DISS. ETH NO. 24915

CALIBRATION, FILTERING AND HEDGING: NON-LINEAR INFORMATION PROCESSING IN MATHEMATICAL FINANCE

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

LUKAS OLIVER GONON

MSc ETH Mathematics

born on 01.11.1989

citizen of Schleitheim, SH

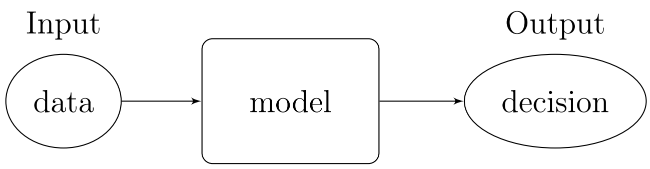
accepted on the recommendation of

Prof. Dr. Josef Teichmann examiner
Prof. Dr. Dan Crisan co-examiner

2018

Abstract

The modeling paradigm is ubiquitous in contemporary financial industry: within a pre-specified class of stochastic processes, one is selected as a model by calibration to market prices of options or estimation based on time-series of historical data. It is then used to calculate prices and hedging strategies for new derivative products. Hence input information is mapped to real-world decisions in a highly non-linear way.



Thus, it is very important to quantify uncertainty about model output and design decision procedures that depend as little as possible on modeling assumptions. However, one faces a dilemma, because a more realistic model usually comes at the price of lower numerical tractability and requires more input data in the model selection step. These challenging tasks have been a major focus of the recent mathematical finance literature. Approaches include combining calibration and estimation within a sophisticated Bayesian framework, i.e. high-dimensional filtering, looking for model-independent bounds for option prices, and designing novel data driven numerical procedures for more realistic models.

This thesis examines these questions in three different areas:

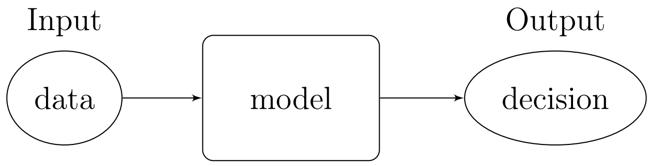
Firstly, we consider the filtering problem for the class of affine processes. While being very popular in financial modeling, their proper use in multi-dimensional models has been limited by the lack of a numerically tractable filtering methodology. We solve this problem with a novel surprisingly simple approximation technique based on generalized stochastic Riccati ordinary differential equations. Our theoretical results are complemented by numerical experiments for Cox–Ingersoll–Ross and Wishart processes illustrating the efficiency of the method.

Secondly, we consider the problem of hedging a derivative in a scenario based discrete-time market with realistic transaction costs. Such a setting is more realistic than most of the models used in practice today, but of course it suffers from numerical intractability. We demonstrate that this has changed thanks to recent technological advances: using hedging strategies built from neural networks and machine learning optimization techniques, optimal hedging strategies can be approximated very well, as our numerical study and our theoretical results show in an impressive way.

Finally, we study the Skorokhod embedding problem for Lévy processes with non-deterministic initial value. Due to correspondence of option price data to marginal distributions, this problem is intimately connected to robust finance, but it is also of independent mathematical interest and no (explicit) solution has been known to date. Using time-change techniques and ideas inspired from local volatility calibration, we construct a solution to this problem, which is new also in the special case of Brownian motion.

Kurzfassung

Das Paradigma der Modellierung ist allgegenwärtig in der heutigen Finanzindustrie: aus einer vorgegebenen Klasse von stochastischen Prozessen wird - mittels Kalibrierung an Marktpreise von Optionen oder Parameterschätzung basierend auf Zeitreihen von historischen Daten - ein Spezifischer als Modell ausgewählt und dann zur Berechnung von Preisen und Hedging-Strategien für zusätzliche Derivate benutzt. Dabei wird Input-Information auf höchst nichtlineare Art und Weise in praktische Entscheidungen übersetzt.



Es ist daher sehr wichtig die Unsicherheit über den Modell-Output zu quantifizieren und Entscheidungsprozesse zu wählen, die so wenig wie möglich von den Modellannahmen abhängen. Dies führt zu einem Dilemma, da ein realistischeres Modell meist numerisch schwieriger zu handhaben ist und bei der Modellauswahl mehr Input-Daten benötigt. Diese Herausforderungen stehen im Zentrum der modernen finanzmathematischen Forschung. Lösungsansätze sind unter anderem Kalibrierung und Schätzung in einem ausgefeilten bayesschen Framework zu kombinieren, d.h. hochdimensionales *filtering*, modellunabhängige Schranken für Optionspreise zu suchen und neue, datenbasierte numerische Methoden für realistischere Modelle zu entwickeln.

Diese Doktorarbeit untersucht obige Fragen in drei verschiedenen Gebieten:

Zuerst wird das Filtering-Problem für die Klasse der affinen Prozesse behandelt. Diese sind sehr beliebt in der praktischen Modellierung, ihr Einsatz in mehrdimensionalen Modellen war aber bisher nur beschränkt möglich mangels einer effizienten numerischen Filtering-Methode. Wir lösen dieses Problem und stellen eine neue, überraschend einfache Approximationstechnik zur Verfügung, die auf verallgemeinerten stochastischen Riccati-Differentialgleichungen basiert. Wir ergänzen unsere theoretischen Ergebnisse mit numerischen Experimenten für Cox-Ingersoll-Ross und Wishart-Prozesse und illustrieren damit die Effizienz der Methode.

Als nächstes beschäftigen wir uns mit dem Hedging-Problem in einem szenariobasierten, zeitdiskreten Markt mit realistischen Transaktionskosten. In der

Praxis sind solche Modelle bis anhin kaum benutzt worden, da numerische Berechnungen nicht möglich waren. Wir zeigen mit numerischen und theoretischen Resultaten, dass sich dies dank des technologischen Fortschritts geändert hat: Mittels Hedging-Strategien, die auf neuronalen Netzwerken basieren, und Optimierungstechniken des maschinellen Lernens können optimale Hedging-Strategien sehr effizient berechnet und approximiert werden.

Zuletzt untersuchen wir das Skorokhod-Embedding-Problem für Lévy-Prozesse mit nichtdeterministischem Startwert. Dieses Problem ist sehr nahe mit Fragen der robusten Finanzmathematik verwandt, es ist aber auch von unabhängigem mathematischen Interesse und bisher gab es keine (explizite) Lösung dazu. Wir benutzen Time-Change-Techniken und Ideen, die von der Kalibrierung von Local-Volatility-Modellen inspiriert sind, um eine Lösung zu konstruieren. Diese ist auch im Spezialfall der brownschen Bewegung neu.

Acknowledgements

Some people might say that writing a doctoral thesis is the last step of a solid academic education. Just like in other types of schooling, a PhD student encounters not only one, but various teachers. Let me express my gratitude to all of them for what they have taught me: most importantly to my supervisor Josef Teichmann and my collaborators Leif Döring, David Prömel and Oleg Reichmann as well as Hans Bühler and Ben Wood. But also to all my friends and colleagues and everyone else with whom I have shared the pleasure of an inspiring discussion.

Special thanks to Dan Crisan for agreeing to act as a co-examiner.

Finally, I would like to thank Emil and his still nameless brother for their sincere and refreshing lack of interest in my work and Cécile for her uncountable support.

Contents

Chapter I

Introduction

Randomness is understood as a quantifiable uncertainty about future outcomes: while the exact future evolution of a system is unknown and can not be forecasted perfectly, one may nevertheless weight different outcomes as more or less likely, according to their probability of occurrence. Stochastic processes mathematically describe the dynamic random evolution of a system. They are widely used for modeling time-dependent quantities in many areas of science, technology, economics and finance. This amounts to choosing a stochastic process that describes in the best possible way some data observed to date and related to this quantity. Based on the chosen process one makes predictions about future outcomes and takes real-world decisions.

A bank might choose, e.g., a Heston model to describe the evolution of the price of a stock and its volatility. The model parameters would be chosen so that the model prices fit as closely as possible the market prices of call options on that stock. Subsequently, the bank would use the model (based on the calibrated parameters) to calculate prices for additional derivative products and sell these products to clients in reality. While this particular model choice is not consistent with all observations, as call option prices for different maturities would require different parameters, the procedure described above is at the basis of day-to-day practice in the financial industry (with varying models, derivatives products, asset classes, ...). Let us elaborate on the modeling paradigm in this specific context of equity options and provide further examples.

As an alternative to the standard Heston model, the bank might choose to calibrate a local volatility extended Heston model. This is a non-parametric model and can be chosen perfectly consistent with all observed prices of European call options on the given stock. However, this model is much less numerically tractable and to fully specify it, knowledge of option prices for all expirations and all strikes would be required. Since this data is available on a discrete grid only, an interpolation (e.g. on the level of implied volatility) not introducing arbitrage has to be chosen. Thus, to compensate the lack of data, further modeling assumptions have to be made and different choices of such assumptions may lead to different prices for the additional derivative products.

Having sold such a derivative product, the bank is now exposed to additional financial risk: it has received a cash premium (the price) and in exchange agreed to make a payment at a future date, the expiry date. The size of the payment is currently unknown and linked to the stock price at the future date. To hedge this exposure, i.e. to minimize the risk of a financial loss caused by this unknown future payment, the bank trades in the underlying (and possibly additional derivatives). It readjusts its holdings in each of these hedging instruments e.g. every day up to the expiry date. The precise amounts to be held on a given day are calculated from the pricing model (see above), which is used to make a prediction of the price of the product on the subsequent day and the sensitivity of this price with respect to changes in the underlying and e.g. its volatility.

In practice the bank also sells products written on several underlyings and not just a single stock as considered above. To calculate prices and determine hedging strategies for such multi-asset products, the bank now also specifies a correlation model for the underlyings. However, even for relatively simple model choices the very limited multi-asset options data that is available may not be sufficient to fully specify the model parameters and, even if there is enough data, calibration may be highly challenging numerically. This could be resolved by choosing e.g. a multivariate generalization of the Heston model and modeling the covariance matrix process by a Wishart process. Alternatively, the bank may decide to estimate model parameters from time-series of past prices instead of calibrating these. In whichever way model selection is performed, as in the previous examples data is translated to pricing and hedging decisions in a non-linear way.

In these examples or any other application leading to real-world decisions it is therefore important to address model (or Knightian) uncertainty, i.e. to take into account that the chosen stochastic model is only an idealization of reality and, to deal with this, aim at decisions and procedures which depend as little as possible on the modeling assumptions. Such questions of robustness with respect to model misspecification have been a major focus of the mathematical finance literature in the past decade. Handling model uncertainty *within a model* can of course never fully resolve the issue, but it can help to replace standard models with significantly more robust ones. We refer to this general (and not sharply separated) task as *dynamic uncertainty modeling*, see also [CKT17]. It addresses one of the key problems of stochastic modeling: different models may lead to inconsistent decisions. As elaborated above, however, a challenge that one faces when addressing this is the opposing second key problem: a good model needs a lot of data and computation time, which may not be given due to real-time evaluation requirements.

This thesis studies three instances of dynamic uncertainty modeling, each of them treated in an independent chapter, addressing a different problem. We now explain each of the chapters very briefly from the perspective of dynamic uncertainty modeling. Subsequently, we will give a more detailed overview of each of the chapters.

Consider a Black-Scholes model under the real-world probability measure. Since the drift parameter is notoriously hard to estimate (see e.g. [Rog01, Section 5]), it is essential to incorporate the uncertainty about the drift into the model. Updating this uncertainty by subjective prior information leads to the seductively attractive Bayesian approach: one models the drift with a (latent) stochastic process. The distribution of this process at time t conditional on the observations (in our case the price process) up until t then *quantifies* the uncertainty about the drift *within the model*. Calculating this distribution is the classical *filtering problem*, see e.g. [BC09a]. For Gaussian models (i.e. when a linear stochastic differential equation is taken as a prior for the drift) this leads to the famous and numerically tractable Kalman filter, but in general (non-Gaussian and higher-dimensional) situations numerical calculation is highly challenging. Chapter II of this thesis studies the filtering problem for the more general class of *affine processes* (see e.g. [DFS03],[CFMT11]) and introduces a new approximation scheme that allows for efficient numerical calculation in high dimension. We provide theoretical results on the associated (stochastic) Riccati equations and illustrate the method by numerical experiments for Cox-Ingersoll-Ross and Wishart processes.

In Chapter II we have developed an efficient numerical approximation method in a classical spirit, but it is not always possible to do this. In the sequel we use novel solution concepts and techniques inspired by machine learning.

The brief summary of Chapter II given above provides us with a perfect example of the key modeling dilemma: a more realistic model usually comes at the price of lower numerical and computational tractability. In parallel, as illustrated in the local volatility example given above, a more sophisticated model also requires more data in the model selection step. Lack of data can only be compensated by additional assumptions, which may interfere with the original aim of a more realistic model.

Thus, even though in reality trading is subject to transaction costs, a bank selling a derivative will base (real-world) hedging decisions on an idealized model neglecting such costs. The reason for this is simply that more realistic models have not been numerically tractable so far. Thanks to recent technological advances, however, this is about to change, as the study presented in Chapter III illustrates. Here the problem of hedging a derivative in a discrete-time (incomplete) market with frictions (e.g. transaction costs or temporary market impact) is considered. Risk-preferences are specified in terms of a convex risk measure. *It is not possible to calculate indifference prices (and associated optimal hedging strategies) in such a framework using classical numerical techniques.* However, in Chapter III we show that, using modern machine learning techniques, numerical calculation *is* feasible, but we dismiss the concept of a classical solution accompanied by an optimal algorithm approximating it. The key idea is to parametrize hedging strategies by neural networks and use machine learning optimization techniques to train them. This is indeed feasible even on a standard laptop thanks to highly efficient numerical packages such as TensorFlow, Theano and Torch. We prove theoretical approximation results and present various numerical experiments in Python. Note that here uncertainty about modeling assumptions is not part of the model, but rather of the decision procedure: the algorithm and the implementation can handle a wide range of market environments in the same framework and so one may run it for a variety of different modeling assumptions in parallel.

An alternative and very popular approach to dynamic uncertainty modeling is taken in the area of *robust finance*, see e.g. [Hob98, BHR01b, CL10, CO11], [BHLP13, DS14, DOR14, GHLT14, BN15] and references therein. One does not incorporate incoming information dynamically, but instead takes a static approach and specifies a family of possible models (with minimal assumptions, e.g. only imposing consistency with prices of initially observed call options) and aims at finding a hedging strategy (and associated price bounds) that super-replicates the payoff of a given derivative in all of these models. In order to prove that these bounds are tight, one often needs to construct a martingale that has prescribed marginal distributions (and potentially additional properties, see e.g. [Hob98], [BHR01b], [Hob11]). By a random time-change, this is equivalent to the *Skorokhod embedding problem* (SEP), see e.g. [Obł04] for an overview. For both of these problems there exist a variety of solutions with continuous trajectories (see e.g. [HPRY11]), but results on processes with jumps are scarce. In particular, so far there has been no (non-randomized) solution for the SEP for general jump processes with non-deterministic starting value. In Chapter IV we solve this problem for the class of Lévy processes ([Ber96],[Sat99],[Kyp14]) and thereby also obtain a new solution for the case of Brownian motion. Our construction is very natural from the point of view of mathematical finance: it is based on inverting the Fokker-Planck equation associated to the Lévy process. This can be seen as calibrating a local volatility model (as in [Dup94], [CGMY04]) and amounts to time-changing the Lévy process. However, making this precise is technically delicate (the resulting local volatility function is time-dependent, it may be unbounded and exhibit zeros) and requires results on time-changes and uniqueness of Fokker-Planck equations that have not yet been available in the literature. These are developed in the general framework of martingale problems of [EK86] in the appendix of Chapter IV.

We now give a more detailed overview on each of the chapters. The three chapters correspond to [GT17],[BGTW17] and [DGPR17], respectively.

Chapter II: Affine Filtering

This chapter is devoted to the filtering problem for affine processes. The results can also be found in [GT17].

More precisely, we set $D = \mathbb{R}_+^m \times \mathbb{R}^{d-m}$ and consider a conservative affine process $((X_t)_{t\geq 0}, (\mathbb{P}_x)_{x\in D})$ with state space D. X is considered a latent signal, about which inference is only made through the observation process Y defined by

$$Y_t = \int_0^t X_s \, \mathrm{d}s + W_t, \quad t \in [0,T], \tag{1.1}$$

for a d-dimensional Brownian motion W independent of X. The filtering problem is to calculate the distribution of X_t conditional on $\mathcal{F}_t^Y := \sigma(Y_s \colon s \in [0,t])$, the *filtering distribution*, for any $t \in [0,T]$.

Theoretically, this problem has been solved in great generality and so also in the present setting we may use [KO88] to characterize the (process of) conditional distribution(s) as the unique solution to the (measure-valued) Kushner-Stratonovich stochastic differential equation. Since X is a $\mathbb{P}_x$-semimartingale, one may take an alternative approach (originating from [Cla78],[Dav80]) and consider the *pathwise filtering functional*, defined for $x \in D$, $t \in [0,T]$, $f \in B(D)$ and $y \in C([0,T],\mathbb{R}^d)$ by

$$\sigma_t^x(f,y) = \mathbb{E}_x\left[f(X_t)\exp\left(y_t^\top X_t - \int_0^t y_s^\top \,\mathrm{d}X_s - \frac{1}{2}\int_0^t |X_s|^2\,\mathrm{d}s\right)\right]. \tag{1.2}$$

Then $\sigma_t^x(f,Y)/\sigma_t^x(1,Y) = \mathbb{E}_x[f(X_t)|\mathcal{F}_t^Y]$, $\mathbb{P}_x$-a.s and so $\sigma_t^x(\cdot,Y)$ can be used to fully describe the filtering distribution at t. See also [BC09a, Chapters 2-5] for more details.

Due to the infinite-dimensional structure of the problem (except in special cases, see [BC09a, Chapter 6]), it has remained highly challenging to numerically calculate the filtering distribution. Well-established methods exist for approximately Gaussian or very low-dimensional settings (see [BC09a, Chapters 8-10]), but this does not cover the case of (general) affine processes. Motivated by the prominent role of affine processes in financial modeling (see e.g. [DFS03], [KRM15, Section 3]), in Chapter II we fill this gap by providing an ordinary differential equation based numerical filtering method.

In more detail, the approach taken in Chapter II is to approximate (1.2) by a *linearized* functional of the form

$$\rho_t^x(f,y) = \mathbb{E}_x\left[f(X_t)\exp\left(y_t^\top X_t - \int_0^t y_s^\top \,\mathrm{d}X_s - \int_0^t \gamma_s^\top X_s - c_s\,\mathrm{d}s\right)\right] \tag{1.3}$$

with appropriately chosen $\gamma \in C([0,T],\mathbb{R}^d)$ and $c \in C([0,T],\mathbb{R})$. The key reason for doing this is that evaluations of $\rho_t^x(\cdot,y)$ at the Fourier basis can be calculated by solving a system of (generalized Riccati) ordinary differential equations:

More precisely, as one of our main contributions, we show that there exists $T_0 > 0$ such that for all $u \in i\mathbb{R}^d$ and $T \le T_0$ the system

$$\begin{aligned} -\partial_t\Phi(t,T,u) &= F(\Psi(t,T,u) - y_t) - c_t, \quad \Phi(T,T,u) = 0 \\ -\partial_t\Psi(t,T,u) &= R(\Psi(t,T,u) - y_t) - \gamma_t, \quad \Psi(T,T,u) = u + y_T, \quad 0 \le t \le T. \end{aligned} \tag{1.4}$$

has a unique solution $\Phi(\cdot,T,u) \in C^1([0,T],\mathbb{R})$, $\Psi(\cdot,T,u) \in C^1([0,T],\mathbb{R}^d)$ and that

$$\rho_T^x(f_u,y) = \exp(\Phi(0,T,u) + x^\top\Psi(0,T,u)), \tag{1.5}$$

where $f_u(z) := \exp(u^\top z)$, $z \in D$.

Here F and R are vector fields associated to the (characteristic function of the) affine process and of Lévy-Khintchine-form. The representation (1.5) can then be used to devise Fourier filtering techniques, analogously to the Fourier pricing techniques used for affine (log-price) models, see e.g. [CM99] and [DFS03]. Note that we need to impose an exponential moment condition on the jump-measures, without which (1.3) may not be finite for any $t > 0$. The proof of the above result is based on a change of measure (relying on [KMK10],[CFY05]) and comparison results for generalized Riccati equations. These are of independent interest and extend results from [KMK10] and [KRM15] to Riccati equations associated to non-conservative time-inhomogeneous affine "processes" that do not necessarily satisfy the admissibility conditions. Finally, we also provide an extension of the above result to the smoothing distribution, non-deterministic initial law and give an interpretation in terms of time-inhomogeneous affine processes (see e.g. [Fil05]).

These theoretical results are complemented by numerical experiments for a Cox–Ingersoll–Ross (CIR) process [CIR85] and a Wishart process [Bru91]. The latter is a matrix-valued extension of the former and a special case of affine processes taking values in the set of symmetric positive semi-definite matrices (see [CFMT11]). For both processes, standard approximate non-linear filters (such as the extended and ensemble Kalman filter, see [BC09a, Chapter 8] and [LSZ15]) do not give accurate results, since they work under a paradigm of Gaussian approximations. However, if the Feller condition is not satisfied, the marginal distributions of CIR and Wishart processes are highly non-Gaussian close to their well-known boundaries. Therefore, such methods fail for these processes. To overcome this, alternative parametric families of distributions (e.g. a Gamma distribution in [Bat06]) have been proposed. However, so far the problem of finding a general procedure applicable also to higher-dimensional and general affine processes has remained open.

To illustrate the accuracy and feasibility of our method, in experiments for the CIR process the filter induced by (1.3) is compared to the benchmark (a bootstrap particle filter) and two standard Gaussian- and Gamma-approximation approaches (extended Kalman filter and [Bat06]) and shown to be more accurate than the two other approximations. For the Wishart process the situation turns out to be even more extreme: numerical experiments (already) for $d = 3$ show that in order to achieve the same level of accuracy (measured in terms of mean square error) by a particle filter as is obtained by the filter induced by (1.3), an outrageous number of particles is necessary.

Chapter III: Deep Hedging

In this chapter we apply deep learning techniques to the problem of pricing and hedging in discrete-time market models with frictions. The chapter is based on [BGTW17].

Let us start by a formal description of the market model. On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with filtration $\mathbb{F} = (\mathcal{F}_{t_k})_{k=0,\ldots,n}$ we consider a discrete-time $\mathbb{F}$-adapted stochastic process $S = (S_{t_k})_{k=0,\ldots n}$ in $\mathbb{R}^d$ representing the prices of available hedging instruments. At time $t_0 = 0$ an agent sells a contingent claim, yielding a (random, $\mathcal{F}_T$-measurable) payoff Z at time $t_n = T$ for the buyer. The agent charges price p_0 and trades dynamically in S according to an $\mathbb{F}$-predictable hedging strategy δ, i.e. she readjusts the number of units she holds in each of the hedging instruments at t_j, $j = 0, \ldots, n-1$. Assuming risk-free borrowing and lending (which just means that we are working with discounted quantities), the value of her position at T is thus given as

$$\mathrm{PL}_T(Z, p_0, \delta) := -Z + p_0 + (\delta \cdot S)_T - C_T(\delta), \tag{1.6}$$

where $\cdot$ denotes discrete-time stochastic integration and $C_T(\delta)$ denote the cumulative transaction costs associated to δ.

This leads to the problem of hedging her exposure to Z: ideally, the agent would like replicate Z, i.e. to choose p_0 and δ such that (1.6) is equal to 0, $\mathbb{P}$-a.s. However, this is not possible in general and so she needs to specify risk-preferences. For example, she could aim at finding δ and p_0 that minimize the variance of (1.6). Since such an approach penalizes both gains and losses, here we follow e.g. [FL00], [Xu06], [KS07], [IJS09] and describe risk-preferences by a convex risk measure ρ, see [FS16]. Denote by $\mathcal{H}$ the set of all available hedging strategies. Optimal pricing and hedging then amounts to calculating the (a posteriori unique) *indifference price* $p(Z)$ defined implicitly by

$$\inf_{\delta \in \mathcal{H}} \rho \left(\mathrm{PL}_T(Z, p(Z), \delta) \right) = \inf_{\delta \in \mathcal{H}} \rho \left(\mathrm{PL}_T(0, 0, \delta) \right), \tag{1.7}$$

and an optimal hedging strategy for Z is a minimizer (if it exists) for the left hand side of (1.7). Thus, her risk when selling the contingent claim at price $p(Z)$ and hedging optimally is the same as when not entering into the transaction.

While such a framework is highly flexible and general (and e.g. includes exponential utility indifference pricing), numerical calculation has been extremely challenging even in low-dimensional and more idealized settings, see e.g. [HN89], [DPZ93],[WW97] and [KMK15]. The approach pursued in Chapter III is to only consider hedging strategies *built from neural networks*, a parametric family of functions that can be used to approximate any multivariate function (surprisingly) efficiently (see [BGKP17]) and for which efficient machine learning optimization algorithms (see [GBC16]) and toolboxes implementing these (we use TensorFlow here) are available. In Chapter III we demonstrate how one may build on this to calculate (1.7) numerically.

In detail, instead of optimizing over the whole of $\mathcal{H}$ in (1.7), we only consider strategies of the form

$$\delta^{\theta}_{t_k} := F^{\theta_k}(S_{t_0}, \ldots, S_{t_{k-1}}), \quad k = 1, \ldots, n, \tag{1.8}$$

where F^{θ_k} is a neural network with a fixed architecture (i.e. a fixed activation function, number of layers and nodes) and weights parametrized by θ_k. Indexing the "complexity" of the architecture by $M \in \mathbb{N}$, the set $\mathcal{H}_M$ of such hedging strategies is thus parametrized by some $\Theta_M \subset \mathbb{R}^p$ (for some $p \in \mathbb{N}$ depending on M) and so problem (1.7) with $\mathcal{H}$ replaced by $\mathcal{H}_M$ is an optimization problem over Θ_M. Denote by $p_M(Z)$ the indifference price for this approximate problem.

As a theoretical justification, we prove that as $M \to \infty$, the approximate indifference price $p_M(Z)$ converges to the price $p(Z)$ of the original problem (1.7).

The proof relies on a suitable version of the universal approximation theorem (see e.g. [Hor91] and the references in [BGKP17]), but some care is needed, since standard results work on compact spaces. Also note that for simplicity of the exposition here we have assumed $\mathcal{F}_{t_k} = \sigma(S_{t_0}, \dots, S_{t_k})$, but the result holds more generally (with a suitable modification of (1.8), see Chapter III). Furthermore, one may also consider the more restricted choice of *recurrent network* hedging strategies $\delta^\theta_{t_k} := F^{\theta_k}(S_{t_{k-1}}, \delta_{t_{k-1}})$ and in fact this turns out to be very efficient in numerical examples.

This simple but important result provides a theoretical justification for hedging based on neural networks. While for a special class of risk measures (see e.g. [BTT07]) one can then directly apply stochastic gradient type-algorithms to train the neural network, i.e. find a parameter $\theta \in \Theta_M$ that is close-to-minimal for the objective $\rho\left(\mathrm{PL}_T(Z, 0, \delta^\theta)\right)$. For general risk measures however, this objective is not directly amenable to these algorithms and so an additional approximation is required. By relying on the dual representation of ρ, we show that one may approximate the optimal hedging problem by solving

$$\inf_{\bar\theta \in \Theta_M} \sup_{\tilde\theta \in \tilde\Theta_M} J(\theta),$$

$$J(\theta) := \mathbb{E}\left[-\mathrm{PL}(Z, 0, \delta^{\bar\theta}) \exp(F^{\tilde\theta} \circ S)\right] - \bar\alpha(F^{\tilde\theta} \circ S) - \lambda_0(\mathbb{E}[\exp(F^{\tilde\theta} \circ S)] - 1) \tag{1.9}$$

where $\theta = (\bar\theta, \tilde\theta)$, $F^{\tilde\theta} \colon \mathbb{R}^{d(n+1)} \to \mathbb{R}$ is a neural network parametrized by $\tilde\theta \in \tilde\Theta_M$, λ_0 is a Lagrange multiplier and $\bar\alpha$ is related to the dual representation of ρ. We prove a similar convergence result as above and conclude the theoretical justifications by arguing that J is amenable to backpropagation and stochastic gradient type algorithms. Throughout the chapter we take a sample based approach and assume that Ω is finite, but this assumption is essential only for this last part.

These theoretical justifications are complemented by a numerical study demonstrating the surprising feasibility and accuracy of the method. To have a benchmark at hand, as a first example we consider a setting without transaction costs and where S is a discretized Heston [Hes93] stochastic volatility model under a risk-neutral measure. Note that in addition to the underlying one may also trade in a variance swap and so $d = 2$ hedging instruments are available. The deep hedge associated to the expected shortfall (also called conditional or average value at risk) risk measure for varying levels of risk-aversion is compared to the model delta-vega hedge and is seen to produce a very accurate approximation. As a second example, we examine the effect of proportional transaction costs on pricing. Denoting $p_\varepsilon = p_\varepsilon(Z)$ the exponential utility indifference price of Z associated to transaction costs of size ε, it has been established in a variety of one-dimensional models (see e.g. [WW97], [KMK15] and the references therein) that

$$p_\varepsilon - p_0 = O(\varepsilon^{2/3}), \qquad \text{as } \varepsilon \downarrow 0. \tag{1.10}$$

As one of our main contributions, using the methodology developed in Chapter III, we are able not only to reproduce (1.10) *in a Black-Scholes model, but also in a Heston model with $d = 2$ hedging instruments. For this case (or any other model with $d > 1$) neither theoretical nor numerical results on* (1.10) *have been available previously.*

Finally, we consider a setting with $d = 10$, built from 5 independent Heston models and show that the algorithm gives accurate approximations of hedging strategies also here. This demonstrates the feasibility also in higher-dimensional setups.

Let us conclude by pointing out the flexibility of the methodology: the algorithm takes as an input a transaction cost structure, a risk measure ρ and samples of the price process S and the payoff Z. Given these specifications, it calculates an approximate indifference price and a close-to-optimal hedging strategy.

Chapter IV: Skorokhod Embedding Problem for Lévy Processes

In this chapter we provide a solution to the Skorokhod embedding problem for Lévy processes with non-deterministic initial value. The chapter is based on [DGPR17].

The Skorokhod embedding problem (SEP) is formulated as follows: given two probability distributions μ_0, μ_1 on $\mathbb{R}$ and a Lévy process L with $L_0 \sim \mu_0$ under $\mathbb{P}$, find an $\mathbb{F}$-stopping time such that $L_\tau \sim \mu_1$ and $\mathbb{E}[\tau] < \infty$. $\mathbb{F}$ denotes the natural augmented filtration of L.

This problem is classical [Sko61, Sko65] in the special case when L is a Brownian motion with $\mu_0 = \delta_0$ and a variety of solutions exist. For background we refer to the survey [Obł04], the recent systematic solution methodology [BCH17b] and the survey [Hob11]. The latter explains in detail the connection of (SEP) to robust finance and provides a list of references in this direction. Studying the problem also for $\mu_0 \neq \delta_0$ is motivated by the interest in multi-marginal Skorokhod embeddings, robust bounds for forward start options and martingale optimal transport, see e.g. [BHR01a], [HN12], [BHLP13], [COT15], [OS17], [BCH17a] and the references therein. For more general (non-continuous) Markov processes, there have only been few results in the literature. Solutions are constructed in [Ros71], [Mon72], [FF91], [BL92], [OP09], but these are either specific to the case $\mu_0 = \delta_0$, the construction only covers transient processes or the stopping times are non-explicit and possibly randomized (i.e. they are stopping times with respect to a larger filtration). In particular, in Chapter IV we provide the first solution for (SEP) with $\mu_0 \neq \delta_0$ that is non-randomized and covers all Lévy processes. This generality comes at the price of some regularity assumptions on the measures μ_0 and μ_1, but the solution is very natural, it is explicit and it appears to be new also in the case of Brownian motion.

Before we state the result in detail, let us provide a summary of the approach. The basic idea is to exploit the connection between stopping-times, random time-changes and Fokker-Planck equations. For $\sigma\colon [0,1] \times \mathbb{R} \to [0,\infty)$ sufficiently regular and bounded, there exists a unique solution to the time-change equation

$$X_t = L_{\delta(t)}, \quad \delta(t) = \int_0^t \sigma(s, L_{\delta(s)})\,\mathrm{d}s \text{ for } t \in [0,1], \tag{1.11}$$

and for any $t \in [0,1]$, $\delta(t)$ is an $\mathbb{F}$-stopping time. Thus, if one finds a choice of σ yielding $X_1 \sim \mu_1$, then one automatically obtains a solution to (SEP) by setting

$\tau := \delta(1)$. Since the marginal distributions of X in (1.11) satisfy the Fokker-Planck equation, one obtains a link between these marginals and σ: supposing that the distribution of X_t admits a density $\phi(t, \cdot)$ with respect to the Lebesgue measure for all $t \in [0, 1]$ and formally inverting the Fokker-Planck equation, one obtains

$$\sigma(t, x) = \frac{((\mathcal{A}^*)^{-1} \partial_t \phi(t, \cdot))(x)}{\phi(t, x)}, \quad (t, x) \in [0, 1] \times \mathbb{R}, \tag{1.12}$$

where $\mathcal{A}^*$ is the (for now formal) adjoint of the infinitesimal generator of L. This can be seen as the general version of the Dupire-formula for calibrating local volatility (or local Lévy) models, see [Dup94], [CGMY04]. Assuming that μ_0 and μ_1 admit Lebesgue-densities h_0 and h_1, our approach is now to *choose* a family of densities $(\phi(t, \cdot))_{t \in [0,1]}$ with $\phi(i, \cdot) = h_i$ (for $i = 0, 1$), define σ by (1.12) and verify that this choice of σ satisfies the regularity and boundedness assumptions guaranteeing a unique solution to (1.11) with $X_1 \sim \mu_1$. By the reasoning given above, this yields a solution to (SEP).

While the approach may appear simple at first sight, the key difficulty is that the properties (non-negativity, regularity and boundedness) of σ depend heavily on the choice of interpolation $(\phi(t, \cdot))_{t \in [0,1]}$. A choice of interpolation that, surprisingly, turns out to work for all Lévy processes is the *linear interpolation* $\phi(t, \cdot) = t h_1 + (1 - t) h_0$. With this choice of interpolation the formal expression for σ in (1.12) can be written as $\sigma = H/\phi$, where H is the solution to the Poisson equation $\mathcal{A}^* H = h_1 - h_0$. In particular, σ is non-negative if and only if H is non-negative. This last condition is in fact necessary for (SEP) to admit a solution, as we also show in Chapter IV.

Finally, since H may admit zeros and σ may be unbounded, the standard time-change results available in the literature do not apply and we have to extend these to our setting. Since these results are of independent interest, they are developed in the general framework of martingale problems [EK86] in a separate appendix.

Having outlined the key ideas, let us now state the main result of Chapter IV in detail. We assume that μ_0 and μ_1 have strictly positive densities h_0 and h_1 with respect to the Lebesgue measure. We distinguish three classes of Lévy processes and impose different regularity assumptions on h_0 and h_1 for each of them, e.g. if L is a symmetric α-stable Lévy process with $\alpha \in (1, 2]$ (which includes Brownian motion for $\alpha = 2$) we assume $h_0, h_1 \in C_0(\mathbb{R})$. Denoting by η the characteristic exponent of the Lévy process L, the main result of Chapter IV is as follows:

Firstly, a solution to (SEP) exists if and only if

$$\frac{\widehat{\mu_1} - \widehat{\mu_0}}{\eta} \in L^1(\mathbb{R}), \quad H \geq 0 \quad \textit{and} \quad H \in L^1(\mathbb{R}), \tag{1.13}$$

where $\widehat{\mu_i}$ is the characteristic function of μ_i and

$$H(x) := \frac{1}{2\pi} \int_{\mathbb{R}} \frac{\widehat{\mu_1}(\xi) - \widehat{\mu_0}(\xi)}{\eta(\xi)} e^{-ix\xi} \, \mathrm{d}\xi, \quad x \in \mathbb{R}. \tag{1.14}$$

Secondly, if (1.13) *is satisfied, then a solution to (SEP) is given by*

$$\tau := \inf \left\{ t \in [0, \rho) : \int_0^t e^{-G(r)} \frac{h_1(L_r)}{H(L_r)} \, \mathrm{d}r \geq 1 \right\} \wedge \rho, \tag{1.15}$$

where, for $t \geq 0$*,*

$$\rho := \inf\{t \in [0,\infty) \,:\, H(L_t) = 0\} \quad \textit{and} \quad G(t) := \int_0^t \frac{h_1(L_r) - h_0(L_r)}{H(L_r)}\,\mathrm{d}r$$

with the usual convention $\inf \emptyset := \infty$*. Finally,* $\mathbb{E}^{\mu_0}[\tau] = \int_{\mathbb{R}} H(x)\,\mathrm{d}x$*.*

Let us conclude by briefly explaining why conditions (1.13) are necessary. To see this, one assumes that a solution τ to (SEP) exists and applies the Riesz representation theorem to find a measure ν on $\mathbb{R}$ such that

$$\mathbb{E}\left[\int_0^\tau g(L_s)\,\mathrm{d}s\right] = \int_{\mathbb{R}} g(x)\nu(\mathrm{d}x), \quad g \in C_c(\mathbb{R}). \tag{1.16}$$

On the other hand, from Dynkin's lemma and properties of the Lévy process one obtains

$$\eta(u)\mathbb{E}\left[\int_0^\tau e^{iuL_s}\,\mathrm{d}s\right] = \widehat{\mu_1}(u) - \widehat{\mu_0}(u), \quad u \in \mathbb{R}. \tag{1.17}$$

By extending (1.16) to $g(x) := e^{iux}$ and using (1.17), one may deduce $\nu(\mathrm{d}x) = H(x)\mathrm{d}x$ and (1.13).

Chapter II

Affine Filtering

1 Introduction

Consider a time-dependent signal which can not be observed directly, but only through noisy measurements. Given the stream of observations made up until today, what can you say about the signal? For example, what is the best estimate for the signal today? There are various mathematical formulations of this problem. Research fields such as time-series analysis, signal processing and (frequentist) non-parametric statistics model the signal process as a deterministic function or focus on a discrete-time setting. In *stochastic filtering* the signal and observation processes are modeled as continuous-time stochastic processes, i.e. a Bayesian perspective is adopted. The classical mathematical formulation of the above problem in this context is the following: consider a D-valued stochastic process X, a p-dimensional Brownian motion W and an observation function $h\colon D \to \mathbb{R}^p$. Define Y as

$$Y_t = \int_0^t h(X_s)\,\mathrm{d}s + W_t, \quad t \geq 0. \tag{1.1}$$

Note that both X and W are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $p \in \mathbb{N}$, D is some set and some regularity on h and the sample paths of X needs to be imposed to make (1.1) well-defined. Here X models the signal and Y the observation process. The *filtering problem* is to calculate π_t, the conditional distribution of X_t given $\mathcal{F}_t^Y := \sigma(Y_s\colon s \in [0,t])$, i.e. the observations up to time t, for each $t \geq 0$.

Starting in the mid-twentieth century, stochastic filtering has received an enormous amount of attention and has influenced many fields of mathematics - we refer to the introductory textbooks [LS01], [BC09a] the historical overview in [Cri14] and the handbook [CR11]. Theoretically the filtering problem has been solved: $(\pi_t)_{t\geq 0}$ can be characterized as the unique solution to a measure-valued stochastic differential equation (the Fujisaki-Kallianpur-Kunita or Kushner-Stratonovich equation). For applications in mathematical finance [BH98] or geophysics [LSZ15] also a numerical calculation of π_t is essential - in fact for any application of a continuous-time stochastic model that features latent factors. It has been shown that apart from a few special cases, e.g. when h is affine and X is an Ornstein-Uhlenbeck process or when the state space D consists of finitely many points, the equation for $(\pi_t)_{t\geq 0}$ is truly infinite-dimensional. As a consequence, devising numerical methods to calculate π_t or even just the conditional mean $\mathbb{E}[X_t|\mathcal{F}_t^Y]$ is very challenging. In most cases it is in-feasible due to computational constraints. Standard numerical methods ([BC09a, Chapters 8-10]) either only work for low-dimensional state spaces or for approximately Gaussian setups.[1] However, post-crisis financial modeling asks for factor processes X which are both high-dimensional and not approximately Gaussian. The lack of numerical filtering methods for such processes has put serious limitations on the modeling flexibility: one has not been able to include latent factors in them.

In the present chapter, we fill this gap and show that the narrow class of processes for which an efficient numerical solution is possible (see above) also includes *affine processes*. More precisely, we consider the case when h is affine and the signal process X is an *affine process* with state space $D = \mathbb{R}_+^m \times \mathbb{R}^{d-m}$ as characterized in [DFS03]. This class of processes includes for example Lévy processes, Cox-Ingersoll-Ross processes [CIR85] or the Heston model [Hes93] and is very widely used in financial applications (see e.g. [DFS03], [KRM15, Section 3] for a list of references). The filtering problem arises naturally in this context; for example, X could model the short rate and Y the observed yields of bond prices as in [GP99], [CS03], see also [BH98].

Let us briefly summarize the key ideas of our approach. As a first step the filtering distribution is rewritten in terms of the pathwise filtering functional as studied by [Dav80], [Cla78]. Although the functional itself is not tractable, it can be approximated by a linearized version thereof. This new *linearized filtering functional* (LFF) is numerically tractable, since the Fourier coefficients can be calculated by solving a system of generalized Riccati equations with vector fields depending on the observation Y. This gives rise to Fourier filtering techniques, analogously to the Fourier pricing techniques used for affine (log-price) models, see e.g. [CM99] and [DFS03]. In addition the (approximate) conditional moments can be calculated by solving a system of ordinary differential equations. In contrast to existing numerical methods (e.g. a particle filter), this is very well-suited to parallel computations and thus promising for high-dimensional filtering.

All of this is explained in detail in Section 3, while Section 2 provides background on affine processes and the filtering problem. The proofs of the statements on the LFF as well as local existence and uniqueness of solutions to the Riccati equations in Section 3 are then given in Section 4. This theoretical analysis is complemented by a numerical study in Sections 5 and 6, where the methodology is applied to the problem of filtering Cox-Ingersoll-Ross (CIR) and Wishart processes.

[1] In high-dimensional geophysical applications for example, only approximate Gaussian filters are routinely used (see the preface of [LSZ15]).

1.1 Notation

Fix a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which all random variables are defined.

Fix $p \in \mathbb{N}$, $m \in \mathbb{N} \cup \{0\}$, $d \in \mathbb{N}$ with $d \geq m$ and set $D = \mathbb{R}_+^m \times \mathbb{R}^{d-m}$. Let $\langle \cdot, \cdot \rangle$ denote the standard inner product on $\mathbb{R}^d$ and $|\cdot|$ the associated norm. Also write $\langle \cdot, \cdot \rangle$ for the linear extension of the inner product to $\mathbb{R}^d + i\mathbb{R}^d$, but without complex conjugation. Set

$$I := \{1, \dots, m\}, \quad J := \{m+1, \dots, d\}.$$

For $k \in \mathbb{N}$, write

$$\mathbb{C}_-^k = \{u \in \mathbb{C}^k : \operatorname{Re} u_i \leq 0, \forall i\}, \quad \mathbb{C}_{--}^k = \{u \in \mathbb{C}^k : \operatorname{Re} u_i < 0, \forall i\}$$

and define $\mathcal{U} = \mathbb{C}_-^m \times i\mathbb{R}^n$.

Denote by $B(D)$ and $C_b(D)$ the sets of bounded measurable functions and bounded continuous functions on D and by $\mathcal{P}(D)$ the set of probability measures on D. As usually, $\mathcal{P}(D)$ is equipped with the topology of weak convergence. Let $\mathcal{M}^+(D)$ denote the set of finite measures on the Borel σ-algebra $\mathcal{B}(D)$. Given $\mu \in \mathcal{M}^+(D)$ and a measurable, μ-integrable function f on D, write $\mu f := \int_D f(x)\mu(\mathrm{d}x)$.

Fix a continuous truncation function $\chi\colon \mathbb{R}^d \to [-1, 1]^d$ with $\chi(\xi) = \xi$ in a neighborhood of 0 and bounded away from 0 outside that neighborhood. In fact, in order to be able to rely on a result from [KMK10] for $k = 1, \dots d$ we choose

$$\chi_k(x) = \begin{cases} 0 & \text{if } x_k = 0, \\ (1 \wedge |x_k|)\frac{x_k}{|x_k|} & \text{otherwise.} \end{cases}$$

Let $\pi_0 \in \mathcal{P}(D)$, $D(\mathcal{L}) \subset C_b(D)$ and $\mathcal{L}\colon D(\mathcal{L}) \to C_b(D)$ linear. Recall that a D-valued stochastic process $(X_t)_{t\geq 0}$ defined on some probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ is called a *solution to the martingale problem for* $(D(\mathcal{L}), \mathcal{L}, \pi_0)$, if $\tilde{\mathbb{P}} \circ X_0^{-1} = \pi_0$ and for each $h \in D(\mathcal{L})$, the process

$$h(X_t) - h(X_0) - \int_0^t \mathcal{L}h(X_s)\,\mathrm{d}s, \quad t \geq 0,$$

is a martingale (in its own filtration). The martingale problem for $(D(\mathcal{L}), \mathcal{L}, \pi_0)$ is said to be *well-posed* if there exists a solution and any two solutions have the same finite-dimensional marginal distributions.

2 Background: Affine processes and the filtering problem

2.1 Affine processes

2.1.1 Definition and characterization

Let us review the definition of an affine process and some consequences thereof. We refer to [DFS03], [KRST11] and [CT13] for further details and references.

Consider a D-valued time-homogeneous Markov process $((X_t)_{t\geq 0}, (\mathbb{P}_x)_{x\in D})$ defined on $(\Omega, \mathcal{F})$, see [RW00a, Chapter III]. Denote by $(P_t)_{t\geq 0}$ the associated semigroup on $B(D)$ and assume $P_t 1 = 1$ for all $t \geq 0$ (i.e. the process is conservative). $((X_t)_{t\geq 0}, (\mathbb{P}_x)_{x\in D})$ is called affine, if it is stochastically continuous, X has RCLL-paths ($\mathbb{P}_x$-a.s. for any $x \in D$) and there exist functions $\phi : \mathbb{R}_{\geq 0} \times \mathcal{U} \to \mathbb{C}$ and $\psi : \mathbb{R}_{\geq 0} \times \mathcal{U} \to \mathbb{C}^d$ such that for all $x \in D$, $(t, u) \in \mathbb{R}_{\geq 0} \times \mathcal{U}$:

$$\mathbb{E}_x[e^{\langle X_t, u\rangle}] = \exp(\phi(t, u) + \langle x, \psi(t, u)\rangle). \tag{2.1}$$

Remark 2.1. As shown in [KRST11] this definition implies that for all $u \in \mathcal{U}$,

$$F(u) := \frac{\partial \phi}{\partial t}(t, u)\bigg|_{t=0+}, \quad R(u) := \frac{\partial \psi}{\partial t}(t, u)\bigg|_{t=0+} \tag{2.2}$$

exist and are continuous at $u = 0$. Thus, in the terminology of [DFS03] we are considering a conservative, regular affine process.

Remark 2.2. Alternatively, we could only assume that $((X_t)_{t\geq 0}, (\mathbb{P}_x)_{x\in D})$ is conservative, stochastically continuous and (2.1) holds for $(t, u) \in \mathbb{R}_{\geq 0} \times i\mathbb{R}^d$. Then [KRST11] implies that it is a Feller process and in particular, we may choose an RCLL version of X on D (under $\mathbb{P}_x$, for any $x \in D$).[2] Finally [DFS03, Theorem 2.7] implies that (2.1) can be extended to $\mathbb{R}_{\geq 0} \times \mathcal{U}$.

Let us now review some key properties of affine processes. To formulate these, an additional definition is required: a collection of parameters

$$(a, \alpha, b, \beta, c, \gamma, \mu^0, \mu) \tag{2.3}$$

is called admissible, if it satisfies the following (admissibility) conditions:

$$a \in \mathrm{Sem}^d \text{ with } a_{i,j} = 0 \text{ for } \quad i, j \in I \tag{2.4}$$

$$\alpha = (\alpha^1, \dots, \alpha^m) \text{ with } \alpha^i \in \mathrm{Sem}^d \text{ and } \alpha^i_{k,j} = 0 \text{ for } \quad k, j \in I \setminus \{i\} \tag{2.5}$$

$$b \in \mathbb{R}^d \text{ with } b_i - \int_{D\setminus\{0\}} \chi_i(\xi)\mu^0(\mathrm{d}\xi) \geq 0 \text{ for } \quad i \in I \tag{2.6}$$

$$\beta \in \mathbb{R}^{d\times d} \text{ with } \beta_{i,j} - \int_{D\setminus\{0\}} \chi_i(\xi)\mu^j(\mathrm{d}\xi) \geq 0 \text{ for } \quad i, j \in I \text{ and } i \neq j \tag{2.7}$$

$$\beta_{i,k} = 0 \text{ for } \quad i \in I,\, k \in J \tag{2.8}$$

$$c \in \mathbb{R}_+ \tag{2.9}$$

$$\gamma \in \mathbb{R}^m_+ \tag{2.10}$$

$$\mu = (\mu^1, \dots, \mu^m) \text{ and for } i \in I \cup \{0\}, \mu^i \text{ is a Borel measure on } D \setminus \{0\} \tag{2.11}$$

$$\int_{D\setminus\{0\}} \chi_k(\xi)\mu^i(\mathrm{d}\xi) < \infty \text{ for } \quad i \in I \cup \{0\},\, k \in I \setminus \{i\} \tag{2.12}$$

$$\int_{D\setminus\{0\}} \chi_k(\xi)^2\mu^i(\mathrm{d}\xi) < \infty \text{ for } \quad i \in I \cup \{0\},\, k \in (J \cup \{i\}) \setminus \{0\} \tag{2.13}$$

[2] Since the process is conservative, there is no need to consider the one-point compactification of D.

Remark 2.3. The admissibility conditions are identical to those in [DFS03, Definition 2.6]. We have only changed notation slightly in order to match the semimartingale notation in [KMK10]. The measure m in [DFS03, Definition 2.6] is denoted μ^0 here, the truncation function is arbitrary (as in [Fil05]) and we denote by b, β the parameters $\tilde{b}, \tilde{\beta}$ from [DFS03, Theorem 2.12]. Our conditions (2.6), (2.7) for these are equivalent to conditions (2.6) and (2.7) in [DFS03, Definition 2.6]. This leads to different expressions below for (2.14) and (2.16) than in [DFS03], see also [DFS03, Remark 2.13].

Suppose $((X_t)_{t\geq 0}, (\mathbb{P}_x)_{x\in D})$ is an affine process and denote again by $(P_t)_{t\geq 0}$ the restriction of the associated semigroup to $C_0(D)$. Then (see [DFS03, Theorem 2.7, Theorem 2.12 and Proposition 9.1]) there exists a collection of admissible parameters[3] (2.3) with $c = 0$ and $\gamma = 0$ such that the following properties hold:

- F and R in (2.2) are given as

$$\begin{aligned} F(u) &= \frac{1}{2}\langle u, au\rangle + \langle b, u\rangle + \int_{D\setminus\{0\}} \left(e^{\langle \xi, u\rangle} - 1 - \langle \chi(\xi), u\rangle\right) \mu^0(\mathrm{d}\xi) \\ R_i(u) &= \frac{1}{2}\langle u, \alpha^i u\rangle + \langle \beta^i, u\rangle + \int_{D\setminus\{0\}} \left(e^{\langle \xi, u\rangle} - 1 - \langle \chi(\xi), u\rangle\right) \mu^i(\mathrm{d}\xi) \end{aligned} \tag{2.14}$$

  for $i = 1, \dots, m$ and $R_i(u) = \langle \beta^i, u\rangle$ for $i = m+1, \dots, d$. Here $\beta^i \in \mathbb{R}^d$ is defined via

$$\beta^i_j := \beta_{j,i}, \quad \text{for} \quad 1 \leq i, j \leq d.$$

- ϕ and ψ solve the generalized Riccati equations

$$\begin{aligned} \partial_t \phi(t,u) &= F(\psi(t,u)), \quad \phi(0,u) = 0 \\ \partial_t \psi(t,u) &= R(\psi(t,u)), \quad \psi(0,u) = u \end{aligned} \tag{2.15}$$

  for $t \geq 0, u \in \mathcal{U}$.

- $(P_t)_{t\geq 0}$ is a Feller semigroup (in the sense of [RY99, Chapter III]). Denote by $(D(\mathcal{A}), \mathcal{A})$ its infinitesimal generator. Then $C_c^\infty(D)$ is a core for $\mathcal{A}$, $C_0^2(D) \subset D(\mathcal{A})$ and for any $f \in C_0^2(D)$, $x \in D$,

$$\begin{aligned} \mathcal{A}f(x) = \frac{1}{2}\sum_{k,l=1}^{d} \alpha_{kl}(x)\frac{\partial^2 f(x)}{\partial x_k \partial x_l} + \langle \beta(x), \nabla f(x)\rangle \\ + \int_{D\setminus\{0\}} (f(x+\xi) - f(x) - \langle \chi(\xi), \nabla f(x)\rangle) K(x, \mathrm{d}\xi) \end{aligned} \tag{2.16}$$

[3] Recall that we only consider conservative affine processes here.

where

$$\begin{aligned} \alpha(x) &= a + \sum_{i=1}^{m} \alpha^i x_i \\ \beta(x) &= b + \sum_{i=1}^{d} \beta^i x_i \\ K(x, \mathrm{d}\xi) &= \mu^0(\mathrm{d}\xi) + \sum_{i=1}^{m} x_i \mu^i(\mathrm{d}\xi). \end{aligned} \tag{2.17}$$

- X is a semimartingale (under $\mathbb{P}_x$, for any $x \in D$) admitting characteristics (B, C, ν) with respect to χ given by

$$B_t = \int_0^t \beta(X_s)\,\mathrm{d}s, \quad C_t = \int_0^t \alpha(X_s)\,\mathrm{d}s, \quad \nu(\mathrm{d}t, \mathrm{d}\xi) = K(X_t, \mathrm{d}\xi)\,\mathrm{d}t, \tag{2.18}$$

where α, β, K are as in (2.17).

Finally, let us put (conservative) affine processes into the framework of [EK86]. This is the purpose of Lemma 2.4 below. It is very close to [DFS03, Lemma 10.2], but considers arbitrary initial laws and establishes uniqueness also within the class of solutions to the martingale problem which are not necessarily RCLL. This extension is required to establish uniqueness for evolution equations (as the Zakai equation in Theorem 2.9 below) associated to $\mathcal{A}$.

Lemma 2.4. *Fix a collection of admissible parameters* (2.3) *and define* $\mathcal{A}_0$ *as the restriction of* $\mathcal{A}$ *(see* (2.16)*) to* $C_c^\infty(D)$. *Then for any* $\pi_0 \in \mathcal{P}(D)$, *the martingale problem for* $(C_c^\infty(D), \mathcal{A}_0, \pi_0)$ *is well-posed and the solution has RCLL-sample paths.*

Proof. The statement of [DFS03, Theorem 2.7] that X is a Feller process means that $(P_t)_{t\geq 0}$ is a strongly continuous, positive contraction semigroup on $C_0(D)$ in the terminology of [EK86]. Furthermore, by [EK86, Chap.4, Cor. 2.8] and since X is conservative, $(D(\mathcal{A}), \mathcal{A})$ is conservative (in the terminology of [EK86]). Thus $(P_t)_{t\geq 0}$ is a Feller semigroup (on $C_0(D)$) also in the terminology of [EK86]. Set $D(\mathcal{A}_0) = C_c^\infty(D)$. By [DFS03, Theorem 2.7], $D(\mathcal{A}_0)$ is a core for $(D(\mathcal{A}), \mathcal{A})$ and so the closure of the operator $(D(\mathcal{A}_0), \mathcal{A}_0)$ is again $(D(\mathcal{A}), \mathcal{A})$. Combining [EK86, Chap.4, Thm. 2.2, 2.7 and 4.1] then yields the statement. □

In view of Lemma 2.4 the following terminology is sensible: Fix $\pi_0 \in \mathcal{P}(D)$. We call an RCLL stochastic process X on $(\Omega, \mathcal{F}, \mathbb{P})$ an affine process started from π_0, if it is a solution to the martingale problem for $(C_c^\infty(D), \mathcal{A}_0, \pi_0)$. By Lemma 2.4 this uniquely determines the law of X under $\mathbb{P}$.

2.1.2 Exponential moments of affine processes

For the analysis of this chapter, it will be necessary to extend (2.1) to $U_0 \subset \mathbb{R}_{\geq 0} \times \mathbb{C}^d$, where U_0 is open and $0 \in U_0$. This means that an assumption on exponential moments is required. Suppose that

$$\int_{D\setminus\{|z|\le 1\}} |z| e^{\langle z,u\rangle} \mu^i(\mathrm{d}z) < \infty \quad \text{for all } i = 0,\dots,m \text{ and } u \in \mathbb{R}^d. \tag{2.19}$$

Suppose $((X_t)_{t\ge 0}, (\mathbb{P}_x)_{x\in D})$ is an affine process and define

$$E = \{(t,u) \in \mathbb{R}_{\ge 0} \times \mathbb{R}^d \ : \ \mathbb{E}_x[e^{\langle X_t,u\rangle}] < \infty \ \text{for all} \ x \in D\}. \tag{2.20}$$

By definition, this is the maximal domain on which the left hand side of (2.1) is finite. Under assumption (2.19), E is open, $0 \in E$ and ϕ and ψ can be extended to E. This is summarized in the next Lemma, which directly follows from [KRM15] and [FM09]. See also [SV10] and further references in all these articles.

Lemma 2.5. *Suppose* (2.19) *holds. Then*

(i) E is open in $\mathbb{R}_{\ge 0} \times \mathbb{R}^d$,

(ii) for any $(T,u) \in \mathbb{R}_{\ge 0} \times \mathbb{C}^d$ with $(T, \operatorname{Re} u) \in E$, there exists a unique solution to (2.15) *on $[0,T]$ and* (2.1) *holds.*

Proof. By [DFS03, Lemma 5.3] and (2.19), F and R are analytic functions. Therefore the same reasoning as in the proof of [FM09, Lemma 2.3] shows that for any $u \in \mathbb{C}^d$, there exists $t_+(u) \in (0,\infty]$ such that (2.15) has a unique solution on $[0, t_+(u))$ and the set

$$D_{\mathbb{R}} := \{(t,y) \in \mathbb{R}_{\ge 0} \times \mathbb{R}^d \ : \ t < t_+(y)\}$$

is open in $\mathbb{R}_{\ge 0} \times \mathbb{R}^d$. Furthermore, by [KRM15, Theorem 2.14(b)], [KRM15, Theorem 2.17(b)] and (2.19), one has $D_{\mathbb{R}} \subset E$ and (2.1) holds for all $(t,u) \in D_{\mathbb{R}}$, $x \in D$. [KRM15, Theorem 2.14(a)] implies $E \subset D_{\mathbb{R}}$ and hence $E = D_{\mathbb{R}}$. This shows (i). $(T, \operatorname{Re} u) \in E$ yields $(T, \operatorname{Re} u) \in D_{\mathbb{R}}$ and so [KRM15, Theorem 2.26] implies (ii). ☐

A further consequence of (2.19) is the following:

Lemma 2.6. *Assume* (2.19)*. Then for any $T \ge 0$, $k \in \mathbb{N}$, $x \in D$*

$$\mathbb{E}_x[|X_T|^{2k}] < \infty, \tag{2.21}$$

$$\mathbb{E}_x\left[\int_0^T |X_t|^{2k} \,\mathrm{d}t\right] < \infty. \tag{2.22}$$

Proof. By [DFS03, Lemma 5.3] and (2.19), F and R are analytic functions on $\mathbb{C}^d$. Thus by [DFS03, Lemma 6.5(i)], ϕ and ψ are in $C^\infty(\mathbb{R}_+ \times \mathcal{U})$. Combining this with $i\mathbb{R}^d \subset \mathcal{U}$ and [DFS03, Theorem 2.16(i)] yields (2.21). By [DFS03, Lemma A.1], for any $t \in [0,T]$, $\mathbb{E}_x[|X_t|^{2k}]$ is a sum of partial derivatives (up to order k) of $\psi(t,\cdot)$ and $\phi(t,\cdot)$ at 0. But all of these are continuous (as argued above) and so $t \mapsto \mathbb{E}_x[|X_t|^{2k}]$ is bounded on $[0,T]$. Hence (2.22) follows. ☐

2.1.3 Time-inhomogeneous affine processes

As it turns out, linear filtering of an affine process gives rise to a time-inhomogeneous affine process. This class of time-inhomogeneous Markov processes has been studied in [Fil05]. Similar to the time-homogeneous case (as summarized in Section 2.1.1), [Fil05] has obtained characterizations in terms of a martingale problem or (for conservative processes) semimartingale characteristics. We do not repeat these here; for our purposes it is sufficient to understand the conditions on the parameters that are necessary and sufficient for the existence of such a process. For more details we refer to [Fil05].

