Polynomial Semimartingales and a Deep Learning Approach to Local Stochastic Volatility Calibration

Dissertation zur Erlangung des Doktorgrades

vorgelegt von

Wahid Khosrawi-Sardroudi

an der Fakultät für Mathematik und Physik
der Albert-Ludwigs-Universität Freiburg

April 2019

Dekan:	Prof. Dr. Gregor Herten
Referenten:	Prof. Dr. Thorsten Schmidt Prof. Dr. Martin Keller-Ressel
Datum der mündlichen Prüfung:	29.05.2019

Preface

Financial markets have experienced a precipitous increase in complexity over the past decades, posing a significant challenge from a risk management point of view. This complexity motivates the application and development of sophisticated models based on the theory of stochastic processes and in particular stochastic calculus. In this regard, the contribution of this thesis is twofold, namely by extending the class if tractable stochastic processes in form of polynomial processes and polynomial semimartingales and by showing how efficient calibration of local stochastic volatility models is possible by applying machine learning techniques.

In the first part - the main part - we extend the class of polynomial processes that has previously been established to include beyond stochastic discontinuity. This extension is motivated by the fact that certain events in financial markets take place at a deterministic time point but without foreseeable outcome. Such events consist e.g. of decisions regarding interest rates of central banks or political elections/votes. Since the outcome has a significant impact on markets, it is therefore desirable to consider stochastic processes, that can reproduce such jumps at previously specified time points. Such an extension has already been introduced in the affine framework. We will show that similar modifications hold true in the polynomial case. In particular, we will show how after this extension, computation of mixed moments in a multivariate setting reduces to solving a measure ordinary differential equation, posing a significant reduction in complexity to the measure partial differential case in the context of Kolmogorow equations. A central role in the theory of time-homogeneous polynomial processes is played by the theory of one parameter matrix semigroups. Hence, we will develop a two parameter version of the matrix semigroup theory under lower regularity then what exists in the literature. This accounts for time-inhomogeneity of the stochastic processes we consider. While in the one parameter case, full regularity follows already from very mild assumptions, we will see that this is not the case anymore in the two parameter case.

In the second part of this thesis we investigate a more applied topic, namely

the exact calibration of local stochastic volatility models to financial data. We show how this computationally challenging problem can be efficiently solved by applying machine learning techniques in form of deep neural networks. These methods have dramatically surged in the literature. Since this surge was accompanied by the development of highly efficient machine learning libraries, we can exploit this and make use of sophisticated computational tools such as gpu accelerated numerical computation. We will provide a short exposition to the underlying concepts and give numerical examples in form of toy models. We will further show how this high dimensional problem can be made tractable by the application of an auxiliary machine learning method in the context of variance reduction for Monte-Carlo pricing methods.

