

Bayer, Christian; Belomestny, Denis; Butkovsky, Oleg; Schoenmakers, John

Article — Published Version

A reproducing kernel Hilbert space approach to singular local stochastic volatility McKean–Vlasov models

Finance and Stochastics

Suggested Citation: Bayer, Christian; Belomestny, Denis; Butkovsky, Oleg; Schoenmakers, John (2024) : A reproducing kernel Hilbert space approach to singular local stochastic volatility McKean–Vlasov models, Finance and Stochastics, ISSN 1432-1122, Springer Berlin Heidelberg, Berlin/Heidelberg, Vol. 28, Iss. 4, pp. 1147-1178, <https://doi.org/10.1007/s00780-024-00541-5>

This Version is available at:

<https://hdl.handle.net/10419/315067>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by/4.0/>



A reproducing kernel Hilbert space approach to singular local stochastic volatility McKean–Vlasov models

Christian Bayer¹ · Denis Belomestny² · Oleg Butkovsky¹ · John Schoenmakers¹

Received: 2 March 2022 / Accepted: 27 November 2023 / Published online: 7 August 2024
© The Author(s) 2024

Abstract

Motivated by the challenges related to the calibration of financial models, we consider the problem of numerically solving a singular McKean–Vlasov equation

$$dX_t = \sigma(t, X_t)X_t \frac{\sqrt{v_t}}{\sqrt{\mathbb{E}[v_t|X_t]}} dW_t,$$

where W is a Brownian motion and v is an adapted diffusion process. This equation can be considered as a singular local stochastic volatility model. While such models are quite popular among practitioners, its well-posedness has unfortunately not yet been fully understood and in general is possibly not guaranteed at all. We develop a novel regularisation approach based on the reproducing kernel Hilbert space (RKHS) technique and show that the regularised model is well posed. Furthermore, we prove propagation of chaos. We demonstrate numerically that a thus regularised model is able to perfectly replicate option prices coming from typical local volatility models. Our results are also applicable to more general McKean–Vlasov equations.

Keywords Stochastic volatility models · Singular McKean–Vlasov equations · Reproducing kernel Hilbert space

Mathematics Subject Classification 91G20 · 65C30 · 46E22

JEL Classification C63 · G17

✉ O. Butkovsky
oleg.butkovskiy@gmail.com

C. Bayer
christian.bayer@wias-berlin.de

D. Belomestny
denis.belomestny@uni-due.de

J. Schoenmakers
john.schoenmakers@wias-berlin.de

¹ Weierstrass Institute, Mohrenstrasse 39, 10117 Berlin, Germany

² Duisburg-Essen University, Essen, Germany

1 Introduction

The present article is motivated by Guyon and Henry-Labordère [13], where the authors proposed a particle method for the calibration of local stochastic volatility models (e.g. stock price models). For ease of presentation, let us assume zero interest rates and recall that *local volatility models*

$$dX_t = \sigma(t, X_t)X_t dW_t, \quad (1.1)$$

where W denotes a one-dimensional Brownian motion under a risk-neutral measure and X the price of a stock, can replicate any sufficiently regular implied volatility surface, provided that we choose the local volatility according to *Dupire's formula*, symbolically $\sigma := \sigma_{\text{Dup}}$; see Dupire [8]. (In case of deterministic nonzero interest rates, the discussion below remains virtually unchanged after passing to forward stock and option prices). Unfortunately, it is well understood that Dupire's model exhibits unrealistic random price behaviour despite perfect fits to market prices of options. On the other hand, *stochastic volatility models*

$$dX_t = \sqrt{v_t} X_t dW_t$$

for a suitably chosen stochastic variance process (v_t) , may lead to realistic (in particular, time-homogeneous) dynamics, but are typically difficult or impossible to fit to observed implied volatility surfaces. We refer to Gatheral [11] for an overview of stochastic and local volatility models. Local stochastic volatility models can combine the advantages of both local and stochastic volatility models. Indeed, if the stock price is given by

$$dX_t = \sqrt{v_t} \sigma(t, X_t) X_t dW_t,$$

then it exactly fits the observed market option prices provided that

$$\sigma_{\text{Dup}}^2(t, x) = \sigma^2(t, x) \mathbb{E}[v_t | X_t = x].$$

This is a simple consequence of Gyöngy's celebrated Markovian projection theorem; see Gyöngy [14, Theorem 4.6] and also Brunick and Shreve [4, Corollary 3.7]. With this choice of σ , we have

$$dX_t = \sigma_{\text{Dup}}(t, X_t) X_t \frac{\sqrt{v_t}}{\sqrt{\mathbb{E}[v_t | X_t]}} dW_t. \quad (1.2)$$

Note that v in (1.2) can be any integrable and positive adapted stochastic process. In a sense, (1.2) may be considered as an inversion of the Markovian projection due to [14], applied to Dupire's local volatility model, i.e., (1.1) with $\sigma = \sigma_{\text{Dup}}$.

Thus the stochastic local volatility model of McKean–Vlasov type (1.2) solves the smile calibration problem. However, equation (1.2) is singular in a sense explained below and very hard to analyse and solve. Even the problem of proving existence or uniqueness for (1.2) (under various assumptions on v) turned out to be notoriously difficult and only a few results are available; we refer to Lackner et al. [17] for an

extensive discussion and literature review. Let us recall that the theory of standard McKean–Vlasov equations of the form

$$dZ_t = \tilde{H}(t, Z_t, \mu_t) dt + \tilde{F}(t, Z_t, \mu_t) dW_t$$

with $\mu_t = \text{Law}(Z_t)$ is well understood under appropriate regularity conditions, in particular, Lipschitz-continuity of \tilde{H} and \tilde{F} with respect to the standard Euclidean distances in the first two arguments and with respect to the Wasserstein distance in μ_t ; see Funaki [10], Carmona and Delarue [6, Chap. 4.2], Mishura and Veretennikov [20]. Denoting $Z_t := (X_t, Y_t)$, it is not difficult to see that the conditional expectation $(x, \mu_t) \mapsto \mathbb{E}[A(Y_t) | X_t = x]$ is in general not Lipschitz-continuous in the above sense. Therefore the standard theory does not apply to (1.2).

There are a number of results available in the literature where the Lipschitz condition on drift and diffusion is not imposed. Bossy and Jabir [3] considered singular McKean–Vlasov (MV) systems of the form:

$$dX_t = \mathbb{E}[\ell(X_t)|Y_t]dt + \mathbb{E}[\gamma(X_t)|Y_t]dW_t, \quad (1.3)$$

$$dY_t = b(X_t, Y_t)dt + \sigma(Y_t)dB_t, \quad (1.4)$$

or, alternatively, the seemingly even less regular equation

$$dX_t = \sigma(p(t, X_t))dW_t, \quad (1.5)$$

where $p(t, \cdot)$ denotes the density of X_t . Bossy and Jabir [3] establish well-posedness of (1.3)–(1.5) under suitable regularity conditions (in particular, ellipticity) based on energy estimates of the corresponding nonlinear PDEs. Interestingly, these techniques break down when the roles of X and Y are reversed in (1.3), (1.4), that is, when $\mathbb{E}[\gamma(X_t)|Y_t]$ is replaced by $\mathbb{E}[\gamma(Y_t)|X_t]$ in (1.3), and similarly for the drift term. Hence the results of [3] do not imply well-posedness of (1.2). Lacker et al. [17] studied the two-dimensional SDE

$$dX_t = b_1(X_t) \frac{h(Y_t)}{\mathbb{E}[h(Y_t)|X_t]} dt + \sigma_1(X_t) \frac{f(Y_t)}{\sqrt{\mathbb{E}[f^2(Y_t)|X_t]}} dW_t, \quad (1.6)$$

$$dY_t = b_2(Y_t) dt + \sigma_2(Y_t) dB_t, \quad (1.7)$$

where W and B are two independent one-dimensional Brownian motions. Clearly, this can be seen as a generalisation of (1.2) with a non-zero drift and with the process v chosen in a special way. The authors proved strong existence and uniqueness of solutions to (1.6), (1.7) in the *stationary* case. In particular, this implies strong conditions on b_1 and b_2 , but also requires the initial value (X_0, Y_0) to be random and to have the stationary distribution. Existence and uniqueness of solutions to (1.6), (1.7) in the general case (without the stationarity assumptions) remains open. Finally, let us mention the result of Jourdain and Zhou [16, Theorem 2.2] which established weak existence of the solution to (1.2) for the case when v is a jump process taking finitely many values.

Another question apart from well-posedness of these singular McKean–Vlasov equations is how to solve them numerically (in a certain sense). Let us recall that

even for standard SDEs with singular or irregular drift, where existence/uniqueness is known for quite some time, the convergence of the corresponding Euler scheme with non-vanishing rate has been established only very recently; see Butkovsky et al. [5], Jourdain and Menozzi [15]. The situation with the singular McKean–Vlasov equations presented above is much more complicated and very few results are available in the literature. In particular, the results of Lacker et al. [17] do not provide a way to construct a numerical algorithm for solving (1.2) even in the stationary case considered there.

In this paper, we study the problem of numerically solving singular McKean–Vlasov (MV) equations of a more general form than (1.2), namely

$$dX_t = H(t, X_t, Y_t, \mathbb{E}[A_1(Y_t)|X_t])dt + F(t, X_t, Y_t, \mathbb{E}[A_2(Y_t)|X_t])dW_t, \quad (1.8)$$

where H, F, A_1, A_2 are sufficiently regular functions, W is a d -dimensional Brownian motion and Y is a given stochastic process, for example a diffusion process. A key issue is how to approximate the conditional expectations $\mathbb{E}[A_i(Y_t)|X_t = x]$, $i = 1, 2$, $x \in \mathbb{R}^d$.

In their seminal paper, Guyon and Henry-Labordère [13] suggested an approach to tackle this problem (see also Antonelli and Kohatsu-Higa [1]). They used the “identity”

$$\mathbb{E}[A(Y_t)|X_t = x] \text{ “=” } \frac{\mathbb{E}[A(Y_t)\delta_x(X_t)]}{\mathbb{E}[\delta_x(X_t)]},$$

where δ_x is the Dirac delta function concentrated at x . This suggests the approximation

$$\mathbb{E}[A(Y_t)|X_t = x] \approx \frac{\sum_{i=1}^N A(Y_t^{i,N})k_\varepsilon(X_t^{i,N} - x)}{\sum_{i=1}^N k_\varepsilon(X_t^{i,N} - x)}. \quad (1.9)$$

Here $(X^{i,N}, Y^{i,N})_{i=1,\dots,N}$ is a particle system, $k_\varepsilon(\cdot) \approx \delta_0(\cdot)$ is a regularising kernel and $\varepsilon > 0$ is a small parameter. This technique for solving (1.8) (assuming (1.8) has a solution for a moment) works very well in practice, especially when coupled with interpolation on a grid in x -space. Due to the local nature of the performed regression, the method can be justified under only weak regularity assumptions on the conditional expectation. Note, however, that the interpolation part might require higher order regularity.

On the other hand, the method has an important disadvantage shared by all local regression methods: For any given point x , only points (X^i, Y^i) in a neighbourhood around x of size proportional to ε contribute to the estimate of $\mathbb{E}[A(Y_t)|X_t = x]$ as we have $k_\varepsilon(\cdot - x) \approx 0$ outside that neighbourhood. Hence local regression cannot take advantage of “global” information about the structure of the function $x \mapsto \mathbb{E}[A(Y_t)|X_t = x]$. If for example the conditional expectation can be globally approximated via a polynomial, it is highly inefficient (from a computational point of view) to approximate it locally using (1.9). Taken to the extreme, if we assume a compactly supported kernel k and formally take $\varepsilon = 0$, then the estimator (1.9) collapses to $\mathbb{E}[A(Y_t)|X_t = X_t^{i,N}] \approx A(Y_t^{i,N})$ since only $X_t^{i,N}$ is close enough to itself to

contribute to the estimator. In the context of the stochastic local volatility model (1.2), this means that the dynamics silently collapses to a pure local volatility dynamics if ε is chosen too small.

This disadvantage of local regression methods can be avoided by using *global regression* techniques. Indeed, taking advantage of global regularity and global structural features of the unknown target function, global regression methods are often seen to be more efficient than their local counterparts; see e.g. Bach [2]. On the other hand, the global regression methods require more regularity (e.g. global smoothness) than the minimal assumptions needed for local regression methods. In addition, the choice of basis functions can be crucial for global regression methods.

In fact, the starting point of this work was to replace (1.9) by global regression based on, say, L basis functions. However, it turns out that the Lipschitz constants of the resulting approximation to the conditional expectations in terms of the particle distribution explode as $L \rightarrow \infty$, unless the basis functions are carefully chosen.

As an alternative to Guyon and Henry-Labordère [13], we propose in this paper a novel approach based on ridge regression in the context of reproducing kernel Hilbert spaces (RKHSs) which in particular does not have either of the above mentioned disadvantages, even when the number of basis functions is infinite.

