Finance and Stochastics (2026) 30:397–439
https://doi.org/10.1007/s00780-026-00588-6

Calibration of local volatility models with stochastic interest rates using optimal transport

Benjamin Joseph[1] · Grégoire Loeper[2] · Jan Obłój[3]

Received: 11 August 2023 / Accepted: 17 February 2025 / Published online: 23 February 2026

Abstract
We develop a nonparametric, semimartingale optimal transport, calibration methodology for local volatility models with stochastic interest rates. The method finds a fully calibrated model which is closest, in a way defined by a general cost function, to a given reference model. We establish a general duality result which allows to solve the problem by optimising over solutions to a second-order fully nonlinear Hamilton–Jacobi–Bellman equation. Our methodology is analogous to Guo et al. (SIAM J. Financ. Math. 13:1–31, 2022; Math. Finance 32:46–77, 2022), but features a novel element of solving for discounted densities, or sub-probability measures. As an example, we apply the method to a sequential calibration problem, where a Vašíček model is already given for the interest rates and we seek to calibrate a stock price's local volatility model with a volatility coefficient depending on time, the underlying and the short rate process, and the two processes driven by possibly correlated Brownian motions. The equity model is calibrated to any number of European option prices.

Keywords Semimartingale optimal transport · Model calibration · Stochastic interest rates · Computational methods · Local volatility

This research has been supported by BNP Paribas Global Markets and the EPSRC Centre for Doctoral Training in Mathematics of Random Systems: Analysis, Modelling and Simulation (EP/S023925/1).

✉ J. Obłój
jan.obloj@maths.ox.ac.uk

B. Joseph
benjamin.joseph@maths.ox.ac.uk

G. Loeper
gregoire.loeper@bnpparibas.com

1 Mathematical Institute and Christ Church, University of Oxford, Oxford, UK

2 BNP Paribas Global Markets, Paris, France

3 Mathematical Institute and St John's College, University of Oxford, Oxford, UK

Mathematics Subject Classification 91G20 · 91G80 · 60H30

JEL Classification G13 · C61 · C63

1 Introduction

Modelling involves inevitable trade-offs: "All models are wrong, some models are useful" as Box [13] put it. Models need to capture the important aspects of the system they represent, but also need to be tractable and analytically and/or numerically solvable. In particular, calibration – picking model parameters which recover known outputs – is an essential part of any modelling process. It is a key challenge faced by financial industry practitioners on a daily basis, as their pricing models need to match market prices of liquid instruments before they can be used to price any bespoke or illiquid products.

In practice, models for a key underlying, such as the S&P500 index, need to be calibrated to a large number of options with different maturities and strikes. This may be nigh impossible for a simple parametric model. Dupire [24] derived a formula to calibrate a local volatility model to an arbitrary number of options, establishing it as a benchmark for modelling equities (see also Atlan [4] for its stochastic interest rate extension). However, it came with its own shortcomings. Its calibration poses serious numerical challenges, see Bain et al. [7], and requires interpolation of the data as Dupire's formula assumes a continuum of prices across strikes and maturities. Moreover, it has been criticised for wrong dynamic behaviour compared to suitable stochastic volatility models; see Hagan et al. [39]. This issue can lead to potentially costly mispricing of exotic products. Naturally, this motivated further research and led in particular to the introduction of local stochastic volatility (LSV) models or forward variance stochastic volatility models; see Bergomi [11] for details and historical references. While successful in many ways, these models also came with shortcomings, often related to an inability to calibrate to a new class of products. Prominent among these was the challenge to calibrate jointly to SPX and VIX options; see Guyon [35] for a literature review and discussion.

Monte Carlo methods to calibrate with Dupire's formula have been used in Deelstra and Rayée [23], Hok and Tan [44], Ögetbil et al. [61]. Another approach is to discretise the nonlinear Fokker–Planck equation and directly solve it by finite differences as in Ren et al. [62] or by finite volume methods as in Engelmann et al. [27] and Wyns and Du Toit [70]. Dupire [25] derives a nonlinear McKean SDE for the local volatility, which in fact recovered the mimicking result of Gyöngy [38] by using a financial argument. This has been solved via Monte Carlo methods developed in Henry-Labordère [43] for an approximate calibration, and an exact calibration by the particle method of Guyon and Henry-Labordère [36, 37]. Cozma et al. [21] used variance reduction techniques to implement the particle method in a four-factor model. All these methods require a continuum of strikes and maturities to construct a surface of implied volatilities, which in practice results in interpolation and extrapolation of market data to build these surfaces. We adopt a different approach without this requirement by formulating the exact calibration of European options as discrete constraints within a convex optimisation problem.

Recently, a nonparametric exact calibration method based on optimal transport has been proposed, which aims to address the above challenges. This method uses semimartingale optimal transport as introduced by Tan and Touzi [67], which is the semimartingale version of the celebrated work of Benamou and Brenier [10] (see also Huesmann and Trevisan [45]). This approach has already been used in several contexts: for local volatility calibration in Guo et al. [33], for local stochastic volatility calibration in Guo et al. [34], for joint VIX/SPX calibration in Guo et al. [31] and for optimal investment in Guo et al. [29]. The most general formulation of the method, pointing to the breadth of its applications, is found in Guo and Loeper [30].

An overview of these results is given in Guo et al. [32]. It is worth noting that whilst the link with optimal transport was not recognised, a variational approach to calibrating local volatility models was first constructed already in Avellaneda et al. [5]. The core contribution of the method is to build a fully calibrated model while trying to preserve desirable features of a given model. In effect, we project a given reference model onto the set of calibrated models. As the calibration constraints depend on one-dimensional marginals, classical mimicking results, see Gyöngy [38] and Brunick and Shreve [18], allow us to restrict to Markovian models. This in turn allows us to use PDE methods to solve the dual problem.

The need for a calibrated interest rate model is common to all financial products involving future payments. While in a very low interest rate environment, which roughly held in the financial markets between 2009 and 2021, one may be tempted to ignore this need for short-dated products, it is no longer feasible for the current market conditions. It is therefore natural to extend the semimartingale optimal transport approach to a setup that includes stochastic rates. Our contribution here is to fill this gap and understand how to develop and calibrate joint models for rates and equities. This extension is not trivial, since the computation of the discount factor involves the whole path of the short rate, which renders the problem path-dependent. We show how to overcome this difficulty. We develop suitable duality results and seek to calibrate a stock price local volatility model with a volatility coefficient depending on time, the underlying and the short rate process, and driven by a Brownian motion which can be correlated with the randomness driving the rates process. In a sequel paper Joseph et al. [51], we consider a simultaneous joint calibration problem. A particular difficulty here is in dealing with the path-dependent discount terms while keeping the number of state variables, and thus the dimension of the problem, at $d = 2$. This is important for computational reasons as the numerical methods rely on solving a nonlinear HJB equation and pricing PDEs, which becomes computationally more difficult by standard techniques in dimension $d \geq 3$. However, our duality result covers any interest rate model, at the expense of higher numerical complexity.

2 Preliminaries and notation

We adopt the setup of Guo et al. [31, 34], who in turn used the formulation of Tan and Touzi [67]. Let E be a Polish space equipped with its Borel σ-algebra, let $C(E)$ be the space of continuous functions on E and $C_b(E)$ the space of continuous bounded functions on E. Let $\mathcal{M}(E)$ be the space of finite signed Borel measures endowed with the weak* topology, $\mathcal{M}_+(E) \subseteq \mathcal{M}(E)$ the subset of nonnegative finite Borel measures, and $\mathcal{P}(E)$ the set of Borel probability measures also under the weak* topology. Note that if E is compact, the topological dual of $C_b(E)$ is given by $C_b(E)^* = \mathcal{M}(E)$; but if E is non-compact, then $C_b(E)^*$ is larger than $\mathcal{M}(E)$. Let $L^1(\mathrm{d}\mu)$ be the space of μ-integrable functions. For clarity, we write $C_b(E;\mathbb{R}^d)$, $\mathcal{M}(E;\mathbb{R}^d)$ and $L^1(\mathrm{d}\mu;\mathbb{R}^d)$ for the vector-valued versions of those spaces (with an analogous notation for the matrix-valued versions). Write $\mathbb{S}^d$ for the set of $d\times d$ symmetric matrices and $\mathbb{S}^d_+ \subseteq \mathbb{S}^d$ for the subset of positive semidefinite symmetric matrices. For $a,b\in\mathbb{R}^d$, write $a\cdot b$ for the inner product $a^\top b$, and for $A,B\in\mathbb{S}^d$, write $A:B$ for their inner product $\mathrm{Tr}(A^\top B)$. As a shorthand, we define $\Lambda := [0,T]\times\mathbb{R}^d$ and $\mathcal{X} := \mathbb{R}\times\mathbb{R}^d\times\mathbb{S}^d$, which will be used for the domain and range of the triple representing the law of a semimartingale, its drift and its volatility. Finally, denote the duality bracket between $C_b(E)$ and $C_b(E)^*$ by $\langle\cdot,\cdot\rangle$, and for a Borel curve $t\mapsto\mu_t$ with $\mu_t\in\mathcal{M}_+(E)$ for $t\in[0,T]$, write $\mathrm{d}\mu_t\mathrm{d}t$ as a shorthand for $\mu_t(\mathrm{d}x)\mathrm{d}t$.

We fix a time horizon $T>0$ and dimension $d>1$. We mainly work on the canonical space $\Omega := C([0,T];\mathbb{R}^d)$ of continuous $\mathbb{R}^d$-valued paths on $[0,T]$, but sometimes we need to work on $C([0,T];\mathbb{R}^{d+1})$. The canonical process is denoted in two ways: either as e, with $e_t(\omega)=\omega_t$, and the dimension is clear from the context, or as X, with $X_t(\omega)=\omega_t$, with the latter restricted to the d-dimensional setting. We take the canonical filtration $\mathbb{F} = (\mathcal{F}_t)_{0\le t\le T}$ as the right-continuous version of the natural filtration generated by e. We consider all probability measures $\mathbb{P}$ on $(\Omega,\mathcal{F}_T)$ such that X is an $(\mathbb{F},\mathbb{P})$-semimartingale with decomposition

$$X_t = X_0 + A_t^{\mathbb{P}} + M_t^{\mathbb{P}}, \qquad t\in[0,T],\ \mathbb{P}\text{-a.s.},$$

where $(M_t^{\mathbb{P}})_{t\ge0}$ is a local $(\mathbb{F},\mathbb{P})$-martingale and $(A_t^{\mathbb{P}})_{t\in[0,T]}$ a finite-variation process.

Definition 2.1 We say that $\mathbb{P}$ is *characterised* by $(\alpha^{\mathbb{P}},\beta^{\mathbb{P}})$ if

$$\alpha_t^{\mathbb{P}} = \frac{\mathrm{d}A_t^{\mathbb{P}}}{\mathrm{d}t}, \quad \beta_t^{\mathbb{P}} = \frac{\mathrm{d}\langle M^{\mathbb{P}}\rangle_t}{\mathrm{d}t} \qquad (\mathrm{d}\mathbb{P}\otimes\mathrm{d}t)\text{-a.e.},$$

where $(\alpha_t^{\mathbb{P}},\beta_t^{\mathbb{P}})_{t\in[0,T]}$ is an $(\mathbb{R}^d\times\mathbb{S}^d)$-valued progressively measurable process.

The existence of progressively measurable $(\alpha_t^{\mathbb{P}},\beta_t^{\mathbb{P}})$ is guaranteed when $A^{\mathbb{P}}$ and $\langle M^{\mathbb{P}}\rangle$ are assumed to be Lebesgue-absolutely continuous $\mathbb{P}$-a.s.; see Jacod and Shiryaev [49, Propositions I.3.13 and II.2.9]. Note that $(\alpha^{\mathbb{P}},\beta^{\mathbb{P}})$ is only determined $(\mathrm{d}\mathbb{P}\otimes\mathrm{d}t)$-almost everywhere. The set of probability measures $\mathbb{P}$ satisfying the conditions above is denoted by $\mathcal{P}$. We note that regular conditional probabilities exist on Ω and use these implicitly; e.g. $\mathbb{E}^{\mathbb{P}}_{t,x}[\alpha_t^{\mathbb{P}}]$ or $\mathbb{E}^{\mathbb{P}}[\alpha_t^{\mathbb{P}}|X_t=x]$ denote the conditional expectation seen as a measurable function of (t,X_t) and evaluated at $X_t=x$. Finally, we note that by Doob's martingale representation theorem (see Karatzas and Shreve [53, Theorem 3.4.2]) there exists a Brownian motion $W^{\mathbb{P}}$, possibly on an enlarged probability space, such that

$$X_t = X_0 + \int_0^t \alpha_s^{\mathbb{P}}\,\mathrm{d}s + \int_0^t (\beta_s^{\mathbb{P}})^{1/2}\,\mathrm{d}W_s^{\mathbb{P}}, \qquad t\le T.$$

The process $X := (\mathbf{S}, X^r)$ is composed of the short rate $(X^r_t)_{t\in[0,T]}$ and a $(d-1)$-dimensional process $(\mathbf{S}_t)_{t\in[0,T]}$ corresponding to the underlying asset, such as the S&P 500, and extra state variables, e.g. extra assets, stochastic factors in the volatility functions, or multi-factors in the short rate. In specific examples in Sect. 4.1, the short rate is also denoted by r. The state variable's x-coordinate corresponding to X^r is denoted by x_r to stress that it represents the short rate. The stochastic discount factor is derived from the short rate as $Y_t := \exp(-\int_0^t X^r_s \,\mathrm{d}s)$, and we refer to (X, Y) as the augmented process. Note that the augmented process can be expressed as an updating function in the sense of Brunick and Shreve [18, Definition 3.1]. We further consider the subset $\mathcal{P}^1 \subseteq \mathcal{P}$ of measures $\mathbb{P} \in \mathcal{P}$ which satisfy

$$\mathbb{E}^{\mathbb{P}}\bigg[\int_0^T (|\alpha^{\mathbb{P}}_t| + |\beta^{\mathbb{P}}_t|)\,\mathrm{d}t\bigg] < \infty \quad \text{and} \quad X^r_t > -1 \qquad (\mathrm{d}\mathbb{P}\otimes\mathrm{d}t)\text{-a.e.} \tag{2.1}$$

We are not aware of any relevant model which would require the short rate to be unbounded from below; so this part of the assumption poses no problem. It implies in particular that $0 < Y_t < \mathrm{e}^T$, $t \in [0, T]$, under all measures considered in this paper. The integrability assumption on $(\alpha^{\mathbb{P}}, \beta^{\mathbb{P}})$ in (2.1) would equally be satisfied under any reasonable model, and it is imposed so that we can apply [18, Theorem 3.6]. These mimicking results, extending earlier works of Krylov [54] and Gyöngy [38], allow us to construct a Markov process with the same one-dimensional marginals as a given semimartingale. As our constraints and cost only depend on these marginals, this allows us to restrict our attention to Markov processes. Markovian projection relies on the classical result that the marginal law of a diffusion process is a distributional solution to the corresponding Fokker–Planck equation, with the converse result given in Figalli [28] where existence and uniqueness results are obtained for the corresponding SDE satisfied by a process with a marginal law that is a weak solution to a Fokker–Planck equation. In addition, since we consider a calibration problem with a stochastic interest rate, to avoid path-dependent terms, we augment the process to include stochastic discount factor terms. For simplicity, our arguments consider a one-factor short rate, but it is easy to adapt to a multi-factor or multi-curve setting.

Theorem 2.2 *Let $\mathbb{P} \in \mathcal{P}^1$ be a candidate model. There exist jointly measurable versions of the conditional expectations $\alpha_t(x, y) = \mathbb{E}^{\mathbb{P}}_{t,x,y}[\alpha^{\mathbb{P}}_t]$, $\beta_t(x, y) = \mathbb{E}^{\mathbb{P}}_{t,x,y}[\beta^{\mathbb{P}}_t]$, and $\tilde{\mathbb{P}} \in \mathcal{P}^1$ such that $\alpha^{\tilde{\mathbb{P}}}_t = \alpha_t(X_t, Y_t)$ and $\beta^{\tilde{\mathbb{P}}}_t = \beta_t(X_t, Y_t)$ $(\mathrm{d}\tilde{\mathbb{P}}\otimes\mathrm{d}t)$-a.e. and $\tilde{\mathbb{P}} \circ (X_t, Y_t)^{-1} = \mathbb{P} \circ (X_t, Y_t)^{-1}$. Moreover, possibly on some other probability space with $\mathbb{P}'$, there exists a Markov process (X, Y) satisfying*

$$\begin{cases} \mathrm{d}X_t = \alpha_t(X_t, Y_t)\,\mathrm{d}t + \big(\beta_t(X_t, Y_t)\big)^{\frac{1}{2}}\,\mathrm{d}W^{\mathbb{P}'}_t, & 0 \le t \le T, \qquad X_0 = x_0, \\ \mathrm{d}Y_t = -Y_t X^r_t\,\mathrm{d}t, & 0 \le t \le T, \qquad Y_0 = 1, \end{cases} \tag{2.2}$$

such that $\mathbb{P}' \circ (X_t, Y_t)^{-1} = \mathbb{P} \circ (X_t, Y_t)^{-1}$ for all $t \in [0, T]$ and where $W^{\mathbb{P}'}$ is a Brownian motion. Finally, the marginal distributions $\rho_t = \rho(t, \cdot) = \mathbb{P} \circ (X_t, Y_t)^{-1}$

are a weak solution to the Fokker–Planck equation

$$\begin{cases} \partial_t \rho_t + \nabla_x \cdot (\rho_t \alpha_t) - \frac{1}{2} \nabla_x^2 : (\rho_t \beta_t) - \partial_y (x_r y \rho_t) = 0 & \text{on } (0,T) \times \mathbb{R}^{d+1}, \\ \rho_0^{\mathbb{P}} = \delta_{X_0, Y_0}, & (x,y) \in \mathbb{R}^{d+1}. \end{cases} \tag{2.3}$$

Definition 2.3 We let $\mathcal{P}^1_{\mathrm{loc}}$ denote the subset of measures in $\mathcal{P}^1$ under which the semimartingale characteristics of X are given as measurable functions in (t, X_t, Y_t).

Remark 2.4 The first assertion in Theorem 2.2 follows from Brunick and Shreve [18, Theorem 7.1]. The second then follows via the martingale representation theorem, and is [18, Theorem 3.6]. The last assertion is obtained using Itô's formula. Note that $\tilde{\mathbb{P}}$ is in $\mathcal{P}^1_{\mathrm{loc}}$ and is simply the distribution of X solving (2.2), i.e., $\tilde{\mathbb{P}} = \mathbb{P}' \circ X^{-1}$. With a slight abuse of language, we refer to $\mathcal{P}^1_{\mathrm{loc}}$ as measures under which (X, Y) is a Markov process solving (2.2).

Remark 2.5 A Borel curve $(\nu_t)_{t\in[0,T]} \subseteq \mathcal{M}(\mathbb{R}^d)$ is narrowly continuous if for every $f \in C_b(\mathbb{R}^d)$, the map $t \mapsto \int_{\mathbb{R}^d} f \,\mathrm{d}\nu_t$ is continuous. Trevisan [68, Remark 2.3] (see also Ambrosio et al. [2, Lemma 8.1.2]) shows that any solution $(\rho_t)_{t\in(0,T)}$ to (2.3) has a narrowly continuous representative $(\tilde{\rho}_t)_{t\in[0,T]}$ such that $\rho_t = \tilde{\rho}_t$ for almost all $t \in (0, T)$. Defining $\mathcal{L}$ as the diffusion operator associated with (2.3), we then also have

$$\int f_{t_2} \,\mathrm{d}\tilde{\rho}_{t_2} - \int f_{t_1} \,\mathrm{d}\tilde{\rho}_{t_1} = \int_{t_1}^{t_2} \int \big(\partial_t f + (\mathcal{L} f)_t\big) \,\mathrm{d}\tilde{\rho}_t \,\mathrm{d}t \qquad \text{for } 0 \le t_1 < t_2 \le T.$$

We therefore may assume without loss of generality that the solution to (2.3) is narrowly continuous.

