A Reproducing Kernel Hilbert Space approach to singular local stochastic volatility McKean-Vlasov models

Christian Bayer*[1], Denis Belomestny†[2], Oleg Butkovsky‡[1], and John Schoenmakers§[1]

[1]Weierstrass Institute, Mohrenstrasse 39, 10117 Berlin, Germany.
[2]Duisburg-Essen University, Essen

January 26, 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

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

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. Whilst such models are quite popular among practitioners, unfortunately, its well-posedness has not been fully understood yet and, in general, is possibly not guaranteed at all. We develop a novel regularization approach based on the reproducing kernel Hilbert space (RKHS) technique and show that the regularized model is well-posed. Furthermore, we prove propagation of chaos. We demonstrate numerically that a thus regularized model is able to perfectly replicate option prices due to typical local volatility models. Our results are also applicable to more general McKean–Vlasov equations.

1. Introduction

The present article is motivated by [GHL12], wherein Guyon and Henry-Labordère proposed a particle method for the calibration of local stochastic

*christian.bayer@wias-berlin.de
†denis.belomestny@uni-due.de
‡oleg.butkovskiy@gmail.com
§john.schoenmakers@wias-berlin.de

volatility models (e.g. stock price models). For ease of presentation, let us assume zero interest rates and recall that *local volatility models*

$$dX_t = \sigma(t, X_t) X_t dW_t, \tag{1.1}$$

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, $\sigma \equiv \sigma_{\mathrm{Dup}}$ [Dup94]. (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 behavior despite perfect fits to market prices of options. On the other hand, *stochastic volatility models*

$$dX_t = \sqrt{v_t} X_t dW_t \tag{1.2}$$

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 [Gat11] 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

$$dX_t = \sqrt{v_t} \sigma(t, X_t) X_t dW_t, \tag{1.3}$$

then it exactly fits the observed market option prices provided that

$$\sigma_{\mathrm{Dup}}(t, x)^2 = \sigma(t, x)^2 \mathsf{E}\left[v_t | X_t = x\right]. \tag{1.4}$$

This is a simple consequence of the celebrated Gyöngy's Markovian projection theorem [Gyo86, Theorem 4.6], see also [BS13a, Corollary 3.7]. With this choice of σ we have

$$dX_t = \sigma_{\mathrm{Dup}}(t, X_t) X_t \frac{\sqrt{v_t}}{\sqrt{\mathsf{E}\left[v_t | X_t\right]}} dW_t, \tag{1.5}$$

Note that v in (1.5) can be any integrable and positive adapted stochastic process. In a sense, (1.5) may be considered as an inversion of the Markovian projection due to [Gyo86], applied to Dupire's local volatility model, i.e. (1.1) with $\sigma \equiv \sigma_{\mathrm{Dup}}$.

Thus, the stochastic local volatility model of McKean–Vlasov type (1.5) solves the smile calibration problem. However, equation (1.5) is singular in a sense explained below and very hard to analyze and to solve. Even the problem of proving existence or uniqueness for (1.5) (under various assumptions on v) turned out to be notoriously difficult and only a few results

are available; we refer to [LSZ20] for an extensive discussion and literature review. Let us recall that the theory of standard McKean–Vlasov equations of the form

$$dZ_t = \widetilde{H}\left(t, Z_t, \mu_t\right) dt + \widetilde{F}\left(t, Z_t, \mu_t\right) dW_t \tag{1.6}$$

with $\mu_t = \mathrm{Law}(Z_t)$, is well understood under appropriate regularity conditions, in particular, Lipschitz continuity of $\widetilde{H}$ and $\widetilde{F}$ w.r.t. the standard Euclidean distances in the first two arguments and w.r.t. the Wasserstein distance in μ_t, see [Fun84, CD16a, MV16]. Denoting $Z_t := (X_t, Y_t)$, it is not difficult to see that the conditional expectation $(x, \mu_t) \mapsto \mathsf{E}\left[A(Y_t) \mid X_t = x\right]$ is, in general, not Lipschitz continuous in the above sense. Therefore, the standard theory does not apply to (1.5).

There are a number of results available in the literature where the Lipschitz condition on drift and diffusion is not imposed. Bossy and Jabir [BJ17] considered singular McKean-Vlasov (MV) systems of the form:

$$dX_t = \mathsf{E}[\ell(X_t)|Y_t]dt + \mathsf{E}[\gamma(X_t)|Y_t]dW_t, \tag{1.7a}$$

$$dY_t = b(X_t, Y_t)dt + \sigma(Y_t)dB_t, \tag{1.7b}$$

or, alternatively, the seemingly even less regular equation

$$dX_t = \sigma(p(t, X_t))dW_t, \tag{1.8}$$

where $p(t, \cdot)$ denotes the density of X_t. [BJ17] establishes well-posedness of (1.7) and (1.8) under suitable regularity conditions (in particular, ellipticity) based on energy estimates of the corresponding non-linear PDEs. Interestingly, these techniques break down when the roles of X and Y are reversed in (1.7), that is, when $\mathsf{E}[\gamma(X_t)|Y_t]$ is replaced by $\mathsf{E}[\gamma(Y_t)|X_t]$ in (1.7a) – and similarly for the drift term. Hence, the results of [BJ17] do not imply well-posedness of (1.5). In [LSZ20], the authors studied the following two-dimensional SDE,

$$dX_t = b_1(X_t)\frac{h(Y_t)}{\mathsf{E}[h(Y_t)|X_t]}\, dt + \sigma_1(X_t)\frac{f(Y_t)}{\sqrt{\mathsf{E}[f^2(Y_t)|X_t]}}\, dW_t, \tag{1.9a}$$

$$dY_t = b_2(Y_t)\, dt + \sigma_2(Y_t)\, dB_t, \tag{1.9b}$$

where W and B are two independent one-dimensional Brownian motions. Clearly, this can be seen as a generalization of (1.5) with a non-zero drift and with the process v chosen in a special way. The authors proved strong existence and uniqueness of the solutions to (1.9) 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 (1.9) in the general case (without the stationarity assumptions) remains open. Finally, let us mention [JZ20, Theorem 2.2], which established weak existence of the solutions to (1.5) 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 [BDG19, JM21]. 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 [LSZ20] do not provide a way to construct a numerical algorithm for solving (1.5) 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 (1.5):

$$dX_t = H\left(t, X_t, Y_t, \mathsf{E}[A_1(Y_t)|X_t]\right) dt + F\left(t, X_t, Y_t, \mathsf{E}[A_2(Y_t)|X_t]\right) dW_t, \quad (1.10)$$

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 $\mathsf{E}[A_i(Y_t)|X_t = x]$, $i = 1, 2$, $x \in \mathbb{R}^d$.

One approach to tackle this problem was suggested by Guyon and Henry-Labordère in the seminal paper [GHL12] (see also [AKH02]). They used the "identity"

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

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

$$\mathsf{E}[A(Y_t)|X_t = x] \approx \frac{\sum_{i=1}^N A(Y_t^{i,N}) k_\varepsilon(X_t^{i,N} - x)}{\sum_{i=1}^N k_\varepsilon(X_t^{i,N} - x)}. \quad (1.11)$$

Here $\varepsilon > 0$ is a small parameter, $k_\varepsilon(\cdot) \approx \delta_0(\cdot)$ is a regularizing kernel, and $(X^{i,N}, Y^{i,N})_{i=1\dots N}$ is a particle system. This technique for solving (1.10) (assuming (1.10) 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 regression performed, 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 neighborhood around x of size proportional to ε contribute to the estimate of $\mathsf{E}[A(Y_t)|X_t = x]$ – as $k_\varepsilon(\cdot - x) \approx 0$ outside that neighborhood. Hence, local regression cannot take advantage of "global" information about the structure of the function $x \mapsto \mathsf{E}[A(Y_t)|X_t = x]$. 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 (1.11). Taken to the extremes, if we assume a compactly supported kernel k and formally take $\varepsilon = 0$, then the estimator (1.11) collapses to $\mathsf{E}[A(Y_t)|X_t = X_t^{i,N}] \approx A(Y_t^{i,N})$, 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 (1.5), 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., [Bac19]. 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 (1.11) by global regression based on, say, L basis functions. However, it turns out that Lipschitz constants of the resulting approximation to the conditional expectations in terms of the particle distribution explode as $L \to \infty$, unless the basis functions are carefully chosen.