Recall that an RKHS \mathcal{H} is a Hilbert space of real-valued functions $f : \mathcal{X} \rightarrow \mathbb{R}$ such that the evaluation map $\mathcal{H} \ni f \mapsto f(x)$ is continuous for every $x \in \mathcal{X}$. This crucial property implies that there exists a positive symmetric kernel $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$, i.e., for any $c_1, \dots, c_n \in \mathbb{R}$, $x_1, \dots, x_n \in \mathcal{X}$, one has

$$\sum_{i,j=1}^n c_i c_j k(x_i, x_j) \geq 0,$$

such that $k_x := k(\cdot, x) \in \mathcal{H}$ for every $x \in \mathcal{X}$, and one has $\langle f, k_x \rangle_{\mathcal{H}} = f(x)$ for all $f \in \mathcal{H}$. As a main feature, any positive definite kernel k uniquely determines an RKHS \mathcal{H} and the other way around. In our setting, we consider $\mathcal{X} \subseteq \mathbb{R}^d$. For a detailed introduction and further properties of RKHSs, we refer to the literature, for example Steinwart and Christmann [25, Chap. 4]. We recall that the RKHS framework is popular in machine learning where it is widely used for computing conditional expectations. In the learning context, kernel methods are most prominently used in order to avoid the curse of dimensionality when dealing with high-dimensional features by the *kernel trick*. We stress that this issue is not relevant in the application to calibration of equity models, but it might be interesting for more general, high-dimensional singular McKean–Vlasov systems.

Consider a pair of random variables (X, Y) taking values in $\mathcal{X} \times \mathcal{X}$ with finite second moments and denote $\nu := \text{Law}(X, Y)$. Suppose that $A : \mathcal{X} \rightarrow \mathbb{R}$ is sufficiently regular and \mathcal{H} is large enough so that we have $\mathbb{E}[A(Y)|X = \cdot] \in \mathcal{H}$. Then formally,

$$\begin{aligned} c_A^\nu(\cdot) &:= \int_{\mathcal{X} \times \mathcal{X}} k(\cdot, x) A(y) \nu(dx, dy) = \int_{\mathcal{X}} k(\cdot, x) \nu(dx, \mathcal{X}) \int_{\mathcal{X}} A(y) \nu(dy|x) \\ &= \int_{\mathcal{X}} k(\cdot, x) \mathbb{E}[A(Y)|X = x] \nu(dx, \mathcal{X}) \\ &=: \mathcal{C}^\nu \mathbb{E}[A(Y)|X = \cdot], \end{aligned}$$

where

$$C^v f(\cdot) := \int_{\mathcal{X}} k(\cdot, x) f(x) \nu(dx, \mathcal{X}), \quad f \in \mathcal{H}.$$

Unfortunately, in general, the operator C^v is not invertible. As C^v is positive definite, it is, however, possible to *regularise* the inversion by replacing C^v by $C^v + \lambda I_{\mathcal{H}}$ for some $\lambda > 0$, where $I_{\mathcal{H}}$ is the identity operator on \mathcal{H} . Indeed, it turns out that

$$m_A^\lambda(\cdot; \nu) := (C^v + \lambda I_{\mathcal{H}})^{-1} c_A^v \quad (1.10)$$

is the solution to the minimisation problem

$$m_A^\lambda(\cdot; \nu) := \arg \min_{f \in \mathcal{H}} \left(\mathbb{E}[(A(Y) - f(X))^2] + \lambda \|f\|_{\mathcal{H}}^2 \right); \quad (1.11)$$

see Proposition 3.3. On the other hand, one also has

$$\mathbb{E}[A(Y)|X = \cdot] = \arg \min_{f \in L^2(\mathbb{R}^d, \text{Law}(X))} \mathbb{E}[(A(Y) - f(X))^2],$$

and therefore it is natural to expect that if $\lambda > 0$ is small enough and \mathcal{H} is large enough, then $m_A^\lambda(\cdot; \nu) \approx \mathbb{E}[A(Y)|X = \cdot]$, that is, $m_A^\lambda(\cdot; \nu)$ is close to the true conditional expectation.

The main result of the article is that the regularised MV system obtained by replacing the conditional expectations with their regularised versions (1.10) in (1.8) is well posed and propagation of chaos holds for the corresponding particle system; see Theorems 2.2 and 2.3. To establish these results, we study the joint regularity of $m_A^\lambda(x; \nu)$ in the space variable x and the measure ν for fixed $\lambda > 0$. Such results are almost absent in the literature on RKHSs and we here fill this gap. In particular, we prove that under suitable conditions, $m_A^\lambda(x; \nu)$ is Lipschitz in both arguments, that is, with respect to the standard Euclidean norm in x and the Wasserstein-1-norm in ν , and can be calculated numerically in an efficient way; see Sect. 2. Additionally, in Sect. 3, we study the convergence of $m_A^\lambda(\cdot; \nu)$ in (1.10) to the true conditional expectation for fixed ν as $\lambda \searrow 0$.

Let us note that as a further nice feature of the RKHS approach compared to the kernel method of [13], one may incorporate, at least in principle, global prior information concerning properties of $\mathbb{E}[A(Y)|X = \cdot]$ into the choice of the RKHS-generating kernel k . In a nutshell, if one anticipates beforehand that $\mathbb{E}[A(Y)|X = x] \approx f(x)$ for some known “nice” function f , one may pass to a new kernel given by setting $\tilde{k}(x, y) := k(x, y) + f(x)f(y)$. This degree of freedom is similar to, for example, the possibility of choosing basis functions in line with the problem under consideration in the usual regression methods for American options. We also note that the Lipschitz constants for $m_A^\lambda(x; \nu)$ with respect to both arguments are expressed in bounds related to A and the kernel k only; see Theorem 2.4. In contrast, if we had dealt with standard ridge regression, that is, ridge regression based on a fixed system of basis functions, we should have to impose restrictions on the regression coefficients leading to a nonconvex constrained optimisation problem.

In summary, the contribution of the current work is fourfold. First, we propose an RKHS-based approach to regularise (1.8) and prove the well-posedness of the regularised equation. Second, we show convergence of the approximation (1.11) to the true conditional expectation as $\lambda \searrow 0$. Third, we suggest a particle-based approximation of the regularised equation and analyse its convergence. Finally, we apply our algorithm to the problem of smile calibration in finance and illustrate its performance on simulated data. In particular, we validate our results by solving numerically a regularised version of (1.2) (with m_A^λ in place of the conditional expectation). We show that our system is indeed an approximate solution to (1.2) in the sense that we get very close fits of the implied volatility surface – the final goal of the smile calibration problem.

The rest of the paper is organised as follows. Our main theoretical results are given in Sect. 2. Convergence properties of the regularised conditional expectation m_A^λ are established in Sect. 3. A numerical algorithm for solving (1.8) and an efficient implementable approximation of m_A^λ are discussed in Sect. 4. Section 5 contains numerical examples. The results of the paper are summarised in Sect. 6. Finally, all the proofs are placed in Sect. 7.

Convention on constants. Throughout the paper, C denotes a positive constant whose value may change from line to line. The dependence of constants on parameters, if needed, will be indicated e.g. by $C(\lambda)$.

2 Main results

We begin by introducing the basic notation. For $a \in \mathbb{R}$, we set $a^+ := \max(a, 0)$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. For $d \in \mathbb{N}$, let $\mathcal{X} \subseteq \mathbb{R}^d$ be an open subset and $\mathcal{P}_2(\mathcal{X})$ the set of all probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ with finite second moment. If $\mu, \nu \in \mathcal{P}_2(\mathcal{X})$, $p \in [1, 2]$, we denote the *Wasserstein- p* (Kantorovich) distance between them by

$$\mathbb{W}_p(\mu, \nu) := \inf(\mathbb{E}[|X - Y|^p])^{1/p},$$

where the infimum is taken over all random variables X, Y with $\text{Law}(X) = \mu$ and $\text{Law}(Y) = \nu$. Let $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ be a symmetric, positive definite kernel and \mathcal{H} a reproducing kernel Hilbert space of functions $f : \mathcal{X} \rightarrow \mathbb{R}$ associated with the kernel k . That is, for any $x \in \mathcal{X}$, $f \in \mathcal{H}$, one has

$$f(x) = \langle f, k(x, \cdot) \rangle_{\mathcal{H}}.$$

In particular, $\langle k(x, \cdot), k(y, \cdot) \rangle_{\mathcal{H}} = k(x, y)$ for any $x, y \in \mathcal{X}$. We refer to Steinwart and Christmann [25, Chap. 4] for further properties of RKHSs.

Let $A : \mathcal{X} \rightarrow \mathbb{R}$ be a measurable function such that $|A(x)| \leq C(1 + |x|)$ for some universal constant $C > 0$ and all $x \in \mathcal{X}$. For $\nu \in \mathcal{P}_2(\mathcal{X} \times \mathcal{X})$, $\lambda \geq 0$, consider the optimisation problem (*ridge regression*)

$$m_A^\lambda(\cdot; \nu) := \arg \min_{f \in \mathcal{H}} \left(\int_{\mathcal{X} \times \mathcal{X}} |A(y) - f(x)|^2 \nu(dx, dy) + \lambda \|f\|_{\mathcal{H}}^2 \right). \quad (2.1)$$

We fix $T > 0$, $d \in \mathbb{N}$ and consider the system

$$dX_t = H(t, X_t, Y_t, \mathbb{E}[A_1(Y_t)|X_t])dt + F(t, X_t, Y_t, \mathbb{E}[A_2(Y_t)|X_t])dW_t^X, \quad (2.2)$$

$$dY_t = b(t, Y_t)dt + \sigma(t, Y_t)dW_t^Y, \quad (2.3)$$

where $H: [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}^d$, $F: [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}^d \times \mathbb{R}^d$, $A_i: \mathbb{R}^d \rightarrow \mathbb{R}$, $b: [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$, $\sigma: [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d \times \mathbb{R}^d$ are measurable functions, W^X, W^Y are two (possibly correlated) d -dimensional Brownian motions on $(\Omega, \mathcal{F}, \mathbb{P})$, and $t \in [0, T]$. We note that our choice of Y as a diffusion process in (2.3) is mostly for convenience, and we expect our results to hold in more generality when appropriately modified.

Denote $\mu_t := \text{Law}(X_t, Y_t)$. As mentioned above, the functional

$$(x, \mu_t) \mapsto \mathbb{E}[A_i(Y_t)|X_t = x]$$

is not Lipschitz-continuous even if A_i is smooth. Therefore the classical results on well-posedness of McKean–Vlasov equations are not applicable to (2.2), (2.3). The main idea of our approach is to replace the conditional expectation by the corresponding RKHS approximation (2.1) which has “nice” properties (in particular, it is Lipschitz-continuous). This implies strong existence and uniqueness of the new system. Furthermore, we demonstrate numerically that the solution to the new system is still “close” to the solution of (2.2), (2.3) in a certain sense. Thus we consider the system

$$d\widehat{X}_t = H(t, \widehat{X}_t, Y_t, m_{A_1}^\lambda(\widehat{X}_t; \widehat{\mu}_t))dt + F(t, \widehat{X}_t, Y_t, m_{A_2}^\lambda(\widehat{X}_t; \widehat{\mu}_t))dW_t^X, \quad (2.4)$$

$$dY_t = b(t, Y_t)dt + \sigma(t, Y_t)dW_t^Y, \quad (2.5)$$

$$\widehat{\mu}_t = \text{Law}(\widehat{X}_t, Y_t), \quad (2.6)$$

where $t \in [0, T]$. We need the following assumptions on the kernel k (formulated in a slightly redundant manner for ease of notation).

Assumption 2.1 The kernel k is twice continuously differentiable in both variables, $k(x, x) > 0$ for all $x \in \mathcal{X}$, and

$$D_k^2 := \sup_{\substack{(x, y) \in \mathcal{X} \times \mathcal{X} \\ 1 \leq i, j \leq d}} \max\{|\partial_{x_i} \partial_{y_j} k^2(x, y)|, |\partial_{x_i} \partial_{y_j} k(x, y)|, |\partial_{x_i} k(x, y)|, \\ |\partial_{y_j} k(x, y)|, |k(x, y)|\} < \infty.$$

Let $\mathcal{C}^1(\mathcal{X}; \mathbb{R})$ be the space of all functions $f: \mathcal{X} \rightarrow \mathbb{R}$ such that

$$\|f\|_{\mathcal{C}^1} := \sup_{x \in \mathcal{X}} |f(x)| + \sup_{\substack{x \in \mathcal{X} \\ i=1, \dots, d}} |\partial_{x_i} f(x)| < \infty.$$

Now we are ready to state our main results. Their proofs are given in Sect. 7.

Theorem 2.2 Suppose that Assumption 2.1 is satisfied for the kernel k with $\mathcal{X} = \mathbb{R}^d$ and

- 1) $A_i \in C^1(\mathbb{R}^d; \mathbb{R})$, $i = 1, 2$;
- 2) there exists a constant $C > 0$ such that for any $t \in [0, T]$, $x, y, x', y' \in \mathbb{R}^d$, $z, z' \in \mathbb{R}$, we have

$$\begin{aligned} & |H(t, x, y, z) - H(t, x', y', z')| + |F(t, x, y, z) - F(t, x', y', z')| \\ & \quad + |b(t, y) - b(t, y')| + |\sigma(t, y) - \sigma(t, y')| \\ & \leq C(|x - x'| + |y - y'| + |z - z'|); \end{aligned}$$

- 3) for any fixed $x, y \in \mathbb{R}^d$, $z \in \mathbb{R}$, we have

$$\int_0^T (|H(t, x, y, z)|^2 + |F(t, x, y, z)|^2 + |b(t, y)|^2 + |\sigma(t, y)|^2) dt < \infty;$$

- 4) $\mathbb{E}[|\widehat{X}_0|^2] + \mathbb{E}[|Y_0|^2] < \infty$.

Then for any $\lambda > 0$, the system (2.4)–(2.6) with the initial condition (\widehat{X}_0, Y_0) has a unique strong solution.