A collection of parameters (depending on $t \geq 0$)

$$(a(t), \alpha(t), b(t), \beta(t), c(t), \gamma(t), \mu^0(t), \mu(t)) \tag{2.23}$$

is called admissible (or strongly admissible), if the following (admissibility) conditions are satisfied:

- for any $t \geq 0$, (2.23) satisfies conditions (2.4)-(2.13),
- $(a(t), \alpha(t), b(t), \beta(t), c(t), \gamma(t))$ are continuous in $t \in \mathbb{R}_+$,
- the measures $\chi_k(\cdot)\mu^i(t,\cdot)$ (on $D \setminus \{0\}$) are weakly continuous in $t \in \mathbb{R}_+$ for any $i \in I \cup \{0\}$, $k \in I \setminus \{i\}$,
- the measures $\chi_k(\cdot)^2\mu^i(t,\cdot)$ (on $D \setminus \{0\}$) are weakly continuous in $t \in \mathbb{R}_+$ for any $i \in I \cup \{0\}$, $k \in (J \cup \{i\}) \setminus \{0\}$.

Remark 2.7. As before, b and β here denote $\tilde{b}, \tilde{\beta}$ in [Fil05, Theorem 2.13]. Since χ_k is bounded and continuous, the third continuity condition guarantees that $\tilde{b}, \tilde{\beta}$ in [Fil05, Theorem 2.13] are continuous if and only if b and β in [Fil05, Definition 2.5] are continuous. Together with Remark 2.3 this implies that the present admissibility conditions are identical with [Fil05, Definition 2.5].

Remark 2.8. If $c(t) = 0$, $\gamma(t) = 0$ for all $t \geq 0$, then the admissibility condition here is equivalent to [KMK10, Definition 2.4].

By [Fil05, Theorem 2.13, Lemma 3.1 and Proposition 4.3] for any collection of parameters satisfying these conditions (and only under these), there exists a strongly regular time-inhomogeneous affine process $(\bar{X}, (\mathbb{P}_{(r,x)})_{(r,x)\in\mathbb{R}_+\times D})$ (a time-inhomogeneous, stochastically continuous Markov process with an additional regularity condition as (2.2), see [Fil05]) with transition function $(P_{t,T})$ satisfying for any $u \in \mathcal{U}$, $0 \leq t \leq T$,

$$P_{t,T}\exp(\langle u,\cdot\rangle)(x) = \exp(\Phi(t,T,u) + \langle x, \Psi(t,T,u)\rangle), \quad \forall x \in D, \tag{2.24}$$

where Φ and Ψ solve the generalized Riccati equations

$$\begin{aligned} -\partial_t\Phi(t,T,u) &= F(t,\Psi(t,T,u)), \quad \Phi(T,T,u) = 0 \\ \partial_t\Psi(t,T,u) &= R(t,\Psi(t,T,u)), \quad \Psi(T,T,u) = u, \quad 0 \leq t \leq T \end{aligned} \tag{2.25}$$

with vector fields

$$
\begin{aligned}
F(t,u) &= \frac{1}{2}\langle u, a(t)u\rangle + \langle b(t), u\rangle - c(t) + \int_{D\setminus\{0\}} \left(e^{\langle \xi, u\rangle} - 1 - \langle \chi(\xi), u\rangle\right) \mu^0(t, \mathrm{d}\xi) \\
R_i(t,u) &= \frac{1}{2}\langle u, \alpha^i(t)u\rangle + \langle \beta^i(t), u\rangle - \gamma_i(t) \\
&\quad + \int_{D\setminus\{0\}} \left(e^{\langle \xi, u\rangle} - 1 - \langle \chi(\xi), u\rangle\right) \mu^i(t, \mathrm{d}\xi), \quad i = 1, \dots, m, \\
R_i(t,u) &= \langle \beta^i(t), u\rangle, \quad i = m+1, \dots, d,
\end{aligned}
\tag{2.26}
$$

where $\beta^i_j(t) := \beta_{j,i}(t)$.

Finally, fix $(r,x) \in \mathbb{R}_+ \times D$. As noted in [Fil05] one may assume that $\bar{X}$ has RCLL paths, $\mathbb{P}_{(r,x)}$-a.s. and so the following terminology makes sense: suppose Y is a stochastic process on $(\Omega, \mathcal{F}, \mathbb{P})$ with RCLL paths. We will say that (under $\mathbb{P}$) Y is a time-inhomogeneous affine process started in (r,x) with admissible parameters (2.23), if the law of Y under $\mathbb{P}$ (on the space of RCLL-paths) is identical to the law of $\bar{X}$ under $\mathbb{P}_{(r,x)}$.

2.2 The filtering problem

2.2.1 Problem formulation and the Zakai equation

Fix $\pi_0 \in \mathcal{P}(D)$ and suppose X is an affine process started from π_0 (see Section 2.1.1) on $(\Omega, \mathcal{F}, \mathbb{P})$. Further, suppose $\mathbb{F}$ is a right-continuous filtration on $(\Omega, \mathcal{F}, \mathbb{P})$ with respect to which X is adapted and such that $\mathcal{F}_0$ contains all $\mathbb{P}$-nullsets.

Let us introduce the problem of filtering X given noisy observations Y, as in the standard setup, see [LS01] and [BC09a]. The exposition here follows [KO88].

Define Y as

$$
Y_t = \int_0^t h(X_s)\,\mathrm{d}s + W_t, \quad t \geq 0, \tag{2.27}
$$

where W is a p-dimensional $\mathbb{F}$-Brownian motion independent of X, $h\colon D \to \mathbb{R}^p$ is measurable and

$$
\mathbb{E}\left[\int_0^T |h(X_s)|^2\,\mathrm{d}s\right] < \infty, \tag{2.28}
$$

for all $T \geq 0$. Set

$$
\mathcal{F}^Y_t = \sigma(Y_s \ : \ 0 \leq s \leq t) \vee \mathcal{N}, \quad t \geq 0, \tag{2.29}
$$

where $\mathcal{N}$ denotes the collection of $\mathbb{P}$-nullsets of $(\Omega, \mathcal{F})$.

The goal of filtering theory is to calculate, for $t \in [0, \infty)$, the conditional distribution of X_t given $\mathcal{F}^Y_t$. Formally this is described by a measure-valued process as follows: by [BC09a, Theorem 2.1] there exists a $\mathcal{P}(D)$-valued $(\mathcal{F}^Y_t)_{t\geq 0}$-adapted, RCLL-process $(\pi_t)_{t\geq 0}$ such that for any $f \in B(D)$, $t \geq 0$,

$$
\pi_t f = \mathbb{E}[f(X_t)|\mathcal{F}^Y_t] \quad \mathbb{P}\text{-a.s.}
$$

It can be shown that π satisfies the *Kushner-Stratonovich equation*. This is a stochastic partial differential equation for the process π, usually written in weak form, i.e. applied to test functions $f \in D(\mathcal{A})$.

Alternatively, one may consider an $\mathcal{M}^+(D)$-valued (but not $\mathcal{P}(D)$-valued) process, which leads to the linear Duncan-Mortensen-Zakai equation or shortly *Zakai equation*: define

$$\sigma_t := \exp\left(\int_0^t (\pi_s h)^\top \,\mathrm{d}Y_s - \frac{1}{2}\int_0^t |\pi_s h|^2 \,\mathrm{d}s\right) \pi_t \tag{2.30}$$

which is nonzero $\mathbb{P}$-a.s., for any $t \geq 0$, because

$$\mathbb{E}\left[\int_0^t |\pi_s h|^2 \,\mathrm{d}s\right] < \infty, \tag{2.31}$$

as can be deduced from (2.28).

We are now concerned with the filtering problem on the time interval $[0, T]$, for some $T > 0$ fixed. By (2.28) and independence,

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \exp\left(-\int_0^T h(X_s)^\top \,\mathrm{d}W_s - \frac{1}{2}\int_0^T |h(X_s)|^2 \,\mathrm{d}s\right) \tag{2.32}$$

defines a new probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ that is equivalent to $\mathbb{P}$ on $\mathcal{F}_T$.[4] Furthermore, the law of X under $\mathbb{P}$ is the same as under $\mathbb{Q}$ and, on $[0, T]$ under the measure $\mathbb{Q}$, Y is a Brownian motion independent of X.

It can be shown (see [BC09a, Exercise 3.37]) that $\sigma_t 1$ defined in (2.30) is equal to $\mathbb{E}_\mathbb{Q}[\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}} | \mathcal{F}_t^Y]$. Combining this with the abstract Bayes' rule and the definition (2.30), one obtains (see [BC09a, Proposition 3.16]) that for any $t \in [0, T]$, $f \in B(D)$,

$$\sigma_t f = \mathbb{E}_\mathbb{Q}\left[f(X_t) \exp\left(\int_0^t h(X_s)^\top \,\mathrm{d}Y_s - \frac{1}{2}\int_0^t |h(X_s)|^2 \,\mathrm{d}s\right) \middle| \mathcal{F}_t^Y \right], \tag{2.33}$$

$\mathbb{P}$-a.s., and the *Kallianpur-Striebel* formula

$$\pi_t f = \frac{\sigma_t f}{\sigma_t 1}. \tag{2.34}$$

Furthermore, σ satisfies the *Zakai equation*

$$\sigma_t f = \pi_0 f + \int_0^t \sigma_s(\mathcal{A}f) \,\mathrm{d}s + \int_0^t \sigma_s(hf) \,\mathrm{d}Y_s \quad \text{for any } f \in D(\mathcal{A}). \tag{2.35}$$

By (2.31) and (2.30), h is σ_t-integrable for all $t \leq T$ and $\int_0^T |\sigma_t h|^2 \,\mathrm{d}s < \infty$, $\mathbb{P}$-a.s. Hence all terms in (2.35) are indeed well-defined.

[4] See [LS01, Example I.6.2.4]. Independence is crucial here, otherwise a Nivikov' type assumption would be needed.

2.2.2 Uniqueness for the Zakai equation

The following result is a consequence of [KO88, Theorem 4.2]:

Theorem 2.9 (Well-posedness of the Zakai equation)**.** *Let $h \in C(D)$, $\mathcal{A}$ the generator of a (conservative) affine process (see* (2.16)*) and σ as in* (2.35)*. Assume* (2.28)*.*

Suppose $(\rho_t)_{t\in[0,T]}$ is an $(\mathcal{F}_t^Y)_{t\in[0,T]}$-adapted RCLL $\mathcal{M}^+(D)$-valued process such that h is ρ_t-integrable for all $t \le T$, $\int_0^T |\rho_t h|^2 \,\mathrm{d}s < \infty$, $\mathbb{P}$-a.s. and satisfying

$$\rho_t f = \pi_0 f + \int_0^t \rho_s(\mathcal{A}f)\,\mathrm{d}s + \int_0^t \rho_s(hf)\,\mathrm{d}Y_s, \quad \text{for any } f \in C_c^\infty(D) \tag{2.36}$$

and for $f = 1$ (with $\mathcal{A}1 := 0$). Then $\rho_t = \sigma_t$ for all $t < T$, $\mathbb{P}$-a.s.

Proof. Define $D(\mathcal{A}_0) := C_c^\infty(D)$ and $\mathcal{A}_0$ the restriction of $\mathcal{A}$ to $D(\mathcal{A}_0)$. Then by Lemma 2.4, for any $\pi_0 \in \mathcal{P}(D)$, the martingale problem for $(D(\mathcal{A}_0), \mathcal{A}_0, \pi_0)$ is well-posed. Furthermore, for any $f \in D(\mathcal{A}_0)$, $h_i f \in C_0(D)$ for $i = 1, \dots, p$ and so the assumptions of [KO88, Theorem 4.2] are indeed satisfied. □

Since the assumptions for [KO88, Theorem 4.1] are the same as for [KO88, Theorem 4.2], as a corollary one also obtains a uniqueness result for the Kushner-Stratonovich equation.

Let us point out that Theorem 2.9 holds in the setting considered in Section 3. Taking $h(x) = x$, $\pi_0 = \delta_x$ for some $x \in D$ and assuming that the jump-measures of the affine process satisfy (2.19), one obtains from Lemma 2.6 that (2.28) is indeed satisfied.

2.2.3 Robust filtering

Thanks to the uniqueness result for the Zakai equation in Theorem 2.9, theoretically the filtering problem is settled: one finds a solution to the Zakai equation and uses the Kallianpur-Striebel formula (2.34) to calculate the filter. However, in practice one is given a fixed $y \in C([0,T], \mathbb{R}^p)$ (of *finite variation*), whereas (2.33) only specifies the filter $\mathbb{P}$-a.s. Thus a definition of (2.33) for *all* $y \in C([0,T], \mathbb{R}^p)$ is needed.

Let us briefly review the main result of [Dav80]. See [BC09a, Chapter 5] and [vH07, Section 1.4] for further references on robust filtering. Suppose $h \in D(\mathcal{A})$ so that $h(X)$ is a semimartingale. Since X and Y are independent, one can integrate by parts

$$\int_0^t h(X_s)^\top \,\mathrm{d}Y_s = Y_t^\top h(X_t) - \int_0^t Y_s^\top \,\mathrm{d}h(X_s) \tag{2.37}$$

and rewrite σ in (2.33) as

$$\sigma_t f = \mathbb{E}_\mathbb{Q}\left[f(X_t) \exp\left(Y_t^\top h(X_t) - \int_0^t Y_s^\top \,\mathrm{d}h(X_s) - \frac{1}{2}\int_0^t |h(X_s)|^2 \,\mathrm{d}s \right) \middle| \mathcal{F}_t^Y \right].$$

Recalling that X and Y are independent under $\mathbb{Q}$ and X has the same distribution under $\mathbb{P}$ as under $\mathbb{Q}$, the conditional expectation is actually given as $\mathbb{E}\left[F(X,y)\right]|_{y=Y}$ for a suitable function $F\colon D\times\mathbb{R}^p\to\mathbb{R}$. In fact, the following *robustness* property has been established in [Cla78], [CC05]: define the *pathwise filtering functional* $\sigma_t\colon B(D)\times C([0,T],\mathbb{R}^d)\to\mathbb{R}$ by

$$\sigma_t(f,y)=\mathbb{E}\left[f(X_t)\exp\left(y_t^\top h(X_t)-\int_0^t y_s^\top\,\mathrm{d}h(X_s)-\frac{1}{2}\int_0^t|h(X_s)|^2\,\mathrm{d}s\right)\right], \tag{2.38}$$

then $\sigma_t(f,\cdot)/\sigma_t(1,\cdot)$ is locally Lipschitz continuous and

$$\frac{\sigma_t(f,Y)}{\sigma_t(1,Y)}=\mathbb{E}[f(X_t)|\mathcal{F}_t^Y],\quad \mathbb{P}\text{-a.s.} \tag{2.39}$$

See also [CDFO13] for an extension to multidimensional observation and correlated noise.

In [Dav80] the following observation is made: fix $y\in C[0,T]$ and define a two-parameter semigroup of operators on $B(D)$ by

$$T_{s,t}^y f(x)=\mathbb{E}_x\left[f(X_{t-s})\exp\left(-\int_s^t y_u^\top\,\mathrm{d}h(X_{u-s})-\frac{1}{2}\int_s^t|h(X_{u-s})|^2\,\mathrm{d}u\right)\right], \tag{2.40}$$

for $t\geq s\geq 0$, $x\in D$. Then

$$\sigma_t(f,y)=\int_D T_{0,t}^y(e^{y(t)h}f)(x)\pi_0(\mathrm{d}x) \tag{2.41}$$

and, this is the main result of [Dav80], the (extended) generator $\mathcal{A}_t^y$ of the semigroup $T_{s,t}^y$ is given by

$$\mathcal{A}_t^y f=e^{y(t)h}(\mathcal{A}-\frac{1}{2}h^2)(e^{-y(t)h}f).$$

This is closely related to applying a Doss-Sussmann method (see e.g. [RW00a, Theorem 28.2]) to the Zakai equation, as explained in [Dav11].

3 The linearized filtering functional

In this section we introduce and study a computationally tractable approximation of the pathwise filtering functional (2.38) when both X and h are affine. Throughout this section X is an affine process on $(\Omega,\mathcal{F},\mathbb{P})$ started from $\pi_0\in\mathcal{P}(D)$ with admissible parameters (2.3) and F, R are as in (2.14). If $\pi_0=\delta_x$ for $x\in D$, we write $\mathbb{P}_x$ for $\mathbb{P}$.

3.1 Definition and main results

3.1.1 Definition of the approximate filter

Fix an observation $y \in C([0,\infty),\mathbb{R}^d)$ with $y(0) = 0$ and $\gamma \in C([0,\infty),\mathbb{R}^d)$, $c \in C([0,\infty),\mathbb{R})$. The *linearized filtering functional* (LFF) ρ is defined as

$$\rho_t(f,y) = \mathbb{E}\left[f(X_t)\exp\left(y_t^\top X_t - \int_0^t y_s^\top \,\mathrm{d}X_s - \int_0^t \gamma_s^\top X_s - c_s \,\mathrm{d}s\right)\right] \tag{3.1}$$

for any $t \geq 0$ and $f\colon D \to \mathbb{R}$ measurable such that the right hand side of (3.1) is well-defined (e.g. $f \geq 0$). If $\rho_t(1,y)$ is finite, define the approximate pathwise filter (the *affine functional filter* or AFF) by

$$\bar{\pi}_t(f,y) = \frac{\rho_t(f,y)}{\rho_t(1,y)}. \tag{3.2}$$

If $\pi_0 = \delta_x$ for $x \in D$, we write $\rho_t^x(f,y)$ for $\rho_t(f,y)$ and $\bar{\pi}_t^x(f,y)$ for $\bar{\pi}_t(f,y)$.

3.1.2 Heuristic motivation

The linearized filtering functional (3.1) is the same as the pathwise filtering functional (2.38) for $h(x) = x$, but with $\frac{1}{2}|x|^2$ approximated by the affine function $\gamma_s x + c_s$. The motivation for studying ρ_t is the following: if for some $t > 0$, $x_0 \in D$ and (small) $\varepsilon > 0$, $\mathbb{P}(\{X_s \in B_\varepsilon(x_0)\,\forall s \in [0,t]\})$ is almost 1, then (3.1) and (2.38) (with $\gamma_s = x_0$ and $c_s = \frac{x_0^2}{2}$) are very close. Consequently, (2.39) implies that also the approximate filter $\bar{\pi}_t(f,Y)$ should be close to $\pi_t(f)$.

3.1.3 Fourier filtering

The key point is that (3.1) is computationally tractable, since one can calculate the Fourier coefficients of (3.1) by solving a system of generalized Riccati equations:

Theorem 3.1. *Assume* (2.19) *holds. Let* $u \in \mathbb{C}^d$ *and* $T \in \mathbb{R}_+$. *Suppose* $\Phi \in C^1([0,T],\mathbb{R})$ *and* $\Psi \in C^1([0,T],\mathbb{R}^d)$ *solve*

$$\begin{aligned} -\partial_t\Phi(t,T,u) &= F(\Psi(t,T,u) - y_t) - c_t, \quad && \Phi(T,T,u) = 0 \\ -\partial_t\Psi(t,T,u) &= R(\Psi(t,T,u) - y_t) - \gamma_t, \quad && \Psi(T,T,u) = u + y_T, \quad 0 \leq t \leq T. \end{aligned} \tag{3.3}$$

Then for any $x \in D$, *the Fourier coefficient of* $\rho_T^x(\cdot,y)$ *is well-defined and given as*

$$\rho_T^x(\exp(\langle u,\cdot\rangle),y) = \exp(\Phi(0,T,u) + \langle x, \Psi(0,T,u)\rangle). \tag{3.4}$$

Furthermore, there exists $T_0 > 0$ *such that for all* $u \in i\mathbb{R}^d$ *and* $T \leq T_0$, *the system* (3.3) *has a unique solution on* $[0,T]$.

The proof of Theorem 3.1 is postponed to Section 4.3 below. Let us briefly discuss how to use Theorem 3.1 in practice, relate it to the literature and discuss its assumptions.

Remark 3.2. Suppose $f\colon D \to \mathbb{C}$ is given as

$$f(y) = \int_{\mathbb{R}^d} e^{i\langle v,y\rangle} \hat{f}(v)\,\mathrm{d}v, \quad y \in D$$

for some $\hat{f}\colon \mathbb{R}^d \to \mathbb{C}$ integrable. Then for any $T > 0$ small enough, by Theorem 3.1, definition (3.1) and Fubini's theorem

$$\rho_T^x(f,y) = \int_{\mathbb{R}^d} \rho_T^x(\exp(\langle iv,\cdot\rangle),y)\hat{f}(v)\,\mathrm{d}v = \int_{\mathbb{R}^d} e^{\Phi(0,T,iv)+\langle x,\Psi(0,T,iv)\rangle}\hat{f}(v)\,\mathrm{d}v.$$

This is analogous to the Fourier method used in option pricing in the framework of affine models.

Remark 3.3. Expressions of type (3.4) are called affine transform formulas in the literature, see e.g. [KRM15] and the references therein. Note that the present result is not covered in the literature, since the Riccati equations (3.3) are time-inhomogeneous and correspond to a non-conservative affine "process" for which the admissiblity conditions (2.9) and (2.10) are not necessarily satisfied.

Remark 3.4. In general, it does not hold that $\rho_T(1,y) < \infty$ for all $T > 0$ and so the statement of Theorem 3.1 really just holds up to a finite T_0 (depending on y). To see this, let $u \in \mathbb{R}^d$ and consider $y_s = us$, $c_s = \gamma_s = 0$ for all $s \geq 0$. Then the product rule (as in (2.37)) and $\dot{y}_s = u$ show

$$\rho_T^x(1,y) = \mathbb{E}_x\left[\exp\left(u^\top \int_0^T X_s\,\mathrm{d}s\right)\right]$$

which is not necessarily finite. For example, if $d = m = 1$ and X is a CIR process (see Section 5) with parameters $\beta > 0$, $b \geq 0$ and $\sigma > 0$, then for $u < \beta^2/(2\sigma^2)$ and T large enough (satisfying $\tanh(\gamma T/2) \geq \gamma/\beta$ with $\gamma = \sqrt{\beta^2 - 2\sigma^2 u}$) the expectation is not finite, see [FKR10] or [Duf01].

Finally, note that (2.19) could be weakened to the following assumption: there exists $V \subset \mathbb{R}^d$ open with $0 \in V$ such that (2.19) holds for $u \in V$ (instead of all $u \in \mathbb{R}^d$).

3.1.4 The smoothing distribution

Our approximation (3.1) and (3.2) also gives rise to an approximation of the smoothing distribution, i.e. the distribution of $X_{[0,t]}$ conditional on $\mathcal{F}_t^Y$.

Fix $t > 0$ and denote by $D[0,t]$ the set of RCLL-mappings $[0,t] \to D$. Consider $G\colon D[0,t] \to \mathbb{R}$ bounded, measurable[5] and, analogously to (3.1) and (3.2) define

$$\begin{aligned}\rho_t(G,y) &= \mathbb{E}\left[G(X_{[0,t]})\exp\left(y_t^\top X_t - \int_0^t y_s^\top\,\mathrm{d}X_s - \int_0^t \gamma_s^\top X_s - c_s\,\mathrm{d}s\right)\right] \\ \bar{\pi}_t(G,y) &= \frac{\rho_t(G,y)}{\rho_t(1,y)}\end{aligned} \tag{3.5}$$

[5]More precisely, for $s \in [0,T]$ define $Y_s\colon D[0,t] \to \mathbb{R}$ by $Y_s(\omega) := \omega(s)$ and equip $D[0,t]$ with the σ-algebra generated by $(Y_s)_{s\in[0,t]}$.

for any $t \geq 0$ such that $\rho_t(1, y) < \infty$. Then $\bar{\pi}_t(\cdot, y)$ is a probability measure on $D[0,t]$ and an approximation to the smoothing distribution. Again, if $\pi_0 = \delta_x$ for $x \in D$, we write $\rho_t^x(G, y)$ for $\rho_t(G, y)$ and $\bar{\pi}_t^x(G, y)$ for $\bar{\pi}_t(G, y)$.

The following result shows that $\bar{\pi}_t(\cdot, y)$ coincides with the the distribution on $D[0,t]$ of a time-inhomogeneous affine process. It will be used for the calculation of (approximate) conditional moments in Section 5 and 6 below. To formulate it, define

$$\widehat{\pi_0}(z) = \int_D e^{\langle x,z\rangle}\pi_0(\mathrm{d}x), \qquad \text{for } z \in D_{\pi_0} = \{z \in \mathbb{C}^d \ : \ |\exp(\langle \cdot, z\rangle)| \in L^1(D, \pi_0)\}.$$

Theorem 3.5. *Let $T_0 > 0$, $t \in (0, T_0]$ and $\Psi(\cdot, t, 0)$ as in Theorem 3.1. Suppose $\Psi(0, t, 0) \in D_{\pi_0}$. Then for any $G \in B(D[0,t])$,*

$$\bar{\pi}_t(G, y) = \frac{\int_D e^{\langle x, \Psi(0,t,0)\rangle} \mathbb{E}_{\mathbb{Q}_x^{y,t}}[G(X_{[0,t]})]\pi_0(\mathrm{d}x)}{\widehat{\pi_0}(\Psi(0,t,0))},$$

where under $\mathbb{Q}_x^{y,t}$, X is a time-inhomogeneous affine process started from $(0, x)$ with admissible parameters

$$(a(s), \alpha(s), b(s), \beta(s), 0, 0, \mu^0(s), \mu(s))_{s\geq 0}$$

defined by (4.1) below with $g(s) := \Psi(s \wedge t, t, 0) - y_{s\wedge t}$ for $s \geq 0$.

Remark 3.6. If $\pi_0 = \delta_x$ for $x \in D$, then $D_{\pi_0} = \mathbb{C}^d$, $\widehat{\pi_0}(z) = e^{\langle x,z\rangle}$ and so Theorem 3.5 implies

$$\bar{\pi}_t^x(G, y) = \mathbb{E}_{\mathbb{Q}_x^{y,t}}[G(X_{[0,t]})]. \tag{3.6}$$

Remark 3.7. As a simple example, consider a CIR process (see Section 5) started in $x > 0$. Theorem 3.5 implies that for $t \leq T_0$ the approximate smoothing distribution is given by (3.6). Under $\mathbb{Q}_x^{y,t}$ the process X is the unique solution to

$$dX_s = b + \beta X_s + u(s, X_s)\mathrm{d}s + \sigma\sqrt{X_s}\mathrm{d}B_s, \quad X_0 = x, \tag{3.7}$$

where $u(s, x) := \sigma^2(\Psi_{s\wedge t} - y_{s\wedge t})x$, $\Psi_s := \Psi(s, t, 0)$ solves (the second part of) (3.3) and B is a Brownian motion under $\mathbb{Q}_x^{y,t}$. Thus, the approximate smoothing distribution is the distribution (on path space) of a new process, which is obtained by inserting the additional drift term $u(s, X_s)$ in the original SDE (5.1).

From [vH07, Chapters 1.4.3 and 4.2] one obtains formally a representation analogous to (3.6) for the exact smoothing distribution, the only difference being the choice of u in (3.7). However, calculating the function u in this case requires solving a PDE. For the approximate filter u can be obtained by solving an ODE, which is an enormous reduction of complexity.

3.1.5 An alternative point of view

To clarify further the relation to [Dav80], let $x_0 \in \mathbb{R}^m_{++} \times \mathbb{R}^n$, $c_0 > 0$ and define $H(x) = (x_0)_I^\top x_I + c_0$ and

$$\bar{T}_t^y f(x) = \mathbb{E}_x \left[f(X_t) \exp\left(-\int_0^t y_u^\top \,\mathrm{d}X_u - \int_0^t H(X_u)\,\mathrm{d}u \right) \right], \tag{3.8}$$

so that, in analogy to (2.41) it holds that (with $\gamma^\top = ((x_0)_I^\top, 0))$ and $c = c_0$ in (3.1))

$$\rho_t^x(f, y) = \bar{T}_t^y(\exp(\langle y_t, \cdot\rangle)f)(x).$$

Thus we have approximated $T_{0,t}^y$ in (2.40) by $\bar{T}_t^y$. Now if $\beta^i = 0$ for $i \in J$, then (3.8) corresponds to a non-conservative, time-inhomogeneous affine process:

Proposition 3.8. *Assume* (2.19) *holds and* $\beta^i = 0$ *for* $i \in J$. *Then there exists* $c_0 > 0, T > 0$ *such that for all* $t \in [0, T]$, $\bar{T}_t^y$ *in* (3.8) *satisfies* $\bar{T}_t^y = P_{0,t}^y$, *where* $(P_{s,t}^y)$ *is the transition semigroup of a time-inhomogeneous affine process with (admissible) parameters* (2.23) *defined for all* $t \geq 0$ *by* (4.1) *below with* $g(t) = -y_{t\wedge T}$ *and*

$$\begin{aligned} c(t) &= c_0 - F(-y_t) \\ \gamma^i(t) &= x_0^i - R_i(-y_t), \quad i \in I. \end{aligned} \tag{3.9}$$

3.1.6 Discussion

Remark 3.9. The ordinary differential equation (3.3) is formulated backwards in time, which appears to lead to a non-recursive filter. This can easily be resolved and we now explain how a recursive procedure can be obtained: fix $T_0 > 0$ sufficiently small. Theorem 3.1 guarantees that for any $v \in \mathbb{R}^d$ and $T \leq T_0$ there exists a unique $u_0 \in \mathbb{C}^d$ such that the ODE

$$\begin{aligned} -\partial_t \bar{\Psi}(t, u_0) &= R(\bar{\Psi}(t, u_0) - y_t) - \gamma_t \\ \bar{\Psi}(0, u_0) &= u_0 \end{aligned} \tag{3.10}$$

has a unique solution on $[0, T]$ with $\bar{\Psi}(T, u_0) = iv + y_T$. More specifically, one chooses $u_0 := \Psi(0, T, iv)$ and $\bar{\Psi}(t, u_0) := \Psi(t, T, iv)$. This gives the following recursive procedure to calculate the approximate filter at time $T \leq T_0$:

- solve for all $u_0 \in \mathbb{C}^d$ (for which a solution exists) the ODE (3.10) up to time T.
- for $v \in \mathbb{R}^d$, find the unique solution $u_0 \in \mathbb{C}^d$ to $\bar{\Psi}(T, u_0) = iv + y_T$ and evaluate

$$\rho_T^x(\exp(\langle iv, \cdot\rangle), y) = \exp\left(\int_0^T F(\bar{\Psi}(s, u_0) - y_s) - c_s \, \mathrm{d}s + \langle x, u_0\rangle\right).$$

In order to calculate the approximate filter at time $\tilde{T} \in [T, T_0]$ one only needs to continue solving (3.10) on $[T, \tilde{T}]$ (and then repeat the second step for $v \in \mathbb{R}^d$), hence the procedure is indeed recursive.

Remark 3.10. Consider a p-dimensional Brownian motion W, $C \in \mathbb{R}^{p\times d}$, $\Gamma \in \mathbb{R}^{p\times p}$ invertible and an observation process given as

$$\bar{Y}_t = \int_0^t C X_s \, \mathrm{d}s + \Gamma W_t, \quad t \geq 0. \tag{3.11}$$

The present methodology also provides an approximation for this setup: Since $\bar{Y}$ and $Y = \Gamma^{-1}\bar{Y}$ generate the same filtration, the filtering distribution is given by (2.38) and (2.39) with $h(x) = \Gamma^{-1}Cx$. The pathwise functional $\sigma_t(f, y)$ in (2.38) is approximated naturally by $\rho_t(f, (\Gamma^{-1}C)^\top y))$ (see (3.1)) with $\gamma_s = (\Gamma^{-1}C)^\top \Gamma^{-1}Cx_0$ (corresponding to the linearization of h around x_0) and $x_0 = \int_D x\pi_0(\mathrm{d}x)$. We do not specify c here, since it cancels out in the normalization (3.2).

In fact, this choice of γ has been used in the examples in Section 5.

Remark 3.11. The choice of the functions γ and c is of course essential for how close ρ_t and $\bar{\pi}_t$ are to σ_t and π_t. In the examples we have always made the choice specified in the previous remark. Let us examine the approximation quality in this setting. Using the product rule (2.37), the definition of $\mathbb{Q}$ and applying the change of measure (2.32) in (3.1) yields $\mathbb{P}$-a.s.

$$\begin{aligned}\rho_t(f, Y) &= \mathbb{E}_{\mathbb{Q}}\left[f(X_t)\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}} \exp\left(\frac{1}{2}\int_0^t |h(X_s)|^2 \,\mathrm{d}s - \int_0^t \gamma_s^\top X_s - c_s \,\mathrm{d}s\right)\middle|\, \mathcal{F}_t^Y\right] \\ &= \mathbb{E}\left[f(X_t) \exp\left(\frac{1}{2}\int_0^t |h(X_s)|^2 \,\mathrm{d}s - \int_0^t \gamma_s^\top X_s - c_s \,\mathrm{d}s\right)\middle|\, \mathcal{F}_t^Y\right] \sigma_t 1\end{aligned}$$

and so (in the setting of the previous remark)

$$\bar{\pi}_t(f, Y) = \frac{\mathbb{E}\left[f(X_t)A_t \middle|\, \mathcal{F}_t^Y\right]}{\mathbb{E}\left[A_t \middle|\, \mathcal{F}_t^Y\right]}, \quad A_t = \exp\left(\frac{1}{2}\int_0^t |\Gamma^{-1}C(X_s - x_0)|^2 \,\mathrm{d}s\right).$$

This gives an indication about the approximation quality:

If (with high probability) $\log A_t$ is very small, then the approximation quality is good. This happens for example if $\Gamma = \varepsilon I$ for large $\varepsilon > 0$. If $\varepsilon > 0$ is very small on the other hand, then the approximation quality decreases. However, in this regime there is no need for filtering, since $\int_0^\cdot X_s \,\mathrm{d}s$ can be almost read off from (3.11). For intermediate values of ε this is more difficult to judge and from numerical experiments it appears that there is a range of ε for which the filtering problem is not easy, and nevertheless the approximation is not very good.

Remark 3.12. If the observations arrive only at discrete-time points (as opposed to the continuous-time setting considered here) a similar approximation can be defined. In this case the ordinary differential equations (3.3) are replaced by difference equations.

4 Proofs

4.1 Proof of auxiliary results

In this section we prepare for the proof of the main results. To this end, we study a change of measure, estimates for the function R in (2.14) and properties of $\bar{T}^y$ in (3.8).

4.1.1 Change of measure

One of the key tools in the proofs is a change of measure, which turns the original (time-homogeneous) affine process into a time-inhomogeneous affine process. The next Lemma 4.1 verifies that the associated parameters satisfy the admissibility conditions. Based on this, Proposition 4.2 below will then provide the ingredients for the change of measure.

Lemma 4.1. *Suppose* $g\colon \mathbb{R}_+ \to \mathbb{R}^d$ *is continuous,* (2.3) *are admissible with* $c = 0$, $\gamma = 0$ *and* (2.19) *holds. For* $t \geq 0$, *define parameters* (2.23) *by* $c(t) = 0$, $\gamma(t) = 0$ *and*

$$\begin{aligned} a(t) &= a \\ \alpha(t) &= \alpha \\ b(t) &= b + ag_t + \int_{D\setminus\{0\}} \chi(\xi)(e^{\langle g_t,\xi\rangle} - 1)\mu^0(\mathrm{d}\xi) \\ \beta_{i,j}(t) &= \beta_{i,j} + (\alpha^j g_t)_i + \int_{D\setminus\{0\}} \chi_i(\xi)(e^{\langle g_t,\xi\rangle} - 1)\mu^j(\mathrm{d}\xi), \quad i \in I \cup J,\, j \in I \\ \beta_{i,j}(t) &= \beta_{i,j}, \quad i \in I \cup J,\, j \in J \\ \mu^i(t,\mathrm{d}\xi) &= e^{\langle g_t,\xi\rangle}\mu^i(\mathrm{d}\xi), \quad i \in I \cup \{0\}. \end{aligned} \tag{4.1}$$

Then (2.23) *is strongly admissible and for all* $T \geq 0$,

$$\sup_{t\in[0,T]} \int_{\{\xi_k>1\}} \xi_k \mu^i(t,\mathrm{d}\xi) < \infty, \qquad \textit{for} \quad i,k \in I. \tag{4.2}$$

Proof. **Admissibility for fixed $t \geq 0$:** Firstly, (2.4) implies $a_{i,j} = 0$ for all $i \in I, j \in I \cup J$ (see (2.4) in [DFS03]). Thus, for $i \in I$, definition (4.1), the assumed integrability (2.12) and the non-negativity condition (2.6) yield

$$b_i(t) - \int_{D\setminus\{0\}} \chi_i(\xi)\mu^0(t,\mathrm{d}\xi) = b_i - \int_{D\setminus\{0\}} \chi_i(\xi)\mu^0(\mathrm{d}\xi) \geq 0.$$

Similarly, for $i,j \in I$ with $i \neq j$, (2.5) implies $\alpha^j_{i,k} = 0$ for all $k \in I \cup J$. If this was not the case, i.e. if $\alpha^j_{i,k} \neq 0$ for some $k \in J \cup \{j\}$, then defining $v \in \mathbb{R}^d$ by $v_l = \delta_{lk}$ for $l \in J \cup \{j\}$, $v_l = C\delta_{li}$ for $l \in I \setminus \{j\}$ and using (2.5) would yield

$$0 \leq v^\top \alpha^j v = 2C\alpha^j_{i,k} + \alpha^j_{k,k}$$

for all $C \in \mathbb{R}$ and hence a contradiction. Consequently $(\alpha^j g_t)_i = 0$ and as above one uses (2.7) and (2.12) to obtain

$$\beta_{i,j}(t) - \int_{D\setminus\{0\}} \chi_i(\xi)\mu^j(t,\mathrm{d}\xi) = \beta_{i,j} - \int_{D\setminus\{0\}} \chi_i(\xi)\mu^j(\mathrm{d}\xi) \geq 0.$$

Finally, for $i \in I \cup \{0\}$ and any non-negative $f \in B(D)$ one uses $|e^{\langle g_t,\xi\rangle}| \leq e^{|g_t|}$ on $\{|\xi| \leq 1\}$ to estimate

$$\int_{D\setminus\{0\}} f(\xi)\mu^i(t,\mathrm{d}\xi) \leq e^{|g_t|}\int_{\{|\xi|\leq 1\}\setminus\{0\}} f(\xi)\mu^i(\mathrm{d}\xi) + \|f\|_\infty \int_{D\setminus\{|\xi|\leq 1\}} |\xi| e^{\langle g_t,\xi\rangle}\mu^i(\mathrm{d}\xi). \tag{4.3}$$

Inserting $f = \chi_k$ for $k \in I \setminus \{i\}$ and $f = \chi_k^2$ for $k \in (J \cup \{i\}) \setminus \{0\}$ in (4.3), the integrability conditions for $\mu^i(t, \cdot)$ follow from (2.12), (2.13) and (2.19).

Altogether, it has been verified that (2.23) satisfy for each $t \geq 0$ conditions (2.4)-(2.13).

Continuity in t: Let us first verify the third and fourth admissibility conditions. To do so, note that for any $f\colon D \setminus \{0\} \to \mathbb{R}$ which is μ^i-integrable, dominated convergence and continuity of g yield that

$$t \mapsto \int_{\{|\xi| \leq 1\} \setminus \{0\}} f(\xi) e^{\langle g_t, \xi \rangle} \mu^i(\mathrm{d}\xi) \quad \text{is continuous.} \tag{4.4}$$

Suppose the following is established: for any $f \in C_b(D)$,

$$t \mapsto \int_{D \setminus \{|\xi| \leq 1\}} f(\xi) e^{\langle g_t, \xi \rangle} \mu^i(\mathrm{d}\xi) \quad \text{is continuous.} \tag{4.5}$$

Then for $k \in I \setminus \{i\}$ and any $h \in C_b(D)$, one defines $f := \chi_k h$, notes that $f \in C_b(D)$ (since $\chi \in C_b(D)$) and μ^i-integrable by (2.12) and concludes that

$$t \mapsto \int_{D \setminus \{0\}} h(\xi) \chi_k(\xi) \mu^i(t, \mathrm{d}\xi) \quad \text{is continuous,}$$

by (4.4) and (4.5). Thus $\chi_k(\cdot)\mu^i(t, \cdot)$ is weakly continuous and the last strong admissibility condition follows analogously with $f := \chi_k^2 h$ and (2.13).

To verify (4.5), note that (2.19) and [DFS03, Lemma A.2] yield that the function $G_0\colon \mathbb{R}^d \to \mathbb{R}$ defined via

$$G_0(u) := \int_{D \setminus \{|\xi| \leq 1\}} f(\xi) e^{\langle u, \xi \rangle} \mu^i(\mathrm{d}\xi) \tag{4.6}$$

is analytic. In particular, composing it with the continuous function y preserves continuity and hence (4.5) holds.

Finally, it remains to argue that $b(\cdot)$ and $\beta(\cdot)$ are continuous. To show this, for any $i \in I \cup \{0\}$, $k \in I \cup J$ one uses $\mu^i(D \setminus \{|\xi| \leq 1\}) < \infty$ (since χ is bounded away from 0 on $D \setminus \{|\xi| \leq 1\}$ and by (2.12) and (2.13)) to decompose

$$\begin{aligned} &\int_{D \setminus \{0\}} \chi_k(\xi)(e^{\langle g_t, \xi \rangle} - 1) \mu^i(\mathrm{d}\xi) \\ &\quad = \int_{\{|\xi| \leq 1\} \setminus \{0\}} \chi_k(\xi)(e^{\langle g_t, \xi \rangle} - 1) \mu^i(\mathrm{d}\xi) + \int_{D \setminus \{|\xi| \leq 1\}} \chi_k(\xi) e^{\langle g_t, \xi \rangle} \mu^i(\mathrm{d}\xi) \\ &\qquad - \int_{D \setminus \{|\xi| \leq 1\}} \chi_k(\xi) \mu^i(\mathrm{d}\xi). \end{aligned} \tag{4.7}$$

The second term is continuous in t by (4.5) and so it remains to show that the first integral is continuous in t. But this follows from dominated convergence: for any $T > 0$ one may use Lipschitz continuity of exp, continuity of g, the Cauchy-Schwarz inequality and the properties of χ to find $C_0, C_1, C_2 > 0$ such that for all $t \in [0, T], \xi \in \{|\xi| \leq 1\} \setminus \{0\}$,

$$|\chi_k(\xi)(e^{\langle g_t, \xi \rangle} - 1)| \leq C_0 |\chi_k(\xi)| |\langle g_t, \xi \rangle| \leq C_1 |\chi(\xi)|^2 \frac{|\xi|}{|\chi(\xi)|} \leq C_2 |\chi(\xi)|^2.$$

But (2.12) and (2.13) imply $\int_{\{|\xi|\leq 1\}\setminus\{0\}} |\chi(\xi)|^2 \mu^i(\mathrm{d}\xi) < \infty$ and thus the claim.

Verification of (4.2)**:** Finally, again (2.19) and [DFS03, Lemma A.2] applied to the measure $|\xi|\mu^i(\mathrm{d}\xi)$ on $D \setminus \{|\xi| > 1\}$ (which is finite by (2.19)) shows that the function on $G\colon \mathbb{R}^d \to \mathbb{R}$ defined via

$$G(u) := \int_{D\setminus\{|\xi|\leq 1\}} |\xi| e^{\langle u,\xi\rangle} \mu^i(\mathrm{d}\xi)$$

is analytic and thus for $i, k \in I$,

$$\sup_{t\in[0,T]} \int_{\{\xi_k>1\}} \xi_k e^{g_t^\top \xi} \mu^i(\mathrm{d}\xi) \leq \sup_{t\in[0,T]} \int_{D\setminus\{|\xi|\leq 1\}} |\xi| e^{g_t^\top \xi} \mu^i(\mathrm{d}\xi) = \sup_{u\in K} G(u) < \infty,$$

since $K := \{g_s \,:\, s \in [0,T]\}$ is a compact set by continuity of g. □

Based on Lemma 4.1 and a result from [KMK10] (alternatively, one could use [CFY05]) we can now prove the following key tool:

Proposition 4.2. *Suppose* $g\colon \mathbb{R}_+ \to \mathbb{R}^d$ *is continuous and* (2.19) *holds. Then for any* $x \in D$,

(i) the process

$$E_t := \exp\left(\int_0^t g_u^\top \,\mathrm{d}X_u - \int_0^t F(g_u) + \langle X_u, R(g_u)\rangle \,\mathrm{d}u\right), \quad t \geq 0, \tag{4.8}$$

is a $\mathbb{P}_x$*-martingale,*

(ii) if for some $t \geq 0$ *and all* $s \geq 0$, $g(s) = g(s \wedge t)$, *then* $(E_{s\wedge t})_{s\geq 0}$ *is the density process (w.r.t.* $\mathbb{P}_x$*) of a measure* $\mathbb{Q}$ *on* $(\Omega, \mathcal{F})$ *such that, under* $\mathbb{Q}$, X *is a time-inhomogeneous affine process started from* $(0, x)$ *with parameters as in Lemma 4.1.*

Proof. The proof of Proposition 4.2 is structured as follows: In Step 1, E in (4.8) is rewritten as $\mathcal{E}(M)$ for a suitable local martingale M. In Step 2 it is verified that Lemma 4.1 implies conditions (4.12), (4.13) and (4.14) below. Finally, in Step 3 we combine Step 1 and 2 with [KMK10] and obtain (i) and (ii).

Step 1: We follow the notation and definitions of [JS03].

Denote by μ^X the jump-measure and by X^c the continuous martingale part of X, respectively. By (2.19), $|e^{g^\top x} - 1 - g^\top \chi(x)| * \nu$ is an adapted, continuous, increasing $\mathbb{R}$-valued process and thus (combining [JS03, Lemma I.3.10 and Proposition II.1.28]) $e^{g^\top x} - 1 + g^\top \chi(x) \in G_{loc}(\mu^X)$. By linearity and [JS03, Theorem II.2.34], $g^\top \chi(x) \in G_{loc}(\mu^X)$ and so also $e^{g^\top x} - 1 \in G_{loc}(\mu^X)$. Thus by [JS03, Theorem II.1.8(ii)], the process

$$M_t = \int_0^t g_s^\top \,\mathrm{d}X_s^c + (e^{g^\top x} - 1) * (\mu^X - \nu)_t, \quad t \geq 0 \tag{4.9}$$

is a local martingale. By an argument as above and [JS03, Corollary II.2.38], $g^\top x \in G_{loc}(\mu^X)$ and $W := e^{g^\top x} - 1 - g^\top x \in G_{loc}(\mu^X)$ and thus, using $\Delta M_t = e^{g_t^\top \Delta X_t} - 1 > -1$ one has

$$\begin{aligned}
(\log(1+x)-x) * \mu^M & \\
&= (-g^\top x + e^{g^\top x} - 1) * \mu^X \\
&= W * (\mu^X - \nu) + W * \nu \\
&\overset{(4.9)}{=} (-g^\top x) * (\mu^X - \nu) + M + \int_0^{\cdot} -g_s^\top \, \mathrm{d}X_s^c + W * \nu \\
&= \int_0^{\cdot} -g_s^\top \, \mathrm{d}X_s + M + W * \nu + \int_0^{\cdot} g_s^\top \beta(X_s) \, \mathrm{d}s + g^\top (x - \chi(x)) * \nu \\
&= -\int_0^{\cdot} g_s^\top \, \mathrm{d}X_s + M + (e^{g^\top x} - 1 - g^\top \chi(x)) * \nu + \int_0^{\cdot} g_s^\top \beta(X_s) \, \mathrm{d}s.
\end{aligned} \tag{4.10}$$

Denoting by $\mathcal{E}$ the stochastic exponential, the definition (see also [JS03, Theorem 8.10]) and (2.18) yields

$$\begin{aligned}
\mathcal{E}(M)_t &= \exp\left(M_t - \frac{1}{2} \int_0^t g_s^\top \alpha(X_s) g_s \, \mathrm{d}s - (\log(1+x) - x) * \mu_t^M \right) \\
&\overset{(4.10)}{=} \exp\Bigg(\int_0^t g_u^\top \, \mathrm{d}X_u - \int_0^t g_u^\top \beta(X_u) \, \mathrm{d}u - \frac{1}{2} \int_0^t g_s^\top \alpha(X_s) g_s \, \mathrm{d}s \\
&\qquad + (g^\top \chi(x) - e^{g^\top x} + 1) * \nu_t \Bigg) \\
&= \exp\left(\int_0^t g_u^\top \, \mathrm{d}X_u - \int_0^t F(g_u) + \langle X_u, R(g_u) \rangle \, \mathrm{d}u \right),
\end{aligned} \tag{4.11}$$

where the last step follows by definition (2.14).

Step 2: Define $W\colon \mathbb{R}_+ \times \mathbb{R}^d \to [0, \infty)$ by $W(t,x) := e^{\langle g_t, x \rangle}$. We now show that for all $j \in I \cup \{0\}$, $t \geq 0$,

$$\int_0^t \int_{D \setminus \{0\}} (1 - \sqrt{W(s,x)})^2 \mu^j(\mathrm{d}x) \mathrm{d}s < \infty, \tag{4.12}$$

$$\int_{D \setminus \{0\}} |\chi(x)(W(t,x) - 1)| \mu^j(\mathrm{d}x) < \infty, \tag{4.13}$$

$$t \mapsto \chi_k(W(t,x) - 1)(W(t,x) - 1)\mu^j(\mathrm{d}x) \text{is weakly continuous in } t \in \mathbb{R}_+. \tag{4.14}$$

It remains to argue that (4.12)-(4.14) are indeed satisfied. Since exp is Lipschitz continuous and g is continuous, there exists $C \geq 0$ such that for all $s \in [0,t]$, $|x| \leq 1$,

$$|1 - \sqrt{W(s,x)}| = |1 - e^{-\frac{1}{2}\langle g_s, x \rangle}| \leq C |\langle g_s, x \rangle| \leq C |g_s| |x|.$$

Taking $K \subset \mathbb{R}^d$ compact with $g_s, \frac{1}{2} g_s \in K$ for all $s \in [0,t]$ and splitting the integral in $\{|x| \leq 1\}$ and $\{|x| \geq 1\}$, we obtain (for G_0 as in (4.6) with $f = 1$)

$$\int_0^t \int_{D\setminus\{0\}} (1 - e^{-\frac{1}{2}\langle g_s, x\rangle})^2 \mu^j(\mathrm{d}x)\mathrm{d}s \leq C \int_0^t |g_s|^2 \,\mathrm{d}s \int_{|x|\leq 1} |x|^2 \mu^j(\mathrm{d}x) + 2 \sup_{u\in K} G_0(u) + \int_{D\setminus\{|x|\leq 1\}} \mu^j(\mathrm{d}x),$$

which is finite by the integrability properties of the Lévy-measures (2.12), (2.13) and since G_0 is continuous. Thus (4.12) indeed holds and an analogous reasoning gives (4.13).

To establish (4.14), denote $\tilde{\mu}(t, \mathrm{d}x) := \chi_k(W(t,x)-1)(W(t,x)-1)\mu^j(\mathrm{d}x)$ and again consider $D \setminus \{|\xi| \leq 1\}$ and $\{|\xi| \leq 1\} \setminus \{0\}$ separately, i.e. for $f \in C_b(D)$ write

$$\int_{D\setminus\{0\}} f(\xi)\tilde{\mu}(t,\mathrm{d}\xi) = \int_{D\setminus\{|\xi|\leq 1\}} f(\xi)\tilde{\mu}(t,\mathrm{d}\xi) + \int_{\{|\xi|\leq 1\}\setminus\{0\}} f(\xi)\tilde{\mu}(t,\mathrm{d}\xi). \tag{4.15}$$

The second term is continuous in t by dominated convergence and the same argument used to show that b and β are continuous. The first term in (4.15) is the composition of $F_0\colon \mathbb{R}^d \to \mathbb{R}_+$ defined by

$$F_0(u) := \int_{D\setminus\{|\xi|\leq 1\}} f(\xi)\chi_k(e^{\langle u,\xi\rangle} - 1)(e^{\langle u,\xi\rangle} - 1)\mu^j(\mathrm{d}\xi)$$

and g. To establish (4.14) it thus suffices to show that F_0 is continuous. To see this, assume $f \geq 0$ (for general f apply the subsequent argument to the positive and negative parts of f separately), define $h\colon [-1,\infty) \to \mathbb{R}$ by $h(z) := z^2 - z\chi_k(z)$ and write

$$F_0(u) = G_0(2u) - 2G_0(u) + G_0(0) - \int_{D\setminus\{|\xi|\leq 1\}} f(\xi)h(e^{\langle u,\xi\rangle} - 1)\mu^j(\mathrm{d}\xi)$$

with G_0 as in (4.6). For the truncation function χ chosen in [KMK10], $h(z) = \max(0, z^2 - z)$ for all $z \in [-1,\infty)$ and so h is non-decreasing and convex. In particular for any $\xi \in D\setminus\{|\xi| \leq 1\}$, the function on $\mathbb{R}^d$ defined by $u \mapsto h(e^{\langle u,\xi\rangle} - 1)$ is convex and so

$$u \mapsto \int_{D\setminus\{|\xi|\leq 1\}} f(\xi)h(e^{\langle u,\xi\rangle} - 1)\mu^j(\mathrm{d}\xi)$$

is a ($\mathbb{R}_+$-valued) convex function on $\mathbb{R}^d$. [Roc70, Corollary 10.1.1] implies that it is continuous and so the proof is complete.

Step 3: Recall that (2.3) with $c = 0$ and $\gamma = 0$ is strongly admissible in the sense of [KMK10, Definition 2.4] and by Lemma 4.1 the same holds for (4.1). Furthermore, recall the definition of M in (4.9). Since $g\colon \mathbb{R}_+ \to \mathbb{R}^d$ and $W\colon \mathbb{R}_+ \times \mathbb{R}^d \to [0,\infty)$ (defined above) are continuous, satisfy (by Step 2) conditions (4.12), (4.13) and (4.14) and since (4.2) holds, [KMK10, Theorem 4.1] and its proof show that $\mathcal{E}(M)$ is a martingale and that $\mathcal{E}(M)$ can be used as the density process of a probability measure $\mathbb{Q}$ that is locally absolutely continuous w.r.t. $\mathbb{P}_x$ and has the properties stated in (ii). But $E_t = \mathcal{E}(M)_t$ (by Step 1) and hence the claim. □

4.1.2 Estimates for R

Lemma 4.3. *There exists a function $g \in C(\mathbb{R}^d, \mathbb{R}_+)$ such that $g(x) = g((x_I^+, x_J))$ for all $x \in \mathbb{R}^d$ (with $x_I^+ = (x_1^+, \dots, x_m^+)$) and for any $u \in \mathbb{C}^d$,*

$$\operatorname{Re} \langle \bar{u}_I, R_I(u) \rangle \leq g(\operatorname{Re} u)(1 + |u_J|^2)(1 + |u_I|^2). \tag{4.16}$$

Proof. Inequality (4.16) is derived in [KRM15, Lemma 5.5] with

$$g(x) := c_0(1 + x_I^+) + c_1 e^{x^+} + \sum_{i=1}^m \int_{D \cap |\xi| \geq 1} e^{\langle \xi, x \rangle} \mu^i(\mathrm{d}\xi) + \int_{D \cap |\xi| \leq 1} \xi_i (e^{\xi_i x_i^+} - 1) \mu^i(\mathrm{d}\xi) \tag{4.17}$$

for some $c_0, c_1 > 0$. Since $\xi_k \geq 0$ for all $k \in I$, $e^{\langle \xi, x \rangle} \leq e^{\langle \xi, (x_I^+, x_J) \rangle}$ and so (4.16) remains valid if instead of g one uses $g((x_I^+, x_J))$. Continuity of this function follows by (2.19) and so the lemma is proved. □

Lemma 4.4. *Let $r > 0$ and $S_r := \{u \in \mathbb{C}^d \ : \ \forall \in I \operatorname{Re} u_i \leq r, |u_J| \leq r\}$. Then there exists $C > 0$ such that for all $u \in S_r$,*

$$|R_I(u)| \leq C(1 + |u|^2).$$

Proof. By the triangle inequality it suffices to find for each $i \in I$ a constant $C_i > 0$ such that $|R_i(u)| \leq C_i(1 + |u|^2)$ for all $u \in S_r$. For $i \in I$,

$$|R_i(u)| \leq \left(\frac{1}{2}|\alpha^i| + |\beta^i| \right)(1 + |u|^2) + \int_{D \setminus \{0\}} \left| e^{\langle \xi, u \rangle} - 1 - \langle \chi(\xi), u \rangle \right| \mu^i(\mathrm{d}\xi)$$

and so we only need to analyze the μ^i-integral. Set $B := \{z \in \mathbb{C} \ : \ \operatorname{Re} z \leq (d+1)r\}$, then for all $z \in B$,

$$|\exp(z) - 1 - z| \leq |z| \sup_{t \in (0,1)} |e^{tz} - 1| = |z| \sup_{t \in (0,1)} |z \int_0^t e^{sz} \, \mathrm{d}s| \leq |z|^2 e^{(d+1)r}.$$

Furthermore, for any $\xi \in D$ with $|\xi| \leq 1$ and $u \in S_r$,

$$\operatorname{Re} \langle \xi, u \rangle = \langle \xi, \operatorname{Re} u \rangle \leq \left(\sum_{i \in I} \xi_i + |\xi_J| \right) r \leq (d+1)r$$

implies $\langle \xi, u \rangle \in B$. Combining these two observations with the Cauchy-Schwarz inequality and $\chi(\xi) = \xi$ on $\{|\xi| \leq 1\}$ one obtains with $D_1 := D \cap \{|\xi| \leq 1\}$

$$\int_{D_1} \left| e^{\langle \xi, u \rangle} - 1 - \langle \chi(\xi), u \rangle \right| \mu^i(\mathrm{d}\xi) \leq |u|^2 e^{(d+1)r} \int_{D_1} |\xi|^2 \mu^i(\mathrm{d}\xi) =: |u|^2 C_i^0,$$

where C_i^0 is finite because of (2.12) and (2.13).

By (2.19) and [DFS03, Lemma A.2], the function

$$\tilde{R}_i(u) := \int_{D \setminus \{|\xi| \leq 1\}} e^{\langle \xi, u \rangle} \mu^i(\mathrm{d}\xi), \quad u \in \mathbb{C}^d$$

is analytic. In particular, $C_0 := \sup_{|u|\le r} |\tilde{R}_i(u)|$ is finite.

Since χ is bounded away from 0 on $D \setminus \{|\xi| \le 1\}$, $C := \mu^i(D \setminus \{|\xi| \le 1\})$ is finite (by (2.12) and (2.13)). Combining this with $|\chi(\xi)| \le d$ one obtains for $u \in S_r$

$$\int_{D\setminus\{|\xi|\le 1\}} \left|e^{\langle \xi, u\rangle} - 1 - \langle \chi(\xi), u\rangle\right| \mu^i(\mathrm{d}\xi) \le \int_{D\setminus\{|\xi|\le 1\}} e^{\langle \xi, \mathrm{Re}\, u\rangle} \mu^i(\mathrm{d}\xi) + C(1 + d|u|)$$
$$\le C_0 + C(1 + |u|),$$

where we have used $\xi_I \in \mathbb{R}^m_+$ for the last estimate. Combining all the estimates yields the desired statement. □

4.1.3 Properties of $\bar{T}^y$

To prepare the proof of Proposition 3.8 we provide two additional Lemmas. The first is an application of Itô's lemma and essentially identifies the extended generator of $\bar{T}^y$ in (3.8). The second Lemma rephrases a result from [Fil05].

Recall that $H(x) = (x_0)_I^\top x_I + c_0$ and for $f \in C_0^{1,2}(\mathbb{R}_+ \times D)$, $t \ge 0$, set

$$\begin{aligned} \mathcal{A}_t^y f(t,x) &= \mathcal{A} f(t,x) + f(t,x)[F(-y_t) + \langle x, R(-y_t)\rangle - H(x)] \\ &\quad - \langle \alpha(x) y_t, \nabla_x f(t,x)\rangle \\ &\quad + \int_{D\setminus\{0\}} [f(t, x+\xi) - f(t,x)](e^{-\langle y_t, \xi\rangle} - 1) K(x, \mathrm{d}\xi). \end{aligned} \tag{4.18}$$

Proposition 4.5. *Suppose $y\colon \mathbb{R}_+ \to \mathbb{R}^d$ is continuous,* (2.3) *are admissible with $c = 0$, $\gamma = 0$ and* (2.19) *holds. Define*

$$U_t := \exp\left(-\int_0^t y_u^\top \,\mathrm{d}X_u - \int_0^t H(X_u)\,\mathrm{d}u\right) \tag{4.19}$$

and $\mathcal{A}^y$ as in (4.18)*. Then for any $f \in C_0^{1,2}(\mathbb{R}_+ \times D)$, the process*

$$U_t f(t, X_t) - f(0, X_0) - \int_0^t U_u(\partial_u + \mathcal{A}_u^y) f(u, X_u)\,\mathrm{d}u, \quad t \ge 0, \tag{4.20}$$

is a local martingale.