I want to take this opportunity to express my great gratitude to several people who have enriched my lives over the past years. Let me start with my supervisor Thorsten Schmidt. Thank you for taking a chance on me when giving me this position and thank you for sharing with me your experience over the past years. It would be a futile attempt to express in words just how much I learned during my time here in Freiburg. I also want to thank you for your friendship and advice whenever I needed it. Thank you for the many evenings we spent together with the whole AG Schmidt making food (specially the pizza), drinking coffee and frequenting the various bars here in Freiburg. Thank you also for all the opportunities you provided me that went far beyond anything I would have expected when I started my PhD. This in particular includes the fellow mathematicians I was able to meet thanks to your support, including those I like to call the "Austrians in my life". This includes in particular Josef Teichmann, who not only allowed me to learn from him as a mathematician, shared with me his passion and who helped me take my first steps into the fantastic field of mathematical machine learning, but who also spend numerous hours talking with me about politics, a shared passion of ours. Thank you Josef, for your friendship, your guidance and for the doors you have opened for me. I hope you know how grateful I am for everything you have done for me. I further would like to thank Christa Cuchiero. I become sentimental thinking back to that legendary night on the rooftop bar in Kiel, where you, Thorsten, Josef and I spent a wonderful time talking and discussing about maths, politics and everything interesting we could think of. Wonderful, how this became the starting point of the legendary FWZ-seminar. I hereby want to apologize to you and the others for making you try the "Mexikaner" shots. Being from Hamburg, I was sure you all would enjoy them as much as I do. Thank you also for inviting me to work together with you and Josef on our joint project and thank you for inviting me to Vienna where we made significant progress together. Next, I would like to thank Philipp Harms, another Austrian who enriched my life. Thank you for your friendship and help whenever I needed someone to discuss mathematics. How happy I am for you to have come join us in Freiburg. No question for me that our friendship will continue beyond my time in Freiburg. Let me finish the group of "FWZ-Austrians" by mentioning Irene Klein. Thank you for our discussions and for your work in co-organizing the FWZ-seminar. I apologize to you for my failed attempts to emulate the Austrian dialect. I promise I will not stop practicing in the years to come. I also want to thank Sandrine, for being the best "PhD-sister". Thank you for taking time to discuss with me, for the laughs we had and for the numerous invitations to your and Mo's apartment, you two were always wonderful hosts. Further, I want to thank all members of the mathematical stochastic group in Freiburg. The environment here was always positive, motivating and made it very easy to come to the office every day. This includes in no particular order Tolu, Monika, Lukas, Yves, Maria, Bene, Marcus, Felix, Elli, Ernst-August, Pascal, Johannes, Timo and Joni. Also my thanks to the professors Peter Pfaffelhuber, Angelika Rohde, Ludger Rüschendorf, Ernst Eberlein, Hans Rudolf Lerche and Steffan Tappe. I further would like to thank all the colleagues I have met outside of Freiburg. I want to thank Sam Cohen, Martin Keller-Ressel, Claudio Fontana, Kathrin Glau, Martin Larsson, Sergio Pulido, Damir Filipović and Johannes Muhle-Karbe for all the help and advice they have given me over the past years. Further thanks go to Sara, Lukas, Marvin, Eduardo, Matteo and all other colleagues I was privileged to meet.

Many thanks to all my friends here in Freiburg from outside the world of mathematics. In particular many thanks to Philip and Nadja, you are both amazing and you both made my time in Freiburg so much better.

Finally, I want to thank my family. Thanks goes to my sister Parisa and her husband Christian. Thank you Parisa for always being the first in our family to walk the path I would follow shortly after. Thank you for removing all the obstacles in my way. I hope you know how much I love you and how important you are to me. The greatest imaginable thanks to my parents Ali and Zohreh. What the two of you did for Parisa and me goes beyond what words could describe. I feel privileged knowing that the two of you love me and it is this very love that motivates me to continue my path. Both of you are my role models and everything I am and everything I achieve in my life is thanks to the love the two of you gave me and the many sacrifices you made.

Freiburg, April of 2019

Contents

Part I

Polynomial Semimartingales

Chapter 1

Introduction

The first part of this thesis aims to generalize the class of finite dimensional polynomial processes to include processes beyond stochastic continuity. Stochastic processes are widely used in financial modeling for various applications. In the context of equity modeling, the seminal work of Black and Scholes [BS73] introduced a by today's standard simple model that allows for the valuation of European vanilla options. Other areas of finance have profited from the introduction of financial models based on stochastic calculus, see for example Schönbucher [Sch98] in the context of term structure modeling and the seminal Nobel price winning work by Merton [Mer74] in the context of credit risk management.

When dealing with financial markets, one is often confronted with two challenges. First, the model needs to "capture" empirical properties usually observed in financial data. Second, it is necessary to correctly model dependencies in a market in order to not undervalue risks. Regarding the latter, this motivates the simultaneous modeling of multiple prices where the movement of one price is not independent of the movement of the others. This in particular becomes a necessity when considering financial derivatives that depend on multiple underlying assets.