To analyse a numerical scheme solving (2.4)–(2.6), we consider the particle system

$$\begin{aligned} dX_t^{N,n} &= H(t, X_t^{N,n}, Y_t^{N,n}, m_{A_1}^\lambda(X_t^{N,n}; \mu_t^N))dt \\ & \quad + F(t, X_t^{N,n}, Y_t^{N,n}, m_{A_2}^\lambda(X_t^{N,n}; \mu_t^N))dW_t^{X,n}, \end{aligned} \quad (2.7)$$

$$dY_t^{N,n} = b(t, Y_t^{N,n})dt + \sigma(t, Y_t^{N,n})dW_t^{Y,n}, \quad (2.8)$$

$$\mu_t^N = \frac{1}{N} \sum_{n=1}^N \delta_{(X_t^{N,n}, Y_t^{N,n})}, \quad (2.9)$$

where $N \in \mathbb{N}$, $n = 1, \dots, N$, $t \in [0, T]$ and the pairs of d -dimensional Brownian motions $(W^{X,n}, W^{Y,n})$, $n = 1, \dots, N$, are independent and have the same law as (W^X, W^Y) . The following propagation of chaos result holds; it establishes both weak and strong convergence of $X^{N,n}$.

Theorem 2.3 Suppose that all the conditions of Theorem 2.2 are satisfied. Suppose the initial values $(X_0^{N,n}, Y_0^{N,n})$ are independent and have the same law as (\widehat{X}_0, Y_0) . Moreover, suppose that $\mathbb{E}[|\widehat{X}_0|^q] + \mathbb{E}[|Y_0|^q] < \infty$ for some $q > 4$. Then there exists a constant $C = C(\lambda, T, \mathbb{E}[|\widehat{X}_0|^q], \mathbb{E}[|Y_0|^q]) > 0$ such that for any $n = 1, \dots, N$, $N \in \mathbb{N}$, we have

$$\mathbb{E} \left[\sup_{0 \leq t \leq T} |X_t^{N,n} - \widehat{X}_t^n|^2 \right] + \sup_{0 \leq t \leq T} \mathbb{E}[(\mathbb{W}_2(\mu_t^N, \widehat{\mu}_t))^2] \leq C \epsilon_N, \quad (2.10)$$

where the process \widehat{X}^n solves (2.4)–(2.6) with $W^{X,n}$, $W^{Y,n}$ in place of W^X , W^Y , respectively, and where

$$\epsilon_N = \begin{cases} N^{-1/2} & \text{if } d = 1, \\ N^{-1/2} \log N & \text{if } d = 2, \\ N^{-1/d} & \text{if } d > 2. \end{cases}$$

A crucial step which allows us to obtain these results is the Lipschitz-continuity of m^λ . The following holds.

Theorem 2.4 Assume that the kernel k satisfies Assumption 2.1. Let $A \in C^1(\mathcal{X}; \mathbb{R})$. Then for any $x, y \in \mathcal{X}$, $\mu, \nu \in \mathcal{P}_2(\mathcal{X} \times \mathcal{X})$, we have

$$|m_A^\lambda(x; \mu) - m_A^\lambda(y; \nu)| \leq C_1 \mathbb{W}_1(\mu, \nu) + C_2 |x - y|,$$

where

$$C_1 := \left(\frac{D_k}{\lambda^2} + \frac{1}{\lambda} \right) d D_k^2 \|A\|_{C^1} \quad \text{and} \quad C_2 := \frac{\sqrt{d}}{\lambda} D_k^2 \|A\|_{C^1}$$

may be considered to be (possibly suboptimal) Lipschitz constants with respect to the Wasserstein metric and Euclidean norm, respectively.

This result is interesting for at least two reasons. First, it shows that m_A^λ is Lipschitz-continuous in both arguments, provided that the kernel k is smooth enough. That is, the Lipschitz-continuity property depends on \mathcal{H} only through the smoothness of the kernel k . Second, this result gives an explicit dependence of the corresponding (possibly suboptimal) Lipschitz constants on λ and k .

Remark 2.5 Let us stress that Theorem 2.2 establishes the existence and uniqueness of a solution to (2.2), (2.3) only for a fixed regularisation parameter $\lambda > 0$ and cannot be used to study the limiting case $\lambda \rightarrow 0$. Indeed, it follows from Theorem 2.4 that as $\lambda \rightarrow 0$, the Lipschitz constants of m_A^λ blow up. However, Theorem 2.3 does not imply that the optimal Lipschitz constants blow up for $\lambda \rightarrow 0$, or that the solution to (2.2), (2.3) blows up. We demonstrate numerically in Sect. 5 that for $\lambda \rightarrow 0$, in the examples there, the solution to (2.2), (2.3) does not blow up. On the contrary, it weakly converges to a limit; this suggests that (at least) weak existence of a solution to (2.2), (2.3) may hold. Verifying this theoretically remains however an important open problem.

Remark 2.6 A natural question is whether (2.2), (2.3) can be formulated for a different state space, that is, for X, Y taking values in \mathcal{X}, \mathcal{Y} rather than \mathbb{R}^d . Indeed, for equity models, $\mathcal{X} = \mathcal{Y} = \mathbb{R}_+$ is clearly a more natural choice for both the price process and the variance process. Heuristically, the theory should hold for more general \mathcal{X} and \mathcal{Y} , provided that those sets are invariant under the dynamics (2.2), (2.3) as well as under the regularised dynamics. It is, however, difficult to derive meaningful assumptions guaranteeing this kind of invariance, which prompts us to work with \mathbb{R}^d instead.

3 Approximation of conditional expectations

In this section, we study the approximation m_A^λ introduced in (2.1) in more detail. **Throughout this section**, we fix an open set $\mathcal{X} \subseteq \mathbb{R}^d$ and a measure $\nu \in \mathcal{P}_2(\mathcal{X} \times \mathcal{X})$, and impose the following relatively weak assumptions on the function $A: \mathcal{X} \rightarrow \mathbb{R}$ and the positive kernel $k: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$.

Assumption 3.1 The function A has sublinear growth, i.e., there exists a constant $C > 0$ such that $|A(x)| \leq C(1 + |x|)$ for all $x \in \mathcal{X}$.

Assumption 3.2 The kernel $k(\cdot, \cdot)$ is continuous on $\mathcal{X} \times \mathcal{X}$ and satisfies

$$0 < k(x, x) \leq C(1 + |x|^2)$$

for some $C > 0$.

It is easy to see that Assumption 3.2 implies for any $x \in \mathcal{X}$ that

$$\|k(x, \cdot)\|_{\mathcal{H}}^2 = \langle k(x, \cdot), k(x, \cdot) \rangle_{\mathcal{H}} = k(x, x) \leq C(1 + |x|^2). \quad (3.1)$$

Due to Assumption 3.2 and Steinwart and Christmann [25, Lemma 4.33], \mathcal{H} is a separable RKHS and one has for any $f \in \mathcal{H}$, $x \in \mathcal{X}$ that

$$|f(x)| = |\langle k(x, \cdot), f \rangle_{\mathcal{H}}| \leq \|k(x, \cdot)\|_{\mathcal{H}} \|f\|_{\mathcal{H}} \leq C(1 + |x|) \|f\|_{\mathcal{H}}, \quad (3.2)$$

where we also used (3.1). Hence every $f \in \mathcal{H}$ has sublinear growth and, as a consequence, the objective functional in (2.1) is finite for any fixed $\nu \in \mathcal{P}_2(\mathcal{X} \times \mathcal{X})$. It is also easy to see that (3.2) and (3.1) imply that for any $x, y \in \mathcal{X}$,

$$|k(x, y)| \leq C(1 + |x|) \|k(\cdot, y)\|_{\mathcal{H}} \leq C(1 + |x|)(1 + |y|). \quad (3.3)$$

Therefore, the Bochner integrals

$$c_A^\nu := \int_{\mathcal{X} \times \mathcal{X}} k(\cdot, x) A(y) \nu(dx, dy), \quad C^\nu f := \int_{\mathcal{X} \times \mathcal{X}} k(\cdot, x) f(x) \nu(dx, dy)$$

are well-defined functions in \mathcal{H} for every $f \in \mathcal{H}$. Moreover, it is clear that the operator $C^\nu: \mathcal{H} \rightarrow \mathcal{H}$ is symmetric and positive semidefinite since

$$\langle g, C^\nu f \rangle_{\mathcal{H}} = \int_{\mathcal{X}} \langle g, k(\cdot, x) \rangle f(x) \nu(dx, \mathcal{X}) = \int_{\mathcal{X}} g(x) f(x) \nu(dx, \mathcal{X}).$$

Thus by the Hellinger–Toeplitz theorem (see e.g. Reed and Simon [21, Sect. III.5]), C^ν is a bounded self-adjoint linear operator on \mathcal{H} . As a consequence, for any $\lambda \geq 0$, the operator $C^\nu + \lambda I_{\mathcal{H}}$ is a bounded self-adjoint operator on \mathcal{H} with spectrum contained in the interval $[\lambda, \|C^\nu\| + \lambda]$. Hence if $\lambda > 0$, then $(C^\nu + \lambda I_{\mathcal{H}})^{-1}$ exists and is a bounded self-adjoint operator on \mathcal{H} with norm

$$\|(C^\nu + \lambda I_{\mathcal{H}})^{-1}\|_{\mathcal{H}} \leq \lambda^{-1}.$$

We are now ready to state the following useful representation for the solution to (2.1).

Proposition 3.3 *Under Assumptions 3.1, 3.2, for any fixed $v \in \mathcal{P}_2(\mathcal{X} \times \mathcal{X})$ and $\lambda > 0$, the solution to (2.1) can be represented as*

$$m_A^\lambda(\cdot; v) = (C^v + \lambda I_{\mathcal{H}})^{-1} c_A^v. \quad (3.4)$$

This representation may be seen as an infinite sample version of the usual solution representation for a ridge regression problem based on finite samples. We thus consider it as not essentially new, but in order to keep our paper as self-contained as possible, we present a proof in Sect. 7. Proposition 3.3 allows us to prove Lipschitz-continuity of m_A^λ , that is, Theorem 2.4.

Let us now proceed with investigating when the function $m_A^\lambda = m_A^\lambda(\cdot; v)$ is a “good” approximation to the true conditional expectation

$$m_A = m_A(x; v) := \mathbb{E}_{(X,Y) \sim v}[A(Y)|X = x] \quad (3.5)$$

for small enough $\lambda > 0$. Consider the Hilbert space $\mathcal{L}_2^v := L_2(\mathcal{X}, v(dx, \mathcal{X}))$ with $v(U, \mathcal{X}) := v(U \times \mathcal{X}) > 0$. For $f \in \mathcal{L}_2^v$, put

$$T^v f(\cdot) := \int_{\mathcal{X}} k(\cdot, x) f(x) v(dx, \mathcal{X}). \quad (3.6)$$

Recalling (3.3), it is easy to see that T^v is a linear operator $\mathcal{L}_2^v \rightarrow \mathcal{L}_2^v$. Note that $\mathcal{H} \subseteq \mathcal{L}_2^v$ due to (3.2); thus C^v is the restriction of T^v to \mathcal{H} . Further, since

$$|k(x, y)| \leq \sqrt{k(x, x)} \sqrt{k(y, y)},$$

the kernel k is Hilbert–Schmidt on $\mathcal{L}_2(\mathcal{X} \times \mathcal{X}, v(dx, \mathcal{X}) \otimes v(dy, \mathcal{X}))$, i.e.,

$$\int k^2(x, y) v(dx, \mathcal{X}) v(dy, \mathcal{X}) < \infty$$

due to Assumption 3.2. As a consequence of standard results from functional analysis, one then has (see for example [21, Sect. VI]) that

- (i) the operator T^v is self-adjoint and compact;
- (ii) there exists an orthonormal system $(a_n)_{n \in \mathbb{N}}$ in \mathcal{L}_2^v of eigenfunctions corresponding to nonnegative eigenvalues σ_n of T^v , and $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \dots$;
- (iii) if $J := \{n \in \mathbb{N} : \sigma_n > 0\}$, one has

$$T^v f = \sum_{n \in J} \sigma_n \langle f, a_n \rangle_{\mathcal{L}_2^v} a_n, \quad f \in \mathcal{L}_2^v, \quad (3.7)$$

with $\lim_{n \rightarrow \infty} \sigma_n = 0$ if $J = \mathbb{N}$.

A generalisation of Mercer’s theorem to unbounded domains, see Sun [26], implies the following statement.

Proposition 3.4 *Let k be a kernel satisfying Assumption 3.2 and assume that $v(\cdot, \mathcal{X})$ is a nondegenerate Borel measure, that is, for every open set $U \subseteq \mathcal{X}$, one has*

$\nu(U, \mathcal{X}) > 0$. Then one may take the eigenfunctions a_n in (3.7) to be continuous and k has a series representation

$$k(x, y) = \sum_{n \in J} \sigma_n a_n(x) a_n(y), \quad x, y \in \mathcal{X},$$

with uniform convergence on compact sets. Moreover, $(\tilde{a}_n)_{n \in J}$ with $\tilde{a}_n := \sqrt{\sigma_n} a_n$ is an orthonormal basis of \mathcal{H} , and the scalar product in \mathcal{H} takes the form

$$\langle f, g \rangle_{\mathcal{H}} = \sum_{n \in J} \frac{\langle f, a_n \rangle_{\mathcal{L}_2^\nu} \langle g, a_n \rangle_{\mathcal{L}_2^\nu}}{\sigma_n} \quad \text{for } f, g \in \mathcal{H}. \quad (3.8)$$

Now we are ready to present the main result of this section, which quantifies the convergence properties of $m_A^\lambda(\cdot, \nu)$ as $\lambda \rightarrow 0$ for a fixed measure ν . Recall the notation (3.5). Let $P_{\overline{\mathcal{H}}}$ denote the orthogonal projection in \mathcal{L}_2^ν onto $\overline{\mathcal{H}}$, the closure of \mathcal{H} in \mathcal{L}_2^ν . Then for any $f \in \mathcal{L}_2^\nu$,

$$P_{\overline{\mathcal{H}}} f = \sum_{n \in J} \langle f, a_n \rangle_{\mathcal{L}_2^\nu} a_n \quad \text{and} \quad \langle P_{\overline{\mathcal{H}}} f, a_m \rangle_{\mathcal{L}_2^\nu} = \langle f, a_m \rangle_{\mathcal{L}_2^\nu}, \quad m \in J, \quad (3.9)$$

since $(a_n)_{n \in J}$ is an orthonormal system in \mathcal{L}_2^ν .

