g be a C_b^2 function from $\mathbb{R}^d$ to $\mathbb{R}$. Then, there exists a constant C depending only on $V_{0:d}$ and T such that for all $k = 1, \cdots, N-1$:*

$$\begin{aligned}\left|(P^{\eta}_{T_k,T_{T_{k+1}}} - \tilde{P}^{\hat{\mu}}_{T_k,T_{T_{k+1}}})g(y)\right| \quad &\leq \quad C||g||_{2,\infty}\sum_{i=0}^{d}\int_{T_k}^{T_{k+1}}\left|\langle\eta_t,\varphi_i\rangle - \sum_{p=0}^{q-1}[(t-T_k)^p/p!]\langle\eta_{T_k},(\mathcal{L}^{\eta})^p\varphi_i\rangle\right|dt \\ &+C||g||_{2,\infty}\sum_{i=0}^{d}\sum_{p=0}^{q-1}\Delta_{T_{k+1}}^{p+1}\left|\langle\eta_{T_k},(\mathcal{L}^{\eta})^p\varphi_i\rangle - \langle\hat{\mu}_{T_k},(\mathcal{L}^{\hat{\mu}})^p\varphi_i\rangle\right|. \quad (8.6)\end{aligned}$$

Moreover, if g is C_b^3, there exists a constant C depending only on $V_{0:d}, d$ such that for all $l = 1, \cdots, d$ and $k = 1, \cdots, N-1$:

$$\begin{aligned}&\left|\mathbb{E}\left(\left[g\left(X^{T_k,y,\eta}_{T_{k+1}}\right) - g\left(\tilde{X}^{T_k,y,\hat{\mu}}_{T_{k+1}}\right)\right]I^{T_k,T_{k+1}}_{(l)}\right)\right| \quad (8.7)\\ &\leq C||g||_{3,\infty}\sum_{i=0}^{d}\int_{T_k}^{T_{k+1}}\int_{T_k}^{t}\left|\langle\eta_s,\varphi_i\rangle - \sum_{p=0}^{q-1}[(t-T_k)^p/p!]\langle\eta_{T_k},(\mathcal{L}^{\eta})^p\varphi_i\rangle\right|dsdt\\ &+C||g||_{3,\infty}\sum_{i=0}^{d}\sum_{p=0}^{q-1}\Delta_{T_{k+1}}^{p+1}\left|\langle\eta_{T_k},(\mathcal{L}^{\eta})^p\varphi_i\rangle - \langle\hat{\mu}_{T_k},(\mathcal{L}^{\hat{\mu}})^p\varphi_i\rangle\right|\end{aligned}$$

and

$$\begin{aligned}&\left|\mathbb{E}\left(\left[g\left(X^{T_k,y,\eta}_{T_{k+1}}\right) - g\left(\tilde{X}^{T_k,y,\hat{\mu}}_{T_{k+1}}\right)\right]I^{T_k,T_{k+1}}_{(0,l)}\right)\right| \quad (8.8)\\ &\leq C||g||_{3,\infty}\Delta_{T_{k+1}}\sum_{i=0}^{d}\int_{T_k}^{T_{k+1}}\int_{T_k}^{t}\left|\langle\eta_s,\varphi_i\rangle - \sum_{p=0}^{q-1}[(t-T_k)^p/p!]\langle\eta_{T_k},(\mathcal{L}^{\eta})^p\varphi_i\rangle\right|dsdt\\ &+C||g||_{3,\infty}\sum_{i=0}^{d}\sum_{p=0}^{q-1}\Delta_{T_{k+1}}^{p+2}\left|\langle\eta_{T_k},(\mathcal{L}^{\eta})^p\varphi_i\rangle - \langle\hat{\mu}_{T_k},(\mathcal{L}^{\hat{\mu}})^p\varphi_i\rangle\right|.\end{aligned}$$

Lemma 8.4. *Let $n \geq 1$, g be a C_b^{n+2} function from $\mathbb{R}^d$ to $\mathbb{R}$ and $\mathbb{Q}$ be a cubature measure of order n. Then, there exist constants C, C' depending only on $V_{0:d}, d, n$ such that for all $i = 1, \ldots, d$, for all*

$k = 1, \cdots, N-1$:

$$\left|(\tilde{P}^{\eta}_{T_k,T_{k+1}} - \tilde{Q}^{\eta}_{T_k,T_{k+1}})g(y)\right| = \left|(\mathbb{E} - \mathbb{E}_{\mathbb{Q}})\left[g(\tilde{X}^{T_k,y,\eta}_{T_{k+1}})\right]\right| \leq C \sum_{l=n+1}^{n+2} ||g||_{l,\infty} \Delta^{(l)/2}_{T_{k+1}} \tag{8.9}$$

$$\left|(\mathbb{E} - \mathbb{E}_{\mathbb{Q}})\left[g(\tilde{X}^{T_k,y,\eta}_{T_{k+1}}) I^{T_k,T_{k+1}}_{(i)}\right]\right| \leq C \sum_{l=n}^{n+1} ||g||_{l,\infty} \Delta^{(l+1)/2}_{T_{k+1}} \tag{8.10}$$

$$\left|(\mathbb{E} - \mathbb{E}_{\mathbb{Q}})\left[g(\tilde{X}^{T_k,y,\eta}_{T_{k+1}}) I^{T_k,T_{k+1}}_{(0,i)}\right]\right| \leq C' \sum_{l=n-2}^{n-1} ||g||_{l,\infty} \Delta^{(l+3)/2}_{T_{k+1}}, \tag{8.11}$$

for all $y \in \mathbb{R}^d$.