One the other hand, for the first, such properties can be the geometry of the observed data such as bounded state spaces of prices. An example where this is a needed feature would be the modeling of electricity prices. Such prices show a strong seasonality, motivating the consideration of time-inhomogeneous models. From a simplified point of view, a model defined by an Itô diffusion would need to allow for time depending parameters.

Probably the best known short cumming of the Black-Scholes model is that it is incapable to reproduce the implied volatility smile observed in the market. In the past decades, stochastic volatility models have aimed to address the issue of "fitting the smile". This work was pioneered by Hull and White [HW87] and the most famous example would probably be the Heston model, see Heston [Hes93].

In the last years in particular, stochastic volatility models have gained attention since empirical data suggests that the volatility process is "rough", see in particular Gatheral et al. [Gat+14], motivating the extension to fractional stochastic volatility models. These models were originally considered to capture long memory effects that can not be reproduced by a Markov model, see Comte and Renault [CR98]. Advances in this field were in particular achieved in the context of affine processes, see El Euch and Rosenbaum [EER19], Jaber et al. [Jab+17], Cuchiero and Teichmann [CT18] and Gatheral and Keller-Ressel [GKR18]. In this thesis, we will not deal with an extension of polynomial processes to a fraction setting.

From an applications point of view, one faces the problem that the model needs to be sufficiently tractable so that pricing and calibration is feasible, standing in direct contrast to the above formulated desirable features of financial models. A class of stochastic processes bridging this gap is given by affine processes. The study of such stochastic processes has a long history and we refer to the introduction in Cuchiero [Cuc11] for an overview.

An extension of of this class of processes is given in form of polynomial processes. A first systematic treatment of the multivariate case was given in Cuchiero et al. [Cuc+12] where a full characterization in terms of semimartingale characteristics is provided. In this reference, the authors introduce polynomial processes as time homogeneous Markov processes where they allow for a non-zero killing rate. An alternative but related approach is given in Filipović and Larsson [FL16] and Filipović and Larsson [FL17]. The approach taken there is one from a semimartingale point of view without requiring the considered processes to be Markov. We in particular want to refer to the introduction of the last reference where the authors give examples of applications of polynomial processes in mathematical finance. Let us however mention the comprehensive study of multivariate polynomial processes in a jump diffusion setting and supported on the unit simplex in Cuchiero et al. [Cuc+18b], together with their applications in stochastic portfolio theory, see Cuchiero [Cuc18].

All these approaches have in common that the stochastic processes considered are stochastically continuous which is essentially a consequence of the fact that time homogeneity implies a semigroup structure which in return, already under very mild regularity assumptions, implies strong regularity of the semigroup, see Proposition 4.2 in this thesis. Regularity of the semigroup in return is inherited by the time dependency of conditional moments which by means of the Markov inequality then implies stochastic continuity. It is however a desirable feature of financial models to allow for jump-times that are previously known, see e.g. the introduction of [KR+18] and the references therein.

A first extension to a time-inhomogeneous setting in the spirit of [Cuc+12] was provided by C. A. Hurtado [CAH17], however under strong regularity assumptions on the Markov semigroup that imply stochastic continuity. From an application point of view however, there are numerous examples of events such as interest rate decisions, credit default events or political elections, where the time point of the event is known, but not the outcome. Since the "type" of outcome has a significant impact on markets, being able to model stochastic discontinuity falls under the class of desirable features of financial models.