Our calibration problem below will be written as a minimisation of a cost functional which represents a "distance" to a given favourite reference model and also ensures perfect calibration. We consider a strongly convex cost function $F : \Lambda \times \mathbb{R}^d \times \mathbb{S}^d \to [0, +\infty]$ which is proper and lower semicontinuous in (α, β). It is set to take the value $+\infty$ outside of a set Γ; see for example (4.16). In particular, we always set $F = +\infty$ if $\beta \notin \mathbb{S}^d_+$ to ensure β is a legitimate covariance matrix. We use these properties implicitly, e.g. whenever we restrict to $\beta \in \mathbb{S}^d_+$. We also define the function $\hat{F} : \Lambda \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}^d \to [0, +\infty]$ as $\hat{F}(t, x, y, \alpha, \beta) = f(y) F(t, x, \alpha, \beta)$, where the function f is specified by the discounting term in the payoffs of the option constraints.

The strong convexity of F (see Nesterov [60, Definition 2.1.3]) means that for any subderivative ∇ performed over $(\alpha, \beta) \in \mathbb{R}^d \times \mathbb{S}^d_+$, there exists $C > 0$ such that for all $(t, x, \alpha, \beta, \alpha', \beta') \in \Lambda \times \mathbb{R}^d \times \mathbb{S}^d_+ \times \mathbb{R}^d \times \mathbb{S}^d_+$ with $F(t, x, \alpha, \beta) < \infty$, we have

$$\begin{aligned} F(t, x, \alpha', \beta') \ge F(t, x, \alpha, \beta) &+ \langle \nabla F(t, x, \alpha, \beta), (\alpha' - \alpha, \beta' - \beta) \rangle \\ &+ C(|\alpha - \alpha'|^2 + \|\beta - \beta'\|^2_{\mathrm{Fro}}). \end{aligned} \tag{2.4}$$

Here $\|\cdot\|_{\mathrm{Fro}}$ denotes the Frobenius norm, which for a matrix M is given by $\|M\|_{\mathrm{Fro}} = \sqrt{\sum_{i,j} |m_{i,j}|^2}$. We additionally assume that F is p-coercive, i.e., there exist $p > 1$ and $C > 0$ such that for all $(t, x, \alpha, \beta) \in \Lambda \times \mathbb{R}^d \times \mathbb{S}^d_+$, we have

$$|\alpha|^p + \|\beta\|^p_{\mathrm{Fro}} \le C\big(1 + F(t, x, \alpha, \beta)\big).$$

The Legendre–Fenchel transform of F (see for example Rockafellar [63, §12]) is given by

$$F^*(t, x, a, b) := \sup_{\alpha\in\mathbb{R}^d, \beta\in\mathbb{S}^d_+} \big(\alpha \cdot a + \beta : b - F(t, x, \alpha, \beta)\big),$$

where the supremum is a priori over $(\alpha, \beta) \in \mathbb{R}^d \times \mathbb{S}^d$, but can be restricted to $\beta \in \mathbb{S}^d_+$ by the comment above, or indeed can later be restricted to the set Γ as $F = +\infty$ elsewhere. While F itself need not be differentiable with respect to (α, β), since F is strictly convex in (α, β), we have that $F^*(t, x, a, b)$ is differentiable in (a, b) (see [63, Theorem 26.3]). For convenience, we set $F(\alpha, \beta) := F(t, x, \alpha, \beta)$ and $F^*(a, b) := F^*(t, x, a, b)$.

3 Problem formulation

We want to calibrate our model to market prices of n options. The ith option has maturity $\tau_i \in (0, T)$, payoff function $G_i \in C_b(\mathbb{R}^d; \mathbb{R})$ and price u_i. While European call option payoffs do not satisfy boundedness, as calibrating instruments, they are equivalent to put options via the usual call–put parity, and puts have bounded and continuous payoffs. We let $\tau = (\tau_1, \dots, \tau_n)$, $G(x) = (G_1(x), \dots, G_n(x))$ and $u = (u_1, \dots, u_n)$. We are only interested in modelling up to the horizon covered by the market instruments; so we assume that $T > \max_{i\le n} \tau_i$.

Definition 3.1 Given an initial distribution μ_0, expiry times τ and market prices u corresponding to payoffs G, we introduce the set of *calibrated measures* as

$$\mathcal{P}(\mu_0, \tau, u) = \{\mathbb{P} \in \mathcal{P}^1 : \mathbb{P} \circ X_0^{-1} = \mu_0, \mathbb{E}^{\mathbb{P}}[Y_{\tau_i} G_i(X_{\tau_i})] = u_i \text{ for } i = 1, \dots, n\}.$$

We denote by $\mathcal{P}_{\mathrm{loc}}(\mu_0, \tau, u) = \mathcal{P}(\mu_0, \tau, u) \cap \mathcal{P}^1_{\mathrm{loc}}$ those calibrated models under which the semimartingale characteristics of X are given as measurable functions of (t, X_t, Y_t); see Remark 2.4. We denote by $\widetilde{\mathcal{P}}_{\mathrm{loc}}(\mu_0, \tau, u) \subseteq \mathcal{P}_{\mathrm{loc}}(\mu_0, \tau, u)$ the subset of those measures under which the semimartingale characteristics of X are given as measurable functions of (t, X_t).

We remark that we usually take the initial law μ_0, i.e., the distribution of (X_0, Y_0), as $\delta_{(x_0,1)}$, where $X_0 = x_0$ are the observed initial values of our state variables and $Y_0 = 1$ always.

Our primal problem consists of selecting one of the possible calibrated models in $\mathcal{P}(\mu_0, \tau, u)$ by minimising a cost functional. The following key results assert that when doing so, we can in fact restrict to Markov processes.

Proposition 3.2 *Given an initial distribution μ_0, expiration times τ and market prices u, we have*

$$
\begin{aligned}
V &:= \inf_{\mathbb{P}\in\mathcal{P}(\mu_0,\tau,u)} \mathbb{E}^{\mathbb{P}}\bigg[\int_0^T Y_t F(\alpha_t^{\mathbb{P}}, \beta_t^{\mathbb{P}})\ \mathrm{d}t\bigg] \\
&= \inf_{\mathbb{P}\in\widetilde{\mathcal{P}}_{\mathrm{loc}}(\mu_0,\tau,u)} \mathbb{E}^{\mathbb{P}}\bigg[\int_0^T Y_t F(\alpha_t^{\mathbb{P}}, \beta_t^{\mathbb{P}})\ \mathrm{d}t\bigg]. \qquad (3.1)
\end{aligned}
$$

We prove this in three steps. In Sect. 3.1, we reduce from $\mathbb{P}\in\mathcal{P}(\mu_0,\tau,u)$ to $\mathbb{P}\in\mathcal{P}_{\mathrm{loc}}(\mu_0,\tau,u)$, meaning that it is sufficient to consider Markovian models with drift and diffusion depending on (t, X_t, Y_t). Section 3.2 then introduces a conditioning argument which we refer to as the discounted density transformation. We then prove the "discounted superposition principle", which is an adaptation of Trevisan [68, Theorem 2.5] and shows a correspondence between the solution of the transformed Fokker–Planck equation and the solution of the martingale problem with the discounted coefficients. Section 3.3 then uses the transformation in Sect. 3.2 to reduce from $\mathbb{P}\in\mathcal{P}_{\mathrm{loc}}(\mu_0,\tau,u)$ to $\mathbb{P}\in\widetilde{\mathcal{P}}_{\mathrm{loc}}(\mu_0,\tau,u)$, thus showing that it is sufficient to consider Markovian models with drift and diffusion depending on (t, X_t).

3.1 First Markovian reduction

Similarly to Guo et al. [34, Proposition 3.4], since the market constraints only depend on the marginal distributions and the cost functional is convex, a combination of the Markovian projection in Theorem 2.2 and Jensen's inequality readily shows that we can restrict our attention to Markov process in (t, X_t, Y_t).

Proof of Proposition 3.2 (Step I) Note that $\widetilde{\mathcal{P}}_{\mathrm{loc}}(\mu_0,\tau,u)\subseteq\mathcal{P}(\mu_0,\tau,u)$. So the inequality "$\le$" is trivial, and if $\mathcal{P}(\mu_0,\tau,u)$ is empty, then both sides are $+\infty$. We show now the inequality "$\ge$", but with $\mathcal{P}_{\mathrm{loc}}(\mu_0,\tau,u)$ replacing $\widetilde{\mathcal{P}}_{\mathrm{loc}}(\mu_0,\tau,u)$. Take $\mathbb{P}\in\mathcal{P}(\mu_0,\tau,u)$ and use Theorem 2.2 and Remark 2.4 to find the corresponding $\tilde{\mathbb{P}}\in\mathcal{P}_{\mathrm{loc}}(\mu_0,\tau,u)$ such that (X,Y) has the same marginals under $\mathbb{P}$ and $\tilde{\mathbb{P}}$. Using the tower property and Jensen's inequality as $(\alpha,\beta)\mapsto F(\alpha,\beta)$ is convex, we have

$$
\begin{aligned}
\mathbb{E}^{\mathbb{P}}\bigg[\int_0^T Y_t F(\alpha_t^{\mathbb{P}}, \beta_t^{\mathbb{P}})\ \mathrm{d}t\bigg] &= \mathbb{E}^{\mathbb{P}}\bigg[\int_0^T Y_t \mathbb{E}^{\mathbb{P}}_{t,X_t,Y_t}[F(\alpha_t^{\mathbb{P}}, \beta_t^{\mathbb{P}})]\ \mathrm{d}t\bigg] \\
&\ge \mathbb{E}^{\mathbb{P}}\bigg[\int_0^T Y_t F(\mathbb{E}^{\mathbb{P}}_{t,X_t,Y_t}[\alpha_t^{\mathbb{P}}], \mathbb{E}^{\mathbb{P}}_{t,X_t,Y_t}[\beta_t^{\mathbb{P}}])\ \mathrm{d}t\bigg] \\
&= \mathbb{E}^{\tilde{\mathbb{P}}}\bigg[\int_0^T Y_t F(\alpha_t^{\tilde{\mathbb{P}}}, \beta_t^{\tilde{\mathbb{P}}})\ \mathrm{d}t\bigg].
\end{aligned}
$$

This gives the desired inequality. □

3.2 A "discounted density" transformation and superposition principle

While we could easily reduce to Markovian models in the state process (X, Y), this is not satisfactory. In particular, when we consider in Sect. 5.2 a local volatility model with a short rate, we should obtain a three-dimensional fully nonlinear PDE instead of a two dimensional one, which would entail a substantially increased computational effort. We thus want to further reduce to the Markov process X, i.e., to obtain (3.1). This requires novel arguments to deal with the stochastic discount term. Instead of working with probability measures, we work with discounted densities, i.e., sub-probability measures. When $\mathbb{P} \in \mathcal{P}^1_{\mathrm{loc}}$ is fixed, we write ρ, α, β for $\rho^{\mathbb{P}}, \alpha^{\mathbb{P}}, \beta^{\mathbb{P}}$.

Definition 3.3 Let $\mathbb{P} \in \mathcal{P}^1_{\mathrm{loc}}$ with the semimartingale characteristics of X given by $\alpha^{\mathbb{P}}_t = \alpha_t(X_t, Y_t)$, $\beta^{\mathbb{P}}_t = \beta_t(X_t, Y_t)$. Let $\rho_t(\mathrm{d}x, \mathrm{d}y)$ and $\rho^X_t(\mathrm{d}x)$ be the marginal distribution of (X_t, Y_t) and of X_t, respectively, for $t \in [0, T]$. We also write $\rho_t(\mathrm{d}x, \mathrm{d}y) = \zeta_{t,x}(\mathrm{d}y)\rho^X_t(\mathrm{d}x)$, i.e., $\zeta_{t,x}$ is the law of Y_t conditionally on $X_t = x$. Define the *"discounted" density, drift and volatility* for $(t, x) \in [0, T] \times \mathbb{R}^d$ by

$$\rho^D_t(\mathrm{d}x) = D_t(x)\rho^X_t(\mathrm{d}x), \qquad \text{where } D_t(x) = \int_{\mathbb{R}} y\zeta_{t,x}(\mathrm{d}y), \tag{3.2}$$

$$\alpha^D_t(x) = \frac{\mathbb{E}[Y_t\alpha_t(X_t, Y_t)|X_t = x]}{\mathbb{E}[Y_t|X_t = x]} = \int_{\mathbb{R}} y\alpha_t(x, y)\frac{\zeta_{t,x}(\mathrm{d}y)}{D_t(x)}, \tag{3.3}$$

$$\beta^D_t(x) = \frac{\mathbb{E}[Y_t\beta_t(X_t, Y_t)|X_t = x]}{\mathbb{E}[Y_t|X_t = x]} = \int_{\mathbb{R}} y\beta_t(x, y)\frac{\zeta_{t,x}(\mathrm{d}y)}{D_t(x)}. \tag{3.4}$$

By construction, ρ^D_t inherits narrow continuity from ρ_t. In addition, using (2.1) and (3.2)–(3.4), we have

$$\begin{aligned}
&\int_0^T \int_{\mathbb{R}^d} \big(|\alpha^D_t(x)| + |\beta^D_t(x)|\big)\rho^D_t(\mathrm{d}x)\,\mathrm{d}t \\
&= \int_0^T \int_{\mathbb{R}^{d+1}} \big(|y\alpha_t(x, y)| + |y\beta_t(x, y)|\big)\rho_t(\mathrm{d}x, \mathrm{d}y)\,\mathrm{d}t \\
&\le e^T\mathbb{E}^{\mathbb{P}}\bigg[\int_0^T (|\alpha^{\mathbb{P}}_t| + |\beta^{\mathbb{P}}_t|)\,\mathrm{d}t\bigg] < \infty,
\end{aligned} \tag{3.5}$$

so that α^D and β^D are $(\mathrm{d}t \otimes \rho^D)$-integrable.

A key tool in the dimension reduction is a "discounted" version of the superposition principle of Trevisan [68, Theorem 2.5]. We first define the augmented martingale problem.

Definition 3.4 For $f \in C^{1,2}_b([0, T] \times \mathbb{R}^{d+1}; \mathbb{R})$ and given measurable functions $a : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$, $b : [0, T] \times \mathbb{R}^d \to \mathbb{S}^d_+$ and $c : [0, T] \times \mathbb{R}^d \to \mathbb{R}$, define the operator

$$\mathcal{L}f = a \cdot \nabla_x f + \frac{1}{2}b : \nabla^2_x f - yc\partial_y f.$$

Then $\boldsymbol{\eta} \in \mathcal{P}(C([0,T];\mathbb{R}^{d+1}))$ is a *solution to the augmented martingale problem* aMP(a,b,c) if

$$\int\int_0^T (|a_t|\circ e_t + |b_t|\circ e_t + |c_t|\circ e_t)\,\mathrm{d}t\,\mathrm{d}\boldsymbol{\eta} < \infty \tag{3.6}$$

and for all $f \in C_b^{1,2}([0,T]\times\mathbb{R}^{d+1};\mathbb{R})$, the process

$$[0,T] \ni t \mapsto f(t,\,\cdot\,)\circ e_t - \int_0^t \big(\partial_t f(s,\,\cdot\,) + (\mathcal{L}f)(s,\,\cdot\,)\big)\circ e_s\,\mathrm{d}s$$

is a martingale with respect to the natural filtration on $C([0,T];\mathbb{R}^{d+1})$.

We now state the "discounted superposition principle".

Theorem 3.5 *Let $(\nu_t)_{t\in[0,T]} \in C([0,T];\mathcal{M}(\mathbb{R}^d))$ be a narrowly continuous solution of the general discounted Fokker–Planck equation*

$$\partial_t\nu_t + \nabla_x\cdot\big(a_t(x)\nu_t(x)\big) - \frac{1}{2}\nabla_x^2 : \big(b_t(x)\nu_t(x)\big) + c_t(x)\nu_t(x) = 0 \qquad \text{on } (0,T)\times\mathbb{R}^d,$$

where $a \in L^1(\mathrm{d}\nu_t\,\mathrm{d}t;\mathbb{R}^d)$, $b \in L^1(\mathrm{d}\nu_t\,\mathrm{d}t;\mathbb{S}_+^d)$ and $c \in L^1(\mathrm{d}\nu_t\,\mathrm{d}t;\mathbb{R})$. Then there exists a solution $\boldsymbol{\eta} \in \mathcal{P}(C([0,T];\mathbb{R}^{d+1}))$ to the aMP(a,b,c) *such that for all $t\in[0,T]$, the discounted version of its narrowly continuous curve of marginals $\eta_t = \boldsymbol{\eta}\circ e_t^{-1}$ coincides with ν_t, that is, $\int_{\mathbb{R}} y\eta_t(\,\cdot\,,\mathrm{d}y) = \nu_t$.*

The proof of Theorem 3.5 follows closely the steps in the proof of Trevisan [68, Theorem 2.5], which itself has a similar structure to Ambrosio et al. [2, Theorem 8.2.1], Ambrosio and Figalli [1, Theorem 4.5], Figalli [28, Theorem 2.6], Ambrosio and Trevisan [3, Theorem 7.1]. We defer the proof to the Appendix, but summarise here the main steps. First, in Sect. A.1, we argue that the result holds for smooth coefficients and construct smooth approximating sequences in an analogous way to [68]. Then Sect. A.2 asserts the compactness of solutions for the augmented martingale problem as a direct consequence of [68, Sect. A.2], and establishes convergence via an easy modification of [68, Sect. A.3]. For the generalisation to coefficients which are C^2 in space, the adaptation of the proof is easy with the existence of a solution to aMP(a,b,c) with a given initial law following directly from [68, Theorem A.6], and uniqueness following from a simple adaptation of [68, Theorem A.7], with the approach in obtaining estimates an easy modification. The generalisation to bounded coefficients relies on the de la Vallée-Poussin criterion and [68, Corollary A.5], with our modification again being the handling of the c-term — here we rely on the already established tightness and superposition results for the augmented martingale problem, as it is just the usual martingale problem for (X,Y), and on the fact that Y is strictly positive. The estimates for the x-coordinates are obtained in the same way as in [68], and the estimates for the y-coordinate use the function $\log(y\chi_R)$, where χ_R is some cutoff function in order to eliminate the y-term in the augmented diffusion operator. The case of locally bounded coefficients then follows a similarly modified approach to the case of bounded coefficients as in [68].

3.3 Second Markovian reduction

We use now the tools introduced above to finish the proof of Proposition 3.2.

Proposition 3.6 *Let $\mathbb{P} \in \mathcal{P}^1_{\mathrm{loc}}$ and $\rho^D, \alpha^D, \beta^D$ be given as in Definition 3.3. Then ρ^D_t is a weak solution to the "discounted" version of the Fokker–Planck equation, for $(t,x) \in (0,T) \times \mathbb{R}^d$,*

$$\partial_t \rho^D_t(x) + \nabla_x \cdot \big(\alpha^D_t(x)\rho^D_t(x)\big) - \frac{1}{2}\nabla^2_x : \big(\beta^D_t(x)\rho^D_t(x)\big) + x_r \rho^D_t(x) = 0. \tag{3.7}$$

Moreover, there exists a local volatility model $\tilde{\mathbb{P}} \in \mathcal{P}^1$ under which the semimartingale characteristics of X are given by $\alpha^{\tilde{\mathbb{P}}}_t = \alpha^D_t(X_t)$ and $\beta^{\tilde{\mathbb{P}}}_t = \beta^D_t(X_t)$ $(\mathrm{d}\tilde{\mathbb{P}}\otimes\mathrm{d}t)$-a.e. and such that for any payoff function $G \in C_b(\mathbb{R}^d;\mathbb{R})$ and any $t \in [0,T]$,

$$\mathbb{E}^{\mathbb{P}}[Y_t G(X_t)] = \mathbb{E}^{\tilde{\mathbb{P}}}[Y_t G(X_t)]. \tag{3.8}$$

Furthermore, the value of the objective function decreases, i.e.,

$$\mathbb{E}^{\mathbb{P}}\bigg[\int_0^T Y_t F\big(\alpha_t(X_t,Y_t), \beta_t(X_t,Y_t)\big)\,\mathrm{d}t\bigg] \geq \mathbb{E}^{\tilde{\mathbb{P}}}\bigg[\int_0^T Y_t F\big(\alpha^D_t(X_t), \beta^D_t(X_t)\big)\,\mathrm{d}t\bigg].$$

The notion of solution to (3.7) is in the spirit of [68, Definition 2.2], but with an extra term $-x_r f$ in the diffusion operator of [68, Definition 2.1].