Lemma 8.5. *Let* $n \geq 1$ *and* g *be a* C^{n+2}_b *function from* $\mathbb{R}^d$ *to* $\mathbb{R}$. *Let* $\mathbb{Q}$ *be a cubature measure of order* n, *and* $\hat{X}$ *be the associated cubature tree. Then, there exists a constant* C *depending only on* d, q, $V_{0:d}, n, T, ||\varphi_{0:d}||_{2q+n+2,\infty}$ *such that, for all* $k = 1, \cdots, N-1$:

$$\begin{aligned}\left|\mathbb{E}\left[g\left(X^{T_k,y,\mu}_{T_{k+1}}\right)\right] - \mathbb{E}_{\mathbb{Q}_{T_k,T_{k+1}}}\left[g\left(X^{T_k,y,\hat{\mu}}_{T_{k+1}}\right)\right]\right| &\leq C||g||_{2,\infty}\left[\Delta^{q+1}_{T_{k+1}} + \Delta_{T_{k+1}} N^{-[(n-1)\wedge 2q]/2}\right] \\ &\quad + C||g||_{n+1,\infty}\Delta^{(n+1)/2}_{T_{k+1}} + C||g||_{n+2,\infty}\Delta^{(n+2)/2}_{T_{k+1}} \end{aligned} \tag{8.12}$$

$$\begin{aligned}\left|\mathbb{E}\left[\left(g\left(X^{T_k,y,\mu}_{T_{k+1}}\right) - g\left(\hat{X}^{T_k,y,\hat{\mu}}_{T_{k+1}}\right)\right) I^{T_k,T_{k+1}}_{(l)}\right]\right| &\leq C||g||_{3,\infty}\left[\Delta^{q+1}_{T_{k+1}} + \Delta^{2}_{T_{k+1}} N^{-[(n-1)\wedge 2q]/2}\right] \\ &\quad + C||g||_{n,\infty}\Delta^{(n+1)/2}_{T_{k+1}} + C||g||_{n+1,\infty}\Delta^{(n+2)/2}_{T_{k+1}} \end{aligned} \tag{8.13}$$

and

$$\begin{aligned}\left|\mathbb{E}\left[\left(g\left(X^{T_k,y,\mu}_{T_{k+1}}\right) - g\left(\hat{X}^{T_k,y,\hat{\mu}}_{T_{k+1}}\right)\right) I^{T_k,T_{k+1}}_{(0,l)}\right]\right| &\leq C||g||_{3,\infty}\left[\Delta^{q+2}_{T_{k+1}} + \Delta^{3}_{T_{k+1}} N^{-[(n-1)\wedge 2q]/2}\right] \\ &\quad + C||g||_{n-2,\infty}\Delta^{(n+1)/2}_{T_{k+1}} + C||g||_{n-1,\infty}\Delta^{(n+2)/2}_{T_{k+1}} \end{aligned} \tag{8.14}$$

8.4 Proofs of Lemmas 8.3, 8.4 and 8.5

8.4.1 Proof of Lemma 8.3

Let $k \in \{0, \cdots, N-1\}$, consider the PDE (8.1) with g as boundary condition. It is clear that this PDE admits a unique solution $\tilde{u}$ and that there exists a positive constant $C(T, V_{0:d})$ such that for every multi-index of space derivatives with $\beta \in \mathcal{A}_3$

$$||\tilde{u}||_\infty + ||D_\beta \tilde{u}||_\infty \leq C(T, V_{0:d})||g||_{||\beta||,\infty}. \tag{8.15}$$

Let us write:

$$\left|(P^{\eta}_{T_k,T_{k+1}} - \tilde{P}^{\hat{\mu}}_{T_k,T_{k+1}})g(y)\right| = \left|\mathbb{E}\left[g\left(X^{T_k,y,\eta}_{T_{k+1}}\right) - g\left(\tilde{X}^{T_k,y,\hat{\mu}}_{T_{k+1}}\right)\right]\right|.$$

Now, since $\tilde{u}$ is the solution of (8.1) we have that

$$g\left(X^{T_k,y,\eta}_{T_{k+1}}\right) = \tilde{u}(T_k, y) + \sum_{j=1}^{d} \int_{T_k}^{T_{k+1}} \mathcal{V}^{\eta}_{(j)} \tilde{u}(s, X^{T_k,y,\eta}_s) dB^j_s, \tag{8.16}$$

by Itô's Formula and

$$\begin{aligned} g\left(\tilde{X}^{T_k,y,\hat{\mu}}_{T_{k+1}}\right) &= \tilde{u}(T_k, y) + \int_{T_k}^{T_{k+1}} (\mathcal{L}^{\eta} - \tilde{\mathcal{L}}^{\hat{\mu}})\tilde{u}(s, \tilde{X}^{T_k,y,\hat{\mu}}_s) ds \\ &\quad + \sum_{j=1}^{d} \int_{T_k}^{T_{k+1}} \mathcal{V}^{\hat{\mu}}_{(j)} \tilde{u}(s, \tilde{X}^{T_k,y,\hat{\mu}}_s) dB^j_s. \end{aligned} \tag{8.17}$$