Theorem 3.5 Assume that the kernel k satisfies Assumption 3.2, $\nu(\cdot, \mathcal{X})$ is a nondegenerate Borel measure, and that $m_A(\cdot; \nu) \in \mathcal{L}_2^\nu$ (for instance when A is bounded and measurable). Then for any $\lambda > 0$,

$$\|P_{\overline{\mathcal{H}}} m_A(\cdot; \nu) - m_A^\lambda(\cdot; \nu)\|_{\mathcal{L}_2^\nu}^2 = \sum_{n \in J} \frac{\lambda^2}{(\sigma_n + \lambda)^2} \langle m_A(\cdot; \nu), a_n \rangle_{\mathcal{L}_2^\nu}^2. \quad (3.10)$$

In particular, $\|P_{\overline{\mathcal{H}}} m_A(\cdot; \nu) - m_A^\lambda(\cdot; \nu)\|_{\mathcal{L}_2^\nu} \rightarrow 0$ as $\lambda \searrow 0$. If we have in addition $P_{\overline{\mathcal{H}}} m_A(\cdot; \nu) \in \mathcal{H}$, then

$$\|P_{\overline{\mathcal{H}}} m_A(\cdot; \nu) - m_A^\lambda(\cdot; \nu)\|_{\mathcal{H}}^2 = \sum_{n \in J} \frac{\lambda^2}{(\sigma_n + \lambda)^2 \sigma_n} \langle m_A(\cdot; \nu), a_n \rangle_{\mathcal{L}_2^\nu}^2, \quad (3.11)$$

and thus $\|P_{\overline{\mathcal{H}}} m_A(\cdot; \nu) - m_A^\lambda(\cdot; \nu)\|_{\mathcal{H}} \rightarrow 0$ for $\lambda \searrow 0$.

Theorem 3.5 establishes convergence of $m_A^\lambda(\cdot; \nu)$ as $\lambda \rightarrow 0$, but without a rate. Its proof is placed in Sect. 7. Additional assumptions are needed to guarantee a convergence rate. This is done in the following corollary.

Corollary 3.6 Suppose that the conditions of Theorem 3.5 are satisfied and that, moreover, for some $\theta \in (0, 1]$, we have

$$\sum_{n \in J} \sigma_n^{-\theta} \langle m_A(\cdot; \nu), a_n \rangle_{\mathcal{L}_2^\nu}^2 < \infty. \quad (3.12)$$

Then

$$\begin{aligned} & \|P_{\overline{\mathcal{H}}}m_A(\cdot; v) - m_A^\lambda(\cdot; v)\|_{\mathcal{L}_2^v}^2 \\ & \leq \left(1 - \frac{\theta}{2}\right)^2 \left(\frac{\lambda\theta}{2-\theta}\right)^\theta \sum_{n \in J} \sigma_n^{-\theta} \langle m_A(\cdot; v), a_n \rangle_{\mathcal{L}_2^v}^2. \end{aligned} \quad (3.13)$$

In particular, if $\theta = 1$, then $P_{\overline{\mathcal{H}}}m_A \in \mathcal{H}$ and we get

$$\|P_{\overline{\mathcal{H}}}m_A(\cdot; v) - m_A^\lambda(\cdot; v)\|_{\mathcal{L}_2^v} \leq \frac{\sqrt{\lambda}}{2} \|P_{\overline{\mathcal{H}}}m_A(\cdot; v)\|_{\mathcal{H}}. \quad (3.14)$$

Proof Inequality (3.13) follows from (3.10), (3.12) and the fact that the maximum of the function $x \mapsto \lambda^2 x^\theta / (x + \lambda)^2$, $x > 0$, is equal to $(1 - \theta/2)^2 (\lambda\theta / (2 - \theta))^\theta$. Inequality (3.14) follows from (3.8), (3.9) and (3.13). \square

Remark 3.7 If the operator T^v defined in (3.6) is injective, that is, $T^v f = 0$ for $f \in \mathcal{L}_2^v$ implies $f = 0$ v -a.s., then $P_{\overline{\mathcal{H}}} = I_{\mathcal{L}_2^v}$. In this case, $J = \mathbb{N}$ and Theorem 3.5 and Corollary 3.6 quantify the convergence to the true conditional expectation. A sufficient condition for T^v to be injective is that the kernel k is *integrally strictly positive definite (ispd)* in the sense that

$$\int_{\mathcal{X} \times \mathcal{X}} k(x, y) \mu(dx) \mu(dy) > 0$$

for all non-zero signed Borel measures μ defined on \mathcal{X} . Indeed, for any $f \in \mathcal{L}_2^v$, we may define a signed Borel measure $\mu_f(A) := \int_A f(x) v(dx, \mathcal{X})$, $A \in \mathcal{B}(\mathcal{X})$, which is finite since $|\mu_f(A)| \leq \int f^2(x) v(dx, \mathcal{X}) < \infty$. Hence if k is an ispd kernel, then $T^v f = 0$ implies

$$\begin{aligned} 0 &= \langle T^v f, f \rangle_{\mathcal{L}_2^v} = \int_{\mathcal{X} \times \mathcal{X}} k(y, x) f(x) f(y) v(dx, \mathcal{X}) v(dy, \mathcal{X}) \\ &= \int_{\mathcal{X} \times \mathcal{X}} k(y, x) \mu_f(dx) \mu_f(dy), \end{aligned}$$

which in turn implies $\mu_f = 0$, i.e., $f = 0$ v -a.s. Furthermore, it should be noted that any ispd kernel is strictly positive definite in the usual sense, but the converse is not true. Examples of ispd kernels are Gaussian kernels, Laplace kernels and many more. For details on ispd kernels, we refer to Sriperumbudur et al. [24].

In summary, we have shown in this section that under certain conditions, $m_A^\lambda(\cdot, v)$ may converge at least in the \mathcal{L}_2^v -sense to the true conditional expectation $m_A(\cdot, v)$ as $\lambda \rightarrow 0$. This makes the heuristic discussion around (1.10) and (1.11) in Sect. 1 more rigorous.

Remark 3.8 Note that the measure $\widehat{\mu}_t$ in the solution of (2.4)–(2.6) depends on λ so that in fact $\widehat{\mu}_t = \widehat{\mu}_t^\lambda$. Therefore, even when $m_A^\lambda(\cdot, v) \rightarrow m_A(\cdot, v)$ for fixed

ν and $\lambda \downarrow 0$, the question whether $m_{A_i}^\lambda(\cdot, \widehat{\mu}_t^\lambda)$ converges in some sense is still not answered. We believe that this question is intimately linked to the problem of existence of a solution to (2.2), (2.3). As already explained, this is an unsolved open problem and considered beyond our scope here. However, loosely speaking, assuming that the latter system has indeed a solution (in some sense) with solution measure μ_t , say, it is natural to expect that for a suitable “rich enough” RKHS, we obtain $m_{A_i}^\lambda(\cdot, \mu_t) \rightarrow m_{A_i}(\cdot, \mu_t)$ (the true conditional expectation) as $\lambda \searrow 0$.

4 Numerical algorithm

Let us now describe in detail our numerical algorithm to construct solutions to (1.8). We begin by discussing an efficient way of calculating m_A^λ .

4.1 Estimation of the conditional expectation

Let us recall that in order to solve the particle system (2.7)–(2.9), we need to compute

$$m_A^\lambda(\cdot; \mu_t^N) = \arg \min_{f \in \mathcal{H}} \left(\frac{1}{N} \sum_{n=1}^N |A(Y_t^{N,n}) - f(X_t^{N,n})|^2 + \lambda \|f\|_{\mathcal{H}}^2 \right) \quad (4.1)$$

for t belonging to a certain partition of $[0, T]$ and fixed large $N \in \mathbb{N}$; here $A = A_1$ or $A = A_2$. It follows from the representer theorem for RKHSs in Schölkopf et al. [23, Theorem 1] that m_A^λ has the representation

$$m_A^\lambda(\cdot; \mu_t^N) = \sum_{i=1}^N \alpha_i k(X_t^{N,i}, \cdot) \quad (4.2)$$

for some $\alpha = (\alpha_1, \dots, \alpha_N)^\top \in \mathbb{R}^N$. Note that the optimal α can be calculated explicitly by plugging the representation (4.2) into the minimisation problem (4.1) in place of f and minimising over α . However, computing the optimal α directly takes $O(N^3)$ operations, which is prohibitively expensive keeping in mind that the number of particles N is going to be very large. Furthermore, even evaluating (4.2) at $X_t^{N,n}$, $n = 1, \dots, N$, for a given $\alpha \in \mathbb{R}^N$ is rather expensive; it requires $O(N^2)$ operations and thus is impossible to implement.

To develop an efficient algorithm, let us note that many particles $X_t^{N,i}$ – and as a consequence the implied basis functions $k(X_t^{N,i}, \cdot)$ – will be close to each other. Therefore we can considerably reduce the computational cost by only using $L \ll N$ rather than N basis functions as suggested in (4.2). More precisely, we choose Z^1, \dots, Z^L among $X_t^{N,1}, \dots, X_t^{N,N}$ – e.g. by random choice or by taking every $\frac{N}{L}$ th point among the ordered sequence $X_t^{N,(1)}, \dots, X_t^{N,(N)}$ when X is one-dimensional – and approximate

$$\sum_{i=1}^N \alpha_i k(X_t^{N,i}, \cdot) \approx \sum_{j=1}^L \beta_j k(Z^j, \cdot),$$

where $\beta = (\beta_1, \dots, \beta_L)^\top \in \mathbb{R}^L$. It is easy to see that

$$\begin{aligned} \left\| \sum_{j=1}^L \beta_j k(Z^j, \cdot) \right\|_{\mathcal{H}}^2 &= \left\langle \sum_{j=1}^L \beta_j k(Z^j, \cdot), \sum_{j=1}^L \beta_j k(Z^j, \cdot) \right\rangle_{\mathcal{H}} \\ &= \sum_{j,\ell=1}^L \beta_j \beta_\ell \langle k(Z^j, \cdot), k(Z^\ell, \cdot) \rangle_{\mathcal{H}} \\ &= \sum_{j,\ell=1}^L \beta_j \beta_\ell k(Z^j, Z^\ell) = \beta^\top R \beta, \end{aligned}$$

where $R := (k(Z^j, Z^\ell))_{j,\ell=1,\dots,L}$ is an $L \times L$ matrix. Thus recalling (4.1), we see that we have to solve

$$\arg \min_{\beta \in \mathbb{R}^L} \left(\frac{1}{N} (G - K\beta)^\top (G - K\beta) + \lambda \beta^\top R \beta \right),$$

where $G := (A(Y_t^{N,n}))_{n=1,\dots,N}$ and $K := (k(Z^j, X_t^{N,n}))_{n=1,\dots,N, j=1,\dots,L}$ is an $(N \times L)$ -matrix. Differentiating with respect to β , we obtain that the optimal value $\hat{\beta} = \hat{\beta}((X_t^N), (Y_t^N))$ satisfies

$$(K^\top K + N\lambda R) \hat{\beta} = K^\top G, \quad (4.3)$$

and we approximate the expectation as

$$m_A^\lambda(x; \mu_t^N) \approx \sum_{j=1}^L \hat{\beta}_j k(Z^j, x) =: \hat{m}_A^\lambda(x; \mu_t^N). \quad (4.4)$$

Remark 4.1 The method of choosing basis points Z^1, \dots, Z^L can be seen as a systematic and adaptive approach of choosing basis functions $k(Z^j, \cdot)$, $j = 1, \dots, L$, in a global regression method. We note that the technique of evaluating the conditional expectation only in points on a grid $G_{f,t}$ coupled with spline-type interpolation between grid points suggested in Guyon and Henry-Labordère [13] is motivated by similar concerns regarding the explosion of computational costs.

Remark 4.2 Let us see how many operations we need to calculate $\hat{\beta}$, taking into account that $L \ll N$. We need $O(NL)$ to calculate K , $O(L^2)$ to calculate R , $O(NL^2)$ to calculate $K^\top K$ (this is the bottleneck), $O(L^3)$ to invert $K^\top K + N\lambda R$ and $O(NL)$ to calculate $K^\top G$ and solve (4.3). Thus in total, we need $O(NL^2)$ operations.