As an alternative to [GHL12] we propose in this paper a novel approach based on ridge regression in the context of reproducing kernel Hilbert spaces (RKHS) which, in particular, does not have either of the above mentioned disadvantages, even when the number of basis functions is infinite.

Let us recall that an RKHS $\mathcal{H}$ is a Hilbert space of real valued functions $f : \mathcal{X} \to \mathbb{R}$, 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 $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, i.e. for any $c_1, ..., c_n \in \mathbb{R}$, $x_1, ..., x_n \in \mathcal{X}$ one has

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

such that for every $x \in \mathcal{X}$, $k_x := k(\cdot, x) \in \mathcal{H}$, and one has that $\langle f, k_x \rangle_{\mathcal{H}} = f(x)$, for all $f \in \mathcal{H}$. As a main feature, any positive definite kernel k uniquely determines a RKHS $\mathcal{H}$ and the other way around. In our setting we will consider $\mathcal{X} \subset \mathbb{R}^d$. For a detailed introduction and further properties of RKHS we refer to the literature, for example [SC08, Chapter 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 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 $\nu := \mathrm{Law}(X, Y)$. Suppose that $A\colon \mathcal{X} \to \mathbb{R}$ is sufficiently regular and $\mathcal{H}$ is large enough so that we have $\mathsf{E}\left[A(Y)|X = \cdot\right] \in \mathcal{H}$. Then, formally,

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

where

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

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

$$m_A^\lambda(\cdot; \nu) := (\mathcal{C}^\nu + \lambda I_{\mathcal{H}})^{-1} c_A^\nu, \tag{1.12}$$

is the solution to the minimization problem

$$m_A^\lambda(\cdot; \nu) := \operatorname*{arg\,min}_{f\in\mathcal{H}} \big(\mathsf{E}(A(Y) - f(X))^2 + \lambda\|f\|_{\mathcal{H}}^2\big), \tag{1.13}$$

see Proposition 3.3. On the other hand one also has

$$\mathsf{E}[A(Y)|X = \cdot] = \operatorname*{arg\,min}_{f\in L_2(\mathbb{R}^d, \mathrm{Law}(X))} \mathsf{E}(A(Y) - f(X))^2,$$

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

The main result of the article is that the regularized MV system obtained by replacing the conditional expectations with their regularized versions (1.12) in (1.10) is well-posed and propagation of chaos holds for the corresponding particle system, see Theorem 2.2 and Theorem 2.3. To establish these theorems, we study the joint regularity of $m_A^\lambda(x; \nu)$ in the space variable x, and the measure ν for fixed $\lambda > 0$. These type of results are almost absent in the literature on RKHS 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, w.r.t. the standard Euclidean norm in x and the Wasserstein-1-norm in ν, and, can be calculated numerically in an efficient way, see Section 2. Additionally, in Section 3 we study the convergence of $m_A^\lambda(\cdot; \nu)$ in (1.12) 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 [GHL12], one may incorporate, at least in

principle, global prior information concerning properties of $\mathsf{E}[A(Y)|X = \cdot]$ into the choice of the RKHS generating kernel k. In a nutshell, if one anticipates beforehand that $\mathsf{E}[A(Y)|X = x] \approx f(x)$ for some known "nice" function f, one may pass on to a new kernel $\tilde{k}(x, y) := k(x, y) + f(x)f(y)$. 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 $m_A^\lambda(\cdot; \nu)$ 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 would have dealt with standard ridge regression, that is, ridge regression based on a fixed system of basis functions, we would have to impose restrictions on the regression coefficients leading to a nonconvex constrained optimization problem.

Thus, the contribution of the current work is fourfold. First, we propose a RKHS-based approach to regularize (1.10) and prove the well-posedness of the regularized equation. Second, we show convergence of the approximation (1.13) to the true conditional expectation as $\lambda \searrow 0$. Third, we suggest a particle based approximation of the regularized equation and analyze 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 regularized version of (1.5) (with m_A^λ in place of the conditional expectation). We show that our system is indeed an approximate solution to (1.5) 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 organized as follows. Our main theoretical results are given in Section 2. Convergence properties of the regularized conditional expectation m_A^λ are established in Section 3. A numerical algorithm for solving (1.10) and an efficient implementable approximation of m_A^λ are discussed in Section 4. Section 5 contains numerical examples. The results of the paper are summarized in Section 6. Finally, all the proofs are placed in Section 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, $C(\lambda)$.

Acknowledgements. The authors would like to thank the referees and the associated editor for their helpful comments and feedback. We are also grateful to Peter Friz and Mykhaylo Shkolnikov for useful discussions. D. B. 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 Excel-

lence Strategy — The Berlin Mathematics Research Center MATH+ (EXC-2046/1, project ID: 390685689, sub-project EF1-22).

Competing Interests

The authors declare no competing interests.

2. Main results

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

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

where the infimum is taken over all random variables X, Y with $\mathrm{Law}(X) = \mu$, $\mathrm{Law}(Y) = \nu$. Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a symmetric, positive definite kernel, and $\mathcal{H}$ be a reproducing kernel Hilbert space of functions $f \colon \mathcal{X} \to \mathbb{R}$ 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 [SC08, Chapter 4] for further properties of RKHS.

Let $A \colon \mathcal{X} \to \mathbb{R}$ 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 following optimization problem (*ridge regression*)

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

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

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

$$dY_t = b(t, Y_t)dt + \sigma(t, Y_t)dW_t^Y, \tag{2.2b}$$

where $H \colon [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$, $F \colon [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d \times \mathbb{R}^d$, $A_i \colon \mathbb{R}^d \to \mathbb{R}$, $b \colon [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$, $\sigma \colon [0, T] \times \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$ are measurable functions, W^X, W^Y are two (possibly correlated) d-dimensional Brownian motions on $(\Omega, \mathcal{F}, \mathsf{P})$, and $t \in [0, T]$. We note that our choice of Y as a diffusion process in (2.2b) is mostly for convenience, and we expect our results to hold in more generality, when appropriately modified.

As mentioned above, denoting $\mu_t := \mathrm{Law}(X_t, Y_t)$, we see that the functional $(x, \mu_t) \mapsto \mathsf{E}\left[A_i(Y_t)|X_t = x\right]$ 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). 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 would imply strong existence and uniqueness of the new system. Furthermore, we will demonstrate numerically that the solution to the new system is still "close" to the solution of (2.2) in a certain sense. Thus, we consider the following system:

$$d\widehat{X}_t = H(t, \widehat{X}_t, Y_t, m^{\lambda}_{A_1}(\widehat{X}_t; \widehat{\mu}_t))dt + F(t, \widehat{X}_t, Y_t, m^{\lambda}_{A_2}(\widehat{X}_t; \widehat{\mu}_t))\, dW^X_t, \tag{2.3a}$$

$$dY_t = b(t, Y_t)dt + \sigma(t, Y_t)\, dW^Y_t \tag{2.3b}$$

$$\widehat{\mu}_t = \mathrm{Law}(\widehat{X}_t, Y_t). \tag{2.3c}$$

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

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

Let $\mathcal{C}^1(\mathcal{X}, \mathbb{R})$ be the space of all functions $f\colon \mathcal{X} \to \mathbb{R}$ 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 Section 7.