Proof of Proposition 3.2 (Step II) Proposition 3.6 directly gives that

$$\inf_{\mathbb{P}\in\mathcal{P}_{\mathrm{loc}}(\mu_0,\tau,u)} \mathbb{E}^{\mathbb{P}}\bigg[\int_0^T Y_t F(\alpha^{\mathbb{P}}_t, \beta^{\mathbb{P}}_t)\,\mathrm{d}t\bigg] \geq \inf_{\mathbb{P}\in\widetilde{\mathcal{P}}_{\mathrm{loc}}(\mu_0,\tau,u)} \mathbb{E}^{\mathbb{P}}\bigg[\int_0^T Y_t F(\alpha^{\mathbb{P}}_t, \beta^{\mathbb{P}}_t)\,\mathrm{d}t\bigg],$$

and hence we have equality since $\widetilde{\mathcal{P}}_{\mathrm{loc}}(\mu_0,\tau,u) \subseteq \mathcal{P}_{\mathrm{loc}}(\mu_0,\tau,u)$. This completes the proof. □

Proof of Proposition 3.6 Recall the notation of Definition 3.3. Since ρ_t solves (2.3), we have for $(t,x) \in (0,T)\times\mathbb{R}^d$ that

$$\begin{aligned}\int_{\mathbb{R}} y\bigg(&\partial_t \rho_t(x,y) + \nabla_x \cdot \big(\rho_t(x,y)\alpha_t(x,y)\big) - \frac{1}{2}\nabla^2_x : \big(\rho_t(x,y)\beta_t(x,y)\big) \\ &- \partial_y\big(x_r y \rho_t(x,y)\big)\bigg)\,\mathrm{d}y = 0.\end{aligned}$$

By construction, the first term is given by $\int_{\mathbb{R}} y \partial_t \rho_t(x, y)\,\mathrm{d}y = \partial_t \rho_t^D(x)$. For the drift, we then have

$$\begin{aligned}\int_{\mathbb{R}} y \nabla_x \cdot \big(\rho_t(x, y)\alpha_t(x, y)\big)\,\mathrm{d}y &= \nabla_x \cdot \left(\int_{\mathbb{R}} \alpha_t(x, y)\frac{y}{D_t(x)}\zeta_{t,x}(y)\rho_t^X(x)D_t(x)\,\mathrm{d}y\right)\\ &= \nabla_x \cdot \left(\frac{\int_{\mathbb{R}} \alpha_t(x, y) y \zeta_{t,x}(y)\,\mathrm{d}y}{D_t(x)}\rho_t^X(x)D_t(x)\right)\\ &= \nabla_x \cdot \big(\rho_t^D(x)\alpha_t^D(x)\big).\end{aligned}$$

The same calculation gives

$$\int_{\mathbb{R}} y \nabla_x^2 : \big(\rho_t(x, y)\beta_t(x, y)\big)\,\mathrm{d}y = \nabla_x^2 : \big(\rho_t^D(x)\beta_t^D(x)\big).$$

Finally, integration by parts gives $\int_{\mathbb{R}} y \partial_y(x_r y)\rho_t(x, y)\,\mathrm{d}y = -x_r \rho_t^D(x)$, and so ρ_t^D solves (3.7). Recall also that the integrability condition (3.5) holds. We can thus apply Theorem 3.5 with $\nu = \rho^D$ to obtain another probability measure $\tilde{\boldsymbol{\rho}} \in \mathcal{P}(C([0, T]; \mathbb{R}^{d+1}))$ solving $\mathrm{aMP}(\alpha^D, \beta^D, x_r)$ and with the property that $\tilde{\rho}_t^D(\cdot) = \int_{\mathbb{R}} y \tilde{\rho}_t(\cdot, \mathrm{d}y) = \rho^D(\cdot)$. By Guo and Loeper [30, Lemma 3.4], for the canonical process $(\tilde{X}, \tilde{Y})$ under $\tilde{\boldsymbol{\rho}}$, the coordinate $\tilde{X}$ is a local volatility process with the semimartingale characteristics $\alpha_t^D(\tilde{X}_t)$ and $\beta_t^{\tilde{\mathbb{P}}} = \beta_t^D(\tilde{X}_t)$, and we further have $\tilde{Y}_t = \tilde{Y}_0 \exp(-\int_0^t \tilde{X}_s^r\,\mathrm{d}s)$ $(\mathrm{d}t \otimes \mathrm{d}\tilde{\boldsymbol{\rho}})$-a.e. We let $\tilde{\mathbb{P}}$ be the projection of $\tilde{\boldsymbol{\rho}}$ on the first d coordinates. Note that $\tilde{\mathbb{P}} \in \mathcal{P}^1$ by (3.6), X has the desired semimartingale characteristics under $\tilde{\mathbb{P}}$ and also $\tilde{\rho}_t$ is the distribution of (X_t, Y_t) under $\tilde{\mathbb{P}}$. Then, given a payoff function $G \in C_b(\mathbb{R}^d; \mathbb{R})$, we have for any fixed $t \in [0, T]$ that

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}[Y_t G(X_t)] = \int_{\mathbb{R}^{d+1}} y G(x)\rho_t(x, y)\,\mathrm{d}x\,\mathrm{d}y &= \int_{\mathbb{R}^d} G(x)\rho_t^D(x)\,\mathrm{d}x\\ &= \int_{\mathbb{R}^d} G(x)\tilde{\rho}_t^D(x)\,\mathrm{d}x\\ &= \int_{\mathbb{R}^{d+1}} y G(x)\tilde{\rho}_t(x, y)\,\mathrm{d}x\,\mathrm{d}y\\ &= \mathbb{E}^{\tilde{\mathbb{P}}}[Y_t G(X_t)]\end{aligned}$$

so that (3.8) holds. Finally, we also have

$$\begin{aligned}
&\mathbb{E}^{\mathbb{P}}\left[\int_0^T Y_t F\big(\alpha_t(X_t, Y_t), \beta_t(X_t, Y_t)\big)\,\mathrm{d}t\right]\\
&= \int_0^T \int_{\mathbb{R}^{d+1}} y F\big(\alpha_t(x,y), \beta_t(x,y)\big)\zeta_{t,x}(\mathrm{d}y)\rho_t^X(\mathrm{d}x)\,\mathrm{d}t\\
&= \int_0^T \int_{\mathbb{R}^{d+1}} F\big(\alpha_t(x,y), \beta_t(x,y)\big)\frac{y}{D_t(x)}\zeta_{t,x}(\mathrm{d}y)\, D_t(x)\rho_t^X(\mathrm{d}x)\,\mathrm{d}t\\
&\ge \int_0^T \int_{\mathbb{R}^d} F\left(\frac{\int_{\mathbb{R}} y\alpha_t(x,y)\zeta_{t,x}(\mathrm{d}y)}{D_t(x)}, \frac{\int_{\mathbb{R}} y\beta_t(x,y)\zeta_{t,x}(\mathrm{d}y)}{D_t(x)}\right) D_t(x)\rho_t^X(\mathrm{d}x)\,\mathrm{d}t\\
&= \int_0^T \int_{\mathbb{R}^d} F\big(\alpha_t^D(x), \beta_t^D(x)\big)\rho_t^D(\mathrm{d}x)\,\mathrm{d}t\\
&= \int_0^T \int_{\mathbb{R}^d} F\big(\alpha_t^D(x), \beta_t^D(x)\big)\tilde{\rho}_t^D(\mathrm{d}x)\,\mathrm{d}t \qquad (3.9)\\
&= \mathbb{E}^{\tilde{\mathbb{P}}}\left[\int_0^T Y_t F\big(\alpha_t^D(X_t), \beta_t^D(X_t)\big)\,\mathrm{d}t\right]. \qquad (3.10)
\end{aligned}$$

□

4 The dual problem

Proposition 3.2 asserted that without any loss of generality, we can restrict our attention to $\mathbb{P} \in \widetilde{\mathcal{P}}_{\mathrm{loc}}(\mu_0, \tau, u)$. By conditioning on X_t, we can express the value function as an integral against the discounted density, as shown above in (3.9), (3.10). Accordingly, **from now on**, we always work with the discounted versions of the density, drift and volatility. We drop the superscript D for notational ease. The problem we need to solve on the right-hand side of (3.1) is thus given equivalently as follows:

Problem 4.1 The value function for the primal problem is given by

$$V := \inf_{\rho,\alpha,\beta} \int_0^T \int_{\mathbb{R}^d} F\big(\alpha_t(x), \beta_t(x)\big)\rho_t(\mathrm{d}x)\,\mathrm{d}t, \tag{4.1}$$

where the infimum is taken over

$$(\rho, \alpha, \beta) \in C\big([0,T]; \mathcal{M}(\mathbb{R}^d)\big) \times L^1(\mathrm{d}\rho_t\,\mathrm{d}t; \mathbb{R}^d) \times L^1(\mathrm{d}\rho_t\,\mathrm{d}t; \mathbb{S}^d),$$

subject to the constraints, in the sense of distributions for $(t,x) \in (0,T) \times \mathbb{R}^d$,

$$0 = \partial_t \rho_t(x) + \nabla_x\big(\rho_t(x)\alpha_t(x)\big] - \frac{1}{2}\nabla_x^2 : [\rho_t(x)\beta_t(x)] + x_r\rho_t(x),$$

$$\int_{\mathbb{R}^d} G_i(x)\rho_{\tau_i}(\mathrm{d}x) = u_i \qquad \text{for } i = 1, \dots, n,$$

$$\rho_0(\cdot) = \mu_0.$$

We note that Problem 4.1 is in fact an example of unbalanced optimal transport, with the source term in the continuity equation being $-x_r\rho$; see Chizat et al. [19], Séjourné et al. [65]. To solve this constrained optimisation problem, we use a duality method inspired by Guo et al. [31, 34], Huesmann and Trevisan [45]. This is similar in spirit to the duality argument used in Brenier [16, Proposition 2.7], which considers the dual of a relaxation of the minimal geodesic problem as an alternative approach to solving the Euler equations of an incompressible fluid. This technique has previously been applied in the proof of Brenier [15, Theorem 3.2] when looking for the existence of variational solutions to "homogenised vortex sheet equations". The technique is also used to formulate the dual of a variational problem involving the Euler–Poisson system of equations in Loeper [57]. The proof relies mainly on the Fenchel–Rockafellar duality theorem and an adjustment to make the problem convex. As our primal problem is quite similar to [34], the approach used there can be adapted to our setting. The following result uses the notion of viscosity solution from Definition 4.8 below.

From now on, we denote by BV-$C_b^2([0,T]\times\mathbb{R}^d)$ the space of functions $[0,T]\times\mathbb{R}^d\ni(t,x)\mapsto f(t,x)\in\mathbb{R}$ such that f is of bounded variation in $t\in[0,T]$ and C_b^2 in $x\in\mathbb{R}^d$.

Theorem 4.2 *The dual expression for the value function V is*

$$V=\sup_{\lambda\in\mathbb{R}^n}\left(\lambda\cdot u-\int_{\mathbb{R}^d}\varphi^\lambda(0,x)\,\mathrm{d}\mu_0\right),\tag{4.2}$$

where $\varphi^\lambda=\varphi$ is the viscosity solution to the HJB equation, for $(t,x)\in(0,T)\times\mathbb{R}^d$,

$$\partial_t\varphi-x_r\varphi+\sum_{i=1}^n\lambda_iG_i(x)\delta_{\tau_i}+F^*\left(\nabla_x\varphi,\frac{1}{2}\nabla_x^2\varphi\right)=0\tag{4.3}$$

with terminal condition $\varphi(T,\,\cdot\,)=0$. If V is finite, then the infimum in Problem 4.1 is attained. If the supremum is attained for some $\lambda^\star\in\mathbb{R}^n$, with the corresponding $\varphi^{\lambda^*}\in$ BV-$C_b^2([0,T]\times\mathbb{R}^d)$ solving the corresponding HJB equation with terminal condition $\varphi^{\lambda^*}(T,\,\cdot\,)=0$, and with $(\rho^*,\alpha^*,\beta^*)$ as the solution of Problem 4.1, then (α^*,β^*) is given in terms of φ^{λ^*} by

$$(\alpha_t^*,\beta_t^*)=\nabla F^*\left(\nabla_x\varphi^*(t,\,\cdot\,),\frac{1}{2}\nabla_x^2\varphi^*(t,\,\cdot\,)\right)\qquad \mathrm{d}\rho_t^*\,\mathrm{d}t\text{-almost everywhere.}$$

Remark 4.3 Solving the dual problem in Theorem 4.2 yields the primal optimisers (α^*,β^*). These characterise the distribution of the Markov process (X,Y) solving (2.2) with these coefficients, i.e., ρ^* solves (2.3) and for the associated measure $\mathbb{P}^*$ on Ω, we have $\rho_t^*=\mathbb{P}^*\circ(X_t,Y_t)^{-1}$. We can then compute the associated discount factor $D_t^*(x)$ and apply the transformation (3.2) to obtain that $D_t^*(x)\rho_t^*(x)$ is a solution of (3.7).

Remark 4.4 It is presently not known if the supremum in (4.2) is attained for a generic cost function F — in general, this is an open problem in (semi)martingale optimal transport. In the static case of martingale OT, Beiglböck et al. [9] show that dual attainment is delicate and depends on convexity and growth properties of the cost in the second marginal. In the Bass martingale setting of Backhoff-Veraguas et al. [6, Sect. 7], i.e., under the assumption of irreducibility of the marginals, the authors establish the attainability of the static dual corresponding to a specific weak martingale OT problem. This static problem is equivalent to the dynamic formulation of Huesmann and Trevisan [45], Tan and Touzi [67]; see also Joseph et al. [52].

We prove Theorem 4.2 in the remainder of this section through a series of lemmas. Our first observation is that the objective function (4.1) is not jointly convex in (ρ, α, β). We define the measures $\mathcal{A} := \rho\alpha$, $\mathcal{B} := \rho\beta$ so that $\mathcal{A}$ and $\mathcal{B}$ are absolutely continuous with respect to ρ. Then the objective function is convex in $(\rho, \mathcal{A}, \mathcal{B})$ with constraints that are affine in $(\rho, \mathcal{A}, \mathcal{B})$. The convexity of the objective with respect to $(\rho, \mathcal{A}, \mathcal{B})$ follows from the classical result that the function $\bar{f}(z_1, z_2, z_3) := z_3 f(\frac{z_1}{z_3}, \frac{z_2}{z_3})$ is convex in (z_1, z_2, z_3) whenever f is convex in (z_1, z_2) on the set $\{z_3 > 0\}$. Note also that we write $\mathrm{d}\rho$ for $\rho_t(\mathrm{d}x)\,\mathrm{d}t$, $\mathrm{d}\mathcal{A}$ for $\alpha_t(x)\rho_t(\mathrm{d}x)\,\mathrm{d}t$ and $\mathrm{d}\mathcal{B}$ for $\beta_t(x)\rho_t(\mathrm{d}x)\,\mathrm{d}t$. Moreover, the constraints in Problem 4.1 can be formulated in the weak sense as

$$\begin{aligned}\int_\Lambda \Big(\partial_t\varphi \mathrm{d}\rho + \nabla\varphi\cdot \mathrm{d}\mathcal{A} + \frac{1}{2}\nabla^2\varphi : \mathrm{d}\mathcal{B} - x_r\varphi \mathrm{d}\rho\Big) + \int_{\mathbb{R}^d}\varphi \mathrm{d}\mu_0 &= 0, \qquad (4.4)\\ \int_\Lambda \sum_{i=1}^n \lambda_i G_i(x)\delta_{\tau_i}\mathrm{d}\rho - \sum_{i=1}^n \lambda_i u_i &= 0,\end{aligned}$$

for any smooth compactly supported test function $\varphi \in C_c^\infty(\Lambda)$ with $\varphi(T, \cdot) = 0$ and $\lambda \in \mathbb{R}^n$. The terminal condition on φ in Theorem 4.2 arises when integrating by parts to derive (4.4), since we need the boundary term $\rho_T(\cdot)$ to vanish as we do not have a priori knowledge of $\rho_T(\cdot)$. Therefore we can write Problem 4.1 as the following saddle point problem.

Problem 4.5

$$\begin{aligned}V = \inf_{\rho,\mathcal{A},\mathcal{B}} \sup_{\varphi,\lambda} \bigg(\int_\Lambda \Big(F\Big(\frac{\mathrm{d}\mathcal{A}}{\mathrm{d}\rho}, \frac{\mathrm{d}\mathcal{B}}{\mathrm{d}\rho}\Big)\mathrm{d}\rho - \partial_t\varphi \mathrm{d}\rho - \nabla\varphi\cdot\mathrm{d}\mathcal{A} - \frac{1}{2}\nabla^2\varphi : \mathrm{d}\mathcal{B} + x_r\varphi\mathrm{d}\rho\Big)\\ - \int_{\mathbb{R}^d}\varphi\mathrm{d}\mu_0 - \int_\Lambda \sum_{i=1}^n \lambda_i G_i(x)\delta_{\tau_i}\mathrm{d}\rho + \sum_{i=1}^n \lambda_i u_i\bigg),\end{aligned}$$

where the supremum is taken over $(\varphi, \lambda) \in C_c^\infty(\Lambda; \mathbb{R}) \times \mathbb{R}^n$ and the infimum over $(\rho, \alpha, \beta) \in C([0, T]; \mathcal{M}(\mathbb{R}^d)) \times L^1(\mathrm{d}\rho_t\,\mathrm{d}t; \mathbb{R}^d) \times L^1(\mathrm{d}\rho_t\,\mathrm{d}t; \mathbb{S}^d)$, $\mathcal{A} = \rho\alpha$, $\mathcal{B} = \rho\beta$.

We now want to find a functional equal to the objective function on the right-hand side of (4.1), and another that is the remainder of the right-hand side of V inside the infimum in Problem 4.5. To do this, we use the following terminology from Huesmann and Trevisan [45, Proof of Theorem 4.3]. Define BV_T-$C_b^2([0, T] \times \mathbb{R}^d)$ as the set of $\varphi \in \mathrm{BV}$-$C_b^2([0, T] \times \mathbb{R}^d)$ such that $\varphi(T, \cdot) = 0$.

Definition 4.6 We say that the triple $(\gamma, a, b) \in C_b(\Lambda; \mathcal{X})$ is *represented* by $(\varphi, \lambda) \in \mathrm{BV}_T\text{-}C_b^2([0,T] \times \mathbb{R}^d) \times \mathbb{R}^n$ if

$$\begin{aligned} \gamma + \partial_t \varphi - x_r \varphi + \sum_{i=1}^n \lambda_i G_i(x) \delta_{\tau_i} &= 0, \\ a + \nabla \varphi &= 0, \\ b + \frac{1}{2} \nabla^2 \varphi &= 0. \end{aligned}$$

Since $(\gamma, a, b) \in C_b(\Lambda; \mathcal{X})$, the presence of the Dirac delta functions gives that $t \mapsto \varphi(t, \cdot)$ is of bounded variation on $[0, T]$ with jump discontinuities at $t = \tau_i$, which we denote as $\varphi \in \mathrm{BV}_T\text{-}C_b^2([0,T] \times \mathbb{R}^d)$ since we require φ to be at least C^2 in space. Proceeding in an analogous way as in Guo et al. [34], to obtain the dual problem, first define functionals $\Phi : C_b(\Lambda; \mathcal{X}) \to \mathbb{R} \cup \{+\infty\}$ and $\Psi : C_b(\Lambda; \mathcal{X}) \to \mathbb{R} \cup \{+\infty\}$ by

$$\Phi(\gamma, a, b) = \begin{cases} 0, & \text{if } \gamma + F^*(a,b) \le 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

$$\Psi(\gamma, a, b) = \begin{cases} \int_{\mathbb{R}^d} \varphi(0, x)\, \mathrm{d}\mu_0 - \sum_{i=1}^n \lambda_i u, & \text{if } (\gamma, a, b) \text{ is represented by } (\varphi, \lambda), \\ +\infty, & \text{otherwise.} \end{cases}$$

Lemma 4.7 *The objective value V can be expressed in terms of Φ and Ψ as*

$$V = \inf_{\rho, \mathcal{A}, \mathcal{B}} \big(\Phi^*(\rho, \mathcal{A}, \mathcal{B}) + \Psi^*(\rho, \mathcal{A}, \mathcal{B}) \big),$$

where the infimum is taken over $(\rho, \mathcal{A}, \mathcal{B}) \in C_b(\Lambda; \mathcal{X})^*$.