Proof. Define E by (4.8) with $g := -y$. Then $E_t = \mathcal{E}(M)_t$ for M as in (4.9). Furthermore, $U_t = E_t \exp(V_t)$, where

$$V_t = \int_0^t F(-y_u) + \langle X_u, R(-y_u)\rangle\,\mathrm{d}u - \int_0^t H(X_u)\,\mathrm{d}u$$

is continuous and of bounded variation, $[E, V] = 0$ and thus

$$dU_t = d(\mathcal{E}(M)_t \exp(V_t)) = U_{t-}\,\mathrm{d}M_t + U_t\,\mathrm{d}V_t.$$

For $f \in C_0^{1,2}(\mathbb{R}_+ \times D)$, Itô's formula shows that

$$f(t, X_t) = f(0, X_0) + \int_0^t (\partial_s + \mathcal{A}) f(s, X_s)\,\mathrm{d}s + N_t, \quad t \ge 0,$$

where N is a local martingale with continuous part $\int \nabla_x f(s, X_{s-})^\top \, \mathrm{d}X^c_s$. Combining this with (2.18), the definition (4.9), the fact that f is bounded and $e^{-y^\top x} - 1 \in G_{loc}(\mu^X)$, we obtain

$$\begin{aligned}
[M, N]_t &= \langle M^c, N^c \rangle_t + \sum_{s \le t} (f(s, X_s) - f(s, X_{s-}))(e^{-y_s \Delta X_s} - 1) \\
&\doteq - \int_0^t y_s^\top \alpha(X_s) \nabla_x f(s, X_s) \, \mathrm{d}s \\
&\quad + \int_0^t (f(s, X_{s-} + \xi) - f(s, X_{s-}))(e^{-y_u^\top \xi} - 1) K(X_{u-}, \mathrm{d}\xi) \, \mathrm{d}u,
\end{aligned}$$

where $U \doteq V$ means that $U - V$ is a local martingale.

Putting everything together, Itô's formula written in differential form gives

$$\begin{aligned}
dU_t f(t, X_t) &\doteq U_t (\partial_t + \mathcal{A}) f(t, X_t) \, \mathrm{d}t + f(t, X_{t-}) U_t dV_t + d[U, N]_t \\
&\doteq U_t \mathcal{A}^y_t f(t, X_t) \, \mathrm{d}t,
\end{aligned}$$

which shows that (4.20) is a local martingale. □

In the following Lemma, we allow the function spaces (defined before) to contain complex valued functions.

Lemma 4.6. *There exists a dense subset $L \subset C_0(D)$ with the following property: for any $T > 0$, $h \in L$ there exists $u \in C^{1,2}([0, T] \times D)$ bounded, satisfying*

$$\begin{aligned}
\partial_t u(t, x) + \mathcal{A}^y_t u(t, x) &= 0 \qquad (t, x) \in [0, T) \times D, \\
u(T, x) &= h(x) \qquad\qquad x \in D.
\end{aligned} \tag{4.21}$$

Proof. Denote by $\Theta_0 \subset C_0(D)$ the set of $\mathbb{C}$-valued functions from [DFS03, Proposition 8.2]. Any $h \in \Theta_0$ is of the form

$$h(x) = \int_{\mathbb{R}^n} e^{(v, iq)^\top x} g(q) \, \mathrm{d}q, \quad x \in D,$$

for some $g \in C_c^\infty(\mathbb{R}^n)$ and $v \in \mathbb{C}^m_{--}$. Denote by L the complex linear span of Θ_0. In [DFS03, Lemma 8.4] it is shown that L dense in $C_0(D)$.

Fix $T > 0$ and $h \in \Theta_0$. For $(t, x) \in [0, T] \times D$, define $u(t, x) := P_{t,T} h(x)$. Then u is bounded, satisfies $u(T, \cdot) = h$ and, as established in the proof of [Fil05, Proposition 6.3], $u \in C^{1,2}([0, T] \times D)$ and (4.21) indeed holds. □

4.2 Proof of Proposition 3.8

Proof of Proposition 3.8. Since $x^i_0 > 0$ for $i \in I$, $y_0 = 0$, $R(0) = 0$ and R, y are continuous, there exists $T > 0$ such that $x^i_0 - R_i(-y_t) \ge 0$ for $t \in [0, T]$. Taking $c_0 = \sup_{t \in [0,T]} F(-y_t)$ it follows that $c(t) \ge 0$ and $\gamma(t) \in \mathbb{R}^m_+$ for all $t \in [0, T]$. Combining this with Lemma 4.1, it follows that the parameters are indeed admissible.

To prove the proposition, it suffices to show $\bar{T}^y_t = P^y_{0,t}$ and for this it is sufficient to show $\bar{T}^y_t h = P^y_{0,t} h$ for all h in a dense subset of $C_0(D)$. Taking L from Lemma 4.6, for any $h \in L$ we find $f \in C_0^{1,2}([0,t] \times D)$ such that $f(t,\cdot) = h$ and the du-integral in (4.20) vanishes. Hence

$$N_s := U_{s\wedge t} f(s \wedge t, X_{s\wedge t}), \quad s \geq 0,$$

is a local martingale by Proposition 4.5, where U is as in (4.19). On the other hand,

$$U_t = \exp(V_t) E_t,$$

where E is defined in (4.8) (with $g = -y$) and

$$V_t := \int_0^t F(-y_u) + \langle X_u, R(-y_u)\rangle \,\mathrm{d}u - \int_0^t H(X_u)\,\mathrm{d}u, \quad t \geq 0.$$

Since $\beta^j = 0$ for $j \in J$ and $x^i \geq 0$ for $i \in I$,

$$H(x) - F(-y_t) - \langle x, R(-y_t)\rangle = c(t) + x^\top \gamma(t) \geq 0$$

for all $(t,x) \in [0,T] \times D$. Thus $V_t \leq 0$ for $t \in [0,T]$ and $\exp(V_t)$ is bounded on $[0,T]$. Since E is a martingale by Proposition 4.2, the local martingale N satisfies

$$N_s = \exp(V_{t\wedge s}) E_{t\wedge s} f(s \wedge t, X_{s\wedge t}), \quad s \geq 0,$$

and is the product of a bounded process and a martingale. Thus N is a true martingale and combining this with $f(t,\cdot) = h$, the definition (3.8) and $f(s,\cdot) = P_{s,t}h$ (see Lemma 4.6) yields

$$\bar{T}^y_t h(x) = \mathbb{E}_x[U_t f(t,X_t)] = \mathbb{E}_x[N_t] = \mathbb{E}_x[N_0] = f(0,x) = P_{0,t}h(x). \tag{4.22}$$

□

4.3 Proof of Theorem 3.1 and 3.5

Proof of Theorem 3.1. We proceed in two steps: First (3.4) is verified under the assumption that a solution to (3.3) exists. In the second part, existence and uniqueness for (3.3) is established.

Expression for the Fourier coefficients: Since Ψ is continuously differentiable, each component is of finite variation and thus $[\Psi^j, X^j] = 0$ for all j. By the product rule and (3.3),

$$\begin{aligned}
(u + y_T)^\top X_T - \Psi(0,T,u)^\top X_0 &= \Psi(T,T,u)^\top X_T - \Psi(0,T,u)^\top X_0 \\
&= \int_0^T \Psi(s,T,u)^\top \,\mathrm{d}X_s + \int_0^T X_s^\top \partial_s \Psi(s,T,u)\,\mathrm{d}s \\
&= \int_0^T \Psi(s,T,u)^\top \,\mathrm{d}X_s - \int_0^T X_s^\top (R(\Psi(s,T,u) - y_s) - \gamma_s)\,\mathrm{d}s.
\end{aligned} \tag{4.23}$$

By Proposition 4.2 applied to the continuous function $g\colon \mathbb{R}_+ \to \mathbb{R}^d$ defined by

$$g(t) := \Psi(t \wedge T, T, u) - y_{t\wedge T},$$

the process

$$\tilde{E}_t := \exp\left(\int_0^t g_u^\top \,\mathrm{d}X_u - \int_0^t F(g_u) + \langle X_u, R(g_u)\rangle \,\mathrm{d}u\right), \quad t \geq 0,$$

is a martingale.

Combining this with (4.23) and the definition of ρ we obtain

$$\begin{aligned}
\rho_T^x(f_u, y) &= \mathbb{E}\left[\exp\left((u + y_T)^\top X_T - \int_0^T y_s^\top \,\mathrm{d}X_s - \int_0^T c_s + X_s^\top \gamma_s \,\mathrm{d}s\right)\right] \\
&= \mathbb{E}\Bigg[\exp\Bigg(\Psi(0,T,u)^\top X_0 + \int_0^T (\Psi(s,T,u) - y_s)^\top \,\mathrm{d}X_s \\
&\qquad - \int_0^T c_s X_s^\top R(\Psi(s,T,u) - y_s)\,\mathrm{d}s\Bigg)\Bigg] \\
&= \mathbb{E}\left[\tilde{E}_T \exp\left(\Psi(0,T,u)^\top X_0 + \int_0^T F(\Psi(s,T,u) - y_s) - c_s \,\mathrm{d}s\right)\right] \\
&= \exp(\Phi(0,T,u) + \Psi(0,T,u)^\top x).
\end{aligned} \tag{4.24}$$

Existence and uniqueness of solutions to (3.3)**:** Suppose first for some $T > 0$ there exists $\tilde{\Psi} \in C^1([0,T], \mathbb{R}^d)$ satisfying

$$\partial_t \tilde{\Psi}(t,u) = R(\tilde{\Psi}(t,u) + y_T - y_{T-t}) - \gamma_{T-t}, \quad \tilde{\Psi}(0,u) = u. \tag{4.25}$$

Then a solution to (3.3) is obtained by setting $\Psi(t,T,u) := \tilde{\Psi}(T-t,u) + y_T$ and

$$\Phi(t,T,u) = \int_t^T F(\Psi(s,T,u) - y_s) - c_s \,\mathrm{d}s, \quad t \in [0,T].$$

Conversely, any solution to (3.3) gives rise to $\tilde{\Psi}$ satisfying (4.25) by setting $\tilde{\Psi}(t,u) := \Psi(T-t,T,u) - y_T$. Thus, to prove the theorem it suffices to construct $T > 0$ such that for all $u \in i\mathbb{R}^d$ there exists a unique $\tilde{\Psi}(\cdot,u) \in C^1([0,T],\mathbb{R}^d)$ satisfying (4.25). To do so, we will establish the following statements:

(i) for any $T > 0$, $u \in \mathbb{C}^d$, there exists $t_+(u,T) \in (0,\infty]$ such that (4.25) has a unique solution on $[0, t_+(u,T))$. If $t_+(u,T) < \infty$, then it holds that $\lim_{t\uparrow t_+(u,T)} |\tilde{\Psi}(t,u)| = \infty$.

(ii) there exists $T_0 > 0$ such that $t_+(0,T_0) > T_0$, i.e. the solution to (4.25) with $u = 0$, $T = T_0$ exists on $[0,T_0]$.

(iii) for any $u \in i\mathbb{R}^d$, $t_+(u,T_0) > T_0$.

Then (iii) implies that for any $u \in i\mathbb{R}^d$ there exists a unique solution to (4.25) on $[0, T_0]$, which proves the theorem. We now show (i)-(iii). In what follows, we set $y_r := 0$ for $r < 0$ so that $y \in C(\mathbb{R}, \mathbb{R}^d)$.

(i) By [DFS03, Lemma 5.3] and (2.19), R is an analytic function. In particular it is locally Lipschitz continuous. Combining this with the fact that y is continuous, (i) follows from the global existence and uniqueness result for ordinary differential equations [Ama90, Theorem 7.6].

(ii) For $(t, z, T) \in \mathbb{R} \times \mathbb{C}^d \times \mathbb{R}$, set

$$f(t, z, T) := R(z + y_T - y_{T-t}) - \gamma_{T-t}.$$

Then $f \in C(\mathbb{R} \times \mathbb{C}^d \times \mathbb{R}, \mathbb{C}^d)$ and, since R is locally Lipschitz-continuous, the prerequisites of[Ama90, Theorem 8.3] are satisfied. Thus, the set

$$D := \{(t, \tau, u, T) \in \mathbb{R} \times \mathbb{R} \times \mathbb{C}^d \times \mathbb{R} \ : \ t \in J(\tau, u, T)\}$$

is open, where $J(\tau, u, T)$ is the maximal interval of existence of the (unique) solution to

$$\dot{x}(t) = f(t, x(t), T), \quad x(\tau) = u.$$

Since $(0, 0, 0, 0) \in D$ and D is open, $(T_0, 0, 0, T_0) \in D$ for $T_0 > 0$ small enough. Thus $T_0 \in J(0, 0, T_0)$ and, since the right endpoint of the open interval $J(0, 0, T_0)$ is $t_+(0, T_0)$, the claim follows.

(iii) Fix $u \in i\mathbb{R}^d$. By (ii), $t_+(0, T_0) > T_0$ and so it suffices to show that $t_+(u, T_0) \geq t_+(0, T_0)$ or, by (i), that $|\tilde{\Psi}(t, u)|$ does not explode on $[0, T_0]$. Consider the J-components first. By (2.8), for $j \in J$ (4.25) is given as

$$\partial_t \tilde{\Psi}_j(t, u) = \langle \beta^j, \tilde{\Psi}_J(t, u) + y_J(T) - y_J(T - t) \rangle - (\gamma_{T-t})_j, \quad \tilde{\Psi}_j(0, u) = u_j \tag{4.26}$$

and, as this is a system of first order linear equations, $\tilde{\Psi}_J(t, u)$ exists for all $t \geq 0$. Thus it remains to analyze the I-components. We claim that there exists constants $c_0, c_1 > 0$ such that for all $t \in [0, T_0 \wedge t_+(u, T_0))$

$$\partial_t |\tilde{\Psi}_I(t, u)|^2 \leq c_0(c_1 + |\tilde{\Psi}_I(t, u)|^2). \tag{4.27}$$

Assuming that (4.27) has been established, Gronwall's inequality applied to $c_1 + |\tilde{\Psi}_I(t, u)|^2$ implies

$$|\tilde{\Psi}_I(t, u)|^2 \leq (c_1 + |u_I|^2) \exp(c_0 t) - c_1 \tag{4.28}$$

for all $t \in [0, T_0 \wedge t_+(u, T_0))$. This allows to conclude (iii) by contradiction: If $T_0 \geq t_+(u, T_0)$, then (4.28) holds for all $t \in [0, t_+(u, T_0))$ and the left hand side of (4.28) explodes as $t \uparrow T_0$, whereas the right hand side is bounded by its value at T_0. Hence, by contradiction $T_0 < t_+(u, T_0)$ as claimed.

Therefore it suffices to establish (4.27). To do so, we follow the proof of [DFS03, Proposition 6.1] and [KRM15, Proposition 5.1]. For $t \in \mathbb{R}$, set $\bar{y}(t) := y_{T_0} - y_{T_0 - t}$. As argued above, (4.26) implies that $\tilde{\Psi}_J(t, u)$ exists for all $t \geq 0$. Furthermore, the real part of (4.26) does not depend on u and therefore

$$\operatorname{Re} \tilde{\Psi}_J(t, u) = \operatorname{Re} \tilde{\Psi}_J(t, 0) = \tilde{\Psi}_J(t, 0). \tag{4.29}$$

Set $T := t_+(0,T_0) \wedge t_+(u,T_0)$ and for $(t,x) \in [0,T] \times \mathbb{R}^m$,

$$f(t,x) := R_I((x, \tilde{\Psi}_J(t,0)) + \bar{y}(t)) - (\gamma_{T_0-t})_I.$$

Then by [KRM15, Lemma 5.7], continuity of y and Lipschitz continuity of R_I, f satisfies the conditions of the comparison result [MMKS11, Proposition A.2]. Furthermore, (4.29) and the inequality $\mathrm{Re}\, R_i(z) \le R_i(\mathrm{Re}\,(z))$ (valid for all $z \in \mathbb{C}^d$) yield

$$\partial_t \mathrm{Re}\, \tilde{\Psi}_i(t,u) - f_i(t, \mathrm{Re}\, \tilde{\Psi}_I(t,u)) \le 0 = \partial_t \tilde{\Psi}_i(t,0) - f_i(t, \tilde{\Psi}_I(t,0)),$$

and $\mathrm{Re}\, \tilde{\Psi}_i(0,u) = \tilde{\Psi}_i(0,0)$ for $t \in [0,T)$, $i \in I$. Hence the comparison result [MMKS11, Proposition A.2] implies

$$\mathrm{Re}\, \tilde{\Psi}_i(t,u) \le \tilde{\Psi}_i(t,0), \quad \forall i \in I,\ t \in [0, t_+(0,T_0) \wedge t_+(u,T_0)) \tag{4.30}$$

For $t \in [0, T_0 \wedge t_+(u,T_0))$ one uses (4.25) to write

$$\begin{aligned} \frac{1}{2}\partial_t |\tilde{\Psi}_I(t,u)|^2 &= \mathrm{Re}\, \langle \overline{\tilde{\Psi}_I}(t,u), \partial_t \tilde{\Psi}_I(t,u)\rangle \\ &= \mathrm{Re}\, \langle \overline{\tilde{\Psi}_I}(t,u) + \bar{y}(t), R_I(\tilde{\Psi}(t,u) + \bar{y}(t))\rangle \\ &\quad - \langle \bar{y}(t), \mathrm{Re}\, R_I(\tilde{\Psi}(t,u) + \bar{y}(t))\rangle \\ &\quad - \langle \mathrm{Re}\, \tilde{\Psi}_I(t,u), (\gamma_{T_0-t})_I\rangle \\ &= I_1 - I_2 - I_3, \end{aligned} \tag{4.31}$$

where each I_i denotes an inner product. The three inner products in (4.31) can be estimated separately:

For the first one, denote by $g \in C(\mathbb{R}^d, \mathbb{R}_+)$ the function from Lemma 4.3, write $x^{+,I} := (x_I^+, x_J)$ for $x \in \mathbb{R}^d$ and recall $g(x) = g(x^{+,I})$. By (4.30) and the fact that $\tilde{\Psi}_J(t,u)$ exists for all $t \ge 0$, there exists $K \subset \mathbb{R}^d$ compact such that $(\mathrm{Re}\, \tilde{\Psi}(t,u) + \bar{y}(t))^{+,I} \in K$ for all $t \in [0, T_0 \wedge t_+(u,T_0))$. Hence Lemma 4.3 yields

$$\begin{aligned} I_1 &\le g(\mathrm{Re}\, \tilde{\Psi}(t,u) + \bar{y}(t))(1 + |\tilde{\Psi}_J(t,u) + \bar{y}_J(t)|^2)(1 + |\tilde{\Psi}_I(t,u) + \bar{y}_I(t)|^2) \\ &\le 4g((\mathrm{Re}\, \tilde{\Psi}(t,u) + \bar{y}(t))^{+,I})(1 + |\tilde{\Psi}_J(t,u) + \bar{y}_J(t)|^2)(C_1 + |\tilde{\Psi}_I(t,u)|^2) \\ &\le C_0(C_1 + |\tilde{\Psi}_I(t,u)|^2) \end{aligned} \tag{4.32}$$

where $C_0 := (4 \sup_{x \in K} g(x) \wedge 1) \sup_{t \in [0,T_0]} (1 + |\tilde{\Psi}_J(t,u) + \bar{y}_J(t)|^2)$ and $C_1 := 1 + 2 \sup_{t \in [0,T_0]} |\bar{y}(t)|^2$.

For the second one, Lemma 4.4, the fact that $\tilde{\Psi}_J(t,u)$ exists for all $t \ge 0$ and (4.30) yield that there exists $C > 0$ such that

$$\begin{aligned} -I_2 &\le |\bar{y}(t)||R_I(\tilde{\Psi}(t,u) + \bar{y}(t))| \\ &\le CC_1(1 + |\tilde{\Psi}(t,u) + \bar{y}(t)|^2) \\ &\le 4CC_1(C_0 + C_1 + |\tilde{\Psi}_I(t,u)|^2) \end{aligned} \tag{4.33}$$

and for the last one

$$-I_3 \leq |\gamma_{T_0-t}||\tilde{\Psi}_I(t,u)| \leq \sup_{s\in[0,T_0]} |\gamma_s|(1+|\tilde{\Psi}_I(t,u)|^2). \tag{4.34}$$

Combining (4.31) with the estimates (4.32),(4.33) and (4.34) yields (4.27), as desired. □

Proof of Theorem 3.5. Precisely as in the derivation of (4.24), one combines the definition (3.5), the product rule (4.23) for $u=0$ and the definition of E_t in (4.8) to write

$$\begin{aligned}\rho_t(G,y) &= \mathbb{E}\left[G(X_{[0,t]})E_t \exp\left(\Psi(0,t,0)^\top X_0 + \int_0^t F(g_s) - c_s \,\mathrm{d}s\right)\right] \\ &= \exp(\Phi(0,t,0)) \int_D \exp(\langle x, \Psi(0,t,0)\rangle)\mathbb{E}_x[G(X_{[0,t]})E_t]\pi_0(\mathrm{d}x).\end{aligned}$$

But for any $x \in D$,

$$\mathbb{E}_x\left[G(X_{[0,t]})E_t\right] = \mathbb{E}_{\mathbb{Q}_x^{y,t}}\left[G(X_{[0,t]})\right]$$

with $\mathbb{Q}_x^{y,t} = \mathbb{Q}$ as in Lemma 4.2(ii). Thus the statement follows from the definition of $\bar{\pi}_t(G,y)$ and Lemma 4.2(ii). □

5 Illustration: Filtering a Cox-Ingersoll-Ross process

In this section the methodology developed in Section 3 is applied to the problem of filtering a Cox-Ingersoll-Ross process. We compare the approximation via our linearized filtering functional (LFF) (respectively the induced affine functional filter (AFF)) and other existing approximate filtering methods to the true solution.

5.1 Problem formulation

A Cox-Ingersoll-Ross (CIR) process is a weak solution to the stochastic differential equation

$$dX_t = (b+\beta X_t)dt + \sigma\sqrt{X_t}dB_t, \quad X_0 = x, \tag{5.1}$$

where $b \geq 0$, $\beta \in \mathbb{R}$, $\sigma > 0$ and B is a Brownian motion. Denoting by $\mathbb{P}_x$ the law of X, this gives rise to a conservative affine process with state space $D = \mathbb{R}_+$. The parameters in (2.3) are given as $(0, \sigma^2, b, \beta, 0, 0, 0, 0)$. Let W a Brownian motion independent of X, $\Gamma > 0$ and set

$$Y_t = \int_0^t X_s \,\mathrm{d}s + \Gamma W_t, \quad t \geq 0. \tag{5.2}$$

The goal is to calculate, for any $t \geq 0$, the distribution of X_t conditional on the σ-algebra generated by $(Y_s)_{s\in[0,t]}$ (see Section 2.2.1). In particular, we are interested in the conditional mean and variance

$$\begin{aligned} \hat{x}_t &= \mathbb{E}_x[X_t|\mathcal{F}_t^Y], \\ V_t &= \mathbb{E}_x[(X_t - \hat{x}_t)^2|\mathcal{F}_t^Y], \quad t \geq 0. \end{aligned} \tag{5.3}$$

There are various methods available to numerically approximate (5.3). For any of these methods one has to pass to a setup of discrete-time observations at some stage. To do this we fix $T > 0$, $N \in \mathbb{N}$ and a time-grid $0 = t_0 < t_1 < \ldots < t_N = T$. Instead of observing the entire path (5.2), one observes at time t_i the random variable

$$y_i = X_{t_i}(t_i - t_{i-1}) + \Gamma\sqrt{t_i - t_{i-1}}\varepsilon_i, \tag{5.4}$$

for $i = 1, \ldots N$, where $\varepsilon_1, \ldots \varepsilon_N$ are i.i.d. standard normal random variables. This amounts to discretizing the integral in (5.2) using a Riemann sum and setting $y_i = Y_{t_i}^{disc} - Y_{t_{i-1}}^{disc}$. The filtering distribution is then approximated as $\mathbb{E}_x[f(X_{t_n})|\mathcal{F}_{t_n}^Y] \approx \pi_{t_n}^N(f)$ with

$$\pi_{t_n}^N(f) := \mathbb{E}_x[f(X_{t_n})|\mathcal{F}_{t_n}^{Y,N}], \tag{5.5}$$

for any measurable $f\colon \mathbb{R}_+ \to \mathbb{C}$ satisfying $\mathbb{E}_x[|f(X_t)|] < \infty$, where $\mathcal{F}_{t_n}^{Y,N} = \sigma(y_1, \ldots y_n)$ and $n = 1, \ldots N$. In particular, instead of (5.3) in what follows we will denote

$$\begin{aligned} \hat{x}_t &= \mathbb{E}_x[X_t|\mathcal{F}_t^{Y,N}], \\ V_t &= \mathbb{E}_x[(X_t - \hat{x}_t)^2|\mathcal{F}_t^{Y,N}], \quad t \in \{t_0, \ldots t_N\}. \end{aligned} \tag{5.6}$$

5.2 Numerical solution: Approximate filtering methods

There are various methods at hand to numerically approximate (5.5) and (5.6). To illustrate the quality of these we first generate a sample path of the signal and observation process. More precisely, a sample of $(X_{t_0}, X_{t_1}, \ldots, X_{t_N})$ is generated by (exact) sampling from the transition density (see [Gla04, Section 3.4]). Based on this sample, a sample of $(y_1, \ldots, y_N)$ is generated using (5.4).

For this sample observation we now compare different methods for approximating (5.5) and (5.6). As a benchmark we calculate (5.6) using a (bootstrap) particle filter with sufficiently many particles (10^6 in the examples below), see [BC09a, Chapter 10]. In the plots these results will be denoted by $\hat{x}$ and V by slight abuse of notation.

This benchmark is now compared to the approximation using the linearized filtering functional (LFF, developed in the present chapter) and two standard approximations (explained in more detail below): A Gamma-approximation ([Bat06]) and a normal approximation ([GP99], see also [BH98]). The respective approximations to (5.6) are denoted as follows:

- Normal: $\hat{x}^{(EKF)}$, $V^{(EKF)}$

- Gamma: $\hat{x}^{(G)}$, $V^{(G)}$
- LFF: $\hat{x}^{(LFF)}$, $V^{(LFF)}$.

Firstly, let us explain the approximations from [Bat06] and [GP99] in more detail. In both cases basic idea is to *postulate* that (at each time-step t_n) the conditional distribution in (5.5) belongs to a certain two-parameter family of probability distributions (Normal in [GP99] and Gamma in [Bat06]). Then (at each time-step t_n) one only needs to approximate (5.6) and determine the two parameters from this. In [GP99] the updating procedure for (5.6) is based on the exact formulas for the mean and variance of a CIR process and the Kalman filter. This can be seen as a version of the extended Kalman filter. In [Bat06] numerical integration on the level of characteristic functions is used to update (5.6). We refer to these articles for more details. Both approximations [Bat06] and [GP99] can be viewed as special cases of the projection filter (first introduced in [BHL98]), see [BH98].

Finally, the unconditional mean and variance are denoted by $\bar{x}_t := \mathbb{E}_x[X_t]$ and $v_t := \mathbb{E}_x[(X_t - \bar{x}_t)^2]$. Since these correspond to a situation where no observations are available, a comparison of $(\bar{x}, v)$ and $(\hat{x}, V)$ shows how much information the (sample path of the) observation $(y_1, \ldots, y_N)$ contains about X. Therefore, these are also shown in the plots below.

5.3 Discussion

We now compare the methods introduced above for two sets of parameters. For both settings the following choices have been made:

- instead of a constant x, the signal process X is started in $X_0 = \max(0, Z)$, where $Z \sim \mathcal{N}(x_0, s_0^2)$ is independent of B and W,
- the time horizon is $T = 1$ and the discretization uses an equidistant grid $t_i = iT/N$, $i = 0, \ldots N$,
- $N = 1000$, $\sigma = 0.04$, $\beta = -0.2$ and $s_0 = 2 \cdot 10^{-5}$.

The remaining parameter values differ for the two settings; they are indicated in the caption of the figures.

Case 1 We choose $b = 10^{-6}$, $\Gamma = x_0 = 0.005$. Figures II.1 and II.2 show the same sample path of a CIR process. The sample of observations is not shown in the plot, but one clearly sees that for t sufficiently large the conditional mean $\hat{x}$ is neither very close to X nor very close to the mean $\hat{x}$. Thus, the filtering problem is indeed not trivial: the posterior distribution in (5.5) is neither close to the distribution of X_{t_n} nor concentrated at X_{t_n}.

In both figures the conditional mean $\hat{x}$ is shown along with (dotted) "confidence bounds" given by $\hat{x} + \sqrt{V}$ and $\hat{x} - \sqrt{V}$. This allows to show both conditional mean and variance in the same plot. The analogous bounds are also shown for the unconditional mean and the different approximations.

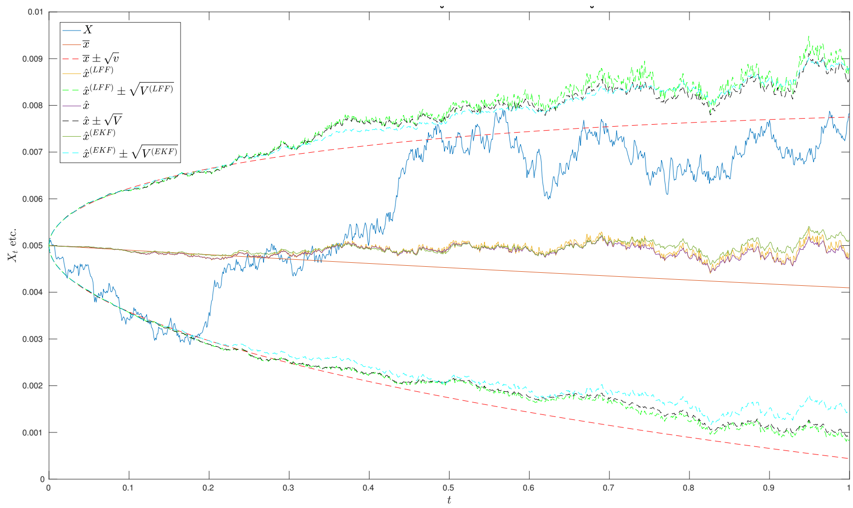


Figure II.1: Case 1: Comparison with extended Kalman filter.

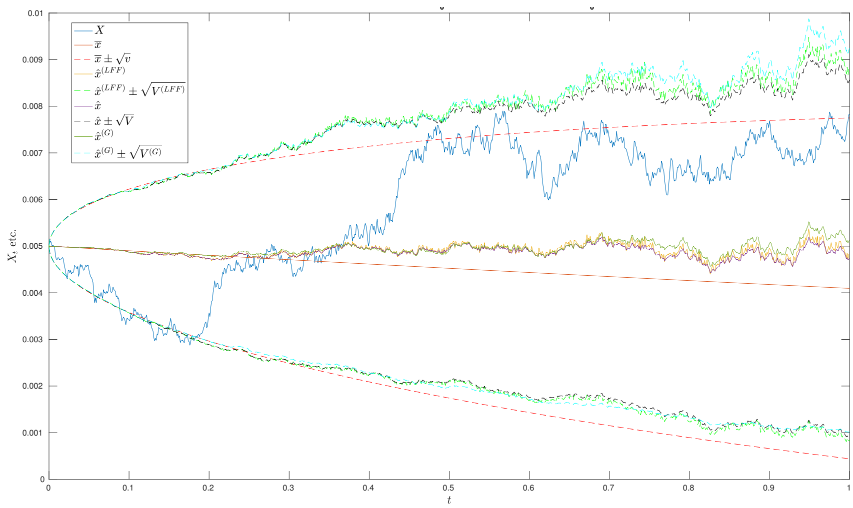


Figure II.2: Case 1: Comparison with Gamma approximation.

The two figures illustrate that the linearized filtering functional provides a more accurate approximation for (5.6) than the standard methods.

Case 2 We choose $b = 2 \cdot 10^{-5}$, $\Gamma = x_0 = 0.0001$. In this case both the approximation using the linearized filtering functional (LFF) and the normal approximation are not very good. However, it appears that the LFF-approximation becomes better as t approaches 1. Although this behaviour is typical in the present parameter regime, a precise explanation (possibly based on ergodicity properties of the CIR process) is presently not available.

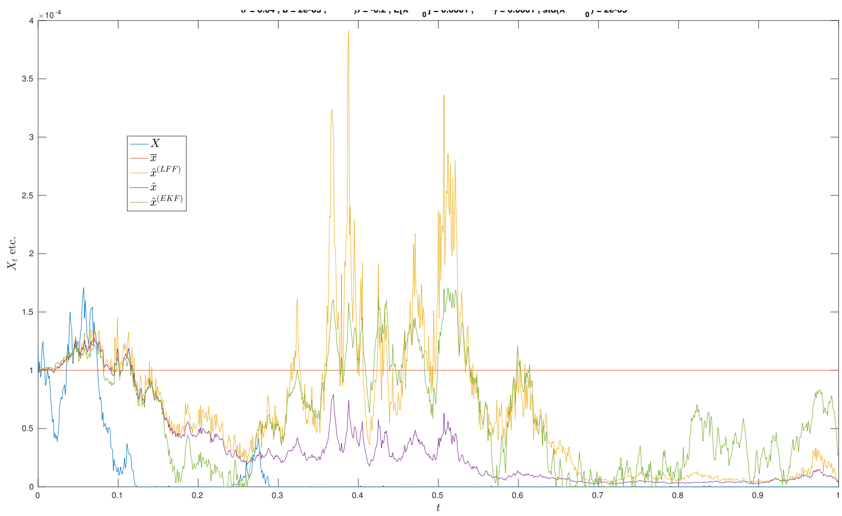


Figure II.3: Case 2: Comparison with extended Kalman filter.

6 Illustration: Filtering a Wishart process

So far this chapter has been concerned with the filtering problem for $\mathbb{R}^m_+ \times \mathbb{R}^{d-m}$-valued affine processes. We now test the methodology on Wishart processes, an S_d^+-valued generalization of the CIR process (as studied in Section 5). Here S_d^+ denotes the set of all symmetric, positive semidefinite $d \times d$ matrices. Wishart processes were introduced in [Bru91] and are commonly used for multivariate stochastic volatility modeling. They are a subclass of S_d^+-valued affine processes as characterized in [CFMT11].

Although in theory sequential Monte Carlo methods can be applied for numerically filtering Wishart processes, in practice this is infeasible for $d \geq 3$ (see below). Hence, so far no numerical method has been available for this problem. We fill this gap by introducing a linearized filtering functional analogous to (3.1) and perform numerical experiments for $d = 3$. This section contains simulation results. A generalization of the theory in Sections 3 and 4 to S_d^+-valued affine processes will be subject of future work.

6.1 The signal process

Denote by S_d^+ the set of all symmetric, positive semidefinite $d \times d$ matrices and set $S_d^- = -S_d^+$. A Wishart process is (an S_d^+-valued) weak solution to

$$dX_t = (b + HX_t + X_t H^\top)\mathrm{d}t + \sqrt{X_t}\mathrm{d}B_t\Sigma + \Sigma^\top \mathrm{d}B_t^\top \sqrt{X_t}, \quad X_0 = x, \qquad (6.1)$$

for B a $d \times d$-matrix of independent standard Brownian motions and suitable $b \in S_d^+$, $x \in S_d^+$, $H \in \mathbb{R}^{d\times d}$, $\Sigma \in \mathbb{R}^{d\times d}$. For simplicity, we assume that $\Sigma \in S_d^+$, $H = 0$, $b = n\Sigma^2$ for some $n \in \mathbb{N}$ with $n \geq d+1$ and that x has distinct eigenvalues. Then [Bru91, Proof of Theorem 2"] ensures that (6.1) has a unique strong solution for all $t \geq 0$. It also ensures that sample paths of X can be simulated easily: Given $z_0 \in \mathbb{R}^{n\times d}$ with $x = z_0^\top z_0$ and an $n \times d$-Brownian motion W, set $Z_t = W_t\Sigma + z_0$ for $t \geq 0$. Then $X := Z^\top Z$ is a weak solution to (6.1). Hence, to simulate a sample path of X one only needs to simulate a sample path of W and apply these two transformations. Finally, for $u, v \in S_d$ (the set of symmetric $d \times d$-matrices) define $\langle u, v\rangle_{S_d} := \mathrm{tr}(uv)$. Then for $t \geq 0$ the Laplace transform of X_t is given by

$$\mathbb{E}[e^{\langle u, X_t\rangle_{S_d}}] = \exp(\phi(t,u) + \langle \psi(t,u), x\rangle_{S_d}), \quad u \in S_d^-$$

for some $\phi : \mathbb{R}_{\geq 0} \times S_d^- \to \mathbb{R}_-$ and $\psi : \mathbb{R}_{\geq 0} \times S_d^- \to S_d^-$. In fact ϕ and ψ solve generalized Riccati equations (2.15) with $R(u) := 2u\Sigma^2 u$, $F(u) := n\mathrm{tr}(\Sigma^2 u)$.

6.2 Numerical solution of the filtering problem

Fix $h\colon S_d^+ \to \mathbb{R}^m$ linear and $\Gamma \in \mathbb{R}^{m\times m}$ symmetric, invertible. The observation process Y is defined as

$$Y_t = \int_0^t h(X_s)\,\mathrm{d}s + \Gamma W_t, \quad t \geq 0,$$

where W is an m-dimensional Brownian motion independent of X, and (the signal process) X is a solution to (6.1) with parameters as specified above (under $\mathbb{P}$). As before our goal is to numerically calculate the distribution of X_t conditional on the σ-algebra generated by $(Y_s)_{s\in[0,t]}$, for any $t \geq 0$. For this two methods are used: firstly a bootstrap particle filter as in [BC09a, Chapter 9] and secondly the approximate affine filter (AFF) induced by the linearized filtering functional (LFF). These are defined analogously to the case of a canonical state space. More precisely, fix $x_0 \in S_d^+$ and for $t \geq 0$, $y \in C(\mathbb{R}_+, \mathbb{R}^m)$ and $f \in B(S_d^+)$ define the LFF $\rho_t(\cdot, y)$ by

$$\rho_t(f,y) = \mathbb{E}\left[f(X_t)\exp\left(y_t^\top \bar{h}(X_t) - \int_0^t y_s^\top\,\mathrm{d}\bar{h}(X_s) - \int_0^t h(x_0)^\top \bar{h}(X_s)\,\mathrm{d}s\right)\right],$$

where $\bar{h} = \Gamma^{-2}h$, and the AFF $\bar{\pi}_t(\cdot, y)$ by (3.2). As in Remark 3.10 the LFF is obtained by linearizing the pathwise filtering functional (associated to the observation process $\Gamma^{-1}Y$ and observation function $\Gamma^{-1}h$) at x_0. Denoting by $h^\top$ the adjoint[6] of h and setting $\bar{y}_s = h^\top(\Gamma^{-2}y_s)$ and $\bar{x}_0 = h^\top(\Gamma^{-2}h(x_0))$ one rewrites $\rho_t(f,y)$ as

$$\rho_t(f,y) = \mathbb{E}\left[f(X_t)\exp\left(\langle \bar{y}_t, X_t\rangle_{S_d} - \int_0^t \langle \bar{y}_s, \mathrm{d}X_s\rangle_{S_d} - \int_0^t \langle \bar{x}_0, X_s\rangle_{S_d}\,\mathrm{d}s\right)\right]$$

and based on Section 3 one expects

$$\bar{\pi}_t(f,y) = \mathbb{E}_{\mathbb{Q}_x^{y,t}}[f(X_t)],$$

[6] By definition, this is the unique linear map $h^\top : \mathbb{R}^m \to S_d$ such that $h(x)^\top y = \langle x, h^\top(y)\rangle_{S_d}$ for all $y \in \mathbb{R}^m$.

where under $\mathbb{Q}_x^{y,t}$, X satisfies $X_0 = x$ and

$$dX_s = (n\Sigma^2 + H_s X_s + X_s H_s^\top)\mathrm{d}s + \sqrt{X_s}\mathrm{d}B_s\Sigma + \Sigma\mathrm{d}B_s^\top\sqrt{X_s}, \quad s \in (0,t]\,, \tag{6.2}$$

with $H_s = 2\Sigma^2(\Psi(s) - \bar{y}_s)$, B a $d \times d$ Brownian motion under $\mathbb{Q}_x^{y,t}$ and Ψ the solution to

$$\begin{aligned} -\partial_s\Psi(s) &= R(\Psi(s) - \bar{y}_s) - \bar{x}_0, \quad s \in [0,t) \\ \Psi(t) &= \bar{y}_t. \end{aligned} \tag{6.3}$$

In particular, (6.2) yields an ordinary differential equation for the approximate conditional mean $\hat{X}_t = \bar{\pi}_t(\mathrm{id}, y)$ at time t: Formally taking expectations in (6.2) one obtains $\hat{X}_0 = x$ and

$$\frac{\mathrm{d}\hat{X}_s}{\mathrm{d}s} = (n\Sigma^2 + H_s\hat{X}_s + \hat{X}_s H_s^\top), \quad s \in (0,t]. \tag{6.4}$$

6.3 Discussion

We now compare the two methods in an example. The following choices have been made: $h(x) := \mathrm{vech}(x)$ is the half-vectorization operator (which takes the elements of x in the lower triangular part and writes them in an m-dimensional column vector) and $m = \frac{1}{2}d(d+1)$. Denote by I_d the $d \times d$ identity matrix. We choose $d = 3$, $\Gamma = \Gamma_0 I_3$, $\Sigma = \sigma I_3$ and the parameter values as shown in the following summary:

$$\begin{aligned} \mathrm{d}X_t &= n\sigma^2 I_3\mathrm{d}t + \sigma X_t^{1/2}\mathrm{d}B_t + \sigma\mathrm{d}B_t^\top X_t^{1/2}, \quad X_0 = x_0 \\ \mathrm{d}Y_t &= \mathrm{vech}(X_t)\mathrm{d}t + \Gamma_0\mathrm{d}W_t, \quad Y_0 = 0 \\ (n, \sigma, x_0, \Gamma_0) &= (4, 0.04, \mathrm{diag}(0.75^2, 0.5^2, 0.25^2), 0.06). \end{aligned} \tag{6.5}$$

The filtering problem is discretized analogously to the case of a CIR process discussed in detail in Section 5. We choose $T = 1$ and equidistant time-points $t_i = iT/N$, $i = 0, \dots N$ with $N = 100$. (Exact) samples of $(X_{t_0}, X_{t_1}, \dots X_{t_N})$ can be generated as explained in Section 6.1 and a spline interpolation is used to generate a continuous observation path y from discrete measurements.

In this setting the conditional mean $\hat{x}_t$ (see (5.3) and (5.6)) is approximated by

- $\hat{x}_t^{(PF)}$ based on a bootstrap particle filter with N_p particles (as in [BC09a, Chapter 10]),
- $\hat{x}_t^{(AFF)} = \hat{X}_t$ in (6.4).

The computation time required to calculate $\hat{x}_t^{(PF)}$ on a standard laptop is enormous already for moderate N_p (e.g. for $N_p = 10^3, 10^4, 10^5$ it takes roughly 10 seconds, 1 minute, 10 minutes, respectively). On the other hand, in all these cases the approximation is very bad and so, in contrast to Section 5, here no benchmark is available. The two approximations are therefore compared based on their mean square error: we generate M sample paths of (6.5), calculate the

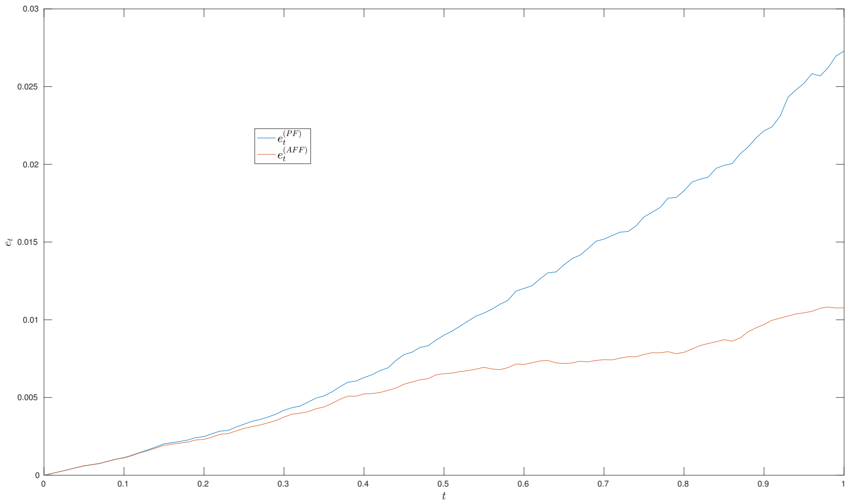


Figure II.4: Comparison of mean square error for conditional mean of particle filter and AFF.

approximate conditional mean with both methods and calculate the average at each time-point,

$$e_t^{(m)} = \frac{1}{M} \sum_{j=1}^{M} \|X_t^j - x_t^{(m),j}\|^2, \quad t = t_0, \ldots, t_N$$

for $m \in \{PF, AFF\}$. Here X^j is the j-th sample path of X, $x^{(m),j}$ is the approximate conditional mean (calculated using method m) associated to it and $\|u\|^2 := \langle u, u \rangle_{S_d}$ for $u \in S_d$. By the law of large numbers and the definition of $\hat{x}$, a smaller value of $e_{t_i}^{(m)}$ indicates that (on average) $\hat{x}_{t_i}$ and $x_{t_i}^{(m)}$ are closer.

Figure II.4 shows a plot of $(t_i, e_{t_i}^{(m)})$, $i = 0, \ldots, N$ for $m \in \{PF, AFF\}$, $N_p = 10^4$ and $M = 100$. For this number of particles the calculation of $\hat{x}^{(PF)}$ takes about 15 times longer than the calculation of $\hat{x}^{(AFF)}$ (on average). Nevertheless, the approximation quality of a bootstrap particle filter is considerably worse than that of the AFF, since the average mean-square error is significantly larger for longer time-periods, as shown in Figure II.4.

Chapter III

Deep Hedging

1 Introduction

The problem of pricing and hedging derivatives is crucial for risk-management in the financial securities industry. In idealized (frictionless, complete) market models mathematical finance provides a tractable solution (risk neutral pricing and delta-hedging) to this problem. Even though real markets are not frictionless (e.g. transaction costs are incurred), in practice hedging decisions are often still based on this idealized solution due to the lack of alternatives: in more realistic market models so far there has not been any general approach to pricing and hedging, which is also numerically feasible. A few examples illustrating this are provided in Section 1.1 below.

Building on recent theoretical insights (see e.g. [BGKP17]) and computational advances in the area of neural networks, this chapter studies hedging strategies built from deep neural networks. It turns out that these provide an efficient parametrization of the space of all hedging strategies and a generic hedging algorithm applicable to a wide range of market environments (different market models and frictions).

More concretely, we build on ideas from e.g. [Xu06], [IJS09], [FL00], [FS16] and use convex risk measures to define prices and optimal hedging strategies. To calculate these numerically, the strategies are approximated by deep neural networks. State-of-the-art machine learning optimization techniques (see [GBC16]) are then used to train these networks, yielding a close-to-optimal *deep hedge*. All of this is implemented in PYTHON using TensorFlow and a numerical study is carried out for the Heston model [Hes93], where trading is allowed in both stock and a variance swap. Experiments with proportional transaction costs show promising results and the approach is feasible also in a high-dimensional setting. The experiments are complemented by theoretical convergence results.

1.1 Related literature

There is a vast literature on hedging in market models with frictions. In addition to those discussed later on in the chapter ([HN89], [DPZ93],[WW97], [KMK15]) here we pick out only a small selection to illustrate the complex character of the problem. For example, [RS10] study a market in which trading a security has a (temporary) impact on its price. The price process is modeled by a one-dimensional Black-Scholes model. The optimal trading strategy can be obtained by solving a system of three coupled (non-linear) PDEs. In [BSV17] a more general tracking problem (covering the temporary price impact hedging problem) is carried out for a Bachelier model and a closed form solution (involving conditional expectations of a time integral over the optimal frictionless hedging strategy) is obtained for the strategy. [SSC95] prove that in a Black-Scholes market with proportional transaction costs, the cheapest superhedging price for a European call option is the spot price of the underlying. Thus, in the one-dimensional case the concept of super-replication is of little interest to practitioners. This is of course different if a larger number of hedging instruments is allowed, but then the problem becomes numerically intractable by classical methods.

1.2 Outline

This chapter is structured as follows. In Sections 2 and 3 we provide the theoretical framework for pricing and hedging using convex risk measures in discrete-time markets with frictions. Section 4 discusses hedging by neural network strategies and provides theoretical results explaining why the approach works. Finally, in Section 5 we present numerical experiments that illustrate the surprising feasibility and accuracy of the method.

2 Setting: Discrete time-market with transaction costs

Consider a discrete-time financial market with finite time horizon T and $n+1$ trading dates $0 = t_0 < t_1 < \ldots < t_n = T$. Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which all random variables are defined and a filtration $\mathbb{F} = (\mathcal{F}_{t_k})_{k=0,\ldots,n}$. $\mathcal{F}_{t_k}$ models the information available at t_k. Assume $\Omega = \{\omega_1, \ldots, \omega_N\}$, $\mathbb{P}(\{\omega_i\}) > 0$ for all i and set $\mathcal{X} := \{X : \Omega \to \mathbb{R}\}$.

The market contains d tradable assets, a risk-free bank account and a contingent claim (modeling a portfolio of derivatives or an exotic option) with maturity T. Prices of the d tradable assets are given by an $\mathbb{R}^d$-valued $\mathbb{F}$-adapted stochastic process $S = (S_{t_k})_{k=0,\ldots,n}$. For notational simplicity, we assume that all prices have been discounted and set interest rates to 0. The payoff of the contingent claim is an $\mathcal{F}_T$-measurable random variable Z.

At time 0 an agent sells the contingent claim at price p_0 and she wants to hedge this exposure at maturity, i.e. she aims at replicating its payoff Z at T by trading in S (and the bank account). A hedging strategy is thus any $\mathbb{R}^d$-valued $\mathbb{F}$-adapted stochastic process $\delta = (\delta_{t_k})_{k=0,\ldots,n-1}$. Let us denote by $\mathcal{H}$ the set of such processes. Here $\delta_{t_k}^{(i)}$ represents the number of units that the agent decides to hold in asset i at time t_k.

In a market without transaction costs the agent's wealth at time T is thus given by $-Z + p_0 + (\delta \cdot S)_T$, where

$$(\delta \cdot S)_T := \sum_{j=0}^{n-1} \delta_{t_j} \cdot (S_{t_{j+1}} - S_{t_j}).$$

However, we are interested in situations where e.g. $S^{(2)}$ is the price of a call option with underlying $S^{(1)}$ and so market frictions cannot be neglected. We assume that any trading activity causes costs as follows: if the agent decides to buy or sell Δ units of asset i at t_k, she will be charged a transaction cost of $c_i(|\Delta|, S_{t_k})$. Thus the total cost incurred up to maturity is

$$C_T(\delta) := \sum_{i=1}^{d} \left[c_i(|\delta_{t_0}^{(i)}|, S_{t_0}^{(i)}) + \sum_{j=1}^{n-1} c_i(|\delta_{t_j}^{(i)} - \delta_{t_{j-1}}^{(i)}|, S_{t_j}^{(i)}) + \bar{c}_i(|\delta_{t_{n-1}}^{(i)}|, S_{t_n}^{(i)}) \right]$$

and the agent's terminal portfolio value at T is

$$\mathrm{PL}_T(Z, p_0, \delta) := -Z + p_0 + (\delta \cdot S)_T - C_T(\delta). \tag{2.1}$$

Throughout, we assume that the non-negative functions $c_i \colon [0, \infty) \times \mathbb{R} \to [0, \infty)$ satisfy $c_i(0, \cdot) = 0$ and that for any $s \in \mathbb{R}$, $x \mapsto c_i(x, s)$ is upper semi-continuous. Furthermore, in the numerical examples we have assumed no transaction costs at maturity, i.e. $\bar{c}_i = 0$.

Example 2.1. This includes the following effects (and by summation also combinations of them):

- Proportional transaction costs: $c_i(x, s) = C_i x s$ for constants $C_i \geq 0$
- Fixed transaction costs: $c_i(x, s) = f_i \mathbb{1}_{\{x \geq \varepsilon\}}$ for constants $f_i \geq 0$ and $\varepsilon > 0$.
- Temporary market impact: $c_i(x, s) = \varepsilon_i x^2 s$ for constants $\varepsilon_i \geq 0$.

Remark 2.2. The assumption that Ω is finite is only essential for the numerical solution of the optimal hedging problem (from Section 4.4 onwards). Alternatively one could start with arbitrary Ω and discretize it for the numerical solution. If one imposed appropriate integrability conditions on Z and the elements of $\mathcal{H}$, the results prior to Section 4.4 would remain valid for general Ω.

3 Pricing and hedging using convex risk measures

While in idealized situations (Brownian market, continuous-time trading and no transaction costs) for any claim Z there exists a (unique) hedging strategy $\delta \in \mathcal{H}$ and a price $p_0 \in \mathbb{R}$ such that Z can be replicated perfectly, i.e. $\mathrm{PL}_T(Z, p_0, \delta) = 0$ holds $\mathbb{P}$-a.s., this is not true in the present setting. The agent therefore specifies an optimality criterion reflecting her risk-preferences and seeks to find a hedging strategy achieving the minimum. We focus here on pricing and hedging based on convex risk measures as studied e.g. in [Xu06],[IJS09]. See also [KS07] and further references therein for a dynamic setting.

Denote by $\rho\colon \mathcal{X} \to \mathbb{R}$ a convex risk measure as defined in [FS16]. Suppose that the agent wants to determine both an optimal hedging strategy and a price p_0. Setting

$$\pi(Z) := \inf_{\delta \in \mathcal{H}} \rho(-Z + (\delta \cdot S)_T - C_T(\delta)), \tag{3.1}$$

we define an optimal hedging strategy as a minimizer $\delta \in \mathcal{H}$ of (3.1). Recalling the interpretation of $\rho(X)$ as the minimal amount of capital that has to be added to the risky position X to make it acceptable for the risk measure ρ, $\pi(Z)$ is simply the minimal amount that the agent needs to charge in order to make her terminal position acceptable, if she hedges optimally. The *indifference price* $p(Z)$ is now defined as the amount of cash that she needs to charge in order to be indifferent between selling Z and not doing so, i.e. as the solution p_0 to $\pi(Z - p_0) = \pi(0)$. By cash-invariance this is equivalent to taking $p_0 = p(Z)$, where

$$p(Z) := \pi(Z) - \pi(0). \tag{3.2}$$

It is easily seen that without trading restrictions and transaction costs, this price coincides with the price of a replicating portfolio (if it exists):

Lemma 3.1. *Suppose $c_i = 0$ for $i = 1, \ldots, d$. If Z is attainable, i.e. there exists $\delta^* \in \mathcal{H}$ and $p \in \mathbb{R}$ such that $Z = p + (\delta^* \cdot S)_T$, then $p(Z) = p$.*

Proof. For any $\delta \in \mathcal{H}$, the assumptions and cash-invariance of ρ imply

$$\rho(-Z + (\delta \cdot S)_T - C_T(\delta)) = p + \rho(([\delta - \delta^*] \cdot S)_T).$$

Taking the infimum over $\delta \in \mathcal{H}$ on both sides and using $\mathcal{H} - \delta^* = \mathcal{H}$ one obtains

$$\pi(Z) = p + \inf_{\delta \in \mathcal{H}} \rho(([\delta - \delta^*] \cdot S)_T) = p + \pi(0).$$

□

Remark 3.2. The methodology developed in this chapter can also be applied to approximate optimal hedging strategies in a setting where the price p_0 is given exogenously: fix a loss function $l\colon \mathbb{R} \to [0, \infty)$. Suppose $p_0 > 0$ is given (e.g. the derivative has already been sold at this price without taking into account risk-management). The agent then wishes to minimize her loss at maturity, i.e. she defines an optimal hedging strategy as a minimizer to

$$\inf_{\delta \in \mathcal{H}} \mathbb{E}\left[l(-Z + p_0 + (\delta \cdot S)_T - C_T(\delta))\right]. \tag{3.3}$$

This problem, i.e. optimal hedging under a capital constraint, is closely related to taking for ρ a shortfall risk measure, see e.g. [FL00]. Similarly, one could treat temporary market impact models as e.g. [RS10] by considering

$$\inf_{\delta \in \mathcal{H}} \mathbb{E}\left[l(-Z + p_0 + (\delta \cdot S)_T)\right] + \mathbb{E}\left[C_T(\delta)\right]. \tag{3.4}$$

3.1 Example: Exponential Utility Indifference Price

The following Lemma shows that the present framework includes exponential utility indifference pricing as studied for example in [HN89], [DPZ93],[WW97] and [KMK15]. Recall that for the exponential utility function (defined for $x \in \mathbb{R}$ by $U(x) := -\exp(-\lambda x)$ for a risk-aversion parameter $\lambda > 0$) the indifference price $p_0(Z) \in \mathbb{R}$ of Z is defined by

$$\sup_{\delta \in \mathcal{H}} \mathbb{E}\left[U(p_0(Z) - Z + (\delta \cdot S)_T + C_T(\delta))\right] = \sup_{\delta \in \mathcal{H}} \mathbb{E}\left[U((\delta \cdot S)_T + C_T(\delta))\right].$$

In other words, if the seller charges a cash amount of $p_0(Z)$, sells Z and trades in the market, she obtains the same expected utility as by not not selling Z at all.

Lemma 3.3. *Define $p_0(Z)$ as above. Choose ρ as the entropic risk measure*

$$\rho(X) = \frac{1}{\lambda} \log \mathbb{E}[\exp(-\lambda X)], \tag{3.5}$$

and define $p(Z)$ by (3.2). *Then $p_0(Z) = p(Z)$.*

Proof. Using the special form of U, one may write the indifference price as

$$p_0(Z) = \frac{1}{\lambda} \log \left(\frac{\sup_{\delta \in \mathcal{H}} \mathbb{E}\left[U(-Z + (\delta \cdot S)_T + C_T(\delta))\right]}{\sup_{\delta \in \mathcal{H}} \mathbb{E}\left[U((\delta \cdot S)_T + C_T(\delta))\right]} \right)$$

and so the claim follows from (3.2) and (3.5). □

3.2 A special class of risk measures: Optimized certainty equivalents

Fix $l \colon \mathbb{R} \to \mathbb{R}$ continuous, non-decreasing and convex. Based on the *loss function* l one may define a convex risk measure ρ by setting

$$\rho(X) := \inf_{w \in \mathbb{R}} \left\{ w + \mathbb{E}[l(-X - w)] \right\}, \quad X \in \mathcal{X}. \tag{3.6}$$

Lemma 3.4. (3.6) *defines a convex risk measure.*

Proof. Let $X, Y \in \mathcal{X}$.

(i) [Monotonicity] Suppose $X \leq Y$. Since l is non-decreasing, for any $w \in \mathbb{R}$ one has $\mathbb{E}[l(-X - w)] \geq \mathbb{E}[l(-Y - w)]$ and thus $\rho(X) \geq \rho(Y)$.

(ii) [Cash invariance] For any $m \in \mathbb{R}$, (3.6) gives

$$\rho(X + m) = \inf_{w \in \mathbb{R}} \left\{ (w + m) - m + \mathbb{E}[l(-X - (w + m))] \right\} = -m + \rho(X).$$

(iii) [Convexity] Let $\lambda \in [0,1]$. Then convexity of l implies

$$\begin{aligned}
&\rho(\lambda X + (1-\lambda)Y) \\
&\quad = \inf_{w \in \mathbb{R}} \{w + \mathbb{E}[l(-\lambda X - (1-\lambda)Y - w)]\} \\
&\quad = \inf_{w_1, w_2 \in \mathbb{R}} \{\lambda w_1 + (1-\lambda) w_2 + \mathbb{E}[l(\lambda(-X - w_1) + (1-\lambda)(-Y - w_2))]\} \\
&\quad \le \inf_{w_1 \in \mathbb{R}} \inf_{w_2 \in \mathbb{R}} \{\lambda(w_1 + \mathbb{E}[l(-X - w_1)]) + (1-\lambda)(w_2 + \mathbb{E}[l(-Y - w_2)])\} \\
&\quad = \lambda \rho(X) + (1-\lambda)\rho(Y).
\end{aligned}$$

□

Taking $l(x) := -u(-x)$ $(x \in \mathbb{R})$ for a utility function $u \colon \mathbb{R} \to \mathbb{R}$, (3.6) coincides with the optimized certainty equivalent as defined (and studied in a lot more detail than here) in [BTT07].

Example 3.5. Fix $\lambda > 0$ and set $l(x) := \exp(\lambda x) - \frac{1+\log(\lambda)}{\lambda}$, $x \in \mathbb{R}$. Then the optimization problem in (3.6) can be solved explicitly and the minimizer w^* satisfies $e^{\lambda w^*} = \lambda \mathbb{E}[\exp(-\lambda X)]$. Inserting this into (3.6), one obtains the *entropic risk measure* defined in (3.5) above.