4.2 Solving the regularised McKean–Vlasov equation

With the function \hat{m}_A^λ in hand, we now consider the Euler scheme for the particle system (2.7)–(2.9). We fix a time interval $[0, T]$, the number M of time steps and for simplicity consider a uniform time increment $\delta := T/M$. Let $\Delta W_i^{X,n}$ and $\Delta W_i^{Y,n}$

denote independent copies of $W_{(i+1)\delta}^X - W_{i\delta}^X$ and $W_{(i+1)\delta}^Y - W_{i\delta}^Y$, respectively, for $n = 1, \dots, N, i = 1, \dots, M$. Note that for stochastic volatility models, the Brownian motions driving the stock price and the variance process are usually correlated. We now define $\tilde{X}_0^n = X_0^n, \tilde{Y}_0^n = Y_0^n$ and for $i = 0, \dots, M - 1$,

$$\begin{aligned}\tilde{X}_{i+1}^n &= \tilde{X}_i^n + H(i\delta, \tilde{X}_i^n, \tilde{Y}_i^n, \hat{m}_{A_1}^\lambda(\tilde{X}_i^n; \tilde{\mu}_i^N))\delta \\ &\quad + F(i\delta, \tilde{X}_i^n, \tilde{Y}_i^n, \hat{m}_{A_2}^\lambda(\tilde{X}_i^n; \tilde{\mu}_i^N))\Delta W_i^{X,n},\end{aligned}\quad (4.5)$$

$$\tilde{Y}_{i+1}^n = \tilde{Y}_i^n + b(i\delta, \tilde{Y}_i^n)\delta + \sigma(i\delta, \tilde{Y}_i^n)\Delta W_i^{Y,n}, \quad (4.6)$$

where $\tilde{\mu}_i^N = \frac{1}{N} \sum_{n=1}^N \delta_{(\tilde{X}_i^{N,n}, \tilde{Y}_i^{N,n})}$. Thus at each discretisation time step of (4.5), (4.6), we need to compute approximations of the conditional expectations $\hat{m}_{A_r}^\lambda(\tilde{X}_i^n; \tilde{\mu}_i^N)$, $r = 1, 2$. This is done using the algorithm discussed in Sect. 4.1 and takes $O(NL^2)$ operations; see Remark 4.2. Thus the total number of operations needed to implement (4.5), (4.6) is $O(MNL^2)$.

5 Numerical examples and applications to local stochastic volatility models

As a main application of the regularisation approach presented above, we consider the problem of calibration of stochastic volatility models to market data. Fix a time period $T > 0$. To simplify the calculations, we suppose that the interest rate is $r = 0$. Let $C(t, K)$, $t \in [0, T]$, $K \geq 0$, be the price at time 0 of a European call option with strike K and maturity t on a non-dividend paying stock. We assume that the market prices $(C(t, K))_{t \in [0, T], K \geq 0}$ are given and satisfy the following conditions: (i) C is continuous and increasing in t and twice continuously differentiable in x , (ii) $\partial_{xx}C(t, x) > 0$, (iii) $C(t, x) \rightarrow 0$ as $x \rightarrow \infty$ for any $t \geq 0$ and $C(t, 0) = \text{const}$. It is known by Lowther [19, Theorem 1.3 and Sect. 2.1] and Dupire [8] that under these conditions, there exists a diffusion process $(S_t)_{t \in [0, T]}$ which is able to perfectly replicate the given call option prices, that is, $\mathbb{E}[(S_t - K)^+] = C(t, K)$. Furthermore, S solves the stochastic differential equation

$$dS_t = \sigma_{\text{Dup}}(t, S_t)S_t dW_t, \quad t \in [0, T], \quad (5.1)$$

where W is a Brownian motion and σ_{Dup} is the Dupire local volatility given by

$$\sigma_{\text{Dup}}^2(t, x) := \frac{2\partial_t C(t, x)}{x^2 \partial_{xx} C(t, x)}, \quad x > 0, t \in [0, T]. \quad (5.2)$$

We study local stochastic volatility (LSV) models. That is, we assume that the stock price X follows the dynamics

$$dX_t = \sqrt{Y_t} \sigma_{\text{LV}}(t, X_t)X_t dW_t^X, \quad t \in [0, T], \quad (5.3)$$

where W^X is a Brownian motion and $(Y_t)_{t \in [0, T]}$ is a strictly positive variance process, both adapted to some filtration $(\mathcal{F}_t)_{t \geq 0}$. If the function σ_{LV} is given by

$$\sigma_{LV}^2(t, x) := \frac{\sigma_{\text{Dup}}^2(t, x)}{\mathbb{E}[Y_t | X_t = x]}, \quad x > 0, t \in [0, T],$$

and $\int_0^T \mathbb{E}[Y_t \sigma_{LV}(t, X_t)^2 X_t^2] dt < \infty$, then the one-dimensional marginal distributions of X coincide with those of S (see Gyöngy [14, Theorem 4.6], Brunick and Shreve [4, Corollary 3.7]). Thus

$$C(T, K) = \mathbb{E}[(X_T - K)^+], \quad T, K > 0. \quad (5.4)$$

In particular, the choice $Y \equiv 1$ recovers the local volatility model. If Y is a diffusion process,

$$dY_t = b(t, Y_t)dt + \sigma(t, Y_t)dW_t^Y, \quad (5.5)$$

where W^Y is a Brownian motion possibly correlated with W^X , we see that the model (5.3)–(5.5) is a special case of the general McKean–Vlasov equation (2.2), (2.3). To solve (5.3)–(5.5), we implement the algorithm described in Sect. 4; see (4.5), (4.6) together with (4.4). We validate our results by doing two different checks. First, we verify that the one-dimensional distribution of \tilde{X}_M^1 is close to the correct marginal distribution $\text{Law}(X_T) = \text{Law}(S_T)$. To do this, we compare the call option prices obtained by the algorithm (that is, $N^{-1} \sum_{n=1}^N (\tilde{X}_M^n - K)^+$) with the given prices $C(T, K)$ for various $T > 0$ and $K > 0$. If the algorithm is correct and if $\tilde{\mu}_M^N \approx \text{Law}(X_T, Y_T)$, then according to (5.4), one must have

$$C(T, K) \approx N^{-1} \sum_{n=1}^N (\tilde{X}_M^n - K)^+ =: \tilde{C}(T, K). \quad (5.6)$$

On the other hand, if the algorithm is not correct and $\text{Law}(X_T, Y_T)$ is very different from $\tilde{\mu}_M^N$, then (5.6) will not hold.

Second, we also control the multivariate distribution of $(\tilde{X}_i)_{i=0, \dots, M}$. Recall that for any $t \in [0, T]$, we have $\text{Law}(X_t) = \text{Law}(S_t)$. We want to make sure that the dynamics of the process \tilde{X} is different from the dynamics of the local volatility process S . As a test case, we compare option values on the quadratic variation of the logarithm of the price. More precisely, for each $K > 0$, we compare European options on quadratic variation,

$$QV_S(K) := \frac{1}{N} \sum_{n=1}^N \left(\sum_{i=0}^M (\log S_{(i+1)T/M}^n - \log S_{iT/M}^n)^2 - K \right)^+,$$

$$QV_{\tilde{X}}(K) := \frac{1}{N} \sum_{n=1}^N \left(\sum_{i=0}^M (\log \tilde{X}_{i+1}^n - \log \tilde{X}_i^n)^2 - K \right)^+,$$

and verify that these two curves are different. Here, $(S_i^n)_{i=1,\dots,N}$ is an Euler approximation of (5.1). We also check that the prices of European options on quadratic variation converge as $N \rightarrow \infty$.

We consider two different ways to generate market prices $C(T, K)$. First, we assume that the stock follows the Black–Scholes (BS) model, that is, we assume that $\sigma_{\text{Dup}} \equiv \text{const} = 0.3$ and $S_0 = 1$. Second, we consider a stochastic volatility model for the market, that is, we set $C(T, K) := \mathbb{E}[(\bar{S}_T - K)^+]$, where $(\bar{S}_t)_{t \geq 0}$ follows the Heston model

$$d\bar{S}_t = \sqrt{v_t} \bar{S}_t dW_t, \quad (5.7)$$

$$dv_t = \kappa(\theta - v_t) dt + \xi \sqrt{v_t} dB_t, \quad (5.8)$$

with the following parameters: $\kappa = 2.19$, $\theta = 0.17023$, $\xi = 1.04$ and correlation $\rho = -0.83$ between the driving Brownian motions W and B , with initial values $\bar{S}_0 = 1$, $v_0 = 0.0045$; cf. similar parameter choices in Lemaire et al. [18, Table 1]. We compute option prices based on (5.7), (5.8) with the COS method; see Fang and Oosterlee [9]. We then calculate σ_{Dup} from $C(T, K)$ using (5.2).

As our baseline stochastic volatility model for Y , we choose a capped-from-below Heston-type model, but with different parameters than the data-generating Heston model. Specifically, we set $b(t, x) = \lambda(\mu - x)$ and $\sigma(t, x) = \eta\sqrt{x}$ in (5.5), where $Y_0 = 0.0144$, $\lambda = 1$, $\mu = 0.0144$ and $\eta = 0.5751$. We cap the solution of (5.5) from below at the level $\varepsilon_{\text{CIR}} = 10^{-3}$ to avoid singularity at 0. Numerical experiments have shown that such capping is necessary. We assume that the correlation between W^X and W^Y is very strong and equals -0.9 , which makes calibration more difficult. Since the variance process has different parameters compared to the price-generating stochastic volatility model, a non-trivial local volatility function is required to match the implied volatility. Hence even though the generating model is of the same class, the calibration problem is still non-trivial and involves a singular MKV SDE.

We took \mathcal{H} to be the RKHS associated with the Gaussian kernel k with variance 0.1. We fix the number of time steps as $M = 500$ and take $\lambda = 10^{-9}$, $L = 100$. At each time step of the Euler scheme, we choose $(Z_m^j)_{j=1,\dots,L}$ by the rule that

$$Z_m^j \text{ is the } j \frac{100}{L+1} \text{-percentile of the sequence } (\tilde{X}_m^n)_{n=1,\dots,N}, \quad (5.9)$$

an approach comparable to the choice of the evaluation grid $G_{f,t}$ suggested in Guyon and Henry-Labordère [13].

Figure 1 compares the theoretical and the calculated prices (in terms of implied volatilities) in the (a) Black–Scholes and (b)–(d) Heston settings for various strikes and maturities. That is, we first calculate $C(T, K)$ using the Black–Scholes model (“Black–Scholes setting”) or (5.7), (5.8) (“Heston setting”); then we calculate σ_{Dup}^2 by (5.2); then we calculate \tilde{X}_M^n , $n = 1, \dots, N$, using the algorithm (4.5), (4.6) with $H \equiv 0$, $A_2(x) = x$ and

$$F(t, x, y, z) := x \sigma_{\text{Dup}}(t, x) \frac{\sqrt{y}}{\sqrt{z \vee \varepsilon}},$$

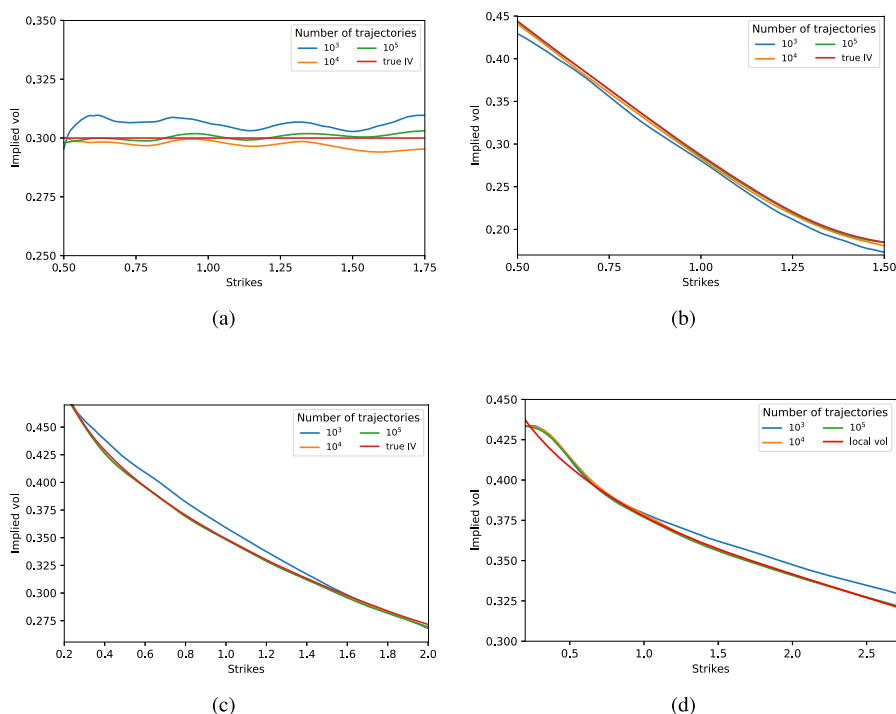


Fig. 1 Fit of the smile for different number of particles: (a) Black–Scholes setting, $T = 1$ year; (b) Heston setting, $T = 1$ year; (c) Heston setting, $T = 4$ years; and (d) Heston setting, $T = 10$ years

where $\varepsilon = 10^{-3}$ (see also Reisinger and Tsianni [22]); then we calculate $\tilde{C}(T, K)$ using (5.6); finally, we transform the prices $C(T, K)$ and $\tilde{C}(T, K)$ to the implied volatilities. We should like to note that this additional capping of the function F is less critical than the capping of the baseline process Y .

We plot in Fig. 1 implied volatilities for a wide range of strikes and maturities. More precisely, we consider all strikes K such that $\mathbb{P}[S_T < K] \in [0.05, 0.95]$ – this corresponds to all but very far in-the-money and out-of-the-money options. One can see from Fig. 1 that already for $N = 10^3$ trajectories, the identity (5.6) holds up to a small error for all the considered strikes and maturities. This error further diminishes as the number of trajectories increases. At $N = 10^5$, the true implied volatility curve and the one calculated from our approximation model become almost indistinguishable.