We remark that switching from functions $\varphi \in C_c^\infty(\Lambda)$ (with $\varphi(T, \cdot) = 0$) to functions $\varphi \in \mathrm{BV}_T\text{-}C_b^2([0,T] \times \mathbb{R}^d)$ does not change the value of the supremum; this will be formalised by the notion of viscosity solutions later in Definition 4.8.

Proof of Lemma 4.7 We first evaluate Φ^* on $(\rho, \mathcal{A}, \mathcal{B}) \in \mathcal{M}(\Lambda; \mathcal{X})$ such that $\rho \in \mathcal{M}_+(\Lambda)$ and $\mathcal{A}, \mathcal{B} \ll \rho$ to obtain, by arguing as in [34, Lemma A.1] to justify the exchange of integral and supremum, that

$$\begin{aligned} \Phi^*(\rho, \mathcal{A}, \mathcal{B}) &= \sup_{\gamma + F^*(a,b) \le 0} \int_\Lambda \left(\gamma + a \cdot \frac{\mathrm{d}\mathcal{A}}{\mathrm{d}\rho} + b : \frac{\mathrm{d}\mathcal{B}}{\mathrm{d}\rho} \right) \mathrm{d}\rho \\ &= \sup_{a,b} \int_\Lambda \left(a \cdot \frac{\mathrm{d}\mathcal{A}}{\mathrm{d}\rho} + b : \frac{\mathrm{d}\mathcal{B}}{\mathrm{d}\rho} - F^*(a,b) \right) \mathrm{d}\rho \\ &= \int_\Lambda \sup_{a,b} \left(a \cdot \frac{\mathrm{d}\mathcal{A}}{\mathrm{d}\rho} + b : \frac{\mathrm{d}\mathcal{B}}{\mathrm{d}\rho} - F^*(a,b) \right) \mathrm{d}\rho \\ &= \int_\Lambda F\left(\frac{\mathrm{d}\mathcal{A}}{\mathrm{d}\rho}, \frac{\mathrm{d}\mathcal{B}}{\mathrm{d}\rho} \right) \mathrm{d}\rho, \end{aligned}$$

Furthermore, the arguments in [34, Lemma A.1] also show that $\Phi^* = \infty$ for $(\rho, \mathcal{A}, \mathcal{B}) \in C_b(\Lambda; \mathcal{X})^*$ which are not of the form above. In summary, we thus have

$$\Phi^*(\rho, \mathcal{A}, \mathcal{B}) = \begin{cases} \int_\Lambda F(\frac{\mathrm{d}\mathcal{A}}{\mathrm{d}\rho}, \frac{\mathrm{d}\mathcal{B}}{\mathrm{d}\rho})\,\mathrm{d}\rho, & \text{if } (\rho, \mathcal{A}, \mathcal{B}) \in \mathcal{M}(\Lambda, \mathcal{X}), \\ & \quad \rho \in \mathcal{M}_+(\Lambda), \mathcal{A}, \mathcal{B} \ll \rho, \\ +\infty, & \text{otherwise.} \end{cases} \tag{4.5}$$

We now compute $\Psi^* : C_b(\Lambda; \mathcal{X})^* \to \mathbb{R} \cup \{+\infty\}$ to obtain

$$\begin{aligned}
&\Psi^*(\rho, \mathcal{A}, \mathcal{B}) \\
&= \sup_{(\gamma, a, b) \in C_b(\Lambda; \mathcal{X})} \left(\langle (\gamma, a, b), (\rho, \mathcal{A}, \mathcal{B}) \rangle - \int_{\mathbb{R}^d} \varphi(0, x) \mathrm{d}\mu_0 + \sum_{i=1}^n \lambda_i u_i \right) \\
&= \sup_{\varphi, \lambda} \Bigg(\bigg\langle \Big(x_r \varphi - \partial_t \varphi - \sum_{i=1}^n \lambda_i G_i(x) \delta_{\tau_i}, -\nabla \varphi, -\frac{1}{2} \nabla^2 \varphi \Big), (\rho, \mathcal{A}, \mathcal{B}) \bigg\rangle \\
&\qquad - \int_{\mathbb{R}^d} \varphi(0, x) \mathrm{d}\mu_0 + \sum_{i=1}^n \lambda_i u_i \Bigg) \\
&= \sup_{\varphi, \lambda} \Bigg(\int_\Lambda \Big(x_r \varphi \mathrm{d}\rho - \partial_t \varphi \mathrm{d}\rho - \sum_{i=1}^n \lambda_i G_i(x) \delta_{\tau_i} \mathrm{d}\rho - \nabla_x \varphi \mathrm{d}\mathcal{A} - \frac{1}{2} \nabla_x^2 \varphi : \mathrm{d}\mathcal{B} \Big) \\
&\qquad - \int_{\mathbb{R}^d} \varphi(0, x) \mathrm{d}\mu_0 + \sum_{i=1}^n \lambda_i u_i \Bigg),
\end{aligned} \tag{4.6}$$

where (γ, a, b) are represented by (φ, λ) in the sense of Definition 4.6. Since Φ^* is independent of (φ, λ), we can thus simply add (4.5) and (4.6) to get the argument of the infimum in Problem 4.5 so that

$$V = \inf_{(\rho, \mathcal{A}, \mathcal{B}) \in \mathcal{M}(\Lambda; \mathcal{X})} \big(\Phi^*(\rho, \mathcal{A}, \mathcal{B}) + \Psi^*(\rho, \mathcal{A}, \mathcal{B}) \big),$$

where the restrictions $\rho \in \mathcal{M}_+(\Lambda)$, $\mathcal{A}, \mathcal{B} \ll \rho$ are ensured automatically since otherwise $\Phi^* = +\infty$. To finish the proof, we invoke Guo et al. [34, Lemma A.2] to conclude that

$$\begin{aligned}
V &= \inf_{(\rho, \mathcal{A}, \mathcal{B}) \in \mathcal{M}(\Lambda; \mathcal{X})} \big(\Phi^*(\rho, \mathcal{A}, \mathcal{B}) + \Psi^*(\rho, \mathcal{A}, \mathcal{B}) \big) \\
&= \inf_{(\rho, \mathcal{A}, \mathcal{B}) \in C_b(\Lambda; \mathcal{X})^*} \big(\Phi^*(\rho, \mathcal{A}, \mathcal{B}) + \Psi^*(\rho, \mathcal{A}, \mathcal{B}) \big).
\end{aligned}$$

□

Now we apply the Fenchel–Rockafellar duality theorem as formulated in Villani [69, Theorem 1.9]. We first note that as the constraints in the functionals Φ and Ψ are affine, the functionals are clearly convex in (γ, a, b). We now check the conditions of the theorem at the point $(0, O_{d\times 1}, O_{d\times d})$ which is represented by

$(\varphi, \lambda) = (e^{tx_r}, O_{n\times 1})$, where O refers to the zero matrix of appropriate dimension. Since F is nonnegative,

$$F^*(O_{d\times 1}) = - \inf_{(\alpha,\beta)\in\mathbb{R}^d\times\mathbb{S}^d_+} F(\alpha, \beta) \le 0.$$

Therefore we have $\Phi(0, O_{d\times 1}, O_{d\times d}) = 0$, and $(0, O_{d\times 1}, O_{d\times d})$ is a point of continuity of Φ since F^* is continuous. Moreover, since $\varphi(t, x) = e^{tx_r}$ and μ_0 is a probability measure,

$$\Psi(0, O_{d\times 1}, O_{d\times d}) = \int_{\mathbb{R}^d} \varphi(0, x)\, \mathrm{d}\mu_0 = 1.$$

Thus as Ψ is finite and Φ is finite and continuous at $(0, O_{d\times 1}, O_{d\times d})$ and both take values in $(-\infty, +\infty]$ we may apply the Fenchel–Rockafellar duality theorem and obtain that

$$\begin{aligned}
&\inf_{(\gamma,a,b)\in C_b(\Lambda;\mathcal{X})} \big(\Phi(-\gamma, -a, -b) + \Psi(\gamma, a, b)\big)\\
&= \sup_{(\rho,\mathcal{A},\mathcal{B})\in C_b(\Lambda;\mathcal{X})^*} \big(- \Phi^*(-\rho, -\mathcal{A}, -\mathcal{B}) - \Psi^*(-\rho, -\mathcal{A}, -\mathcal{B})\big)\\
&= \sup_{(\rho,\mathcal{A},\mathcal{B})\in C_b(\Lambda;\mathcal{X})^*} \big(- \Phi^*(\rho, \mathcal{A}, \mathcal{B}) - \Psi^*(\rho, \mathcal{A}, \mathcal{B})\big)\\
&= - \inf_{(\rho,\mathcal{A},\mathcal{B})\in C_b(\Lambda;\mathcal{X})^*} \big(\Phi^*(\rho, \mathcal{A}, \mathcal{B}) + \Psi^*(\rho, \mathcal{A}, \mathcal{B})\big).
\end{aligned}$$

Thus we obtain after rearranging that

$$\begin{aligned}
V &= \inf_{(\rho,\mathcal{A},\mathcal{B})\in C_b(\Lambda;\mathcal{X})^*} \big(\Phi^*(\rho, \mathcal{A}, \mathcal{B}) + \Psi^*(\rho, \mathcal{A}, \mathcal{B})\big)\\
&= \sup_{(\gamma,a,b)\in C_b(\Lambda;\mathcal{X})} \big(- \Phi(-\gamma, -a, -b) - \Psi(\gamma, a, b)\big)\\
&= \sup_{\substack{(\gamma,a,b)\in C_b(\Lambda;\mathcal{X})\\ -\gamma+F^*(-a,-b)\le 0\\ (\gamma,a,b) \text{ is represented by } (\varphi,\lambda)}} \left(\sum_{i=1}^n \lambda_i u_i - \int_{\mathbb{R}^d} \varphi(0, x)\, \mathrm{d}\mu_0\right)\\
&= \sup_{\substack{(\lambda,\varphi)\in\mathbb{R}^n\times \mathrm{BV}_T\text{-}C_b^2([0,T]\times\mathbb{R}^d)\\ \partial_t\varphi - x_r\varphi + \sum_{i=1}^n \lambda_i G_i(x)\delta_{\tau_i} + F^*(\nabla_x\varphi, \frac{1}{2}\nabla_x^2\varphi)\le 0}} \left(\sum_{i=1}^n \lambda_i u_i - \int_{\mathbb{R}^d} \varphi(0, x)\, \mathrm{d}\mu_0\right). \qquad (4.7)
\end{aligned}$$

Now, to obtain equality in the HJB equation constraint and thus the HJB equation (4.3) and the dual formulation in Theorem 4.2, we adapt the classical notion of viscosity solutions from Lions [55] to include the required jump discontinuities, in analogy to Guo et al. [34]. First define disjoint intervals $I_k := [\tau_{k-1}, \tau_k)$ with $\tau_0 = 0$, with $\bigcup_{k=1}^n I_k = [0, T)$.

Definition 4.8 For any $\lambda \in \mathbb{R}^n$, we say $\varphi \in \mathrm{BV}_T\text{-}C_b^2([0,T]\times\mathbb{R}^d)$ is a *viscosity subsolution (supersolution)* of (4.3) if $\varphi|_{I_k\times\mathbb{R}^d} \in C_b(I_k \times \mathbb{R}^d)$ is a classical (continuous) viscosity subsolution (supersolution) of (4.3) in $I_k \times \mathbb{R}^d$ and for all $k = 1, \dots, n$ has jump discontinuities

$$\varphi(t,x) = \varphi(t-,x) - \sum_{i=1}^{n} \lambda_i G_i(x) \mathbb{1}_{\{t=\tau_i\}},$$

with terminal condition $\varphi(T,\cdot) = 0$. In addition, $\varphi \in \mathrm{BV}_T\text{-}C_b^2([0,T]\times\mathbb{R}^d)$ is a *viscosity solution of* (4.3) if it is both a viscosity subsolution and viscosity supersolution of (4.3).

The expression for V in (4.7) involves supersolutions to (4.3), and the first step is to show that we can restrict to viscosity solutions, as stated in the first part of Theorem 4.2. For this, we follow the proof of [34, Proposition 3.5]. [34, Remark 3.9] provides a comparison principle, which is the classical comparison principle of viscosity solutions applied to φ on $I_k \times \mathbb{R}^d$ for each k. Using this, one can deduce existence and uniqueness of solutions to (4.3) via Crandall et al. [22, Theorem 4.1]. The comparison principle also implies that V in (4.2) is smaller than the supremum over viscosity solutions. We then use the smoothing argument from Bouchard et al. [12], which shows that any viscosity solution of (4.3) can be approached by smooth supersolutions. This together with (4.7) shows that V is larger than the supremum over viscosity solutions. This allows us to conclude the proof of the first point of Theorem 4.2, namely that the dual of Problem 4.1 can be expressed as (4.2), where φ solves the HJB equation (4.3) with terminal condition $\varphi(T,\cdot) = 0$.

We now seek to obtain the form of the optimal (α, β) for the second part of Theorem 4.2.

Lemma 4.9 *If the supremum is attained for some $\lambda^\star \in \mathbb{R}^n$, with the corresponding $\varphi^{\lambda^*} \in \mathrm{BV}_T\text{-}C_b^2([0,T]\times\mathbb{R}^d)$ solving the corresponding HJB equation, and with $(\rho^*, \alpha^*, \beta^*)$ the solution of Problem 4.1, then (α^*, β^*) is given in terms of φ^{λ^*} by*

$$(\alpha_t^*, \beta_t^*) = \nabla F^*\left(\nabla_x \varphi^*(t,\cdot), \frac{1}{2}\nabla_x^2 \varphi^*(t,\cdot)\right) \qquad \mathrm{d}\rho_t^* \mathrm{d}t\text{-almost everywhere}$$

Proof If (α^*, β^*) is the solution of Problem 4.1, then $(\rho^*, \rho^*\alpha^*, \rho^*\beta^*)$ also achieves the infimum in Problem 4.5. Assume that λ^* is the optimiser of (4.2) with φ^{λ^*} solving (4.3). Then λ^* achieves the supremum in Problem 4.5 and using the optimisers of Problem 4.1 and (4.2), we may rewrite Problem 4.5 as

$$V = \int_\Lambda \Bigg(F(\alpha^*, \beta^*) - \partial_t \varphi^* - \nabla_x \varphi^* \cdot \alpha^* - \frac{1}{2}\nabla_x^2 \varphi^* : \beta^* + x_r \varphi^* \\ - \sum_{i=1}^{n} \lambda^* G_i(x) \delta_{\tau_i} \Bigg) \mathrm{d}\rho^* - \int_{\mathbb{R}^d} \varphi^* \, \mathrm{d}\mu_0 + \sum_{i=1}^{n} \lambda_i^* u_i. \tag{4.8}$$

Since φ^*, λ^* are optimal, we have from Theorem 4.2 that

$$V = \sum_{i=1}^{n} \lambda_i^* u_i - \int_{\mathbb{R}^d} \varphi^* \, \mathrm{d}\mu_0.$$

Therefore (4.8) is equivalent to

$$\begin{aligned} 0 &= \int_\Lambda \Big(F(\alpha^*, \beta^*) - \partial_t \varphi^* - \nabla_x \varphi^* \cdot \alpha^* - \frac{1}{2} \nabla_x^2 \varphi^* : \beta^* + x_r \varphi^* \\ &\qquad - \sum_{i=1}^{n} \lambda^* G_i(x) \delta_{\tau_i} \Big) \, \mathrm{d}\rho^* \\ &= \int_\Lambda \Big(F(\alpha^*, \beta^*) + F^*\Big(\nabla_x \varphi^*, \frac{1}{2} \nabla_x^2 \varphi^*\Big) \\ &\qquad - \nabla_x \varphi^* \cdot \alpha^* - \frac{1}{2} \nabla_x^2 \varphi^* : \beta^* \Big) \, \mathrm{d}\rho^*. \end{aligned} \tag{4.9}$$

Now define $(\bar{\alpha}, \bar{\beta})$ as

$$(\bar{\alpha}, \bar{\beta}) = \nabla F^* \Big(\nabla_x \varphi^*, \frac{1}{2} \nabla_x^2 \varphi^* \Big).$$

Note that

$$\begin{aligned} (\bar{\alpha}, \bar{\beta}) &= \nabla F^* \Big(\nabla_x \varphi^*, \frac{1}{2} \nabla_x^2 \varphi^* \Big) \\ &= \operatorname*{arg\,max}_{(a,b) \in \mathbb{R}^d \times \mathbb{S}_+^d} \Big(a \cdot \nabla_x \varphi^* + b : \frac{1}{2} \nabla_x^2 \varphi^* - F(a, b) \Big). \end{aligned} \tag{4.10}$$

Thus

$$F^* \Big(\nabla_x \varphi^*, \frac{1}{2} \nabla_x^2 \varphi^* \Big) = \nabla_x \varphi^* \cdot \bar{\alpha} + \frac{1}{2} \nabla_x^2 \varphi^* : \bar{\beta} - F(\bar{\alpha}, \bar{\beta}).$$

Since F is convex, its Legendre transform is an involution (see Rockafellar [63, Corollary 12.2.1]); so taking the double Legendre transform, we have

$$(A, B) := \nabla F(\bar{\alpha}, \bar{\beta}) = \nabla F^{**}(\bar{\alpha}, \bar{\beta}) = \operatorname*{arg\,max}_{(a,b) \in \mathbb{R}^d \times \mathbb{S}_+^d} \big(a \cdot \bar{\alpha} + b : \bar{\beta} - F^*(a, b) \big).$$

Therefore we get

$$\begin{aligned} F^{**}(\bar{\alpha}, \bar{\beta}) &= A \cdot \bar{\alpha} + B : \bar{\beta} - F^*(A, B) \\ &= A \cdot \bar{\alpha} + B : \bar{\beta} - \max_{(x,y) \in \mathbb{R}^d \times \mathbb{S}_+^d} \big(A \cdot x + B : y - F(x, y) \big). \end{aligned}$$

Since $F(\bar\alpha,\bar\beta) = F^{**}(\bar\alpha,\bar\beta)$, we have from (4.10) that for the affine terms to cancel, we need $(A,B) = (\nabla_x\varphi^*, \frac{1}{2}\nabla_x^2\varphi^*)$. So substituting into (4.9) gives

$$0 = \int_\Lambda \left(F(\alpha^*,\beta^*) - F(\bar\alpha,\bar\beta) - \nabla_x\varphi^*\cdot(\alpha^*-\bar\alpha) - \frac{1}{2}\nabla_x^2\varphi^* : (\beta^*-\bar\beta)\right) \mathrm{d}\rho^*. \quad (4.11)$$

Since $F(\alpha,\beta)$ is assumed to be strongly convex in (α,β), we have from (2.4) that for some constant $C>0$,

$$F(\alpha^*,\beta^*) - F(\bar\alpha,\bar\beta) \geq \langle \nabla F(\bar\alpha,\bar\beta), (\alpha^*-\bar\alpha, \beta^*-\bar\beta)\rangle + C(|\alpha^*-\bar\alpha|^2 + \|\beta^*-\bar\beta\|_{\mathrm{Fro}}^2).$$

Applying this inequality to (4.11) and noting that $\nabla F(\bar\alpha,\bar\beta) = (\nabla_x\varphi^*, \frac{1}{2}\nabla_x^2\varphi^*)$ gives

$$0 \geq \int_\Lambda C(|\alpha^*-\bar\alpha|^2 + \|\beta^*-\bar\beta\|_{\mathrm{Fro}}^2)\,\mathrm{d}\rho^* \geq 0.$$

Therefore we have $(\alpha^*,\beta^*) = (\bar\alpha,\bar\beta)$ $\mathrm{d}\rho^*$-almost everywhere. □

4.1 Sequential calibration setup

We now specify the setting in which we seek to apply Theorem 4.2. We first specify the state variables of our model: we consider $d=2$, $X_t = (r_t, Z_t)$, where r_t is the short rate and Z_t is the log-stock price. We start with a setting in which a model for the short rate is fixed and has already been calibrated. We refer to this as a "sequential calibration" problem. Our aim is then to calibrate a local volatility model for the stock price, where the local volatility function can depend on the short rate. For this, we employ the OT methodology developed above. The joint calibration problem in which both the short rate and the stock price are calibrated simultaneously using the OT methodology, and in particular the short-rate model parameters can depend on the stock price, is considered in our sequel paper Joseph et al. [51].

