



DISSERTATION | DOCTORAL THESIS

Titel | Title

Signature methods and polynomial McKean-Vlasov equations
for financial modeling

verfasst von | submitted by

Janka Oona Möller BSc ETH MSc ETH

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Doktorin der Naturwissenschaften (Dr.rer.nat.)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree
programme code as it appears on the
student record sheet:

UA 796 605 405

Dissertationsgebiet lt. Studienblatt | Field of
study as it appears on the student record
sheet:

Mathematik

Betreut von | Supervisor:

Univ.-Prof. Dipl.-Ing. Dr. Christa Cuchiero Privatdoz.

Abstract

The aim of this thesis is to explore novel approaches to modeling and learning from the dynamics of financial markets. In that, the focus lies in incorporating path-dependence and distributional dependencies into the description of financial systems.

The first problem we tackle is portfolio optimization using streams of market data in the context of Stochastic Portfolio Theory (SPT). We extend functionally generated portfolios originally introduced in SPT to the notion of linear path-functional portfolios. The latter are constructed via certain linear combinations of non-anticipative path-functionals serving as feature maps, of which a special case are elements of the signature of paths of relevant factors. We study theoretical aspects of signature portfolios such as universal approximation of any continuous path-dependent portfolio and portfolio selection problems such as expected log-utility maximization and mean-variance type optimizations. We show that both take a convex quadratic form. We illustrate these theoretical findings with various numerical experiments. On one hand, we demonstrate in simulated markets that the trained signature portfolios are indeed remarkably close to the growth-optimal portfolios. On the other hand, we train signature portfolios using real market data of the NASDAQ, S&P500 and SMI, and find that the learned portfolios out-perform the respective market portfolio in almost all cases, over the considered out-of-sample period, even under transaction costs.

In a second line of research regarding path-dependence, we develop a path-dependent stochastic volatility model and calibrate it to VIX and SPX implied volatility surfaces jointly. To achieve both tractability and universality, we propose to model the volatility process via the signature of an underlying polynomial process. This enables us to derive closed form expressions for the VIX and SPX index, respectively. Moreover, we show that the calibration task can be split into an offline sampling and online learning phase, that Fourier pricing techniques are available and propose variance reduction techniques in the case of Monte-Carlo simulations. We consolidate these findings to calibrate the model to real market data, and find that the calibrated implied volatilities are indeed highly accurate.

Finally, we address distributional dependencies in the evolution of financial systems by introducing so-called polynomial and affine McKean-Vlasov stochastic differential equations (SDEs). We study them using the theory of time-inhomogeneous polynomial and affine processes. In the polynomial case, we let the dependence on the law of the SDEs coefficients break down to a dependence on the (conditional) moments of the process, emulating the form of a standard polynomial process. A special case of such McKean-Vlasov SDEs were previously used to model capitalization curves in the context of SPT. In the affine case, we can allow the coefficients to depend on the characteristic function more generally. Both approaches allow us to study solutions of the SDEs in a Markovian setting and to establish new existence and uniqueness results which are, to the best of our knowledge, not obtainable using standard techniques.

Kurzfassung

Das Ziel dieser Arbeit ist es, neue Ansätze zu erforschen, die zur Modellierung von und zum Lernen aus den Dynamiken der Finanzmärkte dienen. Dabei liegt der Fokus darauf, Pfadabhängigkeit und Verteilungsabhängigkeit in die Untersuchung von Finanzsystemen einzubeziehen.

Die erste Fragestellung, die wir angehen, ist Portfolio-Optimierung unter Verwendung von Zeitreihen im Kontext der Stochastischen Portfoliotheorie (SPT). Wir erweitern die ursprünglich in der SPT eingeführten funktional generierten Portfolios auf lineare Pfad-Funktional-Portfolios. Letztere werden durch bestimmte lineare Kombinationen von nicht-antizipierenden Pfadfunktionalen konstruiert, die als Feature-Maps dienen, wobei die Elemente der Signatur von Pfaden relevanter Faktoren ein Spezialfall davon sind. Wir untersuchen theoretische Aspekte von Signatur-Portfolios, wie universelle Approximation jedes stetigen pfadabhängigen Portfolios, und Portfolioselektion, wie die Maximierung des erwarteten logarithmischen Nutzens und Mean-Variance-Optimierung. Wir zeigen, dass beide eine konvexe quadratische Form annehmen. Diese theoretischen Erkenntnisse illustrieren wir mit verschiedenen numerischen Experimenten. Einerseits zeigen wir in simulierten Märkten, dass die trainierten Signatur-Portfolios tatsächlich bemerkenswert nahe an den Growth-Optimal Portfolios liegen. Andererseits trainieren wir Signatur-Portfolios mit realen Marktdaten von NASDAQ, S&P500 und SMI und stellen fest, dass die gelernten Portfolios das jeweilige Marktportfolio in fast allen Fällen über den betrachteten Out-of-Sample-Zeitraum, selbst unter Berücksichtigung von Transaktionskosten, übertreffen.

In einer zweiten Forschungslinie zur Pfadabhängigkeit entwickeln wir ein pfadabhängiges stochastisches Volatilitätsmodell und führen eine gemeinsame Kalibrierung von den impliziten Volatilitätsflächen von VIX und SPX durch. Um sowohl gute analytische Handhabbarkeit als auch Universalität zu erreichen, modellieren wir den Volatilitätsprozess über die Signatur eines zugrunde liegenden polynomiellen Prozesses. Dies ermöglicht es uns, geschlossene Formeln für den VIX- und den SPX-Index abzuleiten. Darüber hinaus zeigen wir, dass die Kalibrierungsaufgabe in eine Offline-Sampling-Phase und eine Online-Lernphase unterteilt werden kann, dass Fourier-Bepreisungstechniken verfügbar sind und schlagen Varianz-Reduktionstechniken im Fall von Monte-Carlo-Simulationen vor. Wir konsolidieren diese Erkenntnisse, indem wir das Modell mit realen Marktdaten kalibrieren, und stellen fest, dass die kalibrierten implizierten Volatilitäten tatsächlich sehr genau sind.

Schließlich adressieren wir Verteilungsabhängigkeit in der Beschreibung von Finanzsystemen, indem wir sogenannte polynomielle und affine McKean-Vlasov stochastische Differenzialgleichungen (SDEs) einführen. Dabei untersuchen wir diese unter Verwendung der Theorie zeitinhomogener polynomialer und affiner Prozesse. Im polynomiellen Fall brechen wir die Abhängigkeit der SDE-Koeffizienten von der gesamten Verteilung auf eine Abhängigkeit von den (bedingten) Momenten des Prozesses hinunter, in einer Form, welche die eines klassischen

polynomiellen Prozesses nachahmt. Ein Spezialfall solcher McKean-Vlasov SDEs wurde zuvor verwendet, um Kapitalisierungskurven im Kontext von SPT zu modellieren. Im affinen Fall können wir die Koeffizienten allgemeiner von der charakteristischen Funktion abhängig machen. Beide Ansätze ermöglichen es uns, Lösungen der SDEs in einem Markov-Setting zu untersuchen und neue Existenz- und Eindeutigkeitsresultate zu etablieren, die unseres Wissens nach, mit Standardtechniken nicht erreichbar sind.

Acknowledgements

First and foremost, I would like to thank Christa Cuchiero, for her guidance, support and encouragement. Her knowledge and enthusiasm for mathematics have provided a very stimulating environment with plenty of opportunities to grow, and I cannot put in words how much I have learned from her.

I am also very grateful to Xin Guo for inviting me to visit her at UC Berkeley and for many fruitful discussions.

Moreover, I would like to express my appreciation to Hao Ni and Thorsten Schmidt for reviewing my thesis.

Many thanks to everyone of the research group for being such great company, whether for chats over a cup of coffee or mathematical discussions. For that, special thanks go to Yuan Chen, Guido Gazzani, Luca Gonzato and Francesca Primavera.

I am very thankful to my family for their support throughout these years.

My deepest gratitude goes to Benjamin for his unconditional love and support. I could not have undertaken this journey without you.

Contents

Abstract	i
Kurzfassung	iii
Acknowledgements	v
Introduction	1
1.1 Signatures, polynomial processes, and the combination of the two	3
1.2 Signature methods in Stochastic Portfolio Theory	3
1.3 Joint calibration to SPX and VIX options with signature-based models	5
1.4 Polynomial and affine McKean-Vlasov SDEs	6
Bibliographic information on the papers contained in this thesis	9
1 Signature methods in Stochastic Portfolio Theory	11
1.1 Introduction	11
1.2 Preliminaries	15
1.2.1 The signature of a continuous semimartingale	15
1.2.2 Universal approximation theorem for linear functions on the signature	18
1.2.3 Global universal approximation theorem for linear functions on the	21
signature	21
1.2.4 The Johnson-Lindenstrauss lemma	25
1.2.5 The market and important portfolio specifications	25
1.3 (Linear) path-functional portfolios	29
1.4 Approximation properties of signature portfolios	33
1.4.1 Universal approximation of the growth-optimal portfolio	35
1.4.2 A class of non-Markovian markets where the growth-optimal portfolio	37
is a signature portfolio	37
1.5 Optimization tasks for linear path-functional portfolios	38
1.5.1 Mean-variance optimization	40
1.5.2 Optimizing the (expected) log-(relative)-wealth	43
1.5.3 Constraints and regularization	46
1.6 Transaction costs for portfolios with short-selling and without bank account .	50
1.7 Numerical results	52
1.7.1 Simulated data	52
1.7.2 Real market data	55
Appendix of Chapter 1	73

Contents

1.A	Proofs of Section 1.2	73
1.B	Proofs of Section 1.4	76
1.C	Proofs of Section 1.5	77
1.D	Proofs of Section 1.6	78
2	Joint calibration to SPX and VIX options with signature-based models	79
2.1	Introduction	79
2.1.1	State of the art	82
2.2	Signature: definition and properties	83
2.3	The model	87
2.4	Expected signature of polynomial diffusion	90
2.5	VIX options with signatures	94
2.5.1	Explicit formulas for the VIX	95
2.5.2	Variance reduction for pricing VIX options	99
2.5.3	Calibration to VIX options	101
2.5.4	The case of time-varying parameters	105
2.6	SPX as a signature-based model	106
2.6.1	Exploiting the affine nature of the signature: Fourier pricing of SPX and VIX options	110
2.6.2	The case of time-varying parameters	113
2.7	Joint calibration of SPX and VIX options	114
2.7.1	Numerical results	115
	Appendix of Chapter 2	123
2.A	Numerical results for the Brownian motion case	123
2.B	On the stability of the calibrated parameters	124
3	Polynomial and Affine McKean-Vlasov SDEs	129
3.1	Introduction	129
3.2	Preliminaries	132
3.2.1	Time-inhomogeneous polynomial processes	132
3.2.2	Time-inhomogeneous affine processes	137
3.3	Existence and uniqueness of solutions to time-inhomogeneous diffusions	138
3.4	Polynomial McKean-Vlasov equations	143
3.4.1	Primal point of view	144
3.4.2	Dual point of view	156
3.4.3	Adding common noise	160
3.5	Affine McKean-Vlasov equations	172
3.5.1	Outline of the affine case	172
3.5.2	Moments	175
3.5.3	Density	178
3.5.4	Logarithm of conditional exponential moments	179
	Appendix of Chapter 3	181
3.A	ODE theory	181
3.B	Proofs	181

Bibliography

183

Introduction

If one were to tell a tale about the developments in mathematical finance, it may start like this: *In the beginning, there were constants*. This alludes to both the Black-Scholes model, where the drift and volatility are assumed to be constant and Markowitz' approach to portfolio optimization, in which the portfolio weights are chosen to be constant. Moreover, the log-optimal portfolio in a Black-Scholes market has constant weights. Although these approaches are elegant and lead to high tractability, it is well-understood that the reality is far more complex.

One of many examples where this over-simplification becomes apparent is the fact that the Black-Scholes model leads to constant implied volatility surfaces for options on the modelled underlying, which – from the beginning on – was not in line with the observed volatility-smiles in real financial markets. This led to the development of *local* volatility models and *stochastic* volatility models. Some of those are of an *affine* form such as the Heston model and thereby retained tractability in option pricing and calibration thanks to the availability of Fourier pricing techniques. Since the inception of VIX derivatives in the early 2000s and their increase in popularity to hedge volatility exposures, a new modeling challenge arose: the joint calibration to VIX and SPX options. However, the aforementioned models are not rich enough to model the implied volatility surfaces of VIX and SPX options jointly. Hence, this task called for another leap in complexity. Indeed, the so-called *joint calibration problem* has only recently been "solved" by models involving rough volatility, jumps, Markovian approaches with the market's forward variance curve as additional input or path-dependence (see Subsection 2.1.1 for an overview). Our contribution falls under the latter class, by introducing a tractable path-dependent stochastic volatility model based on the combination of signature methods and polynomial processes. We elaborate on our findings in Subsection 1.3.

A second stream of developments was initiated by Fernholz, the founder of the so-called *Stochastic Portfolio Theory* (SPT), see Fernholz (2002). A core concept of SPT are *functionally generated portfolios*, which are systematically constructed via the observed relative capitalizations of the stocks in the market. It can be shown that these portfolios lead, under assumptions such as "market-diversity and non-degeneracy of the diffusion matrix" or "sufficient volatility" of the underlying market model, to relative-arbitrage strategies. Although the objective at hand is different from the one of Markowitz, it can conceptually be seen as another example of the evolution from constants towards dependence on market quantities, modeled via stochastic processes. What we mean by that is, that the portfolio weights are no-longer assumed to be constant but rather a function of the vector of relative market capitalizations. Not surprisingly, recent developments seek to incorporate even more complex dependencies, namely path-dependence of the relative market capitalizations and additionally

a process of finite variation [Schied et al. \(2018\)](#); [Karatzas and Kim \(2019\)](#). Our research is related to these most recent developments. In that, our contribution lies in generalizing the functional form of portfolios to a path-functional one, and to provide a framework for optimizing such portfolios, which is computationally tractable. We explain our approach and main results in more detail in Subsection [1.2](#).

Another key aspect of SPT is the observation that the relative market capitalizations of the companies in ranked order have remained remarkably stable over time see [Fernholz \(2002\)](#); [Campbell and Wong \(2022\)](#). By a company's rank we here mean its relative size in the market, i.e. the largest company has rank 1, the second-largest stock's rank is 2, etc. Modeling these market capitalization curves via interactions between the stocks as in so-called "volatility stabilized markets models" (see [Fernholz and Karatzas \(2005\)](#)) and taking the limit as the number of companies tends to infinity (see e.g. [Shkolnikov \(2012\)](#); [Cuchiero and Huber \(2024\)](#)) yields a market model in the form of an SDE, where the coefficients depend on the marginal law of its solution. Equations of such a form are called *McKean-Vlasov SDEs*. This leads us to another key aspect of modern financial modeling, namely the incorporation of distributional dependencies. We study the existence and uniqueness of solutions to such SDEs using the theory of time-inhomogeneous affine and polynomial processes, introducing the notion of *polynomial and affine McKean-Vlasov SDEs*. In that, we do not only embed a specific instance of a McKean-Vlasov SDE which was first studied in [Cuchiero and Huber \(2024\)](#) into a general framework, but also propose a novel solution technique for McKean-Vlasov SDEs. We present a more detailed summary of our contributions in Subsection [1.4](#).

As outlined here, incorporating path-dependence and distributional dependencies into the study of financial systems is of significant interest in modern financial modeling while it poses at the same time substantial challenges. The complexities arise primarily from the need to deal with infinite dimensional objects (paths or distributions) while the system to be described is finite dimensional (e.g. the price processes of stocks). In particular, such models are in general no longer Markovian in finite dimensions. The overarching theme of the approaches presented in this thesis is to find finite-dimensional representations of such infinite dimensional objects which are on one hand rich enough to emulate the distributional resp. path-dependencies and on the other hand lead to tractability of the problems. To do so, we use so-called signature methods to deal with path-dependencies (Chapter [1](#) & [2](#)), leverage the polynomial technology to handle distributional dependencies (Chapter [3](#)) and combine the two methods to achieve tractability in our path-dependent stochastic volatility model (Chapter [2](#)).

Before we turn to presenting the results of our research in the respective areas, we give a brief (conceptual) introduction to signatures and polynomial processes; two fascinating concepts that lie at the heart of this thesis.

I.1 Signatures, polynomial processes, and the combination of the two

Conceptually, the signature of a path was first introduced by [Chen \(1957\)](#) and has become an important mathematical object, especially in the field of rough path theory, see e.g. [Lyons et al. \(2007\)](#); [Friz and Victoir \(2010\)](#). For an $\mathbb{R}^d$ -valued continuous semimartingale $(X_t)_{t \geq 0}$ its signature $(\mathbb{X}_t)_{t \geq 0}$ is defined as the collection of all iterated integrals of X . Let us make this a bit more precise; the signature $\mathbb{X}_t$ at some point $t \geq 0$ is given by

$$\mathbb{X}_t := \left(1, X_t - X_0, \int_0^t \int_0^s \otimes \circ dX_s \otimes \circ dX_t, \dots, \int_0^t \int_0^{s_N} \dots \int_0^{s_1} \otimes \circ dX_{s_1} \dots \otimes \circ dX_t, \dots \right)$$

where $\circ$ denotes the Stratonovic integral. Conceptually, the signature (and linear functions thereof) can be seen as a path-space analogue of polynomials. One aspect where this analogy becomes apparent, is that linear functions on the signature are dense in the set of continuous (w.r.t. an appropriate variational metric) path-functionals, similar to the density of polynomials in continuous functions on compact subsets of $\mathbb{R}^d$. This is known as the *universal approximation property* of linear functions on signatures, and we will make this statement more precise in Section 1.2.2. However, it should not come as a surprise that this remarkable property makes signatures very useful in modeling path-dependencies. Let us also refer the reader to [Ni \(2012\)](#); [Lyons et al. \(2024\)](#) for more results on fascinating properties of the signature of stochastic processes.

In one sentence, one may introduce polynomial processes as a class of Markov processes whose moments can be obtained by solving a particular ODE. In the time-homogeneous case, this ODE reduces to a linear one and thus admits a solution in the form of a simple matrix-exponential. The time-inhomogeneous case calls for a so-called Magnus expansion. This result is referred to as the *moment-formula*. First introduced in [Cuchiero et al. \(2012\)](#), polynomial processes have since then gained importance in stochastic analysis and mathematical finance due to their desirable properties. In the context of this thesis, we are primarily interested in polynomial diffusions, which we will introduce in more detail in Chapter 2 for the time-homogeneous case and in Chapter 3 for a time-inhomogeneous setting. Moreover, we show in Section 2.4 that the truncated signature of a polynomial diffusion is itself a polynomial diffusion. This result makes the combination of the two concepts particularly appealing and will enable us to ensure the tractability of our path-dependent stochastic volatility model in Chapter 2.

I.2 Signature methods in Stochastic Portfolio Theory

In Chapter 1, we use signature methods for systematic portfolio construction using streams of market information. We take inspiration from Fernholz' *functionally generated portfolios* and generalize them in several aspects:

- instead of taking as an input the current observation of the market weights, we construct portfolio weights from a stream of market information (including historical observations);

- we extend the log-gradient of a function in Fernholz' construction to any continuous path-functional;
- we construct the portfolio weights via an auxiliary portfolio τ which can be seen as a benchmark portfolio, as it will always be attainable in our portfolio optimization tasks. This means that the portfolio τ is included in the set of portfolios over which we optimize.

To this end, we introduce the notion of *path-functional portfolios* which are portfolios whose weights $(\pi_t^1, \dots, \pi_t^d)_{t \in [0, T]}$ are of the form

$$\pi_t^i(\tau, X) = \tau_t^i \left(f^i(t, X_{[0, t]}) + 1 - \sum_{j=1}^d \tau_t^j f^j(t, X_{[0, t]}) \right)$$

for $1 \leq i \leq d$ where d are the number of stocks in the market, $\{f^i\}_{1 \leq i \leq d}$ are continuous (in some variation-metric) non-anticipative path-functionals, X is assumed to be a multidimensional continuous semimartingale and τ is an auxiliary portfolio. In terms of portfolio optimization, we focus on portfolios where each $f^i(t, X_{[0, t]}) = \sum_{\nu \in \mathcal{V}} l_\nu^i \phi^\nu(t, X_{[0, t]})$ i.e. it is given by a linear combination of feature maps $\{\phi^\nu\}_{\nu \in \mathcal{V}}$ which are themselves continuous non-anticipative path-functionals, some set of feature-labels $\mathcal{V}$ and constant, deterministic (optimization) parameters $\{l_\nu^i\}_{\nu \in \mathcal{V}, 1 \leq i \leq d}$. This opens the door to not only using signatures for portfolio construction but various other machine learning tools, such as random neural networks [Herrera et al. \(2024\)](#); [Huang et al. \(2006\)](#); [Gallicchio and Scardapane \(2020\)](#); [Cao et al. \(2018\)](#), reservoir computing [Grigoryeva and Ortega \(2018b\)](#) and echo-state neural networks [Jaeger \(2001\)](#); [Grigoryeva and Ortega \(2018a\)](#); [Gonon and Ortega \(2021\)](#). We show that linear path-functional portfolio lead to remarkably tractable optimization problems, even when incorporating transaction costs. More specifically, we show that both the log-utility maximization and a mean-variance type optimization problem lead to a convex quadratic form. To incorporate proportional transaction costs, while preserving the convex quadratic structure, we propose a "regularization" for transaction costs, which indeed proves to increase the performance (w.r.t. the non-regularized portfolios) under transaction costs in our numerical experiments. When choosing the feature maps to be elements of the signatures, we call the resulting portfolios *signature portfolios*. These are of special interest to us due to their approximation properties. Building on the (*global*) *universal approximation theorem* for linear functions on signatures, we show that:

- signature portfolios can approximate any continuous path-functional portfolio arbitrarily well, on compact sets;
- in a large class of non-Markovian markets, signature portfolios can uniformly approximate the growth-optimal portfolio, on compact sets.

To underline the latter, we run several numerical experiments in simulated market settings, and find that indeed the out-of-sample realizations of the signature portfolios' weights are remarkably close to those of the respective growth-optimal portfolio.

Moreover, we train signature portfolios in various settings using real market data of the NASDAQ, S&P500 and SMI. The large number of stocks in those markets, however, pose problems in the computation of the signature, as the number of components grows prohibitively large. To deal with such high-dimensional environments, we propose two randomized versions of the signature portfolios, one based on randomized signatures and the other on a Johnson-Lindenstrauss projection of the signature. For the latter, we develop a memory-efficient algorithm to enable computational tractability. In our numerical experiments using real market data, we find that in most cases the signature portfolios out-perform the respective market portfolio over the out-of-sample period, even under transaction costs.

In summary; signature methods enabled us to extract expressive features from streams of data to successfully optimize portfolios in a computationally tractable manner.

1.3 Joint calibration to SPX and VIX options with signature-based models

As outlined above, the joint calibration to VIX and SPX options is a major challenge in stochastic volatility modeling. The two crucial aspects in this task are:

- a model needs to be rich enough to be able to produce the VIX and SPX smiles jointly;
- by definition, the VIX index itself is given by a conditional expectation which poses challenges for numerical tractability, as in many models this calls for nested Monte-Carlo procedures.

We address the former by proposing a class of path-dependent stochastic volatility models using signature methods. More specifically, we choose a primary process $(X_t)_{t \geq 0}$ and model the volatility $(\sigma_t^S)_{t \geq 0}$ as a linear function ℓ on the signature of the time-augmented process $(\hat{X}_t)_{t \geq 0}$:

$$dS_t(\ell) = S_t(\ell) \sigma_t^S(\ell) dB_t \text{ with } \sigma_t^S(\ell) := \ell(\hat{\mathbb{X}}_t) \text{ and for } S_0 = s_0 > 0,$$

where $(B_t)_{t \geq 0}$ is a one-dimensional Brownian motion. In order to obtain an efficient pricing formula for the VIX index, we choose the primary process $(X_t)_{t \geq 0}$ to be a polynomial diffusion, which allows us to compute the conditional expectation of $\ell(\hat{\mathbb{X}}_t)$ in closed-form. By that, we obtain a closed-form expression for the VIX itself, in form of a simple matrix exponential, and have therefore alleviated the need for a nested Monte-Carlo procedure. Moreover, the log-price $\log(S_t(\ell))$ can also be expressed as a function of the signature of $(t, X_t, B_t)_{t \geq 0}$. The structure of our model allows us to split the calibration task into an *offline sampling* part, where we sample realizations of $(\hat{\mathbb{X}}_t)_{t \geq 0}$, and an *online optimization* part, where we perform the joint calibration by optimizing the parameters of the linear map ℓ . Apart from these properties which are beneficial for numerical tractability, our class of models benefits from the universal approximation property of the signature, meaning that we can approximate any stochastic volatility model, whose volatility is given by a continuous path-functional of $(X_t)_{t \geq 0}$. Additionally, our class nests several classical models, see Remark 2.3.4. By combining signature methods with polynomial diffusions, we achieve another desirable feature; the

availability of Fourier-pricing techniques, which we outline in more detail in Subsection 2.6.1.

Given these theoretical results, we then turn to our model's performance when calibrated to real VIX and SPX implied volatility surfaces. We perform the joint calibration in various settings, for several short maturities as well as for a wider range of maturities, using time-varying parameters which allow us to capture regime-switching effects. In all of these experiments, we find that our model's calibrated implied volatility surfaces are remarkably close to the true ones. Given the large number of calibration parameters in our model, we additionally perform a parameter-stability analysis. That means, we re-calibrate our model weekly and analyze the change in the optimal parameters. We observe that while the parameters of course change when re-calibrated, the parameters do not vary drastically but rather change in a continuous fashion.

Conceptually, our class of models can be seen as a factor model, where the factors are elements of the signature. Thanks to the richness of linear functions on the signature, we have found a finite-dimensional representation for path-dependence, which is rich enough to model the VIX and SPX options jointly and have retained tractability using the polynomial technology.

I.4 Polynomial and affine McKean-Vlasov SDEs

In Chapter 3, we aim to establish existence and uniqueness of McKean-Vlasov SDEs using the theory of time-inhomogeneous polynomial processes, a class of Markov processes first introduced in [Hurtado and Schmidt \(2020\)](#). To do so, we study McKean-Vlasov SDEs in which the dependence on the law takes the form of a dependence on the moments. This particular structure then allows us to utilize the theory of Markov processes by augmenting the solution by its moments. This can be seen as a Markovian lift, where the one-dimensional solution to the SDE itself is non-Markovian, but together with its moments it is. Let us make this more clear by stating one of our main results:

Theorem. *Let $(W_t)_{t \geq 0}$ be a one-dimensional Brownian motion and fix some $N > 0$. We call an SDE of the form*

$$dZ_t = \left(b(\mathbb{E}[Z_t], \dots, \mathbb{E}[Z_t^N]) + \beta(\mathbb{E}[Z_t], \dots, \mathbb{E}[Z_t^N]) Z_t \right) dt + \sqrt{c(\mathbb{E}[Z_t], \dots, \mathbb{E}[Z_t^N]) + \gamma(\mathbb{E}[Z_t], \dots, \mathbb{E}[Z_t^N]) Z_t + \Gamma(\mathbb{E}[Z_t], \dots, \mathbb{E}[Z_t^N]) Z_t^2} dW_t \quad (\text{poly-MV})$$

a polynomial McKean-Vlasov SDE for some initial condition $Z_0 = z \in \mathbb{R}$ and some continuous maps $b, \beta, c, \gamma, \Gamma : \mathbb{R}^N \rightarrow \mathbb{R}$ which are such that the term under the square-root is non-negative. If the set of Assumptions A1 or Assumptions A2 hold and $b, \beta, c, \gamma, \Gamma$ are locally Lipschitz-continuous on $\mathbb{R}^N$, then (poly-MV) admits a path-wise unique strong solution for all $t \in [0, \infty)$. Moreover, $(Z_t)_{t \in [0, \infty)}$ is a time-inhomogeneous polynomial process.

We would like to point out that we present various extensions of the above theorem in Chapter 3, in particular we study other state-spaces and formulate an alternative set of assumptions on the maps $b, \beta, c, \gamma, \Gamma$.

As alluded to before, we augment the process by its moments to establish existence of a solution to (poly-MV). Therefore, a crucial step is to derive the system of ODEs which is satisfied by the moments. We can do so in a primal and a dual approach. In the primal approach, we essentially apply Ito's formula together with some Gronwall inequalities to find the corresponding ODE and conclude *a posteriori* that the process $(Z_t)_{t \geq 0}$ is a time-inhomogeneous polynomial process. In the dual view-point, however, we start by studying a particular time-inhomogeneous polynomial process and establish *a posteriori* that this process is a solution to (poly-MV). Although the two approaches lead to the same sufficient conditions for existence of solutions to (poly-MV), there are some essential differences in the approaches, which we outline in Remark 3.4.17.

In addition to our study of McKean-Vlasov SDEs of the form (poly-MV) we propose two further extensions:

- another main result lies in our study of polynomial McKean-Vlasov SDEs *with common noise*. In that, we generalize a particular case studied in Cuchiero and Huber (2024) which arises in the context of Stochastic Portfolio Theory. In the case with common noise, the coefficients of the SDE now depend on the *conditional* moments of $(Z_t)_{t \geq 0}$ and, in addition to establishing existence of solutions, we provide conditions under which the process and its conditional moments are jointly polynomial.
- we propose so-called *affine McKean-Vlasov SDEs*, where we leverage the theory of time-inhomogeneous affine processes. In contrast to the polynomial case, the SDE's coefficients may now depend more generally on the characteristic function of $(Z_t)_{t \geq 0}$ and not just on the moments. Our procedure for establishing solutions is similar to the dual approach in the polynomial case. However, instead of a system of ODE's for the moments, we now obtain a PDE for the characteristic function. We explore in Section 3.5 various examples which allow for explicit computations, but have not yet worked out a well-rounded theory for the affine case.

Bibliographic information on the papers contained in this thesis

This thesis is a collection of the following papers. We made minor modifications to the respective paper in order for them to better fit the format of this thesis.

Chapter 1:

Signature methods in Stochastic Portfolio Theory

Christa Cuchiero, Janka Möller

Not published elsewhere yet.

Contribution of the author of this dissertation:

The paper was worked out in close collaboration.

Online versions available at: <https://arxiv.org/abs/2310.02322>

The paper is submitted to the SIAM Journal on Financial Mathematics.

Chapter 2:

Joint calibration to SPX and VIX options with signature-based models

Christa Cuchiero, Guido Gazzani, Janka Möller, Sara Svaluto-Ferro

Published in Mathematical Finance, 2024, pages 1–53.

Contribution of the author of this dissertation:

The paper was worked out in close collaboration.

Online versions available at: <https://arxiv.org/abs/2301.13235> and
<https://onlinelibrary.wiley.com/doi/10.1111/mafi.12442?af=R>

Chapter 3:

Polynomial and affine McKean-Vlasov SDEs

Christa Cuchiero, Janka Möller

Not published elsewhere yet.

Contribution of the author of this dissertation:

The paper was worked out in close collaboration.

1 Signature methods in Stochastic Portfolio Theory

Abstract

In the context of Stochastic Portfolio Theory, we introduce a novel class of portfolios which we call linear path-functional portfolios. These are portfolios which are determined by certain transformations of linear functions of a collection of feature maps that are non-anticipative path functionals of an underlying semimartingale. As main example for such feature maps we consider the signature of the (ranked) market weights. We prove that these portfolios are universal in the sense that every continuous, possibly path-dependent, portfolio function of the market weights can be uniformly approximated by signature portfolios. We also show that signature portfolios can approximate the growth-optimal portfolio in several classes of non-Markovian market models arbitrarily well and illustrate numerically that the trained signature portfolios are remarkably close to the theoretical growth-optimal portfolios. Besides these universality features, the main numerical advantage lies in the fact that several optimization tasks like maximizing (expected) logarithmic wealth or mean-variance optimization within the class of linear path-functional portfolios reduce to a convex quadratic optimization problem, thus making it computationally highly tractable. We apply our method also to real market data based on several indices. Our results point towards out-performance on the considered out-of-sample data, also in the presence of transaction costs.

1.1 Introduction

Portfolio theory and portfolio optimization are a core topic in mathematical finance and have been the subject of vivid research for decades. The most prominent example is of course the mean-variance optimization problem as formulated by Harry Markowitz, the founder of *Modern Portfolio Theory* (see [Markovitz \(1959\)](#)). His framework takes into account risk-preferences of investors while offering a tractable form of optimization, namely being convex and quadratic. Although the choice of constant portfolio weights is arguably oversimplified, the elegance of the Markowitz model can not be denied. Ever since, researchers have been striving for more and more realistic models with fewer limiting assumptions while hoping to preserve tractability.

In the spirit of coming up with more realistic assumptions, for instance without the need of claiming a specific form of the hardly measurable drift, Robert Fernholz developed *Stochastic Portfolio Theory* (SPT) (see [Fernholz \(2002\)](#) and also [Fernholz and Karatzas \(2010\)](#)). A key feature of SPT, again in the realm of more realistic modeling assumptions, is the rather

relaxed no-arbitrage condition. Indeed, the price process is only assumed to be a (continuous) semimartingale satisfying *No Unbounded Profit with Bounded Risk* and not necessarily the stronger condition of *No free lunch with vanishing risk* (see [Delbaen and Schachermayer \(1994\)](#)). It is also central that the portfolios' performance is measured with respect to a benchmark portfolio, which is usually chosen to be the market portfolio, i.e. large indices like the S&P500. At least since the introduction of very liquid ETFs replicating market indices the out-performance of the market is indeed a major challenge for investors. In this context Fernholz introduced *functionally generated portfolios* and derived the so-called *master formula* describing the relative wealth process of such portfolios and allowing for the detection for relative arbitrages with respect to the market portfolio. These portfolios are constructed via the log-gradient of a function of the market weights. While appreciating the beauty of the framework and also its robustness in view of certain optimal worst case long-run growth rates (see [Kardaras and Robertson \(2021\)](#); [Itkin et al. \(2022\)](#)), one may point out that the log-gradient form as well as the neglect of past information is somewhat restrictive. To reduce these limitations various generalizations of the original framework have been developed, e.g., towards replacing the market weights by market-to-book ratios [Kim \(2023\)](#), considering trading strategies generated by Lyapunov functions [Karatzas and Ruf \(2017\)](#), adding extra information in the form of a finite variation process [Ruf and Xie \(2020\)](#) and incorporating past information in a path-dependent setting [Schied et al. \(2018\)](#); [Karatzas and Kim \(2019\)](#). In this context it is worth mentioning that the latter two papers do not work in a stochastic setting but rather in a path-wise one, using functional Itô-calculus as in [Cont and Fournie \(2010\)](#); [Dupire \(2019\)](#) based on Föllmer integration. An alternative path-wise formulation allowing for more general strategies (beyond gradient type) and building on the theory of rough path has been studied in [Allan et al. \(2023\)](#).

In this chapter, we consider for simplicity the stochastic setting of continuous semimartingales (even though a path-wise formulation in the spirit of [Allan et al. \(2023\)](#) would also be possible), take inspiration from the functionally generated portfolios and generalize them to what we call *path-functional portfolios*. These are portfolios constructed via an auxiliary portfolio τ and non-anticipative path-functionals with a general semimartingale as input. In doing so, we get rid of any gradient-type form of the functionals and allow for a general benchmark portfolio τ and past information of the market or other exogenous relevant factors in form of a general continuous semimartingale. The choice of continuous semimartingales is here only for the ease of exposition, but could be replaced by other (predictable stochastic) processes. This generalization can be achieved by using rough path theory, where similar universal approximation results hold for more general processes, see e.g. [Cuchiero et al. \(2022\)](#) for rough paths with jumps.

One of our main interest concerns *tractable portfolio optimization* in the framework of SPT using such path-functional portfolios. To do so, we focus on *linear path-functional portfolios*, which are constructed via a linear combination of possibly path-dependent feature maps and constant optimization parameters.

This allows us to incorporate modern machine learning techniques for time-series data, in particular *signature methods* (see e.g. [Kalsi et al. \(2020\)](#); [Bayer et al. \(2021\)](#); [Futter et al.](#)

(2023) in the area of optimal control, Bühler et al. (2020); Ni et al. (2022) on applications to data generation/simulation, Lyons et al. (2019); Cuchiero et al. (2023a, 2022) for topics in financial modelling, Compagnoni et al. (2022); Cuchiero et al. (2021a,b); Akyildirim et al. (2023) on randomization techniques and their application, Levin et al. (2013); Kiraly and Oberhauser (2019); Chevyrev and Oberhauser (2022); Andrès et al. (2024); Akyildirim et al. (2022); Wu et al. (2021); Wang et al. (2019); Lyons et al. (2014) as well as Chapter 2 for other applications in mathematical finance and machine learning), but also other tools, like random neural networks Herrera et al. (2024); Huang et al. (2006); Gallicchio and Scardapane (2020); Cao et al. (2018). In the current chapter, our focus lies on signature methods, which play an important role in rough path theory (see e.g. Lyons (1998); Friz and Victoir (2010)) and whose appeal stems from the *(global) universal approximation theorem*. It states that linear functions on the signature can approximate continuous (with respect to certain variation distances) path-functionals arbitrarily well on compact sets of paths or even globally when using the setup of weighted spaces (see Cuchiero et al. (2023b)). This result then motivates the notion of *signature portfolios*, which are linear path-functional portfolios with feature maps being elements of the signature. Indeed, with this choice we can approximate generic path-functional portfolios including functionally generated portfolios and the growth-optimal portfolio in a large class of non-Markovian markets, arbitrary well.

Despite this versatility, signature portfolios and linear path-functional portfolios in general lead to remarkably tractable optimization tasks. More concretely, we show that maximizing the (expected) logarithmic wealth as well as the mean-variance optimization of portfolio returns lead to *convex quadratic optimization problems*. We would like to highlight that the Markowitz portfolio itself is included in the class of signature portfolios and moreover, that any chosen benchmark portfolio τ can always be *attained* by signature portfolios constructed via τ .

In addition to our theoretical results, we optimize signature portfolios in numerical experiments using simulated and real market data. By means of simulated data generated from the Black-Scholes model, volatility stabilized models Fernholz and Karatzas (2005) and so-called signature market models, we illustrate that the trained signature portfolios (of small degree) are remarkably close to the theoretical growth-optimal portfolios.

In the real market situation, we additionally incorporate transaction costs in our optimization by proposing a regularization under which the optimization problems remain convex and quadratic. To deal with high-dimensional market indices such as the NASDAQ and S&P500, we use dimension reduction techniques leading to what we call randomized signature portfolios and JL-signature portfolios, where the former rely on randomized signatures (see Akyildirim et al. (2022); Compagnoni et al. (2022); Cuchiero et al. (2021a,b) and also Muca Cirone et al. (2023) for a connection to neural signature kernels), while the latter are based on an explicit Johnson-Lindenstrauss projection. Note that we do not reduce the dimension of the underlying process, but only of its signature. To make the computation of the JL-signature portfolios feasible, we present a memory efficient algorithm to do so. We train the portfolios in both a log-relative wealth and mean-variance optimization in the NASDAQ, the SMI and S&P500 markets. We use randomized- and JL-signature portfolios in the high-dimensional markets (NASDAQ and S&P500) and signature portfolios for the SMI. In the NASDAQ market, the

portfolios are computed from the ranked market weights, while in the other two cases the name-based market weights are used. We incorporate proportional transaction costs of 1% and 5% respectively in the name-based markets. In almost all configurations, our trained portfolios outperform the market portfolios during the out-of-sample period, even under 5% of proportional transaction costs. It is worth noting that in the setting with transaction costs, we still re-balance our portfolios daily, whereas the market portfolio *does not* pay any transaction costs.

To provide more context to our findings, we highlight two papers which are particularly related to our research, namely [Campbell and Wong \(2022\)](#) and [Futter et al. \(2023\)](#). In [Campbell and Wong \(2022\)](#), the authors study portfolio optimization in SPT over a family of rank-based functionally generated portfolios parameterized by an exponentially concave function. Their objective is to maximize the relative logarithmic growth rate, which in their parameterization leads to a convex optimization problem. In their empirical study, they use (ranked) market data of the 100 largest US stocks, which they manage to out-perform during the out-of-sample period with their trained portfolios. As they do not include transaction costs in the learning procedure (which however could be done by adopting e.g. similar transaction cost treatments as ours), the out-performance does no longer work with 0.45% of proportional transaction costs. This could also be related to the fact that the considered portfolios are long only functionally generated and thus a rather small class. Indeed, the class of portfolios we consider is more general in several respects: we allow for short-selling, the inclusion of information from the past and from exogenous signals as well as for a general benchmark portfolio τ in the construction of the portfolios. At the same time, the portfolios of [Campbell and Wong \(2022\)](#) can be approximated arbitrarily well using signature portfolios, which is a result of the universal approximation theorem.

Very recently, the work of [Futter et al. \(2023\)](#) appeared, where the authors study mean-variance optimization of the wealth process using signature methods. In contrast to our multiplicative setting, they consider an additive approach where the trading strategies correspond to numbers of shares, which are then directly parameterized as linear functions on the signature. This means in particular that the strategies are only self-financing if a bank-account is included, while in our case the strategies are always self-financing *without* a bank-account, since we use portfolio weights. Even though the formulation of [Futter et al. \(2023\)](#), where the *wealth process* itself is the quantity of interest, is the standard one in the literature of the continuous-time mean-variance portfolio selection problem (see e.g. [Zhou and Li \(2000\)](#)), it differs from the original approach of Markowitz [Markowitz \(1959\)](#), where mean-variance optimization of the *returns* was actually the objective. The latter is advantageous when optimizing (in a model free setup) along the observed trajectory, since the returns are more likely to be stationary than the assets themselves. For these reasons, we use returns and a multiplicative approach, which allows in particular to include the *classical Markowitz portfolio* (as a very special case) in our setup (which is not possible in [Futter et al. \(2023\)](#)). Another difference is that we consider general *linear* path-functionals and not only signature portfolios, while still obtaining a convex quadratic optimization problem. Additionally, we include transaction costs and dimension reduction techniques to be able to deal with 500-dimensional price processes, whereas in the numerical implementations of [Futter](#)

et al. (2023) only two-dimensional price processes are considered. Let us also point out that in Fütter et al. (2023) the signal is augmented with the two-dimensional lead-lag process to express Itô-integrals as linear functions on the corresponding signature. We usually only augment with time, which – under certain conditions on the quadratic variation – still allows us to get linear expressions for the Itô-integrals (see Remark 1.5.12). As the computation of the signature becomes expensive for higher dimensions, this might be of relevance.

The remainder of the article is structured as follows. In Section 1.2 we recall the notion of the signature of continuous semimartingales, provide two versions of the universal approximation theorem, state the Johnson-Lindenstrauss Lemma and specify the financial market setting as well as classes of important portfolios. Section 1.3 is dedicated to the introduction of (linear) path-functional portfolios with signature portfolios as special case, while Section 1.4 addresses their approximation properties, in particular in view of approximating the growth optimal portfolio. In Section 1.5 the convex quadratic optimization tasks for the mean-variance and (expected) log-(relative) wealth are discussed, and Section 1.6 explains the treatment of proportional transaction costs. Section 1.7 concludes the chapter with a presentation of the numerical results. Some proofs are gathered in the Appendix.

1.2 Preliminaries

1.2.1 The signature of a continuous semimartingale

In this section, we first introduce the signature of an $\mathbb{R}^n$ -valued continuous semimartingale X and then state the so-called *universal approximation theorem* (see Theorem 1.2.12), which motivates the use of signature methods. In the following, we denote by X an $\mathbb{R}^n$ -valued continuous semimartingale defined on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]})$ with $T > 0$ finite. We start by introducing several preliminary notions needed to define the signature. Similar introductions can for instance be found in Lyons et al. (2007); Friz and Victoir (2010). Readers who are neither interested in a mathematical introduction of the signature nor the associated approximation theorems of linear functions of the signature in general or signature portfolios in particular, may choose to skip Sections 1.2.1, 1.2.2, 1.2.3 and 1.4.

Definition 1.2.1 ((Truncated) Extended Tensor Algebra and Group). The signature of X , denoted by $\mathbb{X}$, is an element of the extended tensor algebra $T((\mathbb{R}^n))$ defined as

$$T((\mathbb{R}^n)) := \left\{ (a_0, a_1, \dots) \mid a_k \in (\mathbb{R}^n)^{\otimes k}, k \geq 0 \right\}$$

with $(\mathbb{R}^n)^{\otimes 0} := \mathbb{R}$. For $\mathbf{a}, \mathbf{b} \in T((\mathbb{R}^n))$ we define the following operations:

$$\mathbf{a} + \mathbf{b} = (a_0 + b_0, a_1 + b_1, \dots),$$

$$\mathbf{a} \otimes \mathbf{b} = (c_0, c_1, \dots) \text{ for } c_k = \sum_{j=0}^k a_j \otimes b_{k-j},$$

$$\lambda \mathbf{a} = (\lambda a_0, \lambda a_1, \dots) \text{ for } \lambda \in \mathbb{R}.$$

Note that $T((\mathbb{R}^n))$ forms a real non-commutative unital algebra under these operations with unit element $\mathbf{1} = (1, 0, 0, \dots)$. Moreover, $T_1((\mathbb{R}^n)) \subset T((\mathbb{R}^n))$, defined as

$$T_1((\mathbb{R}^n)) := \{\mathbf{a} \mid \mathbf{a} \in T((\mathbb{R}^n)) \text{ with } a_0 = 1\}$$

is a group under the operation $\otimes$ with unit element $\mathbf{1} = (1, 0, 0, \dots)$. For $\mathbf{a} \in T_1((\mathbb{R}^n))$ its inverse is defined as

$$\mathbf{a}^{-1} = \sum_{k \geq 0} (\mathbf{1} - \mathbf{a})^k.$$

We define the *truncated* tensor algebra at level N by

$$T^N(\mathbb{R}^n) := \{(a_0, a_1, \dots, a_N) \mid a_k \in (\mathbb{R}^n)^{\otimes k}, 0 \leq k \leq N\}$$

and denote by $T_1^N(\mathbb{R}^n)$ the corresponding *truncated* tensor group. We denote by $(e_1, \dots, e_n)$ the canonical basis of $\mathbb{R}^n$. Let $I = (i_1, \dots, i_m)$ be a multi-index and denote its length by $|I| = m$, then $e_I := e_{i_1} \otimes \dots \otimes e_{i_m}$ a basis element of $(\mathbb{R}^n)^{\otimes m}$. We use the convention that $|I| = 0$ means that $I = \emptyset$, the empty word. In this formulation, we can write any $\mathbf{a} \in T^N(\mathbb{R}^n)$ as

$$\mathbf{a} = \sum_{0 \leq |I| \leq N} \langle e_I, \mathbf{a} \rangle e_I,$$

where $\langle \cdot, \cdot \rangle$ is defined as the inner product of $(\mathbb{R}^n)^{\otimes k}$ for each $k \geq 0$. Finally, we introduce the two canonical projections

$$\begin{aligned} \Pi_k : T((\mathbb{R}^n)) &\rightarrow (\mathbb{R}^n)^{\otimes k} \\ \mathbf{a} &\mapsto a_k, \end{aligned}$$

$$\begin{aligned} \Pi_{\leq N} : T((\mathbb{R}^n)) &\rightarrow T^N(\mathbb{R}^n) \\ \mathbf{a} &\mapsto (a_0, \dots, a_N). \end{aligned}$$

Of course, these projections can also be applied on the truncated tensor algebra, i.e. $\Pi_k : T^N(\mathbb{R}^n) \rightarrow (\mathbb{R}^n)^{\otimes k}$ for $k \leq N$ and $\Pi_{\leq N} : T^M(\mathbb{R}^n) \rightarrow T^N(\mathbb{R}^n)$ for $M \geq N$.

We define accordingly $T_0^N(\mathbb{R}^d) := \{\mathbf{a} \in T^N(\mathbb{R}^d) \mid a_0 = 0\}$. In fact, it holds that $T_1^N(\mathbb{R}^d)$ is a Lie group under the tensor product $\otimes$ and $T_0^N(\mathbb{R}^d)$ is a Lie algebra with Lie bracket $T_0^N(\mathbb{R}^d) \times T_0^N(\mathbb{R}^d) \ni (\mathbf{a}, \mathbf{b}) \mapsto [\mathbf{a}, \mathbf{b}] := \Pi_{\leq N}(\mathbf{a} \otimes \mathbf{b} - \mathbf{b} \otimes \mathbf{a}) \in T_0^N(\mathbb{R}^d)$. The two are related by the exponential map

$$\begin{aligned} \exp : T_0^N(\mathbb{R}^d) &\rightarrow T_1^N(\mathbb{R}^d) \\ \mathbf{a} &\mapsto \Pi_{\leq N} \left(1 + \sum_{k=1}^N \frac{\mathbf{a}^{\otimes k}}{k!} \right). \end{aligned}$$

We call $\mathfrak{g}^N(\mathbb{R}^d) = \mathbb{R}^d \oplus [\mathbb{R}^d, \mathbb{R}^d] \oplus \dots \oplus [\mathbb{R}^d, [\dots, [\mathbb{R}^d, \mathbb{R}^d]]]$ the *free step- N nilpotent Lie algebra* and $G^N(\mathbb{R}^d) = \exp(\mathfrak{g}^N(\mathbb{R}^d))$ the *free nilpotent group of step N* .

We are now ready to define the signature of a continuous semimartingale. For an introduction of the signature of a continuous semimartingale in a similar spirit, see [Cuchiero et al. \(2023a\)](#). To this end, consider a finite time-horizon $T > 0$ and some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, \mathbb{P})$.

Definition 1.2.2 (Signature of a Continuous Semimartingale). For an $\mathbb{R}^n$ -valued continuous semimartingale X , we denote its signature by $\mathbb{X}$ and define $\mathbb{X}$ as the process in $T_1((\mathbb{R}^n))$ whose value at time t is given via

$$\Pi_k(\mathbb{X}_t) = \int_{0 \leq u_1 \leq \dots \leq u_k \leq t} \otimes \circ dX_{u_1} \cdots \otimes \circ dX_{u_k},$$

where $\circ$ denotes Stratonovic integration. Accordingly, we define the truncated signature at level N , which we denote by $\mathbb{X}^N$ as

$$\mathbb{X}_t^N = \Pi_{\leq N}(\mathbb{X}_t).$$

Note that $\mathbb{X}_t^N$ takes values in $G^N(\mathbb{R}^n)$ for all $t \in [0, T]$. We define the increments of the signature $\mathbb{X}_{s,t}$ by $\mathbb{X}_{s,t} := \mathbb{X}_s^{-1} \otimes \mathbb{X}_t$ and note that

$$\Pi_k(\mathbb{X}_{s,t}) = \int_{s \leq u_1 \leq \dots \leq u_k \leq t} \otimes \circ dX_{u_1} \cdots \otimes \circ dX_{u_k},$$

in particular, $\Pi_1(\mathbb{X}_{s,t}) = X_t - X_s$.

Let us now state a few important and useful properties about the signature.

Lemma 1.2.3 (Shuffle Product Property). For two multiindices I, J we define the shuffle product $\sqcup$ as

$$\begin{aligned} e_I \sqcup e_J &= e_{\emptyset} \sqcup e_I = e_I \\ (e_I \otimes e_i) \sqcup (e_J \otimes e_j) &= (e_I \sqcup (e_J \otimes e_j)) \otimes e_i + ((e_I \otimes e_i) \sqcup e_J) \otimes e_j. \end{aligned}$$

Then, for any two multiindices I, J and any continuous semimartingale X , it holds that

$$\langle e_I, \mathbb{X} \rangle \langle e_J, \mathbb{X} \rangle = \langle e_I \sqcup e_J, \mathbb{X} \rangle.$$

Proof. This follows directly from the properties of the Stratonovic integral. $\square$

Lemma 1.2.4 (Uniqueness of the Signature). Let X be a continuous semimartingale with at least one strictly monotone component. Then the signature uniquely determines X up to vertical translations of the trajectories.

Proof. This follows from the fact, that the signature determines the underlying semimartingale uniquely up to tree-like equivalences, see [Boedihardjo et al. \(2016\)](#). A more direct proof, in the case where that component is the time, can be found e.g. in [Cuchiero et al. \(2022\)](#). See also [Chang et al. \(2017\)](#); [Boedihardjo et al. \(2014\)](#) for further considerations related to this. $\square$

We now define a class of functions which plays a major role in this chapter, namely *linear functions on the signature*:

Definition 1.2.5 (Linear Functions on the Signature). Let X be an $\mathbb{R}^n$ -valued continuous semimartingale and $\mathbb{X}$ its signature. We call a function $L : T((\mathbb{R}^n)) \rightarrow \mathbb{R}$ a *linear function on the signature* if L is given by

$$L(\mathbb{X}_t) = \sum_{0 \leq |I| \leq N} l_I \langle e_I, \mathbb{X}_t \rangle$$

for some coefficients $l_I \in \mathbb{R}$ and any $N \in \mathbb{N}$. It is important to note that all l_I are *constant in time*. We denote the set of all linear functions on the signature by $\mathfrak{L}$.

1.2.2 Universal approximation theorem for linear functions on the signature

Linear functions on the signature are dense in a set of certain continuous path-functionals defined on compact sets of paths. We call this the *universal approximation property* and state it rigorously in Theorem 1.2.12. To this end we need the following preparations which have been formulated in a similar manner in Bayer et al. (2021); Kalsi et al. (2020); Litterer and Oberhauser (2011).

Let us first define for some $p \in (2, 3)$ the *p-variation metric* of paths $\mathbf{x}, \mathbf{y}$ with values in the group $T_1^N(\mathbb{R}^n)$ by

$$d_{p-var;[0,t]}(\mathbf{x}, \mathbf{y}) = \max_{k \in 1, \dots, N} \sup_{D \in \mathcal{D}} \left(\sum_{t_i, t_{i-1} \in D} |\Pi_k(\mathbf{x}_{t_{i-1}, t_i} - \mathbf{y}_{t_{i-1}, t_i})|_k^{\frac{p}{k}} \right)^{\frac{k}{p}}, \quad (1.2.1)$$

where $\mathcal{D}$ is the set of all partitions of $[0, t]$ and $|\cdot|$ is the Euclidean distance. The definition of increments of paths with values in the group $T_1^N(\mathbb{R}^n)$ is as for the signature in Definition 1.2.2, namely $\mathbf{x}_{s,t} := \mathbf{x}_s^{-1} \otimes \mathbf{x}_t$.

Definition 1.2.6 (Lifted Paths). Let $\varphi : [0, T] \rightarrow \mathbb{R}$ be a strictly monotone increasing function. For $N \geq 2$ and $t \in [0, T]$ we define by

$$\mathcal{G}_t^N(\varphi, x) = \bigcup_{X \text{ cont. semimart.}} \{\hat{\mathbb{X}}_{[0,t]}^N(\omega) \mid \hat{X}_s = (\varphi(s), X_s), \omega \in \Omega_X, X_0 = x\}$$

the space of *lifted paths* $\hat{\mathbb{X}}_{[0,t]}^N(\omega)$ of all $\mathbb{R}^n$ -valued continuous semimartingales with initial value x and for each X , $\Omega_X \subset \Omega$ is taken to be a probability-one set on which the metric given in (1.2.1) is well-defined for all $t \in [0, T]$. Note that we denote by $\hat{\mathbb{X}}$ the signature of $\hat{X}$ and we use the subscript $[0, t]$ to emphasize that we take the *entire path* between $[0, t]$. We equip $\mathcal{G}_t^N(\varphi, x)$ with the metric $d_{p-var;[0,t]}$ (now for paths with values in $T_1^N(\mathbb{R}^{n+1})$).

Remark 1.2.7. We would like to point out that using a strictly monotone increasing function φ other than the identity can be useful in practice. For example, we use the time-augmentation $\varphi_T(t) := \frac{t}{T}$ for our experiments using real market data in Section 1.7.2 to account for the different length of the training and testing period.

Definition 1.2.8 (Space of Lifted (Stopped) Paths). We define the space of *lifted stopped paths* via

$$\Lambda_T^N(\varphi, x) = \bigcup_{t \in [0, T]} \mathcal{G}_t^N(\varphi, x)$$

We then equip $\Lambda_T^N(\varphi, x)$ with the metric d_Λ with its form being based on [Cont and Fournie \(2010\)](#); [Dupire \(2019\)](#)

$$\begin{aligned} d_\Lambda \left(\hat{\mathbb{X}}_{[0, t]}^N(\omega_1), \hat{\mathbb{Y}}_{[0, s]}^N(\omega_2) \right) \\ = |t - s| + d_{p-var; [0, t \vee s]} \left(\widehat{\mathbb{X}}_{[0, t \vee s]}^t(\omega_1), \widehat{\mathbb{Y}}_{[0, t \vee s]}^s(\omega_2) \right), \end{aligned}$$

where we denote by $\widehat{X}^t$ the process $(\widehat{X}^t)_u = (\varphi(u), X_u^t)$, i.e. the process where we stop the semimartingale X at time t but not the time-augmentation φ and write $\widehat{\mathbb{X}}^t$ for the signature of $\widehat{X}^t$. See [Bayer et al. \(2021\)](#) for a similar definition in the setting of general rough paths.

Remark 1.2.9. Note that if $\hat{\mathbb{X}}_{[0, t]}^N(\omega) \in \Lambda_T^N(\varphi, x)$, then $\widehat{\mathbb{X}}_{[0, T]}^t(\omega) \in \mathcal{G}_T^N(\varphi, x)$ and hence $\widehat{\mathbb{X}}_{[0, T]}^t(\omega) \in \Lambda_T^N(\varphi, x)$. This is the reason, we do not stop the time-augmentation φ . Indeed, $(\hat{\mathbb{X}}^N)_{[0, T]}^t(\omega) \notin \mathcal{G}_T^N(\varphi, x)$, since φ is required to be strictly monotone.

Based on [Cont and Fournie \(2010\)](#) we formulate the notion of non-anticipative path-functionals in our setting:

Definition 1.2.10 (Non-Anticipative Path-Functionals). We call functionals $f : \Lambda_T^2(\varphi, x) \rightarrow \mathbb{R}$ *non-anticipative path-functionals*.

The following continuity result for the signature is crucial for the universal approximation property.

Corollary 1.2.11 (Lyon's Lift). *For all $t \in [0, T]$ and $N \geq 2$ there exists a unique bijection*

$$S^N : \mathcal{G}_t^2(\varphi, x) \mapsto \mathcal{G}_t^N(\varphi, x)$$

with inverse $\Pi_{\leq 2}$ and S^N is continuous on bounded sets. We call S^N Lyons' lift. Moreover, for each X and all $\omega \in \Omega_X$ it holds that

$$S^N \left(\hat{\mathbb{X}}_{[0, t]}^2(\omega) \right) = \hat{\mathbb{X}}_{[0, t]}^N(\omega).$$

Proof. The first part of the theorem follows from ([Friz and Victoir, 2010](#), Theorem 8.10, Theorem 9.5 and Corollary 9.11) and the second statement follows from ([Friz and Victoir, 2010](#), Proposition 17.1 and Exercise 17.2). $\square$

Relying on two lemmas proved in Appendix 1.A, we can now show the universal approximation theorem for linear functions on the signature. We here state a version that holds uniformly in time. For this we also provide a rigorous proof in Appendix 1.A which is – up to our knowledge – not available in the literature. For similar assertions we however refer to Bayer et al. (2021); Kalsi et al. (2020). In the setting of càdlàg rough paths an analogous result has been proved in Cuchiero et al. (2022), however with different topologies.

Theorem 1.2.12. *Let $K \subset \mathcal{G}_T^2(\varphi, x)$ be a compact subset. For any continuous non-anticipative path-functional $f : \Lambda_T^2(\varphi, x) \rightarrow \mathbb{R}$ and for every $\epsilon > 0$ there exists a linear function on the signature L such that*

$$\sup_{(t, \hat{\mathbb{X}}_{[0,t]}^2(\omega)) \in [0,T] \times K} |f(\hat{\mathbb{X}}_{[0,t]}^2(\omega)) - L(\hat{\mathbb{X}}_t)(\omega)| < \epsilon.$$

Let us point out that, conceptually, the reason for Theorem 1.2.12 to hold uniformly in time is due to the time-augmentation. To show the applicability of this result let us give some examples of non-anticipative path-functionals.

Example 1.2.13 (Examples of Continuous Non-Anticipative Path-Functionals).

- (i) **Solutions of SDEs driven by X :** For $\gamma > 2$, an $\mathbb{R}^n$ -valued continuous semimartingale X and an $\mathbb{R}^m$ -valued continuous process of bounded variation H , consider the SDE

$$dY_t = \sum_{i=1}^n V_i(Y_t) \circ dX_t^i + \sum_{j=1}^m W_j(Y_t) dH_t^j, \quad Y_0 = y_0 \in \mathbb{R}^d,$$

where $\{V_i\}_{1 \leq i \leq n}$ is a collection of vector fields in $\text{Lip}^\gamma(\mathbb{R}^d)$ and $\{W_i\}_{1 \leq i \leq m}$ is a collection of $\text{Lip}^1(\mathbb{R}^d)$ vector fields.¹ Set $Z = (X, H)$ and $z = (X_0, H_0)$ and define D_Z to be the set

$$D_Z = \bigcup_{t \in [0, T]} \left\{ \hat{\mathbb{Z}}_{[0,t]}^2(\omega) \mid \omega \in \Omega_Z \right\} \subset \Lambda_T^2(\varphi, z).$$

Then the components of the solution to this SDE $(Y_t^i)_{t \in [0, T]}$, can be seen as a map

$$\begin{aligned} Y_t^i(\omega) : D_Z &\rightarrow \mathbb{R} \\ \hat{\mathbb{Z}}_{[0,t]}^2(\omega) &\mapsto Y_t^i(\omega) \end{aligned}$$

and it holds almost surely (i.e. on Ω_Z) that $Y_t^i \in C(D_Z; \mathbb{R})$ for every $t \in [0, T]$. The continuity follows from the fact that the solution to the corresponding rough differential equation solves the SDE almost surely by (Friz and Victoir, 2010, Theorem 17.3) and the RDE solution is continuous (Friz and Victoir, 2010, Corollary 10.28 & Theorem

¹Following the definition in (Friz and Victoir, 2010, Definition 10.2) a vector field V is in Lip^γ if V is $\lfloor \gamma \rfloor$ times continuously differentiable and if the supremum norm of V and the supremum norm of its first $\lfloor \gamma \rfloor$ derivatives admits a common bound.

12.10) in the lifted input path. Hence, we can apply Theorem 1.2.12 and obtain that for any compact set $K \subset \{\hat{Z}_{[0,T]}^2(\omega) | \omega \in \Omega_Z\}$ and all $\epsilon > 0$ it holds that for all $i \in \{1, \dots, d\}$ there exists a linear function L^i on the signature such that

$$\sup_{(t \in [0,T], \hat{Z}_{[0,T]}^2(\omega)) \in K} \left| Y_t^i(\omega) - L^i(\hat{Z}_t(\omega)) \right| < \epsilon.$$

- (ii) **Continuous Functions Without Path-Dependence:** Consider an arbitrary but fixed $\mathbb{R}^n$ -valued continuous semimartingale X with initial value $X_0 = x$. For any $t \in [0, T]$ let

$$\begin{aligned} h : \mathbb{R}^{d+1} &\rightarrow \mathbb{R} \\ \hat{X}_t(\omega) &\mapsto h\left(\hat{X}_t(\omega)\right) \end{aligned}$$

be a continuous function. Such a function can for almost all $\omega \in \Omega$ be extended to a non-anticipative path-functional $\tilde{h}$ on $\Lambda_T^2(\varphi, x)$. To be precise, simply define $\tilde{h}$ as

$$\begin{aligned} \tilde{h} : \Lambda_T^2(\varphi, x) &\rightarrow \mathbb{R} \\ \hat{\mathbb{X}}_{[0,t]}^2(\omega) &\mapsto \tilde{h}\left(\hat{\mathbb{X}}_{[0,t]}^2(\omega)\right) := h\left((\hat{X}_{[0,t]})_{\odot}(\omega)\right) = h\left(\hat{X}_t(\omega)\right). \end{aligned}$$

In particular, $\tilde{h} \in C(\Lambda_T^2(\varphi, x); \mathbb{R})$. To see this, note that for any $\hat{\mathbf{x}}_{[0,t]}, \hat{\mathbf{y}}_{[0,s]} \in \Lambda_T^2(\varphi, x)$ it holds that

$$d_{\Lambda}(\hat{\mathbf{x}}, \hat{\mathbf{y}}) \geq |\hat{x}_{0,t} - \hat{y}_{0,s}| = |\hat{x}_t - \hat{y}_s|$$

with $\hat{x} = \Pi_1(\hat{\mathbf{x}})$, $\hat{y} = \Pi_1(\hat{\mathbf{y}})$, where we replaced the $\max_{k \in \{1, \dots, N\}}$ by $k = 1$ and $\sup_{D \in \mathcal{D}}$ by $D = [0, t \vee s]$ in the definition of $d_{p-var; [0, t \vee s]}$ to obtain the inequality. The last equality follows from the fact that all paths in $\Lambda_T^2(\varphi, x)$ have the same initial value. Due to the continuity of h , for every $\epsilon > 0$ there exists a $\delta > 0$ such that for all $t_1, t_2 \in [0, T]$ and all paths $X(\omega_1), Y(\omega_2)$ satisfying $|\hat{X}_{t_1}(\omega_1) - \hat{Y}_{t_2}(\omega_2)| < \delta$ we have

$$\left| h\left(\hat{X}_{t_1}(\omega_1)\right) - h\left(\hat{X}_{t_2}(\omega_2)\right) \right| < \epsilon.$$

By the above estimate, $d_{\Lambda}\left(\hat{\mathbb{X}}_{[0,t_1]}^2(\omega_1), \hat{\mathbb{X}}_{[0,t_2]}^2(\omega_2)\right) < \delta$ therefore implies

$$\left| \tilde{h}\left(\hat{\mathbb{X}}_{[0,t_1]}^2(\omega_1)\right) - \tilde{h}\left(\hat{\mathbb{X}}_{[0,t_2]}^2(\omega_2)\right) \right| < \epsilon$$

and thus proves the claim.

1.2.3 Global universal approximation theorem for linear functions on the signature

In this subsection, we state an alternative universal approximation theorem which is due to Cuchiero et al. (2023b). We will refer to Theorem 1.2.20 also as *global* universal approximation

theorem because we are no longer limited to compact sets, however this comes at the price of approximating with respect to weighted norms instead of the uniform norm. The following is based on the statements and the results of [Cuchiero et al. \(2023b\)](#) adapted to the setting considered in this chapter. However, we do formulate everything in the path-functional setting here, leading to a approximation *uniformly in time* which has not been done in [Cuchiero et al. \(2023b\)](#).

Let us start by introducing the following metrics:

Definition 1.2.14 (The Carnot-Carathéodory Norm and Homogeneous α -Hölder Metric).

(i) Let $\mathbf{a} \in G^N(\mathbb{R}^n)$, we define the *Carnot-Carathéodory* norm $\|\cdot\|_{cc}$ by

$$\|\mathbf{a}\|_{cc} := \inf \left\{ \int_0^T \|dX_t\| \mid X : [0, T] \rightarrow \mathbb{R}^d \text{ continuous and of finite variation,} \right. \\ \left. \text{such that } \mathbb{X}_T^N = \mathbf{a} \right\}.$$

Moreover, for $\mathbf{a}, \mathbf{b} \in G^N(\mathbb{R}^n)$ the corresponding metric $d_{cc}(\cdot, \cdot)$ is induced by

$$d_{cc}(\mathbf{a}, \mathbf{b}) := \|\mathbf{a}^{-1} \otimes \mathbf{b}\|_{cc}.$$

(ii) Recall the space of lifted path $\mathcal{G}_t^N(\varphi, x)$ for $N \geq 2$ and $t \in [0, T]$. For two lifted paths $\hat{\mathbb{X}}_{[0,t]}^N(\omega_1), \hat{\mathbb{Y}}_{[0,t]}^N(\omega_2) \in \mathcal{G}_t^N(\varphi, x)$ we define the *homogeneous α -Hölder metric* $d_{\alpha, [0,t]}(\cdot, \cdot)$ for $0 \leq \alpha < \frac{1}{2}$ by

$$d_{\alpha, [0,t]}(\hat{\mathbb{X}}_{[0,t]}^N(\omega_1), \hat{\mathbb{Y}}_{[0,t]}^N(\omega_2)) = \sup_{\substack{r, s \in [0,t] \\ r < s}} \frac{d_{cc}(\hat{\mathbb{X}}_{r,s}^N(\omega_1), \hat{\mathbb{Y}}_{r,s}^N(\omega_2))}{|r - s|^\alpha}.$$

Notation 1.2.15. In the following we shall denote by $\mathcal{G}_t^{N, \alpha}(\varphi, x) := (\mathcal{G}_t^N(\varphi, x), d_{\alpha, [0,t]})$ the space of lifted paths equipped with the homogeneous α -Hölder metric for some $\frac{1}{3} < \alpha < \frac{1}{2}$. Moreover, we write $\Lambda_T^{N, \alpha}(\varphi, x) := (\Lambda_T^N(\varphi, x), d_{\Lambda^\alpha})$ for the corresponding space of lifted stopped paths equipped with

$$d_{\Lambda^\alpha}(\hat{\mathbb{X}}_{[0,t]}^N(\omega_1), \hat{\mathbb{Y}}_{[0,s]}^N(\omega_2)) := |t - s| + d_{\alpha, [0, t \vee s]}(\widehat{\mathbb{X}}_{[0, t \vee s]}^t(\omega_1), \widehat{\mathbb{Y}}_{[0, t \vee s]}^s(\omega_2)).$$

We shall assume without loss of generality that for each continuous semimartingale X , the associated probability-one set Ω_X in the definition of $\mathcal{G}_t^N(\varphi, x)$ is such that the homogeneous α -Hölder norm is finite for all lifted paths and all $0 \leq \alpha < \frac{1}{2}$.

Remark 1.2.16. Due to the fact that we are now considering different topologies on the space of lifted (stopped) paths, the continuity statements of Lemma 1.A.2, Corollary 1.2.11 and Lemma 1.A.3 need to be proved again. In fact, they do hold analogously:

- The proof that the map λ defined in Lemma 1.A.2 is again continuous in the α -Hölder topology works very similar to the one of Lemma 1.A.2, where one ultimately chooses $\delta > 0$ such that

$$\delta + d_{\alpha, [0, t+\delta]} \left(\widehat{\mathbb{Y}}_{[0, t+\delta]}^s(\omega_1), \widehat{\mathbb{Y}}_{[0, t+\delta]}^t(\omega_1) \right) < \epsilon.$$

- The continuity of Lyons' lift holds by (Friz and Victoir, 2010, Theorem 9.5, Corollary 9.11).
- The proof of Lemma 1.A.3 holds analogously in the α -Hölder topology.

In order to formulate the global universal approximation theorem we need the notion of a weighted space and certain weighted function spaces being generalizations of continuous functions.

Definition 1.2.17 (Weighted Space and Weighted Function Space). Let (X, τ_X) be a completely regular Hausdorff space and consider a function $\psi : X \rightarrow (0, \infty)$. We call the pair (X, ψ) a *weighted space* if it holds that $K_R := \psi^{-1}((0, R]) = \{x \in X \mid \psi(x) \leq R\}$ is τ_X -compact for all $R > 0$. Moreover, for a weighted space (X, ψ) , we define

$$B_\psi(X) = \left\{ f : X \rightarrow \mathbb{R} \mid \sup_{x \in X} \frac{|f(x)|}{\psi(x)} < \infty \right\}$$

and equip it with $\|f\|_{\mathcal{B}_\psi(X)} := \sup_{x \in X} \frac{|f(x)|}{\psi(x)}$. We denote by $\mathcal{B}_\psi(X)$ the closure of the space of bounded continuous functions $C_b(X)$ under $\|f\|_{\mathcal{B}_\psi(X)}$. We refer to $\mathcal{B}_\psi(X)$ as weighted function space.

The global universal approximation theorem via linear functions on the signature hinges on an application of the weighted Stone-Weierstrass theorem (Cuchiero et al., 2023b, Theorem 3.6). In contrast to (Cuchiero et al., 2023b, Theorem 5.4) we want to extend the result to be uniform in time and hence to the space of lifted stopped paths. While $\Lambda_T^2(\varphi, x)$ is *not* a weighted space, we can show that a larger space which contains $\Lambda_T^2(\varphi, x)$ is a weighted space. In fact, we need to consider the "stopped paths" of all *weakly geometric α -Hölder rough path* and not just those stemming from a continuous semimartingale. To this end we introduce the space of *weakly geometric α -Hölder rough paths* for $\frac{1}{3} < \alpha < \frac{1}{2}$ starting in $x \in \mathbb{R}^n$ and time-augmented by a continuous and strictly increasing function φ and denote it by

$$\begin{aligned} \hat{C}_t^\alpha(\varphi, x) := \left\{ \hat{\mathbf{X}}_{[0, t]} \in C^\alpha([0, T]; G^2(\mathbb{R}^{n+1})) \mid \langle e_1, \hat{\mathbf{X}}_s \rangle = \varphi(s) \text{ for all } s \in [0, t] \right. \\ \left. \text{and } \langle e_i, \hat{\mathbf{X}}_0 \rangle = x^i \text{ for all } i \in \{2, \dots, n+1\} \right\}. \end{aligned}$$

The paths which are a realization of a continuous semimartingale, i.e. $\mathcal{G}_t^{2, \alpha}(\varphi, x)$, is a subspace of $\hat{C}_t^\alpha(\varphi, x)$. In the following, we will slightly abuse notation and extend $d_{\alpha, [0, t]}$ to elements of $\hat{C}_t^\alpha(\varphi, x)$, as well as d_{Λ^α} to elements of $\bigcup_{t \in [0, T]} \hat{C}_t^\alpha(\varphi, x)$. To do so, we need a notion of how to stop a weakly geometric rough path such that this is compatible with how we defined the signature of the stopped path in the semimartingale setting.

Definition 1.2.18. For any $\frac{1}{3} < \alpha < \frac{1}{2}$, $t \in [0, t]$ and $\hat{\mathbf{X}}_{[0,t]} \in \hat{C}_t^\alpha(\varphi, x)$ we define the associated *stopped rough path* $\widehat{\mathbf{X}}_{[0,T]}^t$ in the following way; for all $s \in [0, t]$, $(\widehat{\mathbf{X}}^t)_s := \hat{\mathbf{X}}_s$ and for all $r \in [t, T]$

$$\langle e_I, (\widehat{\mathbf{X}}^t)_r \rangle := \begin{cases} \varphi(r) & \text{for } I = (1) \\ \frac{1}{2}(\varphi(r))^2 & \text{for } I = (11) \\ \langle e_I, \hat{\mathbf{X}}_t \rangle & \text{for } I = (i) \text{ or } I = (ji) \text{ where } i, j \in \{1, \dots, n+1\}, i \neq 1 \\ \varphi(r) \cdot \langle e_i, \hat{\mathbf{X}}_t \rangle - \langle e_{(1i)}, \hat{\mathbf{X}}_t \rangle & \text{for } I = (i1) \text{ where } i \in \{2, \dots, n+1\}. \end{cases}$$

Let us point out that the above definition is constructed such that any stopped weakly geometric rough path remains weakly geometric and such that it coincides with how we constructed the lifts of stopped semimartingales.

We are now ready to formulate the following lemma which we prove in Appendix 1.A.

Lemma 1.2.19. Let $\frac{1}{3} < \alpha < \frac{1}{2}$ and define the weight function

$$\psi(\hat{\mathbf{X}}_{[0,t]}) := \exp\left(\zeta \|\widehat{\mathbf{X}}_{[0,T]}^t\|_{\alpha, [0,T]}^\xi\right) \quad (1.2.2)$$

for $\hat{\mathbf{X}}_{[0,t]} \in \bigcup_{t \in [0, T]} \hat{C}_t^\alpha(\varphi, x)$, $\zeta > 0$, $\xi > 2$, $\alpha < \frac{1}{2}$ and $\|\cdot\|_{\alpha, [0,T]}$ being the norm induced by $d_{\alpha, [0,T]}$. Then $\bigcup_{t \in [0, T]} \hat{C}_t^\alpha(\varphi, x)$ equipped with $d_{\Lambda^{\alpha'}}$ for $0 \leq \alpha' < \alpha$ is a weighted space.

We have thus shown that $\left(\bigcup_{t \in [0, T]} \hat{C}_t^\alpha(\varphi, x), d_{\Lambda^{\alpha'}}\right)$ is a weighted space for the weight function given by (1.2.2) and $0 \leq \alpha' < \alpha$. We would like to emphasise that to achieve this result it is crucial to equip $\bigcup_{t \in [0, T]} \hat{C}_t^\alpha(\varphi, x)$ with a weaker topology than the α -Hölder topology. We are finally ready to formulate the global universal approximation theorem for non-anticipative path-functionals. Its proof is also given in Appendix 1.A.

Theorem 1.2.20 (Global Universal Approximation Theorem). *Let $\frac{1}{3} < \alpha < \frac{1}{2}$, $0 \leq \alpha' < \alpha$ and consider the weight function ψ given in (1.2.2). Then, for every non-anticipative path-functional $f \in \mathcal{B}_\psi\left(\left(\bigcup_{t \in [0, T]} \hat{C}_t^\alpha(\varphi, x), d_{\Lambda^{\alpha'}}\right)\right)$ (i.e. $\bigcup_{t \in [0, T]} \hat{C}_t^\alpha(\varphi, x)$ equipped with $d_{\Lambda^{\alpha'}}$) and for every $\epsilon > 0$ there exists a linear function on the signature L such that*

$$\sup_{\hat{\mathbb{X}}_{[0,t]}^2(\omega) \in \Lambda_T^{2, \alpha'}(\varphi, x)} \frac{|f(\hat{\mathbb{X}}_{[0,t]}^2(\omega)) - L(\hat{\mathbb{X}}_t)(\omega)|}{\psi(\hat{\mathbb{X}}_{[0,t]}^2(\omega))} < \epsilon.$$

Remark 1.2.21. Note that the approximation stated in Theorem 1.2.20 is indeed uniform in time, as the supremum is taken over all paths in $\Lambda_T^{2, \alpha'}(\varphi, x)$, i.e. over all lifted stopped paths. One could equivalently take the supremum over $(t, \hat{\mathbb{X}}_{[0,T]}^2(\omega)) \in [0, T] \times \mathcal{G}_T^{2, \alpha'}(\varphi, x)$. In fact, as shown in the proof of Theorem 1.2.20, we can even take the supremum over $\bigcup_{t \in [0, T]} \hat{C}_t^\alpha(\varphi, x)$.

1.2.4 The Johnson-Lindenstrauss lemma

The Johnson-Lindenstrauss Lemma [Johnson \(1984\)](#) is an important mathematical result, which is interesting for machine learning and data science, as it can be seen as a dimension reduction technique. There are many variants of the Johnson-Lindenstrauss lemma, we will here use a version stated in ([Matoušek, 2008](#), Theorem 3.1):

Theorem 1.2.22. *Let $n \in \mathbb{N}$, $\epsilon \in (0, \frac{1}{2})$, $\delta \in (0, 1)$ and set $k = C\epsilon^{-2} \log(\frac{\delta}{2})$ for a suitable constant C . Then consider a random matrix A of dimensions $k \times n$, whose entries A_{ij} are i.i.d. random variables with $\mathbb{E}[A_{ij}] = 0$, $\text{Var}[A_{ij}] = \frac{1}{k}$ and sub-Gaussian tails. Then it holds for all $x \in \mathbb{R}^n$ that*

$$\mathbb{P}[(1 - \epsilon)\|x\| \leq \|Ax\| \leq (1 + \epsilon)\|x\|] \geq 1 - \delta.$$

Remark 1.2.23. Note that since the above holds for all $x \in \mathbb{R}^n$, it immediately follows that it holds in particular for all $x - y$, where $x, y \in \mathbb{R}^n$. Hence, Theorem 1.2.22 can be understood to preserve distances up to $(1 \pm \epsilon)$.

Remark 1.2.24. In this chapter, we will use Theorem 1.2.22 to obtain a random projection of the truncated signature. We will choose the components A_{ij} to be i.i.d. with $A_{ij} \sim \mathcal{N}(0, \frac{1}{k})$.

1.2.5 The market and important portfolio specifications

Consider a finite time-horizon $T > 0$ and some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, \mathbb{P})$. We assume that the market consists of d companies, with market capitalizations given by a vector $(S_t)_{t \in [0, T]} = (S_t^1, \dots, S_t^d)_{t \in [0, T]}$ of d positive continuous semimartingales with fixed initial value $S_0 = s_0$. We assume this market fulfils the *No Unbounded Profit with Bounded Risk (NUPBR)* [Hulley and Schweizer \(2010\)](#) condition. Furthermore, we assume that each company has issued a single infinitely divisible share, hence $S_t^1, \dots, S_t^d$ correspond to the prices of these shares. We denote the relative capitalization of the companies, also called *market weights* by $\mu_t = (\mu_t^1, \dots, \mu_t^d)$ with

$$\mu_t^i = \frac{S_t^i}{S_t^1 + \dots + S_t^d}.$$

Definition 1.2.25. We call a vector $\pi_t = (\pi_t^1, \dots, \pi_t^d)$ of predictable processes fulfilling $\sum_{i=1}^d \pi_t^i \equiv 1$ and being integrable with respect to R , with $R_t^i := \int_0^t \frac{dS_u^i}{S_u^i}$, an *(R-integrable) portfolio*.

The component π_t^i denotes the proportion of wealth invested in stock i at time t , hence every portfolio is self-financing. Consider investing with a portfolio π over a time-horizon $[t_0, t_1]$ for $0 \leq t_0 < t_1 \leq T$. We denote its wealth process portfolio π with initial wealth w by $(W_t^{\pi, w})_{t \in [t_0, t_1]}$. Without loss of generality we will assume from now on that the initial investment is always one unit of currency, i.e. that $w = 1$ and therefore denote $W_t^\pi := W_t^{\pi, 1}$. The wealth process of a portfolio fulfils the stochastic differential equation

$$\frac{dW_t^\pi}{W_t^\pi} = \sum_{i=1}^d \pi_t^i \frac{dS_t^i}{S_t^i} \quad (1.2.3)$$

which is understood in Itô-sense.

A special portfolio is the *market portfolio* whose portfolio weights are given by the relative capitalizations μ_t , whence we also refer to μ_t as the market weights. The wealth of the market portfolio is given by

$$W_t^\mu = \frac{S_t^1 + \dots + S_t^d}{S_{t_0}^1 + \dots + S_{t_0}^d}$$

since we require $W_{t_0}^\mu = 1$.

We define the relative wealth process of a portfolio π to be

$$V_t^\pi := \frac{W_t^\pi}{W_t^\mu}.$$

By using Itô's-formula, it follows directly from (1.2.3) that

$$\frac{dV_t^\pi}{V_t^\pi} = \sum_{i=1}^d \pi_t^i \frac{d\mu_t^i}{\mu_t^i}. \quad (1.2.4)$$

Furthermore, by applying Itô's formula again, we obtain

$$d \ln(V_t^\pi) = \sum_{i=1}^d \frac{\pi_t^i}{\mu_t^i} d\mu_t^i - \frac{1}{2} \sum_{i,j=1}^d \frac{\pi_t^i \pi_t^j}{\mu_t^i \mu_t^j} d[\mu^i, \mu^j]_t, \quad (1.2.5)$$

where $[X, Y]_t$ denotes the co-variation process of continuous semimartingales X, Y .

Notation 1.2.26. We denote by Δ^n the unit simplex of $\mathbb{R}^n$ and by Δ_+^n its interior. We call a portfolio π long-only, if $\pi_t \in \Delta^d$ for all $t \in [0, T]$ almost surely.

We now turn to some portfolios which are of special interest, those are the *numeraire*, *log-optimal* and *growth-optimal* portfolio.

Definition 1.2.27 (Numeraire Portfolio). We call a portfolio ρ the *numeraire portfolio* if $\frac{W^\pi}{W^\rho}$ is a supermartingale for any other portfolio π .