We plot the prices of the options on the logarithms of quadratic variation in Fig. 2. It is immediate to see that in the Black–Scholes model ((5.1) with $\sigma_{\text{Dup}} \equiv \sigma$), we have $\langle \log S \rangle_T = \sigma^2 T$ and thus $\mathbb{E}[(\langle \log S \rangle_T - K)^+] = (\sigma^2 T - K)^+$. As shown in Fig. 2(a), the prices of the options on the quadratic variation of X are vastly different. This implies that despite the marginal distributions of X and S being identical, their dynamics are markedly dissimilar. We also see that these curves converge as the number of particles increases to infinity. This shows that the dynamics of (\tilde{X}^n) is stable with respect to n . Options on the logarithm of quadratic variation for the Heston

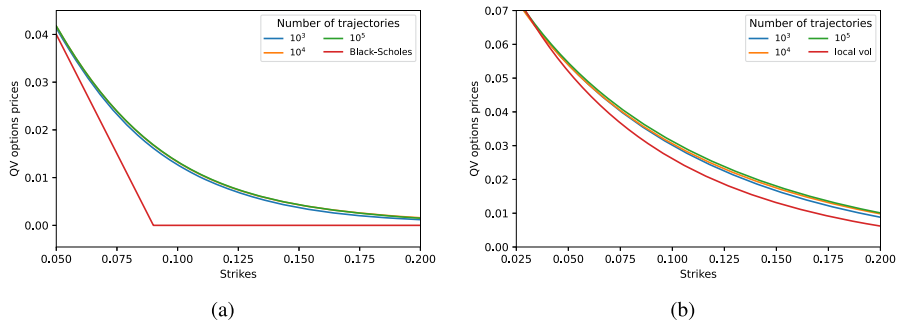


Fig. 2 Prices of options on log of quadratic variation for different number of particles: (a) Black–Scholes setting, $T = 1$ year; (b) Heston setting, $T = 1$ year

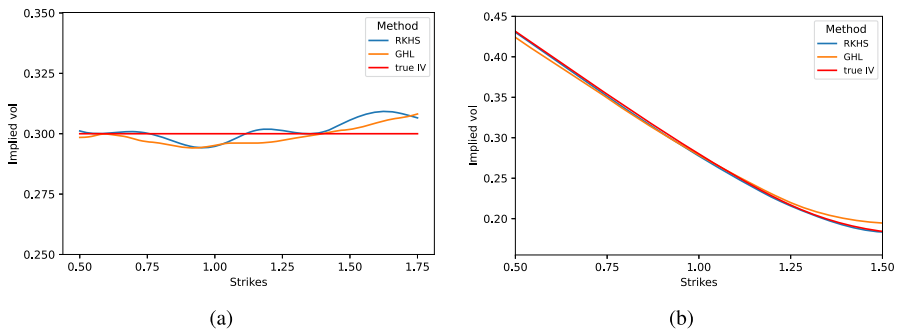


Fig. 3 Comparison with Guyon and Henry-Labordère [13]: (a) Black–Scholes setting, $T = 1$ year; (b) Heston setting, $T = 1$ year

setting are presented in Fig. 2(b). We see that in this case, the dynamics of X and S are different as expected, and the dynamics of (\tilde{X}^n) is also stable.

It is interesting to compare our approach with the algorithms of Guyon and Henry-Labordère [13, 12]. We consider a numerical setup similar to [12, p. 10], taking $N = 10^6$ particles to calculate implied volatilities. However, we calibrate our model and calculate the approximation of conditional expectation using only $N_1 = 1000$ of these particles. We compare our results in the Black–Scholes (a) and Heston (b) settings against implied volatilities calculated via the Euler method for the local volatility model S . Figure 3 shows great agreement between the results of the two methods.

Remark 5.1 The computational time needed for running our algorithm is comparable with the algorithm of [12], but highly dependent on implementation details in both cases.

Figure 4 shows that not only do the marginal distributions of X calculated with our method and [13] agree with each other, but so do the distributions of the processes. We also observe that in both settings, the dynamics of X are different from the dynamics of S .

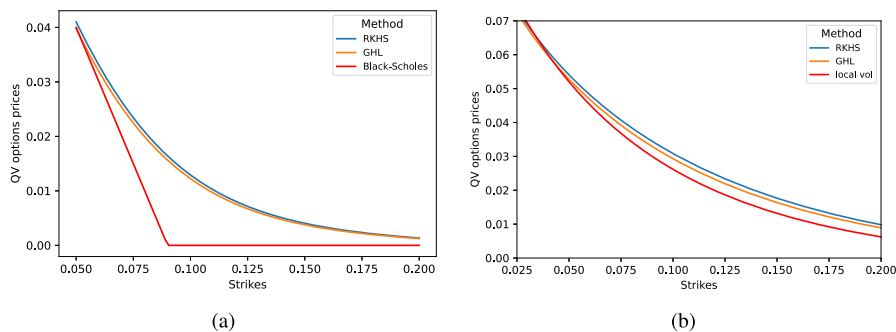


Fig. 4 Comparison with Guyon and Henry-Labordère [13]. Options on quadratic variation: (a) Black–Scholes setting, $T = 1$ year; (b) Heston setting, $T = 1$ year

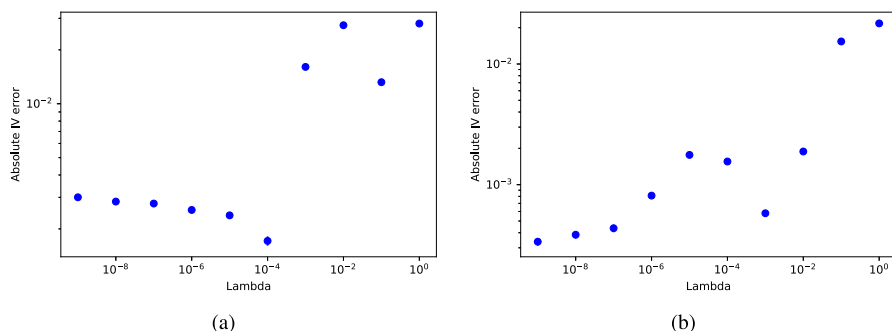


Fig. 5 Mean absolute implied volatility error for different values of λ : (a) Black–Scholes setting; (b) Heston setting

Now let us discuss the stability of our model as the regularisation parameter $\lambda \rightarrow 0$. We studied the absolute error in the implied volatility of the 1-year ATM call option for various $\lambda \in [10^{-9}, 1]$ in the Black–Scholes and Heston settings described above. We used $N = 10^6$ trajectories and $L = 100$ at each step according to (5.9), and performed 100 repetitions at each considered value of λ . The results are presented in Fig. 5. The vertical lines in Figs. 5–7 denote the standard deviation in the absolute errors of the implied volatilities. We observe that in both settings, the error initially drops as λ decreases and then stabilises around $\lambda \approx 10^{-9}$. Therefore, for all of our calculations, we took $\lambda = 10^{-9}$. It is evident that the error does not blow up as λ becomes very small.

Let us examine how the error in call option prices in (5.6) (and therefore the distance between the laws of the true and approximated solutions) depends on the number N of trajectories. Recall that it follows from Theorem 2.3 that this error should decrease as $N^{-1/4}$ (note the square in the left-hand side of (2.10)). Figure 6 shows how the absolute error in the implied volatility of a 1-year ATM call option decreases as the number of trajectories increases in (a) the Black–Scholes setting and (b) the Heston setting. We took $\lambda = 10^{-9}$, $L = 100$, $N \in [250, 2^8 \times 250]$ and performed

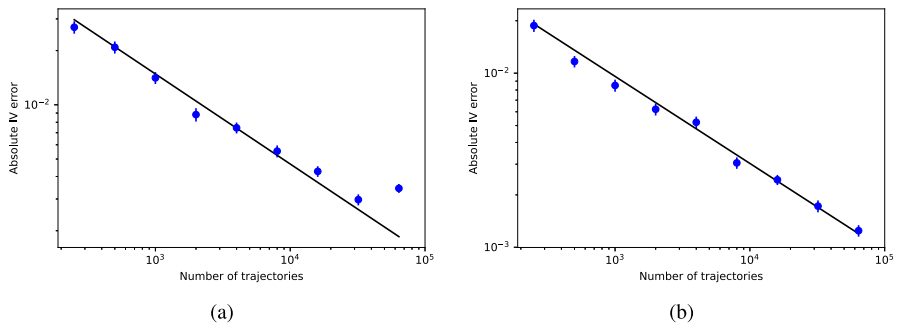


Fig. 6 Mean absolute implied volatility error versus number of trajectories. The black line is the approximation: $\text{error} = C N^{-1/2}$; (a) Black–Scholes setting, $C = 0.469$; (b) Heston setting, $C = 0.303$

Table 1 Average error in implied volatility of 1-year options with given strike. Heston setting

Strike K	0.6	0.8	1	1.2	1.4	1.6
$\mathbb{P}[S_T < K]$	0.0990	0.2258	0.4443	0.7475	0.9558	0.9961
True IV	0.3999	0.3383	0.2795	0.2273	0.1927	0.1803
IV error	0.0011	0.0018	0.0006	0.0032	0.0043	0.0011

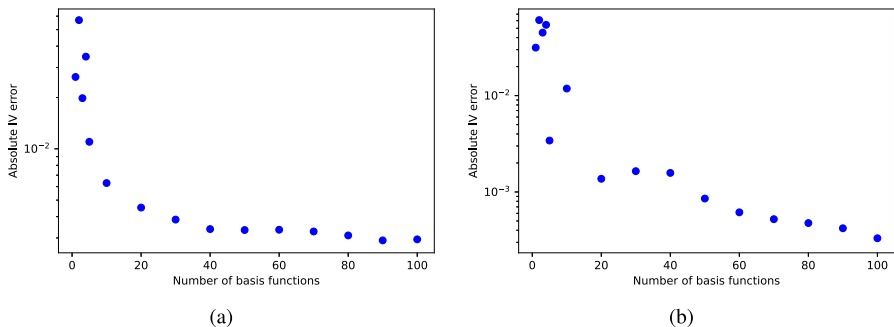


Fig. 7 Mean absolute implied volatility error versus number of basis functions: (a) Black–Scholes setting; (b) Heston setting

100 repetitions at each value of N . We see that the error decreases as $O(N^{-1/2})$ in both settings, which is even better than predicted by theory.

We collect average errors in implied volatilities of 1-year European call options for different strikes in Table 1. We considered the Heston setting and as above, we used $\lambda = 10^{-9}$, $L = 100$, $N = 10^5$.

We also investigate the dependence of the error in the implied volatility on the number L of basis functions in the representation (4.4). Recall that since the number of operations depends on L quadratically (it equals $O(MNL^2)$), it is extremely expensive to set L to be large. In Fig. 7, we plotted the dependence on L of the absolute error in the implied volatility of a 1-year ATM call option. We used $N = 10^6$

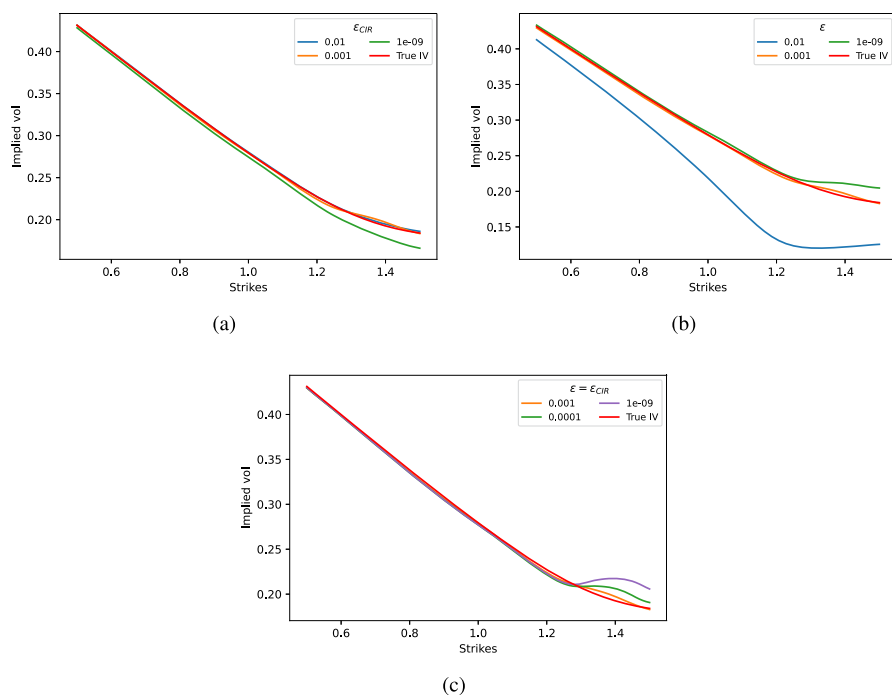


Fig. 8 Fit of the smile of 1-year call options for different truncation levels: (a) $\varepsilon = 10^{-3}$, ε_{CIR} varies; (b) $\varepsilon_{CIR} = 10^{-3}$, ε varies; and (c) $\varepsilon = \varepsilon_{CIR}$ varies

trajectories, $\lambda = 10^{-9}$, $L \in [1, \dots, 100]$ and did 100 repetitions at each value of the number of basis functions. We see that as the number of basis functions increases, the error first drops significantly, but then stabilises at $L \approx 80$.

5.1 On the choice of $(\varepsilon, \varepsilon_{CIR})$

We recall that there are two different truncations involved in the model. First, we cap the CIR process from below at the level of $\varepsilon_{CIR} = 10^{-3}$. Second, in the Euler scheme (4.5), (4.6), we take as a diffusion coefficient

$$F(t, x, y, z) := x \sigma_{Dup}(t, x) \frac{\sqrt{y}}{\sqrt{z \vee \varepsilon}}$$

with $\varepsilon = 10^{-3}$. We claim that both of these truncations are necessary.