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

(1) $A_i \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{R})$, $i = 1, 2$;

(2) *there exists a constant $C > 0$ such that for any $t \in [0, T]$, $x, y, x', y' \in \mathbb{R}^d$, $z, z' \in \mathbb{R}$,*

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

(3) *for any fixed* $x, y, \in \mathbb{R}^d$, $z \in \mathbb{R}$ *one has*

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

(4) $\mathsf{E}|\widehat{X}_0|^2 < \infty$, $\mathsf{E}|Y_0|^2 < \infty$.

Then for any $\lambda > 0$ *the system* (2.3) *with the initial condition* $(\widehat{X}_0, Y_0)$ *has a unique strong solution.*

To analyze a numerical scheme solving (2.3), we consider a particle system

$$\begin{aligned} dX_t^{N,n} &= H\big(t, X_t^{N,n}, Y_t^{N,n}, m_{A_1}^{\lambda}(X_t^{N,n}; \mu_t^N)\big)dt \\ &\quad + F\big(t, X_t^{N,n}, Y_t^{N,n}, m_{A_2}^{\lambda}(X_t^{N,n}; \mu_t^N)\big)dW_t^{X,n}, \qquad (2.4a) \\ dY_t^{N,n} &= b(t, Y_t^{N,n})\, dt + \sigma(t, Y_t^{N,n})\, dW_t^{Y,n}, \qquad (2.4b) \\ \mu_t^N &= \frac{1}{N}\sum_{n=1}^N \delta_{(X_t^{N,n}, Y_t^{N,n})}, \qquad (2.4c) \end{aligned}$$

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 jointly 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 that the initial values* $(X_0^{N,n}, Y_0^{N,n})$ *are jointly independent and have the same law as* $(\widehat{X}_0, Y_0)$. *Moreover, suppose that* $\mathsf{E}|\widehat{X}_0|^q < \infty$, $\mathsf{E}|Y_0|^q < \infty$ *for some* $q > 4$. *Then there exists a constant* $C = C(\lambda, T, \mathsf{E}|\widehat{X}_0|^q, \mathsf{E}|Y_0|^q) > 0$ *such that for any* $n = 1, \dots, N$, $N \in \mathbb{N}$,

$$\mathsf{E}\left[\sup_{0\le t\le T} |X_t^{N,n} - \widehat{X}_t^n|^2\right] + \sup_{0\le t\le T} \mathsf{E}[\mathbb{W}_2(\mu_t^N, \widehat{\mu}_t)^2] \le C\epsilon_N, \qquad (2.5)$$

where the process $\widehat{X}^n$ *solves* (2.3) *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 allowed 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 $A \in \mathcal{C}^1(\mathcal{X}, \mathbb{R})$. Then for any $x, y \in \mathcal{X}$, $\mu, \nu \in \mathcal{P}_2(\mathcal{X} \times \mathcal{X})$ on has*

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

where

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

may be considered to be (possibly suboptimal) Lipschitz constants with respect to the Wasserstein metric and the Euclidian 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 (2.2) only for a fixed regularization parameter $\lambda > 0$ and can not be used to study the limiting case $\lambda \to 0$. Indeed, it follows from Theorem 2.4, that as $\lambda \to 0$, the Lipschitz constants of m_A^λ blow up. Yet, Theorem 2.3 does not imply that for $\lambda \to 0$ the *optimal* Lipschitz constants blow up, nor that the solution to (2.2) blows up. We will demonstrate numerically in Section 5 that for $\lambda \to 0$, in the examples there, the solution to (2.2) 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) may hold. Verifying this theoretically remains however an important open problem.

Remark 2.6. A natural question is whether (2.2) 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 will hold for more general $\mathcal{X}$ and $\mathcal{Y}$, provided that those sets are invariant under the dynamics (2.2) – as well as under the regularized 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 $\mathcal{X} \subset \mathbb{R}^d$, a measure $\nu \in \mathcal{P}_2(\mathcal{X} \times \mathcal{X})$, and impose the following relatively weak assumptions on the function $A\colon \mathcal{X} \to \mathbb{R}$ and the positive kernel $k\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$.

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

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 $x \in \mathcal{X}$

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

Due to Assumption 3.2 and [SC08, Lemma 4.33], $\mathcal{H}$ is a separable RKHS and one has for any $f \in \mathcal{H}$, $x \in \mathcal{X}$,

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

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

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

Therefore, the Bochner integrals

$$c_A^\nu := \int_{\mathcal{X}\times\mathcal{X}} k(\cdot, x)A(y)\nu(dx, dy), \quad \text{and} \quad \mathcal{C}^\nu f := \int_{\mathcal{X}\times\mathcal{X}} k(\cdot, x)f(x)\nu(dx, dy). \tag{3.4}$$

are well defined functions in $\mathcal{H}$ for every $f \in \mathcal{H}$. Moreover, the operator $\mathcal{C}^\nu \colon \mathcal{H} \to \mathcal{H}$ is symmetric and positive semidefinite since

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

Thus, by the Hellinger-Toeplitz theorem (see, e.g., [RS80, Section III.5]), $\mathcal{C}^\nu$ is a bounded self-adjoint linear operator on $\mathcal{H}$. As a consequence, for any $\lambda \geq 0$, the operator $\mathcal{C}^\nu + \lambda I_{\mathcal{H}}$ is a bounded self-adjoint operator on $\mathcal{H}$ with spectrum contained in the interval $[\lambda, \|\mathcal{C}^\nu\| + \lambda]$. Hence, if $\lambda > 0$, then $(\mathcal{C}^\nu + \lambda I_{\mathcal{H}})^{-1}$ exists and is a bounded self-adjoint operator on $\mathcal{H}$ with norm

$$\|(\mathcal{C}^\nu + \lambda I_{\mathcal{H}})^{-1}\|_{\mathcal{H}} \leq \lambda^{-1}. \tag{3.5}$$

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 $\nu \in \mathcal{P}_2(\mathcal{X}\times\mathcal{X})$ and $\lambda > 0$, the solution to (2.1) can be represented as*

$$m_A^\lambda(\cdot; \nu) = (\mathcal{C}^\nu + \lambda I_{\mathcal{H}})^{-1} c_A^\nu. \tag{3.6}$$

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 of it in Section 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 $m_A^\lambda = m_A^\lambda(\cdot;\nu)$ is a "good" approximation to the true conditional expectation

$$m_A = m_A(x;\nu) := \mathsf{E}_{(X,Y)\sim\nu}[A(Y)|X=x] \tag{3.7}$$

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

$$T^\nu f := \int_{\mathcal{X}} k(\cdot,x) f(x)\nu(dx,\mathcal{X}). \tag{3.8}$$

Recalling (3.3), it is easy to see that T^ν is a linear operator $\mathcal{L}_2^\nu \to \mathcal{L}_2^\nu$. Note that that $\mathcal{H} \subset \mathcal{L}_2^\nu$ due to (3.2); thus, $\mathcal{C}^\nu$ is the restriction of T^ν to $\mathcal{H}$. Further, since $|k(x,y)| \le \sqrt{k(x,x)}\sqrt{k(y,y)}$, the kernel k is Hilbert-Schmidt on $\mathcal{L}_2(\mathcal{X}\times\mathcal{X}, \nu(dx,\mathcal{X})\otimes\nu(dy,\mathcal{X}))$, i.e.