Theorem 1.2.28. In a semimartingale market the NUPBR condition (see e.g. [Hulley and Schweizer \(2010\)](#)) holds if and only if the numeraire portfolio ρ exists and $W_T^\rho < \infty$ almost surely.

Proof. See ([Karatzas and Kardaras, 2007](#), Theorem 4.12), where in our case the predictable closed convex constraints $\mathfrak{C}$ are

$$\mathfrak{C} = \left\{ \pi \in \mathbb{R}^d \mid \sum_{i=1}^d \pi^i = 1 \right\}.$$

Note that ([Karatzas and Kardaras, 2007](#), Theorem 4.12) considers the setting with bank-account, however our constraint set $\mathfrak{C}$ prohibits us from investing into the bank account, i.e. $\pi^0 \equiv 0$ (in the notation of [Karatzas and Kardaras \(2007\)](#)). $\square$

To connect the numeraire portfolio with the log-optimal one introduced in Definition 1.2.30 below, we recall Proposition 4.19 from Karatzas and Kardaras (2007).

Proposition 1.2.29. Let $U : (0, \infty) \rightarrow \mathbb{R}$ be concave and strictly increasing. The utility maximization problem

$$\sup_{\pi} \mathbb{E} [U(W_T^{\pi})]$$

has no solution or has infinitely many solutions, if the NUPBR condition does not hold.

Definition 1.2.30. Consider the log-utility optimization problem

$$\sup_{\pi} \mathbb{E} [\log(W_T^{\pi})].$$

If it is well-posed, we call the corresponding optimal portfolio *log-optimal* portfolio.

The following lemma can also be found in Karatzas and Kardaras (2007).

Lemma 1.2.31. Under the NUPBR condition, the log-utility optimization problem admits a solution, if it is finite, in which case the numeraire portfolio is the log-optimal portfolio.

Finally, let us introduce the notion of the growth-optimal portfolio.

Definition 1.2.32. Consider a market model

$$dS_t = \text{diag}(S_t)(a_t dt + \Sigma_t dB_t),$$

where a_t is a d -dimensional vector, Σ_t a $d \times m$ -matrix for $m \geq d$ and B a m -dimensional Brownian motion. Assume a_t, B_t are predictable processes and satisfy

$$\sum_{i=1}^d \int_0^T |a_t^i| dt + \sum_{i=1}^d \sum_{j=1}^m \int_0^T |\Sigma_t^{ij}|^2 dt < \infty \quad \mathbb{P}\text{-a.s.}$$

and $\Sigma_t \Sigma_t^T$ is almost surely invertible for all $t \in [0, T]$. The log-wealth-process of a portfolio π in this market model is then given by

$$d \ln(W_t^{\pi}) = \left(\pi_t^T a_t - \frac{1}{2} \pi_t^T \Sigma_t \Sigma_t^T \pi_t \right) dt + \pi_t^T \Sigma_t dW_t.$$

We call $g_t^{\pi} = \pi_t^T a_t - \frac{1}{2} \pi_t^T \Sigma_t \Sigma_t^T \pi_t$ the portfolios *growth-rate* and the portfolio with maximal growth-rate, if it exists, the *growth-optimal* portfolio.

Lemma 1.2.33. Consider a market model given in Definition 1.2.32 and let us define the (instantaneous) market price of risk

$$\theta_t := \Sigma_t^T (\Sigma_t \Sigma_t^T)^{-1} a_t.$$

The growth-optimal portfolio exists if and only if $\int_0^T \|\theta_t\|^2 dt < \infty$ almost surely. Moreover, the log-optimal portfolio exists if and only if $\mathbb{E} \left[\int_0^T \|\theta_t\|^2 dt \right] < \infty$. Moreover, if the growth-optimal portfolio $\pi^{(g)}$ exists, it is given by

$$\pi_t^{(g)} = (\Sigma_t \Sigma_t^T)^{-1} (a_t - \kappa \mathbf{1}) \quad \text{for } \kappa = \frac{\sum_i ((\Sigma_t \Sigma_t^T)^{-1} a_t)_i - 1}{\sum_{i,j} (\Sigma_t \Sigma_t^T)^{-1}_{i,j}}$$

where $\mathbf{1} = (1, \dots, 1)^T$.

Proof. Regarding the existence of the growth-optimal and log-optimal portfolios, see [Hulley and Schweizer \(2010\)](#), also for the general case of continuous semimartingale markets. The specific form of the growth-optimal portfolio is due to [Filipovic and Platen \(2009\)](#). $\square$

Lemma 1.2.34. Consider the setting of a market as in Definition 1.2.32. Then the growth-optimal portfolio exists if and only if the numeraire portfolio exists in which case they are the same². Moreover, if the log-utility maximization problem is finite, then they also coincide with the log-optimal portfolio.

Proof. See [Filipovic and Platen \(2009\)](#) and [Hulley and Schweizer \(2010\)](#). The last assertion then follows from Lemma 1.2.31 $\square$

In SPT, there is another class of portfolios of special interest, those are *functionally generated portfolios*, see [Fernholz \(2002\)](#).

Definition 1.2.35 (Functionally Generated Portfolios). Let U be a neighbourhood of Δ^d and consider a C^2 -function $G : U \rightarrow \mathbb{R}_+$ such that $x_i D_i \log G(x)$ is bounded on Δ^d . Then G defines the functionally generated portfolio via

$$\pi_t^i = \mu_t^i \left(D_i \log G(\mu_t) + 1 - \sum_{j=1}^d \mu_t^j D_j \log G(\mu_t) \right).$$

The function G is called the *portfolio generating function* and if it is concave, π is a long-only portfolio.

Before we conclude this section, we introduce the *ranked market-weights* which are of particular interest in SPT, due to the remarkable stability of the capital distribution curves, see for example [Fernholz \(2002\)](#); [Karatzas and Ruf \(2017\)](#).

Definition 1.2.36 (Ranked Capitalization and Market Weight Processes). For an $\mathbb{R}_+^n$ -valued continuous semimartingale X we denote its *ranked process* by $\mathbf{X} = (X^{(1)}, \dots, X^{(n)})$ which is defined as

$$\max_{i \in \{1, \dots, n\}} X_t^i = X_t^{(1)} \geq X_t^{(2)} \geq \dots \geq X_t^{(n-1)} \geq X_t^{(n)} = \min_{i \in \{1, \dots, n\}} X_t^i,$$

where we break ties by allocating a lower rank to smaller labels. To be precise, if $X_t^i = X_t^j$ with $i \geq j$, we set $X_t^{(r_i)} = X_t^i$ and $X_t^{(r_j)} = X_t^j$ where $r_i \geq r_j$.

Accordingly, we denote by μ the *ranked market weights* and by $\mathbf{S}$ the *ranked capitalization process*.

Corollary 1.2.37. Given an $\mathbb{R}^n$ -valued continuous semimartingale X , its ranked process is again an $\mathbb{R}^n$ -valued continuous semimartingale.

Proof. See ([Banner and Ghomrasni, 2008](#), Theorem 2.2). $\square$

²Assuming that their generated wealth is almost surely finite this holds if and only if NUPBR is satisfied (see Theorem 2.25).

1.3 (Linear) path-functional portfolios

Before we present our application of signature methods in the context of SPT and portfolio optimization, we generalize the concept of functionally generated portfolios by introducing so-called *path-functional portfolios*.

Definition 1.3.1 (Path-Functional Portfolios). We introduce *path-functional portfolios* which are generated by a family of non-anticipating path functionals $\{f^i\}_{1 \leq i \leq d}$ depending on time and the path of X and which are constructed via an auxiliary portfolio τ , where τ is required to be uniformly bounded. Moreover, we assume that the non-anticipating path functionals $\{f^i\}_{1 \leq i \leq d}$ fulfill the necessary integrability conditions such that the following processes π are portfolios in the sense of Definition 1.2.25. We consider two types of such portfolios:

$$(I) \quad \pi_t^i(\tau, X) = \tau_t^i \left(f^i \left(t, X_{[0,t]} \right) + 1 - \sum_{j=1}^d \tau_t^j f^j \left(t, X_{[0,t]} \right) \right)$$

$$(II) \quad \pi_t^i(\tau, X) = f^i \left(t, X_{[0,t]} \right) + \tau_t^i \left(1 - \sum_{j=1}^d f^j \left(t, X_{[0,t]} \right) \right)$$

where we use the notation of the subscript $[0, t]$ to make the dependence of on the entire path explicit. Moreover, we refer to the functionals $\{f^i\}_{1 \leq i \leq d}$ as the *portfolio controlling functions*.

Remark 1.3.2 (Relation to (Classical) Stochastic Portfolio Theory). In the above definition, we consider a general continuous semimartingale and a general auxiliary portfolio τ in the construction of the path-functional portfolios. In the spirit of SPT, one would choose $\tau = \mu$ the market portfolio and $X = \mu$, the process of market weights. One can then recover the classical functionally generated portfolios from the path-functional portfolios of type *I* by setting $X = \mu$ and choosing $f^i(t, \mu_{[0,t]}) = f^i(\mu_t) = D_i \log(G(\mu_t))$. We would like to highlight that beyond this relation to classical SPT, constructing the path-functional portfolios in this form (i.e. via additive normalization) is crucial for our results in Sections 1.4 and 1.5, in particular in view of getting quadratic optimization problems in the current multiplicative setting where the strategies are in terms fractions of wealth.

Remark 1.3.3 (Role of the process X). In relation to classical SPT, we will often choose $X = \mu$, i.e. construct our portfolios via information on the relative market weights. However, we would like to point out that other information may be of interest such as the absolute capitalization, market-to-book ratios Kim (2023) or earnings data, to just name a few. Adding information beyond the market weights to the portfolio construction, has also appeared e.g. in Ruf and Xie (2019) in the context of SPT. However, we would like to point out that in Ruf and Xie (2019) the additional information has to be a continuous process of finite variation, while we allow it to be a general continuous semimartingale.

Remark 1.3.4 (Path-Functional Portfolios of Type *II*). A simple example of a path-functional portfolio of type *II* would be any portfolio generated by a deep (recurrent) neural network, see for example Zhang et al. (2020). Denote by $\mathcal{NN}^i(t, X_{[0,t]})$ the i -th output of a deep (recurrent) neural network. If

- the last layer of the neural network is for example a softmax function and the portfolio weights are defined by $\pi_t^{(\mathcal{NN}),i} = \mathcal{NN}^i(t, X_{[0,t]})$, or,
- the portfolio weights are constructed as $\pi_t^{(\mathcal{NN}),i} = \frac{\mathcal{NN}^i(t, X_{[0,t]})}{\sum_{j=1}^d \mathcal{NN}^j(t, X_{[0,t]})}$,

then, the portfolios $\pi^{(\mathcal{NN})}$ are in both cases path-functional portfolios of type *II* but not of type *I*. This is by identifying $f^{(\mathcal{NN}),i}(t, X_{[0,t]}) = \mathcal{NN}^i(t, X_{[0,t]})$ or $f^{(\mathcal{NN}),i}(t, X_{[0,t]}) = \frac{\mathcal{NN}^i(t, X_{[0,t]})}{\sum_{j=1}^d \mathcal{NN}^j(t, X_{[0,t]})}$ respectively. Note that the choice of τ is irrelevant because

$$\sum_{j=1}^d f^{(\mathcal{NN}),j}(t, X_{[0,t]}) = 1$$

by definition.

Remark 1.3.5 (Conversion between Path-Functional Portfolios of Type I&II). Note that for a fixed auxiliary portfolio τ with weight-processes which are continuous semimartingales and for a continuous semimartingale $(X_t)_{t \in [0,T]}$ any path-functional portfolio of type *I* with portfolio controlling functions $f^{(I),i} : X_{[0,t]} \mapsto f^{(I),i}(X_{[0,t]})$ is a path-functional portfolio of type *II* with portfolio controlling function $f^{(II),i} : (\tau, X)_{[0,t]} \mapsto \tau_t^i f^{(I),i}(X_{[0,t]})$. This also holds vice versa, if the components of τ are all non-zero. Indeed, any path-functional portfolio of type *II* with portfolio controlling function $f^{(II),i} : X_{[0,t]} \mapsto f^{(II),i}(X_{[0,t]})$ is a path-functional portfolio of type *I* with portfolio controlling function $f^{(I),i} : (\tau, X)_{[0,t]} \mapsto \frac{f^{(II),i}(X_{[0,t]})}{\tau_t^i}$. However, note that this conversion always involves adding the process of the auxiliary portfolio τ to the input. Not only does this enlarge the dimensions of the inputs, but also is τ then required to be a continuous semimartingale. Hence, this conversion is not always possible.

We now introduce a special class of path-functional portfolios which are particularly useful for optimizing path functional portfolios, as we will make more explicit in Section 1.5. This class is the one of *linear path-functional portfolios*.

Definition 1.3.6 (Linear Path-Functional Portfolios). For a path-functional portfolio of type *I* or *II*, we call it a *linear* path-functional portfolio of the corresponding type, if the portfolio controlling functions are of the form

$$f^i(t, X_{[0,t]}) = \sum_{\nu \in \mathcal{V}} l_\nu^i \phi^\nu(t, X_{[0,t]})$$

where $\{\phi^\nu\}_{\nu \in \mathcal{V}}$ is a collection of feature maps which are themselves non-anticipating path functionals, $\mathcal{V}$ is a finite set of features and $l_\nu^i \in \mathbb{R}$ for each i, ν are constant (optimization) parameters.

Remark 1.3.7 (Significance of the Auxiliary Portfolio τ). Considering linear path-functional portfolios gives us a first idea on how to do portfolio optimization in that context. Namely, optimizing the parameters $\{l_\nu^i\}_{1 \leq i \leq d, \nu \in \mathcal{V}}$ for a given collection of feature maps. Note, that the auxiliary portfolio τ can always be attained in this way by

1.3 (Linear) path-functional portfolios

- setting $l_\nu^i = l_\nu^j$ for all $i, j \in \{1, \dots, d\}$, $\nu \in \mathcal{V}$ for portfolios of type I
- setting $l_\nu^i = 0$ for all $i \in \{1, \dots, d\}$, $\nu \in \mathcal{V}$ for portfolios of type II .

Hence, if one wants to learn a linear path-functional portfolio which outperforms a given benchmark portfolio $\bar{\pi}$, one should use the benchmark portfolio as the auxiliary portfolio, i.e. set $\tau = \bar{\pi}$, since then the benchmark portfolio is included in the family of portfolios one optimizes over. In the context of stochastic portfolio theory, we therefore often choose $\tau = \mu$, because we aim to outperform the market portfolio.

Among all linear path-functional portfolios, we are in particular interested in those, whose feature maps are (randomized) elements of the signature, resulting in *(JL- or randomized-) signature portfolios*. We make this more precise in the following definition:

Definition 1.3.8 (Signature Portfolios, JL- and Randomized-Signature Portfolios). For a given auxiliary portfolio τ and any $\mathbb{R}^n$ -valued continuous semimartingale X , we call linear path-functional portfolios *(JL- or randomized-)signature portfolios* if the feature maps are elements of the (JL- or randomized-)signature of the time-augmented semimartingale $\hat{X}$. More precisely, let $\hat{X}_t = (\varphi(t), X_t)$ be the time-augmented process of X , where φ is a strictly increasing function. Recall that $\hat{\mathbb{X}}^N$ denotes the signature of $\hat{X}$ truncated at level N . We then consider the following portfolios:

- *Signature Portfolios*: A *signature portfolio of degree N* is a linear path-functional portfolio with $\mathcal{V} = \{I \mid I = (i_1, \dots, i_m) \in \{1, \dots, n\}^m \text{ for } 0 \leq m \leq N\}$ (the set of multiindices up to length N) and with feature maps

$$\phi^I(t, X_{[0,t]}) = \langle e_I, \hat{\mathbb{X}}_t^N \rangle = \langle e_I, \hat{\mathbb{X}}_t \rangle.$$

In other words, signature portfolios are path-functional portfolios, where the portfolio controlling functions are linear functions on the signature.

- *JL-Signature Portfolios*: A *JL-signature portfolio of dimension (P, N)* is a linear path-functional portfolio with $\mathcal{V} = \{1, \dots, P\}$, where P is the dimension of the projection, and for each $p \in \{1, \dots, P\}$ with feature maps

$$\phi^p(t, X_{[0,t]}) = \langle A^p, \hat{\mathbb{X}}_t^N \rangle,$$

where A^p is the p -th column of the Johnson-Lindenstrauss projection introduced in Theorem 1.2.22. Hence, JL-signature portfolios are path-functional portfolios, where the portfolio controlling functions are linear functions on the Johnson-Lindenstrauss projected signature.

- *Randomized-Signature Portfolios*: Let $\mathcal{S}$ denote the solution to

$$d\mathcal{S}_t = \sum_{i=1}^{n+1} \sigma \left(b^{(i)} + A^{(i)} \mathcal{S}_t \right) \circ d\hat{X}_t^i \quad \mathcal{S}_0 = (1, 0, 0, \dots) \in \mathbb{R}^p,$$

where σ is an activation function, randomly chosen $b^{(i)} \in \mathbb{R}^p$, $A^{(i)} \in \mathbb{R}^{p \times p}$ for all $i \in \{1, \dots, n+1\}$ and $\hat{X}$ is the time-augmented process of an $\mathbb{R}^n$ -valued continuous semimartingale X . We call $\mathcal{S}(\hat{X})$ the randomized signature of dimension p of $\hat{X}$, see [Akyildirim et al. \(2022\)](#); [Compagnoni et al. \(2022\)](#); [Cuchiero et al. \(2021a,b\)](#). For a connection with neural signature kernels and controlled ResNets we refer to [Muca Cirone et al. \(2023\)](#). A *randomized-signature portfolio of dimension P* is a linear path-functional portfolio with $\mathcal{V} = \{1, \dots, P\}$, where P is the dimension of the randomized signature, and with feature maps

$$\phi^P(t, X_{[0,t]}) = \langle e_P, \mathcal{S}_t \rangle.$$

Hence, randomized-signature portfolios are path-functional portfolios, where the portfolio controlling functions are linear combinations of the elements of the randomized signature.

Remark 1.3.9. We have highlighted signature, JL- and randomized-signature portfolios as linear path-functional portfolios of special interest in this chapter. However, we would like to mention some other examples of linear path-functional portfolios whose portfolio controlling functions are

- linear functions on increments of the signature. That is, for a fixed time span Θ , we observe a rolling window of length $\Theta < T$ and forget the information before that, i.e.

$$\phi^I(t, X_{[0,t]}) = \langle e_I, \hat{\mathbb{X}}_{0 \vee (t-\Theta), t} \rangle.$$

We will refer to such portfolios as signature portfolios *with rolling windows*. We shall not put particular emphasis on these portfolios in this chapter, since our approximation results in Section 1.4 are tailored to signature portfolios *without* rolling windows and since portfolios with rolling windows are numerically less tractable, as we outline in Section 1.7.

- random neural networks, which are neural networks where the parameters of the hidden layers are not trained but randomly sampled and only the linear read-out layer is trained, see [Herrera et al. \(2024\)](#). Training the read-out layer, exactly amounts to training the parameters $\{l_\nu^i\}_{1 \leq i \leq d, \nu \in \mathcal{V}}$ of path-functional portfolios. Such neural networks may have an infinite-dimensional input space, corresponding in our context to path spaces, as e.g. considered in [Cuchiero et al. \(2023b\)](#) or [Benth et al. \(2023\)](#).
- given by reservoir computers with linear read-out layers, see for instance [Grigoryeva and Ortega \(2018b\)](#). A special class of which are Echo State Networks [Jaeger \(2001\)](#); [Grigoryeva and Ortega \(2018a\)](#); [Gonon and Ortega \(2021\)](#) or Quantum Reservoir Computers [Negoro et al. \(2018\)](#); [Fujii and Nakajima \(2017\)](#). Note, that there is an interesting connection between the Johnson-Lindenstrauss projection of signatures, randomized signature and reservoir computing, which is worked out in [Cuchiero et al. \(2021a\)](#).

- given by a constant, which leads to the case of constant portfolio weights when using a constant auxiliary portfolio τ . In this case we recover the Markowitz portfolio optimization as a very special case in Subsection 1.5.1. Moreover, since the first element of the signature is constant, this class of portfolios is also included in the class of signature portfolios.

1.4 Approximation properties of signature portfolios

This section is dedicated to prove universal approximation properties of signature portfolios. Let us recall that for a compact subset $K \subset \mathcal{G}_T^2(\varphi, x)$ the set

$$\mathcal{K}_\Lambda = \lambda([0, T] \times K) = \left\{ \hat{\mathbb{X}}_{[0,t]}^2(\omega) \mid t \in [0, T] \text{ and } \hat{\mathbb{X}}_{[0,T]}^2(\omega) \in K \right\} \quad (1.4.1)$$

is a compact subset of $\Lambda_T^2(\varphi, x)$ by the continuity of λ , see Lemma 1.A.2. Furthermore, recall that in Definition 1.3.1 we have introduced portfolio controlling functions $\{f^i\}_{1 \leq i \leq d}$ as non-anticipative path-functionals taking as input the paths of some semimartingale X , i.e. the $\{f^i\}_{1 \leq i \leq d}$ were of form $f^i(t, X_{[0,t]})$. In the following, we will require a more specific form, namely portfolio controlling functionals depending on the lifted path, that is $f^i(\hat{\mathbb{X}}_{[0,t]}^2)$, where the dependence on the lifted path is important for the continuity statements. However, this is not a contradiction to the general definition, in particular since $\hat{\mathbb{X}}^2$ is a continuous semimartingale itself. Note that the functions $\{f^i\}_{1 \leq i \leq d}$ can of course only depend on the first level $X_{[0,t]}$ as well.

Theorem 1.4.1. *Consider an arbitrary but fixed auxiliary portfolio τ and a path-functional portfolio $\pi(\tau, \cdot)$ of type I (type II) with portfolio controlling functions $f^i \in C(\mathcal{K}_\Lambda; \mathbb{R})$ for all $i \in \{1, \dots, d\}$. Then for every $\epsilon > 0$ there exists a signature portfolio $\pi^*(\tau, \cdot)$ of type I (type II) such that it holds for all $i \in \{1, \dots, d\}$*

$$\sup_{(t, \hat{\mathbb{X}}_{[0,T]}^2(\omega)) \in [0, T] \times K} |\pi^i(\tau, \hat{X}_{[0,t]})(\omega) - \pi^{*,i}(\tau, \hat{X}_{[0,t]})(\omega)| < \epsilon.$$

Moreover, take $\frac{1}{3} < \alpha < \frac{1}{2}$, $0 \leq \alpha' < \alpha$, an arbitrary but fixed auxiliary portfolio τ and a path-functional portfolio $\pi(\tau, \cdot)$ of type I (type II) with portfolio controlling functions such that for all $i \in \{1, \dots, d\}$ $f^i \in \mathcal{B}_\psi \left(\left(\bigcup_{t \in [0, T]} \hat{C}_t^\alpha(\varphi, x), d_{\Lambda^{\alpha'}} \right) \right)$. Then for every continuous semimartingale X starting in x and fulfilling $\mathbb{E} \left[\exp(\zeta \|\hat{\mathbb{X}}_{[0,T]}^2\|_{\alpha, [0, T]}^\xi) \right] < \infty$ with $\zeta > 0$ and $\xi > 2$, it holds that for every $\epsilon, \delta > 0$ there exists a signature portfolio $\pi^*(\tau, \cdot)$ of type I (type II) such that it holds for all $i \in \{1, \dots, d\}$

$$\mathbb{P} \left[\sup_{t \in [0, T]} \left| \pi^i(\tau, \hat{X}_{[0,t]}) - \pi^{*,i}(\tau, \hat{X}_{[0,t]}) \right| > \epsilon \right] < \delta.$$

Proof. We present the proof for the case of portfolios of type I, however the proof is analogous for the case of type II.

Let us start by showing the first statement of the theorem. By the universal approximation result (Theorem 1.2.12), we know that for each $\{f^i\}_{1 \leq i \leq d}$ there exists a linear function on the signature of $\hat{X}$ that approximates f^i arbitrarily well, on compact sets. Those linear functions can be chosen as the portfolio controlling functions f^i , of a signature portfolio π^* . Hence, there exists for each $\epsilon' > 0$ a signature portfolio π^* of type I such that for all $i \in \{1, \dots, d\}$ it holds that

$$\sup_{(t, \hat{X}_{[0,T]}^2(\omega)) \in [0,T] \times K} |f^i(\hat{X}_{[0,t]}^2(\omega)) - f^{*,i}(\hat{X}_{[0,t]}^2(\omega))| < \epsilon' \quad (1.4.2)$$

for almost all $\omega \in \Omega$.

Now, we show that this translates to the portfolio weights. The statement is trivial for all $\omega \in \Omega$ where $\hat{X}_{[0,T]}^2 \notin K$.

On $\Omega^{(K)} := \{\omega \mid \hat{X}_{[0,T]}^2(\omega) \in K\}$, it is implied by (1.4.2) that

$$\begin{aligned} \sup_{t \in [0,T]} |\pi_t^i - \pi_t^{*,i}| &= \sup_{t \in [0,T]} \left| \tau_t^i \left(f^i(\hat{X}_{[0,t]}^2) + 1 - \sum_{j=1}^d \tau_t^j f^j(\hat{X}_{[0,t]}^2) \right) \right. \\ &\quad \left. - \tau_t^i \left(f^{*,i}(\hat{X}_{[0,t]}^2) + 1 - \sum_{j=1}^d \tau_t^j f^{*,j}(\hat{X}_{[0,t]}^2) \right) \right| \\ &= \sup_{t \in [0,T]} \left| \tau_t^i \left(f^i(\hat{X}_{[0,t]}^2) - f^{*,i}(\hat{X}_{[0,t]}^2) \right) - \sum_{j=1}^d \tau_t^j \left(f^j(\hat{X}_{[0,t]}^2) - f^{*,j}(\hat{X}_{[0,t]}^2) \right) \right| \\ &\leq \sup_{t \in [0,T]} \left| \tau_t^i \left(f^i(\hat{X}_{[0,t]}^2) - f^{*,i}(\hat{X}_{[0,t]}^2) \right) \right| + \sum_{j=1}^d \sup_{t \in [0,T]} \left| \tau_t^j \left(f^j(\hat{X}_{[0,t]}^2) - f^{*,j}(\hat{X}_{[0,t]}^2) \right) \right| \\ &\leq M \sup_{t \in [0,T]} \left| f^i(\hat{X}_{[0,t]}^2) - f^{*,i}(\hat{X}_{[0,t]}^2) \right| + M \sum_{j=1}^d \sup_{t \in [0,T]} \left| f^j(\hat{X}_{[0,t]}^2) - f^{*,j}(\hat{X}_{[0,t]}^2) \right| \\ &< M(d+1)\epsilon' \end{aligned}$$

holds almost surely, where M is the bound of τ . The result follows by choosing $\epsilon' = \frac{\epsilon}{M(d+1)}$.

The proof of the second approximation statement follows by similar arguments. Recall Theorem 1.2.20 and use the same reasoning as above to translate the approximation result from the portfolio controlling functions to the portfolio weights itself, i.e. we obtain that there exists for each $\epsilon' > 0$ a signature portfolio π^* of type I such that for all $i \in \{1, \dots, d\}$ it holds that

$$\sup_{t \in [0,T]} \frac{|\pi^i(\tau, \hat{X}_{[0,t]}) - \pi^{*,i}(\tau, \hat{X}_{[0,t]})|}{\exp(\zeta \|\hat{X}_{[0,T]}^2\|_{\alpha, [0,T]}^\xi)} < \epsilon' \quad (1.4.3)$$

almost surely. Therefore,

$$\begin{aligned} \mathbb{P} \left[\sup_{t \in [0, T]} \left| \pi^i(\tau, \hat{X}_{[0, t]}) - \pi^{*,i}(\tau, \hat{X}_{[0, t]}) \right| > \epsilon \right] &\leq \epsilon^{-1} \mathbb{E} \left[\sup_{t \in [0, T]} \left| \pi^i(\tau, \hat{X}_{[0, t]}) - \pi^{*,i}(\tau, \hat{X}_{[0, t]}) \right| \right] \\ &< \frac{\epsilon'}{\epsilon} \mathbb{E} \left[\sup_{t \in [0, T]} \exp(\zeta \|\hat{\mathbb{X}}_{[0, T]}^2\|_{\alpha, [0, T]}^\xi) \right], \end{aligned}$$

where the first inequality is due to Markov's inequality and the second one follows from (1.4.3).

Note that by the properties of $\|\cdot\|_{cc}$ and $\|\cdot\|_{\alpha, [0, T]}$ it holds that $\sup_{t \in [0, T]} \|\hat{\mathbb{X}}_{[0, T]}^2\|_{\alpha, [0, T]} = \|\hat{\mathbb{X}}_{[0, T]}^2\|_{\alpha, [0, T]}$. Hence the statement is obtained by setting $\delta = \frac{\epsilon'}{\epsilon} \mathbb{E} \left[\exp(\zeta \|\hat{\mathbb{X}}_{[0, T]}^2\|_{\alpha, [0, T]}^\xi) \right]$ and choosing ϵ' accordingly. $\square$

Remark 1.4.2. Note, that by Corollary 1.2.13, Theorem 1.4.1 also ensures that signature portfolios approximate classical functionally generated portfolios arbitrarily well, on compact sets. In particular, this includes the functionally generated portfolios as introduced by Fernholz Fernholz (2002) and those considered for functional portfolio optimization in Campbell and Wong (2022).

Remark 1.4.3. The requirements of Theorem 1.4.1 on f^i is a simple continuity condition and examples for such continuous path-functionals are provided in Example 1.2.13.

Having obtained a universal approximation theorem of path-functional portfolios, we will now apply it to obtain universal approximation results of the growth-optimal portfolio by signature portfolios in several markets.

1.4.1 Universal approximation of the growth-optimal portfolio

In this subsection, we present a large class of markets where the growth-optimal portfolio can be regarded as a path-functional portfolio and show that it can therefore be approximated by a signature portfolio.

The following statements are to be understood in the market setting outlined in Definition 1.2.32 and we therefore assume that the corresponding necessary conditions are satisfied, in particular those ensuring the existence of the growth-optimal portfolio (Lemma 1.2.33). Moreover, recall also the definition of the compact set $\mathcal{K}_\Lambda$ given in (1.4.1).

Theorem 1.4.4. Consider the following class of markets with d stocks:

$$dS_t = \text{diag}(S_t)(a(\hat{\mathbb{X}}_{[0, t]}^2)dt + \Sigma(\hat{\mathbb{X}}_{[0, t]}^2)dB_t), \quad (1.4.4)$$

where the components of a , Σ are non-anticipative path-functionals in $C(\mathcal{K}_\Lambda; \mathbb{R})$ and X some $\mathbb{R}^n$ -valued semimartingale with $n \in \mathbb{N}$. Then for any auxiliary portfolio τ there exists a signature portfolio of type II which approximates the weights of the growth-optimal portfolio arbitrarily well on $\mathcal{K}_\Lambda$. Moreover, if the components of the auxiliary portfolio τ are

deterministic, then there also exists a signature portfolio of type *I* which approximates the weights of the growth-optimal portfolio almost surely arbitrarily well on $\mathcal{K}_\Lambda$.

Alternatively, let $\frac{1}{3} < \alpha < \frac{1}{2}$, $0 \leq \alpha' < \alpha$ and let a, Σ be such that the components of $\pi^{(g)}$ are non-anticipative path-functionals in $\mathcal{B}_\psi \left(\left(\bigcup_{t \in [0, T]} \hat{C}_t^\alpha(\varphi, x), d_{\Lambda^{\alpha'}} \right) \right)$.³ Then for every continuous semimartingale X starting in x and fulfilling $\mathbb{E} \left[\exp(\zeta \|\hat{\mathcal{X}}_{[0, T]}^2\|_{\alpha, [0, T]}^\xi) \right] < \infty$ with $\zeta > 0$ and $\xi > 2$, it holds that for every $\epsilon, \delta > 0$ there exists a signature portfolio $\pi^*(\tau, \cdot)$ for type *II* for any auxiliary portfolio τ (or of type *I* for a deterministic portfolio τ) such that it holds for all $i \in \{1, \dots, d\}$

$$\mathbb{P} \left[\sup_{t \in [0, T]} \left| \pi^{(g), i}(\tau, \hat{X}_{[0, t]}) - \pi^{*, i}(\tau, \hat{X}_{[0, t]}) \right| > \epsilon \right] < \delta,$$

where $\pi^{(g)}$ is the growth-optimal portfolio of the respective market.

Remark 1.4.5. Note that the semimartingale X in (1.4.4) could be (a functions of) S , so that we are actually dealing with path-dependent SDEs. In the particular case where a and Σ are entire functions of the signature as introduced in Cuchiero et al. (2023c) we are then in the tractable setup of *signature SDEs*.

Proof of Theorem 1.4.4. Recalling Lemma 1.2.33, it is straightforward that for any auxiliary portfolio τ , the growth-optimal portfolio $\pi^{(g)}$ is a path-functional portfolio of type *II* with portfolio controlling functions

$$f^{(g, II), i} = \pi^{(g), i}.$$

Likewise, $\pi^{(g)}$ is a path-functional portfolio of type *I* for any deterministic auxiliary portfolio and portfolio controlling functions

$$f^{(g, I), i} = \tau^i \cdot \pi^{(g), i}.$$

By Lemma 1.B.1 and the form of $\pi^{(g), i}$ Theorem 1.4.1 is applicable for $\pi^{(g), i}$ (viewed as a path-functional portfolio of type *I* resp. type *II*). Hence, the statements follow. $\square$

Remark 1.4.6. If $X = \mu$ or $X = S$, the growth optimal is also a path-functional portfolio of type *I* with auxiliary portfolio $\tau = \mu$.

Remark 1.4.7 (Examples of Markovian Markets). We here give two prominent examples of markets, to which Theorem 1.4.4 can be applied.

- Black-Scholes Market:

$$dS_t = \text{diag}(S_t)(adt + \Sigma dB_t),$$

where a and Σ are constant.

³Note that a sufficient condition for this to hold is if the components of a are non-anticipative path-functionals in $\mathcal{B}_\psi \left(\left(\bigcup_{t \in [0, T]} \hat{C}_t^\alpha(\varphi, x), d_{\Lambda^{\alpha'}} \right) \right)$ and Σ is constant.

- Volatility Stabilized Markets:

$$\frac{dS_t^i}{S_t^i} = \frac{1+\alpha}{2} \frac{1}{\mu_t^i} dt + \sqrt{\frac{1}{\mu_t^i}} dB_t^i \text{ for all } 1 \leq i \leq d$$

and $\alpha \geq 0$. Volatility stabilized markets are of great interest in SPT because they reflect the observation in real markets that smaller stocks tend to have greater volatility than larger stocks, see for example [Fernholz and Karatzas \(2010\)](#). See also [Cuchiero \(2019\)](#); [Fernholz and Karatzas \(2005\)](#) for further properties of volatility stabilized markets.

1.4.2 A class of non-Markovian markets where the growth-optimal portfolio is a signature portfolio

Let us now turn to a class of (possibly) non-Markovian markets, where the growth-optimal portfolio can not only be approximated by a signature portfolio but *is* a signature portfolio.

Theorem 1.4.8. *Let X be a continuous semimartingale. For a market of d stocks, consider the class of Sig-market models*

$$dS_t = \text{diag}(S_t)(a(\hat{X}_{[0,t]})dt + \Sigma dB_t)$$

where $a(\hat{X}_{[0,t]})_i = \sum_{0 \leq |I| \leq N} \alpha_I^{(i)} \langle e_I, \hat{X}_t \rangle$ for $N \geq 0$, $\alpha_I^{(i)} \in \mathbb{R}$, Σ is a constant $d \times m$ matrix and B_t is an m -dimensional Brownian motion for $m \geq d$. Then, the growth-optimal portfolio is for any auxiliary portfolio a signature portfolio of type II and a signature portfolio of type I for any constant and deterministic auxiliary portfolio τ .

Proof. The integrability and measurability assumptions required in Definition 1.2.32 and Lemma 1.2.33 follow from the integrability and measurability of elements of the signature. Applying Lemma 1.2.33, it follows that the growth-optimal weights are linear functions on the signature of $\hat{X}$. This is for any auxiliary portfolio a signature portfolio of type II, or a signature portfolio of type I for any constant and deterministic auxiliary portfolio τ . $\square$

Remark 1.4.9 (Existence of Solutions to Sig-Markets). As addressed in Remark 1.4.5, the existence of solutions is non-trivial for Sig-markets if $X = h(S)$, i.e. if the semimartingale of which we construct the signature is a function of the price process itself. In particular, if h is a real analytic function, then we deal with signature-SDEs in the spirit of [Cuchiero et al. \(2023c\)](#). The study of existence of solutions to such equations goes beyond the subject of this chapter, however, we give a simple example below, where the existence is guaranteed. To this end, let us denote by

$$\mathcal{I}_1 := \{I | I = (i_1, \dots, i_m) \text{ for } 0 \leq n \leq N \text{ and } i_2 = i_3 = \dots = i_n = 1\}$$

Let $X = \log S_t$. If $\alpha_I^{(i)} = 0$ for all $I \notin \mathcal{I}_1$, then

$$dS_t = \text{diag}(S_t) \cdot \left(a(\widehat{(\log S)}_{[0,t]}) dt + \Sigma dB_t \right) \quad (1.4.5)$$

admits a unique strong solution. To see this, note that (1.4.5) can be seen a part of a (larger) system of linear equations. Setting $Y = \log S$, $\hat{Y} = (t, Y)$ and denoting by $\hat{Y}^N|_{\mathcal{I}_1} = (\langle e_I, \hat{Y} \rangle)_{I \in \mathcal{I}_1}$ the restriction to elements of the signature corresponding to multiindices in $\mathcal{I}_1$. Let us denote the vectorization of the truncated signature denoted by $\text{vec}(\cdot)$ and fix a labelling function $\mathcal{L} : \mathcal{I}_1 \rightarrow \{1, \dots, |\mathcal{I}_1|\}$ such that $\text{vec}(\hat{Y}^N|_{\mathcal{I}_1})_{\mathcal{L}(I)} = \langle e_I, \hat{Y}^N|_{\mathcal{I}_1} \rangle$. Moreover, for a multiindex $I = (i_1, \dots, i_{|I|})$, we denote by $I' := (i_1, \dots, i_{|I|-1})$ corresponding multiindex shortend by the last letter. Then said linear system of equations is

$$d\left(\text{vec}(\hat{Y}^N|_{\mathcal{I}_1})_t\right) = \left(b + A \cdot \text{vec}(\hat{Y}^N|_{\mathcal{I}_1})_t\right) dt + \tilde{\Sigma} dB_t.$$

where $\tilde{\Sigma}_{j,r} = \Sigma_{s,r} \delta_{j,\mathcal{L}(s)}$ for $1 \leq j \leq |\mathcal{I}_1|$, $1 \leq r \leq m$ and for $k, l \in \{1, \dots, |\mathcal{I}_1|\}$

$$b_k = \begin{cases} -\frac{1}{2} \Sigma^\top \Sigma + a_\emptyset^{(i)} & \text{if } k = \mathcal{L}(i), i \neq 1 \\ 1 & \text{if } k = \mathcal{L}(1) \\ 0 & \text{else} \end{cases}$$

$$A_{k,l} = \begin{cases} a_{\mathcal{L}(J)}^{(i)} & \text{if } k = \mathcal{L}(i), l = \mathcal{L}(J), i \neq 1 \\ 1 & \text{if } |I| \geq 2, k = \mathcal{L}(I), l = \mathcal{L}(I') \\ 0 & \text{else.} \end{cases}$$

Remark 1.4.10 (NUPBR in Sig-Markets). The NUPBR condition in the above Sig-markets holds. This follows from the existence of the growth-optimal portfolio by Lemma 1.2.33, Lemma 1.2.34 and Theorem 1.2.28.

1.5 Optimization tasks for linear path-functional portfolios

We now want to study some tasks for portfolio optimization and their form for linear path-functional portfolios. To formulate them in the most general way, let us introduce what we call a *universe* of stocks.

Definition 1.5.1 (Universe). Consider a market of d stocks with capitalization process $S = (S^1, \dots, S^d)$. We call $\mathcal{U} \subseteq \{1, \dots, d\}$ a *universe* of stocks with capitalization process $S^\mathcal{U} = (S^u)_{u \in \mathcal{U}}$. Moreover, we define the universe weights (resp. universe portfolio) $\mu^\mathcal{U}$ to be

$$\mu_t^{\mathcal{U},i} = \begin{cases} \frac{S_t^{\mathcal{U},i}}{\sum_{j \in \mathcal{U}} S_t^{\mathcal{U},j}} & \text{if } i \in \mathcal{U} \\ 0 & \text{otherwise} \end{cases}.$$

Moreover, denote by $(W_t^\mathcal{U})_{t \in [0,T]}$ the wealth process of universe $\mathcal{U}$, which is given by

$$W_t^\mathcal{U} = \frac{\sum_{i \in \mathcal{U}} S_t^i}{\sum_{i \in \mathcal{U}} S_{t_0}^i}.$$

1.5 Optimization tasks for linear path-functional portfolios

This notion of a universe of stocks is useful if one does not necessarily want to invest in all the stocks in the market but just in a subset of stocks. Of course, this can always be achieved by fixing certain weights of a portfolio to be zero, however using our notion of a universe of stocks is particularly useful, if one wants to compare the wealth process of a portfolio to the wealth process of the universe. Let us make this precise:

Corollary 1.5.2. *Take a universe $\mathcal{U} \subseteq \{1, \dots, d\}$ and consider a portfolio π where $\pi^i \equiv 0$ for $i \notin \mathcal{U}$. Then, the relative wealth process of π with respect to $\mu^\mathcal{U}$ is given by*

$$d\left(\frac{W_t^\pi}{W_t^\mathcal{U}}\right) = \left(\frac{W_t^\pi}{W_t^\mathcal{U}}\right) \sum_{i \in \mathcal{U}} \pi_t^i \frac{d\mu_t^{\mathcal{U},i}}{\mu_t^{\mathcal{U},i}}. \quad (1.5.1)$$

Or, equivalently

$$d\log\left(\frac{W_t^\pi}{W_t^\mathcal{U}}\right) = \sum_{i \in \mathcal{U}} \pi_t^i \frac{d\mu_t^{\mathcal{U},i}}{\mu_t^{\mathcal{U},i}} - \frac{1}{2} \sum_{i,j \in \mathcal{U}} \frac{\pi_t^i}{\mu_t^{\mathcal{U},i}} \frac{\pi_t^j}{\mu_t^{\mathcal{U},j}} d[\mu^{\mathcal{U},i}, \mu^{\mathcal{U},j}]_t. \quad (1.5.2)$$

Proof. Using the definition of π and $W^\mathcal{U}$, the statement follows directly from (1.2.3) and applying Itô's formula. $\square$

Remark 1.5.3. When we consider path-functional portfolios investing in a universe $\mathcal{U}$, it is convenient to construct them via the universe portfolio $\mu^\mathcal{U}$. Note that for path-functional portfolios $\pi^{(I)}$ of type *I* it directly follows that $\pi^{(I),i}(\mu^\mathcal{U}, \cdot) = 0$ for $i \notin \mathcal{U}$. However, for path-functional portfolios $\pi^{(II)}$ of type *II*, we have to require additionally that $\pi^{(II),i}(\mu^\mathcal{U}, \cdot) = 0$ for $i \notin \mathcal{U}$ by setting $f^i \equiv 0$ for $i \notin \mathcal{U}$.

Remark 1.5.4 (Investments in Ranked Markets). Rank-based portfolios are of particular interest in SPT, see Fernholz (2002). We aim to therefore incorporate those into our considerations. Recall that our very basic assumptions on the financial market were that the stocks' capitalizations are positive continuous semimartingales. Considering a market M which fulfills these assumptions, we know that there exists another market which also fulfills these assumptions and for which the stocks' capitalizations are the same as the ranked capitalizations of the market M . This follows simply from the fact that the ranked capitalizations are again positive continuous semimartingales. However, the assumption of NUPBR does not translate directly (see (Karatzas and Ruf, 2017, Section 3.2) for a comment on the potential lack of a local martingale deflator in ranked markets and Kardaras (2012) for a connection between the existence of a local martingale deflator and the NUPBR condition). For the optimization problems to be meaningful in the ranked market, we have to assume NUPBR again additionally. Hence, under the assumption of NUPBR for the ranked market weights, the optimization problems we present in the following hold also for investments in the ranked market.

It is sometimes useful to write optimization tasks in a vectorized form. To this end, let us introduce what we call a *labelling function*.

Definition 1.5.5. For a given linear path-functional portfolio π investing in a universe of stocks $\mathcal{U}$, let $\mathcal{V}$ be the set of features, $|\mathcal{V}|$ the number of features and $|\mathcal{U}|$ the number of stocks in a universe. A bijective function $\mathcal{L}$ of the form

$$\begin{aligned}\mathcal{L} : \mathcal{U} \times \mathcal{V} &\longrightarrow \{1, \dots, |\mathcal{U}| \cdot |\mathcal{V}|\} \\ (i, \nu) &\longmapsto \mathcal{L}(i, \nu)\end{aligned}$$

is called *labelling function*.

Before studying two concrete optimization tasks for linear path-functional portfolios, let us make the following statement about a more general class of optimization tasks, which turn out to be *quadratic optimization problems* for linear path-functional portfolios:

Proposition 1.5.6. Let us denote by $\mathbb{S}^d$ the space of symmetric matrices of dimension $d \times d$. For a universe $\mathcal{U}$, consider a class of linear path-functional portfolios Π (of either type) given by a feature space $\mathcal{V}$, a collection of feature maps $\{\phi^\nu\}_{\nu \in \mathcal{V}}$ consisting of non-anticipative path-functionals, a continuous semimartingale X being the underlying process and an auxiliary portfolio τ . Recall that the portfolio controlling functionals $\{f^i\}_{i \in \mathcal{U}}$ of such portfolios are given by $f^i(t, X_{[0,t]}) = \sum_{\nu \in \mathcal{V}} l_\nu^i \phi^\nu(t, X_{[0,t]})$ for $i \in \mathcal{U}$. Then, any optimization problem of the form

$$\inf_{\pi \in \Pi} \mathbb{E} \left[\int_{t_0}^t \pi_s^\top C_s \pi_s \lambda_1(ds) - \int_{t_0}^t b_s^\top \pi_s \lambda_2(ds) \right]$$

where b , C are stochastic processes with values in $\mathbb{R}^d$, $\mathbb{S}^d$ respectively and λ^i are signed measures on $[t_0, t]$, is a quadratic optimization problem in the linear coefficients $\{l_\nu^i\}_{\nu \in \mathcal{V}, 1 \leq i \leq d}$.

Proof. Recall the form of linear path-functional portfolios of type *I* and *II* respectively. Therefore, it is clear that the parameters $\{l_\nu^i\}$ appear at most in quadratic terms. Moreover, since they are constant in time and deterministic, they can be pulled outside of the integrals. $\square$

Let us now, present two prominent optimization tasks which fall into the above class, namely *mean-variance and log-wealth optimization*. We assume that the following optimization problems are well-posed and that the expectations are finite.

1.5.1 Mean-variance optimization

In this optimization problem, we consider a strategy, where we invest at time t with portfolio weights π_t of a linear path-functional portfolio into a universe $\mathcal{U}$ and follow a buy-and-hold strategy over the time-span Δ . The (relative) return of this strategy over the time-horizon $[t, t + \Delta]$ is given by

$$\mathcal{R}_{t,t+\Delta}^{W,\pi} = \frac{W_{t+\Delta}^\pi}{W_t^\pi} - 1 = \sum_{j \in \mathcal{U}} \pi_t^j \frac{S_{t+\Delta}^j}{S_t^j} - 1 \quad (1.5.3)$$

and respectively the relative return (relative to the universe portfolio $\mu^\mathcal{U}$)

$$\mathcal{R}_{t,t+\Delta}^{V,\pi} = \frac{V_{t+\Delta}^\pi}{V_t^\pi} - 1 = \sum_{j \in \mathcal{U}} \pi_t^j \frac{\mu_{t+\Delta}^{\mathcal{U},j}}{\mu_t^{\mathcal{U},j}} - 1. \quad (1.5.4)$$

1.5 Optimization tasks for linear path-functional portfolios

In this setting, we can state the following corollary about the mean-variance optimization problem:

Corollary 1.5.7. *Let $\lambda \in \mathbb{R}_+$ be a risk-tolerance parameter, specifying the investor's risk-preferences. Then, for a linear path-functional portfolio $\pi(\tau, X)$ the following two optimization problems are convex quadratic optimization problems in the parameters $\{l_\nu^i\}_{i \in \mathcal{U}, \nu \in \mathcal{V}}$. More precisely, for the*

(i) *mean-variance optimization of returns we have*

$$\begin{aligned} & \min_{\pi} \text{Var} \left(\mathcal{R}_{t,t+\Delta}^{W,\pi} \right) - \lambda \mathbb{E} \left(\mathcal{R}_{t,t+\Delta}^{W,\pi} \right) \\ & \Leftrightarrow \min_{\{l_\nu^i\}} \mathbf{1}^\top \text{Var} \left(\mathbf{Y}^W \right) \mathbf{1} - \mathbf{1}^\top \left(\lambda \mathbb{E} \left(\mathbf{Y}^W \right) - \boldsymbol{\sigma}^W \right) \end{aligned}$$

for $\mathbf{l}_{\mathcal{L}(i,\nu)} = l_\nu^i$, $\boldsymbol{\sigma}_{\mathcal{L}(i,\nu)}^W = 2\text{Cov}(\mathbf{Y}_{\mathcal{L}(i,\nu)}^W, \mathcal{R}_{t,t+\Delta}^{W,\tau})$,

$$\mathbf{Y}_{\mathcal{L}(i,\nu)}^W = \begin{cases} \tau_t^i \phi^\nu(t, X_{[0,t]}) \left(\frac{S_{t+\Delta}^i}{S_t^i} - \mathcal{R}_{t,t+\Delta}^{W,\tau} - 1 \right) & \text{if } \pi \text{ is of type I} \\ \phi^\nu(t, X_{[0,t]}) \left(\frac{S_{t+\Delta}^i}{S_t^i} - \mathcal{R}_{t,t+\Delta}^{W,\tau} - 1 \right) & \text{if } \pi \text{ is of type II} \end{cases}$$

and where $\text{Var}(\mathbf{Y}^W)$ refers to the covariance matrix of $\mathbf{Y}^W$.

(ii) *mean-variance optimization of relative return we have*

$$\begin{aligned} & \min_{\pi} \text{Var} \left(\mathcal{R}_{t,t+\Delta}^{V,\pi} \right) - \lambda \mathbb{E} \left(\mathcal{R}_{t,t+\Delta}^{V,\pi} \right) \\ & \Leftrightarrow \min_{\{l_\nu^i\}} \mathbf{1}^\top \text{Var} \left(\mathbf{Y}^V \right) \mathbf{1} - \mathbf{1}^\top \left(\lambda \mathbb{E} \left(\mathbf{Y}^V \right) - \boldsymbol{\sigma}^V \right) \end{aligned}$$

for $\mathbf{l}_{\mathcal{L}(i,\nu)} = l_\nu^i$, $\boldsymbol{\sigma}_{\mathcal{L}(i,\nu)}^V = 2\text{Cov}(\mathbf{Y}_{\mathcal{L}(i,\nu)}^V, \mathcal{R}_{t,t+\Delta}^{V,\tau})$ and

$$\mathbf{Y}_{\mathcal{L}(i,\nu)}^V = \begin{cases} \tau_t^i \phi^\nu(t, X_{[0,t]}) \left(\frac{\mu_{t+\Delta}^{\mathcal{U},i}}{\mu_t^{\mathcal{U},i}} - \mathcal{R}_{t,t+\Delta}^{V,\tau} - 1 \right) & \text{if } \pi \text{ is of type I} \\ \phi^\nu(t, X_{[0,t]}) \left(\frac{\mu_{t+\Delta}^{\mathcal{U},i}}{\mu_t^{\mathcal{U},i}} - \mathcal{R}_{t,t+\Delta}^{V,\tau} - 1 \right) & \text{if } \pi \text{ is of type II} \end{cases}$$

and where $\text{Var}(\mathbf{Y}^V)$ refers to the covariance matrix of $\mathbf{Y}^V$.

Proof. We proof the above statement for the case of $\mathcal{R}^{W,\pi}$ and for π of type I. The proof of the other cases is completely analogous.

$$\begin{aligned} \mathcal{R}_{t,t+\Delta}^{W,\pi} &= \sum_{i \in \mathcal{U}} \tau_t^i \left(f^i(t, X_{[0,t]}) + 1 - \sum_{j \in \mathcal{U}} \tau_t^j f^j(t, X_{[0,t]}) \right) \left(\frac{S_{t+\Delta}^i}{S_t^i} \right) - 1 \\ &= \sum_{i \in \mathcal{U}} \tau_t^i f^i(t, X_{[0,t]}) \left(\frac{S_{t+\Delta}^i}{S_t^i} - \mathcal{R}_{t,t+\Delta}^{W,\tau} - 1 \right) + \mathcal{R}_{t,t+\Delta}^{W,\tau} + 1 - 1 \\ &= \mathbf{1}^\top \mathbf{Y}^W + \mathcal{R}_{t,t+\Delta}^{W,\tau} \end{aligned}$$

Hence,

$$\mathbb{E}(\mathcal{R}_{t,t+\Delta}^{W,\pi}) = \mathbf{I}^\top \mathbb{E}(\mathbf{Y}^W) + \mathbb{E}(\mathcal{R}_{t,t+\Delta}^{W,\tau}). \quad (1.5.5)$$

Note that the rightmost term of (1.5.5) do not contribute in the optimization problem, as they are independent of $\mathbf{l}$. And,

$$\begin{aligned} \text{Var}(\mathcal{R}_{t,t+\Delta}^{W,\pi}) &= \text{Var}(\mathbf{I}^\top \mathbf{Y}^W) + 2\text{Cov}(\mathbf{I}^\top \mathbf{Y}^W, \mathcal{R}_{t,t+\Delta}^{W,\tau}) + \text{Var}(\mathcal{R}_{t,t+\Delta}^{W,\tau}) \\ &= \mathbf{I}^\top \text{Var}(\mathbf{Y}^W) \mathbf{l} + \mathbf{I}^\top \boldsymbol{\sigma}^W + \text{Var}(\mathcal{R}_{t,t+\Delta}^{W,\tau}), \end{aligned} \quad (1.5.6)$$

where the rightmost term of (1.5.6) does not contribute in the optimization problem. The fact that the optimization problem is convex follows directly from the fact that the covariance matrix $\text{Var}(\mathbf{Y}^W)$ is positive semi-definite. $\square$

Remark 1.5.8. Note that the expression of the optimization problem for the relative return simplifies considerably if we use a linear path-functional portfolio of type I and choose the universe portfolio as the auxiliary portfolio, i.e. $\tau = \mu^{\mathcal{U}}$. To be precise, we obtain

$$\begin{aligned} &\min_{\pi} \text{Var}(\mathcal{R}_{t,t+\Delta}^{V,\pi}) - \lambda \mathbb{E}(\mathcal{R}_{t,t+\Delta}^{V,\pi}) \\ &\Leftrightarrow \min_{\{l_\nu^i\}} \mathbf{I}^\top \text{Var}(\mathbf{Y}^V) \mathbf{l} - \lambda \mathbf{I}^\top \mathbb{E}(\mathbf{Y}^V) \end{aligned}$$

for $\mathbf{l}_{\mathcal{L}(i,\nu)} = l_\nu^i$ and $\mathbf{Y}_{\mathcal{L}(i,\nu)}^V = \phi^\nu(t, X_{[0,t]})(\mu_{t+\Delta}^{\mathcal{U},i} - \mu_t^{\mathcal{U},i})$, since $\mathcal{R}_{t,t+\Delta}^{V,\mu^{\mathcal{U}}} = 0$.

Remark 1.5.9. If $\text{Var}(\mathbf{Y}^V)$ or $\text{Var}(\mathbf{Y}^W)$ respectively are almost surely invertible, that is the optimization problem is *strictly* convex and hence admits a unique set of optimal parameters, then the optimal vector $\mathbf{l}^*$ can be found via first-order conditions, i.e.

$$\mathbf{l}^* = \text{Var}(\mathbf{Y}^V)^{-1} (\lambda \mathbb{E}(\mathbf{Y}^V) - \boldsymbol{\sigma}^V)$$

and likewise for the case of optimizing returns. Note that the problem is never strictly convex for *any* linear path-functional portfolio of type I , because due to the form of such portfolios, the weights are invariant under scalar shifts of the parameters $\{l_\nu^i\}_{\nu \in \mathcal{V}, 1 \leq i \leq d}$. More precisely, for linear path-functional portfolios of type I , let $\pi^*(\tau, X)$ be optimal portfolio weights corresponding to an optimal set $\{l_\nu^{*,i}\}_{\nu \in \mathcal{V}, 1 \leq i \leq d}$, then the set $\{l_\nu^{*,i} + \alpha\}_{\nu \in \mathcal{V}, 1 \leq i \leq d}$ for $\alpha \in \mathbb{R}$ results in *the same* portfolio weights $\pi^*(\tau, X)$.

Remark 1.5.10. Let us here outline how the mean-variance optimization problem (of the relative returns) of Corollary 1.5.7 can be related to a (model-free) real market situation where only the observation of one trajectory is available (see also Section 1.7.2). Tailored to Corollary 1.5.7 we assume a discrete time situation with time steps of length Δ and suppose that the conditional law of $\mu_{i\Delta}$ is time-homogeneous and depends on the last $n \wedge i$ values of μ for some $n \in \mathbb{N}$. More precisely, given $(\mu_{(i-n)\Delta \wedge 0}, \dots, \mu_{(i-1)\Delta}) = (x_0, \dots, x_{(n-1) \wedge (i-1)}) \in (\Delta^d)^{n \wedge i}$, we know the conditional law $\rho(x_0, \dots, x_{(n-1) \wedge (i-1)}, \cdot)$ of $\mu_{i\Delta}$. Consider now a portfolio π

1.5 Optimization tasks for linear path-functional portfolios

that is a non-anticipative path functional depending also only on the past n values of μ , i.e. a portfolio with rolling window in the spirit of Remark 1.3.9. Then for $i \geq n$

$$\begin{aligned} \mathbb{E} \left[\frac{V_{i\Delta}^\pi}{V_{(i-1)\Delta}^\pi} \middle| \mathcal{F}_{(i-1)\Delta} \right] &= \mathbb{E}[(R_{(i-1)\Delta, i\Delta}^{V, \pi} + 1) | \mathcal{F}_{(i-1)\Delta}] \\ &= \int_{\Delta^d} \sum_{j=1}^d \pi^j(x_0, \dots, x_{n-1}) \frac{y^j}{x_{n-1}^j} \rho(x_0, \dots, x_{n-1}, dy) \end{aligned}$$

and similarly for $i < n$, taking just the history starting from 0 into account. Assume now further that we have an invariant measure for the distribution of the rolling path segments of length n , which we denote by ρ . Then under this invariant measure we have for each $i \geq n$

$$\begin{aligned} \mathbb{E} \left[\frac{V_{i\Delta}^\pi}{V_{(i-1)\Delta}^\pi} \right] &= \int_{(\Delta^d)^n} \int_{\Delta^d} \sum_{j=1}^d \pi^j(x_0, \dots, x_{n-1}) \frac{y^j}{x_{n-1}^j} \rho(x_0, \dots, x_{n-1}, dy) \rho(dx_0, \dots, dx_{n-1}) \\ &=: M^\pi \end{aligned}$$

and accordingly for $i < n$ assuming compatible invariant measures for the paths of shorter lengths. Under these assumptions we then obtain the following ergodicity result

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \frac{V_{i\Delta}^\pi}{V_{(i-1)\Delta}^\pi} = 1 + \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N R_{(i-1)\Delta, i\Delta}^{V, \pi} = M^\pi.$$

This can be deduced from Birkhoff's ergodic theorem for discrete time Markov processes (see e.g. [Theorem 2.2, Section 2.1.4] Eberle (2016) and also (Cuchiero et al., 2019, Section 3) for the standard Markov situation with $n = 1$). Indeed, this becomes applicable by identifying the states of the Markov process as the past path segments of length n , i.e. a state $z_{i\Delta}$ is given by $z_{i\Delta} = (\mu_{(i-n)\Delta}, \dots, \mu_{i\Delta})$. The transition probabilities from $z_{i\Delta}$ to $z_{(i+1)\Delta}$ must then be chosen such that they assign 0 probability to elements where there exists some $j \in \{1, \dots, n-1\}$ with $z_{(i+1)\Delta}^j \neq z_{i\Delta}^{j+1}$.

Clearly the above reasoning also holds true for $\text{Var} \left[V_{i\Delta}^\pi / V_{(i-1)\Delta}^\pi \right] = \text{Var} [R_{(i-1)\Delta, i\Delta}^{V, \pi}]$. Therefore, under such ergodicity assumptions all the expected values in Corollary 1.5.7 can be approximated by time averages, which will be important in our numerical implementations as outlined in Section 1.7.2.

1.5.2 Optimizing the (expected) log-(relative)-wealth

Similarly as the mean-variance portfolio optimization problem also the log-optimal portfolio over linear path functional portfolios can be found by solving a convex quadratic optimization task. We state it here in form of the relative log-utility optimization problem, i.e. the goal is to solve

$$\sup_{\pi} \mathbb{E} \left[\log \left(\frac{W_t^\pi}{W_t^{\mathcal{U}}} \right) \right] = \sup_{\pi} \mathbb{E} [\log(V_t^\pi)],$$

where V_t^π denotes again the relative wealth with respect to the universe $\mathcal{U}$. Note however that under similar ergodicity conditions as in Remark 1.5.10 the use of time averages

$$\frac{1}{N} \log(V_{N\Delta}^\pi) = \frac{1}{N} \sum_{i=1}^N \log \left(\frac{V_{i\Delta}^\pi}{V_{(i-1)\Delta}^\pi} \right)$$

is justified to approximate $\frac{1}{N} \mathbb{E}[\log(V_{N\Delta}^\pi)]$. This can also be translated to continuous time, where under appropriate assumptions (see e.g. (Cuchiero et al., 2019, Section 4.2.3 and Theorem 4.9)) we also have that

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log(V_t^\pi) = \lim_{t \rightarrow \infty} \frac{1}{t} \mathbb{E}[\log(V_t^\pi)],$$

whence

$$\frac{1}{t} \log(V_t^\pi) \approx \frac{1}{t} \mathbb{E}[\log(V_t^\pi)].$$

In this case the expected values in the subsequent theorem, i.e. $\mathbb{E}[\mathbf{Q}(t)]$ and $\mathbb{E}[\mathbf{c}(t)]$ can therefore be just replaced by $\mathbf{Q}(t)$ and $\mathbf{c}(t)$ computed along the observed trajectory. This will play again an important role in our empirical implementations.

Theorem 1.5.11 (Relative Log-Utility Optimization). *Consider a universe $\mathcal{U} \subseteq \{1, \dots, d\}$ and a linear path functional portfolio $\pi(\mu^\mathcal{U}, X)$. Let $t_0 \geq 0$ be the time at which we start investing. Denote by $\mathcal{L}$ an arbitrary but fixed labelling function. Then, for the (relative) log-utility optimization problem we have*

$$\sup_{\{l_\nu^{(i)}\}_{i \in \mathcal{U}, \nu \in \mathcal{V}}} \mathbb{E} \left[\log \left(\frac{W_t^\pi}{W_t^\mathcal{U}} \right) \right] \Leftrightarrow \inf_{\mathbf{l} \in \mathbb{R}^{|\mathcal{U}| \cdot |\mathcal{V}|}} \frac{1}{2} \mathbf{l}^\top \mathbb{E}[\mathbf{Q}(t)] \mathbf{l} - \mathbb{E}[\mathbf{c}(t)^\top] \mathbf{l}. \quad (1.5.7)$$

where $\mathbf{l}_{\mathcal{L}(i, \nu)} = l_\nu^i$ for $i \in \mathcal{U}$, $\nu \in \mathcal{V}$ and

$$(\mathbf{c}(t))_{\mathcal{L}(i, \nu)} = \int_{t_0}^t \gamma_s^i \phi^\nu(s, X_{[0, s]}) d\mu_s^{\mathcal{U}, i}$$

$$(\mathbf{Q}(t))_{\mathcal{L}(i, \nu), \mathcal{L}(j, \rho)} = \int_{t_0}^t \gamma_s^i \gamma_s^j \phi^\nu(s, X_{[0, s]}) \phi^\rho(s, X_{[0, s]}) d[\mu^{\mathcal{U}, i}, \mu^{\mathcal{U}, j}]_s$$

for

$$\gamma_s^i = \begin{cases} 1 & \text{if } \pi \text{ is of type I} \\ (\mu_s^{\mathcal{U}, i})^{-1} & \text{if } \pi \text{ is of type II} \end{cases}.$$

Moreover, it is a convex quadratic optimization problem.

Remark 1.5.12. Let us highlight how the expression of $\mathbf{c}$, $\mathbf{Q}$ simplify in the case of a signature portfolio of type I, $t_0 = 0$, $\mathcal{U} = \{1, \dots, d\}$, $X = \mu$ and μ is such that it holds for some $N_0 > 0$

$$d[\mu^i, \mu^j]_t = \sum_{0 \leq |I| \leq N_0} \alpha_I^{i, j} \langle e_I, \hat{\mu}_t \rangle,$$

1.5 Optimization tasks for linear path-functional portfolios

where $\alpha_I^{i,j} \in \mathbb{R}$ for all $i, j \in \{1, \dots, d\}$, $0 \leq |I| \leq N_0$. Then we obtain

$$\begin{aligned} (\mathbf{c}(t))_{\mathcal{L}(i,I)} &= \langle e_I \otimes e_i, \hat{\mu}_t \rangle \\ (\mathbf{Q}(t))_{\mathcal{L}(i,I), \mathcal{L}(j,J)} &= \sum_{0 \leq |K| \leq N_0} \alpha_K^{i,j} \langle (e_I \sqcup e_J) \otimes e_K, \hat{\mu}_t \rangle. \end{aligned}$$

More generally, consider a signature portfolio $\pi(\mu, \check{\mu})$ of type I where $\check{\mu}_t := (t, \mu_t, [\mu, \mu]_t)$ for all $t \in [0, T]$ in a vectorized form, i.e. fix a labelling function $\kappa(\cdot, \cdot)$ such that $\check{\mu}^{\kappa(i,j)} = [\mu^i, \mu^j]$ for $1 \leq i, j \leq d$ and set $t_0 = 0$, $\mathcal{U} = \{1, \dots, d\}$. Then the elements of $\mathbf{c}$ and $\mathbf{Q}$ are linear functions on the signature of $\check{\mu}$, namely

$$\begin{aligned} (\mathbf{c}(t))_{\mathcal{L}(i,I)} &= \langle e_I \otimes e_i, \check{\mu}_t \rangle \\ (\mathbf{Q}(t))_{\mathcal{L}(i,I), \mathcal{L}(j,J)} &= \langle (e_I \sqcup e_J) \otimes e_{\kappa(i,j)}, \check{\mu}_t \rangle. \end{aligned}$$

Remark 1.5.13. Note that although we start trading at time $t_0 \geq 0$ we feed to the feature maps ϕ^ν the semimartingale X starting at time 0. The interpretation of this is that we start “observing” the market at time 0 and but only start trading at time $t_0 \geq 0$. This time-difference between observing and starting to invest can be important in non-Markovian settings.

Before we prove Theorem 1.5.11, let us state the following lemma.

Lemma 1.5.14. For a path-functional portfolio $\pi(\mu^{\mathcal{U}}, X)$ it holds that

$$\begin{aligned} \log \left(\frac{W_t^\pi}{W_t^{\mathcal{U}}} \right) &= \sum_{i \in \mathcal{U}} \int_{t_0}^t \gamma_s^i f^i(s, X_{[0,s]}) d\mu_s^{\mathcal{U},i} \\ &\quad - \frac{1}{2} \sum_{i,j \in \mathcal{U}} \int_{t_0}^t \gamma_s^i \gamma_s^j f^i(s, X_{[0,s]}) f^j(s, X_{[0,s]}) d[\mu^{\mathcal{U},i}, \mu^{\mathcal{U},j}]_s, \end{aligned} \quad (1.5.8)$$

where

$$\gamma_s^i = \begin{cases} 1 & \text{if } \pi \text{ is of type } I \\ (\mu_s^{\mathcal{U},i})^{-1} & \text{if } \pi \text{ is of type } II \end{cases}.$$

We refer the reader to Appendix 1.C for a proof of Lemma 1.5.14.

Proof of Theorem 1.5.11. To show the equivalence in (1.5.7), we use Lemma 1.5.14 and make use of the fact that we are using linear path-functional portfolios.

$$\log \left(\frac{W_t^\pi}{W_t^{\mathcal{U}}} \right) = \sum_{i \in \mathcal{U}} \sum_{\nu \in \mathcal{V}} l_\nu^i \int_{t_0}^t \gamma_s^i \phi^\nu(s, X_{[0,s]}) d\mu_s^{\mathcal{U},i} \quad (1.5.9)$$

$$- \frac{1}{2} \sum_{i,j \in \mathcal{U}} \sum_{\nu, \rho \in \mathcal{V}} l_\rho^j l_\nu^i \int_{t_0}^t \gamma_s^i \gamma_s^j \phi^\nu(s, X_{[0,s]}) \phi^\rho(s, X_{[0,s]}) d[\mu^{\mathcal{U},i}, \mu^{\mathcal{U},j}]_s \quad (1.5.10)$$

Hence, we have shown that

$$\log \left(\frac{W_t^\pi}{W_t^\mathcal{U}} \right) = \mathbf{c}(t)^\top \mathbf{l} - \frac{1}{2} \mathbf{l}^\top \mathbf{Q}(t) \mathbf{l} \quad (1.5.11)$$

from which the equivalence in (1.5.7) follows directly.

To show that the optimization problem given in (1.5.7) is convex, we show that the matrix $\mathbf{Q}(t)$ is positive semidefinite. Consider the stochastic process $Y_t = \sum_{j=1}^d \int_{t_0}^t \gamma_s^j f^j(s, X_{[0,s]}) d\mu_s^{\mathcal{U},j}$ whose quadratic variation $[Y, Y]_t$ is given by it holds that

$$[Y, Y]_t = \sum_{i=1}^d \sum_{j=1}^d \int_{t_0}^t \gamma_s^i \gamma_s^j f^i(s, X_{[0,s]}) f^j(s, X_{[0,s]}) d[\mu^{\mathcal{U},i}, \mu^{\mathcal{U},j}]_s.$$

On the other hand

$$\mathbf{l}^\top \mathbf{Q}(t) \mathbf{l} = \sum_{i=1}^d \sum_{j=1}^d \int_{t_0}^t \gamma_s^i \gamma_s^j f^i(s, X_{[0,s]}) f^j(s, X_{[0,s]}) d[\mu^{\mathcal{U},i}, \mu^{\mathcal{U},j}]_s.$$

Hence, $\mathbf{l}^\top \mathbf{Q}(t) \mathbf{l} = [Y, Y]_t \geq 0$ for all $t \in [t_0, T]$ and for all $\mathbf{l} \in \mathbb{R}^{(|\mathcal{U}| \cdot |\mathcal{V}|)}$, proving that $\mathbf{Q}(t)$ is positive semidefinite for every $t \in [t_0, T]$. $\square$

Remark 1.5.15. If the log-utility maximization problem is finite, it holds that

$$\sup_{\{l_\nu^{(i)}\}_{i \in \mathcal{U}, \nu \in \mathcal{V}}} \mathbb{E} [\log (W_t^\pi)] \Leftrightarrow \sup_{\{l_\nu^{(i)}\}_{i \in \mathcal{U}, \nu \in \mathcal{V}}} \mathbb{E} \left[\log \left(\frac{W_t^\pi}{W_t^\mathcal{U}} \right) \right]. \quad (1.5.12)$$

This simply follows from the following equivalences

$$\begin{aligned} \sup_{\{l_\nu^{(i)}\}_{i \in \mathcal{U}, \nu \in \mathcal{V}}} \mathbb{E} [\log (W_t^\pi)] &\Leftrightarrow \sup_{\{l_\nu^{(i)}\}_{i \in \mathcal{U}, \nu \in \mathcal{V}}} \mathbb{E} \left[\log (W_t^\pi) - \log (W_t^\mathcal{U}) \right] \\ &\Leftrightarrow \sup_{\{l_\nu^{(i)}\}_{i \in \mathcal{U}, \nu \in \mathcal{V}}} \mathbb{E} \left[\log \left(\frac{W_t^\pi}{W_t^\mathcal{U}} \right) \right]. \end{aligned}$$

Remark 1.5.16. Similar to Remark 1.5.9, the optimal vector $\mathbf{l}^*$ is given by

$$\mathbf{l}^* = 2\mathbb{E}[\mathbf{Q}(t)]^{-1} \mathbb{E}[\mathbf{c}(t)],$$

if $\mathbb{E}[\mathbf{Q}(t)]$ is almost surely invertible. As mentioned this can never hold for linear path-functional portfolios of type I .

1.5.3 Constraints and regularization

Given an optimization task involving a linear path-functional portfolio which is convex and quadratic, we can make the following statements about adding constraints or regularization:

Proposition 1.5.17. Given an optimization problem

$$\min_{\mathbf{l}} -\mathbf{l}^T \mathbf{c}(t) + \frac{1}{2} \mathbf{l}^T \mathbf{Q}(t) \mathbf{l}$$

which is convex quadratic, it remains convex quadratic under the following modifications.

- (i) **Bounds Constraints:** $|l_\nu^i|^2 \leq b_{\mathcal{L}(i,\nu)}$ for $b_{\mathcal{L}(i,\nu)} \geq 0$ for all $i \in \mathcal{U}$, $\nu \in \mathcal{V}$.
- (ii) **L2-Regularization:** we can add an L^2 -regularization term of the form $\gamma \mathbf{l}^T \mathbf{l}$ such that the optimization problem becomes

$$\min_{\mathbf{l}} -\mathbf{l}^T \mathbf{c}(t) + \frac{1}{2} \mathbf{l}^T \mathbf{Q}(t) \mathbf{l} + \gamma \mathbf{l}^T \mathbf{l}. \quad (1.5.13)$$

Note that, this corresponds to a new matrix $\tilde{\mathbf{Q}}(t) = \mathbf{Q}(t) + \gamma \mathbb{I}$, where $\mathbb{I}$ is the identity matrix.

Moreover, the above modifications can be combined and the optimization task remains convex.

Proof. (i) $|l_\nu^i| \leq b_{\mathcal{L}(i,\nu)}$ for $b_{\mathcal{L}(i,\nu)} \geq 0$ corresponds to the constraint

$$\mathbf{l}^T \delta_{\mathcal{L}(i,\nu), \mathcal{L}(i,\nu)} \mathbf{l} - b_{\mathcal{L}(i,\nu)}^2 \leq 0,$$

where $\delta_{\mathcal{L}(i,\nu), \mathcal{L}(i,\nu)}$ is the matrix

$$\delta_{jk} = \begin{cases} 1 & \text{if } j = k = \mathcal{L}(i, \nu) \\ 0 & \text{otherwise} \end{cases},$$

which is obviously positive semi-definite.

- (ii) Since $\mathbf{Q}(t)$ and $\mathbb{I}$ are positive semi-definite and the sum of positive semi-definite matrices is positive semi-definite, $\tilde{\mathbf{Q}}(t)$ is again positive semi-definite. □

Remark 1.5.18 (Long-Only and Leverage Constraints). We can add several forms of short-selling constraints under which the convex quadratic optimization problems remain convex and quadratic for linear path-functional portfolios. However, we would like to emphasise that these constraints are only effective during training and there is no theoretical guarantee that the portfolio-weights respect such constraints during the out-of-sample period. The reason for this is that different dynamics of the underlying process X may lead to violation of the constraints during the testing period.

- (i) **Long-Only:** $\pi^i(t) \geq 0$ for all $i \in \mathcal{U}$;
- (ii) **Bounds for Short-Selling:** $-a \leq \pi^i(t) \leq b$ for all $i \in \mathcal{U}$ and some $a, b \in \mathbb{R}$;

- (iii) **Leverage constraints** (see [Ni et al. \(2023\)](#)): $\pi^i(t) \geq 0$ for all $i \in \mathcal{U}_L \subseteq \mathcal{U}$ i.e., impose long-only constraints on a subset $\mathcal{U}_L$ of stocks and impose a leverage constraint $\sum_{i \in \mathcal{U}_L} \pi_t^i \leq b$ for some $b > 0$.

Proposition 1.5.19 (Long-Only Path-Functional Portfolios). When choosing τ to be long-only, we can formulate the following sufficient conditions for a path-functional portfolio to be long-only which increase in strength in the sense that $(iv) \Rightarrow (iii) \Rightarrow (ii) \Rightarrow (i)$. The following are required to hold $\mathbb{P}$ -a.s. for all $t \in [0, T]$.

- (i) $|f^i(t, X_{[0,t]}) - f^{max}(t, X_{[0,t]})| \leq 1$ where

$$f^{max}(t, X_{[0,t]}) := \max\{f^1(t, X_{[0,t]}), \dots, f^d(t, X_{[0,t]})\}$$

for all $i \in \{1, \dots, d\}$

- (ii) $|f^i(t, X_{[0,t]}) - f^j(t, X_{[0,t]})| \leq 1$ for all $i, j \in \{1, \dots, d\}$

- (iii) $0 \leq f^i(t, X_{[0,t]}) \leq 1$ for all $i \in \{1, \dots, d\}$

- (iv) $f^i(t, X_{[0,t]}) \in [\alpha, \alpha + 1]$, $\alpha \in \mathbb{R}$ for all $i \in \{1, \dots, d\}$

Proof. By definition, we know that for each $i \in \{1, \dots, d\}$ for path-functional portfolios of type *I* it holds:

$$\begin{aligned} \pi^i(t) \geq 0 &\Leftrightarrow \left(f^i(t, X_{[0,t]}) + 1 - \sum_{j=1}^d f^j(t, X_{[0,t]}) \mu_t^j \right) \geq 0 \\ &\Leftrightarrow \left(f^i(t, X_{[0,t]}) - \sum_{j=1}^d f^j(t, X_{[0,t]}) \mu_t^j \right) \geq -1. \end{aligned}$$

One sufficient condition for this to hold is

$$|f^i(t, X_{[0,t]}) - f^{max}(t, X_{[0,t]})| \leq 1$$

where $f^{max}(t, X_{[0,t]}) := \max\{f^1(t, X_{[0,t]}), \dots, f^d(t, X_{[0,t]})\}$. That is because

$$\begin{aligned} \left(f^i(t, X_{[0,t]}) - \sum_{j=1}^d f^j(t, X_{[0,t]}) \mu_t^j \right) &\geq f^i(t, X_{[0,t]}) - f^{max}(t, X_{[0,t]}) \sum_{j=1}^d \mu_t^j \\ &= f^i(t, X_{[0,t]}) - f^{max}(t, X_{[0,t]}). \end{aligned}$$

Similarly, for path-functional portfolio of type *II* we find:

$$\pi^i(t) \geq 0 \Leftrightarrow f^i(t, X_{[0,t]}) - \tau_t^i \sum_{j=1}^d f^j(t, X_{[0,T]}) \geq -\tau_t^i \geq -1$$

1.5 Optimization tasks for linear path-functional portfolios

and also

$$\begin{aligned} f^i(t, X_{[0,t]}) - \tau_t^i \sum_{j=1}^d f^j(t, X_{[0,T]}) &\geq f^i(t, X_{[0,t]}) - d \cdot \tau_t^i f^{max}(t, X_{[0,T]}) \\ &\geq f^i(t, X_{[0,t]}) - f^{max}(t, X_{[0,t]}). \end{aligned}$$

□

The previous sufficient conditions are formulated on the level of portfolio controlling functions. How they translate to the optimization parameters $\{l_\nu^i\}_{\nu \in \mathcal{V}, 1 \leq i \leq d}$ for linear path-functional portfolios depends on the form of the feature maps and on the properties of X . The following example holds for a particular case of signature-portfolios with $X = \mu$.

Example 1.5.20. Let us choose $X = \mu$, $\tau = \mu$ and a linear path-functional portfolio of either type *I* or *II* with feature maps

$$\phi^{i_1 \sqcup \dots \sqcup i_m}(t, \mu_{[0,t]}) = \langle e_{i_1 \sqcup \dots \sqcup i_m}, \hat{\mu}_t \rangle = \hat{\mu}_t^{i_1} \dots \hat{\mu}_t^{i_m}$$

for $0 \leq m \leq N$ and $i_1, \dots, i_m \in \{1, \dots, d+1\}$. Hence, by the form of the above the feature maps are simply polynomials and it holds that

$$0 \leq \phi^{i_1 \sqcup \dots \sqcup i_m}(t, \mu_{[0,t]}) \leq 1 \text{ for all } t \in [0, T] \text{ and all } i_1, \dots, i_m \in \{1, \dots, d+1\}, m \leq N$$

assuming w.l.o.g that the time-augmentation $\varphi(t) \in [0, 1]$ for all $t \in [0, T]$ (recalling that $\hat{\mu}_t^1 = \varphi(t)$). In this case it follows that

$$\begin{aligned} |f^j(t, X_{[0,t]}) - f^k(t, X_{[0,t]})| &= \left| \sum_{m=0}^N \sum_{1 \leq i_1 \leq \dots \leq i_m \leq d+1} (l_{i_1 \sqcup \dots \sqcup i_m}^{(j)} - l_{i_1 \sqcup \dots \sqcup i_m}^{(k)}) \langle e_{i_1 \sqcup \dots \sqcup i_m}, \hat{\mu}_t \rangle \right| \\ &\leq \sum_{m=0}^N \sum_{1 \leq i_1 \leq \dots \leq i_m \leq d+1} |l_{i_1 \sqcup \dots \sqcup i_m}^{(j)} - l_{i_1 \sqcup \dots \sqcup i_m}^{(k)}| \cdot |\langle e_{i_1 \sqcup \dots \sqcup i_m}, \hat{\mu}_t \rangle| \\ &\leq \sum_{m=0}^N \sum_{1 \leq i_1 \leq \dots \leq i_m \leq d+1} |l_{i_1 \sqcup \dots \sqcup i_m}^{(j)} - l_{i_1 \sqcup \dots \sqcup i_m}^{(k)}| \end{aligned}$$

We denote by $\tilde{N}$ the number of words of length $0 \leq m \leq N$ with letters in $\{1, \dots, d+1\}$, where the order does not matter, more precisely,

$$\tilde{N} = 1 + \sum_{m=1}^N \binom{d+m}{m}.$$

Then, $|l_{i_1 \sqcup \dots \sqcup i_N}^{(j)} - l_{i_1 \sqcup \dots \sqcup i_N}^{(k)}| \leq 1/\tilde{N}$ implies that

$$\sum_{N=0}^n \sum_{1 \leq i_1 \leq \dots \leq i_N \leq d} |l_{i_1 \sqcup \dots \sqcup i_N}^{(j)} - l_{i_1 \sqcup \dots \sqcup i_N}^{(k)}| \leq 1. \quad (1.5.14)$$

And hence the quadratic constraint, preserving the convex quadratic structure of the optimization problem,

$$(l_{i_1 \sqcup \dots \sqcup i_N}^{(j)} - l_{i_1 \sqcup \dots \sqcup i_N}^{(k)})^2 \leq \frac{1}{\tilde{N}^2}$$

is sufficient to ensure that the signature portfolios are long-only. Of course this is a particular case of a signature portfolio based on multivariate polynomials whose coefficients should not be too far away from each other.

1.6 Transaction costs for portfolios with short-selling and without bank account

When trading in the presence of (proportional) transaction costs, we need to consider re-balancing at discrete times. Let t be an arbitrary but fixed time at which we re-balance. At time t^- (just before re-balancing) our portfolio has weights π_{t^-} and a wealth of $W_{t^-}^\pi$. We want to re-balance to some fixed target weights π_t . This situation has been studied in [Ruf and Xie \(2020\)](#) for the case of long-only weights. However, we allow for short-selling in our portfolios which makes the problem significantly more involved, as we illustrate below.