Therefore,

$$\mathbb{E}\left[g\left(X_{T_{k+1}}^{T_k,y,\eta}\right)-g\left(\tilde{X}_{T_{k+1}}^{T_k,y,\hat{\mu}}\right)\right]=-\mathbb{E}\int_{T_k}^{T_{k+1}}\left(\mathcal{L}^{\eta}-\tilde{\mathcal{L}}^{\hat{\mu}}\right)\tilde{u}(t,\tilde{X}_t^{T_k,y,\hat{\mu}})dt, \tag{8.18}$$

since $\partial_t\tilde{u}=-\mathcal{L}^{\eta}\tilde{u}$. As (8.15) implies that $\tilde{u}$ and its two first derivatives are bounded, we may control the term above by the difference between the two generators, then by the difference between the frozen (in space of probability measure) $(\eta_t)_{T_K\leq t\leq T_{k+1}}$ and the approximate and frozen (in time) measure $\hat{\mu}_{T_k}$. Hence, taking into account the particular dependence on the measure in our framework, we deduce:

$$\begin{aligned}
&\left|\mathbb{E}\left[g\left(X_{T_{k+1}}^{T_k,y,\eta}\right)-g\left(\tilde{X}_{T_{k+1}}^{T_k,y,\hat{\mu}}\right)\right]\right|\\
&\quad\leq C||g||_{2,\infty}\int_{T_k}^{T_{k+1}}\left[\sum_{i=0}^{d}\left|\langle\eta_t,\varphi_i\rangle-\sum_{p=0}^{q-1}[(t-T_k)^p/p!]\langle\eta_{T_k},(\mathcal{L}^{\eta})^p\varphi_i\rangle\right|\right.\\
&\qquad\qquad\left.+\sum_{i=0}^{d}\sum_{p=0}^{q-1}\Delta_{T_{k+1}}^{p}\left|\langle\eta_{T_k},(\mathcal{L}^{\eta})^p\varphi_i\rangle-\langle\hat{\mu}_{T_k},(\mathcal{L}^{\hat{\mu}})^p\varphi_i\rangle\right|\right]dt\\
&\quad\leq C(T,V_{0:d})||g||_{2,\infty}\int_{T_k}^{T_{k+1}}\sum_{i=0}^{d}\left|\langle\eta_t,\varphi_i\rangle-\sum_{p=0}^{q-1}[(t-T_k)^p/p!]\langle\eta_{T_k},(\mathcal{L}^{\eta})^p\varphi_i\rangle\right|dt\\
&\qquad+C(T,V_{0:d})||g||_{2,\infty}\sum_{i=0}^{d}\sum_{p=0}^{q-1}\Delta_{T_{k+1}}^{p}\Delta_{T_{k+1}}\left|\langle\eta_{T_k},(\mathcal{L}^{\eta})^p\varphi_i\rangle-\langle\hat{\mu}_{T_k},(\mathcal{L}^{\hat{\mu}})^p\varphi_i\rangle\right|
\end{aligned}$$

This concludes the proof of the first assertion. Now, we deduce from (8.16) and (8.18) and integration by parts, that:

$$\begin{aligned}
&\mathbb{E}\left(\left[g\left(X_{T_{k+1}}^{T_k,y,\eta}\right)-g\left(\tilde{X}_{T_{k+1}}^{T_k,y,\hat{\mu}}\right)\right]I_{(l)}^{T_k,T_{k+1}}\right) \qquad (8.19)\\
&=\mathbb{E}\int_{T_k}^{T_{k+1}}\left(\mathcal{L}^{\eta}-\tilde{\mathcal{L}}^{\hat{\mu}}\right)\tilde{u}(t,\tilde{X}_t^{T_k,y,\hat{\mu}})I_{(l)}^{T_k,t}dt\\
&\quad+\mathbb{E}\int_{T_k}^{T_{k+1}}\left[\mathcal{V}_{(l)}(t,X_t^{T_k,y,\eta},\langle\eta_t,\varphi_l\rangle)-\mathcal{V}_{(l)}(t,X_t^{T_k,y,\eta},\langle\hat{\mu}_{T_k},\varphi_l\rangle)\right]\tilde{u}\left(t,X_t^{T_k,y,\eta}\right)dt\\
&\quad+\frac{1}{2}\mathbb{E}\int_{T_k}^{T_{k+1}}\left[\mathcal{V}_{(l)}(t,X_t^{T_k,y,\eta},\langle\hat{\mu}_{T_k},\varphi_l\rangle)\tilde{u}(t,X_t^{T_k,y,\eta})-\mathcal{V}_{(l)}(t,\tilde{X}_t^{T_k,y,\hat{\mu}},\langle\hat{\mu}_{T_k},\varphi_l\rangle)\tilde{u}(t,\tilde{X}_t^{T_k,y,\hat{\mu}})\right]dt.
\end{aligned}$$