A first extension in the affine framework, allowing for stochastic discontinuity while preserving tractability, was given in Keller-Ressel et al. [KR+18]. In this part of the thesis, we want to follow their lead and extend the class of polynomial processes to include examples with stochastic discontinuity while preserving tractability. Tractability in the context of polynomial processes means that one should be able to compute conditional mixed moments in arbitrary but finite dimension in an efficient way. While in the time-homogeneous setting, this boils down to the computation of a matrix exponential, the time-inhomogeneous setting in [CAH17] requires solving a potentially high-dimensional ordinary differential equation. We will show that in our framework, the computation of mixed moments is achieved by solving a measure differential equation, where the time derivative is replaced by a Radon-Nikodým derivative. This is due to the fact that the Lebesgue measure is replaced with a measure that charges points, corresponding to deterministic jump-times. This is very much similar to the results presented in [KR+18], where instead of Riccati equations, the authors show that the conditional characteristic function is given as the solution of a measure Riccati equation. Let us further note, that our characterization results wrt semimartingale characteristics differ from the affine case. A central role in this characterization is played by the existence of an increasing càdlàg function A wrt which the differential characteristics are given. While in [KR+18, Theorem 3.2] the characteristics are given wrt the continuous part of A so that predictable jumps are "taken care of" by the discontinuous part of the predictable compensator of the associated jump measure, we formulate our results without decomposing A to continuous and pure jump part.

Before giving a brief overview of our approach, let us briefly mention the work on polynomial processes in infinite dimensions, namely Cuchiero et al. [Cuc+18c] and Benth et al. [Ben+18].

Outline of this part of the thesis

We want to take the viewpoint expressed in the introduction of Filipović and Larsson [FL17] and opt to not assume a priory a Markov process since from an application point of view, it seems preferable to consider a semimartingale framework. However, we want to give an (almost) complete characterization for stochastic processes that have the fundamental property that conditional moments are expressed by a polynomial of a corresponding degree, we call this the polynomial property. Hence we will consider càdlàg processes on a fixed stochastic basis and show how the polynomial property is related to a Markovian type structure on the space of multivariate polynomials of at most a certain degree, allowing us to give results similar to those in [Cuc+12] without assuming an underlying Markov process. We will repeatedly take advantage of finite dimensionality of the space of polynomials of at most a given degree, allowing us to develop a theory under significantly fewer regularity assumptions. Due to time-inhomogeneity, we can not rely on well established results of matrix semigroups. Hence after fixing some notation and reviewing some basic results from probability in the subsequent two chapters, we will consider the time-inhomogeneous version we will refer to as matrix evolution systems in Chapter 4. While evolution systems have been considered in both, finite and infinite dimensional settings ([EN99] and [Böt14] to name just two), usually the assumptions are under strong regularity and rely on space-time homogenization. Since such an approach is not reasonable in our case as we do not want to leave a finite dimensional setting, the presentation we give under the regularity assumptions we make, to the best of our knowledge, is not available in the literature. In the same chapter, we will introduce a concept of infinitesimal description of such matrix evolution systems that is suitable for our regularity assumptions. We will then show how such an infinitesimal description, named extended generator and motivated by extended generators of Markov processes, relates to measure differential equations (or rather integral equations) that will play a central role in our characterization results. As extended generators for matrix evolution systems and Markov processes share the same name, this could lead to confusion. In this thesis however, it will always be clear from context which of the two concepts is meant, avoiding any ambiguity on this regard.

In Chapter 5, we will relate certain invariance properties of matrix evolution systems to invariance properties of the extended generators. While a triviality in the regular case, we will see that these results are not as clear anymore in our situation. For that, we will in particular formalize the well known relationship between polynomials of bounded degree and Euclidean spaces.

In Chapter 6, we will finally introduce a class of processes we call (k-)polynomial processes in the spirit of [Cuc+12]. We will show how these processes relate to matrix evolution systems and how the extended generator of the stochastic process corresponds to the extended generator of a related matrix evolution system. We will see that these processes are special semimartingales and their characteristics satisfy certain properties. We will then introduce a class of semimartingales called polynomial semimartingales in the spirit of [FL17]. These are special semimartingales with characteristics that satisfy assumptions motivated by the results from the previously introduced class of polynomial processes and we show that under reasonable assumptions, these semimartingales indeed belong to the class of polynomial processes.