Example 3.6. Let $\alpha \in (0,1)$ and set $l(x) := \frac{1}{1-\alpha} \max(x, 0)$. The associated risk measure (3.6) is called *average value at risk at level* $1-\alpha$ (see [FS16, Definition 4.48, Proposition 4.51] with $\lambda := 1-\alpha$) or also *conditional value at risk* or *expected shortfall*.

Proposition 3.7. *Suppose S is a $\mathbb{P}$-martingale, ρ is defined as in* (3.6) *and π, p as in* (3.1), (3.2). *Then*

(i) $\pi(0) = \rho(0)$,

(ii) $p(Z) \ge \mathbb{E}[Z]$ *for any* $Z \in \mathcal{X}$.

Proof. Since $0 \in \mathcal{H}$ and $C_T(0) = 0$, one has $\pi(0) \le \rho(0)$ for any choice of risk measure ρ in (3.1). Under the present assumptions the converse inequality is also true: By first applying Jensen's inequality (recall that l is convex) and then using that S is a martingale, that $C_T(\delta) \ge 0$ for any $\delta \in \mathcal{H}$ and that l is non-decreasing, one obtains

$$\begin{aligned}
\pi(Z) &= \inf_{w \in \mathbb{R}} \inf_{\delta \in \mathcal{H}} \{w + \mathbb{E}[l(Z - (\delta \cdot S)_T + C_T(\delta) - w)]\} \\
&\ge \inf_{w \in \mathbb{R}} \inf_{\delta \in \mathcal{H}} \{w + l(\mathbb{E}[Z - (\delta \cdot S)_T + C_T(\delta) - w])\} \\
&\ge \inf_{w \in \mathbb{R}} \{w + l(\mathbb{E}[Z] - w)\} = \rho(-\mathbb{E}[Z]) = \mathbb{E}[Z] + \rho(0).
\end{aligned} \tag{3.7}$$

Inserting $Z = 0$ yields the converse inequality $\pi(0) \ge \rho(0)$ and thus (i). Combining (i), (3.2) and (3.7) then directly gives (ii). □

4 Approximating strategies by deep neural networks

The key idea that we pursue in this chapter is to approximate hedging strategies by neural networks. Before describing this approach in more detail we recall the definition and approximation properties of neural networks and prove some basic results on hedging strategies built from them. While these results show that the approach is theoretically well-founded, they are only one half of the answer to the question why we have used neural networks (and not some other parametric family of functions) to approximate hedging strategies. The other half is that numerically, optimal hedging strategies built from neural networks can be calculated very efficiently. This is explained first for the case of OCE risk measures. Finally, an extension to general risk measures is presented.

4.1 Universal approximation by neural networks

Let us first recall the definition of a (feed forward) neural network:

Definition 4.1. Let $L, N_0, N_1, \dots, N_L \in \mathbb{N}$ with $L \geq 2$, let $\bar{\rho}\colon \mathbb{R} \to \mathbb{R}$ and for any $l = 1, \dots, L$, let $W_l\colon \mathbb{R}^{N_{l-1}} \to \mathbb{R}^{N_l}$ an affine function. A function $F\colon \mathbb{R}^{N_0} \to \mathbb{R}^{N_L}$ defined as

$$F(x) = W_L \circ F_{L-1} \circ \cdots \circ F_1 \text{ with } F_l = \bar{\rho} \circ W_l \text{ for } l = 1, \dots, L-1$$

is called a (feed forward) neural network. Here the *activation function* $\bar{\rho}$ is applied componentwise. L denotes the number of layers, $N_1, \dots, N_{L-1}$ denote the dimensions of the hidden layers and N_0, N_L of the input and output layers, respectively. For any $l = 1, \dots, L$ the affine function W_l is given as $W_l(x) = A^l x + b^l$ for some $A^l \in \mathbb{R}^{N_l \times N_{l-1}}$ and $b^l \in \mathbb{R}^{N_l}$. For any $i = 1, \dots N_l, j = 1, \dots, N_{l-1}$ the number A^l_{ij} is interpreted as the weight of the edge connecting the node i of layer $l-1$ to node j of layer l. The number of non-zero weights of a network is the sum of the number of non-zero entries of the matrices A^l, $l = 1, \dots, L$ and vectors b^l, $l = 1, \dots, L$.

Denote by $\mathcal{NN}_{\bar{\rho}, d_0, d_1}$ the set of neural networks mapping from $\mathbb{R}^{d_0} \to \mathbb{R}^{d_1}$ and with activation function $\bar{\rho}$. The next result ([Hor91, Theorems 1 and 2]) illustrates that neural networks approximate multivariate functions arbitrarily well.

Theorem 4.2 (Universal approximation, [Hor91])**.** *Suppose $\bar{\rho}$ is bounded and non-constant. The following statements hold:*

- *For any finite measure μ on $(\mathbb{R}^{d_0}, \mathcal{B}(\mathbb{R}^{d_0}))$ and $1 \leq p < \infty$, the set $\mathcal{NN}_{\bar{\rho}, d_0, 1}$ is dense in $L^p(\mathbb{R}^{d_0}, \mu)$.*
- *If in addition $\bar{\rho} \in C(\mathbb{R})$, then $\mathcal{NN}_{\bar{\rho}, d_0, 1}$ is dense in $C(\mathbb{R}^{d_0})$ for the topology of uniform convergence on compact sets.*

Since each component of an $\mathbb{R}^{d_1}$-valued neural network is an $\mathbb{R}$-valued neural network, this result easily generalizes to $\mathcal{NN}_{\bar{\rho},d_0,d_1}$ with $d_1 > 1$, see also [Hor91]. A variety of other results with different assumptions on $\bar{\rho}$ or emphasis on approximation rates are available, see e.g. [BGKP17] for further references.

In what follows, fix a bounded and non-constant activation function $\bar{\rho}$ and denote by $\{\mathcal{NN}_{M,d_0,d_1}\}_{M\in\mathbb{N}}$ a sequence of subsets of $\mathcal{NN}_{\bar{\rho},d_0,d_1}$ with the following properties:

- $\mathcal{NN}_{M,d_0,d_1} \subset \mathcal{NN}_{M+1,d_0,d_1}$ for all $M \in \mathbb{N}$,
- $\bigcup_{M\in\mathbb{N}} \mathcal{NN}_{M,d_0,d_1} = \mathcal{NN}_{\bar{\rho},d_0,d_1}$,
- for any $M \in \mathbb{N}$, one has $\mathcal{NN}_{M,d_0,d_1} = \{F^\theta \colon \theta \in \Theta_{M,d_0,d_1}\}$ with $\Theta_{M,d_0,d_1} \subset \mathbb{R}^q$ for some $q \in \mathbb{N}$ (depending on M).

Remark 4.3. We have two classes of examples in mind: the first one is to take for $\mathcal{NN}_{M,d_0,d_1}$ the set of all neural networks in $\mathcal{NN}_{\bar{\rho},d_0,d_1}$ with an arbitrary number of layers and nodes, but at most M non-zero weights. The second one is to take for $\mathcal{NN}_{M,d_0,d_1}$ the set of all neural networks in $\mathcal{NN}_{\bar{\rho},d_0,d_1}$ with a *fixed architecture*, i.e. a fixed number of layers $L^{(M)}$ and fixed input and output dimensions for each layer. These are specified by d_0, d_1 and some non-decreasing sequences $\{L^{(M)}\}_{M\in\mathbb{N}}$ and $\{N_1^{(M)}\}_{M\in\mathbb{N}}, \ldots, \{N_{L^{(M)}-1}^{(M)}\}_{M\in\mathbb{N}}$. In both cases the set $\mathcal{NN}_{M,d_0,d_1}$ is parametrized by matrices A^l and vectors b^l.

4.2 Hedging strategy input parametrization

Suppose now the information at t_k is described by some $\mathbb{R}^p$-valued stochastic factor process X, i.e.

$$\mathcal{F}_{t_k} = \sigma(X_{t_0}, \ldots, X_{t_k}), \quad k = 0, \ldots, n. \tag{4.1}$$

The role of X is twofold: Firstly, as indicated by (4.1), it models the flow of information. Recall that S is $\mathbb{F}$-adapted, so X includes at least all information on S (e.g. $X = S$). However, X could also include information on additional quantities that are observed but not traded.

Secondly, X describes a parametrization of the input variables for hedging strategies: one could have $d = p$ and $X_{t_k} = \Phi_k(S_{t_k})$ for some $\Phi_k \colon \mathbb{R}^d \to \mathbb{R}^d$ invertible (so that X and S contain precisely the same information), but by expressing the (optimal) hedging strategy in terms of X instead of S one obtains a computationally simpler problem.

Example 4.4. Let $\sigma > 0$, $K > 0$, $d = 1$ and denote by N the cdf of a standard normal. If S follows a (discretely sampled) Black-Scholes model and Z is a European call option with strike K and maturity T, the hedging strategy

$$\delta_{t_k}^{BS} = N\left(\frac{1}{\sigma\sqrt{T - t_k}}\left[\log(S_{t_k}) - \log(K) + \frac{\sigma^2}{2}(T - t_k)\right]\right) \tag{4.2}$$

is close to optimal. Setting $X_{t_k} = \log(S_{t_k})$, one reads off from (4.2) that $\delta^{BS}_{t_k} = F^{(k)}(X_{t_k})$ for a neural network with one hidden layer (with $W_2(x) = x$, $\bar{\rho} = N$ and W_1 a suitable affine function). Thus, if one did not know the exact form of (4.2), it could nevertheless be learnt easily by the methodology developed in the present chapter, using a very simple neural network architecture. If one used $X_{t_k} = S_{t_k}$ instead, an exact representation by a neural network is not possible. The methodology of course still works, but a more complex network structure has to be used, since $x \mapsto \log(x)$ also needs to be approximated by a neural network.

4.3 Optimal hedging using deep neural networks

Motivated by the universal approximation results stated above, we now consider neural network hedging strategies: we define

$$\begin{aligned} \mathcal{H}_M &= \{(\delta_{t_k})_{k=0,\dots,n-1} \in \mathcal{H} \,:\, \delta_{t_k} = F^k(X_{t_0},\dots,X_{t_k})\,, F^k \in \mathcal{NN}_{M,p(k+1),d}\} \\ &= \{(\delta^\theta_{t_k})_{k=0,\dots,n-1} \in \mathcal{H} \,:\, \delta^\theta_{t_k} = F^{\theta_k}(X_{t_0},\dots,X_{t_k})\,, \theta_k \in \Theta_{M,p(k+1),d}\} \end{aligned} \tag{4.3}$$

and replace the set $\mathcal{H}$ in (3.1) by $\mathcal{H}_M \subset \mathcal{H}$, i.e. we aim at calculating

$$\begin{aligned} \pi^M(Z) &:= \inf_{\delta \in \mathcal{H}_M} \rho(-Z + (\delta \cdot S)_T - C_T(\delta)) \\ &= \inf_{\theta \in \Theta_M} \rho(-Z + (\delta^\theta \cdot S)_T - C_T(\delta^\theta)), \end{aligned} \tag{4.4}$$

where $\Theta_M = \prod_{k=0}^{n-1} \Theta_{M,p(k+1),d}$. Thus, the infinite-dimensional problem of finding an optimal hedging strategy is reduced to the finite-dimensional problem of finding optimal parameters.

Remark 4.5. In problems (3.3) and (3.4) one would analogously replace $\mathcal{H}$ by $\mathcal{H}_M$.

Remark 4.6. In certain situations (e.g. if S is an $(\mathbb{F},\mathbb{P})$-Markov process and $Z = g(S_T)$ for $g\colon \mathbb{R}^d \to \mathbb{R}$) one knows (or expects) that the optimal strategy in (3.1) is of the form $\delta_{t_k} = F_k(X_{t_k}, \delta_{t_{k-1}})$ for some $F_k\colon \mathbb{R}^{p+d} \to \mathbb{R}^d$ measurable, for $k = 0,\dots,n-1$. Then one would replace $\mathcal{H}_M$ in (4.4) by the smaller set of *recurrent neural network* strategies defined by (with the convention $\delta_{-1} := 0$)

$$\begin{aligned} \mathcal{H}_M &= \{(\delta_{t_k})_{k=0,\dots,n-1} \in \mathcal{H} \,:\, \delta_{t_k} = F^k(X_{t_k}, \delta_{t_{k-1}})\,, F^k \in \mathcal{NN}_{M,p+d,d}\} \\ &= \{(\delta^\theta_{t_k})_{k=0,\dots,n-1} \in \mathcal{H} \,:\, \delta^\theta_{t_k} = F^{\theta_k}(X_{t_k}, \delta^\theta_{t_{k-1}})\,, \theta_k \in \Theta_{M,p+d,d}\}. \end{aligned} \tag{4.5}$$

The set Θ_M in (4.4) is then given by $\Theta_M := (\Theta_{M,p+d,d})^n$.

The next proposition shows that thanks to (4.1) and the universal approximation theorem, strategies in $\mathcal{H}$ are approximated arbitrarily well by strategies in $\mathcal{H}_M$. Consequently, the neural network price $\pi^M(Z) - \pi^M(0)$ converges to the exact price $p(Z)$.

Proposition 4.7. *Define* $\mathcal{H}_M$ *as in* (4.3) *and* π^M *as in* (4.4). *Then*

$$\lim_{M \to \infty} \pi^M(Z) = \pi(Z).$$

Proof. Since $\mathcal{H}_M \subset \mathcal{H}_{M+1} \subset \mathcal{H}$, for all $M \in \mathbb{N}$ it holds that $\pi^M(Z) \geq \pi^{M+1}(Z) \geq \pi(Z)$. Thus it suffices to show that for any $\varepsilon > 0$ there exists $M \in \mathbb{N}$ such that $\pi^M(Z) \leq \pi(Z) + \varepsilon$. By definition, there exists $\delta \in \mathcal{H}$ such that

$$\rho(-Z + (\delta \cdot S)_T - C_T(\delta)) \leq \pi(Z) + \frac{\varepsilon}{2}. \tag{4.6}$$

Since δ is $\mathbb{F}$-adapted and by (4.1), for each $k = 0, \dots, n-1$ there exists a measurable function $F^k \colon \mathbb{R}^{p(k+1)} \to \mathbb{R}^d$ such that $\delta_{t_k} = F^k(X_{t_0}, \dots, X_{t_k})$. Since Ω is finite, δ_{t_k} is bounded and so $F_i^k \in L^1(\mathbb{R}^{p(k+1)}, \mu)$ for any $i = 1, \dots d$, where μ is the law of $(X_{t_0}, \dots, X_{t_k})$ under $\mathbb{P}$. Thus one may use Theorem 4.2 to find $F_i^{k,n} \in \mathcal{NN}_{\bar{\rho}, p(k+1), 1}$ such that $\delta_{t_k}^{(i),n} := F_i^{k,n}(X_{t_0}, \dots, X_{t_k})$ converges to $F_i^k(X_{t_0}, \dots, X_{t_k})$ in $L^1(\mathbb{P})$ as $n \to \infty$. By passing to a suitable subsequence, this convergence holds $\mathbb{P}$-a.s. (and simultaneously for all i, k) and so, by assumption on Ω, for all $\omega \in \Omega$. Thus for any fixed $\omega \in \Omega$ one may combine this with upper semi-continuity of $c_i(\cdot, s)$ (for any $s \geq 0$, $i = 1, \dots, d$) to find $n_0(\omega) \in \mathbb{N}$ such that for all $n \geq n_0(\omega)$ one has

$$-(\delta^n \cdot S)_T(\omega) + C_T(\delta^n)(\omega) \leq -(\delta \cdot S)_T(\omega) + C_T(\delta)(\omega) + \frac{\varepsilon}{2}.$$

Combining this with monotonicity of ρ and then using cash-invariance of ρ and (4.6) one obtains for all $n \geq \max_{i=1,\dots,N} n_0(\omega_i)$ that

$$\rho(-Z + (\delta^n \cdot S)_T - C_T(\delta^n)) \leq \rho(-Z + (\delta \cdot S)_T - C_T(\delta) - \frac{\varepsilon}{2}) \leq \pi(Z) + \varepsilon. \tag{4.7}$$

Since $\delta^n \in \mathcal{H}_M$ for all M large enough, one obtains $\pi^M(Z) \leq \pi(Z) + \varepsilon$ by (4.4) and (4.7), as desired. □

4.4 Numerical solution for OCE-risk measures

While Theorem 4.2 and Proposition 4.7 give a theoretical justification for using hedging strategies built from neural networks, we now turn to computational considerations: how can we calculate a (close-to) optimal parameter $\theta \in \Theta_M$ for (4.4)?

To explain the key ideas we focus on the case when ρ is an OCE risk measure (see (3.6)), a class of general risk measures is treated below.

Inserting the definition of ρ, see (3.6), into (4.4), the optimization problem can be rewritten as

$$\pi^M(Z) = \inf_{\bar{\theta} \in \Theta_M} \inf_{w \in \mathbb{R}} \left\{ w + \mathbb{E}[l(Z - (\delta^{\bar{\theta}} \cdot S)_T + C_T(\delta^{\bar{\theta}}) - w)] \right\} = \inf_{\theta \in \Theta} J(\theta),$$

where $\Theta = \mathbb{R} \times \Theta_M$ and for $\theta = (w, \bar{\theta}) \in \Theta$,

$$J(\theta) := w + \mathbb{E}[l(Z - (\delta^{\bar{\theta}} \cdot S)_T + C_T(\delta^{\bar{\theta}}) - w)]. \tag{4.8}$$

Generally, to find a local minimum of a differentiable function J, one may use a *gradient descent* algorithm: Starting with an initial guess $\theta^{(0)}$, one iteratively defines

$$\theta^{(j+1)} = \theta^{(j)} - \eta_j \nabla J_j(\theta^{(j)}), \tag{4.9}$$

for some (small) $\eta_j > 0$, $j \in \mathbb{N}$ and with $J_j = J$. Under suitable assumptions on J and the sequence $\{\eta_j\}_{j\in\mathbb{N}}$, $\theta^{(j)}$ converges to a local minimum of J as $j \to \infty$. Of course, the success and feasibility of this algorithm crucially depends on two points: Firstly, can one avoid finding a local minimum instead of a global one? Secondly, can ∇J be calculated efficiently?

One of the key insights of deep learning is that for cost functions J built based on neural networks both of these problems can be dealt with simultaneously by using a variant of *stochastic gradient descent* and the *(error) backpropagation* algorithm. What this means in our context is that in each step j the expectation in (4.8) (which is in fact a weighted sum over all elements of the finite, but potentially very large sample space Ω) is replaced by an expectation over a randomly (uniformly) chosen subset of Ω of size $N_{\text{batch}} \ll N$, so that J_j used in the update (4.9) is now given as

$$J_j(\theta) = w + \sum_{m=1}^{N_{\text{batch}}} l(Z(\omega_m^{(j)}) - (\delta^{\bar{\theta}} \cdot S)_T(\omega_m^{(j)}) + C_T(\delta^{\bar{\theta}})(\omega_m^{(j)}) - w) \frac{N}{N_{\text{batch}}} \mathbb{P}(\{\omega_m^{(j)}\})$$

for some $\omega_1^{(j)}, \ldots, \omega_{N_{\text{batch}}}^{(j)} \in \Omega$. This is the simplest form of the (minibatch) stochastic gradient algorithm. Not only does it make the gradient computation a lot more efficient (or possible at all, if N is large), but it also avoids getting stuck in local minima: even if $\theta^{(j)}$ arrives at a local minimum at some j, it moves on afterwards (due to the randomness in the gradient). In order to calculate the gradient of J_j for each of the terms in the sum, one may now rely on the compositional structure of neural networks. If l, $\{c_i\}_{i=1,\ldots,d}$ and $\bar{\rho}$ are differentiable and the derivatives are available in closed form, then one may use the chain rule to calculate the gradient of $F^{\bar{\theta}_k}$ with respect to θ analytically and the same holds for the gradient of J_j. Furthermore, these analytical expressions can be evaluated very efficiently using the so called backpropagation algorithm (see subsequent section).

While this certainly answers the second question posed above (efficiency), the first one (local minima) is only partially resolved, as there is no general result guaranteeing convergence to the global minimum in a reasonable amount of time. However, it is common belief that for sufficiently large neural networks, it is possible to arrive at a sufficiently low value of the cost function in a reasonable amount of time, see [GBC16, Chapter 8].

Finally, note that for the experiments in Section 5 below we have used Adam, a more refined version of the stochastic gradient algorithm, as introduced in [KB15] and also discussed in [GBC16, Chapter 8.5.3].

Remark 4.8. In the experiments in Section 5 below, the functions l, $\{c_i\}_{i=1,\ldots,d}$ and $\bar{\rho}$ are continuous, but have only piecewise continuous derivatives. Nevertheless, similar techniques can be applied.

4.5 Extension to general risk measures

As explained in Section 4.4, for OCE risk measures the optimal hedging problem (4.4) is amenable to deep learning optimization techniques (i.e. variants of stochastic gradient descent) via (4.8). The key ingredient for this is that the objective J satisfies

(ML1) the gradient of J decomposes into a sum over the samples, i.e. $\nabla_\theta J(\theta) = \sum_{m=1}^{N} \nabla_\theta J(\theta, \omega_m)$ and

(ML2) $\nabla_\theta J(\theta, \omega_m)$ can be calculated efficiently for each m, i.e. using backpropagation.

The goal of the present section is to show that for a general class of convex risk measures (including all coherent ones) one can approximate (3.1) by a minimax problem over neural networks and that the objective functional of this approximate problem also has these two key properties, making it amenable to deep learning optimization techniques.

Denote by $\mathcal{P}$ the set of probability measures on $(\Omega, \mathcal{F})$. The following result serves as a starting point:

Theorem 4.9 (Robust representation of convex risk measures)**.** *Suppose* $\rho\colon \mathcal{X} \to \mathbb{R}$ *is a convex risk measure. Then* ρ *can be written as*

$$\rho(X) = \max_{\mathbb{Q}\in\mathcal{P}} \left(\mathbb{E}_{\mathbb{Q}}[-X] - \alpha(\mathbb{Q})\right), \quad X \in \mathcal{X}, \tag{4.10}$$

where $\alpha(\mathbb{Q}) := \sup_{X\in\mathcal{X}} \left(\mathbb{E}_{\mathbb{Q}}[-X] - \rho(X)\right)$.

Proof. Since for Ω finite the set of probability measures $\mathcal{P}$ coincides with the set of finitely additive, normalized set functions (appearing in [FS16, Theorem 4.16]), the present statement follows directly from the cited theorem and [FS16, Remark 4.17]. □

The function $\alpha\colon \mathcal{P} \to \mathbb{R}$ is called the (minimal) penalty function of the risk measure ρ.

Since Ω is finite, $\mathcal{P}$ can be identified with the standard $N-1$ simplex in $\mathbb{R}^N$ and so (4.10) is an optimization over $\mathbb{R}^N$. However, N is very large in our context and so the representation (4.10) is of little use for numerical calculations. The next result shows that $\rho(X)$ can be approximated by an optimization problem over a lower-dimensional space. To state it, let us define the set $\mathcal{L} \subset \mathcal{X}$ of log-likelihoods by

$$\mathcal{L} := \{f \in \mathcal{X} \ : \ \mathbb{E}[\exp(f)] = 1\},$$

define $\bar{\alpha}\colon \mathcal{L} \to \mathbb{R}$ by $\bar{\alpha}(f) = \alpha(\exp(f)\mathrm{d}\mathbb{P})$ for any $f \in \mathcal{L}$ and write $\mathcal{P}_{eq}$ for the set of probability measures on $(\Omega, \mathcal{F})$, which are equivalent to $\mathbb{P}$. Furthermore, one may view $\bar{X} = (X_{t_0}, \ldots, X_{t_n})$ as a map $\Omega \to \mathbb{R}^{p(n+1)}$.

Theorem 4.10. *Suppose*

(i) $\alpha(\mathbb{Q}) < \infty$ *for some* $\mathbb{Q} \in \mathcal{P}_{eq}$,

(ii) $\bar{\alpha}$ *is continuous,*

(iii) $\mathcal{F} = \mathcal{F}_T$.

Then for any $X \in \mathcal{X}$*,* $\rho(X) = \lim_{M\to\infty} \rho^M(X)$*, where*

$$\rho^M(X) := \sup_{\substack{\theta\in\Theta_{M,p(n+1),1} \\ \mathbb{E}[\exp(F^\theta\circ\bar{X})]=1}} \left(\mathbb{E}[-X\exp(F^\theta\circ\bar{X})] - \bar{\alpha}(F^\theta\circ\bar{X})\right). \tag{4.11}$$

Proof. We proceed in two steps. In a first step we show that for any $X \in \mathcal{X}$ one may write

$$\rho(X) = \sup_{\substack{\bar{f}\in\mathcal{M} \\ \mathbb{E}[\exp(\bar{f}\circ\bar{X})]=1}} \left(\mathbb{E}[-X\exp(\bar{f}\circ\bar{X})] - \bar{\alpha}(\bar{f}\circ\bar{X})\right), \tag{4.12}$$

where $\mathcal{M}$ denotes the set of measurable functions mapping from $\mathbb{R}^{p(n+1)} \to \mathbb{R}$. In the second step we rely on (4.12) to prove the statement.

Step 1: Since $\mathbb{P}(\{\omega_i\}) > 0$ for all i, $\mathcal{X}$ coincides with $L^\infty(\Omega,\mathcal{F},\mathbb{P})$ and ρ is law-invariant. Thus by (i) and [FS16, Theorem 4.43] one may write

$$\rho(X) = \sup_{\mathbb{Q}\in\mathcal{P}_{eq}} \left(\mathbb{E}_\mathbb{Q}[-X] - \alpha(\mathbb{Q})\right), \quad X \in \mathcal{X}. \tag{4.13}$$

Note that $\mathcal{P}_{eq}$ may be written in terms of $\mathcal{L}$ as

$$\mathcal{P}_{eq} = \{\exp(f)\mathrm{d}\mathbb{P} \ : \ f \in \mathcal{L}\}. \tag{4.14}$$

Furthermore, using (iii) one obtains

$$\mathcal{X} = \{\bar{f}\circ\bar{X} \ : \ \bar{f} \in \mathcal{M}\}. \tag{4.15}$$

Combining (4.13), (4.14) and the definition of $\bar{\alpha}$ one obtains

$$\rho(X) = \sup_{f\in\mathcal{L}} \left(\mathbb{E}[-X\exp(f)] - \bar{\alpha}(f)\right),$$

which can be rewritten as (4.12) by using (4.15).

Step 2: Note that one may also write (4.11) as

$$\rho^M(X) = \sup_{\substack{f\in\mathcal{NN}_{M,p(n+1),1} \\ \mathbb{E}[\exp(f\circ\bar{X})]=1}} \left(\mathbb{E}[-X\exp(f\circ\bar{X})] - \bar{\alpha}(f\circ\bar{X})\right). \tag{4.16}$$

Combining (4.16) with (4.12) and using $\mathcal{NN}_{M,p(n+1),1} \subset \mathcal{NN}_{M+1,p(n+1),1} \subset \mathcal{M}$, one obtains that $\rho^M(X) \leq \rho^{M+1}(X) \leq \rho(X)$ for all $M \in \mathbb{N}$. Thus it suffices to show that for any $\varepsilon > 0$ there exists $M \in \mathbb{N}$ such that $\rho^M(X) \geq \rho(X) - \varepsilon$.

By (4.12), for any $\varepsilon > 0$ one finds $\bar{f} \in \mathcal{M}$ such that

$$\mathbb{E}[\exp(\bar{f}\circ\bar{X})] = 1, \tag{4.17}$$

$$\rho(X) - \frac{\varepsilon}{2} \leq \mathbb{E}[-X\exp(\bar{f}\circ\bar{X})] - \bar{\alpha}(\bar{f}\circ\bar{X}). \tag{4.18}$$

Precisely as in the proof of Proposition 4.7, one may use Theorem 4.2 to find $f^{(n)} \in \mathcal{NN}_{\bar{\rho},p(n+1),1}$ such that $\mathbb{P}$-a.s., $f^{(n)}\circ\bar{X}$ converges to $\bar{f}\circ\bar{X}$ as $n \to \infty$. Combining this with (4.17), one obtains that for all n large enough, $c_n := \log(\mathbb{E}[\exp(f^{(n)} \circ \bar{X})])$ is well-defined and that $\bar{f}^{(n)} \circ \bar{X}$ also converges $\mathbb{P}$-a.s. to $\bar{f} \circ \bar{X}$, as $n \to \infty$, where $\bar{f}^{(n)} := f^{(n)} - c_n$. Using this, (4.18) and assumption (ii), for some (in fact all) $n \in \mathbb{N}$ large enough one obtains

$$\rho(X) - \varepsilon \leq \mathbb{E}[-X \exp(\bar{f}^{(n)} \circ \bar{X})] - \bar{\alpha}(\bar{f}^{(n)} \circ \bar{X}). \tag{4.19}$$

From $\mathcal{NN}_{\bar{\rho},p(n+1),1} - c_n = \mathcal{NN}_{\bar{\rho},p(n+1),1}$ and from the choice of $\mathcal{NN}_{M,p(n+1),1}$, one has $\bar{f}^{(n)} \in \mathcal{NN}_{M,p(n+1),1}$ for M large enough. By combining this with (4.19) and the choice of c_n one obtains

$$\rho(X) - \varepsilon \leq \rho^M(X),$$

as desired. □

Combining (4.4) and (4.11), one thus approximates (3.1) by solving

$$\inf_{\theta_0 \in \Theta_M} \sup_{\theta_1 \in \Theta_{M,p(n+1),1}} J(\theta), \tag{4.20}$$

where $\theta = (\theta_0, \theta_1)$,

$$J(\theta) := \mathbb{E}\left[-\mathrm{PL}(Z, 0, \delta^{\theta_0}) \exp(F^{\theta_1} \circ \bar{X})\right] - \bar{\alpha}(F^{\theta_1} \circ \bar{X}) - \lambda_0(\mathbb{E}[\exp(F^{\theta_1} \circ \bar{X})] - 1)$$

and λ_0 is a Lagrange multiplier.

We conclude this section by arguing that the objective J in (4.20) indeed satisfies (ML1) and (ML2). This is standard (c.f. Section 4.4) for all terms in the sum except for $\bar{\alpha}(F^{\theta_1} \circ \bar{X})$ and so we only consider this term.

Recall that Ω is finite and consists of N elements, thus $\mathcal{X} = \{X \colon \Omega \to \mathbb{R}\}$ can be identified with $\mathbb{R}^N$. As for standard backpropagation the compositional structure can be used for efficient computation:

Proposition 4.11. *Suppose $\bar{\alpha}$ can be extended to $\bar{\alpha} \colon \mathcal{X} \to \mathbb{R}$ continuously differentiable, $\bar{\rho}$ is continuously differentiable and $\mathcal{NN}_{M,p(n+1),1}$ is the set of neural networks with a fixed architecture (see Remark 4.3). Then $J(\theta_1) := \bar{\alpha}(F^{\theta_1} \circ \bar{X})$, $\theta_1 \in \Theta_{M,p(n+1),1}$ is continuously differentiable and satisfies (ML1).*

Proof. Note that $F = F^{\theta_1}$ is parametrized by the matrices A^l and vectors b^l, $l = 1, \ldots, L$, and that one may consider all partial derivatives separately. Given $\bar{\alpha} \colon \mathcal{X} \to \mathbb{R}$ and $\nabla\bar{\alpha} \colon \mathcal{X} \to \mathcal{X}$, one thus aims at calculating $\partial_{A^l_{i,j}} \bar{\alpha}(F \circ \bar{X})$ and $\partial_{b^l_i} \bar{\alpha}(F \circ \bar{X})$ for $l = 1, \ldots, L, i = 1, \ldots, N_l, j = 1, \ldots, N_{l-1}$. This can be done by the chain rule: For $\theta \in \{A^l_{i,j}, b^l_i\}$, one has

$$\partial_\theta \bar{\alpha}(F \circ \bar{X}) = \sum_{m=1}^{N} \nabla\bar{\alpha}(F \circ \bar{X})(\omega_m) \partial_\theta F(\bar{X}(\omega_m))$$

and in particular (ML1) holds. □

Furthermore, in the notation of the proof, for any $m = 1, \dots N$ the derivative $\partial_\theta F(\bar{X}(\omega_m))$ can be calculated using standard backpropagation algorithm (preceded by a forward iteration) and so (ML2) holds as well. For the reader's convenience we state it here: One sets $x^0 = \bar{X}(\omega_m)$, iteratively calculates $x^l := F_l(x^{l-1})$ for $l = 1, \dots, L-1$ and $x^L := W_L(x^{L-1})$. Then (this is the backward pass) one sets $J^L := A^L$ and calculates iteratively $J^l = J^{l+1} dF_l(x^{l-1})$ for $l = L-1, \dots, 1$, where

$$dF_l(x^{l-1}) = \mathrm{diag}(\bar{\rho}'(W_l x^{l-1}))A^l.$$

From this one may use again the chain rule to obtain for any $l = 1, \dots L, i = 1, \dots, N_l, j = 1, \dots, N_{l-1}$ the derivatives of F with respect to the parameters as

$$\partial_{A^l_{i,j}} F(\bar{X}(\omega_m)) = J^{l+1}_i \bar{\rho}'((W_l x^{l-1})_i) x^{l-1}_j$$
$$\partial_{b^l_i} F(\bar{X}(\omega_m)) = J^{l+1}_i \bar{\rho}'((W_l x^{l-1})_i).$$

5 Numerical experiments and results

After having introduced the optimal hedging problem (3.1) in Section 3 and described in Section 4 how one may numerically approximate the solution by (4.4) using neural networks, we now turn to numerical experiments to illustrate the feasibility of the approach. We start by explaining in Section 5.1 the modeling choices in detail. The remainder of this section will then be devoted to examining the following three questions:

- Section 5.2: How does neural network hedging (with different risk-preferences) compare to the benchmark in a Heston model without transaction costs?
- Section 5.3: What is the effect of proportional transaction costs on the exponential utility indifference price?
- Section 5.4: Is the numerical method scalable to higher dimensions?

5.1 Setting and Implementation

For the results presented here we have chosen a time horizon of 30 trading days with daily rebalancing. Thus, $T = 30/365$, $n = 30$ and the trading dates are $t_i = i/365$, $i = 0, \dots, n$. As explained in Section 4 and Remark 4.6, the number of units $\delta_{t_i} \in \mathbb{R}^d$ that the agent decides to hold in each of the instruments at t_i is parametrized by a recurrent neural network: we set $\delta^\theta_{t_k} = F^{\theta_k}(X_{t_k}, \delta^\theta_{t_{k-1}})$ where F^{θ_k} is a feed forward neural network with two hidden layers and $X_{t_k} = \Phi(S_{t_0}, \dots, S_{t_k})$ for some $\Phi \colon \mathbb{R}^{(k+1)d} \to \mathbb{R}^d$ specified below. More precisely, in the notation of Definition 4.1, F^{θ_k} is a neural network with $L = 3$, $N_0 = 2d$, $N_1 = N_2 = d + 15$, $N_3 = d$ and the activation function is always chosen as $\bar{\rho}(x) = \max(x, 0)$. The weight matrices and biases are the parameters to be optimized in (4.4). Note that these are different for each k.

Having made these choices, the algorithm outlined in Section 4 can now be used for approximate pricing and hedging *in any market situation*: given sample trajectories of the hedging instruments $S(\omega_m)$, samples of the payoff $Z(\omega_m)$ and associated weights $\mathbb{P}(\{\omega_m\})$ for $m = 1, \dots, N$ (on a finite probability space $\Omega = \{\omega_1, \dots, \omega_N\}$), for any choice of transaction cost structure $\{c_i\}_{i=1,\dots,d}$ and any risk measure ρ one may now use the algorithm outlined in Section 4 to calculate close-to optimal hedging strategies and approximate prices. Of course, for a path-dependent derivative with payoff $Z = G(S_0, \dots, S_T)$ with $G\colon (\mathbb{R}^d)^{n+1} \to \mathbb{R}$ one obtains samples of the payoff by simply evaluating G on the sample trajectories of S.

Different risk measures ρ, transaction cost functions $\{c_i\}_{i=1,\dots,d}$ and payoffs Z will be used in the examples and so these are described separately in each of the subsequent sections. To illustrate the feasibility of the algorithm and have a benchmark at hand for comparison (at least in the absence of transaction costs), we have chosen to generate the sample paths of S from a standard stochastic volatility model under a risk-neutral measure $\mathbb{P}$. Thus in most of the examples below, the process S follows (a discretization of) a Heston model, see the beginning of Section 5.2 below. But we stress again that, as explained above, the algorithm is *model independent* in the sense that no information about the Heston model is used except for the (weighted) samples of the price and variance process.

The algorithm has been implemented in Python, using TensorFlow to build and train the neural networks. To allow for a larger learning rate, the technique of batch normalization (see [IS15] and [GBC16, Chapter 8.7.1]) is used in each layer of each network right before applying the activation function. The network parameters are initialized randomly (drawn from uniform and normal distribution). For network training the Adam algorithm (see [KB15], [GBC16, Chapter 8.5.3]) with a learning rate of 0.005 and a batch size of 256 has been used. Finally, the model hedge for the benchmark in Section 5.2 has been calculated using Quantlib.

5.2 Benchmark: No transaction costs

As a first example, we consider hedging without transaction costs in a Heston model. In this example the risk measure ρ is chosen as the average value at risk (also called conditional value at risk or expected shortfall), defined for any random variable X by

$$\rho(X) := \frac{1}{1-\alpha} \int_0^{1-\alpha} \mathrm{VaR}_\gamma(X) \mathrm{d}\gamma \tag{5.1}$$

for some $\alpha \in [0, 1)$, where $\mathrm{VaR}_\gamma(X) := \inf\{m \in \mathbb{R} : \mathbb{P}(X < -m) \leq \gamma\}$. An alternative representation of ρ of type (3.6) is discussed in Example 3.6. We refer to [FS16, Section 4.4] for further details. Note that different levels of α correspond to different levels of risk-aversion, ranging from risk-neutral for α close to 0 to very risk-averse for α close to 1. The limiting cases are $\rho(X) = -\mathbb{E}[X]$ for $\alpha = 0$ and $\lim_{\alpha \uparrow 1} \rho(X) = -\mathrm{essinf}(X)$, see [FS16, p.234 and Remark 4.50].

A brief reminder on the Heston model

Recall that a Heston model is specified by the stochastic differential equations

$$\begin{aligned} \mathrm{d}S_t^{(1)} &= \sqrt{V_t}S_t^{(1)}\mathrm{d}B_t, \quad \text{for } t>0 \text{ and } S_0^{(1)} = s_0 \\ \mathrm{d}V_t &= \alpha(b-V_t)\mathrm{d}t + \sigma\sqrt{V_t}\mathrm{d}W_t, \quad \text{for } t>0 \text{ and } V_0 = v_0, \end{aligned} \tag{5.2}$$

where B and W are one-dimensional Brownian motions (under a probability measure $\mathbb{Q}$) with correlation $\rho \in [-1,1]$ and α, b, σ, v_0 and s_0 are positive constants. Below we have chosen $\alpha = 1$, $b = 0.04$, $\rho = -0.7$, $\sigma = 2$, $v_0 = 0.04$ and $s_0 = 100$, reflecting a typical situation in an equity market.

Here $S^{(1)}$ is the price of a liquidly tradeable asset and V is the (stochastic) variance process of $S^{(1)}$, modeled by a Cox-Ingersoll-Ross (CIR) process. V itself is not tradable directly, but only through options on variance. In our framework this is modeled by an idealized variance swap with maturity T, i.e. we set $\mathcal{F}_t^H := \sigma((S_s^{(1)}, V_s) \,:\, s \in [0,t])$,

$$S_t^{(2)} := \mathbb{E}_{\mathbb{Q}}\left[\int_0^T V_s\,\mathrm{d}s \,\middle|\, \mathcal{F}_t^H\right], \quad t \in [0,T], \tag{5.3}$$

and consider $(S^{(1)}, S^{(2)})$ as the prices of liquidly tradeable assets. A standard calculation[1] shows that (5.3) is given as

$$S_t^{(2)} = \int_0^t V_s\,\mathrm{d}s + L(t, V_t) \tag{5.4}$$

where

$$L(t,v) = \frac{v-b}{\alpha}(1 - e^{-\alpha(T-t)}) + b(T-t).$$

Consider now a European option with payoff $g(S_T^{(1)})$ at T for some $g\colon \mathbb{R} \to \mathbb{R}$. Its price (under $\mathbb{Q}$) at $t \in [0,T]$ is given as $H_t := \mathbb{E}_{\mathbb{Q}}[g(S_T^{(1)})|\mathcal{F}_t^H]$. By the Markov property of $(S^{(1)}, V)$, one may write the option price at t as $H_t = u(t, S_t^{(1)}, V_t)$ for some $u\colon [0,T] \times [0,\infty)^2 \to \mathbb{R}$. Assuming that u is sufficiently smooth, one may apply Itô's formula to H and use (5.4) to obtain

$$g(S_T^{(1)}) = p + \int_0^T \delta_t^{(1)}\,\mathrm{d}S_t^{(1)} + \int_0^T \delta_t^{(2)}\,\mathrm{d}S_t^{(2)} \tag{5.5}$$

where $p = \mathbb{E}_{\mathbb{Q}}[g(S_T^{(1)})]$ and

$$\delta_t^{(1)} := \partial_s u(t, S_t^{(1)}, V_t) \text{ and } \delta_t^{(2)} := \frac{\partial_v u(t, S_t^{(1)}, V_t)}{\partial_v L(t, V_t)}. \tag{5.6}$$

Thus, if continuous-time trading was possible, (5.5) shows that the option payoff can be replicated perfectly by trading in $(S^{(1)}, S^{(2)})$ according to the strategy (5.6).

[1] For example, one may use that $(\log(S^{(1)}), V)$ is an affine process to see that the conditional expectation in (5.3) can be taken only with respect to $\sigma(V_t, s \in [0,t])$. This conditional expectation can then be calculated by using the SDE for V or by directly inserting the expression from e.g. [Duf01, Section 3].

Remark 5.1. The strategy (5.6) depends on V_t. Although not observable directly, an estimate can be obtained by estimating $\int_0^t V_s \,\mathrm{d}s$ and solving (5.4) for V_t.

Setting: Discretized Heston model

In addition to the setting explained in detail in Section 5.1, here we set $d = 2$, consider no transaction costs (i.e. $c_1 = c_2 = 0$) and generate sample trajectories of the price process of the hedging instruments from a discretely sampled Heston model. Thus, $S = (S_{t_0}, \ldots, S_{t_n})$ and for any $k = 0, \ldots, n$, $S_{t_k} = (S^{(1)}_{t_k}, S^{(2)}_{t_k})$ is given by (5.2) and (5.4) under $\mathbb{Q}$. The sample paths of S are generated by (exact) sampling from the transition density of the CIR process (see [Gla04, Section 3.4]) and then using the (simplified) Brodie-Kaya scheme (see [AJK10] and [BK06]).[2] Generating independent samples of S according to this scheme can now be viewed as sampling from a uniform distribution on a (huge) finite probability space Ω.[3] Thus, in the notation of Section 5.1 one has $\mathbb{P}(\omega_m) = 1/N$ for all $m = 1, \ldots, N$ with each $S(\omega_m)$ corresponding to a sample of the Heston model generated as explained above.

If continuous-time trading was possible, any European option could be replicated perfectly by following the strategy (5.6). However, in the present setup the hedging portfolio can only be adjusted at discrete time-points. Nevertheless one may choose $\delta^H_{t_k} := (\delta^{(1)}_{t_k}, \delta^{(2)}_{t_k})$ for $k = 0 \ldots n-1$ with $\delta^{(i)}$ defined by (5.6) and charge the risk-neutral price p. This will be referred to as the model-delta hedging strategy (or simply model hedge) and serves as a benchmark.

Finally, in order to compare the neural network strategies to this benchmark, the network input is chosen as $X_{t_k} = (\log(S^{(1)}_{t_k}), V_{t_k})$. One could also replace V_{t_k} by $S^{(2)}_{t_k}$ instead. The network structure at time-step t_k is illustrated in Figure III.1.

Results

We now compare the model hedge δ^H to the deep hedging strategies δ^θ corresponding to different risk-preferences, captured by different levels of α in the average value at risk (5.1).

As a first example, consider a European call option, i.e. $Z = (S^{(1)}_T - K)^+$ with $K = s_0$. Following the methodology outlined in Section 5.1, we calculate a (close-to) optimal parameter θ for (4.4) and denote by δ^θ and p^θ the (close-to) optimal hedging strategy and value of (4.4), respectively. By definition of the indifference price (3.2), the approximation property Proposition 4.7, Proposition 3.7 and $\rho(0) = 0$, p^θ is an approximation to the indifference price $p(Z)$. As an out-of-sample test, one can then simulate another set of sample trajectories (here 10^6) and evaluate the terminal hedging errors $p + Z + (\delta^H \cdot S)_T$ (model hedge) and $p^\theta + Z + (\delta^\theta \cdot S)_T$ (CVar) on each of them. In fact, since the risk-adjusted price p^θ is higher than the risk-neutral price $p = 1.69$ (as shown in Proposition 3.7(ii)),

[2] This corresponds to replacing V in the SDE for $S^{(1)}$ in (5.2) by a piecewise constant process and the integral in (5.4) by a sum.

[3] To be more precise, one replaces the normal distributions appearing in the simulation scheme for S by (arbitrarily fine) discrete distributions.

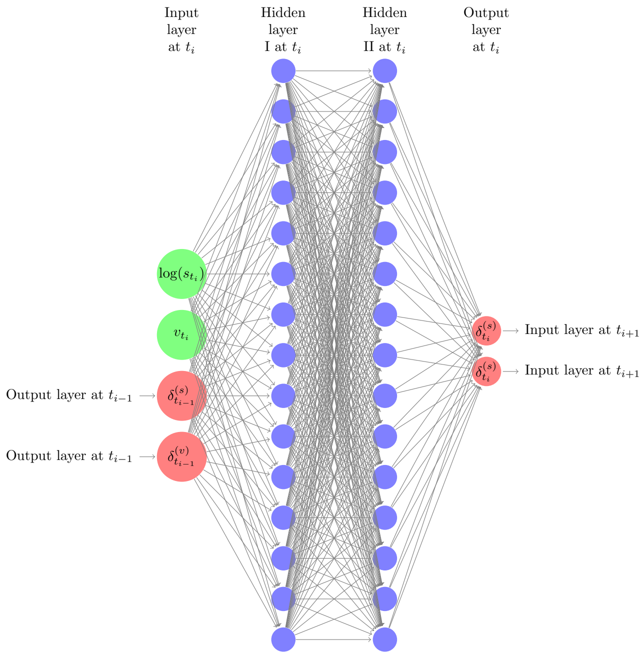


Figure III.1: Recurrent network structure

for (CVar) we have evaluated $p + Z + (\delta^\theta \cdot S)_T$, i.e. the hedging error from using the optimal strategy associated to ρ, but only charging the risk-neutral price p. This is shown in a histogram in Figure III.2 for $\alpha = 0.5$, yielding a risk-adjusted price $p^\theta = 1.94$. As one can see, the hedging performance of δ^H and δ^θ is very similar. In particular

- for this choice of risk-preferences (ρ as in (5.1) with $\alpha = 0.5$) the optimal strategy in (3.1) is close to the model hedge δ^H,
- the neural network strategy δ^θ is able to approximate very well the optimal strategy in (3.1).

This is also illustrated by Figure III.3, where the strategies δ^θ_t and δ^H_t at a fixed time-point t are plotted conditional on $(S^{(1)}_t, V_t) = (s, v)$ on a grid of values for (s, v). To make this last comparison fully sensible instead of the recurrent network structure $\delta^\theta_{t_k} = F^{\theta_k}(X_{t_k}, \delta^\theta_{t_{k-1}})$ here a simpler structure $\delta^\theta_{t_k} = F^{\theta_k}(X_{t_k})$ is used. The hedging performance for this simpler structure is, however, very similar, see Figure III.4. Of course, this is also expected from (5.6).[4]

A more extreme case is shown in Figure III.6, where instead of the model hedge the 99%-CVar criterion is used, i.e. $\alpha = 0.99$. This results in a significantly higher risk-adjusted price $p^\theta = 3.49$. If both the 50% and 99%-CVar optimal strategies are used, but only the risk-neutral price is charged (see Figure III.7) one can

[4] For non-zero transaction costs this is not true anymore, i.e. the recurrent network structure is needed. For example, Figure III.5 is generated for precisely the same parameters as Figure III.4, except that $\alpha = 0.99$ and proportional transaction costs are incurred, i.e. $c_i(x, s) = \varepsilon x s$ with $\varepsilon = 0.01$.

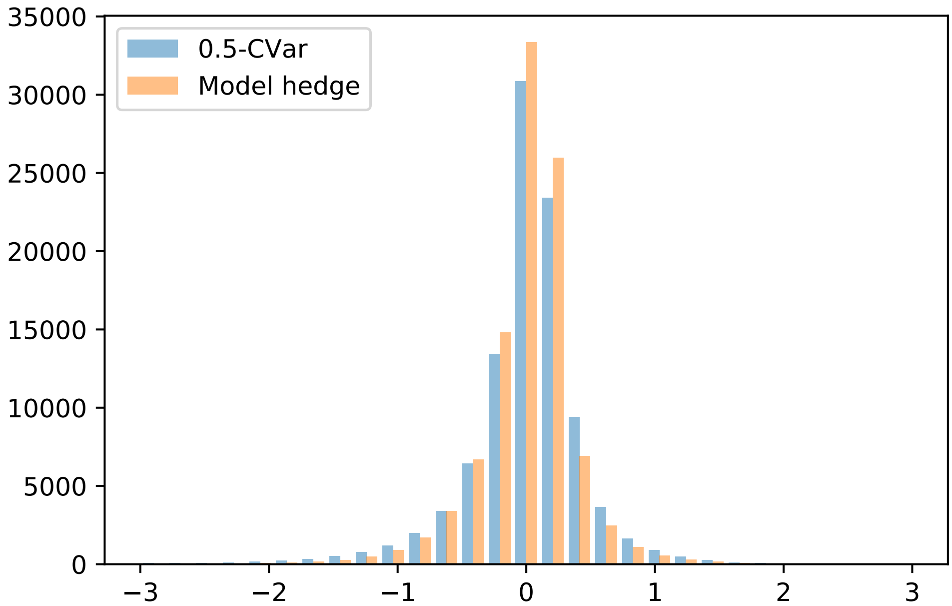


Figure III.2: Comparison of model hedge and deep hedge associated to 50%-expected shortfall criterion.

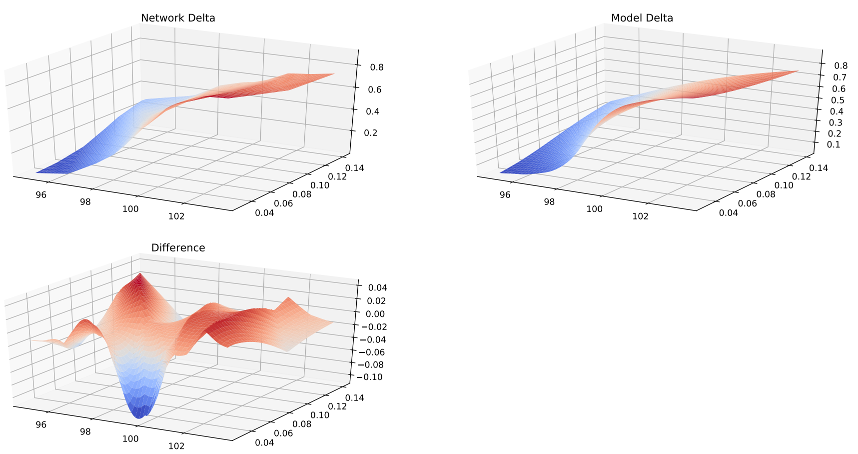


Figure III.3: $\delta_t^{H,(1)}$ and neural network approximation as a function of (s_t, v_t) for $t = 15$ days

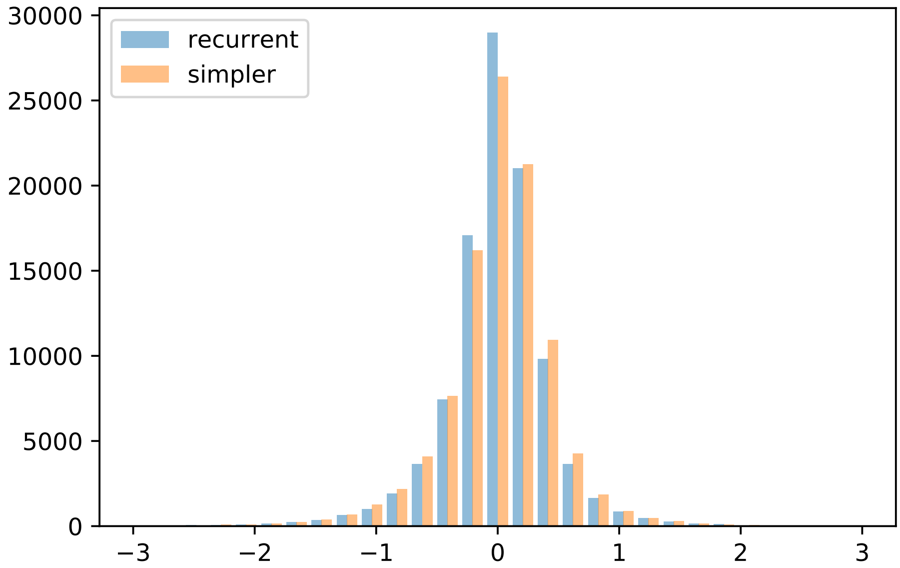
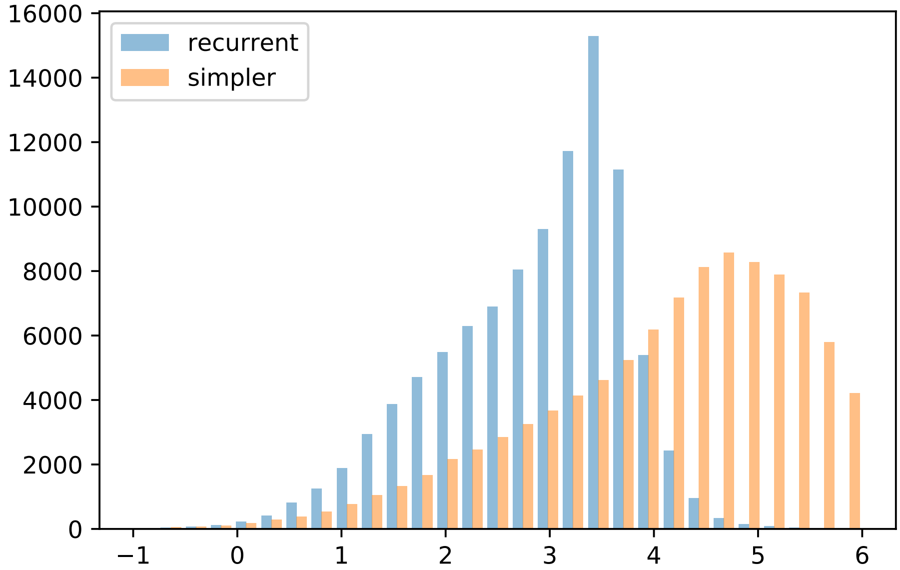


Figure III.4: Comparison of recurrent and simpler network structure (no transaction costs).



	Mean Loss	Price	Realized CVar
recurrent	0.0018	5.5137	-0.0022
simpler	0.0022	6.7446	-0.0

Figure III.5: Network architecture matters: Comparison of recurrent and simpler network structure (with transaction costs and 99%-CVar criterion).

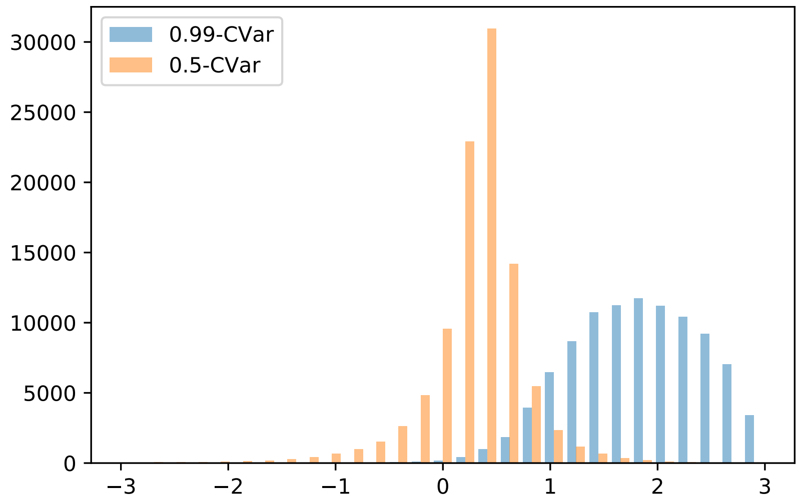
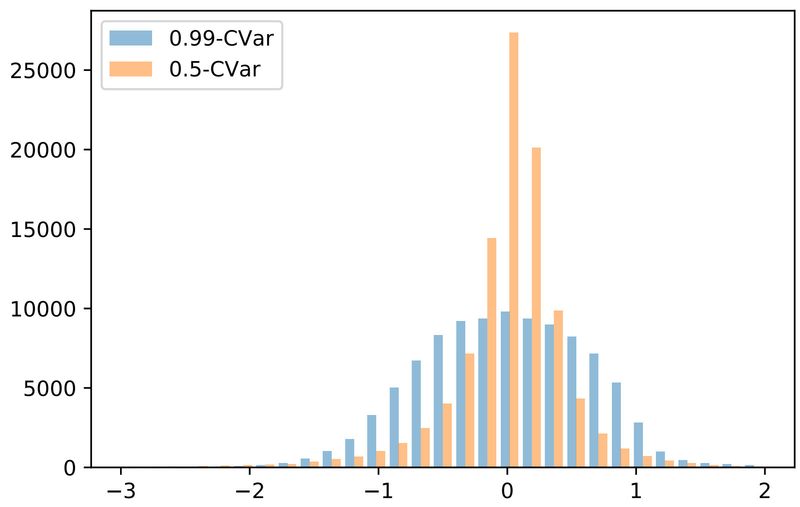


Figure III.6: Comparison of 99%-CVar and 50%-CVar optimiality criterion.



	Mean Loss	Realized 0.5-CVar	Realized 0.99-CVar
0.99-CVar	0.2635	0.527	1.8034
0.5-CVar	0.1514	0.2531	2.3631

Figure III.7: Comparison of 99%-CVar and 50%-CVar optimiality criterion, normalized to risk-neutral price.

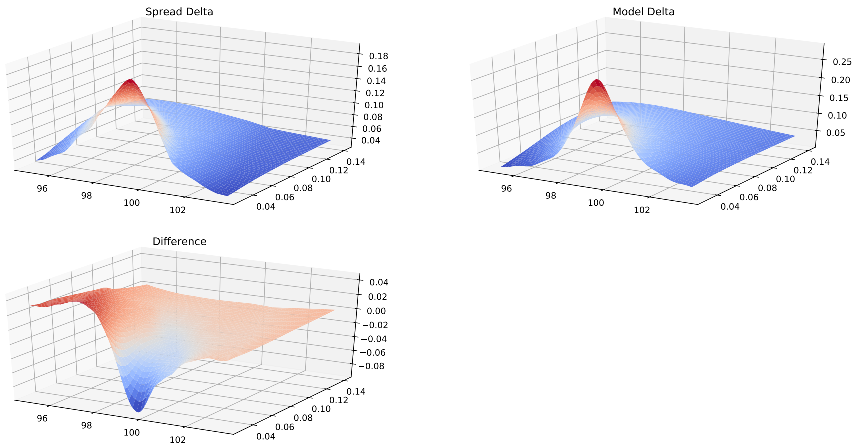


Figure III.8: Call spread $\delta_t^{H,(1)}$ and neural network approximation as a function of (s_t, v_t) for $t = 15$ days

clearly see the risk preferences: the 50%-CVar strategy is more centered at 0 and also has a smaller mean hedging error, but the 99%-expected shortfall strategy yields smaller extreme losses (c.f. also the realized 99%-CVar *loss* value realized on the test sample, shown in the table below Figure III.7).

To further illustrate the implications of risk-preferences on hedging, as a last example we consider selling a call-spread, i.e. $Z = [(S_T^{(1)} - K_1)^+ - (S_T^{(1)} - K_2)^+]/(K_2 - K_1)$ for $K_1 < K_2$. Here we have chosen $K_1 = s_0$, $K_2 = 101$. Proceeding as above, we compare the model hedge to the more risk-averse hedging strategies associated to $\alpha = 0.95$ and $\alpha = 0.99$. The strategies (on a grid of values for spot and variance) are shown in Figures III.8 and III.9. The model hedge would again correspond to $\alpha = 0.5$. As one can see for higher levels of risk-aversion, the strategy flattens. From a practical perspective, this precisely corresponds to a barrier shift, i.e. a more risk-averse hedge for a call spread with strikes K_1 and K_2 actually aims at hedging a spread with strikes $\tilde{K}_1$ and K_2 for $\tilde{K}_1 < K_1$.

5.3 Price asymptotics under proportional transaction costs

In Section 5.2 we have seen that in a market without transaction costs, deep hedging is able to recover the model hedge and can be used to calculate risk-adjusted optimal hedging strategies.