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

due to Assumption 3.2. As a consequence of the standard results from functional analysis, one then has (see, for example, [RS80, Section VI]):

(i) the operator T^ν is self-adjoint and compact;

(ii) there exists an orthonormal system $(a_n)_{n\in\mathbb{N}}$ in $\mathcal{L}_2^\nu$ of eigenfunctions corresponding to nonnegative eigenvalues σ_n of T^ν and $\sigma_1 \ge \sigma_2 \ge \sigma_3 \ge \ldots$;

(iii) If $J := \{n \in \mathbb{N} : \sigma_n > 0\}$, one has

$$T^\nu f = \sum_{n\in J} \sigma_n \langle f, a_n\rangle_{\mathcal{L}_2^\nu} a_n, \quad f \in \mathcal{L}_2^\nu \tag{3.9}$$

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

A generalization of Mercer's theorem to unbounded domains [Sun05] implies the following statement.

Proposition 3.4. *Let k be a kernel satisfying Assumption 3.2 and assume that $\nu(\cdot,\mathcal{X})$ is a nondegenerate Borel measure. That is, for every open set $U \subset \mathcal{X}$ one has $\nu(U,\mathcal{X}) > 0$. Then one may take the eigenfunctions a_n in (3.9) to be continuous and k has a series representation*

$$k(x,y) = \sum_{n\in J} \sigma_n a_n(x) a_n(y), \quad x,y \in \mathcal{X} \tag{3.10}$$

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

$$\langle f, g\rangle_{\mathcal{H}} = \sum_{n\in J} \frac{\langle f, a_n\rangle_{\mathcal{L}_2^\nu} \langle g, a_n\rangle_{\mathcal{L}_2^\nu}}{\sigma_n} \quad \text{for} \quad f, g \in \mathcal{H}. \tag{3.11}$$

Now we are ready to present the main result of this section, which quantifies the convergence properties of $m_A^\lambda(\cdot, \nu)$ as $\lambda \to 0$ for a fixed measure ν. Recall the notation (3.7). Let $P_{\overline{\mathcal{H}}}$ denote the orthogonal projection in $\mathcal{L}_2^\nu$ onto $\overline{\mathcal{H}}$, i.e. the closure of $\mathcal{H}$ in $\mathcal{L}_2^\nu$. Hence for any $f \in \mathcal{L}_2^\nu$,

$$P_{\overline{\mathcal{H}}} f = \sum_{n\in J} \langle f, a_n\rangle_{\mathcal{L}_2^\nu} a_n \qquad \text{and} \qquad \left\langle P_{\overline{\mathcal{H}}} f, a_m\right\rangle_{\mathcal{L}_2^\nu} = \langle f, a_m\rangle_{\mathcal{L}_2^\nu}, \quad m \in J, \tag{3.12}$$

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

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, because* A *is bounded measurable). Then for any* $\lambda > 0$

$$\left\| P_{\overline{\mathcal{H}}} m_A(\cdot;\nu) - m_A^\lambda(\cdot;\nu)\right\|_{\mathcal{L}_2^\nu}^2 = \sum_{n\in J} \frac{\lambda^2}{(\sigma_n+\lambda)^2} \langle m_A(\cdot;\nu), a_n\rangle_{\mathcal{L}_2^\nu}^2 . \tag{3.13}$$

In particular, $\left\| P_{\overline{\mathcal{H}}} m_A(\cdot;\nu) - m_A^\lambda(\cdot;\nu)\right\|_{\mathcal{L}_2^\nu} \to 0$ *as* $\lambda \searrow 0$. *If, moreover,* $P_{\overline{\mathcal{H}}} m_A(\cdot;\nu) \in \mathcal{H}$ *one has*

$$\left\| P_{\overline{\mathcal{H}}} m_A(\cdot;\nu) - m(\cdot;\nu)_A^\lambda(\cdot;\nu)\right\|_{\mathcal{H}}^2 = \sum_{n\in J} \frac{\lambda^2}{(\sigma_n+\lambda)^2 \sigma_n} \langle m_A(\cdot;\nu), a_n\rangle_{\mathcal{L}_2^\nu}^2 , \tag{3.14}$$

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

Theorem 3.5 establishes convergence of $m_A^\lambda(\cdot;\nu)$ as $\lambda \to 0$ though without a rate. Its proof is placed in Section 7. Additional assumptions are needed to guarantee a certain 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* $\theta \in (0, 1]$,

$$\sum_{n\in J} \sigma_n^{-\theta} \langle m_A(\cdot;\nu), a_n\rangle_{\mathcal{L}_2^\nu}^2 < \infty. \tag{3.15}$$

Then

$$\left\| P_{\overline{\mathcal{H}}} m_A(\cdot;\nu) - m_A^\lambda(\cdot;\nu)\right\|_{\mathcal{L}_2^\nu}^2 \le \left(1 - \frac{\theta}{2}\right)^2 \left(\frac{\lambda\theta}{2-\theta}\right)^\theta \sum_{n\in J} \sigma_n^{-\theta} \langle m_A(\cdot;\nu), a_n\rangle_{\mathcal{L}_2^\nu}^2 . \tag{3.16}$$

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

$$\left\| P_{\overline{\mathcal{H}}} m_A(\cdot;\nu) - m_A^\lambda(\cdot;\nu) \right\|_{\mathcal{L}_2^\nu} \leq \frac{\sqrt{\lambda}}{2} \left\| P_{\overline{\mathcal{H}}} m_A(\cdot;\nu) \right\|_{\mathcal{H}}. \tag{3.17}$$

Proof. Inequality (3.16) follows from (3.13), (3.15), 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 (3.17) follows from (3.11), (3.12) and (3.16). □

Remark 3.7. If operator T^ν defined in (3.8) is injective, that is, $T^\nu f = 0$ for $f \in \mathcal{L}_2^\nu$ implies $f = 0$, ν-a.s., then $P_{\overline{\mathcal{H}}} = I_{\mathcal{L}_2^\nu}$. 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^ν 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 $f \in \mathcal{L}_2^\nu$ we may define a signed Borel measure $\mu_f(A) := \int_A f(x)\nu(dx,\mathcal{X})$, $A \in \mathcal{B}(\mathcal{X})$, which is finite since $|\mu_f(A)| \leq \int f^2(x)\nu(dx,\mathcal{X}) < \infty$. Hence, if k is an ispd kernel then $T^\nu f = 0$ implies

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

which in turn implies $\mu_f = 0$, i.e. $f = 0$, ν-a.s. Further 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 [SGF+10].

Thus, in this section we have shown that, under certain conditions, $m_A^\lambda(\cdot,\nu)$ may converge at least in $\mathcal{L}_2^\nu$-sense to the true conditional expectation $m_A(\cdot,\nu)$ as $\lambda \to 0$. This makes the heuristic discussion around (1.12) and (1.13) in Section 1 more rigorous.

Remark 3.8. Note that the measure $\widehat{\mu}_t$ in the solution of (2.3) depends on λ, so in fact $\widehat{\mu}_t = \widehat{\mu}_t^\lambda$. Therefore, even when $m_A^\lambda(\cdot,\nu) \to m_A(\cdot,\nu)$ 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). As already explained, this is an unsolved open problem and therefore considered out of our scope. 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, $m_{A_i}^\lambda(\cdot,\mu_t) \to m_{A_i}(\cdot,\mu_t)$ (the true conditional expectation) as $\lambda \searrow 0$.