First, note that the first two terms in the right hand side above may be controlled by the difference between the coefficients times the supremum norm of the first and second order derivatives of $\tilde{u}$ times the order of the integrals, as we did for (8.18). Second, note that the first assertion of Lemma 8.3 can be applied to the function $\mathcal{V}_{(l)}(t,.,\langle\hat{\mu}_{T_k},\varphi_l\rangle)\tilde{u}(t,.)$ in the last term on the right hand side above. These arguments, together with the bound (8.15) lead to:

$$
\begin{aligned}
&\left|\mathbb{E}\left(\left[g\left(X_{T_{k+1}}^{T_k,y,\eta}\right)-g\left(\tilde{X}_{T_{k+1}}^{T_k,y,\hat{\mu}}\right)\right]I_{(l)}^{T_k,T_{k+1}}\right)\right| \\
&\leq C(T,V_{0:d})\int_{T_k}^{T_{k+1}}(||g||_{1,\infty}+||g||_{2,\infty}(t-T_k)^{1/2})\sum_{i=0}^{d}\left|\langle\eta_t,\varphi_i\rangle-\sum_{p=0}^{q-1}[(t-T_k)^p/p!]\langle\eta_{T_k},(\mathcal{L}^\eta)^p\varphi_i\rangle\right|dt \\
&\quad+C(T,V_{0:d})\sum_{i=0}^{d}\sum_{p=0}^{q-1}\Delta_{T_{k+1}}^{p}\Delta_{T_{k+1}}(||g||_{1,\infty}+||g||_{2,\infty}\Delta_{T_{k+1}}^{1/2})\left|\langle\eta_{T_k},(\mathcal{L}^\eta)^p\varphi_i\rangle-\langle\hat{\mu}_{T_k},(\mathcal{L}^{\hat{\mu}})^p\varphi_i\rangle\right| \\
&\quad+C(T,V_{0:d})||g||_{3,\infty}\Bigg\{\int_{T_k}^{T_{k+1}}\int_{T_k}^{t}\sum_{i=0}^{d}\left|\langle\eta_s,\varphi_i\rangle-\sum_{p=0}^{q-1}[(s-T_k)^p/p!]\langle\eta_{T_k},(\mathcal{L}^\eta)^p\varphi_i\rangle\right|dsdt \\
&\qquad\qquad+\sum_{i=0}^{d}\sum_{p=0}^{q-1}\Delta_{T_{k+1}}^{p}\Delta_{T_{k+1}}^{2}\left|\langle\eta_{T_k},(\mathcal{L}^\eta)^p\varphi_i\rangle-\langle\hat{\mu}_{T_k},(\mathcal{L}^{\hat{\mu}})^p\varphi_i\rangle\right|\Bigg\},
\end{aligned}
$$

and this concludes the proof of the second assertion. Finally, (8.16) and (8.18) and integration by parts, give:

$$
\begin{aligned}
&\mathbb{E}\left(\left[g\left(X_{T_{k+1}}^{T_k,y,\eta}\right)-g\left(\tilde{X}_{T_{k+1}}^{T_k,y,\hat{\mu}}\right)\right]I_{(0,l)}^{T_k,T_{k+1}}\right) \\
&=\mathbb{E}\int_{T_k}^{T_{k+1}}\left(\mathcal{L}^\eta-\tilde{\mathcal{L}}^{\hat{\mu}}\right)\tilde{u}(t,\tilde{X}_t^{T_k,y,\eta})I_{(0,l)}^{T_k,t}dt \\
&\quad+\mathbb{E}\int_{T_k}^{T_{k+1}}\left[\mathcal{V}_{(l)}(t,X_t^{T_k,y,\eta},\langle\eta_t,\varphi_l\rangle)-\mathcal{V}_{(l)}(t,X_t^{T_k,y,\eta},\langle\hat{\mu}_{T_k},\varphi_l\rangle)\right]\tilde{u}\left(t,X_t^{T_k,y,\eta}\right)I_{(0)}^{T_k,t}dt \\
&\quad+\frac{1}{2}\mathbb{E}\int_{T_k}^{T_{k+1}}\left[\mathcal{V}_{(l)}(t,X_t^{T_k,y,\eta},\langle\hat{\mu}_{T_k},\varphi_l\rangle)\tilde{u}(t,X_t^{T_k,y,\eta})-\mathcal{V}_{(l)}(t,\tilde{X}_t^{T_k,y,\hat{\mu}},\langle\hat{\mu}_{T_k},\varphi_l\rangle)\tilde{u}(t,\tilde{X}_t^{T_k,y,\hat{\mu}})\right]I_{(0)}^{T_k,t}dt.
\end{aligned}
$$

Note the similarity with (8.19). So that a similar development gives

$$
\begin{aligned}
&\left|\mathbb{E}\left(\left[g\left(X_{T_{k+1}}^{T_k,y,\eta}\right)-g\left(\tilde{X}_{T_{k+1}}^{T_k,y,\hat{\mu}}\right)\right]I_{(0,l)}^{T_k,T_{k+1}}\right)\right| \\
&\leq C(T,V_{0:d})\int_{T_k}^{T_{k+1}}(||g||_{1,\infty}(t-T_k)+||g||_{2,\infty}(t-T_k)^{3/2})\sum_{i=0}^{d}\left|\langle\eta_t,\varphi_i\rangle-\sum_{p=0}^{q-1}[(t-T_k)^p/p!]\langle\eta_{T_k},(\mathcal{L}^\eta)^p\varphi_i\rangle\right|dt \\
&\quad+C(T,V_{0:d})\Delta_{T_{k+1}}\sum_{i=0}^{d}\sum_{p=0}^{q-1}\Delta_{T_{k+1}}^{p}\Delta_{T_{k+1}}(||g||_{1,\infty}+||g||_{2,\infty}\Delta_{T_{k+1}}^{1/2})\left|\langle\eta_{T_k},(\mathcal{L}^\eta)^p\varphi_i\rangle-\langle\hat{\mu}_{T_k},(\mathcal{L}^{\hat{\mu}})^p\varphi_i\rangle\right| \\
&\quad+C(T,V_{0:d})\Delta_{T_{k+1}}||g||_{3,\infty}\Bigg\{\int_{T_k}^{T_{k+1}}\int_{T_k}^{t}\sum_{i=0}^{d}\left|\langle\eta_s,\varphi_i\rangle-\sum_{p=0}^{q-1}[(s-T_k)^p/p!]\langle\eta_{T_k},(\mathcal{L}^\eta)^p\varphi_i\rangle\right|dsdt \\
&\qquad+\sum_{i=0}^{d}\sum_{p=0}^{q-1}\Delta_{T_{k+1}}^{p}\Delta_{T_{k+1}}^{2}\left|\langle\eta_{T_k},(\mathcal{L}^\eta)^p\varphi_i\rangle-\langle\hat{\mu}_{T_k},(\mathcal{L}^{\hat{\mu}})^p\varphi_i\rangle\right|\Bigg\}
\end{aligned}
$$