As the transaction costs are not paid on the portfolio weights, but on the dollar amounts invested in a stock, we define $\psi_{t^-}^i, \psi_t^i$ to be the dollar amounts invested in stock i directly before and after re-balancing. Note, that of course it holds

$$\psi_t^i = \pi_t^i \cdot W_t^\pi$$

and the transaction costs to be paid are

$$TC_t = c \sum_{i=1}^d |\psi_t^i - \psi_{t^-}^i|$$

where c is the percentage of transaction costs to be paid. The difficulty lies in solving for ψ_t^i , since the amount which we can invest in stock i depends on the amount of transaction costs to be paid *and vice-versa*. This difficulty arises from the fact, that we do not have a bank-account. Since we know the portfolio weights at each time, we need to solve for W_t^π . We can use the self-financing identity which requires

$$W_t^\pi = W_{t^-}^\pi - TC_t = W_{t^-}^\pi - c \sum_{i=1}^d |\psi_t^i - \psi_{t^-}^i|. \quad (1.6.1)$$

Moreover, it is convenient to define $\alpha(t) := \frac{W_t^\pi}{W_{t^-}^\pi}$ and rewrite (1.6.1) as

$$1 - \alpha_t = c \sum_{i=1}^d |\alpha_t \pi_t^i - \pi_{t^-}^i|. \quad (1.6.2)$$

1.6 Transaction costs for portfolios with short-selling and without bank account

Hence, we reduced the problem to solving for α_t . Since we consider an arbitrary but fixed re-balancing time t , let us set $\alpha := \alpha_t$ in the following. We can make the following statement about solutions for α :

Proposition 1.6.1. Let α^* be a solution to the equation

$$1 - \alpha = c \sum_{i=1}^d |\alpha \pi_t^i - \pi_{t-}^i|$$

and denote by $L(\alpha) = 1 - \alpha$ and $R(\alpha) = c \sum_{i=1}^d |\alpha \pi_t^i - \pi_{t-}^i|$. We can make the following statements about α^* :

- (i) If $\sum_{i=1}^d |\pi_{t-}^i| < \frac{1}{c}$ a unique solution $\alpha^* \in [0, 1]$ exists.
- (ii) If $\sum_{i=1}^d |\pi_{t-}^i| \geq \frac{1}{c}$ there *may not exist a solution*, or at least none for $\alpha \in [0, 1]$. If a solution exists, $\alpha^* \leq 1$ and it *may not be unique*.

Remark 1.6.2. If we have long-only portfolios (and if $c < 1$ i.e. less than 100% transaction costs) only the first case of Proposition 1.6.1 is relevant. Hence, in that case one always has a unique solution $\alpha^* \in [0, 1]$. However, when we allow for short-selling, the second case of Proposition 1.6.1 becomes important.

We give a proof of Proposition 1.6.1 in Appendix 1.D and move directly to the interpretation of situation in the second case of Proposition 1.6.1.

Remark 1.6.3 (Interpretation of No or Non-Unique Solutions for α). We want to give some interpretation of the three possible outcomes of the second case of Proposition 1.6.1.

- (a) *no solution*: The trade is infeasible. In order to pay the transaction costs, we have to reduce the dollar amounts we invest. However, reducing the dollar amounts ψ_t leads to more transaction costs (because the difference to the previous investment amount grows, i.e. $|\psi_t - \psi_{t-}|$ increases). If the transaction costs grow faster than the money we gain from reducing the dollar amount ψ_t , the trade is infeasible.
- (b) *no solution in $[0, 1]$* : Again, if we need to pay transaction costs without a bank-account we have to free-up the money by selling stocks. Unlike case a) this might be feasible, but the amount of transaction costs to be paid is more than the total wealth of the portfolio. This leads to a solution with negative α .
- (c) *no unique solution in $[0, 1]$* : Here, we are in the case where the transaction costs do not exceed our total wealth. Since reducing the dollar amount leads to an increase in transaction costs, there might be several strategies with respect to dollar amounts invested leading to the desired portfolio weights after transaction costs. For example, one could do a more expensive trade (smaller ψ_t) and pay more transaction costs or a less expensive trade (larger ψ_t) where one pays less transaction costs.

In practice, we treat the above cases as follows. We consider cases a) and b) to lead to ruin, because both are caused by strategies which are infeasible given our wealth. If one of those cases occurs, we set our wealth to zero and terminate the investment. In case c) we choose the solution with highest α . This is what any reasonable investor would do: choose the trade with the least transaction costs.

1.7 Numerical results

In this section we present our numerical results using simulated and real market data.⁴

1.7.1 Simulated data

In this subsection, we aim to learn the signature portfolios which maximize $\mathbb{E}[\log \frac{W_T^\pi}{W_T^\mu}]$ in three financial market models respectively, by using simulations of those market models. This corresponds to the optimization task described in Theorem 1.5.11. To apply this in a numerical setting, we use the following Monte-Carlo optimization to approximate the expected values:

Corollary 1.7.1. *Let $\tilde{\Omega}$ be a probability-one set on which $\mathbf{Q}(t)$ is positive-semidefinite. Consider a Monte-Carlo type optimization of the expected log-relative wealth. Recall that $V_t^\pi = \frac{W_t^\pi}{W_t^\mu}$, hence we optimize for*

$$\max_{\pi} \frac{1}{M} \sum_{m=1}^M \log(V_t^\pi)(\omega_m)$$

for $\omega_1, \dots, \omega_m \in \tilde{\Omega}$. Again, this can be formulated as a convex quadratic optimization problem, namely

$$\min_{\mathbf{l}} \frac{1}{2} \mathbf{l}^T \hat{\mathbf{Q}}(t) \mathbf{l} - \mathbf{l}^T \hat{\mathbf{c}}(t), \quad (1.7.1)$$

where $\hat{\mathbf{Q}}(t) = \frac{1}{M} \sum_{m=1}^M \mathbf{Q}(t)(\omega_m)$ and $\hat{\mathbf{c}}(t) = \frac{1}{M} \sum_{m=1}^M \mathbf{c}(t)(\omega_m)$.

Proof. We have shown in the proof of Theorem 1.5.11 that $\mathbf{Q}(t)$ is almost surely positive semi-definite for all $t \in [t_0, T]$, hence such a set $\tilde{\Omega}$ exists. Since the sum of positive semi-definite matrices is positive semi-definite, $\hat{\mathbf{Q}}(t)$ is positive semi-definite and the statement follows. $\square$

Moreover, we will compare the performance of the learned portfolio with the growth-optimal portfolio of the respective markets. This is justified by Lemma 1.2.34, which states that the growth-optimal portfolio (which is also the numeraire portfolio) solves the relative log-utility optimization problem.

For each market, we simulate M_{train} trajectories over the time-horizon $[0, 1]$ of a given market model and consider investments over the whole time-horizon, i.e. $t_0 = 0$, $t = 1$. We

⁴The code corresponding to this section is available at <https://github.com/janka-moeller/Sig-SPT>

train a signature portfolio using the Monte-Carlo type optimization and the formulas for $\mathbf{Q}(t)(\omega)$ and the vector $\mathbf{c}(t)(\omega)$ provided in Theorem 1.5.11, where we use signature portfolios $\pi(\mu, \hat{\mu})$, i.e. choose $X = \mu$ to construct the portfolios. Note, that calculating the matrix $\mathbf{Q}(t)(\omega)$ and vector $\mathbf{c}(t)(\omega)$ for each training-sample can be parallellized and computed offline. Having obtained the respective $\hat{\mathbf{Q}}(t)$ and $\hat{\mathbf{c}}(t)$, we use the `gurobipy.model.optimize()`⁵ method to solve the convex quadratic optimization problem. We want to emphasize again, that we only need to solve the convex quadratic optimization task once and we never have to update $\hat{\mathbf{Q}}(t)$ and $\hat{\mathbf{c}}(t)$, which is very beneficial for the computational tractability of the optimization task.

After the optimization, we compare the weights and performance of the trained signature portfolio and the theoretical growth-optimal portfolio on M_{test} out-of-sample trajectories.

More concretely, we consider the following three market models and signature portfolios:

(i) **Correlated Black-Scholes Model:**

$$dS_t = \text{diag}(S_t)(adt + \Sigma dB_t),$$

where $a \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d \times d}$, B_t is a d -dimensional Brownian motion. We here take $d = 3$ and train a signature portfolio $\pi(\mu, \hat{\mu})$ of type I of degree three. We simulate 1'000 time-steps.

(ii) **Volatility Stabilized Market:**

$$\frac{dS_t^i}{S_t^i} = \frac{1 + \alpha}{2} \frac{1}{\mu_t^i} dt + \sqrt{\frac{1}{\mu_t^i}} dB_t^i, \quad 1 \leq i \leq d.$$

We choose again $d = 3$, so that B is a three-dimensional Brownian motion. We take $\alpha = 10$ and train a signature portfolio $\pi(\mu, \hat{\mu})$ of type I of degree three. We simulate 10'000 time-steps.

(iii) **Signature Market:**

$$dS_t = \text{diag}(S_t)(\mathbf{a}_t dt + \Sigma dB_t).$$

We choose again $d = 3$, so that B is a three-dimensional Brownian motion. Moreover, $(\mathbf{a}_t)_i = \sum_{0 \leq |I| \leq 2} \alpha_I^{(i)} \langle e_I, \hat{\mu} \rangle_t$, $\alpha_I^{(i)} \in \mathbb{R}$ with the restriction that $\alpha_I^{(i)} = 0$ if $I = (i_1, i_2)$ and $i_2 \neq 1$ and Σ is a constant 3×3 matrix. Here we train a signature portfolio $\pi(\mu, \hat{\mu})$ of type II of degree two and simulate 1'000 time-steps.

We use $M_{train} = 100'000$ in-sample trajectories for training and $M_{test} = 100'000$ out-of-sample trajectories for evaluation, i.e. computing the average logarithmic relative wealth of the respective portfolios on the test-samples. Furthermore, we used the bound constraints $|l_I^i| \leq 10'000$ for all I and for all $1 \leq i \leq d$ for each signature portfolio.

⁵for more information, see https://www.gurobi.com/documentation/9.1/refman/py_python_api_overview.html#sec:Python.

Remark 1.7.2. In a volatility stabilized market, signature portfolios $\pi(\mu, \hat{\mu})$ of type I , with $\hat{\mu} = (t, \mu)$ fall into the first case of Remark 1.5.12, where

$$d[\mu^i, \mu^j]_t = \begin{cases} \mu_t^i(1 - \mu_t^i)dt & \text{for } i = j \\ -\mu_t^i\mu_t^jdt & \text{for } i \neq j. \end{cases}$$

Remark 1.7.3. In Lemma 1.2.34 we have already established that the growth-optimal portfolio solves the relative log-utility optimization problem. We now want to outline that the growth-optimal portfolio exists *and* that the relative log-utility optimization for signature portfolios is well-posed, i.e. that the expectation of this optimization task is finite for all signature portfolios. Let us denote the growth-optimal portfolio by $\pi^{(g)}$, recall its form given by Lemma 1.2.33 and note that a sufficient condition for both of the above to be satisfied is if $|\mathbb{E}[\log W_T^{\pi^{(g)}}]| < \infty$. We will now present the arguments for the existence of $\pi^{(g)}$ and the well-posedness of the optimization problem for each of the markets.

- (i) Correlated Black-Scholes Model: It can easily be checked that for the growth-optimal portfolio $\pi^{(g)}$ it holds that $|\mathbb{E}[\log W_T^{\pi^{(g)}}]| < \infty$.
- (ii) Volatility Stabilized Market: Our optimization problem is well-posed for signature portfolios, because volatility stabilized market models are polynomial processes, see Cuchiero (2019) and all coefficients of $\mathbf{c}$ and $\mathbf{Q}$ are linear functions on the signature of $\hat{\mu}$, as established in Remarks 1.5.12 and 1.7.2. Moreover, we know that the truncated signature of a polynomial process is again a polynomial process (see Chapter 2) and hence their expectation is finite. Therefore, the expectation of any coefficient of $\mathbf{c}$ and $\mathbf{Q}$ is finite and $|\mathbb{E}[\frac{W_T^{\pi}}{W_T^{\mu}}]| < \infty$ holds for any signature portfolio π of type I .
- (iii) Signature Market: Assuming that a solution exists, the coefficients $\mathbf{a}_i$ are all uniformly bounded on $[0, 1]$. This follows from the fact that μ^i are positive and uniformly bounded for all $i = 1, \dots, d$, hence so is $\int_0^T \mu_t^i dt$. It thereby follows $|\mathbb{E}[\log W_T^{\pi^{(g)}}]| < \infty$ for the growth-optimal portfolio ρ .

Remark 1.7.4. Note that the growth-optimal portfolios of the Black-Scholes and volatility stabilized markets are also functionally generated portfolios in the classical sense. Indeed, in the Black-Scholes case the portfolio generating function is

$$G^{(BS)}(\mu_t) = \prod_{i=1}^d (\mu_t^i)^{c_i},$$

with portfolio weights $\pi_t^{(BS),i} = c_i$, where

$$c_i = ((\Sigma^T \Sigma)^{-1} \mathbf{a})_i - \sum_{j=1}^d ((\Sigma^T \Sigma)^{-1})_{ij} \left(\frac{\sum_{k=1}^d (\Sigma^T \Sigma)^{-1} \mathbf{a}_k - 1}{\sum_{j,k=1}^d ((\Sigma^T \Sigma)^{-1})_{jk}} \right).$$

In the volatility stabilized market case the generating function of the growth optimal portfolio is given by

$$G^{(Vol)}(\mu_t) = \prod_{i=1}^d (\mu_t^i)^{\frac{1+\alpha}{2}} \exp \left(\mu_t^i \frac{d}{2} (\alpha - 1) \right),$$

Mean log-relative wealth of	growth-optimal portfolio	signature portfolio
Black-Scholes market:	9.0115	9.0122
Volatility stabilized market:	8.7619	8.7417
Signature market:	0.4399	0.4398

Table 1.1: Mean logarithmic relative wealth of the theoretical growth-optimal and signature portfolios, evaluated on 100'000 test-samples, in each market respectively.

with portfolio weights

$$\pi_t^{(Vol),i} = \frac{1 + \alpha}{2} + \mu_t^i \frac{d}{2}(\alpha - 1).$$

Therefore, these two examples are not only examples of signature portfolios approximating the respective growth-optimal portfolios but also examples for signature portfolios approximating functionally generated portfolios.

Figure 1.1 shows the trained signature portfolios (right) and the growth-optimal portfolios (left) evaluated at one out-of-sample trajectory in each market respectively (rows). As expected from the theoretical results, the signature weights are very similar to the theoretical growth-optimal weights in all markets. We want to emphasize, that the growth-optimal weights were never shown to the signature portfolio during training, but the signature portfolio was trained to maximize the expected logarithmic relative wealth.

Furthermore, we want to highlight, that although the growth-optimal portfolio in the Black-Scholes market has constant weights, this approximation task is far from trivial for a signature portfolio of type I . Because signature portfolios of type I approximate the controlling functions of the growth-optimal portfolio, the approximation task was actually

$$f^{(BS),i}(t, \mu_{[0,t]}) \approx \frac{c_i}{\mu_t^i}.$$

And likewise, in the volatility stabilized market, the approximation task was

$$f^{(Vol),i}(t, \mu_{[0,t]}) \approx \frac{\alpha + 1}{2\mu_t^i} + \frac{d}{2}(\alpha - 1).$$

We quantify the performance by comparing the average logarithmic relative wealth over the M_{test} test samples of the growth-optimal and signature portfolios, where in both cases we compute the average logarithmic relative wealth numerically on the test samples. This explains why the signature portfolio sometimes even leads to higher values. We present these results in Table 1.1.

1.7.2 Real market data

Details on the optimization and investment procedure

Before we present the performance of signature portfolios, as well as, JL- and randomized-signature portfolios in real markets, we want to give some details on the optimization and investment procedure.

Approximating expectations by time-averages When working with real market data, we only have one realization available. Therefore, as already discussed in Remark 1.5.10 and at the beginning of Section 1.5.2 we replace the expectation by a time-average. More precisely, consider $[0, \Delta, 2\Delta, \dots, N\Delta = t]$ to be a equally spaced grid of $[0, t]$ with distance Δ . Then, for $s \in [0, t - \Delta]$ and analogously to Remark 1.5.10 we consider the following approximations:

$$\mathbb{E} \left[\log \left(\frac{V_{s+\Delta}^\pi}{V_s^\pi} \right) \right] \approx \frac{1}{N} \sum_{i=1}^N \log \left(\frac{V_{i\Delta}^\pi}{V_{(i-1)\Delta}^\pi} \right) = \frac{1}{N} \log(V_t^\pi)$$

$$\mathbb{E}[\mathcal{R}_{s,s+\Delta}^{V,\pi}] \approx \frac{1}{N} \sum_{i=1}^N \mathcal{R}_{(i-1)\Delta, i\Delta}^{V,\pi} =: \bar{\mathcal{R}}^{V,\pi},$$

$$\text{Var}[\mathcal{R}_{s,s+\Delta}^{V,\pi}] \approx \frac{1}{N-1} \sum_{i=1}^N (\mathcal{R}_{(i-1)\Delta, i\Delta}^{V,\pi} - \bar{\mathcal{R}}^{V,\pi})^2.$$

Recall that sufficient ergodicity/stationarity conditions guaranteeing that these approximations are justified have been discussed in Remark 1.5.10.

Note however that we cannot expect that stationarity of the relative returns and log-relative returns is satisfied for every portfolio in our optimization class. Consider for example signature portfolios of type I with $\tau = (\frac{1}{d}, \dots, \frac{1}{d})^\top$, $\hat{X}_t = (t, X_t)$ and $\mathcal{U} = \{1, \dots, d\}$. It is easy to see that the stationarity assumption can not hold for all portfolios in this class, just by considering the following portfolio π^1 where the only parameters that are non-zero are chosen to be $l_{(1)}^d$. Then

$$\mathbb{E}[\mathcal{R}_{t,t+\Delta}^{V,\pi^1}] = \frac{1}{d} \mathbb{E} \left[t \cdot \frac{\mu_{t+\Delta}^d}{\mu_t^d} \right] + \left(\frac{1}{d} - \frac{t}{d^2} \right) \mathbb{E} \left[\sum_{j=1}^d \frac{\mu_{t+\Delta}^j}{\mu_t^j} \right] - \frac{t+1}{d}.$$

However, there are of course classes of linear path-functional portfolios where the stationarity assumptions holds for all portfolios in the class. Indeed, the simplest example is the class of Markowitz-type portfolios, i.e. with τ being constant and constant portfolio maps. A far-reaching generalization thereof are signature portfolios with *rolling windows*, as defined in Remark 1.3.9 and for which stationarity holds under the conditions of Remark 1.5.10. Computing these portfolios is slightly more expensive, since one needs to compute the increments of the signature. Therefore we did not consider them in our implementations.

Finally, let us emphasize that the optimization using time-averages does make sense in practice, even if the stationarity of returns may not hold for every portfolio. For example consider $\Delta = 1$ day, then the mean-variance optimization using time-averages can be seen as looking for a portfolio with high average daily returns (in time) but without the daily returns varying too much over time and likewise for the log-relative wealth optimization. Moreover, we argue that the relative returns and log-relative returns of our optimized portfolios do exhibit some "stability" in time, as almost all the learnt portfolios perform very well in the out-of-sample period, as we will present in the following. Note, that all of the above optimization problems remain convex quadratic optimization problems if we replace expectations by time-averages.

Optimization under transaction costs In practice, we re-balance our portfolio-weights once a day since our market data is also of daily frequency. In this setting, we would like to incorporate transaction costs during optimization.

As explained in detail in Section 1.6, including transaction costs is not trivial in our setting. In particular, since we have shown that (1.6.1) does not necessarily have a (unique) solution, we cannot aim for any closed-form. Therefore, incorporating the exact amount of transaction-costs to be paid during the optimization would render the optimization task infeasible. However, we propose the following penalization which preserves the form of a convex quadratic optimization problem. Moreover, our empirical results in Subsection 1.7.2 verify that this penalization is effective in all cases and indeed useful for learning portfolios which perform well even under 5% of transaction costs.

Corollary 1.7.5. *Given a convex quadratic optimization problem of the form*

$$\min_{\mathbf{l}} \frac{1}{2} \mathbf{l}^\top \mathbf{Q}(T) \mathbf{l} + \mathbf{c}(T)^\top \mathbf{l}$$

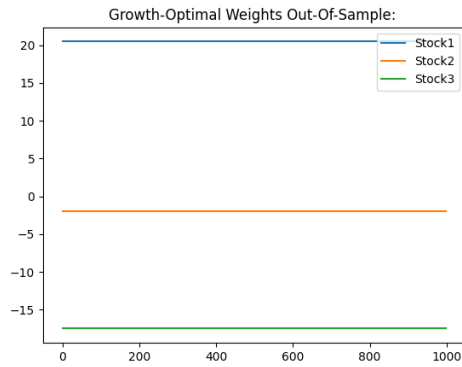
adding the penalization for linear path-functional portfolios π

$$+ \frac{\beta}{T} \sum_{t=0}^{T-1} \sum_{i \in \mathcal{U}} \left(\frac{\pi_{t+1}^i}{\mu_{t+1}^{\mathcal{U},i}} - \frac{\pi_t^i}{\mu_t^{\mathcal{U},i}} \right)^2$$

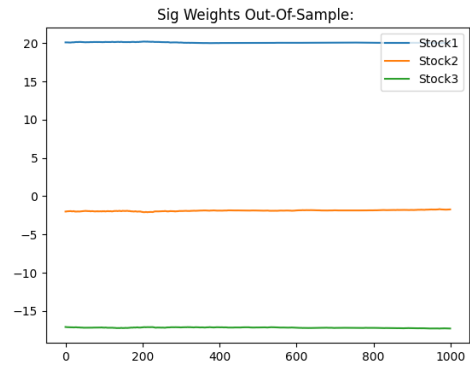
preserves the form of a convex quadratic optimization problem. Here, β is a hyperparameter that has to be chosen appropriately.

Proof. Clearly the penalization is quadratic in π and hence quadratic in the parameters $\{l_\nu^i\}_{i \in \mathcal{U}, \nu \in \mathcal{V}}$. Moreover, the penalization is positive for all $\{l_\nu^i\}_{i \in \mathcal{U}, \nu \in \mathcal{V}}$ and hence the convexity is preserved as well. $\square$

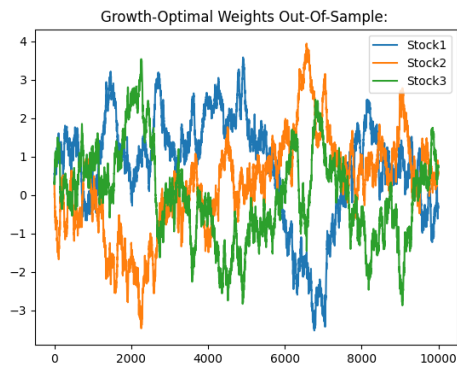
The motivation for this penalization is, first of all, that the universe portfolio $\mu^{\mathcal{U}}$ is not punished, which should be the case because this portfolio has no transaction costs at all. Moreover, the penalization punishes changes in the weights which exceed changes in the market weights. Note that those are exactly the changes that lead to transaction costs.



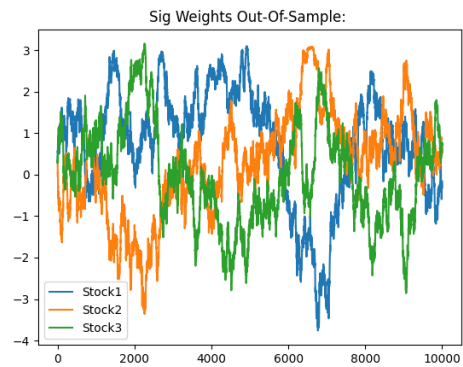
(a) Black-Scholes market: growth-optimal weights



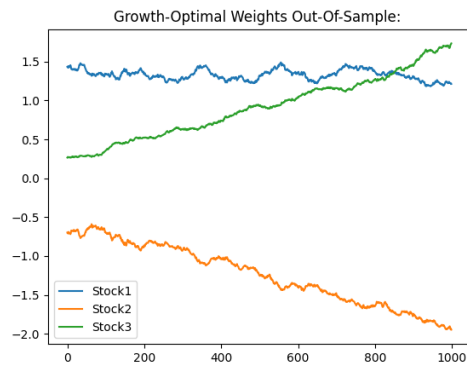
(b) Black-Scholes market: signature weights



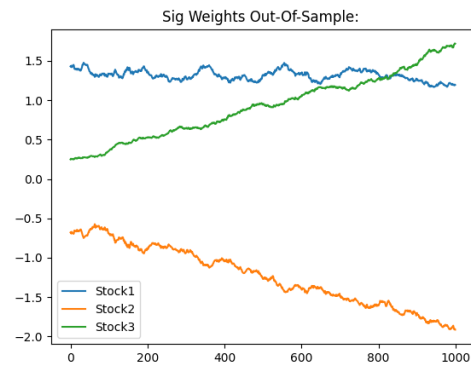
(c) Vol. stabilized market: growth-optimal weights



(d) Vol. stabilized Market: signature weights



(e) Signature market: growth-optimal weights



(f) Signature market: signature weights

Figure 1.1: The theoretical growth-optimal weights (left) and the signature portfolio's weights (right) for the Black-Scholes, volatility stabilized and signature market respectively, evaluated at one test sample.

Algorithm 1 Optimization Under Transaction Costs

```

1: function OPTIMIZE_UNDER_TC(beta):
2:    $l \leftarrow \text{SOLVE\_CONVEX\_QUADRATIC}(\text{beta})$ 
3:    $\text{weights} \leftarrow \text{GET\_INSAMPLE\_WEIGHTS}(l)$ 
4:    $\text{wealth} \leftarrow \text{WEALTH\_UNDER\_TC}(\text{weights})$ 
5:   return  $-1 \cdot \text{wealth}$ 
6: end function
7: while  $(-1 \cdot \text{OPTIMIZE\_UNDER\_TC}(\text{beta}_0)) < 10^{-4}$  do
8:    $\text{beta}_0 += 0.5$ 
9: end while
10:  $\text{optimal\_beta} \leftarrow \text{MINIMIZE}(\text{OPTIMIZE\_UNDER\_TC}, \text{beta}_0)$ 
11:  $\text{optimal\_l} \leftarrow \text{SOLVE\_CONVEX\_QUADRATIC}(\text{optimal\_beta})$ 

```

The downside of this penalization is, that for a given level of transaction costs, we do not know how to choose β . To find an appropriate β , we propose the procedure described in Algorithm 1. That is, we solve the quadratic optimization problem for a given β and calculate the in-sample performance *under the true transaction costs* for the found optimal portfolio. We then minimize over β in order to find the one which best penalizes for transaction costs over the in-sample period. Note that choosing the initial value β_0 can be delicate. Namely, if in a neighbourhood around β_0 all optimal strategies lead to ruin under transaction costs, one may not move away from the initial value. Therefore, we test if β_0 leads to ruin under transaction cost and, if so, increase it, in Algorithm 1. Note that the condition in Line 8 must be false for some β_0 because the universe portfolio itself is included in the set of portfolios we optimize over.

The optimization problems In the following, we consider optimization problems of various signature-type linear path-functional portfolios with L^2 -regularization, with and without a regularization for transaction costs. The two types of optimization problems are maximizing the expected log-relative wealth

$$\max_{\{l_\nu^{(i)}\}_{i \in \mathcal{U}, \nu \in \mathcal{V}}} \mathbb{E} \left[\log \left(\frac{W_t^\pi}{W_t^\mathcal{U}} \right) \right] - \gamma_{\text{LO}} \sum_{\substack{i \in \mathcal{U} \\ \nu \in \mathcal{V}}} (l_\nu^{(i)})^2 - \frac{\beta}{T} \sum_{t=0}^{T-1} \sum_{i \in \mathcal{U}} \left(\frac{\pi_{t+1}^i}{\mu_{t+1}^{\mathcal{U},i}} - \frac{\pi_t^i}{\mu_t^{\mathcal{U},i}} \right)^2 \quad (\text{Log-Opt.})$$

and maximizing mean-variance

$$\max_{\{l_\nu^{(i)}\}_{i \in \mathcal{U}, \nu \in \mathcal{V}}} \lambda \mathbb{E} \left(\mathcal{R}_{t,t+\Delta}^{V,\pi} \right) - \text{Var} \left(\mathcal{R}_{t,t+\Delta}^{V,\pi} \right) - \gamma_{\text{MV}} \sum_{\substack{i \in \mathcal{U} \\ \nu \in \mathcal{V}}} (l_\nu^{(i)})^2 - \frac{\beta}{T} \sum_{t=0}^{T-1} \sum_{i \in \mathcal{U}} \left(\frac{\pi_{t+1}^i}{\mu_{t+1}^{\mathcal{U},i}} - \frac{\pi_t^i}{\mu_t^{\mathcal{U},i}} \right)^2 \quad (\text{MV-}\lambda)$$

where in both cases we added bound constraints $|l_\nu^{(i)}| \leq 10'000$ for all $i \in \mathcal{U}, \nu \in \mathcal{V}$.

Hyperparameters and cross-validation The main hyperparameters of our optimization procedure are:

- t_0 : this is the time at which we start to invest, with respect to the time when we start calculating the signature. The interpretation of this is that we can observe the

market for some time before we start to invest, which is particularly relevant in the non-Markovian setting. In the following we choose $t_0 = 100$, hence, we always start calculating the signature 100 days before the respective investment period starts.

- γ : this is the parameter of the L_2 -regularization described in Proposition 1.5.17. For the case of maximizing the (expected) log-relative wealth, we choose our γ during cross-validation. We do not train this regularization parameter in the mean-variance optimization but fix it apriori to a small value, as the minimization of the variance itself can already be regarded as regularizing.

For performing the cross-validation, we first split our data into a in-sample-period of T_{ins} days, a cross-validation of T_{cv} (consecutive to the in-sample period) and a testing period of T_{test} days, starting at the end of the cross-validation period. Once we found the optimal hyper-parameters during cross-validation, we learn the parameters of our portfolios again on the T_{ins} days prior to the testing period (using the obtained hyper-parameters). We then evaluate the performance of the learned portfolios during the testing-period.

Calculating the JL-signature In order to calculate the JL-signature, the true truncated signature needs to be calculated before applying the JL-projection. However, calculating the true signature can be computationally too heavy in large markets. We therefore propose the following memory-efficient procedure (Algorithm 2), which is based on the observation that for each component of the signature at level l , at most l components of the underlying path appear in the integrals. Hence, one can instead compute the signature of combinations of l components of the path and apply the projection "batch-wise". It is important to note that by doing so, some words are computed multiple times, for example

$$\int_0^t \int_0^s \circ dX_u^1 \circ dX_s^1$$

can arise from the combination (X^1, X^2, X^3) and from (X^1, X^4, X^5) and from many more. Therefore, one needs to be careful which words to keep and our proposed algorithm takes care of that. Moreover, it is important to store the random projection matrix, once a realization is computed.

Parametrization of time-components We parameterize the time-component of the time augmentation in the following way: Let $\hat{X}_t = (\varphi_T(t), X_t)$. For a given trading horizon T_{hor} , we consider the parametrization $\varphi_{T_{hor}}(t) = \frac{t}{T_{hor}}$. Concretely, for the in-sample period of 2000 days, we set $T_{hor} = 2000$ and for the out-of-sample period we set $T_{hor} = 750$. This time-augmentation therefore contains information about the amount of time that is left (or has passed) in the current trading period. This is to compensate for the different training and testing periods. Another way to deal with this would be to use signature portfolios with rolling windows as defined in Remark 1.3.9.

Algorithm 2 Calculate JL-Signature

```

1: function JL_SIGNATURE(path, level):
2:   Initialize A_list=list()
3:   Initialize JL_sig=0
4:   for  $l \in \{1, \dots, \text{level}\}$  do
5:     combos  $\leftarrow$  COMBNATIONS(dim(path), l)
6:     for  $c \in \text{combos}$  do
7:       A_slice  $\sim N(0, 1/k)$ 
8:       sig, words  $\leftarrow$  SIGNATURE(path[c], level)
9:       words_to_keep= list()
10:      for word  $\in$  words do
11:        if length(set(word))== l then
12:          words_to_keep.append(word)
13:        end if
14:      end for
15:      sig= sig[words_to_keep]
16:      A_slice= A_slice[ : , words_to_keep]
17:      JL_sig+=matrix_product(A_slice, sig)
18:    end for
19:  end for
20:  return JL_sig
21: end function

```

Performance metrics Recall that we re-balance our portfolios once a day and follow a buy-and-hold strategy in between. We evaluate the out-of-sample performance of portfolios by their log-relative wealth

$$\log(V_T^{b\&h}) = \log\left(\prod_{t=t_0}^{T-1} \sum_{i \in \mathcal{U}} \pi_t^i \frac{\mu_{t+1}^{\mathcal{U},i}}{\mu_t^{\mathcal{U},i}}\right)$$

and their log-wealth

$$\log(W_T^{b\&h}) = \log\left(\prod_{t=t_0}^{T-1} \sum_{i \in \mathcal{U}} \pi_t^i \frac{S_{t+1}^{\mathcal{U},i}}{S_t^{\mathcal{U},i}}\right).$$

Survivorship bias In our name-based approach we only consider a universe of stocks to invest in which were present in the market for the entire in- and out-of-sample period. This opens up the issue of survivorship bias. Note that the same holds true for the ranked NASDAQ market, since there was always a 100-largest stock present, even though the company holding that rank may have changed over time. As a first measure to reduce this bias, we compare the performance of our trained portfolios only to the universe portfolio of the surviving stocks and not to the entire market portfolio. Nevertheless, our trained portfolios may benefit from the fact that stocks do not go bankrupt in other ways which the universe portfolio can not exploit. Let us address them in more detail:

- **Leverage:** If bankruptcy of stocks is not possible, leverage (i.e. short-selling of some stocks in order to invest more heavily in others) is less risky. We do not impose any

Portfolio Reg. TC	Signature Portfolio			JL-Signature Portfolio			Rand.-Signature Portfolio		
	0%	1%	5%	0%	1%	5%	0%	1%	5%
Log-Opt.	-0.5	-0.45	-0.13	-0.32	-0.32	-0.12	-0.01	-0.01	$3 \cdot 10^{-8}$
MV $\lambda = 1$	-3.31	-1.45	-0.16	-3.70	-1.67	-0.18	-3.37	-2.26	-0.20
MV $\lambda = 0.75$	-2.43	-1.72	-0.16	-2.72	-2.10	-0.18	-2.48	-2.22	-0.21
MV $\lambda = 0.5$	-1.56	-1.54	-0.17	-1.75	-1.75	-0.19	-1.59	-1.54	-0.22
MV $\lambda = 0.05$	-0.05	-0.05	-0.04	-0.05	-0.05	-0.04	-0.04	-0.04	-0.03

(a) Minimum portfolio weight observed during the out-of-sample period for the SMI universe for the respective portfolios with and without regularization for transaction costs.

Portfolio Reg. TC	JL-Signature Portfolio			Rand.-Signature Portfolio		
	0%	1%	5%	0%	1%	5%
Log-Opt.	-1.67	-0.66	-0.11	-0.80	-0.63	-0.15
MV $\lambda = 1$	-0.14	-0.14	-0.11	-0.04	-0.04	-0.04
MV $\lambda = 0.75$	-0.10	-0.10	-0.10	-0.03	-0.03	-0.03
MV $\lambda = 0.5$	-0.06	-0.06	-0.06	-0.01	-0.01	-0.01
MV $\lambda = 0.05$	$8 \cdot 10^{-5}$	$8 \cdot 10^{-5}$	$8 \cdot 10^{-5}$	$8 \cdot 10^{-5}$	$8 \cdot 10^{-5}$	$8 \cdot 10^{-5}$

(b) Minimum portfolio weight observed during the out-of-sample period for the S&P500 universe for the respective portfolios with and without regularization for transaction costs.

Table 1.2: The tables show the minimum portfolio weights observed during the out-of-sample period of the respective universe. The ones which are positive (i.e. no short-selling) are highlighted in bold.

leverage constraints in our optimization. Nevertheless we do not observe extreme short-selling during the out-of-sample period. We report the minimum observed out-of-sample weights observed in Tables 1.2a and 1.2b, where we highlight those in bold which are long-only. At least for those which do not short-sell we can exclude leveraging as an unfair advantage. In the cases where only small negative weights were observed, we dare to say that an unfair advantage through leveraging is unlikely to have been the driving factor for potential out-performance of the universe portfolio. Aside from leveraging, we would like to point out that regularization for transaction costs reduces short-selling.

- **Over-weighting small stocks:** If stocks' capitalization cannot go to zero, a potential strategy may be to put a lot of weight on stocks with capitalizations close to zero, because the upside is much higher than the downside. To investigate this, we point to Figures 1.3c and 1.3d, which show the average out-of-sample weights of the ranked NASDAQ universe. We observe that the trained portfolios *underweight* small stocks on average compared to the universe portfolio.

Choice of a benchmark In the following numerical experiments we will use the universe weights as a auxiliary portfolio τ . Recall from Remark 1.3.7 that the auxiliary portfolio τ is the natural benchmark for linear path-functional portfolios because it can always be attained by them. As an additional benchmark we included the equally-weighted portfolio. We want

Optimization Task	JL-Signature		Randomized Signature	
	$\log V_T^{b\&h}$	$\log W_T^{b\&h}$	$\log V_T^{b\&h}$	$\log W_T^{b\&h}$
Log-Opt.	0.1045	0.6622	0.0002	0.5579
MV $\lambda = 4$	0.3395	0.8972	-0.0045	0.5531
MV $\lambda = 3$	0.3173	0.8750	0.0081	0.5658
MV $\lambda = 2$	0.2531	0.8108	0.0131	0.5707
MV $\lambda = 1$	0.1472	0.7049	0.0104	0.5680
MV $\lambda = 0.75$	0.1143	0.6720	0.0085	0.5662
MV $\lambda = 0.5$	0.0788	0.6364	0.0061	0.5638
MV $\lambda = 0.05$	0.0083	0.5660	0.0007	0.5584

Table 1.3: Performance of the trained portfolios over the out-of-sample period in terms of log-(relative)-wealth. We ran the mean-variance optimization task for seven different risk-tolerances λ . All but one of the optimized portfolios out-performed the universe portfolio over the out-of-sample period.

to emphasise that of course other auxiliary portfolios and hence benchmark portfolios could be used.

Rank-based approach: NASDAQ

In this part of our empirical analysis we study a rank-based approach to portfolio optimization. We consider the *ranked* NASDAQ market and choose as a universe the stocks with ranks $1, \dots, 100$. We obtained the data from the CRSP database.⁶ We train JL-signature portfolios of dimension $(50, 3)$ of type I and randomized signature portfolios of dimension 50 of type I investing in the universe of the 100 largest stocks and use as input the universe weights i.e. $\hat{X} = \hat{\mu}^U$ and the universe portfolio as the auxiliary portfolio (i.e. $\tau = \mu^U$). The portfolio is daily re-balanced. We train such portfolios in two ways, once in the log-wealth optimization and once in the mean-variance optimization, where in both optimization problems, we measure the performance of the portfolios by the log-relative wealth and the log-wealth achieved in the out-of-sample period.

We choose $t_0 = 100$ and take as an in-sample period 2000 trading days and as an out-of-sample period the following 750 trading days. For the mean-variance optimization task we set the L^2 -regularization parameter $\gamma_{MV} = 10^{-6}$ and for the log-wealth optimization task we found $\gamma_{LO} = 4.849 \cdot 10^{-3}$ for JL-signature portfolios and $\gamma_{LO} = 1 \cdot 10^{-2}$ for randomized-signature portfolios during the cross-validation. The grid-search for γ_{LO} was performed over 100 equally-distant points in $[10^{-6}, 10^{-2}]$.

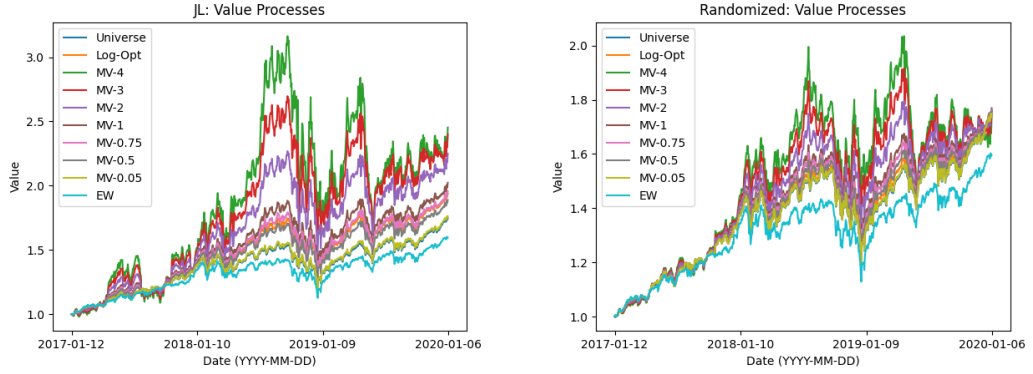
We report the out-of-sample performance of the trained portfolios in Table 1.3 in terms of log-(relative)-wealth with respect to the universe portfolio. All but one of the trained portfolios outperform the universe portfolio.

In Figure 1.2a and Figure 1.2b we display the wealth processes of the trained portfolio, the one of the universe portfolio and as a benchmark also of the equally-weighted portfolio.

⁶The raw/processed data required to reproduce our findings cannot be shared at this time due to legal reasons.

The wealth processes of the mean-variance portfolio are the more volatile, the higher the risk-tolerance is, as one would expect. Moreover, the wealth-processes of the randomized signature portfolios are more tamed than their JL-counterparts.

We present the average values (in time) of the portfolio weights for each rank in Figure 1.3. The trained portfolios mainly take over- and under-weighted positions in the largest stocks. While the positions are extremer, the higher the risk-tolerance. Although we do not enforce any long-only constraints or regularization, we do not observe any extreme short-selling positions.



(a) Wealth processes of the learned JL-signature portfolios. (b) Wealth processes of the randomized signature portfolios.

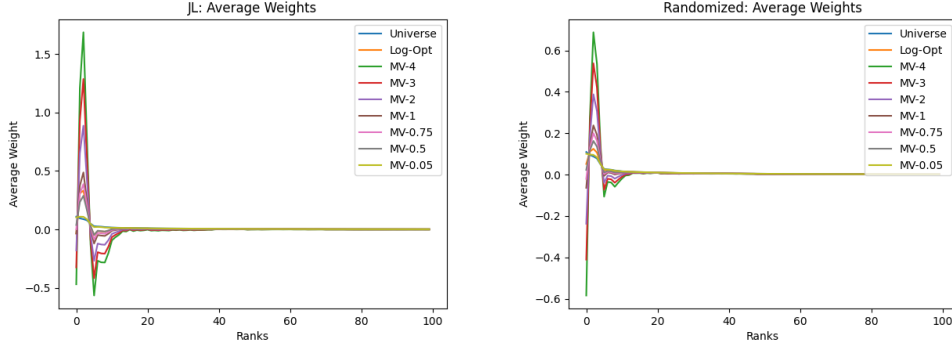
Figure 1.2: Wealth processes of NASDAQ the learned signature portfolios, the universe portfolio and the equally-weighted portfolio, over the out-of-sample period.

Name-based approach: SMI and S&P500

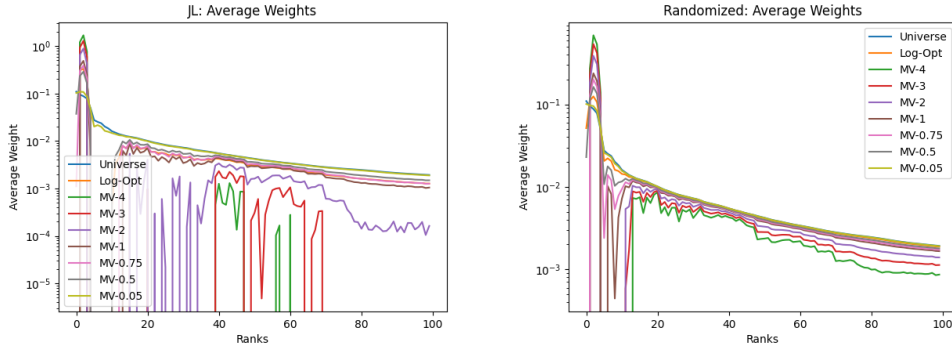
In the name-based setting we tackle the mean-variance and log-relative wealth optimization problems under transaction costs. We do this in two markets; the Swiss Market Index (SMI) and S&P500 Index. For the SMI, we consider the universe of stocks which survived between 2000-2022 and for the S&P500 those that survived between 2001-2022. This amounts to 17 stock in the SMI universe and to 378 stocks in the S&P500 universe. For both markets, we obtained the data from Reuters Datastream.⁷ As inputs and as auxiliary portfolio we use the respective universe weights, i.e. $X = \tau = \mu^U$. We compare the performance of the trained portfolios with the universe portfolio. Again, we choose $t_0 = 100$, an in-sample period of 2000 days and an out-of-sample period of 750 days. When we include the regularization cost, we choose $\beta_0 = 0.5$ and enforce a lower bound $\beta \geq 10^{-8}$. We want to emphasize, that even with transaction costs our portfolios are re-balanced daily.

⁷The raw/processed data required to reproduce our findings cannot be shared at this time due to legal reasons.

1.7 Numerical results



(a) Average weights of the JL-signature portfolios. (b) Average weights of the randomized signature portfolios.



(c) Plot of the logarithm of the average weights of the JL-signature portfolios. (d) Plot of the logarithm of the average weights of the rand.-signature portfolios.

Figure 1.3: Average weights in the NASDAQ of the signature portfolios and the universe portfolio, averaged over the out-of-sample period.

SMI market We train signature portfolios $\pi(\mu^{\mathcal{U}}, \hat{\mu}^{\mathcal{U}})$ of type I with three configurations of feature maps

- the true signature up to degree two;
- the JL-signature of dimension (30,2);
- the randomized signature of dimension 30.

We set $\gamma_{MV} = 10^{-8}$ and found $\gamma_{LO} = 2.02 \cdot 10^{-3}$ for the true signature, $\gamma_{LO} = 2.42 \cdot 10^{-3}$ for the JL-signature and $\gamma_{LO} = 1 \cdot 10^{-2}$ for the randomized signature respectively during cross-validation, where the grid-search was performed over 100 equally-distant points in $[10^{-8}, 10^{-2}]$. We present the out-of-sample performance in terms of log-relative wealth in Tables 1.4a, 1.4b, 1.4c as well as the annualized Sharpe-ratios in Table 1.5a. The first three columns of Tables 1.4a, 1.4b, 1.4c show the performance of portfolios with and without transaction costs, where we added no regularization for transaction costs. The next four

columns show the performance with and without transactions costs, but with a regularization for transaction costs at the respective level. The corresponding regularization parameters β which we found during the in-sample training are shown in Table 1.6. We note that the lower the risk-tolerance λ the less regularization for transaction costs is needed. However, adding such a regularization, the signature portfolios trained under the mean-variance optimization often outperform the universe portfolio under transaction costs during the out-of-sample period. This is remarkable, since the universe portfolio does not pay any transaction costs. The portfolio trained under log-relative wealth optimization does not out-perform the universe portfolio under transaction costs for the true and JL-signature portfolios and neither do some of the portfolios trained using the JL- and randomized signature under mean-variance optimization with higher risk-tolerances. Nevertheless, the regularization for transaction cost proved effective, since all portfolios performed better with the regularization than without, under the respective level of transaction costs.

We show the wealth processes of the signature portfolios for the SMI universe in Figure 1.4. It is obvious that the higher λ is, the more volatile the portfolios are, which lead to some signature portfolios under-performing the universe portfolio under transaction costs during some parts of the out-of-sample period. The signature portfolio with $\lambda = 0.05$, however, did rather well over the entire period even under transaction costs, which we highlight in Figures 1.4d & 1.4e.

Turning to the Sharpe-ratios in Table 1.5a we see that while the Sharpe-ratios of the optimized portfolios are often higher than the one of the universe portfolio without transaction costs, the universe portfolio is less often out-performed under transaction costs in terms of Sharpe-ratios. Only the mean-variance optimized portfolio with $\lambda = 0.05$ achieves a consistently higher Sharpe ratio than both universe and equally-weighted portfolio. It is also worth highlighting, that often the Sharpe-ratio increases for higher transaction costs (and thereby higher regularization for transaction costs).

S&P500 market We trained JL-signature portfolios $\pi(\mu^{\mathcal{U}}, \hat{\mu}^{\mathcal{U}})$ of dimension (30,2) and type I as well as randomized-signature portfolios $\pi(\mu^{\mathcal{U}}, \hat{\mu}^{\mathcal{U}})$ of dimension 30 and type I in the S&P500 market. For the mean-variance optimization, we set $\gamma_{MV} = 10^{-6}$ and for obtained $\gamma_{LO} = 1.02 \cdot 10^{-4}$ for both the JL-signature portfolio and the randomized signature portfolio during cross-validation over an equally-spaced grid of 100 points in $[10^{-6}, 10^{-2}]$. We report the out-of-sample results of the optimized portfolios in Table 1.4d & 1.4e for the JL- and randomized-signature portfolios respectively. It is remarkable that many portfolios also out-performed the universe portfolio without any regularization for transaction costs. Indeed, in many cases the optimization yielded $\beta = 10^{-8}$, as we report in Table 1.6. In the cases where such a regularization was needed, the performance with transaction costs was always improved with the regularization. Apart from the log-wealth optimized portfolio under 5% of transaction costs, the learned portfolio all out-performed the universe portfolio under transaction costs. We show the corresponding wealth processes in Figure 1.6.

The out-of-sample annualized Sharpe-ratios are reported in Table 1.5b. The mean-variance optimized portfolios manage to out-perform the universe and equally-weighted portfolios also in terms of Sharpe-ratios, even under transaction costs. Only the log-wealth optimized

portfolio fails to achieve a higher Sharpe-ratio than the universe portfolio under transaction costs.

Conclusion We would like to offer some conclusion on the performance of our optimized portfolios in particular comparing the signature, JL-signature and randomized signature portfolios. We highlight that computing the true signature portfolio of the S&P500 universe would not have been feasible, since the number of stocks would have been too high. However, the JL-signature portfolios did consistently achieve a better out-of-sample performance both in terms of log-relative wealth and Sharpe-ratios than the randomized signature portfolios. For the SMI universe we are impressed with the performance of the JL- and randomized-signature portfolios compared to those of the true signature portfolio, especially due to the vast difference in number of optimization parameters (510 for JL- & rand.-signature each vs. 10'290 for the true signature portfolios!). In terms of Sharpe-ratio the JL- and randomized-signature portfolios even slightly out-performed the true signature portfolios, but no significant difference between the two randomization approaches was observable. In terms of log-relative wealth, the true signature portfolio performed better than the JL-signature portfolios, which in turn achieved better results than the randomized-signature ones, however we would like to point out that the differences were small. Overall, in terms of feasibility and performance, we consider the JL-signature portfolios to be the most favorable.

Opt. Task Prop. TC	log V_T without Reg.			log V_T with Reg. $\beta_{1\%}$		log V_T with Reg. $\beta_{5\%}$	
	0%	1%	5%	0%	1%	0%	5%
Log-Opt.	0.0458	-0.0494	-0.4333	0.0461	-0.0293	0.0540	-0.0158
MV $\lambda = 1$	1.7896	-1.9354	$-\infty$	1.0514	0.1760	0.0257	0.2804
MV $\lambda = 0.75$	1.4767	-0.6214	-14.6792	1.1547	0.0601	0.2724	0.0247
MV $\lambda = 0.5$	1.0748	0.1155	-4.3536	1.0539	0.1409	0.2620	0.0220
MV $\lambda = 0.05$	0.1240	0.0973	-0.0096	0.1229	0.0979	0.1056	0.0381

- (a) Log-relative wealth of the signature portfolios with and without transaction costs trading in the SMI universe.

Opt. Task Prop. TC	log V_T without Reg.			log V_T with Reg. $\beta_{1\%}$		log V_T with Reg. $\beta_{5\%}$	
	0%	1%	5%	0%	1%	0%	5%
Log-Opt.	0.0620	0.0093	-0.2009	0.0619	0.0094	0.0420	-0.0201
MV $\lambda = 1$	1.6656	-1.6336	$-\infty$	0.9654	0.1402	0.2061	0.0187
MV $\lambda = 0.75$	1.3676	-0.4919	-12.2094	1.1439	-0.0583	0.2051	0.0178
MV $\lambda = 0.5$	0.9913	0.1398	-3.7521	0.9913	0.1398	0.2074	0.0144
MV $\lambda = 0.05$	0.1137	0.0894	-0.0072	0.1137	0.0894	0.0947	0.0254

- (b) Log-relative wealth of the JL-signature portfolios with and without transaction costs trading in the SMI universe.

Opt. Task Prop. TC	log V_T without Reg.			log V_T with Reg. $\beta_{1\%}$		log V_T with Reg. $\beta_{5\%}$	
	0%	1%	5%	0%	1%	0%	5%
Log-Opt.	0.0397	0.0331	0.0070	0.0396	0.0331	0.0217	0.0050
MV $\lambda = 1$	1.4602	-1.3840	$-\infty$	0.9378	-0.0191	0.1334	-0.0005
MV $\lambda = 0.75$	1.1976	-0.4161	-11.9226	1.0101	-0.1188	0.1381	0.0004
MV $\lambda = 0.5$	0.8668	0.1176	-3.3104	0.8157	0.1504	0.1498	0.0006
MV $\lambda = 0.05$	0.0991	0.0752	-0.0195	0.0970	0.0749	0.0659	0.0177

- (c) Log-relative wealth of the randomized-signature portfolios with and without transaction costs trading in the SMI universe.

Opt. Task Prop. TC	log V_T without Reg.			log V_T with Reg. $\beta_{1\%}$		log V_T with Reg. $\beta_{5\%}$	
	0%	1%	5%	0%	1%	0%	5%
Log-Opt.	1.7442	-2.5700	$-\infty$	0.9282	0.0402	0.2287	-0.1564
MV $\lambda = 1$	0.3089	0.2657	0.0944	0.3089	0.2657	0.2292	0.0987
MV $\lambda = 0.75$	0.2342	0.2092	0.1101	0.2342	0.2092	0.2179	0.1078
MV $\lambda = 0.5$	0.1578	0.1460	0.0990	0.1578	0.1460	0.1578	0.0990
MV $\lambda = 0.05$	0.0161	0.0156	0.0137	0.0161	0.0156	0.0161	0.0137

- (d) Log-relative wealth of the JL-signature portfolios with and without transaction costs trading in the S&P500 universe.

Opt. Task Prop. TC	log V_T without Reg.			log V_T with Reg. $\beta_{1\%}$		log V_T with Reg. $\beta_{5\%}$	
	0%	1%	5%	0%	1%	0%	5%
Log-Opt.	1.0296	0.1415	-3.9870	0.7823	0.2411	0.1789	-0.0820
MV $\lambda = 1$	0.0988	0.0903	0.0566	0.0988	0.0903	0.0988	0.0566
MV $\lambda = 0.75$	0.0743	0.0686	0.0455	0.0743	0.0686	0.0743	0.0455
MV $\lambda = 0.5$	0.0497	0.0462	0.0319	0.0497	0.0462	0.0497	0.0319
MV $\lambda = 0.05$	0.0050	0.0047	0.0033	0.0050	0.0047	0.0050	0.0033

- (e) Log-relative wealth of the randomized-signature portfolios with and without transaction costs trading in the S&P500 universe.

Table 1.4: The first three columns show the performance with and without transaction costs of signature portfolios with no regularization for transaction costs. The next four column show the performance with and without transaction costs of portfolios trained with a regularization for transaction costs at the respective level.

1.7 Numerical results

Portfolio Prop. TC	Signature Portfolio			JL-Signature Portfolio			Rand.-Signature Portfolio		
	0%	1%	5%	0%	1%	5%	0%	1%	5%
Universe	0.6061	0.6061	0.6061	0.6061	0.6061	0.6061	0.6061	0.6061	0.6061
Equally-weighted	0.7487	0.6287	0.1482	0.7487	0.6287	0.1482	0.7487	0.6287	0.1482
Log-Opt.	0.5786	0.4353	0.55	0.6735	0.5459	0.5389	0.7246	0.7052	0.6239
MV $\lambda = 1$	1.282	0.5191	0.5998	1.263	0.4941	0.6144	1.2029	0.3374	0.5805
MV $\lambda = 0.75$	1.3051	0.4052	0.6004	1.2873	0.3081	0.6125	1.2237	0.2585	0.5812
MV $\lambda = 0.5$	1.3396	0.4849	0.5979	1.3226	0.4906	0.6016	1.2513	0.5202	0.5773
MV $\lambda = 0.05$	0.946	0.8718	0.7073	0.9211	0.8515	0.6726	0.8707	0.8028	0.6499

- (a) Annualized Sharpe-ratios of the signature portfolios, universe portfolio and equally weighted portfolio with and without transaction costs trading in the SMI universe. Note that for the results with transaction costs, the corresponding regularization for transaction costs was included in the training.

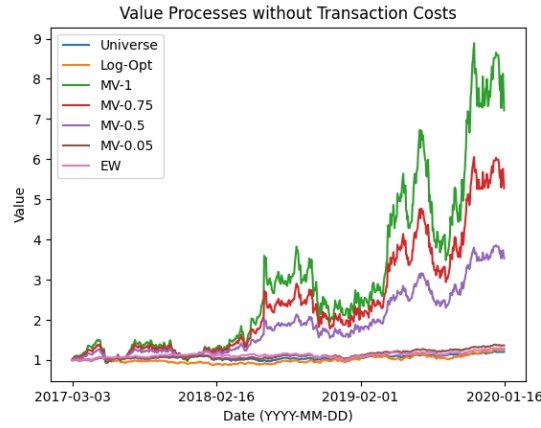
Portfolio Prop. TC	JL-Signature Portfolio			Rand.-Signature Portfolio		
	0%	1%	5%	0%	1%	5%
Universe	0.8530	0.8530	0.8530	0.8530	0.8530	0.8530
Equally-weighted	0.9292	0.7466	0.0164	0.9292	0.7466	0.0164
Log-Opt.	1.223	0.4901	0.3653	1.1904	0.6777	0.5330
MV $\lambda = 1$	1.1817	1.1040	0.8753	1.0077	0.9879	0.9087
MV $\lambda = 0.75$	1.1402	1.0905	0.9048	0.9754	0.9614	0.9054
MV $\lambda = 0.5$	1.0781	1.0521	0.9488	0.9391	0.9302	0.8945
MV $\lambda = 0.05$	0.8834	0.8821	0.8772	0.8625	0.8616	0.8581

- (b) Annualized Sharpe-ratios of the signature portfolios, universe portfolio and equally weighted portfolio with and without transaction costs trading in the S&P500 universe. Note that for the results with transaction costs, the corresponding regularization for transaction costs was included in the training.

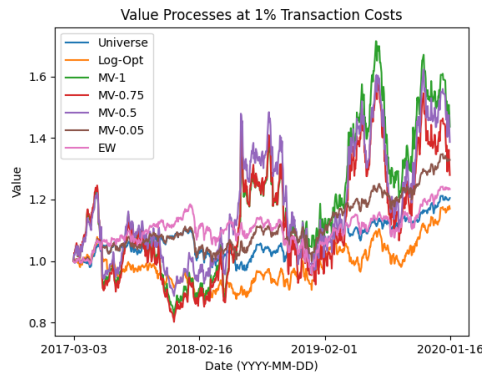
Table 1.5: The tables show the annualized out-of-sample Sharpe-ratios of the trained signature portfolios as well as universe and equally-weighted portfolio in the SMI and SPX universe respectively. The numbers marked in bold are those that are higher than the Sharpe-ratio of the respective universe portfolio.

Universe Beta	SMI		SMI (JL)		SMI (rand.)		S&P500 (JL)		S&P500 (rand.)	
	$\beta_{1\%}$	$\beta_{5\%}$	$\beta_{1\%}$	$\beta_{5\%}$	$\beta_{1\%}$	$\beta_{5\%}$	$\beta_{1\%}$	$\beta_{5\%}$	$\beta_{1\%}$	$\beta_{5\%}$
Log-Opt.	27.958	847.61	0.5	619.60	0.4999	960.77	9.6437	97.160	2.9908	65.285
MV $\lambda = 1$	0.0879	1.1661	0.0609	0.9611	0.0193	0.5510	10^{-8}	0.0059	10^{-8}	10^{-8}
MV $\lambda = 0.75$	0.0240	0.8063	0.0101	0.6598	0.0040	0.3692	10^{-8}	0.0012	10^{-8}	10^{-8}
MV $\lambda = 0.5$	0.0008	0.4537	10^{-8}	0.3645	0.0010	0.1954	10^{-8}	10^{-8}	10^{-8}	10^{-8}
MV $\lambda = 0.05$	0.0003	0.0095	10^{-8}	0.0076	0.0003	0.0111	10^{-8}	10^{-8}	10^{-8}	10^{-8}

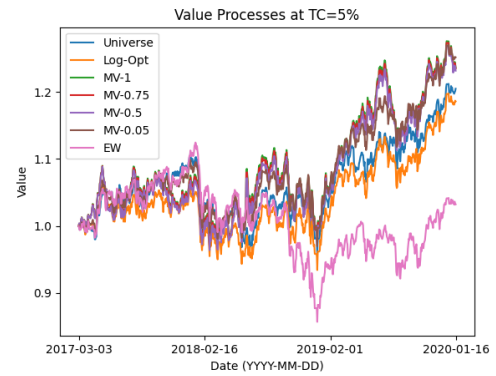
Table 1.6: Optimal regularization parameters β found during in-sample training.



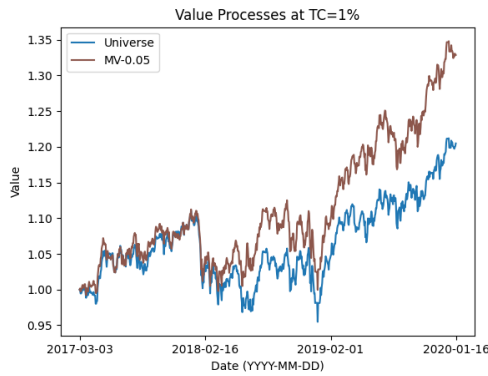
(a) Wealth process without transaction costs



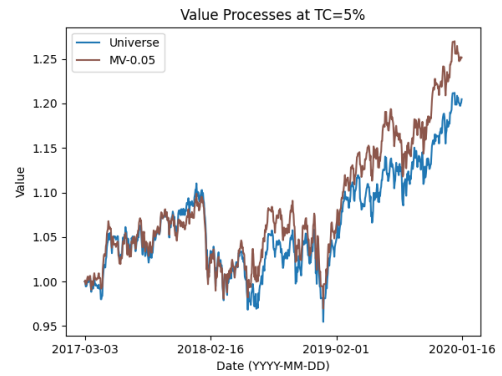
(b) Wealth process with 1% proportional transaction costs



(c) Wealth process with 5% proportional transaction costs



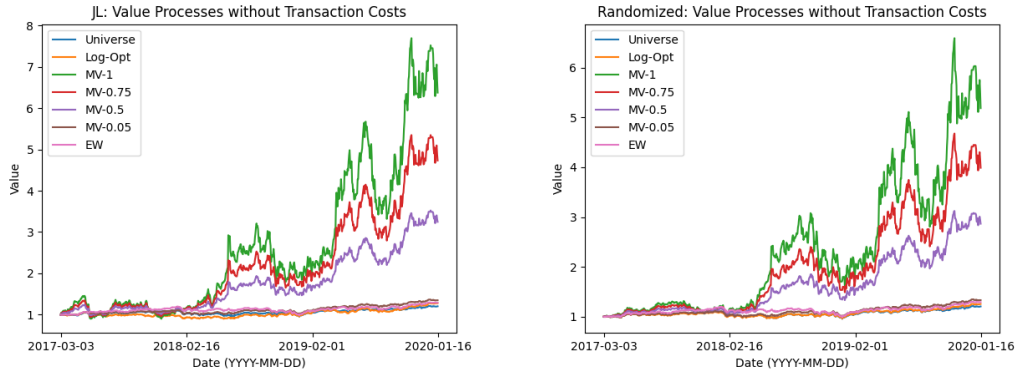
(d) Wealth process with 1% proportional transaction costs



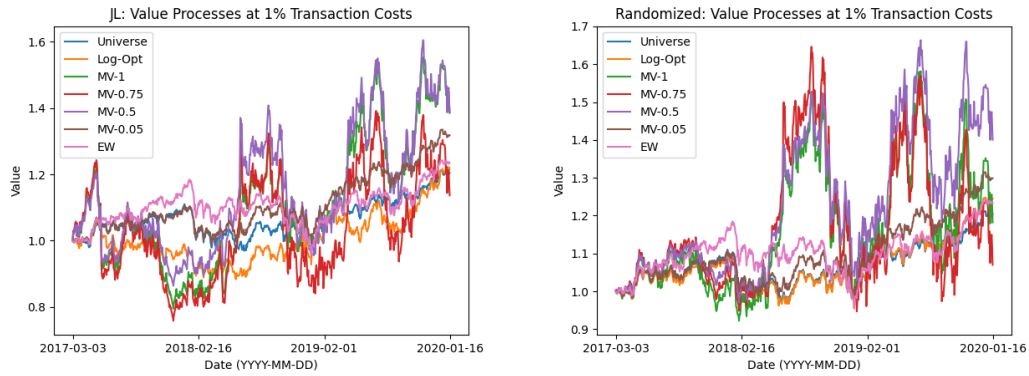
(e) Wealth Process with 5% Proportional Transaction Costs

Figure 1.4: SMI wealth processes of the signature portfolios, the universe portfolio and the equally-weighted portfolio with and without transaction costs. Note that signature portfolios in the settings with transaction cost are those, where we included the regularization for transaction costs.

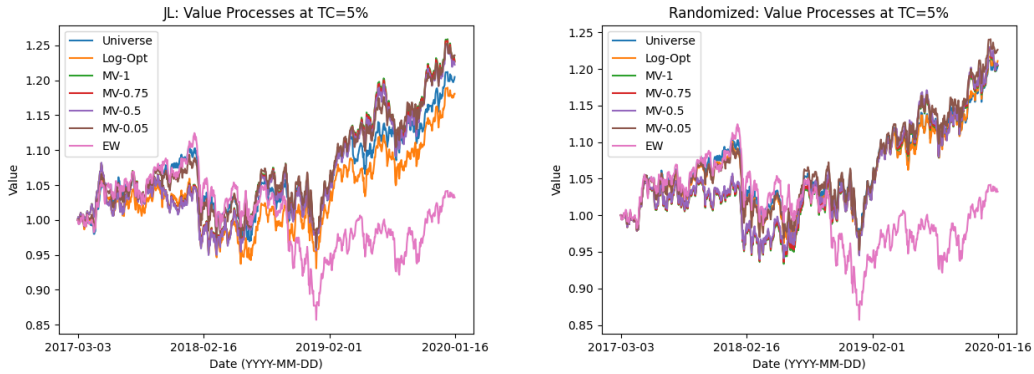
1.7 Numerical results



(a) Wealth process without transaction costs for JL-signature portfolios (left) and randomized-signature portfolios (right)

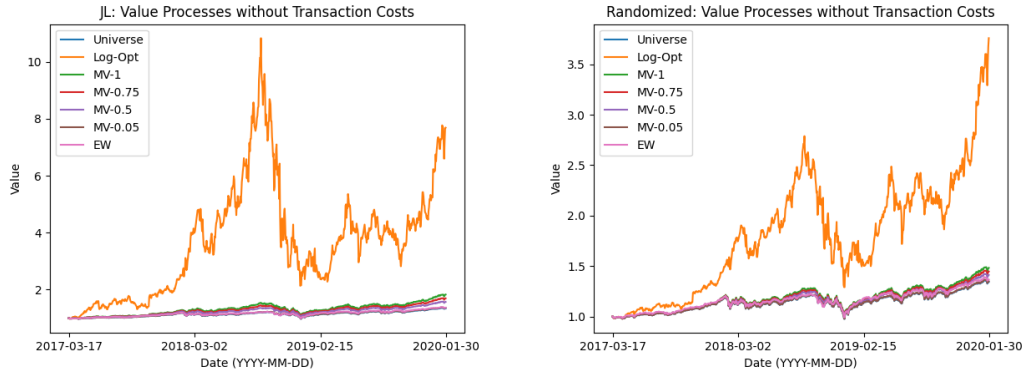


(b) Wealth process with 1% proportional transaction costs for JL-signature portfolios (left) and randomized-signature portfolios (right)

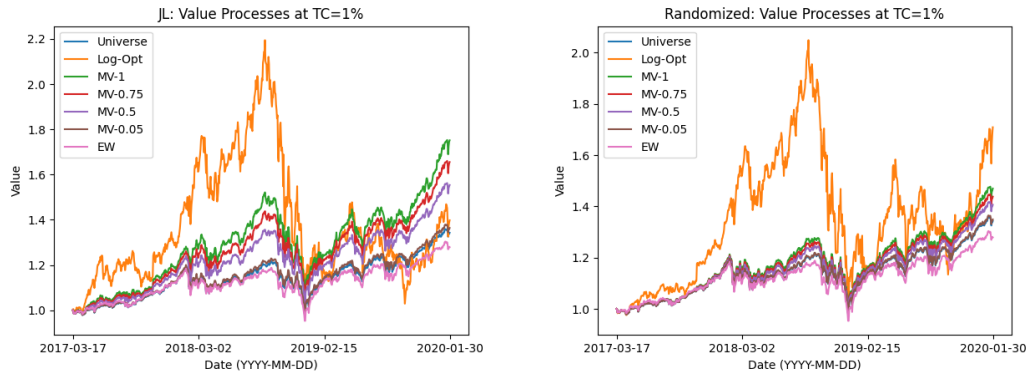


(c) Wealth process with 5% proportional transaction costs for JL-signature portfolios (left) and randomized-signature portfolios (right)

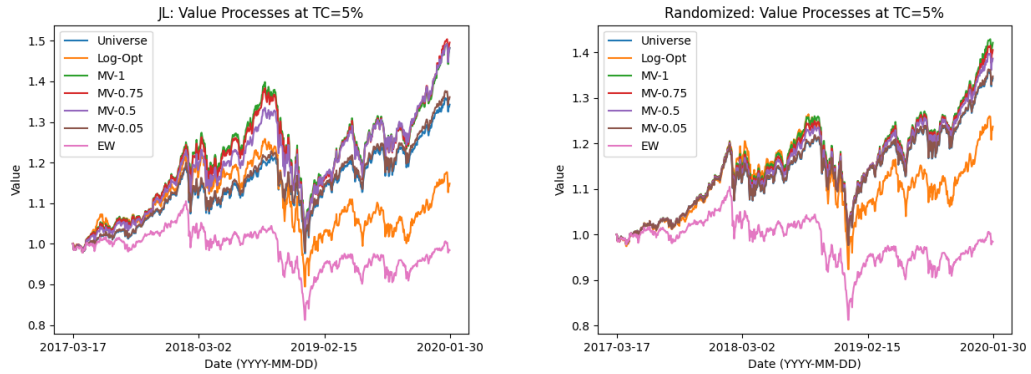
Figure 1.5: SMI wealth processes of the JL-signature portfolios (left), randomized-signature portfolios (right), universe portfolio and equally-weighted portfolio with and without transaction costs. Note that signature portfolios in the settings with transaction cost are those, where we included the regularization for transaction costs.



(a) Wealth process without transaction costs for JL-signature portfolios (left) and randomized-signature portfolios (right)



(b) Wealth process with 1% proportional transaction costs for JL-signature portfolios (left) and randomized-signature portfolios (right)



(c) Wealth process with 5% proportional transaction costs for JL-signature portfolios (left) and randomized-signature portfolios (right)

Figure 1.6: S&P500 wealth processes of the JL-signature portfolios (left), randomized-signature portfolios (right), universe portfolio and equally-weighted portfolio with and without transaction costs. Note that signature portfolios in the settings with transaction cost are those, where we included the regularization for transaction costs.

Appendix of Chapter 1

1.A Proofs of Section 1.2

Since the space of lifted stopped paths as introduced in Definition 1.2.8 contains paths defined on different time-intervals, we need the following definition.

Definition 1.A.1 (Projection on the final value). We denote by $(\cdot)_\odot$ the projection to the value at the final time, i.e. for each $t \in [0, T]$

$$(x_{[0,t]})_\odot := x_t.$$

For this projection on the final value we have the following continuity result.

Lemma 1.A.2. The function

$$\begin{aligned} \lambda : [0, T] \times \mathcal{G}_T^N(\varphi, x) &\rightarrow \Lambda_T^N(\varphi, x) \\ (t, \hat{\mathbb{X}}_{[0,T]}^N(\omega)) &\mapsto \hat{\mathbb{X}}_{[0,t]}^N(\omega) \end{aligned}$$

is continuous.

Proof. Take an arbitrary but fixed $(s, \hat{\mathbb{Y}}_{[0,T]}^N(\omega_1))$. We show that λ is continuous at $(s, \hat{\mathbb{Y}}_{[0,T]}^N(\omega_2))$ but since this point is arbitrary, λ is continuous everywhere. For simplicity, we write $\hat{\mathbf{x}}_{[0,t]} = \hat{\mathbb{X}}_{[0,t]}^N(\omega_2)$, $\hat{\mathbf{y}}_{[0,t]} = \hat{\mathbb{Y}}_{[0,t]}^N(\omega_1)$ and $\widehat{\mathbf{y}}_{[0,t]}^s = \widehat{\mathbb{Y}}_{[0,t]}^s(\omega_1)$. For any $(t, \hat{\mathbf{x}}_{[0,T]}) \in [0, T] \times \mathcal{G}_T^N$ it holds that

$$\begin{aligned} &d_\Lambda \left(\lambda \left(t, \hat{\mathbf{x}}_{[0,T]} \right), \lambda \left(s, \hat{\mathbf{y}}_{[0,T]} \right) \right) \\ &= d_\Lambda \left(\hat{\mathbf{x}}_{[0,t]}, \hat{\mathbf{y}}_{[0,s]} \right) \\ &\leq d_\Lambda \left(\hat{\mathbf{x}}_{[0,t]}, \hat{\mathbf{y}}_{[0,t]} \right) + d_\Lambda \left(\hat{\mathbf{y}}_{[0,t]}, \hat{\mathbf{y}}_{[0,s]} \right) \\ &= d_{p-var;[0,t]} \left(\hat{\mathbf{x}}_{[0,t]}, \hat{\mathbf{y}}_{[0,t]} \right) + |t - s| + d_{p-var;[0,t \vee s]} \left(\widehat{\mathbf{y}}_{[0,t \vee s]}^t, \widehat{\mathbf{y}}_{[0,t \vee s]}^s \right) \\ &\leq d_{p-var;[0,t]} \left(\hat{\mathbf{x}}_{[0,t]}, \hat{\mathbf{y}}_{[0,t]} \right) + |t - s| + \|\hat{\mathbf{y}}\|_{p-var;[t \wedge s, t \vee s]}, \end{aligned}$$

where $\|\cdot\|_{p-var;[0,T]}$ is the norm induced by $d_{p-var;[0,T]}$. Let now $\epsilon > 0$ and choose $\delta > 0$ such that

$$\delta + \|\hat{\mathbf{y}}\|_{p-var;[s-\delta, s+\delta]} < \epsilon.$$

Note that this is possible because $\|\hat{\mathbf{y}}\|_{p-var;[s-\delta, s+\delta]} \rightarrow 0$ as $\delta \rightarrow 0$.

Therefore, for all $(t, \hat{\mathbf{x}}_{[0,T]}) \in [0, T] \times \mathcal{G}_T^N$ (equipped with the product topology) satisfying

$$|t - s| + d_{p-var;[0,T]}(\hat{\mathbf{x}}_{[0,T]}, \hat{\mathbf{y}}_{[0,T]}) < \delta,$$

we have

$$\begin{aligned} |t - s| + d_{p-var;[0,t]}(\hat{\mathbf{x}}_{[0,t]}, \hat{\mathbf{y}}_{[0,t]}) + \|\hat{\mathbf{y}}\|_{p-var;[t \wedge s, t \vee s]} &< \delta + \|\hat{\mathbf{y}}\|_{p-var;[t \wedge s, t \vee s]} \\ &\leq \delta + \|\hat{\mathbf{y}}\|_{p-var;[s-\delta, s+\delta]} \\ &< \epsilon, \end{aligned}$$

which proves continuity of λ . □

The following lemma can be proved by means of Corollary 1.2.11.

Lemma 1.A.3. For each multiindex I the map

$$\Lambda_T^2(\varphi, x) \ni \hat{\mathbb{X}}_{[0,t]}^2(\omega) \mapsto \langle e_I, S^{|I|}(\hat{\mathbb{X}}_{[0,t]}^2(\omega))_{\odot} \rangle = \langle e_I, \hat{\mathbb{X}}_t^N(\omega) \rangle \in \mathbb{R}$$

is continuous on bounded sets.

Proof. Let us show the statement for an arbitrary but fixed multiindex I . For $N = 2$ we apply the same notation as in the proof of Lemma 1.A.2 above, i.e. $\hat{\mathbf{x}}_{[0,t]} = \hat{\mathbb{X}}_{[0,t]}^2(\omega_x)$, etc. We show continuity at some fixed $\hat{\mathbf{y}}_{[0,s]}$ in some bounded set of $\Lambda_T^2(\varphi, x)$. Recall that by Corollary 1.2.11

$$\hat{\mathbf{x}}_{[0,t]} \mapsto \langle e_I, S^{|I|}(\hat{\mathbf{x}}_{[0,t]}) \rangle$$

is continuous on bounded sets of $\mathcal{G}_t^2(\varphi, x)$ for each $t \in [0, T]$. Thus, for every $\epsilon > 0$ there exists some $\delta_1 > 0$ such that for all $\hat{\mathbf{x}}_{[0,t]}$ satisfying

$$d_{p-var;[0,t]}(\hat{\mathbf{x}}_{[0,t]}, \hat{\mathbf{y}}_{[0,t]}) < \delta_1$$

we have

$$\|\langle e_I, S^{|I|}(\hat{\mathbf{x}}_{[0,t]}) \rangle - \langle e_I, S^{|I|}(\hat{\mathbf{y}}_{[0,t]}) \rangle\|_{p-var;[0,t]} < \frac{\epsilon}{2}.$$

Here, $\|\cdot\|_{p-var;[0,t]}$ denotes the p -variation distance on $\mathbb{R}$. Moreover, the map

$$[0, T] \ni t \mapsto \langle e_I, S^{|I|}(\hat{\mathbf{y}}_{[0,t]})_{\odot} \rangle \in \mathbb{R}$$

is continuous for each fixed $\hat{\mathbf{y}}$. Hence, there exists some $\delta_2 > 0$ such that for all t with $|t - s| < \delta_2$

$$|\langle e_I, S^{|I|}(\hat{\mathbf{y}}_{[0,t]})_{\odot} \rangle - \langle e_I, S^{|I|}(\hat{\mathbf{y}}_{[0,s]})_{\odot} \rangle| < \frac{\epsilon}{2}.$$

Choose now $\delta = \min(\delta_1, \delta_2)$ and let $\hat{\mathbf{x}}_{[0,t]}$ (in the bounded set) be such that

$$\begin{aligned} \delta &> d_{\Lambda}(\hat{\mathbf{x}}_{[0,t]}, \hat{\mathbf{y}}_{[0,s]}) = |t - s| + d_{p-var;[0,t \vee s]}(\widehat{\mathbf{x}}_{[0,t \vee s]}^t, \widehat{\mathbf{y}}_{[0,t \vee s]}^s) \\ &\geq |t - s| + d_{p-var;[0,t]}(\hat{\mathbf{x}}_{[0,t]}, \widehat{\mathbf{y}}_{[0,t]}^{t \wedge s}). \end{aligned}$$

Then, by the above

$$\begin{aligned}
|\langle e_I, S^{[I]}(\hat{\mathbf{x}}_{[0,t]})_\odot \rangle - \langle e_I, S^{[I]}(\hat{\mathbf{y}}_{[0,s]})_\odot \rangle| &\leq |\langle e_I, S^{[I]}(\hat{\mathbf{x}}_{[0,t]})_\odot \rangle - \langle e_I, S^{[I]}(\widehat{\mathbf{y}^{t \wedge s}}_{[0,t]})_\odot \rangle| \\
&\quad + |\langle e_I, S^{[I]}(\widehat{\mathbf{y}^{t \wedge s}}_{[0,t]})_\odot \rangle - \langle e_I, S^{[I]}(\hat{\mathbf{y}}_{[0,s]})_\odot \rangle| \\
&\leq \|\langle e_I, S^{[I]}(\hat{\mathbf{x}}_{[0,t]})_\odot \rangle - \langle e_I, S^{[I]}(\widehat{\mathbf{y}^{t \wedge s}}_{[0,t]})_\odot \rangle\|_{p\text{-var}, [0,t]} \\
&\quad + |\langle e_I, S^{[I]}(\hat{\mathbf{y}}_{[0,t]})_\odot \rangle - \langle e_I, S^{[I]}(\hat{\mathbf{y}}_{[0,s]})_\odot \rangle| \\
&< \epsilon,
\end{aligned}$$

which proves the assertion. $\square$

We are now prepared to prove Theorem 1.2.12.