4. Numerical algorithm

Let us now describe in detail our numerical algorithm to construct solutions to (1.10). 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 (2.4), we need to compute

$$m_A^\lambda(\cdot;\mu_t^N) = \operatorname*{arg\,min}_{f\in\mathcal{H}}\Big\{\frac{1}{N}\sum_{n=1}^N |A(Y_t^{N,n}) - f(X_t^{N,n})|^2 + \lambda\,\|f\|_{\mathcal{H}}^2\Big\}. \tag{4.1}$$

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 RKHS [SHS01, Theorem 1] that m_A^λ has the following representation:

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

for some $\alpha = (\alpha_1,\dots,\alpha_N)^T \in \mathbb{R}^N$. Note that the optimal α can be calculated explicitly by plugging the representation (4.2) into the above minimization problem in place of f and minimizing 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 taking every $\frac{N}{L}$th point among the ordered sequence $X_t^{N,(1)},\dots,X_t^{N,(N)}$ in case when X is one-dimensional – and approximate

$$\sum_{i=1}^N \alpha_i k(X_t^{N,i},\cdot) \approx \sum_{j=1}^L \beta_j k(Z^j,\cdot), \tag{4.3}$$

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

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

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

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

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

$$(K^\top K + N\lambda R)\widehat{\beta} = K^\top G, \tag{4.4}$$

and we approximate expectation as

$$m_A^\lambda(x; \mu_t^N) \approx \sum_{j=1}^{L} \widehat{\beta}_j k(Z^j, x) =: \widehat{m}_A^\lambda(x; \mu_t^N). \tag{4.5}$$

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 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 [GHL12] is motivated by similar concerns regarding explosion of computational cost.

Remark 4.2. Let us see how many operations we need to calculate $\widehat{\beta}$, 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 (4.4). Thus, in total we would need $O(NL^2)$ operations.

4.2 Solving the regularized McKean–Vlasov equation

With the function $\widehat{m}_A^\lambda$ in hand, we now consider the Euler scheme for the particle system (2.4). We fix a time interval $[0, T]$, the number of time steps

M, and, for simplicity, we consider a uniform time increment $\delta := T/M$. Let $\Delta W_i^{X,n}$ and $\Delta W_i^{Y,n}$ denote independent copies of $W^X_{(i+1)\delta} - W^X_{i\delta}$ and $W^Y_{(i+1)\delta} - W^Y_{i\delta}$, respectively, $n = 1, \dots, N$, $i = 1, \dots, M$. Note that for stochastic volatility models, the Brownian motions driving the stock price and the variance process are usually correlated. We now define $\widetilde{X}_0^n = X_0^n$, $\widetilde{Y}_0^n = Y_0^n$, and for $i = 0, \dots, M-1$,

$$\begin{aligned} \widetilde{X}_{i+1}^n &= \widetilde{X}_i^n + H\left(i\delta, \widetilde{X}_i^n, \widetilde{Y}_i^n, \widehat{m}^\lambda_{A_1}(\widetilde{X}_i^n; \widetilde{\mu}_i^N)\right)\delta \\ &\quad + F\left(i\delta, \widetilde{X}_i^n, \widetilde{Y}_i^n, \widehat{m}^\lambda_{A_2}(\widetilde{X}_i^n; \widetilde{\mu}_i^N)\right)\Delta W_i^{X,n} \end{aligned} \tag{4.6a}$$

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

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

5. Numerical examples and applications to local stochastic volatility models

As a main application of the regularization approach presented above, we consider the problem of calibration of stochastic volatility models to market data. Fix time period $T > 0$. To simplify the calculations, we suppose that the interest rate $r = 0$. Let $C(t, K)$, $t \in [0, T]$, $K \geq 0$, be the price at time 0 of a European call option on a non-dividend paying stock with strike K and maturity t. We assume that the market prices $(C(t, K))_{t\in[0,T], K\geq 0}$ are given and satisfy the following conditions: C is continuous and increasing in t, twice continuously differentiable in x, $\partial_{xx}C(t, x) > 0$, $C(t, x) \to 0$ as $x \to \infty$ for any $t \geq 0$, $C(t, 0) = const$. It is known [Low08, Theorem 1.3 and Section 2.1], [Dup94] 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 $\mathsf{E}(S_t - K)_+ = C(t, K)$. Furthermore, S solves the following stochastic differential equation

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

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

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

We study Local Stochastic Volatility (LSV) models. That is, we assume that the stock price X follows the dynamics

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

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

$$\sigma^2_{\mathrm{LV}}(t,x) := \frac{\sigma^2_{\mathrm{Dup}}(t,x)}{\mathsf{E}\left[Y_t \middle| X_t = x\right]}, \quad x > 0,\, t \in [0,T],$$

and $\int_0^T \mathsf{E}\left[Y_t \sigma_{\mathrm{LV}}(t,X_t)^2 X_t^2\right] dt < \infty$, then marginal distributions of X_t coincide with marginal distributions of S_t ([Gyo86, Theorem 4.6], [BS13b, Corollary 3.7]). Thus,

$$C(T,K) = \mathsf{E}(X_T - K)_+, \quad T, K > 0. \tag{5.4}$$

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

$$dY_t = b(t,Y_t)dt + \sigma(t,Y_t)dW^Y_t, \tag{5.5}$$

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). To solve (5.3)-(5.5), we implement the algorithm described in Section 4: see (4.6) together with (4.5). We validate our results, by doing two different checks. First, we verify that 1-dimensional marginal distributions $(\widetilde{X}_M)$ are close to the correct marginal distribution $\mathrm{Law}(X_T) = \mathrm{Law}(S_T)$. To do it we compare the call option prices obtained by the algorithm (that is $N^{-1}\sum_{n=1}^N (\widetilde{X}^n_M - K)_+$) with the given prices $C(T,K)$ for various $T > 0$ and $K > 0$. If the algorithm is correct and if $\widetilde{\mu}^N_M \approx \mathrm{Law}(X_T, Y_T)$, then, according to (5.4), one must have

$$C(T,K) \approx N^{-1}\sum_{n=1}^N (\widetilde{X}^n_M - K)_+ =: \widetilde{C}(T,K). \tag{5.6}$$

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

Second, we also control the multivariate distribution of $(\widetilde{X}_i)_{i=0,\dots,M}$. Recall that for any $t \in [0,T]$ we have $\mathrm{Law}(X_t) = \mathrm{Law}(S_t)$. We want to make sure that the dynamics of process $\widetilde{X}$ 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:

$$QV_S(K) := \frac{1}{N}\sum_{n=1}^N \Big(\sum_{i=0}^M (\log(S^n_{(i+1)T/M}) - \log(S^n_{iT/M}))^2 - K\Big)_+$$

$$QV_{\widetilde{X}}(K) := \frac{1}{N}\sum_{n=1}^N \Big(\sum_{i=0}^M (\log(\widetilde{X}^n_{i+1}) - \log(\widetilde{X}^n_i))^2 - K\Big)_+$$

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

We will 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 $\sigma_{\text{Dup}} \equiv const$ for $const = 0.3$ and $S_0 = 1$. Second, we consider a stochastic volatility model for the market, that is, we set $C(T, K) := \mathsf{E}\left[(\overline{S}_T - K)_+\right]$, where $\overline{S}_t$, $t > 0$, follows the Heston model

$$d\overline{S}_t = \sqrt{v_t}\overline{S}_t \, dW_t, \tag{5.7a}$$

$$dv_t = \kappa(\theta - v_t) \, dt + \xi\sqrt{v_t} \, dB_t, \tag{5.7b}$$

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 $\overline{S}_0 = 1$, $v_0 = 0.0045$, cf. similar parameter choices in [LMP22, Table 1]. We compute option prices based on (5.7) with the COS method, see [FO09]. 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 $\varepsilon_{CIR} = 10^{-3}$ 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 RKHS associated with the Gaussian kernel k with variance 0.1. We fix the number of time steps $M = 500$, $\lambda = 10^{-9}$, $L = 100$. At each time step of the Euler scheme we choose $(Z^j)_{j=1,\dots,L}$ by the following rule:

$$Z_j \text{ is the } j \cdot 100/(L+1) \text{ percentile of the sequence } \{\widetilde{X}^n_m\}_{n=1,\dots,N}, \tag{5.8}$$

an approach comparable to the choice of the evaluation grid $G_{f,t}$ suggested in [GHL12].