from where the last claim is deduced. □

8.4.2 Proof of Lemma 8.4

Let $k\in\{0,\cdots,N-1\}$, once again, we consider the unique infinitely differentiable solution $\tilde{u}$ of PDE (8.1) with g as boundary condition. Recall that for every $\beta\in\mathcal{M}$ there exists a positive constant $C(T,V_{0:d})$ such that:

$$
||\tilde{u}||_\infty+||D_\beta\tilde{u}||_\infty\leq C(T,V_{0:d})||g||_{||\beta||,\infty}. \tag{8.20}
$$

The result then follows from Stratonovich-Taylor expansion of $(t,y) \mapsto \tilde{u}(t,y)$ around (T_k, X_{T_k}) by Theorem 5.6.1 in [KP92] and bounding the remainder as in Proposition 2.1 of [LV04]. □

8.4.3 Proof of Lemma 8.5

Note that

$$\begin{aligned}\left|\mathbb{E}\left[g\left(X_{T_{k+1}}^{T_k,y,\mu}\right)\right] - \mathbb{E}_{\mathbb{Q}_{T_k,T_{k+1}}}\left[g\left(X_{T_{k+1}}^{T_k,y,\hat{\mu}}\right)\right]\right| &= \left|\mathbb{E}\left[g\left(X_{T_{k+1}}^{T_k,y,\mu}\right)\right] - \mathbb{E}\left[g\left(\tilde{X}_{T_{k+1}}^{T_k,y,\hat{\mu}}\right)\right]\right| \\ &+ \left|\mathbb{E}\left[g\left(\tilde{X}_{T_{k+1}}^{T_k,y,\mu}\right)\right] - \mathbb{E}_{\mathbb{Q}_{T_k,T_{k+1}}}\left[g\left(X_{T_{k+1}}^{T_k,y,\hat{\mu}}\right)\right]\right|.\end{aligned} \tag{8.21}$$

Combining estimate (8.6) of Lemma 8.3 with Claim 7.3 and (3.2) in Theorem 3.1, we get that the first term in the right hand side is bounded by:

$$C||g||_{2,\infty}\left[\Delta_{T_{k+1}}^{q+1} + \Delta_{T_{k+1}} N^{-[(n-1)\wedge 2q]/2}\right].$$

The second term in the right hand side of (8.21) can be estimated by combining this bound with the estimate (8.9) in 8.4 (when choosing $\eta = \mu$).

The other assertion follows from the same procedure, substituting (8.7) (resp. (8.8)) to (8.6) and (8.10) (resp. (8.11)) to (8.9).

9 Proofs of Corollary 3.2 and 3.3

9.1 Proof of Corollary 3.2

Many practical applications, particularly in finance, require the algorithm to be able to solve problems in which the boundary condition ϕ is less regular, e.g. when ϕ is just Lipschitz. In this section, we prove how the results obtained in the regular case extend to the case when assumption **(LB)** holds as Corollary 3.2 state.

A preliminary result. We use in addition an auxiliary result shown in the proof of Theorem 8 in [CG07]:

Lemma 9.1. *There exists a positive constant C such that:*

$$\sum_{j=0}^{N-2} \Delta_{T_{j+1}}^{(m+1)/2} (T - T_j)^{-m/2} \leq C L(\gamma, m),$$

where L is defined in (3.6).

We are now ready to examine the error convergence for the forward and backward components of the algorithm.

Proof of the forward approximation in Corollary 3.2

The regularity of the solution of the linear associated linear PDE is essential to our analysis. We start by stating a result in this sense under **(LB)**. This is summarized by

Claim 9.2. *Under **(LB)**, there exists a unique solution ψ to the PDE (8.1) and for every multi-index $\beta \in \mathcal{M}$ there exists a constant C such that:*

$$||D_\beta \psi(t,\cdot)||_\infty \leq C(T-t)^{-(||\beta||-1)/2} \tag{9.1}$$

Proof. This follows from classical results of parabolic equations with parameter, see Chapter 9, Section 3 of [Fri08].