Proof of Theorem 1.2.12. Consider the function λ as introduced in Lemma 1.A.2 and note that for any compact subset $K \subset \mathcal{G}_T^2(\varphi, x)$ the set

$$\mathcal{K}_\Lambda = \lambda([0, T] \times K) = \left\{ \hat{\mathbb{X}}_{[0,t]}^2(\omega) \mid t \in [0, T] \text{ and } \hat{\mathbb{X}}_{[0,T]}^2(\omega) \in K \right\}$$

is a compact subset of $\Lambda_T^2(\varphi, x)$ due to the continuity of λ as proved in Lemma 1.A.2. For each $\hat{\mathbb{X}}_{[0,t]}^2(\omega) \in \Lambda_T^2(\varphi, x)$ and $N \geq 2$ it holds that $S^N(\hat{\mathbb{X}}_{[0,t]}^2(\omega))_\odot = \hat{\mathbb{X}}_t^N(\omega)$ a.s. Hence, we can associate with every linear function on the signature L a non-anticipative path-functional $f_L : \hat{\mathbb{X}}_{[0,t]}^2(\omega) \mapsto L(\hat{\mathbb{X}}_t(\omega))$. Moreover, for every linear function on the signature L it holds that $f_L \in C(\Lambda_T^2(\varphi, x); \mathbb{R})$, which follows from Lemma 1.A.3. Let us denote by

$$\mathfrak{F}_\mathfrak{L} := \left\{ f_L \mid f_L : \hat{\mathbb{X}}_{[0,t]}^2(\omega) \mapsto L(\hat{\mathbb{X}}_t(\omega)) \text{ for } \hat{\mathbb{X}}_{[0,t]}^2(\omega) \in \Lambda_T^2(\varphi, x), L \in \mathfrak{L} \right\}$$

the set of path functionals induced by linear functions on the signature and by

$$\mathfrak{F}_\mathfrak{L}|_{\mathcal{K}_\Lambda} := \left\{ f_L \mid f_L : \hat{\mathbb{X}}_{[0,t]}^2(\omega) \mapsto L(\hat{\mathbb{X}}_t(\omega)) \text{ for } \hat{\mathbb{X}}_{[0,t]}^2(\omega) \in \mathcal{K}_\Lambda, L \in \mathfrak{L} \right\}$$

its restriction to lifted paths in $\mathcal{K}_\Lambda$. We will now apply the Stone-Weierstrass Theorem to show that $\mathfrak{F}_\mathfrak{L}|_{\mathcal{K}_\Lambda}$ is dense in $C(\mathcal{K}_\Lambda; \mathbb{R})$. To this end we need to show that $\mathfrak{F}_\mathfrak{L}|_{\mathcal{K}_\Lambda}$ is a sub-algebra of $C(\mathcal{K}_\Lambda; \mathbb{R})$ which separates points and vanishes nowhere. The algebra property follows from the fact that for each $\hat{\mathbb{X}}_{[0,t]}^2(\omega) \in \Lambda_T^2(\varphi, x)$ and for any two $f_{L_1}, f_{L_2} \in \mathfrak{F}_\mathfrak{L}$ it holds that

$$\begin{aligned}
(f_{L_1} \cdot f_{L_2})\left(\hat{\mathbb{X}}_{[0,t]}^2(\omega)\right) &= f_{L_1}\left(\hat{\mathbb{X}}_{[0,t]}^2(\omega)\right) f_{L_2}\left(\hat{\mathbb{X}}_{[0,t]}^2(\omega)\right) \\
&= L_1\left(\hat{\mathbb{X}}_t(\omega)\right) L_2\left(\hat{\mathbb{X}}_t(\omega)\right) = L\left(\hat{\mathbb{X}}_t(\omega)\right) = f_L\left(\hat{\mathbb{X}}_{[0,t]}^2(\omega)\right),
\end{aligned}$$

where the second last equality follows from the shuffle product property. We now show that $\mathfrak{F}_\mathfrak{L}$ separates points. To this end take $\hat{\mathbb{X}}_{[0,t]}^2(\omega_x), \hat{\mathbb{Y}}_{[0,s]}^2(\omega_y) \in \Lambda_T^2(\varphi, x)$ with $\hat{\mathbb{X}}_{[0,t]}^2(\omega_x) \neq \hat{\mathbb{Y}}_{[0,s]}^2(\omega_y)$. If $t \neq s$, the two points are separated in the time-component, i.e. $\langle e_1, \hat{\mathbb{X}}_t(\omega_x) \rangle \neq$

$\langle e_1, \hat{\mathbb{Y}}_s(\omega_y) \rangle$. If $t = s$, the point-separation follows from Lemma 1.2.4. The algebra vanishes nowhere because $f_\emptyset(\hat{\mathbb{X}}_{[0,t]}^2(\omega)) = \langle e_\emptyset, \hat{\mathbb{X}}_t(\omega) \rangle \equiv 1$ for all $\hat{\mathbb{X}}_{[0,t]}^2(\omega) \in \Lambda_T^2$, where $\emptyset$ denotes the empty word. The desired density statement follows by the Stone-Weierstrass Theorem, i.e. for any continuous non-anticipative path-functional $f : \Lambda_T^2(\varphi, x) \rightarrow \mathbb{R}$ and for every $\epsilon > 0$ there exists a path-functional $f_L \in \mathfrak{F}_{\mathcal{L}}$ such that for almost all $\omega \in \Omega$

$$\sup_{(t, \hat{\mathbb{X}}_{[0,T]}^2(\omega)) \in [0, T] \times K} |f(\hat{\mathbb{X}}_{[0,t]}^2(\omega)) - f_L(\hat{\mathbb{X}}_t)(\omega)| < \epsilon,$$

which is equivalent to the claim of Theorem 1.2.12. $\square$

Proof of Lemma 1.2.19. We have to prove that $K_R = \psi^{-1}((0, R])$ is compact for any $R > 0$. Let us first note that the map λ can be easily extended to weakly geometric α -Hölder rough paths and is continuous again by the exact same arguments as before. Moreover, note that the set

$$K'_R = \left\{ \widehat{\mathbf{X}}_{[0,T]}^t \left| \psi(\hat{\mathbf{X}}_{[0,t]}) \in [0, R) \right. \right\}$$

equipped with the $d_{\alpha', [0, T]}$ -topology is for any $R > 0$ a $\|\cdot\|_{\alpha, [0, T]}$ -bounded subset of $\hat{C}_T^\alpha(\varphi, x)$ and hence compact for $0 \leq \alpha' < \alpha$ due to the compact embedding of α -Hölder spaces into α' -Hölder spaces, see (Cuchiero et al., 2023b, Theorem A.3). Moreover, $K_R = \lambda([0, T] \times K'_R)$ and is therefore compact for any $R > 0$ due to Tychonoff's theorem and the continuity of λ . $\square$

Proof of Theorem 1.2.20. Using Lemma 1.2.19 the proof is analogous to the proof of (Cuchiero et al., 2023b, Theorem 5.4). In fact it even holds that

$$\sup_{\hat{\mathbf{X}}_{[0,t]} \in \bigcup_{t \in [0, T]} \hat{C}_t^\alpha(\varphi, x)} \frac{|f(\hat{\mathbf{X}}_{[0,t]}) - L(\hat{\mathbb{X}}_t)|}{\psi(\hat{\mathbf{X}}_{[0,t]})} < \epsilon$$

and since $\Lambda_T^2(\varphi, x) \subset \bigcup_{[0, t]} \hat{C}_t^\alpha(\varphi, x)$ for every $\frac{1}{3} < \alpha \leq \frac{1}{2}$ the claim follows. $\square$

1.B Proofs of Section 1.4

The following lemma is needed in the proof of Theorem 1.4.4.

Lemma 1.B.1. Let D be a (subset of a) metric space. Consider a matrix-valued function $M : D \ni x \mapsto M(x)$ whose coefficients are continuous functions from D to $\mathbb{R}$ and which is non-singular for all $x \in D$, then the coefficients of its matrix-inverse M^{-1} are again continuous functions.

Proof. Let us prove the statement by induction:

- Let $M : D \rightarrow \mathbb{R}^{2 \times 2}$. Then the inverse of

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is given by

$$M^{-1} = \frac{1}{(ad - bc)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

By assumption the determinant $(ad - bc)$ is non-zero, which implies continuity of the coefficients of M^{-1} .

- Assume the statement holds for an arbitrary but fixed $n \geq 2$.
- Let $M : D \rightarrow \mathbb{R}^{(n+1) \times (n+1)}$. Then we can write M in block matrix form

$$M = \begin{pmatrix} \tilde{M} & m^1 \\ m^2 & m^3 \end{pmatrix},$$

where $\tilde{M}$ takes values in $\mathbb{R}^{n \times n}$, $m^1, (m^2)^\top$ in $\mathbb{R}^n$ and m^3 in $\mathbb{R}$. The inverse is then computed via

$$M^{-1} = \begin{pmatrix} \tilde{M}^{-1} + \frac{1}{\Delta} \tilde{M}^{-1} m^1 m^2 \tilde{M}^{-1} & -\frac{1}{\Delta} \tilde{M}^{-1} m^1 \\ \frac{1}{\Delta} m^2 \tilde{M}^{-1} & \frac{1}{\Delta} \end{pmatrix}$$

where $\Delta = m^3 - m^2 \tilde{M}^{-1} m^1$ with values in $\mathbb{R} \setminus \{0\}$ by assumption on invertability. As (the components of) $\tilde{M}^{-1}, m^1, m^2$ and m^3 are all continuous functions, continuity of the components of M^{-1} follows as well.

□

1.C Proofs of Section 1.5

Proof of Lemma 1.5.14. By Equation (1.5.2), we know that

$$\begin{aligned} \log \left(\frac{W_t^\pi}{W_t^\mathcal{U}} \right) &= \sum_{i \in \mathcal{U}} \int_{t_0}^t \gamma_s^i f^i(s, X_{[0,s]}) d\mu_s^{\mathcal{U},i} \\ &- \frac{1}{2} \sum_{i,j \in \mathcal{U}} \int_{t_0}^t \gamma_s^i \gamma_s^j f^i(s, X_{[0,s]}) f^j(s, X_{[0,s]}) d[\mu^{\mathcal{U},i}, \mu^{\mathcal{U},j}]_s \\ &+ \sum_{i \in \mathcal{U}} \int_{t_0}^t \left(1 - \sum_{k=1}^d \lambda_s^k f^k(s, X_{[0,s]}) \right) d\mu_s^{\mathcal{U},i} \end{aligned} \quad (1.C.1)$$

$$- \frac{1}{2} \sum_{i,j \in \mathcal{U}} \int_{t_0}^t \left(1 - \sum_{k=1}^d \lambda_s^k f^k(s, X_{[0,s]}) \right) \left(1 - \sum_{m=1}^d \lambda_s^m f^m(s, X_{[0,s]}) \right) d[\mu^{\mathcal{U},i}, \mu^{\mathcal{U},j}]_s, \quad (1.C.2)$$

for

$$\lambda_s^k = \begin{cases} \mu_s^{\mathcal{U},k} & \text{if } \pi \text{ is of type } I \\ 1 & \text{if } \pi \text{ is of type } II \end{cases}.$$

Note that the terms in (1.C.1) and (1.C.2) are both zero due to $\sum_{i \in \mathcal{U}} \mu_t^{\mathcal{U},i} \equiv 1$. □

1.D Proofs of Section 1.6

Proof of Proposition 1.6.1. (i) This case always holds for long-only portfolios and is discussed in Ruf and Xie (2020). The argument in the long-only case is that $L(1) < R(1)$ and $L(0) > R(0)$, while both sides are obviously continuous and monotone. While the former statement still holds in the case with short-selling, the monotonicity is no longer true. However, one can show by a tedious case-by-case study that where $R'(\alpha)$ exists it is piece-wise constant and increasing. Therefore a unique solution for $\alpha^* \in [0, 1]$ is still assured in this case.

(ii) Here the above argument does not hold anymore. This is because now $L(0) \leq R(0)$, while it still holds that $L(1) < R(1)$. Let us give three examples, one for each of the cases a) no solution b) no solution in $[0, 1]$, c) no unique solution in $[0, 1]$.

(a) *no solution*: consider a market of two stocks and $c = 0.05$, i.e. 5% transaction costs. For the portfolio weights

$$\pi_{t-} = (13.1, -12.1)^T, \quad \pi_t = (13, -12)^T$$

there does not exist a solution for α .

(b) *no solution in $[0, 1]$* : again for $c = 0.05$ and a market of two stocks. For the portfolio weights

$$\pi_{t-} = (11, -10)^T, \quad \pi_t = (10, -9)^T$$

the solution is $\alpha = -1$.

(c) *no unique solution in $[0, 1]$* : consider $c = 0.05$ and a market of three stocks with portfolio weights

$$\pi_{t-} = (5, 6, -10)^T, \quad \pi_t = (5.5, 6.5, -11)^T$$

then the solutions are $\alpha = 0.3333$ and $\alpha = 0.9535$ (numbers are rounded). □

2 Joint calibration to SPX and VIX options with signature-based models

Abstract

We consider a stochastic volatility model where the dynamics of the volatility are described by a linear function of the (time extended) signature of a primary process which is supposed to be a polynomial diffusion. We obtain closed form expressions for the VIX squared, exploiting the fact that the truncated signature of a polynomial diffusion is again a polynomial diffusion. Adding to such a primary process the Brownian motion driving the stock price, allows then to express both the log-price and the VIX squared as linear functions of the signature of the corresponding augmented process. This feature can then be efficiently used for pricing and calibration purposes. Indeed, as the signature samples can be easily precomputed, the calibration task can be split into an offline sampling and a standard optimization. We also propose a Fourier pricing approach for both VIX and SPX options exploiting that the signature of the augmented primary process is an infinite dimensional affine process. For both the SPX and VIX options we obtain highly accurate calibration results, showing that this model class allows to solve the joint calibration problem without adding jumps or rough volatility.

2.1 Introduction

The joint calibration of option pricing models to SPX and VIX¹ options is a problem that has gained a lot of attention in quantitative finance since several years. One main reason for the increased interest is that the VIX index has become an important underlying for many derivatives. In fact, futures and options written on it are extensively used to hedge the volatility exposure of option portfolios, see e.g., [Rhoads \(2011\)](#). We address the reader to the official website of CBOE² for details on the history of the VIX, how it is computed, traded and used as underlying for derivatives, which emphasizes in particular the need to jointly calibrate to both, the prices of options on the S&P 500 index and to prices of VIX derivatives. In this respect the main challenge is to reconcile the large negative skew of SPX options' implied volatilities with relatively lower implied volatilities arising from the VIX options, especially for short maturities (see e.g., [Guyon \(2020a\)](#)).

Inspired by [Arribas et al. \(2020\)](#) and [Cuchiero et al. \(2023a\)](#), we consider here a new type of stochastic volatility model for the discounted, dividend-adjusted price process $S = (S_t)_{t \geq 0}$.

¹With SPX and VIX we refer to the index tickers of the S&P 500 and its volatility index, respectively. In the sequel we will use SPX and S&P 500 interchangeably.

²www.cboe.com/tradable_products/vix/

It is given by

$$dS_t(\ell) = S_t(\ell)\sigma_t^S(\ell)dB_t,$$

for an initial condition $S_0 \in \mathbb{R}_+$, a standard Brownian motion B , and a volatility process σ^S satisfying

$$\sigma_t^S(\ell) := \ell(\widehat{\mathbb{X}}_t), \quad (2.1.1)$$

where ℓ is a linear map of the signature $\widehat{\mathbb{X}}_t$ of a process $\widehat{X}$. Specifically, the main ingredient in this framework is a d -dimensional polynomial diffusion process $X = (X_t^1, \dots, X_t^d)_{t \geq 0}$ (see [Cuchiero et al. \(2012\)](#); [Filipović and Larsson \(2016\)](#)), which we call here *primary process* and whose augmentation with time t is denoted by $(\widehat{X}_t)_{t \geq 0} = (t, X_t^1, \dots, X_t^d)_{t \geq 0}$. By modeling σ^S via (2.1.1) we assume that the signature of $\widehat{X}$, denoted by $\widehat{\mathbb{X}}$ (and rigorously introduced in Section 2.2), serves as a linear regression basis for the volatility process, while the parameters of the linear map ℓ have to be learned from (option price) data. Note that the parameters of X are prespecified beforehand and can thus be seen – in analogy to machine learning terminology – as *hyperparameters* (that of course can be optimized over some validation set). As outlined below this is one of the crucial features that allows for the split of the calibration task into precomputable samples and parameters ℓ to be optimized.

Let us now highlight the implications of this modeling approach and the novelty of the present work.

- The current framework can be seen as *universal* in a large class of continuous (non-rough) stochastic volatility models in the following sense: for a stochastic volatility model, whose volatility is given by a continuous path-functional depending on a polynomial diffusion $(X_t)_{t \geq 0}$, it follows from Proposition 2.2.3 that this path-functional can be approximated by a linear function of the signature of $(\widehat{X}_t)_{t \geq 0}$. A concrete example for a volatility process of such a form is an Itô-diffusion with sufficiently regular coefficients, as in this case, the volatility is indeed given by a continuous path-functional of the time-augmented Brownian motion driving the diffusion.
- Additionally, our model truly nests several classical models (see Remark 2.3.4) and for instance also the ‘quintic Ornstein-Uhlenbeck volatility model’, recently proposed by [Abi Jaber et al. \(2022b\)](#), which – with an additional input curve – is shown to fit SPX and VIX smiles well.
- By choosing the parameters of ℓ appropriately, the modeling framework incorporates both, Markovian (in (S, X)) and path-dependent models.
- Up to our knowledge, it is the first signature-based model that is employed for pricing and calibration of VIX options as well as joint calibration, together with SPX options.
- We illustrate that the joint calibration problem can be solved in this framework without jumps and rough volatility (compare also [Rømer \(2022\)](#); [Abi Jaber et al. \(2022b\)](#); [Guyon and Lekeufack \(2023\)](#)).

- By using time-varying parameters we can go beyond short maturities both for SPX and VIX options (as classically tackled in the literature) and achieve a joint calibration also for longer maturities.

In order to achieve the highly accurate calibration results, illustrated in Section 2.5.3 and Section 2.7, we exploit the following mathematical and numerical properties.

- Defining $Z := (X, B)$, then not only $\sigma^S(\ell)$ but also the log-price $\log(S(\ell))$ can be expressed as a linear function of the signature of $\hat{Z}$. The computational benefit is immediate, since no (Euler) simulation scheme is needed to sample from the marginals of the price process. In terms of the parameters ℓ , $\log(S(\ell))$ is the sum of a quadratic function and a linear one, see Proposition 2.6.4.
- Since $\hat{X}$ is additionally assumed to be a *polynomial diffusion* (see Cuchiero et al. (2012); Filipović and Larsson (2016)), the VIX under our model can be computed analytically via matrix exponentials. Indeed, in this case the forward variance can be represented by a quadratic form in the parameters ℓ and the corresponding matrix can be computed by polynomial technology, i.e. via matrix exponentials, see Theorem 2.5.1. This tractability property is a consequence of the fact that *the truncated signature of a polynomial diffusion is again a polynomial diffusion* (see Section 2.4).
- We can efficiently apply a Monte Carlo approach (potentially with variance reduction) for option pricing and calibration, since the signature samples of $\hat{Z}$ can be computed offline and therefore the simulation and optimization step can be completely separated. Indeed, due to the representations of VIX and $\log(S(\ell))$ described above, the same samples can be used for every linear map ℓ . Therefore, the calibration task can be split into an offline sampling and a standard optimization, as no simulation is needed during the latter. Moreover, due to the fact that we can obtain a closed-form expression for the VIX (thanks to the polynomial technology) we can avoid a nested Monte Carlo procedure to evaluate the conditional expectation.
- Alternatively, a Fourier pricing approach for both VIX and SPX options can be used. Indeed, by building on the fact that the signature of $\hat{Z}$ is an affine process (with values in the extended tensor algebra) as proved in Cuchiero et al. (2023c), its Fourier-Laplace transform can be computed by solving an (extended tensor algebra valued) Riccati equation, which in turn can be used for Fourier pricing as outlined in Section 2.6.1.

The remainder of the chapter is organized as follows. Section 2.1.1 gives a review over the different contributions in the literature concerning the joint calibration problem. In Section 2.2 we introduce the signature in the context of continuous semimartingales, its main properties as well as notation used throughout this chapter. Section 2.3 is dedicated to the introduction of our signature-based model and the connections to classical and also recent stochastic volatility models in the literature. Section 2.4 is then devoted to the discussion and proof of the matrix exponential formula for the (conditional) truncated expected signature of a polynomial diffusion. This result is at the core of Section 2.5, where we derive a tractable

formula for the VIX, needed for pricing VIX options and VIX futures. Building on these formulas, our calibration results to VIX options are presented in Section 2.5.3. In Section 2.6, we then prove, similarly as for the VIX, a tractable expression for S . Additionally, we exploit in Section 2.6.1 the affine nature of the signature process (as proved in Cuchiero et al. (2023c)), to obtain a Fourier pricing approach within our modeling choice for both VIX and SPX options. We finally present the numerical results of the joint calibration problem in Section 2.7, both in the case of constant parameters and with time-varying parameters, where the latter are introduced in Section 2.5.4 and Section 2.6.2.

The data used in Section 2.5.3 and Section 2.7.1 were purchased from OptionMetrics³. An implementation of the model for the joint calibration can be found in [jankamoeller/joint_calib_SPX_VIX](https://github.com/jankamoeller/joint_calib_SPX_VIX).

2.1.1 State of the art

This section is primarily dedicated to a literature review on the joint calibration problem and secondly, to a brief overview on signature methods in finance.

First attempts to solve the joint calibration problem appear in Gatheral (2008), with a double constant elasticity of variance model (CEV), which despite being rather flexible cannot fit accurately the implied volatilities of SPX and VIX options jointly. Later on, the introduction of models with jumps in the SPX (or additionally also in the volatility) led to different contributions, for instance the forward variance model of Cont and Kokholm (2013) described as an exponential of an affine process with Lévy jumps, the regime-switching enhancement of the classical Heston model by Papanicolaou and Sircar (2014), the 3/2 model with jumps in the asset price of Baldeaux and Badran (2014), in the volatility (Kokholm and Stisen (2015)), or with co-jumps and idiosyncratic jumps in the volatility (Pacati et al. (2018)).

Continuous stochastic volatility models based on Markovian semimartingales have also been employed to solve the joint calibration problem. For instance, in Fouque and Saporito (2018) a Heston model with stochastic vol-of-vol has been calibrated, however only for maturities above 4 months where VIX options are less liquid. More recently, Rømer (2022) considered a model where the volatility is driven by two Ornstein-Uhlenbeck (OU) processes using a non-standard transformation function. This choice of two OU-processes has been an inspiration for our concrete numerical implementations. We also point out that the (non-rough) model introduced in Abi Jaber et al. (2022a,b), where the volatility is described by a polynomial of order five in one single OU-process, falls (apart from the additional input curve) into this class of continuous Markovian models and is a particular instance of our framework. Let us also refer to the paper by Guyon and Mustapha (2023), where a neural SDE model has been successfully jointly calibrated. Within the class of continuous, however not necessarily Markovian models, Guyon and Lekeufack (2023) conduct an empirical and statistical analysis as well as a joint calibration for a family of models where the volatility depends on the paths of the asset. These models can be turned into Markovian ones by using exponential kernels instead of general ones, see also Gazzani and Guyon (2024) for their joint

³<https://optionmetrics.com/>

calibration.

Two further distinct lines of research are worth being mentioned as well: first, martingale optimal transport and second rough volatility.

The martingale optimal transport approach is used to calibrate discrete-time models as proposed in [Guyon \(2020b, 2024\)](#). These models are closely related to Schrödinger bridge problems, where the idea is to calibrate only the drift of the volatility while keeping the volatility of volatility unchanged, see e.g. [Guo et al. \(2022a\)](#) as well as the references therein regarding an optimal transport approach. Although the calibration within that setting is accurate, it is also computationally rather expensive and not amenable to calibrate to several maturities jointly. These computational challenges have been tackled recently in [Guyon and Bourgey \(2022\)](#).

In the area of rough volatility modeling, initiated by the seminal paper of [Gatheral et al. \(2018\)](#), the main idea is to replace the standard Brownian motion in the volatility process by a fractional Brownian motion. Even though the roughness of the trajectories found in [Gatheral et al. \(2018\)](#), can also be related to the estimation procedure as discussed e.g. in [Cont and Das \(2023\)](#), the non-Markovianity given by the fractional Brownian motion with Hurst parameter $H < 0.5$, is well-suited to reproduce certain stylized facts arising in financial data, e.g. volatility persistence or multiple scales of mean reversion; see [Bayer et al. \(2016\)](#). Several classical models have been enhanced with rougher noise, but for simplicity we here only mention those employed in the SPX/VIX calibration. One example is the quadratic rough Heston model introduced in [Gatheral et al. \(2020\)](#), which was in turn calibrated in [Rosenbaum and Zhang \(2021\)](#) by relying on neural networks approaches, also exploited in e.g. [Bayer et al. \(2019\)](#). In [Rømer \(2022\)](#) an exhaustive study of the flexibility of different rough and non-rough volatility models for the joint SPX/VIX calibration is carried out, including the rough Bergomi and the rough Heston model. Some of these, for instance the rough Heston model, have an affine structure i.e., can be embedded in the class of affine Volterra processes. In particular they allow for Fourier pricing after solving the related fractional Riccati equations. This underlying structure is the building block of an extension with jumps investigated in [Bondi et al. \(2024a,b\)](#). We refer additionally to [Di Nunno et al. \(2023\)](#); [Gazzani and Guyon \(2024\)](#) for a very recent literature review on volatility modeling.

Concerning our framework, signature-based methods provide a generic non-parametric way to extract characteristic features (linearly) and path-dependency from data, which is essential in (machine) learning and calibration tasks in finance. This explains why these techniques become more and more popular in mathematical finance, see e.g., [Bühler et al. \(2020\)](#); [Kalsi et al. \(2020\)](#); [Arribas et al. \(2020\)](#); [Lyons et al. \(2019\)](#); [Ni et al. \(2022\)](#); [Bayer et al. \(2021\)](#); [Min and Hu \(2021\)](#); [Cuchiero et al. \(2022\)](#); [Akyildirim et al. \(2023\)](#); [Ning et al. \(2023\)](#); [Wiese et al. \(2023\)](#); [Cohen et al. \(2023\)](#); [Lemahieu et al. \(2023\)](#), Chapter 1, and the references therein.

2.2 Signature: definition and properties

We start by introducing basic notions related to the definition of the signature of an $\mathbb{R}^d$ -valued continuous semimartingale. This is similar as in [Cuchiero et al. \(2023a\)](#) or [Bayer et al. \(2021\)](#),

but to keep the chapter self-contained we recall the essential definitions and properties.

For each $n \in \mathbb{N}_0$ we define recursively the n -fold tensor product of $\mathbb{R}^d$,

$$(\mathbb{R}^d)^{\otimes 0} := \mathbb{R}, \quad (\mathbb{R}^d)^{\otimes n} := \underbrace{\mathbb{R}^d \otimes \cdots \otimes \mathbb{R}^d}_n.$$

For $d \in \mathbb{N}$, we define the extended tensor algebra on $\mathbb{R}^d$ as

$$T((\mathbb{R}^d)) := \{\mathbf{a} := (a_0, \dots, a_n, \dots) : a_n \in (\mathbb{R}^d)^{\otimes n}\}.$$

Similarly we introduce the truncated tensor algebra of order $n \in \mathbb{N}$

$$T^{(n)}(\mathbb{R}^d) := \{\mathbf{a} \in T((\mathbb{R}^d)) : a_m = 0, \forall m > n\},$$

and the tensor algebra $T(\mathbb{R}^d) := \bigcup_{n \in \mathbb{N}} T^{(n)}(\mathbb{R}^d)$. Note that $T^{(n)}(\mathbb{R}^d)$ has dimension

$$d_n := (d^{n+1} - 1)/(d - 1). \quad (2.2.1)$$

For each $\mathbf{a}, \mathbf{b} \in T((\mathbb{R}^d))$ and $\lambda \in \mathbb{R}$ we set

$$\begin{aligned} \mathbf{a} + \mathbf{b} &:= (a_0 + b_0, \dots, a_n + b_n, \dots), \\ \lambda \cdot \mathbf{a} &:= (\lambda a_0, \dots, \lambda a_n, \dots), \\ \mathbf{a} \otimes \mathbf{b} &:= (c_0, \dots, c_n, \dots), \end{aligned}$$

where $c_n := \sum_{k=0}^n a_k \otimes b_{n-k}$. Observe that $(T((\mathbb{R}^d)), +, \cdot, \otimes)$ is a real non-commutative algebra.

For a multi-index $I := (i_1, \dots, i_n)$ we set $|I| := n$. We also consider the empty index $I := \emptyset$ and set $|I| := 0$. If $n \geq 1$ or $n \geq 2$ we set $I' := (i_1, \dots, i_{n-1})$, and $I'' := (i_1, \dots, i_{n-2})$, respectively. We also use the notation

$$\{I : |I| = n\} := \{1, \dots, d\}^n,$$

omitting the parameter d whenever this does not introduce ambiguity. Observe that multi-indices can be identified with words, as it is done for instance in [Lyons et al. \(2019\)](#).

Next, for each $|I| \geq 1$ we set

$$e_I := e_{i_1} \otimes \cdots \otimes e_{i_n}.$$

Observe that the set $\{e_I : |I| = n\}$ is an orthonormal basis of $(\mathbb{R}^d)^{\otimes n}$. Denoting by $e_\emptyset$ the basis element corresponding to $(\mathbb{R}^d)^{\otimes 0}$, each element of $\mathbf{a} \in T((\mathbb{R}^d))$ can thus be written as

$$\mathbf{a} = \sum_{|I| \geq 0} \mathbf{a}_I e_I,$$

for some $\mathbf{a}_I \in \mathbb{R}$. Note that if $a_n \in (\mathbb{R}^d)^{\otimes n}$ we use non-bold notation whereas for the components $\mathbf{a}_I \in \mathbb{R}$ we write them bold. Finally, for each $\mathbf{a} \in T(\mathbb{R}^d)$ and each $\mathbf{b} \in T((\mathbb{R}^d))$ we set

$$\langle \mathbf{a}, \mathbf{b} \rangle := \sum_{|I| \geq 0} \langle \mathbf{a}_I, \mathbf{b}_I \rangle.$$

Observe in particular that $\mathbf{b}_I = \langle e_I, \mathbf{b} \rangle$.

In the present work it will be useful to enumerate the elements of the truncated tensor algebra. To this extent we introduce the isomorphism $\mathbf{vec} : T^{(n)}(\mathbb{R}^d) \rightarrow \mathbb{R}^{d_n}$ and an injective labeling function $\mathcal{L} : \{I : |I| \leq n\} \rightarrow \{1, \dots, d_n\}$, such that

$$\mathbf{vec}(\mathbf{u}) := \sum_{|I| \leq n} e_{\mathcal{L}(I)} \mathbf{u}_I, \quad (2.2.2)$$

where d_n is as in (2.2.1).

Throughout the chapter we fix a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{Q})$ on which we consider the stochastic processes to be defined. We are now ready to introduce the signature of an $\mathbb{R}^d$ -valued continuous semimartingale.

Definition 2.2.1. Let X be a continuous $\mathbb{R}^d$ -valued semimartingale with $d \geq 1$. The *signature* of X is the $T((\mathbb{R}^d))$ -valued process $(s, t) \mapsto \mathbb{X}_{s,t}$ whose components are recursively defined as

$$\langle e_\emptyset, \mathbb{X}_{s,t} \rangle := 1, \quad \langle e_I, \mathbb{X}_{s,t} \rangle := \int_s^t \langle e_{I'}, \mathbb{X}_{s,r} \rangle \circ dX_r^{i_n},$$

for each $I = (i_1, \dots, i_n)$, $I' = (i_1, \dots, i_{n-1})$ and $0 \leq s \leq t$, where $\circ$ denotes the Stratonovich integral. Its projection $\mathbb{X}^n$ on $T^{(n)}(\mathbb{R}^d)$ is given by

$$\mathbb{X}_{s,t}^n = \sum_{|I| \leq n} \langle e_I, \mathbb{X}_{s,t} \rangle e_I$$

and is called *signature of X truncated at level n* . If $s = 0$, we use the notation $\mathbb{X}_t$ and $\mathbb{X}_t^n$, respectively.

Observe that the signature of X and the signature of $X - c$ coincide for each $c \in \mathbb{R}$. Moreover, with an equivalent notation we can write

$$\begin{aligned} \mathbb{X}_t = & \left(1, \int_0^t 1 \circ dX_s^1, \dots, \int_0^t 1 \circ dX_s^d, \int_0^t \left(\int_0^s 1 \circ dX_r^1 \right) \circ dX_s^1, \right. \\ & \left. \int_0^t \left(\int_0^s 1 \circ dX_r^1 \right) \circ dX_s^2, \dots, \int_0^t \left(\int_0^s 1 \circ dX_r^d \right) \circ dX_s^d, \dots \right). \end{aligned}$$

A well-known and extremely useful property of the signature is that every polynomial function in the signature has a linear representation. For the precise statement we first need to introduce the following concept (see also Definition 2.4 in [Lyons et al. \(2019\)](#) or Section 2.2. in [Bayer et al. \(2021\)](#)).

Definition 2.2.2. For every two multi-indices I and J the *shuffle product* is defined recursively as

$$e_I \sqcup e_J := (e_{I'} \sqcup e_J) \otimes e_{i_{|I|}} + (e_I \sqcup e_{J'}) \otimes e_{j_{|J|}},$$

with $e_I \sqcup e_\emptyset := e_\emptyset \sqcup e_I = e_I$. It extends to $\mathbf{a}, \mathbf{b} \in T(\mathbb{R}^d)$ as

$$\mathbf{a} \sqcup \mathbf{b} = \sum_{|I|, |J| \geq 0} \mathbf{a}_I \mathbf{b}_J (e_I \sqcup e_J).$$

Observe that $(T(\mathbb{R}^d), +, \sqcup)$ is a commutative algebra, which in particular means that the shuffle product is associative and commutative.

In the following proposition we summarize some useful properties of the signature. These results have been developed in the rough paths literature (see for instance [Ree \(1958\)](#) or [Lyons et al. \(2007\)](#) for the shuffle property, [Boedihardjo et al. \(2016\)](#) for the uniqueness of the signature, and [Chen \(1957, 1977\)](#) for Chen's identity) and have then been refined in the context of semimartingales (see e.g., [Bayer et al. \(2021\)](#) and Chapter 1). For a more detailed exposition and proofs we refer to [Cuchiero et al. \(2023a\)](#).

Proposition 2.2.3. Let X and Y be two continuous $\mathbb{R}^d$ -valued semimartingales with $X_0 = Y_0 = 0$. Then the following properties hold.

Shuffle property For each two multi-indices I, J and each $0 \leq s \leq t$ it holds

$$\langle e_I, \mathbb{X}_{s,t} \rangle \langle e_J, \mathbb{X}_{s,t} \rangle = \langle e_I \sqcup e_J, \mathbb{X}_{s,t} \rangle. \quad (2.2.3)$$

Uniqueness of the signature Set $\hat{X}_t := (t, X_t)$, $\hat{Y}_t := (t, Y_t)$ and let $\hat{\mathbb{X}}$ and $\hat{\mathbb{Y}}$ be the corresponding signature processes. Then the signature $\hat{\mathbb{X}}_T = \hat{\mathbb{Y}}_T$ if and only if $X_t = Y_t$ for each $t \in [0, T]$.

Chen's identity For each $0 \leq s \leq u \leq t$ it holds

$$\mathbb{X}_{s,t} = \mathbb{X}_{s,u} \otimes \mathbb{X}_{u,t}. \quad (2.2.4)$$

This can equivalently be written as

$$\langle e_I, \mathbb{X}_{s,t} \rangle = \sum_{e_{I_1} \otimes e_{I_2} = e_I} \langle e_{I_1}, \mathbb{X}_{s,u} \rangle \langle e_{I_2}, \mathbb{X}_{u,t} \rangle, \quad (2.2.5)$$

for each multi-index I .

Universal approximation theorem For each $n \in \mathbb{N}$ consider the sets

$$\mathcal{S}^{(n)} := \{(\hat{\mathbb{X}}_t^n)_{t \in [0, T]}(\omega) : \omega \in \Omega\}$$

and let $S^{(n)} : \mathcal{S}^{(2)} \rightarrow \mathcal{S}^{(n)}$ denote the corresponding Lyons lift. Then it holds that $S^{(n)}((\hat{\mathbb{X}}_t^2)_{t \in [0, T]}) = (\hat{\mathbb{X}}_t^n)_{t \in [0, T]}$ almost surely. Consider then a generic distance $d_{\mathcal{S}^{(2)}}$ on the set of trajectories given by $\mathcal{S}^{(2)}$, with respect to which the map from $\mathcal{S}^{(2)}$ to $\mathbb{R}$ given by

$$\hat{\mathbf{x}}^2 \mapsto \langle e_I, S^{(|I|)}(\hat{\mathbf{x}}^2)_t \rangle$$

is continuous for each multi-index I and every $t \in [0, T]$. Let K be a compact subset of $\mathcal{S}^{(2)}$ and consider a continuous map $f : K \rightarrow \mathbb{R}$. Then for every $\varepsilon > 0$ there exists some $\ell \in T(\mathbb{R}^d)$ such that

$$\sup_{(\hat{\mathbb{X}}_t^2)_{t \in [0, T]} \in K} |f((\hat{\mathbb{X}}_t^2)_{t \in [0, T]}) - \langle \ell, \hat{\mathbb{X}}_T \rangle| < \varepsilon,$$

almost surely.

2.3 The model

We start by introducing the concept of polynomial diffusions (see [Cuchiero et al. \(2012\)](#); [Filipović and Larsson \(2016\)](#)) which will play a key role for the computation of the conditional expected signature. Here we denote by $\sqrt{\cdot}$ the matrix square root.

Definition 2.3.1. Suppose that an $\mathbb{R}^d$ -valued process $X = (X_t)_{t \geq 0}$ is a weak solution of

$$dX_t = b(X_t)dt + \sqrt{a(X_t)}dW_t, \quad X_0 = x_0$$

for some d -dimensional Brownian motion W and some maps $a : \mathbb{R}^d \rightarrow \mathbb{S}_+^d$ and $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that a_{ij} is a polynomial of degree at most 2 and b_j is a polynomial of degree at most 1 for each $i, j \in \{1, \dots, d\}$. Then we call X *polynomial diffusion*.

We are now ready to introduce the model $(S_t)_{t \geq 0}$ for the discounted, dividend-adjusted dynamics of the S&P 500 index already outlined in the introduction. Its dynamics under a risk-neutral probability measure $\mathbb{Q}$ are given by

$$dS_t = S_t \sigma_t^S dB_t, \quad (2.3.1)$$

where $S_0 \in \mathbb{R}^+$, $\sigma^S = (\sigma_t^S)_{t \geq 0}$ is the volatility process to be specified and $B = (B_t)_{t \geq 0}$ is a one-dimensional Brownian motion, correlated with σ^S . We define additionally the instantaneous variance via $V_t := (\sigma_t^S)^2$ for every $t \geq 0$. Our modeling choice is to parametrize the volatility process σ^S as a linear function of the time-extended signature of a primary process X , namely

$$\sigma_t^S(\ell) := \ell_\emptyset + \sum_{0 < |I| \leq n} \ell_I \langle e_I, \widehat{X}_t \rangle, \quad (2.3.2)$$

where

- $(X_t)_{t \geq 0}$ and thus also $\widehat{X} = (t, X_t)_{t \geq 0}$ is a polynomial diffusion (with values in $\mathbb{R}^d$ and $\mathbb{R}^{d+1}$ respectively) in the sense of Definition 2.3.1.
- $\ell := \{\ell_I \in \mathbb{R} : |I| \leq n\}$ denotes the collection of parameters of the model, i.e., $\ell \in \mathbb{R}^{(d+1)_n}$.

We then denote by ρ the correlation matrix process between the components of X , i.e.

$$\rho_{ij} = \frac{[X^i, X^j]}{\sqrt{[X^i]} \sqrt{[X^j]}} \in [-1, 1],$$

for all $i, j = 1, \dots, d$, where $[\cdot, \cdot]$ denotes the quadratic covariation.

In order to simplify the notation we will drop the dependence on ℓ for the processes $S = (S_t)_{t \geq 0}$ and $(\sigma_t^S)_{t \geq 0}$ as in (2.3.1), whenever this does not cause any confusion.

Remark 2.3.2. As an alternative definition for the volatility process $(\sigma_t^S)_{t \geq 0}$ one can set

$$\sigma_t^S(\ell) := \ell_\emptyset + \sum_{0 < |I| \leq n} \ell_I \langle e_I, \widehat{\mathbb{X}}_{t-\varepsilon, t} \rangle,$$

for some fixed $\varepsilon > 0$. In this case the value of the volatility process σ^S at time t does not depend on the whole trajectory of the primary process X , but just on its evolution from $t - \varepsilon$ to t . For an economically reasonable choice for ε the lags used in Section 3.4 of [Guyon and Lekeufack \(2023\)](#) can be adapted to the current setting.

Remark 2.3.3 (Interest rates and dividends). In the model given by (2.3.1) we describe the discounted, dividend-adjusted prices and construct the VIX from them, in line with the definition of the CBOE for the computation of the VIX. However, contingent claims are often expressed in terms of undiscounted, unadjusted prices. If the dynamics of the discounted, dividend-adjusted price process are given by (2.3.1), the undiscounted, unadjusted one is denoted by $\tilde{S}$ and fulfills

$$d\tilde{S}_t = (r - q)\tilde{S}_t dt + \tilde{S}_t \sigma_t^S(\ell) dB_t,$$

where here $r, q \in \mathbb{R}$ denote the interest rate and the dividend, respectively. Therefore $\tilde{S}_t(\ell) = e^{(r-q)t} S_t(\ell)$ and the price of a call option on the S&P 500 index under our model, reads

$$C(T, K) = \mathbb{E} \left[e^{-rT} (\tilde{S}_T(\ell) - K)^+ \right] = \mathbb{E} [e^{-rT} (e^{(r-q)T} S_T(\ell) - K)^+]$$

where $T > 0$ denotes the maturity time and $K \in \mathbb{R}$ the undiscounted strike price.

It is worth mentioning that the pool of eligible primary processes is rather wide, including for example correlated Brownian motions, geometric Brownian motions, OU processes, Cox-Ingersoll-Ross (CIR) processes, Jacobi processes, and all continuous affine processes.

The reason why we require the primary process to be a polynomial diffusion is due to the tractability properties of the truncated signature $\widehat{\mathbb{X}}^n$ under this assumption. We will indeed see in Section 2.4 that in this case the (conditional) truncated expected signature of $\widehat{X}$ can be computed by solving a finite-dimensional ODE, i.e., can be written in terms of a matrix exponential.

Remark 2.3.4. We illustrate here that several classical and also recently considered stochastic volatility models are nested within our modeling choice (2.3.2).

- Suppose that $(X_t)_{t \geq 0}$ is a 1-dimensional OU process and let the order of the signature be $n = 1$, with $\ell_\emptyset = \ell_{(0)} = 0$ and $\ell_{(1)} \neq 0$. Then the process $S = (S_t)_{t \geq 0}$ coincides with the Stein-Stein model, as introduced in [Stein and Stein \(1991\)](#).
- Suppose that $(X_t)_{t \geq 0}$ is a 1-dimensional geometric Brownian motion without drift and let the order of the signature be $n = 1$, with $\ell_\emptyset = \ell_{(0)} = 0$ and $\ell_{(1)} \neq 0$. Then the process $S = (S_t)_{t \geq 0}$ coincides with the SABR model, as introduced in initially in [Hagan et al. \(2002\)](#) with $\beta = 1$.

- Suppose that $(X_t)_{t \geq 0}$ is a 1-dimensional OU process and let the order of the signature be $n = 5$, with $\ell_\emptyset, \ell_{(1)}, \ell_{(1,1,1)}, \ell_{(1,1,1,1,1)}$ non-zero and $\ell_I = 0$ otherwise. Then the process $S = (S_t)_{t \geq 0}$ coincides with the model considered in [Abi Jaber et al. \(2022a,b\)](#) with an exponential kernel (a part from the deterministic input curve considered there additionally). Going beyond the assumption of $\hat{X}$ being a polynomial diffusion we may allow for $(X_t)_{t \geq 0}$ to be a one-dimensional fractional Brownian motion, thus leaving the semimartingale setting. And if we do not consider the time augmentation, we can also include fractional kernels and therefore the whole class of Gaussian polynomial volatility models introduced in [Abi Jaber et al. \(2022a\)](#) within our framework.

Remark 2.3.5. As indicated in the last point of the previous remark, our framework can be extended beyond the semimartingale case as long as the trajectories of the corresponding process can be enhanced to be almost surely a weakly geometric p -rough path. This holds for instance true for the case of time-augmented multidimensional fractional Brownian motion when $H \in (1/4, 1)$, since for any $p \in (1/H, 4)$ there exists an almost surely weakly geometric p -rough path, such that the projection on the first component coincides with the process' increments. For this result we refer to [Coutin and Qian \(2002\)](#), Theorem 2. Observe that the case considered in [Abi Jaber et al. \(2022a\)](#) is simpler since it is a one dimensional setting, meaning that the corresponding signature boils down to Taylor polynomials of fractional Brownian motion.

Note however, while our framework can be extended beyond the semimartingale case as long as signatures can be defined, our methodology to compute conditional truncated expected signatures via finite dimensional matrix exponentials only works in the polynomial diffusion setting. The same applies to the linear representation of the log-price provided in Section 2.6.

Remark 2.3.6. Let X be a 1-dimensional OU-process, such that without loss of generality $X_0 = 0$, i.e.,

$$dX_t = \kappa(\theta - X_t)dt + \sigma dW_t.$$

Then, for $n = 2$ the instantaneous dynamics of the volatility process are given by

$$\begin{aligned} d\sigma_t &= \ell_0 dt + \ell_1 dX_t + \ell_{00} t dt + \ell_{01} t dX_t + \ell_{10} X_t dt + \ell_{11} dX_t^2 \\ &= (\ell_0 + \ell_{00} t + \ell_1 \kappa(\theta - X_t) + \ell_{01} t \kappa(\theta - X_t) + \ell_{10} X_t + 2\ell_{11} X_t(\kappa(\theta - X_t) + \sigma^2))dt \\ &\quad + (\ell_1 \sigma + \ell_{01} t \sigma + 2\ell_{11} \sigma X_t) dW_t, \end{aligned}$$

which can be rewritten as

$$d\sigma_t = (f_1(t) + cX_t(g(t) - X_t))dt + (f_2(t) + \tilde{c}X_t)dW_t,$$

where f_1, f_2, g are affine functions of time and $c, \tilde{c} \in \mathbb{R}$, all depending on the model parameters $\{\ell_I, |I| \leq n\}$. The previous simple derivation implies:

- If $n = 1$ the instantaneous vol of vol is constant and given by $|\ell_1 \sigma|$.
- If $n \geq 2$ the instantaneous vol of vol is stochastic, depending explicitly on X_t .

- For $n = 2$, the instantaneous volatility exhibits a stochastic mean reversion rate given by the term cX_t , with a time-dependent long-run mean by the affine function $g(t)$. We will see in the subsequent sections that this type of model with a 3-dimensional OU-process is flexible enough to solve the joint calibration problem.
- Notice that even for $n = 2$, the choice $X = W$, i.e. choosing just a Brownian motion (as for instance in [Arribas et al. \(2020\)](#); [Cuchiero et al. \(2023a\)](#) for the price process), would lead to restrictive dynamics of the instantaneous volatility.

2.4 Expected signature of polynomial diffusion

Let $(Y_t)_{t \geq 0}$ be a polynomial diffusion in sense of Definition 2.3.1 whose dynamics are given by

$$dY_t = b(Y_t)dt + \sigma(Y_t)dW_t, \quad Y_0 = y_0, \quad (2.4.1)$$

where $\sigma(Y_t)$ denotes the matrix square root of $a(Y_t)$. Recall that in this case the components of $a : \mathbb{R}^d \rightarrow \mathbb{S}_+^d$ are polynomials of degree at most 2, the components of $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ are polynomials of degree at most 1, and $W = (W_t)_{t \geq 0}$ is a d -dimensional Brownian motion. Denote then by $\mathbb{Y}$ the corresponding signature.

We now explain how to employ the polynomial technology to compute the conditional truncated expected signature of $(Y_t)_{t \geq 0}$. The corresponding code is available at [saras-valuto/AffPolySig](#). Several representations of related quantities in particular for the Brownian case can be found in the literature, see for instance [Fawcett \(2003\)](#), [Lyons and Victoir \(2004\)](#), [Ni \(2012\)](#), [Lyons and Ni \(2015\)](#), [Boedihardjo et al. \(2021\)](#), [Cass and Ferrucci \(2024\)](#), [Lyons et al. \(2024\)](#). Our approach follows [Cuchiero et al. \(2023c\)](#) and is based on the classical theory of polynomial processes (see [Cuchiero et al. \(2012\)](#) and [Filipović and Larsson \(2016\)](#)). Even though results for the corresponding infinite dimensional stochastic processes (see for instance [Cuchiero and Svaluto-Ferro \(2021\)](#); [Cuchiero et al. \(2024\)](#)) are needed in the case of general signature SDEs considered in [Cuchiero et al. \(2023c\)](#), the polynomial property of $(Y_t)_{t \geq 0}$ here permits to stay in the finite dimensional setting.

Lemma 2.4.1. Let $(Y_t)_{t \geq 0}$ be the polynomial diffusion given by (2.4.1) and b and a be the corresponding drift and diffusion coefficients. Then

$$b_j(y) = b_j^c + \sum_{k=1}^d b_j^k y_k \quad \text{and} \quad a_{ij}(y) = a_{ij}^c + \sum_{k=1}^d a_{ij}^k y_k + \sum_{k,h=1}^d a_{ij}^{kh} y_k y_h,$$

for some $b_j^c, b_j^k, a_{ij}^c, a_{ij}^k, a_{ij}^{kh} = a_{ij}^{hk} \in \mathbb{R}$. Moreover, $b_j(Y_t) = \langle \mathbf{b}_j, \mathbb{Y}_t^1 \rangle$ and $a_{ij}(Y_t) = \langle \mathbf{a}_{ij}, \mathbb{Y}_t^2 \rangle$

for

$$\begin{aligned} \mathbf{b}_j &= \left(b_j^c + \sum_{k=1}^d b_j^k Y_0^k \right) e_\emptyset + \sum_{k=1}^d b_j^k e_k \quad \text{and} \\ \mathbf{a}_{ij} &= \left(a_{ij}^c + \sum_{k=1}^d a_{ij}^k Y_0^k + \sum_{k,h=1}^d a_{ij}^{kh} Y_0^k Y_0^h \right) e_\emptyset + \sum_{k=1}^d \left(a_{ij}^k + 2 \sum_{h=1}^d a_{ij}^{kh} Y_0^h \right) e_k \\ &\quad + \sum_{k,h=1}^d a_{ij}^{kh} e_k \sqcup e_h. \end{aligned}$$

Observe that the upper index on Y_0^k and Y_0^h refers to Y 's components and not to powers.

Proof. The first part follows by the observation that by definition of polynomial diffusion b and a are polynomials of degree at most 1 and 2, respectively. For the second part it then suffices to note that $\langle e_\emptyset, \mathbb{Y}_t^1 \rangle = \langle e_\emptyset, \mathbb{Y}_t^2 \rangle = 1$, $\langle e_k, \mathbb{Y}_t^1 \rangle = \langle e_k, \mathbb{Y}_t^2 \rangle = (Y_t^k - Y_0^k)$, and $\langle e_k \sqcup e_h, \mathbb{Y}_t^2 \rangle = (Y_t^k - Y_0^k)(Y_t^h - Y_0^h)$. $\square$

Lemma 2.4.2. Let $(Y_t)_{t \geq 0}$ be the polynomial diffusion given by (2.4.1) and let $\mathbf{b}$ and $\mathbf{a}$ as in Lemma 2.4.1. The truncated signature $(\mathbb{Y}_t^n)_{t \geq 0}$ is a polynomial diffusion in the sense of Definition 2.3.1 and for each $|I| \leq n$ it holds that

$$\langle e_I, \mathbb{Y}_t^n \rangle = \int_0^t \langle L e_I, \mathbb{Y}_s^n \rangle ds + \int_0^t \langle e_{I'}, \mathbb{Y}_s^n \rangle \sigma_{i_{|I|}}(Y_s) dW_s,$$

where the operator $L : T((\mathbb{R}^d)) \rightarrow T((\mathbb{R}^d))$ satisfies $L(T^{(n)}(\mathbb{R}^d)) \subseteq T^{(n)}(\mathbb{R}^d)$ and is given by

$$L e_I = e_{I'} \sqcup \mathbf{b}_{i_{|I|}} + \frac{1}{2} e_{I''} \sqcup \mathbf{a}_{i_{|I|-1} i_{|I|}}. \quad (2.4.2)$$

Proof. Let $\sigma_j(Y_t)$ denote the j -th row of $\sigma(Y_t)$. By definition of the signature, Stratonovich integral and by the shuffle property we can compute

$$\begin{aligned} \langle e_I, \mathbb{Y}_t \rangle &= \int_0^t \langle e_{I'}, \mathbb{Y}_s \rangle \circ d \langle e_{i_{|I|}}, Y_s \rangle \\ &= \int_0^t \langle e_{I'}, \mathbb{Y}_s \rangle d \langle e_{i_{|I|}}, Y_s \rangle + \frac{1}{2} \left[\langle e_{I'}, \mathbb{Y} \rangle, \langle e_{i_{|I|}}, Y \rangle \right]_t \\ &= \int_0^t \langle e_{I'}, \mathbb{Y}_s \rangle d \langle e_{i_{|I|}}, Y_s \rangle + \frac{1}{2} \int_0^t \langle e_{I''}, \mathbb{Y}_s \rangle d [\langle e_{i_{|I|-1}}, Y \rangle, \langle e_{i_{|I|}}, Y \rangle]_s \\ &= \int_0^t \langle e_{I'}, \mathbb{Y}_s \rangle \langle \mathbf{b}_{i_{|I|}}, \mathbb{Y}_s \rangle ds + \int_0^t \langle e_{I'}, \mathbb{Y}_s \rangle \sigma_{i_{|I|}}(Y_s) dW_s \\ &\quad + \frac{1}{2} \int_0^t \langle e_{I''}, \mathbb{Y}_s \rangle \langle \mathbf{a}_{i_{|I|-1} i_{|I|}}, \mathbb{Y}_s \rangle ds \\ &= \int_0^t \langle e_{I'} \sqcup \mathbf{b}_{i_{|I|}}, \mathbb{Y}_s \rangle ds + \int_0^t \langle e_{I'}, \mathbb{Y}_s \rangle \sigma_{i_{|I|}}(Y_s) dW_s + \frac{1}{2} \int_0^t \langle e_{I''} \sqcup \mathbf{a}_{i_{|I|-1} i_{|I|}}, \mathbb{Y}_s \rangle ds \\ &= \int_0^t \langle L e_I, \mathbb{Y}_s \rangle ds + \int_0^t \langle e_{I'}, \mathbb{Y}_s \rangle \sigma_{i_{|I|}}(Y_s) dW_s, \end{aligned}$$

for each $|I| \geq 0$. Since $|I \sqcup J| = |I| + |J|$ it holds that $L(T^{(n)}(\mathbb{R}^d)) \subseteq T^{(n)}(\mathbb{R}^d)$. For $|I| \leq n$ we thus get that the corresponding drift's components are linear maps in $\mathbb{Y}^n$. Similarly, since $\mathbf{a}_{ij} = \sum_{|I|, |J| \leq 1} \lambda_{ij}^{IJ} e_I \sqcup e_J$ for some $\lambda_{ij}^{IJ} \in \mathbb{R}$ and for $|I| \leq n$ we can compute

$$\begin{aligned} \langle e_{I'}, \mathbb{Y}_s \rangle \sigma_{i_{|I|}}(Y_s) (\langle e_{J'}, \mathbb{Y}_s \rangle \sigma_{i_{|J|}}(Y_s))^\top &= \langle e_{I'}, \mathbb{Y}_s \rangle \langle e_{J'}, \mathbb{Y}_s \rangle \langle \mathbf{a}_{i_{|I|} j_{|J|}}, \mathbb{Y}_s \rangle \\ &= \sum_{|H_1|, |H_2| \leq 1} \lambda_{i_{|I|} j_{|J|}}^{H_1 H_2} \langle e_{I'} \sqcup e_{H_1}, \mathbb{Y}_s \rangle \langle e_{J'} \sqcup e_{H_2}, \mathbb{Y}_s \rangle, \end{aligned}$$

we also have that the components of the corresponding diffusion matrix are polynomials of degree 2 in $\mathbb{Y}^n$. Lemma 2.2 in Filipović and Larsson (2016) yields the polynomial property. $\square$

Since the linear operator L maps the finite dimensional vector space $T^{(n)}(\mathbb{R}^d)$ to itself, it admits a matrix representation.

Definition 2.4.3. We call the operator L defined in (2.4.2) *dual operator corresponding to* $\mathbb{Y}$. For each $|I| \leq n$ set then $\eta_{IJ} \in \mathbb{R}$ such that

$$Le_I = \sum_{|J| \leq n} \eta_{IJ} e_J,$$

and fix a labelling injective function $\mathcal{L} : \{I : |I| \leq n\} \rightarrow \{1, \dots, d_n\}$ as introduced before (2.2.2). We then call the matrix $G \in \mathbb{R}^{d_n \times d_n}$ given by

$$G_{\mathcal{L}(I)\mathcal{L}(J)} := \eta_{IJ}, \quad (2.4.3)$$

the d_n -dimensional matrix representative of L .

Observe that using the notation of (2.2.2), for each $\mathbf{u} \in T^{(n)}(\mathbb{R}^d)$ the matrix representative G of L satisfies

$$\mathbf{vec}(L\mathbf{u}) = G\mathbf{vec}(\mathbf{u}).$$

Theorem 2.4.4. Let $(Y_t)_{t \geq 0}$ be the polynomial diffusion given by (2.4.1), $(\mathcal{F}_t)_{t \geq 0}$ be the filtration generated by $(Y_t)_{t \geq 0}$ and let G be the d_n -dimensional matrix representative of the dual operator corresponding to $\mathbb{Y}$. Then for each $T, t \geq 0$ and each $|I| \leq n$ it holds

$$\mathbb{E}[\mathbf{vec}(\mathbb{Y}_{T+t}^n) | \mathcal{F}_T] = e^{tG^\top} \mathbf{vec}(\mathbb{Y}_T^n),$$

or equivalently,

$$\mathbb{E}[\langle e_I, \mathbb{Y}_{T+t}^n \rangle | \mathcal{F}_T] = \sum_{|J| \leq n} (e^{tG^\top})_{\mathcal{L}(I)\mathcal{L}(J)} \langle e_J, \mathbb{Y}_T^n \rangle,$$

where $e^{(\cdot)}$ denotes the matrix exponential.

Proof. By Lemma 2.4.2 we know that $\mathbf{vec}(\mathbb{Y}^n)$ is a polynomial diffusion and Theorem 3.1 in Filipović and Larsson (2016) for polynomials of degree 1 yields the claim. $\square$

Example 2.4.5. For the present chapter a crucial role is played by the polynomial diffusion given by time, a d -dimensional OU process, and a Brownian motion. Specifically, we consider the process $\widehat{Z}_t := (\widehat{X}_t, B_t)$ where B is a Brownian motion and $\widehat{X}_t = (t, X_t)$ with

$$dX_t^j = \kappa^j(\theta^j - X_t^j)dt + \sqrt{a(X_t)}dW_t, \quad X_0 = x_0,$$

for $a_{ij}(X_t) = \sigma^i \sigma^j \rho_{ij}$, and W being a d -dimensional Brownian motion. We denote by $\rho_{j(d+1)}$ the correlation between X^j and B . Setting $\kappa^{d+1} := 0$ and $\sigma^{d+1} := 1$ we can see that $\widehat{Z}$ satisfies (2.4.1) in $d+2$ dimensions for

$$b_j(\widehat{Z}_t) = 1_{\{j=0\}} + \kappa^j(\theta^j - \widehat{Z}_t^j)1_{\{j \neq 0\}} \quad \text{and} \quad a_{ij}(\widehat{Z}_t) = \sigma^i \sigma^j \rho_{ij} 1_{\{i,j \neq 0\}}.$$

The corresponding $\mathbf{b}$ and $\mathbf{a}$ are given by $\mathbf{b}_j = e_0(1_{\{j=0\}} + \kappa^j(\theta^j - \widehat{Z}_0^j)1_{\{j \neq 0\}}) - e_j \kappa^j 1_{\{j \neq 0\}}$ and $\mathbf{a}_{ij} = e_0 \sigma^i \sigma^j \rho_{ij} 1_{\{i,j \neq 0\}}$ and we thus get

$$\begin{aligned} Le_I &= e_{I'}(1_{\{i_{|I|}=0\}} + \kappa^{i_{|I|}}(\theta^{i_{|I|}} - \widehat{Z}_0^{i_{|I|}})1_{\{i_{|I|} \neq 0\}}) - (e_{I'} \sqcup e_{i_{|I|}})\kappa^{i_{|I|}}1_{\{i_{|I|} \neq 0\}} \\ &\quad + \frac{1}{2}e_{I''}\sigma^{i_{|I|-1}}\sigma^{i_{|I|}}\rho_{i_{|I|-1},i_{|I|}}1_{\{i_{|I|-1},i_{|I|} \neq 0\}}. \end{aligned}$$

An application of L to the first basis elements yields the following results:

- $L(e_1) = e_0 \kappa^1(\theta^1 - X_0^1) - e_1 \kappa^1$;
- $L(e_I \otimes e_0) = e_I \sqcup \mathbf{b}_0 + \frac{1}{2}e_{I'} \sqcup \mathbf{a}_{i_{|I|}0} = e_I$;
- $L(e_0 \otimes e_1 \otimes e_2) = e_0 \otimes e_1 \kappa^2(\theta^2 - X_0^2) - (e_0 \otimes e_1) \sqcup e_2 \kappa^2 + \frac{1}{2}e_0 \sigma^1 \sigma^2 \rho_{12}$.

Letting $(\mathcal{F}_t)_{t \geq 0}$ be the filtration generated by $(\widehat{Z}_t)_{t \geq 0}$ by Theorem 2.4.4 we can conclude that

$$\mathbb{E}[\mathbf{vec}(\widehat{Z}_{T+t}^n) | \mathcal{F}_T] = e^{tG^\top} \mathbf{vec}(\widehat{Z}_T^n), \quad (2.4.4)$$

or equivalently,

$$\mathbb{E}[\langle e_I, \widehat{Z}_{T+t}^n \rangle | \mathcal{F}_T] = \sum_{|J| \leq n} (e^{tG^\top})_{\mathcal{L}(I)\mathcal{L}(J)} \langle e_J, \widehat{Z}_T^n \rangle, \quad (2.4.5)$$

where G denotes the $(d+2)_n$ -dimensional matrix representative of L . In order to work with the VIX it will be convenient to restrict our attention to the signature components of $(\widehat{Z}_t)_{t \geq 0}$ not involving B . The following remark will be useful.

Remark 2.4.6. Observe that given a subset $E \subseteq \{0, \dots, d+1\}$, setting $\mathcal{I}_E := \{I : i_j \in E\}$ it holds $L(\mathcal{I}_E) \subseteq \mathcal{I}_E$. This in particular implies that

$$Le_I = \sum_{J \in \mathcal{I}_E} \eta_{IJ} e_J$$

for each $I \in \mathcal{I}_E$. Choosing $E = \{0, \dots, d\}$, letting $\mathcal{L}_E : \mathcal{I}_E \rightarrow \{1, \dots, (d+1)_n\}$ be a labelling injective function, and setting $G_{\mathcal{L}_E(I)\mathcal{L}_E(J)}^E := \eta_{IJ}$ we can see that (2.4.5) reduces to

$$\mathbb{E}[\langle e_I, \widehat{X}_{T+t}^n \rangle | \mathcal{F}_T] = \sum_{|J| \leq n} (e^{t(G^E)^\top})_{\mathcal{L}_E(I)\mathcal{L}_E(J)} \langle e_J, \widehat{X}_T^n \rangle.$$

To simplify the notation we often drop the E from G^E whenever this does not introduce any confusion.

Remark 2.4.7. Let $(Y_t)_{t \geq 0}$ be a polynomial diffusion and let $\mathbb{Y}^{-1}$ be defined via $e_\emptyset = \mathbb{Y}_s^{-1} \otimes \mathbb{Y}_s$, i.e. $\langle e_\emptyset, \mathbb{Y}_s^{-1} \rangle = 1$ and

$$\sum_{e_{I_1} \otimes e_{I_2} = e_I} \langle e_{I_1}, \mathbb{Y}_s^{-1} \rangle \langle e_{I_2}, \mathbb{Y}_s \rangle = 0,$$

for each $|I| > 0$. Observe that it can be defined recursively on $|I|$ and each component of $\mathbb{Y}_s^{-1}$ corresponds to a linear combination of components of $\mathbb{Y}_s$ of the same length or shorter.

Since by Chen's identity (see (2.2.4) or (2.2.5)) we have $\mathbb{Y}_s \otimes \mathbb{Y}_{s,t} = \mathbb{Y}_t$, for each $s \leq u \leq t$ and $|I| \leq n$ we then get

$$\begin{aligned} \mathbb{E}[\langle e_I, \mathbb{Y}_{s,t} \rangle | \mathcal{F}_u] &= \mathbb{E}[\langle e_I, \mathbb{Y}_s^{-1} \otimes \mathbb{Y}_t \rangle | \mathcal{F}_u] = \sum_{e_{I_1} \otimes e_{I_2} = e_I} \langle e_{I_1}, \mathbb{Y}_s^{-1} \rangle \mathbb{E}[\langle e_{I_2}, \mathbb{Y}_t \rangle | \mathcal{F}_u] \\ &= \sum_{e_{I_1} \otimes e_{I_2} = e_I} \langle e_{I_1}, \mathbb{Y}_s^{-1} \rangle \mathbf{vec}(e_{I_2})^\top e^{(t-u)G^\top} \mathbf{vec}(\mathbb{Y}_u^n), \end{aligned}$$

where G denotes the d_n -dimensional matrix representative of the dual operator of $\mathbb{Y}$.

2.5 VIX options with signatures

In this section we discuss the implication on pricing VIX options under the model (2.3.1)-(2.3.2). The VIX index is a popular measure of the market's expected volatility of the S&P 500, calculated and published by the Chicago Board Options Exchange (CBOE). The current VIX value quotes the expected annualized change in the S&P 500 over the following 30 days, based on options-based theory and current options-market data. As stylized definition we consider

$$\text{VIX}_T^2 = \text{Price}_T \left[-\frac{2}{\Delta} \log \left(\frac{S_{T+\Delta}}{F_T^{T+\Delta}} \right) \right], \quad (2.5.1)$$

where $\Delta = 30$ days, $F_T^{T+\Delta}$ denotes the price at time T of the SPX future with maturity $T + \Delta$ and with Price_T we refer to the market price at time T of the log-contract, i.e. the payoff in (2.5.1). Hence, under a given model we define the VIX,

$$\text{VIX}_T = \sqrt{\mathbb{E} \left[-\frac{2}{\Delta} \log \left(\frac{S_{T+\Delta}}{S_T} \right) | \mathcal{F}_T \right]}, \quad (2.5.2)$$

where $\Delta = 30$ days and S_T denotes the price process at time $T > 0$. Recall that under a diffusion model, the previous expression is equivalent to

$$\text{VIX}_T = \sqrt{\frac{1}{\Delta} \mathbb{E} \left[\int_T^{T+\Delta} V_t dt | \mathcal{F}_T \right]}, \quad (2.5.3)$$

as long as $\mathbb{E}[\int_0^t V_s ds] < \infty$ for all $t \geq 0$, see e.g. Neuberger (1994); Gatheral (2011). With VIX options we here usually refer to either put or calls written on VIX. In the present work we will consider without loss of generality only call options.

2.5.1 Explicit formulas for the VIX

This section is dedicated to one of the main implication of our modeling framework, namely an explicit formula for the VIX expression (2.5.2) for S following (2.3.1)-(2.3.2). In particular we show in the next theorem that the computation of the VIX squared reduces to a quadratic form in the parameters ℓ . The entries of the corresponding positive semidefinite matrix can be computed by polynomial technology, i.e. by matrix exponential as proved in Section 2.4.

Theorem 2.5.1. *Let $S = (S_t)_{t \geq 0}$ be a price process described by*

$$dS_t = S_t \sigma_t^S dB_t,$$

where $\sigma^S = (\sigma_t^S)_{t \geq 0}$ denotes the volatility process, $B = (B_t)_{t \geq 0}$ a one-dimensional Brownian motion. Assume that σ^S satisfies (2.3.2). Following (2.2.2), fix an injective labeling function $\mathcal{L} : \{I : |I| \leq n\} \rightarrow \{1, \dots, (d+1)_{2n+1}\}$ and let G be the $(d+1)_{(2n+1)}$ -dimensional matrix representative of the dual operator corresponding to $\hat{\mathbb{X}}$. Then,

$$\mathbb{E} \left[\int_0^t V_s ds \right] < \infty \quad (2.5.4)$$

holds for every $t \geq 0$ and

$$VIX_T(\ell) = \sqrt{\frac{1}{\Delta} \ell^\top Q(T, \Delta) \ell}, \quad (2.5.5)$$

where

$$Q_{\mathcal{L}(I)\mathcal{L}(J)}(T, \Delta) = \mathbf{vec}((e_I \sqcup e_J) \otimes e_0)^\top (e^{\Delta G^\top} - \text{Id}) \mathbf{vec}(\hat{\mathbb{X}}_T^{2n+1}), \quad (2.5.6)$$

and $\text{Id} \in \mathbb{R}^{(d+1)_{2n+1} \times (d+1)_{2n+1}}$ denotes the identity matrix. More explicitly without the vectorisation this reads

$$Q_{\mathcal{L}(I)\mathcal{L}(J)}(T, \Delta) = \sum_{e_K = (e_I \sqcup e_J) \otimes e_0} \sum_{|H| \leq 2n+1} (e^{\Delta G^\top} - \text{Id})_{\mathcal{L}(K)\mathcal{L}(H)} \langle e_H, \hat{\mathbb{X}}_T \rangle.$$

Proof. Observe that

$$V_t(\ell) = \left(\sum_{|I| \leq n} \ell_I \langle e_I, \hat{\mathbb{X}}_t \rangle \right)^2 = \sum_{|I|, |J| \leq n} \ell_I \ell_J \langle e_I \sqcup e_J, \hat{\mathbb{X}}_t \rangle.$$

Since continuous polynomial diffusions have finite moments of every degree, (2.5.4) is satisfied due to Lemma 2.4.2. Under (2.5.3), the expression for $V_t(\ell)$ yields then

$$\begin{aligned} VIX_T^2(\ell) &= \frac{1}{\Delta} \sum_{|I|, |J| \leq n} \ell_I \ell_J \mathbb{E} \left[\int_T^{T+\Delta} \langle e_I \sqcup e_J, \hat{\mathbb{X}}_t \rangle dt \middle| \mathcal{F}_T \right] \\ &= \frac{1}{\Delta} \ell^\top Q(T, \Delta) \ell, \end{aligned}$$

where for each $T > 0$ the matrix Q is given by

$$\begin{aligned}
Q_{\mathcal{L}(I)\mathcal{L}(J)}(T, \Delta) &:= \mathbb{E} \left[\int_T^{T+\Delta} \langle e_I \sqcup e_J, \widehat{\mathbb{X}}_t \rangle dt | \mathcal{F}_T \right] \\
&= \mathbb{E} \left[\int_0^{T+\Delta} \langle e_I \sqcup e_J, \widehat{\mathbb{X}}_t \rangle dt - \int_0^T \langle e_I \sqcup e_J, \widehat{\mathbb{X}}_t \rangle dt | \mathcal{F}_T \right] \\
&= \mathbb{E} [\langle (e_I \sqcup e_J) \otimes e_0, \widehat{\mathbb{X}}_{T+\Delta} \rangle - \langle (e_I \sqcup e_J) \otimes e_0, \widehat{\mathbb{X}}_T \rangle | \mathcal{F}_T] \\
&= \mathbb{E} [\langle (e_I \sqcup e_J) \otimes e_0, \widehat{\mathbb{X}}_{T+\Delta} \rangle | \mathcal{F}_T] - \langle (e_I \sqcup e_J) \otimes e_0, \widehat{\mathbb{X}}_T \rangle.
\end{aligned}$$

By Theorem 2.4.4 we can rewrite the matrix Q as

$$\begin{aligned}
Q_{\mathcal{L}(I)\mathcal{L}(J)}(T, \Delta) &= \mathbf{vec}((e_I \sqcup e_J) \otimes e_0)^\top e^{\Delta G^\top} \mathbf{vec}(\widehat{\mathbb{X}}_T^{2n+1}) \\
&\quad - \mathbf{vec}((e_I \sqcup e_J) \otimes e_0)^\top \mathbf{vec}(\widehat{\mathbb{X}}_T^{2n+1}) \\
&= \mathbf{vec}((e_I \sqcup e_J) \otimes e_0)^\top (e^{\Delta G^\top} - \text{Id}) \mathbf{vec}(\widehat{\mathbb{X}}_T^{2n+1}),
\end{aligned}$$

and the claim follows. $\square$

Remark 2.5.2. Consider now the model described in Remark 2.3.2 and set for simplicity $\varepsilon \geq \Delta$. Then the results of Theorem 2.5.1 still hold however with

$$Q_{\mathcal{L}(I)\mathcal{L}(J)}(T, \Delta) = \sum_{e_{I_1} \otimes e_{I_2} = e_I \sqcup e_J} \int_T^{T+\Delta} \langle e_{I_1}, \widehat{\mathbb{X}}_{t-\varepsilon}^{-1} \rangle \mathbf{vec}(e_{I_2})^\top e^{(t-T)G^\top} \mathbf{vec}(\widehat{\mathbb{X}}_T) dt,$$

where G denotes the $(d+1)_{2n+1}$ -dimensional matrix representative of the dual operator corresponding to $\widehat{\mathbb{X}}$. To adapt the proof we just need to note that for each $t \in [T, T+\Delta]$ Remark 2.4.7 yields

$$\mathbb{E}[\langle e_I \sqcup e_J, \widehat{\mathbb{X}}_{t-\varepsilon, t} \rangle | \mathcal{F}_T] = \sum_{e_{I_1} \otimes e_{I_2} = e_I \sqcup e_J} \langle e_{I_1}, \widehat{\mathbb{X}}_{t-\varepsilon}^{-1} \rangle \mathbf{vec}(e_{I_2})^\top e^{(t-T)G^\top} \mathbf{vec}(\widehat{\mathbb{X}}_T).$$

Note that since the integration's variable t appears twice in this expression the time integral cannot be incorporated in the signature.

Remark 2.5.3. Observe that accounting for the scaling factor of 100, conventionally introduced by CBOE, the VIX index squared can equivalently be redefined (see e.g., Rosenbaum and Zhang (2021); Rømer (2022)) as

$$\text{VIX}_T^2 = \frac{100^2}{\Delta} \mathbb{E} \left[\int_T^{T+\Delta} V_t dt | \mathcal{F}_T \right], \quad (2.5.7)$$

where $T, t > 0$ and $\Delta = \frac{1}{12}$, i.e., approximately 30 days. Notice that since the expressions (2.5.3) and (2.5.7) differ only by a scaling factor, all the theoretical results of the present work hold true disregarding this scaling. For sake of simplicity we will always use (2.5.3). We address the reader to Chapter 11 in Gatheral (2011) for further details about the conventions of CBOE and its link with (2.5.2).

We observe that the expression (2.5.6) is computationally appealing as we can unpack the computation in three parts: compute the coordinate vector $\text{vec}((e_I \sqcup e_J) \otimes e_0)$, which depends just on $d > 0$ and $n > 0$, calculate the matrix exponential of $G^\top$ which depends on the choice of the primary process X , and finally sample $\widehat{\mathbb{X}}_T^{2n+1}$ which is the only part that depends on the chosen maturity time T . In order to compute the matrix exponential we rely on [Bader et al. \(2019\)](#) who developed a Padé-inspired approximation to reduce the matrix multiplications, see also [Moler and Van Loan \(2003\)](#) for further possible methods. For the implementation of the signature samples and its computational complexity we refer to [Reizenstein and Graham \(2018\)](#); [Kidger and Lyons \(2020\)](#).

Remark 2.5.4. In general the computation of $G \in \mathbb{R}^{(d+1)2n+1 \times (d+1)2n+1}$, even if done only once, can be costly. For this reason it can sometimes be interesting to avoid the last time integral and to consider the following equivalent expression of the matrix Q , for $|I|, |J| \leq n$:

$$Q_{\mathcal{L}(I)\mathcal{L}(J)}(T, \Delta) = \left(\int_T^{T+\Delta} \text{vec}(e_I \sqcup e_J)^\top e^{(t-T)G^\top} dt \right) \text{vec}(\widehat{\mathbb{X}}_T^{2n}), \quad (2.5.8)$$

where now $G \in \mathbb{R}^{(d+1)2n \times (d+1)2n}$ and where we use the fact that we can interchange the conditional expectation with the time-integral by dominated convergence. As G is singular, this time integral has to be computed numerically, in general. We propose here two possible methods that can be used in order to compute it efficiently.

- (i) **Approximation of the time integral:** e.g., via the trapezoidal rule also applied for VIX² in [Bourgey and De Marco \(2022\)](#). Hence if we consider the shuffled coordinates $\text{vec}(e_I \sqcup e_J)$ of the exponential matrix we can use the symmetry of the shuffle to reduce the number of integrals to be solved from $((d+1)2n)^2$ to $\frac{(d+1)n((d+1)n+1)}{2} \cdot (d+1)2n$, instead of $(d_{2n})^2$. Observe that for our integral the error of such an approximation is given by

$$\text{Err}(N) = -\frac{\Delta^2}{12N^2} G^\top (e^{G^\top \Delta} - I) + \mathcal{O}(N^{-3}),$$

as $N \rightarrow +\infty$. As a further dimension reduction one can exploit the polynomial nature of $\widehat{\mathbb{X}}^n$ to obtain a matrix representation of its second order moments. Without entering into details, the matrix G would then be the matrix corresponding to the linear operator acting on coefficients of polynomials of degree 2 in $\widehat{\mathbb{X}}^n$. Its dimension would thus be $\frac{(d+1)n((d+1)n+1)}{2}$.

- (ii) **Approximation of the matrix exponential:** we can avoid to approximate the integral by approximating the matrix exponential. Assuming that

$$\lim_{N \rightarrow +\infty} (G^\top \Delta)^N = 0, \quad (2.5.9)$$

this can for instance be done via its Taylor expansion:

$$\int_T^{T+\Delta} e^{tG^\top} dt = \Delta \left(I + \frac{G^\top \Delta}{2!} + \cdots + \frac{(G^\top \Delta)^N}{N+1!} + \mathcal{O}((G^\top \Delta)^{N+1}) \right).$$

Observe that (2.5.9) holds true whenever the spectral radius, i.e., the maximal eigenvalue in absolute value, of the matrix $G^\top \Delta$ is less than 1 (see for instance Theorem 1.5 in [Quarteroni et al. \(2010\)](#)). This requirement suggests that for numerical purposes the parameters of the primary process have to be chosen accordingly.

An interesting example is given by the case where X is a d -dimensional correlated Brownian motion, as considered for instance in [Cuchiero et al. \(2023a\)](#). In this case the process has no linear drift and the corresponding matrix G is nilpotent, meaning that $G^n = 0$, for each n big enough.

In general, this Taylor approach permits to avoid a numerical integration and produces an accurate approximation, allocating as few memory as possible.

Remark 2.5.5. A further step in the direction of a fast evaluation of $\text{VIX}_T(\ell)$ can be taken by noticing that the matrix Q in (2.5.6) admits a Cholesky decomposition. Indeed since Q is positive semidefinite and symmetric by the shuffle property, we know that there exists an upper triangular matrix $U_T \in \mathbb{R}^{(d+1)_n \times (d+1)_n}$, with possible zero elements on the diagonal, such that

$$Q(T, \Delta) = U_T U_T^\top,$$

where for sake of simplicity we drop the dependence on Δ of U_T . Hence the evaluation of the $\text{VIX}_T(\ell)$ reduces to

$$\text{VIX}_T(\ell) = \sqrt{\frac{1}{\Delta} \ell^\top U_T U_T^\top \ell} = \frac{1}{\sqrt{\Delta}} \sqrt{(U_T^\top \ell)^2} = \frac{1}{\sqrt{\Delta}} \|U_T^\top \ell\|,$$

where here $\|\cdot\|$ denotes the Euclidean norm. We stress the fact that the Cholesky decomposition can be carried out offline, and the computational benefit is immediate if several samples of the signature are considered.

In the following remark we discuss a possible dimension reduction technique from which one can benefit computationally. Inspired by the approach of [Cuchiero et al. \(2021a\)](#); [Compagnoni et al. \(2023\)](#), we employ the Johnson-Lindenstrauss Lemma and consider a random projection of the signature. A first way to use this tool is the following.

Remark 2.5.6. Let $d_< \in \mathbb{N}$ be the dimension of the space to which we would like to project the signature of order $n > 0$, such that $d_< \ll (d+1)_n$. Consider $A = (\alpha_{ij}) \in \mathbb{R}^{d_< \times (d+1)_n}$, such that $\alpha_{ij} \sim \mathcal{N}(0, 1/d_<)$. Then a possible way to employ the randomised signature is to parametrize the volatility process as follows,

$$\sigma_t^S(\ell) := \tilde{\ell}^\top A \cdot \text{vec}(\hat{\mathbb{X}}_t^n)$$

where with $\tilde{\ell} = \ell \cdot A^\top \in \mathbb{R}^{d_<}$ we denote the randomised parameters. Due to the linearity of integral and conditional expectation in (2.5.3) this modeling choice is equivalent to consider the randomised matrix $\tilde{Q} \in \mathbb{R}^{d_< \times d_<}$ given by

$$\tilde{Q}_{\mathcal{L}(I)\mathcal{L}(J)}(T, \Delta) := A Q_{\mathcal{L}(I)\mathcal{L}(J)}(T, \Delta) A^\top,$$

which leads to the following representation of $VIX_T(\ell)$:

$$VIX_T(\ell) = \sqrt{\frac{1}{\Delta} \tilde{\ell}^\top \tilde{Q}(T, \Delta) \tilde{\ell}}.$$

Observe that even if this procedure does not reduce the number iterated integrals to be computed offline, it reduces the number of parameters to calibrate, yielding in general to a faster evaluation of $VIX_T(\ell)$.

Remark 2.5.7 (Options on VIX). Note that VIX options are written on VIX futures. The price process of a VIX future contract with maturity $T > 0$, is given by

$$F_t(T) := \mathbb{E}[VIX_T | \mathcal{F}_t], \quad (2.5.10)$$

and we write in particular $F(T) := F_0(T)$ to simplify notation. We point out that the VIX index does not pay dividends. The correct implied volatility for VIX options can then be obtained by inverting the Black-Scholes formula with interest rate $r > 0$ and $e^{-r(T-t)} F_t(T)$ as underlying. When calibrating to VIX options, we stress that we additionally calibrate to VIX futures' prices, see Section 2.5.3. This is important since futures prices under the calibrated model are employed to compute its implied volatility surface. Including VIX futures in the calibration leads to a consistent model, both for VIX options and VIX futures, see e.g. Pacati et al. (2018); Guo et al. (2022a); Guyon (2020a, 2024). Using market prices of the VIX futures to invert the implied volatility surface could lead to inconsistencies if one would like to price further derivatives with the calibrated model.

2.5.2 Variance reduction for pricing VIX options

We here discuss variance reduction techniques (see e.g. Glasserman (2004)) that can speed up the calibration in the subsequently applied Monte Carlo approach further. The key idea is to introduce a control variate, namely an easy to evaluate random variable Φ^{cv} such that given $T > 0$ and $K > 0$,

$$\mathbb{E}[\Phi^{cv}] = 0, \quad \text{Var}((VIX_T(\ell) - K)^+ - \Phi^{cv}) < \text{Var}((VIX_T(\ell) - K)^+).$$

A well-working example of control variates used for pricing and calibrating neural SDE models can be found in Gierjatowicz et al. (2022), where Φ^{cv} is constructed from hedging strategies.

In the following we describe two possible choices of control variates, which consist of polynomials on VIX futures. We stress the fact that these can be seen as linear functions of the signature of the primary process $\hat{X}$, hence they belong to the class of sig-payoffs, see Lyons et al. (2019); Arribas et al. (2020) and Section 4.2.2 in Cuchiero et al. (2023a).

- The first example is to employ the VIX squared as main ingredient, see for instance Bourgey and De Marco (2022); Guerreiro and Guerra (2023) for a similar choice within a rough Bergomi model for pricing VIX options. This is particularly easy to treat in our set up, as for any given maturity $T > 0$ we have

$$\mathbb{E}[VIX_T^2(\ell)] = \frac{1}{\Delta} \ell^\top Q^{cv}(T, \Delta) \ell,$$

with $Q^{cv}(T, \Delta) := \mathbb{E}[Q_{\mathcal{L}(I)\mathcal{L}(J)}(T, \Delta)]$. By Theorem 2.5.1 and Theorem 2.4.4 we indeed have

$$\begin{aligned} Q_{\mathcal{L}(I)\mathcal{L}(J)}^{cv}(T, \Delta) &= \mathbf{vec}((e_I \sqcup e_J) \otimes e_0)^\top (e^{\Delta G^\top} - \text{Id}) \mathbb{E}[\mathbf{vec}(\widehat{\mathbb{X}}_T^{2n+1})] \\ &= \mathbf{vec}((e_I \sqcup e_J) \otimes e_0)^\top (e^{\Delta G^\top} - \text{Id}) e^{TG^\top} \mathbf{vec}(\widehat{\mathbb{X}}_0^{2n+1}) \\ &= \mathbf{vec}((e_I \sqcup e_J) \otimes e_0)^\top (e^{(T+\Delta)G^\top} - e^{TG^\top}) \mathbf{vec}(\widehat{\mathbb{X}}_0^{2n+1}) \end{aligned}$$

where G denotes the $(d+1)_{2n+1}$ -dimensional matrix representative of the dual operator corresponding to $\widehat{\mathbb{X}}$ and $\mathbf{vec}(\widehat{\mathbb{X}}_0^{2n+1}) = e_\emptyset \in \mathbb{R}^{(d+1)_{2n+1}}$.