Figure 1 compares the theoretical and the calculated prices (in terms of implied volatilities) in the Black-Scholes (a) and Heston (b-d) 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) ("Heston setting"); then we calculate σ^2_{Dup} by (5.2); then we calculate $\widetilde{X}^n_M$, $n = 1, \dots, N$, using

the algorithm (4.6) with $H \equiv 0$, $A_2(x) = x$, and

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

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

We plot at Figure 1 implied volatilities for a wide range of strikes and maturities. More precisely, we consider all strikes K such that $\mathsf{P}(S_T < K) \in [0.05, 0.95]$ — this corresponds to all but very far in–the–money and out–of–the–money options. One can see from Figure 1 that already for $N = 10^3$ trajectories, 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.

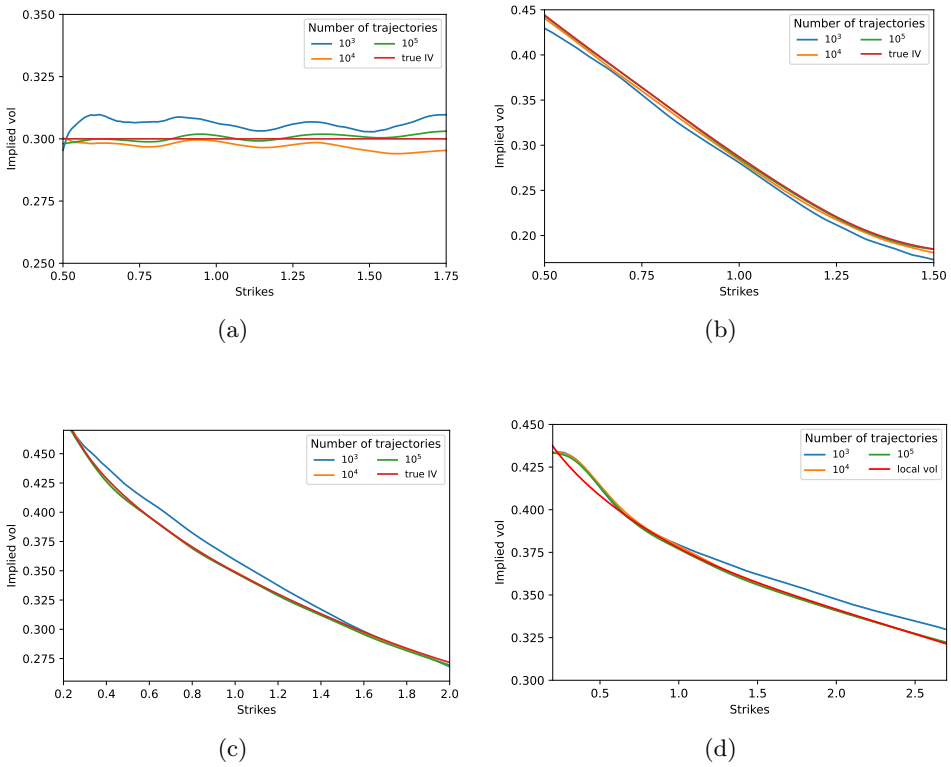
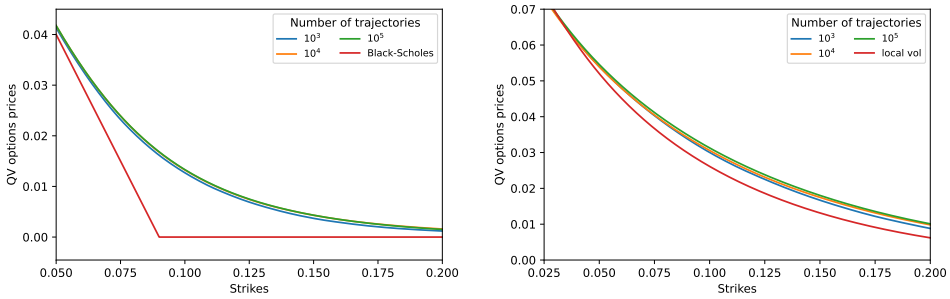


Figure 1: Fit of the smile for different number of particles. (a): Black-Scholes setting, $T = 1$ year. (b): Heston setting, $T = 1$ year. (c): Heston setting, $T = 4$ years. (d): Heston setting, $T = 10$ years.

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 (equation (5.1) with $\sigma_{Dup} = \sigma$), we have $\langle \log S \rangle_T = \sigma^2 T$, and thus $\mathsf{E}[\langle \log S \rangle_T - K]_+ = (\sigma^2 T - K)_+$. 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 converges as the number of particles increases to infinity. This shows, that the dynamics of $(\widetilde{X}^n)$ is stable with respect to n. Options on the logarithms of quadratic variation for the Heston 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 $(\widetilde{X}^n)$ is also stable.

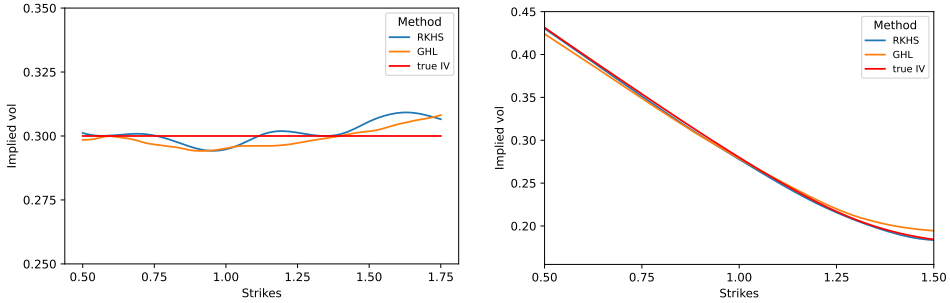


(a) (b)

Figure 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.

It is interesting to compare our approach with the algorithm of [GHL12] and [GHL11]. We consider a numerical setup similar to [GHL11, 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. Fig. 3 shows great agreement between the results of the two methods.

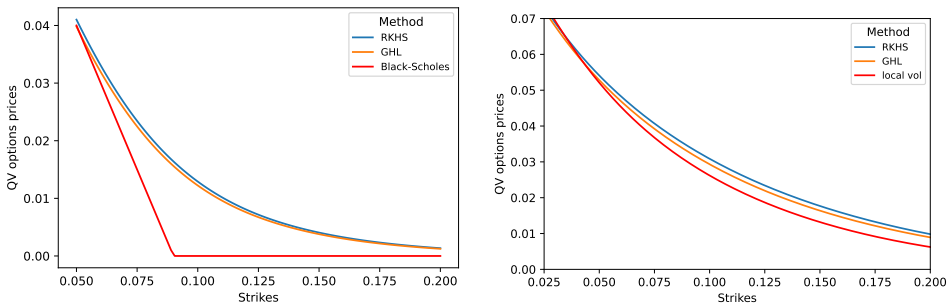
Remark 5.1. The computational time needed for running our algorithm is comparable with the algorithm of [GHL11], but highly dependent on implementation details in both cases.



(a) (b)

Figure 3: Comparison with [GHL12]. (a): Black-Scholes setting, $T = 1$ year. (b): Heston setting, $T = 1$ year.