□

Thanks to the uniform ellipticity assumption, even if the terminal condition is not differentiable, we know that the solution of the PDE (7.2) is smooth except at the boundary. Precisely, the gradient bounds (7.8) are now given by

$$||\nabla_y^n \psi(t,\cdot)||_\infty \leq C(T, V_{0:d})||\phi||_{1,\infty}(T-t)^{(1-n)/2}, \tag{9.2}$$

where ψ is defined in (7.1). With this in hand, we can follow the proof exactly as the one of the corresponding forward part in Theorem 3.1 up to estimate (7.7) but where we separate the error on the last step, since there is no smoothing effect there. Then, plugging estimate (9.2) in (7.7) instead of (7.8), we get:

$$\begin{aligned}
&\left|(P_{T_0,T_N} - Q^{\hat{\mu}}_{T_0,T_N})\phi(x)\right| \\
&\leq C(T, V_{0:d})||\phi||_{1,\infty} \sum_{j=0}^{N-2}\sum_{p=0}^{q-1} \Delta_{T_{j+1}}^{p+1}(T-T_{j+1})^{-1/2} \sum_{i=0}^{d} \left|(P_{T_0,T_N} - Q^{\hat{\mu}}_{T_0,T_N})\varphi_i(x)\right| \\
&\quad + C(T, V_{0:d}, d)||\phi||_{1,\infty}\|\varphi\|_{2q,\infty} \sum_{j=0}^{N-2} \Delta_{T_{j+1}}^{q+1}(T-T_{j+1})^{-1/2} \\
&\quad + C(V_{0:d}, d, m)||\phi||_{1,\infty} \sum_{j=0}^{N-2}\sum_{l=m+1}^{m+2} \Delta_{T_{j+1}}^{\frac{l}{2}}(T-T_{j+1})^{(1-l)/2} + \left|(P_{T_0,T_N} - P^{\eta}_{T_0,T_N})\phi(x)\right|.
\end{aligned}$$

We conclude the proof by using Lemma 9.1 on the sums and by combining Lipschitz property of ϕ and adapted time-step on the last step error. □

Proof of the backward approximation in Corollary 3.2

Just as in the forward case, our analysis relies on the regularization properties of the associated non-linear PDE under **(LB)**. We have

Claim 9.3. *Under **(LB)**, there exists a unique solution u of (8.3), for all $(t,y) \in [0,T]\times\mathbb{R}^d$, it is given by*

$$u(t,y) = Y_t^{t,y,\mu},$$

where $Y_t^{t,y,\mu}$ is defined in (1.5). Moreover, for ϕ Lipschitz and bounded, for every multi-index $\beta \in \mathcal{M}$ there exists a positive constant C depending on the regularity of $V_{0:d}, f, \varphi$ and T such that:

$$||D_\beta u(t,\cdot)||_\infty \leq C||\phi||_{1,\infty}^{|\beta|}(T-t)^{-(||\beta||-1)/2} \tag{9.3}$$

Proof. To prove Claim 9.3, we follow the same arguments given for Lemma 8.2. First, due to the regularity properties of the diffusion under the elliptic case, we have similar properties as those used in the paper of Crisan and Delarue [CD12c], even in the non-homogeneous case (notably, the integration by parts property as shown in [Ma02]). Hence, we get the control on derivatives result for space derivatives, and extend it, as before, to time derivatives. □

Armed with the regularity of the function u, we can repeat the proof of the backward approximation in Theorem (3.1). We recover (7.22) for the first order scheme and (7.42) for the second order scheme, i.e.

$$\mathcal{E}_u^1(k) \leq \left(1 + C\Delta_{T_{k+1}}\right)\mathcal{E}_u^1(k+1) + \bar{\epsilon}(k+1), \tag{9.4}$$

$$\mathcal{E}_u^2(k) + \Delta_{T_{k+1}}\mathcal{E}_f(k) \leq (1 + C\Delta_{T_{k+1}})\left[\mathcal{E}_u^2(k+1) + \Delta_{T_k}\mathcal{E}_f(k+1)\right] + C'(\bar{\epsilon}_2(k) + \Delta_{T_{k+1}}N^{-[(m-1)\wedge q]/2}), \tag{9.5}$$

where $\bar{\epsilon}, \bar{\epsilon}_2$ are respectively defined in (7.23), (7.43). Now, if we show that

$$\sum_{k=0}^{N-1} \bar{\epsilon}(k) \leq N^{-1}; \quad \text{and} \quad \sum_{k=0}^{N-2} \bar{\epsilon}_2(k) \leq N^{-2}; \tag{9.6}$$

then, as in the smooth setting, we can apply Gronwall lemma on (9.4) and (9.5) and conclude on the desired rates of convergence for the approximation of u. The arguments for the rate of the approximation of v are exactly as in the smooth setting thus completing the proof of the claimed result.

Therefore, we only need to prove (9.6). But, Claim 9.3 and the definition of M_u given in (7.10) imply

$$M_u(n, T_{k+1}) \leq (T - T_{k+1})^{(1-n)/2} ||\phi||^n_{1,\infty},$$

which together with Claim 9.3 show that under the Lipschitz boundary setup,

$$\begin{aligned}
\bar{\epsilon}(k+1) &\leq C\Big((T-T_k)^{-1/2} \Delta_{T_{k+1}} \left[\Delta^3_{T_{k+1}} + \Delta^q_{T_{k+1}} + N^{-[(m-1)\wedge 2q]/2} \right] \\
&\quad + (T-T_k)^{-m/2} \Delta^{(m+1)/2}_{T_{k+1}} + (T-T_k)^{-(m+1)/2} \Delta^{(m+2)/2}_{T_{k+1}} \\
&\quad + (T-T_k)^{-1} \Delta^{3/2}_{T_{k+1}} \left[\Delta^{1/2}_{T_{k+1}} + \Delta^{q-1/2}_{T_{k+1}} + \Delta^{1/2}_{T_{k+1}} N^{-[(m-1)\wedge 2q]/2} \right] \Big). \\
&\leq C\Big((T-T_k)^{-1/2} \Delta_{T_{k+1}} N^{-1} + (T-T_k)^{-m/2} \Delta^{(m+1)/2}_{T_{k+1}} \\
&\quad + (T-T_k)^{-(m+1)/2} \Delta^{(m+2)/2}_{T_{k+1}} + (T-T_k)^{-1} \Delta^{3/2}_{T_{k+1}} N^{-1/2} \Big);
\end{aligned} \tag{9.7}$$