The goal of this section is to illustrate the power of the methodology by numerically calculating the indifference price (3.2) in a multi-asset market with transaction costs.

So far, this has been regarded a highly challenging problem, see e.g. the introduction of [KMK15]. For example, calculating the exponential utility indifference price for a call option in a Black-Scholes model involves solving a multidimensional nonlinear free boundary problem, see e.g. [HN89], [DPZ93]. Motivated by this [WW97] have studied asymptotically optimal strategies and price asymptot-

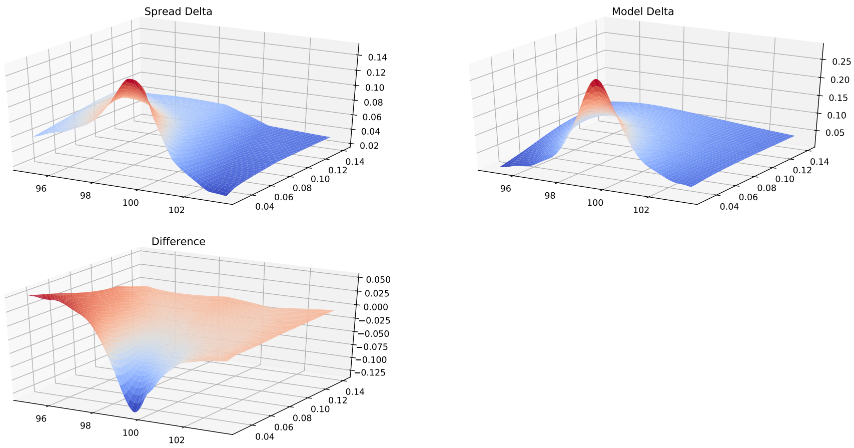


Figure III.9: Call spread $\delta_t^{H,(1)}$ and neural network approximation as a function of (s_t, v_t) for $t = 15$ days

ics for small proportional transaction costs, i.e. for $c_i(y,s) = \varepsilon y s$ and as $\varepsilon \downarrow 0$. One of the results in the asymptotic analysis is that

$$p_\varepsilon - p_0 = O(\varepsilon^{2/3}), \qquad \text{as } \varepsilon \downarrow 0, \tag{5.7}$$

where $p_\varepsilon = p_\varepsilon(Z)$ is the utility indifference price of Z associated to transaction costs of size ε. In fact (5.7) is true in more general one-dimensional models, see [KMK15], and the rate $2/3$ also emerges in a variety of related problems with proportional transaction costs, see e.g. [Rog04], [MKRS17] and the references therein.

Here we numerically verify (5.7) using the deep hedging algorithm, first for a Black-Scholes model (for which (5.7) is known to hold) and then for a Heston model (with $d = 2$ hedging instruments). For this latter case (or any other model with $d > 1$) there have been neither numerical nor theoretical results on (5.7) previously in the literature.

Black-Scholes model

Consider first $d = 1$ and $S_t = s_0 \exp(-t\sigma^2/2 + \sigma W_t)$, where $\sigma > 0$ and W is a one-dimensional Brownian motion. We choose $\sigma = 0.2$, $s_0 = 100$ and use the explicit form of S to generate sample trajectories. Setting $X_{t_k} = \log(S_{t_k})$ and proceeding precisely as in the Heston case (see Sections 5.1 and 5.2), we may use the deep hedging algorithm to calculate the exponential utility indifference price p_ε for different values of ε. Recall that we choose proportional transaction costs $c_1(x,s) = \varepsilon x s$ and ρ is the entropic risk measure (3.5) (see Lemma 3.3). For the numerical example we take $\lambda = 1$ and $Z = (S_T - K)^+$ with $K = s_0$ and we calculate p_{ε_i} for $\varepsilon_i = 2^{-i+5}$, $i = 1, \ldots, 5$.

Figure III.10 shows the pairs $(\log(\varepsilon_i), \log(p_{\varepsilon_i} - p_0))$ (in red) and the closest (in squared distance) straight line with slope $2/3$ (in blue). Thus, in this range of ε the relation $\log(p_\varepsilon - p_0) = 2/3 \log(\varepsilon) + C$ for some $C \in \mathbb{R}$ indeed holds true and hence also (5.7).

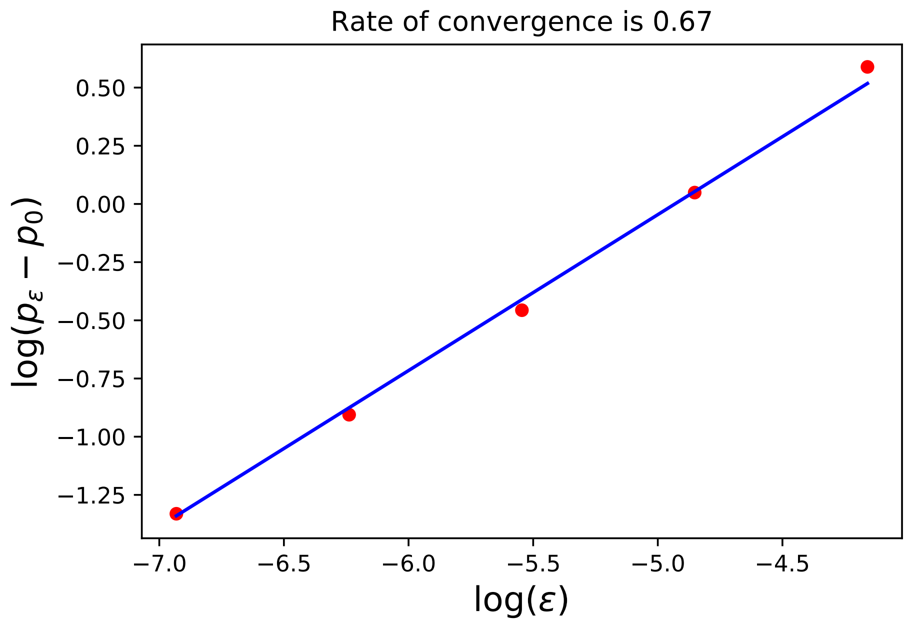


Figure III.10: Black-Scholes model price asymptotics.

Note that trading is only possible at discrete time-points and so the indifference price and the risk-neutral price do not coincide. Since (5.7) is a result for continuous-time trading (where $p = p_0$), we have compared to the risk-neutral price p here (thus neglecting the discrete-time friction in p_ε for $\varepsilon > 0$).

Heston model

We now consider a Heston model with two hedging instruments, i.e. $d = 2$ and the setting is precisely as in Section 5.2, except that here ρ is chosen as (3.5) and $c_1(x, s) = c_2(x, s) = \varepsilon xs$. Choosing $\lambda = 1$, $Z = (S_T^{(1)} - K)^+$ and ε_i as in the Black-Scholes case above, one can again calculate the exponential utility indifference prices and show the difference to p_0 in a log-log plot (see above) in a graph. These are shown as red dots in Figure III.11. Here the blue line in Figure III.11 is the regression line, i.e. the least squares fit of the red dots. The rate is very close to $2/3$ and so it appears that the relation (5.7) also holds in this case.

5.4 High-dimensional example

As a last example consider a model built from 5 separate Heston models, i.e. $d = 10$ and $(S^{(h)}, S^{(h+1)})$ is the price process of spot and variance swap in a Heston model (specified by (5.2) and (5.4)) for $h = 1, \ldots, 5$. To have a benchmark at hand the 5 models are assumed independent and each of them has parameters as specified in Section 5.2. This choice is of course no restriction for the algorithm and is only made for convenience. The payoff is a sum of call options on each of the underlyings, i.e. $Z = \sum_{h=1}^{5} Z_h$ with $Z_h = (S_T^{(2h-1)} - K)^+$ and $K =$

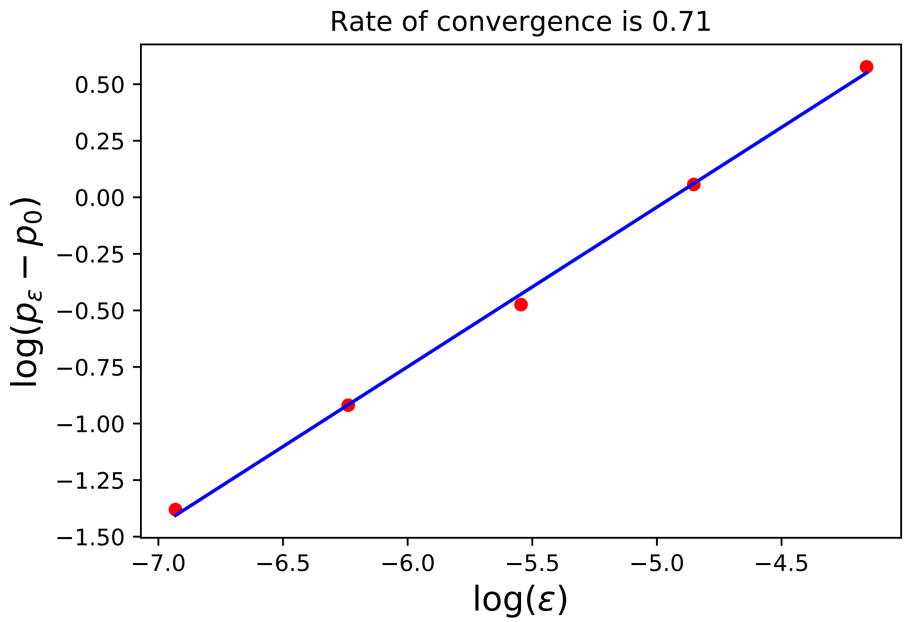


Figure III.11: Heston model price asymptotics

$s_0 = 100$. In a market with continuous-time trading and no transaction costs, Z can be replicated perfectly by trading according to strategy (5.6) in each of the models. In particular, this strategy is decoupled, i.e. the optimal holdings in $(S^{(h)}, S^{(h+1)})$ only depend on $(S^{(h)}, S^{(h+1)})$. While in the present setup trading is only possible at discrete time steps and so the strategy optimizing (3.1) leads to a non-deterministic terminal hedging error (2.1), by independence one still expects that the optimal strategy is decoupled as above, at least for certain classes of risk measures. To see this most prominently, here we consider *variance optimal hedging*: the objective is chosen as (3.3) for $l(x) = x^2$ and $p_0 = 5p$, where $p = \mathbb{E}[Z_1]$.

Let $\delta \in \mathcal{H}$ and write $\delta^{(2h-1:2h)} := (\delta^{(2h-1)}, \delta^{(2h)})$ for $h = 1, \ldots, 5$ (and analogously for S). If δ is decoupled, i.e. such that $\delta^{(2h-1:2h)}$ is independent of $S^{(2j-1:2j)}$ for $j \neq h$, then by independence and since S is a martingale one has

$$\mathbb{E}\left[(-Z + p_0 + (\delta \cdot S)_T)^2\right] = \sum_{h=1}^{5} \mathrm{Var}\left(-Z_i + (\delta^{(2h-1:2h)} \cdot S^{(2h-1:2h)})_T\right). \qquad (5.8)$$

By building δ from the (discrete-time) variance optimal strategies for each of the 5 models, one sees from (5.8) that the minimal value of (3.3) over *all* $\delta \in \mathcal{H}$ is at most 5 times the minimal value of (3.3) associated to a single Heston model. This consideration serves as a guideline for assessing the approximation quality of the neural network strategy.

To assess the scalability of the algorithm, we now calculate the close-to-optimal neural network hedging strategy associated to (3.3) in both instances (i.e. for $n_H = 5$ models and for a single one, $n_H = 1$) and compare the results. Unless specified otherwise, the parameters are as in Section 5.1. Since for $n_H = 5$

we are actually solving 5 problems at once, we allow for a network with more hidden nodes by taking $N_1 = N_2 = 12n_H$. We then train both networks for a fixed number of iterations (here 2×10^5) and measure the performance in terms of both training time and realized loss (evaluated on a test set of $n_H \times 10^5$ sample paths): the training times on a standard Lenovo X1 Carbon laptop are 5.75 and 2.1 hours for $n_H = 5$ and $n_H = 1$, respectively and the realized losses are 1.13 and 0.20. In view of the considerations above, this indicates that the approximation quality is roughly the same for both instances (and close-to-optimal).

While far from a systematic study, this last example nevertheless demonstrates the potential of the algorithm for high-dimensional hedging problems.

Chapter IV

Skorokhod Embedding Problem for Lévy Processes

1 Introduction

Consider a one-dimensional Brownian motion L and denote by $\mathbb{F}$ the filtration generated by L. The classical Skorokhod embedding problem is as follows: given a probability distribution μ_1, find an $\mathbb{F}$-stopping time τ such that $L_\tau \sim \mu_1$ and $\mathbb{E}[\tau] < \infty$. The last condition is imposed to exclude non-meaningful solutions, see e.g. [RW00b, Remark 51.7]. The problem was originally formulated and solved by Skorokhod [Sko61, Sko65]. Since then a variety of solutions with different properties and additional inherent structure have been found, see the comprehensive survey [Obł04]. For further background we refer to the introduction of this thesis, Section 1.1 below and e.g. [DGPR17], [BCH17b], [Hob11].

In the present chapter we study the Skorokhod embedding problem (SEP) for a Lévy process L (which includes the case of a Brownian motion) with non-deterministic initial distribution μ_0. Under some regularity assumptions on μ_0, μ_1 (but none on L), we provide conditions on μ_0, μ_1 and L which are necessary for a solution to exist and, assuming these, give an explicit solution. Our solution is new also in the special case when L is a Brownian motion and it is the first solution that is non-randomized and covers *all* Lévy processes (and not only e.g. those which are transient or admit local times).

In more detail, our approach is to choose an interpolation $(\mu_t)_{t\in[0,1]}$ of μ_0 and μ_1 and construct a process with marginal distributions $(\mu_t)_{t\in[0,1]}$ by a time-change of L, which then also gives rise to a stopping time solving the SEP. Given the choice of interpolation, the function σ determining the time-change is found by inverting the Fokker-Planck equation. This can be seen as calibrating a local volatility model (as in [Dup94], [CGMY04]) driven by L. However, since σ is determined through our choice of interpolation, here we cannot simply assume regularity (and positivity) of σ and existence of the associated process, but we need to derive it from assumptions on $(\mu_t)_{t\in[0,1]}$. Even for sufficiently regular μ_0 and μ_1, this turns out to be delicate for general L and leads to a σ that is time-dependent, possibly unbounded and may exhibit zeros. Proving that the desired process exists and indeed has marginal distributions $(\mu_t)_{t\in[0,1]}$ thus requires results on time-changes and uniqueness of Fokker-Planck equations that have not been available in the literature. These are developed in the general framework of martingale problems of [EK86] in the appendix of the present chapter.

1.1 Related literature

An overview of classical results on the Skorkohod embedding problem can be found in the comprehensive survey [Obł04]. We mention in particular the solution due to [Bas83], which has served as an inspiration. The stopping time in [Bas83] is constructed from a random time-change associated to a suitable martingale. The elegance of the solution in [Bas83] appears to be specific to the case of Brownian motion, see e.g. the introduction of [DGPR17].

For more general Markov processes, there have been less results in the literature. Bertoin and Le Jan [BL92] consider $\mu_0 = \delta_0$ and deal with necessary and sufficient conditions for the solvability of the Skorokhod embedding problem for a large class of Markov processes, including Lévy processes that admit local times. Sufficient conditions for different types of Lévy processes and $\mu_0 = \delta_0$ were also obtained in [Mon72] and [OP09]. Namely, Monroe [Mon72] adresses symmetric α-stable Lévy processes with $\alpha \in (1,2]$ and Obłój and Pistorius [OP09] the case of spectrally negative Lévy processes. In a more abstract setting Falkner and Fitzsimmons [FF91] provide even necessary and sufficient conditions for general but transient Markov processes, which cover only partially the class of Lévy processes. For a relaxed version of the SEP (allowing for randomized stopping times, i.e. allowing for stopping times which are measurable w.r.t. a larger filtration than the natural filtration generated by the underlying Markov process) Rost [Ros71] shows necessary and sufficient conditions for general Markov processes. A discussion about differences between randomized and non-randomized solutions to the SEP can be found for instance in [FF91].

Recent motivation to deal with various versions of the classical Skorokhod embedding problem stems from its applications in mathematical finance starting with the seminal work of Hobson [Hob98], where model-independent pricing bounds and hedging techniques for lookback options were studied by means of Skorokhod embedding. The link between robust financial mathematics and the classical SEP was utilized by many authors to determine robust price bounds for exotic options, see [Hob11] for a more detailed introduction to this area. More recently, additional interest in the Skorokhod embedding problem was also caused by new applications in game theory (e.g. [SS13, FH16]) and in numerical analysis (e.g. [GMO15, AKU16]).

1.2 Outline

The chapter is structured as follows. In Section 2 we formulate the problem and state our main result, Theorem 2.1. Section 3 contains preliminary results on the Poisson equation for Lévy processes and the proof of Theorem 2.1. The proof relies on results on random time-changes and uniqueness for Fokker-Planck equations, which are developed in Appendices A, B and C.

2 Main result

In this section we state and discuss Theorem 2.1, the main result of the present chapter. We start with the precise formulation of the Skorokhod embedding problem and a reminder on Lévy processes.

Let μ_0 and μ_1 be two given probability distributions. On a complete probability space $(\Omega, \mathcal{F}, \mathbb{P}^{\mu_0})$ we consider a stochastic process L with $L_0 \sim \mu_0$ under $\mathbb{P}^{\mu_0}$ and denote by $(\mathcal{F}_t)_{t\geq 0}$ the $\mathbb{P}^{\mu_0}$-augmented natural filtration of L. In this setting the Skorokhod embedding problem is formulated as follows:

(SEP). *Find an $(\mathcal{F}_t)_{t\geq 0}$-stopping time τ such that $L_\tau \sim \mu_1$ and $\mathbb{E}^{\mu_0}[\tau] < \infty$.*

Throughout this chapter, L will be a Lévy process with initial distribution μ_0 under $\mathbb{P}^{\mu_0}$. For the special case $\mu_0 = \delta_0$ we always abbreviate $\mathbb{P} = \mathbb{P}^{\delta_0}$.

Recall that a continuous-time process $(L_t)_{t\geq 0}$ with values in $\mathbb{R}$ is called Lévy process if it has almost surely RCLL sample paths, is almost surely issued from 0, is stochastically continuous and has stationary and independent increments. Due to the Lévy-Khintchine representation, there exist $\alpha \geq 0$, $\gamma \in \mathbb{R}$ and a measure ν on $\mathbb{R}$ with $\nu(\{0\}) = 0$ and $\int_{\mathbb{R}} (x^2 \wedge 1)\, \nu(\mathrm{d}x) < \infty$ such that

$$\mathbb{E}[e^{iuL_t}] = e^{t\eta(u)}, \quad u \in \mathbb{R},\ t \geq 0, \tag{2.1}$$

with the characteristic exponent

$$\eta(u) = -\frac{1}{2}\alpha^2 u^2 + iu\gamma + \int_{\mathbb{R}\setminus\{0\}} (e^{iuy} - 1 - iuy\mathbb{1}_{\{|y|\leq 1\}})\, \nu(\mathrm{d}y), \quad u \in \mathbb{R}. \tag{2.2}$$

The triplet (α^2, γ, ν) is called Lévy triplet and fully characterizes L. We exclude the trivial case of a constant Lévy process, i.e. $\alpha = \gamma = \nu = 0$. For more background information we refer for instance to the introductory texts of Bertoin [Ber96] and Kyprianou [Kyp14] and to [Sat99]. A Lévy process with initial distribution μ_0 is defined as $L = \tilde{L} + X_0$, where $X_0 \sim \mu_0$ is independent from the Lévy process $\tilde{L}$.

We are now ready to state the main result. Condition (2.3) with H as in (2.4) also appears in [BL92], where $\mu_0 = \delta_0$ and Lévy processes with local times are considered. In particular, Theorem 2.1 shows that this is the necessary and sufficient condition also for a wide class of measures and Lévy processes without local times. Allowing the Lévy process to be more general forces us on the other hand to assume a priori regularity on μ_0, μ_1. We will always assume that μ_0, μ_1 have positive densities with respect to the Lebesgue measure. Additional smoothness will be imposed (e.g. $h_0, h_1 \in C_0(\mathbb{R})$ for the Brownian motion). Assumptions on the densities are different for different Lévy processes; we state the precise assumptions in Section 2.1 below.

Theorem 2.1. *Suppose L is a Lévy process with initial distribution μ_0 and characteristic exponent η. Suppose μ_0, μ_1 have strictly positive densities h_0, h_1 which are "sufficiently smooth" (specified below in Assumption 2.4).*

(i) The necessary and sufficient condition for the existence of a finite mean Skorokhod embedding is

$$\frac{\widehat{\mu_1} - \widehat{\mu_0}}{\eta} \in L^1(\mathbb{R}), \quad H \geq 0 \quad \text{and} \quad H \in L^1(\mathbb{R}), \tag{2.3}$$

where

$$H(x) := \frac{1}{2\pi} \int_{\mathbb{R}} \frac{\widehat{\mu_1}(\xi) - \widehat{\mu_0}(\xi)}{\eta(\xi)} e^{-ix\xi} \, \mathrm{d}\xi, \quad x \in \mathbb{R}. \tag{2.4}$$

(ii) If (2.3) is satisfied, then an explicit solution under $\mathbb{P}^{\mu_0}$ is given as follows:

$$\tau := \inf \left\{ t \in [0, \rho) \,:\, \int_0^t e^{-G(r)} \frac{h_1(L_r)}{H(L_r)} \, \mathrm{d}r \geq 1 \right\} \wedge \rho,$$

where, for $t \geq 0$,

$$\rho := \inf\{t \in [0, \infty) \,:\, H(L_t) = 0\} \quad \text{and} \quad G(t) := \int_0^t \frac{h_1(L_r) - h_0(L_r)}{H(L_r)} \, \mathrm{d}r$$

with the usual convention $\inf \emptyset := \infty$.

(iii) With τ from (ii) it holds that $\mathbb{E}^{\mu_0}[\tau] = \int_{\mathbb{R}} H(x) \, \mathrm{d}x$.

The conditions might look complicated at first sight but they are explicit since they only involve the given densities and the given characteristic exponent of the Lévy process. Also the stopping time is fairly explicit: it only involves the process and explicit functions but no further stochastic quantities (e.g. local times).

For the case of a Brownian motion starting from an initial law μ_0 with finite second moment, it is a classical result that there is a finite mean embedding for μ_1 if and only if μ_0 and μ_1 have the same first moment, finite second moments and μ_0 is smaller than μ_1 in convex order, i.e.

$$\int_{\mathbb{R}} \varphi(x) \, \mu_0(\mathrm{d}x) \leq \int_{\mathbb{R}} \varphi(x) \, \mu_1(\mathrm{d}x) \quad \text{for all } \varphi \text{ convex.}$$

Sufficiency follows e.g. by [BC74], necessity by the optional sampling theorem and Jensen's inequality.

Even though the three conditions in (2.3) cannot be considered separately from each other, each of them has an interpretation in analogy to the Brownian case: Since $\eta(0) = 0$, the integrability at zero of $(\widehat{\mu_1} - \widehat{\mu_0})/\eta$ forces a decay of $\widehat{\mu_1} - \widehat{\mu_0}$ in relation to the behavior of η at zero. Since the behavior at zero of a characteristic function relates to the moments, the integrability of $(\widehat{\mu_1} - \widehat{\mu_0})/\eta$ is an abstract condition for equal first moments of μ_0, μ_1. Non-negativity of

H is a generalization of the convex order condition for Brownian motion and integrability of H corresponds to finite second moments.

Finally, note that the function H appearing in (2.4) is the unique solution in $C_0(\mathbb{R})$ to the Poisson equation

$$\mathcal{A}^* H = h_1 - h_0, \tag{2.5}$$

where $\mathcal{A}^*$ is the generator of the dual Lévy process $-L$. This is shown in Section 3.1 below and can be seen informally by taking the Fourier transform in (2.5) and recalling the Fourier representation $\widehat{\mathcal{A}^* H}(u) = \eta(u)\widehat{H}(u)$ of $\mathcal{A}^*$. Using that for a Brownian motion the solution to (2.5) can also be represented as

$$H(x) = \int_{\mathbb{R}} (h_1(y) - h_0(y))|y - x| \,\mathrm{d}y, \quad x \in \mathbb{R}, \tag{2.6}$$

see e.g. [Sat72, Corollary 5.1], and that $\int_{\mathbb{R}} H(x)\,\mathrm{d}x = \int_{\mathbb{R}} x^2\,\mu_1(\mathrm{d}x) - \int_{\mathbb{R}} x^2\,\mu_0(\mathrm{d}x)$ as shown e.g. in [Cha77, Lemma 7.1], one may recover from (2.3) the classical conditions for Brownian motion stated above.

Remark 2.2. Note that for lattice type Lévy processes there exist $u_0 \neq 0$ with $\eta(u_0) = 0$. For such u_0 the condition $(\widehat{\mu_1} - \widehat{\mu_0})/\eta \in L^1(\mathbb{R})$ in (2.3) thus requires a decay of $\widehat{\mu_1}(u) - \widehat{\mu_0}(u)$ as $u \to u_0$.

2.1 Regularity Assumptions

For different Lévy processes the necessary and sufficient conditions for the solvability of the Skorokhod embedding problem (SEP) provided in Theorem 2.1 require different regularity assumptions on the initial density h_0 and the target density h_1. In order to state these assumptions, we distinguish between the following types of Lévy processes.

Definition 2.3. We say a Lévy process with characteristic exponent η is of type

S if it is symmetric and $\int_1^\infty \frac{1}{|\eta(u)|}\,\mathrm{d}u < \infty$,

0 if $\liminf_{u\to\infty} |\eta(u)| \in (0, \infty]$,

D if $\liminf_{u\to\infty} |\eta(u)| = 0$.

Notice that these three types cover all Lévy processes as in particular any Lévy process is either of type 0 or of type D. Based on this classification, we make the following regularity assumptions on the densities h_0, h_1.

Assumption 2.4 (Regularity Assumptions)**.**

- *If L is of type S, then $h_0, h_1 \in C_0(\mathbb{R})$.*
- *If L is of type 0, then $h_i \in C_0^2(\mathbb{R})$ with $h_i^{(k)} \in L^1(\mathbb{R})$ for $k = 1, 2$, $i = 0, 1$.*
- *If L is of type D, then $\widehat{h_1} - \widehat{h_0} \in C_c(\mathbb{R})$.*

The Lévy processes considered by Bertoin and Le Jan are of type 0 as we will see in the next remark. The subclass of processes considered in the Appendix of [BL92] (for which conditions (2.3) were proved) are even of type S.

Remark 2.5. In [BL92], the Lévy processes are assumed to be recurrent and satisfy that 0 is regular for 0. Excluding the compound Poisson case, the last condition is equivalent to condition (i) of Lemma 2.7 below and

$$\sigma^2 > 0 \quad \text{or} \quad \int_{\mathbb{R}} (|x| \wedge 1)\, \nu(\mathrm{d}x) = \infty,$$

see [RW00a, Section I.30]. Hence, these processes are of type 0 by Lemma 2.7.

Let us give further examples:

Example 2.6. Type S: Symmetric α-stable Lévy processes with index $\alpha \in (1,2]$ are of type S. In particular, a Brownian motion is of type S and so for a Brownian motion Theorem 2.1 provides a solution to the SEP for any positive, continuous densities $h_0, h_1 \in C_0(\mathbb{R})$ which have the same first moment, finite second moments and which are in convex order.

Type 0: Symmetric α-stable Lévy processes with index $\alpha \in (0,1]$ are of type 0, but not of type S.

Type D: Lattice-type compound Poisson processes are of type D. Other examples of processes of type D can be found in [Ber96, Exercise I.7] and [Sat99, Example 41.23].

In fact, Lévy processes of type 0 form a large class as demonstrated by the sufficient conditions presented in the next lemma.

Lemma 2.7. *Suppose that either*

(i) $\int_{\mathbb{R}} \mathrm{Re}\left(\frac{1}{1-\eta(\xi)}\right) \mathrm{d}\xi < \infty$ *or*

(ii) for some $t > 0$*, the distribution of* $L_t - L_0$ *has a non-trivial absolutely continuous part.*

Then L *is of type 0.*

Proof. We argue by contraposition. Suppose there exists $\{u_k\}_{k\in\mathbb{N}} \subset \mathbb{R}$ such that $\lim_{k\to\infty} |u_k| = \infty$ and $\lim_{k\to\infty} \eta(u_k) = 0$. Then condition (ACP) in [Sat99] cannot be satisfied (see [Sat99, Example 41.23]). Combining this with [Sat99, Thm. 43.3] and [Sat99, Remark 43.6], condition (i) does not hold. Similarly, for any $t > 0$, $\lim_{k\to\infty} \mathbb{E}^{\mu_0}[e^{iu_k(L_t - L_0)}] = 1$ and by the Riemann-Lebesgue Theorem it follows that the law of $L_t - L_0$ does not have an absolutely continuous part. Hence, (ii) does not hold either. □

3 Proofs

In this section we prove Theorem 2.1. The exposition is structured as follows: Firstly, in Section 3.1 preliminary results on Lévy processes and the associated

Poisson equation are presented. Secondly, in Section 3.2 it is proved that (2.3) is indeed necessary for the existence of a finite mean Skorokhod embedding. Finally, in Section 3.3 it is established that τ in Theorem 2.1 is a finite mean solution to the Skorokhod embedding problem (SEP), thereby also proving sufficiency of (2.3).

3.1 The Poisson Equation for Lévy Processes

In this section we lay the foundation for the proof of Theorem 2.1. As sketched in the introduction of this thesis, it is crucial to understand the solvability of the Poisson equation $\mathcal{A}^*H = h_1 - h_0$ and properties of the solution H. As sketched below Theorem 2.1, the Poisson equation for Lévy processes can be tackled with Fourier transforms. Throughout this section we take $\mu_0 := \delta_0$ and set $\mathbb{P} := \mathbb{P}^{\mu_0}$.

3.1.1 Preliminaries

Recall from the introduction that (α^2, γ, ν) denotes the Lévy triplet and

$$\eta(u) = -\frac{1}{2}\alpha^2 u^2 + iu\gamma + \int_{\mathbb{R}\setminus\{0\}} (e^{iuy} - 1 - iuy\mathbb{1}_{\{|y|\le 1\}})\, \nu(\mathrm{d}y), \quad u \in \mathbb{R},$$

is the characteristic exponent, i.e. $\mathbb{E}[e^{iuL_t}] = e^{t\eta(u)}$ for $u \in \mathbb{R}$ and $t \ge 0$. In what follows we collect the machinery that we need to study the Poisson equation for Lévy processes in the next section.

For $t \ge 0$ and $f \in C_0(\mathbb{R})$ define the transition semigroup $P_t f(x) := \mathbb{E}[f(L_t + x)], x \in \mathbb{R}$, and, for $q > 0$, $f \in C_0(\mathbb{R})$ the resolvent operators

$$U^q f(x) := \int_0^\infty e^{-qt} P_t f(x)\,\mathrm{d}t, \quad x \in \mathbb{R}.$$

By dominated convergence, $f \in C_0(\mathbb{R})$ implies $P_t f \in C_0(\mathbb{R})$ for any $t \ge 0$ and thus

$$D(\mathcal{A}) := \left\{ f \in C_0(\mathbb{R}) \,:\, \lim_{t\to 0} t^{-1}(P_t f - f) \text{ exists in } C_0(\mathbb{R}) \right\}$$

is well-defined. For $f \in D(\mathcal{A})$ define $\mathcal{A}f := \lim_{t\to 0} t^{-1}(P_t f - f)$. Then, see [Sat99, Thm. 31.5], the generator $\mathcal{A}\colon D(\mathcal{A}) \to C_0(\mathbb{R})$ is linear, $C_0^2(\mathbb{R}) \subset D(\mathcal{A})$ and for $u \in C_0^2(\mathbb{R})$ it holds that

$$\mathcal{A}u(x) = \frac{1}{2}\alpha^2 u''(x) + \gamma u'(x) + \int_{\mathbb{R}\setminus\{0\}} [u(x+y) - u(x) - yu'(x)\mathbb{1}_{\{|y|\le 1\}}]\, \nu(\mathrm{d}y), \quad x \in \mathbb{R}. \tag{3.1}$$

Furthermore, $C_c^\infty(\mathbb{R})$ is a core for $\mathcal{A}$. This means by definition that for any $f \in D(\mathcal{A})$, there exists a sequence $\{f_n\}_{n\in\mathbb{N}} \subset C_c^\infty(\mathbb{R})$ such that $\lim_{n\to\infty} f_n = f$ and $\lim_{n\to\infty} \mathcal{A}f_n = \mathcal{A}f$ in $C_0(\mathbb{R})$. In particular, we note the following:

Lemma 3.1. *Suppose L is a Lévy process with Lévy triplet (α^2, γ, ν). Set $\mathcal{D} := C_c^\infty(\mathbb{R})$ and define $\mathcal{A}\colon \mathcal{D} \to C_0(\mathbb{R})$ as (3.1) for $u \in \mathcal{D}$. Then $(\mathcal{D}, \mathcal{A})$ satisfies Assumption A.7.*

Proof. This follows by general theory (as explained in Example A.8) since the transition semigroup is a positive, strongly continuous contraction semigroup on $C_0(\mathbb{R})$ by [Sat99, Thm. 31.5] and since $\mathcal{D}$ is a core (see above) for the infinitesimal generator of $(P_t)_{t\geq 0}$.[1] □

Two objects which are less popular in the study of Lévy processes, but central for our purposes, are the potential operator and the adjoint, both of which we introduce next. From [Sat99, Remark 31.10] or [Sat72, Thm. 4.1] it follows that the Lévy process admits a potential operator. By definition this means that $\mathcal{A}$ is injective, the domain $D(V) := \{\mathcal{A}f : f \in D(\mathcal{A})\}$ of the potential operator $V := -\mathcal{A}^{-1}$ is dense in $C_0(\mathbb{R})$ and, for $f, g \in C_0(\mathbb{R})$,

$$g \in D(V) \text{ and } Vg = f \quad \Longleftrightarrow \quad U^q g \to f \text{ in } C_0(\mathbb{R}) \text{ as } q \to 0. \tag{3.2}$$

Furthermore, for $t \geq 0$, set $L_t^* := -L_t$. Then L^* is also a Lévy process, called the dual Lévy process, and its Lévy triplet is given by $(\alpha^2, -\gamma, \nu^*)$ where $\nu^*(A) := \nu(\{-x : x \in A\})$ for $A \in \mathcal{B}(\mathbb{R})$. In other words, the characteristic exponent η^* of L^* is given for $u \in \mathbb{R}$ as $\eta^*(u) = \eta(-u)$ where η is the characteristic exponent (2.2) of L. Since L^* is also a Lévy process, the transition semigroup, resolvent operator, infinitesimal generator and potential operator have been defined above. We will denote them by P_t^*, $(U^q)^*$, $\mathcal{A}^*$ and V^*, respectively.

For example, we denote by $\mathcal{A}^*$ the infinitesimal generator associated to the dual Lévy process L^* and refer to it as the dual of $\mathcal{A}$. Recall from the above that $C_0^2(\mathbb{R}) \subset D(\mathcal{A}^*)$ and for $u \in C_0^2(\mathbb{R})$, $x \in \mathbb{R}$,

$$\mathcal{A}^* u(x) = \frac{1}{2}\alpha^2 u''(x) - \gamma u'(x) + \int_{\mathbb{R}\setminus\{0\}} [u(x-y) - u(x) + y u'(x) \mathbb{1}_{\{|y|\leq 1\}}]\, \nu(\mathrm{d}y). \tag{3.3}$$

Remark 3.2. The semigroup $(P_t)_{t\geq 0}$ and the operator $\mathcal{A}$ are defined on (a subset of) $C_0(\mathbb{R})$ in the present context. We define $\mathcal{A}^*$ also on $C_0(\mathbb{R})$ (and not on the dual space of $C_0(\mathbb{R})$ as in [Sat72]). The next lemma justifies the $*$-notation.

The following lemma is immediate, it identifies the dual generator $\mathcal{A}^*$ as the adjoint operator of $\mathcal{A}$. For the proof of Theorem 2.1 we will not need all the cases, but we have included the other ones for completeness.

Lemma 3.3. *Suppose L is a Lévy process with Lévy triplet (α^2, γ, ν) and denote by $\mathcal{A}$ its infinitesimal generator $\mathcal{A}\colon D(\mathcal{A}) \to C_0(\mathbb{R})$ and $\mathcal{A}^*\colon D(\mathcal{A}^*) \to C_0(\mathbb{R})$ the infinitesimal generator of $-L$. Then*

$$\int_{\mathbb{R}} \mathcal{A}f(x) g(x)\, \mathrm{d}x = \int_{\mathbb{R}} f(x) \mathcal{A}^* g(x)\, \mathrm{d}x, \tag{3.4}$$

for any $f \in D(\mathcal{A})$, $g \in D(\mathcal{A}^*)$ such that either $f, g \in L^1(\mathbb{R})$ or $f, \mathcal{A}f \in L^1(\mathbb{R})$ or $g, \mathcal{A}^* g \in L^1(\mathbb{R})$ or $\mathcal{A}f, \mathcal{A}^* g \in L^1(\mathbb{R})$.

[1] One could also verify Assumption A.7 (i) by hand taking $\phi \in C_c^\infty(\mathbb{R})$ with $\phi(x) = 1$ for $x \in [-1, 1]$ and $\phi(x) = 0$ for $x \notin [-2, 2]$ and setting $\phi_n(x) := \phi(x/n)$ for $x \in \mathbb{R}$, $n \in \mathbb{N}$. Then $\{\phi_n\}_{n\in\mathbb{N}} \subset \mathcal{D}$ and $\text{bp-lim}_{n\to\infty}(\phi_n, \mathcal{A}\phi_n) = (1, 0)$.

Proof. **Case 1**, $f, g \in L^1(\mathbb{R})$: For $t \geq 0$ and $x \in \mathbb{R}$ denote by P^* the transition semigroup of $-L$, i.e. $P_t^* g(x) = \mathbb{E}[g(x - L_t)]$. By Fubini's Theorem $g \in L^1(\mathbb{R})$ implies $P_t^* g \in L^1(\mathbb{R})$ for any $t \geq 0$. By [Ber96, Chap. II, Prop. 1] for any $t \geq 0$ it holds that

$$\int_{\mathbb{R}} P_t f(x) g(x) \, \mathrm{d}x = \int_{\mathbb{R}} f(x) P_t^* g(x) \, \mathrm{d}x. \tag{3.5}$$

To be precise, in [Ber96, Chap. II, Prop. 1] f and g are assumed non-negative, but by considering positive and negative parts separately and using $g \in L^1(\mathbb{R})$ and $P_t^* g \in L^1(\mathbb{R})$, [Ber96, Chap. II, Prop. 1] implies (3.5).

Using the definition of $\mathcal{A}$, $f \in D(\mathcal{A})$ and $g \in L^1(\mathbb{R})$ to apply dominated convergence in the first step and $g \in D(\mathcal{A}^*)$ and $f \in L^1(\mathbb{R})$ in the last step, one obtains

$$\begin{aligned}
\int_{\mathbb{R}} \mathcal{A} f(x) g(x) \, \mathrm{d}x &= \lim_{t \to 0} \frac{1}{t} \int_{\mathbb{R}} (P_t f(x) - f(x)) g(x) \, \mathrm{d}x \\
&\overset{(3.5)}{=} \lim_{t \to 0} \frac{1}{t} \int_{\mathbb{R}} f(x) (P_t^* g(x) - g(x)) \, \mathrm{d}x \\
&= \int_{\mathbb{R}} f(x) \mathcal{A}^* g(x) \, \mathrm{d}x
\end{aligned}$$

and so (3.4) has been established under the assumption $f, g \in L^1(\mathbb{R})$.

Case 2, $f, \mathcal{A} f \in L^1(\mathbb{R})$ or $g, \mathcal{A}^* g \in L^1(\mathbb{R})$: Suppose $g, \mathcal{A}^* g \in L^1(\mathbb{R})$, the other case can be treated by the same argument. Since $C_c^\infty(\mathbb{R})$ is a core for $\mathcal{A}$, there exists $\{f_n\}_{n \in \mathbb{N}} \subset C_c^\infty(\mathbb{R})$ with $\lim_{n \to \infty} f_n = f$ and $\lim_{n \to \infty} \mathcal{A} f_n = \mathcal{A} f$ in $C_0(\mathbb{R})$. But $f_n, g \in L^1(\mathbb{R})$ and so (3.4) holds for f_n and g for any $n \in \mathbb{N}$. Furthermore, the assumptions $g, \mathcal{A}^* g \in L^1(\mathbb{R})$ allow us to apply dominated convergence and so (3.4) also holds for f and g, as desired.

Case 3, $\mathcal{A} f$, $\mathcal{A}^* g \in L^1(\mathbb{R})$: For the proof of the last part, denote by V and V^* the potential operators associated to $\mathcal{A}$ and $\mathcal{A}^*$, respectively. We claim that for any $\tilde{f} \in D(V)$, $\tilde{g} \in D(V^*)$ with $\tilde{f}, \tilde{g} \in L^1(\mathbb{R})$ it holds that

$$\int_{\mathbb{R}} V \tilde{f}(x) \tilde{g}(x) \, \mathrm{d}x = \int_{\mathbb{R}} \tilde{f}(x) V^* \tilde{g}(x) \, \mathrm{d}x. \tag{3.6}$$

Once this is established, we may set $\tilde{f} := \mathcal{A} f$, $\tilde{g} := \mathcal{A}^* g$ and apply (3.6) to deduce (3.4).

So assume $\tilde{f} \in D(V)$, $\tilde{g} \in D(V^*)$ and $\tilde{f}, \tilde{g} \in L^1(\mathbb{R})$. For $q > 0$ and $x \in \mathbb{R}$ denote by $(U^q)^*$ the resolvent operator of $-L$, i.e. $(U^q)^* \tilde{g}(x) = \int_0^\infty e^{-qt} P_t^* \tilde{g}(x) \, \mathrm{d}t$ where P^* is the transition semigroup of $-L$ as above. By Fubini's Theorem, $\tilde{g} \in L^1(\mathbb{R})$ implies $(U^q)^* \tilde{g} \in L^1(\mathbb{R})$ for any $q > 0$. By [Ber96, Chap. II, Prop. 1] for any $q > 0$ it holds that

$$\int_{\mathbb{R}} U^q \tilde{f}(x) \tilde{g}(x) \, \mathrm{d}x = \int_{\mathbb{R}} \tilde{f}(x) (U^q)^* \tilde{g}(x) \, \mathrm{d}x. \tag{3.7}$$

By the same argument as in Case 2, [Ber96, Chap. II, Prop. 1] implies (3.7).

Using (3.2), $\tilde{f} \in D(V)$ and $\tilde{g} \in L^1(\mathbb{R})$ one may let $q \to 0$ and apply dominated convergence to obtain that the left-hand side of (3.7) converges to the left-hand side of (3.6) and analogously for the right-hand side. Thus (3.6) is indeed established. □

3.1.2 Solving the Poisson Equation using the Fourier Transform

By definition of the potential operator $V^* = -(\mathcal{A}^*)^{-1}$ in the previous section, for $g \in D(V^*)$ the function $H = -V^* g$ is the unique solution to the Poisson equation

$$\mathcal{A}^* H = g \tag{3.8}$$

in $C_0(\mathbb{R})$. In this section we study the solvability of (3.8) and further properties of solutions.

The first proposition justifies the heuristic given in the introduction below Theorem 2.1 and, hence, the occurrence of the function H in Theorem 2.1. Note that the appearing assumption $g \in L^1(\mathbb{R})$ will not pose any restriction as in our applications $g = h_1 - h_0$ and h_0, h_1 are probability densities.

Proposition 3.4. *If $g \in C_0(\mathbb{R}) \cap L^1(\mathbb{R})$ and $\xi \mapsto \frac{\hat{g}(\xi)}{\eta(\xi)} \in L^1(\mathbb{R})$, then there is a unique solution $H \in D(\mathcal{A}^*) \subset C_0(\mathbb{R})$ to the Poisson equation $\mathcal{A}^* H = g$ and*

$$H(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \frac{\hat{g}(\xi)}{\eta(\xi)} e^{-ix\xi} \, \mathrm{d}\xi, \quad x \in \mathbb{R}. \tag{3.9}$$

Proof. We start with some preliminary facts that do not use the assumption of the proposition. Note that $g \in C_0(\mathbb{R}) \cap L^1(\mathbb{R})$ implies $g \in L^\infty(\mathbb{R}) \cap L^1(\mathbb{R})$. Hence, one can use Fubini's Theorem to see that for any $q > 0$, $(U^q)^* g(x) = \int_0^\infty e^{-qt} P_t^* g(x) \, \mathrm{d}t$ is in $L^1(\mathbb{R})$. Taking the Fourier transform we obtain (see e.g. [Ber96, Chap. I, Prop. 9]) for any $q > 0$

$$\widehat{(U^q)^* g}(\xi) = (q - \eta(\xi))^{-1} \hat{g}(\xi), \quad \xi \in \mathbb{R}. \tag{3.10}$$

By (2.1) it holds that $|e^{\eta(\xi)}| = |\mathbb{E}[e^{i\xi L_1}]| \leq 1$ and thus

$$\mathrm{Re}(\eta(\xi)) \leq 0 \quad \text{for any } \xi \in \mathbb{R}, \tag{3.11}$$

where $\mathrm{Re}(z)$ denotes the real part of $z \in \mathbb{C}$. In particular, the right-hand side of (3.10) is indeed well-defined and

$$\forall \xi \in \mathbb{R}, q > 0 \, : \, |\eta(\xi)| \leq |q - \eta(\xi)|. \tag{3.12}$$

By assumption, the inverse Fourier transform of $\xi \mapsto \frac{\hat{g}(\xi)}{\eta(\xi)}$, given by H in (3.9), is well-defined. Furthermore, by (3.12) and our assumption, for any $q > 0$ the right-hand side of (3.10) is integrable and so, as $(U^q)^* g \in C_0(\mathbb{R}) \cap L^1(\mathbb{R})$ and $\widehat{(U^q)^* g} \in L^1(\mathbb{R})$, by Fourier inversion and (3.10), $(U^q)^* g$ can be represented as

$$(U^q)^* g(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \frac{\hat{g}(\xi) e^{-ix\xi}}{q - \eta(\xi)} \, \mathrm{d}\xi, \quad x \in \mathbb{R}. \tag{3.13}$$

Thus, one has

$$\begin{aligned}\sup_{x\in\mathbb{R}} |-H(x) - (U^q)^* g(x)| &= \sup_{x\in\mathbb{R}} \left| \frac{1}{2\pi} \int_{\mathbb{R}} \hat{g}(\xi) e^{-ix\xi} \left(\frac{1}{\eta(\xi)} + \frac{1}{q-\eta(\xi)} \right) \mathrm{d}\xi \right| \\ &\le \frac{1}{2\pi} \int_{\mathbb{R}} |\hat{g}(\xi)| \left| \frac{1}{\eta(\xi)} + \frac{1}{q-\eta(\xi)} \right| \mathrm{d}\xi.\end{aligned}$$

However, by (3.12) the last integrand is bounded from above by $\xi \mapsto 2\hat{g}(\xi)/\eta(\xi)$, which is integrable by assumption, and so we can let $q \to 0$ and apply dominated convergence to conclude $(U^q)^* g \to -H$ as $q \to 0$ in $C_0(\mathbb{R})$. By (3.2) (for the dual Lévy process L^*), this implies $g \in D(V^*)$ and $V^* g = -H$ or, in other words, $\mathcal{A}^* H = g$. □

In the next proposition we show that under the assumptions of Theorem 2.1, it holds that $H \in D(\mathcal{A}^*) \cap D(\mathcal{A})$. This property will be crucial to guarantee uniqueness for the time-change in the proof of sufficiency of Theorem 2.1 and forces us to assume that the densities h_0, h_1 are "sufficiently smooth" in the next sections.

Proposition 3.5. *Suppose h_0, h_1 are as in Theorem 2.1 and* (2.3) *holds with H as in* (2.4). *Then $H \in D(\mathcal{A}) \cap D(\mathcal{A}^*)$.*

Proof. Set $g := h_1 - h_0$. By Proposition 3.4, $H \in D(\mathcal{A}^*)$ and so we only need to verify $H \in D(\mathcal{A})$. If L is of type S, then it is symmetric. In particular $\mathcal{A} = \mathcal{A}^*$ and hence the claim. If L is of type 0 or D, then (as established in the proof of Lemma 3.9 below), $\hat{g} \in L^1(\mathbb{R})$ and we now show that this implies $H \in D(\mathcal{A})$.

Since the complex conjugate of $\eta(\xi)$ is given by $\eta(-\xi)$ for all $\xi \in \mathbb{R}$ and since $\hat{g} \in L^1(\mathbb{R})$, the function

$$f(x) := \frac{1}{2\pi} \int_{\mathbb{R}} \frac{\hat{g}(\xi)\eta(-\xi)}{\eta(\xi)} e^{-ix\xi} \,\mathrm{d}\xi, \quad x \in \mathbb{R},$$

is well-defined and, by the Riemann-Lebesgue Theorem, $f \in C_0(\mathbb{R})$. Inserting (3.9) in the definition, applying Fubini's Theorem and using (2.1) and the definition of f yields

$$\begin{aligned}&\sup_{x\in\mathbb{R}} \left| \frac{1}{t}(P_t H(x) - H(x)) - f(x) \right| \\ &= \sup_{x\in\mathbb{R}} \left| \frac{1}{t}(\mathbb{E}[H(L_t + x)] - H(x)) - f(x) \right| \\ &= \sup_{x\in\mathbb{R}} \left| \frac{1}{2\pi t} \int_{\mathbb{R}} \frac{\hat{g}(\xi)}{\eta(\xi)} (\mathbb{E}[e^{-iL_t\xi}] - 1) e^{-ix\xi} \,\mathrm{d}\xi - f(x) \right| \\ &\le \frac{1}{2\pi} \int_{\mathbb{R}} \frac{|\hat{g}(\xi)|}{|\eta(\xi)|} \left| \frac{e^{t\eta(-\xi)} - 1}{t} - \eta(-\xi) \right| \mathrm{d}\xi,\end{aligned}$$

which tends to 0 as $t \downarrow 0$, by dominated convergence. By definition, this implies $H \in D(\mathcal{A})$ and $\mathcal{A}H = f$. To see that dominated convergence can be applied, recall (3.11) and so for all $\xi \in \mathbb{R}$, $t > 0$,

$$\frac{|\hat{g}(\xi)|}{|\eta(\xi)|}\left|\frac{e^{t\eta(-\xi)}-1}{t}-\eta(-\xi)\right| \le 2|\hat{g}(\xi)|.$$

□

The rest of this section may be skipped on first reading, all following propositions are not needed for the proof of Theorem 2.1.

We give conditions on g and η so that Proposition 3.4 implies the existence of the solution H, conditions that imply $H \in L^1(\mathbb{R})$ and that H is Lipschitz continuous. For future applications, those might be useful to verify the conditions of Theorem 2.1.

Proposition 3.6. *Assume the non-degeneracy condition $\eta(u) \neq 0$ for all $u \neq 0$ (i.e. L is non-lattice) and either $\nu \neq 0$ or $\alpha \neq 0$. If $g \in C_0(\mathbb{R}) \cap L^1(\mathbb{R})$, $x \mapsto x^i g(x) \in L^1(\mathbb{R})$ for $i = 1, 2$,*

$$\int_{\mathbb{R}} g(x)x^j \,\mathrm{d}x = 0, \quad j = 0, 1, \tag{3.14}$$

and there exists $R > 0$ such that

$$\int_{|\xi|>R} \frac{|\hat{g}(\xi)|}{|\eta(\xi)|}\,\mathrm{d}\xi < \infty, \tag{3.15}$$

then $\xi \mapsto \frac{\hat{g}(\xi)}{\eta(\xi)} \in L^1(\mathbb{R})$.

Proof. First note that in the present one-dimensional setup, the law of L_1 is degenerate (in the sense of [Sat99, Def. 24.16]) if and only if there exists $a \in \mathbb{R}$ with $L_1 = a$, $\mathbb{P}$-a.s. Since we have assumed $\alpha \neq 0$ or $\nu \neq 0$, this is not the case here (see [Sat99, Thm. 24.3]). In particular, we may apply [Sat99, Prop. 24.19] and obtain that there exist $\varepsilon' > 0$ and $c > 0$ such that

$$|\mathbb{E}[e^{i\xi L_1}]| \le 1 - c|\xi|^2 \quad \text{for } |\xi| < \varepsilon'. \tag{3.16}$$

By (2.1), the left-hand side of (3.16) is greater or equal than $e^{\mathrm{Re}(\eta(\xi))}$ and thus there exist $C > 0$ and $\varepsilon > 0$ such that

$$-\mathrm{Re}(\eta(\xi)) \ge -\log(1 - c|\xi|^2) \ge C|\xi|^2 \tag{3.17}$$

for all $\xi \in B_\varepsilon(0)$.

On the other hand, for $g \neq 0$ (if $g = 0$, then the claim trivially holds), we may decompose $g = g^+ - g^-$ with $g^+ \ge 0$, $g^- \ge 0$. Setting $c_0 := \int_{\mathbb{R}} g^+(x)\,\mathrm{d}x$, (3.14) with $i = 0$ implies $c_0 = \int_{\mathbb{R}} g^-(x)\,\mathrm{d}x$ and thus $c_0 > 0$. Setting $h_1 := g^+/c_0$ and $h_0 := g^-/c_0$, both h_0 and h_1 are probability densities and so we may apply [Sat99, Prop. 2.5 (ix)] to h_0 and h_1 to obtain that, by our moment assumptions, $\hat{g} \in C^2(\mathbb{R})$ and, by (3.14), $\hat{g}(0) = \hat{g}'(0) = 0$. Taylor expanding around 0, we therefore obtain

$$|\hat{g}(\xi)| \le C_0\xi^2 \tag{3.18}$$

for some $C_0 > 0$ and all $\xi \in B_\varepsilon(0)$. Combining (3.17) and (3.18) yields

$$|\hat{g}(\xi)| \leq C_0\xi^2 \leq -\mathrm{Re}(\eta(\xi))\frac{C_0}{C} \leq \frac{C_0}{C}|\eta(\xi)|$$

for all $\xi \in B_\varepsilon(0)$ and thus $\xi \mapsto \frac{\hat{g}(\xi)}{\eta(\xi)}$ is locally bounded at zero. Since $\eta(u) \neq 0$ for $u \neq 0$ and $\hat{g}$ and η are continuous, it follows that $\xi \mapsto \frac{\hat{g}(\xi)}{\eta(\xi)}$ is bounded on any compact subset of $\mathbb{R}$. Combining this with (3.15) yields the claim. □

Proposition 3.7. *Suppose g and H are as in Proposition 3.4. If in addition for some $R > 0$,*

$$\int_{|\xi|>R} \frac{|\xi||\hat{g}(\xi)|}{|\eta(\xi)|}\,\mathrm{d}\xi < \infty, \tag{3.19}$$

then H is Lipschitz continuous.

Proof. By assumption,

$$\int_{|\xi|\leq R} \frac{|\xi||\hat{g}(\xi)|}{|\eta(\xi)|}\,\mathrm{d}\xi \leq R\int_{\mathbb{R}} \frac{|\hat{g}(\xi)|}{|\eta(\xi)|}\,\mathrm{d}\xi < \infty$$

and combining this with (3.19) yields

$$L := \int_{\mathbb{R}} \frac{|\xi||\hat{g}(\xi)|}{|\eta(\xi)|}\,\mathrm{d}\xi < \infty. \tag{3.20}$$

On the other hand, precisely as in the proof of Proposition 3.4 we may apply Fourier inversion to write, for any $q > 0$, $(U^q)^*$ as (3.13). Using $|e^{iu} - e^{iv}| \leq |u-v|$ for $u, v \in \mathbb{R}$ yields

$$\begin{aligned}|(U^q)^*g(x) - (U^q)^*g(y)| &\overset{(3.13)}{=} \frac{1}{2\pi}\left|\int_{\mathbb{R}} \frac{\hat{g}(\xi)(e^{-ix\xi} - e^{-iy\xi})}{q - \eta(\xi)}\,\mathrm{d}\xi\right| \\ &\overset{(3.12)}{\leq} \frac{1}{2\pi}|x-y|\int_{\mathbb{R}} \frac{|\xi||\hat{g}(\xi)|}{|\eta(\xi)|}\,\mathrm{d}\xi \overset{(3.20)}{=} \frac{L}{2\pi}|x-y|\end{aligned}$$

for any $q > 0$ and $x, y \in \mathbb{R}$. Letting $q \to 0$ and using that $(U^q)^*g \to V^*g$ pointwise (even in $C_0(\mathbb{R})$) by (3.2), this last estimate implies the result. □

The next result shows that if a solution of the Poisson equation exists (e.g. if the conditions of Proposition 3.4 hold, but here we impose a slightly weaker assumption), then $H \geq 0$ implies $H \in L^1(\mathbb{R})$. This is useful for verifying the conditions of Theorem 2.1.

Proposition 3.8. *If $g \in C_0(\mathbb{R}) \cap L^1(\mathbb{R})$, $\xi \mapsto \frac{\hat{g}(\xi)}{\eta(\xi)}$ is locally bounded at zero, there is a solution $H \in C_0(\mathbb{R})$ to the Poisson equation $\mathcal{A}^*H = g$ and $H \geq 0$, then $H \in L^1(\mathbb{R})$.*

Proof. Using (3.10) which did not rely on the stronger assumptions of Proposition 3.4 and the local boundedness of $\xi \mapsto \frac{\hat{g}(\xi)}{\eta(\xi)}$ in the first and (3.12) in the second inequality, there is some $\varepsilon > 0$ and $C > 0$ with

$$|\widehat{(U^q)^* g}(\xi)| \leq C|(q-\eta(\xi))|^{-1}|\eta(\xi)| \leq C \quad \text{for } \xi \in B_\varepsilon(0). \tag{3.21}$$

Let us take a function φ such that

$$\varphi \in L^1(\mathbb{R}) \text{ satisfies } \varphi = 0 \text{ on } \mathbb{R} \setminus B_\varepsilon(0), \varphi \geq 0, \hat{\varphi} \in L^1(\mathbb{R}) \text{ and } \hat{\varphi} \geq 0. \tag{3.22}$$

Since φ, $\hat{\varphi}, (U^q)^* g \in L^1(\mathbb{R})$, see the beginning of the proof of Proposition 3.4, Fubini's Theorem gives

$$\int_{\mathbb{R}} (U^q)^* g(x) \hat{\varphi}(x) \, \mathrm{d}x = \int_{\mathbb{R}} \int_{\mathbb{R}} (U^q)^* g(x) e^{i\xi x} \varphi(\xi) \, \mathrm{d}\xi \, \mathrm{d}x = \int_{\mathbb{R}} \widehat{(U^q)^* g}(\xi) \varphi(\xi) \, \mathrm{d}\xi. \tag{3.23}$$

Furthermore, $H\hat{\varphi} \geq 0$ and $H\hat{\varphi} \in L^1(\mathbb{R})$ since $H \in C_0(\mathbb{R})$ and we have assumed $H \geq 0$ and (3.22). Thus, recalling $H = -V^* g$, we may estimate

$$\begin{aligned} 0 \leq \int_{\mathbb{R}} H(x) \hat{\varphi}(x) \, \mathrm{d}x &\overset{(3.2)}{=} -\lim_{q \to 0} \int_{\mathbb{R}} (U^q)^* g(x) \hat{\varphi}(x) \, \mathrm{d}x \\ &\overset{(3.23)}{=} -\lim_{q \to 0} \int_{\mathbb{R}} \widehat{(U^q)^* g}(\xi) \varphi(\xi) \, \mathrm{d}\xi \overset{(3.21)}{\leq} C \int_{\mathbb{R}} \varphi(\xi) \, \mathrm{d}x, \end{aligned} \tag{3.24}$$

where the first equality uses dominated convergence and the last step relies on our assumption (3.22) that $\varphi = 0$ outside $B_\varepsilon(0)$.

We now claim that there exists $\{\varphi_n\}_{n \in \mathbb{N}} \subset L^1(\mathbb{R})$ and $I \in (0, \infty)$ such that for each n, φ_n satisfies (3.22) and $\lim_{n \to \infty} \hat{\varphi_n}(x) = I$ for any $x \in \mathbb{R}$. Assuming that such a sequence can indeed be constructed, the following argument will finish the proof: inserting φ_n in (3.24), letting $n \to \infty$ and using Fatou's Lemma yields

$$\begin{aligned} 0 \leq \int_{\mathbb{R}} H(x) \, \mathrm{d}x &= \frac{1}{I} \int_{\mathbb{R}} \liminf_{n \to \infty} H(x) \hat{\varphi_n}(x) \, \mathrm{d}x \\ &\leq \liminf_{n \to \infty} \frac{1}{I} \int_{\mathbb{R}} H(x) \hat{\varphi_n}(x) \, \mathrm{d}x \\ &\leq \liminf_{n \to \infty} \frac{C}{I} \int_{\mathbb{R}} \varphi_n(x) \, \mathrm{d}x = \liminf_{n \to \infty} \frac{C}{I} \hat{\varphi_n}(0) = C < \infty \end{aligned}$$

and therefore indeed $H \in L^1(\mathbb{R})$.