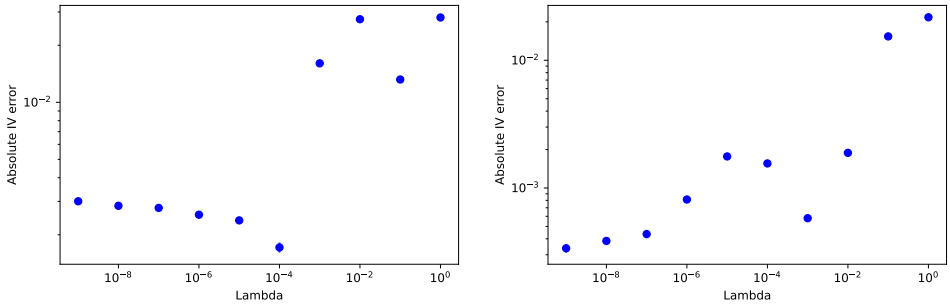
Fig. 4 shows that not only do the marginal distributions of X calculated with our method and [GHL12] agree with each other, but so do the cumulative distributions. We also observe that in both settings, the dynamics of X are different from the dynamics of S.



(a) (b)

Figure 4: Comparison with [GHL12]. Options on quadratic variation (a): Black-Scholes setting, $T = 1$ year. (b): Heston setting, $T = 1$ year.

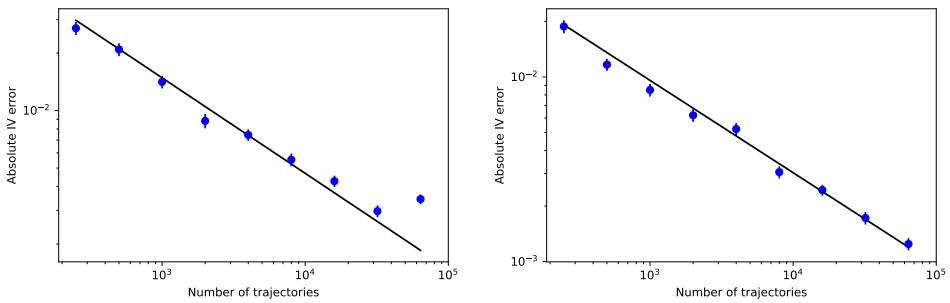
Now, let us discuss the stability of our model as the regularization parameter $\lambda \to 0$. We studied the absolute error in the implied volatility of the 1-year ATM call option for various $\lambda \in [10^{-9}, 1]$ in the Black-Scholes and Heston settings described above. We used $N = 10^6$ trajectories and $L = 100$ Z_js at each step according to (5.8), and performed 100 repetitions at each considered value of λ. The results are presented in Figure 5. The vertical lines in Fig. 5–Fig. 7 denote the standard deviation in the absolute errors of the implied volatilities. We observe that in both settings, initially, the error drops as λ decreases, then it stabilizes 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.



(a) (b)

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

Let's 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 of trajectories N. 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 (2.5)). 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 $\lambda = 10^{-9}$, $L = 100$, $N \in [250, 2^8 \cdot 250]$, and performed 100 repetitions at each value of N. We see the error decreases as $O(N^{-1/2})$ in both settings, which is even better than predicted by theory.



(a) (b)

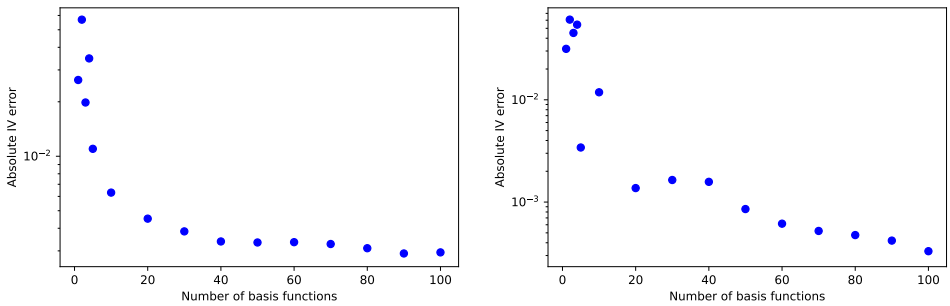
Figure 6: Mean absolute implied volatility error vs number of trajectories. The black line is the approximation: error$= CN^{-1/2}$ (a): Black-Scholes setting; $C = 0.469$. (b): Heston setting; $C = 0.303$.

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 $\lambda = 10^{-9}$, $L = 100$, $N = 10^5$.

Strike K	0.6	0.8	1	1.2	1.4	1.6
$\mathsf{P}(S_T < K)$	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

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

We also investigate the dependence of the error in the implied volatility on the number of basis functions L in the representation (4.5). 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 Figure 7, we plotted the dependence of the absolute error in the implied volatility of 1-year ATM call option on L. We used $N = 10^6$ 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 stabilizes at $L \approx 80$.



(a) (b)

Figure 7: Mean absolute implied volatility error vs number of basis functions. (a): Black-Scholes setting. (b): Heston setting.

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 (4.6) we take as a diffusion

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

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

Fig. 8 below shows fit of the smile for 1-year European call options depending on ε and ε_{CIR}. We use the model of Section 5 with $M = 500$

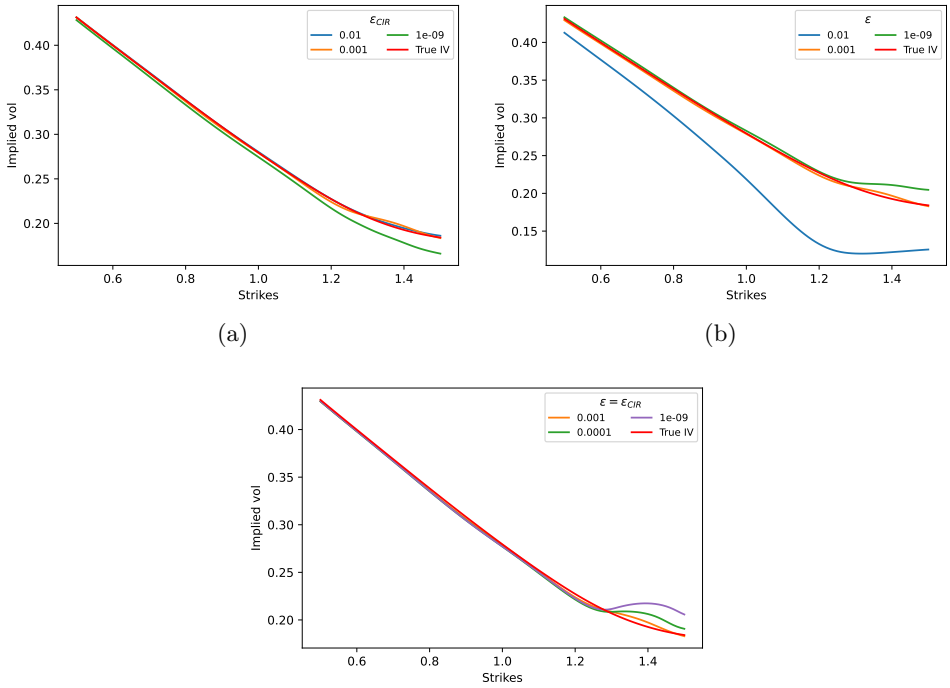


Figure 8: Fit of the smile of 1-year call options for different truncation levels (a): $\varepsilon = 10^{-3}$, ε_{CIR} varies; (b): $\varepsilon_{CIR} = 10^{-3}$, ε varies; (c): $\varepsilon = \varepsilon_{CIR}$ varies.

timesteps 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 [GHL12]. We suggest a novel RKHS based regularization method and prove that this regularization 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 [GHL12]. There are still some questions left open here. First, it remains unclear whether the regularized McKean-Vlasov SDE remains well-posed when the regularization 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 Section 2 and Section 3.