In this thesis, we do not allow for a positive killing rate, unlike [Cuc+12]. Hence, in particular regarding [Cuc+12, Proposition 2.12], our setting corresponds to a killing rate $\gamma = 0$ and by that hitting time of the "coffin state" $T_\Delta = \infty$. On the other hand, the setting in [FL16] and [FL17] is entirely covered by our approach. We do not however provide results such as uniqueness to corresponding martingale problems, preserving properties under subordination or boundary attainment of polynomial (jump)-diffusions.

Chapter 2

Preliminaries

Let us briefly fix some notation we will use in this thesis. We will in particular make conventions that will hold for the whole thesis unless explicitly stated otherwise. We will adapt most of the notation from Jacod and Shiryaev [JS13] which should be to the benefit of readers, familiar with this reference.

2.1 Basic notational conventions

Throughout this thesis, we have

(i) $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ with $\{0\} \notin \mathbb{N}$

(ii) $\mathbb{R}_+ = [0, \infty)$, and $\mathbb{R}_{>0} = (0, \infty)$

(iii) $\mathbb{T} = \{(s,t) \in \mathbb{R}^2 : 0 \le s \le t\}$

(iv) $\dot{\bigcup}_{i \in I} A_i$ is the disjoint union of the sets A_i for $i \in I$

(v) $\inf_\emptyset := \infty$ and $\sup_\emptyset = 0$.

(vi) $a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$, where $a, b \in A$ for a totally ordered set $(A, \le)$

(vii) $x^+ = x \vee 0$ and $x^- = x \wedge 0$ for $x \in \mathbb{R}$

(viii) $A^c =$ the complement of a set A

(ix) $\lfloor x \rfloor = \sup\{n \in \mathbb{N}_0 : n \le x\}$ for $x \in \mathbb{R}_+$

(x) $t_n \downarrow t$ a sequence of real numbers converging to $t \in \mathbb{R}$ with $t_n > t$ for all $n \ge 1$

(xi) $t_n \uparrow t$ a sequence of real numbers converging to $t \in \mathbb{R}$ with $t_n < t$ for all $n \geq 1$

(xii) $\lim_{\varepsilon \downarrow 0} = \lim_{\varepsilon \to 0, \varepsilon > 0}$

(xiii) $\ll$ absolute continuity between measures

(xiv) a.s. = almost surely and a.e. = almost everywhere

(xv) $\mathbb{1}_A$ indicator function of the set A

(xvi) $\times$ the Cartesian product

(xvii) $\otimes$ product σ-algebra

(xviii) $\|x\|_{\mathbb{R}^d}$ any norm on $\mathbb{R}^d$ for $x \in \mathbb{R}^d$

(xix) $\|x\|_p = \left(\sum_{i=1}^d |x_i|^p\right)^{\frac{1}{p}}$ for $x \in \mathbb{R}^d$ and $p \in [1, \infty)$

(xx) $\bigvee_{[a,b]}(f)$ the total variation of a function f on the interval $[a, b]$

(xxi) Pol_k the set of polynomials over $\mathbb{R}^d$ of at most degree k

(xxii) $\deg f =$ the degree of a polynomial f, hence the smallest k for which $f \in \mathrm{Pol}_k$

(xxiii) $\mathrm{M}_n(\mathbb{R}) = \mathbb{R}^{n \times n}$, the set of square matrices over $\mathbb{R}$ in dimension n