Thus, the remainder of the proof will be devoted to construct a sequence $\{\varphi_n\}_{n \in \mathbb{N}} \subset L^1(\mathbb{R})$ with the desired properties. Take $\chi_0 \in C_c^\infty(\mathbb{R}) \setminus \{0\}$ with $\chi_0 \geq 0$, $\chi_0(-x) = \chi_0(x)$ for all $x \in \mathbb{R}$ and $\chi_0(x) = 0$ for $x \notin B_{\varepsilon/3}(0)$. Set $\chi(y) := \int_{\mathbb{R}} \chi_0(y-x)\chi_0(x) \, \mathrm{d}x = \chi_0 * \chi_0(y)$. Then $\chi(y) = 0$ for $y \notin B_\varepsilon(0)$ and since the Fourier transform turns convolution into products, $\hat{\chi}(\xi) = (\hat{\chi_0}(\xi))^2$ for all $\xi \in \mathbb{R}$. In particular, $\hat{\chi} \geq 0$. Furthermore, $\hat{\chi_0} \neq 0$ implies that $I := \int_{\mathbb{R}} \hat{\chi}(\xi) \, \mathrm{d}\xi = \int_{\mathbb{R}} (\hat{\chi_0}(\xi))^2 \, \mathrm{d}\xi$ satisfies $I > 0$ and $\chi_0 \in C_c^\infty(\mathbb{R})$ implies $\hat{\chi_0} \in L^2(\mathbb{R})$, thus $I \in (0, \infty)$. Finally, since $\hat{\chi} \in L^1(\mathbb{R})$, Fourier inversion gives $\chi(x) = (2\pi)^{-1}\hat{\hat{\chi}}(-x)$ for all $x \in \mathbb{R}$ (see [Sat99, Prop. 37.2]).

For $n \in \mathbb{N}$, define

$$\varphi_n(x) := 2\pi n \chi(-x) \exp\Big(-\frac{1}{2}n^2x^2\Big), \quad x \in \mathbb{R},$$

and note that $\varphi_n(x) = \hat{\hat{\chi}}(x) n \hat{\psi}_n(x)$, where $\psi_n(x) := \frac{1}{n\sqrt{2\pi}} \exp(-x^2/(2n^2))$ is the density of a normal with mean zero and variance n. In particular, $\varphi_n = n(\widehat{\hat{\chi} * \psi_n})$ and using $\hat{\hat{f}}(x) = f(-x)$ for $f \in L^1(\mathbb{R})$ with $\hat{f} \in L^1(\mathbb{R})$, as above we obtain

$$\hat{\varphi_n}(\xi) = n(\widehat{\widehat{\hat{\chi} * \psi_n}})(\xi) = n\hat{\chi} * \psi_n(-\xi) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \hat{\chi}(y) \exp\Big(-\frac{1}{2n^2}(\xi + y)^2\Big)\, \mathrm{d}y. \tag{3.25}$$

Thus, for any $n \in \mathbb{N}$, φ_n indeed satisfies (3.22) and applying dominated convergence (and noting that the integrand on the right-hand side converges pointwise to $\hat{\chi}$) in (3.25) gives for any $\xi \in \mathbb{R}$, $\lim_{n\to\infty} \hat{\varphi_n}(\xi) = I$ as desired. □

3.2 Necessity of Conditions

In this section we assume that τ is a finite mean solution for the Skorokhod embedding problem corresponding to $\mu_i(\mathrm{d}x) = h_i(x)dx$ and study the associated Poisson equation $\mathcal{A}^*H = h_1 - h_0$. We show that

- there is a solution H. Using Proposition 3.4 we need to show $(\hat{h}_1 - \hat{h}_0)/\eta \in L^1(\mathbb{R})$. Integrability at infinity is only a consequence of the smoothness assumption on h_0, h_1 (Lemma 3.9) without using τ. Integrability at zero is a consequence of Dynkin's formula and the existence of τ (Lemma 3.10).
- $H \geq 0$ and $H \in L^1(\mathbb{R})$. This is a consequence of Dynkin's formula and the Riesz representation theorem.

The first lemma is the source of our assumptions on the regularity for h_0 and h_1.

Lemma 3.9. *Under Assumption 2.4 on h_0 and h_1 (as in Theorem 2.1) we have*

$$\int_{|u|>R} \left| \frac{\widehat{h_1}(u) - \widehat{h_0}(u)}{\eta(u)} \right| \mathrm{d}u < \infty$$

for some $R > 0$.

Proof. We consider all cases listed in Assumption 2.4 separately.

Type S: Since $|\widehat{h_i}(u)| \leq 1$ for $i = 0, 1$, for $R \geq 1$ one can use symmetry and the integrability assumption for $1/\eta$ to estimate

$$\int_{|u|>R} \left| \frac{\widehat{h_1}(u) - \widehat{h_0}(u)}{\eta(u)} \right| \mathrm{d}u \leq \int_{|u|>R} \frac{2}{|\eta(u)|}\, \mathrm{d}u \leq 4 \int_R^\infty \frac{1}{|\eta(u)|}\, \mathrm{d}u < \infty.$$

Type 0: By assumption there exists $R > 0, C > 0$ with

$$|\eta(u)| \geq C \quad \text{for} \quad |u| \geq R. \tag{3.26}$$

On the other hand, the regularity assumptions for type 0 guarantee that $h_i^{(2)} \in L^1(\mathbb{R})$ and so standard Fourier analysis gives

$$\left| u^2(\widehat{h}_1(u) - \widehat{h}_0(u)) \right| = \left| \widehat{h_1^{(2)}}(u) - \widehat{h_0^{(2)}}(u) \right| \leq \tilde{C} \tag{3.27}$$

for all $u \in \mathbb{R}$, where $\tilde{C} := \int_{\mathbb{R}} |h_1^{(2)}(x)| + |h_0^{(2)}(x)| \, \mathrm{d}x$.

Using (3.26) in the first and (3.27) in the second step yields

$$\begin{aligned} \int_{|u|>R} \left| \frac{\widehat{h}_1(u) - \widehat{h}_0(u)}{\eta(u)} \right| \mathrm{d}u &\leq \frac{1}{C} \int_{|u|>R} \left| \widehat{h}_1(u) - \widehat{h}_0(u) \right| \mathrm{d}u \\ &\leq \frac{\tilde{C}}{C} \int_{|u|>R} \frac{1}{|u|^2} \, \mathrm{d}u < \infty. \end{aligned}$$

Type D: By assumption there exists $R > 0$ such that $\widehat{h}_1(u) - \widehat{h}_0(u) = 0$ for all $|u| > R$ and so the integral is 0. □

Lemma 3.9 was independent of the Skorokhod embedding problem whereas the integrability around the origin indeed is a consequence of the SEP. The crucial ingredient of the proof is the use of Dynkin's formula for the complex exponential function:

Lemma 3.10. *Suppose τ is a finite mean solution to the Skorokhod embedding problem for $\mu_i(\mathrm{d}x) = h_i(x)\,\mathrm{d}x$. Then*

$$\eta(u)\mathbb{E}^{\mu_0}\left[\int_0^\tau e^{iuL_s} \, \mathrm{d}s\right] = \widehat{h}_1(u) - \widehat{h}_0(u) \tag{3.28}$$

for all $u \in \mathbb{R}$ and

$$\int_{|u|\leq R} \left| \frac{\widehat{h}_1(u) - \widehat{h}_0(u)}{\eta(u)} \right| \mathrm{d}u < \infty$$

for any $R > 0$.

Proof. Let $f, g \in C_b(\mathbb{R})$ be such that

$$M_t^f := f(L_t) - f(L_0) - \int_0^t g(L_s) \, \mathrm{d}s, \quad t \geq 0,$$

is a martingale. The optional sampling theorem implies that also $(M_{t\wedge\tau}^f)_{t\geq 0}$ is a martingale. In particular, for any $t \geq 0$,

$$\mathbb{E}^{\mu_0}\left[\int_0^{\tau\wedge t} g(L_s) \, \mathrm{d}s\right] = \mathbb{E}^{\mu_0}[f(L_{\tau\wedge t})] - \mathbb{E}^{\mu_0}[f(L_0)].$$

It holds that $\mathbb{P}^{\mu_0}(\lim_{t\to\infty} L_{\tau\wedge t} = L_\tau) = 1$, since $\mathbb{P}^{\mu_0}(\tau < \infty) = 1$ and L is quasi-left continuous. Using dominated convergence, $\mathbb{E}^{\mu_0}[\tau] < \infty$ and $f, g \in C_b(\mathbb{R})$, one may let $t \to \infty$ to obtain Dynkin's formula,

$$\mathbb{E}^{\mu_0}\left[\int_0^\tau g(L_s)\,\mathrm{d}s\right] = \mathbb{E}^{\mu_0}[f(L_\tau)] - \mathbb{E}^{\mu_0}[f(L_0)]. \tag{3.29}$$

For $u \in \mathbb{R}$, set

$$M_t^u := e^{iuL_t} - e^{iuL_0} - \eta(u)\int_0^t e^{iuL_r}\,\mathrm{d}r, \quad t \geq 0.$$

Then M^u is $(\mathcal{F}_t)_{t\geq 0}$-adapted and for any $t \geq 0$, M_t^u is a bounded random variable. Furthermore,

$$\begin{aligned}
\mathbb{E}^{\mu_0}[M_t^u - M_s^u|\mathcal{F}_s] &= e^{iuL_s}\mathbb{E}^{\mu_0}\left[e^{iu(L_t-L_s)} - 1 - \eta(u)\int_s^t e^{iu(L_r-L_s)}\,\mathrm{d}r\,\middle|\,\mathcal{F}_s\right] \\
&\overset{(2.1)}{=} e^{iuL_s}\left(e^{(t-s)\eta(u)} - 1 - \eta(u)\int_s^t e^{(r-s)\eta(u)}\,\mathrm{d}r\right) \\
&= 0
\end{aligned}$$

and therefore M^u is a complex-valued $(\mathcal{F}_t)_{t\geq 0}$-martingale. Thus Dynkin's formula (3.29) can be applied to $f(x) := e^{iux}$, $g(x) := \eta(u)f(x)$ and so

$$\begin{aligned}
\eta(u)\mathbb{E}^{\mu_0}\left[\int_0^\tau e^{iuL_r}\,\mathrm{d}r\right] &= \mathbb{E}^{\mu_0}\left[\int_0^\tau g(L_r)\,\mathrm{d}r\right] \\
&= \mathbb{E}^{\mu_0}[f(L_\tau)] - \mathbb{E}^{\mu_0}[f(L_0)] \\
&= \mathbb{E}^{\mu_0}[e^{iuL_\tau}] - \mathbb{E}^{\mu_0}[e^{iuL_0}].
\end{aligned}$$

This proves the first claim of the lemma. We can now deduce that $(\hat{h}_1 - \hat{h}_0)/\eta$ is integrable in compact sets. By the above we obtain

$$\left|\widehat{h_1}(u) - \widehat{h_0}(u)\right| = \left|\eta(u)\mathbb{E}^{\mu_0}\left[\int_0^\tau e^{iuL_r}\,\mathrm{d}r\right]\right| \leq |\eta(u)|\mathbb{E}^{\mu_0}[\tau],$$

and this implies

$$\int_{|u|\leq R}\left|\frac{\widehat{h_1}(u) - \widehat{h_0}(u)}{\eta(u)}\right|\,\mathrm{d}u \leq 2R\,\mathbb{E}^{\mu_0}[\tau] < \infty.$$

□

Combining the previous lemmas we proved that a finite mean solution to the Skorokhod embedding problem for "sufficiently smooth" densities implies $(\widehat{h_1} - \widehat{h_0})/\eta \in L^1(\mathbb{R})$ which, solving in Fourier domain, implies there is a solution H to the Poisson equation $\mathcal{A}^*H = h_1 - h_0$.

Now we can finish the proof by showing that existence of a finite mean solution to the Skorokhod embedding problem implies $H \geq 0$ and $H \in L^1(\mathbb{R})$.

Proof of Theorem 2.1 (necessity): We showed that $(\widehat{h_1} - \widehat{h_0})/\eta \in L^1(\mathbb{R})$ so the Poisson equation $\mathcal{A}^* H = h_1 - h_0$ can be solved using Proposition 3.4 as

$$H(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \frac{\widehat{h_1}(\xi) - \widehat{h_0}(\xi)}{\eta(\xi)} e^{-ix\xi} \, \mathrm{d}\xi, \quad x \in \mathbb{R}.$$

It remains to prove $H \geq 0$ and $H \in L^1(\mathbb{R})$:

Define the functional $\Lambda \colon C_c(\mathbb{R}) \to \mathbb{R}$ by

$$\Lambda(g) := \mathbb{E}^{\mu_0} \left[\int_0^\tau g(L_s) \, \mathrm{d}s \right].$$

Then $\Lambda(g) \geq 0$ for $g \geq 0$, Λ is linear and $|\Lambda(g)| \leq \|g\|_\infty \mathbb{E}^{\mu_0}[\tau]$. By the Riesz Representation Theorem (e.g. [Rud87, Theorem 2.14]), there exists a measure ν on $\mathcal{B}(\mathbb{R})$ such that for all $g \in C_c(\mathbb{R})$,

$$\mathbb{E}^{\mu_0} \left[\int_0^\tau g(L_s) \, \mathrm{d}s \right] = \Lambda(g) = \int_{\mathbb{R}} g(x) \, \nu(\mathrm{d}x). \tag{3.30}$$

Choosing $\{g_n\}_{n \in \mathbb{N}} \subset C_c(\mathbb{R})$, increasing monotonically to 1 with $g_n \geq 0$ and applying monotone convergence gives

$$\nu(\mathbb{R}) = \lim_{n \to \infty} \int_{\mathbb{R}} g_n(x) \, \nu(\mathrm{d}x) = \lim_{n \to \infty} \mathbb{E}^{\mu_0} \left[\int_0^\tau g_n(L_s) \, \mathrm{d}s \right] = \mathbb{E}^{\mu_0}[\tau] < \infty.$$

Thus ν is a finite measure and by dominated convergence, (3.30) holds for all $g \in C_b(\mathbb{R})$. Inserting $g(x) := e^{iux}$ for $u \in \mathbb{R}$ in (3.30) and using (3.28) yields

$$\hat{\nu}(u) = \int_{\mathbb{R}} e^{iux} \, \nu(\mathrm{d}x) = \mathbb{E}^{\mu_0} \left[\int_0^\tau e^{iuL_s} \, \mathrm{d}s \right] = \frac{\widehat{h_1}(u) - \widehat{h_0}(u)}{\eta(u)},$$

which is integrable by Lemma 3.9 and Lemma 3.10. Hence, e.g. by [Sat99, Proposition 2.5 (xii)], ν is absolutely continuous with respect to the Lebesgue measure and has a (non-negative) bounded continuous density given by

$$x \mapsto \frac{1}{2\pi} \int_{\mathbb{R}} \frac{\widehat{h_1}(\xi) - \widehat{h_0}(\xi)}{\eta(\xi)} e^{-ix\xi} \, \mathrm{d}\xi, \quad x \in \mathbb{R}.$$

But this function is identical to H and thus $\nu(\mathrm{d}x) = H(x) \, \mathrm{d}x$. In particular, $H \geq 0$ and $H \in L^1(\mathbb{R})$. □

3.3 Sufficiency of Conditions

Under the assumptions of Theorem 2.1 we now construct a finite mean stopping time with $L_\tau \sim h_1(x) \, \mathrm{d}x$ under the initial condition $L_0 \sim h_0(x) \, \mathrm{d}x$. We refer the reader to the sketch in the introduction of this thesis to follow more easily the construction of τ. During the proof we refer to auxiliary time-change arguments and uniqueness results for Fokker-Planck equations as developed in the appendix (Appendix B and C). Also we refer to assumptions as formulated in Appendix A.

Let $\mathcal{D} := C_c^\infty(\mathbb{R})$ and define the action of the Lévy generator $\mathcal{A}\colon \mathcal{D} \to C_0(\mathbb{R})$ for $u \in \mathcal{D}$ via (3.1). Furthermore, taking into account the definitions from Theorem 2.1, let

$$\phi(t,x) := (1-t)h_0(x) + th_1(x), \tag{3.31}$$

$$\sigma(t,x) := \frac{H(x)}{(1-t)h_0(x) + th_1(x)}, \quad (t,x) \in [0,1] \times \mathbb{R}, \tag{3.32}$$

and, under $\mathbb{P}^{\mu_0}$,

$$\delta(s) := \inf\big\{t \in [0,\rho) \,:\, \Delta(t) \geq s\big\} \wedge \rho, \quad s \in [0,1], \tag{3.33}$$

where

$$\Delta(t) := 1 - e^{G(t)} + \int_0^t e^{(G(t)-G(r))} \frac{h_1(L_r)}{H(L_r)} \,\mathrm{d}r, \quad t \in [0,\rho), \tag{3.34}$$

with

$$\rho := \inf\{t \in [0,\infty) \,:\, H(L_t) = 0\} \quad \text{and} \quad G(t) := \int_0^t \frac{h_1(L_r) - h_0(L_r)}{H(L_r)} \,\mathrm{d}r.$$

The proof is split in two main steps: First we assume in addition that h_0 and h_1 are such that σ is bounded and argue as sketched in the introduction. Then, for σ unbounded, we approximate h_i by $h_i^{(\varepsilon)}$ with associated $\sigma^{(\varepsilon)}$ bounded and deduce Theorem 2.1.

Proof of Theorem 2.1 (sufficiency ***if*** $\boldsymbol{\sigma}$ ***is bounded****).* For the proof the following statements are established:

(i) $(\delta(s))_{s\in[0,1]}$ constitutes a family of $(\mathcal{F}_t)_{t\geq 0}$-stopping times satisfying $\mathbb{P}^{\mu_0}$-*a.s.*,

$$\delta(s) = \int_0^s \sigma(u, L_{\delta(u)}) \,\mathrm{d}u, \quad s \in [0,1], \tag{3.35}$$

(ii) for any $s \in [0,1]$, the law of $L_{\delta(s)}$ under $\mathbb{P}^{\mu_0}$ is $\phi(s,x)\,\mathrm{d}x$,

(iii) $\mathbb{E}^{\mu_0}[\delta(1)] = \int_{\mathbb{R}} H(x)\,\mathrm{d}x < \infty$.

Theorem 2.1 can then be deduced from (i)-(iii) by setting $\tau := \delta(1)$ because $\phi(1,\cdot) = h_1$ by construction. Note that the stopping time looks slightly different here than in the statement of Theorem 2.1. Both representations are equal because from (3.34) one obtains

$$\begin{aligned} \Delta(t) \geq 1 \quad &\Longleftrightarrow \quad -e^{G(t)} + \int_0^t e^{(G(t)-G(r))} \frac{h_1(L_r)}{H(L_r)} \,\mathrm{d}r \geq 0 \\ &\Longleftrightarrow \quad \int_0^t e^{-G(r)} \frac{h_1(L_r)}{H(L_r)} \,\mathrm{d}r \geq 1, \end{aligned}$$

and, hence, the claimed representation of $\tau = \delta(1)$ as generalized inverse in Theorem 2.1 and in (3.33) are equal.

The proof of (i)-(iii) proceeds roughly as follows: Firstly, it is proved that L and σ satisfy the assumptions of Lemma B.2 and that (3.34) is the solution to the differential equation (B.5), so Lemmas B.2 and B.6 imply (i). Based on Lemma B.9, one then verifies that $(\phi(s,x)\,\mathrm{d}x)_{s\in[0,1]}$ and the marginals of $L_{\delta(\cdot)}$ are solutions to the Fokker-Planck equation (C.2). Then from the uniqueness result Theorem C.1 it follows that $L_{\delta(s)}$ indeed has the law $\phi(s,x)\,\mathrm{d}x$, i.e. (ii). Combining the representation of $\delta(1)$ established in (i) with (ii) and the fact that $\sigma(t,x)\phi(t,x) = H(x)$ for all $t \in [0,1]$ and $x \in \mathbb{R}$ one easily obtains (iii).

Verification of (i): The claim is a consequence of Lemma B.2 and its proof (set $t_0 = 1$ and compare (B.5) and (B.6) for the formula of δ in terms of Δ) and Lemma B.6. We only need to verify the conditions of Lemma B.2 and then solve

$$\Delta(s) = \int_0^s \sigma(\Delta(r), L_r)^{-1}\,\mathrm{d}r, \quad s \in [0, \delta(t_0)),$$

for our choice of σ from (3.32). It is the particular form of the denominator which allows us to solve (B.5) explicitly and get the formula for Δ as claimed in (3.34).

By Proposition 3.4, $H \in D(\mathcal{A}^*) \subset C_0(\mathbb{R})$, where $\mathcal{A}^*\colon D(\mathcal{A}^*) \to C_0(\mathbb{R})$ denotes the adjoint (see Subsection 3.1.1 and (3.3)). Since also $H \in D(\mathcal{A})$ by Proposition 3.5, H is regular for (the law of) L by Proposition B.5. Since h_0 and h_1 are assumed positive and continuous, for any $K \subset \mathbb{R}$ compact, there exist $C_0, C_1 > 0$ such that

$$C_0 \le h_i(x) \le C_1, \quad x \in K,\, i = 0, 1. \tag{3.36}$$

Set $\tilde{\sigma}(t,x) := 1/\phi(t,x)$ for $(t,x) \in [0,1] \times K$. Then (3.36) implies that for any $(t,x) \in [0,1] \times K$, one has $1/C_1 \le \tilde{\sigma}(t,x) \le 1/C_0$ and

$$|\tilde{\sigma}(t,x) - \tilde{\sigma}(s,x)| = \frac{1}{\phi(s,x)\phi(t,x)}|\phi(s,x) - \phi(t,x)| \le \frac{2C_1}{C_0^2}|t-s|.$$

In particular, Assumption A.6 (ii) holds and $\sigma = H\tilde{\sigma}$ (see (3.32) above) satisfies Assumption A.6. Since σ is also assumed to be bounded, Lemma B.2 and Lemma B.6 can be applied. Hence, the random times defined by (B.6), (B.5) and (B.3) are stopping times and (B.1) holds. As the definitions (B.6) and (3.33) coincide, in order to show (i), it thus suffices to show that ρ and Δ in (3.33) coincide $\mathbb{P}^{\mu_0}$-a.s. with (B.3) and (B.5).

Since H is regular for the law of L (as argued above), ρ in (3.33) is equal to (B.3) (see Definition A.4). Furthermore, as shown in Lemma B.2, the solution to the Carathéodory differential equation (B.5) is $\mathbb{P}^{\mu_0}$-a.s. unique. Thus it suffices to show that $\mathbb{P}^{\mu_0}$-a.s., Δ defined by (3.34) is a solution to (B.5), i.e. that $\mathbb{P}^{\mu_0}$-a.s.

$$\Delta(t) = \int_0^t \sigma(\Delta(r), L_r)^{-1}\,\mathrm{d}r, \quad t \in [0, \delta(1)), \tag{3.37}$$

holds. Inserting (3.32) in (3.37) yields

$$\Delta(t) = \int_0^t \frac{\Delta(r)(h_1(L_r) - h_0(L_r))}{H(L_r)}\,\mathrm{d}r + \int_0^t \frac{h_0(L_r)}{H(L_r)}\,\mathrm{d}r, \quad t \in [0, \delta(1)). \tag{3.38}$$

On the other hand, $\delta(1) \le \rho$ and $\mathbb{P}^{\mu_0}$-a.s. the candidate solution Δ in (3.34) is absolutely continuous on every closed subinterval of $[0,\rho)$ and

$$\begin{aligned}\dot{\Delta}(t) &= -\dot{G}(t)e^{G(t)} + \dot{G}(t)\int_0^t e^{(G(t)-G(r))}\frac{h_1(L_r)}{H(L_r)}\,\mathrm{d}r + \frac{h_1(L_t)}{H(L_t)}\\ &= \dot{G}(t)(\Delta(t)-1) + \frac{h_1(L_t)}{H(L_t)}\\ &= \Delta(t)\frac{h_1(L_t)-h_0(L_t)}{H(L_t)} + \frac{h_0(L_t)}{H(L_t)},\end{aligned}$$

for almost every $t \in [0,\rho)$. This is equivalent to (3.34) being a solution to (3.38) on $[0,\rho)$ as desired.

Verification of (ii): Firstly, in (i) it has been verified that Assumption A.6 holds and by Lemma 3.1, Assumption A.7 is satisfied. Secondly, Assumption A.11 holds and, as argued above, Assumption A.12 is satisfied. Thus, Lemma B.9 and Theorem C.1 can be applied. This shows that $(\tilde{p}(s,\cdot))_{s\in[0,1]}$ is the unique solution to the Fokker-Planck equation (B.16), where $\tilde{p}(s,\cdot)$ is the law of $L_{\delta(s)}$ under $\mathbb{P}^{\mu_0}$. Thus, in order to establish (ii), it suffices to verify that $(p(s,\cdot))_{s\in[0,1]}$ with $p(s,\mathrm{d}x) = \phi(s,x)\,\mathrm{d}x$ also satisfies the Fokker-Planck equation (C.2) with (C.1).

Inserting $p(s,\mathrm{d}x) = \phi(s,x)\,\mathrm{d}x$ with ϕ from (3.31) into the left-hand side of (C.2), using $H \in D(\mathcal{A}^*)$, $\mathcal{A}^*H = h_1 - h_0$ (by (2.4) and Proposition 3.4) and $H \in L^1(\mathbb{R})$, $h_1 - h_0 \in L^1(\mathbb{R})$ (by assumption), Lemma 3.3 gives

$$\begin{aligned}\int_{\mathbb{R}} f(x)\,p(t,\mathrm{d}x) - \int_{\mathbb{R}} f(x)\,\mu_0(\mathrm{d}x) &\overset{(3.31)}{=} \int_{\mathbb{R}} f(x)t(h_1-h_0)(x)\,\mathrm{d}x\\ &= \int_0^t\int_{\mathbb{R}} \mathcal{A}^*H(x)f(x)\,\mathrm{d}x\,\mathrm{d}s\\ &\overset{(3.4)}{=} \int_0^t\int_{\mathbb{R}} H(x)\mathcal{A}f(x)\,\mathrm{d}x\,\mathrm{d}s\\ &= \int_0^t\int_{\mathbb{R}} \sigma(s,x)\mathcal{A}f(x)\,p(s,\mathrm{d}x)\,\mathrm{d}s,\end{aligned}$$

where the last step is just the definition (3.32) and (3.31). Hence, by Theorem C.1 and our argument above we may indeed conclude (ii).

Verification of (iii): Having verified that the marginals of $L_{\delta(\cdot)}$ are given as in (3.31), we may apply the representation of δ as solution to an integral equation (established in (i)), Tonelli's Theorem and the definition of σ and ϕ to see, using (3.35),

$$\begin{aligned}\mathbb{E}^{\mu_0}[\delta(1)] &= \mathbb{E}^{\mu_0}\left[\int_0^1 \sigma(u, L_{\delta(u)})\,\mathrm{d}u\right]\\ &= \int_0^1 \mathbb{E}^{\mu_0}[\sigma(u, L_{\delta(u)})]\,\mathrm{d}u\\ &= \int_0^1\int_{\mathbb{R}} \sigma(u,x)\phi(u,x)\,\mathrm{d}x\,\mathrm{d}u\\ &= \int_{\mathbb{R}} H(x)\,\mathrm{d}x.\end{aligned}$$

The right-hand side is finite by assumption and so $\tau = \delta(1)$ has finite mean. □

Proof of Theorem 2.1 (sufficiency). To finish the proof of sufficiency we need to remove the assumption that h_0 and h_1 are such that $\sigma = H/\phi$ is bounded using a truncation procedure. For this sake we shift the densities in order to shift down ϕ.

Approximate stopping times $\boldsymbol{\delta^{(\varepsilon)}}$: Since $H \in L^1(\mathbb{R})$ by assumption, $C := \int_{\mathbb{R}} H(x)\,\mathrm{d}x$ is well-defined and $p := H/C$ is a probability density on $\mathbb{R}$. For any $\varepsilon \in (0,1)$ we define

$$\begin{aligned} h_i^{(\varepsilon)}(x) &:= (1-\varepsilon)h_i(x) + \varepsilon p(x), \quad i = 0,1, \\ \phi^{(\varepsilon)}(t,x) &:= (1-t)h_0^{(\varepsilon)}(x) + t h_1^{(\varepsilon)}(x), \\ H^{(\varepsilon)}(x) &:= (1-\varepsilon)H(x), \end{aligned} \tag{3.39}$$

for $(t,x) \in [0,1] \times \mathbb{R}$, and the approximation to σ by $\sigma^{(\varepsilon)} := \frac{H^{(\varepsilon)}}{\phi^{(\varepsilon)}}$. Then, for any $\varepsilon \in (0,1)$, $\phi^{(\varepsilon)} \geq \varepsilon p$ and so, for any $t \in [0,1]$ and $x \in \mathbb{R}$ with $H(x) \neq 0$,

$$\sigma^{(\varepsilon)}(t,x) \leq \frac{H^{(\varepsilon)}(x)}{\varepsilon p(x)} = \frac{(1-\varepsilon)C}{\varepsilon}.$$

Thus, for any $\varepsilon \in (0,1)$, $\sigma^{(\varepsilon)}$ is bounded. Furthermore, $h_i^{(\varepsilon)} \in C_0(\mathbb{R})$, $h_i^{(\varepsilon)}(x) > 0$ and

$$h_1^{(\varepsilon)}(x) - h_0^{(\varepsilon)}(x) = (1-\varepsilon)(h_1(x) - h_0(x)) \tag{3.40}$$

for any $x \in \mathbb{R}$, $i = 0,1$ and $\varepsilon \in [0,1)$. In particular, $H^{(\varepsilon)}$ satisfies the Poisson equation $\mathcal{A}^* H^{(\varepsilon)} = h_1^{(\varepsilon)} - h_0^{(\varepsilon)}$. Since $H^{(\varepsilon)} = (1-\varepsilon)H$, the following properties are inherited from H: $H^{(\varepsilon)}$ is non-negative, $H^{(\varepsilon)} \in L^1(\mathbb{R})$ and $H^{(\varepsilon)} \in D(\mathcal{A})$ by Proposition 3.5.

Thus Step (i) of the bounded case applied with $h_0^{(\varepsilon)}, h_1^{(\varepsilon)}$ instead of h_0, h_1 shows that, for any $\varepsilon \in (0,1)$,

$$\delta^{(\varepsilon)}(s) := \inf\{t \in [0,\rho) \,:\, \Delta^{(\varepsilon)}(t) \geq s\} \wedge \rho, \quad s \in [0,1], \tag{3.41}$$

constitutes a family of $(\mathcal{F}_t)_{t\geq 0}$-stopping times, where

$$\begin{aligned} \rho^{(\varepsilon)} &:= \inf\{t \in [0,\infty) \,:\, H^{(\varepsilon)}(L_t) = 0\}, \\ \Delta^{(\varepsilon)}(t) &:= 1 - e^{G^{(\varepsilon)}(t)} + \int_0^t e^{(G^{(\varepsilon)}(t) - G^{(\varepsilon)}(r))} \frac{h_1^{(\varepsilon)}(L_r)}{H^{(\varepsilon)}(L_r)}\,\mathrm{d}r, \\ G^{(\varepsilon)}(t) &:= \int_0^t \frac{h_1^{(\varepsilon)}(L_r) - h_0^{(\varepsilon)}(L_r)}{H^{(\varepsilon)}(L_r)}\,\mathrm{d}r, \end{aligned} \tag{3.42}$$

for $t \in [0,\rho)$ and we note that $\rho = \rho^{(\varepsilon)}$, since $H^{(\varepsilon)} = (1-\varepsilon)H$.

Some simplifications: The choice of $\delta^{(\varepsilon)}$ is very convenient as the mean and $\Delta^{(\varepsilon)}$ simplify in a neat way. By Step (ii) of the bounded case, for any $\varepsilon \in (0,1)$, $L_{\delta^{(\varepsilon)}(s)}$ has law $\phi^{(\varepsilon)}(s,x)\,\mathrm{d}x$ under $\mathbb{P}^{\mu_0}$, for any $s \in [0,1]$ and by (iii),

$$\mathbb{E}^{\mu_0}[\delta^{(\varepsilon)}(1)] = \int_{\mathbb{R}} H^{(\varepsilon)}(x)\,\mathrm{d}x = (1-\varepsilon)\int_{\mathbb{R}} H(x)\,\mathrm{d}x. \tag{3.43}$$

Next, (3.40) and $H^{(\varepsilon)} = (1-\varepsilon)H$ imply that $G^{(\varepsilon)} = G$ and thus from (3.42) one obtains, for any $t \in [0,\rho)$, a simple formula for $\Delta^{(\varepsilon)}$:

$$\begin{aligned}\Delta^{(\varepsilon)}(t) &= 1 - e^{G(t)} + \int_0^t e^{(G(t)-G(r))} \frac{h_1^{(\varepsilon)}(L_r)}{(1-\varepsilon)H(L_r)} \,\mathrm{d}r \\ &\overset{(3.39)}{=} 1 - e^{G(t)} + \int_0^t e^{(G(t)-G(r))} \frac{h_1(L_r)}{H(L_r)} \,\mathrm{d}r + \frac{\varepsilon}{(1-\varepsilon)C} e^{G(t)} \int_0^t e^{-G(r)} \,\mathrm{d}r \\ &= \Delta(t) + \frac{\varepsilon}{(1-\varepsilon)C} e^{G(t)} \int_0^t e^{-G(r)} \,\mathrm{d}r. \end{aligned} \tag{3.44}$$

Limiting stopping time $\boldsymbol{\delta}$: Set $\Delta^{(0)} := \Delta$ and $\delta^{(0)} := \delta$ (from (3.33)). We need to show that $\delta^{(0)}$ is a stopping time and we need to compute the distribution of L_δ.

Since $f\colon [0,1) \to \mathbb{R}$, $f(x) := x/(1-x)$, is increasing and $f(0) = 0$ the last decomposition of $\Delta^{(\varepsilon)}$ shows that $\mathbb{P}^{\mu_0}$-a.s. for any $0 \le \varepsilon < \tilde{\varepsilon} < 1$ and all $t \in [0,\rho)$, $\Delta^{(\varepsilon)}(t) \le \Delta^{(\tilde{\varepsilon})}(t)$ and hence, from (3.41), $\delta^{(\varepsilon)}(s) \ge \delta^{(\tilde{\varepsilon})}(s)$ for all $s \in [0,1]$. In particular, for any $s \in [0,1]$, $(\delta^{(\frac{1}{n})}(s))_{n\in\mathbb{N}}$ is a sequence of stopping times with $\delta^{(\frac{1}{n})}(s) \le \delta^{(\frac{1}{n+1})}(s) \le \delta(s)$ for any $n \in \mathbb{N}$. Thus $\tilde{\delta}(s) := \lim_{n\to\infty} \delta^{(\frac{1}{n})}(s) \in [0,\infty]$ exists $\mathbb{P}^{\mu_0}$-a.s. and $\tilde{\delta}(s) \le \delta(s)$. Since $(\mathcal{F}_t)_{t\ge 0}$ is right-continuous, $\tilde{\delta}(s)$ is an $(\mathcal{F}_t)_{t\ge0}$-stopping time. By (3.43) and monotone convergence,

$$\mathbb{E}^{\mu_0}[\tilde{\delta}(1)] = \lim_{n\to\infty} \mathbb{E}^{\mu_0}[\delta^{(\frac{1}{n})}(1)] = \int_{\mathbb{R}} H(x)\,\mathrm{d}x < \infty. \tag{3.45}$$

In particular, $\tilde{\delta}(s) \le \tilde{\delta}(1) < \infty$, $\mathbb{P}^{\mu_0}$-a.s.

If $\tilde{\delta}(s) \ge \rho$, then $\delta(s) \le \rho$ and $\tilde{\delta}(s) \le \delta(s)$ imply $\tilde{\delta}(s) = \delta(s)$. Otherwise $\tilde{\delta}(s) < \rho$ and thus

$$\int_0^{\tilde{\delta}(s)} \frac{1}{H(L_s)} \,\mathrm{d}s < \infty. \tag{3.46}$$

Using continuity and the decomposition (3.44) one obtains

$$\begin{aligned}\Delta(\tilde{\delta}(s)) &= \lim_{n\to\infty} \Delta(\delta^{(\frac{1}{n})}(s)) \\ &= \lim_{n\to\infty} \left(\Delta^{(\frac{1}{n})}(\delta^{(\frac{1}{n})}(s)) - \frac{e^{G(\delta^{(\frac{1}{n})}(s))}\frac{1}{n}}{(1-\frac{1}{n})C} \int_0^{\delta^{(\frac{1}{n})}(s)} e^{-G(r)}\,\mathrm{d}r \right) \ge s, \end{aligned} \tag{3.47}$$

where the last inequality follows from $\Delta^{(\frac{1}{n})}(\delta^{(\frac{1}{n})}(s)) \ge s$ (by definition), the fact that on $\{\tilde{\delta}(s) < \rho\}$, G is bounded on the compact interval $[0, \tilde{\delta}(s)]$ (which follows directly from (3.46)) and $\delta^{(\frac{1}{n})}(s) \le \tilde{\delta}(s)$ for all $n \in \mathbb{N}$. The definition (3.33) and inequality (3.47) imply $\tilde{\delta}(s) \ge \delta(s)$ also on $\{\tilde{\delta}(s) < \rho\}$. We conclude that $\mathbb{P}^{\mu_0}$-a.s., $\tilde{\delta}(s) = \delta(s)$ and the sequence of stopping times $\{\delta^{(\frac{1}{n})}(s)\}_{n\in\mathbb{N}}$ increases monotonically to $\delta(s)$. Hence, $\delta^{(0)} = \delta$ is a stopping time and by quasi-left continuity of L, [EK86, Chap. 4, Thm. 3.12] implies

$$\lim_{n\to\infty} L_{\delta^{(\frac{1}{n})}(s)} = L_{\delta(s)}, \quad \mathbb{P}^{\mu_0}\text{-a.s.}$$

In particular, for any $f \in C_0(\mathbb{R})$,

$$\begin{aligned}\mathbb{E}^{\mu_0}[f(L_{\delta(s)})] = \lim_{n\to\infty} \mathbb{E}^{\mu_0}[f(L_{\delta^{(\frac{1}{n})}(s)})] &= \lim_{n\to\infty} \int_{\mathbb{R}} f(x)\phi^{(\frac{1}{n})}(s,x)\,\mathrm{d}x \\ &= \int_{\mathbb{R}} f(x)\phi(s,x)\,\mathrm{d}x,\end{aligned}$$

which implies that (ii) (and (iii), as seen from (3.45)) has been established also without the assumption that σ is bounded. □

A Notation and Definitions

The aim of the appendix is to generalize time-change methods and uniqueness results for Fokker-Planck equations from time-homogeneous generators to the case of time-inhomogeneous generators of the form $\sigma\mathcal{A}$, where $\mathcal{A}$ is time-independent but σ is time-dependent. To the best of our knowledge the results we need are not available in the literature apart from special cases such as $\mathcal{A}$ being the Laplacian (see Appendix C for a discussion).

In what follows the reader might keep in mind the time-inhomogeneous example to which the theory is applied in the main body of the chapter: $\mathcal{L} = \sigma\mathcal{A}$, where $\mathcal{A}$ is the generator of a Lévy process, H solves the Poisson equation $\mathcal{A}^*H = h_1 - h_0$ and

$$\sigma(t,x) = \frac{H(x)}{th_1(x) + (1-t)h_0(x)}. \tag{A.1}$$

The example motivates us to study the very specific choice $\sigma = \tilde{\sigma}H$ below with separate hypothesis on $\tilde{\sigma}$ and H.

Even though the main body of this chapter is specialized to Lévy processes the appendix is in more generality to allow for later extensions of our approach to other classes of processes, e.g. diffusion processes or Markov chains.

A.1 Notation

The stochastic basis consists of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a filtration $(\mathcal{F}_t)_{t\geq 0}$ satisfying the usual conditions of completeness and right-continuity. The space of continuous functions $f\colon \mathbb{R} \to \mathbb{R}$ satisfying $\lim_{|x|\to\infty} f(x) = 0$ is denoted by $C_0(\mathbb{R})$. For $n \in \mathbb{N}$ let $C_0^n(\mathbb{R})$ be the subset of functions $f \in C_0(\mathbb{R})$ such that f is n-times differentiable and all derivatives of order less or equal to n belong to $C_0(\mathbb{R})$ and we set $C_0^\infty(\mathbb{R}) := \bigcap_{n\in\mathbb{N}} C_0^n(\mathbb{R})$. The spaces of functions with compact support $C_c(\mathbb{R})$, $C_c^n(\mathbb{R})$ and $C_c^\infty(\mathbb{R})$ are defined analogously.

Next, fix a locally compact, complete, separable metric space (E,d) and denote by $\mathcal{B}(E)$ its Borel σ-algebra. The space $D_E[0,\infty)$ stands for all maps $\omega\colon [0,\infty) \to E$ which are right-continuous and have a left-limit at each point $t \in [0,\infty)$ (short: RCLL paths). For $x \in E$ and $\varepsilon > 0$, set $B_\varepsilon(x) := \{y \in E : d(x,y) < \varepsilon\}$. $\mathcal{P}(E)$ denotes the set of probability measures on $(E, \mathcal{B}(E))$. $B(E)$

denotes the space of real-valued, bounded, measurable functions on E and $\|\cdot\|$ is the sup-norm. $C_0(E)$ denotes the space of continuous functions that vanish at infinity and $C_b(E)$ the space of bounded continuous functions on E. A sequence $\{f_n\}_{n\in\mathbb{N}} \subset B(E)$ converges *boundedly and pointwise* to $f \in B(E)$ (denoted by $\text{bp-lim}_{n\to\infty} f_n = f$) if $\sup_n \|f_n\| < \infty$ and $\lim_{n\to\infty} f_n(x) = f(x)$ for all $x \in E$. $U \subset B(E)$ is called *bp-closed*, if $\{f_n\}_{n\in\mathbb{N}} \subset U$ and $\text{bp-lim}_{n\to\infty} f_n = f$ implies $f \in U$. For $V \subset B(E)$, we define bp-closure(V) as the smallest subset of $B(E)$ which is bp-closed and contains V. A sequence $\{(f_n, g_n)\}_{n\in\mathbb{N}} \subset B(E) \times B(E)$ is said to converge to $(f,g) \in B(E) \times B(E)$ boundedly and pointwise (denoted by $\text{bp-lim}_{n\to\infty}(f_n, g_n) = (f,g)$) if $\text{bp-lim}_{n\to\infty} f_n = f$ and $\text{bp-lim}_{n\to\infty} g_n = g$. The definitions of bp-closed and bp-closure are then defined analogously for subsets of $B(E) \times B(E)$.

Let $D(\mathcal{L}) \subset C_b(E)$ and $\mathcal{L}\colon D(\mathcal{L}) \to C_b(E)$ linear. $(D(\mathcal{L}), \mathcal{L})$ is said to be *conservative* if

$$\text{there exists } \{h_n\}_{n\in\mathbb{N}} \subset D(\mathcal{L}) \text{ such that } (1,0) = \text{bp-lim}_{n\to\infty}(h_n, \mathcal{L}h_n). \tag{A.2}$$

For a stochastic process $(X_t)_{t\geq 0}$ we set $\mathcal{F}_t^X := \sigma(X_s : s \leq t)$. A *solution to the martingale problem for* $(\mathcal{L}, \mu)$ is a progressively measurable E-valued stochastic process $(X_t)_{t\geq 0}$ defined on some probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ such that for each $h \in D(\mathcal{L})$, the process

$$h(X_t) - h(X_0) - \int_0^t \mathcal{L}h(X_s)\,\mathrm{d}s, \quad t \geq 0, \tag{A.3}$$

is an $(\mathcal{F}_t^X)_{t\geq 0}$-martingale and $\tilde{\mathbb{P}} \circ X_0^{-1} = \mu$. *Uniqueness* is said to hold for the martingale problem for $(\mathcal{L}, \mu)$ if any two solutions X, $\tilde{X}$ have the same finite-dimensional distributions. The martingale problem for $(\mathcal{L}, \mu)$ is said to be *well-posed* if there exists a solution and uniqueness holds. A solution to the $D_E[0,\infty)$-martingale problem (or RCLL-martingale problem) for $(\mathcal{L}, \mu)$ is an RCLL process that is a solution to the martingale problem for $(\mathcal{L}, \mu)$. *Uniqueness* is said to hold for the RCLL-martingale problem for $(\mathcal{L}, \mu)$ if any two solutions to the RCLL-martingale problem for $(\mathcal{L}, \mu)$ have the same law on $D_E[0,\infty)$. The RCLL-martingale problem for $(\mathcal{L}, \mu)$ is said to be *well-posed* if there exists a solution and uniqueness holds.

Similarly, for a linear operator $(D(\mathcal{A}), \mathcal{A})$ and $\sigma\colon [0,\infty) \times E \to [0,\infty)$ measurable, a solution to the (time-inhomogeneous) martingale problem for $(\sigma\mathcal{A}, \mu)$ is a progressively measurable E-valued process X defined on some probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ such that for each $f \in D(\mathcal{A})$ the process

$$f(X_t) - f(X_0) - \int_0^t \sigma(s, X_s)\mathcal{A}f(X_s)\,\mathrm{d}s, \quad t \geq 0, \tag{A.4}$$

is an $(\mathcal{F}_t^X)_{t\geq 0}$-martingale and $\tilde{\mathbb{P}} \circ X_0^{-1} = \mu$. The function σ is called the *local speed function* of X. Uniqueness, well-posedness and the corresponding concepts among $D_E[0,\infty)$-processes are defined analogously to the time-homogeneous case. For $\mu \in \mathcal{P}([0,\infty))$ and $\nu \in \mathcal{P}(E)$ we write $\mu \otimes \nu$ for the product measure generated by

μ and ν on $[0,\infty)\times E$. If F is a measurable space, δ_x denotes the Dirac measure at $x \in F$. For $V \subset B(E)$, span(V) denotes the smallest linear subspace of $B(E)$ containing V, i.e. the set of all finite linear combinations of elements of V.

Remark A.1. Our definitions and notation are precisely as in [EK86] with the following two exceptions: Firstly, in [EK86] $(D(\mathcal{L}),\mathcal{L})$ is said to be conservative if

$$(1,0) \in \text{bp-closure}(\{(h,\mathcal{L}h) \,:\, h \in D(\mathcal{L})\}). \tag{A.5}$$

While our requirement (A.2) implies (A.5), the converse is not true in general, cf. [EK86, Chap. 3, Sec. 4].

Secondly, (not necessarily RCLL) solutions to martingale problems are required to be progressively measurable in our context, as in [Kur98].

Remark A.2. The motivation for the definition of bp-closure is as follows: Suppose $(D(\mathcal{L}),\mathcal{L})$ is a linear operator on $C_b(E)$ and X is a solution to the martingale problem for $(\mathcal{L},\mu)$ for some $\mu \in \mathcal{P}(E)$. Then by the dominated convergence theorem, the set of functions $(h,g) \in B(E)\times B(E)$, for which

$$h(X_t) - h(X_0) - \int_0^t g(X_s)\,\mathrm{d}s, \quad t \geq 0, \tag{A.6}$$

is an $(\mathcal{F}_t^X)_{t\geq 0}$-martingale, is bp-closed. Therefore, (A.6) is a martingale for all $(h,g) \in \text{bp-closure}(\{(h,\mathcal{L}h) \,:\, h \in D(\mathcal{L})\})$.

Finally, we provide two definitions. The first one is a (strong) recurrence property, the second is essential for studying uniqueness of time-change equations. Denote by Z the coordinate process on $D_E[0,\infty)$.

Definition A.3. A probability measure P on $(D_E[0,\infty),\mathcal{B}(D_E[0,\infty))$ is called recurrent if

$$P\left(\int_0^\infty \mathbb{1}_{B_a(Z_0)}(Z_t)\,\mathrm{d}t = \infty\right) = 1 \quad \text{for every } a > 0. \tag{A.7}$$

Definition A.4. Let $H\colon E \to [0,\infty)$ measurable and $P \in \mathcal{P}(D_E[0,\infty))$. H is called regular for P if P-a.s.

$$\inf\left\{s \in [0,\infty) \,:\, \int_0^s H(Z_u)^{-1}\,\mathrm{d}u = \infty\right\} = \rho \quad \text{and} \quad H(Z_\rho) = 0 \text{ on } \{\rho < \infty\}$$

where

$$\rho := \inf\left\{s \in [0,\infty) \,:\, H(Z_s) = 0\right\}.$$

Example A.5. Let $E = \mathbb{R}$ and Z a P-Brownian motion. Suppose $H\colon \mathbb{R} \to [0,\infty)$ satisfies

$$\{x \in \mathbb{R} \,:\, H(x) = 0\} = I(H), \tag{A.8}$$

where $I(H)$ is the closed set

$$I(H) := \left\{x \in \mathbb{R} \,:\, \forall\varepsilon > 0 : \int_{x-\varepsilon}^{x+\varepsilon} \frac{\mathrm{d}y}{H(y)} = \infty\right\}. \tag{A.9}$$

Then H is regular for P. Indeed, since $I(H)$ in (A.9) is closed and (A.8) holds, $H^{-1}(\{0\})$ is closed. Hence, for P-a.e. ω with $\rho(\omega) < \infty$, $H(Z_{\rho(\omega)}(\omega)) = 0$ by (right-)continuity. Hence, the second part of the definition is established and the first part follows directly from (A.8) and [KS91, Chap. 5, Lem. 5.2].

A.2 Assumptions

The following assumptions are used at different places throughout the chapter. Recall that our standard application will assume $\mathcal{A}$ to be the generator of a Lévy process, H the solution to the Poisson equation $\mathcal{A}^*H = h_1 - h_0$ and $\tilde{\sigma}(t,x) = 1/(th_1(x) + (1-t)h_0(x))$.

Our set of assumptions is split in such a way that we can distinguish as good as possible between assumptions on the stochastic process and on the densities h_0, h_1.

Assumption A.6 (Regularity of σ)**.** *Let $t_0 > 0$ and $\sigma\colon [0,\infty) \times E \to [0,\infty)$ be of the form $\sigma(t,x) := H(x)\tilde{\sigma}(t,x)$ for $(t,x) \in [0,\infty) \times E$ with $\tilde{\sigma}(t,x) \equiv 0$ for $t > t_0$ and such that*

(i) $H\colon E \to [0,\infty)$ is measurable,

(ii) $\tilde{\sigma}\colon [0,t_0] \times E \to (0,\infty)$ is measurable and satisfies the following: for each compact set $K \subset E$ and $S \in (0,t_0)$ there exists $C_1, C_2, C_3 > 0$ such that

$$|\tilde{\sigma}(t,x) - \tilde{\sigma}(s,x)| \leq C_1|t-s| \quad \text{and} \quad C_2 \leq \tilde{\sigma}(t,x) \leq C_3,$$

for all $s,t \in [0,S]$ and for all $x \in K$, where C_3 does not depend on S (but it may depend on K).

Assumption A.7. *Let $\mathcal{D} \subset C_0(E)$ and $\mathcal{A}\colon \mathcal{D} \to C_0(E)$ be linear so that*

(i) $(\mathcal{D},\mathcal{A})$ is conservative, $\mathcal{D}$ is dense in $C_0(E)$ and an algebra in $C_0(E)$,

(ii) for any $\mu_0 \in \mathcal{P}(E)$, the RCLL-martingale problem for $(\mathcal{A},\mu_0)$ is well-posed.[2]

Example A.8. Let $\bar{\mathcal{A}}$ be the generator of a Feller semigroup on $C_0(E)$ with domain $D(\bar{\mathcal{A}})$ and $\mathcal{D}$ a core for $\bar{\mathcal{A}}$ (see [EK86]). Suppose $\mathcal{D}$ is an algebra in $C_0(E)$ and denote by $\mathcal{A}$ the restriction of $\bar{\mathcal{A}}$ to $\mathcal{D}$. Then $\mathcal{D}$ is dense, $(\mathcal{D},\mathcal{A})$ satisfies the assumptions of [EK86, Chap. 4, Thm. 2.2] and thus, by [EK86, Chap. 4, Cor. 2.8], $(\mathcal{D},\mathcal{A})$ is conservative in the sense of (A.5) (from the proof of [EK86, Chap. 4, Cor. 2.8] also conservative in our sense (A.2)). Furthermore, by [EK86, Chap. 4, Thm. 2.7 and Thm. 4.1], for any $\mu_0 \in \mathcal{P}(E)$ the martingale problem for $(\mathcal{A},\mu_0)$ is well-posed and the solution has sample paths in $D_E[0,\infty)$. Hence, Assumption A.7 indeed holds.

Assumption A.9 (Recurrence and boundedness)**.** *Let $(\mathcal{D},\mathcal{A})$ and $(P_x)_{x\in E}$ as in Assumption A.7 and footnote 2. For $t_0 > 0$ and $\sigma = H\tilde{\sigma}$ as in Assumption A.6 assume that*

[2] The most important special case of Assumption A.7 (ii) is $\mu_0 = \delta_x$. In this case the corresponding law on $D_E[0,\infty)$ of the RCLL-solution to the martingale problem is denoted by P_x.

(i) for any $x \in E$, P_x is recurrent (in the sense of Definition A.3) and H is bounded on compacts,

(ii) $\sigma\mathcal{A}f \in C_0([0,t_0] \times E)$ for all $f \in \mathcal{D}$.

Assumption A.9 (ii) can be seen as a weak locality assumption, which is always satisfied for Brownian motion:

Example A.10. Set $\mathcal{D} := C_c^\infty(\mathbb{R})$ and $\mathcal{A}f(x) := \frac{1}{2}f''(x)$ for $f \in \mathcal{D}$. By [EK86, Chap. 5, Prop. 1.1] and Example A.8, Assumption A.7 is satisfied and under P_x the canonical process Z is a Brownian motion started from x so (the first part of) Assumption A.9 (i) holds. A sufficient condition for Assumption A.9 (ii) to hold is that σ is continuous on $[0,t_0] \times \mathbb{R}$: If this is true, then $(t,x) \mapsto \sigma(t,x)\mathcal{A}f(x)$ is continuous and even compactly supported for all $f \in \mathcal{D}$.

Assumption A.11 (Boundedness of σ)**.** *$\sigma \colon [0,\infty) \times E \to [0,\infty)$ is bounded.*

Assumption A.9 (i) or A.11 ensure that there exists a non-exploding solution to the time-change equation (B.1) below.

Assumption A.12 (Regularity of H)**.** *Let $(P_x)_{x\in E}$ as in Assumption A.7 and $H \colon E \to [0,\infty)$ measurable. Assume that for any $x \in E$, H is regular for P_x (in the sense of Definition A.4).*

Assumption A.12 is needed to guarantee uniqueness of the time-change equations. Proposition B.5 provides a useful criterion to verify it.

B Time-changes for Markov processes

Given a Markov process M with generator $\mathcal{A}$ and a sufficiently regular time-inhomogeneous coefficient σ, our aim is to obtain a Markov process X with generator $\sigma\mathcal{A}$. The new Markov process X is identified as a time-change of M, where the time-change δ is characterized by the pathwise Carathéodory differential equation

$$\delta(t) = \int_0^t \sigma(s, M_{\delta(s)})\,\mathrm{d}s, \quad t \in [0,t_0]. \tag{B.1}$$

B.1 Constructing the Time-Change

In order to ensure the existence of the time-change δ, we first need to provide an auxiliary lemma concerning so-called Carathéodory differential equations.

Lemma B.1. *Let $t_0 > 0$ and consider the Carathéodory differential equation*

$$\Delta(t) = \int_0^t \gamma(\Delta(s), s)\,\mathrm{d}s, \tag{B.2}$$

where $\gamma \colon [0,t_0] \times [0,\infty) \to [0,\infty)$ and the integral is understood in the Lebesgue sense. For some $S \in (0,t_0]$ and $T > 0$ suppose that $\gamma(r,\cdot)$ is measurable for each

$r \in [0, S]$, $\gamma(0, \cdot)$ is integrable on $[0, T]$ and there exists an integrable function $f\colon [0, T] \to [0, \infty)$ such that

$$|\gamma(r, t) - \gamma(s, t)| \leq f(t)|r - s| \quad \text{for all} \quad r, s \in [0, S],\, t \in [0, T].$$

Then there exists a unique absolutely continuous function $\Delta\colon I \to [0, S]$ satisfying (B.2) *for some interval $I \subset [0, \infty)$, where either there exists $T_0 \in (0, T]$ such that we may take $I = [0, T_0]$ and we have $\Delta(T_0) = S$ or we may take $I = [0, T]$ and have $\Delta(t) < S$ for all $t \leq T$.*

Proof. Since

$$|\gamma(r, t)| \leq |\gamma(0, t)| + |f(t)|S \quad \text{for all} \quad r \in [0, S]$$

and the right-hand side is integrable on $[0, T]$, γ satisfies the Carathéodory Conditions in [Fil88, Chap. 1] and thus [Fil88, Chap. 1, Thm. 1] guarantees the existence of a solution Δ on an interval $[0, T_0]$ for some $T_0 > 0$. The solution Δ can be extended either to the whole interval $[0, T]$ provided $\Delta(t) \leq S$ for all $t \in [0, T]$ or to the interval $[0, T_0]$ for some $T_0 \in (0, T]$ with $\Delta(T_0) = S$ (see e.g. [Fil88, Chap. 1, Thm. 4]). Uniqueness of the solution follows by [Fil88, Chap. 1, Thm. 2]. □

Due to the time-inhomogeneity of the coefficient σ in the differential equation (B.1), we need to include the time variable t in the state space of the time-changed process X. This time-inhomogeneity prevents us to directly rely on well-known results as for example [EK86, Chap. 6, Thm. 1.1]. Therefore, we verify as a first step that equation (B.1) indeed has a solution.

Lemma B.2. *Let M be an E-valued process on $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{P}$-almost surely RCLL sample paths. Denote by P the law on $D_E[0, \infty)$ of M. Assume that*

- *$t_0 > 0$ and $\sigma = H\tilde{\sigma}$ are given as in Assumption A.6,*
- *H is regular for P (in the sense of Definition A.4),*
- *either H is bounded on compacts and P is recurrent (see Definition A.3) or σ is bounded.*

Then there exists a family of random times $(\delta(t))_{t \in [0, t_0]}$ such that

(i) $\delta\colon [0, t_0] \to [0, \infty)$ is non-decreasing and absolutely continuous, $\mathbb{P}$-a.s.,

(ii) $\delta(t_0)$ is finite, $\mathbb{P}$-a.s.,

(iii) δ solves the Carathéodory differential equation (B.1) *for M.*

Proof. Using the conventions $\inf \emptyset := \infty$ and $[0, 0) := \{0\}$, we define the random time

$$\rho := \inf\left\{ s \in [0, \infty) \,:\, \int_0^s H(M_u)^{-1}\, \mathrm{d}u = \infty \right\} \tag{B.3}$$

and the random times

$$\delta(t) := \begin{cases} \inf\{s \in [0,\rho) \,:\, \Delta(s) \geq t\} \wedge \rho & \text{if } t \in [0,t_0) \\ \sup_{s\in[0,t_0)} \inf\{r \in [0,\rho) \,:\, \Delta(r) \geq s\} \wedge \rho & \text{if } t = t_0 \end{cases} \tag{B.4}$$

where Δ is given by the Carathéodory differential equation

$$\Delta(s) = \int_0^s \sigma(\Delta(r), M_r)^{-1}\,\mathrm{d}r, \quad s \in [0, \delta(t_0)). \tag{B.5}$$

In other words, δ is the right inverse of Δ until σ becomes 0, then one sets $\delta = \rho$. To prove that $(\delta(t))_{t\in[0,t_0]}$ is well-defined, it is sufficient to show that equation (B.5) has indeed a unique solution on the interval $[0, \delta(t) \wedge T]$ for every $t \in [0,t_0)$ and $T \in [0,\rho)$. Let us fix $t \in (0,t_0)$, an RCLL sample path of M denoted by $(M_s(\omega))_{s\in[0,T]}$ for $\omega \in \Omega$ and $T \in [0, \rho(\omega))$. By the RCLL property, the path $(M_s(\omega))_{s\in[0,T]}$ is contained in a compact set $K \subset E$, i.e. $\{M_s(\omega) \,:\, s \in [0,T]\} \subset K$, and $N(\omega) := \{s \in [0,T] \,:\, H(M_s(\omega)) = 0\}$ is a Lebesgue null set by definition of ρ. Therefore, the function $\gamma(r,s) := \sigma(r, M_s(\omega))^{-1}$ is well-defined for $r \in [0,t]$ and $s \in [0,T] \setminus N(\omega)$ and we set $\gamma(\cdot, s) := 1$ for $s \in N(\omega)$. Since $\{M_s(\omega) \,:\, s \in [0,T]\} \subset K$, by Assumption A.6 on $\sigma = H\tilde{\sigma}$ there exist $C_1, C_2 > 0$ such that $C_2 H(M_t)^{-1} \leq \gamma(u,s)$ and

$$\begin{aligned} |\gamma(u,s) - \gamma(v,s)| &= H(M_s(\omega))^{-1} \left| \frac{\tilde{\sigma}(u, M_s(\omega)) - \tilde{\sigma}(v, M_s(\omega))}{\tilde{\sigma}(u, M_s(\omega))\tilde{\sigma}(v, M_s(\omega))} \right| \\ &\leq C_2^2 C_1 H(M_s(\omega))^{-1} |u - v| \end{aligned}$$

for all $u, v \in [0,t]$ and $s \in [0,T]$. Thus, γ satisfies the assumptions of Lemma B.1, which says that there exists a unique solution Δ of the Carathéodory differential equation (B.5) on the interval $[0, \delta(t) \wedge T]$. Moreover, since now $\delta(s)$ is well-defined for all $s \in [0,t_0)$ and $\delta(t_0) = \sup_{s<t_0} \delta(s)$, $\delta(t_0)$ is also well-defined.