Proof of Proposition 3.3. Since $\mathcal{H}$ is separable, let $I \subset \mathbb{N}$ 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

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

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

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

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

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

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

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

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

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

which shows (3.6). □

Proof of Theorem 2.4. Let us write

$$|m_A^\lambda(x;\mu) - m_A^\lambda(y;\nu)| \le |m_A^\lambda(x;\mu) - m_A^\lambda(x;\nu)| + |m_A^\lambda(x;\nu) - m_A^\lambda(y;\nu)| \\ = I_1 + I_2. \tag{7.4}$$

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)

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

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 they are bounded $\ell_2(I) \to \ell_2(I)$ operators with bounded inverses, it easy to see that

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

Therefore

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

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

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

Hence, by using the identity

$$\sum_{i\in I} e_i(x)e_i(y) = \sum_{i\in I} \big\langle k(x,\cdot), e_i\big\rangle_{\mathcal{H}} \big\langle k(y,\cdot), e_i\big\rangle_{\mathcal{H}} = \big\langle k(x,\cdot), k(y,\cdot)\big\rangle_{\mathcal{H}} = k(x,y), \tag{7.7}$$

we get

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

By the Kantorovich-Rubinstein duality formula ([Vil21]), for every $h : \mathcal{X} \to \mathbb{R}$ with $h \in C^1(\mathcal{X})$ one has

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

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

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

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

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

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

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

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

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

where again Assumption 2.1 was used. Next note that

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

due to Assumption 2.1. Substituting now (7.10), (7.11), and (7.12) into (7.6) and then into (7.5), we finally get

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

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

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

Note that

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

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

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

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 of the theorem. □

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

Proof of Theorem 2.2. It follows from Theorem 2.4, and the assumptions of the theorem, and the fact that $\mathbb{W}_1$-metric can be bounded from above by the $\mathbb{W}_2$-metric, that the drift and diffusion of (2.3) are Lipschitz and satisfy the conditions of [CD16a, Theorem 4.21]. Hence it has a unique strong solution. □

Proof of Theorem 2.3. We see that Theorem 2.4 and the conditions of the theorem implies that all the assumptions of [CD16b, Theorem 2.12] hold (note that the total state dimension is $2d$ in our case). This implies (2.5). $\square$

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

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

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

$$(\mathcal{C}^\nu + \lambda I_{\mathcal{H}})^{-1}\widetilde{a}_i = (\sigma_i + \lambda)^{-1}\widetilde{a}_i. \tag{7.16}$$

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

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

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

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

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

$$P_{\overline{\mathcal{H}}} m_A = \sum_{i\in J}\left\langle \mathsf{E}_{(X,Y)\sim\nu}\left[A(Y)|X=\cdot\right], a_i\right\rangle_{\mathcal{L}_2^\nu} a_i. \tag{7.18}$$

Further, for $i \in J$ we deduce

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

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

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

Thus

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

which is (3.13). Similarly, recalling (3.11), we get

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

which is finite whenever $P_{\overline{\mathcal{H}}}m_A \in \mathcal{H}$, that is, $\sum_{i\in J}\langle m_A, a_i\rangle^2_{\mathcal{L}_2^\nu}\sigma_i^{-1} < \infty$. This shows (3.14). It is easily seen by dominated convergence that the l.h.s. of (3.13) goes to zero, and, in the case $P_{\overline{\mathcal{H}}}m_A \in \mathcal{H}$ the l.h.s. of (3.14) goes to zero as well. □

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(2):423–476, 2002.

[Bac19] Francis Bach. Are all kernels cursed? Available at https://francisbach.com/cursed-kernels/, 2019.

[BDG19] Oleg Butkovsky, Konstantinos Dareiotis, and Máté Gerencsér. Approximation of sdes – a stochastic sewing approach. *arXiv preprint arXiv:1909.07961*, 2019.

[BJ17] Mireille Bossy and Jean-François Jabir. On the wellposedness of some McKean models with moderated or singular diffusion coefficient. In *International Symposium on BSDEs*, pages 43–87. Springer, 2017.

[BS13a] Gerard Brunick and Steven Shreve. Mimicking an Itô process by a solution of a stochastic differential equation. *Ann. Appl. Probab.*, 23(4):1584–1628, 2013.

[BS13b] Gerard Brunick and Steven Shreve. Mimicking an Itô process by a solution of a stochastic differential equation. *Ann. Appl. Probab.*, 23(4):1584–1628, 2013.

[CD16a] René Carmona and François Delarue. *Probabilistic Theory of Mean Field Games with Applications I.* Springer, Probability Theory and Stochastic Modelling 83, 2016.

[CD16b] René Carmona and François Delarue. *Probabilistic Theory of Mean Field Games with Applications II.* Springer, Probability Theory and Stochastic Modelling 84, 2016.

[Dup94] Bruno Dupire. Pricing with a smile. *Risk*, 7:18–20, 1994.

[FO09] Fang Fang and Cornelis W Oosterlee. A novel pricing method for European options based on Fourier-cosine series expansions. *SIAM Journal on Scientific Computing*, 31(2):826–848, 2009.

[Fun84] Tadahisa Funaki. A certain class of diffusion processes associated with nonlinear parabolic equations. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 67(3):331–348, 1984.

[Gat11] Jim Gatheral. *The volatility surface: a practitioner's guide*, volume 357. John Wiley & Sons, 2011.

[GHL11] Julien Guyon and Pierre Henry-Labordere. The smile calibration problem solved. *Available at SSRN 1885032*, 2011.

[GHL12] Julien Guyon and Pierre Henry-Labordère. Being particular about calibration. *Risk Magazine*, 2012.

[Gyo86] István Gyongy. Mimicking the one-dimensional marginal distributions of processes having an Itô differential. *Probab. Theory Relat. Fields*, 71(4):501–516, 1986.

[JM21] Benjamin Jourdain and Stéphane Menozzi. Convergence rate of the Euler-Maruyama scheme applied to diffusion processes with $L_Q - L_\rho$ drift coefficient and additive noise. *arXiv preprint arXiv:2105.04860*, 2021.

[JZ20] Benjamin Jourdain and Alexandre Zhou. Existence of a calibrated regime switching local volatility model. *Mathematical Finance*, 30(2):501–546, 2020.

[LMP22] Vincent Lemaire, Thibaut Montes, and Gilles Pagès. Stationary Heston model: calibration and pricing of exotics using product recursive quantization. *Quant. Finance*, 22(4):611–629, 2022.

[Low08] George Lowther. Fitting martingales to given marginals. *arXiv preprint arXiv:0808.2319*, 2008.

[LSZ20] Daniel Lacker, Mykhaylo Shkolnikov, and Jiacheng Zhang. Inverting the Markovian projection, with an application to local stochastic volatility models. *Annals of Probability*, 48(5):2189–2211, 2020.

[MV16] Yuliya S. Mishura and Alexander Yu. Veretennikov. Existence and uniqueness theorems for solutions of McKean–Vlasov stochastic equations. *arXiv preprint arXiv:1603.02212*, 2016.

[RS80] Michael Reed and Barry Simon. *Functional analysis. Revised and Enlarged Edition.* Methods of Modern Mathematical Physics, Academic Press, 1980.

[SC08] Ingo Steinwart and Andreas Christmann. *Support vector machines.* Springer Science & Business Media, 2008.

[SGF+10] Bharath K. Sriperumbudur, Arthur Gretton, Kenji Fukumizu, Bernhard Schölkopf, and Gert R. G. Lanckriet. Hilbert space embeddings and metrics on probability measures. *J. Mach. Learn. Res.*, 11:1517–1561, 2010.

[SHS01] Bernhard Schölkopf, Ralf Herbrich, and Alex J Smola. A generalized representer theorem. In *International conference on computational learning theory*, pages 416–426. Springer, 2001.

[Sun05] Hongwei Sun. Mercer theorem for RKHS on noncompact sets. *Journal of Complexity*, 21:337 – 349, 2005.

[Vil21] Cédric Villani. *Topics in optimal transportation*, volume 58. American Mathematical Soc., 2021.