Figure 8 shows the fit of the smile for 1-year European call options depending on ε and ε_{CIR} . We use the model of Sect. 5 with $M = 500$ time steps and $N = 10^6$ trajectories. We see from these plots that if ε or ε_{CIR} are either too small or too large, the smile produced by the model may not closely match the true implied volatility curve. Therefore a certain lower capping of the CIR process is indeed necessary.

6 Conclusion and outlook

In this paper, we study the problem of calibrating local stochastic volatility models via the particle approach pioneered in Guyon and Henry-Labordère [13]. We suggest a novel RKHS-based regularisation method and prove that this regularisation guarantees well-posedness of the underlying McKean–Vlasov SDE and the propagation of chaos property. Our numerical results suggest that the proposed approach is rather efficient for the calibration of various local stochastic volatility models and can obtain similar efficiency as widely used local regression methods; see [13]. There are still some questions left open here. First, it remains unclear whether the regularised McKean–Vlasov SDE remains well posed when the regularisation parameter λ tends to zero. This limiting case needs a separate study. Another important issue is the choice of RKHS and the number of basis functions which ideally should be adapted to the problem at hand. This problem of adaptation is left for future research.

7 Proofs

In this section, we present the proofs of the results from Sects. 2 and 3.

Proof of Proposition 3.3 Since \mathcal{H} is separable, let $I \subseteq \mathbb{N}$ and let $e := (e_i)_{i \in I}$ be a total orthonormal system in \mathcal{H} (note that I is finite if \mathcal{H} is finite-dimensional). Define the vector $\gamma^\nu \in \ell_2(I)$ by

$$\begin{aligned} \gamma_i^\nu &:= \langle e_i, c_A^\nu \rangle_{\mathcal{H}} = \int_{\mathcal{X} \times \mathcal{X}} \langle e_i, k(\cdot, x) \rangle_{\mathcal{H}} A(y) \nu(dx, dy) \\ &= \int_{\mathcal{X} \times \mathcal{X}} e_i(x) A(y) \nu(dx, dy), \quad i \in I. \end{aligned} \quad (7.1)$$

Since the operator \mathcal{C}^ν is bounded, it may be described by the (possibly infinite) symmetric matrix

$$B^\nu := (\langle e_i, \mathcal{C}^\nu e_j \rangle_{\mathcal{H}})_{(i,j) \in I \times I} = \left(\int_{\mathcal{X}} e_i(x) e_j(x) \nu(dx, \mathcal{X}) \right)_{(i,j) \in I \times I}, \quad (7.2)$$

which acts as a bounded positive semidefinite operator on $\ell_2(I)$. Denote

$$\beta^\nu = (B^\nu + \lambda I)^{-1} \gamma^\nu. \quad (7.3)$$

For $f \in \mathcal{H}$, write $f = \sum_{i \in I} \beta_i e_i$. Then, recalling (7.1) and (7.2), we derive

$$\begin{aligned}
 & \arg \min_{f \in \mathcal{H}} \left(\int_{\mathcal{X} \times \mathcal{X}} |A(y) - f(x)|^2 v(dx, dy) + \lambda \|f\|_{\mathcal{H}}^2 \right) \\
 &= \arg \min_{\beta \in \ell_2(I)} \left(\int_{\mathcal{X} \times \mathcal{X}} |A(y) - \sum_{i \in I} \beta_i e_i|^2 v(dx, dy) + \lambda \|\beta\|_{\ell_2(I)}^2 \right) \\
 &= \arg \min_{\beta \in \ell_2(I)} \left(-2\langle \beta, \gamma^v \rangle_{\ell_2(I)} + \langle \beta, (B^v + \lambda I)\beta \rangle_{\ell_2(I)} \right) \\
 &= \arg \min_{\beta \in \ell_2(I)} \left(\langle \beta - \beta^v, (B^v + \lambda I)(\beta - \beta^v) \rangle_{\ell_2(I)} \right) \\
 &= \beta^v,
 \end{aligned}$$

where we inserted the definition (7.3) and used the fact that $B^v + \lambda I$ is strictly positive definite for $\lambda > 0$. To complete the proof, it remains to note that

$$\sum_{i=1}^{\infty} \beta_i^v e_i = (C^v + \lambda I_{\mathcal{H}})^{-1} c_A^v,$$

which shows (3.4). \square

Proof of Theorem 2.4 Let us write

$$\begin{aligned}
 |m_A^\lambda(x; \mu) - m_A^\lambda(y; \nu)| &\leq |m_A^\lambda(x; \mu) - m_A^\lambda(x; \nu)| + |m_A^\lambda(x; \nu) - m_A^\lambda(y; \nu)| \\
 &=: I_1 + I_2.
 \end{aligned} \tag{7.4}$$

Working with respect to the orthonormal basis introduced in the proof of Proposition 3.3, see (7.3), we derive for the first term in (7.4) that

$$\begin{aligned}
 I_1 &= |\langle k(x, \cdot), m_A^\lambda(\cdot; \mu) - m_A^\lambda(\cdot; \nu) \rangle_{\mathcal{H}}| \\
 &\leq \|k(x, \cdot)\|_{\mathcal{H}} \|m_A^\lambda(\cdot; \mu) - m_A^\lambda(\cdot; \nu)\|_{\mathcal{H}} \\
 &\leq \sqrt{k(x, x)} \|\beta^\mu - \beta^\nu\|_{\ell_2(I)} \\
 &\leq D_k \|\beta^\mu - \beta^\nu\|_{\ell_2(I)},
 \end{aligned} \tag{7.5}$$

where we used (3.1) and Assumption 2.1.

Denote $Q^\nu := B^\nu + \lambda I$ and $Q^\mu := B^\mu + \lambda I$. Recalling that these are bounded operators from $\ell_2(I)$ to $\ell_2(I)$ with bounded inverses, it is easy to see that

$$\|(Q^\mu)^{-1} - (Q^\nu)^{-1}\|_{\ell_2(I)} \leq \|(Q^\mu)^{-1}\|_{\ell_2(I)} \|(Q^\nu)^{-1}\|_{\ell_2(I)} \|Q^\mu - Q^\nu\|_{\ell_2(I)}.$$

Therefore we get

$$\begin{aligned}
 \|\beta^\mu - \beta^\nu\|_{\ell_2(I)} &= \|(\mathcal{Q}^\mu)^{-1}\gamma^\mu - (\mathcal{Q}^\nu)^{-1}\gamma^\nu\|_{\ell_2(I)} \\
 &\leq \|((\mathcal{Q}^\mu)^{-1} - (\mathcal{Q}^\nu)^{-1})\gamma^\mu\|_{\ell_2(I)} + \|(\mathcal{Q}^\nu)^{-1}(\gamma^\mu - \gamma^\nu)\|_{\ell_2(I)} \\
 &\leq \|(\mathcal{Q}^\mu)^{-1}\|_{\ell_2(I)} \|(\mathcal{Q}^\nu)^{-1}\|_{\ell_2(I)} \|\mathcal{Q}^\mu - \mathcal{Q}^\nu\|_{\ell_2(I)} \|\gamma^\mu\|_{\ell_2(I)} \\
 &\quad + \|(\mathcal{Q}^\nu)^{-1}\|_{\ell_2(I)} \|\gamma^\mu - \gamma^\nu\|_{\ell_2(I)} \\
 &\leq \frac{1}{\lambda^2} \|B^\mu - B^\nu\|_{\ell_2(I)} \|\gamma^\mu\|_{\ell_2(I)} + \frac{1}{\lambda} \|\gamma^\mu - \gamma^\nu\|_{\ell_2(I)}. \quad (7.6)
 \end{aligned}$$

Now observe that for any $i, j \in I$, we have

$$\begin{aligned}
 (B_{ij}^\mu - B_{ij}^\nu)^2 &= \left(\int_{\mathcal{X}} e_i(x) e_j(x) (\mu(dx, \mathcal{X}) - \nu(dx, \mathcal{X})) \right)^2 \\
 &= \int_{\mathcal{X}} \int_{\mathcal{X}} e_i(x) e_j(x) e_i(y) e_j(y) \\
 &\quad \times (\mu(dx, \mathcal{X}) - \nu(dx, \mathcal{X})) (\mu(dy, \mathcal{X}) - \nu(dy, \mathcal{X})).
 \end{aligned}$$

Hence by using the identity

$$\begin{aligned}
 \sum_{i \in I} e_i(x) e_i(y) &= \sum_{i \in I} \langle k(x, \cdot), e_i \rangle_{\mathcal{H}} \langle k(y, \cdot), e_i \rangle_{\mathcal{H}} \\
 &= \langle k(x, \cdot), k(y, \cdot) \rangle_{\mathcal{H}} = k(x, y), \quad (7.7)
 \end{aligned}$$

we get

$$\begin{aligned}
 \|B^\mu - B^\nu\|_{\ell_2(I)}^2 &\leq \|B^\mu - B^\nu\|_{\text{HS}}^2 \\
 &= \int_{\mathcal{X}} (\mu(dx, \mathcal{X}) - \nu(dx, \mathcal{X})) \int_{\mathcal{X}} k^2(x, y) (\mu(dy, \mathcal{X}) - \nu(dy, \mathcal{X})). \quad (7.8)
 \end{aligned}$$

By the Kantorovich–Rubinstein duality formula (see Villani [27, Chap. 1]), for every $h : \mathcal{X} \rightarrow \mathbb{R}$ with $h \in C^1(\mathcal{X})$, one has

$$\begin{aligned}
 \left| \int_{\mathcal{X}} h(x) (\mu(dx, \mathcal{X}) - \nu(dx, \mathcal{X})) \right| &= \left| \int_{\mathcal{X} \times \mathcal{X}} h(x) (\mu(dx, dy) - \nu(dx, dy)) \right| \\
 &\leq \mathbb{W}_1(\mu, \nu) \sup_{x \in \mathcal{X}} |\partial_x h(x)|,
 \end{aligned}$$

where ∂_x denotes the gradient with respect to x . So we continue (7.8) with

$$\|B^\mu - B^\nu\|_{\ell_2(I)}^2 \leq \mathbb{W}_1(\mu, \nu) \sup_{x \in \mathcal{X}} \left| \int_{\mathcal{X}} \partial_x k^2(x, y) (\mu(dy, \mathcal{X}) - \nu(dy, \mathcal{X})) \right|, \quad (7.9)$$

and for each particular $x \in \mathcal{X}$, we have similarly

$$\begin{aligned} & \left| \int_{\mathcal{X}} \partial_x k^2(x, y) (\mu(dy, \mathcal{X}) - \nu(dy, \mathcal{X})) \right| \\ & \leq \sum_{i=1}^d \left| \int_{\mathcal{X}} \partial_{x_i} k^2(x, y) (\mu(dy, \mathcal{X}) - \nu(dy, \mathcal{X})) \right| \\ & \leq \mathbb{W}_1(\mu, \nu) \sum_{i=1}^d \sup_{y \in \mathcal{X}} |\partial_y \partial_{x_i} k^2(x, y)| \\ & \leq d^2 D_k^2 \mathbb{W}_1(\mu, \nu), \end{aligned}$$

where the last inequality follows from Assumption 2.1. Combining this with (7.9), we deduce

$$\|B^\mu - B^\nu\|_{\ell_2(I)} \leq D_k \mathbb{W}_1(\mu, \nu) d. \quad (7.10)$$

By a similar argument, using (7.7), we derive

$$\begin{aligned} & \|\gamma^\mu - \gamma^\nu\|_{\ell_2(I)}^2 \\ & \leq \sum_{i \in I} \int_{\mathcal{X} \times \mathcal{X}} \int_{\mathcal{X} \times \mathcal{X}} e_i(x) e_i(x') A(y) A(y') (\mu - \nu)(dx, dy) (\mu - \nu)(dx', dy') \\ & \leq \int_{\mathcal{X} \times \mathcal{X}} \int_{\mathcal{X} \times \mathcal{X}} k(x, x') A(y) A(y') (\mu - \nu)(dx, dy) (\mu - \nu)(dx', dy') \\ & \leq d^2 \mathbb{W}_1^2(\mu, \nu) \|A\|_{\mathcal{C}^1}^2 D_k^2, \end{aligned} \quad (7.11)$$

where again Assumption 2.1 was used. Next note that

$$\begin{aligned} \|\gamma^\mu\|_{\ell_2(I)}^2 &= \int_{\mathcal{X} \times \mathcal{X}} \int_{\mathcal{X} \times \mathcal{X}} k(x, x') A(y) A(y') \mu(dx, dy) \mu(dx', dy') \\ &\leq \int_{\mathcal{X} \times \mathcal{X}} \int_{\mathcal{X} \times \mathcal{X}} |A(y)| \sqrt{k(x, x')} |A(y')| \sqrt{k(x', x')} \mu(dx, dy) \mu(dx', dy') \\ &= \left(\int_{\mathcal{X} \times \mathcal{X}} |A(y)| \sqrt{k(x, x)} \mu(dx, dy) \right)^2 \\ &\leq \int_{\mathcal{X} \times \mathcal{X}} |A(y)|^2 \mu(dx, dy) \int_{\mathcal{X} \times \mathcal{X}} k(x, x) \mu(dx, dy) \\ &\leq D_k^2 \|A\|_{\mathcal{C}^1}^2 \end{aligned} \quad (7.12)$$

due to Assumption 2.1. Substituting now (7.10)–(7.12) into (7.6) and then into (7.5), we finally get

$$I_1 \leq (\lambda^{-1} D_k + 1) \lambda^{-1} D_k^2 \mathbb{W}_1(\mu, \nu) d \|A\|_{\mathcal{C}^1}. \quad (7.13)$$