(xxiv) $\det M$, the determinant of a matrix $M \in \mathrm{M}_n(\mathbb{R})$

Fix a dimension $d \in \mathbb{N}$. Let $I \subset \mathbb{R}_+$. For a function $u : I \to \mathbb{R}^d$, define

$$u_t^* := \sup_{s \in I, s \leq t} \|u(s)\|_{\mathbb{R}^d}.$$

For $I = \mathbb{R}_{>0}$, we extend a function u to $\mathbb{R}_+$ unless otherwise stated by setting $u(0) = 0$. If u has left limits, we fix

$$u_{t-} := \lim_{\varepsilon \downarrow 0} u_{t-\varepsilon}, \quad \Delta u_t := u_t - u_{t-},$$

where we make the convention $u_{0-} = u_0$ which implies $\Delta u_0 = 0$, compare the conventions made in [JS13, p. I.1.8]. We extend these conventions to stochastic processes $X : \mathbb{R}_+ \times \Omega \to \mathbb{R}^d$ by making them for each $\omega \in \Omega$ fixed, i.e. for each function $t \mapsto X(t, \omega)$. We will interchangeably use the notation $X_t = X(t)$ for both, stochastic processes and deterministic functions.

For a matrix $M \in \mathbb{R}^{d \times d}$ we denote by $\|M\| = \sup_{\|x\|_{\mathbb{R}^d} = 1} \|Mx\|$ the induced operator norm for $\|\cdot\|_{\mathbb{R}^d}$. It will be clear from context which vector norm is meant or a statement of the form $\|M\| < \infty$ will be independent of the concrete choice of the norm by means of equivalence of norms.

2.2 Notation and conventions for multivariate polynomials

Following Filipović and Larsson [FL16], we consider the set Pol_k of polynomials over $\mathbb{R}^d$ with real coefficients and of most degree k, i.e.

$$\mathrm{Pol}_k = \left\{ x \mapsto \sum_{|\mathbf{l}|=0}^{k} \alpha_{\mathbf{l}} x^{\mathbf{l}} \mid \alpha_{\mathbf{l}} \in \mathbb{R},^{l} = (l_1, \dots, l_d), x^{\mathbf{l}} = x_1^{l_1} \dots x_d^{l_d} \right\}.$$

Denote by $N_k = N$ the dimension of Pol_k. Note that Pol_k is a linear space of finite dimension. Fix a basis $\beta = \{\beta_1, \dots, \beta_N\}$ of Pol_k and define the vector valued function

$$\mathbf{H}_\beta(x) := (\beta_1(x), \dots, \beta_N(x))^\top,$$

such that we then have that given $f \in \mathrm{Pol}_k$, there is a unique coordinate representation $\mathbf{f} \in \mathbb{R}^N$ such that $f(x) = \mathbf{f}^\top \mathbf{H}_\beta(x)$ for all $x \in \mathbb{R}^d$. It is clear that $\mathbf{f}$ depends on the concrete choice of β. In Chapter 5, we will introduce additional conventions regarding β that allow for a convenient notation when considering multiple basis β for several degrees l of Pol_l.

Throughout, we will use the following notation. For $x \in \mathbb{R}^d$, define $f_i(x) = x_i, i = 1, \dots, d$, the projection on the i-th coordinate. Hence, $f_i \in \mathrm{Pol}_1$. We will further use the multi-index notation. For $\mathbf{k} = (k_1, \dots, k_d) \in \mathbb{N}_0^d$, define the polynomial

$$f_{\mathbf{k}}(x) = f_{(k_1,\dots,k_d)}(x) = \prod_{i=1}^{d} f_i(x)^{k_i}.$$

Hence, for $|\mathbf{k}| = \sum_{i=1}^d k_i$, we have $f_{\mathbf{k}} \in \mathrm{Pol}_{|\mathbf{k}|}$. When $d = 1$, we want to distinguish f_i from $f_{\mathbf{k}}$ by making the convention

$$f_i(x) = x \text{ while } f_{(i)}(x) = x^i,$$

even though the left expression only makes sense for $i = 1$ while the one on the right for all $i \in \mathbb{N}_0$. We denote by $\leq$ the partial order on $\mathbb{N}^d$, i.e.

$$\mathbf{l} \leq \mathbf{k} \Leftrightarrow l_i \leq k_i \ \forall i \in \{1, \dots, d\}.$$