Observe that Q^{cv} can again be computed offline similarly to the matrix Q . Thus to compute the expectation of $\text{VIX}_T^2(\ell)$ we only have to evaluate the previous quadratic form. To apply this now for pricing a call option with maturity $T > 0$ and strike $K > 0$, we set

$$\begin{aligned} \Phi^{cv}(\ell, T, K) &:= c_{T,K}(\Delta \text{VIX}_T^2(\ell) - \ell^\top Q^{cv}(T, \Delta)\ell), \\ &= c_{T,K}(\ell^\top (Q(T, \Delta) - Q^{cv}(T, \Delta))\ell), \end{aligned}$$

where the constant $c_{T,K}$ maximizing the variance reduction is given by:

$$c_{T,K}^* = \frac{\text{Cov}((\text{VIX}_T(\ell) - K)^+, \ell^\top Q(T, \Delta)\ell)}{\text{Var}(\ell^\top Q(T, \Delta)\ell)}.$$

Notice that also in this case both Q and Q^{cv} satisfy the condition for applying the Cholesky decomposition, leading to a faster evaluation of the control variate as discussed in Remark 2.5.5. Note that the Cholesky decomposition cannot be applied to $Q - Q^{cv}$, as this is in general an indefinite matrix.

- As a second example we consider a generic polynomial in VIX^2 as control variate by defining

$$Y_m^{cv}(\ell, T, K) = \sum_{i=0}^m \alpha_i(T, K) (\text{VIX}_T^2(\ell))^i \quad (2.5.11)$$

where $\alpha_i(T, K)$ are chosen to approximate the payoff $(\text{VIX}_T - K)^+$ with strike price K for some $m \geq 1$. The corresponding control-variate is then defined as $\Phi^{cv}(\ell, T, K) := c_{T,K}(Y_m^{cv}(\ell, T, K) - \mathbb{E}[Y_m^{cv}(\ell, T, K)])$. Regarding the computational effort, let us remark the following.

- (i) VIX_T^2 is computed anyway for every realisation and is hence already available, therefore the computation of $Y_m^{cv}(\ell, T, K)$ is not expensive.
- (ii) It is possible to calculate $\mathbb{E}[Y_m^{cv}(\ell, T, K)]$ analytically relying on the moment formula, see Theorem 2.4.4.
- (iii) The choice of $c_{T,K} \in \mathbb{R}$ is important and the optimal one, i.e., the one leading the highest variance reduction, is given by the following expression

$$c_{T,K}^* = \frac{\text{Cov}((\text{VIX}_T(\ell) - K)^+, Y_m^{cv}(\ell, T, K))}{\text{Var}(Y_m^{cv}(\ell, T, K))},$$

see for instance Section 4.1.1 in Glasserman (2004).

We stress the fact that for $m = 1$ the two control variates introduced coincide.

2.5.3 Calibration to VIX options

In this section we focus on the calibration to VIX options only. Let $\mathcal{T}$ be a set of maturities and $\mathcal{K}$ a collection of strikes. Consider the model given by (2.3.1) and (2.3.2).

Using Monte Carlo compute an approximation of option and futures' prices with $N_{MC} > 0$ samples, i.e.

$$\pi_{\text{VIX}}^{\text{model}}(\ell, T, K) \approx \frac{e^{-rT}}{N_{MC}} \sum_{i=1}^{N_{MC}} (\text{VIX}_T(\ell, \omega_i) - K)^+, \quad F_{\text{VIX}}^{\text{model}}(\ell, T) \approx \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} \text{VIX}_T(\ell, \omega_i), \quad (2.5.12)$$

where

$$\text{VIX}_T(\ell, \omega) = \sqrt{\frac{1}{\Delta} \ell^\top Q(T, \Delta)(\omega) \ell} = \frac{1}{\sqrt{\Delta}} \|U_T^\top(\omega) \ell\|.$$

It is crucial to note that in this framework a Monte Carlo approach is tractable since for every ℓ the same samples can be used. This means that we do not need to carry out any simulation during the optimization task. Indeed, the matrix Q can be simulated offline while only the products with $\ell \in \mathbb{R}^{(d+1)n}$ enter in the calibration step.

Observe that an auxiliary randomization can be employed in every optimisation step as discussed in Remark 2.5.6. Moreover, if we want to use control variates to reduce the variance of the Monte Carlo estimator as described in the previous section, we would consider

$$\pi_{\text{VIX}}^{\text{model}}(\ell, T, K) \approx \frac{e^{-rT}}{N_{VR}} \sum_{i=1}^{N_{VR}} (\text{VIX}_T(\ell, \omega_i) - K)^+ - \Phi^{cv}(\ell, T, K)(\omega_i).$$

Due to the variance reduction the number of samples needed is $N_{VR} \ll N_{MC}$ and Φ^{cv} is as in Section 2.5.2

The calibration to VIX call options and the corresponding futures on $\mathcal{T}$ and $\mathcal{K}$ consists in minimizing the functional

$$L_{\text{VIX}}(\ell) := \sum_{T \in \mathcal{T}, K \in \mathcal{K}} \mathcal{L} \left(\pi_{\text{VIX}}^{\text{model}}(\ell, T, K), \pi_{\text{VIX}}^{b,a}(T, K), \sigma_{\text{VIX}}^{b,a}(T, K), F_{\text{VIX}}^{\text{model}}(\ell, T), F_{\text{VIX}}^{\text{mkt}}(T) \right), \quad (2.5.13)$$

where $\mathcal{L}$ denotes a real-valued loss function, $F_{\text{VIX}}^{\text{mkt}}(T)$ the market's futures' prices and

$$\pi_{\text{VIX}}^{b,a}(T, K) := \{\pi_{\text{VIX}}^{\text{mkt},b}(T, K), \pi_{\text{VIX}}^{\text{mkt},a}(T, K)\}, \quad \sigma_{\text{VIX}}^{b,a}(T, K) := \{\sigma_{\text{VIX}}^{\text{mkt},b}(T, K), \sigma_{\text{VIX}}^{\text{mkt},a}(T, K)\},$$

the market's option bid/ask prices $\pi_{\text{VIX}}^{\text{mkt},b}(T, K)$, $\pi_{\text{VIX}}^{\text{mkt},a}(T, K)$, and bid/ask implied volatilities $\sigma_{\text{VIX}}^{\text{mkt},b}(T, K)$, $\sigma_{\text{VIX}}^{\text{mkt},a}(T, K)$, respectively. We will specify the choice of the function $\mathcal{L}$ in Section 2.5.3 and Section 2.7.1. In both sections we employ the same optimizer, i.e. BFGS with default parameters in *scipy.optimize*.

Remark 2.5.8 (Initial guess search). Since within our model choice we are given a quadratic function in ℓ to be minimized, a stochastic optimization with an initial guess is employed. In order to achieve faster convergence we consider an hyperparameter search to choose the starting parameters. The steps are outlined as follows.

- Find the magnitude of the coefficients returning Monte Carlo prices of the VIX options close to the one observable on the market. To this extent we sample $N_\ell > 0$ times parameters $\ell \in J_i = [-10^{-i}, 10^{-i}]^{(d+1)n}$, for $i = 1, \dots, m$ with $m > 0$.
- Select $J^* \in (J_i)_{i=1}^m$ such that

$$J^* \in \operatorname{argmin}_{i: \ell \in J_i} L_{\text{VIX}}(\ell).$$

- Choose the initial guess to be

$$\ell_{\text{initial}} \in \operatorname{argmin}_{\ell \in J^*} L_{\text{VIX}}(\ell).$$

Numerical results

In the present section we report the results of the calibration to VIX options only. Here we consider call options written on the VIX on the trading day 02/06/2021, the same as in [Guyon and Lekeufack \(2023\)](#). We stress that for such recent dates the bid-ask spreads for VIX options are rather tight with respect to older dated options as considered for instance in [Gatheral et al. \(2020\)](#); [Bondi et al. \(2024b\)](#). The maturities are reported in the following table with the corresponding range of strikes (in percentage) with respect to the market's futures prices.

$T_1 = 0.0383$	$T_2 = 0.0767$	$T_3 = 0.1342$	$T_4 = 0.2108$	$T_5 = 0.2875$	$T_6 = 0.3833$
(90%,250%)	(90%,250%)	(80%,310%)	(80%,300%)	(75%,395%)	(80%,405%)

We underline that the shortest maturity considered is 14 days. Regarding our modeling choice we fix $d = 2$, $n = 3$, which means to calibrate 40 parameters. For X we choose a 2-dimensional Ornstein-Uhlenbeck processes, see Example [2.4.5](#), with the following (hyperparameter) configuration:

$$\kappa = (0.1, 25)^\top, \quad \theta = (0.1, 4)^\top, \quad \sigma = (0.7, 10)^\top, \quad \rho = \begin{pmatrix} 1 & -0.577 & 0.3 \\ \cdot & 1 & -0.6 \\ \cdot & \cdot & 1 \end{pmatrix},$$

where we slightly abuse notation and denote by ρ the correlation matrix of (X, B) . This implies that its last column describes the correlations of X with the Brownian motion B driving the price process S .

These hyperparameters are chosen randomly. Indeed, in spirit of reservoir computing, the idea is to view the OU-process' signature as (randomly chosen) reservoir, while a simple

readout mechanism is trained, i.e. the linear function defined by $\{\ell_I : |I| \leq n\}$, to map the state of the reservoir to the desired output (in our case instantaneous volatilities). However, it is of course possible to perform a hyperparameter optimization or to add expert knowledge, e.g. that a high mean reversion rate is important. We tried the latter by mimicking a rough or strong mean-reverting model as suggested in Rogers (2023); Rømer (2022).

We also refer to Appendix 2.A for numerical results where we use only a correlated 2-dimensional Brownian motion as primary process, which yields significantly worse results. Note that the second simplest choice after Brownian motion within the family of polynomial diffusions (also with exact simulation) is the Ornstein-Uhlenbeck process which we thus applied.

Before stating the loss function $\mathcal{L}$ that we employed in the calibration task, let us make the following remark.

Remark 2.5.9. Let $f : \mathbb{R}^+ \times \mathbb{R}^+ \rightarrow \mathbb{R}^+$ be the call pricing functional in the Black-Scholes model, depending on the volatility σ^{BS} and the spot price ξ , i.e., $f : (\sigma^{\text{BS}}, \xi) \mapsto f(\sigma^{\text{BS}}, \xi)$. By Taylor expansion in an appropriate neighbourhood of $(\sigma^{\text{mkt}}, \xi^{\text{mkt}})$ we obtain

$$f(\sigma^{\text{BS}}, \xi) \approx f(\sigma^{\text{mkt}}, \xi^{\text{mkt}}) + \frac{\partial f}{\partial \sigma}(\sigma^{\text{mkt}}, \xi^{\text{mkt}})(\sigma^{\text{BS}} - \sigma^{\text{mkt}}) + \frac{\partial f}{\partial \xi}(\sigma^{\text{mkt}}, \xi^{\text{mkt}})(\xi - \xi^{\text{mkt}}),$$

which equivalently gives

$$(\sigma^{\text{BS}} - \sigma^{\text{mkt}}) \approx \frac{1}{\frac{\partial f}{\partial \sigma}(\sigma^{\text{mkt}}, \xi^{\text{mkt}})} (f(\sigma^{\text{BS}}, \xi) - f(\sigma^{\text{mkt}}, \xi^{\text{mkt}})) - \frac{\frac{\partial f}{\partial \xi}(\sigma^{\text{mkt}}, \xi^{\text{mkt}})}{\frac{\partial f}{\partial \sigma}(\sigma^{\text{mkt}}, \xi^{\text{mkt}})} (\xi - \xi^{\text{mkt}}), \quad (2.5.14)$$

where we recognize for the derivatives with respect to σ and ξ , the Greeks Vega and Delta, respectively.

Motivated by Remark 2.5.9 we propose, for a fixed maturity and strike price, the following loss-function for $\beta \in \{0, 1\}$

$$\mathcal{L}^\beta(\pi, \pi^{\text{mkt}, b, a}, \sigma^{\text{mkt}, b, a}, F, F^{\text{mkt}}) = \left(\frac{(\beta \tilde{1}_{\{\pi \notin [\pi^{\text{mkt}, b}, \pi^{\text{mkt}, a}]\}} + (1 - \beta)) |\pi - (\pi^{\text{mkt}, a} + \pi^{\text{mkt}, b})/2| + |\delta^{\text{mkt}} e^{-rT} (F - F^{\text{mkt}})|}{v^{\text{mkt}}(\sigma^{\text{mkt}, a} - \sigma^{\text{mkt}, b})} \right)^2, \quad (2.5.15)$$

where

- v^{mkt} and δ^{mkt} denote the Vega and Delta of the option under the Black-Scholes model which depend on the maturity and on the strike price;
- F and F^{mkt} denote futures with maturity T such that the variables ξ, ξ^{mkt} appearing in Remark 2.5.9 are $\xi = e^{-rT} F$ and $\xi^{\text{mkt}} = e^{-rT} F^{\text{mkt}}$, respectively;
- $\tilde{1}_{\{x \notin [y^b, y^a]\}} := s(y^b - x) + s(x - y^a)$ for $s(x) := \frac{1}{2} \tanh(100x) + \frac{1}{2}$ a smooth version of the indicator function.

Remark 2.5.10. (i) We observe that by Remark 2.5.9 minimizing $\mathcal{L}^0$ is equivalent to minimizing an upper bound of the square of the right-hand side of (2.5.14) normalized by the bid-ask spread of the implied volatilities. Note that we slightly abused notation, since v^{mkt} and δ^{mkt} of course depend on the strike and the maturity.

(ii) Note that as $\ell \mapsto \text{VIX}_T(\ell, \omega) = \frac{1}{\sqrt{\Delta}} \|U_T^\top \ell\|$ is convex and the call payoff is convex and increasing, the model option and futures prices are convex in ℓ . If $\beta = 0$ and the initialization of ℓ is such that both the model and futures prices are higher than the market ones, then we actually deal with a convex optimization problem.

(iii) If our aim does not consist in calibrating to the mid-price or mid-implied-volatility precisely, but we merely want to be within the bid-ask spreads we can set $\beta = 1$

For the next calibration result we minimize $\mathcal{L}^1$ as introduced above with $N_{MC} = 80000$ Monte Carlo samples for the previous maturities and strikes.

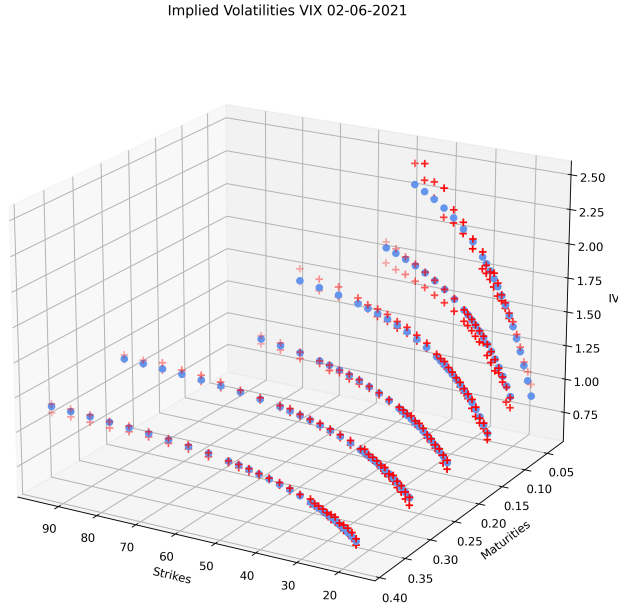


Figure 2.1: The red crosses denote the bid-ask spreads (of the implied volatilities) for each maturity, while the azure dots denote the calibrated implied volatilities of the model. On the x -axis we find the strikes and on the y -axis we find the maturities.

We observe that the calibrated VIX smiles fall systematically in the bid-ask interval for all the maturities considered. We report additionally in the next tables the relative error between the market futures prices and the calibrated ones for each maturity, i.e.,

$$\varepsilon_T := \frac{|F^{mkt}(T) - F_{\text{VIX}}^{\text{model}}(\ell^*, T)|}{F^{mkt}(T)}, \quad (2.5.16)$$

where $\ell^* \in \mathbb{R}^{40}$ denotes the calibrated parameters and here $F_{\text{VIX}}^{\text{model}}(\ell^*, T)$ stands for the calibrated future model price. In Figure 2.2 we can find an illustration of the calibrated and the market futures' term structure.

$T_1 = 0.0383$	$T_2 = 0.0767$	$T_3 = 0.1342$
$\varepsilon_{T_1} = 7.0 \times 10^{-6}$	$\varepsilon_{T_2} = 2.1 \times 10^{-3}$	$\varepsilon_{T_3} = 1.3 \times 10^{-5}$
$T_4 = 0.2108$	$T_5 = 0.2875$	$T_6 = 0.3833$
$\varepsilon_{T_4} = 1.5 \times 10^{-4}$	$\varepsilon_{T_5} = 1.9 \times 10^{-6}$	$\varepsilon_{T_6} = 1.3 \times 10^{-6}$

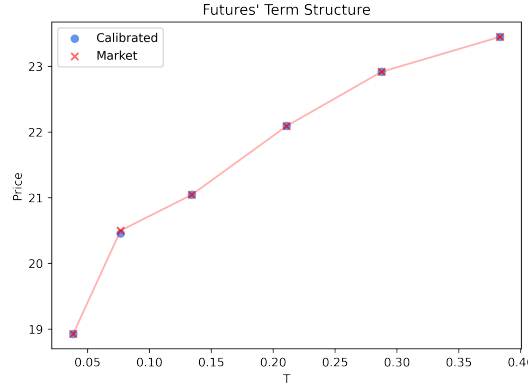


Figure 2.2: The blue circles denote the calibrated futures prices and the red crosses the futures prices on the market, in between a linear interpolation is reported.

2.5.4 The case of time-varying parameters

We now consider the case of maturity dependent parameters as for instance employed in [Gierjatowicz et al. \(2022\)](#); [Cuchiero et al. \(2023a\)](#). Since it will be important later on to distinguish maturities of options written on the VIX index from maturities of options written on the SPX index, we introduce the two sets $\mathcal{T}^{\text{VIX}}$ and $\mathcal{T}^{\text{SPX}}$. Let us fix here $\mathcal{T}^{\text{VIX}} = \{T_1, \dots, T_N\}$, where $T_i < T_{i+1}$ for any in $i = 1, \dots, N-1$ and denote by $\ell(T_i) \in \mathbb{R}^{d_n}$ the parameters depending on the maturity $T_i > 0$. We set $T_0 = 0$ and $T_{N+1} = +\infty$. Then, we consider for any $t \geq 0$ the volatility process to be

$$\sigma_t^S(\ell) = \sum_{i=0}^N \sum_{|I| \leq n} \ell_I(T_i) 1_{[T_i, T_{i+1})}(t) \langle e_I, \widehat{\mathbb{X}}_t \rangle. \quad (2.5.17)$$

Therefore the variance process reads as follows,

$$V_t(\ell) = \sum_{i=0}^N \sum_{|J|, |I| \leq n} \ell_I(T_i) \ell_J(T_i) 1_{[T_i, T_{i+1})}(t) \langle e_I \sqcup e_J, \widehat{\mathbb{X}}_t \rangle. \quad (2.5.18)$$

Assumption 2.5.11. Assume that for a set of maturities $\mathcal{T}^{\text{VIX}}$ it holds that $|T_i - T_j| \geq \Delta$ for all $i \neq j$.

Proposition 2.5.12. Let $\mathcal{T}^{\text{VIX}}$ be a set of maturities on the VIX index and let $Q(T, \tau)$ be the matrix as defined in (2.5.6) (here for general $\tau > 0$ instead of Δ). Then, under (2.5.18) the VIX squared at time $T_i \in \mathcal{T}^{\text{VIX}}$ is given by

$$\text{VIX}_{T_i}^2(\ell) = \frac{1}{\Delta} \left(\sum_{j=i}^N \ell(T_j)^\top (Q(T_i, (T_{j+1} - T_i) \wedge \Delta) - Q(T_i, (T_j - T_i) \wedge \Delta)) \ell(T_j) \right).$$

Note that, if $T_{i+1} - T_i > \Delta$ (which is in particular holds under Assumption 2.5.11) then,

$$\text{VIX}_{T_i}^2(\ell) = \frac{1}{\Delta} \ell(T_i)^\top Q(T_i, \Delta) \ell(T_i).$$

Proof. By the definition of the VIX, it holds that

$$\begin{aligned} \text{VIX}_{T_i}^2(\ell) &= \frac{1}{\Delta} \mathbb{E} \left[\int_{T_i}^{T_i+\Delta} \sum_{j=i}^N \sum_{|J|, |I| \leq n} \ell_I(T_j) \ell_J(T_j) 1_{[T_j, T_{j+1})}(t) \langle e_I \sqcup e_J, \widehat{\mathbb{X}}_t \rangle dt \middle| \mathcal{F}_{T_i} \right] \\ &= \frac{1}{\Delta} \sum_{j=i}^N \sum_{|J|, |I| \leq n} \ell_I(T_j) \ell_J(T_j) \mathbb{E} \left[\int_{T_j \wedge (T_i+\Delta)}^{T_{j+1} \wedge (T_i+\Delta)} \langle e_I \sqcup e_J, \widehat{\mathbb{X}}_t \rangle dt \middle| \mathcal{F}_{T_i} \right] \\ &= \frac{1}{\Delta} \sum_{j=i}^N \sum_{|J|, |I| \leq n} \ell_I(T_j) \ell_J(T_j) \left(\mathbb{E} \left[\int_{T_i}^{T_{j+1} \wedge (T_i+\Delta)} \langle e_I \sqcup e_J, \widehat{\mathbb{X}}_t \rangle dt \middle| \mathcal{F}_{T_i} \right] \right. \\ &\quad \left. - \mathbb{E} \left[\int_{T_i}^{T_j \wedge (T_i+\Delta)} \langle e_I \sqcup e_J, \widehat{\mathbb{X}}_t \rangle dt \middle| \mathcal{F}_{T_i} \right] \right) \end{aligned}$$

and hence the first statement follows by the definition of Q in (2.5.6). $\square$

Notice that also in the case of Proposition 2.5.12, Remark 2.5.5 applies.

2.6 SPX as a signature-based model

The goal of this section is to express the discounted, dividend-adjusted price of the SPX, modeled via (2.3.1)-(2.3.2)

$$dS_t(\ell) = S_t(\ell) \sigma_t^S(\ell) dB_t,$$

in terms of the signature of $(t, X_t, B_t)_{t \geq 0}$, allowing again to precompute its samples and use the same ones for every ℓ . This is in the same spirit as in Cuchiero et al. (2023a), even though there the asset price was directly modeled as linear function of the signature of some primary process.

Recall that by (2.3.2) σ^S is parametrized as follows

$$\sigma_t^S(\ell) := \ell_\emptyset + \sum_{0 < |I| \leq n} \ell_I \langle e_I, \widehat{\mathbb{X}}_t \rangle,$$

where $\widehat{X}_t = (t, X_t)$ with X a d -dimensional polynomial diffusion X in the sense of Definition 2.3.1. Before addressing a more tractable expression for S , that allows to avoid (Euler) simulation schemes, we recall the following well-known integrability result.

Lemma 2.6.1. Assume that $\mathbb{E}[S_0] < \infty$. Then, the process $(S_t)_{t \geq 0}$ is a (non-negative) supermartingale and in particular $\mathbb{E}[S_t] < \infty$ for each $t \geq 0$.

Proof. Note that $S_t = S_0 \mathcal{E} \left(\int_0^t \sigma_s^S dB_s \right)$ for all $t \geq 0$. Moreover $(\int_0^t \sigma_s^S dB_s)_{t \geq 0}$ is a local martingale and hence, by the properties of the stochastic exponential, S_t is a non-negative local martingale. It follows from Fatou's Lemma that non-negative local martingales are supermartingales. $\square$

In the following we suppose without loss of generality that $S_0 = 1$.

Remark 2.6.2. Recall that if Novikov's condition is satisfied, then a stochastic exponential of the form $S_t = \mathcal{E} \left(\int_0^t \sigma_s^S dB_s \right)$ for $t \in [0, T]$ is a true martingale. For σ_s^S as in (2.3.2), such condition reads

$$\mathbb{E} \left[\exp \left\{ \frac{1}{2} \int_0^T V_t(\ell) dt \right\} \right] < +\infty.$$

Observe that

$$\begin{aligned} \mathbb{E} \left[\exp \left\{ \frac{1}{2} \int_0^T V_t(\ell) dt \right\} \right] &= \mathbb{E} \left[\exp \left\{ \frac{1}{2} \sum_{|I|, |J| \leq n} \ell_I \ell_J \int_0^T \langle e_I \sqcup e_J, \widehat{X}_t \rangle dt \right\} \right] \\ &= \mathbb{E} \left[\exp \left\{ \frac{1}{2} \ell^\top Q^0(T) \ell \right\} \right], \end{aligned} \quad (2.6.1)$$

where for $\mathcal{L} : \{I : |I| \leq n\} \rightarrow \{1, \dots, (d+1)_n\}$,

$$Q_{\mathcal{L}(I)\mathcal{L}(J)}^0(T) := \langle (e_I \sqcup e_J) \otimes e_0, \widehat{X}_T \rangle.$$

We point out that the previous condition is not necessarily satisfied for all $\ell \in \mathbb{R}^{(d+1)_n}$. Indeed, let us consider X to be a one-dimensional Brownian motion and choose ℓ such that the only non trivial component is the last one, i.e.,

$$\ell_I := \begin{cases} c \in \mathbb{R}, & \text{if } I = (1, \dots, 1), |I| = n, \\ 0, & \text{otherwise.} \end{cases} \quad (2.6.2)$$

Then, (2.6.1) translates into

$$\mathbb{E} \left[\exp \left\{ \frac{c^2}{2} \int_0^T \frac{2n!}{n!n!} \frac{X_t^{2n}}{2n!} dt \right\} \right] = \mathbb{E} \left[\exp \left\{ \frac{c^2}{2(n!)^2} \int_0^T X_t^{2n} dt \right\} \right],$$

which is not finite in general, e.g. if $n = 2$, $c = \sqrt{2}(n!)$ then by Jensen's inequality it follows

$$\mathbb{E} \left[\exp \left\{ \int_0^T X_t^4 dt \right\} \right] \geq \mathbb{E} \left[\frac{1}{T} \int_0^T e^{TX_t^4} dt \right] = \frac{1}{T} \int_0^T \mathbb{E}[e^{TX_t^4}] dt = +\infty.$$

Remark 2.6.3. As well known from the results of [Delbaen and Schachermayer \(1994\)](#) the existence of an equivalent local martingale measure is sufficient for NFLVR, and risk neutral pricing works, too. This is important when the process $S(\ell)$ is a non-negative true *local* martingale such that $\mathbb{E}[S_T(\ell)] < S_0$. If one could go short in the asset, and thus gets $S_0(\ell)$, and long in the ‘call option with strike 0’ (corresponding to the payoff $S_T(\ell)$) with price $\mathbb{E}[S_T(\ell)] < S_0(\ell)$, an arbitrage would be created. But the latter is simply not allowed as trading strategy under NFLVR. We address the reader to [Kardaras et al. \(2015\)](#) for further details.

The key idea is to rewrite (2.3.1) as a type of signature-based model in sense of [Cuchiero et al. \(2023a\)](#) including $B = (B_t)_{t \geq 0}$ as part of the primary process. This is possible since Itô integrals with respect to primary process’ components can be rewritten as linear functions of the signature of the primary process itself. Before stating the result we need to introduce some auxiliary notation. We denote by $(Z_t)_{t \geq 0}$ the $(d+1)$ -dimensional process given by

$$Z_t = (X_t, B_t), \quad (2.6.3)$$

by $(\hat{Z}_t)_{t \geq 0}$ its time extension, and by $(\hat{\mathbb{Z}}_t)_{t \geq 0}$ the signature of $(\hat{Z}_t)_{t \geq 0}$. With a slight abuse of notation we again denote by ρ the correlation matrix process between the components of Z . Observe that ρ encodes in particular the correlation between X and B . Finally, we let $a_{ij}^J \in \mathbb{R}$ denote the coefficients satisfying

$$d[Z^i, Z^j]_t = \sum_{|J| \leq 2} a_{ij}^J \langle e_J, \hat{\mathbb{Z}}_t \rangle dt,$$

for each $i, j \in \{1, \dots, d+1\}$.

Proposition 2.6.4. Let $S = (S_t)_{t \geq 0}$ satisfy (2.3.1) with $S_0 = 1$, and $\sigma^S = (\sigma_t^S)_{t \geq 0}$ satisfy (2.3.2). Then,

$$\log(S_t(\ell)) = -\frac{1}{2} \ell^\top Q^0(t) \ell + \sum_{|I| \leq n} \ell_I \langle \tilde{e}_I^B, \hat{\mathbb{Z}}_t \rangle, \quad (2.6.4)$$

where

$$\tilde{e}_\emptyset^B := e_{d+1}, \quad \tilde{e}_I^B := e_I \otimes e_{d+1} - \sum_{|J| \leq 2} \frac{a_{i|I|(d+1)}^J}{2} (e_{I' \sqcup e_J} \otimes e_0),$$

for each $|I| > 0$, and the components of the matrix $Q^0(t) \in \mathbb{R}^{(d+1)_n \times (d+1)_n}$ are given by

$$Q_{\mathcal{L}(I)\mathcal{L}(J)}^0(t) = \langle (e_I \sqcup e_J) \otimes e_0, \hat{\mathbb{Z}}_t \rangle,$$

for a labeling function $\mathcal{L} : \{I : |I| \leq n\} \rightarrow \{1, \dots, (d+1)_n\}$.

Proof. We can compute

$$\begin{aligned}
 \log(S_t(\ell)) &= -\frac{1}{2} \int_0^t V_s(\ell) ds + \int_0^t \sigma_s^S(\ell) dB_s \\
 &= -\frac{1}{2} \sum_{|I|, |J| \leq n} \ell_I \ell_J \int_0^t \langle e_I \sqcup e_J, \widehat{\mathbb{X}}_s \rangle ds + \sum_{|I| \leq n} \ell_I \int_0^t \langle e_I, \widehat{\mathbb{X}}_s \rangle dB_s \\
 &\stackrel{(*)}{=} -\frac{1}{2} \sum_{|I|, |J| \leq n} \ell_I \ell_J \langle (e_I \sqcup e_J) \otimes e_0, \widehat{\mathbb{X}}_t \rangle + \sum_{|I| \leq n} \ell_I \langle \tilde{e}_I^B, \widehat{\mathbb{Z}}_t \rangle \\
 &= -\frac{1}{2} \ell^\top Q^0(t) \ell + \sum_{|I| \leq n} \ell_I \langle \tilde{e}_I^B, \widehat{\mathbb{Z}}_t \rangle,
 \end{aligned}$$

where for $(*)$ we used that $\int_0^t \langle e_I, \widehat{\mathbb{X}}_s \rangle dB_s = \langle \tilde{e}_I^B, \widehat{\mathbb{Z}}_t \rangle$ by Lemma 3.10 in [Cuchiero et al. \(2023a\)](#). $\square$

Remark 2.6.5. Consider again the model described in Remark 2.3.2. Then the results of Proposition 2.6.4 still hold with

$$Q_{\mathcal{L}(I)\mathcal{L}(J)}^0(t) := \int_0^t \langle e_I \sqcup e_J, \widehat{\mathbb{X}}_{s-\varepsilon, s} \rangle ds,$$

and $\int_0^t \langle e_I, \widehat{\mathbb{X}}_{s-\varepsilon, s} \rangle dB_s$ instead of $\langle \tilde{e}_I^B, \widehat{\mathbb{Z}}_t \rangle$. Since the proof follows closely the proof of the original result, we omit it.

Remark 2.6.6. ■ Observe that since the matrix $(\langle e_I \sqcup e_J, \widehat{\mathbb{X}}_t \rangle)_{|I|, |J| \leq n}$ is positive semi-definite, by monotonicity of the time integral on $[0, t]$ for some $t > 0$, we also have

$$\ell^\top Q^0(t) \ell \geq 0,$$

for all $\ell \in \mathbb{R}^{(d+1)n}$. This means that for any $t > 0$, we can rewrite the log-price as

$$\log(S_t) = -\frac{1}{2} \|(U_t^0)^\top \ell\|^2 + \sum_{|I| \leq n} \ell_I \langle \tilde{e}_I^B, \widehat{\mathbb{Z}}_t \rangle,$$

where U_t^0 is the upper-triangular matrix of the Cholesky decomposition of $Q^0(t)$.

- Notice that the log-price model in (2.6.4), it is not exactly a signature-based model in the sense of [Cuchiero et al. \(2023a\)](#), as here it is given by a linear part in the parameters ℓ and an additional quadratic part. It can also be rewritten as

$$d \log(S_t) = -\frac{1}{2} \ell^\top \tilde{Q}(t) \ell dt + \ell^\top \mathbf{vec}(\widehat{\mathbb{X}}_t^n) dB_t,$$

where $\tilde{Q}$ is given by

$$\tilde{Q}_{\mathcal{L}(I)\mathcal{L}(J)}(t) := \langle e_I \sqcup e_J, \widehat{\mathbb{X}}_t \rangle.$$

Hence, the relevant factors entering in the dynamics of $\log S$ and σ^S are the components of $\widehat{\mathbb{X}}^{2n}$. Note also that $(\log S, \widehat{\mathbb{X}}^{2n})$ is a $1 + (d+1)_{2n}$ dimensional polynomial diffusion

(see (2.2.1)), whence in particular Markovian. This is in spirit of path-dependent factor model, for instance also considered in Guyon and Lekeufack (2023), with the additional tractability feature that $(\log S, \widehat{\mathbb{X}}^{2n})$ is a polynomial diffusion. Therefore all techniques for polynomial processes in view of pricing and hedging can be applied.

- In order to sample the log-price at maturity, consistently with the VIX, we follow the following road map. We simulate $\widehat{\mathbb{Z}}$ and compute $\langle \tilde{e}_I^B, \widehat{\mathbb{Z}} \rangle$ for each I as specified above. Next, we drop from the samples of $\widehat{\mathbb{Z}}$ the terms where B appears, i.e. the components corresponding to indices containing the letter $d+1$. The result coincides with a sampling of $\widehat{\mathbb{X}}$ and is then used to work with both Q and Q^0 .

This is equivalent to sampling $\widehat{\mathbb{X}}$ for the variance process and to compute an additional Itô integral as in (2.3.1).

In the following corollary we state the form of $\tilde{e}_I^B$ when X is d -dimensional OU-process. We omit the proof for sake of brevity.

Corollary 2.6.7. *Let X be a d -dimensional OU-process as in Example 2.4.5 driven by a d -dimensional Brownian motion with correlation matrix ρ . Then $\tilde{e}_I^B$ is given by*

$$\tilde{e}_I^B = e_I \otimes e_{d+1} - \frac{1}{2} 1_{\{i_{|I|} \neq 0\}} (\sigma^{i_{|I|}} \rho_{i_{|I|} d+1}) e_{I'} \otimes e_0,$$

for any multi-index $I \neq \emptyset$.

Remark 2.6.8 (Variance reduction for pricing SPX options). Observe that a possible control variate for reducing the variance of the Monte Carlo estimator for pricing SPX options is the value at maturity of the log-price process. This means,

$$\Phi^{cv}(\ell, T, K) := c_{T,K} \left(\log(S_T(\ell)) + \frac{1}{2} \ell^\top Q^{0,cv}(T) \ell \right),$$

where, using that the linear part (in ℓ) of $\log(S_T(\ell))$ vanishes under the risk-neutral expectation, we have

$$Q_{\mathcal{L}(I)\mathcal{L}(J)}^{0,cv}(T) = \mathbf{vec}((e_I \sqcup e_J) \otimes e_0)^\top e^{TG^\top} \mathbf{vec}(\widehat{\mathbb{X}}_0^{2n+1}),$$

for $G \in \mathbb{R}^{(d+1)2n+1 \times (d+1)2n+1}$ denoting the $(d+1)_{2n+1}$ -dimensional matrix representative of the dual operator corresponding to $\widehat{\mathbb{X}}$. We choose the optimal $c_{T,K}^* \in \mathbb{R}$ as

$$c_{T,K}^* = \frac{\text{Cov}((S_T(\ell) - K)^+, \log(S_T(\ell)))}{\text{Var}(\log(S_T(\ell)))}.$$

2.6.1 Exploiting the affine nature of the signature: Fourier pricing of SPX and VIX options

This section is dedicated to outline how the linear parametrizations of the log-price and the volatility process in $\widehat{\mathbb{Z}}$ can be used for Fourier pricing. Assume that

$$dZ_t^j = \kappa^j(\theta^j - Z_t^j)dt + \sigma^j dW_t^j, \quad Z_0^j = 0,$$

for each $j = 1, \dots, d+1$, where W denotes a $(d+1)$ -dimensional Brownian motion with $W^{d+1} = B$. All parameters $\kappa^j, \theta^j, \sigma^j$ are in $\mathbb{R}$ with $\kappa^{d+1} = \theta^{d+1} = 0$ and $\sigma^{d+1} = 1$ so that $Z^{d+1} = W^{d+1} = B$. Note that we do not account for correlations.

We illustrate now how to apply the results of [Cuchiero et al. \(2023c\)](#) in the present setting. Since $(\widehat{Z}_t)_{t \geq 0}$ is a polynomial diffusion, by Lemma 2.4.1 there are $\mathbf{b} \in (T((\mathbb{R}^{d+2})))^{d+2}$ and $\mathbf{a} \in (T((\mathbb{R}^{d+2})))^{(d+2) \times (d+2)}$ such that

$$d\widehat{Z}_t^j = \langle \mathbf{b}_j, \widehat{Z}_t \rangle dt + \sqrt{\langle \mathbf{a}_{jj}, \widehat{Z}_t \rangle} dW_t^j,$$

where $\mathbf{b}_j = \kappa^j \theta^j e_\emptyset - \kappa_j e_j$ and $\mathbf{a}_{jj} = (\sigma^j)^2 e_\emptyset$, using that (with a small abuse of notation) $\kappa^0 \theta^0 := 1$, $\kappa^j := 0$ and $\sigma^0 := 0$. Consider then the Riccati operator $\mathcal{R}$ given by

$$\mathcal{R}(\mathbf{u}) = \sum_{j=0}^{d+1} \sum_{|I| \geq 0} \left(\kappa^j \theta^j \mathbf{u}_{(Ij)} e_I + \kappa^j \mathbf{u}_{(Ij)} e_j \sqcup e_I + \frac{1}{2} (\sigma^j)^2 (\mathbf{u}_{(Ijj)} e_I + \mathbf{u}_{(Ij)}^2 e_I \sqcup e_I) \right).$$

By Theorem 4.23 in [Cuchiero et al. \(2023c\)](#), we expect that

$$\mathbb{E}[\exp(\langle \mathbf{u}, \widehat{Z}_T \rangle)] = \exp(\psi(T)_\emptyset),$$

where ψ is a solution of the extended tensor algebra valued Riccati equation⁴

$$\partial_t \psi(t) = \mathcal{R}(\psi(t)), \quad \psi(0) = \mathbf{u}. \quad (2.6.5)$$

Choosing $\mathbf{u}$ as

$$\mathbf{u}(\ell) := -\frac{1}{2}(\ell \sqcup \ell) \otimes e_0 + \tilde{\ell}$$

where $\tilde{\ell} := \sum_{|I| \leq n} \ell_I \tilde{e}_I^B$, by Proposition 2.6.4 we get

$$\log(S_t(\ell)) = \langle \mathbf{u}(\ell), \widehat{Z}_t \rangle.$$

The representation of the Fourier-Laplace transform described above can then be used for Fourier pricing. We dedicate the remaining part of this section to illustrate how this can be done.

From Fourier analysis we know that for $K > 0$ and $C < 0$ it holds

$$(K - e^y)^+ = \frac{1}{2\pi} \int_{\mathbb{R}} e^{(i\lambda + C)y} \frac{K^{-C+1-i\lambda}}{(i\lambda + C)(i\lambda + C - 1)} d\lambda.$$

This in particular implies that

$$\begin{aligned} \mathbb{E}[(K - S_T(\ell))^+] &= \frac{1}{2\pi} \int_{\mathbb{R}} \mathbb{E}[e^{(i\lambda + C) \log(S_T(\ell))}] \frac{K^{-C+1-i\lambda}}{(i\lambda + C)(i\lambda + C - 1)} d\lambda \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \mathbb{E}[e^{\langle \mathbf{u}_\lambda, \widehat{Z}_T \rangle}] \frac{K^{-C+1-i\lambda}}{(i\lambda + C)(i\lambda + C - 1)} d\lambda \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} e^{\psi_\lambda(T)_\emptyset} \frac{K^{-C+1-i\lambda}}{(i\lambda + C)(i\lambda + C - 1)} d\lambda, \end{aligned}$$

⁴We refer to [Cuchiero et al. \(2023c\)](#) for the appropriate solution concept and to a numerical treatment in the one dimensional case where (2.6.5) reduces to a sequence-valued Riccati equation.

where $\mathbf{u}_\lambda := (i\lambda + C)\mathbf{u}(\ell)$ and ψ_λ is a solution of the Riccati equation with initial condition $\psi_\lambda(0) = \mathbf{u}_\lambda$.

Let us now consider the case of VIX options where Fourier pricing can be applied by computing the Fourier-Laplace transform of VIX squared, see also [Sepp \(2008\)](#); [Papanicolaou and Sircar \(2014\)](#); [Bondi et al. \(2024b\)](#) and references therein for a Fourier-based approach to pricing VIX options. Fix a labelling injective function

$$\mathcal{L} : \{I : |I| \leq n\} \rightarrow \{1, \dots, (d+1)_{(2n+1)}\}$$

as introduced before (2.2.2) and recall that by Theorem 2.5.1 it holds

$$\text{VIX}_T^2(\ell) = \frac{1}{\Delta} \ell^\top Q(T, \Delta) \ell$$

for

$$Q_{\mathcal{L}(I)\mathcal{L}(J)}(T, \Delta) = \sum_{e_K = (e_I \sqcup e_J) \otimes e_0} \sum_{|H| \leq 2n+1} (e^{\Delta G^\top} - \text{Id})_{\mathcal{L}(K)\mathcal{L}(H)} \langle e_H, \widehat{\mathbb{X}}_T \rangle.$$

where G denotes the $(d+1)_{(2n+1)}$ -dimensional matrix representative of the dual operator corresponding to $\widehat{\mathbb{X}}$.

Setting for $|I|, |J| \leq n$

$$I(\Delta)\mathbf{u} := \sum_{|K|, |H| \leq 2n+1} (e^{\Delta G^\top} - \text{Id})_{\mathcal{L}(K)\mathcal{L}(H)} \mathbf{u}_K e_H$$

we can write

$$\begin{aligned} \text{VIX}_T^2(\ell) &= \frac{1}{\Delta} \sum_{|I|, |J| \leq n} \ell_I \ell_J \sum_{e_K = (e_I \sqcup e_J) \otimes e_0} \sum_{|H| \leq 2n+1} (e^{\Delta G^\top} - \text{Id})_{\mathcal{L}(K)\mathcal{L}(H)} \langle e_H, \widehat{\mathbb{X}}_T \rangle \\ &= \frac{1}{\Delta} \sum_{|K|, |H| \leq 2n+1} (e^{\Delta G^\top} - \text{Id})_{\mathcal{L}(K)\mathcal{L}(H)} \langle e_K, (\ell \sqcup \ell) \otimes e_0 \rangle \langle e_H, \widehat{\mathbb{X}}_T \rangle \\ &= \frac{1}{\Delta} \langle I(\Delta)((\ell \sqcup \ell) \otimes e_0), \widehat{\mathbb{X}}_T \rangle. \end{aligned}$$

Also in this case since $\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{i\lambda y} (K - \sqrt{|y|})^+ dy = \frac{\sqrt{\frac{2}{\pi}} F_S(K\sqrt{|\lambda|})}{|\lambda|^{3/2}}$ is integrable for $F_S(u) = \int_0^u \sin(z^2) dz$, Fourier analysis yields

$$(K - \sqrt{y})^+ = \frac{1}{\pi} \int_{\mathbb{R}} e^{-i\lambda y} \frac{F_S(K\sqrt{|\lambda|})}{|\lambda|^{3/2}} d\lambda,$$

for each $y \geq 0$. This in particular implies that

$$\mathbb{E}[(K - \text{VIX}_T(\ell))^+] = \frac{1}{\pi} \int_{\mathbb{R}} \mathbb{E}[e^{-i\lambda \text{VIX}_T^2(\ell)}] \frac{F_S(K\sqrt{|\lambda|})}{|\lambda|^{3/2}} d\lambda = \frac{1}{\pi} \int_{\mathbb{R}} e^{\psi_\lambda(T)\emptyset} \frac{F_S(K\sqrt{|\lambda|})}{|\lambda|^{3/2}} d\lambda,$$

where ψ_λ is a solution of the Riccati equation with initial condition $\psi_\lambda(0) = -i\lambda \mathbf{v}(\ell)$ for $\mathbf{v}(\ell) := \frac{1}{\Delta} I(\Delta)((\ell \sqcup \ell) \otimes e_0)$.

Analogous formulations in terms of the error function are also possible, see for instance [Bondi et al. \(2024b\)](#). In the same spirit one can also obtain a representation of futures prices. We here do not provide an implementation of this Fourier pricing approach but numerical experiments can be found in [Cuchiero et al. \(2023c\)](#).

2.6.2 The case of time-varying parameters

Analogously to Section 2.5.4, we now further enhance Proposition 2.6.4 by allowing the parameters ℓ to depend on the maturity.

Proposition 2.6.9. Let $S = (S_t)_{t \geq 0}$ satisfy (2.3.1) with $S_0 = 1$, and $(\sigma_t^S)_{t \geq 0}$ satisfy (2.5.17) for a set of maturities $\mathcal{T}^{\text{VIX}} = \{T_1, \dots, T_N\}$. Recall that in this case $V = (V_t)_{t \geq 0}$ satisfy

$$V_t(\ell) = \sum_{i=0}^N \sum_{|J|, |I| \leq n} \ell_I(T_i) \ell_J(T_i) 1_{[T_i, T_{i+1})}(t) \langle e_I \sqcup e_J, \widehat{\mathbb{X}}_t \rangle.$$

Then, with the notation of Proposition 2.6.4 we write the following recursion for the log-price process

$$\begin{aligned} \log(S_t(\ell^{<m+1})) &= \sum_{i=0}^N \left[-\frac{1}{2} \ell(T_i)^\top (Q^0(t \wedge T_{i+1}) - Q^0(t \wedge T_i)) \ell(T_i) \right. \\ &\quad \left. + \sum_{|I| \leq n} \ell_I(T_i) \langle \tilde{e}_I^B, \widehat{\mathbb{Z}}_{t \wedge T_{i+1}} - \widehat{\mathbb{Z}}_{t \wedge T_i} \rangle \right] \end{aligned}$$

for each $t \geq 0$, where $T_0 := 0$, $\ell^{<m+1} := \{\ell(0), \dots, \ell(T_m)\}$, $m = \max\{j : T_j < t\}$, and

$$Q_{\mathcal{L}(I)\mathcal{L}(J)}^0(t) = \langle (e_I \sqcup e_J) \otimes e_0, \widehat{\mathbb{X}}_t \rangle,$$

for a labeling function $\mathcal{L} : \{I : |I| \leq n\} \rightarrow \{1, \dots, (d+1)_{2n+1}\}$.

Proof. We know that

$$\log(S_t(\ell)) = -\frac{1}{2} \int_0^t V_s(\ell) ds + \int_0^t \sigma_s^S(\ell) dB_s$$

and we will calculate each integral separately. We start with the first one.

$$\begin{aligned} \int_0^t V_s(\ell) ds &= \int_0^t \sum_{i=0}^N \sum_{|I|, |J| \leq n} \ell_I(T_i) \ell_J(T_i) 1_{[T_i, T_{i+1})}(s) \langle e_I \sqcup e_J, \widehat{\mathbb{X}}_s \rangle ds \\ &= \sum_{i=0}^N \sum_{|I|, |J| \leq n} \ell_I(T_i) \ell_J(T_i) \int_{t \wedge T_i}^{t \wedge T_{i+1}} \langle e_I \sqcup e_J, \widehat{\mathbb{X}}_s \rangle ds \\ &= \sum_{i=0}^N \sum_{|I|, |J| \leq n} \ell_I(T_i) \ell_J(T_i) \left(\int_0^{t \wedge T_{i+1}} \langle e_I \sqcup e_J, \widehat{\mathbb{X}}_s \rangle ds - \int_0^{t \wedge T_i} \langle e_I \sqcup e_J, \widehat{\mathbb{X}}_s \rangle ds \right) \\ &= \sum_{i=0}^N \sum_{|I|, |J| \leq n} \ell_I(T_i) \ell_J(T_i) \left(\langle (e_I \sqcup e_J) \otimes e_0, \widehat{\mathbb{X}}_{t \wedge T_{i+1}} \rangle - \langle (e_I \sqcup e_J) \otimes e_0, \widehat{\mathbb{X}}_{t \wedge T_i} \rangle \right) \\ &= \sum_{i=0}^N \left[\ell(T_i)^\top (Q^0(t \wedge T_{i+1}) - Q^0(t \wedge T_i)) \ell(T_i) \right]. \end{aligned}$$

Using similar arguments and Lemma 3.10 in [Cuchiero et al. \(2023a\)](#), the second integral yields

$$\begin{aligned}
\int_0^t \sigma_s^S(\ell) dB_s &= \int_0^t \sum_{i=0}^N \sum_{|I| \leq n} \ell_I(T_i) 1_{[T_i, T_{i+1})} \langle e_I, \widehat{\mathbb{X}}_s \rangle dB_s \\
&= \sum_{i=0}^N \sum_{|I| \leq n} \ell_I(T_i) \int_{t \wedge T_i}^{t \wedge T_{i+1}} \langle e_I, \widehat{\mathbb{X}}_s \rangle dB_s \\
&= \sum_{i=0}^N \sum_{|I| \leq n} \ell_I(T_i) \left(\int_0^{t \wedge T_{i+1}} \langle e_I, \widehat{\mathbb{X}}_s \rangle dB_s - \int_0^{t \wedge T_i} \langle e_I, \widehat{\mathbb{X}}_s \rangle dB_s \right) \\
&= \sum_{i=0}^N \sum_{|I| \leq n} \ell_I(T_i) \langle \tilde{e}_I^B, \widehat{\mathbb{Z}}_{t \wedge T_{i+1}} - \widehat{\mathbb{Z}}_{t \wedge T_i} \rangle,
\end{aligned}$$

and the claim follows. $\square$

2.7 Joint calibration of SPX and VIX options

We here consider again the model introduced in (2.3.1)-(2.3.2). Note that we just work with call options, but the setup can easily be extended also to other liquid options on the market. Again we denote by $\mathcal{T}^{\text{VIX}}$ and $\mathcal{T}^{\text{SPX}}$ the maturities set for options written on VIX and SPX, respectively. Similarly we use the notation $\mathcal{K}^{\text{VIX}}$ and $\mathcal{K}^{\text{SPX}}$ for the corresponding strikes. The functional to be minimized in order to achieve a joint calibration of the SPX/VIX options reads as follows:

$$L_{\text{joint}}(\ell, \lambda) := \lambda L_{\text{SPX}}(\ell) + (1 - \lambda) L_{\text{VIX}}(\ell), \quad (2.7.1)$$

where $\lambda \in (0, 1)$ and

- $L_{\text{VIX}}(\ell)$ is as in (2.5.13), i.e.

$$\sum_{T \in \mathcal{T}^{\text{VIX}}, K \in \mathcal{K}^{\text{VIX}}} \mathcal{L} \left(\pi_{\text{VIX}}^{\text{model}}(\ell, T, K), \pi_{\text{VIX}}^{b,a}(T, K), \sigma_{\text{VIX}}^{b,a}(T, K), F_{\text{VIX}}^{\text{model}}(\ell, T), F_{\text{VIX}}^{\text{mkt}}(T) \right),$$

with $\pi_{\text{VIX}}^{\text{model}}$ and $F_{\text{VIX}}^{\text{model}}$ as in (2.5.12) for $\text{VIX}_T(\ell, \omega_i)$ defined as in (2.5.5);

- $L_{\text{SPX}}(\ell)$ is the SPX loss function given by

$$L_{\text{SPX}}(\ell) := \sum_{T \in \mathcal{T}^{\text{SPX}}, K \in \mathcal{K}^{\text{SPX}}} \mathcal{L}(\pi_{\text{SPX}}^{\text{model}}(\ell, T, K), \pi_{\text{SPX}}^{b,a}(T, K), \sigma_{\text{SPX}}^{b,a}(T, K)),$$

for a real-valued function $\mathcal{L}$. Observe that with a slight abuse of notation we denote this function as the one for L_{VIX} , but for SPX options we do not have to calibrate to futures, hence the last term of (2.5.15) vanishes.

By Proposition 2.6.4 the SPX call option payoff with maturity $T > 0$ and a strike price $K > 0$ reads in our model as follows

$$e^{-rT}(\tilde{S}_T(\ell) - K)^+ = e^{-rT} \left(\exp \left\{ (r - q)T - \frac{1}{2} \ell^\top Q^0(t) \ell + \sum_{|I| \leq n} \ell_I \langle \tilde{e}_I^B, \hat{\mathbb{Z}}_T \rangle \right\} - K \right)^+,$$

where $\tilde{S}$ denotes the undiscounted, unadjusted process as discussed in Remark 2.3.3 and $r, q > 0$ the interest rate and the dividends, respectively. Recall also that the call option payoff written on the VIX is given by

$$e^{-rT}(\text{VIX}_T(\ell) - K)^+ = e^{-rT} \left(\sqrt{\frac{1}{\Delta} \ell^\top Q(T, \Delta) \ell} - K \right)^+ = e^{-rT} \left(\frac{1}{\sqrt{\Delta}} \|U_T^\top \ell\| - K \right)^+,$$

where U_T denotes the upper-triangular matrix of the Cholesky decomposition of the symmetric positive semidefinite matrix $Q(T, \Delta)$.

Remark 2.7.1. We report in the table below the (average over 10^3 trials) timings of evaluating $\text{VIX}_T(\ell)$ and $S_T(\ell)$ for $\ell \in \mathbb{R}^{85}$, a fixed $T > 0$ and $N_{MC} = 8 \cdot 10^5$ samples on both CPU (on the left) and GPU (on the right) with PyTorch, respectively:

$(\text{VIX}_T(\ell)(\omega_i))_{i=1}^{N_{MC}}$	$(S_T(\ell)(\omega_i))_{i=1}^{N_{MC}}$	$(\text{VIX}_T(\ell)(\omega_i))_{i=1}^{N_{MC}}$	$(S_T(\ell)(\omega_i))_{i=1}^{N_{MC}}$
0.771s	0.702s	0.04s	0.014s

This evaluations are the relevant operations in the Monte Carlo pricing and in turn in the calibration procedure. Note again that both the sampling of the signature components and the matrix exponential, are achieved offline, as they do not depend on ℓ .

2.7.1 Numerical results

Before presenting our numerical results, let us discuss two different ways of approaching the joint calibration problem that can be found in the recent literature.

- (i) The first approach consists in choosing for instance the first maturity of SPX and VIX to coincide (or differ up to two days), i.e., $T_1^{\text{SPX}} = T_1^{\text{VIX}}$ and then for $j \geq 2$, $T_j^{\text{SPX}} = T_{j-1}^{\text{VIX}} + \Delta$, see for instance Guyon (2024); Guo et al. (2022b); Guyon and Lekeufack (2023).
- (ii) The second approach is to consider $\mathcal{T}^{\text{SPX}} = \mathcal{T}^{\text{VIX}}$, i.e., to choose the same (or close together) maturities both for SPX and VIX options. This perspective has been adopted for instance by Gatheral et al. (2018); Rosenbaum and Zhang (2021); Grzelak (2022); Bondi et al. (2024b).

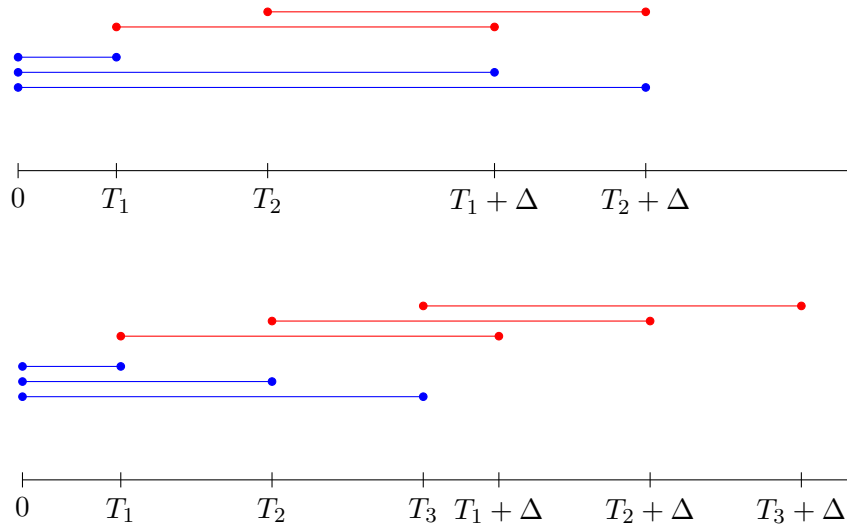


Figure 2.3: The blue lines denote the time interval where the dynamics of the variance process influence the SPX option up to the maturity time. For instance the shortest blue line denotes the time interval where the dynamics of the variance process enter up to maturity T_1 . Similarly the red lines denote the corresponding ones for the VIX, as for instance the variance process enters here in the time integral on $[T_1, T_1 + \Delta]$, see (2.5.3). On the upper graph a representation of the joint calibration approach (i) is given where we notice that the maturities of the VIX are chosen so that there is a maximal overlap with the ones of the SPX. On the lower graph a representation of approach (ii) is given where the maturities $\mathcal{T} = \{T_1, T_2, T_3\}$ are considered. We observe that there is less overlap between the maturities of the SPX and VIX than in approach (i).

Both approaches deal with the joint modeling of SPX and VIX options and in order to be consistent with both viewpoints taken in the literature, we show how our signature-based model solves the joint calibration within both settings. For this reason we split the rest of the section in two subsection and discuss them separately.

First approach

Here we consider call options for both indices on the trading day 02/06/2021, as in Guyon and Lekeufack (2023). Maturities are reported in the following tables with the corresponding range of strikes (in percentage) with respect to the spot and the market's futures prices.

$T_1^{\text{VIX}} = 0.0383$	$T_2^{\text{VIX}} = 0.0767$
(90%,220%)	(90%,220%)

$T_1^{\text{SPX}} = 0.0383$	$T_2^{\text{SPX}} = 0.1205$	$T_3^{\text{SPX}} = 0.1588$
(92%,105%)	(70%,105%)	(80%,120%)

2.7 Joint calibration of SPX and VIX options

We stress that the shortest maturity considered is of 14 days for both SPX and VIX, then the second and third maturity of the SPX are 44 days and 58 days, respectively, and the second one for the VIX is 28 days. Moreover, we consider a high moneyness level (up to 220%) for VIX options, usually rather difficult to fit. Regarding our modeling choice we fix $d = 3$, $n = 3$ and choose the primary process X to be a three dimensional Ornstein-Uhlenbeck process (see Example 2.4.5) with parameters

$$\kappa = (0.1, 25, 10)^\top, \quad \theta = (0.1, 4, 0.08)^\top, \quad \sigma = (0.7, 10, 5)^\top,$$

$$\rho = \begin{pmatrix} 1 & 0.213 & -0.576 & 0.329 \\ \cdot & 1 & -0.044 & -0.549 \\ \cdot & \cdot & 1 & -0.539 \\ \cdot & \cdot & \cdot & 1 \end{pmatrix}, \quad X_0 = (1, 0.08, 2)^\top.$$

Note that with this configuration we need to calibrate 85 parameters, i.e., $\ell \in \mathbb{R}^{85}$. Concerning the calibration task, we solve (2.7.1) with $\lambda = 0.35$ with $N_{MC} = 80000$ Monte Carlo samples for the previous maturities and strikes. Furthermore we specify the loss function $\mathcal{L}$ as in (2.5.15) for $\beta = 1$ both for SPX and VIX.

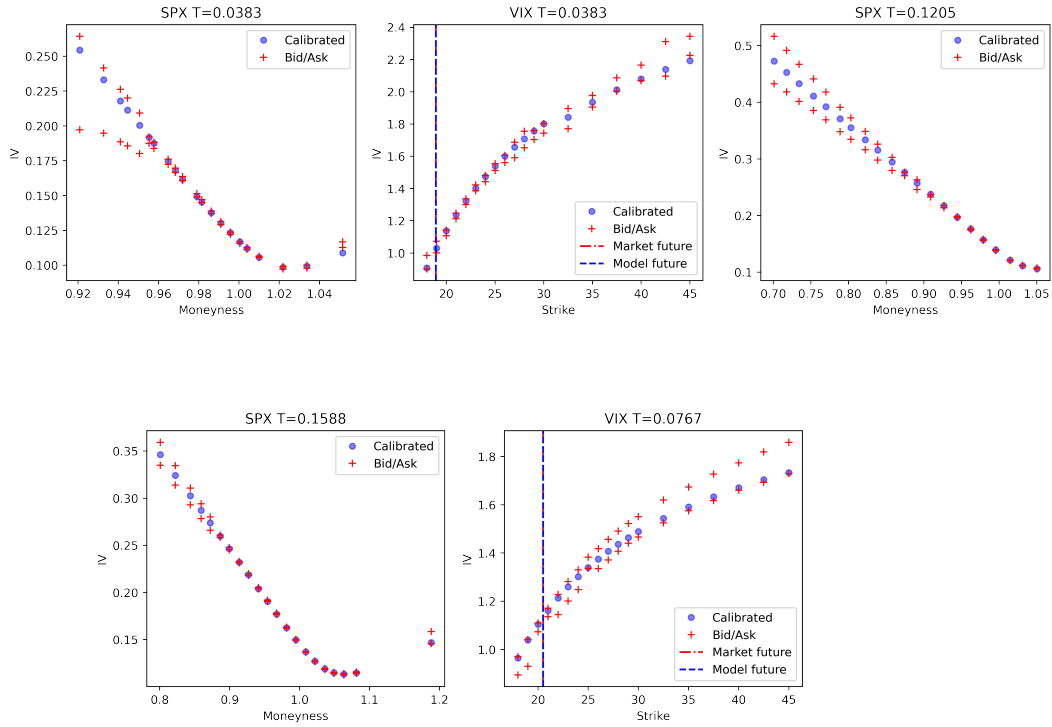


Figure 2.4: In blue the calibrated implied volatility smiles from top-left at maturities $T_1^{\text{SPX}}, T_1^{\text{VIX}}, T_2^{\text{SPX}}, T_3^{\text{SPX}}, T_2^{\text{VIX}}$. In red the corresponding bid-ask spreads. In the graphs of the VIX smiles the red dashed line indicates the market future price at maturity and the blue dashed line the calibrated one.

We report also the relative error between the market futures prices and the calibrated ones as defined in (2.5.16):

$T_1^{\text{VIX}} = 0.0383$	$T_2^{\text{VIX}} = 0.0767$
$\varepsilon_{T_1^{\text{VIX}}} = 9.8 \cdot 10^{-6}$	$\varepsilon_{T_2^{\text{VIX}}} = 6.6 \cdot 10^{-8}$

Simulation of time-series of SPX and VIX Let $\ell^* \in \mathbb{R}^{85}$ be the calibrated parameters already used for Figure 2.4. We then fix $T = 60$ days the longest considered maturity for the SPX and sample a trajectory for $(V_t(\ell^*))_{t \in [0, T]}$, $(\text{VIX}_t(\ell^*))_{t \in [0, T]}$, $(S_t(\ell^*))_{t \in [0, T]}$. Precisely, we sample 12 grid points per day, i.e. we consider a 2 hours sampling per calendar day, for a total of $N = 720$ grid points. The results of this simulation are reported in Figure 2.5.

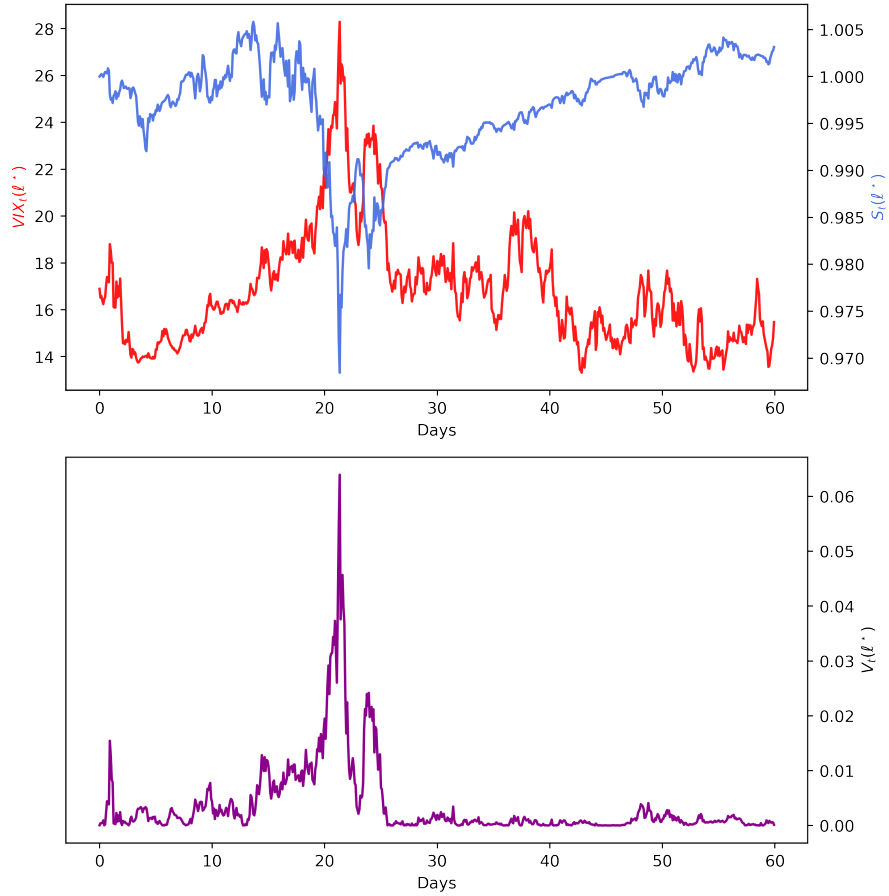


Figure 2.5: On the top: one realization of the calibrated model $S(\ell^*)$ for the SPX (in blue) and the corresponding calibrated VIX (in red). On the bottom: the corresponding realization of the calibrated variance process $V(\ell^*)$.

Observe that even though ℓ^* was only calibrated to option prices, the trajectories produced by the model are economically reasonable and also in line with several stylized facts, such as negative correlation between SPX and VIX or volatility clustering. To obtain the dynamics under the physical measure, these trajectories could still be adjusted by an appropriate market price of risk, but the quantities which are invariant under equivalent measure changes like the volatility of volatility or the correlation stay the same.

The case of time-varying parameters Next, we consider again the case of time-varying parameters as introduced in Section 2.5.4 and Section 2.6.2 for VIX and SPX, respectively. Although the joint calibration is mainly considered for short-dated options in the literature as VIX options are then more liquid, it is even more challenging to provide an accurate fit for both, short and long maturities. Allowing the parameters ℓ of our model to depend on time, in particular on the maturities, we are able to calibrate additionally to longer maturities than the ones considered in Figure 2.4. We consider for the choice of the primary process the same configuration as we used for Figure 2.4. The procedure of our time-varying calibration routine is as follows:

1. Calibrate jointly $T_1^{\text{SPX}}, T_1^{\text{VIX}}$ and T_2^{SPX} .
2. Use the parameters from the calibration of T_j^{SPX} and T_{j-1}^{VIX} to fit jointly the maturities T_{j+1}^{SPX} and T_j^{VIX} for $j = 2, \dots, J$.

We consider $J = 4$, where the last maturity for the SPX is 170 days, and the last maturity for the VIX is 77 days. For the first two maturities of the SPX and the first of the VIX we consider the same moneyness ranges as in Figure 2.4, hence we specify here only the ranges for the longer maturities:

$T_2^{\text{VIX}} = 0.1342$	$T_3^{\text{VIX}} = 0.2875$	$T_4^{\text{VIX}} = 0.3833$
(90%,330%)	(78%,395%)	(80%,405%)

$T_3^{\text{SPX}} = 0.2163$	$T_4^{\text{SPX}} = 0.3696$	$T_5^{\text{SPX}} = 0.4654$
(75%,125%)	(60%,135%)	(50%,145%)

We observe that for this choice of maturities Assumption 2.5.11 is satisfied. Hence the second expression for the time-varying VIX is used from Proposition (2.5.12). On the other hand in order to compute the price of the SPX options in the time-varying case we use the representation of the log-price provided in Proposition 2.6.9. In (2.7.1), we employ $\lambda = 0.25$ for each calibration within the rolling procedure and we consider always as loss function $\mathcal{L}^\beta$ as introduced in (2.5.15) for $\beta = 0$. It is worth mentioning that the initial parameter search discussed in Remark 2.5.8, has been employed for calibrating jointly $T_1^{\text{SPX}}, T_1^{\text{VIX}}$ and T_2^{SPX} , whereas for the next slices we have considered the previously calibrated parameters as starting point of the optimization.

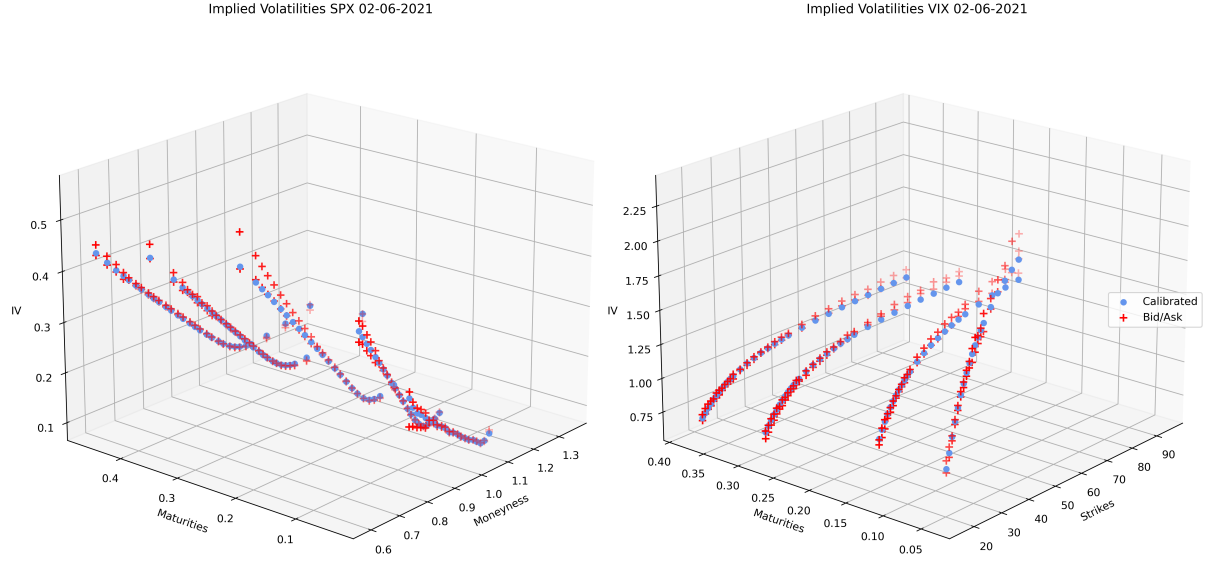


Figure 2.6: On the left-hand side: SPX smiles, in blue the calibrated implied volatilities and in red the bid-ask spreads. On the right-hand side: VIX smiles, in blue the calibrated implied volatilities and in red the bid-ask spreads.

Finally we report the absolute relative error on the VIX futures' prices:

$T_1^{\text{VIX}} = 0.0383$	$T_2^{\text{VIX}} = 0.1205$	$T_3^{\text{VIX}} = 0.1588$	$T_4^{\text{VIX}} = 0.2108$
$\varepsilon_{T_1^{\text{VIX}}} = 1.2 \cdot 10^{-11}$	$\varepsilon_{T_2^{\text{VIX}}} = 2.3 \cdot 10^{-5}$	$\varepsilon_{T_3^{\text{VIX}}} = 3.9 \cdot 10^{-5}$	$\varepsilon_{T_4^{\text{VIX}}} = 2.9 \cdot 10^{-6}$

Second approach

Let us now consider the second approach described at the beginning of Section 2.7.1. Specifically, we consider a unique set of maturities for both SPX and VIX on the trading day of 02/06/2021. For this study, we do not consider time-varying parameters. In the following table we report the moneyness ranges for SPX options in the second row and on the last row the ones for VIX options:

$T_1 = 0.0192$	$T_2 = 0.0383$	$T_3 = 0.0575$	$T_4 = 0.0767$
(97%,104%)	(92%,105%)	(90%,110%)	(85%,110%)
(90%,200%)	(90%,220%)	(90%,290%)	(90%,290%)

We consider $\lambda = 0.5$ and as loss function $\mathcal{L}$ we employ (2.5.15) with $\beta = 1$ for VIX options and the same (without futures) for SPX options.

2.7 Joint calibration of SPX and VIX options

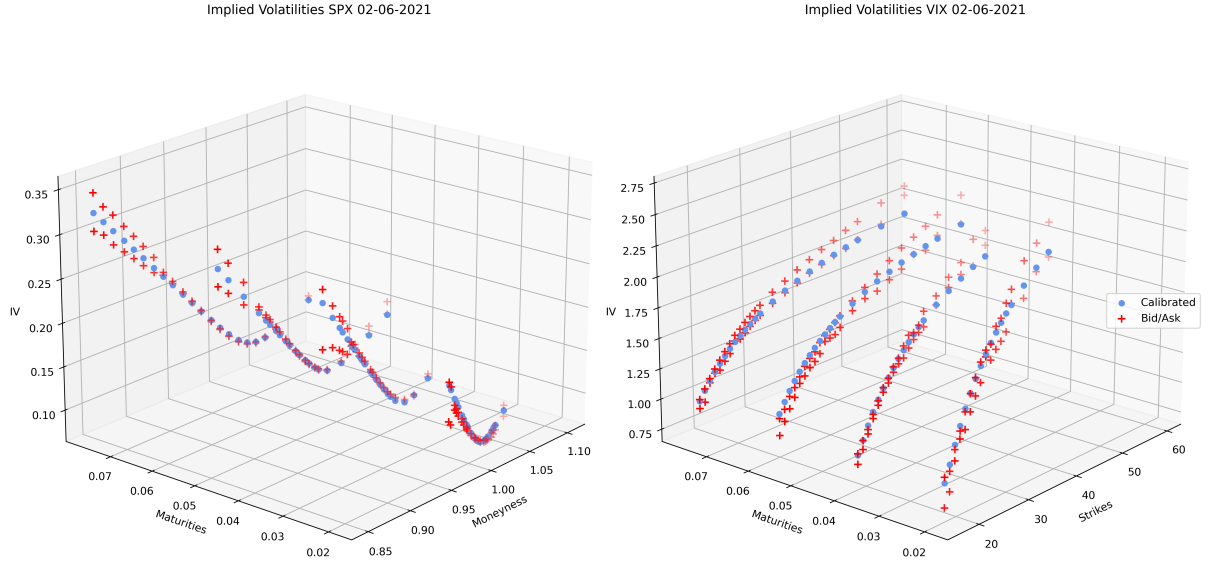


Figure 2.7: On the left-hand side: SPX smiles, in blue the calibrated implied volatilities and in red the bid-ask spreads. On the right-hand side: VIX smiles, in blue the calibrated implied volatilities and in red the bid-ask spreads.

We additionally report the relative error of the calibrated VIX futures:

$T_1^{\text{VIX}} = 0.0192$	$T_2^{\text{VIX}} = 0.0383$	$T_3^{\text{VIX}} = 0.0575$	$T_4^{\text{VIX}} = 0.0767$
$\varepsilon_{T_1^{\text{VIX}}} = 6.2 \cdot 10^{-3}$	$\varepsilon_{T_2^{\text{VIX}}} = 1.2 \cdot 10^{-5}$	$\varepsilon_{T_3^{\text{VIX}}} = 1.3 \cdot 10^{-2}$	$\varepsilon_{T_4^{\text{VIX}}} = 1.6 \cdot 10^{-3}$

Appendix of Chapter 2

2.A Numerical results for the Brownian motion case

This appendix is dedicated to the calibration to VIX options only, similarly as in Section 2.5.3, however with the primary process $(X_t)_{t \geq 0}$ being simply correlated Brownian motions (similarly as in Cuchiero et al. (2023a)) instead of OU-processes.

To be precise, we here model given by (2.3.1)-(2.3.2), where $(X_t)_{t \geq 0}$ is 2-dimensional Brownian motion. The correlation matrix of $Z = (X, B)$ is specified, as in Section 2.5.3, namely by

$$\rho = \begin{pmatrix} 1 & -0.577 & 0.3 \\ \cdot & 1 & -0.6 \\ \cdot & \cdot & 1 \end{pmatrix}.$$

For the other parameters we consider a truncation's level $n = 3$, we sample $N_{MC} = 80000$ trajectories for Monte Carlo pricing, and we minimize the loss function (2.5.15) with $\beta = 1$ to fit the same data-set as in Section 2.5.3.

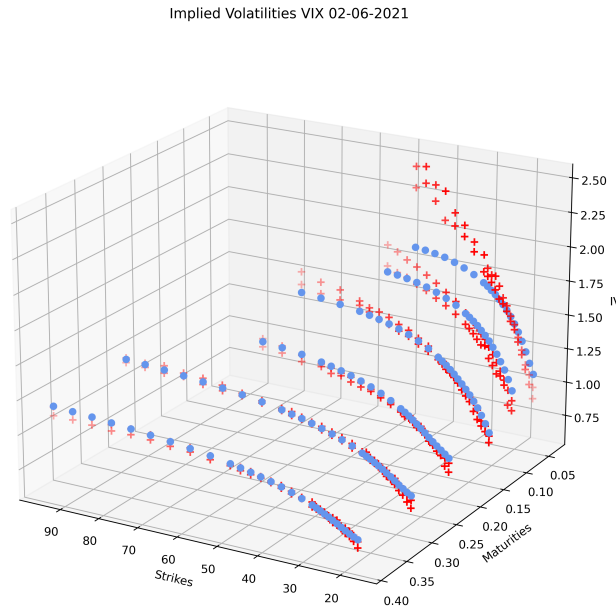


Figure 2.8: The red crosses denote the bid-ask spreads (of the implied volatilities) for each maturity, while the azure dots denote the calibrated implied volatilities of the model.

We observe that with this specification the model is neither able to calibrate to all future market prices (see Figure 2.9 below) nor to fit the market implied volatilities accurately. One can indeed see that the model implied volatilities often do not lie within the bid-ask spreads, in particular for high strikes and short maturities.

$T_1 = 0.0383$	$T_2 = 0.0767$	$T_3 = 0.1342$
$\varepsilon_{T_1} = 2.1 \times 10^{-5}$	$\varepsilon_{T_2} = 2.7 \times 10^{-2}$	$\varepsilon_{T_3} = 2.1 \times 10^{-7}$
$T_4 = 0.2108$	$T_5 = 0.2875$	$T_6 = 0.3833$
$\varepsilon_{T_4} = 1.6 \times 10^{-6}$	$\varepsilon_{T_5} = 6.8 \times 10^{-3}$	$\varepsilon_{T_6} = 2.1 \times 10^{-3}$

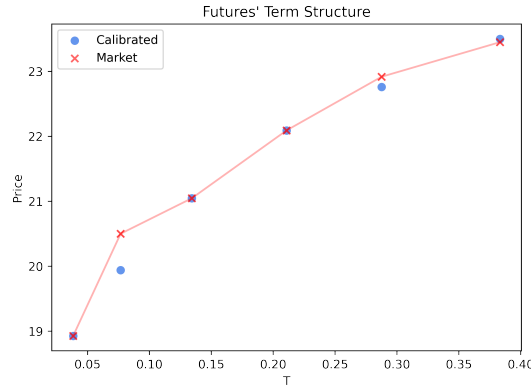


Figure 2.9: The blue circles denote the calibrated futures prices and the red crosses the market futures prices, in between a linear interpolation is applied.

2.B On the stability of the calibrated parameters

We report here an analysis of the stability of the calibrated parameters, which is for example essential for hedging since unstable parameters can lead to oscillating hedge ratios and high transaction cost.

For this purpose we have considered call options with the same time to maturity every trading week in the month of June 2021, i.e. we take the following five dates: June 2, 2021; June 9, 2021; June 16, 2021; June 23, 2021 and June 30, 2021. The times to maturity are on one hand $T_1^{\text{SPX}} = 14$ days and $T_2^{\text{SPX}} = 44$ days for SPX options and on the other hand $T_1^{\text{VIX}} = 14$ days for VIX (weekly) options. We employ the same primary process as of Section 7.1.1 for all 5 trading days. The goodness of fit of the respective implied volatilities is reported in Figures 2.13–2.16, omitting June 2, 2021 as it is already presented in Section 2.7.1. The

2.B On the stability of the calibrated parameters

obtained parameters are reported in Figures 2.10–2.12. The labels of the x -axis refer to the signature's index I and on the y -axis we have the corresponding coefficients ℓ_I .

Although no regularization on the model parameters has been enforced during the calibration, we observe from Figures 2.10–2.12 that most of the parameters which are close to zero in the first trading day, are kept in a neighbourhood of zero in the subsequent trading days. Likewise the more relevant parameters are mostly stable over the trading days or just slightly and continuously adjusted to fit the corresponding smiles, leading to an inherently stable calibration procedure. This experiment also indicates that – even though we consider on purpose an overparametrized model – it does not suffer from overfitting (which would be the case if the parameters were highly oscillatory).

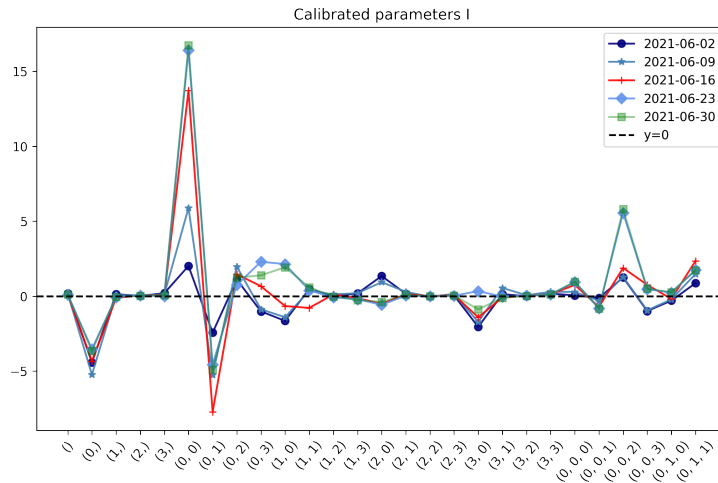


Figure 2.10: First subset of calibrated parameters ℓ_I .

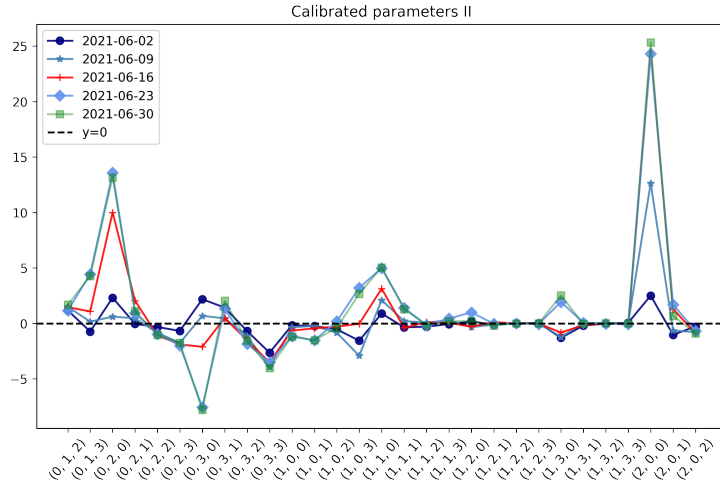


Figure 2.11: Second subset of calibrated parameters ℓ_I .