Now let us bound I_2 in (7.4). We clearly have

$$I_2 = |\langle k(x, \cdot) - k(y, \cdot), m_A^\lambda(\cdot; \nu) \rangle| \leq \|k(x, \cdot) - k(y, \cdot)\|_{\mathcal{H}} \|m_A^\lambda(\cdot; \nu)\|_{\mathcal{H}}. \quad (7.14)$$

Note that

$$\begin{aligned} & \|k(x, \cdot) - k(y, \cdot)\|_{\mathcal{H}}^2 \\ &= \langle k(x, \cdot) - k(y, \cdot), k(x, \cdot) - k(y, \cdot) \rangle_{\mathcal{H}} \\ &= k(x, x) - k(x, y) - (k(y, x) - k(y, y)) \\ &= \left(\int_0^1 \partial_2 k(x, x + \xi(y - x)) d\xi \right)^\top (x - y) \\ &\quad - \left(\int_0^1 \partial_2 k(y, x + \xi(y - x)) d\xi \right)^\top (x - y) \\ &= (x - y)^\top \left(\int_0^1 \int_0^1 \partial_1 \partial_2 k(x + \eta(y - x), x + \xi(y - x)) d\xi d\eta \right)^\top (x - y), \end{aligned}$$

with ∂_1, ∂_2 denoting the vector of derivatives of k with respect to the first and second argument, respectively. Recalling Assumption 2.1, we derive

$$\|k(x, \cdot) - k(y, \cdot)\|_{\mathcal{H}}^2 \leq d D_k^2 |x - y|^2. \quad (7.15)$$

Further, using (7.12), we see that

$$\|m_A^\lambda(\cdot; \nu)\|_{\mathcal{H}} = \|\beta^\nu\|_{\ell_2(I)} \leq \|(B^\nu + \lambda I)^{-1}\|_{\ell_2(I)} \|\gamma^\nu\|_{\ell_2(I)} \leq \lambda^{-1} D_k \|A\|_{\mathcal{C}^1}.$$

Combining this with (7.15) and substituting into (7.14), we get

$$I_2 \leq \sqrt{d} \lambda^{-1} D_k^2 \|A\|_{\mathcal{C}^1} |x - y|.$$

This together with (7.13) and (7.4) finally yields

$$|m_A^\lambda(x; \mu) - m_A^\lambda(y; \nu)| \leq C_1 \mathbb{W}_1(\mu, \nu) + C_2 |x - y|,$$

where $C_1 = (\lambda^{-1} D_k + 1) \lambda^{-1} D_k^2 d \|A\|_{\mathcal{C}^1}$ and $C_2 = \sqrt{d} \lambda^{-1} D_k^2 \|A\|_{\mathcal{C}^1}$. This completes the proof. \square

Now we are ready to prove the main results of Sect. 2. They follow from Theorem 2.4 obtained above.

Proof of Theorem 2.2 It follows from Theorem 2.4, the assumptions of Theorem 2.2 and the fact that the \mathbb{W}_1 -metric can be bounded from above by the \mathbb{W}_2 -metric that the drift and diffusion of the system (2.4)–(2.6) are Lipschitz and satisfy the conditions of Carmona and Delarue [6, Theorem 4.21]. Hence it has a unique strong solution. \square

Proof of Theorem 2.3 We see that Theorem 2.4 and the conditions of Theorem 2.3 imply that all the assumptions of Carmona and Delarue [7, Theorem 2.12] hold (note that the total state dimension is $2d$ in our case). This implies (2.10). \square

Proof of Theorem 3.5 Consider the operator C^v in the orthonormal basis $(\tilde{a}_n)_{n \in J}$ of \mathcal{H} . Put

$$D^v := (\langle \tilde{a}_i, C^v \tilde{a}_j \rangle_{\mathcal{H}})_{(i,j) \in J \times J} = (\langle \tilde{a}_i, T^v \tilde{a}_j \rangle_{\mathcal{H}})_{(i,j) \in J \times J} = (\sigma_j \delta_{ij})_{(i,j) \in J \times J},$$

since \tilde{a}_j is an eigenvector of T^v with eigenvalue σ_j . Since C^v is diagonal in this basis, we see that for $\lambda > 0$, one has for $i \in J$ that

$$(C^v + \lambda I_{\mathcal{H}})^{-1} \tilde{a}_i = (\sigma_i + \lambda)^{-1} \tilde{a}_i. \quad (7.16)$$

Consider also the function c_A^v in this basis. We write for $i \in J$ similarly to (7.1)

$$\eta_i^v := \langle c_A^v, \tilde{a}_i \rangle_{\mathcal{H}} = \int_{\mathcal{X} \times \mathcal{X}} \tilde{a}_i(x) A(y) v(dx, dy), \quad i \in I,$$

and we clearly have $c_A^v = \sum_{i \in J} \eta_i^v \tilde{a}_i$. Then, using Proposition 3.3 and (7.16), we derive for $\lambda > 0$ that

$$\begin{aligned} m_A^\lambda(\cdot; v) &= (C^v + \lambda I_{\mathcal{H}})^{-1} c_A^v = \sum_{i \in J} \eta_i^v (C^v + \lambda I_{\mathcal{H}})^{-1} \tilde{a}_i \\ &= \sum_{i \in J} \eta_i^v (\sigma_i + \lambda)^{-1} \tilde{a}_i. \end{aligned} \quad (7.17)$$

Next, since $m_A \in \mathcal{L}_2^v$, we have

$$P_{\mathcal{H}} m_A = \sum_{i \in J} \langle \mathbb{E}_{(X,Y) \sim v} [A(Y) | X = \cdot], a_i \rangle_{\mathcal{L}_2^v} a_i. \quad (7.18)$$

Further, for $i \in J$, we deduce that

$$\begin{aligned} \langle \mathbb{E}_{(X,Y) \sim v} [A(Y) | X = \cdot], a_i \rangle_{\mathcal{L}_2^v} &= \int_{\mathcal{X}} \mathbb{E}_{(X,Y) \sim v} [A(Y) | X = x] a_i(x) v(dx, \mathcal{X}) \\ &= \mathbb{E}_{(X,Y) \sim v} [a_i(X) \mathbb{E}[A(Y) | X]] \\ &= \mathbb{E}_{(X,Y) \sim v} [a_i(X) A(Y)] \\ &= \sigma_i^{-1/2} \eta_i^v, \end{aligned}$$

where we used that $\tilde{a}_n = \sqrt{\sigma_n} a_n$. Substituting this into (7.18) and combining with (7.17), we get

$$P_{\mathcal{H}} m_A - m_A^\lambda = \sum_{i \in J} (\eta_i^v \sigma_i^{-1} - \eta_i^v (\sigma_i + \lambda)^{-1}) \tilde{a}_i = \sum_{i \in J} \eta_i^v \frac{\lambda}{\sigma_i (\sigma_i + \lambda)} \tilde{a}_i.$$

Thus the orthonormality of the \tilde{a}_i gives

$$\|P_{\mathcal{H}} m_A - m_A^\lambda\|_{\mathcal{L}_2^v}^2 = \sum_{i \in J} (\eta_i^v)^2 \frac{\lambda^2}{\sigma_i (\sigma_i + \lambda)^2} = \sum_{i \in J} \langle m_A, a_i \rangle_{\mathcal{L}_2^v}^2 \frac{\lambda^2}{(\sigma_i + \lambda)^2},$$

which is (3.10). Similarly, recalling (3.8), we get

$$\|P_{\mathcal{H}}m_A - m_A^\lambda\|_{\mathcal{H}}^2 = \sum_{i \in J} (\eta_i^v)^2 \frac{\lambda^2}{\sigma_i^2(\sigma_i + \lambda)^2} = \sum_{i \in J} \langle m_A, a_i \rangle_{\mathcal{L}_2^v}^2 \frac{\lambda^2}{\sigma_i(\sigma_i + \lambda)^2},$$

which is finite whenever $P_{\mathcal{H}}m_A \in \mathcal{H}$, that is, $\sum_{i \in J} \langle m_A, a_i \rangle_{\mathcal{L}_2^v}^2 \sigma_i^{-1} < \infty$. This shows (3.11). It is easily seen by dominated convergence that the left-hand side of (3.10) goes to zero, and if $P_{\mathcal{H}}m_A \in \mathcal{H}$, the left-hand side of (3.11) goes to zero as well. \square

Acknowledgements The authors would like to thank the referees and the Associate Editor for their helpful comments and feedback. We are also grateful to Peter Friz and Mykhaylo Shkolnikov for useful discussions. OB would like to thank Vadim Sukhov for very helpful conversations regarding implementing parallelised algorithms in Python.

Funding Open Access funding enabled and organized by Projekt DEAL. DB acknowledges the financial support from Deutsche Forschungsgemeinschaft (DFG), Grant Nr. 497300407. CB, OB and JS are supported by the DFG Research Unit FOR 2402. OB is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – The Berlin Mathematics Research Center MATH+ (EXC-2046/1, project ID: 390685689, sub-project EF1-22).

Declarations

Competing Interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Antonelli, F., Kohatsu-Higa, A.: Rate of convergence of a particle method to the solution of the McKean–Vlasov equation. *Ann. Appl. Probab.* **12**, 423–476 (2002)
2. Bach, F.: Are all kernels cursed? (2019). Available online at <https://francisbach.com/cursed-kernels/>
3. Bossy, M., Jabir, J.-F.: On the wellposedness of some McKean models with moderated or singular diffusion coefficient. In: Cohen, S.N., et al. (eds.) *Frontiers in Stochastic Analysis – BSDEs, SPDEs and Their Applications*, Edinburgh, July 2017, pp. 43–87. Springer, Cham (2019)
4. Brunick, G., Shreve, S.: Mimicking an Itô process by a solution of a stochastic differential equation. *Ann. Appl. Probab.* **23**, 1584–1628 (2013)
5. Butkovsky, O., Dareiotis, K., Gerencsér, M.: Approximation of SDEs – a stochastic sewing approach. *Probab. Theory Relat. Fields* **181**, 975–1034 (2021)
6. Carmona, R., Delarue, F.: *Probabilistic Theory of Mean Field Games with Applications I. Mean Field FBSDEs, Control, and Games*. Springer, Cham (2018)
7. Carmona, R., Delarue, F.: *Probabilistic Theory of Mean Field Games with Applications II. Mean Field Games with Common Noise and Master Equations*. Springer, Cham (2018)
8. Dupire, B.: Pricing with a smile. *Risk* **7**, 18–20 (1994)
9. Fang, F., Oosterlee, C.W.: A novel pricing method for European options based on Fourier-cosine series expansions. *SIAM J. Sci. Comput.* **31**, 826–848 (2009)

10. Funaki, T.: A certain class of diffusion processes associated with nonlinear parabolic equations. *Z. Wahrscheinlichkeitstheor. Verw. Geb.* **67**, 331–348 (1984)
11. Gatheral, J.: *The Volatility Surface: A Practitioner's Guide*. Wiley, New York (2011)
12. Guyon, J., Henry-Labordère, P.: The smile calibration problem solved (2011). Preprint. Available online at <https://ssrn.com/abstract=1885032>
13. Guyon, J., Henry-Labordère, P.: Being particular about calibration. *Risk Mag.* **25**(1), 88–93 (2012)
14. Gyöngy, I.: Mimicking the one-dimensional marginal distributions of processes having an Itô differential. *Probab. Theory Relat. Fields* **71**, 501–516 (1986)
15. Jourdain, B., Menozzi, S.: Convergence rate of the Euler–Maruyama scheme applied to diffusion processes with L_Q – L_ρ drift coefficient and additive noise. *Ann. Appl. Probab.* **34**, 1663–1697 (2024)
16. Jourdain, B., Zhou, A.: Existence of a calibrated regime switching local volatility model. *Math. Finance* **30**, 501–546 (2020)
17. Lackner, D., Shkolnikov, M., Zhang, J.: Inverting the Markovian projection, with an application to local stochastic volatility models. *Ann. Probab.* **48**, 2189–2211 (2020)
18. Lemaire, V., Montes, T., Pagès, G.: Stationary Heston model: calibration and pricing of exotics using product recursive quantization. *Quant. Finance* **22**, 611–629 (2022)
19. Lowther, G.: Fitting martingales to given marginals (2008). Preprint, Available online at <https://arxiv.org/abs/0808.2319>
20. Mishura, Y.S., Veretennikov, A.Y.: Existence and uniqueness theorems for solutions of McKean–Vlasov stochastic equations. *Theory Probab. Math. Stat.* **103**, 59–101 (2020)
21. Reed, M., Simon, B.: *Functional Analysis. Revised and Enlarged Edition*. Academic Press, San Diego (1980)
22. Reisinger, C., Tsianni, M.O.: Convergence of the Euler–Maruyama particle scheme for a regularised McKean–Vlasov equation arising from the calibration of local-stochastic volatility models (2023). Preprint, Available online at <https://arxiv.org/abs/2302.00434>
23. Schölkopf, B., Herbrich, R., Smola, A.J.: A generalized representer theorem. In: Helmbold, D., Williamson, B. (eds.) *International Conference on Computational Learning Theory*, pp. 416–426. Springer, Berlin (2001)
24. Sriperumbudur, B.K., Gretton, A., Fukumizu, K., Schölkopf, B., Lanckriet, G.R.G.: Hilbert space embeddings and metrics on probability measures. *J. Mach. Learn. Res.* **11**, 1517–1561 (2010)
25. Steinwart, I., Christmann, A.: *Support Vector Machines*. Springer, Berlin (2008)
26. Sun, H.: Mercer theorem for RKHS on noncompact sets. *J. Complex.* **21**, 337–349 (2005)
27. Villani, C.: *Topics in Optimal Transportation*. Am. Math. Soc., Providence (2021)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.