Hence for a sequence $(c_{\mathbf{l}})_{\mathbf{l} \in \mathbb{N}^d}$, the sum $\sum_{\mathbf{l} \leq \mathbf{k}} c_{\mathbf{l}}$ is taken over all multi-indices satisfying $\mathbf{l} \leq \mathbf{k}$. This implies $|\mathbf{l}| < |\mathbf{k}|$ for $\mathbf{l} \neq \mathbf{k}$. Further, if $j \leq k$ are two elements of $\mathbb{N}_0$, we denote with $\sum_{|\mathbf{l}|=j}^{k} c_{\mathbf{l}}$ the sum over all multi indices $\mathbf{l}$ satisfying $j \leq |\mathbf{l}| \leq k$. Hence if $k = |k|$, the sum contains all elements with degree k, not just $\mathbf{k}$ in contrast to the first notation for sums we introduced above. It will further be convenient to use the multi index notation for binomial coefficients, also referred to as multi-binomial coefficients. These are defined by

$$\binom{\mathbf{k}}{\mathbf{l}} := \prod_{i=1}^{d} \binom{k_i}{l_i}.$$

Note that we have that if $\mathbf{l} \not\leq \mathbf{k}$, then $\binom{\mathbf{k}}{\mathbf{l}} = 0$. Hence, for $k = |k|$ we get the equality

$$\sum_{\mathbf{l} \leq \mathbf{k}} \binom{\mathbf{k}}{\mathbf{l}} c_{\mathbf{l}} = \sum_{|\mathbf{l}|=0}^{k} \binom{\mathbf{k}}{\mathbf{l}} c_{\mathbf{l}}.$$

This is in particular useful in the context of the multi-binomial theorem. Recall that for $x, y \in \mathbb{R}^d$ we have

$$f_{\mathbf{k}}(x+y) = (x+y)^{\mathbf{k}} = \sum_{\mathbf{l} \leq \mathbf{k}} \binom{\mathbf{k}}{\mathbf{l}} x^{\mathbf{k}-\mathbf{l}} y^{\mathbf{l}} = \sum_{|\mathbf{l}|=0}^{k} \binom{\mathbf{k}}{\mathbf{l}} x^{\mathbf{k}-\mathbf{l}} y^{\mathbf{l}}. \tag{2.1}$$

Let us make the convention that a sum, indexed by an empty set, is zero. The following lemma will be useful when we characterize k-polynomial processes as semimartingales in Chapter 6. A proof can be found in [CAH17], but we present a shorter and more direct proof for the readers convenience. We denote with $\mathbf{e}_i \in \mathbb{N}^d$ the element with all entries zero except the i-th being one. In particular we have $f_{\mathbf{e}_i}(x) = f_i(x)$. Further, we denote with D_i and D_{ij} the partial derivatives of a function $f : \mathbb{R}^d \to \mathbb{R}$ as is customary in the literature.

Lemma 2.1. *Let $x, y \in \mathbb{R}^2$ and $\mathbf{k} \in \mathbb{N}_0^d$ with $k = |\mathbf{k}|$. We then have*

$$(x+y)^{\mathbf{k}} - x^{\mathbf{k}} - \sum_{i=1}^{d} D_i f_{\mathbf{k}}(x) f_i(y) = \sum_{|\mathbf{l}|=2}^{k} \binom{\mathbf{k}}{\mathbf{l}} x^{\mathbf{k}-\mathbf{l}} y^{\mathbf{l}} \tag{2.2}$$

Proof. Note that we have

$$\{\mathbf{l} \in \mathbb{N}^d : |\mathbf{k}| = 0\} = \{(0, \dots, 0)\} \text{ and } \{\mathbf{l} \in \mathbb{N}^d : |\mathbf{k}| = 1\} = \{\mathbf{e}_i, i \in \{1, \dots, d\}\}.$$