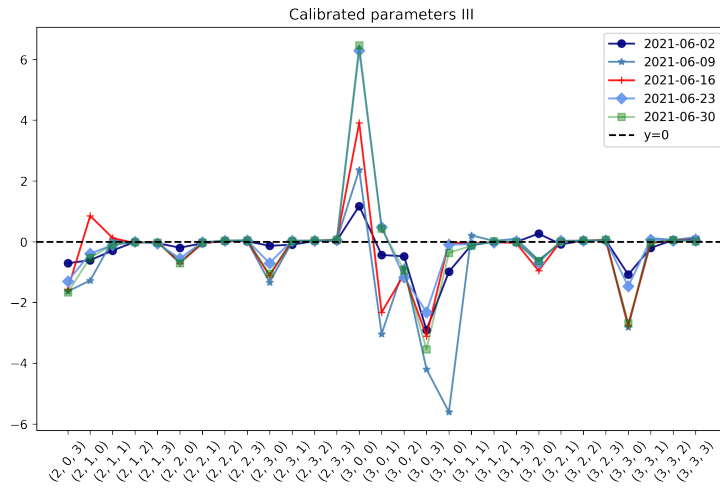


Figure 2.12: Third subset of calibrated parameters ℓ_I .

2.B On the stability of the calibrated parameters

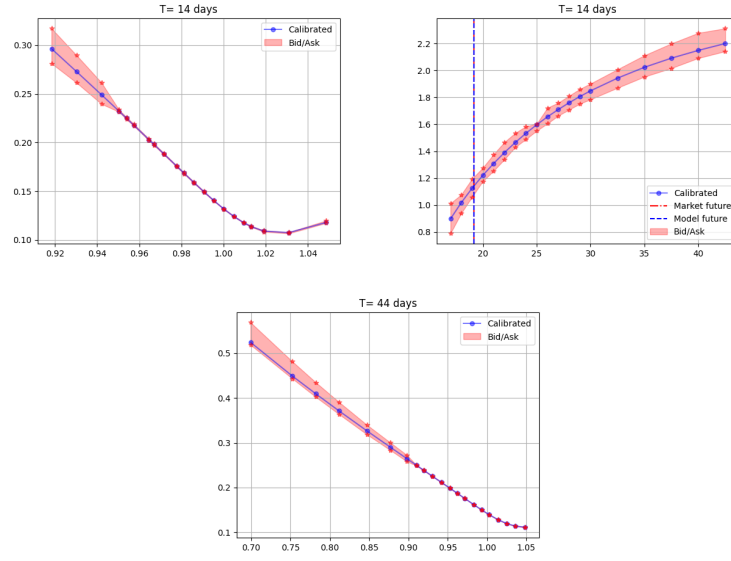


Figure 2.13: Implied volatilities as of June 9, 2021. In blue the calibrated implied volatility smiles from top-left at maturities T_1^{SPX} , T_1^{VIX} , T_2^{SPX} . In red the corresponding bid-ask spreads. In the graphs of the VIX smile the red dashed line indicates the market future price at maturity and the blue dashed line the calibrated one.

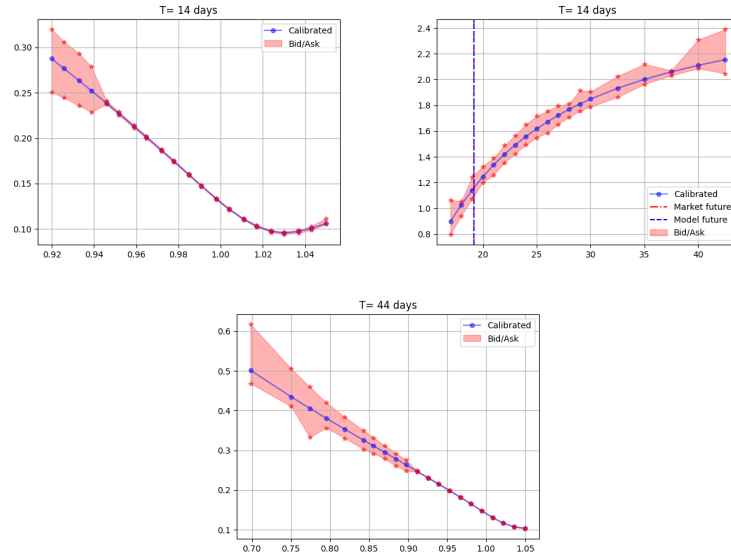


Figure 2.14: Implied volatilities as of June 16, 2021

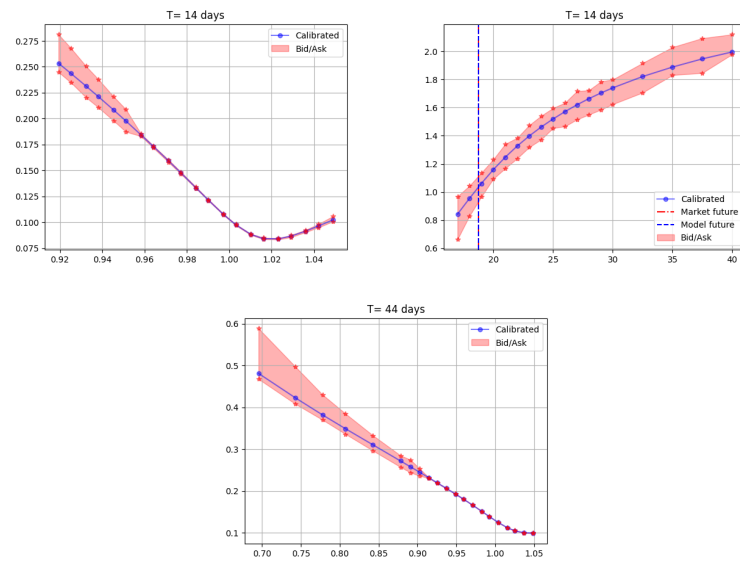


Figure 2.15: Implied volatilities as of June 23, 2021

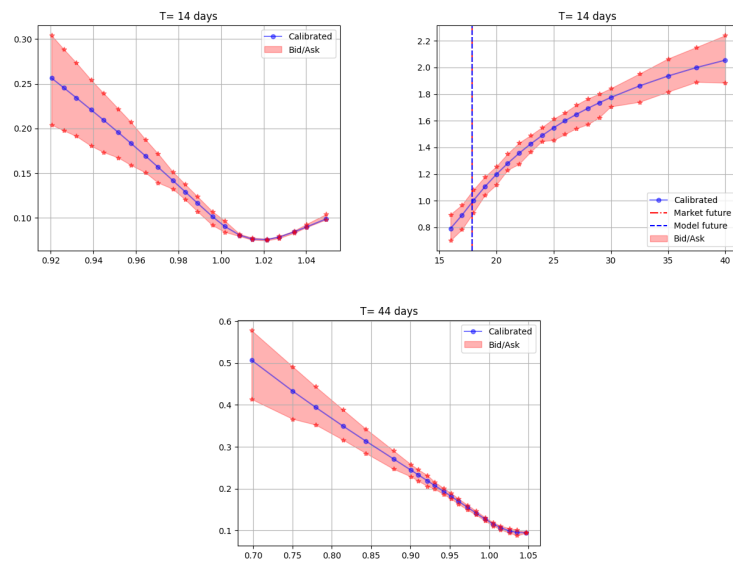


Figure 2.16: Implied volatilities as of June 30, 2021

3 Polynomial and Affine McKean-Vlasov SDEs

Abstract

We study a new class of McKean-Vlasov stochastic differential equations (SDEs) applying the theory of time-inhomogeneous polynomial and affine processes. In the polynomial case, the drift and volatility coefficients depend on the state variables themselves as well as their (conditional) moments in a way that mimics the standard polynomial structure. Such McKean-Vlasov SDEs arise for instance in the context of Stochastic Portfolio Theory when interacting particle systems are used to model the capital distribution curves. In the affine case, the drift and volatility may depend more generally on the characteristic function of the process, opening the door to numerous applications in mathematical finance where the affine characteristics are enhanced by law-dependent factors. Our approach leads to new results on the existence and uniqueness of solutions to (conditional) McKean-Vlasov SDEs which were, to the best of our knowledge, not obtainable using standard methods. Moreover, we show in the polynomial case (without common noise) that the moments of these McKean-Vlasov SDEs can be computed by non-linear ODEs. Additionally, in the affine case, the characteristic function can be obtained by solving a PDE.

3.1 Introduction

Stochastic differential equations (SDEs) of McKean-Vlasov type are a class of SDEs where the drift and diffusion coefficient depends on the marginal law of the solution. To make this precise, let us consider SDEs of the form

$$\begin{cases} dX_t = B(t, X_t, \mu_t)dt + \Sigma(t, X_t, \mu_t)dW_t, \\ \mu_t = \text{Law}(X_t), \end{cases} \quad (3.1.1)$$

where $(W_t)_{t \geq 0}$ is a one-dimensional Brownian motion. We refer to such SDEs as *McKean-Vlasov SDEs* in the following. Of course, McKean-Vlasov SDEs can be extended to multiple dimensions, i.e., such that $X_t \in \mathbb{R}^d$ for all $t \geq 0$. We, however, restrict ourselves to the one-dimensional case for the purpose of this work. McKean-Vlasov equations were introduced in McKean (1966) and have since then been widely studied. Amongst other fields, they gained importance in various applications in finance, see e.g. Guyon and Henry-Labordere (2013); Cuchiero and Huber (2024); Zhang (2021); Shkolnikov (2012); Carmona et al. (2015); Jourdain and Zhou (2020); Lacker et al. (2020); Djete (2024) as well as Guo et al. (2023);

[Pham and Wei \(2017\)](#); [Lacker \(2017\)](#) for literature on McKean-Vlasov optimal control. In financial systems, McKean-Vlasov SDEs often arise from a particle system, when the particles' dynamics depend on the empirical measure of the system and the number of particles goes to infinity. In this limit, one can then describe a "typical" particle by McKean-Vlasov dynamics, see [Cuchiero and Huber \(2024\)](#); [Shkolnikov \(2012\)](#); [Djete \(2024\)](#) for concrete examples.

In our work, we are interested in building a connection between the theory of affine and polynomial processes and McKean-Vlasov SDEs. This allows us to develop new solutions techniques and results on existence and uniqueness for certain classes of McKean-Vlasov SDEs. Before we elaborate more on our approach, we would like to provide a brief - but by no means complete - overview on existence and uniqueness results of solutions to McKean-Vlasov SDEs. The common approach consists in considering a space of probability measures $\mathcal{P}$ (often Wasserstein-2-space) and studying the maps B, Σ on the space $[0, \infty) \times \mathbb{R}^d \times \mathcal{P}$. A standard result in the literature is that existence and uniqueness of solutions to McKean-Vlasov SDEs can be established by requiring B, Σ to be Lipschitz-continuous in all arguments and then applying a fixed-point argument, see [Sznitman \(1991\)](#); [Chaintron and Diez \(2022\)](#). Since such Lipschitz assumptions are often too restrictive, establishing existence and uniqueness of solutions under milder assumptions is an active field of research, and we would like to briefly point out a few such approaches. For examples, the authors of [Bahlali et al. \(2020\)](#) go beyond Lipschitz-continuity requiring instead an Osgood-type condition, however not considering any law-dependence of the diffusion term. In [Ding and Qiao \(2021\)](#), the authors use Euler-Maruyama approximations to show existence of a solution to a martingale problem, for which they establish in turn path-wise uniqueness leading to a strong solution. The approach taken in [Hammersley et al. \(2021\)](#) is based on Lyapunov functions. Another route is taken in [Mishura and Veretennikov \(2024\)](#) where the authors study the existence of weak and strong solutions to McKean-Vlasov SDEs, however requiring non-degeneracy in the diffusion term for weak solutions and additionally Lipschitz-continuity in the state-variable of the diffusion term to construct a strong solution. Very recently, [Djete \(2024\)](#) have established existence of McKean-Vlasov SDEs under mild continuity assumptions, in particular, without requiring the dependence on the law to be Lipschitz continuous. However, they do not obtain uniqueness of the solution.

In contrast to those approaches we consider a more specific dependence of the SDEs' coefficients on the law, that is, we let them primarily depend on the marginal moments. By this restriction, we do not only get rid of Lipschitz-requirements on the diffusion term, but we avoid having to work on the space of probability measures. Instead, we can fully treat the McKean-Vlasov SDEs under our specification using the theory of time-inhomogeneous Markov processes, without explicit dependence on any marginal law. In particular, we apply the theory of (time-inhomogeneous) affine and polynomial processes. The theory of time-homogeneous affine and polynomial processes was first introduced in [Duffie et al. \(2003\)](#) and [Cuchiero et al. \(2012\)](#) respectively and has become increasingly popular in the theory of stochastic processes and application to mathematical finance [Cuchiero and Svaluto-Ferro \(2021\)](#); [Cuchiero \(2019\)](#); [Filipović and Larsson \(2016\)](#); [Cuchiero and Huber \(2024\)](#); [Gatheral and Keller-Ressel \(2019\)](#); [Thorsten Schmidt \(2020\)](#); [Fontana et al. \(2024\)](#); [Geuchen et al. \(2024\)](#) and Chapter 2. We

also refer the reader to [Larsson \(2017\)](#) for an introduction to the topic. The appeal of this class of processes lies in the fact that one has an analytic expression for their (conditional) moments and, in the affine case, the same holds additionally for the characteristic function of the process' marginals. For our purpose, we need to work in a time-inhomogeneous setting, which was first introduced in [Hurtado and Schmidt \(2020\)](#) for polynomial and in [Filipović \(2005\)](#) for affine processes, respectively. For a Markov process to be polynomial is a property of its transition operator or infinitesimal generator. More specifically, they are required to map polynomials to (time-dependent) polynomials of the same or lower degree. In this context, it is therefore natural to establish also existence of solutions to SDEs via properties of their generators. This is the approach we pursue in this work for the case of *polynomial McKean-Vlasov SDEs*, outlined in detail in Sections 3.3 and 3.4. In the affine case, existence and uniqueness of solutions can be established via certain admissibility conditions on the parameters describing the infinitesimal generator, see ([Filipović, 2005](#), Section 5).

To me more specific, we treat, in the polynomial case without common noise, McKean-Vlasov SDEs of the following form and establish existence of solutions thereof:

Theorem 3.1.1. *Consider a McKean-Vlasov SDE of the form*

$$dZ_t = (b(\mathbb{E}[Z_t]) + \beta(\mathbb{E}[Z_t])Z_t) dt + \sqrt{c(\mathbb{E}[Z_t]) + \gamma(\mathbb{E}[Z_t])Z_t + \Gamma(\mathbb{E}[Z_t])Z_t^2} dW_t$$

for $N \geq 1$, fixed $\mathbb{E}[Z_0] \in \mathbb{R}^N$, maps $b, \beta, c, \gamma, \Gamma : \mathbb{R}^N \rightarrow \mathbb{R}$ and a Brownian motion $(W_t)_{t \geq 0}$. If $b, \beta, c, \gamma, \Gamma$ are locally Lipschitz continuous on $\mathbb{R}^N$, satisfy either the set of Assumptions A1 or Assumptions A2, and the term under the square-root is non-negative, then there exists a unique strong solution $(Z_t)_{t \geq 0}$. Moreover, $(Z_t)_{t \geq 0}$ is a time-inhomogeneous polynomial process whose moments can be computed by the above non-linear ODE.

Additionally, we would like to point out that studying McKean-Vlasov SDEs from an affine or polynomial perspective has benefits beyond establishing the existence of solutions. Indeed, the solutions are in particular time-inhomogeneous polynomial or affine processes themselves, which means that we obtain analytic expressions for their (conditional) moments, and additionally for their characteristic functions in the affine case. We believe that this can be of importance for modeling purposes and other applications. Moreover, we can characterize state-spaces of the solutions. In particular, we can construct McKean-Vlasov SDEs such that the solution stays in a compact set in the polynomial case, which is another aspect we believe may be useful in modeling in finance but also in areas like population genetics.

We also present an extension of polynomial McKean-Vlasov SDEs to systems with a common noise term and coefficients depending on the moments conditional on the common noise. We call these processes *conditional polynomial McKean-Vlasov SDEs* and present the corresponding results in Subsection 3.4.3. A particular instance of such an SDE has already appeared in [Cuchiero and Huber \(2024\)](#) as a model for capital distribution curves in the context of Stochastic Portfolio Theory. In our work, we study *conditional polynomial McKean-Vlasov SDEs* more broadly and establish general sufficient conditions for existence and uniqueness.

In summary, this chapter is structured as follows. In Section 3.2, we briefly introduce the theory of time-inhomogeneous polynomial and affine processes, with an emphasis on the aspects that are relevant for our main results. After that, in Section 3.3, we first recall what it means to establish existence of a solution to an SDE via the solution to the associated martingale problem, following the work of Ethier and Kurtz (2005) and then extend some results of Ethier and Kurtz (2005) to milder growth-conditions on time-dependent coefficients of SDEs. We are finally ready to present our main results on McKean-Vlasov SDEs in the polynomial setting in Section 3.4. There, we present two approaches, called primal and dual, to establish existence of solutions. The primal approach is then extended to the setting with common noise. We proceed to the affine setting in Section 3.5 which allows us to study McKean-Vlasov SDEs that depend on the characteristic function of the process' marginals and thus potentially on the full law and not only on the moments.

3.2 Preliminaries

3.2.1 Time-inhomogeneous polynomial processes

In this subsection, we introduce the notion of time-inhomogeneous polynomial processes and state some of their properties, which will be crucial for the main results of this chapter. We mainly follow Hurtado and Schmidt (2020). However, while in Hurtado and Schmidt (2020) general time-inhomogeneous polynomial processes possibly with jumps are studied, we restrict ourselves to the case of continuous processes, which is the most relevant setting for us. On the other hand, the authors of Hurtado and Schmidt (2020) formulate everything on a finite time-horizon, while we consider a possibly infinite time-horizon. This extension is straightforward, and we therefore only comment briefly on the required adjustments in some of the proofs.

Before moving to polynomial processes, we first briefly recall the definition of a Markov process as well as its associated family of transition functions and transition operators, where we mostly follow Eberle (2016). For further details on Markov processes, we refer to Eberle (2016); Ethier and Kurtz (2005).

Throughout, we fix a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, \infty)}, \mathbb{P})$. We consider a state-space $S \subseteq \mathbb{R}^d$ and denote its Borel σ -algebra by $\mathcal{S}$.

Definition 3.2.1 (Markov Process). We call a stochastic process $(X_t)_{t \in [0, \infty)}$ taking values in S a *Markov process* if it is $(\mathcal{F}_t)_{t \in [0, \infty)}$ -adapted and

$$\mathbb{P}[X_t \in B | \mathcal{F}_s] = \mathbb{P}[X_t \in B | X_s] \quad \mathbb{P}\text{-a.s. for any } B \in \mathcal{S} \text{ and } 0 \leq s \leq t < \infty$$

Definition 3.2.2 (Transition Function). Let us define $\Delta = \{(s, t) | 0 \leq s \leq t < \infty\}$. We associate with a Markov process X a family of functions, called *transition function*, $(p_{s,t})_{(s,t) \in \Delta}$ such that $p_{s,t} : (S, \mathcal{S}) \rightarrow [0, 1]$ for all $(s, t) \in \Delta$ and

$$\mathbb{P}[X_t \in B | \mathcal{F}_s] = p_{s,t}(X_s, B) \quad \mathbb{P}\text{-a.s. for any } (s, t) \in \Delta \text{ and } B \in \mathcal{S}.$$

Definition 3.2.3 (Transition Operator). We can describe the Markov process $(X_t)_{t \in [0, \infty)}$ via transition operators $(P_{s,t})_{(s,t) \in \Delta}$ with

$$P_{s,t}f(x) = \int_S f(\xi) p_{s,t}(x, d\xi)$$

for all $x \in S$ and $\mathcal{S}$ -measurable functions $f : S \rightarrow \mathbb{R}_{\geq 0}$. Moreover, these transition operators satisfy the *Chapman-Kolmogorov* equations

$$P_{s,u}f(x) = P_{s,t}P_{t,u}f(x)$$

for all $0 \leq s \leq t \leq u < \infty$ and f as above.

Following [Hurtado and Schmidt \(2020\)](#), time-inhomogeneous polynomial processes are a particular type of Markov process, for which we shall need the subsequent notions.

We denote by $\mathbf{n} := (n_1, \dots, n_d) \in \{0, 1, \dots\}^d$ a multiindex, where $|\mathbf{n}| := n_1 + \dots + n_d$ and for $x \in \mathbb{R}^d$, $x^{\mathbf{n}} := x_1^{n_1} \dots x_d^{n_d}$. Let us introduce the following spaces of (time-dependent) polynomials up to degree $m \geq 0$:

$$\begin{aligned} \mathcal{P}_m(S) &:= \left\{ x \mapsto \sum_{|\mathbf{n}|=0}^m a_{\mathbf{n}} x^{\mathbf{n}} \mid x \in S, a_{\mathbf{n}} \in \mathbb{R} \right\} \\ \tilde{\mathcal{P}}_m(S) &:= \left\{ (s, t, x) \mapsto \sum_{|\mathbf{n}|=0}^m a_{\mathbf{n}}(s, t) x^{\mathbf{n}} \mid (s, t, x) \in \Delta \times S, a_{\mathbf{n}}(\cdot) \in C^1(\Delta, \mathbb{R}) \right\}. \end{aligned}$$

Definition 3.2.4 (Polynomial Process). We call a family of transition operators $(P_{s,t})_{(s,t) \in \Delta}$ *polynomial* if for all $m \geq 0$ and $f \in \mathcal{P}_m(S)$ it holds that

$$((s, t, x) \mapsto P_{s,t}f(x)) \in \tilde{\mathcal{P}}_m(S).$$

Moreover, we call a Markov process $(X_t)_{t \in [0, \infty)}$ a (*time-inhomogeneous*) *polynomial process* if its associated transition operators are polynomial.

Remark 3.2.5. Let us note that in the above definition, we have implicitly assumed that polynomials are $\mathcal{S}$ -measurable and, more crucially, that all moments exist.

Definition 3.2.6 (Infinitesimal Generator). For a Markov process with transition operators $(P_{s,t})_{(s,t) \in \Delta}$ we define the family of *infinitesimal generators* $(\mathcal{H}_s)_{s \in [0, \infty)}$ as the linear operators given by,

$$(\mathcal{H}_s f)(x) := \lim_{h \searrow 0} \frac{(P_{s,s+h}f - f)(x)}{h} \quad \text{for all } x \in S,$$

whenever this limit exists. We denote the set of $\mathcal{S}$ -measurable functions $f : S \rightarrow \mathbb{R}$ for which the limit exists by $\mathcal{D}(\mathcal{H}_s)$.

Proposition 3.2.7. Let $(\mathcal{H}_s)_{s \in [0, \infty)}$ be the infinitesimal generator associated with a polynomial transition operator, then it holds for all $m \geq 0$ and $s \in [0, \infty)$ that $\mathcal{P}_m(S) \subseteq \mathcal{D}(\mathcal{H}_s)$ and $\mathcal{H}_s(\mathcal{P}_m(S)) \subseteq \mathcal{P}_m(S)$.

Proof. The result follows from the fact that a polynomial transition operator maps polynomials to polynomials with continuously differentiable time-dependent coefficients. Hence, the limit translates to taking the derivative of these coefficients. Note that the coefficients of $\mathcal{H}_s f$ for $f \in \mathcal{P}_m(S)$ depend on $s \in [0, \infty)$. See (Hurtado and Schmidt, 2020, Proof of Lemma 3.3) for further details. $\square$

Having introduced the most important concepts of time-inhomogeneous polynomial processes, we focus now on *time-inhomogeneous polynomial diffusions*, which will be important for the main results of this chapter.

Definition 3.2.8 (Polynomial Diffusion). Let $(X_t)_{t \in [0, \infty)}$ be a continuous semimartingale with unique decomposition $X_t = X_0 + A_t + M_t$ where $(A_t)_{t \in [0, \infty)}$ is a process of finite variation and $(M_t)_{t \in [0, \infty)}$ is a local martingale. We denote by $(\mathfrak{b}, \mathfrak{c})$ the differential characteristics of X , that is,

$$A^i = \int_0^\cdot \mathfrak{b}^i(s, X_s) ds$$

$$[X^i, X^j]_\cdot = \int_0^\cdot \mathfrak{c}^{i,j}(s, X_s) ds$$

where $[\cdot, \cdot]$ denotes the quadratic variation. We then call $(X_t)_{t \in [0, \infty)}$ a *polynomial diffusion* if for all $i, j \in \{1, \dots, d\}$, $t \in [0, \infty)$, $x \in S$ it holds that

- (i) $\mathfrak{b}^i(t, \cdot) \in \mathcal{P}_1(S)$ and $\mathfrak{b}^i(\cdot, x) \in C([0, \infty))$, and,
- (ii) $\mathfrak{c}^{i,j}(t, \cdot) \in \mathcal{P}_2(S)$ and $\mathfrak{c}^{i,j}(\cdot, x) \in C([0, \infty))$.

Proposition 3.2.9 ((Hurtado and Schmidt, 2020, Corollary 5.5)). A polynomial diffusion $(X_t)_{t \in [0, \infty)}$ is a polynomial process with infinitesimal generator given by

$$(\mathcal{H}_s f)(x) = \sum_{i=1}^d D_i f(x) \mathfrak{b}^i(s, x) + \frac{1}{2} \sum_{i,j=1}^d D_{i,j} f(x) \mathfrak{c}^{i,j}(s, x)$$

for all $f \in \mathcal{P}_m(S)$, $m \geq 0$, $s \in [0, \infty)$, $x \in S$.

Proof. The proof of this hinges on (Hurtado and Schmidt, 2020, Theorem 3.7, Theorem 5.4 and Lemma 5.4) and the extension for these results to hold for all $t \in [0, \infty)$ deserves a brief discussion. First of all, regarding (Hurtado and Schmidt, 2020, Lemma 5.4), the coefficients $\tilde{C}_1, \tilde{C}_2$ now depend on time, but they remain finite over all of $[0, \infty)$ and hence (Hurtado and Schmidt, 2020, Lemma 5.4) holds analogously on $[0, \infty)$. Moreover, Theorem (Hurtado and Schmidt, 2020, Lemma 5.3) holds on $[0, \infty)$ by noting that the Kolmogorov forward and backward equations hold analogously. For the forward equation, see Rüschemdorf et al. (2016). A solution for the backward equation can be constructed exactly as described in the proof of (Hurtado and Schmidt, 2020, Theorem 3.7); knowing a solution to the forward equation exists for all $t \in [s, \infty)$, we can apply (Hurtado and Schmidt, 2020, Lemma A.3) and obtain a solution to the backward equation for all $s \in [0, t)$ where t is arbitrary. $\square$

Let us also point out that a diffusion as defined above can always be realized (in law) via an SDE, see [Jacod and Shiryaev \(2003\)](#). One example of an SDE in this setting is the following:

Example 3.2.10. Let us consider an SDE of the form:

$$dX_t = (b(t) + \beta(t)X_t)dt + \sqrt{c(t) + \gamma(t)X_t + \Gamma(t)X_t^2}dW_t,$$

where $X_0 = x$, the maps $b, \beta, c, \gamma, \Gamma : [0, \infty) \rightarrow \mathbb{R}$ are continuous and $(W_t)_{t \geq 0}$ denotes a one-dimensional Brownian motion. Moreover, assume the term under the square-root is non-negative, β, Γ are uniformly bounded and there exists a continuous and strictly increasing map $h : [0, \infty) \rightarrow [0, \infty)$ such that b, c, γ are all dominated by h . Then there exists a path-wise unique strong solution $(X_t)_{t \geq 0}$ to this SDE, by Theorem 3.3.4 and Proposition 3.3.3. Moreover, looking at the differential characteristics

$$\begin{aligned} \mathbf{b}(t, x) &= b(t) + \beta(t)x \\ \mathbf{c}(t, x) &= c(t) + \gamma(t)x + \Gamma(t)x^2 \end{aligned}$$

$(X_t)_{t \geq 0}$ is a time-inhomogeneous polynomial process, in particular a polynomial diffusion, by Definition 3.2.8.

An important property of time-inhomogeneous polynomial processes is that we have an analytic expression for the conditional moments of the process. Before we state this so-called *moment-formula*, let us make explicit that the transition operators and infinitesimal generators of polynomial processes act *linearly* on polynomials and therefore admit a matrix representation with respect to a polynomial basis.

To this end, let us consider polynomials up to an arbitrary but fixed degree $m > 0$. Moreover, we denote by $(v_0, \dots, v_N)$ a basis such that $\mathcal{P}_m(S) = \text{span}\{(v_0, \dots, v_{N(m)})\}$, where $N(m)$ is the dimension of the degree- m polynomials on S . For a linear operator $A : \mathcal{P}_m(S) \rightarrow \mathcal{P}_m(S)$ we denote its matrix representation by $(\mathbf{A}_{ij})_{0 \leq i, j \leq N(m)}$ such that for all $i \in \{0, \dots, N(m)\}$

$$Av_i = \sum_{j=0}^{N(m)} \mathbf{A}_{ji} v_j. \quad (3.2.1)$$

Indeed, for polynomial processes and for fixed time-points $s \in [0, \infty)$ (resp. $(s, t) \in \Delta$) the operators $\mathcal{H}_s$ (resp. $P_{s,t}$) act linearly on the space of polynomials. Therefore, we associate with $\mathcal{H}_s$ a matrix $\mathbf{H}_s$ for each $s \in [0, \infty)$ and likewise with $P_{s,t}$ a matrix $\mathbf{P}_{s,t}$ for each $(s, t) \in \Delta$. We do want to emphasize that these matrices are time-dependent. Let the following example illustrate the structure of the matrix $\mathbf{P}_{s,t}$:

Example 3.2.11. Let us denote by $(\mathbf{P}_{s,t})_{:,i} := \mathbf{P}_{s,t} \cdot e_i$ the i -th column of $\mathbf{P}_{s,t}$. Moreover, let us consider $S = \mathbb{R}$ and choose the canonical basis $v_i = (x \mapsto x^i)$ for all $i \in \{0, \dots, N\}$. Then, by definition, it holds that

$$\mathbb{E}[X_t^i | X_s = x] = \mathbb{E}[v_i(X_t) | X_s = x] = (1, x, \dots, x^N) \cdot (\mathbf{P}_{s,t})_{:,i}.$$

Hence, the columns of $\mathbf{P}_{s,t}$ contain the *coefficients* of the polynomials to obtain the i -th moment.

We are now ready to state the moment formula:

Theorem 3.2.12 (Moment Formula). *For a polynomial process $(X_t)_{t \in [0, T]}$ it holds for any $f \in \mathcal{P}_m(S)$ where $f = \sum_{i=0}^{N(m)} u_i v_i$ that*

$$\mathbb{E}[f(X_t)|\mathcal{F}_s] = \sum_{i,j=0}^{N(m)} v_i(X_s) (e^{\Omega(s,t)})_{ij} u_j$$

for $u \in \mathbb{R}^{N(m)}$ and where $\Omega(s, t) = \sum_{k=0}^{\infty} \Omega_k(s, t)$ is given by the Magnus expansion of $(\mathbf{H}_s)_{s \in [0, T]}$ assuming we choose a basis such that $\int_0^T \|\mathbf{H}_s\|_2 ds < \pi$ for the spectral norm $\|\cdot\|_2$.¹ In particular - denoting by $[\cdot, \cdot]$ the commutator of two matrices, i.e. $[\mathbf{A}, \mathbf{B}] = \mathbf{A}\mathbf{B} - \mathbf{B}\mathbf{A}$ - the first three elements of the Magnus expansion are given by

$$\begin{aligned} \Omega_1(s, t) &= \int_s^t \mathbf{H}_r dr \\ \Omega_2(s, t) &= -\frac{1}{2} \int_s^t \int_s^{t_1} [\mathbf{H}_{t_1}, \mathbf{H}_{t_2}] dt_2 dt_1 \\ \Omega_3(s, t) &= \frac{1}{6} \int_s^t \int_s^{t_1} \int_s^{t_2} [\mathbf{H}_{t_1}, [\mathbf{H}_{t_2}, \mathbf{H}_{t_3}]] + [\mathbf{H}_{t_3}, [\mathbf{H}_{t_2}, \mathbf{H}_{t_1}]] dt_3 dt_2 dt_1. \end{aligned}$$

Hence, if $[\mathbf{H}_s, \mathbf{H}_t] = 0$ for all $s, t \geq 0$, then the above simplifies to

$$\mathbb{E}[f(X_t)|\mathcal{F}_s] = \sum_{i,j=0}^{N(m)} v_i(X_s) (e^{\int_s^t \mathbf{H}_r dr})_{ij} \cdot u_j.$$

Proof. Using Lemma 3.2.14 below, the result follows from standard ODE theory. We refer the reader to (Hurtado and Schmidt, 2020, Section 7) for more details. $\square$

Remark 3.2.13. In the time-homogeneous case, the infinitesimal generator is constant in time, i.e. $\mathbf{H}_s \equiv \mathbf{H}$ for all $s \geq 0$. Hence, the moment formula simplifies to

$$\mathbb{E}[f(X_t)|\mathcal{F}_s] = \sum_{i,j=0}^{N(m)} v_i(X_s) (e^{(t-s)\mathbf{H}})_{ij} \cdot u_j.$$

Lemma 3.2.14. For a polynomial process $(X_t)_{t \in [0, T]}$ its transition operator and infinitesimal generator are described by the Kolmogorov forward and backward ODEs. We denote by $\frac{d^+}{dt}$ the *right-hand-derivative*, that is $\frac{d^+}{dt} f(t) := \lim_{h \searrow 0} \frac{f(t+h) - f(t)}{h}$.

(i) for every fixed $t \in [0, \infty)$ it holds on $[0, t)$ that

$$\frac{d^+}{ds} \mathbf{P}_{s,t} = -\mathbf{H}_s \mathbf{P}_{s,t} \text{ with } \mathbf{P}_{t,t} = id$$

¹Note that the spectral norm depends on the choice of the basis. Let a matrix $\mathbf{A}$ have spectral norm $\alpha := \|\mathbf{A}\|_2$ under basis $(v_0, \dots, v_N)$, then it has spectral norm $\epsilon\alpha$ under the re-scaled basis $(\epsilon^{-1}v_0, \dots, \epsilon^{-1}v_N)$ for any $\epsilon > 0$. See (Hurtado and Schmidt, 2020, Section 7 & 9) for more details.

(ii) for every fixed $s \in [0, \infty)$ it holds on $[s, \infty)$ that

$$\frac{d^+}{dt} \mathbf{P}_{s,t} = \mathbf{P}_{s,t} \mathbf{H}_t \text{ with } \mathbf{P}_{s,s} = id.$$

Proof. This follows essentially from the definition of the infinitesimal generator and the Chapman-Kolmogorov equation. See (Hurtado and Schmidt, 2020, Propostion 7.1) and Rüschendorf et al. (2016). See also our comments in the proof of 3.2.9 regarding the extension to $[0, \infty)$. $\square$

3.2.2 Time-inhomogeneous affine processes

In this section, we introduce the theory of *time-inhomogeneous affine processes*, following the work of Filipović (2005). Before we introduce affine processes in detail, let us point out that we are again dealing with time-inhomogeneous Markov processes and that affine diffusions are in particular polynomial. In terms of state-spaces, we are mostly concerned with $S = \mathbb{R}$ or $S = \mathbb{R}_{\geq 0}$ for the purpose of this chapter. As it was the case for polynomial processes, we can define affine processes via their transition operators $(P_{s,t})_{(s,t) \in \Delta}$ which we do in the following:

Definition 3.2.15 (Time-inhomogeneous Affine Process). Let $(X_t)_{t \in [0, \infty)}$ be a time-inhomogeneous Markov process with state-space $S = \mathbb{R}$ or $S = \mathbb{R}_{\geq 0}$ and transition operators $(P_{s,t})_{(s,t) \in \Delta}$. Let us denote by $f_u(x) := e^{ux}$ for $x \in S$, $u \in \mathcal{U}$ where

$$\begin{cases} \mathcal{U} = i\mathbb{R} & \text{if } S = \mathbb{R} \\ \mathcal{U} = \{z + iv \mid z \in \mathbb{R}_{\leq 0}, v \in \mathbb{R}\} & \text{if } S = \mathbb{R}_{\geq 0}. \end{cases} \quad (3.2.2)$$

We call $(P_{s,t})_{(s,t) \in \Delta}$ *affine* if for all $(s, t) \in \Delta$ and $u \in \partial\mathcal{U} = i\mathbb{R}$ there exist $\phi(t, s, u), \psi(t, s, u) \in \mathbb{C}$ such that

$$(P_{s,t} f_u)(x) = e^{\phi(t, s, u) + \psi(t, s, u)x} \quad \text{for all } x \in S.$$

Remark 3.2.16. It is important to note that although (3.2.2) is formulated to hold on $\partial\mathcal{U}$ in the above definition, it is in fact possible to extend (3.2.2) to hold on all of $\Delta \times \mathcal{U}$, see (Filipović, 2005, Lemma 3.1, Proposition 4.3)

Necessary conditions for a one-dimensional diffusion to be affine are:

- $\mathbf{b}(t, \cdot) \in \mathcal{P}_1(S)$ and $\mathbf{b}(\cdot, x) \in C([0, \infty))$
- $\mathbf{c}(t, \cdot) \in \mathcal{P}_1(S)$ and $\mathbf{c}(\cdot, x) \in C([0, \infty))$

however, these are not sufficient, and we therefore need the notion of *strongly admissible parameters*:

Definition 3.2.17 (Strongly Admissible Parameters). Again, we consider as state-space to be either $S = \mathbb{R}$ or $S = \mathbb{R}_{\geq 0}$. We call a set of time-dependent parameters $b, \beta, c, \gamma : [0, \infty) \rightarrow \mathbb{R}$ *strongly admissible* if they are continuous maps, and it holds for all $t \geq 0$ that

- $b(t) \in S$;
- $\beta(t) \in \mathbb{R}$;
- $c(t) \equiv 0$ if $S = \mathbb{R}_{\geq 0}$ and $c(t) \geq 0$ if $S = \mathbb{R}$;
- $\gamma(t) \geq 0$ if $S = \mathbb{R}_{\geq 0}$ and $\gamma(t) \equiv 0$ if $S = \mathbb{R}$.

We are now ready to state the following about diffusion processes:

Corollary 3.2.18. *A diffusion taking values in S with differential characteristics given by*

$$\mathfrak{b}(t, x) = b(t) + \beta(t)x$$

$$\mathfrak{c}(t, x) = c(t) + \gamma(t)x$$

where (b, β, c, γ) are strongly admissible parameters on S , then the diffusion is a time-inhomogeneous affine process.

Proof. This follows directly from a more general statement which we present in Proposition 3.2.19 below. $\square$

Proposition 3.2.19 (See (Filipović, 2005, Theorem 2.13)). Let $(X_t)_{t \in [0, \infty)}$ be a time-inhomogeneous Markov process with state-space $S = \mathbb{R}$ or $S = \mathbb{R}_{\geq 0}$, transition operators $(P_{s,t})_{(s,t) \in \Delta}$ and an associated set of strongly admissible parameters (b, β, c, γ) . Then, the space-time process $(t, X_t)_{t \geq 0}$ is a Feller process with infinitesimal generator $\mathcal{H}^0$ and $C^{1,2}([0, \infty) \times S) \subset \mathcal{D}(\mathcal{H})$. It holds for all $t \geq 0, x \in S$ that

$$\mathcal{H}^0 f(t, x) = \partial_t f(t, x) + \mathcal{H}_t f(t, x) \quad \text{for all } f \in C^{1,2}([0, \infty) \times S)$$

and $(\mathcal{H}_s)_{s \in [0, \infty)}$

$$\mathcal{H}_s f(s, x) = \frac{1}{2} (c(s) + \gamma(s)x) \partial_{xx} f(s, x) + (b(s) + \beta(s)x) \partial_x f(s, x) \quad \text{for all } s \geq 0.$$

Moreover, the other direction holds as well; given a set of strongly admissible parameters (b, β, c, γ) there exists a unique affine Markov process $(X_t)_{t \geq 0}$ with transition operators $(P_{s,t})_{(s,t) \in \Delta}$ and where the time-augmented process has infinitesimal generator $\mathcal{H}^0$ given as above.

3.3 Existence and uniqueness of solutions to time-inhomogeneous diffusions

In this section, we present the approach to establish existence of weak solutions to SDEs via existence of solution to the *martingale problem*. First, we briefly outline the concept behind this approach and refer the reader to (Ethier and Kurtz, 2005, Chapter 4 and Chapter 5.3) for more details. In Proposition 3.3.3 and Theorem 3.3.4 we present sufficient conditions on the coefficients of the SDE for a solution to the martingale problem to exist

3.3 Existence and uniqueness of solutions to time-inhomogeneous diffusions

in a time-inhomogeneous setting. While Proposition 3.3.3 follows the same reasoning as in the time-homogeneous setting outlined e.g. in Larsson (2017), Theorem 3.3.4 is an extension of (Ethier and Kurtz, 2005, Proposition 5.3.5 and Theorem 5.3.10) allowing milder growth-conditions of them time-dependencies. Let us start by the definition of a solution to the martingale problem (see (Ethier and Kurtz, 2005, Chapter 4.3) for more details).

Definition 3.3.1 (Solution to the Martingale Problem). Let $(X_t)_{t \geq 0}$ be a stochastic process with state-space $S \subseteq \mathbb{R}^d$ on some filtered probability space and let $A : \mathcal{D}(A) \subseteq C_b(S) \rightarrow C_b(S)$ be a linear operator. We call $(X_t)_{t \geq 0}$ a *solution to the martingale problem for A* if

$$f(X_t) - \int_0^t (Af)(X_s) ds$$

is a martingale for all $f \in \mathcal{D}(A)$.

For our purpose, we are interested in the solution to martingale problems for a given (possibly time-extended) *generator of an SDE* $\mathcal{H}'$ with $\mathcal{D}(\mathcal{H}^0) = C_c^\infty(S)$.

Let us from now on consider a one-dimensional SDE of the form

$$dX_t = B(t, X_t)dt + \Sigma(t, X_t)dW_t \quad (3.3.1)$$

for continuous functions $B, \sigma : [0, \infty) \times S \rightarrow \mathbb{R}$, $S \subseteq \mathbb{R}$, some given initial distribution for X_0 and $(W_t)_{t \geq 0}$ denoting a one-dimensional Brownian motion. We can associate with this SDE a family of *time-dependent* generators $(\mathcal{H}_s)_{s \geq 0}$ acting on $f \in C_c^\infty(S)$ as

$$(\mathcal{H}_s f)(x) = B(s, x)f'(x) + \frac{1}{2}\Sigma^2(s, x)f''(x)$$

for all $s \geq 0$ and $x \in S$. We refer the reader to (Ethier and Kurtz, 2005, Proposition 5.3.1 and Lemma 5.3.2) on how to construct a weak solution of an SDE given a solution to the martingale problem for the generator of the SDE. For brevity, we merely note here that a weak solution to an SDE can be constructed, if a solution to the martingale problem of its generator exists.

Let us point out, that in the time-inhomogeneous case, we need to study the time-extended martingale problem to establish existence of a solution, see (Ethier and Kurtz, 2005, Theorem 4.7.1 & Proposition 5.3.5). For an SDE with time-dependent coefficients and associated generators $(\mathcal{H}_s)_{s \geq 0}$ we define the generator of the time-extended process $(t, X_t)_{t \geq 0}$ as

$$\begin{aligned} \mathcal{H}^0 : \{\tau f \mid (\tau, f) \in C_c^1([0, \infty)) \times C_c^\infty(S)\} &\rightarrow C([0, \infty) \times S) \\ \tau f &\mapsto \mathcal{H}^0 \tau f \end{aligned}$$

such that

$$\begin{aligned} \mathcal{H}^0 \tau f : [0, \infty) \times S &\rightarrow \mathbb{R} \\ (t, x) &\mapsto \tau(t)\mathcal{H}_t f(x) + \tau'(t)f(x). \end{aligned}$$

As in (Ethier and Kurtz, 2005, Proposition 5.3.5) we study the existence of the solution fully in the time-extended setting first and then project the solution down to the space-component.

Establishing the existence of a solution $(t, X_t)_{t \geq 0}$ to the martingale problem for $\mathcal{H}^0$ can be broken down to the following two essential steps (see (Ethier and Kurtz, 2005, Proposition 5.3.5 and Theorem 5.3.10)):

- (i) showing that $\mathcal{H}^0$ satisfies the *positive maximum principle*² (see (Ethier and Kurtz, 2005, Theorem 4.5.4))
- (ii) verifying that the solution $(t, X_t)_{t \geq 0}$ has continuous sample paths with values in $[0, \infty) \times S$ (see (Ethier and Kurtz, 2005, Proposition 3.9))

Clearly, both of the above steps will impose some conditions on the generator $\mathcal{H}^0$ which translate to conditions on the SDE's coefficients B and σ . Let now us tackle step (i):

Definition 3.3.2 (Positive Maximum Principle, see (Ethier and Kurtz, 2005, p. 165)). We say a linear operator A satisfies the *positive maximum principle (PMP)* on S if for $x_0 \in S$ such that $f(x_0) = \sup_{x \in S} f(x)$ for $f \in \mathcal{D}(A)$ it holds that $(Af)(x_0) \leq 0$.

We now establish conditions on the coefficients B, Σ for the positive maximum principle to be fulfilled on a given state-space $S \subseteq \mathbb{R}$ in the next proposition. Although we consider the time-inhomogeneous setting here, the reasoning is very similar to the time-homogeneous case outlined in (Larsson, 2017, Examples 4.15, 4.16 & 4.18).

Proposition 3.3.3. Let

$$dX_t = B(t, X_t)dt + \Sigma(t, X_t)dW_t$$

and denote by $\mathcal{H}^0$ the associated time-extended generator.

- (i) For $S = \mathbb{R}$, the generator $\mathcal{H}^0$ satisfies the PMP without further assumptions as $\Sigma^2(\cdot, \cdot) \geq 0$ on all of $[0, \infty) \times \mathbb{R}$.
- (ii) For $S = \mathbb{R}_{\geq 0}$, $\mathcal{H}^0$ satisfies the PMP, if $\Sigma^2(\cdot, 0) \equiv 0$ and $B(\cdot, 0) \geq 0$ for $x = 0$
- (iii) For $S = [0, 1]$, $\mathcal{H}^0$ satisfies the PMP, if $\Sigma^2(\cdot, 0) \equiv \Sigma^2(\cdot, 1) \equiv 0$ and the drift term fulfills $B(\cdot, 0) \geq 0$, $B(\cdot, 1) \leq 0$.

Proof. Let $(t_0, x_0) \in [0, \infty) \times S$ such that $\tau(t_0)f(x_0) = \sup_{(t,x) \in [0,\infty) \times S} \tau(t)f(x)$. We note that $\tau(t_0)f(x_0) \geq 0$ since the functions have compact support on $[0, \infty) \times S$. Denoting by S° the interior of S it clearly holds that for $(t_0, x_0) \in (0, \infty) \times S^\circ$

$$\tau'(t_0)f(x_0) = 0, \quad \tau(t_0)f'(x_0) = 0, \quad \tau(t_0)f''(x_0) \leq 0.$$

Hence, since $\Sigma^2(\cdot, \cdot) \geq 0$ it is straightforward to check that $\mathcal{H}^0$ satisfies the PMP on $(0, \infty) \times S^\circ$. Moreover, if $t_0 = 0$, it holds that

$$\tau'(t_0)f(x_0) \leq 0, \quad \tau(t_0)f'(x_0) = 0, \quad \tau(t_0)f''(x_0) \leq 0$$

and it is straightforward to check that $\mathcal{H}^0$ therefore fulfills the PMP on $[0, \infty) \times S^\circ$. We are now left to check the cases where S contains some boundary points, i.e. $\partial S \cap S \neq \emptyset$.

²We will introduce the positive maximum principle in Definition 3.3.2.

3.3 Existence and uniqueness of solutions to time-inhomogeneous diffusions

This case is relevant for $S = \mathbb{R}_{\geq 0}$ and $S = [0, 1]$ and yields the additional conditions. If $x_0 \in \partial S \cap S$ and $t_0 \in [0, \infty)$, it holds that

$$\begin{cases} \tau(t_0)f'(x_0) \leq 0 & \text{if } x_0 = \inf x \in S \\ \tau(t_0)f'(x_0) \geq 0 & \text{if } x_0 = \sup x \in S. \end{cases}$$

Note that we can not say anything about the second order derivative of f at the boundary and therefore require Σ^2 to vanish on boundary points of S . It is easy to check that the conditions on B at the boundary points are exactly what we need for the PMP to be satisfied. $\square$

Having taken care of (i), we address step (ii) in the following theorem:

Theorem 3.3.4. *Consider a time-inhomogeneous SDE of the form*

$$dX_t = B(t, X_t)dt + \Sigma(t, X_t)dW_t,$$

with state-space $S = \mathbb{R}$, $S = \mathbb{R}_{\geq 0}$ or $S = [0, 1]$ and $B, \Sigma : [0, \infty) \times S \rightarrow \mathbb{R}$ continuous such that the infinitesimal generator $\mathcal{H}^0$ satisfies the positive maximum principle on the corresponding state space. Then there exists a solution to the corresponding martingale problem, if there is a map $h(t) : [0, \infty) \rightarrow [0, \infty)$ which is continuous and strictly increasing s.t.

$$\Sigma^2(t, x) \leq K^\Sigma(1 + h^2(t) + h(t)x + x^2) \quad (3.3.2)$$

and

$$x \cdot B(t, x) \leq K^B(1 + h(t)x + x^2) \quad (3.3.3)$$

for all $x \in S$ and $t \in [0, \infty)$ and some constants $K^\Sigma, K^B \geq 0$. Moreover, this solution is path-wise unique if $\Sigma^2(t, x), B(t, x)$ are locally Lipschitz in x uniformly in t on bounded intervals, that is: for any open subset $U \subset S$ and $T > 0$ there is a constant K such that

$$|\Sigma^2(t, x) - \Sigma^2(t, y)| \vee |B(t, x) - B(t, y)| \leq K|x - y|$$

for all $x, y \in U$ and $t \in [0, T]$.

Proof. The proof regarding the existence of solutions is analogous to the proof of (Ethier and Kurtz, 2005, Theorem 5.3.10 and Proposition 5.3.5). The only difference is that we define

$$f_n(t, x) := \varphi\left(\frac{t}{n}\right) \varphi\left(\frac{x^2}{h^2(n)}\right)$$

for some $\varphi \in C_c^\infty([0, \infty))$ s.t. $\chi_{[0,1]} \leq \varphi \leq \chi_{[0,2]}$. We say a sequence of functions $\{f_n\}$ converges *boundedly and point-wise* to f (in short $\text{bp-lim}_{n \rightarrow \infty} f_n = f$) if it holds that $\sup_n |f_n| < \infty$ and $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ for all $x \in S$. Clearly, for the sequence $\{f_n\}_{n \in \mathbb{N}}$ as defined above, it holds that

$$\text{bp-lim}_{n \rightarrow \infty} f_n = \chi_{[0, \infty) \times S}.$$

Let Δ be the point at infinity of $[0, \infty) \times S$ and denote by $([0, \infty) \times S)^\Delta$ the one-point-compactification of $[0, \infty) \times S$. For a generator $\mathcal{H}^0$ we define its extension $(\mathcal{H}^0)^\Delta$ acting on $C([0, \infty) \times S)^\Delta$ by

$$(\mathcal{H}^0)^\Delta f = f(\Delta) + \mathcal{H}^0(f - f(\Delta))$$

for $f \in C([0, \infty) \times S)^\Delta$. Then, for $g_n := (\mathcal{H}^0)^\Delta f_n$, it holds that $g_n \rightarrow 0$ point-wise. The crucial step, where the growth conditions (3.3.2) and (3.3.3) are needed, is to ensure that $\inf_n \inf_{(t,x)} g_n(t, x) > -\infty$. By the definition of $g_n := (\mathcal{H}^0)^\Delta f_n$ and the properties of φ , we find that:

$$\begin{aligned} \inf_{(t,x)} g_n(t, x) &\geq \inf_{(t,x)} \underbrace{\frac{1}{n} \varphi' \left(\frac{t}{n} \right)}_{\leq 0} \underbrace{\varphi \left(\frac{x^2}{h^2(n)} \right)}_{\geq 0} + \inf_{(t,x)} \left[\underbrace{\frac{2}{h^2(n)} x \cdot B(t, x)}_{\geq 0} \underbrace{\varphi \left(\frac{t}{n} \right)}_{\geq 0} \underbrace{\varphi' \left(\frac{x^2}{h^2(n)} \right)}_{\leq 0} \right. \\ &\quad \left. + \frac{1}{h^2(n)} \sigma^2(t, x) \left(\underbrace{\varphi' \left(\frac{x^2}{h^2(n)} \right)}_{\leq 0} \underbrace{\varphi \left(\frac{t}{n} \right)}_{\geq 0} + \frac{2x^2}{h^4(n)} \varphi'' \left(\frac{x^2}{h^2(n)} \right) \underbrace{\varphi \left(\frac{t}{n} \right)}_{\geq 0} \right) \right]. \end{aligned}$$

Considering the first term only, the infimum in (t, x) is attained for some $x \in [0, h(n)]$ (resp. $x \in [-h(n), -h(n)]$ if $S = \mathbb{R}$ or $x \in [0, \min\{h(n), 1\}]$ if $S = [0, 1]$) and $t \in [n, 2n]$. Then

$$\inf_n \inf_{(t,x)} \underbrace{\frac{1}{n} \varphi' \left(\frac{t}{n} \right)}_{\leq 0} \underbrace{\varphi \left(\frac{x^2}{h^2(n)} \right)}_{\geq 0} > -\infty,$$

and we are only left to study the second term. We first treat the cases $S = \mathbb{R}_{\geq 0}$ and $S = \mathbb{R}$ for which it follows that the infimum of the second term is attained for $t = n$ and $x = \alpha h(n)$ for some $\alpha \in [1, \sqrt{2}]$ (resp. $\alpha \in [-\sqrt{2}, -1] \cup [1, \sqrt{2}]$ if $S = \mathbb{R}$) fixed. Plugging this in and using (3.3.2), (3.3.3), we obtain

$$\begin{aligned} &\inf_n \left[\frac{2\alpha h(n) B(n, \alpha h(n))}{h^2(n)} \underbrace{\varphi(1)}_{\geq 0} \underbrace{\varphi'(\alpha^2)}_{\leq 0} + \frac{\sigma^2(n, \alpha h(n))}{h^2(n)} \underbrace{\varphi(1)}_{\geq 0} \left(\underbrace{\varphi'(\alpha^2)}_{\leq 0} + 2\alpha^2 \varphi''(\alpha^2) \right) \right] \\ &\geq \inf_n \left[\frac{2C^B \cdot (1 + h^2(n))}{h^2(n)} \underbrace{\varphi(1)}_{\geq 0} \underbrace{\varphi'(\alpha^2)}_{\leq 0} \right. \\ &\quad \left. + \frac{C^\sigma \cdot (1 + h^2(n))}{h^2(n)} \left(\underbrace{\varphi'(\alpha^2)}_{\leq 0} \underbrace{\varphi(1)}_{\geq 0} + 2\alpha^2 \varphi''(\alpha^2) \underbrace{\varphi(1)}_{\geq 0} \right) \right] \\ &> -\infty. \end{aligned}$$

For the case $S = [0, 1]$ the reasoning above is exactly the same, except $x = \min\{\alpha h(n), 1\}$ for some $\alpha \in [1, \sqrt{2}]$.

Secondly, the path-wise uniqueness follows by (Ethier and Kurtz, 2005, Theorem 5.3.8 and Remark 5.3.9). $\square$

3.4 Polynomial McKean-Vlasov equations

In this section, we study sufficient conditions for existence and uniqueness of solutions to one-dimensional McKean-Vlasov SDEs. In particular, we are concerned with studying McKean-Vlasov SDEs which are of the following form:

Definition 3.4.1 (Polynomial McKean-Vlasov SDEs). Let $(W_t)_{t \in [0, \infty)}$ be a one-dimensional Brownian motion. We call an SDE a *polynomial McKean-Vlasov SDE* if it is given by

$$dZ_t = \left(b(\mathbb{E}[Z_t], \dots, \mathbb{E}[Z_t^N]) + \beta(\mathbb{E}[Z_t], \dots, \mathbb{E}[Z_t^N])Z_t \right) dt + \sqrt{c(\mathbb{E}[Z_t], \dots, \mathbb{E}[Z_t^N]) + \gamma(\mathbb{E}[Z_t], \dots, \mathbb{E}[Z_t^N])Z_t + \Gamma(\mathbb{E}[Z_t], \dots, \mathbb{E}[Z_t^N])Z_t^2} dW_t \quad (3.4.1)$$

for some $N \geq 1$, continuous maps $b, \beta, c, \gamma, \Gamma : \mathbb{R}^N \rightarrow \mathbb{R}$ and Z_0 with $\mathbb{E}[Z_0], \dots, \mathbb{E}[Z_0^N] < \infty$.

For now the SDE in (3.4.1) is only formal, since we have not established the existence of solutions, which is the main goal of this section. We can distinguish between several state-spaces for $(Z_t)_{t \geq 0}$ which lead to some restrictions on the maps $b, \beta, c, \gamma, \Gamma$. One obvious condition is that the term under the square-root must be positive, as we want the solution to be real-valued. We discuss possible state spaces in more detail in Theorem 3.4.8, Remark 3.4.9 and Corollary 3.4.10. We would like to point out that certain SDEs that are contained in the class of 3.4.1 have already been studied in the 1980s, see Scheutzow (1985), in the context of diffusions with a period law. Note that the form of (3.4.1) strongly resembles the one of the SDE in Example 3.2.10. This resemblance is by no means a coincidence. In fact, we want to solve SDEs of the form of (3.4.1) using the theory of time-inhomogeneous polynomial processes. To do so, we can follow two approaches, a *primal* and a *dual* one, which we explore in detail in Section 3.4.1 and 3.4.2 respectively.

The primal approach is rather straightforward: we apply Itô's lemma to find a system of ODEs describing the moments $(\mathbb{E}[Z_t], \dots, \mathbb{E}[Z_t^N])_{t \geq 0}$. We then establish conditions under which we obtain a unique global solution to the corresponding system of ODEs and hence have a full description of the (deterministic) process $(\mathbb{E}[Z_t], \dots, \mathbb{E}[Z_t^N])_{t \geq 0}$. Plugging this process back into the original SDE (3.4.1), we establish existence and uniqueness of solutions to the SDE via a solution of the associated martingale problem and conclude that this solution is a time-inhomogeneous polynomial process.

The dual approach is slightly more involved: we start by considering a time-inhomogeneous polynomial process, depending on a deterministic auxiliary process. At first glance, this has nothing to do with the McKean-Vlasov SDEs. However, the auxiliary process is required to satisfy a system of ODEs, which are in fact associated with the Kolmogorov forward and backward equations of the process' transition operator. Using the moment formula, we can then conclude that the auxiliary process coincides with $(\mathbb{E}[Z_t], \dots, \mathbb{E}[Z_t^N])_{t \geq 0}$.

As a third consideration of this section, we extend the primal approach to a common-noise setting, where the moments are replaced by the conditional moments (conditional on the common noise). We present these results in Section 3.4.3.

3.4.1 Primal point of view

Recall the form of the polynomial McKean-Vlasov SDEs given by (3.4.1). To shorten the notation, we write $\mathbb{Z}_t := (Z_t, \dots, Z_t^N)$ and accordingly $\mathbb{E}[\mathbb{Z}_t] := (\mathbb{E}[Z_t], \dots, \mathbb{E}[Z_t^N])$. Using Itô's formula, it is then straightforward to derive for $k \in \{2, \dots, N\}$

$$dZ_t^k = \left[k \cdot (b(\mathbb{E}[\mathbb{Z}_t]) + \beta(\mathbb{E}[\mathbb{Z}_t])Z_t) Z_t^{k-1} \right. \quad (3.4.2)$$

$$\left. + \frac{k(k-1)}{2} (c(\mathbb{E}[\mathbb{Z}_t]) + \gamma(\mathbb{E}[\mathbb{Z}_t])Z_t + \Gamma(\mathbb{E}[\mathbb{Z}_t])Z_t^2) Z_t^{k-2} \right] dt \quad (3.4.3)$$

$$+ k \cdot \sqrt{c(\mathbb{E}[\mathbb{Z}_t]) + \gamma(\mathbb{E}[\mathbb{Z}_t])Z_t + \Gamma(\mathbb{E}[\mathbb{Z}_t])Z_t^2} Z_t^{k-1} dW_t. \quad (3.4.4)$$

We would then like to take the expectation and obtain a non-linear ODE for the moments. However, at this stage we do not know whether

$$M_t^{(k)} := \int_0^t \sqrt{c(\mathbb{E}[\mathbb{Z}_s]) + \gamma(\mathbb{E}[\mathbb{Z}_s])Z_s + \Gamma(\mathbb{E}[\mathbb{Z}_s])Z_s^2} Z_s^{k-1} dW_s$$

is a true martingale or just a local martingale. Note that $(M_t^{(k)})_{t \geq 0}$ is a true martingale if

$$\int_0^t \mathbb{E} \left[\left| (c(\mathbb{E}[\mathbb{Z}_s]) + \gamma(\mathbb{E}[\mathbb{Z}_s])Z_s + \Gamma(\mathbb{E}[\mathbb{Z}_s])Z_s^2) Z_s^{2k-2} \right| \right] ds < \infty \quad (3.4.5)$$

for $t \geq 0$. To establish this for all $k \in \{1, \dots, N\}$, we impose certain growth-conditions on the maps $b, \beta, c, \gamma, \Gamma$. We present two sets of assumptions on these maps under which we can establish the true martingale property and later the existence of a solution to (3.4.1).

Definition 3.4.2. Let us define $\|x\|_N := (\sum_{i=1}^N |x_i|^{\frac{2}{i}})^{\frac{1}{2}}$ for any $x \in \mathbb{R}^N$.

Assumptions A1. Let the following hold for all $x = (x_1, \dots, x_N)^T \in \mathbb{R}^N$

- (i) $\beta : \mathbb{R}^N \rightarrow [-\beta_0, \beta_0]$ and $\Gamma : \mathbb{R}^N \rightarrow [-\Gamma_0, \Gamma_0]$ for some $\beta_0, \Gamma_0 \geq 0$;
- (ii) there is a $c_0 \in \mathbb{R}$ and a continuous increasing function $f_c : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ such that $|c(x)| \leq c_0(1 + f_c(|x_1|) + \|x\|_N^2)$;
- (iii) there are $b_0, \gamma_0 \in \mathbb{R}$ and a continuous increasing function $f_\gamma : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ such that $|\gamma(x)| \leq \gamma_0(1 + f_\gamma(|x_1|) + \|x\|_N)$ and $|b(x)| \leq b_0(1 + |x_1|)$.

Assumptions A2. We assume that for all $x = (x_1, \dots, x_N)^T \in \mathbb{R}^N$

- (i) $\beta : \mathbb{R}^N \rightarrow [-\beta_0, \beta_0]$ and $\Gamma : \mathbb{R}^N \rightarrow [-\Gamma_0, \Gamma_0]$ for some $\beta_0, \Gamma_0 \geq 0$;
- (ii) there is a $c_0 \in \mathbb{R}$ such that $|c(x)| \leq c_0(1 + \|x\|_N^2)$;
- (iii) there are $b_0, \gamma_0 \in \mathbb{R}$ such that $|\gamma(x)| \leq \gamma_0(1 + \|x\|_N)$ and $|b(x)| \leq b_0(1 + \|x\|_N)$.

Lemma 3.4.3. Assuming a solution $(Z_t)_{t \geq 0}$ to (3.4.1) exists, if either Assumptions A1 or Assumptions A2 hold, then $(M_t)_{t \geq 0}$ is a true martingale.

Proof. We begin with the proof for the case of Assumption A1 and later turn to Assumption A2.

Our first step is to show that $|\mathbb{E}[Z_t]| \leq (|\mathbb{E}[Z_0]| + b_0 t) \exp((b_0 + \beta_0)t)$ for all $t \geq 0$. To do so, we can proceed similarly as in (Cuchiero et al., 2012, Proof of Theorem 2.10). Let us introduce a sequence of stopping times $(\tau_n)_{n \in \mathbb{N}}$ with $\tau_n = \inf\{t > 0 \mid |Z_t| \geq n\}$ and note that for any $n \in \mathbb{N}$

$$\begin{aligned} |\mathbb{E}[Z_{t \wedge \tau_n}]| &\leq |\mathbb{E}[Z_0]| + \left| \mathbb{E} \left[\int_0^{t \wedge \tau_n} b(\mathbb{E}[Z_s]) + \beta(\mathbb{E}[Z_s]) Z_s ds \right] \right| + |\mathbb{E}[M_{t \wedge \tau_n}^{(1)}]| \\ &\leq |\mathbb{E}[Z_0]| + \int_0^t |b(\mathbb{E}[Z_{s \wedge \tau_n}])| + |\beta(\mathbb{E}[Z_{s \wedge \tau_n}])| |\mathbb{E}[Z_{s \wedge \tau_n}]| ds \\ &\leq |\mathbb{E}[Z_0]| + \int_0^t b_0(1 + |\mathbb{E}[Z_{s \wedge \tau_n}]|) + \beta_0 |\mathbb{E}[Z_{s \wedge \tau_n}]| ds, \end{aligned}$$

where we used that $(M_{t \wedge \tau_n}^{(1)})_{t \geq 0}$ is a martingale. Using Gronwall's inequality, i.e. the fact that if

$$X_t \leq \alpha(t) + \int_0^t \beta(s) X_s ds \Rightarrow X_t \leq \alpha(t) \exp \left(\int_0^t \beta(s) ds \right)$$

for $\alpha : [0, \infty) \rightarrow \mathbb{R}$ continuous and increasing and $\beta : [0, \infty) \rightarrow [0, \infty)$, we find that

$$|\mathbb{E}[Z_{t \wedge \tau_n}]| \leq (|\mathbb{E}[Z_0]| + b_0 t) \exp((b_0 + \beta_0)t).$$

Since $(Z_t)_{t \geq 0}$ is continuous and adapted, it is locally bounded and $\lim_{n \rightarrow \infty} \tau_n = \infty$ and it thus follows by Fatou's lemma that for all $t \geq 0$

$$|\mathbb{E}[Z_t]| \leq \liminf_{n \rightarrow \infty} |\mathbb{E}[Z_{t \wedge \tau_n}]| \leq (|\mathbb{E}[Z_0]| + b_0 t) \exp((b_0 + \beta_0)t) =: \zeta(t),$$

where we abbreviate this upper bound by ζ for future convenience.

We now move on to the case $k \geq 2$ and aim to show that (3.4.5) holds. Note that for $2 \leq k \leq N$ we find

$$\begin{aligned} &\int_0^t \mathbb{E} \left[\left| (c(\mathbb{E}[Z_s]) + \gamma(\mathbb{E}[Z_s]) Z_s + \Gamma(\mathbb{E}[Z_s]) Z_s^2) Z_s^{2k-2} \right| \right] ds \\ &\leq \int_0^t |c(\mathbb{E}[Z_s])| \cdot \mathbb{E}[|Z_s|^{2k-2}] + |\gamma(\mathbb{E}[Z_s])| \cdot \mathbb{E}[|Z_s|^{2k-1}] + |\Gamma(\mathbb{E}[Z_s])| \cdot \mathbb{E}[|Z_s|^{2k}] ds \\ &\leq \int_0^t c_0(1 + \|\mathbb{E}[Z_s]\|_N^2 + f_c(|\mathbb{E}[Z_s]|)) \cdot \mathbb{E}[|Z_s|^{2k-2}] ds \\ &\quad + \int_0^t \gamma_0(1 + \|\mathbb{E}[Z_s]\|_N + f_\gamma(|\mathbb{E}[Z_s]|)) \cdot \mathbb{E}[|Z_s|^{2k-1}] + \Gamma_0 \cdot \mathbb{E}[|Z_s|^{2k}] ds \end{aligned}$$

Let us recall that $|\mathbb{E}[Z_t]| \leq \zeta(t)$ for all $t \geq 0$ and $\sup_{s \in [0, t]} f_c(\zeta(s)) = f_c(\zeta(t)) < \infty$ and that the same holds for f_γ by the definition of being a continuous increasing function. Moreover, let us note that for all $i \in \{1, \dots, N\}$

$$|\mathbb{E}[Z_t^i]|^{\frac{2}{i}} = |\mathbb{E}[Z_t^i]|^{\frac{2N}{iN}} \leq \mathbb{E}[|Z_t|^{2N}]^{\frac{1}{N}} \leq 1 + \mathbb{E}[|Z_t|^{2N}]$$

by Jensen's inequality and the fact that $x \leq 1 + x^N$ for all $x \in \mathbb{R}$. Hence,

$$\|\mathbb{E}[Z_t]\|_N \leq \sqrt{N + N\mathbb{E}[|Z_t|^{2N}]} \leq 1 + N + N\mathbb{E}[|Z_t|^{2N}].$$

Using the same reasoning, it holds for all $k \in \{1, \dots, N\}$ that

$$\mathbb{E}[|Z_t|^{2k-2}] \leq 1 + \mathbb{E}[|Z_t|^{2N}], \quad \mathbb{E}[|Z_t|^{2k-1}] \leq 1 + \mathbb{E}[|Z_t|^{2N}] \quad \text{and} \quad \mathbb{E}[|Z_t|^{2k}] \leq 1 + \mathbb{E}[|Z_t|^{2N}].$$

Hence it is sufficient to show that for all $t \geq 0$, $\sup_{s \in [0, t]} \mathbb{E}[|Z_s|^{2N}] < \infty$ in order for (3.4.5) to hold for all $k \in \{1, \dots, N\}$. In the following we will denote by $\lesssim$ an inequality up to a multiplicative constant and if we write e.g. $\lesssim_N$ we mean that this multiplicative constant depends on N . We apply the same reasoning as above to show that

$$|\mathbb{E}[Z_t^i]|^{\frac{2}{i}} = |\mathbb{E}[Z_t^i]|^{\frac{2N}{iN}} \leq \mathbb{E}[|Z_t|^N]^{\frac{2}{N}}$$

and hence,

$$\|\mathbb{E}[Z_t]\|_N \lesssim_N \mathbb{E}[|Z_t|^N]^{\frac{1}{N}}.$$

Again, we denote by $(\tau_n)_{n \in \mathbb{N}}$ the sequence of stopping times with $\tau_n := \inf\{t > 0 \mid |Z_t| \geq n\}$. We are now ready to show that

$$\begin{aligned} \mathbb{E}[|Z_{t \wedge \tau_n}|^{2N}] &\leq \mathbb{E}\left[\left|Z_0 + \int_0^{t \wedge \tau_n} (b(\mathbb{E}[Z_r]) + \beta(\mathbb{E}[Z_r])Z_r)dr \right. \right. \\ &\quad \left. \left. + \int_0^{t \wedge \tau_n} \sqrt{c(\mathbb{E}[Z_r]) + \gamma(\mathbb{E}[Z_r])Z_r + \Gamma(\mathbb{E}[Z_r])Z_r^2} dW_r \right|^{2N}\right] \\ &\lesssim_N \mathbb{E}[|Z_0|^{2N}] + t^{2N-1} \mathbb{E}\left[\int_0^t |b(\mathbb{E}[Z_{r \wedge \tau_n}]) + \beta(\mathbb{E}[Z_{r \wedge \tau_n}])Z_{r \wedge \tau_n}|^{2N} dr\right] \\ &\quad + \mathbb{E}\left[\int_0^t |c(\mathbb{E}[Z_{r \wedge \tau_n}]) + \gamma(\mathbb{E}[Z_{r \wedge \tau_n}])Z_{r \wedge \tau_n} + \Gamma(\mathbb{E}[Z_{r \wedge \tau_n}])Z_{r \wedge \tau_n}^2|^{2N} dr\right] \\ &\lesssim_N \mathbb{E}[|Z_0|^{2N}] + t^{2N-1} \int_0^t |b(\mathbb{E}[Z_{r \wedge \tau_n}])|^{2N} + (\beta_0)^{2N} \cdot \mathbb{E}[|Z_{r \wedge \tau_n}|^{2N}] dr \\ &\quad + \int_0^t |c(\mathbb{E}[Z_{r \wedge \tau_n}])|^N + |\gamma(\mathbb{E}[Z_{r \wedge \tau_n}])|^N \mathbb{E}[|Z_{r \wedge \tau_n}|^N] + (\Gamma_0)^N \mathbb{E}[|Z_{r \wedge \tau_n}|^{2N}] dr \\ &\lesssim_N \mathbb{E}[|Z_0|^{2N}] + t^{2N-1} \int_0^t (b_0)^{2N} (1 + |\mathbb{E}[Z_{r \wedge \tau_n}]|^{2N}) + (\beta_0)^{2N} \mathbb{E}[|Z_{r \wedge \tau_n}|^{2N}] dr \\ &\quad + \int_0^t \left((c_0)^N (1 + \mathbb{E}[|Z_{r \wedge \tau_n}|^N]^2 + f_c(|\mathbb{E}[Z_{r \wedge \tau_n}]|)^N) \right. \\ &\quad \left. + (\gamma_0)^N (1 + \mathbb{E}[|Z_r|^N] + f_\gamma(|\mathbb{E}[Z_{r \wedge \tau_n}]|)^N) \mathbb{E}[|Z_{r \wedge \tau_n}|^N] + (\Gamma_0)^N \mathbb{E}[|Z_{r \wedge \tau_n}|^{2N}] \right) dr \\ &\leq \mathbb{E}[|Z_0|^{2N}] + (b_0 t)^{2N} + \left((b_0)^{2N} + (\beta_0)^{2N} \right) t^{2N-1} \int_0^t \mathbb{E}[|Z_{r \wedge \tau_n}|^{2N}] dr \\ &\quad + (c_0)^N (1 + f_c(\zeta(t))^N) t + (\gamma_0)^N (1 + f_\gamma(\zeta(t))^N) t \\ &\quad + \int_0^t \left((c_0)^N + (\gamma_0)^N (2 + f_\gamma(\zeta(r))) + (\Gamma_0)^N \right) \mathbb{E}[|Z_{r \wedge \tau_n}|^{2N}] dr \end{aligned}$$

where we have used the BDG inequality to obtain the second inequality. Moreover, we used Jensen's inequality and the fact that $(a + b)^n \leq 2^{n-1}(|a|^n + |b|^n)$ several times. Again, using Gronwall's inequality, it follows that

$$\begin{aligned} \mathbb{E}[|Z_{t \wedge \tau_n}|^{2N}] &\leq C_N \left(\mathbb{E}[|Z_0|^{2N}] + (b_0 t)^{2N} + (c_0)^N (1 + f_c(\zeta(t))^N) t + (\gamma_0)^N (1 + f_\gamma(\zeta(t))^N) t \right) \\ &\quad \cdot \exp \left(C_N \left((b_0 + \beta_0)^{2N} t^{2N-1} + (c_0)^N + (\gamma_0)^N (2 + f_\gamma(\zeta(t))) + (\Gamma_0)^N \right) t \right) =: \zeta_1^{(2N)}(t) \end{aligned} \quad (3.4.6)$$

for some constant $C_N \in \mathbb{R}_{\geq 0}$. By Fatou's lemma, $\mathbb{E}[|Z_t|^{2N}] \leq \zeta_1^{(2N)}(t)$ and hence also $\sup_{s \in [0, t]} \mathbb{E}[|Z_s|^{2N}] \leq \zeta_1^{(2N)}(t)$. Therefore, we have shown that (3.4.5) holds for all $k \in \{1, \dots, N\}$ and $t \geq 0$. Note, for the above proof, we adapted the reasoning of the linear-growth conditions on coefficients of Itô-SDEs (see e.g. (Cohen and Elliot, 2015, Lemma 16.1.4)) and translated them to the maps $b, \beta, c, \gamma, \Gamma : \mathbb{R}^N \rightarrow \mathbb{R}$.

We have now established the case of the set of Assumptions A1. The reasoning is very similar to the previous case. In particular, it is enough to show that $\sup_{s \in [0, t]} \mathbb{E}[|Z_s|^{2N}] < \infty$ for all $t \in [0, \infty)$ for (3.4.5) to hold. To this end, we introduce the sequence of stopping times $(\tau_n)_{n \in \mathbb{N}}$ with $\tau_n := \inf\{t > 0 \mid |Z_t| \geq n\}$. Using similar arguments as before, we find

$$\begin{aligned} \mathbb{E}[|Z_{t \wedge \tau_n}|^{2N}] &\leq \mathbb{E} \left[\left| Z_0 + \int_0^{t \wedge \tau_n} (b(\mathbb{E}[Z_r]) + \beta(\mathbb{E}[Z_r]) Z_r) dr \right. \right. \\ &\quad \left. \left. + \int_0^{t \wedge \tau_n} \sqrt{c(\mathbb{E}[Z_r]) + \gamma(\mathbb{E}[Z_r]) Z_r + \Gamma(\mathbb{E}[Z_r]) Z_r^2} dW_r \right|^{2N} \right] \\ &\lesssim_N \mathbb{E}[|Z_0|^{2N}] + t^{2N-1} \int_0^t |b(\mathbb{E}[Z_{r \wedge \tau_n}])|^{2N} + (\beta_0)^{2N} \cdot \mathbb{E}[|Z_{r \wedge \tau_n}|^{2N}] dr \\ &\quad + \int_0^t |c(\mathbb{E}[Z_{r \wedge \tau_n}])|^N + |\gamma(\mathbb{E}[Z_{r \wedge \tau_n}])|^N \mathbb{E}[|Z_{r \wedge \tau_n}|^N] + (\Gamma_0)^N \mathbb{E}[|Z_{r \wedge \tau_n}|^{2N}] dr \\ &\lesssim_N \mathbb{E}[|Z_0|^{2N}] + t^{2N-1} \int_0^t (b_0)^{2N} (1 + \mathbb{E}[|Z_{r \wedge \tau_n}|^N]^2) + (\beta_0)^{2N} \mathbb{E}[|Z_{r \wedge \tau_n}|^{2N}] dr \\ &\quad + \int_0^t \left((c_0)^N (1 + \mathbb{E}[|Z_{r \wedge \tau_n}|^N]^2) + (\gamma_0)^N (1 + \mathbb{E}[|Z_r|^N]) \mathbb{E}[|Z_{r \wedge \tau_n}|^N] \right. \\ &\quad \left. + (\Gamma_0)^N \mathbb{E}[|Z_{r \wedge \tau_n}|^{2N}] \right) dr \\ &\leq \mathbb{E}[|Z_0|^{2N}] + (b_0 t)^{2N} + \left((b_0)^{2N} + (\beta_0)^{2N} \right) t^{2N-1} \int_0^t \mathbb{E}[|Z_{r \wedge \tau_n}|^{2N}] dr \\ &\quad + (c_0)^N t + (\gamma_0)^N t + \int_0^t \left((c_0)^N + (\gamma_0)^N + (\Gamma_0)^N \right) \mathbb{E}[|Z_{r \wedge \tau_n}|^{2N}] dr \end{aligned}$$

and hence, again by Gronwall's inequality, it holds

$$\begin{aligned} \mathbb{E}[|Z_{t \wedge \tau_n}|^{2N}] &\leq C_N \cdot \left(\mathbb{E}[|Z_0|^{2N}] + (b_0 t)^{2N} + (c_0)^N t + (\gamma_0)^N t \right) \\ &\cdot \exp \left(C'_N \cdot \left((b_0 + \beta_0)^{2N} t^{2N-1} + (c_0)^N + (\gamma_0)^N + (\Gamma_0)^N \right) t \right) =: \zeta_2^{(2N)}(t) \end{aligned} \quad (3.4.7)$$

for some constant $C'_N \in \mathbb{R}_{\geq 0}$. By Fatou's lemma, it holds $\mathbb{E}[|Z_t|^{2N}] \leq \zeta_2^{(2N)}(t)$ for all $t \in [0, \infty)$. Since this implies $\sup_{s \in [0, t]} \mathbb{E}[|Z_s|^{2N}] \leq \zeta_2^{(2N)}(t)$ for all $t \in [0, \infty)$ and we have shown the local martingale property also for the set of Assumptions A2. $\square$

Example 3.4.4. Let us present some examples for the maps $c, b, \gamma : \mathbb{R}^N \rightarrow \mathbb{R}$ which fulfill the set of Assumptions A1:

- (i) $c(x) = c_0 + \sum_{k=1}^N c_k |x_k|^{\frac{2}{k}} + c_{N+1} \exp(x_1)$ for $c_0, \dots, c_{N+1} \in \mathbb{R}$;
- (ii) $b(x) = b_0 + b_1 |x_1|$ for $b_0, b_1 \in \mathbb{R}$;
- (iii) $\gamma(x) = \gamma_0 + \sum_{k=1}^N \gamma_k |x_k|^{\frac{1}{k}} + \gamma_{N+1} \exp(x_1)$ for $\gamma_0, \dots, \gamma_{N+1} \in \mathbb{R}$.

Similarly, the following maps fulfill the set of Assumptions A2:

- (i) $c(x) = c_0 + \sum_{k=1}^N c_k |x_k|^{\frac{2}{k}}$ for $c_0, \dots, c_N \in \mathbb{R}$;
- (ii) $b(x) = b_0 + \sum_{k=1}^N b_k |x|^{\frac{1}{k}}$ for $b_0, \dots, b_N \in \mathbb{R}$;
- (iii) $\gamma(x) = \sum_{k=1}^N \gamma_k |x|^{\frac{1}{k}}$ for $\gamma_1, \dots, \gamma_N \in \mathbb{R}$.

Proposition 3.4.5. Under the Assumptions A1 or Assumptions A2 on the maps $b, \beta, c, \gamma, \Gamma : \mathbb{R}^N \rightarrow \mathbb{R}$ and denoting $\bar{Z}_t := (1, Z_t, \dots, Z_t^N)^\top$, the process $(\mathbb{E}[\bar{Z}_t])_{t \geq 0}$ satisfies the following $N + 1$ dimensional non-linear ODE:

$$\bar{z}'(t) = \mathbf{L}(z(t)) \bar{z}(t) \text{ and } \bar{z}(0) = \mathbb{E}[\bar{Z}_0] \quad (3.4.8)$$

with $\bar{z}(t) = (1, z(t))$ for $z(t) \in \mathbb{R}^N$ such that $\bar{z}(t)_0 \equiv 1$, $\bar{z}(t)_i = z(t)_i$ for $i \in \{1, \dots, N\}$ and where for $j, k = 0, \dots, N$

$$\mathbf{L}(z(t))_{kj} = \begin{cases} k \cdot \beta(z(t)) + \frac{k(k-1)}{2} \Gamma(z(t)) & \text{if } k = j, k \neq 0 \\ k \cdot b(z(t)) + \frac{k(k-1)}{2} \gamma(z(t)) & \text{if } k = j + 1 \\ \frac{k(k-1)}{2} c(z(t)) & \text{if } k = j + 2 \\ 0 & \text{else.} \end{cases}$$

Moreover, for a fixed initial condition $z(0) \in \mathbb{R}^N$, the ODE given in (3.4.8) has a unique solution (up to its maximal time of existence) if $b, \beta, c, \gamma, \Gamma$ are locally Lipschitz on an open set $A \subseteq \mathbb{R}^N$ with $z(0) \in A$, in short we write $b, \beta, c, \gamma, \Gamma \in \text{Lip}(A, \mathbb{R})$ for the local Lipschitz continuity requirement.

Before we proceed to the proof, let us note for the statement regarding existence and uniqueness of solutions, we can start from any $z(0) \in \mathbb{R}^N$. However, for this solution to describe the process $(\mathbb{E}[Z_t])_{t \geq 0}$, we can only consider starting points of the form $z(0) = (z_0, z_0^2, \dots, z_0^N)$ for some $z_0 \in \mathbb{R}$.