Note that if Assumption A.6 (ii) holds also for $S = t_0$, we set

$$\delta(t) := \inf\{s \in [0,\rho) \,:\, \Delta(s) \geq t\} \wedge \rho, \quad t \in [0,t_0], \tag{B.6}$$

instead of (B.4) and the above argument works for $t = t_0$ as well.

(i) By definition of Δ through equation (B.5), Δ is absolutely continuous and strictly increasing on $[0, \delta(t))$ for every $t \in [0,t_0)$ and thus invertible with $\delta(s) = \Delta^{-1}(s)$ for $s \in [0,t)$. This implies that δ is also non-decreasing and absolutely continuous on $[0,t_0]$ (cf. [Leo09, Thm. 1.7 and Ex. 3.21]).

(ii) To verify that $\delta(t_0) < \infty$, $\mathbb{P}$-a.s., suppose first H is bounded on compacts and P is recurrent. Let $N := \{\omega \in \Omega \,:\, \delta(t_0)(\omega) = \infty\}$. Then for $\omega \in N$, equation (B.5) has a solution on $[0,\infty)$ and $\Delta(t)(\omega) < t_0$ for all $t \geq 0$. Fixing some $a > 0$, we notice that by Assumption A.6 and since H is bounded on compacts, there exists a constant $C_1 > 0$ such that $\sigma(\Delta(s), M_s)^{-1} \geq C_1$ for $s \in \{t \geq 0 \,:\, M_t(\omega) \in B_a(M_0(\omega))\}$ and $\omega \in N$. We therefore have

$$\begin{aligned} t_0 \geq \lim_{t\to\infty} \Delta(t) = \int_0^\infty \sigma(\Delta(s), M_s)^{-1}\,\mathrm{d}s &\geq \int_0^\infty \mathbb{1}_{B_a(M_0)}(M_s)\sigma(\Delta(s), M_s)^{-1}\,\mathrm{d}s \\ &\geq C_1 \int_0^\infty \mathbb{1}_{B_a(M_0)}(M_s)\,\mathrm{d}s \end{aligned} \tag{B.7}$$

on N. However, the right-hand side of (B.7) is infinite, $\mathbb{P}$-a.s., by the recurrence assumption and hence (B.7) can only hold on a null set. Thus N is a $\mathbb{P}$-null set, i.e. $\delta(t_0) < \infty$, $\mathbb{P}$-a.s., as claimed. Supposing Assumption A.11 holds, a similar argument works.

(iii) For every $t \in [0, t_0]$ such that $\delta(t) < \rho$, one observes that

$$1 = \frac{\mathrm{d}}{\mathrm{d}s}(\Delta(\delta(s)) = \frac{\mathrm{d}}{\mathrm{d}s}\Delta(\delta(s))\frac{\mathrm{d}}{\mathrm{d}s}\delta(s) = \sigma(\delta(s), M_{\delta(s)})^{-1}\frac{\mathrm{d}}{\mathrm{d}s}\delta(s) \tag{B.8}$$

for almost all $s \in [0, \Delta(\delta(t))]$, where the chain rule (c.f. [Leo09, Thm. 3.44]) and (B.5) was used. Therefore, the absolutely continuity of δ and identity (B.8) show that δ indeed solves the desired integral equation (B.1) since

$$\delta(t) = \int_0^t \frac{\mathrm{d}}{\mathrm{d}s}\delta(s)\,\mathrm{d}s = \int_0^t \sigma(s, M_{\delta(s)})\,\mathrm{d}s.$$

For every $t \in [0, t_0]$ such that $\delta(t) \geq \rho$, we denote $\chi := \inf\{t \in [0, t_0] : \delta(t) \geq \rho\}$. Notice that δ is constant (equal to ρ) on $[\chi, t_0]$ and, $\mathbb{P}$-a.s., $H(M_\rho) = 0$, by the assumption that H is regular for the law of M.[3] In particular, $\delta(t)$ satisfies equation (B.1) for every $t \in [\chi, t_0]$ as well. Recall that the assumption on H means that ρ defined in (B.3) satisfies

$$H(M_\rho) = 0 \text{ on } \{\rho < \infty\}, \quad \mathbb{P}\text{-}a.s., \tag{B.9}$$

$$\rho = \inf\{s \in [0, \infty) : H(M_s) = 0\}, \quad \mathbb{P}\text{-}a.s. \tag{B.10}$$

□

In order to create a better understanding of the time-change δ (defined by the Carathéodory differential equation (B.1)) and the assumptions of Lemma B.2, two remarks are provided for the special case of Brownian motion.

Remark B.3. As we have seen in the proof of Lemma B.2, (A.7) or Assumption A.11 is required to ensure that the random time $\delta(t_0)$ is $\mathbb{P}$-almost surely finite. More precisely, we used

$$\Delta(t) = \int_0^t \sigma(\Delta(s), M_s)^{-1}\,\mathrm{d}s = t_0 \tag{B.11}$$

for some finite $t \in [0, \infty)$, $\mathbb{P}$-a.s. For example, if M is a Brownian motion under $\mathbb{P}$, it is possible to verify condition (B.11) a posteriori by the uniqueness in law of the time-changed process $X_s := M_{\delta(s)}$ for $s \in [0, t_0]$ as done by Bass [Bas83].

Remark B.4. Let us consider the time-homogeneous case where M is a Brownian motion under $\mathbb{P}$. Suppose that $\sigma = H$ and we have a unique finite solution $(\delta(t))_{t\in[0,t_0]}$ to the differential equation (B.1). Then the time-changed process $X_t := M_{\delta(t)}$ is a weak solution to the Brownian stochastic differential equation (SDE)

$$\mathrm{d}X_t = \sqrt{H(X_t)}\,\mathrm{d}M_t, \quad X_0 = x_0 \in \mathbb{R}, \quad t \in [0, t_0].$$

[3] In fact, here only (B.9) is used and so the statement of the Lemma B.2 could be modified accordingly. (B.10) is only used for uniqueness of the time-change in Lemma B.6 below.

From Engelbert and Schmidt [ES85] (see also [KS91, Chap. 5, Thm. 5.7]) we know that a solution to this SDE exists and uniqueness in law holds for this SDE if and only if $\{x \in \mathbb{R} : H(x) = 0\} = I(H)$. Such a function H is regular for (the law of) M under $\mathbb{P}$ in the sense of Definition A.4, see Example A.5. For example, if $H(x) = |x|^\alpha$ for $\alpha \in (0,1)$, then there is no uniqueness in law. However, the so-called fundamental solution in sense of [ES85] is unique in law: The weak solution of the SDE satisfying $H(X_s(\omega)) > 0$ Leb $\otimes \mathbb{P}$-almost surely is unique in law, i.e. a solution that *does not spend time at the zeros of* H. This is exactly how we construct the time-change: No time is spent at the zeros of H until the first time at which $1/H$ is not integrable anymore, then we stop.

Finally, we provide a condition that is useful in verifying regularity of H (see Definition A.4) needed for the existence of the time-change in Lemma B.2 and the uniqueness in Lemma B.6 below.

Proposition B.5. *Let $\mathcal{D} \subset C_0(E)$ dense in $C_0(E)$ and $\mathcal{A}\colon \mathcal{D} \to C_0(E)$ be linear. Suppose M is a solution on $(\Omega, \mathcal{F}, \mathbb{P})$ to the RCLL-martingale problem for $(\mathcal{A}, \mu_0)$, for some $\mu_0 \in \mathcal{P}(E)$. Denote by P the law on $D_E[0,\infty)$ of M. Then any $H \in \mathcal{D}$ with $H \geq 0$ is regular for P.*

Proof. Define ρ as in (B.3) and recall that, by Definition A.4, (B.9) and (B.10) have to be verified. Set

$$\rho_0 := \inf\{s \in [0,\infty) : H(M_s) = 0\}.$$

Since H is continuous and M is RCLL, $H(M_\rho) = 0$ on $\{\rho < \infty\}$ and $\rho_0 \leq \rho$, $\mathbb{P}$-a.s. In particular, $\rho_0 = \rho$ on $\{\rho_0 = \infty\}$, and if $\{\rho_0 < \infty\}$ is a $\mathbb{P}$-null set, this already establishes the claim. Otherwise the probability measure $\tilde{\mathbb{P}}(\,\cdot\,) := \mathbb{P}(\,\cdot\,|\{\rho_0 < \infty\})$ is well-defined, $\rho_0 < \infty$, $\tilde{\mathbb{P}}$-a.s. and to prove the proposition we only need to show $\tilde{\mathbb{P}}(\rho_0 \geq \rho) = 1$. To do so, on $\{\rho_0 < \infty\}$ define for any $t \geq 0$ the random time

$$\delta(t) := \inf\left\{s \in [0,\infty) : \int_0^s H(M_{u+\rho_0})^{-1}\,\mathrm{d}u \geq t\right\}.$$

Since $H(M_{\rho_0}) = 0$ and $\rho_0 \leq \rho$, $\tilde{\mathbb{P}}$-a.s., it suffices to establish that $\tilde{\mathbb{P}}$-a.s. for any $t \geq 0$, $\delta(t) = 0$.

For the proof of the last statement one proceeds as follows: Since H is bounded, $H(M_\rho) = 0$ on $\{\rho < \infty\}$, $\tilde{\mathbb{P}}$-a.s., and by footnote 3, Lemma B.2 can be applied to the RCLL process $(M_{u+\rho_0})_{u \geq 0}$ on $(\Omega, \mathcal{F}, \tilde{\mathbb{P}})$ with $\tilde{\sigma} = 1$ and $\sigma = H$. This yields $\tilde{\mathbb{P}}$-a.s.

$$\delta(t) = \int_0^t H(M_{\delta(u)+\rho_0})\,\mathrm{d}u, \quad t \geq 0, \tag{B.12}$$

and $\delta(t) < \infty$ for any $t \geq 0$. Denote by $(\mathcal{F}_t)_{t\geq 0}$ the $\mathbb{P}$-usual augmentation of the filtration generated by M. Then ρ and ρ_0 (possibly modified on a $\mathbb{P}$-null set, see [EK86, Chap. 4, Cor. 3.13]) are $(\mathcal{F}_t)_{t\geq 0}$-stopping times and thus

$$\{\delta(t) + \rho_0 \leq s\} = \{\rho_0 \leq s\} \cap \left(\left\{\int_{\rho_0}^s H(M_u)^{-1}\,\mathrm{d}u \geq t\right\} \cup \{\rho \leq s - \rho_0\}\right) \in \mathcal{F}_s$$

shows that also $\delta(t)+\rho_0$ is a stopping time. By assumption on H, M and $\mathcal{A}$ (and [RW00a, Lem. II.67.10]) the process

$$N_t := H(M_t) - H(M_0) - \int_0^t \mathcal{A}H(M_s)\,\mathrm{d}s, \quad t \geq 0,$$

is a $(\mathbb{P}, (\mathcal{F}_t)_{t\geq 0})$-martingale. By the optional sampling theorem, for any $r \geq 0$, $\mathbb{P}$-a.s.,

$$\mathbb{E}[N_{(\delta(t)+\rho_0)\wedge r} | \mathcal{F}_{\rho_0 \wedge r}] = N_{\rho_0\wedge r}$$

or equivalently

$$\mathbb{E}[H(M_{(\delta(t)+\rho_0)\wedge r})|\mathcal{F}_{\rho_0\wedge r}] = H(M_{\rho_0\wedge r}) + \mathbb{E}\left[\int_{\rho_0\wedge r}^{(\delta(t)+\rho_0)\wedge r} \mathcal{A}H(M_u)\,\mathrm{d}u \middle| \mathcal{F}_{\rho_0\wedge r}\right].$$

Multiplying by $\mathbb{1}_{\{\rho_0\leq r\}}$, using $\{\rho_0 \leq r\} \in \mathcal{F}_{\rho_0\wedge r}$ and taking expectations gives

$$\begin{aligned}\mathbb{E}[H(M_{(\delta(t)+\rho_0)\wedge r})\mathbb{1}_{\{\rho_0\leq r\}}] =&\mathbb{E}[H(M_{\rho_0\wedge r})\mathbb{1}_{\{\rho_0\leq r\}}] \\ &+ \mathbb{E}\left[\int_{\rho_0\wedge r}^{(\delta(t)+\rho_0)\wedge r} \mathcal{A}H(M_u)\,\mathrm{d}u \mathbb{1}_{\{\rho_0\leq r\}}\right].\end{aligned}$$

By quasi-left continuity, [EK86, Chap. 4, Thm. 3.12],

$$\lim_{r\to\infty} M_{(\delta(t)+\rho_0)\wedge r}\mathbb{1}_{\{\rho_0\leq r\}} = M_{\delta(t)+\rho_0}\mathbb{1}_{\{\rho_0<\infty\}}, \quad \mathbb{P}\text{-a.s.},$$

and so using dominated convergence, boundedness and non-negativity of H, $H(M_{\rho_0}) = 0$ on $\{\rho < \infty\}$ and setting $C := \|\mathcal{A}H\|$, one estimates

$$\begin{aligned}\mathbb{E}[H(M_{\delta(t)+\rho_0})\mathbb{1}_{\{\rho_0<\infty\}}] &= \lim_{r\to\infty} \mathbb{E}[H(M_{(\delta(t)+\rho_0)\wedge r})\mathbb{1}_{\{\rho_0\leq r\}}] \\ &= \lim_{r\to\infty} \left|\mathbb{E}\left[\int_{\rho_0}^{(\delta(t)+\rho_0)\wedge r} \mathcal{A}H(M_u)\,\mathrm{d}u \mathbb{1}_{\{\rho_0\leq r\}}\right]\right| \\ &\leq C\mathbb{E}[\delta(t)\mathbb{1}_{\{\rho_0<\infty\}}].\end{aligned} \tag{B.13}$$

Using (B.12) and Tonelli's Theorem for the first and (B.13) for the second equality yields

$$\mathbb{E}[\delta(t)\mathbb{1}_{\{\rho_0<\infty\}}] = \int_0^t \mathbb{E}[H(M_{\delta(u)+\rho_0})\mathbb{1}_{\{\rho_0<\infty\}}]\,\mathrm{d}u \leq C\int_0^t \mathbb{E}[\delta(u)\mathbb{1}_{\{\rho_0<\infty\}}]\,\mathrm{d}u \tag{B.14}$$

and so Gronwall's lemma implies that the left-hand side of (B.14) is 0 for any $t \geq 0$. But this implies that $\tilde{\mathbb{P}}$-a.s., $\delta(t) = 0$ for all $t \geq 0$ as desired. □

B.2 Pathwise Uniqueness and Martingale Problem

To verify that the random times $(\delta(t))_{t\in[0,t_0]}$ solving the Carathéodory differential equation (B.1) are indeed stopping times with respect to the filtration generated by the process M, we show pathwise uniqueness of the time-changed Markov process $X_t := M_{\delta(t)}$ for $t \in [0, t_0]$.

Lemma B.6. *Let σ and M be given as in Lemma B.2. $(\delta(t))_{t\in[0,t_0]}$ is the family of random times from Lemma B.2 with $\delta(t) := \delta(t_0)$ for $t > t_0$ and the time-changed process X is given by $X_t := M_{\delta(t)}$ for $t \geq 0$. Suppose M is $(\mathcal{F}_t)$-adapted. Then the following holds:*

(i) The time-changed process X has RCLL sample paths, $\mathbb{P}$-a.s.

(ii) Any RCLL process $\tilde{X}$ satisfying

$$\tilde{X}_t = M_{\int_0^t \sigma(u,\tilde{X}_u)\,\mathrm{d}u}, \quad t \in [0,\infty),\ \mathbb{P}\text{-a.s.}, \tag{B.15}$$

is indistinguishable from X.

(iii) The random times $(\delta(t))_{t\in[0,t_0]}$ are $(\mathcal{F}_t)$-stopping times.

Proof. (i) Since M has RCLL sample paths and δ is non-decreasing and absolutely continuous by Lemma B.2, the time-changed process X has RCLL sample paths.

(ii) Let $\tilde{X}$ be an RCLL process satisfying equation (B.15). Define the random time

$$\tilde{\rho} := t_0 \wedge \inf\left\{t \geq 0 \,:\, H(\tilde{X}_t) = 0\right\}$$

and set

$$\tilde{\delta}(s) := \int_0^s \sigma(u,\tilde{X}_u)\,\mathrm{d}u, \quad s \in [0,\infty).$$

Notice that the integral is well-defined since σ is bounded on compacts and $\tilde{X}$ is RCLL. Since $X_t = M_{\delta(t)}$ and $\tilde{X}_t = M_{\tilde{\delta}(t)}$, to verify that X and $\tilde{X}$ are indistinguishable, it is sufficient to show that $\delta(t) = \tilde{\delta}(t)$ for every $t \in [0,\infty)$, $\mathbb{P}$-a.s.

By [Leo09, Lem. 3.31] $\tilde{\delta}$ is absolutely continuous with weak derivative $\tilde{\delta}'(u) = \sigma(u,\tilde{X}_u)$ for $u \in [0,\infty)$ and invertible on $[0,\tilde{\rho}\wedge t_0]$ by the definition of $\tilde{\rho}$. The inverse of $\tilde{\delta}$ is denoted by $\tilde{\Delta}$ with domain $[0,\tilde{\delta}(\tilde{\rho}\wedge t_0)]$. Because $\tilde{\Delta}$ is also strictly increasing and absolutely continuous, the chain rule (see [Leo09, Thm. 3.44]) gives

$$1 = \frac{\mathrm{d}}{\mathrm{d}t}\tilde{\delta}(\tilde{\Delta}(t)) = \sigma(\tilde{\Delta}(t),\tilde{X}_{\tilde{\Delta}(t)})\frac{\mathrm{d}}{\mathrm{d}t}\tilde{\Delta}(t) \quad \text{for almost all } t \in [0,\tilde{\delta}(\tilde{\rho}\wedge t_0)].$$

Combining this with fundamental theorem of calculus (see [Leo09, Thm. 3.30]), one has that $\tilde{\Delta}$ satisfies the integral equation

$$\tilde{\Delta}(t) = \int_0^t \sigma(\tilde{\Delta}(s),\tilde{X}_{\tilde{\Delta}(s)})^{-1}\,\mathrm{d}s, \quad t \in [0,\tilde{\delta}(\tilde{\rho}\wedge t_0)].$$

Moreover, notice that $M_t = M_{\tilde{\delta}(\tilde{\Delta}(t))} = \tilde{X}_{\tilde{\Delta}(t)}$ for $t \in [0,\tilde{\delta}(\tilde{\rho}\wedge t_0)]$. Therefore, $\Delta(t) = \tilde{\Delta}(t)$ for $t \in [0,\tilde{\delta}(\tilde{\rho}\wedge t_0)\wedge\delta(t_0)]$ since the solution to this equation is unique on $[0,\delta(t_0)]$, see (B.5). Furthermore, we have $\tilde{\delta}(\tilde{\rho}\wedge t_0) \leq \rho\wedge\tilde{\delta}(t_0)$ since $t < \tilde{\delta}(\tilde{\rho}\wedge t_0)$ implies $\tilde{\Delta}(t) < \tilde{\rho}\wedge t_0$ and thus $t < \rho\wedge\tilde{\delta}(t_0)$, where we recall ρ from (B.3) and that (B.9), (B.10) holds. In conclusion, $\Delta(t) = \tilde{\Delta}(t)$ for $t \in [0,\tilde{\delta}(\tilde{\rho}\wedge t_0)]$, which leads to $\tilde{\delta}(s) = \delta(s)$ for $s \in [0,\tilde{\rho}\wedge t_0]$.

To see $\tilde{\delta}(s) = \delta(s)$ for $s > \tilde{\rho}\wedge t_0$, we first observe that $1/(H(M_s)\vee\varepsilon)$ is bounded for every $\varepsilon > 0$ and $\tilde{\sigma}$ is bounded on compacts by Assumption A.6. Applying a change of variables ([Leo09, Cor. 3.57]) and using monotone convergence gives

$$Ct \geq \lim_{\varepsilon\to 0}\int_0^t \frac{\sigma(s,\tilde{X}_s)}{H(\tilde{X}_s)\vee\varepsilon}\,\mathrm{d}s = \lim_{\varepsilon\to 0}\int_0^{\tilde{\delta}(t)} \frac{1}{H(M_s)\vee\varepsilon}\,\mathrm{d}s = \int_0^{\tilde{\delta}(t)} \frac{1}{H(M_s)}\,\mathrm{d}s,$$

which ensures $\tilde{\delta}(t) \leq \rho$ for all $t \geq 0$. Assuming $\tilde{\rho} < t_0$, there exist $\{t_n\}_{n\in\mathbb{N}} \subset [\tilde{\rho}, t_0]$ with $t_n \downarrow \tilde{\rho}$ and $H(\tilde{X}_{t_n}) = 0$ and so $\rho \leq \tilde{\delta}(\tilde{\rho})$ by (B.15) and (B.9), (B.10). In this case $\tilde{\delta}(t) = \tilde{\delta}(\tilde{\rho}) = \rho = \delta(t)$ for all $t \geq \tilde{\rho}$. Assuming $\tilde{\rho} \geq t_0$, we have $\tilde{\delta}(t) = \tilde{\delta}(t_0)$ for all $t \geq t_0$ due to $\sigma(t,\cdot) = 0$ for $t > t_0$ and in particular $\tilde{\delta}(t) = \tilde{\delta}(t_0) = \delta(t)$ for $t \geq t_0$.

(iii) In order to apply a result from [EK86], we consider the two-dimensional process $Y_t := (t, M_t)$ and the time-changed process (t, X_t) for $t \in [0,T]$. Hence, [EK86, Chap. 6, Thm. 2.2 (b)] implies that $\delta(t)$ is a stopping time with respect to the usual augmentation of the filtration generated by M, and thus also an $(\mathcal{F}_t)$-stopping time, where we keep in mind that the first component of Y generates a trivial filtration. □

Corollary B.7. *Let σ, M and $\tilde{X}$ be given as in Lemma B.6 and denote by P the law (on $D_E[0,\infty)$) of M under $\mathbb{P}$. Then the law of $\tilde{X}$ under $\mathbb{P}$ is uniquely determined by P and σ.*

Proof. By Lemma B.6 (ii), the law of $\tilde{X}$ is identical to the law of X under $\mathbb{P}$. To show explicitly that the latter is uniquely determined by P and σ, one proceeds as follows: Let $n \in \mathbb{N}$, $t_1,\dots,t_n \in [0,\infty)$, $B_1,\dots,B_n \in \mathcal{B}(E)$ and let $\pi_1\colon D_E[0,\infty)\times D_E[0,\infty) \to D_E[0,\infty)$ be the projection map on the first component. We have seen in the proof of Lemma B.6 that there exist a unique solution to the time-change equation for $\mathbb{P}$-a.e. sample path $M(\omega)$. Hence, as in the proof of [EK86, Chap. 6, Lem. 2.1], the map

$$\gamma\colon D_E[0,\infty)\times D_E[0,\infty) \to D_E[0,\infty), \quad \gamma(M,X) := M_{\int_0^{\cdot}\sigma(u,X_u)\,\mathrm{d}u},$$

is Borel measurable and the set

$$C := \{(m,x) \in D_E[0,\infty)\times D_E[0,\infty) \,:\, \gamma(m,x) = x, x_{t_1}\in B_1,\dots,x_{t_n}\in B_n\}$$

is in $\mathcal{B}(D_E[0,\infty)^2)$. Then [EK86, Appendix 11, Thm. 11.3] implies that $\pi_1 C$ is in the P-completion of $\mathcal{B}(D_E[0,\infty))$ and thus

$$\mathbb{P}(X_{t_1}\in B_1,\dots,X_{t_n}\in B_n) = P(\pi_1 C)$$

is indeed uniquely determined by P and σ. □

In the next proposition we link the martingale problem for the given process M to the martingale problem for the time-changed process X.

Proposition B.8. *Let σ, M and $(\mathcal{F}_t)_{t\geq 0}$ be given as in Lemma B.2. Define the process $(X_t)_{0\leq t\leq t_0}$ by $X_t := M_{\delta(t)}$, where $(\delta(t))_{t\in[0,t_0]}$ is as in Lemma B.2. Suppose that for some f, $g \in C_0(E)$ the process*

$$M_t^{f,g} := f(M_t) - f(M_0) - \int_0^t g(M_s)\,\mathrm{d}s, \quad t \in [0,t_0],$$

is an $(\mathcal{F}_t)$-martingale and σg is bounded. Then the process $(\tilde{M}_s^{f,g})_{0\leq s\leq t_0}$, given by

$$\tilde{M}_t^{f,g} := f(X_t) - f(X_0) - \int_0^t \sigma(s,X_s)g(X_s)\,\mathrm{d}s, \quad t \in [0,t_0],$$

is a martingale w.r.t. the right-continuous completion of the filtration generated by X.

Proof. First observe that $\tilde{M}_t^{f,g} = M_{\delta(t)}^{f,g}$ for $t \in [0,t_0]$. Indeed, since $\delta(\cdot)$ is monotone and absolutely continuous on $[0,t_0]$ and the function $s \mapsto g(M_s)$ is integrable on $[0,\delta(t_0)]$, $\mathbb{P}$-a.s., a change of variables (cf. [Leo09, Cor. 3.57]) leads to

$$\int_0^{\delta(t)} g(M_s)\,\mathrm{d}s = \int_0^t g(M_{\delta(s)})\sigma(s,M_{\delta(s)})\,\mathrm{d}s = \int_0^t \sigma(s,X_s)g(X_s)\,\mathrm{d}u, \quad t \in [0,t_0],$$

and thus $\tilde{M}_t^{f,g} = M_{\delta(t)}^{f,g}$.

Therefore, it is sufficient to verify that $(M_{\delta(t)}^{f,g})_{t\in[0,t_0]}$ is a martingale. For this purpose we rely on the optional sampling theorem (see e.g. [EK86, Chap. 2, Thm. 2.13]) and check its conditions: Because $f \in C_0(E)$ and σg is bounded, there exists a constant $C := C(f,g) > 0$ with $|\tilde{M}_t^{f,g}| \leq C$ for $t \in [0,t_0]$ and in particular $\sup_{t\in[0,t_0]} \mathbb{E}[|\tilde{M}_t^{f,g}|] < \infty$. Since $\delta(s)$ is finite for every $s \in [0,t_0]$, $\mathbb{P}$-a.s., and since for every $T < \delta(s)$ there exists $\tilde{s} \in [0,s)$ such that $\delta(\tilde{s}) = T$, we have $|M_T^{f,g}| = |M_{\delta(\tilde{s})}^{f,g}| = |\tilde{M}_{\tilde{s}}^{f,g}| \leq C$ and

$$\lim_{T\to\infty} \mathbb{E}\big[|M_T^f|\mathbb{1}_{\{\delta(s)>T\}}\big] \leq C \lim_{T\to\infty} \mathbb{P}(\delta(s) > T) = 0.$$

Hence, for $u, v \in [0,t_0]$ with $u \leq v$ the optional sampling theorem gives

$$\mathbb{E}\big[M_{\delta(v)}^{f,g}\big|\mathcal{F}_{\delta(u)}\big] = M_{\delta(u)}^{f,g},$$

which means that $(\tilde{M}_t^{f,g})_{t\in[0,t_0]}$ is an $(\mathcal{F}_{\delta(t)})$-martingale. Because $\tilde{M}_t^{f,g}$ is measurable with respect to $\sigma(X_s : s \leq t)$ for $t \in [0,t_0]$, $\tilde{M}^{f,g}$ is also a martingale with respect to the usual augmentation of the filtration generated by X, see e.g. [RY99, Thm. II.2.8] or [RW00a, Thm. II.67.10]. □

Based on the previous Proposition B.8, the time-changed process X is a solution to the "time-changed" martingale problem and the marginal distributions of X satisfy the corresponding Fokker-Planck equation:

Lemma B.9. *Let $\mu_0 \in \mathcal{P}(E)$, $\mathcal{D} \subset C_0(E)$ and $\mathcal{A} \colon \mathcal{D} \to C_0(E)$ be linear. Let σ and M be given as in Lemma B.2 and assume in addition that M is a solution on $(\Omega, \mathcal{F}, \mathbb{P})$ to the RCLL-martingale problem for $(\mathcal{A}, \mu_0)$ and either Assumption A.9 (ii) or Assumption A.11 holds. Let us denote by $p(t,\cdot)$ the law of $X_t = M_{\delta(t)} = M_{\int_0^t \sigma(s,X_s)\,\mathrm{d}s}$ as constructed in Proposition B.8 with $p(0,\cdot) = \mu_0$ and $\delta(t) := \delta(t_0)$ for $t > t_0$, where t_0 is as in Assumption A.6. Then one has:*

- *X is a solution to the (time-inhomogeneous) $D_E[0,\infty)$-martingale problem for $(\sigma\mathcal{A}, \mu_0)$,*
- *for any $g \in B([0,\infty)\times E)$ the function $s \mapsto \int_E g(s,x)\,p(s,\mathrm{d}x)$ is measurable,*
- *$(p(s,\mathrm{d}x))_{s\in[0,t_0]}$ satisfies the Fokker-Planck equation, i.e. for any $f \in \mathcal{D}$, $t \in [0,t_0]$,*

$$\int_E f(x)\,p(t,\mathrm{d}x) - \int_E f(x)\,\mu_0(\mathrm{d}x) = \int_0^t \int_E \sigma(s,x)\mathcal{A}f(x)\,p(s,\mathrm{d}x)\,\mathrm{d}s. \tag{B.16}$$

Proof. Let $f \in \mathcal{D}$, then $\sigma\mathcal{A}f$ is bounded by Assumption A.9 (ii) or A.11. Combining this with our assumption that M is a solution to the RCLL-martingale problem for $(\mathcal{A}, \mu_0)$ and with Proposition B.8, we obtain that

$$\tilde{M}_t^f := f(X_t) - f(X_0) - \int_0^t \sigma(s,X_s)\mathcal{A}f(X_s)\,\mathrm{d}s, \quad t \geq 0,$$

is a martingale. In particular, one has $\mathbb{E}[\tilde{M}_t^f] = 0$ for all $t \in [0,t_0]$. Since $\sigma\mathcal{A}f$ is bounded, applying Fubini's Theorem yields (B.16). Finally, $X \colon \Omega \times [0,\infty) \to E$ is measurable and thus so is $(\omega,s) \mapsto (s, X_s(\omega))$. Hence, for $g \in B([0,\infty)\times E)$ also $(\omega,s) \mapsto g(s,X_s(\omega))$ is measurable, and so, by the measurability statement in Fubini's Theorem, also $s \mapsto \int_E g(s,x)\,p(s,\mathrm{d}x)$ is measurable. □

C A Uniqueness Result for Degenerate Fokker-Planck Equations

If X is a solution to the martingale problem for $(\sigma\mathcal{A}, \mu_0)$ and $p(t,\cdot)$ is the law of X_t, then according to the proof of Lemma B.9,

$$\text{for any } g \in B([0,\infty)\times E),\ s \mapsto \int_E g(s,x)\,p(s,\mathrm{d}x) \text{ is measurable} \tag{C.1}$$

and for all f nice enough it holds that

$$\int_E f(x)\,p(t,\mathrm{d}x) - \int_E f(x)\,\mu_0(\mathrm{d}x) = \int_0^t \int_E \sigma(s,x)\mathcal{A}f(x)\,p(s,\mathrm{d}x)\,\mathrm{d}s, \quad t \in [0,t_0]. \tag{C.2}$$

Conversely, one may ask if solutions $(p(t,\cdot))_{t\in[0,t_0]}$ to (C.2) can arise differently. In this section, we provide sufficient conditions which guarantee that the Fokker-Planck equation (C.2) (also called Kolmogorov forward equation) uniquely characterizes the law of X, i.e. the one-dimensional marginal laws of X are the only family of probability measures that satisfy (C.2) for a large class of functions f. More precisely, we prove the following result on uniqueness to the Fokker-Planck equation for time-inhomogeneous operators:

Theorem C.1. *Suppose σ and $(\mathcal{D},\mathcal{A})$ satisfy Assumptions A.6, A.7, A.12 and either A.9 or A.11. Then uniqueness holds for* (C.2)*: If both $(q(t,\cdot))_{0\le t\le t_0}$ and $(p(t,\cdot))_{0\le t\le t_0}$ are families of probability measures on E which satisfy* (C.1) *and* (C.2) *for all $f\in\mathcal{D}$ and $q(0,\cdot)=\mu_0=p(0,\cdot)$, then $q(s,\cdot)=p(s,\cdot)$ for all $s\in[0,t_0]$, where t_0 is as in Assumption A.6 and $\mu_0\in\mathcal{P}(E)$.*

As we will see, existence and uniqueness of solutions to the time-inhomogeneous Fokker-Planck equation (C.2) is closely related to existence and uniqueness of solutions to the martingale problem for the time-homogeneous operator $\sigma\mathcal{A}+\partial_t$ on $C_0([0,\infty)\times E)$ defined in equation (C.6) below. We show that the martingale problem for this operator is well-posed and the associated time-homogeneous Fokker-Planck equation determines the marginal laws of the solution uniquely.

In the present context, mainly two difficulties arise: Firstly, σ is only locally bounded, time-inhomogeneous and $\{(t,x)\in[0,\infty)\times E : \sigma(t,x)=0\}\neq\emptyset$. Secondly, even if well-posedness for the martingale problem associated to $\sigma\mathcal{A}$ can be established, it is not automatic that any solution to the Fokker-Planck equation corresponds to a solution to the martingale problem for $\sigma\mathcal{A}$.

The question tackled in Theorem C.1 is classical. It has been studied in various situations that, to the best of our knowledge, do not cover our assumptions on the operator $\sigma\mathcal{A}$. For instance, for σ bounded away from 0, we refer to [Str75], [Bas88], [BC09b] and further references therein. If σ is globally bounded, see [Fig08].

Classical results on multiplicative perturbations of Feller generators and time-changed Lévy processes allow for less regularity on σ, see e.g. [BSW13, Thm. 4.1], the original reference [Lum73], [ES85] and further references therein. However, these results are all time-homogeneous and do not deal with the question whether the Fokker-Planck equation uniquely determines the law of the time-changed process.

The proof of Theorem C.1 relies on the following theorem on uniqueness for the Fokker-Planck equation corresponding to time-homogeneous operators cited[4] from [Kur98] (see also [BK93, Thm. 4.1]).

Theorem C.2 ([Kur98, Thm. 2.6 (c)])**.** *Let (E_0,d_0) be a locally compact, complete, separable metric space, $D(\mathcal{L})\subset C_b(E_0)$ and $\mathcal{L}\colon D(\mathcal{L})\to C_b(E_0)$ be linear. Let $\nu\in\mathcal{P}(E)$ and suppose that*

[4] More precisely, instead of our hypothesis (iii), in [Kur98] the weaker requirement that $\mathcal{L}$ is a *pre-generator* is imposed. However, as explained in [Kur98, Sec. 2] or [KS01, Remark 1.1], (iii) implies that $\mathcal{L}$ is a pre-generator.

(i) $D(\mathcal{L})$ *is an algebra and separates points,*

(ii) *there exists a countable subset* $\{h_k : k \geq 1\} \subset D(\mathcal{L})$ *such that*

$$\textit{bp-closure}(\textit{span}(\{(h_k, \mathcal{L}h_k) : k \geq 1\})) \supset \{(h, \mathcal{L}h) : h \in D(\mathcal{L})\}, \tag{C.3}$$

(iii) *for each* $y \in E_0$*, there exists a RCLL-solution to the martingale problem for* $(\mathcal{L}, \delta_y)$*,*

(iv) *uniqueness holds for the martingale problem for* $(\mathcal{L}, \nu)$*.*

Then uniqueness holds for the Fokker-Planck equation for $(\mathcal{L}, \nu)$*:*
Suppose $\{\nu_t\}_{t \geq 0} \subset \mathcal{P}(E_0)$ *is such that*

$$\text{for any } g \in B(E_0)\,,\ s \mapsto \int_{E_0} g(y)\,\nu_s(\mathrm{d}y) \text{ is measurable} \tag{C.4}$$

and

$$\int_{E_0} h\,\mathrm{d}\nu_t = \int_{E_0} h\,\mathrm{d}\nu + \int_0^t \int_{E_0} \mathcal{L}h\,\mathrm{d}\nu_s\,\mathrm{d}s, \quad t \geq 0, \tag{C.5}$$

for all $h \in D(\mathcal{L})$*. If* $\{\mu_t\}_{t \geq 0} \subset \mathcal{P}(E_0)$ *also satisfies* (C.4) *and* (C.5)*, then* $\mu_t = \nu_t$ *for all* $t \geq 0$*.*

The rest of Appendix C is devoted to the proof of Theorem C.1. The argument is split into three parts and we will only get to the actual proof in the third part. The procedure is as follows:

- In Section C.1 the time-inhomogeneous problem is put into the time-homogeneous setup by including time as an additional state variable. The associated generator $\mathcal{L}$ is defined in (C.6).
- In Section C.2 well-posedness of the martingale problem for $\mathcal{L}$ is proved.
- In Section C.3 the results from Section C.1 and C.2 are used to show that Theorem C.2 can indeed be applied to prove Theorem C.1.

C.1 Reducing to the Time-Homogeneous Setup

Fix $(\mathcal{D}, \mathcal{A})$ as in Assumption A.7 and a measurable function $\sigma\colon [0,\infty) \times E \to [0,\infty)$. For $f \in \mathcal{D}$ and $\gamma \in C_c^1[0,\infty)$, define the operator $\mathcal{L}$ by

$$\mathcal{L}(f\gamma)(t,x) := \gamma(t)\sigma(t,x)\mathcal{A}f(x) + f(x)\gamma'(t), \quad t \in [0,\infty),\, x \in E, \tag{C.6}$$

and linearly extend $\mathcal{L}$ to $D(\mathcal{L}) := \operatorname{span}\{f\gamma : f \in \mathcal{D}, \gamma \in C_c^1[0,\infty)\} \subset C_0([0,\infty) \times E)$.

The following lemma relates the Fokker-Planck equation (C.2) and the martingale problem for the time-inhomogeneous operator $\sigma\mathcal{A}$ to the Fokker-Planck equation and the martingale problem for the time-homogeneous operator $\mathcal{L}$ on $C_0([0,\infty) \times E)$. Furthermore, sufficient conditions for (C.3) are provided.

Lemma C.3. *Suppose* $\sigma\colon [0,\infty)\times E\to[0,\infty)$ *is measurable,* $(\mathcal{D},\mathcal{A})$ *is as in Assumption A.7 and* $(D(\mathcal{L}),\mathcal{L})$ *as in* (C.6)*. Further suppose* $\sigma\mathcal{A}f$ *is bounded for all* $f\in\mathcal{D}$*. Let* $s_0\geq 0$*,* $\mu_0\in\mathcal{P}(E)$ *and define* $\sigma_{s_0}(s,x):=\sigma(s_0+s,x)$*. Then the following hold:*

(i) If σ *is bounded, then* $(D(\mathcal{L}),\mathcal{L})$ *is conservative, i.e.* (A.2) *holds.*

(ii) If X *is a solution to the (time-inhomogeneous) RCLL-martingale problem for* $(\sigma_{s_0}\mathcal{A},\mu_0)$*, then* $(s_0+t,X_t)_{t\geq 0}$ *is a solution to the RCLL-martingale problem for* $(\mathcal{L},\delta_{s_0}\otimes\mu_0)$*.*

(iii) If σ *is bounded and* (T,X) *is a solution to the RCLL-martingale problem for* $(\mathcal{L},\delta_{s_0}\otimes\mu_0)$*, then* X *is a solution to the (time-inhomogeneous) RCLL-martingale problem for* $(\sigma_{s_0}\mathcal{A},\mu_0)$ *and* T *is indistinguishable from* $(s_0+t)_{t\geq 0}$*.*

(iv) Suppose $(p(t,\cdot))_{t\in[0,t_0]}$ *satisfies* (C.1) *and* (C.2) *for all* $f\in\mathcal{D}$ *and define* $p(t,\cdot):=p(t_0,\cdot)$ *and* $\sigma(t,\cdot):=0$ *for* $t>t_0$ *and for all* $t\in[0,\infty)$ *the measures* $\nu_t:=\delta_t\otimes p(t,\cdot)$ *on* $E_0:=[0,\infty)\times E$*. Then* $\{\nu_t\}_{t\geq 0}$ *satisfies* (C.4) *and* (C.5) *for all* $h\in D(\mathcal{L})$*.*

(v) Suppose either $\sigma\mathcal{A}f\in C_0([0,t_0]\times E)$ *for all* $f\in\mathcal{D}$ *or* σ *is bounded. Then there exists a countable subset* $\{h_k : k\in\mathbb{N}\}\subset D(\mathcal{L})$ *such that* (C.3) *holds.*

Proof. *(i)* Note that by Assumption A.7 (i) and (A.2), there exists $\{f_n\}_{n\in\mathbb{N}}\subset\mathcal{D}$ such that $\text{bp-lim}_{n\to\infty} f_n=1$ and $\text{bp-lim}_{n\to\infty}\mathcal{A}f_n=0$. Furthermore, for each $n\in\mathbb{N}$ there exist $\gamma_n\in C_c^1[0,\infty)$ with $\gamma_n=1$ on $[0,n]$, $\gamma_n=0$ on $[n+1,\infty)$ and $\sup_n\|\gamma_n'\|<\infty$. In particular, $\text{bp-lim}_{n\to\infty}\gamma_n'=0$, $\text{bp-lim}_{n\to\infty} f_n\gamma_n=1$ and, since σ is bounded, also $\text{bp-lim}_{n\to\infty}\mathcal{L}(f_n\gamma_n)=0$. Thus, (A.2) holds with $h_n:=f_n\gamma_n$.

(ii) By assumption $\sigma\mathcal{A}f$ is bounded for all $f\in\mathcal{D}$, thus $\mathcal{L}h\in B([0,\infty)\times E)$ for all $h\in D(\mathcal{L})$. Therefore, [EK86, Chap. 4, Thm. 7.1] implies that $(t,X_t)_{t\geq 0}$ is a solution to the martingale problem for $(D(\mathcal{L}),\mathcal{L}^{\sigma_{s_0}})$, where $\mathcal{L}^{\sigma_{s_0}}$ is given in (C.6) with σ replaced by σ_{s_0}. Inserting $h=f\tilde\gamma$ and $\mathcal{L}^{\sigma_{s_0}}$ in (A.4), this implies that the process defined for any $t\geq 0$ by

$$\tilde\gamma(t)f(X_t)-\tilde\gamma(0)f(X_0)-\int_0^t\tilde\gamma(s)\sigma(s_0+s,X_s)\mathcal{A}f(X_s)\,\mathrm{d}s-\int_0^t\tilde\gamma'(s)f(X_s)\,\mathrm{d}s, \quad (\text{C.7})$$

is an $(\mathcal{F}_t^X)_{t\geq 0}$-martingale for all $f\in\mathcal{D}$ and $\tilde\gamma\in C_c^1[0,\infty)$. In particular, for given $\gamma\in C_c^1[0,\infty)$, we can use $\tilde\gamma:=\gamma(\cdot+s_0)$ (which is again in $C_c^1[0,\infty)$) in (C.7) to see that

$$\gamma(s_0+t)f(X_t)-\gamma(s_0)f(X_0)-\int_0^t\mathcal{L}(f\gamma)(s_0+s,X_s)\,\mathrm{d}s,\quad t\geq 0,$$

is an $(\mathcal{F}_t^X)_{t\geq 0}$-martingale for all $f\in\mathcal{D}$ and $\gamma\in C_c^1[0,\infty)$. By linearity, this extends to all $h\in D(\mathcal{L})$ and therefore $(s_0+t,X_t)_{t\geq 0}$ is a solution to the RCLL-martingale problem for $(\mathcal{L},\delta_{s_0}\otimes\mu_0)$.

(iii) Firstly, both $A_1:=\{(\gamma,\gamma') : \gamma\in C_c^1[0,\infty)\}$ and $A_2:=\{(f,\sigma\mathcal{A}f) : f\in\mathcal{D}\}$ (viewed as subsets of $B([0,\infty)\times E)$) are contained in bp-closure($\{(h,\mathcal{L}h)$:

$h \in D(\mathcal{L})\}$). For A_1, this follows by fixing γ and using $\{f_n\}_{n\in\mathbb{N}} \subset \mathcal{D}$ from the proof of (i) to obtain bp-$\lim_{n\to\infty}(f_n\gamma) = \gamma$ and bp-$\lim_{n\to\infty}\mathcal{L}(f_n\gamma) = \gamma'$ since σ is bounded. For A_2, this follows analogously by using γ_n as defined in the proof of (i) and by noting that for any $f \in \mathcal{D}$, bp-$\lim_{n\to\infty}(\gamma_n f, \mathcal{L}(\gamma_n f)) = (f, \sigma\mathcal{A}f)$. So, if (T, X) is a solution to the RCLL-martingale problem for $(\mathcal{L}, \delta_{s_0} \otimes \mu_0)$, then by Remark A.2

$$h(T_t, X_t) - h(T_0, X_0) - \int_0^t g(T_s, X_s)\,\mathrm{d}s, \quad t \geq 0, \tag{C.8}$$

is an $(\mathcal{F}_t^{(T,X)})_{t\geq 0}$-martingale for all $(h, g) \in$ bp-closure$(\{(h, \mathcal{L}h) : h \in D(\mathcal{L})\})$ and thus in particular for all $(h, g) \in A_1 \cup A_2$. Inserting $(\gamma, \gamma') \in A_1$ in (C.8) thus yields that

$$\gamma(T_t) - \gamma(T_0) - \int_0^t \gamma'(T_s)\,\mathrm{d}s, \quad t \geq 0,$$

is an $(\mathcal{F}_t^{(T,X)})_{t\geq 0}$-martingale and, since it is $(\mathcal{F}_t^T)_{t\geq 0}$-adapted, also a martingale with respect to $(\mathcal{F}_t^T)_{t\geq 0}$. Thus, T is a solution to the RCLL-martingale problem for $(\partial_t, \delta_{s_0})$, where ∂_t has domain $D(\partial_t) := C_c^1[0, \infty)$ and is defined as $\partial_t\gamma := \gamma'$ for $\gamma \in D(\partial_t)$. However, $(s_0 + t)_{t\geq 0}$ is also a solution to the RCLL-martingale problem for $(\partial_t, \delta_{s_0})$ since $\int_0^t \gamma'(s_0+s)\,\mathrm{d}s = \gamma(s_0+t)-\gamma(s_0)$ for all $t \geq 0, \gamma \in D(\partial_t)$. By [EK86, Chap. 4, Thm. 4.1] uniqueness holds for the martingale problem for $(\partial_t, \delta_{s_0})$ and in particular T is indistinguishable from $(s_0 + t)_{t\geq 0}$.[5]

On the other hand, (C.8) is a martingale for all $(h, g) \in A_2$ as deduced above and so for each $f \in \mathcal{D}$,

$$f(X_t) - f(X_0) - \int_0^t \sigma(T_s, X_s)\mathcal{A}f(X_s)\,\mathrm{d}s, \quad t \geq 0,$$

is a martingale. Since T is indistinguishable from $(s_0 + t)_{t\geq 0}$, the claim follows.

(iv) First notice that (C.2) actually holds for all $t \geq 0$ since $p(t, \cdot) = p(t_0, \cdot)$ and $\sigma(t, \cdot) = 0$ for $t > t_0$. Moreover, for any $f \in \mathcal{D}$, $\sigma\mathcal{A}f$ is bounded by assumption and so the function $s \mapsto \int_E \sigma(s, x)\mathcal{A}f(x)\,p(s, \mathrm{d}x)$ is measurable by (C.1) and bounded. Thus, from (C.2) we see that $t \mapsto F(t) := \int_E f(x)\,p(t, \mathrm{d}x)$ is absolutely continuous (c.f. [Leo09, Lem. 3.31]) with $F'(t) = \int_E \sigma(t, x)\mathcal{A}f(x)\,p(t, \mathrm{d}x)$ for a.e. $t \geq 0$. Hence, for any $f \in \mathcal{D}$ and $\gamma \in C_c^1[0, \infty)$, we may integrate by parts (see

[5] To check the assumptions of [EK86, Chap. 4, Thm. 4.1] in more detail (see [EK86] for unexplained definitions), note that $[0, \infty)$ is locally compact, separable, $D(\partial_t)$ is dense in $C_0[0, \infty)$ and $C_0[0, \infty)$ is convergence determining (see [EK86, Chap. 3, Prop. 4.4]), hence separating. Furthermore, whenever $\gamma \in D(\partial_t)$, $t^* \geq 0$ satisfy $\gamma(t^*) = \sup_{t\geq 0}\gamma(t)$, then $\gamma'(t^*) = 0$ and thus ∂_t is dissipative by [EK86, Chap. 4, Lem. 2.1]. Finally, fix $\lambda > 0$, then for any $g \in C_c^1[0, \infty)$, the function $\gamma(t) := \exp(\lambda t)\int_t^\infty g(s)\exp(-\lambda s)\,\mathrm{d}s$ satisfies $\gamma \in D(\partial(t))$ and $\lambda\gamma - \partial_t\gamma = g$ so that the range of the operator $\lambda - \partial_t$ is $C_c^1[0, \infty)$ and in particular dense in $C_0[0, \infty)$.

[Leo09, Cor. 3.37]) to obtain

$$\begin{aligned}\gamma(t)F(t)-\gamma(0)F(0) &= \int_0^t \gamma(s)F'(s)\,\mathrm{d}s + \int_0^t \gamma'(s)F(s)\,\mathrm{d}s \\ &= \int_0^t \int_E [\gamma(s)\sigma(s,x)\mathcal{A}f(x)+\gamma'(s)f(x)]\,p(s,\mathrm{d}x)\,\mathrm{d}s \\ &= \int_0^t \int_E \mathcal{L}(\gamma f)(s,x)\,p(s,\mathrm{d}x)\,\mathrm{d}s = \int_0^t \int_{E_0} \mathcal{L}(\gamma f)\,\mathrm{d}\nu_s\,\mathrm{d}s\end{aligned}$$

for any $t \geq 0$. But $\gamma(t)F(t) = \int_{E_0}(f\gamma)\,\mathrm{d}\nu_t$ by definition, thus $\{\nu_t\}_{t\geq 0}$ satisfies (C.5) for all $h = f\gamma$ and by linearity also for all $h \in D(\mathcal{L})$. Finally, note that by definition of $\{\nu_t\}_{t\geq 0}$ the integral in (C.4) is the same as in (C.1) and so the result follows.

(v) Assume first $\sigma\mathcal{A}f \in C_0([0,t_0]\times E)$ for all $f \in \mathcal{D}$. Set $E_0 := [0,t_0]\times E$ and note that the spaces $C_0(E_0)$ and $C_0(E_0)\times C_0(E_0)$ are separable since E, $[0,t_0]$ and E_0 are separable and because products of separable spaces are separable. For $f \in \mathcal{D}$ and $\gamma \in C^1[0,t_0]$ our assumption and $\mathcal{D} \subset C_0(E)$ imply $f\gamma' \in C_0(E_0)$ and $\gamma\sigma\mathcal{A}f \in C_0(E_0)$. Hence, also $\mathcal{L}(f\gamma) \in C_0(E_0)$ and by linearity, $\mathcal{L}h \in C_0(E_0)$ for any $h \in D(\mathcal{L})$. Setting $G_0 := \{(h,\mathcal{L}h) : h \in D(\mathcal{L})\}$, this shows $G_0 \subset C_0(E_0)\times C_0(E_0)$. Since the latter space is separable (as argued above) and any subspace of a separable metric space is separable, we conclude that there exists $H_0 \subset G_0$, H_0 countable, such that each $(h,\mathcal{L}h) \in G_0$ is the limit in sup-norm of a sequence in H_0. In particular, $G_0 \subset$ bp-closure(H_0), i.e. (C.3) holds.

Secondly, assume that σ is bounded. The same separability reasoning as above shows that there exist $\{\gamma_k\}_{k\in\mathbb{N}} \subset C_c^1[0,\infty)$ and $\{f_l\}_{l\in\mathbb{N}} \subset \mathcal{D}$ with the property that for any $\gamma \in C_c^1[0,\infty)$ and $f \in \mathcal{D}$, there exist $\{k_n\}_{n\in\mathbb{N}}$, $\{l_n\}_{n\in\mathbb{N}} \subset \mathbb{N}$ such that $\gamma = \lim_{n\to\infty}\gamma_{k_n}$, $\gamma' = \lim_{n\to\infty}\gamma'_{k_n}$, $f = \lim_{n\to\infty} f_{l_n}$ and $\mathcal{A}f = \lim_{n\to\infty}\mathcal{A}f_{l_n}$ in sup-norm. Since σ is bounded, this also implies bp-$\lim_{n\to\infty}\sigma\mathcal{A}f_{l_n} = \sigma\mathcal{A}f$ and thus bp-$\lim_{n\to\infty}\gamma_{k_n}f_{l_n} = \gamma f$ and bp-$\lim_{n\to\infty}\mathcal{L}(\gamma_{k_n}f_{l_n}) = \mathcal{L}(\gamma f)$. Thus, we have shown

$$\{(f\gamma,\mathcal{L}(f\gamma)) : f \in \mathcal{D}, \gamma \in C_c^1[0,\infty)\} \subset \text{bp-closure}(\{(\gamma_k f_l, \mathcal{L}(\gamma_k f_l))\}_{k,l\in\mathbb{N}})$$

and by linearity this implies (C.3). □

C.2 Well-Posedness of the Martingale Problem

In this section, we show that the martingale problem for $(D(\mathcal{L}),\mathcal{L})$, see (C.6) above, is well-posed. The proof is split into three parts: Existence is established in Proposition C.4, uniqueness is proved in Proposition C.5 under the assumption that σ is bounded. Finally, in Proposition C.7 the assumption of boundedness is removed.

To prove existence, we use the time-change construction from Lemma B.2. Extra work is needed to incorporate the time-inhomogeneity.

Proposition C.4. *Suppose σ and $(\mathcal{D}, \mathcal{A})$ are as in Theorem C.1 and $(D(\mathcal{L}), \mathcal{L})$ is defined as in* (C.6)*. Then for any $s_0 \geq 0$ and $x_0 \in E$ there exists a solution to the RCLL-martingale problem for $(\mathcal{L}, \delta_{s_0} \otimes \delta_{x_0})$.*

Proof. Define $\sigma_{s_0}(s, x) := \sigma(s_0 + s, x)$ for $s \geq 0$, $x \in E$. Assumption A.9 (ii) or A.11 implies that $\sigma \mathcal{A} f$ is bounded for any $f \in \mathcal{D}$ and so by Lemma C.3 (ii) it suffices to show that there exists a solution to the (time-inhomogeneous) RCLL-martingale problem for $(\sigma_{s_0}\mathcal{A}, \delta_{x_0})$.

If $s_0 \geq t_0$, then $\sigma_{s_0}(s, x) = 0$ for all $(s, x) \in (0, \infty) \times E$, since $\sigma(t, \cdot) = 0$ for $t > t_0$. Setting $X_t := x_0$ for $t \geq 0$, by definition (c.f. (A.4)) it follows that X is a solution to the (time-inhomogeneous) RCLL-martingale problem for $(\sigma_{s_0}\mathcal{A}, \delta_{x_0})$.

If $s_0 < t_0$, set $\tilde{t_0} := t_0 - s_0$ and note that σ_{s_0} satisfies Assumption A.6 on $[0, \tilde{t_0}]$ (and $\sigma_{s_0}(t, \cdot) = 0$ for $t > \tilde{t_0}$), since σ satisfies Assumption A.6. Furthermore, σ_{s_0} satisfies $\sigma_{s_0}\mathcal{A} f \in C_0([0, \tilde{t_0}] \times E)$ for all $f \in \mathcal{D}$ or σ_{s_0} is bounded, since σ satisfies Assumption A.9 (ii) or A.11. Let M denote the coordinate process on $D_E[0, \infty)$ and $\mathbb{P} = P_{x_0}$ (as defined in Assumption A.7). Then σ_{s_0}, $(\mathcal{D}, \mathcal{A})$ and M satisfy the assumptions of Lemma B.9, which implies that there exists a solution to the RCLL-martingale problem for $(\sigma_{s_0}\mathcal{A}, \delta_{x_0})$. □

The next step is to prove uniqueness under the assumption that σ is bounded. Combined with Proposition C.4, well-posedness of the RCLL-martingale problem for $(D(\mathcal{L}), \mathcal{L})$ follows.

The main idea of the proof is to show that any solution $\tilde{X}$ to the RCLL-martingale problem for $(\mathcal{L}, \delta_{(s_0,x)})$ can be written as a time-change

$$\tilde{X}_t = M_{\int_0^t \sigma(s_0+u, \tilde{X}_u)\,\mathrm{d}u}, \quad t \geq 0, \ \mathbb{P}\text{-}a.s.,$$

for M which is a solution to the martingale problem for $(\mathcal{A}, \delta_x)$. Corollary B.7 and Assumption A.7 (ii) then allow us to conclude uniqueness.

Note that if $\sigma(t, x)$ did not depend on t, the proof could be simplified significantly by relying on [EK86, Chap. 6, Thm. 1.4].

Proposition C.5. *Suppose σ and $(\mathcal{D}, \mathcal{A})$ satisfy Assumptions A.6, A.7, A.11 and A.12. Define $(D(\mathcal{L}), \mathcal{L})$ as in* (C.6)*, then for each $\nu \in \mathcal{P}([0, \infty) \times E)$ the RCLL-martingale problem for $(\mathcal{L}, \nu)$ is well-posed.*

Proof. Firstly, note that it suffices to show that for each $(s_0, x_0) \in [0, \infty) \times E$ the RCLL-martingale problem for $(\mathcal{L}, \delta_{(s_0,x_0)})$ is well-posed: If this is established, we can combine [BK93, Thm. 2.1] and Lemma C.3 (v) to conclude that also for any $\nu \in \mathcal{P}([0, \infty) \times E)$ the RCLL-martingale problem for $(\mathcal{L}, \nu)$ is well-posed.

From Proposition C.4 it follows that for any $(s_0, x_0) \in [0, \infty) \times E$ there exists a solution to the RCLL-martingale problem for $(\mathcal{L}, \delta_{(s_0,x_0)})$. In order to prove the current proposition, by the above it is therefore sufficient to prove that for any $s_0 \in [0, \infty)$ and any $x_0 \in E$ uniqueness holds for the RCLL-martingale problem for $(\mathcal{L}, \delta_{s_0} \otimes \delta_{x_0})$. This will now be established.

Set $\mu_0 := \delta_{x_0}$ and suppose (T, X) is a solution RCLL-martingale problem for $(\mathcal{L}, \delta_{s_0} \otimes \mu_0)$ defined on some probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$. By Lemma C.3 (iii) it follows that, $\tilde{\mathbb{P}}$-a.s., $T_t = t + s_0$ for all $t \geq 0$ and that X is a solution to the (time-inhomogeneous) RCLL-martingale problem for $(\sigma(s_0 + \cdot)\mathcal{A}, \mu_0)$, i.e. for each $f \in \mathcal{D}$ the process

$$M_t^f := f(X_t) - f(X_0) - \int_0^t \sigma(s_0 + s, X_s)\mathcal{A}f(X_s)\,\mathrm{d}s, \quad t \geq 0, \tag{C.9}$$

is a martingale. In the following, we show that this implies that X solves the time-change equation (B.15) for some process M that is a solution to the RCLL-problem for $(\mathcal{A}, \mu_0)$. By the uniqueness result for the time-change equation in Corollary B.7, it then follows that the law of X is determined by σ and the law of M. Since uniqueness holds for the RCLL-martingale problem for $(\mathcal{A}, \mu_0)$, it then follows that the law of X is uniquely determined by σ and $(\mathcal{A}, \mu_0)$ and thus the claim follows.

Before we start, let us consider the case $s_0 \geq t_0$. Since $\sigma(t, x) = 0$ for all $t > t_0, x \in E$, the integral term in (C.9) vanishes for any $f \in \mathcal{D}$. In particular, $f(X)$ is a martingale for any $f \in \mathcal{D}$. Combining this with our Assumption A.7 (i) that $\mathcal{D}$ is dense in $C_0(E)$, this implies (see [EK86, Chap. 3, Ex. 7]) that $X_t = X_0$, $\tilde{\mathbb{P}}$-a.s., for any $t \geq 0$. However, X is RCLL and thus X is constant $\tilde{\mathbb{P}}$-a.s. In particular, the law of (T, X) is uniquely determined.