Specifically, we now take a given pre-calibrated Hull–White model for the interest rate, see Hull and White [46], and local volatility dynamics for the log-price; so

$$\mathrm{d}Z_t = r_t - \frac{1}{2}\sigma^2(t,Z_t,r_t) + \sigma(t,Z_t,r_t)\,\mathrm{d}W_t^1, \quad (4.12)$$

$$\mathrm{d}r_t = \big(b(t) - ar_t\big)\,\mathrm{d}t + \sigma_r\,\mathrm{d}W_t^2, \quad (4.13)$$

$$\mathrm{d}\langle W^1, W^2\rangle_t = \xi(t,Z_t,r_t)\,\mathrm{d}t, \quad (4.14)$$

where $a>0$, $\sigma_r>0$ are constants and $b(\cdot)$ is a function of time, calibrated so that the dynamics of $(r_t)_{0\leq t\leq T}$ match the market data (e.g. suitable interest rate caps and floors). Both a and σ_r being positive constants is not a particularly restrictive constraint as remarked in Brigo and Mercurio [17, Sect. 3.3] and Hull and White [47]. Our aim is to calibrate $\sigma(t,Z_t,r_t)$ and $\xi(t,Z_t,r_t)$ using our OT methodology. In order to do this, we want to find a cost function that forces $\alpha_t^{\mathbb{P}}$ and $\beta_t^{\mathbb{P}}$ to take the form (4.12)–(4.14) above. As discussed before, we achieve this by using a functional form for F as long as $(\alpha,\beta)\in\Gamma$ for some convex set Γ, with $F=+\infty$ otherwise. The set Γ will enforce in particular that β is positive semidefinite and symmetric and

$\beta_{22} = \sigma_r^2$, which defines a set that is convex in β. With no further constraints on β_{12}, we obtain a non-explicit form of the Legendre transform F^* of the cost function, which in general requires a numerical solver at each (t, x). To reduce the computational difficulty of the numerical solution, as in Guo et al. [34], we first restrict the correlation to a form which still keeps the set Γ convex, namely

$$\xi(t, Z_t, r_t) = \frac{\sigma_r}{\sigma(t, Z_t, r_t)} \xi_{\text{ref}}(t, Z_t, r_t) \qquad \text{for } t \in [0, T], \tag{4.15}$$

where $\xi_{\text{ref}}(t, Z_t, r_t)$ valued in $\mathbb{R}$ is a fixed reference function. We then set

$$\Gamma(t, Z_t, r_t) = \Big\{ (\alpha, \beta) \in \mathbb{R}^2 \times \mathbb{S}^2 : \alpha_1 = r_t - \frac{1}{2}\beta_{11}, \alpha_2 = b(t) - a r_t, \\ \beta_{12} = \beta_{21} = \xi_{\text{ref}} \sigma_r^2, \beta_{22} = \sigma_r^2 \Big\}, \tag{4.16}$$

where one could change the set Γ suitably if one wishes to use a different short-rate model. We also require the inequality $\xi_{\text{ref}}^2(t, Z_t, r_t)\sigma_r^2 \leq \sigma^2(t, Z_t, r_t)$ to keep $\xi(t, Z_t, r_t) \in [-1, 1]$ as a correlation should be. Since r_t is on a much lower scale than Z_t, we have $\sigma_r^2 \ll \sigma^2$; so the above condition on ξ_{ref} is not financially restrictive. To enforce it, we define a convex function $H : \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ with a parameter $p > 2$ by

$$H(x, \bar{x}, s) := \begin{cases} (p-1)(\frac{x-s}{\bar{x}-s})^{1+p} + (p+1)(\frac{x-s}{\bar{x}-s})^{1-p} - 2p, & \text{if } x, \bar{x} > s, \\ +\infty, & \text{otherwise.} \end{cases}$$

The coefficients of each term ensure that $H(x, \bar{x}, s)$ is minimised over x at $x = \bar{x}$ with $\min_x H(x, \bar{x}, s) = 0$. We fix a reference local volatility function $\bar{\sigma}^2(t, Z_t, r_t)$ that represents the desired model. Our aim is to find a calibrated model which does not deviate too much from the reference one. To achieve this, we set

$$F(t, z, r, \alpha, \beta) = \begin{cases} H(\beta_{11}, \bar{\sigma}^2, \xi_{\text{ref}}^2 \sigma_r^2), & \text{if } (\alpha, \beta) \in \Gamma(t, z, r), \\ +\infty, & \text{otherwise.} \end{cases} \tag{4.17}$$

Remark 4.10 It is easy to check that the function F defined in (4.17) satisfies the assumptions for the duality result of Theorem 4.2.

The cost function F ensures that we retain the Hull–White model in the interest rate, while also matching the market prices for call options by calibrating the volatility of the stock. Additionally, we wish to enforce that the matrix β from our model characteristics is positive definite and that ξ_{ref} remains a correlation function, and we achieve this by setting $s = \xi_{\text{ref}}^2 \sigma_r^2$ as an argument of H in the definition of F. Applying Theorem 4.2, we have the following dual formulation with the given cost function $F(t, z, r, \alpha, \beta)$.

Problem 4.11 In the Hull–White setting, the dual formulation of Theorem 4.2 becomes

$$V = \sup_{\lambda} \big(\lambda \cdot u - \varphi^{\lambda}(0, Z_0, r_0)\big),$$

where $\varphi^{\lambda} = \varphi(t, z, r)$ solves the HJB equation, for $(t, z, r) \in (0, T) \times \mathbb{R}^2$,

$$\begin{aligned}
&\sum_{i=1}^{n} \lambda_i \big(\exp(z) - K_i\big)^{+} \delta_{\tau_i} \\
&+ \partial_t \varphi + \sup_{\beta_{11}} \Bigg(\Big(r - \frac{1}{2}\beta_{11}\Big)\partial_z \varphi + (b(t) - ar)\partial_r \varphi + \frac{1}{2}\beta_{11}\partial^2_{zz}\varphi \\
&\qquad + \xi_{\mathrm{ref}}\sigma_r^2 \partial^2_{zr}\varphi + \frac{1}{2}\sigma_r^2 \partial^2_{rr}\varphi - r\varphi - H(\beta_{11}, \bar{\sigma}^2, \xi^2_{\mathrm{ref}}\sigma_r^2)\Bigg) = 0. \qquad (4.18)
\end{aligned}$$

Lemma 4.12 *The optimal characteristic β^*_{11} in the HJB equation (4.18) is given by*

$$\begin{aligned}
&\beta^*_{11}(t, z, r) \\
&= \xi^2_{\mathrm{ref}}\sigma_r^2 + \Bigg(\frac{(\bar{\sigma}^2 - \xi^2_{\mathrm{ref}}\sigma_r^2)^p(\varphi_{zz} - \varphi_z)}{4(p^2 - 1)} \\
&\qquad + \frac{1}{2}\sqrt{\Big(\frac{(\bar{\sigma}^2 - \xi^2_{\mathrm{ref}}\sigma_r^2)^p(\varphi_{zz} - \varphi_z)}{2(p^2 - 1)}\Big)^2 + 4(\bar{\sigma}^2 - \xi^2_{\mathrm{ref}}\sigma_r^2)^{2p}}\Bigg)^{\frac{1}{p}}. \qquad (4.19)
\end{aligned}$$

Proof By differentiating the argument of the supremum in (4.18) with respect to β_{11}, we notice that solving for β_{11} in (4.18) is equivalent to solving for x the equation

$$\frac{1}{2}(\partial^2_{zz}\varphi - \partial_z\varphi) = \partial_x H(x, \bar{\sigma}^2, \xi^2_{\mathrm{ref}}\sigma_r^2).$$

Computing the right-hand term and rearranging, we have

$$\frac{\partial}{\partial x} H(x, \bar{\sigma}^2, \xi^2_{\mathrm{ref}}\sigma_r^2) = (p^2 - 1)\Bigg(\Big(\frac{x - \xi^2_{\mathrm{ref}}\sigma_r^2}{\bar{\sigma}^2 - \xi^2_{\mathrm{ref}}\sigma_r^2}\Big)^p - \Big(\frac{x - \xi^2_{\mathrm{ref}}\sigma_r^2}{\bar{\sigma}^2 - \xi^2_{\mathrm{ref}}\sigma_r^2}\Big)^{-p}\Bigg).$$

We again rearrange and arrive at the equation, quadratic in $(x - \xi^2_{\mathrm{ref}}\sigma_r^2)^p$,

$$(x - \xi^2_{\mathrm{ref}}\sigma_r^2)^{2p} - (x - \xi^2_{\mathrm{ref}}\sigma_r^2)^p(\bar{\sigma}^2 - \xi^2_{\mathrm{ref}}\sigma_r^2)^p \frac{\varphi_{zz} - \varphi_z}{2(p^2 - 1)} - (\bar{\sigma}^2 - \xi^2_{\mathrm{ref}}\sigma_r^2)^{2p} = 0.$$

Solving this and taking the positive root since $x > \xi^2_{\mathrm{ref}}\sigma_r^2$ gives the desired answer. □

Once the optimal β^*_{11} is obtained, the full model dynamics are specified and we can compute the model price $\psi(0, z, r)$ of an instrument with payoff $\tilde{G}$ and maturity

$\tilde{\tau} \le T$ by solving the standard pricing PDE, for $(t, z, r) \in [0, \tilde{\tau}) \times \mathbb{R}^2$,

$$\begin{cases} 0 = \partial_t \psi + \left(r - \frac{1}{2}\beta_{11}^*\right)\partial_z \psi + \big(b(t) - ar\big)\partial_r \psi + \frac{1}{2}\beta_{11}^* \partial_{zz}^2 \psi \\ \qquad + \overline{\xi}\sigma_r^2 \partial_{zr}^2 \psi + \sigma_r^2 \partial_{rr}^2 \psi - r\psi, \\ \psi(\tilde{\tau}, z, r) = \tilde{G}(z, r). \end{cases} \tag{4.20}$$

If we denote by $\mathbb{P}^*$ the measure corresponding to the dynamics of the selected model, then the Feynman–Kac formula gives

$$\mathbb{E}^{\mathbb{P}^*}[e^{-\int_0^{\tilde{\tau}} r_s \, \mathrm{d}s} \tilde{G}(Z_{\tilde{\tau}}, r_{\tilde{\tau}}) | Z_0 = z, r_0 = r] = \psi(0, z, r).$$

5 Numerical method

In this section, we outline the numerical method used to solve the dual formulation in Theorem 4.2. We use and adapt the methods presented in Guo et al. [34] and Guo et al. [31] for ease of implementation. Analogously to those two works, in order to compute the supremum in Theorem 4.2, we must solve an HJB equation for a given λ and then update λ by using an optimisation algorithm. To speed up the convergence, we compute the gradient with respect to λ of the dual objective function.

Lemma 5.1 *Suppose Problem 4.1 is admissible and define the dual objective function as*

$$L(\lambda) = \lambda \cdot u - \varphi^{\lambda}(0, Z_0, r_0). \tag{5.1}$$

Then for each $i = 1, \dots, n$, the derivatives of the dual objective function are given by

$$\partial_{\lambda_i} L(\lambda) = u_i - \mathbb{E}^{\mathbb{P}}[e^{-\int_0^{\tau_i} r_s \, \mathrm{d}s} G_i(Z_{\tau_i})]. \tag{5.2}$$

The gradient is obtained via the same method in [34, Lemma 4.5], but with the stochastic discount factor appearing inside the expectation in (5.2) as a result of the term $-r\varphi$ in the HJB equation, and from an application of the Feynman–Kac formula. The interpretation is the same here that the gradient represents the difference between the model and market prices. We briefly outline the numerical method from [34] and [31] which can be directly applied here. We are first given an initial guess λ, which is usually taken to be a zero vector, and then solve (4.18) to obtain $\varphi(0, Z_0, r_0)$. Since the HJB equation (4.18) has Dirac deltas at the times τ_i, $i = 1, \dots, n$, the solution also has jump discontinuities in time. However, between these jump discontinuities, the solution is continuous and hence can be found by using standard techniques backward in time, and then the jump discontinuities can be incorporated into a terminal condition at the expiration time. In other words, with all of the expiries ordered $0 < \bar{\tau}_1 < \cdots < \bar{\tau}_N$, we take the jump discontinuity at the last option expiry $\bar{\tau}_N$ as a terminal condition and solve the HJB equation up to the next calibrating option expiry $\bar{\tau}_{N-1}$. We then add the jump discontinuity corresponding to the payoffs of the options expiring at $\bar{\tau}_{N-1}$ to the solution of the HJB equation at $\bar{\tau}_{N-1}+$ and treat it as a new terminal condition at $\bar{\tau}_{N-1}$, and proceed in the same way between each of the $\bar{\tau}_i$. Then given $\varphi(0, Z_0, r_0)$, we calculate the objective value (5.1) and apply an optimisation algorithm to update λ to a new guess. As in [34] and [31], the L-BFGS algorithm of Liu and Nocedal [56] displayed good convergence, and with the gradient calculated in Lemma 5.1, the convergence can be made faster. We remark that to compute (5.2), the term $\mathbb{E}^{\mathbb{P}}[e^{-\int_0^{\tau_i} r_s \, \mathrm{d}s} G_i(Z_{\tau_i})]$ is simply the model price of the ith option. This expectation can be computed either via Monte Carlo methods or via standard pricing PDE techniques by solving (4.20) with β_{11}^* obtained from the solution of the HJB equation. This process is then repeated until the sup norm is sufficiently small, i.e., $\|\nabla_\lambda L(\lambda)\|_\infty < \varepsilon$ for some specified tolerance $\varepsilon > 0$, which corresponds to the model prices matching all the market prices.

We perform a constant rescaling $r_t \mapsto Rr_t$ in the short-rate variable, where we choose $R = 100$ for stability reasons in the finite-difference approximations to make it of the same order as the other coordinate. In addition, we rescale the calibrating option prices and their payoffs by their vegas computed from their Black–Scholes implied volatility. This not only helps the stability of the numerical method by reducing the magnitude of the jump discontinuities, but also converts pricing errors into implied volatility errors since the vega represents how much the option price changes as the volatility changes by 1%.

5.1 Numerically solving the HJB equation

We now outline how to solve the HJB equation (4.18). In Barles and Souganidis [8], it has been shown that monotonicity, stability and consistency of a numerical scheme guarantees convergence locally uniformly as long as there exists a comparison principle for the analytic solution. We use a policy iteration method (see Ma and Forsyth [58]) similar to that in [34] and [31], where we solve the HJB equation by using an implicit finite-difference method, with central-difference approximations for the spatial derivatives from In't Hout and Foulon [48]. We choose a boundary far away enough such that the boundary conditions do not have a significant effect on the HJB equation solution, and our boundary conditions are such that the second derivative of φ does not change with time between each maturity. That is, for a subsequence τ_{i_k}, $k = 1, \dots, m$, of the calibrating option maturity times such that all the τ_{i_k} are distinct, with $\tau_{i_0} = 0$, we set for $t \in (\tau_{i_{k-1}}, \tau_{i_k}]$ and each $k = 1, \dots, m$

$$\partial^2_{zz}\varphi(t, z, r) = \partial^2_{zz}\varphi(\tau_i, z, r) \qquad \text{for } z \in \{z_{\min}, z_{\max}\}, r \in [r_{\min}, r_{\max}], \tag{5.3}$$

$$\partial^2_{rr}\varphi(t, z, r) = \partial^2_{rr}\varphi(\tau_i, z, r) \qquad \text{for } z \in [z_{\min}, z_{\max}], r \in \{r_{\min}, r_{\max}\}. \tag{5.4}$$

We started with this boundary condition and compared it with constant Dirichlet boundary conditions $\varphi(t, x) = C_{i_k}$ for some constant $C_{i_k} \in \mathbb{R}$, all $t \in (\tau_{i_{k-1}}, \tau_{i_k}]$ and x on the boundary of our rectangular domain, and observed no significant change. In addition, we use the boundary conditions (5.3), (5.4) in the linearised pricing PDE (4.20). At each time step, we start with the value of φ at the previous time step, approximate the PDE coefficients α and β by using Lemmas 4.9 and 4.12, and

Algorithm 1: Policy iteration algorithm

Data: Input an initial λ and market prices u_i
Result: Calibrated model prices, optimal characteristics
while $\|\nabla_\lambda L(\lambda)\|_\infty > \varepsilon_1$ **do**
  `/* Solve the HJB equation backward in time */`
  for $k = N-1, \ldots, 0$ **do**
    `/* Terminal Conditions - adding λ multiplied by the payoff */`
    if $t_{k+1} = \tau_i$ for some $i = 1, \ldots, n$ **then**
      $\varphi_{t_{k+1}} \leftarrow \varphi_{t_{k+1}} + \sum_{i=1}^n \lambda_i G_i \mathbb{1}_{\{t_{k+1}=\tau_i\}}$
    end
    `/* Policy iteration to approximate the optimal characteristics */`
    $\varphi_{t_k}^{\text{new}} \leftarrow \varphi_{t_{k+1}}$ `// Approximate using previous time step`
    while $|\varphi_{t_k}^{\text{new}} - \varphi_{t_k}^{\text{old}}| > \varepsilon_2$ **do**
      $\varphi_{t_k}^{\text{old}} \leftarrow \varphi_{t_k}^{\text{new}}$ `// Store the old value of φ`
      Approximate β_{11}^* using Lemmas 4.9 and 4.12 with $\varphi_{t_k}^{\text{old}}$ `// Use old values to approximate optimal characteristics`
      Use β_{11}^* to compute α_1^* and β_{12}^*, then plug into (4.18) to remove the supremum and solve using one step of an implicit finite-difference method, and set the solution to $\varphi_{t_k}^{\text{new}}$
    end
    $\varphi_{t_k} \leftarrow \varphi_{t_k}^{\text{new}}$ `// Save the result once φ has converged to the solution`
  end
  `/* Computing the model prices and gradient */`
  Compute the model prices by solving the pricing PDE (4.20) using the ADI method
  Compute the gradient in (5.2)
  Use the L-BFGS algorithm to update λ
end

then solve the HJB equation at that time step using one step of a fully implicit spatially second-order finite-difference scheme. We then use Lemmas 4.9 and 4.12 again at the same time step to approximate the PDE coefficients again with the new value of φ. This process is repeated until convergence within some specified tolerance, after which we proceed to the next time step. Once we have solved the HJB equation at all time steps, we then use the value of β_{11} computed from solving the HJB equation, compute β_{12} from the convexity adjustment, and use them to solve the linearised model pricing PDE via the ADI method to generate the model prices. In the simulated data example, we use a discretisation on a uniform 100×100 spatial grid of $[4, 5] \times [0, 5]$ for the log-stock and rescaled interest rates, and partition the time interval into year fractions at the resolution of one day so that $\mathrm{d}t = \frac{1}{365}$. We use the minimisation package `minfunc` of Schmidt [64] for the implementation of L-BFGS.

The above numerical method is summarised in Algorithm 1. We choose a discretisation $0 = t_0 < t_1 < \cdots < t_N = T$ of the time interval $[0, T]$ such that the expiration times τ_i, $i = 1, \ldots, n$, of the calibrating instruments are a subset of the discretisation times. We let ε_1 be the tolerance of the model prices to the calibrating prices so that $\|\nabla_\lambda L(\lambda)\|_\infty < \varepsilon_1$, where the gradient is given in Lemma 5.1, and we let ε_2 be the tolerance of the policy iteration for approximating the optimal characteristics.

5.2 Numerical results

We present numerical results showcasing the performance of our proposed calibration method. We use simulated data to investigate the advantages and drawbacks of our method, and in particular its dependence on the reference model $\bar{\sigma}$ in (4.17). We use the CEV (constant elasticity of variance) model of Cox [20], with different sets of parameters for the model used to generate the option prices and for our reference model. The underlying $S_t = \exp(Z_t)$, $0 \leq t \leq T$, thus solves the stochastic differential equation

$$\mathrm{d}S_t = r_t S_t \,\mathrm{d}t + \sigma(t, S_t) S_t \,\mathrm{d}W_t^1$$

with $\sigma(t, S_t) = \sigma S_t^{\gamma-1}$, where $\sigma > 0$ and $\gamma \geq 0$ are both constants. We remark that this is a special case of the SABR ("stochastic α, β, ρ") model derived in Hagan et al. [39] as a stochastic volatility extension of the CEV model.