Proof of Proposition 3.4.5. By Lemma 3.4.3 the process $(M_t^{(k)})_{t \geq 0}$ is a martingale for all $k \in \{1, \dots, N\}$ and hence the form of (3.4.8) follows by an application of Itô's lemma. Moreover, the existence and uniqueness of solutions follows from Theorem 3.A.1. $\square$

Proposition 3.4.6. Note that, if $b, \beta, c, \gamma, \Gamma$ are locally Lipschitz on $\mathbb{R}^N$ and either satisfy the Assumptions A1 or the Assumptions A2, then the solution to (3.4.8) is necessarily global.

Proof. The existence of a unique solution up to its maximal lifetime is assured by the Lipschitz condition. To see the global existence for $A = \mathbb{R}^{N+1}$, we apply Gronwall's inequality. Clearly, the zeroth component is constant, i.e. $\bar{z}_0(t) \equiv \bar{z}_0(0) = 1$.

Let us first treat the set of Assumptions A1 to prove the statement, and later adapt the proof for the set of Assumptions A2. To this end, we start by studying the first component and denote by $(t_n)_{n \in \mathbb{N}}$ a (deterministic) sequence of time-points such that $t_n := \inf\{t > 0 \mid \|z(t)\|_N^{2N} \geq n\}$. Recall that we have already established that a solution exists at least on some interval $[0, t^*)$ where $t^* = \inf\{t > 0 \mid \|z(t)\|_N^{2N} = \infty\}$. For global existence, we would like to show that $t^* = \infty$, for which a sufficient condition is that $\lim_{n \rightarrow \infty} t_n \rightarrow \infty$. Then, considering the first component, we obtain

$$\begin{aligned} \sup_{s \in [0, t \wedge t_n]} |z_1(s)| &= \sup_{s \in [0, t \wedge t_n]} \left| z_1(0) + \int_0^s \beta(z(r)) z_1(r) + b(z(r)) dr \right| \\ &= |Z_0| + \int_0^{t \wedge t_n} |\beta(z(r))| |z_1(r)| + |b(z(r))| dr \\ &\leq |z_1(0)| + \int_0^{t \wedge t_n} \beta_0 |z_1(r \wedge t_n)| + b_0 + K_b |z_1(r \wedge t_n)| dr \\ &\leq |Z_0| + b_0(t \wedge t_n) + \int_0^{t \wedge t_n} (\beta_0 + K_b) \sup_{u \in [0, r \wedge t_n]} |z_1(u)| dr \end{aligned}$$

and hence by Gronwall's inequality, we find

$$\sup_{s \in [0, t \wedge t_n]} |z_1(s)| \leq (|z_1(0)| + b_0(t \wedge t_n)) \cdot \exp((\beta_0 + K_b)(t \wedge t_n)) =: h_1(t \wedge t_n).$$

We will not treat the following components one-by-one, but instead study $\|z(t)\|_N$. Note that the first component plays a special role in this set of assumptions as it only depends on β, b which themselves are bounded by the first component, while the other maps' bounds depend on all components. Clearly, if $\|z(t)\|_N^{2N} < \infty$ for all $t \in [0, \infty)$ this implies $|Z_1(t)|, |z_2(t)|, \dots, |z_N(t)| < \infty$ for all $t \in [0, \infty)$. In preparation to that, let us note that for all $k \in \{1, \dots, N\}$ and $x, y \in \mathbb{R}$ it holds that $|x|^k + |y|^k \leq (|x| + |y|)^k$. Hence, it also holds

that for all $k \in \{1, \dots, N\}$ and $t \geq 0$ that

$$|z_k(t)|^2 = \left(|z_k(t)|^{\frac{2}{k}}\right)^k \leq \left(|z_k(t)|^{\frac{2}{k}}\right)^k + \left(\sum_{\substack{j=1 \\ j \neq k}}^N |z_j(t)|^{\frac{2}{j}}\right)^k \leq \left(\sum_{j=1}^N |z_j(t)|^{\frac{2}{j}}\right)^k = \left(\|z(t)\|_N\right)^{2k}.$$

We are now ready to study $\|z(t \wedge t_n)\|_N^{2N}$.

$$\begin{aligned} \|z(t \wedge t_n)\|_N^{2N} &\lesssim_N \sum_{j=1}^N |z_j(t)|^{\frac{2N}{j}} \\ &\leq \sum_{j=1}^N \left| z_j(0) + \int_0^{t \wedge t_n} \left(j \cdot \beta(z(s)) + \frac{j(j-1)}{2} \Gamma(z(s)) \right) \bar{z}_j(s) ds \right. \\ &\quad + \int_0^{t \wedge t_n} \left(j \cdot b(z(s)) + \frac{j(j-1)}{2} \gamma(z(s)) \right) \bar{z}_{j-1}(s) ds \\ &\quad \left. + \mathbf{1}_{\{j>1\}} \int_0^{t \wedge t_n} \frac{j(j-1)}{2} c(z(s)) \bar{z}_{j-2}(s) ds \right|^{\frac{2N}{j}} \\ &\lesssim_N \sum_{j=1}^N |z_j(0)|^{\frac{2N}{j}} + (t \wedge t_n)^{\frac{2N}{j}-1} \int_0^{t \wedge t_n} \left(|\beta(z(s))|^{\frac{2N}{j}} + \mathbf{1}_{\{j>1\}} |\Gamma(z(s))|^{\frac{2N}{j}} \right) |z_j(s)|^{\frac{2N}{j}} ds \\ &\quad + (t \wedge t_n)^{\frac{2N}{j}-1} \int_0^{t \wedge t_n} \left(|b(z(s))|^{\frac{2N}{j}} + \mathbf{1}_{\{j>1\}} |\gamma(z(s))|^{\frac{2N}{j}} \right) |\bar{z}_{j-1}(s)|^{\frac{2N}{j}} ds \\ &\quad + \mathbf{1}_{\{j>1\}} (t \wedge t_n)^{\frac{2N}{j}-1} \int_0^{t \wedge t_n} |c(z(s))|^{\frac{2N}{j}} |\bar{z}_{j-2}(s)|^{\frac{2N}{j}} ds \\ &\lesssim_N \|z(0)\|_N^{2N} + (1 + \beta_0^{2N} + \Gamma_0^{2N})(1 + t \wedge t_n)^{2N} \int_0^{t \wedge t_n} \|z(s)\|_N^{2N} ds \\ &\quad + (b_0(1 + h_1(t)))^{2N} \sum_{j=1}^N (t \wedge t_n)^{\frac{2N}{j}-1} \int_0^{t \wedge t_n} |\bar{z}_{j-1}(s)|^{\frac{2N}{j}} ds \\ &\quad + \sum_{j=2}^N (t \wedge t_n)^{\frac{2N}{j}-1} \int_0^{t \wedge t_n} (\gamma_0(1 + f_\gamma(|z_1(s)|) + \|z(s)\|_N)) \frac{2N}{j} |\bar{z}_{j-1}(s)|^{\frac{2N}{j}} ds \\ &\quad + \sum_{j=2}^N (t \wedge t_n)^{\frac{2N}{j}-1} \int_0^{t \wedge t_n} (c_0(1 + f_c(|z_1(s)|) + \|z(s)\|_N^2)) \frac{2N}{j} |\bar{z}_{j-2}(s)|^{\frac{2N}{j}} ds \\ &\lesssim_N \|z(0)\|_N^{2N} + (1 + \beta_0^{2N} + \Gamma_0^{2N})(1 + t \wedge t_n)^{2N} \int_0^{t \wedge t_n} \|z(s)\|_N^{2N} ds \\ &\quad + (b_0(1 + h_1(t \wedge t_n)))^{2N} \sum_{j=1}^N (t \wedge t_n)^{\frac{2N}{j}-1} \int_0^{t \wedge t_n} \left(1 + \|z(s)\|_N^j \right) \frac{2N}{j} ds \\ &\quad + \sum_{j=2}^N (t \wedge t_n)^{\frac{2N}{j}-1} \int_0^{t \wedge t_n} (\gamma_0(1 + f_\gamma(|z_1(s)|) + \|z(s)\|_N)) \frac{2N}{j} (1 + \|z(s)\|_N^{j-1}) \frac{2N}{j} ds \end{aligned}$$

$$\begin{aligned}
 & + \sum_{j=2}^N (t \wedge t_n)^{\frac{2N}{j}-1} \int_0^{t \wedge t_n} (c_0(1 + f_c(|z_1(s)|) + \|z(s)\|_N^{2N})^{\frac{2N}{j}} (1 + \|z(s)\|_N^{j-2})^{\frac{2N}{j}} ds \\
 & \lesssim_N \|z(0)\|_N^{2N} + (1 + \beta_0^{2N} + \Gamma_0^{2N})(1 + t \wedge t_n)^{2N} \int_0^{t \wedge t_n} \|z(s)\|_N^{2N} ds \\
 & + (1 + t \wedge t_n)^{2N} \left[(b_0 + \gamma_0 + c_0) \left(1 + h_1(t \wedge t_n) + f_\gamma(h_1(t \wedge t_n)) + f_c(h_1(t \wedge t_n)) \right) \right]^{2N} \\
 & \cdot \int_0^{t \wedge t_n} (2 + \|z(s)\|_N^{2N}) ds \\
 & + \sum_{j=2}^N \left[(1 + \gamma_0^{2N})(t \wedge t_n)^{\frac{2N}{j}-1} \int_0^{t \wedge t_n} (\|z(s)\|_N^{\frac{2N}{j}} + \|z(s)\|_N^{2N}) ds \right. \\
 & \quad \left. + (1 + c_0^{2N})(t \wedge t_n)^{\frac{2N}{j}-1} \int_0^{t \wedge t_n} (\|z(s)\|_N^{\frac{4N}{j}} + \|z(s)\|_N^{2N}) ds \right] \\
 & \lesssim_N \|z(0)\|_N^{2N} + \left[(1 + \beta_0 + \Gamma_0 + b_0 + \gamma_0 + c_0) \right. \\
 & \quad \cdot \left. \left(1 + h_1(t \wedge t_n) + f_\gamma(h_1(t \wedge t_n)) + f_c(h_1(t \wedge t_n)) \right) (1 + t \wedge t_n) \right]^{2N} \\
 & \quad \cdot \left[t + \int_0^t \|z(s \wedge t_n)\|_N^{2N} ds \right]
 \end{aligned}$$

Hence, by an application of Gronwall's inequality, we can show that $\|z(t \wedge t_n)\|_N^{2N} < h_{2N}(t \wedge t_n)$ for all $n \in \mathbb{N}$ and $t \geq 0$ and a monotonically increasing continuous function $h_{2N} : [0, \infty) \rightarrow [0, \infty)$. Now, let's assume $\lim_{n \rightarrow \infty} t_n = t^* < \infty$. However,

$$\lim_{n \rightarrow \infty} \|z(t \wedge t_n)\|_N^{2N} \leq \lim_{n \rightarrow \infty} h_{2N}(t \wedge t_n) = h_{2N}(t^*) < \infty$$

which is a contradiction because by definition $\lim_{n \rightarrow \infty} \|z(t \wedge t_n)\|_N^{2N} \rightarrow \infty$. Hence, we have shown that indeed a global solution exists under the set of Assumptions A1.

The proof using the set of Assumptions A2 is structurally very similar. Since the first component does not play a special role in this set of assumptions, we can start directly by treating $\|z(t)\|_N^{2N}$, where we introduce again the sequence $(t_n)_{n \in \mathbb{N}}$ with

$$t_n := \inf\{t > 0 \mid \|z(t)\|_N^{2N} \geq n\}.$$

Using the same arguments as before, we can show

$$\begin{aligned}
 \|z(t \wedge t_n)\|_N^{2N} & \lesssim_N \sum_{j=1}^N |z_j(t)|^{\frac{2N}{j}} \\
 & \lesssim_N \sum_{j=1}^N |z_j(0)|^{\frac{2N}{j}} + (t \wedge t_n)^{\frac{2N}{j}-1} \int_0^{t \wedge t_n} \left(|\beta(z(s))|^{\frac{2N}{j}} + \mathbf{1}_{\{j>1\}} |\Gamma(z(s))|^{\frac{2N}{j}} \right) |z_j(s)|^{\frac{2N}{j}} ds \\
 & + (t \wedge t_n)^{\frac{2N}{j}-1} \int_0^{t \wedge t_n} \left(|b(z(s))|^{\frac{2N}{j}} + \mathbf{1}_{\{j>1\}} |\gamma(z(s))|^{\frac{2N}{j}} \right) |\bar{z}_{j-1}(s)|^{\frac{2N}{j}} ds \\
 & + \mathbf{1}_{\{j>1\}} (t \wedge t_n)^{\frac{2N}{j}-1} \int_0^{t \wedge t_n} |c(z(s))|^{\frac{2N}{j}} |\bar{z}_{j-2}(s)|^{\frac{2N}{j}} ds
 \end{aligned}$$

$$\begin{aligned}
& \lesssim_N \|z(0)\|_N^{2N} + (1 + \beta_0^{2N} + \Gamma_0^{2N})(1 + t \wedge t_n)^{2N} \int_0^{t \wedge t_n} \|z(s)\|_N^{2N} ds \\
& + \sum_{j=1}^N (t \wedge t_n)^{\frac{2N}{j}-1} \int_0^{t \wedge t_n} (b_0(1 + \|z(s)\|_N))^{\frac{2N}{j}} |\bar{z}_{j-1}(s)|^{\frac{2N}{j}} ds \\
& + \sum_{j=2}^N (t \wedge t_n)^{\frac{2N}{j}-1} \int_0^{t \wedge t_n} (\gamma_0(1 + \|z(s)\|_N))^{\frac{2N}{j}} |\bar{z}_{j-1}(s)|^{\frac{2N}{j}} ds \\
& + \sum_{j=2}^N (t \wedge t_n)^{\frac{2N}{j}-1} \int_0^{t \wedge t_n} (c_0(1 + \|z(s)\|_N^2))^{\frac{2N}{j}} |\bar{z}_{j-2}(s)|^{\frac{2N}{j}} ds \\
& \lesssim_N \|z(0)\|_N^{2N} + (1 + \beta_0^{2N} + \Gamma_0^{2N})(1 + t \wedge t_n)^{2N} \int_0^{t \wedge t_n} \|z(s)\|_N^{2N} ds \\
& + \sum_{j=2}^N (t \wedge t_n)^{\frac{2N}{j}-1} \int_0^{t \wedge t_n} ((b_0 + \gamma_0)(1 + \|z(s)\|_N))^{\frac{2N}{j}} (1 + \|z(s)\|_N^{j-1})^{\frac{2N}{j}} ds \\
& + \sum_{j=2}^N (t \wedge t_n)^{\frac{2N}{j}-1} \int_0^{t \wedge t_n} (c_0(1 + \|z(s)\|_N^2))^{\frac{2N}{j}} (1 + \|z(s)\|_N^{j-2})^{\frac{2N}{j}} ds \\
& \lesssim_N \|z(0)\|_N^{2N} + (1 + \beta_0^{2N} + \Gamma_0^{2N})(1 + t \wedge t_n)^{2N} \int_0^{t \wedge t_n} \|z(s)\|_N^{2N} ds \\
& + (1 + t \wedge t_n)^{2N} (b_0 + \gamma_0 + c_0)^{2N} \int_0^{t \wedge t_n} (1 + \|z(s)\|_N^{2N}) ds \\
& + \sum_{j=2}^N \left[(1 + \gamma_0^{2N} + b_0^{2N})(t \wedge t_n)^{\frac{2N}{j}-1} \int_0^{t \wedge t_n} (\|z(s)\|_N^{\frac{2N}{j}} + \|z(s)\|_N^{2N}) ds \right. \\
& \quad \left. + (1 + c_0^{2N})(t \wedge t_n)^{\frac{2N}{j}-1} \int_0^{t \wedge t_n} (\|z(s)\|_N^{\frac{4N}{j}} + \|z(s)\|_N^{2N}) ds \right] \\
& \lesssim_N \|z(0)\|_N^{2N} + \left[(1 + \beta_0 + \Gamma_0 + b_0 + \gamma_0 + c_0)(1 + t \wedge t_n) \right]^{2N} \left[t + \int_0^t \|z(s \wedge t_n)\|_N^{2N} ds \right]
\end{aligned}$$

By Gronwall's inequality, we find that $\|z(t \wedge t_n)\|_N^{2N} < g_{2N}(t \wedge t_n)$ for some monotonically increasing continuous function $g_{2N} : [0, \infty) \rightarrow [0, \infty)$. The existence of a global solution follows by the same arguments as before. $\square$

Note that the maps given in Example 3.4.4 are not locally Lipschitz on all of $\mathbb{R}^N$ which is required in Proposition 3.4.6 to get a global solution to (3.4.8). However, the following corollary nevertheless allows us to this under some different conditions that are satisfied by the maps in Example 3.4.4.

Corollary 3.4.7. *Let $b, c, \gamma : \mathbb{R}^N \rightarrow \mathbb{R}_{\geq 0}$ satisfy the conditions of Assumptions A1 or Assumptions A2. If $z(0) > 0$, additionally $b, c, \gamma \in \text{Lip}(\mathbb{R}_{>0}^N; \mathbb{R}_{\geq 0})$ and β, Γ are constant, then the solution to (3.4.8) is global.*

Proof. Let us first note that by assumption $z(0) > 0$. Moreover, it holds that

$$\begin{aligned} z_1(t) &= z_1(0) + \int_0^t b(z(s))ds + \int_0^t \beta \cdot z_1(s)ds \\ &\geq z_1(0) + \int_0^t \beta \cdot z_1(s)ds, \end{aligned}$$

where we used the fact that $b(\cdot)$ is non-negative. Hence, we can use Gronwall's inequality for

$$-z_1(t) \leq -z_1(0) + \int_0^t \beta \cdot (-z_1(s))ds$$

and thereby obtain

$$-z_1(t) \leq -z_1(0) \cdot \exp(\beta t) \Leftrightarrow z_1(t) \geq z_1(0) \cdot \exp(\beta t).$$

Therefore, we have shown that if $z_1(0) > 0$ the first component stays positive for all $t \geq 0$. Similarly, this holds for higher components. We show this by induction and therefore assume it holds up to some $1 \leq k \leq N - 1$ that $(z_i(t))_{1 \leq i \leq k} > 0$ for all $t \geq 0$, then it holds

$$\begin{aligned} z_{k+1}(t) &= z_{k+1}(0) + \int_0^t kb(z(s))z_k(s) + \frac{k(k-1)}{2}(c(z(s))z_{k-1}(s) + \gamma(z(s))z_k(s))ds \\ &\quad + \int_0^t \left(k\beta + \frac{k(k-1)}{2}\Gamma\right) z_{k+1}(s)ds \\ &\geq z_{k+1}(0) + \int_0^t \left(k\beta + \frac{k(k-1)}{2}\Gamma\right) z_{k+1}(s)ds. \end{aligned}$$

Again, using Gronwall's inequality, we obtain

$$z_{k+1}(t) \geq z_{k+1}(0) \exp\left(t\left(\frac{k(k-1)}{2}\Gamma + k\beta\right)\right).$$

□

Theorem 3.4.8. *Let us consider a McKean-Vlasov SDE of the form*

$$dZ_t = (b(\mathbb{E}[Z_t]) + \beta(\mathbb{E}[Z_t])Z_t)dt + \sqrt{c(\mathbb{E}[Z_t]) + \gamma(\mathbb{E}[Z_t])Z_t + \Gamma(\mathbb{E}[Z_t])Z_t^2}dW_t \quad (3.4.9)$$

for some $N \geq 1$, fixed $\mathbb{E}[Z_0] \in \mathbb{R}^N$, continuous maps $b, \beta, c, \gamma, \Gamma : \mathbb{R}^N \rightarrow \mathbb{R}$ and a Brownian motion $(W_t)_{t \geq 0}$. If $b, \beta, c, \gamma, \Gamma$ are locally Lipschitz continuous on $\mathbb{R}^N$, satisfy either the set of Assumptions A1 or Assumptions A2, and the term under the square-root is non-negative, then there exists a unique strong solution $(Z_t)_{t \geq 0}$. Moreover, $(Z_t)_{t \geq 0}$ is a time-inhomogeneous polynomial process.

Proof. Let $(z(t))_{t \geq 0}$ be the unique global solution to (3.4.8) with $z(0) = \mathbb{E}[Z_0]$. Note that by the assumptions on $b, \beta, c, \gamma, \Gamma$ in the theorem such a global solution exists by Proposition 3.4.6. Let us now study the following SDE:

$$dZ_t = (b(z(t)) + \beta(z(t))Z_t)dt + \sqrt{c(z(t)) + \gamma(z(t))Z_t + \Gamma(z(t))Z_t^2}dW_t. \quad (3.4.10)$$

We now show that a solution to the martingale problem associated with (3.4.10) exists, by the assumptions on $b, \beta, c, \gamma, \Gamma$. Let us elaborate on this point for the two cases of Assumptions A1 and Assumptions A2 separately.

Under the set of Assumptions A1 we have established in the proof of Proposition 3.4.6 that for all $t \in [0, \infty)$ it holds that

$$\sup_{s \in [0, t]} |z_1(s)| \leq h_1(t) \text{ and } \|z(t)\|_N^{2N} \leq h_{2N}(t).$$

This allows us to verify the growth requirements stated in Theorem 3.3.4 for (3.4.10):

$$\begin{aligned} \sigma^2(t, x) &= c(z(t)) + \gamma(z(t))x + \Gamma(z(t))x^2 \\ &\leq c_0(1 + f_c(h_1(t)) + h_{2N}(t)) + \gamma_0(1 + f_\gamma(h_1(t)) + h_{2N}(t))x + \Gamma_0x^2 \\ xB(t, x) &= b(z(t))x + \beta(z(t))x^2 \\ &\leq b_0(1 + h_1(t))x + \beta_0x^2 \end{aligned}$$

and indeed the growth conditions (3.3.3), (3.3.2) are satisfied for $h(t) = f_c(h_1(t)) + h_{2N}(t) + h_1(t)$, $K^\sigma = c_0 + \gamma_0 + \Gamma_0$ and $K^B = b_0 + \beta_0$.

The arguments in the case of Assumptions A2 are very similar, recalling we have established $\|z(t)\|_N^{2N} \leq g_{2N}(t)$ in the proof of Proposition 3.4.6 and hence for (3.4.10) it holds that

$$\begin{aligned} \sigma^2(t, x) &= c(z(t)) + \gamma(z(t))x + \Gamma(z(t))x^2 \\ &\leq c_0(1 + g_{2N}(t)) + \gamma_0(1 + g_{2N}(t))x + \Gamma_0x^2 \\ xB(t, x) &= b(z(t))x + \beta(z(t))x^2 \\ &\leq b_0(1 + g_{2N}(t))x + \beta_0x^2. \end{aligned}$$

In this case, the growth conditions (3.3.3), (3.3.2) are satisfied for $h(t) = g_{2N}(t)$, $K^\sigma = c_0 + \gamma_0 + \Gamma_0$ and $K^B = b_0 + \beta_0$. The assumption that c, γ, Γ are such that the term under the square-root is non-negative ensures that the PMP holds. Therefore, there exists a solution to the martingale problem associated with (3.4.10). Moreover, $\sigma^2(t, x), B(t, x)$ fulfil the local Lipschitz continuity requirement of Theorem 3.3.4. Therefore, Theorem 3.3.4 is applicable to (3.4.10) and we conclude that (3.4.10) admits a unique strong solution.

Applying Itô's lemma to the solution $(Z_t)_{t \geq 0}$ of (3.4.10) and noting that the local martingales are true martingales due to the growth-bounds on $b, \beta, c, \gamma, \Gamma$, it is apparent that $\mathbb{E}[Z_t]$ is a solution to (3.4.8). On the other hand, $(z(t))_{t \geq 0}$ is the unique global solution to (3.4.8) and therefore $\mathbb{E}[Z_t] = z(t)$ for all $t \in [0, \infty)$. Hence, we can conclude that (3.4.1) and (3.4.10) coincide, and we have found a unique strong solution to (3.4.9). Moreover, the solution of (3.4.9) clearly satisfies the definition of a time-inhomogeneous polynomial process, see Definition 3.2.8. $\square$

Remark 3.4.9. Concrete sufficient conditions on the maps c, γ, Γ for the term under the square-root to be positive are:

- $c \geq 0$, $\Gamma \geq 0$ and $\gamma \equiv 0$, or,

- the term under the square-root is of the form $c(\cdot) + \gamma(\cdot)z + \Gamma(\cdot)z^2 = (\tilde{\gamma}(\cdot) + \tilde{\Gamma}(\cdot)z)$.

In addition to the general case of Theorem 3.4.8 where the state space is $\mathbb{R}$, we can study the state-spaces $\mathbb{R}_{\geq 0}$ and $[0, 1]$.

Corollary 3.4.10. *Let $S = \mathbb{R}_{\geq 0}$ or $S = [0, T]$ and $\mathbb{E}[Z_0] \in S$. Under the set of Assumptions A1 or Assumptions A2 and if $b, \beta, c, \gamma, \Gamma$ are locally Lipschitz continuous on $\mathbb{R}_{\geq 0}$, sufficient conditions for existence of a strong solution to (3.4.9) on the respective state spaces are:*

- $S = \mathbb{R}_{\geq 0}$: in this case a sufficient condition is $b : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ and $c \equiv 0$, $\gamma, \Gamma : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$.
- $S = [0, 1]$: here sufficient conditions are $b : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$, $b + \beta : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\leq 0}$, and $c \equiv 0$, $\gamma \equiv -\Gamma \geq 0$.

In each case, the solution to (3.4.9) is a time-inhomogeneous polynomial process on the respective state-space S .

Proof. Note that by Assumptions A1 and A2 respectively, the growth-conditions of the SDE's coefficients are unchanged to the case of Theorem 3.4.8. The only conditions that differ in this situation are the ones relating to the PMP. Indeed, additional conditions above correspond to the requirements given in Proposition 3.3.3 for the PMP to hold. Then, existence of a unique strong solution follows directly by Theorem 3.3.4. $\square$

Corollary 3.4.11 (Moment Formula). *Let $(Z_t)_{t \geq 0}$ be the solution of (3.4.9) and the assumptions of Theorem 3.4.8 be satisfied. Recall the notation $\bar{Z}_t = (1, Z_t, Z_t^2, \dots, Z_t^N)^\top$ for all $t \geq 0$. Note that for $u = (u_0, \dots, u_N)^\top \in \mathbb{R}^{N+1}$ we can denote polynomials in Z_t by $\langle \bar{Z}_t, u \rangle = u_0 + u_1 Z_t + \dots + u_N Z_t^N$. Then, the following moment formula holds on any arbitrary but fixed time horizon $[0, T]$:*

$$\mathbb{E}[\langle \bar{Z}_t, u \rangle | \mathcal{F}_s] = (1, Z_s, Z_s^2, \dots, Z_s^N) \cdot e^{\Omega(s, t)} \cdot u,$$

any $0 \leq s \leq t \leq T$, any $u \in \mathbb{R}^{N+1}$ and where $\Omega(s, t) = \sum_{k=0}^{\infty} \Omega_k(s, t)$ is the Magnus expansion $\mathcal{E}(t)$ with $\mathcal{E}(t) = \varepsilon \cdot \mathbf{L}(\mathbb{E}[\bar{Z}]_t)^\top$ for $\varepsilon > 0$ such that $\int_0^T \|\mathcal{E}(s)\|_2 ds < \pi$.

Proof. See Theorem 3.2.12. To find the infinitesimal generators $(\mathbf{H}_s)_{s \geq 0}$ of the process and show that $\mathbf{H}_s = \mathbf{L}(\mathbb{E}[\bar{Z}]_s)^\top$ for all $s \geq 0$ we apply Proposition 3.2.9 to elements of a polynomial basis. In particular, let us consider the basis $(v_0, \dots, v_N)$ with $v_i := (x \mapsto x^i)$ for all $i \in \{0, \dots, N\}$. Then, applying Proposition 3.2.9 we find for $k \geq 2$ that

$$\begin{aligned} (\mathbf{H}_s v_k)(x) = & v_{k-2}(x) \cdot \left(\frac{k(k-1)}{2} c(\mathbb{E}[Z_t]) \right) + v_{k-1}(x) \cdot \left(k \cdot b(\mathbb{E}[Z_t]) + \frac{k(k-1)}{2} \gamma(\mathbb{E}[Z_t]) \right) \\ & + v_k(x) \left(k \cdot \beta(\mathbb{E}[Z_t]) + \frac{k(k-1)}{2} \Gamma(\mathbb{E}[Z_t]) \right). \end{aligned}$$

Recalling the convention we introduced in (3.2.1), it follows that $(\mathbf{H}_s)_{jk} = \mathbf{L}(\mathbb{E}[Z_s])_{kj}$ for all $k, j \in \{0, \dots, N\}$. Having found the infinitesimal generators, the rest of the claim follows directly from and application of Theorem 3.2.12. $\square$

Remark 3.4.12. Note that the above expression for $\mathbb{E}[\langle \bar{Z}_t, u \rangle | \mathcal{F}_s]$ can also be obtained by solving Equation 3.4.8 with the initial value given by $\bar{z}(0) = (1, Z_s(\omega), \dots, Z_s(\omega)^N)^\top$.

3.4.2 Dual point of view

We now approach polynomial McKean-Vlasov SDEs using a dual approach. Recall our notation $\bar{Z}_t = (1, Z_t, Z_t^2, \dots, Z_t^N)^\top$ for all $t \geq 0$. We consider a process $(Z_t)_{t \geq 0}$ given by SDEs of the form

$$\begin{aligned} dZ_t = & \left(\tilde{b}(\langle \bar{Z}_0, c(t, 0, e_0) \rangle, \dots, \langle \bar{Z}_0, c(t, 0, e_N) \rangle) + \tilde{\beta}(\langle \bar{Z}_0, c(t, 0, e_0) \rangle, \dots, \langle \bar{Z}_0, c(t, 0, e_N) \rangle) Z_t \right) dt \\ & + \left(\tilde{c}(\langle \bar{Z}_0, c(t, 0, e_0) \rangle, \dots, \langle \bar{Z}_0, c(t, 0, e_N) \rangle) + \tilde{\gamma}(\langle \bar{Z}_0, c(t, 0, e_0) \rangle, \dots, \langle \bar{Z}_0, c(t, 0, e_N) \rangle) Z_t \right. \\ & \left. + \tilde{\Gamma}(\langle \bar{Z}_0, c(t, 0, e_0) \rangle, \dots, \langle \bar{Z}_0, c(t, 0, e_N) \rangle) Z_t^2 \right)^{\frac{1}{2}} dW_t \end{aligned} \quad (3.4.11)$$

where $Z_0 = z_0 \in \mathbb{R}$ and $c : \Delta \times \mathbb{R}^{N+1} \rightarrow \mathbb{R}^{N+1}$ is continuous and $(e_0, \dots, e_N)$ the canonical basis of $\mathbb{R}^{N+1}$. To simplify notation and improve readability, let us write for all $t \geq 0$

$$\mathbf{vec}^{\bar{Z}_0, c}(t) := \begin{pmatrix} \langle \bar{Z}_0, c(t, 0, e_0) \rangle \\ \langle \bar{Z}_0, c(t, 0, e_1) \rangle \\ \vdots \\ \langle \bar{Z}_0, c(t, 0, e_N) \rangle \end{pmatrix}.$$

In this notation, (3.4.11) reads:

$$\begin{aligned} dZ_t = & \left(\tilde{b}(\mathbf{vec}^{\bar{Z}_0, c}(t)) + \tilde{\beta}(\mathbf{vec}^{\bar{Z}_0, c}(t)) Z_t \right) dt \\ & + \sqrt{\tilde{c}(\mathbf{vec}^{\bar{Z}_0, c}(t)) + \tilde{\gamma}(\mathbf{vec}^{\bar{Z}_0, c}(t)) Z_t + \tilde{\Gamma}(\mathbf{vec}^{\bar{Z}_0, c}(t)) Z_t^2} dW_t \end{aligned} \quad (3.4.12)$$

For such an SDE, we can now formulate the following theorem:

Theorem 3.4.13. Assume that a unique solution to the martingale problem associated with (3.4.11) exists and that the maps $c(\cdot)$ are the unique solution to the following ODEs

$$\frac{d}{dt} \left(c(t, 0, e_0), \dots, c(t, 0, e_N) \right) = \left(c(t, 0, e_0), \dots, c(t, 0, e_N) \right) \cdot \tilde{\mathbf{H}}(\mathbf{vec}^{\bar{Z}_0, c}(t)) \quad (3.4.13)$$

$$\frac{d}{dt} \left(c(T, t, e_0), \dots, c(T, t, e_N) \right) = -\tilde{\mathbf{H}}(\mathbf{vec}^{\bar{Z}_0, c}(t)) \cdot \left(c(T, t, e_0), \dots, c(T, t, e_N) \right) \quad (3.4.14)$$

where the matrix $\tilde{\mathbf{H}}$ is given by

$$\tilde{\mathbf{H}}(\text{vec}^{\mathbb{Z}_0, \mathbf{C}}(t))_{kj} = \begin{cases} j \cdot \tilde{\beta}(\text{vec}^{\mathbb{Z}_0, \mathbf{C}}(t)) + \frac{j(j-1)}{2} \tilde{\Gamma}(\text{vec}^{\mathbb{Z}_0, \mathbf{C}}(t)), & \text{if } k = j, j \neq 0 \\ j \cdot \tilde{b}(\text{vec}^{\mathbb{Z}_0, \mathbf{C}}(t)) + \frac{j(j-1)}{2} \tilde{\gamma}(\text{vec}^{\mathbb{Z}_0, \mathbf{C}}(t)), & \text{if } j = k + 1 \\ \frac{j(j-1)}{2} \tilde{c}(\text{vec}^{\mathbb{Z}_0, \mathbf{C}}(t)), & \text{if } j = k + 2 \\ 0 & \text{else} \end{cases}$$

for $j, k = 0, \dots, N$.

Setting for $u = (u_0, \dots, u_N)^\top \in \mathbb{R}^{N+1}$, $\mathbf{C}(t, s, u) := \sum_{i=0}^N u_i \mathbf{C}(t, s, e_i)$ for all $0 \leq s \leq t \leq T$, then

$$\mathcal{M}_t := \langle \bar{\mathbb{Z}}_t, \mathbf{C}(T, t, u) \rangle$$

is a martingale on $[0, T]$ for all $T \geq 0$ for all $u \in \mathbb{R}^{N+1}$ and therefore

$$\mathbb{E}[\mathcal{M}_T | \mathcal{F}_t] = \mathbb{E}[\langle \bar{\mathbb{Z}}_T, u \rangle | \mathcal{F}_t] = \langle \bar{\mathbb{Z}}_t, \mathbf{C}(T, t, u) \rangle.$$

Proof. Recall that by the Definitions 3.2.3 and 3.2.6 the process $(Z_t)_{t \geq 0}$ given by (3.4.11) is a time-inhomogeneous polynomial process with associated transition operators $(P_{s,t})_{(s,t) \in \Delta}$ and infinitesimal generators $(\mathcal{H}_s)_{s \geq 0}$. This is due to the assumption of a unique solution to the associated martingale problem.

Recall that by the definition of polynomial processes, the transition operators and the infinitesimal generators are linear operators on the space of polynomials, and we can associate with them matrices $(\mathbf{P}_{s,t})_{(s,t) \in \Delta}$ and $(\mathbf{H}_s)_{s \geq 0}$ with respect to some basis of polynomials of degree N , see Section 3.2.1. In fact, it is straightforward to check that

$$\mathbf{H}_t = \tilde{\mathbf{H}}(\text{vec}^{\mathbb{Z}_0, \mathbf{C}}(t)).$$

Note that the structure of $\tilde{\mathbf{H}}(\cdot)$ is the same as $\mathbf{L}(\cdot)^\top$ introduced in Subsection 3.4.1. Although this was not explicit in the primal derivation, recall that we did establish in the proof of Corollary 3.4.11 that $\mathbf{L}^\top(\cdot)$ corresponds to the matrix-form of the infinitesimal generator of 3.4.9.

By the definition of time-inhomogeneous polynomial processes, we know that there exists a $\mathbf{C}^* : \Delta \times \mathbb{R}^{N+1} \rightarrow \mathbb{R}^{N+1}$ such that for all $u \in \mathbb{R}^{N+1}$

$$\mathbf{C}^*(T, t, u) := \mathbf{P}_{t,T} \cdot u$$

with

$$\mathbb{E}[Z_T^i | \mathcal{F}_t] = \langle \bar{\mathbb{Z}}_t, \mathbf{P}_{t,T} \cdot e_i \rangle = \langle \bar{\mathbb{Z}}_t, \mathbf{C}^*(T, t, e_i) \rangle$$

as well as

$$\mathbb{E}[\langle \bar{\mathbb{Z}}_T, u \rangle | \mathcal{F}_t] = \sum_{i=0}^N u_i \mathbb{E}[\langle \bar{\mathbb{Z}}_T, e_i \rangle | \mathcal{F}_t] = \langle \bar{\mathbb{Z}}_t, \sum_{i=0}^N u_i \mathbf{P}_{t,T} \cdot e_i \rangle = \langle \bar{\mathbb{Z}}_t, \mathbf{C}^*(T, t, u) \rangle.$$

Let us also recall Example 3.2.11 for an illustration of the columns of $\mathbf{P}_{t,T}$. Even though the explicit moment formula (Theorem 3.2.12) only holds on finite time horizons (due to convergence issues of the Magnus expansion), the above relation holds on all of Δ .

By Lemma 3.2.14 it holds that

$$\begin{aligned}
\frac{d^+}{dt} C^*(t, 0, e_i) &= \frac{d^+}{dt} \mathbf{P}_{0,t} \cdot e_i = \mathbf{P}_{0,t} \cdot \tilde{\mathbf{H}}(\mathbf{vec}^{\mathbb{Z}_0, \mathbf{C}}(t)) \cdot e_i \\
&= \mathbf{P}_{0,t} \cdot \left(\sum_{j=0}^N e_j \cdot e_j^\top \right) \cdot \tilde{\mathbf{H}}(\mathbf{vec}^{\mathbb{Z}_0, \mathbf{C}}(t)) \cdot e_i \\
&= \sum_{j=0}^N \mathbf{P}_{0,t} \cdot e_j \cdot e_j^\top \cdot \tilde{\mathbf{H}}(\mathbf{vec}^{\mathbb{Z}_0, \mathbf{C}}(t)) \cdot e_i \\
&= \sum_{j=0}^N C^*(t, 0, e_j) \tilde{\mathbf{H}}(\mathbf{vec}^{\mathbb{Z}_0, \mathbf{C}}(t))_{ji} \quad \text{for } C^*(0, 0, e_i) = e_i \\
\\
\frac{d^+}{dt} C^*(T, t, e_i) &= \frac{d^+}{dt} \mathbf{P}_{t,T} \cdot e_i = -\tilde{\mathbf{H}}(\mathbf{vec}^{\mathbb{Z}_0, \mathbf{C}}(t)) \cdot \mathbf{P}_{t,T} \cdot e_i \\
&= -\tilde{\mathbf{H}}(\mathbf{vec}^{\mathbb{Z}_0, \mathbf{C}}(t)) \cdot C^*(T, t, e_i) \quad \text{for } C^*(T, T, e_i) = e_i.
\end{aligned}$$

But since we assumed that the map $\mathbf{C}$ is the unique solution to (3.4.13) and (3.4.14), then it holds that $C^* = \mathbf{C}$ and we proved the claim. $\square$

Remark 3.4.14. Considering the SDEs given by 3.4.11 under the assumptions of Theorem 3.4.13, then these SDEs are in fact of McKean-Vlasov type, since the coefficients $\tilde{b}, \tilde{\beta}, \tilde{c}, \tilde{\gamma}, \tilde{\Gamma}$ depend on the moments of $(Z_t)_{t \geq 0}$ due to

$$\langle \bar{\mathbb{Z}}_0, C(t, 0, e_i) \rangle = \mathbb{E}[Z_t^i], \quad i \in \{0, \dots, N\}.$$

Relating this to the vectorized notation, this means that $\mathbf{vec}^{\mathbb{Z}_0, \mathbf{C}}(t)_i = \mathbb{E}[Z_t^i]$ for all $t \in [0, \infty)$ and $0 \leq i \leq N$. However, we have neither established if and when (global) solutions for the ODEs (3.4.13) and (3.4.14) exist, nor if and when we have a unique solution to the associated martingale problem. Both of which will impose conditions on the maps $\tilde{b}, \tilde{\beta}, \tilde{c}, \tilde{\gamma}, \tilde{\Gamma}$. Note that this the existence of solutions to (3.4.14) and (3.4.13) is far from trivial, because $\mathbf{vec}^{\mathbb{Z}_0, \mathbf{C}}(t)$ depends on all of $c(t, 0, e_0), \dots, c(t, 0, e_N)$.

At this point, we would also like to emphasize again the connection between $\tilde{\mathbf{H}}$ and $\mathbf{L}$, namely that structurally $\tilde{\mathbf{H}} = \mathbf{L}^\top$, which should be even more clear now.

Proposition 3.4.15. Assume all the maps

$$\tilde{b}, \tilde{\beta}, \tilde{c}, \tilde{\gamma}, \tilde{\Gamma} : \mathbb{R}^{N+1} \rightarrow \mathbb{R}$$

to be locally Lipschitz-continuous on some set A such that $\mathbb{Z}_0 \in A$. Then a unique solution to (3.4.13), (3.4.14) exists up to its maximal lifetime. Moreover, if $A = \mathbb{R}^N$ and the maps

$\tilde{b}, \tilde{\beta}, \tilde{c}, \tilde{\gamma}, \tilde{\Gamma}$ either satisfy the set of Assumptions A1 or the set of Assumptions A2, then the unique solution to (3.4.13) is global and hence the same holds for the solution to (3.4.14). If, in addition to $A = \mathbb{R}^N$ and Assumptions A1 or A2, the term under the square-root in (3.4.11) remains non-negative, there exists a unique strong solution to (3.4.11).

Proof. Let us fix an arbitrary $Z_0 \in \mathbb{R}$ without loss of generality and note that, using (3.4.13) we obtain the following ODE for $\text{vec}^{\mathbb{Z}_0, c}$:

$$\frac{d}{dt} \text{vec}^{\mathbb{Z}_0, c}(t) = \tilde{\mathbf{H}}(\text{vec}^{\mathbb{Z}_0, c}(t))^T \cdot \text{vec}^{\mathbb{Z}_0, c}(t) \text{ with } \text{vec}^{\mathbb{Z}_0, c}(0) = \bar{Z}_0. \quad (3.4.15)$$

The existence of a unique solution to (3.4.15) up to its maximal lifetime is assured by the local Lipschitz conditions. Note, that (3.4.15) is the same system of ODEs as the one given in (3.4.8) by the fact that $\mathbf{H}^T = \mathbf{L}$. Hence, the existence of a global solution to (3.4.15) under the assumption that $A = \mathbb{R}^N$ and either Assumptions A1 or A2 being fulfilled, follows by an application of Proposition 3.4.6. Having obtained a solution to (3.4.15) given by $(\text{vec}^{\mathbb{Z}_0, c}(t))_{t \in [0, \infty)}$ we can plug this solution in to $\tilde{\mathbf{H}}(\cdot)$ to study (3.4.13) and (3.4.14) respectively to solve for $((c(T, t, e_i))_{i=0}^N)_{T \geq t \geq 0}$. Indeed, it holds that the right-hand-side of (3.4.13) is Lipschitz-continuous in $((c(t, 0, e_i))_{i=0}^N)_{t \in [0, \infty)}$. To see this, recall the form of (3.4.13). Having obtained that $\langle \bar{Z}_0, c(t, 0, e_i) \rangle < \infty$ for all $i = 0, \dots, N$ and $t \in \mathbb{R}_{\geq 0}$ for arbitrary $Z_0 \in \mathbb{R}$, it follows that $\|c(t, 0, e_i)\| < \infty$ all $i = 0, \dots, N$ and $t \in \mathbb{R}_{\geq 0}$. Similarly, for the backward ODE, we can simply plug in $(\tilde{\mathbf{H}}(\text{vec}^{\mathbb{Z}_0, c}(t)))_{0 \leq t \leq T}$ and recalling (3.4.14) it is clearly Lipschitz again, due to the bounds on $\tilde{\mathbf{H}}(\text{vec}^{\mathbb{Z}_0, c}(t))$. Since $\tilde{\mathbf{H}}(\text{vec}^{\mathbb{Z}_0, c}(t))$ is upper-triangular, one can recursively bound the components of $c(T, t, e_i)$ for each $i = 0, \dots, N$. More precisely, for each $i = 0, \dots, N$ apply Gronwall's inequality first on $c(T, t, e_i)_N$ (i.e. the N -th component of $c(T, t, e_i)$) using the bound on the coefficients of $\tilde{\mathbf{H}}(\cdot)$ and next repeat the same for $c(T, t, e_i)_{N-1}$ and so on. We have therefore shown that a global unique solution exists to the forward and backward ODE.

Let us now turn to the existence of a solution to the martingale problem and path-wise uniqueness. Let us note by (3.4.15) and the fact that in this notation (3.4.11) corresponds to (3.4.12), we are in fact exactly in the same situation as in the proof of Theorem 3.4.8. Hence, the arguments for existence and path-wise uniqueness of a solution are exactly those presented in the proof of Theorem 3.4.8. $\square$

Corollary 3.4.16. *As in the primal case, we can study the existence of solution on other state-spaces than $S = \mathbb{R}$. Sufficient conditions for the positive maximum principle to hold on the respective state-spaces are:*

- for state-space $S = \mathbb{R}_{\geq 0}$: $\tilde{b} : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ and $\tilde{c} \equiv 0, \tilde{\gamma}, \tilde{\Gamma} : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$.
- for the case of $S = [0, 1]$: $\tilde{b} : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}, \tilde{b} + \tilde{\beta} : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\leq 0}$, and $\tilde{c} \equiv 0, \tilde{\gamma} \equiv -\tilde{\Gamma} \geq 0$.

Proof. The proof is completely analogous to the one of Corollary 3.4.10. $\square$

Remark 3.4.17. Let us make the following comment on the difference between the primal and dual point of view: although we have ultimately used the same assumptions to establish existence of solutions, the derivation was profoundly different. In the primal case, we have required the set of Assumptions A1 or A2 to obtain the system of ODEs (3.4.8). It then turned out, that these growth-assumptions are also sufficient conditions for the existence of global solutions. In the dual case, however, we obtained the system of ODEs (3.4.15) via the Kolmogorov forward and backward equations. We had thus far not imposed any growth-conditions. Those only came into play as sufficient conditions for the existence of global solutions to (3.4.15). Although this difference may seem subtle, the following SDE shows it is not:

$$dZ_t = \sqrt{\mathbb{E}[Z_t^2]} Z_t dW_t \text{ with } Z_0 = z > 0. \quad (3.4.16)$$

Clearly, $\gamma(x) = x_2$ violates both Assumptions A1 and A2. Moreover, our approach to establish (3.4.8) in the primal approach via Gronwall's inequalities fails. However, in the dual case, neither Assumptions A1 nor A2 are required to find 3.4.15. And although, in this case, γ does not satisfy the sufficient conditions for the existence of a global solution, it is important to note that these are not necessary conditions. Indeed, in this case $N = 2$ and the system of ODEs boils down to:

$$\begin{aligned} \frac{d}{dt}(\text{vec}^{\mathbb{Z}_0, \mathbb{C}}(t))_1 &= 0 \text{ with } (\text{vec}^{\mathbb{Z}_0, \mathbb{C}}(0))_1 = z \\ \frac{d}{dt}(\text{vec}^{\mathbb{Z}_0, \mathbb{C}}(t))_2 &= (\text{vec}^{\mathbb{Z}_0, \mathbb{C}}(t))_2 (\text{vec}^{\mathbb{Z}_0, \mathbb{C}}(t))_1 \text{ with } (\text{vec}^{\mathbb{Z}_0, \mathbb{C}}(0))_2 = z^2 \end{aligned}$$

which admits a global unique solution

$$\text{vec}^{\mathbb{Z}_0, \mathbb{C}}(t) = \begin{pmatrix} z \\ z^2 e^{zt} \end{pmatrix}.$$

Using this it is straightforward to solve for $\mathbb{C}$. Moreover, for the state-space $S = \mathbb{R}_{\geq 0}$ the PMP holds, and we can apply Theorem 3.3.4 for $h(t) = e^{zt}$ to establish existence of path-wise unique solutions. *Ex post* we also find that in this example $(Z_t)_{t \in [0, \infty)}$ is indeed a true martingale, due to the fact that we have an explicit expression for $(\mathbb{E}[Z_t^2])_{t \in [0, \infty)}$ and can therefore check square-integrability.

Let us also point out that we will encounter this example again in Section 3.5.

3.4.3 Adding common noise

In this section, we study a class of SDEs which include an additional *common noise* term $(+\dots dW_t^0)$ and whose coefficients depend on the conditional expectation of the process, that is, conditional on the common noise. Again, we require a form which resembles that of polynomial processes to study this family of SDEs. Let us make this more precise in the following definition:

Definition 3.4.18 (Conditional Polynomial McKean-Vlasov SDEs). Let $(W_t)_{t \geq 0}$ and $(W_t^0)_{t \geq 0}$ denote two independent Brownian motions and denote by $\mathcal{W}_t^0$ the σ -algebra generated by

3.4 Polynomial McKean-Vlasov equations

W_t^0 for each $t \geq 0$. We slightly abuse notation and write for all $t \geq 0$

$$\mathbb{E}[Z_t | \mathcal{W}_t^0] := (\mathbb{E}[Z_t | \mathcal{W}_t^0], \dots, \mathbb{E}[Z_t^N | \mathcal{W}_t^0])$$

for some arbitrary but fixed $N > 0$ and some process $(Z_t)_{t \geq 0}$. We call the following SDE of *conditional polynomial McKean-Vlasov type*:

$$\begin{aligned} dZ_t = & \left(b(\mathbb{E}[Z_t | \mathcal{W}_t^0]) + \beta(\mathbb{E}[Z_t | \mathcal{W}_t^0])Z_t \right) dt \\ & + \sqrt{c(\mathbb{E}[Z_t | \mathcal{W}_t^0]) + \gamma(\mathbb{E}[Z_t | \mathcal{W}_t^0])Z_t + \Gamma(\mathbb{E}[Z_t | \mathcal{W}_t^0])Z_t^2} dW_t \\ & + \left(l(\mathbb{E}[Z_t | \mathcal{W}_t^0]) + \Lambda(\mathbb{E}[Z_t | \mathcal{W}_t^0])Z_t \right) dW_t^0 \end{aligned} \quad (3.4.17)$$

where $b, \beta, c, \gamma, \Gamma, l, \Lambda : \mathbb{R}^N \rightarrow \mathbb{R}$ are some continuous maps.

We would like to point out that a particular case of such a condition polynomial McKean-Vlasov SDE has been studied in [Cuchiero and Huber \(2024\)](#) in the context of market capitalizations curve modeling. The idea to approach this class of processes is similar to the primal approach to polynomial McKean-Vlasov SDEs of Section 3.4.1, where we used Itô's lemma to find an ODE for $\mathbb{E}[Z_t]$ and sufficient conditions for the existence of solutions to this system ODEs. Here, we aim to describe $\mathbb{E}[Z_t | \mathcal{W}_t^0]$ by a system of SDEs via Itô's lemma and determine sufficient conditions for the existence of solutions.

Before we derive this system of SDEs, let us state the following lemma, which will prove useful. We do not claim Lemma 3.4.19 to be a new result, but nevertheless state it to be self-contained and present a proof in the appendix.

Lemma 3.4.19. Let $(W_t)_{t \geq 0}$ and $(W_t^0)_{t \geq 0}$ be two independent Brownian motions, and write $(\mathcal{W}_t^0)_{t \geq 0}$ for the filtration generated by $(W_t^0)_{t \geq 0}$. Moreover, denote by $\mathbb{F} := (\mathcal{F}_t)_{t \geq 0}$ the filtration generated by both Brownian motions jointly. Then, for any locally square-integrable, $\mathbb{F}$ -adapted process $(H_t)_{t \geq 0}$ it holds that

$$\mathbb{E} \left[\int_0^t H_s dW_s \middle| \mathcal{W}_t^0 \right] = 0 \quad \text{and} \quad \mathbb{E} \left[\int_0^t H_s dW_s^0 \middle| \mathcal{W}_t^0 \right] = \int_0^t \mathbb{E} [H_s | \mathcal{W}_t^0] dW_s^0.$$

We are now ready to derive the system of SDEs which is fulfilled by $\mathbb{E}[Z_t | \mathcal{W}_t^0]$. Note that, we do not state anything about existence of solutions yet, but postpone this to Proposition 3.4.21, 3.4.22 and Theorem 3.4.24 and the discussion thereafter.

Similar to the case without common noise, we can formulate assumptions under which we can establish existence of and uniqueness of solutions to (3.4.17). Note that in the case with common-noise, however, we only have one set of assumptions. This is because the procedure of first bounding the first component, as we did for the set of Assumptions A1 does not work in the case with common noise.

Assumptions B1. Recall the notation $\|\cdot\|_N$ as introduced in Definition 3.4.2. We require the maps $b, \beta, c, \gamma, \Gamma, l, \Lambda$ to satisfy the following for all $x \in \mathbb{R}^N$:

- $|\beta(x)| \leq \beta_0, |\Gamma(x)| \leq \Gamma_0, |\Lambda(x)| \leq \Lambda_0$ for some $\beta_0, \Gamma_0, \Lambda_0 \in \mathbb{R}$;

- $|b(x)| \leq b_0(1 + \|x\|_N)$, $|\gamma(x)| \leq \gamma_0(1 + \|x\|_N)$ and $|l(x)| \leq l_0(1 + \|x\|_N)$ for some $b_0, \gamma_0, l_0 \in \mathbb{R}$;
- $|c(x)| \leq c_0(1 + \|x\|_N^2)$ for some $c_0 \in \mathbb{R}$.

Using this set of assumptions, we are ready to state the following lemma:

Lemma 3.4.20. Let us fix an arbitrary $N > 0$ and assume there exists a solution $(Z_t)_{t \in [0, \infty)}$ to (3.4.17). Moreover, let the maps $b, \beta, c, \gamma, \Gamma, l, \Lambda : \mathbb{R}^N \rightarrow \mathbb{R}$ satisfy Assumptions B1. Then, for all $1 \leq k \leq N$ it holds that

$$\begin{aligned} d\mathbb{E}[Z_t^k | \mathcal{W}_t^0] = & \left\{ \left(k \cdot b(\mathbb{E}[Z_t | \mathcal{W}_t^0]) + \frac{k(k-1)}{2} (\gamma(\mathbb{E}[Z_t | \mathcal{W}_t^0]) \right. \right. \\ & + 2l(\mathbb{E}[Z_t | \mathcal{W}_t^0])\Lambda(\mathbb{E}[Z_t | \mathcal{W}_t^0])) \mathbb{E}[Z_t^{k-1} | \mathcal{W}_t^0] \\ & + \mathbf{1}_{\{k \geq 2\}} \frac{k(k-1)}{2} (c(\mathbb{E}[Z_t | \mathcal{W}_t^0]) + l^2(\mathbb{E}[Z_t | \mathcal{W}_t^0])) \mathbb{E}[Z_t^{k-2} | \mathcal{W}_t^0] \\ & + (k \cdot \beta(\mathbb{E}[Z_t | \mathcal{W}_t^0]) + \frac{k(k-1)}{2} (\Gamma(\mathbb{E}[Z_t | \mathcal{W}_t^0]) + \Lambda^2(\mathbb{E}[Z_t | \mathcal{W}_t^0])) \mathbb{E}[Z_t^k | \mathcal{W}_t^0] \Big\} dt \\ & + \left(l(\mathbb{E}[Z_t | \mathcal{W}_t^0]) \mathbb{E}[Z_t^{k-1} | \mathcal{W}_t^0] + \Lambda(\mathbb{E}[Z_t | \mathcal{W}_t^0]) \mathbb{E}[Z_t^k | \mathcal{W}_t^0] \right) dW_t^0. \end{aligned}$$

Proof. Using Itô's formula, we obtain that

$$\begin{aligned} \mathbb{E}[Z_T^k | \mathcal{W}_T^0] = & \mathbb{E}[Z_0^k | \mathcal{W}_T^0] + \int_0^T \left\{ \mathbf{1}_{\{k \geq 2\}} \frac{k(k-1)}{2} (c(\mathbb{E}[Z_t | \mathcal{W}_t^0]) + l^2(\mathbb{E}[Z_t | \mathcal{W}_t^0])) \mathbb{E}[Z_t^{k-2} | \mathcal{W}_t^0] \right. \\ & + \left(k \cdot b(\mathbb{E}[Z_t | \mathcal{W}_t^0]) + \frac{k(k-1)}{2} (\gamma(\mathbb{E}[Z_t | \mathcal{W}_t^0]) + 2l(\mathbb{E}[Z_t | \mathcal{W}_t^0])\Lambda(\mathbb{E}[Z_t | \mathcal{W}_t^0])) \mathbb{E}[Z_t^{k-1} | \mathcal{W}_t^0] \right. \\ & + \left. \left. (k \cdot \beta(\mathbb{E}[Z_t | \mathcal{W}_t^0]) + \frac{k(k-1)}{2} (\Gamma(\mathbb{E}[Z_t | \mathcal{W}_t^0]) + \Lambda^2(\mathbb{E}[Z_t | \mathcal{W}_t^0])) \mathbb{E}[Z_t^k | \mathcal{W}_t^0] \right\} dt \right. \\ & + \mathbb{E} \left[\int_0^T \left(l(\mathbb{E}[Z_t | \mathcal{W}_t^0]) Z_t^{k-1} + \Lambda(\mathbb{E}[Z_t | \mathcal{W}_t^0]) Z_t^k \right) dW_t^0 \middle| \mathcal{W}_T^0 \right] \\ & + \mathbb{E} \left[\int_0^T \left(c(\mathbb{E}[Z_t | \mathcal{W}_t^0]) Z_t^{2k-2} + \gamma(\mathbb{E}[Z_t | \mathcal{W}_t^0]) Z_t^{2k-1} + \Gamma(\mathbb{E}[Z_t | \mathcal{W}_t^0]) Z_t^{2k} \right)^{\frac{1}{2}} dW_t \middle| \mathcal{W}_T^0 \right]. \end{aligned}$$

We know that the last term vanishes if the integrand is square-integrable, thanks to Lemma 3.4.19. Likewise, we can exchange integration and conditional expectation in the second-last term, if the integrand is square-integrable, again due to Lemma 3.4.19. Let us note that $\mathbb{E}[Z_t | \mathcal{W}_T^0] = \mathbb{E}[Z_t | \mathcal{W}_t^0]$ for all $t \leq T$. We also recall the following inequality, which we have derived in Subsection 3.4.1 for the case of the expectation, namely that for all $i \in \{1, \dots, N\}$

$$|\mathbb{E}[Z_t^i | \mathcal{W}_t^0]|^{\frac{2}{i}} = |\mathbb{E}[Z_t^i | \mathcal{W}_t^0]|^{\frac{2N}{iN}} \leq \mathbb{E}[|Z_t|^N | \mathcal{W}_t^0]^{\frac{2}{N}}$$

and hence

$$\|\mathbb{E}[Z_t|\mathcal{W}_t^0]\|_N \lesssim_N \mathbb{E}[|Z_t|^N|\mathcal{W}_t^0]^{\frac{1}{N}}.$$

We now check square-integrability of the integrand of the last term.

$$\begin{aligned} & \int_0^T \mathbb{E} \left[|c(\mathbb{E}[Z_t|\mathcal{W}_t^0])Z_t^{2k-2} + \gamma(\mathbb{E}[Z_t|\mathcal{W}_t^0])Z_t^{2k-1} + \Gamma(\mathbb{E}[Z_t|\mathcal{W}_t^0])Z_t^{2k}| \right] dt \\ & \lesssim_N \int_0^T c_0 \left(\mathbb{E}[|Z_t|^{2k-2}] + \mathbb{E} \left[|\mathbb{E}[|Z_t|^N|\mathcal{W}_t^0]^{\frac{2}{N}} Z_t^{2k-2}| \right] \right) dt \\ & \quad + \int_0^T \gamma_0 \left(\mathbb{E}[|Z_t|^{2k-1}] + \mathbb{E} \left[|\mathbb{E}[|Z_t|^N|\mathcal{W}_t^0]^{\frac{1}{N}} Z_t^{2k-1}| \right] \right) + \Gamma_0 \mathbb{E}[|Z_t^{2k}|] dt \end{aligned}$$

Note that for $k > 1$

$$\begin{aligned} & \mathbb{E} \left[|\mathbb{E}[|Z_t|^N|\mathcal{W}_t^0]^{\frac{2}{N}} Z_t^{2k-2}| \right] = \mathbb{E} \left[\mathbb{E}[|Z_t|^N|\mathcal{W}_t^0]^{\frac{2}{N}} \mathbb{E}[|Z_t|^{2k-2}|\mathcal{W}_t^0] \right] \\ & \leq \mathbb{E} \left[\mathbb{E}[|Z_t|^{2N}|\mathcal{W}_t^0]^{\frac{1}{N}} \mathbb{E}[|Z_t|^{2(k-1)}|\mathcal{W}_t^0]^{\frac{(k-1)N}{(k-1)N}} \right] \leq \mathbb{E} \left[\mathbb{E}[|Z_t|^{2N}|\mathcal{W}_t^0]^{\frac{1}{N}} \mathbb{E}[|Z_t|^{2N}|\mathcal{W}_t^0]^{\frac{k-1}{N}} \right] \\ & = \mathbb{E} \left[\mathbb{E}[|Z_t|^{2N}|\mathcal{W}_t^0]^{\frac{k}{N}} \right] \leq (1 + \mathbb{E}[\mathbb{E}[|Z_t|^{2N}|\mathcal{W}_t^0]]) = (1 + \mathbb{E}[|Z_t|^{2N}]), \end{aligned}$$

while for $k = 1$

$$\mathbb{E} \left[|\mathbb{E}[|Z_t|^N|\mathcal{W}_t^0]^{\frac{2}{N}} Z_t^{2k-2}| \right] = \mathbb{E} \left[\mathbb{E}[|Z_t|^N|\mathcal{W}_t^0]^{\frac{2}{N}} \right] \leq \mathbb{E} \left[1 + \mathbb{E}[|Z_t|^{2N}|\mathcal{W}_t^0] \right] = (1 + \mathbb{E}[|Z_t|^{2N}])$$

and likewise for $k \geq 1$:

$$\begin{aligned} & \mathbb{E} \left[|\mathbb{E}[|Z_t|^N|\mathcal{W}_t^0]^{\frac{1}{N}} Z_t^{2k-1}| \right] = \mathbb{E} \left[\mathbb{E}[|Z_t|^N|\mathcal{W}_t^0]^{\frac{1}{N}} \mathbb{E}[|Z_t|^{2k-1}|\mathcal{W}_t^0] \right] \\ & \leq \mathbb{E} \left[\mathbb{E}[|Z_t|^N|\mathcal{W}_t^0]^{\frac{1}{N}} \mathbb{E}[|Z_t|^{2(k-1)}|\mathcal{W}_t^0]^{\frac{(2k-1)N}{(2k-1)N}} \right] \leq \mathbb{E} \left[\mathbb{E}[|Z_t|^N|\mathcal{W}_t^0]^{\frac{1}{N}} \mathbb{E}[|Z_t|^N|\mathcal{W}_t^0]^{\frac{2k-1}{N}} \right] \\ & = \mathbb{E} \left[\mathbb{E}[|Z_t|^N|\mathcal{W}_t^0]^{\frac{2k}{N}} \right] \leq (1 + \mathbb{E}[\mathbb{E}[|Z_t|^{2N}|\mathcal{W}_t^0]]) = (1 + \mathbb{E}[|Z_t|^{2N}]). \end{aligned}$$

Similarly, it holds for the second last term that

$$\begin{aligned} & \int_0^T \mathbb{E} \left[|l(\mathbb{E}[Z_t|\mathcal{W}_t^0])Z_t^{k-1} + \Lambda(\mathbb{E}[Z_t|\mathcal{W}_t^0])Z_t^k|^2 \right] dt \\ & \leq \int_0^T l_0 \left(\mathbb{E}[|Z_t|^{2k-2}] + \mathbb{E} \left[|\mathbb{E}[|Z_t|^N|\mathcal{W}_t^0]^{\frac{2}{N}} Z_t^{2k-2}| \right] \right) + \Lambda_0 \mathbb{E}[|Z_t^{2k}|] dt \end{aligned}$$

which can be treated using the same inequalities as above.

Hence, it is enough to show that $\sup_{t \in [0, T]} \mathbb{E}[|Z_t|^{2N}] < \infty$ for all $T \in [0, \infty)$ in order to show square-integrability. To prove this, we can proceed exactly like in the proof of Lemma 3.4.3 and show the statement using a Gronwall-type inequality. Again, we introduce

a sequence of stopping-times $(\tau_n)_{n \in \mathbb{N}}$ with $\tau_n = \inf\{t \geq 0 \mid |Z_t| \geq n\}$. We obtain that

$$\begin{aligned}
\mathbb{E} \left[|Z_{t \wedge \tau_n}|^{2N} \right] &= \mathbb{E} \left[\left| Z_0 + \int_0^{t \wedge \tau_n} b(\mathbb{E}[Z_s | \mathcal{W}_s^0]) + \beta(\mathbb{E}[Z_s | \mathcal{W}_s^0]) Z_s ds \right. \right. \\
&\quad + \int_0^{t \wedge \tau_n} \sqrt{c(\mathbb{E}[Z_s | \mathcal{W}_s^0]) + \gamma(\mathbb{E}[Z_s | \mathcal{W}_s^0]) Z_s + \Gamma(\mathbb{E}[Z_s | \mathcal{W}_s^0]) Z_s^2} dW_s \\
&\quad \left. \left. + \int_0^{t \wedge \tau_n} l(\mathbb{E}[Z_s | \mathcal{W}_s^0]) + \Lambda(\mathbb{E}[Z_s | \mathcal{W}_s^0]) Z_s dW_s^0 \right|^{2N} \right] \\
&\lesssim_N \mathbb{E} \left[|Z_0|^{2N} \right] + \mathbb{E} \left[t^{2N-1} \int_0^t |b(\mathbb{E}[Z_{s \wedge \tau_n} | \mathcal{W}_s^0]) + \beta(\mathbb{E}[Z_{s \wedge \tau_n} | \mathcal{W}_s^0]) Z_{s \wedge \tau_n}|^{2N} ds \right] \\
&\quad + \mathbb{E} \left[\int_0^t |c(\mathbb{E}[Z_{s \wedge \tau_n} | \mathcal{W}_s^0]) + \gamma(\mathbb{E}[Z_{s \wedge \tau_n} | \mathcal{W}_s^0]) Z_{s \wedge \tau_n} + \Gamma(\mathbb{E}[Z_{s \wedge \tau_n} | \mathcal{W}_s^0]) Z_{s \wedge \tau_n}^2|^{2N} ds \right] \\
&\quad + \mathbb{E} \left[\int_0^t |l(\mathbb{E}[Z_{s \wedge \tau_n} | \mathcal{W}_s^0]) + \Lambda(\mathbb{E}[Z_{s \wedge \tau_n} | \mathcal{W}_s^0]) Z_{s \wedge \tau_n}|^{2N} ds \right] \\
&\lesssim_N \mathbb{E} \left[|Z_0|^{2N} \right] + t^{2N-1} \int_0^t b_0^{2N} \left(1 + \mathbb{E} \left[\mathbb{E}[|Z_{s \wedge \tau_n}|^N | \mathcal{W}_s^0]^2 \right] \right) + \beta_0^{2N} \mathbb{E}[|Z_{s \wedge \tau_n}|^{2N}] ds \\
&\quad + \int_0^t c_0^N \left(1 + \mathbb{E} \left[\mathbb{E}[|Z_{s \wedge \tau_n}|^N | \mathcal{W}_s^0]^2 \right] \right) + \Gamma_0^N \mathbb{E}[|Z_{s \wedge \tau_n}|^{2N}] ds \\
&\quad + \int_0^t \gamma_0^N \left(\mathbb{E}[|Z_{s \wedge \tau_n}|^N] + \mathbb{E} \left[\mathbb{E}[|Z_{s \wedge \tau_n}|^N | \mathcal{W}_s^0] |Z_{s \wedge \tau_n}|^N \right] \right) ds \\
&\quad + \int_0^t l_0^{2N} \left(1 + \mathbb{E} \left[\mathbb{E}[|Z_{s \wedge \tau_n}|^N | \mathcal{W}_s^0]^2 \right] \right) + \Lambda_0^{2N} \mathbb{E}[|Z_{s \wedge \tau_n}|^{2N}] ds \\
&\lesssim_N \mathbb{E} \left[|Z_0|^{2N} \right] + (b_0 t)^{2N} + c_0^{2N} t + \gamma_0^{2N} t + l_0^{2N} t \\
&\quad + \left(b_0^{2N} t^{2N-1} + \beta_0^{2N} t^{2N-1} + c_0^N + \gamma_0^N + \Gamma_0^N + l_0^{2N} + \Lambda_0^{2N} \right) \int_0^t \mathbb{E}[|Z_{s \wedge \tau_n}|^{2N}] ds
\end{aligned}$$

where we used that

$$\mathbb{E} \left[\mathbb{E}[|Z_{s \wedge \tau_n}|^N | \mathcal{W}_s^0]^2 \right] \leq \mathbb{E} \left[\mathbb{E}[|Z_{s \wedge \tau_n}|^{2N} | \mathcal{W}_s^0] \right] = \mathbb{E} \left[|Z_{s \wedge \tau_n}|^{2N} \right]$$

and

$$\mathbb{E} \left[\mathbb{E}[|Z_{s \wedge \tau_n}|^N | \mathcal{W}_s^0] |Z_{s \wedge \tau_n}|^N \right] = \mathbb{E} \left[\mathbb{E}[|Z_{s \wedge \tau_n}|^N | \mathcal{W}_s^0]^2 \right] \leq \mathbb{E} \left[|Z_{s \wedge \tau_n}|^{2N} \right].$$

Hence, by an application of Gronwall's inequality, we have indeed found an upper bound for $\mathbb{E} \left[|Z_{t \wedge \tau_n}|^{2N} \right] \leq \Theta(t)$ for a continuous increasing function $\Theta : [0, \infty) \rightarrow [0, \infty)$. And moreover, by the fact that $(\tau_n)_{n \in \mathbb{N}}$ and Fatou's lemma, we obtain that also $\mathbb{E} \left[|Z_t|^{2N} \right] \leq \Theta(t)$. $\square$

We are now finally ready to tackle the question of existence of solutions. To do so, we will establish further conditions on the structure of the drift and volatility conditions. Our

first approach is to study, under which assumptions, the process $(Z_t, \mathbb{E}[Z_t|\mathcal{W}_t^0])_{t \geq 0}$ is a multidimensional polynomial process. Note that we are now in a time-homogeneous setting, see [Cuchiero et al. \(2012\)](#) for polynomial processes that are time-homogeneous. We first formulate this for a general state-space S under the assumption of the PMP being satisfied. Later, we will provide examples of state-spaces as well as sufficient conditions for the PMP to hold.

Proposition 3.4.21. Let us assume that the maps $b, \beta, \gamma, \Gamma, l, \Lambda$ are all constant, denoted by $b_0, \beta_0, \gamma_0, \Gamma_0, l_0, \Lambda_0$ respectively, as well as $c(x) = c_0 + c_1 x_1 + c_2 x_2$ for some $c_0, c_1, c_2 \in \mathbb{R}$ and all $x \in \mathbb{R}^N$. If the PMP is satisfied on the desired state-space $S \subseteq \mathbb{R}^3$, then for $N = 2$, $(Z_t, \mathbb{E}[Z_t|\mathcal{W}_t^0])_{t \geq 0}$ (3.4.17) is jointly polynomial on S and there exists a path-wise unique solution. More explicitly, the process $(Z_t, \mathbb{E}[Z_t|\mathcal{W}_t^0])_{t \geq 0}$ satisfies the following system of SDEs:

$$\begin{aligned} dZ_t &= (b_0 + \beta_0 Z_t) dt + \sqrt{c_0 + c_1 \mathbb{E}[Z_t|\mathcal{W}_t^0] + c_2 \mathbb{E}[Z_t^2|\mathcal{W}_t^0] + \gamma_0 Z_t + \Gamma_0 Z_t^2} dW_t \\ &\quad + (l_0 + \Lambda_0 Z_t) dW_t^0 \\ d\mathbb{E}[Z_t|\mathcal{W}_t^0] &= \{b_0 + \beta_0 \mathbb{E}[Z_t|\mathcal{W}_t^0]\} dt + \{l_0 + \Lambda_0 \mathbb{E}[Z_t|\mathcal{W}_t^0]\} dW_t^0 \\ d\mathbb{E}[Z_t^2|\mathcal{W}_t^0] &= \{c_0 + l_0^2 + (2b_0 + \gamma_0 + 2l_0 \Lambda_0 + c_1) \mathbb{E}[Z_t|\mathcal{W}_t^0] \\ &\quad + (2\beta_0 + \Gamma_0 + \Lambda_0^2 + c_2) \mathbb{E}[Z_t^2|\mathcal{W}_t^0]\} dt + \{l_0 \mathbb{E}[Z_t|\mathcal{W}_t^0] + \Lambda_0^0 \mathbb{E}[Z_t^2|\mathcal{W}_t^0]\} dW_t^0 \end{aligned}$$

Proof. Under these assumptions, it is straightforward to check that the system of SDEs outlined in Lemma 3.4.20 reduces to the system of SDEs stated in the claim. Moreover, the process $(Z_t, \mathbb{E}[Z_t|\mathcal{W}_t^0], \mathbb{E}[Z_t^2|\mathcal{W}_t^0])_{t \geq 0}$ is clearly polynomial, which can for example be seen from the differential characteristics. Regarding the path-wise uniqueness, we first consider $(\mathbb{E}[Z_t|\mathcal{W}_t^0], \mathbb{E}[Z_t^2|\mathcal{W}_t^0])_{t \geq 0}$ alone. Indeed, the corresponding drift and diffusion term are locally Lipschitz continuous such that ([Ethier and Kurtz, 2005](#), Theorem 3.7) is applicable and hence a path-wise unique solution exists for $(\mathbb{E}[Z_t|\mathcal{W}_t^0], \mathbb{E}[Z_t^2|\mathcal{W}_t^0])_{t \geq 0}$. We can now plug this solution to solve for $(Z_t) \geq 0$ which is now a one-dimensional SDE and hence we can apply ([Ethier and Kurtz, 2005](#), Theorem 3.8) to establish path-wise uniqueness. $\square$

Alternatively, by changing the form of the map $c(\cdot)$ to not include a term $\mathbb{E}[Z_t^2|\mathcal{W}_t^0]$ but instead a term $(\mathbb{E}[Z_t|\mathcal{W}_t^0])^2$ we gain more flexibility on the form of the rest of the coefficients, since we now only need for $(Z_t, \mathbb{E}[Z_t|\mathcal{W}_t^0])_{t \geq 0}$ to be polynomial instead of $(Z_t, \mathbb{E}[Z_t|\mathcal{W}_t^0], \mathbb{E}[Z_t^2|\mathcal{W}_t^0])_{t \geq 0}$. Let us make this more precise in the next proposition.

Proposition 3.4.22. Assume that the following hold for all $x \in \mathbb{R}^N$

- $\beta(x) \equiv \beta_0, \Gamma(x) \equiv \Gamma_0, l(x) \equiv l_0, \Lambda(x) \equiv \Lambda_0$ for some $\beta_0, \Gamma_0, l_0, \Lambda_0 \in \mathbb{R}$
- $b(x) = b_0 + \tilde{b}_1 x_1$ and $\gamma(x) = \gamma_0 + \gamma_1 x_1$ for some $b_0, \tilde{b}_1, \gamma_0, \gamma_1 \in \mathbb{R}$;
- $c(x) = c_0 + c_1 x_1 + \tilde{c}_2 (x_1)^2$ for some $c_0, c_1, \tilde{c}_2 \in \mathbb{R}$.

If the PMP is satisfied on $S \subseteq \mathbb{R}^2$, then for $(Z_t)_{t \geq 0}$ given by (3.4.17), $(Z_t, \mathbb{E}[Z_t | \mathcal{W}_t^0])_{t \geq 0}$ is a polynomial process on S given by the solution to the SDEs:

$$\begin{aligned} dZ_t &= \left(b_0 + b_1 \mathbb{E}[Z_t | \mathcal{W}_t^0] + \beta_0 Z_t \right) dt + (l_0 + \Lambda_0 Z_t) dW_t^0 \\ &\quad + \sqrt{c_0 + c_1 \mathbb{E}[Z_t | \mathcal{W}_t^0] + c_2 (\mathbb{E}[Z_t | \mathcal{W}_t^0])^2 + (\gamma_0 + \gamma_1 \mathbb{E}[Z_t | \mathcal{W}_t^0]) Z_t + \Gamma_0 Z_t^2} dW_t \\ d\mathbb{E}[Z_t | \mathcal{W}_t^0] &= \left(b_0 + (b_1 + \beta_0) \mathbb{E}[Z_t | \mathcal{W}_t^0] \right) dt + \left(l_0 + \Lambda_0 \mathbb{E}[Z_t | \mathcal{W}_t^0] \right) dW_t^0. \end{aligned}$$

In particular, this solution is path-wise unique.

Proof. It is straightforward to derive this system of SDEs from Lemma 3.4.20 and to see that the solution is a time-homogenous polynomial process. Regarding path-wise uniqueness, we can proceed exactly as in Proposition 3.4.21. Solving first for $(\mathbb{E}[Z_t | \mathcal{W}_t^0])_{t \geq 0}$ and then for $(Z_t)_{t \geq 0}$. $\square$

Remark 3.4.23. The above results hold under the assumption that the PMP is satisfied, and we will provide explicit conditions to ensure that in Proposition 3.4.26.

The fact that the above processes are jointly polynomial gives us access to the full polynomial theory, in particular the moment formula, which may be desirable in applications.

Instead of requiring the processes to be jointly polynomial, we can of course also treat the general case of equation (3.4.17), by showing existence of solutions via the associated martingale problem. This is the purpose of the next theorem. Again, we first formulate existence on some state space S under the assumption that the PMP holds and provide concrete state spaces as well as sufficient conditions for the PMP later.

Theorem 3.4.24. *Let $N \geq 1$ and assume that Assumptions B1 are satisfied. Moreover, let $S \subseteq \mathbb{R}^{N+1}$ be a state space for which the PMP is fulfilled. Then (3.4.17) admits a solution with values in S in the sense of a solution to the martingale problem.*

Proof. To see this, we describe $(Z_t, \mathbb{E}[Z_t | \mathcal{W}_t^0])_{t \geq 0}$ as a joint system of SDEs and recall (Ethier and Kurtz, 2005, Theorem 3.3.10) regarding sufficient conditions for the existence of solutions to SDEs via solutions to the martingale problem. To this end, let us write $X_t = (Z_t, \mathbb{E}[Z_t | \mathcal{W}_t^0])$ for all $t \geq 0$ which is described by:

$$dX_t = B(X_t)dt + \Sigma(X_t)d\tilde{W}_t$$

where $B : \mathbb{R}^{N+1} \rightarrow \mathbb{R}^{N+1}$, $\Sigma : \mathbb{R}^{N+1} \rightarrow \mathbb{R}^{(N+1) \times 2}$ and $\tilde{W}_t = (W_t, W_t^0)$ for all $t \geq 0$. In particular, we emphasize that we are now in a time-homogeneous setting. By (Ethier and Kurtz, 2005, Theorem 3.3.10) we are left to check that there exists a $K \geq 0$, such that

$$(X_t)_j \cdot B(X_t)_j \leq K(1 + \|X_t\|^2) \text{ and } \left| \left(\Sigma(X_t) \Sigma(X_t)^\top \right)_{ij} \right| \leq K(1 + \|X_t\|^2)$$

for all $0 \leq i, j \leq N$. Note that we use the notation $(X_t)_0 := Z_t$ and $(X_t)_k := \mathbb{E}[Z_t^k | \mathcal{W}_t^0]$ for all $1 \leq k \leq N$. We assume w.l.o.g that N is even³. Note, that

$$\begin{aligned} B(X_t)_0 &= b(\mathbb{E}[Z_t | \mathcal{W}_t^0]) + \beta(\mathbb{E}[Z_t | \mathcal{W}_t^0])Z_t \\ B(X_t)_k &= \left(k \cdot b(\mathbb{E}[Z_t | \mathcal{W}_t^0]) + \frac{k(k-1)}{2} (\gamma(\mathbb{E}[Z_t | \mathcal{W}_t^0]) \right. \\ &\quad \left. + l(\mathbb{E}[Z_t | \mathcal{W}_t^0])\Lambda(\mathbb{E}[Z_t | \mathcal{W}_t^0])) \right) \mathbb{E}[Z_t^{k-1} | \mathcal{W}_t^0] \\ &\quad + \mathbf{1}_{\{k \geq 2\}} \frac{k(k-1)}{2} (c(\mathbb{E}[Z_t | \mathcal{W}_t^0]) + l^2(\mathbb{E}[Z_t | \mathcal{W}_t^0])) \mathbb{E}[Z_t^{k-2} | \mathcal{W}_t^0] \\ &\quad + (k \cdot \beta(\mathbb{E}[Z_t | \mathcal{W}_t^0]) + \frac{k(k-1)}{2} (\Gamma(\mathbb{E}[Z_t | \mathcal{W}_t^0]) + \Lambda^2(\mathbb{E}[Z_t | \mathcal{W}_t^0]))) \mathbb{E}[Z_t^k | \mathcal{W}_t^0] \end{aligned}$$

Clearly it holds that

$$(X_t)_0 \cdot B(X_t)_0 \lesssim_N Z_t \cdot \left(b_0(1 + \mathbb{E}[|Z_t|^N | \mathcal{W}_t^0]^{\frac{1}{N}}) + \beta_0 Z_t \right) \leq K(1 + \|X_t\|^2)$$

since $ab \leq (a+b)^2 \lesssim |a|^2 + |b|^2$ for any $a, b \in \mathbb{R}$. Similarly, it holds for $k \in \{1, \dots, N\}$

$$\begin{aligned} (X_t)_k \cdot B(X_t)_k &\leq \left(kb_0 + \frac{k(k-1)}{2} (\gamma_0 + l_0 \Lambda_0) \right) \left| \mathbb{E}[Z_t^{k-1} | \mathcal{W}_t^0] \mathbb{E}[Z_t^k | \mathcal{W}_t^0] \right| \left(1 + \|\mathbb{E}[Z_t | \mathcal{W}_t^0]\|_N \right) \\ &\quad + \mathbf{1}_{\{k \geq 2\}} \frac{k(k-1)}{2} \left(c_0 + l_0^2 (1 + \|\mathbb{E}[Z_t | \mathcal{W}_t^0]\|_N)^2 \right) \left| \mathbb{E}[Z_t^{k-2} | \mathcal{W}_t^0] \mathbb{E}[Z_t^k | \mathcal{W}_t^0] \right| \\ &\quad + \left(k\beta_0 + \frac{k(k-1)}{2} (\Gamma_0 + \Lambda_0^2) \right) \left| \mathbb{E}[Z_t^k | \mathcal{W}_t^0] \right|^2 \\ &\leq K(1 + \|X_t\|^2) \end{aligned}$$

where we used that $\|\mathbb{E}[Z_t | \mathcal{W}_t^0]\|_N \lesssim_N \mathbb{E}[|Z_t|^N | \mathcal{W}_t^0]^{\frac{1}{N}}$ together with

$$\begin{aligned} \left| \mathbb{E}[Z_t^{k-1} | \mathcal{W}_t^0] \mathbb{E}[Z_t^k | \mathcal{W}_t^0] \right| &= \left| \mathbb{E}[Z_t^{k-1} | \mathcal{W}_t^0] \right|^{\frac{N(k-1)}{N(k-1)}} \left| \mathbb{E}[Z_t^k | \mathcal{W}_t^0] \right|^{\frac{Nk}{Nk}} \\ &\leq \mathbb{E}[|Z_t|^N | \mathcal{W}_t^0]^{\frac{(k-1)}{N}} \mathbb{E}[|Z_t|^N | \mathcal{W}_t^0]^{\frac{k}{N}} \\ &= \mathbb{E}[|Z_t|^N | \mathcal{W}_t^0]^{\frac{2k-1}{N}} \end{aligned}$$

and similarly

$$\left| \mathbb{E}[Z_t^{k-2} | \mathcal{W}_t^0] \mathbb{E}[Z_t^k | \mathcal{W}_t^0] \right| \leq \mathbb{E}[|Z_t|^N | \mathcal{W}_t^0]^{\frac{2k-2}{N}}$$

and hence

$$\begin{aligned} \left| \mathbb{E}[Z_t^{k-1} | \mathcal{W}_t^0] \mathbb{E}[Z_t^k | \mathcal{W}_t^0] \right| \|\mathbb{E}[Z_t | \mathcal{W}_t^0]\|_N &\lesssim \mathbb{E}[|Z_t|^N | \mathcal{W}_t^0]^{\frac{2k}{N}} \leq (1 + \mathbb{E}[|Z_t|^N | \mathcal{W}_t^0])^2 \\ \left| \mathbb{E}[Z_t^{k-2} | \mathcal{W}_t^0] \mathbb{E}[Z_t^k | \mathcal{W}_t^0] \right| \|\mathbb{E}[Z_t | \mathcal{W}_t^0]\|_N^2 &\lesssim \mathbb{E}[|Z_t|^N | \mathcal{W}_t^0]^{\frac{2k}{N}} \leq (1 + \mathbb{E}[|Z_t|^N | \mathcal{W}_t^0])^2. \end{aligned}$$

³If the minimal N we consider would actually be odd, it is no problem to go up to $N+1$ instead.