Thus, we may assume $s_0 < t_0$ and in analogy to (B.1) define

$$\delta(t) := \int_0^t \sigma(s_0 + u, X_u)\,\mathrm{d}u,$$

for each $t \geq 0$, and set $\Delta(u) := \inf\{t \geq 0 : \delta(t) \geq u\}$ for each $u \geq 0$ and $Y_u := X_{\Delta(u)}$ for $u \leq \delta(t_0 - s_0)$. Note that σ is bounded and $\sigma(s, \cdot) = 0$ for $s > t_0$, thus $\tilde{\mathbb{P}}$-a.s. $\delta(t_0 - s_0) < \infty$ and $\delta(s) = \delta(t_0 - s_0)$ for all $s > t_0 - s_0$. We now claim that

(i) with probability one, X is constant on any interval $[t, u]$ with $\int_t^u \sigma(s_0 + s, X_s)\,\mathrm{d}s = 0$,

(ii) $\tilde{\mathbb{P}}$-a.s., X satisfies the time-change equation $X_t = Y_{\delta(t)}$ for all $t \geq 0$ and Y is RCLL,

(iii) on an extended probability space there exists a process $\tilde{Y}$ that is a solution to the RCLL-martingale problem for $(\mathcal{A}, \mu_0)$ that satisfies $\tilde{Y}_u = Y_u$ for $u \leq \delta(t_0 - s_0)$ and such that $X = \tilde{Y}_{\delta(t)}$ still holds a.s. Furthermore, this implies the claim.

To prove (i), we define

$$\gamma_t := \inf\{u > t : \delta(u) > \delta(t)\} \quad \text{for each } t \geq 0$$

so that for any $u > t$, $\int_t^{\gamma_t \wedge u} \sigma(s_0 + s, X_s)\,\mathrm{d}s = 0$ and thus (recall $\sigma \geq 0$) also

$$\sigma(s_0 + s, X_s) = 0 \quad \text{for a.e. } s \in [t, \gamma_t \wedge u],\ \tilde{\mathbb{P}}\text{-a.s.} \tag{C.10}$$

Furthermore, for $u < t$, $\{\gamma_t \le u\} = \emptyset$ and for $u \ge t$,

$$\{\gamma_t < u\} = \{\delta(u) > \delta(t)\} \tag{C.11}$$

since δ is non-decreasing and continuous (which implies $\delta(\gamma_t) = \delta(t)$). Let us denote by $(\mathcal{F}_t)_{t\ge 0}$ the usual augmentation (in the sense of [RW00a, II.67.3]) of $(\mathcal{F}_t^X)_{t\ge 0}$. Since δ is adapted, (C.11) implies $\{\gamma_t < u\} \in \mathcal{F}_u$ for any $u \ge 0$ and so γ_t is an $(\mathcal{F}_t)_{t\ge 0}$-stopping time. By right-continuity of X and [RW00a, II.67.10], (C.9) is also a martingale with respect to $(\mathcal{F}_t)_{t\ge 0}$. Hence, we can apply optional sampling to the martingales (C.9) to obtain for any $u > t$ and $f \in \mathcal{D}$ that

$$\begin{aligned} 0 &= \mathbb{E}\big[M^f_{\gamma_t\wedge u} - M^f_t \big| \mathcal{F}_t\big] \\ &= \mathbb{E}\left[f(X_{\gamma_t\wedge u}) - \int_t^{\gamma_t\wedge u} \sigma(s_0+s, X_s)\mathcal{A}f(X_s)\,\mathrm{d}s \middle| \mathcal{F}_t\right] - f(X_t) \\ &= \mathbb{E}\left[f(X_{\gamma_t\wedge u})|\mathcal{F}_t\right] - f(X_t), \end{aligned}$$

where the last step follows from (C.10). Since $f \in \mathcal{D}$ was arbitrary and by Assumption A.7 (i) $\mathcal{D}$ is dense in $C_0(E)$, this implies (see [EK86, Chap. 3, Ex. 7]) that for fixed $u > t$, $X_{\gamma_t\wedge u} = X_t$, $\tilde{\mathbb{P}}$-a.s. Thus we can find $\Omega_0 \in \tilde{\mathcal{F}}$ such that $\tilde{\mathbb{P}}(\Omega_0) = 1$ and on Ω_0, we have $X_{\gamma_t\wedge u} = X_t$ for all $u > t \ge 0$ with $t, u \in \mathbb{Q}$. But then on Ω_0 this extends to all $u > t \ge 0$ by a standard argument: for $u > t \ge 0$, we find $\{u_n\} \subset \mathbb{Q}$, $\{t_n\} \subset \mathbb{Q}$ with $u_n \downarrow u$, $t_n \downarrow t$ as $n \to \infty$. Then $\gamma_{t_n} \wedge u_n \downarrow \gamma_t \wedge u$ as $n \to \infty$ and so we can use right-continuity of X for the first and last equality and our choice of Ω_0 for the second equality to obtain

$$X_t = \lim_{n\to\infty} X_{t_n} = \lim_{n\to\infty} X_{\gamma_{t_n}\wedge u_n} = X_{\gamma_t\wedge u}. \tag{C.12}$$

Thus, if $\omega \in \Omega_0$, $u > t \ge 0$ and $\int_t^u \sigma(s_0+s, X_s(\omega))\,\mathrm{d}s = 0$, then $u \le \gamma_t(\omega)$ and so by (C.12) indeed $X_t(\omega) = X_{\gamma_t(\omega)\wedge u} = X_u(\omega)$.

To prove (ii), set $\Delta_+(u) := \lim_{v\downarrow u} \Delta(v)$ and notice $\Delta_+(u) = \inf\{t \ge 0 : \delta(t) > u\}$. Then from (i) we get that, $\tilde{\mathbb{P}}$-a.s., X is constant on the interval $[\Delta(u), \Delta_+(u)]$ for all $u \ge 0$. Hence, from $\Delta(\delta(t)) \le t \le \Delta_+(\delta(t))$ we obtain

$$X_t = X_{\Delta(\delta(t))} = Y_{\delta(t)}$$

for all $t \ge 0$. Furthermore, $u \downarrow u_0$ implies $\Delta(u) \downarrow \Delta_+(u_0)$ and so by right-continuity of X also

$$Y_u = X_{\Delta(u)} \to X_{\Delta_+(u_0)} = X_{\Delta(u_0)} = Y_{u_0}$$

and since Δ is left-continuous and X has left-limits, the same reasoning shows that Y also has left-limits. Hence, Y is indeed RCLL.

For (iii), notice that $\sigma(s_0+s, X_s) = 0$ for $s \in [\Delta(\delta(t_0-s_0)), t_0-s_0]$ and $s \in [\Delta(v), \Delta_+(v)]$ for any $v \ge 0$. Combining this with $\Delta(u) \le t \Leftrightarrow \delta(t) \ge u$, one obtains

$$\begin{aligned} \int_0^{\Delta(u)\wedge(t_0-s_0)} \sigma(s_0+s, X_s)\mathcal{A}f(X_s)\,\mathrm{d}s &= \int_0^{\Delta(u\wedge\delta(t_0-s_0))} \sigma(s_0+s, X_s)\mathcal{A}f(X_s)\,\mathrm{d}s \\ &= \int_0^{\Delta_+(u\wedge\delta(t_0-s_0))} \sigma(s_0+s, X_s)\mathcal{A}f(X_s)\,\mathrm{d}s, \end{aligned} \tag{C.13}$$

for $u \geq 0$, $f \in \mathcal{D}$. Since $\Delta_+(v) = \inf\{t \geq 0 : \delta(t) > u\}$, a change of variables as in [EK86, Chap. 6, Ex. 12] now allows to rewrite the last expression as

$$\int_0^{\Delta_+(u\wedge\delta(t_0-s_0))} \sigma(s_0+s, X_s)\mathcal{A}f(X_s)\,\mathrm{d}s = \int_0^{u\wedge\delta(t_0-s_0)} \mathcal{A}f(X_{\Delta_+(v)})\,\mathrm{d}v. \quad \text{(C.14)}$$

As argued in (ii), $X_{\Delta_+(v)} = Y_v$ for all $v \geq 0$ and by inserting this in the right-hand side of (C.14) and combining with (C.13), we get

$$\int_0^{\Delta(u)\wedge(t_0-s_0)} \sigma(s_0+s, X_s)\mathcal{A}f(X_s)\,\mathrm{d}s = \int_0^{u\wedge\delta(t_0-s_0)} \mathcal{A}f(Y_s)\,\mathrm{d}s \quad \text{(C.15)}$$

for any $u \geq 0$, $f \in \mathcal{D}$. Furthermore, for any $t, u \geq 0$, $\{\Delta(u) \leq t\} = \{\delta(t) \geq u\} \in \mathcal{F}_t$ so that for each $u \geq 0$, $\Delta(u)$ is a stopping time. Using this, (C.15) and applying the optional sampling theorem to the martingales in (C.9), we therefore get that for any $f \in \mathcal{D}$ the process

$$M^f_{\Delta(u)\wedge(t_0-s_0)} = f(Y_{u\wedge\delta(t_0-s_0)}) - f(Y_0) - \int_0^{u\wedge\delta(t_0-s_0)} \mathcal{A}f(Y_s)\,\mathrm{d}s, \quad u \geq 0, \quad \text{(C.16)}$$

is a martingale with respect to the filtration $(\mathcal{F}_{\Delta(u)\wedge t_0-s_0})_{u\geq 0}$ and thus also with respect to the filtration generated by $Y_{\cdot\wedge\delta(t_0-s_0)}$. Let us denote by W the coordinate process on $D_E[0,\infty)$ and for $(\omega,\omega') \in \tilde{\Omega} \times D_E[0,\infty)$ define

$$\tilde{Y}_u(\omega,\omega') := \begin{cases} Y_u(\omega) & \text{for } u < \delta(t_0-s_0)(\omega) \\ W_{u-\delta(t_0-s_0)(\omega)}(\omega') & \text{for } u \geq \delta(t_0-s_0)(\omega). \end{cases}$$

From (C.16) and Lemma C.6 below (applied to the process $(Y_{u\wedge\delta(t_0-s_0)})_{u\geq 0}$ and the random variable $\delta(t_0-s_0)$) it follows that the process $\tilde{Y}$ is a solution to the RCLL-martingale problem for $(\mathcal{A},\mu_0)$ under a measure Q with $Q(A\times D_E[0,\infty)) = \tilde{\mathbb{P}}(A)$ for all $A \in \tilde{\mathcal{F}}$ and such that $\tilde{Y}_{s\wedge\delta(t_0-s_0)} = Y_{s\wedge\delta(t_0-s_0)}$ for all $s \geq 0$, Q-a.s. Combining this with (ii) and $\delta(\cdot) \leq \delta(t_0-s_0)$, it follows that

$$X_t = Y_{\delta(t)} = Y_{\delta(t)\wedge\delta(t_0-s_0)} = \tilde{Y}_{\delta(t)\wedge\delta(t_0-s_0)} = \tilde{Y}_{\delta(t)} = \tilde{Y}_{\int_0^t \sigma_{s_0}(s,X_s)\,\mathrm{d}s}, \quad t \geq 0,\ Q\text{-a.s.},$$

where $\sigma_{s_0}(s,x) := \sigma(s+s_0,x)$ for $s \geq 0, x \in E$. In particular, X satisfies a time-change equation (B.15) (with M replaced by $\tilde{Y}$ and σ replaced by σ_{s_0}). By our assumptions on σ, σ_{s_0} satisfies Assumption A.6 on $[0, t_0-s_0]$ (and $\sigma_{s_0}(t,\cdot) = 0$ for $t > t_0-s_0$) and A.11. Since uniqueness holds for the RCLL-martingale problem for $(\mathcal{A},\mu_0)$, the law on $D_E[0,\infty)$ of $\tilde{Y}$ under Q is given as P_{x_0} and thus, by Assumption A.12, H is regular for Q. Altogether, Corollary B.7 can be applied to σ_{s_0} and $\tilde{Y}$, which implies that the law of X under Q is uniquely determined by σ_{s_0} and P_{x_0}. But the law of X under $\tilde{\mathbb{P}}$ is the same as under Q and so the claim follows. □

For the well-posedness of the RCLL-martingale problem (see Proposition C.5), we used the following auxiliary lemma. As the authors are not aware of a suitable reference, we also present its complete proof here.

Lemma C.6. *Let (E, d) be a locally compact, complete, separable metric space, $\mathcal{D} \subset C_0(E)$ and $\mathcal{A}\colon \mathcal{D} \to C_0(E)$ linear. Suppose that the $D_E[0,\infty)$-martingale problem for $(\mathcal{A}, \mu)$ is well-posed for any $\mu \in \mathcal{P}(E)$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, τ be a $[0,\infty)$-valued random variable, and Z be an E-valued RCLL process on $(\Omega, \mathcal{F}, \mathbb{P})$ with $Z_u = Z_{u\wedge\tau}$ for all $u \geq 0$, such that*

$$f(Z_u) - f(Z_0) - \int_0^{u\wedge\tau} \mathcal{A}f(Z_s)\,\mathrm{d}s, \quad u \geq 0,$$

is a martingale for all $f \in \mathcal{D}$. Let us denote by X the coordinate process on $D_E[0,\infty)$. On $\Omega' := \Omega \times D_E[0,\infty)$ define the process Y via

$$Y_t(\omega, \omega') := \begin{cases} Z_t(\omega) & \text{for } t < \tau(\omega) \\ X_{t-\tau(\omega)}(\omega') & \text{for } t \geq \tau(\omega) \end{cases}$$

for $(\omega, \omega') \in \Omega'$ and $t \geq 0$. Furthermore, for each $x \in E$, denote by P_x the law of the solution of the RCLL-martingale problem for $(\mathcal{A}, \delta_x)$ and by $\mathcal{S}_E$ the Borel σ-algebra in $D_E[0,\infty)$. Let us define the measure Q on $\mathcal{F} \times \mathcal{S}_E$ by

$$Q(A \times C) := \mathbb{E}[\mathbb{1}_A P_{Z_\tau}(C)] \tag{C.17}$$

for $A \in \mathcal{F}$, $C \in \mathcal{S}_E$ (and extend this to the product σ-algebra). Then under Q, Y is a solution to the RCLL-martingale problem for $(\mathcal{A}, \mu_0)$, where μ_0 is the law of Z_0. Furthermore, $Q(A \times D_E[0,\infty)) = \mathbb{P}(A)$ for all $A \in \mathcal{F}$ and $Z_t = Y_{t\wedge\tau}$ for all $t \geq 0$, Q-a.s.

Proof. Essentially this is [EK86, Chap. 4, Lem. 5.16], the only difference is that we construct Y on $\Omega \times D_E[0,\infty)$ (instead of $D_E[0,\infty) \times [0,\infty) \times D_E[0,\infty)$).

To prove the claim, first notice that by [EK86, Chap. 4, Thm. 4.6], the map $x \mapsto P_x(C)$ is measurable for each $C \in \mathcal{S}_E$ and so Q is indeed well-defined. Furthermore, denoting by μ the law of Z_τ, also the measure $\tilde{Q}$ defined on product sets as

$$\tilde{Q}(B \times C) := \int_E \mathbb{E}[\mathbb{1}_B(Z, \tau)|Z_\tau = x] P_x(C)\,\mu(\mathrm{d}x),$$

for $B \in \mathcal{S}_E \times \mathcal{B}([0,\infty))$, $C \in \mathcal{S}_E$ is well-defined. Thus, denoting by $(X^{(1)}, \eta, X^{(2)})$ the coordinate random variable on $D_E[0,\infty) \times [0,\infty) \times D_E[0,\infty)$, from [EK86, Chap. 4, Lem. 5.16, (5.52) and (5.53)] it follows that under $\tilde{Q}$ the process $(\tilde{Y}_t)_{t\geq 0}$ defined as

$$\tilde{Y}_t := \begin{cases} X^{(1)}_t & \text{for } t < \eta \\ X^{(2)}_{t-\eta} & \text{for } t \geq \eta \end{cases}$$

is a solution to the RCLL-martingale problem for $(\mathcal{A}, \mu_0)$. Thus, it remains to show that the law of Y under Q is the same as the law of $\tilde{Y}$ under $\tilde{Q}$.

Firstly note that for $B \in \mathcal{S}_E \times \mathcal{B}([0,\infty))$ and $C \in \mathcal{S}_E$ one obtains

$$\begin{aligned} Q(\{(Z,\tau) \in B\} \times \{X \in C\}) &= \mathbb{E}[\mathbb{1}_{\{(Z,\tau)\in B\}} P_{Z_\tau}(C)] \\ &= \mathbb{E}[\mathbb{E}[\mathbb{1}_B(Z,\tau)|Z_\tau]\mathbb{E}[P_{Z_\tau}(C)|Z_\tau]] \\ &= \tilde{Q}(B \times C), \end{aligned}$$

where the last step uses [EK86, Chap. 4, Thm. 4.2 (c)]. Hence, $\tilde{Q}$ coincides with the law of (Z, τ, X) on $D_E[0,\infty) \times [0,\infty) \times D_E[0,\infty)$ under Q. In particular, for any $0 \le t_1 < \cdots < t_n$, $n \in \mathbb{N}$, $A_1, \cdots, A_n \in \mathcal{B}(E)$,

$$\begin{aligned}
&Q(Y_{t_1} \in A_1, \ldots, Y_{t_n} \in A_n) \\
&\quad = \sum_{I \subset \{1,\ldots,n\}} Q(Z_{t_i} \in A_i\,, t_i < \tau\,\forall i \in I, X_{t_j - \tau} \in A_j, t_j \ge \tau\,\forall j \in I^c) \\
&\quad = \sum_{I \subset \{1,\ldots,n\}} \tilde{Q}(X^{(1)}_{t_i} \in A_i\,, t_i < \eta\,\forall i \in I, X^{(2)}_{t_j - \eta} \in A_j, t_j \ge \eta\,\forall j \in I^c) \\
&\quad = \tilde{Q}(\tilde{Y}_{t_1} \in A_1, \ldots, \tilde{Y}_{t_n} \in A_n)
\end{aligned}$$

and thus it follows that the law of $\tilde{Y}$ under $\tilde{Q}$ is the same as the law of Y under Q, hence the claim.

Finally, $Y_u = Z_u$ for $u < \tau$ and so we only need to show $Z_\tau = Y_\tau$, Q-a.s. This should be clear but we still give a formal argument: For any $A, B \in \mathcal{B}(E)$ we have, by (C.17) and $P_x(X_0 \in B) = \delta_x(B)$,

$$Q(\{Z_\tau \in A\} \times \{X_0 \in B\}) = \mathbb{E}[\mathbb{1}_A(Z_\tau) P_{Z_\tau}(X_0 \in B)] = \mathbb{P}(Z_\tau \in A \cap B). \quad \text{(C.18)}$$

Now denote by $\{x_m\}_{m\in\mathbb{N}}$ a countable dense subset of E. Then, setting $A_{n,m,k} = \{Z_\tau \in B_{1/k}(x_n)\} \cap \{X_0 \in B_{1/k}(x_m)\}$ for any $n \in \mathbb{N}$, $m \in \mathbb{N} \setminus \{n\}$ and $k \in \mathbb{N}$ with $d(x_n, x_m) > 2/k$, from (C.18) we have

$$Q(A_{n,m,k}) = \mathbb{P}(Z_\tau \in B_{1/k}(x_n) \cap B_{1/k}(x_m)) = \mathbb{P}(\emptyset) = 0.$$

Writing

$$\{Z_\tau \ne Y_\tau\} = \{Z_\tau \ne X_0\} = \bigcup_{n\in\mathbb{N}} \bigcup_{m \in \mathbb{N}\setminus\{n\}} \bigcup_{k \in \mathbb{N}\,:\, d(x_n,x_m) > 2/k} A_{n,m,k},$$

we see that $\{Z_\tau \ne Y_\tau\}$ is a countable union of Q-null sets and so the claim follows. □

Finally, Proposition C.5 is extended in two directions: Firstly, we localize to prove uniqueness for the RCLL-martingale problem when σ is unbounded, and secondly, we show that any progressively measurable (but not necessarily RCLL) solution to the martingale problem has an RCLL modification. Note that the last statement is not true in general (as discussed in [BK03]), but it has to be established to apply Theorem C.2.

In the proof, the notion of a *stopped martingale problem* is used: Let (F, d) be a complete metric space, $(D(\mathcal{L}), \mathcal{L})$ an operator on $C_b(F)$ and U open in F. If X is an F-valued, RCLL process, then $\tau := \inf\{t \ge 0 \,:\, X_t \notin U \text{ or } X_{t-} \notin U\}$ is an $(\mathcal{F}^X_t)_{t\ge0}$-stopping time by [EK86, Chap. 2, Prop. 1.5 a)]. X is called a solution to the stopped martingale problem for $(\mathcal{L}, \nu_0, U)$ if $X_0 \sim \nu_0$, $X^\tau = X$ and

$$h(X_{\tau\wedge t}) - h(X_0) - \int_0^{\tau\wedge t} \mathcal{L}h(X_s)\,\mathrm{d}s, \quad t \ge 0,$$

is an $(\mathcal{F}^X_t)_{t\ge0}$-martingale for any $h \in D(\mathcal{L})$.

Proposition C.7. *Suppose σ and $(\mathcal{D},\mathcal{A})$ are as in Theorem C.1 and $(D(\mathcal{L}),\mathcal{L})$ is defined as in* (C.6). *Then for any $\mu_0 \in \mathcal{P}(E)$ the martingale problem for $(\mathcal{L},\delta_0\otimes\mu_0)$ is well-posed.*

Proof. Suppose first $\mu_0 = \delta_{x_0}$ for some $x_0 \in E$. By Proposition C.4 and our assumptions, there exists a solution Z to the RCLL-martingale problem for $(\mathcal{L},\delta_0\otimes\mu_0)$. Therefore, it suffices to show that if $\tilde{Z}$ is any (progressively measurable) solution to the martingale problem for $(\mathcal{L},\delta_0\otimes\mu_0)$, then $\tilde{Z}$ has the same finite-dimensional marginal distributions as Z. In order to prove this, we proceed as follows:

To start, by [EK86, Chap. 4, Cor. 3.7] and since $D(\mathcal{L})$ is dense in $C_0(\mathcal{L})$, $\tilde{Z}$ has a modification (which we also denote by $\tilde{Z}$) with sample paths in $D_{E_0^\Delta}[0,\infty)$, where E_0^Δ is the one-point compactification of $E_0 := [0,\infty)\times E$. Furthermore, (A.3) remains vaild for $\tilde{Z}$ and all $h \in D(\mathcal{L})$, where we extend h to $C(E_0^\Delta)$ by $h(\Delta) := 0$. By assumption on E, there exists $\{V_n\}_{n\in\mathbb{N}} \subset E$ such that for any n, V_n is open, $\overline{V_n}$ is compact and $\cup_n V_n = E$. Define $U_n := [0,\infty)\times V_n$, $\tau_n := \inf\{t \geq 0 : Z_t \notin U_n \text{ or } Z_{t-} \notin U_n\}$ and $\tilde{\tau}_n =: \inf\{t\geq 0 : \tilde{Z}_t \notin U_n \text{ or } \tilde{Z}_{t-}\notin U_n\}$. Since U_n is open in E_0, it is also open in E_0^Δ and thus τ_n is an $(\mathcal{F}_t^Z)_{t\geq 0}$-stopping time and $\tilde{\tau}_n$ is an $(\mathcal{F}_t^{\tilde{Z}})_{t\geq 0}$-stopping time.

Suppose Assumption A.9 holds. Then for any $n \in \mathbb{N}$ there exists $C_n > 0$ such that $|H(x)| \leq C_n$ and $|\tilde{\sigma}(t,x)| \leq C_n$ for all $(t,x) \in U_n$. Set $\tilde{\sigma}_n(t,x) := \min(\tilde{\sigma}(t,x),C_n)$, $H_n(x) := \min(H(x),C_n)$ and $\sigma_n := H_n\tilde{\sigma}_n$. In the other case, i.e. if Assumption A.11 holds, set $\sigma_n := \sigma$. Then, in both cases, σ_n is bounded and coincides with σ in U_n. Define

$$\mathcal{L}^n(f\gamma)(t,x) := \gamma(t)\sigma_n(t,x)\mathcal{A}f(x) + f(x)\gamma'(t), \quad t\in[0,\infty),\, x \in E,$$

for $f \in \mathcal{D}$ and $\gamma \in C_c^1[0,\infty)$ and linearly extended to $D(\mathcal{L}^n) := \operatorname{span}\{f\gamma : f \in \mathcal{D}, \gamma \in C_c^1[0,\infty)\}$ (and thus $D(\mathcal{L}^n) = D(\mathcal{L})$).

We now claim that:

(i) Z^{τ_n} is a solution to the stopped martingale problem for $(\mathcal{L}^n,\delta_0\otimes\mu_0,U_n)$ and this solution is unique in law,

(ii) $\tilde{Z}^{\tilde{\tau}_n}$ takes values in $D_{E_0}[0,\infty)$ and is also a solution to the stopped martingale problem for $(\mathcal{L}^n,\delta_0\otimes\mu_0,U_n)$ and thus, combining this with (i), we get that the finite-dimensional marginals of $\tilde{Z}^{\tilde{\tau}_n}$ and Z^{τ_n} agree,

(iii) from (ii) it can be deduced that $\tilde{\tau}_n \to \infty$ as $n\to\infty$ and that $\tilde{Z}$ and Z have the same distribution.

To show (i), notice that σ_n is bounded and satisfies Assumption A.6 and Assumption A.12 since they hold for H and $\tilde{\sigma}$. In particular, σ_n and $(\mathcal{D},\mathcal{A})$ satisfy Assumptions A.6, A.7, A.11 and A.12 and thus by Proposition C.5 the RCLL-martingale problem for $(\mathcal{L}^n,\delta_0\otimes\mu_0)$ is well-posed. Therefore, by [EK86, Chap. 4, Thm. 6.1] for each $U\subset E_0$ open there exists a unique solution to the stopped martingale problem for $(\mathcal{L}^n, \delta_0 \otimes \mu_0, U)$. Applying optional sampling to τ_n and the martingales (A.3) and noticing

$$\int_0^{\tau_n \wedge t} \mathcal{L}h(Z_s)\,\mathrm{d}s = \int_0^{\tau_n \wedge t} \mathcal{L}^n h(Z_s)\,\mathrm{d}s, \tag{C.19}$$

we see that Z^{τ_n} is a (and hence the unique) solution to the stopped martingale problem for $(\mathcal{L}^n, \delta_0 \otimes \mu_0, U_n)$.

To show (ii), note that, by definition of $\tilde{\tau}_n$, $\tilde{Z}$ is RCLL and U_n-valued on $[0, \tilde{\tau}_n)$. Let us first show that actually $\tilde{\mathbb{P}}$-a.s., for each n, $\tilde{Z}_{\tilde{\tau}_n} \in E_0$ (a priori, we could have $\tilde{Z}_{\tilde{\tau}_n} = \Delta$). To do so, note that by applying optional sampling and taking expectations in (A.3), we obtain

$$\mathbb{E}[h(\tilde{Z}_{\tilde{\tau}_n \wedge t})] = \mathbb{E}[h(\tilde{Z}_0)] + \mathbb{E}\left[\int_0^{\tilde{\tau}_n \wedge t} \mathcal{L}h(\tilde{Z}_s)\,\mathrm{d}s\right] = \mathbb{E}[h(\tilde{Z}_0)] + \mathbb{E}\left[\int_0^{\tilde{\tau}_n \wedge t} \mathcal{L}^n h(\tilde{Z}_s)\,\mathrm{d}s\right] \tag{C.20}$$

for all $h \in D(\mathcal{L})$ (with $C_0(E_0)$ extended to E_0^Δ as above). The second step in (C.20) follows as in (C.19). Since σ_n is bounded, $(D(\mathcal{L}^n), \mathcal{L}^n)$ is conservative by Lemma C.3 (i) and so (for any $n \in \mathbb{N}$) there exists $\{h_k\}_{k \in \mathbb{N}} \subset D(\mathcal{L})$ such that $\text{bp-lim}_{k \to \infty} h_k = 1$ and $\text{bp-lim}_{k \to \infty} \mathcal{L}^n h_k = 0$. In particular, $\text{bp-lim}_{k \to \infty} h_k = \mathbb{1}_{E_0}$ in $C(E_0^\Delta)$. Inserting h_k in (C.20) and letting $k \to \infty$, dominated convergence gives

$$\mathbb{P}(\tilde{Z}_{\tilde{\tau}_n \wedge t} \in E_0) = \lim_{k \to \infty} \mathbb{E}[h_k(\tilde{Z}_{\tilde{\tau}_n \wedge t})] = \lim_{k \to \infty} \mathbb{E}[h_k(\tilde{Z}_0)] = \mathbb{P}(\tilde{Z}_0 \in E_0) = 1.$$

Therefore, for any $n \in \mathbb{N}$, $\tilde{Z}$ is RCLL and U_n-valued on $[0, \tilde{\tau}_n]$. Thus, for any $n \in \mathbb{N}$, we may view $\tilde{Z}^{\tilde{\tau}_n}$ as a $D_{E_0}[0,\infty)$-valued process and optional sampling applied to (A.3) (and the analogon of (C.19) for $\tilde{Z}$) shows that $\tilde{Z}^{\tilde{\tau}_n}$ is a solution to the stopped martingale problem for $(\mathcal{L}^n, \delta_0 \otimes \mu_0, U_n)$. Thus, by (i), the laws of $\tilde{Z}^{\tilde{\tau}_n}$ and Z^{τ_n} coincide.

To show (iii), first note that $\tilde{\tau}_n \leq \tilde{\tau}_{n+1}$ for all $n \in \mathbb{N}$ and thus $\tau := \lim_{n \to \infty} \tilde{\tau}_n$ is well-defined. Since $\tilde{Z}$ has left-limits in E_0^Δ, also $Y_t := \lim_{n \to \infty} \tilde{Z}_{\tilde{\tau}_n \wedge t}$ is well-defined in E_0^Δ. Furthermore, $Y_t = \Delta$ if and only if $\tau \leq t$. Since Z has sample paths in $D_{E_0}[0,\infty)$, it holds that $\tau_n \to \infty$, $\mathbb{P}$-a.s., and so (ii) implies

$$\mathbb{E}[h(Y_t)] = \lim_{n \to \infty} \mathbb{E}[h(\tilde{Z}_t^{\tilde{\tau}_n})] = \lim_{n \to \infty} \mathbb{E}[h(Z_t^{\tau_n})] = \mathbb{E}[h(Z_t)] \tag{C.21}$$

for any $t \geq 0$ and $h \in C_0(E_0)$. Taking $\{h_k\}_{k \in \mathbb{N}} \subset C_0(E_0)$ with $\text{bp-lim}_{k \to \infty} h_k = 1$ (and thus $\text{bp-lim}_{k \to \infty} h_k = \mathbb{1}_{E_0}$ in $C(E_0^\Delta)$), inserting h_k in (C.21) and letting $k \to \infty$, one obtains

$$\mathbb{P}(\tau > t) = \mathbb{P}(Y_t \in E_0) = \lim_{k \to \infty} \mathbb{E}[h_k(Y_t)] = \lim_{k \to \infty} \mathbb{E}[h_k(Z_t)] = \mathbb{P}(Z_t \in E_0) = 1$$

for all $t \geq 0$. Hence, $\mathbb{P}(\tau = \infty) = 1$, $\tilde{\tau}_n \to \infty$ a.s. and $\tilde{Z}$ does not explode, i.e it has sample paths in $D_{E_0}[0,\infty)$. In particular, for any choice of $0 \leq t_0 < \cdots < t_m$,

$m \in \mathbb{N}$, $h_0, \dots, h_m \in C_0(E_0)$,

$$\mathbb{E}\left[\prod_{k=0}^{m} h_k(\tilde{Z}_{t_k})\right] = \lim_{n\to\infty} \mathbb{E}\left[\prod_{k=0}^{m} h_k(\tilde{Z}^{\tilde{\tau}_n}_{t_k})\right] \overset{(ii)}{=} \lim_{n\to\infty} \mathbb{E}\left[\prod_{k=0}^{m} h_k(Z^{\tau_n}_{t_k})\right]$$
$$= \mathbb{E}\left[\prod_{k=0}^{m} h_k(Z^{\tau_n}_{t_k})\right].$$

Therefore, also the finite-dimensional marginals of $\tilde{Z}$ and Z coincide.

Finally, since now well-posedness of the RCLL-martingale problem for $(\mathcal{L}, \delta_0 \otimes \mu_0)$ in the case $\mu_0 = \delta_{x_0}$ and $x_0 \in E$ is established, Lemma C.3 (v) and [BK93, Thm. 2.1] imply that the RCLL-martingale problem for $(\mathcal{L}, \delta_0 \otimes \mu_0)$ is well-posed also for arbitrary $\mu_0 \in \mathcal{P}(E)$. The exact same argument as above now shows that uniqueness holds even in the class of progressively measurable solutions. □

C.3 From Uniqueness of the Martingale Problem to Uniqueness for the Fokker-Planck Equation

Finally, we put together all results obtained in the previous sections. When σ is continuous, Theorem C.2 can be applied. When σ is not continuous, the following extension of Theorem C.2 will be required:[6]

Theorem C.8 ([Kur98, Thm. 2.7]). *Let E_0 and F be locally compact, separable metric spaces, $D(\mathcal{L}^0) \subset C_b(E_0)$ and $\mathcal{L}^0\colon D(\mathcal{L}^0) \to C_b(E_0 \times F)$ linear. Let $\eta\colon E_0 \times \mathcal{B}(F) \to [0,1]$ be a transition kernel and define*

$$\mathcal{L}_\eta f(x) := \int_F \mathcal{L}^0 f(x,y)\, \eta(x, \mathrm{d}y), \quad f \in D(\mathcal{L}^0). \tag{C.22}$$

For any $y \in F$, define the linear operator $\mathcal{L}^0_y$ with domain $D(\mathcal{L}^0)$ on $C_b(E_0)$ by $f \mapsto \mathcal{L}^0 f(\cdot, y)$. Let $\nu \in \mathcal{P}(E_0 \times F)$ and suppose that

(i) $D(\mathcal{L}^0)$ is an algebra and separates points,

(ii) for any $x \in E_0$ and $y \in F$ there exists a solution to the RCLL-martingale problem for $(\mathcal{L}^0_y, \delta_x)$,

(iii) $\mathcal{L}_\eta$ and $(\mathcal{L}_\eta, \nu)$ satisfy the conditions (ii) and (iv) of Theorem C.2.

Then the conclusion of Theorem C.2 is valid, i.e. uniqueness holds for the forward equation for $(\mathcal{L}_\eta, \nu)$.

In the next lemma we show how to obtain uniqueness for the Fokker-Planck equation for $(\mathcal{D}, \mathcal{A})$ from uniqueness for $(D(\mathcal{L}), \mathcal{L})$.

[6] See footnote 4 for a discussion why (ii) implies that $\mathcal{L}^0_y$ is a pre-generator (in the terminology of [Kur98]).

Lemma C.9. *Suppose σ and $(\mathcal{D},\mathcal{A})$ are as in Theorem C.1 and $(D(\mathcal{L}),\mathcal{L})$ is defined as in* (C.6)*. Let $\mu_0 \in \mathcal{P}(E)$, t_0 as in Assumption A.6 and define $\nu := \delta_0 \otimes \mu_0$. Then the following statements hold:*

(i) If $\sigma\mathcal{A}f \in C_0([0,t_0]\times E)$ for all $f \in \mathcal{D}$, then the assumptions of Theorem C.2 are satisfied.

(ii) Suppose there exists $R > 0$ such that $|\sigma(t,x)| \leq R$ for all $(t,x) \in [0,\infty)\times E$. Define $E_0 := [0,\infty) \times E$, $F := [0,R]$, $D(\mathcal{L}^0) := D(\mathcal{L})$ and for $f \in \mathcal{D}$, $\gamma \in C_c^1[0,\infty)$, set

$$\mathcal{L}^0(f\gamma)((t,x),v) := \gamma(t)v\mathcal{A}f(x) + f(x)\gamma'(t), \quad (t,x) \in E_0,\ v \in F,$$

and linearly extend this definition of $\mathcal{L}^0$ to $D(\mathcal{L}^0)$. Finally, set $\eta((t,x),\cdot) := \delta_{\sigma(t,x)}(\cdot)$ and $\mathcal{L}_\eta$ as in (C.22)*. Then the assumptions of Theorem C.8 are satisfied and $(D(\mathcal{L}),\mathcal{L})$ coincides with $(D(\mathcal{L}^0),\mathcal{L}_\eta)$.*

In particular, in both cases the conclusion of Theorem C.2 is valid, i.e. uniqueness holds for the forward equation (C.5) *for $(\mathcal{L},\nu)$.*

Proof. To prove (i), firstly note that by Assumption A.7 (i) $\mathcal{D}$ is an algebra and dense. Hence, $D(\mathcal{L})$ is an algebra and separates points. Secondly, by Lemma C.3 (v) the condition (C.3) is indeed satisfied. Thirdly, by Proposition C.4 existence holds and fourthly, by Proposition C.7 uniqueness holds. Therefore, assumptions (i)-(iv) of Theorem C.2 are indeed satisfied.

To prove (ii), notice that σ is a measurable function, and thus η is indeed a transition kernel. Furthermore, by definition we have

$$\mathcal{L}h(t,x) = \int_0^R \mathcal{L}^0 h((t,x),v)\,\eta((t,x),\mathrm{d}v) = \mathcal{L}_\eta h(t,x).$$

Hence, $(D(\mathcal{L}),\mathcal{L})$ and $(D(\mathcal{L}^0),\mathcal{L}_\eta)$ indeed coincide and it only remains to show that the assumptions of Theorem C.8 are satisfied. Firstly, $D(\mathcal{L}^0) = D(\mathcal{L})$ is an algebra and separates points as argued in (i). Secondly, for any $v \in [0,R]$ and $(s_0,x) \in E_0$ a solution to the RCLL-martingale problem for $(\mathcal{L}_v^0, \delta_{(s_0,x)})$ can be constructed as follows: Let M be a solution to the RCLL-martingale problem for $(\mathcal{A},\delta_x)$ and set $X_t := M_{vt}$. Then by elementary change of variable,

$$f(X_t) - \int_0^t v\mathcal{A}f(X_s)\,\mathrm{d}s = f(M_{tv}) - \int_0^{tv} \mathcal{A}f(M_s)\,\mathrm{d}s$$

for all $f \in \mathcal{D}, t \geq 0$, and thus X is a solution to the (time-inhomogeneous) RCLL-martingale problem for $(v\mathcal{A},\delta_x)$. Therefore, by Lemma C.3, $(t+s_0,X_t)_{t\geq 0}$ is a solution to the RCLL-martingale problem for $(\mathcal{L}_v^0,\delta_{(s_0,x)})$. Finally, since σ is bounded, $(D(\mathcal{L}),\mathcal{L})$ satisfies (C.3) by Lemma C.3 (v) and the martingale problem for $(\mathcal{L},\nu)$ is well-posed by Proposition C.7. □

After these preparations, we are now ready to prove the main result in this section, Theorem C.1.

Proof of Theorem C.1. Suppose $(p(t,\cdot))_{t\in[0,t_0]}$ and $(\tilde{p}(t,\cdot))_{t\in[0,t_0]}$ both satisfy (C.1) and (C.2). Defining for any $t \geq 0$ the measures $\nu_t := \delta_t \otimes p(t,\cdot)$ and $\tilde{\nu}_t := \delta_t \otimes \tilde{p}(t,\cdot)$, by Lemma C.3 (iv), $(\nu_t)_{t\geq 0}$ and $(\tilde{\nu}_t)_{t\geq 0}$ both satisfy (C.4) and (C.5). Under our assumptions, Lemma C.9 implies that uniqueness holds for (C.5), i.e. $\tilde{\nu}_t = \nu_t$ for all $t \geq 0$ or $\delta_t \otimes p(t,\cdot) = \delta_t \otimes q(t,\cdot)$ for all $t \geq 0$. In particular, $\int_E f(x)\, p(s, \mathrm{d}x) = \int_E f(x)\, \tilde{p}(s, \mathrm{d}x)$ for all $f \in C_0(E)$ and all $s \in [0, t_0]$ and thus the assertion follows. $\square$

Bibliography

[AJK10] L.B.G. Andersen, P. Jäckel, and C. Kahl, *Simulation of square-root processes*, Encyclopedia of Quantitative Finance, John Wiley & Sons, Ltd, 2010.

[AKU16] S. Ankirchner, T. Kruse, and M. Urusov, *Numerical approximation of irregular SDEs via Skorokhod embeddings*, Journal of Mathematical Analysis and Applications **440** (2016), no. 2, 692–715.

[Ama90] H. Amann, *Ordinary differential equations: an introduction to nonlinear analysis*, vol. 13, Walter de Gruyter, 1990.

[Bas83] R.F. Bass, *Skorokhod imbedding via stochastic integrals*, Seminar on probability, XVII, Lecture Notes in Math., vol. 986, Springer, Berlin, 1983, pp. 221–224.

[Bas88] ______, *Uniqueness in law for pure jump Markov processes*, Probability Theory and Related Fields **79** (1988), no. 2, 271–287.

[Bat06] D.S. Bates, *Maximum likelihood estimation of latent affine processes*, Review of Financial Studies **19** (2006), no. 3, 909–965.

[BC74] J.R. Baxter and R.V. Chacon, *Potentials of stopped distributions*, Illinois Journal of Mathematics **18** (1974), no. 4, 649–656.

[BC09a] A. Bain and D. Crisan, *Fundamentals of stochastic filtering*, Springer, 2009.

[BC09b] A. Bentata and R. Cont, *Mimicking the marginal distributions of a semimartingale*, Preprint arXiv:0910.3992 (2009).

[BCH17a] M. Beiglböck, A.M.G. Cox, and M. Huesmann, *The geometry of multi-marginal skorokhod embedding*, Preprint arXiv:1705.09505 (2017).

[BCH17b] ______, *Optimal transport and skorokhod embedding*, Inventiones mathematicae **208** (2017), no. 2, 327–400.

[Ber96] J. Bertoin, *Lévy processes*, Cambridge Tracts in Mathematics, vol. 121, Cambridge University Press, Cambridge, 1996.

[BGKP17] H. Bölcskei, P. Grohs, G. Kutyniok, and P. Petersen, *Optimal approximation with sparsely connected deep neural networks*, Preprint arXiv:1705.01714 (2017).

[BGTW17] H. Bühler, L. Gonon, J. Teichmann, and B. Wood, *Deep hedging*, Preprint (2017).

[BH98] D. Brigo and B. Hanzon, *On some filtering problems arising in mathematical finance*, Insurance: Mathematics and Economics **22** (1998), no. 1, 53–64.

[BHL98] D. Brigo, B. Hanzon, and F. LeGland, *A differential geometric approach to nonlinear filtering: the projection filter*, IEEE Transactions on Automatic Control **43** (1998), no. 2, 247–252.

[BHLP13] M. Beiglböck, P. Henry-Labordère, and F. Penkner, *Model-independent bounds for option prices—a mass transport approach*, Finance and Stochastics **17** (2013), no. 3, 477–501.

[BHR01a] H. Brown, D. Hobson, and L.C.G. Rogers, *The maximum maximum of a martingale constrained by an intermediate law*, Probability Theory and Related Fields **119** (2001), no. 4, 558–578.

[BHR01b] ______, *Robust hedging of barrier options*, Mathematical Finance **11** (2001), no. 3, 285–314.

[BK93] A.G. Bhatt and R.L. Karandikar, *Invariant Measures and Evolution Equations for Markov Processes Characterized Via Martingale Problems*, Annals of Probability **21** (1993), no. 4, 2246–2268.

[BK03] ______, *Martingale Problems and Path Properties of Solutions*, Sankhyā: The Indian Journal of Statistics (2003-2007) **65** (2003), no. 4, 733–743.

[BK06] M. Broadie and Ö. Kaya, *Exact simulation of stochastic volatility and other affine jump diffusion processes*, Operations Research **54** (2006), no. 2, 217–231.

[BL92] J. Bertoin and Y. Le Jan, *Representation of measures by balayage from a regular recurrent point*, Annals of Probability **20** (1992), no. 1, 538–548.

[BN15] B. Bouchard and M. Nutz, *Arbitrage and duality in nondominated discrete-time models*, Annals of Applied Probability **25** (2015), no. 2, 823–859.

[Bru91] M.-F. Bru, *Wishart processes*, Journal of Theoretical Probability **4** (1991), no. 4, 725–751.

[BSV17] P. Bank, H.M. Soner, and M. Voß, *Hedging with temporary price impact*, Mathematics and Financial Economics **11** (2017), no. 2, 215–239.

[BSW13] B. Böttcher, R. Schilling, and J. Wang, *Lévy matters III*, Springer, 2013.

[BTT07] A. Ben-Tal and M. Teboulle, *An old-new concept of convex risk measures: The optimized certainty equivalent*, Mathematical Finance **17** (2007), no. 3, 449–476.

[CC05] J.M.C. Clark and D. Crisan, *On a robust version of the integral representation formula of nonlinear filtering*, Probability Theory and Related Fields **133** (2005), no. 1, 43–56.

[CDFO13] D. Crisan, J. Diehl, P.K. Friz, and H. Oberhauser, *Robust filtering: Correlated noise and multidimensional observation*, Annals of Applied Probability **23** (2013), 2139–2160.

[CFMT11] C. Cuchiero, D. Filipović, E. Mayerhofer, and J. Teichmann, *Affine processes on positive semidefinite matrices*, Annals of Applied Probability **21** (2011), no. 2, 397–463.

[CFY05] P. Cheridito, D. Filipović, and M. Yor, *Equivalent and absolutely continuous measure changes for jump-diffusion processes*, Annals of Applied Probability **15** (2005), no. 3, 1713–1732.

[CGMY04] P. Carr, H. Geman, D.B. Madan, and M. Yor, *From local volatility to local Lévy models*, Quantitative Finance **4** (2004), no. 5, 581–588.

[Cha77] R.V. Chacon, *Potential Processes*, Transactions of the American Mathematical Society **226** (1977), 39–58.

[CIR85] J.C. Cox, J.E. Ingersoll, and S. Ross, *A theory of the term structure of interest rates*, Econometrica **53** (1985), no. 2, 385–407.

[CKT17] C. Cuchiero, I. Klein, and J. Teichmann, *A fundamental theorem of asset pricing for continuous time large financial markets in a two filtration setting*, Preprint arXiv:1705.02087 (2017).

[CL10] P. Carr and R. Lee, *Hedging variance options on continuous semimartingales*, Finance and Stochastics **14** (2010), no. 2, 179–207.

[Cla78] J.M.C. Clark, *The design of robust approximations to the stochastic differential equations of nonlinear filtering*, Communication systems and random process theory (Proc. 2nd NATO Advanced Study Inst., Darlington, 1977), NATO Advanced Study Inst. Ser., Ser. E: Appl. Sci., vol. 25, Sijthoff & Noordhoff, 1978, p. 721–734.

[CM99] P. Carr and D.B. Madan, *Option valuation using the fast fourier transform*, Journal of Computational Finance **2** (1999), no. 4, 61–73.

[CO11] A.M.G. Cox and J. Obłój, *Robust pricing and hedging of double no-touch options*, Finance and Stochastics **15** (2011), no. 3, 573–605.

[COT15] A.M.G. Cox, J. Obłój, and N. Touzi, *The Root solution to the multi-marginal embedding problem: an optimal stopping and time-reversal approach*, Preprint arXiv:1505.03169 (2015).

[CR11] D. Crisan and B. Rozovskii (eds.), *The oxford handbook of nonlinear filtering*, Oxford Handbooks in Mathematics, Oxford University Press, 2011.

[Cri14] D. Crisan, *The stochastic filtering problem: a brief historical account*, Journal of Applied Probability **51A** (2014), 13–22.

[CS03] R.-R. Chen and L. Scott, *Multi-factor cox-ingersoll-ross models of the term structure: Estimates and tests from a kalman filter model*, The Journal of Real Estate Finance and Economics **27** (2003), no. 2, 143–172.

[CT13] C. Cuchiero and J. Teichmann, *Path properties and regularity of affine processes on general state spaces*, Séminaire de Probabilités XLV (Catherine Donati-Martin, Antoine Lejay, and Alain Rouault, eds.), Springer International Publishing, Heidelberg, 2013, pp. 201–244.

[Dav80] M.H.A. Davis, *On a multiplicative functional transformation arising in nonlinear filtering theory*, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **54** (1980), no. 2, 125–139.

[Dav11] ———, *Pathwise nonlinear filtering with correlated noise*, The Oxford Handbook of Nonlinear Filtering, Oxford University Press, 2011, pp. 403–42.

[DFS03] D. Duffie, D. Filipović, and W. Schachermayer, *Affine processes and applications in finance*, Annals of Applied Probability **13** (2003), no. 3, 984–1053.

[DGPR17] L. Döring, L. Gonon, D. Prömel, and O. Reichmann, *On skorokhod embeddings and poisson equations*, Preprint arXiv:1703.05673 (2017).

[DOR14] M.H.A. Davis, J. Obłój, and V. Raval, *Arbitrage bounds for prices of weighted variance swaps*, Mathematical Finance **24** (2014), no. 4, 821–854.

[DPZ93] M.H.A. Davis, V.G. Panas, and T. Zariphopoulou, *European option pricing with transaction costs*, SIAM Journal on Control and Optimization **31** (1993), no. 2, 470–493.

[DS14] Y. Dolinsky and H.M. Soner, *Martingale optimal transport and robust hedging in continuous time*, Probability Theory and Related Fields **160** (2014), no. 1, 391–427.

[Duf01] D. Dufresne, *The integrated square-root process*, Centre for Actuarial Studies, University of Melbourne, 2001, Research Paper no. 90.

[Dup94] B. Dupire, *Pricing with a Smile*, Risk Magazine (1994), 18–20.

[EK86] S.N. Ethier and T.G. Kurtz, *Markov processes. Characterization and convergence*, John Wiley & Sons, 1986.

[ES85] H.-J. Engelbert and W. Schmidt, *On solutions of one-dimensional stochastic differential equations without drift*, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **68** (1985), no. 3, 287–314.

[FF91] N. Falkner and P.J. Fitzsimmons, *Stopping distributions for right processes*, Probability Theory and Related Fields **89** (1991), no. 3, 301–318.

[FH16] H. Feng and D. Hobson, *Gambling in contests with random initial law*, Annals of Applied Probability **26** (2016), no. 1, 186–215.

[Fig08] A. Figalli, *Existence and uniqueness of martingale solutions for SDEs with rough or degenerate coefficients*, Journal of Functional Analysis **254** (2008), no. 1, 109–153.

[Fil88] A.F. Filippov, *Differential Equations with Discontinuous Righthand Sides*, Mathematics and Its Applications, vol. 18, Springer Netherlands, 1988.

[Fil05] D. Filipovic, *Time-inhomogeneous affine processes*, Stochastic Processes and their Applications **115** (2005), no. 4, 639–659.

[FKR10] P.K. Friz and M. Keller-Ressel, *Moment explosions*, John Wiley & Sons, Ltd, 2010.

[FL00] H. Föllmer and P. Leukert, *Efficient hedging: Cost versus shortfall risk*, Finance and Stochastics **4** (2000), 117–146.

[FM09] D. Filipović and E. Mayerhofer, *Affine diffusion processes: theory and applications*, Advanced financial modelling, Radon Ser. Comput. Appl. Math., vol. 8, Walter de Gruyter, Berlin, 2009, pp. 125–164.

[FS16] H. Föllmer and A. Schied, *Stochastic finance: An introduction in discrete time*, Berlin: De Gruyter, 2016.

[GBC16] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*, MIT Press, 2016, `http://www.deeplearningbook.org`.

[GHLT14] A. Galichon, P. Henry-Labordère, and N. Touzi, *A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options*, Annals of Applied Probability **24** (2014), no. 1, 312–336.

[Gla04] P. Glasserman, *Monte carlo methods in financial engineering*, Applications of mathematics : stochastic modelling and applied probability, Springer, 2004.

[GMO15] P. Gassiat, A. Mijatović, and H. Oberhauser, *An integral equation for Root's barrier and the generation of Brownian increments*, Annals of Applied Probability **25** (2015), no. 4, 2039–2065.

[GP99] A.L.J. Geyer and S. Pichler, *A state-space approach to estimate and test multifactor cox-ingersoll-ross models of the term structure*, Journal of Financial Research **22** (1999), no. 1, 107–130.

[GT17] L. Gonon and J. Teichmann, *Affine filtering*, Preprint (2017).

[Hes93] S.L. Heston, *A closed-form solution for options with stochastic volatility with applications to bond and currency options*, The Review of Financial Studies **6** (1993), no. 2, 327–343.

[HN89] S. Hodges and A. Neuberger, *Optimal replication of contingent claims under transaction costs*, The Review of Futures Markets **8** (1989), no. 2, 222–239.

[HN12] D. Hobson and A. Neuberger, *Robust bounds for forward start options*, Mathematical Finance **22** (2012), no. 1, 31–56.

[Hob98] D. Hobson, *Robust hedging of the lookback option*, Finance and Stochastics **2** (1998), no. 4, 329–347.

[Hob11] ———, *The Skorokhod embedding problem and model-independent bounds for option prices*, Paris-Princeton Lectures on Mathematical Finance 2010, Lecture Notes in Math., vol. 2003, Springer, Berlin, 2011, pp. 267–318.

[Hor91] K. Hornik, *Approximation capabilities of multilayer feedforward networks*, Neural Networks **4** (1991), no. 2, 251–257.

[HPRY11] F. Hirsch, C. Profeta, B. Roynette, and M. Yor, *Peacocks and associated martingales, with explicit constructions*, Bocconi & Springer Series, vol. 3, Springer, Milan, 2011.

[IJS09] A. İlhan, M. Jonsson, and R. Sircar, *Optimal static-dynamic hedges for exotic options under convex risk measures*, Stochastic Processes and their Applications **119** (2009), no. 10, 3608–3632.

[IS15] S. Ioffe and C. Szegedy, *Batch normalization: Accelerating deep network training by reducing internal covariate shift*, Proceedings of the 32nd International Conference on Machine Learning, 2015, pp. 448–456.

[JS03] J. Jacod and A.N. Shiryaev, *Limit theorems for stochastic processes*, 2nd ed., Springer, 2003.

[KB15] D.P. Kingma and J. Ba, *Adam: a method for stochastic optimization*, Proceedings of the International Conference on Learning Representations (ICLR) (2015).

[KMK10] J. Kallsen and J. Muhle-Karbe, *Exponentially affine martingales, affine measure changes and exponential moments of affine processes*, Stochastic Processes and their Applications **120** (2010), no. 2, 163–181.

[KMK15] J. Kallsen and J. Muhle-Karbe, *Option pricing and hedging with small transaction costs*, Mathematical Finance **25** (2015), no. 4, 702–723.

[KO88] T.G. Kurtz and D.L. Ocone, *Unique characterization of conditional distributions in nonlinear filtering*, Annals of Probability **16** (1988), no. 1, 80–107.

[KRM15] M. Keller-Ressel and E. Mayerhofer, *Exponential moments of affine processes*, Annals of Applied Probability **25** (2015), no. 2, 714–752.

[KRST11] M. Keller-Ressel, W. Schachermayer, and J. Teichmann, *Affine processes are regular*, Probability Theory and Related Fields **151** (2011), no. 3, 591–611.

[KS91] I. Karatzas and S.E. Shreve, *Brownian motion and stochastic calculus*, second ed., Springer, 1991.

[KS01] T. Kurtz and R. Stockbridge, *Stationary Solutions and Forward Equations for Controlled and Singular Martingale Problems*, Electronic Journal of Probability **6** (2001), no. 17, 1–52.

[KS07] S. Klöppel and M. Schweizer, *Dynamic indifference valuation via convex risk measures*, Mathematical Finance **17** (2007), no. 4, 599–627.

[Kur98] T. Kurtz, *Martingale Problems for Conditional Distributions of Markov Processes*, Electronic Journal of Probability **3** (1998), no. 9, 1–29.

[Kyp14] A.E. Kyprianou, *Fluctuations of Lévy processes with applications*, second ed., Universitext, Springer, Heidelberg, 2014, Introductory lectures.

[Leo09] G. Leoni, *A first course in Sobolev spaces*, Graduate Studies in Mathematics, vol. 105, American Mathematical Society, Providence, RI, 2009.

[LS01] R.S. Liptser and A.N. Shiryaev, *Statistics of Random Processes, Volumes I and II.*, 2nd ed., Springer, 2001.

[LSZ15] K. Law, A. Stuart, and K. Zygalakis, *Data assimilation*, Springer-Verlag, New York, 2015.

[Lum73] G. Lumer, *Perturbation de générateurs infinitésimaux du type "changement de temps"*, Annales de l'institut Fourier **23** (1973), no. 4, 271–279.

[MKRS17] J. Muhle-Karbe, M. Reppen, and H.M. Soner, *A primer on portfolio choice with small transaction costs*, Annual Review of Financial Economics **9** (2017), no. 1, 301–331.

[MMKS11] E. Mayerhofer, J. Muhle-Karbe, and A.G. Smirnov, *A characterization of the martingale property of exponentially affine processes*, Stochastic Processes and their Applications **121** (2011), no. 3, 568–582.

[Mon72] I. Monroe, *Using additive functionals to embed preassigned distributions in symmetric stable processes*, Transactions of the American Mathematical Society **163** (1972), 131–146.

[Obł04] J. Obłój, *The Skorokhod embedding problem and its offspring*, Probability Surveys **1** (2004), 321–390.

[OP09] J. Obłój and M. Pistorius, *On an explicit Skorokhod embedding for spectrally negative Lévy processes*, Journal of Theoretical Probability **22** (2009), no. 2, 418–440.

[OS17] J. Obłój and P. Spoida, *An iterated azéma–yor type embedding for finitely many marginals*, Annals of Probability **45** (2017), no. 4, 2210–2247.

[Roc70] R.T. Rockafellar, *Convex analysis*, Princeton University Press, 1970.

[Rog01] L.C.G. Rogers, *The relaxed investor and parameter uncertainty*, Finance and Stochastics **5** (2001), no. 2, 131–154.

[Rog04] L.C.G. Rogers, *Why is the effect of proportional transaction costs* $O(\delta^{2/3})$, Mathematics of Finance (G. Yin and Q. Zhang, eds.), American Mathematical Society, Providence, RI, 2004, pp. 303–308.

[Ros71] H. Rost, *The stopping distributions of a Markov process*, Inventiones mathematicae **14** (1971), 1–16.

[RS10] L.C.G. Rogers and S. Singh, *The cost of illiquidity and its effects on hedging*, Mathematical Finance **20** (2010), no. 4, 597–615.

[Rud87] W. Rudin, *Real and complex analysis*, third ed., McGraw-Hill Book Co., New York, 1987.

[RW00a] L.C.G. Rogers and D. Williams, *Diffusions, Markov Processes, and Martingales*, second ed., vol. 1, Cambridge University Press, 2000.

[RW00b] L.C.G. Rogers and David Williams, *Diffusions, Markov Processes, and Martingales*, second ed., vol. 2, Cambridge University Press, 2000.

[RY99] D. Revuz and M. Yor, *Continuous martingales and Brownian motion*, 3rd ed., Springer, 1999.

[Sat72] K.-I. Sato, *Potential operators for Markov processes*, Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 3: Probability Theory (Berkeley, Calif.), University of California Press, 1972, pp. 193–211.

[Sat99] ______, *Lévy processes and infinitely divisible distributions*, Cambridge University Press, 1999.

[Sko61] A.V. Skorohod, *Issledovanija po teorii slučajnyh processov: (stohastičaskie differencial'nye uravnenija i predel'nye teoremy dlja processov Markova)*, Izdatel'stvo Kievskogo universiteta, 1961 (Russian).

[Sko65] A.V. Skorokhod, *Studies in the theory of random processes*, Translated from the Russian by Scripta Technica, Inc, Addison-Wesley Publishing Co., Inc., Reading, Mass., 1965.

[SS13] C. Seel and P. Strack, *Gambling in contests*, Journal of Economic Theory **148** (2013), no. 5, 2033–2048.

[SSC95] H.M. Soner, S.E. Shreve, and J. Cvitanić, *There is no nontrivial hedging portfolio for option pricing with transaction costs*, Annals of Applied Probability **5** (1995), no. 2, 327–355.

[Str75] Daniel W. Stroock, *Diffusion processes associated with Lévy generators*, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **32** (1975), no. 3, 209–244.

[SV10] P. Spreij and E. Veerman, *The affine transform formula for affine jump-diffusions with general closed convex state spaces.*, Preprint arXiv:1005.1099 (2010).

[vH07] R. van Handel, *Filtering, stability, and robustness*, Ph.D. thesis, California Institute of Technology, 2007.

[WW97] A.E. Whalley and P. Wilmott, *An asymptotic analysis of an optimal hedging model for option pricing with transaction costs*, Mathematical Finance **7** (1997), no. 3, 307–324.

[Xu06] M. Xu, *Risk measure pricing and hedging in incomplete markets*, Annals of Finance **2** (2006), no. 1, 51–71.