We solve a pricing PDE to compute the generating model prices and consider the following instruments:

1) calls on the underlying with an expiration of 60 days;
2) calls on the underlying with an expiration of 120 days.

We note that the payoffs of these options are not smooth and cause instabilities when computing the derivatives in the terminal conditions of the HJB equation. Thus we use a smoothed version of the call option payoffs, which for $\varepsilon \ll K$ is given by

$$(S_T - K)^+ \approx \frac{S_T - K}{2}\left(\tanh\left(\frac{S_T - K}{\varepsilon}\right) + 1\right).$$

We keep the interest-rate model parameters the same throughout since they are assumed to be given.

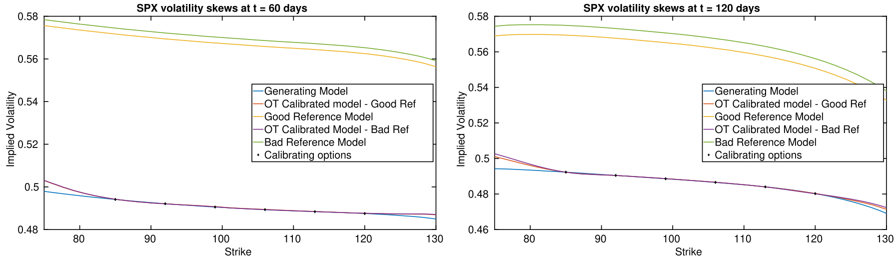
We present two numerical examples — the "good" reference model where the initial reference model is parametrically close to the generating model, and the "bad" reference model where the initial reference model is not parametrically close to the generating model and in particular has a correlation with a different sign. Note that in the extreme case when the generating model and the reference model are the same, the calibration procedure will stop instantly and recover the generating model. The parameters for all models are summarised in Table 1.

Table 1 Parameter values for the simulated data example

Parameter	Value	Interpretation
Z_0	$\log 92$	Initial log-underlying price
r_0	0.025×100	Initial short rate scaled by $R = 100$
ε_1	10^{-4}	Tolerance for the difference in model and market implied volatility
ε_2	10^{-8}	Tolerance for the policy iteration approximation of the optimal characteristics
p	4	Exponent in the cost function
σ	0.78	Volatility scaling of generating CEV model
γ	0.9	Power law in generating CEV model
$\theta(t)$	$ar_0 + \frac{\sigma_r^2}{2a}(1 - e^{-2at})$	Initial term structure of Hull–White generating model
a	0.4	Speed of mean reversion of Hull–White generating model
σ_r	0.05	Volatility of Hull–White generating model
ξ	-0.6	Instantaneous correlation between short rate and log-stock in generating model
$\overline{\sigma}_{\text{good}}$	0.9	Volatility scaling of "good" reference CEV model
$\overline{\gamma}_{\text{good}}$	0.9	Power law in "good" reference CEV model
$\overline{\xi}_{\text{good}}$	-0.4	Instantaneous correlation between short rate and log-stock in "good" reference model
$\overline{\sigma}_{\text{bad}}$	1.2	Volatility scaling of "bad" reference CEV model
$\overline{\gamma}_{\text{bad}}$	0.78	Power law in "bad" reference CEV model
$\overline{\xi}_{\text{bad}}$	0.4	Instantaneous correlation between short rate and log-stock in "bad" reference model

Our numerical methods converged for both the "good" and "bad" reference model case, calibrating all the call options to a tolerance of 10^{-4}, with the implied volatility in the calibrated model replicating the prices of the generating model. The implied volatility skews in the generating and calibrated models are indistinguishable for both maturities within the range $K \in [85, 120]$ in which our option strikes were taken, indicating that our optimal transport model replicates the observed simulated data both at and between the calibrating option strikes. A dependence on the reference model is only observed outside of this range, which is to be expected. The plots are given in Fig. 1, with the exact results summarised in Table 2.

A feature of our method is that it is designed to find a calibrated model closest to the reference model, in the sense of minimising (4.1). This results in a spiky volatility surface – the method tries to stick to the reference model while making sharp deviations needed to calibrate to the given option prices. This is not necessarily a desired feature, and we propose to smooth the surface to obtain a reasonable volatility surface while matching the market data. We used an iterative procedure: once we obtained convergence of the model prices to the generating model prices within the tolerance of ε_1, we stored those characteristics as the new reference model, applied an interpolation method using cubic splines to smooth the surface and then restarted the process with the smoothed surface as our reference model. We remark that the final



(a) (b)

Fig. 1 Implied volatility skews under generating model, reference model and OT-calibrated models with both reference models and across two maturities

Table 2 Table of generating and calibrated model prices and implied volatilities

Option type	Strike	Generating model		Calibrated model: good reference		Calibrated model: bad reference	
		Price	IV	Price	IV	Price	IV
SPX call options at $t = 60$ days	85	11.3666	0.4941	11.3666	0.4941	11.3668	0.4941
	92	7.5389	0.4921	7.5398	0.4922	7.5396	0.4922
	99	4.7538	0.4906	4.7549	0.4906	4.7537	0.4905
	106	2.8616	0.4893	2.8625	0.4894	2.8613	0.4893
	113	1.6523	0.4884	1.6532	0.4885	1.6526	0.4884
	120	0.9189	0.4875	0.9192	0.4876	0.9192	0.4876
SPX call options at $t = 120$ days	85	14.2787	0.4923	14.2787	0.4923	14.2780	0.4923
	92	10.7017	0.4905	10.7007	0.4904	10.7009	0.4904
	99	7.8563	0.4886	7.8580	0.4887	7.8575	0.4886
	106	5.6560	0.4866	5.6575	0.4866	5.6568	0.4866
	113	3.9917	0.4840	3.9918	0.4840	3.9910	0.4840
	120	2.7493	0.4802	2.7495	0.4802	2.7483	0.4802

epoch of the calibration algorithm involved no interpolation as this spoils calibration; so the overall algorithm provided the OT-calibrated surface after iterating through smoothed reference models. In addition to providing more reasonable model dynamics, the smoothed reference model iteration also improved the numerical stability of the algorithm, and thus allowed us to achieve a better calibration. This came at a significant computational cost due to the iterations of the calibration routine; however, both ran on a standard notebook laptop in a matter of hours. Both reference models were smoothed around 10 times before changes in the volatility surface plots were no longer noticeable.

We remark that the above reference model iteration could in principle be circumvented via a regularisation technique by encoding a smoothing penalty into the cost function, e.g. via a second-order Tikhonov regularisation method. Clearly this would change the optimal characteristics given in (4.19), but the optimal surface would then have this smoothing penalty encoded into it. However, this would require at all grid-

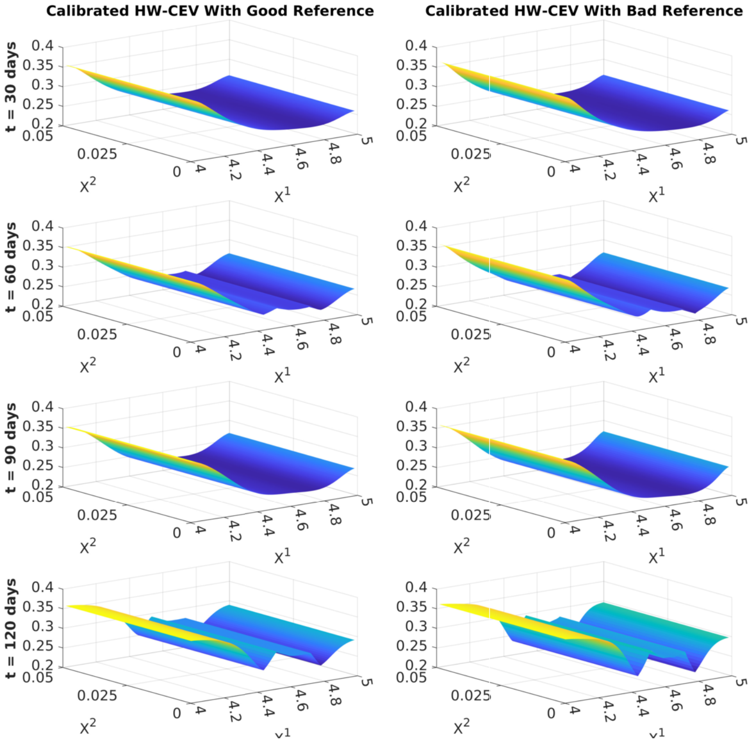


Fig. 2 Plots of β_{11} at $t = 30$, 60, 90 and 120 days for calibrated model

points to estimate the first and second derivatives of β_{11} via a central-difference method, which would substantially increase the computational load beyond that of the reference model iteration method, and additionally potentially require extra state variables to account for the spatial derivatives of β.

To better illustrate the features of the OT-calibrated model, we present plots of the surfaces of the model characteristics, both on their own – see Figs. 2 and 3 – and superimposed with the characteristics of the generating and reference models for comparison; see Figs. 4 and 5. Broadly speaking, as discussed above, the OT-calibrated model is close to the generating one in the region specified by the data and close to the reference one otherwise. In addition, the surfaces we obtain for the correlation ξ are highly dependent on the reference model, as one may expect from the convexity adjustment in (4.15). The choice of reference value therefore means that ξ is always close to the reference model with some fluctuations obtained through the change in the value β_{11}. Consequently, the modeller's (or trader's) insight in specifying a correct correlation (not least its sign) is relatively more important than specifying a correct initial volatility.

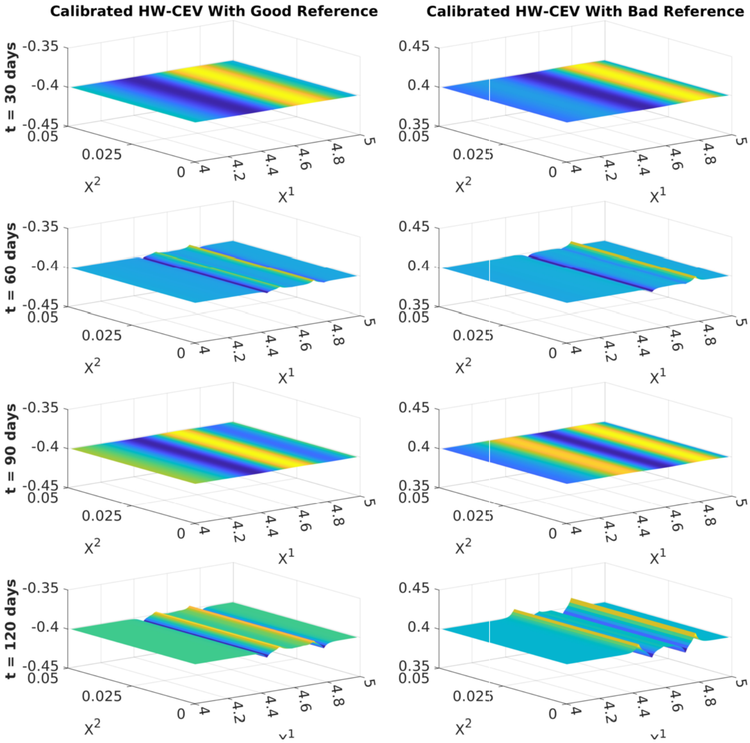


Fig. 3 Plots of ξ at $t = 30$, 60, 90 and 120 days for calibrated model

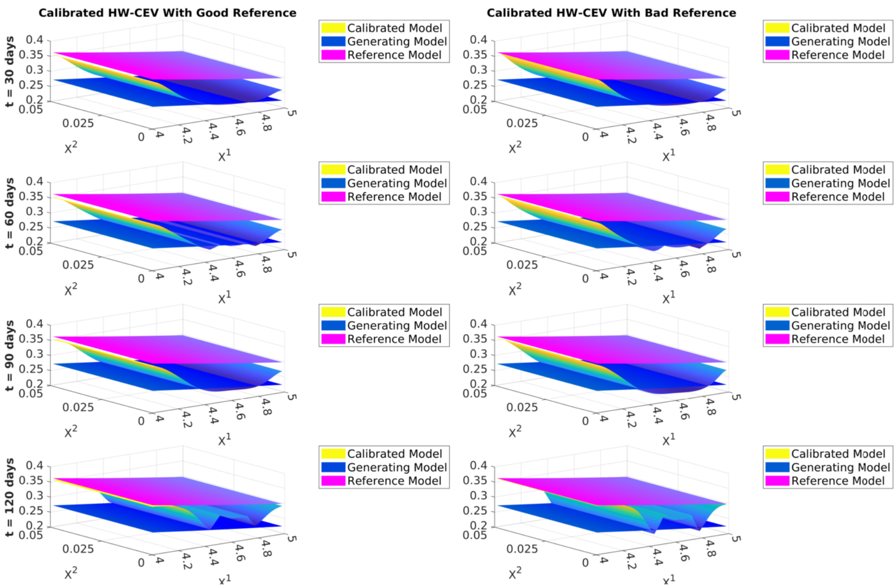


Fig. 4 Plots of β_{11} at $t = 30$, 60, 90 and 120 days for calibrated model compared with generating model

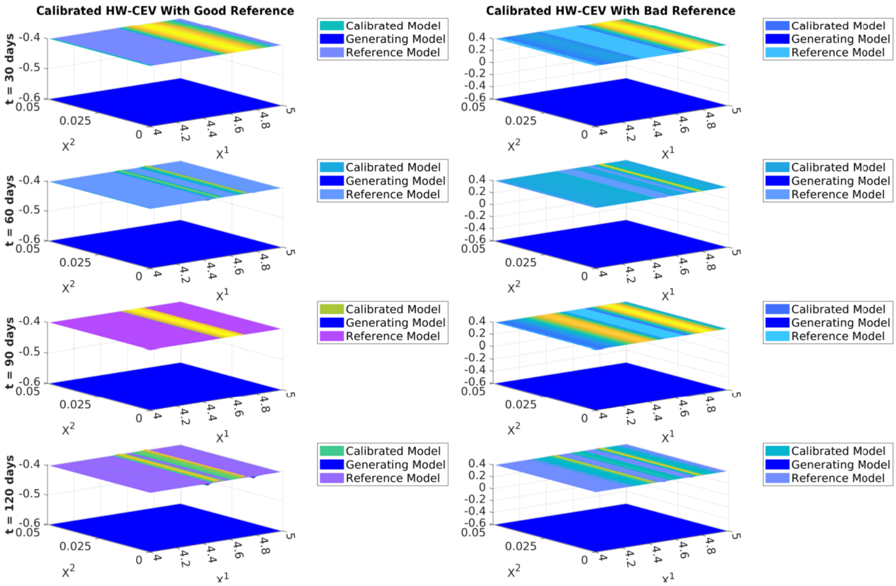


Fig. 5 Plots of ξ at $t = 30$, 60, 90 and 120 days for calibrated model compared with generating model

6 Conclusions

This paper has developed a nonparametric optimal-transport-driven method to calibrate a stock price model in a stochastic interest-rate environment. By switching to sub-probability measures representing discounted densities, the method manages not to increase the dimension of the state variables. The method is flexible and can accommodate different calibrating instruments. We developed a generic duality result and applied it in the context of a Hull–White short rate model and a local volatility stock price model with short-rate dependence. We illustrated the numerical performance of the method by considering a sequential calibration problem: the short-rate model is calibrated first in a parametric setting and then feeds into our nonparametric local volatility calibration. We highlighted how some model features are pinned down well via the market prices of options. However others, notably the correlation between the Brownian motions driving the rates and the stock price, are much more sensitive to the modeller's choice of their reference model. In a follow-up paper Joseph et al. [51], we apply the duality result developed here to consider simultaneous OT-driven calibration for a joint model for stochastic interest rates and the stock price process and demonstrate its performance on real market data. We note that here and in [51], we restrict ourselves to short-rate models to keep the dimension of the HJB equation to two state variables. It would be natural and interesting to extend this setup to include more stochastic factors, e.g. stochastic volatility for the stock prices. Each extra factor would raise the dimension of the HJB and imply a significant numerical challenge. It would also be of interest to consider more realistic fixed-income models, such as market models as in Brace et al. [14], Miltersen et al. [59], Jamshidian [50] or more recent multi-curve framework models, see Henrard [41, 42], where the post-2008 financial crisis effects of the LIBOR–OIS spreads are taken into account. This, however, would drastically increase the dimension of our state process. It would likely necessitate novel tools to solve the resulting PDEs, such as recent machine learning methods, see E et al. [26], Han et al. [40]. We believe that these are very interesting avenues for future research.

Appendix: Proof of the discounted superposition principle (Theorem 3.5)

We first define two operators associated with the discounted Fokker–Planck equation (3.7) and the augmented Fokker–Planck equation (2.3). Given functions $a:[0,T]\times\mathbb{R}^d\to\mathbb{R}^d$, $b:[0,T]\times\mathbb{R}^d\to\mathbb{S}^d_+$, $c:[0,T]\times\mathbb{R}^d\to\mathbb{R}$, set

$$C([0,T];\mathbb{R}^d)\ni f\mapsto \mathcal{L}^D f := a\cdot\nabla_x f+\frac{1}{2}b:\nabla_x^2 f - cf,$$

$$C([0,T];\mathbb{R}^{d+1})\ni f\mapsto \mathcal{L}^A f := a\cdot\nabla_x f+\frac{1}{2}b:\nabla_x^2 f + yc\partial_y f.$$

Note that if $(\nu_t)_{t\in[0,T]}\subseteq\mathcal{P}(\mathbb{R}^{d+1})$ solves the augmented Fokker–Planck equation (2.3) associated with $\mathcal{L}^A$, denoted aFP(a,b,c), then $(\nu_t^D)_{t\in[0,T]}\subseteq\mathcal{M}(\mathbb{R}^d)$ with $\nu_t^D=\int_\mathbb{R} y\nu_t(\,\cdot\,,\mathrm{d}y)$ solves the discounted Fokker–Planck equation in Theorem 3.5 associated with $\mathcal{L}^D$, denoted dFP(a,b,c). This is easily seen since the coefficients a,b,c have no y-dependence.

We construct a sequence $(\nu_t^n)_{t\in[0,T]}\subseteq\mathcal{M}(\mathbb{R}^d)$ of smooth solutions to (3.7) with coefficients a^n, b^n, c^n all in $\in C^\infty$ for which we can easily show Theorem 3.5. We therefore can construct a sequence of diffusion operators $\mathcal{L}^{A,n}$ corresponding to aMP(a^n,b^n,c^n) which has a solution $\boldsymbol{\eta}^n$ such that $\int_\mathbb{R} y\eta_t^n(\,\cdot\,,\mathrm{d}y)=\nu_t^n$. The tightness and convergence of $(\boldsymbol{\eta}^n)$ is then shown, and then we use the arguments of Trevisan [68, Sect. A.4] to generalise the coefficients a,b,c.