where we have used the fact that $\Delta_{T_k} \leq CN^{-1}$ even on the decreasing discretization. We can proceed similarly for $\bar{\epsilon}_2$ from inequality (7.44) , to get

$$\begin{aligned}
\bar{\epsilon}_2(k) \leq & C'' \Big((T-T_k)^{-3/2} \Delta^3_{T_{k+1}} + (T-T_k)^{-3/2} \left[\Delta^{q+1}_{T_{k+1}} + \Delta^2_{T_{k+1}} N^{-[(m-1)\wedge 2q]/2} \right] \\
& + \sum_{i=m-2}^{m+3} (T-T_k)^{-(i-1)/2} \Delta^{(i+1)/2}_{T_{k+1}} \Big).
\end{aligned} \tag{9.8}$$

Then, (9.6) follows by applying Lemma 9.1 to (9.7) and (9.8) □

9.2 Proof of Corollary 3.3

On a first hand, by following the proof of (3.2) in Theorem 3.1 we get (7.6), where the difference between the integral of the φ_i, $i = 1, \ldots, d$ against the measures in the right hand side are replaced by the distance $d_{\mathcal{F}}$. Since for all $T_k < t < T$ $d_{\mathcal{F}}(\mu_t, \mu_{T_k}) \leq C(t - T_k)$ (resp. $(t - T_k)^{1/2}$) in the case (3.7) (resp. (3.8)), the result follows from Gronwall's Lemma. This gives the rate of approximation of the law of the forward process.

On a second hand, the backward errors are then obtained by the same arguments already developed in the proof of (3.3) in Theorem 3.1, using the new forward approximation (3.7) (resp. (3.8)) instead of (3.2) (resp. (3.5)) in the proofs of Claims 7.5 and 7.6.

Acknowledgements

The authors would like to thank the CNRS/Royal Society project 'Promising" for partial financial support during the preparation of this work. We thank as well the members of the project for the

discussions on the treated topics. Finally, we would like to thank François Delarue for his valuable suggestions and careful reading of the paper.

References

[AKH02] Fabio Antonelli and Arturo Kohatsu-Higa, *Rate of convergence of a particle method to the solution of the McKean–Vlasov equation*, The Annals of Applied Probability **12** (2002), no. 2, 423–476, Mathematical Reviews number (MathSciNet): MR1910635; Zentralblatt MATH identifier: 1015.60048.

[BDH+03] Ph. Briand, B. Delyon, Y. Hu, E. Pardoux, and L. Stoica, *Lp solutions of backward stochastic differential equations*, Stochastic Processes and their Applications **108** (2003), no. 1, 109–129.

[Ber09] Lorenzo Bergomi, *Smile dynamics III*, SSRN Scholarly Paper ID 1493308, Social Science Research Network, Rochester, NY, October 2009.

[BHH+11] N. Bush, B. M. Hambly, H. Haworth, L. Jin, and C. Reisinger, *Stochastic evolution equations in portfolio credit modelling*, SIAM Journal on Financial Mathematics **2** (2011), no. 1, 627–664.

[BJPR09] Tomasz R. Bielecki, Monique Jeanblanc-Picqué, and Marek Rutkowski, *Credit risk modeling*, Osaka University Press Osaka, 2009.

[BLP09] Rainer Buckdahn, Juan Li, and Shige Peng, *Mean-field backward stochastic differential equations and related partial differential equations*, Stochastic Processes and their Applications **119** (2009), no. 10, 3133–3154.

[Bos05] Mireille Bossy, *Some stochastic particle methods for nonlinear parabolic PDEs*, GRIP—Research Group on Particle Interactions, ESAIM Proc., vol. 15, EDP Sci., Les Ulis, 2005, p. 18–57.

[BT08] Christian Bayer and Josef Teichmann, *Cubature on wiener space in infinite dimension*, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science **464** (2008), no. 2097, 2493–2516 (en).

[Car10] Pierre Cardaliaguet, *Notes on mean field games.*, Technical report (2010).

[CD12a] René Carmona and Francois Delarue, *Probabilistic analysis of mean-field games*, Tech. report, 2012.

[CD12b] René Carmona and François Delarue, *Mean field forward-backward stochastic differential equations*, Tech. report, 2012.

[CD12c] Dan Crisan and François Delarue, *Sharp derivative bounds for solutions of degenerate semi-linear partial differential equations*, Journal of Functional Analysis **263** (2012), no. 10, 3024–3101.

[CDL12] René Carmona, François Delarue, and Aimé Lachapelle, *Control of McKean–Vlasov dynamics versus mean field games*, Mathematics and Financial Economics (2012), 1–36 (en).

[CG07] D. Crisan and S. Ghazali, *On the convergence rates of a general class of weak approximations of SDEs*, Stochastic differential equations: theory and applications, Interdiscip. Math. Sci., vol. 2, World Sci. Publ., Hackensack, NJ, 2007, p. 221–248.

[CM02] Patrick Cattiaux and Laurent Mesnager, *Hypoelliptic non-homogeneous diffusions*, Probability Theory and Related Fields **123** (2002), no. 4, 453–483 (en).