On the other hand, for $j, k \in \{1, \dots, N\}$ we find

$$\begin{aligned} \left(\Sigma(X_t) \Sigma(X_t)^\top \right)_{jk} &= \underbrace{\Sigma(X_t)_{j0} \Sigma(X_t)_{k0}}_{=0} + \Sigma(X_t)_{j1} \Sigma(X_t)_{k1} \\ &= \left(l(\mathbb{E}[Z_t | \mathcal{W}_t^0]) \mathbb{E}[Z_t^{k-1} | \mathcal{W}_t^0] + \Lambda(\mathbb{E}[Z_t | \mathcal{W}_t^0]) \mathbb{E}[Z_t^k | \mathcal{W}_t^0] \right) \\ &\quad \cdot \left(l(\mathbb{E}[Z_t | \mathcal{W}_t^0]) \mathbb{E}[Z_t^{j-1} | \mathcal{W}_t^0] + \Lambda(\mathbb{E}[Z_t | \mathcal{W}_t^0]) \mathbb{E}[Z_t^j | \mathcal{W}_t^0] \right). \end{aligned}$$

Using the growth conditions of l, Λ and the fact that

$$\|\mathbb{E}[Z_t | \mathcal{W}_t^0]\|_N \mathbb{E}[|Z_t|^{k-1} | \mathcal{W}_t^0] \lesssim_N \mathbb{E}[|Z_t|^N | \mathcal{W}_t^0]^{\frac{k}{N}} \leq 1 + \mathbb{E}[|Z_t|^N | \mathcal{W}_t^0] \quad (3.4.18)$$

it is clear that

$$\left| \left(\Sigma(X_t) \Sigma(X_t)^\top \right)_{jk} \right| \leq K(1 + \|X_t\|^2)$$

for all $i, j \in \{1, \dots, N\}$. Whereas

$$\begin{aligned} \left| \left(\Sigma(X_t) \Sigma(X_t)^\top \right)_{00} \right| &= \left| (\Sigma(X_t)_{00})^2 + (\Sigma(X_t)_{01})^2 \right| \\ &= \left| c(\mathbb{E}[Z_t | \mathcal{W}_t^0]) + \gamma(\mathbb{E}[Z_t | \mathcal{W}_t^0]) Z_t + \Gamma(\mathbb{E}[Z_t | \mathcal{W}_t^0]) Z_t^2 \right. \\ &\quad \left. + l(\mathbb{E}[Z_t | \mathcal{W}_t^0])^2 + 2l(\mathbb{E}[Z_t | \mathcal{W}_t^0]) \Lambda(\mathbb{E}[Z_t | \mathcal{W}_t^0]) Z_t + \Lambda(\mathbb{E}[Z_t | \mathcal{W}_t^0])^2 Z_t^2 \right| \\ &\leq K(1 + \|X_t\|^2) \end{aligned}$$

by assumptions on the maps $c, l, \gamma, \Gamma, \Lambda$. We are now only left with the terms $\left(\Sigma(X_t) \Sigma(X_t)^\top \right)_{0k}$ for $k \in \{1, \dots, N\}$, where

$$\begin{aligned} \left| \left(\Sigma(X_t) \Sigma(X_t)^\top \right)_{0k} \right| &= \left| \Sigma(X_t)_{01} \Sigma(X_t)_{k1} \right| = \left| \left(l(\mathbb{E}[Z_t | \mathcal{W}_t^0]) + \Lambda(\mathbb{E}[Z_t | \mathcal{W}_t^0]) Z_t \right) \right. \\ &\quad \left. \cdot \left(l(\mathbb{E}[Z_t | \mathcal{W}_t^0]) \mathbb{E}[Z_t^{k-1} | \mathcal{W}_t^0] + \Lambda(\mathbb{E}[Z_t | \mathcal{W}_t^0]) \mathbb{E}[Z_t^k | \mathcal{W}_t^0] \right) \right| \\ &\lesssim_N \left| \left(l_0 + l_0 \mathbb{E}[|Z_t|^N | \mathcal{W}_t^0]^{\frac{1}{N}} + \Lambda_0 |Z_t| \right) \right. \\ &\quad \left. \cdot \left(l_0 \mathbb{E}[|Z_t|^N | \mathcal{W}_t^0]^{\frac{k-1}{N}} + l_0 \mathbb{E}[|Z_t|^N | \mathcal{W}_t^0]^{\frac{k}{N}} + \Lambda_0 \mathbb{E}[|Z_t|^k | \mathcal{W}_t^0] \right) \right| \\ &\leq K(1 + \|X_t\|^2). \end{aligned}$$

□

We are left to formulate sufficient conditions for the PMP to be satisfied on the desired state space. However, in contrast to Proposition 3.3.3, we are now in a multidimensional but time-homogeneous setting. The reasoning is the same as the one applied in Proposition 3.3.3. We nevertheless briefly summarize sufficient conditions for the multidimensional setting in the next lemma for completeness.

Lemma 3.4.25. Let us consider a state space $S \subseteq \mathbb{R}^d$ such that $S = S^{(1)} \otimes S^{(2)} \otimes \dots \otimes S^{(d)}$ with $S^{(1)}, \dots, S^{(d)} \in \{\mathbb{R}, \mathbb{R}_{\geq 0}\}$ for a multidimensional SDE driven by $m \geq 1$ independent Brownian motions $(W_t)_{t \geq 0} = (W_t^1, \dots, W_t^m)_{t \geq 0}$

$$dX_t = B(X_t)dt + \Sigma(X_t)dW_t \quad (3.4.19)$$

where $B : \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\Sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m}$. This SDE has a generator given by

$$(\mathcal{H}f)(x) = \sum_{i=1}^d B(x)_i \partial_i f(x) + \frac{1}{2} \sum_{i,j=1}^d (\Sigma(x)\Sigma(x)^\top)_{ij} \partial_{ij} f(x)$$

for $f \in C_c^\infty(\mathbb{R}^d)$. Let $x_0 \in S$ be such that $f(x_0) = \sup_{x \in S} f(x)$. Then $\mathcal{H}$ satisfies the positive maximum principle on $S = \mathbb{R}^d$ directly. For all other cases, we need some additional conditions on the boundary. Let us denote the set $I(x_0) := \{i \in \{1, \dots, d\} | (x_0)_i = 0 \text{ and } S^{(i)} = \mathbb{R}_{\geq 0}\}$. If $I(x_0) \neq \emptyset$, i.e. at least for one coordinate the maximum lies on the boundary, we require for all $i \in I(x_0)$ and $j \in \{1, \dots, d\}$

$$(\Sigma(x_0)\Sigma(x_0)^\top)_{ij} = 0 \text{ and } B(x_0)_i \geq 0.$$

Proof. The reasoning is completely analogous to the proof of Proposition 3.3.3. Let us first treat the case that x_0 is in the interior of S . In this case we know by the definition of $f(x_0)$ being the maximum that

$$\partial_i f(x_0) = 0 \text{ for all } i \in \{1, \dots, d\}$$

$$(\partial_{ij} f(x_0))_{1 \leq i, j \leq d} \text{ is negative semidefinite.}$$

Hence, we can easily verify that $(\mathcal{H}f)(x_0) \leq 0$ in this case, noting that

$$\begin{aligned} (\mathcal{H}f)(x_0) &= \frac{1}{2} \sum_{i,j=1}^d (\Sigma(x_0)\Sigma(x_0)^\top)_{ij} \partial_{ij} f(x_0) = \frac{1}{2} \sum_{i,j,k=1}^d \Sigma(x_0)_{ik} \Sigma(x_0)_{kj}^\top \partial_{ij} f(x_0) \\ &= \frac{1}{2} \sum_{i,j,k=1}^d \Sigma(x_0)_{kj}^\top (\partial_{ji} f(x_0)) \Sigma(x_0)_{ik} \leq 0 \end{aligned}$$

by the negative semidefiniteness of the Hessian. Now for the cases that in some coordinates (potentially all) the maximum lies on the boundary, let us treat separately the sets of indices $I(x_0)$ and $\{1, \dots, d\} \setminus I(x_0)$. For all $i \in I(x_0)$ it holds that $\partial_i f(x_0) \leq 0$ and by the requirements of the theorem we find

$$\begin{aligned} \sum_{i \in I(x_0)} B(x_0)_i \partial_i f(x_0) &\leq 0 \\ \sum_{\substack{i \in I(x_0) \\ j \in \{1, \dots, d\}}} (\Sigma(x_0)\Sigma(x_0)^\top)_{ij} \partial_{ij} f(x_0) &+ \sum_{\substack{j \in I(x_0) \\ i \in \{1, \dots, d\}}} (\Sigma(x_0)\Sigma(x_0)^\top)_{ij} \partial_{ij} f(x_0) = 0, \end{aligned}$$

where the second sum follows by symmetry.

On the other hand, $\partial_i f(x_0) = 0$ for all $i \notin I(x_0)$ and $(\partial_{ij} f(x_0))_{i,j \notin I(x_0)}$ needs to be negative semidefinite, as $f(x_0)$ is concave in these coordinates. Hence, it follows as above that

$$\sum_{i \notin I(x_0)} B(x_0)_i \partial_i f(x_0) + \frac{1}{2} \sum_{i,j \notin I(x_0)} (\Sigma(x_0) \Sigma(x_0)^\top)_{ij} \partial_{ij} f(x_0) \leq 0$$

and we have proven the claim. $\square$

We now outline possible state-spaces and sufficient conditions for the PMP to be satisfied in the case of Proposition 3.4.21, 3.4.22 and Theorem 3.4.24, respectively. Note that what we mean by "state-spaces" here is slightly different than what is sometimes referred to as the state-space of a Markov process in the literature. We mean that the process takes values in the respective space and will not leave the state-space in finite time. However, the process will not explore the entire space, and it is also not possible for the process to start at every point in the space. Namely, if $Z_0 = z \in \mathbb{R}$ the process $(Z_t, \mathbb{E}[Z_t | \mathcal{W}_t^0], \mathbb{E}[Z_t^2 | \mathcal{W}_t^0])_{t \geq 0}$ can only start in points like (z, z, z^2) . Keeping this in mind, we are now ready to state the following theorems:

Proposition 3.4.26 (State-spaces and PMP for Proposition 3.4.21 & 3.4.22). We consider the state-space $S = \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}$ and $S = \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}$, respectively. In both cases, the PMP holds if $b, \gamma, \Gamma_0 \geq 0$ and $l_0 = c = 0$. For the case $S = \mathbb{R} \times \mathbb{R}$ (Proposition 3.4.22) we require $c_0, c_2, \Gamma_0 \geq 0$ and $\gamma, c_1 = 0$. And in addition to that, it must hold $l_0 = b_0 = 0$ for the PMP to be satisfied on $\mathbb{R} \times \mathbb{R} \times \mathbb{R}_{\geq 0}$ in the context of Proposition 3.4.21.

Proof. Let us recall the statements of Lemma 3.4.25. We can then treat the individual cases separately:

- We start with Proposition 3.4.21 and $S = \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}$. In this case we require for $x_0 \in \mathbb{R}_{> 0} \times \mathbb{R}_{> 0} \times \mathbb{R}_{> 0}$ that $\Sigma(x_0)$ is real-valued, i.e. the term under the square-root to be non-negative, which is indeed the case for $\gamma_0, \Gamma_0 \geq 0$ and $l_0 = c = 0$. Next we check the case that the boundary is attained in at least one coordinate, i.e. $I(x_0)$. It is easy to see that the relevant drift terms indeed point inwards and that the diffusion coefficients vanish at these points, if $b, \gamma, \Gamma_0 \geq 0$ and $l_0 = c = 0$.
- In the case of Proposition 3.4.21 and $S = \mathbb{R} \times \mathbb{R} \times \mathbb{R}_{\geq 0}$. We first check the interior, i.e. $x_0 \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}_{> 0}$ where again the only requirement is for the term under the square-root to be non-negative, which indeed holds if $c_1 = \gamma_0 = 0$ and $\Gamma_0, c_0, c_2 \geq 0$. In this case, the boundary can only be attained in the last coordinate, i.e. $I(x_0) = \{2\}$. Indeed, $B(x_0)_2 \geq 0$ if $b_0 = \gamma_0 = l_0 = c_1 = 0$ and $c_0 \geq 0$ and $\Sigma(x_0)_{22} = 0$ if $l_0 = 0$.
- Turning to Proposition 3.4.22 and the case $S = \mathbb{R} \times \mathbb{R}$ the term under the square-root is indeed positive if $c_1 = \gamma = 0$ and $c_0, c_2, \Gamma_0 \geq 0$. In this case there is no boundary to be checked.

- Lastly, in the case of Proposition 3.4.22 and $S = \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}$ the term under the square-root is non-negative on all of $\mathbb{R}_{> 0} \times \mathbb{R}_{> 0}$ if $c, \gamma, \Gamma_0 \geq 0$. If $0 \in I(x_0)$ it holds that $B(x_0)_0 \geq 0$ if $b \geq 0$ and the corresponding diffusion terms vanish if $l_0 = c = 0$. Alternatively, if $1 \in I(x_0)$, the drift $B(x_0)_1 \geq 0$ if $b_0 \geq 0$ and the diffusion vanishes if $l_0 = 0$.

□

Remark 3.4.27. Let us point out the connection between the above statements and (Filipović and Larsson, 2016, Proposition 6.4). In contrast to our considerations, (Filipović and Larsson, 2016, Proposition 6.4) prohibits absorption of the process by the boundary, while we allow the process to remain at the boundary. This difference leads to the requirement in (Filipović and Larsson, 2016, Proposition 6.4) that the drift remains strictly positive on the boundary. We are able to accommodate this for the case $S = \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}$ by enforcing $b, \beta > 0$. If we allow for absorption at the boundary and therefore slightly weaken the conditions of (Filipović and Larsson, 2016, Proposition 6.4), the conditions we found in Proposition 3.4.26 are indeed also sufficient conditions for (Filipović and Larsson, 2016, Proposition 6.4).

Proposition 3.4.28 (State-spaces and PMP for Theorem 3.4.24). First, we consider the state-space $S = \mathbb{R}_{\geq 0} \times \cdots \times \mathbb{R}_{\geq 0}$ on which the PMP holds if $l \equiv c \equiv 0$, $b, \gamma, \Gamma, \Lambda \geq 0$. On the other hand, for the state-space $S = \mathbb{R} \times \mathbb{R} \times \mathbb{R}_{\geq 0} \times \cdots \times \mathbb{R} \times \mathbb{R}_{\geq 0}$ alternating or (or $S = \mathbb{R} \times \mathbb{R} \times \mathbb{R}_{\geq 0} \times \cdots \times \mathbb{R} \times \mathbb{R}_{\geq 0} \times \mathbb{R}$ alternating) sufficient conditions for the PMP to hold are $\gamma \equiv 0$, $\Gamma, c \geq 0$ and for all $x_0 \in S$ it must hold that $c(x_0) = l(x_0) = 0$, $b(x_0) \geq 0$. Moreover, if $b, \beta, c, \gamma, \Gamma, l, \Lambda$ are locally Lipschitz-continuous, then the solutions are path-wise unique.

Proof. We first study the case $S = \mathbb{R}_{\geq 0} \times \cdots \times \mathbb{R}_{\geq 0}$ and start with the interior, i.e. $x_0 \in \mathbb{R}_{> 0} \times \cdots \times \mathbb{R}_{> 0}$. Recalling (3.4.17), the term under the square-root is positive if $c, \gamma, \Gamma \geq 0$. Now turning to the boundary, let us consider $0 \in I(x_0)$. Clearly, $B(x_0)_0 > 0$ if $b \geq 0$ and the diffusion terms vanish if $l \equiv c \equiv 0$. For the cases $k \in I(x_0)$ for any $k \geq 1$, let us recall Lemma 3.4.20. Then $B(x_0)_k \geq 0$ if $b, \gamma, l, \Lambda \geq 0$ and $\Sigma(x_0)_{k2} = 0$ if $l \equiv 0$.

Now for alternating state-spaces $S = \mathbb{R} \times \mathbb{R} \times \mathbb{R}_{\geq 0} \times \cdots \times \mathbb{R} \times \mathbb{R}_{\geq 0}$ the interior is $S^\circ = \mathbb{R} \times \mathbb{R} \times \mathbb{R}_{> 0} \times \cdots \times \mathbb{R} \times \mathbb{R}_{> 0}$ which leads us to require $\gamma \equiv 0$ and $c, \Gamma \geq 0$ for the term under the square-root to be non-negative for all $x_0 \in S$. Moreover, if some $k \in I(x_0)$ (where k is even), we find $B(x_0)_k \geq 0$ if $\gamma(x_0) = c(x_0) = l(x_0) = 0$ and $b(x_0) \geq 0$ for all $x_0 \in \partial S$, whereas the diffusion term $\Sigma(x_0)_{k2} = 0$ if $l(x_0) = 0$ for all $x_0 \in \partial S$.

We now turn to the claim of path-wise uniqueness. Let us recall the form of the SDE describing $(\mathbb{E}[Z_t | \mathcal{W}_t^0])_{t \geq 0}$ as stated in Lemma 3.4.20. It is clear that if the maps $b, \beta, c, \gamma, \Gamma, l, \Lambda$ are locally Lipschitz-continuous, so are the associated drift and volatility terms and hence (Ethier and Kurtz, 2005, Theorem 3.7) is applicable.

Hence, the obtained path-wise unique solution of $(\mathbb{E}[Z_t | \mathcal{W}_t^0])_{t \geq 0}$ can be plugged into the original SDE for $(Z_t)_{t \geq 0}$. While the volatility coefficient is not locally Lipschitz-continuous due to the square-root, this is now a one-dimensional SDE, and we can apply (Ethier and Kurtz, 2005, Theorem 3.8) to conclude path-wise uniqueness. □

3.5 Affine McKean-Vlasov equations

3.5.1 Outline of the affine case

Let us recall that for time-inhomogeneous affine process $(Z_t)_{t \geq 0}$ there exist maps ϑ, ϕ, ψ such that

$$\vartheta(T, t, u) = \mathbb{E}[e^{uZ_T} | \mathcal{F}_t] = e^{\phi(T, t, u) + Z_t \psi(T, t, u)} \text{ for any } u \in \mathcal{U}, \quad (3.5.1)$$

where the set $\mathcal{U}$ depends on the state-space of $(Z_t)_{t \geq 0}$ and is specified below.

Let us now introduce what we call an *affine McKean-Vlasov SDE*:

Definition 3.5.1 (Affine McKean-Vlasov SDE). We call a time-inhomogeneous affine process of the form

$$dZ_t = (b(t) + \beta(t)Z_t)dt + \sqrt{c(t) + \gamma(t)Z_t}dW_t \quad (3.5.2)$$

where $(W_t)_{t \geq 0}$ of McKean-Vlasov type if the maps b, c, β, γ in fact depend on the law of the process Z . Depending on the state-space S of the above process, the set of admissible parameters differ, recall Definition 3.2.17. Since we focus on the one-dimensional setting here, we distinguish the cases:

- state space $S = \mathbb{R}_{\geq 0}$: $c(t) \equiv 0$, $b(t) \geq 0$, $\gamma(t) \geq 0$ for all $t \geq 0$ and $\mathcal{U} = \{u \in \mathbb{C} \mid \operatorname{Re}(u) \leq 0\}$
- state space $S = \mathbb{R}$: $\gamma(t) \equiv 0$, $c(t) \geq 0$ for all $t \geq 0$ and $\mathcal{U} = i\mathbb{R}$.

Remark 3.5.2. The following results hold in both cases, i.e. for $S = \mathbb{R}$ or $S = \mathbb{R}_{\geq 0}$. However, we always assume the parameters are such that we are either in the case $S = \mathbb{R}$ or $S = \mathbb{R}_{\geq 0}$ and everything is well-defined. In particular, $c \neq 0$ and $\gamma \neq 0$ is not allowed simultaneously.

We would like to point out that this section is a collection of examples which can be treated in the affine setting. In contrast to the polynomial case, we are yet to develop a theory of the same scope as we did in Section 3.4. Let us start by giving some examples of how the coefficients of the SDE in (3.5.2) may depend on the law of the process:

Example 3.5.3. In the next subsections, we will study the following cases of affine McKean-Vlasov SDEs:

- in Subsection 3.5.2, we consider coefficients depending on the moments of the process, i.e. the maps b, β, c, γ are maps of the form

$$t \mapsto f(\mathbb{E}[Z_t^m]) \text{ for some } m \geq 1$$

where the specific function f and moment m may be different for b, β, γ respectively. ⁴ We would like to emphasize that this is a special case of the polynomial setting. However, due to the affine structure, it is additionally possible to obtain the characteristic function by solving the transport PDE. We will present several examples where this can be obtained analytically.

⁴It is in principle also possible to let each map depend on more than one moment, but we don't consider this case here for better readability.

- in Subsection 3.5.3, we allow the coefficients to depend on the density at some point $x \in S$ of Z_t which we denote by $\rho(t, x)$;
- and, in Subsection 3.5.4, we consider a fixed finite time horizon $T > 0$ and coefficients depending on the logarithm of conditional (with respect to $\mathcal{F}_t$) exponential moments, i.e. $\log(\mathbb{E}[e^{uZ_T} | \mathcal{F}_t])$ for some $u \in \mathbb{R}_-$.

We are now ready to provide a sketch of the proofs for the respective cases, which we will give in detail in the following subsections for each case respectively. However, the proofs of all cases follow the same basic procedure, which is similar to the dual point of view of Subsection 3.4.2 for the polynomial case:

- (i) First, we identify the maps b, β, γ with functions

$$\begin{aligned} b(t) &\equiv \tilde{b}(u' \mapsto v(t, u')), \quad \beta(t) \equiv \tilde{\beta}(u' \mapsto v(t, u')), \\ c(t) &\equiv \tilde{c}(u' \mapsto v(t, u')), \quad \gamma(t) \equiv \tilde{\gamma}(u' \mapsto v(t, u')) \end{aligned}$$

where the map $u' \mapsto v(t, u')$ itself is the unique, global (in time) solution to some differential equation.

- (ii) We conclude that the resulting SDE is a time-inhomogeneous affine process and establish the existence of a unique solution to the SDE via results from the theory of (time-inhomogeneous) affine processes.
- (iii) So far we have not verified that the SDE proposed in step (i) is of McKean-Vlasov type. It is in this last step that we conclude $(u \mapsto v(t, u)) = (u \mapsto \vartheta(t, 0, u))$ (or $(u \mapsto v(t, u)) = (u \mapsto \vartheta(T, t, u))$ in the case of conditional moments). This conclusion follows from the fact that $u \mapsto v(t, u)$ was the *unique* solution to some differential equation. Of course, the differential equation was intentionally chosen to make this conclusion.

Without having seen the specific examples, this procedure might be a bit abstract, but we believe this general outline is helpful to capture the spirit of our approach. Before we finally study the different cases in detail, let us formulate the following lemma, which will be crucial in the proofs:

Lemma 3.5.4. Consider a time-inhomogeneous process

$$dZ_t = b(t) + \beta(t)Z(t) + \sqrt{c(t) + \gamma(t)Z(t)}dW_t \text{ with } Z_0 = z$$

with strongly admissible parameters $(b(t), \beta(t), c(t), \gamma(t))$. Then the function $\vartheta(t, 0, u) = \mathbb{E}[e^{uZ_t}]$ (for $u \in \mathcal{U}$) satisfies the PDE

$$\left(\frac{1}{2}c(t)u^2 + b(t)u\right) \vartheta(t, 0, u) + \left(\frac{1}{2}\gamma(t)u^2 + \beta(t)u\right) \partial_u \vartheta(t, 0, u) = \partial_t^+ \vartheta(t, 0, u) \quad (3.5.3)$$

with initial conditions $\vartheta(t, 0, 0) = 1$ and $\vartheta(0, 0, u) = e^{uz}$. We will sometimes refer to (3.5.3) as *transport equation* or *transport PDE* in the following.

Proof. By Proposition 3.2.19 we know that for strongly admissible parameters $(a(t), \alpha(t), \sigma(t))$ there exists a unique strongly regular affine Markov process with transition kernels $(P_{t,T})_{0 \leq t \leq T}$ and infinitesimal generators $(\mathcal{H}_s)_{s \geq 0}$. Moreover, the corresponding space-time process is Feller and therefore from (Filipović, 2005, Equation (6.5)) and (Duffie et al., 2003, Proposition 8.2) we know that for all $\theta \in C_c^\infty([0, \infty))$ and $f_u : S \ni x \mapsto e^{xu} \in \mathbb{C}$ it holds that

$$\begin{aligned} & \lim_{h \searrow 0} \frac{\theta(t+h)P_{t,t+h}f_u(x) - \theta(t)f_u(x)}{h} \\ &= \frac{(\theta(t+h) - \theta(t))P_{t,t+h}f_u(x) + \theta(t)(P_{t,t+h}f_u(x) - f_u(x))}{h} \\ &= \partial_t^+ \theta(t)f_u(x) + \theta(t)\partial_T^+(P_{t,T}f_u(x))|_{T=t} = \partial_t^+ \theta(t)f_u(x) + \theta(t)\mathcal{H}_t f_u(x) \end{aligned}$$

where the Feller property of the space-time process was used in the last equality. Hence, we conclude

$$\begin{aligned} \partial_T^+(P_{t,T}f_u(x))|_{T=t} &= \mathcal{H}_t f_u(x) \\ \text{for } \mathcal{H}_t f_u(x) &= \frac{1}{2}(c(t) + \gamma(t)x)\partial_x^2 f_u(x) + (b(t) + \beta(t)x)\partial_x f_u(x). \end{aligned}$$

From this, we can derive the Kolmogorov forward equation for all $0 \leq s \leq t$

$$\partial_t^+ P_{s,t}f_u(x) = P_{s,t}\mathcal{H}_t f_u(x)$$

see (Rüschendorf et al., 2016, Lemma 2.1) for a derivation. In order to compute the right-hand-side of the Kolmogorov forward equation, let us note that

$$\begin{aligned} \mathcal{H}_t f_u(x) &= \frac{1}{2}c(t)u^2 e^{ux} + \frac{1}{2}\gamma(t)xu^2 e^{ux} + b(t)ue^{ux} + \beta(t)uxe^{ux} \\ &= \frac{1}{2}c(t)u^2 e^{ux} + \frac{1}{2}\gamma(t)u^2 \partial_u e^{ux} + b(t)ue^{ux} + \beta(t)u\partial_u e^{ux} \end{aligned}$$

and

$$P_{0,t}(\partial_u e^{ux}) = \mathbb{E}[\partial_u e^{uZ_t}] = \mathbb{E}[Z_t e^{uZ_t}].$$

Since the action of $P_{0,t}$ on e^{ux} is well-understood for affine processes, we would like to exchange the derivative and expectation. Indeed, this is possible due to Leibniz' rule and the fact that for each $t \geq 0$, $\sup_{u \in \mathcal{U}} |Z_t e^{uZ_t}| \leq |Z_t|$. Let us make the last inequality clear for the specific sets $\mathcal{U} = i\mathbb{R}$ and $\mathcal{U} = \{z + iv \mid z \in \mathbb{R}_{\leq 0}, v \in R\}$ respectively. In the case $\mathcal{U} = i\mathbb{R}$, clearly $|e^{uZ_t}| \leq 1$ and the inequality follows as claimed. On the other hand, for $\mathcal{U} = \{z + iv \mid z \in \mathbb{R}_{\leq 0}, v \in R\}$ we find $|e^{uZ_t}| = |e^{(z+iv)Z_t}| = |e^{zZ_t}| \cdot |e^{ivZ_t}| \leq 1$. Since $|Z_t| \leq 1 + Z_t^2$ and since in the diffusion case $(Z_t)_{t \geq 0}$ is also a polynomial process it holds that $\mathbb{E}[|Z_t|] = 1 + \mathbb{E}[Z_t^2] < \infty$ for all $t \geq 0$ and hence by dominated convergence the expectation and differentiation can be exchanged and $\vartheta(t, 0, u)$ indeed satisfies the transport PDE (3.5.3). $\square$

3.5.2 Moments

In this subsection, we study affine McKean-Vlasov SDEs where the coefficients depend on the moments of the process, i.e.

$$dZ_t = \left(\tilde{b}(\mathbb{E}[Z_t^{m_1}]) + \tilde{\beta}(\mathbb{E}[Z_t^{m_2}])Z_t \right) dt + \sqrt{\tilde{c}(\mathbb{E}[Z_t^{m_3}]) + \tilde{\gamma}(\mathbb{E}[Z_t^{m_4}])Z_t} dW_t$$

for some integers $m_1, m_2, m_3, m_4 \geq 0$ and functions $\tilde{b}, \tilde{\beta}, \tilde{\gamma}$. The main results of this subsection is the following proposition:

Proposition 3.5.5. Consider the SDE

$$dZ_t = \left(\tilde{b}(\mathbb{E}[Z_t^{m_1}]) + \tilde{\beta}(\mathbb{E}[Z_t^{m_2}])Z_t \right) dt + \sqrt{\tilde{c}(\mathbb{E}[Z_t^{m_3}]) + \tilde{\gamma}(\mathbb{E}[Z_t^{m_4}])Z_t} dW_t \quad (3.5.4)$$

with $Z_0 = z \in D$, $\tilde{b}, \tilde{\beta}, \tilde{c}, \tilde{\gamma}$ some continuous functions and either

- $S = \mathbb{R}_{\geq 0}$, $\tilde{c} \equiv 0$ and $\tilde{b}, \tilde{\gamma} : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$, or,
- $S = \mathbb{R}$, $\tilde{\gamma} \equiv 0$ and $\tilde{c} : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$.

A solution to (3.5.4) exists if $\tilde{b}, \tilde{\beta}, \tilde{c}, \tilde{\gamma}$ are such that the following system of ODEs admits a unique global solution $\mathbf{v}(\cdot) = (v_n(\cdot))_{n=0}^{\max(m_1, m_2, m_3, m_4)}$:

$$\begin{aligned} \partial_t (v_n(t)) = & n\tilde{\beta}(v_{m_2}(t))v_n(t) + \left(n\tilde{b}(v_{m_1}(t)) + \frac{n(n-1)}{2}\tilde{\gamma}(v_{m_4}(t)) \right) v_{n-1}(t)\mathbf{1}_{\{n \geq 2\}} \\ & + \frac{n(n-1)}{2}\tilde{c}(v_{m_3}(t))v_{n-2}(t)\mathbf{1}_{\{n \geq 2\}} \end{aligned} \quad (3.5.5)$$

for each $0 \leq n \leq \max(m_1, m_2, m_3, m_4)$ and where $v_n(0) = z^n$.

Moreover, the solution to (3.5.4) is a time inhomogeneous affine and polynomial process. Thanks to the affine nature of the solution, its characteristic function can be obtained by solving the transport PDE.

Proof. Before we study (3.5.4) directly, let us first study a different SDE, namely:

$$dZ_t = \left(\tilde{b}(v_{m_1}(t)) + \tilde{\beta}(v_{m_2}(t))Z_t \right) dt + \sqrt{\tilde{c}(v_{m_3}(t)) + \tilde{\gamma}(v_{m_4}(t))Z_t} dW_t \quad (3.5.6)$$

with $Z_0 = z$ and where the process $\mathbf{v}(\cdot) = (v_n(\cdot))_{n=0}^{\max(m_1, m_2, m_3, m_4)}$, is the unique global solution to (3.5.5). Note that $(\mathbf{v}(t))_{t \geq 0}$ is continuous in time, and hence either

$$(\tilde{b}(v_{m_1}(t)), \tilde{\beta}(v_{m_2}(t)), \tilde{\gamma}(v_{m_4}(t)))_{t \geq 0}$$

or

$$(\tilde{b}(v_{m_1}(t)), \tilde{\beta}(v_{m_2}(t)), \tilde{c}(v_{m_3}(t)))_{t \geq 0}$$

is a set of strongly admissible parameters. Hence, we know that a unique solution to (3.5.6) exists, and it is a time-inhomogeneous affine process, see (Filipović, 2005, Theorem 2.13). Having established the existence of a solution to (3.5.6) we are now left to make the connection to (3.5.4). To this end, we will show that $v_n(\cdot) = \mathbb{E}[Z^n]$ for each $n \geq 1$. Recall from Lemma 3.5.4 that $\vartheta(t, 0, u) = \mathbb{E}[e^{uZ_t}]$ satisfies (3.5.3). It is straightforward to show by induction that $(\partial_u^n \vartheta(t, 0, 0))_{n \geq 1}$ satisfies the system of ODEs

$$\begin{aligned} \partial_t (\partial_u^n \vartheta(t, 0, 0)) &= n\tilde{\beta}(v_{m_2}(t)) \partial_u^n \vartheta(t, 0, 0) \\ &\quad + \left(n\tilde{b}(v_{m_1}(t)) + \frac{n(n-1)}{2} \tilde{\gamma}(v_{m_4}(t)) \right) \partial_u^{n-1} \vartheta(t, 0, 0) \mathbf{1}_{\{n \geq 2\}} \\ &\quad + \frac{n(n-1)}{2} \tilde{c}(v_{m_3}(t)) \partial_u^{n-2} \vartheta(t, 0, 0) \mathbf{1}_{\{n \geq 2\}} \end{aligned} \quad (3.5.7)$$

for each $1 \leq n \leq \max(m_1, m_2, m_3, m_4)$, $\partial_u^n \vartheta(0, 0, 0) = z^n$ and $\partial_u^0 \vartheta(t, 0, 0) = \vartheta(t, 0, 0)$.

Note that $\partial_u^n \vartheta(t, 0, 0)$ indeed exists for all $n \geq 0$ by the fact that the every affine process is also a polynomial process and hence all moments of $(Z_t)_{t \geq 0}$ are finite if Z_0 is deterministic. However, by assumption that $\mathbf{v}(\cdot)$ is the unique solution to (3.5.5) it follows that $v_m(t) = \partial_u^m \vartheta(t, 0, 0)$ for all $m \in \{m_1, m_2, m_3, m_4\}$ and $t \geq 0$. Moreover, note that $\partial_u^m \vartheta(t, 0, 0) = \mathbb{E}[Z_t^m]$ and hence $v_m(t) = \mathbb{E}[Z_t^m]$ for all $m \in \{m_1, m_2, m_3, m_4\}$ and $t \geq 0$. This proves the claim. $\square$

Let us remark that this system of ODEs is of course a particular instance of (3.4.8) (with $\Gamma \equiv 0$) which we obtained in the polynomial case. This is not surprising because affine diffusions are also polynomial. In the affine case we can obtain the system of ODEs via the transport equation and do not need to make any growth-assumptions on the coefficients as in the primal viewpoint for polynomial processes. Again, what we mean by that is similar to the discussion in Remark 3.4.17, i.e. the growth-assumptions are not required to *derive* the system of ODEs here. However, the requirement of existence of a unique global solution implicitly imposes assumptions on the maps b, β, c, γ . See also Example 3.5.8 below. Moreover, in the affine case, we can obtain the characteristic function as a solution to the transport PDE.

Remark 3.5.6. Note that in general, the system of ODEs (3.5.5) is coupled. However, in the case $m_1 = m_2 = 1$ and $m_3, m_4 \leq 2$ it simplifies to

$$\begin{aligned} \partial_t(v_1(t)) &= \tilde{b}(v_1(t)) + \tilde{\beta}(v_1(t)) v_1(t) \\ \partial_t(v_2(t)) &= 2\tilde{b}(v_1(t)) v_1(t) + 2\tilde{\beta}(v_1(t)) v_2(t) + \tilde{\gamma}(v_{m_4}(t)) v_1(t) + \tilde{c}(v_{m_3}(t)) \end{aligned}$$

Remark 3.5.7. For better readability, we present the case where b, β, c, γ depend on one moment only. However, it is straightforward to extend this to dependencies on multiple moments.

Let us now make a few examples of SDEs which fall under Proposition 3.5.5.

Example 3.5.8. Our goal is to establish the existence of a unique solution to the following SDE:

$$dZ_t = \sqrt{\mathbb{E}[Z_t^2]} Z_t dW_t \text{ with } Z_0 = z > 0. \quad (3.5.8)$$

Note that the existence of a solution to this SDE can be established using Proposition 3.4.15, however in the affine setting we additionally obtain an analytic expression for the characteristic function $\vartheta(t, 0, u)$ of the solution. To this end, set $\tilde{b} \equiv 0$, $\tilde{\beta} \equiv 0$ and $\tilde{\gamma}(v_2(t)) = v_2(t)$. Then the system of ODEs (3.5.5) reduces to

$$\begin{aligned}\partial_t(v_1(t)) &= 0 \text{ with } v_1(0) = z \\ \partial_t(v_2(t)) &= v_2(t) \cdot v_1(t) \text{ with } v_2(0) = z_2 \\ \Rightarrow v_2(t) &= z^2 e^{zt}.\end{aligned}$$

And hence by Proposition 3.5.5 we have already established existence of a solution to (3.5.8). As a by-product, we have furthermore obtained $\mathbb{E}[Z_t^2] = z^2 e^{zt}$. Additionally, we can explicitly solve the transport equation (3.5.3) by the methods of characteristics, that is we parameterize the solution surface and thereby reduce the PDE to a system of ODEs. The associated characteristic equation of (3.5.3) are

$$\frac{d}{d\tau} t(\tau) = -1 \text{ with } t(0) = 0 \quad (3.5.9)$$

$$\frac{d}{d\tau} u(\tau) = \frac{1}{2} \gamma(\tau) u(\tau)^2 + \beta(\tau) u(\tau) \text{ with } u(0) = s \quad (3.5.10)$$

$$\frac{d}{d\tau} \vartheta(t(\tau), 0, u(\tau)) = b(\tau) u(\tau) \vartheta(t(\tau), 0, u(\tau)) \text{ with } \vartheta(0, 0, u(0)) = e^{sz}. \quad (3.5.11)$$

Assuming b, β, γ are at least Lipschitz, a unique solution to the ODEs in (3.5.9) exists at least locally. By the fact that the transversality condition is satisfied by our PDE for any b, β, γ , and therefore the PDE has a unique solution near the initial curve. The existence of a global solution to (3.5.9) then leads to a unique global solution of our PDE. Plugging in $b \equiv 0$, $\beta \equiv 0$ and $\gamma(\tau) = z^2 e^{-z\tau}$ we obtain the characteristic equations of the transport equation associated to (3.5.8). We obtain

$$\vartheta(t, 0, u) = \exp\left(\frac{zu}{1 - \frac{1}{2}uz(e^{zt} - 1)}\right) \text{ for } t \geq 0, u \in \mathbb{C}_-.$$

Example 3.5.9. Let us now consider the SDE:

$$dZ_t = -\mathbb{E}[Z_t]Z_t dt + \sqrt{\mathbb{E}[Z_t]} dW_t \text{ with } Z_0 = z > 0. \quad (3.5.12)$$

This SDE can again be treated using Proposition 3.5.5 but not using Proposition 3.4.15, since there β was required to be uniformly bounded. Concretely, we set $m_2 = m_3 = 1$, $\tilde{b} \equiv 0$, $\tilde{\beta}(v_1(t)) = -v_1(t)$ and $\tilde{\gamma}(v_1(t)) = v_1(t)$. It suffices to solve the ODE

$$\partial_t(v_1(t)) = -(v_1(t))^2 \text{ with } v_1(0) = z$$

and the solution to that is $v_1(t) = \frac{1}{t+z^{-1}}$ which is indeed global in time. Again, using the method of characteristics, we obtain the following solution to the transport equation (3.5.3):

$$\vartheta(t, 0, u) = \exp\left(\frac{u}{t + z^{-1} - tu}\right) \text{ for } t \geq 0, u \in \mathbb{C}_-.$$

Example 3.5.10. As already established in Proposition 3.4.15, when $\tilde{\gamma}$ only depends on the first moment, we can choose a rather "aggressive" form, e.g. $\tilde{\gamma}(\cdot) = \exp(\cdot)$. Let's consider for example the following SDEs:

- $dZ_t = \mu dt + \sqrt{\exp(\mathbb{E}[Z_t])} Z_t dW_t$ with $Z_0 = z \geq 0$ and constant $\mu > 0$, where we apply Proposition 3.5.5 with $\tilde{b} \equiv \mu$ and $\tilde{\gamma} = \exp(v_1(t))$ to find $\mathbb{E}[Z_t] = z + \mu t$.
- $dZ_t = \sqrt{\exp(\mathbb{E}[Z_t])} Z_t dW_t$ with $Z_0 = z > 0$ and we use again Proposition 3.5.5, this time with $\tilde{b} \equiv 0$ and $\tilde{\gamma} = \exp(v_1(t))$ to find $\mathbb{E}[Z_t] = z$.

Applying the same reasoning as in Proposition 3.5.5, we can make the following statement about local existence of solutions:

Corollary 3.5.11. *The SDE (3.5.4) admits a local solution on the interval $t \in [0, T_\infty)$ if a unique local solution to (3.5.5) exists with explosion time T_∞ .*

Example 3.5.12. For example, the SDE

$$dZ_t = \sqrt{\exp(\mathbb{E}[Z_t^2])} Z_t dW_t \text{ with } Z_0 = z > 0$$

admits a local solution up to $T_\infty = z^{-1}e^{-z^2}$. Moreover, $(Z_t)_{t \in [0, t_0]}$ is a true martingale for all $t_0 < T_\infty$.

3.5.3 Density

Let us consider an S -valued stochastic process $(Z_t)_{t \geq 0}$ and assume it admits a probability density. We denote for each $t \geq 0, x \in S$ by $\rho(t, x)$ the probability density of Z_t at point x . In this subsection, we study affine McKean-Vlasov SDEs whose coefficients depend on $\rho(\cdot, x)$ for an arbitrary but fixed point $x \in S$.

Theorem 3.5.13. *For state-spaces $S = \mathbb{R}$ or $S = \mathbb{R}_{\geq 0}$ let us consider the SDE:*

$$dZ_t = (\tilde{b}(\rho(t, x)) + \tilde{\beta}(\rho(t, x))Z_t) dt + \sqrt{\tilde{c}(\rho(t, x)) + \tilde{\gamma}(\rho(t, x))Z_t} dW_t \quad (3.5.13)$$

with initial density $\rho(0, x) = \rho_0(x)$ and $\tilde{b}, \tilde{\beta}, \tilde{c}, \tilde{\gamma}$ continuous maps satisfying the strong admissibility conditions. There exists a unique solution to (3.5.13) if there exists a global unique solution to the following PDE:

$$\partial_t v(t, u) = \left(\frac{1}{2} \tilde{c} \left(\int_{i\mathbb{R}} e^{ux} v(t, u) du \right) u^2 + \tilde{b} \left(\int_{i\mathbb{R}} e^{ux} v(t, u) du \right) u \right) v(t, u) \quad (3.5.14)$$

$$+ \left(\frac{1}{2} \tilde{\gamma} \left(\int_{i\mathbb{R}} e^{ux} v(t, u) du \right) u^2 + \tilde{\beta} \left(\int_{i\mathbb{R}} e^{ux} v(t, u) du \right) u \right) \partial_u v(t, u) \quad (3.5.15)$$

Moreover, if a solution $(Z_t)_{t \geq 0}$ exists to (3.5.13), then it is a time-inhomogeneous affine process.

Proof. We follow our usual procedure in the dual approach to first establish existence of a solution to an associated SDE which we later conclude coincides with (3.5.13). That is, we start by studying the SDE

$$dZ_t = \left(\tilde{b} \left(\int_{i\mathbb{R}} e^{ux} v(t, u) du \right) + \tilde{\beta} \left(\int_{i\mathbb{R}} e^{ux} v(t, u) du \right) Z_t \right) dt \quad (3.5.16)$$

$$+ \sqrt{\tilde{c} \left(\int_{i\mathbb{R}} e^{ux} v(t, u) du \right) + \tilde{\gamma} \left(\int_{i\mathbb{R}} e^{ux} v(t, u) du \right) Z_t} dW_t \quad (3.5.17)$$

with $x \in S$ and $v(\cdot, \cdot)$ being the unique global solution to (3.5.14). Note however that (3.5.14) coincides with the transport equation of (3.5.16) and by uniqueness of solution we conclude $v(t, u) \equiv \vartheta(t, 0, u)$ for all $t \geq 0$ and $u \in \mathcal{U}$. Moreover, recall that the density is given by the Fourier transform of the characteristic function, i.e. $\rho(t, x) = \int_{i\mathbb{R}} e^{ux} \vartheta(t, 0, u) du$. Hence, the SDEs (3.5.16) and (3.5.13) indeed coincide. $\square$

Remark 3.5.14. The type of PDE appearing in (3.5.14) is called non-local PDE. Studying existence and uniqueness of solutions to such PDEs is beyond the scope of this chapter. However, we would like to point out the connection to a subclass of non-local PDEs, so-called non-local balance laws appearing in traffic flow models, see e.g. Keimer and Pflug (2017). Indeed, our PDE in (3.5.14) can be brought to the form of a non-local balance law for example by setting $\tilde{\gamma} \equiv \tilde{b}$, $\tilde{c} \equiv \tilde{\beta} \equiv 0$ and assuming the required regularity conditions hold.

3.5.4 Logarithm of conditional exponential moments

Let us now turn to a particular family of SDEs whose coefficients depend on the conditional expectation of its solution. Although it is not obvious as first glance that this SDE falls into the setting of affine process, it turns out that it can be fully treated using the theory of affine processes.

Proposition 3.5.15. For a finite time-horizon $T > 0$ the SDE

$$dZ_t = \beta \log(\mathbb{E}[e^{uZ_T} | \mathcal{F}_t]) dt + \sqrt{\gamma Z_t} dW_t \text{ with } Z_0 = z \in \mathbb{R}_{\geq 0} \quad (3.5.18)$$

admits a solution $(Z_t)_{0 \leq t \leq T}$ for any $\beta, \gamma \in \mathbb{R}_{\geq 0}$ and $u \in \mathbb{R}_-$. Moreover, the solution $(Z_t)_{0 \leq t \leq T}$ is a time-inhomogeneous affine process.

Proof. First, we study a different SDE than the one stated in the claim and will later conclude that they in fact coincide. To this end, let us consider the SDE

$$dY_t = \beta f(T, t, u) Y_t dt + \sqrt{\gamma Y_t} dW_t \text{ with } Y_0 = z \in \mathbb{R}_{\geq 0} \quad (3.5.19)$$

where we assume $(f(T, t, u))_{0 \leq t \leq T}$ to be the global unique solution to the backward ODE

$$\partial_t f(T, t, u) = -(\beta + \frac{\gamma}{2}) f^2(T, t, u) \text{ with } f(T, T, u) = u \in \mathbb{R}_-.$$

Clearly, $(Y_t)_{0 \leq t \leq T}$ is a time-inhomogeneous affine process, and it holds that

$$\mathbb{E}[e^{uY_T} | \mathcal{F}_t] = e^{\phi(T, t, u) + \psi(T, t, u) Y_t},$$

where $\phi(\cdot), \psi(\cdot)$ are solutions to the following Riccati-ODEs:

$$\begin{aligned}\partial_t \phi(T, t, u) &= 0 \text{ with } \phi(T, T, u) = 0 \\ \partial_t \psi(T, t, u) &= -\beta f(T, t, u) \psi(T, t, u) - \frac{\gamma}{2} \psi^2(T, t, u) \text{ with } \psi(T, T, u) = u.\end{aligned}$$

However, this means that $f(T, t, u) = \psi(T, t, u)$ for all $t \in [0, T]$ and moreover

$$\log(\mathbb{E}[e^{uY_T} | \mathcal{F}_t]) = \psi(T, t, u) Y_t = f(T, t, u) Y_t.$$

Hence (3.5.18) and (3.5.19) actually coincide. □

Remark 3.5.16. This example is of course explicitly constructed to be of this form. In particular, we would like to point out that in case of a dependence on $\log(\mathbb{E}[e^{uZ_T} | \mathcal{F}_t])$ under the square-root is, one would not obtain the appropriate Riccati-ODE.

Appendix of Chapter 3

3.A ODE theory

Theorem 3.A.1 (Existence and Uniqueness of Solutions to non-linear ODEs, see ([Hirsch and Smale, 1974](#), page 171)). *Let us consider the following autonomous non-linear ODE:*

$$du_t = F(u_t)dt \quad u_0 \in A \subseteq \mathbb{R}^n$$

where $u : [0, t^*) \rightarrow \mathbb{R}^n$ and $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$. If F is locally Lipschitz-continuous on A , i.e. $F \in \text{Lip}(A; \mathbb{R}^n)$, then there exists a unique solution $u : [0, t^*) \rightarrow \mathbb{R}^n$. We call the time t^* the maximal time of existence which is defined as $t^* := \inf_{t \geq 0} \{u_t \in \partial A\}$ and $t^* = \infty$ if $u_t \in \partial A$ does not occur. In the latter case, we call a solution global.

3.B Proofs

Proof of Lemma 3.4.19. We recall the L^2 -convergence of the Itô integral as formulated e.g. in ([Karatzas and Shreve, 1998](#), Section 3.2, particulary Definition 2.9). That is, for any square-integrable $(H_t)_{t \geq 0}$ there exists a sequence of simple processes $((H_t^{(n)})_{t \geq 0})_{n \geq 0}$ such that

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\left(\int_0^t H_s^{(n)} dW_s - \int_0^t H_s dW_s \right)^2 \right] = 0.$$

Let us also note the simple fact that for any sequence $X^{(n)} \xrightarrow{L^2} X$ it also holds that $\mathbb{E}[X^{(n)} | \mathcal{M}] \xrightarrow{L^2} \mathbb{E}[X | \mathcal{M}]$ for any sub- σ -algebra $\mathcal{M}$ since

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E}[(\mathbb{E}[X^{(n)} | \mathcal{M}] - \mathbb{E}[X | \mathcal{M}])^2] &= \lim_{n \rightarrow \infty} \mathbb{E}[(\mathbb{E}[X^{(n)} - X | \mathcal{M}])^2] \\ &\leq \lim_{n \rightarrow \infty} \mathbb{E}[\mathbb{E}[(X^{(n)} - X)^2 | \mathcal{M}]] = \lim_{n \rightarrow \infty} \mathbb{E}[(X^{(n)} - X)^2] \end{aligned}$$

by Jensen's inequality and the tower property.

We now proceed to show the first statement. Clearly, for all $n > 0$ it holds that

$$\mathbb{E} \left[\int_0^t H_s^{(n)} dW_s \middle| \mathcal{W}_t^0 \right] = 0.$$

To see this, we use the tower property, i.e.

$$\begin{aligned}
\mathbb{E} \left[\int_0^t H_s^{(n)} dW_s \middle| \mathcal{W}_t^0 \right] &= \sum_{i=0}^{n-1} \mathbb{E} [H_{t_i}^{(n)} (W_{t \wedge t_{i+1}} - W_{t \wedge t_i}) | \mathcal{W}_t^0] \\
&= \sum_{i=0}^{n-1} \mathbb{E} \left[\mathbb{E} [H_{t_i}^{(n)} (W_{t \wedge t_{i+1}} - W_{t \wedge t_i}) | \sigma(W_{t_i}, W_t^0)] \middle| \mathcal{W}_t^0 \right] \\
&= \sum_{i=0}^{n-1} \mathbb{E} \left[H_{t_i}^{(n)} \underbrace{\mathbb{E} [(W_{t \wedge t_{i+1}} - W_{t \wedge t_i}) | \sigma(W_{t_i}, W_t^0)]}_{=0} \middle| \mathcal{W}_t^0 \right] = 0,
\end{aligned}$$

where $\sigma(W_{t_i}, W_t^0)$ denotes the sigma algebra generated by W_{t_i}, W_t^0 . Hence,

$$\mathbb{E} \left[\int_0^t H_s^{(n)} dW_s \middle| \mathcal{W}_t^0 \right] \xrightarrow{L^2} 0$$

and on the other hand

$$\mathbb{E} \left[\int_0^t H_s^{(n)} dW_s \middle| \mathcal{W}_t^0 \right] \xrightarrow{L^2} \mathbb{E} \left[\int_0^t H_s dW_s \middle| \mathcal{W}_t^0 \right]$$

and it follows $\mathbb{E} \left[\int_0^t H_s dW_s \middle| \mathcal{W}_t^0 \right] = 0$ almost surely.

Similarly, we show the second statement. Note that

$$\begin{aligned}
\mathbb{E} \left[\int_0^t H_s^{(n)} dW_s^0 \middle| \mathcal{W}_t^0 \right] &= \sum_{i=0}^{n-1} \mathbb{E} [H_{t_i}^{(n)} (W_{t \wedge t_{i+1}}^0 - W_{t \wedge t_i}^0) | \mathcal{W}_t^0] \\
&= \sum_{i=0}^{n-1} \mathbb{E} [H_{t_i}^{(n)} | \mathcal{W}_t^0] (W_{t \wedge t_{i+1}}^0 - W_{t \wedge t_i}^0) = \int_0^t \mathbb{E} [H_s^{(n)} | \mathcal{W}_t^0] dW_s^0
\end{aligned}$$

Moreover, for all $s, t \geq 0$ it obviously holds that $\mathbb{E} [H_s^{(n)} | \mathcal{W}_t^0] \xrightarrow{L^2} \mathbb{E} [H_s | \mathcal{W}_t^0]$ by assumption that $H^{(n)} \xrightarrow{L^2} H$. Hence, by the L^2 -convergence of the Itô integral

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\int_0^t H_s^{(n)} dW_s^0 \middle| \mathcal{W}_t^0 \right] = \lim_{n \rightarrow \infty} \int_0^t \mathbb{E} [H_s^{(n)} | \mathcal{W}_t^0] dW_s^0 \xrightarrow{L^2} \int_0^t \mathbb{E} [H_s | \mathcal{W}_t^0] dW_s^0.$$

On the other hand, we know that

$$\mathbb{E} \left[\int_0^t H_s^{(n)} dW_s^0 \middle| \mathcal{W}_t^0 \right] \xrightarrow{L^2} \mathbb{E} \left[\int_0^t H_s dW_s^0 \middle| \mathcal{W}_t^0 \right]$$

and thereby it follows that

$$\mathbb{E} \left[\int_0^t H_s dW_s^0 \middle| \mathcal{W}_t^0 \right] = \int_0^t \mathbb{E} [H_s | \mathcal{W}_t^0] dW_s^0$$

holds almost surely.

□

Bibliography

- E. Abi Jaber, C. Illand, and S. Li. Joint SPX-VIX calibration with Gaussian polynomial volatility models: deep pricing with quantization hints. *Preprint arXiv:2212.08297*, 2022a.
- E. Abi Jaber, C. Illand, and S. Li. The quintic Ornstein-Uhlenbeck volatility model that jointly calibrates SPX & VIX smiles. *Preprint arXiv:2212.10917*, 2022b.
- E. Akyildirim, M. Gambarara, J. Teichmann, and S. Zhou. Applications of signature methods to market anomaly detection. *Preprint arXiv:2201.02441*, 2022.
- E. Akyildirim, M. Gambarara, J. Teichmann, and S. Zhou. Randomized signature methods in optimal portfolio selection. *Preprint arxiv:2312.16448*, 2023.
- A. Allan, C. Cuchiero, C. Liu, and D. Prömel. Model-free portfolio theory: A rough path approach. *Math. Finance*, 33:709–765, 2023.
- H. Andrès, A. Boumezoued, and B. Jourdain. Signature-based validation of real-world economic scenarios. *ASTIN Bulletin*, 24:410–440, 2024.
- I. P. Arribas, C. Salvi, and L. Szpruch. Sig-sdes model for quantitative finance, 2020. In: *Proceedings of the First ACM International Conference on AI in Finance*.
- P. Bader, S. Blanes, and F. Casas. Computing the matrix exponential with an optimized taylor polynomial approximation. *Mathematics*, 7(12):1174, 2019.
- K. Bahlali, M. A. Mezerdi, and B. Mezerdi. Stability of McKean-Vlasov stochastic differential equations and applications. *Stochastics and Dynamics*, 20(1), 2020.
- J. Baldeaux and A. Badran. Consistent modelling of VIX and equity derivatives using a $3/2$ plus jumps model. *Applied Mathematical Finance*, 21(4):299–312, 2014.
- A. Banner and R. Ghomrasni. Local times of ranked continuous semimartingales. *Stochastic Process. Appl.*, 118:1244–1253, 2008.
- C. Bayer, P. Friz, and J. Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6):887–904, 2016.
- C. Bayer, B. Horvath, A. Muguruza, B. Stemper, and M. Tomas. On deep calibration of (rough) stochastic volatility models. *Preprint arXiv:1908.08806*, 2019.
- C. Bayer, P. Hager, S. Riedel, and J. Schoenmakers. Optimal stopping with signatures. *Ann. Appl. Probab.*, 33(1), 2021.

Bibliography

- F. E. Benth, N. Detering, and L. Galimberti. Neural networks in Fréchet spaces. *Ann. Math. Artif. Intell.*, 91:75–103, 2023.
- H. Boedihardjo, H. Ni, and Z. Qian. Uniqueness of signature for simple curves. *Journal of Functional Analysis*, 267(6):1778–1806, 2014.
- H. Boedihardjo, X. Geng, T. Lyons, and D. Yang. The signature of a rough path: Uniqueness. *Adv. Math.*, 293:720–737, 2016.
- H. Boedihardjo, J. Diehl, M. Mezzarobba, and H. Ni. The expected signature of Brownian motion stopped on the boundary of a circle has finite radius of convergence. *Bulletin of the London Mathematical Society*, 53(1):285–299, 2021.
- A. Bondi, G. Livieri, and S. Pulido. Affine volterra processes with jumps. *Stochastic Processes and their Applications*, 168:104264, 2024a.
- A. Bondi, S. Pulido, and S. Scotti. The rough Hawkes Heston stochastic volatility model. *Mathematical Finance*, 34(4):1197–1241, 2024b.
- F. Bourgey and S. De Marco. Multilevel Monte Carlo simulation for VIX options in the rough Bergomi model. *Journal of Computational Finance*, 26(2):53–82, 2022.
- H. Bühler, B. Horvath, T. Lyons, I. P. Arribas, and B. Wood. A data-driven market simulator for small data environments, 2020.
- S. Campbell and T.-K. L. Wong. Functional portfolio optimization in Stochastic Portfolio Theory. *SIAM J. Financial Math.*, 13(2):576–618, 2022.
- W. Cao, X. Wang, Z. Ming, and J. Gao. A review on neural networks with random weights. *Neurocomputing*, 275:278–287, 2018.
- R. Carmona, J.-P. Fouque, and L.-H. Sun. Mean field games and systemic risk. *Communications in Mathematical Sciences*, 13(4):911–933, 2015.
- T. Cass and E. Ferrucci. On the Wiener chaos expansion of the signature of a Gaussian process. *Probability Theory and Related Fields*, 189(3):909–947, 2024.
- L.-P. Chaintron and A. Diez. Propagation of chaos: A review of models, methods and applications. I. Models and methods. *Kinetic and Related Models*, 15(6):895–1015, 2022.
- J. Chang, N. Duffield, H. Ni, and W. Xu. Signature inversion for monotone paths. *Electronic Communications in Probability*, 22:1 – 11, 2017.
- K.-T. Chen. Integration of Paths, geometric invariants and a generalized Baker- Hausdorff formula. *Annals of Mathematics*, 65(1):163–178, 1957.
- K. T. Chen. Iterated path integrals. *Bulletin of the American Mathematical Society*, 83: 831–879, 1977.

- I. Chevyrev and H. Oberhauser. Signature moments to characterize laws of stochastic processes. *J. Mach. Learn. Res.*, 23(176):1–42, 2022.
- S. N. Cohen and R. J. Elliot. *Stochastic Calculus and Applications - Second Edition*. Springer, 2015.
- S. N. Cohen, S. Lui, W. Malpass, G. Mantoan, L. Nesheim, A. de Paula, A. Reeves, C. Scott, E. Small, and L. Yang. Nowcasting with signature methods. *Preprint arXiv:2305.10256*, 2023.
- E. M. Compagnoni, L. Biggio, A. Orvieto, T. Hofmann, and J. Teichmann. Randomized signature layers for signal extraction in time series data. *Preprint arxiv:2201.00384*, 2022.
- E. M. Compagnoni, A. Scampicchio, L. Biggio, A. Orvieto, T. Hofmann, and J. Teichmann. On the effectiveness of randomized signatures as reservoir for learning rough dynamics. In *2023 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2023.
- R. Cont and P. Das. Quadratic variation and quadratic roughness. *Bernoulli*, 29(1):496–522, 2023.
- R. Cont and D. Fournie. A functional extension of the Ito formula. *Comptes Rendus Mathematique*, 348(1):57–61, 2010.
- R. Cont and T. Kokholm. A consistent pricing model for index options and volatility derivatives. *Mathematical Finance*, 23(2):248–274, 2013.
- L. Coutin and Z. Qian. Stochastic analysis, rough path analysis and fractional Brownian motions. *Probability theory and related fields*, 122(1):108–140, 2002.
- C. Cuchiero. Polynomial processes in Stochastic Portfolio Theory. *Stochastic Process. Appl.*, 129(5):1829–1872, 2019.
- C. Cuchiero and F. Huber. Polynomial interacting particle systems and non-linear SPDEs for market capitalization curves. *Preprint arxiv:2404.10744*, 2024.
- C. Cuchiero and S. Svaluto-Ferro. Infinite-dimensional polynomial processes. *Finance and Stochastics*, 25(2):383–426, 2021.
- C. Cuchiero, M. Keller-Ressel, and J. Teichmann. Polynomial processes and their applications to mathematical finance. *Finance and Stochastics*, 16:711–740, 2012.
- C. Cuchiero, W. Schachermayer, and T.-K. L. Wong. Cover’s universal portfolio, Stochastic Portfolio Theory and the numeraire portfolio. *Math. Finance*, 29:773–803, 2019.
- C. Cuchiero, L. Gonon, L. Grigoryeva, J.-P. Ortega, and J. Teichmann. Discrete-time signatures and randomness in reservoir computing. *IEEE Trans. Neural Netw. Learn. Syst.*, 33(11):6321–6330, 2021a.

Bibliography

- C. Cuchiero, L. Gonon, L. Grigoryeva, J.-P. Ortega, and J. Teichmann. Expressive power of randomized signature. In *The Symbiosis of Deep Learning and Differential Equations*, 2021b.
- C. Cuchiero, F. Primavera, and S. Svaluto-Ferro. Universal approximation theorems for continuous functions of càdlàg paths and lévy-type signature models. *Preprint arxiv:2208.02293*, 2022.
- C. Cuchiero, G. Gazzani, and S. Svaluto-Ferro. Signature-based models: Theory and calibration. *SIAM J. Financial Math.*, 14(3):910–957, 2023a.
- C. Cuchiero, P. Schmock, and J. Teichmann. Global universal approximation of functional input maps on weighted spaces. *Preprint arxiv:2306.03303*, 2023b.
- C. Cuchiero, S. Svaluto-Ferro, and J. Teichmann. Signature SDEs from an affine and polynomial perspective. *Preprint arxiv:2302.01362*, 2023c.
- C. Cuchiero, L. Di Persio, F. Guida, and S. Svaluto-Ferro. Measure-valued affine and polynomial diffusions. *Stochastic Processes and their Applications*, 175:104392, 2024.
- F. Delbaen and W. Schachermayer. A general version of the fundamental theorem of asset pricing. *Math. Ann.*, 300(1):463–520, 1994.
- G. Di Nunno, K. Kubilius, Y. Mishura, and A. Yurchenko-Tytarenko. From constant to rough: A survey of continuous volatility modeling. *Mathematics*, 11(19):4201, 2023.
- X. Ding and H. Qiao. Euler-Maruyama approximations for stochastic McKean-Vlasov equations with non-Lipschitz coefficients. *Journal of Theoretical Probability*, 34(1):1408–1425, 2021.
- M. F. Djete. Non-regular McKean-Vlasov equations and calibration problem in local stochastic volatility models. *Preprint arXiv:2208.09986*, 2024.
- D. Duffie, D. Filipović, and W. Schachermayer. Affine processes and applications in finance. *The Annals of Applied Probability*, 13(3):984 – 1053, 2003.
- B. Dupire. Functional Itô calculus. *Quant. Finance*, 19(5):721–729, 2019.
- A. Eberle. Markov processes. *Lecture Notes at University of Bonn*, 2016. URL https://wt.iam.uni-bonn.de/fileadmin/WT/Inhalt/people/Andreas_Eberle/Markov_Processes_1617/MPSkript1617.pdf.
- S. N. Ethier and T. G. Kurtz. *Markov Processes: Characterization and Convergence*. John Wiley & Sons, Inc., 2005.
- T. Fawcett. *Problems in stochastic analysis. Connections between rough paths and non-commutative harmonic analysis*. PhD thesis, University of Oxford, 2003. URL <https://ora.ox.ac.uk/objects/uuid:3e14ec19-422a-4a66-96dc-5705001b0fb8>.
- R. Fernholz. *Stochastic Portfolio Theory*. Springer, 2002.

- R. Fernholz and I. Karatzas. Relative arbitrage in volatility-stabilized markets. *Ann. Finance*, 1(2):149–177, 2005.
- R. Fernholz and I. Karatzas. Stochastic Portfolio Theory: an overview, 2010.
- D. Filipović and M. Larsson. Polynomial diffusions and applications in finance. *Finance and Stochastics*, 20(4):931–972, 2016.
- D. Filipovic and E. Platen. Consistent market extensions under the benchmark approach. *Math. Finance*, 19:41–52, 2009.
- D. Filipović. Time-inhomogeneous affine processes. *Stochastic Processes and their Applications*, 115(4):639–659, 2005.
- D. Filipović and M. Larsson. Polynomial diffusions and applications in finance. *Finance and Stochastics*, 20:931– 972, 2016.
- C. Fontana, Z. Grbac, and T. Schmidt. Term structure modeling with overnight rates beyond stochastic continuity. *Mathematical Finance*, 34(1):151–189, 2024.
- J.-P. Fouque and Y. Saporito. Heston stochastic vol-of-vol model for joint calibration of VIX and S&P 500 options. *Quantitative Finance*, 18(6):1003–1016, 2018.
- P. K. Friz and N. Victoir. *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications*. Cambridge University Press, 2010.
- K. Fujii and K. Nakajima. Harnessing disordered-ensemble quantum dynamics for machine learning. *Physical Review Applied*, 8(2):024030, 2017.
- O. Futter, B. Horvath, and M. Wiese. Signature trading: A path-dependent extension of the mean-variance framework with exogenous signals. *Preprint arxiv: 2308.15135*, 2023.
- C. Gallicchio and S. Scardapane. *Deep Randomized Neural Networks*, pages 43–68. Springer, 2020.
- J. Gatheral. Consistent modeling of SPX and VIX options. *Bachelier Congress*, 2008.
- J. Gatheral. *The volatility surface: a practitioner’s guide*. John Wiley & Sons, 2011.
- J. Gatheral and M. Keller-Ressel. Affine forward variance models. *Finance and Stochastics*, 23(3):501–533, 2019.
- J. Gatheral, T. Jaisson, and M. Rosenbaum. Volatility is rough. *Quantitative Finance*, 18(6): 933–949, 2018.
- J. Gatheral, P. Jusselin, and M. Rosenbaum. The quadratic rough Heston model and the joint S&P 500/VIX smile calibration problem. *Preprint arXiv:2001.01789*, 2020.
- G. Gazzani and J. Guyon. Pricing and calibration in the 4-factor path-dependent volatility model. *Preprint arXiv:2406.02319*, 2024.

Bibliography

- B. Geuchen, K. Oberpriller, and T. Schmidt. Affine models with path-dependence under parameter uncertainty and their application in finance. *International Journal of Theoretical and Applied Finance*, 27(2), 2024.
- P. Gierjatowicz, M. Sabate-Vidales, D. Siska, L. Szpruch, and Z. Zuric. Robust pricing and hedging via neural SDEs. *Journal of Computational Finance*, 3(26), 2022.
- P. Glasserman. *Monte Carlo methods in financial engineering*, volume 53. Springer, 2004.
- L. Gonon and J.-P. Ortega. Fading memory echo state networks are universal. *Neural networks : the official journal of the International Neural Network Society*, 138:10–13, 2021.
- L. Grigoryeva and J.-P. Ortega. Echo state networks are universal. *Neural networks : the official journal of the International Neural Network Society*, 108:495–508, 2018a.
- L. Grigoryeva and J.-P. Ortega. Universal discrete-time reservoir computers with stochastic inputs and linear readouts using non-homogeneous state-affine systems. *J. Mach. Learn. Res.*, 19(24):1–40, 2018b.
- L. A. Grzelak. On randomization of affine diffusion processes with application to pricing of options on VIX and S&P 500. *Preprint arXiv:2208.12518*, 2022.
- H. Guerreiro and J. Guerra. VIX pricing in the rBergomi model under a regime switching change of measure. *Quantitative Finance*, 23(5):721–738, 2023.
- I. Guo, G. Loeper, J. Obłój, and S. Wang. Joint modeling and calibration of SPX and VIX by optimal transport. *SIAM Journal on Financial Mathematics*, 13(1):1–31, 2022a.
- I. Guo, G. Loeper, and S. Wang. Calibration of local-stochastic volatility models by optimal transport. *Mathematical Finance*, 32(1):46–77, 2022b.
- X. Guo, H. Pham, and X. Wei. Itô’s formula for flows of measures on semimartingales. *Stochastic Processes and their Applications*, 159:350–390, 2023.
- J. Guyon. Inversion of convex ordering in the VIX market. *Quantitative Finance*, 20(10):1597–1623, 2020a.
- J. Guyon. The joint S&P 500/VIX smile calibration puzzle solved. *Risk*, April, 2020b.
- J. Guyon. Dispersion-constrained martingale Schrödinger problems and the exact joint S&P 500/VIX smile calibration puzzle. *Finance and Stochastics*, 28:27–79, 2024.
- J. Guyon and F. Bourgey. Fast exact joint S&P 500/VIX smile calibration in discrete and continuous time. *Available at SSRN 4315084*, 2022.
- J. Guyon and P. Henry-Labordere. *Nonlinear Option Pricing*. Chapman and Hall/CRC, 2013.
- J. Guyon and J. Lekeufack. Volatility is (mostly) path-dependent. *Quantitative Finance*, 23(9):1221–1258, 2023.

- J. Guyon and S. Mustapha. Neural joint S&P 500/VIX smile calibration. *Risk*, December, 2023.
- P. S. Hagan, D. Kumar, A. S. Lesniewski, and D. Woodward. Managing smile risk. *The Best of Wilmott*, 1:249–296, 2002.
- W. Hammersley, D. Šiška, and L. Szpruch. McKean-Vlasov SDEs under measure dependent Lyapunov conditions. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 57(2):1032 – 1057, 2021.
- C. Herrera, F. Krach, P. Ruysen, and J. Teichmann. Optimal stopping via randomized neural networks. *Frontiers of Mathematical Finance*, 3(1):31–77, 2024.
- M. W. Hirsch and S. Smale. *Differential Equations, Dynamical Systems, and Linear Algebra*. Elsevier, 1974.
- G.-B. Huang, L. Chen, and C.-K. Siew. Universal approximation using incremental constructive feedforward networks with random hidden nodes. *Trans. Neur. Netw.*, 17(4):879–892, 2006.
- H. Hulley and M. Schweizer. *M6—On minimal market models and minimal martingale measures*, pages 35–51. Springer Berlin Heidelberg, 2010.
- M. F. D. C. A. Hurtado and T. Schmidt. Time-inhomogeneous polynomial processes. *Stochastic Analysis and Applications*, 38(3):527–564, 2020.
- D. Itkin, B. Koch, M. Larsson, and J. Teichmann. Ergodic robust maximization of asymptotic growth under stochastic volatility. *Preprint arXiv:2211.15628*, 2022.
- J. Jacod and A. N. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer, 2003.
- H. Jaeger. The "echo state" approach to analysing and training recurrent neural networks-with an erratum note'. *German National Research Center for Information Technology GMD Technical Report*, 148, 2001.
- W. B. Johnson. Extensions of Lipschitz mappings into Hilbert space. *Contemporary mathematics*, 26:189–206, 1984.
- B. Jourdain and A. Zhou. Existence of a calibrated regime switching local volatility model. *Mathematical Finance*, 30(2):501–546, 2020.
- J. Kalsi, T. Lyons, and I. P. Arribas. Optimal execution with rough path signatures. *SIAM J. Financial Math.*, 11(2):470–493, 2020.
- I. Karatzas and C. Kardaras. The numéraire portfolio in semimartingale financial models. *Finance Stoch.*, 11(4):447–493, 2007.
- I. Karatzas and D. Kim. Trading strategies generated pathwise by functions of market weights. *Finance Stoch.*, 24(2):423–463, 2019.

Bibliography

- I. Karatzas and J. Ruf. Trading strategies generated by lyapunov functions. *Finance Stoch.*, 21(3):753–787, 2017.
- I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, 1998.
- C. Kardaras. Market viability via absence of arbitrage of the first kind. *Finance and Stoch.*, 16(4):651–667, 2012.
- C. Kardaras and S. Robertson. Ergodic robust maximization of asymptotic growth. *Ann. Appl. Probab.*, 31(4):1787–1819, 2021.
- C. Kardaras, D. Kreher, and A. Nikeghbali. Strict local martingales and bubbles. *The Annals of Applied Probability*, 25(4):1827–1867, 2015.
- A. Keimer and L. Pflug. Existence, uniqueness and regularity results on nonlocal balance laws. *Journal of Differential Equations*, 263(7):4023–4069, 2017.
- P. Kidger and T. Lyons. Signatory: differentiable computations of the signature and log-signature transforms, on both CPU and GPU. In *International Conference on Learning Representations*, 2020.
- D. Kim. Market-to-book ratio in Stochastic Portfolio Theory. *Finance Stoch.*, 27(2):401–434, 2023.
- F. J. Kiraly and H. Oberhauser. Kernels for sequentially ordered data. *J. Mach. Learn. Res.*, 20(31):1–45, 2019.
- T. Kokholm and M. Stisen. Joint pricing of VIX and SPX options with stochastic volatility and jump models. *The Journal of Risk Finance*, 2015.
- D. Lacker. Limit theory for controlled mckean–vlasov dynamics. *SIAM Journal on Control and Optimization*, 55(3):1641–1672, 2017.
- D. Lacker, M. Shkolnikov, and J. Zhang. Inverting the Markovian projection, with an application to local stochastic volatility models. *The Annals of Probability*, 48(5):2189 – 2211, 2020.
- M. Larsson. Notes on polynomial and affine jump-diffusions. *Lecture Notes at ETH Zurich*, 2017. URL https://metaphor.ethz.ch/x/2017/hs/401-4923-67L/notes/PJDAF_Notes.pdf.
- E. Lemahieu, K. Boudt, and M. Wyna. Generating drawdown-realistic financial price paths using path signatures. *Preprint arXiv:2309.04507*, 2023.
- D. Levin, T. Lyons, and H. Ni. Learning from the past, predicting the statistics for the future, learning an evolving system. *Preprint arXiv:1309.0260*, 2013.
- C. Litterer and H. Oberhauser. On a Chen–Fliess approximation for diffusion functionals. *Monatshefte für Mathematik*, 175:577–593, 2011.

- T. Lyons and H. Ni. Expected signature of Brownian motion up to the first exit time from a bounded domain. *The Annals of Probability*, 43(5):2729–2762, 2015.
- T. Lyons and N. Victoir. Cubature on Wiener space. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 460(2041):169–198, 2004.
- T. Lyons, M. Caruana, and T. Lévy. *Differential Equations Driven by Rough Paths*. Springer, 2007.
- T. Lyons, H. Ni, and H. Oberhauser. A feature set for streams and an application to high-frequency financial tick data. In *Proceedings of the 2014 International Conference on Big Data Science and Computing*, 2014.
- T. Lyons, S. Nejad, and I. P. Arribas. Non-parametric pricing and hedging of exotic derivatives. *Appl. Math. Finance*, 27:457 – 494, 2019.
- T. Lyons, H. Ni, and J. Tao. A PDE approach for solving the characteristic function of the generalised signature process. *Preprint arXiv:2401.02393*, 2024.
- T. J. Lyons. Differential equations driven by rough signals. *Revista Matemática Iberoamericana*, 14(2):215–310, 1998.
- H. M. Markovitz. *Portfolio selection: Efficient diversification of investments*. John Wiley, 1959.
- J. Matoušek. On variants of the Johnson–Lindenstrauss lemma. *Random Structures Algorithms*, 33(2):142–156, 2008.
- H. P. McKean. A class of Markov processes associated with nonlinear parabolic equations. *Proceedings of the National Academy of Sciences of the United States of America*, 56: 1907 – 1911, 1966.
- M. Min and R. Hu. Signed deep fictitious play for mean field games with common noise. In *International Conference on Machine Learning*, pages 7736–7747. PMLR, 2021.
- Y. S. Mishura and A. Y. Veretennikov. Existence and uniqueness theorems for solutions of McKean–Vlasov stochastic equations. *Preprint arxiv:1603.02212*, 2024.
- C. Moler and C. Van Loan. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM review*, 45(1):3–49, 2003.
- N. Muca Cirone, M. Lemercier, and C. Salvi. Neural signature kernels as infinite-width-depth-limits of controlled ResNets. *Proceedings of Machine Learning Research*, 202:25358–25425, 2023.
- M. Negoro, K. Mitarai, K. Fujii, K. Nakajima, and M. Kitagawa. Machine learning with controllable quantum dynamics of a nuclear spin ensemble in a solid. *Preprint arxiv:1806.10910*, 2018.

Bibliography

- A. Neuberger. The log contract. *Journal of portfolio management*, 20:74–74, 1994.
- C. Ni, Y. Li, and P. A. Forsyth. Neural network approach to portfolio optimization with leverage constraints: A case study on high inflation investment. *Quantitative Finance*, 24(6):753–777, 2023.
- H. Ni. *The expected signature of a stochastic process*. Phd thesis, University of Oxford, 2012. URL <https://ora.ox.ac.uk/objects/uuid:e0b9e045-4c09-4cb7-ace9-46c4984f16f6>.
- H. Ni, L. Szpruch, M. Sabate-Vidales, B. Xiao, M. Wiese, and S. Liao. Sig-Wasserstein GANs for Time Series Generation. In *Proceedings of the Second ACM International Conference on AI in Finance*, 2022.
- B. Ning, P. Chakraborty, and K. Lee. Optimal entry and exit with signature in statistical arbitrage. *Preprint arXiv:2309.16008*, 2023.
- C. Pacati, G. Pompa, and R. Renò. Smiling twice: the Heston++ model. *Journal of Banking & Finance*, 96:185–206, 2018.
- A. Papanicolaou and R. Sircar. A regime-switching Heston model for VIX and S&P 500 implied volatilities. *Quantitative Finance*, 14(10):1811–1827, 2014.
- H. Pham and X. Wei. Dynamic programming for optimal control of stochastic McKean-Vlasov dynamics. *SIAM Journal on Control and Optimization*, 55(2):1069–1101, 2017.
- A. Quarteroni, R. Sacco, and F. Saleri. *Numerical mathematics*. Springer Science & Business Media, 2010.
- R. Ree. Lie elements and an algebra associated with shuffles. *Annals of Mathematics*, pages 210–220, 1958.
- J. Reizenstein and B. Graham. The iisignature library: Efficient calculation of iterated-integral signatures and log signatures. *Preprint arXiv:2006.00218*, 2018.
- R. Rhoads. *Trading VIX derivatives: Trading and hedging strategies using VIX futures, options, and exchange-traded notes*, volume 503. John Wiley & Sons, 2011.
- L. Rogers. Things we think we know. In *Options — 45 Years since the Publication of the Black–Scholes–Merton Model*, pages 173–184. World Scientific, 2023.
- S. Rømer. Empirical analysis of rough and classical stochastic volatility models to the SPX and VIX markets. *Quantitative Finance*, 22(10):1805–1838, 2022.
- M. Rosenbaum and J. Zhang. Deep calibration of the quadratic rough Heston model. *Preprint arXiv:2107.01611*, 2021.
- J. Ruf and K. Xie. Generalised Lyapunov functions and functionally generated trading strategies. *Applied Mathematical Finance*, 26(4):293–327, 2019.

- J. Ruf and K. Xie. The impact of proportional transaction costs on systematically generated portfolios. *SIAM J. Financial Math.*, 11(3):881–896, 2020.
- L. Rüschendorf, A. Schnurr, and V. Wolf. Comparison of time-inhomogeneous Markov processes. *Advances in Applied Probability*, 48(4):1015–1044, 2016.
- M. Scheutzow. Some examples of nonlinear diffusion processes having a time-periodic law. *The Annals of Probability*, 13(2):379 – 384, 1985.
- A. Schied, L. Speiser, and I. Voloshchenko. Model-free portfolio theory and its functional master formula. *SIAM J. Finan. Math.*, 9(3):1074–1101, 2018.
- A. Sepp. VIX option pricing in a jump-diffusion model. *Risk*, pages 84–89, 2008.
- M. Shkolnikov. Large systems of diffusions interacting through their ranks. *Stochastic Processes and their Applications*, 122(4):1730–1747, 2012.
- E. M. Stein and J. C. Stein. Stock price distributions with stochastic volatility: an analytic approach. *The review of financial studies*, 4(4):727–752, 1991.
- A.-S. Sznitman. Topics in propagation of chaos. In *Ecole d’Eté de Probabilités de Saint-Flour XIX — 1989*, pages 165–251. Springer Berlin Heidelberg, 1991.
- W. Y. Thorsten Schmidt, Stefan Tappe. Infinite dimensional affine processes. *Stochastic Processes and their Applications*, 130:7131–7169, 2020.
- B. Wang, M. Liakata, H. Ni, T. Lyons, A. Nevado-Holgado, and K. Saunders. A path signature approach for speech emotion recognition. In *Interspeech 2019*, pages 1661–1665, 2019.
- M. Wiese, P. Murray, and R. Korn. Sig-Splines: Universal approximation and convex calibration of time series generative models. *Preprint arXiv:2307.09767*, 2023.
- Y. Wu, H. Ni, T. J. Lyons, and R. L. Hudson. Signature features with the visibility transformation . In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 4665–4672, 2021.
- J. Zhang. *Topics in McKean-Vlasov equations: Rank-based dynamics and Markovian projection with applications in finance and stochastic control*. Phd thesis, Princeton University, 2021. URL <https://dataspace.princeton.edu/handle/88435/dsp01fn1072022?mode=full>.
- Z. Zhang, S. Zohren, and S. Roberts. Deep learning for portfolio optimization. *The Journal of Financial Data Science*, 2020.
- X. Y. Zhou and D. Li. Continuous-time mean-variance portfolio selection: A stochastic LQ framework. *Appl. Math. Optim.*, 42:19–33, 2000.