A.1 Smooth approximation of the augmented martingale problem

Lemma A.1 *Let $(\nu_t)_{t\in[0,T]}\subseteq\mathcal{M}(\mathbb{R}^d)$ be a smooth solution of* dFP(a,b,c) *with coefficients a,b,c in $\in C^\infty$. Then there exists a solution $\boldsymbol{\eta}\in\mathcal{P}(C([0,T];\mathbb{R}^{d+1}))$ to* aMP(a,b,c) *such that $\int_\mathbb{R} y\eta_t(\,\cdot\,,\mathrm{d}y)=\nu_t(\,\cdot\,)$.*

Proof Let $\mathbb{P}\in\mathcal{P}_1$ and define the process

$$\mathrm{d}\tilde{X}_t=a_t(\tilde{X}_t)\,\mathrm{d}t+\big(b_t(\tilde{X}_t)\big)^{\frac{1}{2}}\,\mathrm{d}W_t^{\tilde{\mathbb{P}}},\quad 0\le t\le T,\qquad \tilde{X}_0=x_0. \tag{A.1}$$

Since a and b are smooth, we automatically have existence and uniqueness of a solution to (A.1) (see Stroock and Varadhan [66, Corollary 6.3.3]). Define $\tilde{Y}_t=\exp(-\int_0^t c_s(\tilde{X}_s)\,\mathrm{d}s)$. Then by the same reasoning as in Theorem 2.2, the curve of marginal laws of the augmented process $(\tilde{X},\tilde{Y})$ at t, given by $\eta_t=(e_t)_\#\boldsymbol{\eta}$, solves the Fokker–Planck equation, for $(t,x,y)\in(0,T)\times\mathbb{R}^{d+1}$,

$$\partial_t\eta_t(x,y)+\nabla_x\cdot\big(\eta_t(x,y)a_t(x)\big)-\frac{1}{2}\nabla_x^2:\big(\eta_t(x,y)b_t(x)\big)-\partial_y\big(yc_t(x)\eta_t(x,y)\big)=0.$$

We now define the discounted version of η_t by $\eta_t^D(x) := \int_{\mathbb{R}} y\eta_t(x, y)\,\mathrm{d}y$. Then by a similar calculation as in the proof of Lemma 3.6, $(\eta_t^D)_{t\in[0,T]}$ solves the discounted Fokker–Planck equation (3.7). By an argument similar to Figalli [28, Proposition 4.1], we have uniqueness of the solution to (3.7); so $\eta_t^D = \nu_t$. By Itô's formula, it is immediate that $\boldsymbol{\eta} \in \mathcal{P}(C([0, T]; \mathbb{R}^{d+1}))$ solves aMP(a, b, c), and so the result follows. $\square$

We now let $(\nu_t)_{t\in[0,T]} \subseteq \mathcal{M}(\mathbb{R}^d)$ be a narrowly continuous solution of dFP(a, b, c) for $a \in L^1(\mathrm{d}\nu_t\,\mathrm{d}t; \mathbb{R}^d)$, $b \in L^1(\mathrm{d}\nu_t\,\mathrm{d}t; \mathbb{S}^d)$ and $c \in L^1(\mathrm{d}\nu_t\,\mathrm{d}t; \mathbb{R})$. We first build a smooth sequence (ν_t^n) approximating ν_t for all $t \in [0, T]$, associated with diffusion operators $\mathcal{L}^{D,n}$. Then by Lemma A.1, there exists a sequence $(\boldsymbol{\eta}^n)$ of solutions to the augmented martingale problem associated with $\mathcal{L}^{A,n}$ such that $\int_{\mathbb{R}} y\eta_t^n(\,\cdot\,, \mathrm{d}y) = \nu_t^n$. As in Trevisan [68, Sect. A.1], we construct the approximations in two different ways which we later use in Sect. A.3 when generalising the regularity of the coefficients. In both variants, we emphasise the specific form of the discounted superposition principle with the type of smoothing applied.

Variant 1: Pushforward via smooth maps. Take functions $\pi = (\pi^1, \ldots, \pi^d)$ in $C^2(\mathbb{R}^d; \mathbb{R}^d)$ with uniformly bounded first and second derivatives. Define the standard diffusion operator $\mathcal{L}^S$ by $C([0, T]; \mathbb{R}^d) \ni f \mapsto \mathcal{L}^S = a \cdot \nabla_x f + \frac{1}{2} b : \nabla_x^2 f$. Then we can define approximations of (a, b, c) by

$$
\begin{aligned}
{}_\pi a_t^i &:= \frac{\mathrm{d}\pi_\#(\mathcal{L}^S(\pi^i)\nu_t)}{\mathrm{d}\pi_\#\nu_t},\\
{}_\pi b_t^{i,j} &:= \frac{\mathrm{d}\pi_\#(b_t : (\nabla\pi^i \otimes \nabla\pi^j)\nu_t)}{\mathrm{d}\pi_\#\nu_t},\\
{}_\pi c_t^i &:= \frac{\mathrm{d}\pi_\#(c_t\nu_t)}{\mathrm{d}\pi_\#\nu_t}.
\end{aligned}
$$

Note that uniform bounds on a, b and c are preserved with ${}_\pi a, {}_\pi b, {}_\pi c$. Additionally, by an application of the chain rule to compute $\mathcal{L}^D(f \circ \pi)$, it is easy to see that $(\pi_\#\nu_t)_{t\in[0,T]} \subseteq \mathcal{M}(\mathbb{R}^d)$ solves dFP$({}_\pi a, {}_\pi b, {}_\pi c)$, and therefore by Lemma A.1, there exists a solution $\boldsymbol{\eta}$ to aMP$({}_\pi a, {}_\pi b, {}_\pi c)$ such that $\pi_\#\nu_t = \int_{\mathbb{R}} y\eta_t(\,\cdot\,, \mathrm{d}y)$.

Variant 2: Mollification by convolution. Let $\kappa \geq 0$ be a smooth probability density on $\mathbb{R}^d$. Define

$$
a_t^\kappa := \frac{\mathrm{d}((a_t\nu_t) * \kappa)}{\mathrm{d}(\nu_t * \kappa)}, \qquad b_t^\kappa := \frac{\mathrm{d}((b_t\nu_t) * \kappa)}{\mathrm{d}(\nu_t * \kappa)}, \qquad c_t^\kappa := \frac{\mathrm{d}((c_t\nu_t) * \kappa)}{\mathrm{d}(\nu_t * \kappa)}.
$$

By Trevisan [68, Lemma A.1], a^κ, b^κ and c^κ preserve integrability and have uniformly bounded first and second spatial derivatives on compact sets. Therefore $(\nu_t * \kappa)_{t\in[0,T]} \subseteq \mathcal{M}(\mathbb{R}^d)$ solves dFP$(a^\kappa, b^\kappa, c^\kappa)$. Additionally, by Lemma A.1, there exists a solution $\boldsymbol{\eta}$ to aMP$(a^\kappa, b^\kappa, c^\kappa)$ such that $\nu_t * \kappa = \int_{\mathbb{R}} y\eta_t(\,\cdot\,, \mathrm{d}y)$.

A.2 Compactness and convergence of the augmented martingale problem

We now assume that $(\boldsymbol{\eta}^n)$ is a sequence of solutions of augmented martingale problems obtained via the discounted superposition principle (*superposition solutions*) corresponding to an approximation $\nu_t^n = \pi^n_{\#}\nu_t$ in the pushforward case or $\nu_t^n = \nu_t * \kappa^n$ in the mollification case. In the mollification case, we assume that $\kappa^n \to \delta_0$ narrowly as $n \to \infty$. In the pushforward case, we assume that $\pi_n \to \mathrm{Id}$ locally uniformly, and $\nabla\pi^n \to \mathrm{Id}$ and $\nabla^2\pi^n \to 0$ pointwise. Assume also that the sequences of these derivatives are uniformly bounded; so there exists $C \geq 0$ such that $|\nabla\pi^n|, |\nabla^2\pi^n| \leq C$. Compactness for solutions of the augmented martingale problem then follows directly from [68, Sect. A.2] since we can recast the superposition principle in the setting of [68, Theorem 2.5] by considering superposition solutions to aFP(a, b, c), and then the tightness arguments are directly applicable.

We are looking for solutions of aMP(a, b, c) whose discounted density is equal to a given solution of dFP(a, b, c). We therefore consider the class of test functions $C_b^{1,2}([0, T] \times \mathbb{R}^{d+1}; \mathbb{R}) \ni \tilde{f}(x, y) = yf(x)$, where $f \in C_b^{1,2}([0, T] \times \mathbb{R}^d; \mathbb{R})$. For $f \in C_b^{1,2}([0, T] \times \mathbb{R}^d; \mathbb{R})$, define the norm $\|f\|_{C_b^{1,2}}$ by

$$\|f\|_{C_b^{1,2}([0,T]\times\mathbb{R}^d)} = \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \big(|f(t, x)| + |\partial_t f(t, x)| + |\nabla f(t, x)| + |\nabla^2 f(t, x)|\big).$$

We note that we always have

$$\int\int_s^t (\mathcal{L}_r^A \tilde{f}) \circ e_r \,\mathrm{d}r \,\mathrm{d}\boldsymbol{\eta} = \int_s^t \int_{\mathbb{R}^d} (\mathcal{L}_r^D f_r) \int_{\mathbb{R}} y\eta_r(\mathrm{d}x, \mathrm{d}y) \,\mathrm{d}r.$$

From now on, we use the shorthand $\mathrm{d}\nu = \mathrm{d}\nu_t \,\mathrm{d}t$ for notational ease. Now let $\overline{\mathcal{L}^D}$ be a discounted diffusion operator with coefficients $\overline{a}, \overline{b}, \overline{c}$ which are continuous and compactly supported. In view of the arguments of [68, Sect. A.3] and due to the narrow convergence of $(\boldsymbol{\eta}^n)$ to some limit $\boldsymbol{\eta}$, in order to show that $\nu := \lim_{n\to\infty} \nu^n$ is a discounted superposition solution of $\boldsymbol{\eta}$ (in both variants of smooth approximating sequences), we only need to provide bounds on

$$\limsup_{n\to\infty} \int |\mathcal{L}^{D,n} f - \overline{\mathcal{L}^D} f| \,\mathrm{d}\nu^n + \int |\mathcal{L}^D f - \overline{\mathcal{L}^D} f| \,\mathrm{d}\nu. \tag{A.2}$$

Variant 1: Pushforward via smooth maps. Write ${}_\pi\mathcal{L}^D$ for the discounted diffusion operator associated with the coefficients $({}_\pi a, {}_\pi b, {}_\pi c)$. Then for fixed n, we have

$$\begin{aligned}\int |{}_{\pi^n}\mathcal{L}^{D,n} f - \overline{\mathcal{L}^D} f| \,\mathrm{d}\pi^n_{\#}\nu &\leq \int |\mathcal{L}^{D,n}(f \circ \pi^n) - (\overline{\mathcal{L}^D} f) \circ \pi^n| \,\mathrm{d}\nu, \\ &\leq \int \frac{1}{2} \sum_{i,j=1}^d |a : (\nabla\pi_i \otimes \nabla\pi_j) - \overline{a}^{i,j} \circ \pi| \,\mathrm{d}\nu \\ &\quad + \int \sum_{i=1}^d |\mathcal{L}^S(\pi_i) - \overline{a}^i \circ \pi| \,\mathrm{d}\nu + \int |c - \overline{c} \circ \pi| \,\mathrm{d}\nu.\end{aligned}$$

Since (π^n) converges to the identity, (A.2) is bounded above by

$$\int \sum_{i,j=1}^{d} |a : (\nabla\pi_i \otimes \nabla\pi_j) - \overline{a}^{i,j} \circ \pi|\,\mathrm{d}\nu + 2\int \sum_{i=1}^{d} |\mathcal{L}^S(\pi_i) - \overline{a}^i \circ \pi|\,\mathrm{d}\nu$$
$$+ 2\int |c - \overline{c} \circ \pi|\,\mathrm{d}\nu.$$

Since continuous and compactly supported functions are dense in $L^1(\nu)$, the above can be made arbitrarily small by optimising over $\overline{a}$, $\overline{b}$, $\overline{c}$.

Variant 2: Mollification by convolution. Let $\overline{\omega}$ be a common bounded and continuous modulus of continuity for $\overline{a}$, $\overline{b}$, $\overline{c}$. Let $\overline{\mathcal{L}^{D,n}}$ be the discounted diffusion operator associated with the coefficients

$$\overline{a}^n := \frac{\mathrm{d}((\overline{a}\nu) * \kappa^n)}{\mathrm{d}(\nu * \kappa^n)}, \qquad \overline{b}^n := \frac{\mathrm{d}((\overline{b}\nu) * \kappa^n)}{\mathrm{d}(\nu * \kappa^n)}, \qquad \overline{c}^n := \frac{\mathrm{d}((\overline{c}\nu) * \kappa^n)}{\mathrm{d}(\nu * \kappa^n)}.$$

We first show $\lim_{n\to\infty} \int |\overline{\mathcal{L}^{D,n}} f - \overline{\mathcal{L}^{D}} f|\,\mathrm{d}\nu^n = 0$. Since $\|f\|_{C^{1,2}} \le 1$, we have

$$\begin{aligned}\int |\overline{\mathcal{L}^{D,n}} f - \overline{\mathcal{L}^{D}} f|\,\mathrm{d}\nu^n &\le \int |\overline{a}^n - \overline{a}|\,\mathrm{d}\nu^n + \int |\overline{b}^n - \overline{b}|\,\mathrm{d}\nu^n + \int |\overline{c}^n - \overline{c}|\,\mathrm{d}\nu^n \\ &\le 3\int \overline{\omega}(z)\kappa^n(z)\,\mathrm{d}z \longrightarrow 0.\end{aligned}$$

Therefore we have

$$\begin{aligned}\limsup_{n\to\infty} \int |\mathcal{L}^{D,n} f - \overline{\mathcal{L}^{D}} f|\,\mathrm{d}\nu^n &= \limsup_{n\to\infty} \int |\mathcal{L}^{D,n} f - \overline{\mathcal{L}^{D,n}} f|\,\mathrm{d}\nu^n \\ &\le \int (|a - \overline{a}| + |b - \overline{b}| + |c - \overline{c}|)\,\mathrm{d}\nu.\end{aligned}$$

By the same argument as before, we can make the right-hand side arbitrarily small which concludes the proof of Lemma A.1. □

A.3 Generalisation of drift and diffusion coefficients

We now use the approximation and convergence described in Sects. A.1 and A.2 to generalise the coefficients a, b and c and prove Theorem 3.5.

A.3.1 Smooth and bounded coefficients

Let $a : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$, $b : [0,T] \times \mathbb{R}^d \to \mathbb{S}^d_+$, $c : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ be Borel maps such that

$$\int_0^T (\|a_t\|_{C_b^2(\mathbb{R}^d)} + \|b_t\|_{C_b^2(\mathbb{R}^d)} + \|c_t\|_{C_b^2(\mathbb{R}^d)})\,\mathrm{d}t < \infty, \tag{A.3}$$

where for $f \in C_b^2(\mathbb{R}^d)$, the norm $\|f\|_{C_b^2(\mathbb{R}^d)}$ is given by

$$\|f\|_{C_b^2(\mathbb{R}^d)} = \sup_{x\in\mathbb{R}^d} \big(|f(x)| + |\nabla f(x)| + |\nabla^2 f(x)|\big). \tag{A.4}$$

Lemma A.2 *Let a, b, c be Borel maps satisfying (A.3). Then for every $\overline{\nu} \in \mathcal{P}(\mathbb{R}^d)$, there exists $\boldsymbol{\eta} \in \mathcal{P}(C([0,T];\mathbb{R}^{d+1}))$ solving* aMP(a,b,c) *and with the property that $\int_{\mathbb{R}} y\eta_0(\,\cdot\,, \mathrm{d}y) = \overline{\nu}(\,\cdot\,)$.*

The proof is the same as for [68, Theorem A.6], but by considering the augmented process (X, Y), where the regular martingale problem (see [68, Definition 2.4]) is given by aMP(a,b,c).

Lemma A.3 *Let a, b, c be Borel maps such that for every bounded open set $B \subseteq \mathbb{R}^d$,*

$$\int_0^T (\|a_t\|_{C^2(B)} + \|b_t\|_{C^2(B)} + \|c_t\|_{C^2(B)})\,\mathrm{d}t < \infty,$$

where for $f \in C^2(\mathbb{R}^d)$, the norm $\|f\|_{C^2(B)}$ is as in (A.4), but with the supremum taken over $x \in B$. Let $(\nu_t)_{t\in[0,T]} \subseteq \mathcal{M}(\mathbb{R}^d)$ be a narrowly continuous solution of dFP(a,b,c). *If $\nu_0 \le 0$, then $\nu_t \le 0$ for all $t \in [0,T]$. Therefore for $\overline{\nu} \in \mathcal{M}(\mathbb{R}^d)$, there exists a unique narrowly continuous solution ν such that $\nu_0 = \overline{\nu}$.*

The proof is similar to that of [68, Theorem A.7]; so we only provide a sketch.

Proof of Lemma A.3 Let $g \in C_c^\infty([0,T]\times\mathbb{R}^d)$ be nonnegative. Let $\chi_R : \mathbb{R}^d \to [0,1]$ be a cutoff function and $a_R^\varepsilon, b_R^\varepsilon, c_R^\varepsilon$ mollifications with respect to time and space of $a\chi_R, b\chi_R, c\chi_R$, so that $a_R^\varepsilon, b_R^\varepsilon, c_R^\varepsilon$ satisfy (A.3). Let $\mathcal{L}_R^{D,\varepsilon}$ be the diffusion operator associated with dFP$(a_R^\varepsilon, b_R^\varepsilon, c_R^\varepsilon)$. Let f^ε solve

$$\partial_t f^\varepsilon = -\mathcal{L}_R^{D,\varepsilon} f^\varepsilon + g, \qquad f_T^\varepsilon = 0.$$

Recall the definition of $\mathcal{L}^S$ in Sect. A.1 in Variant 1. Since ν solves dFP(a,b,c) and $f^\varepsilon \in C_b^{1,2}(\mathbb{R}^d)$ and $f^\varepsilon, \nu_0 \le 0$, we have

$$\begin{aligned}
0 &\ge -\int f^\varepsilon \chi_R \,\mathrm{d}\nu_0,\\
&= \int \big(\chi_R \partial_t f^\varepsilon + \mathcal{L}^D(f^\varepsilon\chi_R)\big)\,\mathrm{d}\nu,\\
&= \int \big(\chi_R(g - \mathcal{L}_R^{D,\varepsilon} f^\varepsilon + \mathcal{L}^S f^\varepsilon) + f^\varepsilon \mathcal{L}^S\chi_R + cf^\varepsilon\chi_R + b : \nabla f^\varepsilon \otimes \nabla\chi_R\big)\,\mathrm{d}\nu,\\
&\ge \int g\,\mathrm{d}\nu - \sup_{t\in[0,T]} \|f_t^\varepsilon\|_{C_b^2(\mathbb{R}^d)} \int \big(\chi_R(|a_R^\varepsilon - a| + |b_R^\varepsilon - b| + |c_R^\varepsilon - c|)\\
&\qquad\qquad + |\mathcal{L}^S\chi_R| + |b|\,|\nabla\chi_R|\big)\,\mathrm{d}\nu.
\end{aligned}$$

The sup is uniformly bounded in $\varepsilon > 0$, and $(a^\varepsilon_R, b^\varepsilon_R, c^\varepsilon_R) \to (a, b, c)$ for $|x| \le R$ as $\varepsilon \to 0$. Then letting $R \to \infty$, the right-hand side converges to $\int g \, \mathrm{d}\nu$ since $\nabla \chi_R \to 0$ and $\mathcal{L}^S \chi_R \to 0$. $\square$

A.3.2 Bounded coefficients

This case follows in the same manner as the bounded case in Trevisan [68]. We now assume that a, b, c satisfy

$$\int_0^T \Big(\sup_{x\in\mathbb{R}^d} |a_t(x)| + \sup_{x\in\mathbb{R}^d} |b_t(x)| + \sup_{x\in\mathbb{R}^d} |c_t(x)| \Big)\, \mathrm{d}t < \infty. \tag{A.5}$$