[CM10a] D. Crisan and K. Manolarakis, *Solving backward stochastic differential equations using the cubature method. application to nonlinear pricing*, Progress in analysis and its applications, World Sci. Publ., Hackensack, NJ, 2010, p. 389–397.

[CM10b] Dan Crisan and Konstantinos Manolarakis, *Second order discretization of backward SDEs*, arXiv:1012.5650 (2010).

[DTV12] Philipp Doersek, Josef Teichmann, and Dejan Veluscek, *Cubature methods for stochastic (partial) differential equations in weighted spaces*, arXiv:1201.4024 (2012).

[Fri08] Avner Friedman, *Partial differential equations of parabolic type*, Dover Publications, April 2008 (en).

[FS06] Wendell H. Fleming and H. Mete Soner, *Controlled markov processes and viscosity solutions*, second ed., Stochastic Modelling and Applied Probability, vol. 25, Springer, New York, 2006.

[GHL11] Julien Guyon and Pierre Henry-Labordere, *The smile calibration problem solved*, SSRN Scholarly Paper ID 1885032, Social Science Research Network, Rochester, NY, July 2011.

[GL11] L. G. Gyurkó and T. Lyons, *Efficient and practical implementations of cubature on wiener space*, Stochastic analysis 2010, Springer, Heidelberg, 2011, p. 73–111.

[IDR10] Peter Imkeller and Gonçalo Dos Reis, *Path regularity and explicit convergence rate for BSDE with truncated quadratic growth*, Stochastic Processes and their Applications **120** (2010), no. 3, 348–379.

[KP92] Peter E. Kloeden and Eckhard Platen, *Numerical solution of stochastic differential equations*, Applications of Mathematics (New York), vol. 23, Springer-Verlag, Berlin, 1992.

[KS87] S. Kusuoka and D. Stroock, *Applications of the malliavin calculus. III*, Journal of the Faculty of Science. University of Tokyo. Section IA. Mathematics **34** (1987), no. 2, 391–442.

[Kus01] S. Kusuoka, *Approximation of expectation of diffusion process and mathematical finance*, Taniguchi Conference on Mathematics Nara '98, Adv. Stud. Pure Math., vol. 31, Math. Soc. Japan, Tokyo, 2001, p. 147–165.

[LL06a] J. M. Lasry and P.L. Lions, *Jeux à champ moyen. i. le cas stationnaire*, Comptes Rendus Mathématique. Académie des Sciences. Paris **343** (2006), no. 9, 619–625.

[LL06b] ———, *Jeux à champ moyen. II. horizon fini et contrôle optimal*, Comptes Rendus Mathématique. Académie des Sciences. Paris **343** (2006), no. 10, 679–684.

[LL07a] _____, *Large investor trading impacts on volatility*, Annales de l'Institut Henri Poincaré. Analyse Non Linéaire **24** (2007), no. 2, 311–323.

[LL07b] J.M. Lasry and P.L. Lions, *Mean field games*, Japanese Journal of Mathematics **2** (2007), no. 1, 229–260.

[LV04] Terry Lyons and Nicolas Victoir, *Cubature on wiener space*, Proceedings of The Royal Society of London. Series A. Mathematical, Physical and Engineering Sciences **460** (2004), no. 2041, 169–198, Stochastic analysis with applications to mathematical finance.

[Ma02] Jin Ma, *Representation theorems for backward stochastic differential equations*, The Annals of Applied Probability **12** (2002), no. 4, 1390–1418, Mathematical Reviews number (MathSciNet): MR1936598; Zentralblatt MATH identifier: 1017.60067.

[Pha09] Huyên Pham, *Continuous-time stochastic control and optimization with financial applications*, vol. 61, Springer, 2009.

[PP92] É. Pardoux and S. Peng, *Backward stochastic differential equations and quasilinear parabolic partial differential equations*, Stochastic partial differential equations and their applications (Charlotte, NC, 1991), Lecture Notes in Control and Inform. Sci., vol. 176, Springer, Berlin, 1992, p. 200–217.

[Ric11] Adrien Richou, *Numerical simulation of BSDEs with drivers of quadratic growth*, The Annals of Applied Probability **21** (2011), no. 5, 1933–1964.

[Szn91] Alain-Sol Sznitman, *Topics in propagation of chaos*, École d'Été de Probabilités de Saint-Flour XIX—1989, Lecture Notes in Math., vol. 1464, Springer, Berlin, 1991, p. 165–251.

[Tei06] J. Teichmann, *Calculating the greeks by cubature formulae*, Proceedings of The Royal Society of London. Series A. Mathematical, Physical and Engineering Sciences **462** (2006), no. 2066, 647–670.

[TV03] Denis Talay and Olivier Vaillant, *A stochastic particle method with random weights for the computation of statistical solutions of McKean-Vlasov equations*, The Annals of Applied Probability **13** (2003), no. 1, 140–180, Mathematical Reviews number (MathSciNet): MR1951996; Zentralblatt MATH identifier: 1026.60110.

UNIVERSITE NICE SOPHIA ANTIPOLIS, LABORATOIRE JEAN-ALEXANDRE DIEUDONNE, PARC VALROSE, 06108 CEDEX 02, NICE, FRANCE

E-mail address, P.E. Chaudru de Raynal: `deraynal@unice.fr`

E-mail address, C.A. Garcia Trillos: `camilo@unice.fr`