Using the regularisation kernel $\kappa = \exp(-\sqrt{1+|x|^2})$ and $\kappa^\varepsilon = \varepsilon^{-n}\kappa(x/\varepsilon)$, we have with $\nu^\varepsilon_t = \nu_t * \kappa^\varepsilon$ that $(\nu^\varepsilon_t)_{t\in[0,T]}$ solves dFP$(a^\varepsilon, b^\varepsilon, c^\varepsilon)$ with coefficients satisfying (A.3). Thus existence of $\boldsymbol{\eta}^\varepsilon \in \mathcal{P}(C([0,T];\mathbb{R}^{d+1}))$ solving aMP$(a^\varepsilon, b^\varepsilon, c^\varepsilon)$ as a discounted superposition solution to ν^ε follows. Since (ν^ε_0) is a narrowly convergent family, there exists $\theta : \mathbb{R} \to \mathbb{R}$ increasing with $\lim_{x\to\infty}\theta(x) = \infty$ such that $\sup_{\varepsilon>0} \int \theta(|x|)\, \mathrm{d}\nu^\varepsilon_0 \le 1$. The proof of tightness of $(\boldsymbol{\eta}^\varepsilon)$ is then an easy modification of that in [68], where the de la Vallée-Poussin criterion modifies (A.5) to

$$\int_0^T \Big(\Theta\Big(\sup_{x\in\mathbb{R}^d} |a_t(x)| \Big) + \Theta\Big(\sup_{x\in\mathbb{R}^d} |b_t(x)| \Big) + \Theta\Big(\sup_{x\in\mathbb{R}^d} |c_t(x)| \Big) \Big)\, \mathrm{d}t < \infty$$

for $\Theta : [0,\infty) \to [0,\infty)$ convex increasing with $\lim_{x\to\infty} \Theta(x)/x = \infty$. Thus we proceed in analogy to [68, Corollary 2.5], with $f^i = (x_i \chi_R) \circ e_t$ for $i = 1, \ldots, d$ and $f^{d+1} = (\log(y\chi_R)) \circ e_t$. Note that the function f^{d+1} is always well defined and real-valued since Y is assumed to be supported on $(0, \exp(T))$. By [68, Corollary A.5], there exists some coercive functional $\Psi : C([0,T];\mathbb{R}) \to [0,\infty]$ such that for $i = 1, \ldots, d$, with $x^i_R = x_i \chi_R$, we have

$$\begin{aligned} \int \Psi(x^i_R \circ \gamma)\, \mathrm{d}\boldsymbol{\eta}^\varepsilon(\gamma) &\le \int \theta(|x^i_R|)\, \mathrm{d}\eta^\varepsilon_0 \\ &\quad + \int_0^T \int \big(\Theta(|\mathcal{L}^{S,\varepsilon}_t x^i_R|) + \Theta(b^\varepsilon_t : \nabla x^i_R \otimes \nabla x^i_R) \big)\, \mathrm{d}\eta^\varepsilon_t \, \mathrm{d}t, \\ \int \Psi(f^{d+1} \circ \gamma)\, \mathrm{d}\boldsymbol{\eta}^\varepsilon(\gamma) &\le \int \theta(|f^{d+1}|)\, \mathrm{d}\eta^\varepsilon_0 + \int_0^T \int \Theta(|c^\varepsilon_t|)\, \mathrm{d}\eta^\varepsilon_t \, \mathrm{d}t. \end{aligned}$$

We now note that $\eta^\varepsilon_t = \nu^{A,\varepsilon}_t$, where $(\nu^{A,\varepsilon}_t)_{t\in[0,T]}$ solves aFP$(a^\varepsilon, b^\varepsilon, c^\varepsilon)$, and further that $\nu^{A,\varepsilon}_0 = \nu^\varepsilon_0$ by construction. Thus letting $R \to \infty$ and noting that the y-marginal of η^ε_0 is δ_1, the regular superposition principle ([68, Theorem 2.5]) gives the uniform bounds

$$\int \Psi(x_R^i \circ \gamma)\, \mathrm{d}\boldsymbol{\eta}^\varepsilon(\gamma) \le \int \theta(|x_R^i|)\, \mathrm{d}\nu_0^\varepsilon$$
$$+ \int_0^T \int \left(\Theta(|\mathcal{L}_t^{S,\varepsilon} x_R^i|) + \Theta(b_t^\varepsilon : \nabla x_R^i \otimes \nabla x_R^i)\right) \mathrm{d}\nu_t^{A,\varepsilon}\, \mathrm{d}t,$$
$$\int \Psi(f^{d+1} \circ \gamma)\, \mathrm{d}\boldsymbol{\eta}^\varepsilon(\gamma) \le \theta(0) + \int_0^T \int \Theta(|c_t^\varepsilon|)\, \mathrm{d}\nu_t^{A,\varepsilon}\, \mathrm{d}t.$$

Thus by [68, Lemma A.1], we have the uniform bounds

$$\int \Psi(\gamma^i)\, \mathrm{d}\boldsymbol{\eta}^\varepsilon(\gamma) \le 1 + \int_0^T \left(\Theta\Big(\sup_{x\in\mathbb{R}^d} |a_t(x)|\Big) + \Theta\Big(\sup_{x\in\mathbb{R}^d} |b_t(x)|\Big)\right) \mathrm{d}t,$$
$$\int \Psi(\log \gamma^{d+1})\, \mathrm{d}\boldsymbol{\eta}^\varepsilon(\gamma) \le \theta(0) + \int_0^T \Theta\Big(\sup_{x\in\mathbb{R}^d} |c_t(x)|\Big)\, \mathrm{d}t.$$

Since $\gamma \mapsto \Psi(\log \gamma^{d+1}) + \sum_{i=1}^d \Psi(\gamma^i)$ is coercive in $C([0,T];\mathbb{R}^{d+1})$, we have tightness of $\boldsymbol{\eta}^\varepsilon$. That the result also holds for the limit of this sequence follows directly from [68, Sects. A.2 and A.3] and Sect. A.2. Therefore we have shown Theorem 3.5 for bounded coefficients a, b and c.

A.3.3 Locally bounded coefficients

We now assume that a, b, c satisfy for every bounded Borel set $B \subseteq \mathbb{R}^d$ that

$$\int_0^T \sup_{x\in B} \big(|a_t(x)| + |b_t(x)| + |c_t(x)|\big)\, \mathrm{d}t < \infty. \tag{A.6}$$

Let $M \ge 1$, χ_M be a cutoff function and define $\pi_M : \mathbb{R}^d \to \mathbb{R}^d$ by $\pi_M(x) = x\chi_M(x)$. Thus by (A.6), we have

$$|\mathcal{L}^D(\pi_M^i)| \le \|\pi_M^i\|_{C_b^2} \sup_{|x|\le 2M} \big(|a(x)| + |b(x)| + |x_r c(x)|\big),$$
$$|b : \nabla\pi_M^i \otimes \nabla\pi_M^j| \le \|\pi_M^i\|_{C^1} \sup_{|x|\le MR} |b(x)|,$$

where for $f \in C_b^1(\mathbb{R}^d)$, the norm $\|f\|_{C_b^1}$ is given by

$$\|f\|_{C_b^1} = \sup_{x\in\mathbb{R}^d} \big(|f(x)| + |\nabla f(x)|\big). \tag{A.7}$$

Thus with $\nu_t^M := \pi_\#^M \nu_t$, the family $(\nu_t^M)_{t\in[0,T]}$ solves $\mathrm{dFP}({}_{\pi^M}a, {}_{\pi^M}b, {}_{\pi^M}c)$ with coefficients satisfying (A.5), and so we have $\boldsymbol{\eta}^M \in \mathcal{P}(C([0,T];\mathbb{R}^{d+1}))$ solving $\mathrm{aMP}({}_{\pi^M}a, {}_{\pi^M}b, {}_{\pi^M}c)$ as a discounted superposition solution to ν^M. The argument of tightness is similar to that of Sect. A.3.2 with the modified bound

$$\int_0^T \int \big(\Theta_1(|a_t|) + \Theta_2(|b_t|) + \Theta_3(|c_t|)\big)\, \mathrm{d}\nu_t\, \mathrm{d}t < \infty$$

instead of (A.4). As before, we obtain the inequalities

$$\begin{aligned}\int \Psi(x_R^i \circ \gamma)\, \mathrm{d}\boldsymbol{\eta}^M(\gamma) \le \int \theta(|x_R^i|)\, \mathrm{d}\eta_0^M &\\ + \int_0^T \int \big(\Theta(|\mathcal{L}_t^{S,M} x_R^i|) &\\ + \Theta(b_t^M : \nabla x_R^i \otimes \nabla x_R^i)\big)\, \mathrm{d}\eta_t^M\, \mathrm{d}t&,\\ \int \Psi(f^{d+1} \circ \gamma)\, \mathrm{d}\boldsymbol{\eta}^M(\gamma) \le \int \theta(|f^{d+1}|)\, \mathrm{d}\eta_0^M + \int_0^T \int \Theta(|c_t^M|)\, \mathrm{d}\eta_t^M\, \mathrm{d}t&.\end{aligned}$$

By a reasoning similar to Sect. A.3.1, we obtain via Jensen's inequality the uniform bounds

$$\int \Psi(\gamma^i)\, \mathrm{d}\boldsymbol{\eta}^M(\gamma) \le 1 + \int_0^T \int \big(\Theta_1(|a_t^i|) + \Theta_2(|b_t^{i,i}|)\big)\, \mathrm{d}\nu_t^A\, \mathrm{d}t,$$

$$\int \Psi(\log \gamma^{d+1})\, \mathrm{d}\boldsymbol{\eta}^M(\gamma) \le \theta(0) + \int_0^T \int \Theta_3(|c_t|)\, \mathrm{d}\nu_t^A\, \mathrm{d}t.$$

That the result also holds for the limit of this sequence follows directly from [68, Sects. A.2 and A.3] and Sect. A.2. Therefore we have shown Theorem 3.5 for locally bounded coefficients a, b and c.

A.3.4 General case

The approximation is achieved via convolution, with the approximating ν^ε having superposition solutions since the approximating coefficients satisfy (A.6). The tightness follows an identical argument, with bounds similar to those obtained in Sect. A.3.2 for Ψ with respect to Θ. That the result also holds for the limit of this sequence follows directly from [68, Sects. A.2 and A.3] and Sect. A.2. Therefore we have shown Theorem 3.5 for integrable coefficients a, b and c. This concludes the proof. □

Declarations

Competing interests The authors declare no competing interests.

References

1. Ambrosio, L., Figalli, A.: On flows associated to Sobolev vector fields in Wiener spaces: an approach à la DiPerna–Lions. J. Funct. Anal. **256**, 179–214 (2009)
2. Ambrosio, L., Gigli, N., Savaré, G.: Gradient Flows: In Metric Spaces and in the Space of Probability Measures. Springer, Berlin (2005)
3. Ambrosio, L., Trevisan, D.: Well-posedness of Lagrangian flows and continuity equations in metric measure spaces. Anal. PDE **7**, 1179–1234 (2014)
4. Atlan, M.: Localizing volatilities. Preprint (2006). Available online at https://arxiv.org/abs/math/0604316
5. Avellaneda, M., Friedman, C., Holmes, R., Samperi, D.: Calibrating volatility surfaces via relative-entropy minimization. Appl. Math. Finance **4**, 37–64 (1997)
6. Backhoff-Veraguas, J., Beiglböck, M., Schachermayer, W., Tschiderer, B.: The structure of martingale Benamou–Brenier in $\mathbb{R}^d$. Preprint (2023). Available online at https://arxiv.org/abs/2306.11019
7. Bain, A., Mariapragassam, M., Reisinger, C.: Calibration of local-stochastic and path-dependent volatility models to vanilla and no-touch options. J. Comput. Finance **24**(4), 115–161 (2021)
8. Barles, G., Souganidis, P.: Convergence of approximation schemes for fully nonlinear second order equations. Asymptot. Anal. **4**, 271–283 (1991)
9. Beiglböck, M., Lim, T., Obłój, J.: Dual attainment for the martingale transport problem. Bernoulli **25**, 1640–1658 (2019)
10. Benamou, J.-D., Brenier, Y.: A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. Numer. Math. **84**, 375–393 (2000)
11. Bergomi, L.: Stochastic Volatility Modeling. CRC Press, Boca Raton (2015)
12. Bouchard, B., Loeper, G., Zou, Y.: Hedging of covered options with linear market impact and gamma constraint. SIAM J. Control Optim. **55**, 3319–3348 (2017)
13. Box, G.E.P.: Statistics as a catalyst to learning by scientific method part II–a discussion. J. Qual. Technol. **31**, 16–19 (1999)
14. Brace, A., Gątarek, D., Musiela, M.: The market model of interest rate dynamics. Math. Finance **7**, 127–155 (1997)
15. Brenier, Y.: A homogenized model for vortex sheets. Arch. Ration. Mech. Anal. **138**, 319–353 (1997)
16. Brenier, Y.: Minimal geodesics on groups of volume-preserving maps and generalized solutions of the Euler equations. Commun. Pure Appl. Math. **52**, 411–452 (1999)
17. Brigo, D., Mercurio, F.: Interest Rate Models – Theory and Practice: With Smile, Inflation and Credit. Springer, Berlin (2007)
18. Brunick, G., Shreve, S.: Mimicking an Itô process by a solution of a stochastic differential equation. Ann. Appl. Probab. **23**, 1584–1628 (2013)
19. Chizat, L., Peyré, G., Schmitzer, B., Vialard, F.-X.: Unbalanced optimal transport: dynamic and Kantorovich formulations. J. Funct. Anal. **274**, 3090–3123 (2018)
20. Cox, J.: The constant elasticity of variance option pricing model. J. Portf. Manag. **23**(5), 15–17 (1996)
21. Cozma, A., Mariapragassam, M., Reisinger, C.: Calibration of a hybrid local-stochastic volatility stochastic rates model with a control variate particle method. SIAM J. Financ. Math. **10**, 181–213 (2019)
22. Crandall, M., Ishii, H., Lions, P.-L.: User's guide to viscosity solutions of second order partial differential equations. Bull. Am. Math. Soc. **27**, 1–67 (1992)
23. Deelstra, G., Rayée, G.: Local volatility pricing models for long-dated FX derivatives. Appl. Math. Finance **20**, 380–402 (2013)
24. Dupire, B.: Pricing with a smile. Risk **7**(1), 18–20 (1994)
25. Dupire, B.: A unified theory of volatility. In: Carr, P. (ed.) Derivatives Pricing: The Classic Collection, pp. 185–196. Risk (2004)
26. E, W., Han, J., Jentzen, A.: Algorithms for solving high dimensional PDEs: from nonlinear Monte Carlo to machine learning. Nonlinearity **35**, 278–310 (2021)
27. Engelmann, B., Koster, F., Oeltz, D.: Calibration of the Heston stochastic local volatility model: a finite volume scheme. Int. J. Financ. Eng. **8**, 2050048 (2021)
28. Figalli, A.: Existence and uniqueness of martingale solutions for SDEs with rough or degenerate coefficients. J. Funct. Anal. **254**, 109–153 (2008)
29. Guo, I., Langrené, N., Loeper, G., Ning, W.: Portfolio optimization with a prescribed terminal wealth distribution. Quant. Finance **22**, 333–347 (2022)
30. Guo, I., Loeper, G.: Path dependent optimal transport and model calibration on exotic derivatives. Ann. Appl. Probab. **31**, 1232–1263 (2021)

31. Guo, I., Loeper, G., Obłój, J., Wang, S.: Joint modeling and calibration of SPX and VIX by optimal transport. SIAM J. Financ. Math. **13**, 1–31 (2022)
32. Guo, I., Loeper, G., Obłój, J., Wang, S.: Optimal transport for model calibration. Risk Mag. (2022). Available online at https://ora.ox.ac.uk/objects/uuid:72e2acbc-1163-49e4-ae5e-cb1afabf1e3c/files/sz890rv08v
33. Guo, I., Loeper, G., Wang, S.: Local volatility calibration by optimal transport. In: de Gier, J. et al. (eds.) 2017 MATRIX Annals, pp. 51–64. Springer, Cham (2019)
34. Guo, I., Loeper, G., Wang, S.: Calibration of local-stochastic volatility models by optimal transport. Math. Finance **32**, 46–77 (2022)
35. Guyon, J.: The joint S&P 500/VIX smile calibration puzzle solved. Risk Mag. (2020). https://doi.org/10.2139/ssrn.3397382
36. Guyon, J., Henry-Labordère, P.: The smile calibration problem solved. Preprint (2011). Available online at https://doi.org/10.2139/ssrn.1885032
37. Guyon, J., Henry-Labordère, P.: Being particular about calibration. Risk **25**, 88–93 (2012)
38. Gyöngy, I.: Mimicking the one-dimensional marginal distributions of processes having an Itô differential. Probab. Theory Relat. Fields **71**, 501–516 (1986)
39. Hagan, P., Kumar, D., Lesniewski, A., Woodward, D.: Managing smile risk. Wilmott Mag., 84–108 (2002)
40. Han, J., Jentzen, A., E, W.: Solving high-dimensional partial differential equations using deep learning. Proc. Natl. Acad. Sci. USA **115**, 8505–8510 (2018)
41. Henrard, M.: The irony in the derivatives discounting. Wilmott Mag., 92–98 (2007)
42. Henrard, M.: The irony in derivatives discounting part II: the crisis. Wilmott Mag. **2**, 301–316 (2010)
43. Henry-Labordère, P.: Calibration of local stochastic volatility models to market smiles: a Monte-Carlo approach. Risk **22**, 112–117 (2009)
44. Hok, J., Tan, S.-H.: Calibration of local volatility model with stochastic interest rates by efficient numerical PDE methods. Decis. Econ. Finance **42**, 609–637 (2019)
45. Huesmann, M., Trevisan, D.: A Benamou–Brenier formulation of martingale optimal transport. Bernoulli **25**, 2729–2757 (2019)
46. Hull, J., White, A.: Pricing interest-rate-derivative securities. Rev. Financ. Stud. **3**, 573–592 (1990)
47. Hull, J., White, A.: A note on the models of Hull and White for pricing options on the term structure: response. J. Fixed Income **5**(2), 97–102 (1995)
48. In't Hout, K., Foulon, S.: ADI finite difference schemes for option pricing in the Heston model with correlation. Int. J. Numer. Anal. Model. **7**, 303–320 (2010)
49. Jacod, J., Shiryaev, A.: Limit Theorems for Stochastic Processes, 2nd edn. Springer, Berlin (2003)
50. Jamshidian, F.: LIBOR and swap market models and measures. Finance Stoch. **1**, 293–330 (1997)
51. Joseph, B., Loeper, G., Obłój, J.: Joint calibration of local volatility models with stochastic interest rates using semimartingale optimal transport. Quant. Finance **24**, 1597–1620 (2024)
52. Joseph, B., Loeper, G., Obłój, J.: The measure preserving martingale Sinkhorn algorithm. Preprint (2023). Available online at https://arxiv.org/abs/2310.13797
53. Karatzas, I., Shreve, S.: Brownian Motion and Stochastic Calculus, 2nd edn. Springer, Berlin (1998)
54. Krylov, N.: Once more about the connection between elliptic operators and Itô's stochastic equations. In: Krylov, N. et al. (eds.) Statistics and Control of Stochastic Processes, Steklov Seminar, pp. 214–229. Optimization Software, New York (1985)
55. Lions, P.-L.: Optimal control of diffusion processes and Hamilton–Jacobi–Bellman equations part 2: viscosity solutions and uniqueness. Commun. Partial Differ. Equ. **8**, 1229–1276 (1983)
56. Liu, D., Nocedal, J.: On the limited memory BFGS method for large scale optimization. Math. Program. **45**, 503–528 (1989)
57. Loeper, G.: The reconstruction problem for the Euler–Poisson system in cosmology. Arch. Ration. Mech. Anal. **179**, 153–216 (2006)
58. Ma, K., Forsyth, P.: An unconditionally monotone numerical scheme for the two-factor uncertain volatility model. IMA J. Numer. Anal. **37**, 905–944 (2017)
59. Miltersen, K., Sandmann, K., Sondermann, D.: Closed form solutions for term structure derivatives with log-normal interest rates. J. Finance **52**, 409–430 (1997)
60. Nesterov, Y.: Lectures on Convex Optimization. Springer, Berlin (2018)
61. Ögetbil, O., Ganesan, N., Hientzsch, B.: Calibrating local volatility models with stochastic drift and diffusion. Int. J. Theor. Appl. Finance **25**, 2250011 (2022)
62. Ren, Y., Madan, D., Qian, M.: Calibrating and pricing with embedded local volatility models. Risk Mag. **20**, 138–143 (2007)

63. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
64. Schmidt, M.: minfunc: unconstrained differentiable multivariate optimization in MATLAB (2005). Available online at https://www.cs.ubc.ca/~schmidtm/Software/minFunc.html
65. Séjourné, T., Peyré, G., Vialard, F.-X.: Unbalanced optimal transport, from theory to numerics. In: Trélat, E. and Zuazua, E. (eds.) Handbook of Numerical Analysis, vol. 24, pp. 407–471. Elsevier, Amsterdam (2023)
66. Stroock, D., Varadhan, S.: Multidimensional Diffusion Processes. Springer, Berlin (1979)
67. Tan, X., Touzi, N.: Optimal transportation under controlled stochastic dynamics. Ann. Probab. **41**, 3201–3240 (2013)
68. Trevisan, D.: Well-posedness of multidimensional diffusion processes with weakly differentiable coefficients. Electron. J. Probab. **21**, 1–41 (2016)
69. Villani, C.: Topics in Optimal Transportation, vol. 58. Am. Math. Soc., Providence (2003)
70. Wyns, M., Du Toit, J.: A finite volume–alternating direction implicit approach for the calibration of stochastic local volatility models. Int. J. Comput. Math. **94**, 2239–2267 (2017)