Since for $\mathbf{l}$ with $|\mathbf{l}| = 0$, we get $D_i f_{\mathbf{l}}(x) = 0$, the case $|\mathbf{k}| = 0$ is clear. Consider now $|\mathbf{k}| = 1$, hence $\mathbf{k} = \mathbf{e}_i$ for some i. Then

$$(x+y)^{\mathbf{e}_i} - x^{e_i} - \sum_{j=1}^{d} D_j f_{\mathbf{e}_i}(x) f_j(y) = x_i + y_i - x_i - y_j = 0,$$

hence the case $|\mathbf{k}| = 1$ is shown too. Consider now $|\mathbf{k}| \geq 2$. By the multi-binomial theorem (2.1) we have

$$(x+y)^{\mathbf{k}} = \sum_{|\mathbf{l}|=0}^{k} \binom{\mathbf{k}}{\mathbf{l}} x^{\mathbf{k}-\mathbf{l}} y^{\mathbf{l}}.$$

We now have

$$\sum_{|\mathbf{l}|=0}^{0} \binom{\mathbf{k}}{\mathbf{l}} x^{\mathbf{k}-\mathbf{l}} y^{\mathbf{l}} = x^{\mathbf{k}}.$$

Further, we have

$$\sum_{|\mathbf{l}|=1}^{1} \binom{\mathbf{k}}{\mathbf{l}} x^{\mathbf{k}-\mathbf{l}} y^{\mathbf{l}} = \sum_{i=1}^{d} \binom{\mathbf{k}}{\mathbf{e}_i} x^{\mathbf{k}-\mathbf{e}_i} y^{\mathbf{e}_i} = \sum_{i=1}^{d} k_i x^{\mathbf{k}-\mathbf{e}_i} f_i(y).$$

On the other hand, we have

$$\sum_{i=1}^{d} D_i f_{\mathbf{k}}(x) f_i(y) = \sum_{i=1}^{d} k_i f_{\mathbf{k}-\mathbf{e}_i}(x) f_i(y) = \sum_{i=1}^{d} k_i x^{\mathbf{k}-\mathbf{e}_i} f_i(y),$$

which shows (2.2) as claimed. □

By introducing an order on the set of multi-indices, we will be able to identify a multi-index with a basis function β_i. That will become convenient later. Consider the sets

$$\Lambda_k = \left\{ \alpha \in \mathbb{N}^d : |\alpha| \leq k \right\}.$$

Then $\{f_\alpha : \alpha \in \Lambda_k\}$ is a basis of Pol_k. We define the ***graded reverse-lexicographic order*** $\overset{\text{glex}}{\leq}$ on Λ_k the following way. Let $\overset{\text{lex}}{\leq}$ be the lexicographic order on Λ_k. For $\alpha, \gamma \in \Lambda_k$, we define the total order $\overset{\text{glex}}{\leq}$ by:

- If $|\alpha| \neq |\gamma|$, then $\alpha \overset{\text{glex}}{\leq} \gamma$ for $|\alpha| < |\gamma|$, else $\gamma \overset{\text{glex}}{\leq} \alpha$.
- If $|\alpha| = |\gamma|$, then $\alpha \overset{\text{glex}}{\leq} \gamma$ for $\alpha \overset{\text{lex}}{\geq} \gamma$, else $\gamma \overset{\text{glex}}{\leq} \alpha$.

Since $\alpha \overset{\text{lex}}{=} \gamma$ implies $\alpha = \gamma$, this order is well defined. Under this order, we can enumerate the elements of Λ_k. Let $\{\alpha_i : i = 1, \ldots, |\Lambda_k|\}$ be this enumeration. Then $\beta_i = f_{\alpha_i}$ defines a basis of Pol_k.

Example 2.2. *Consider the case $d = 2$ and $k = 2$. Then*

$$\alpha_1 = (0,0),\ \alpha_2 = (1,0),\ \alpha_3 = (0,1),\ \alpha_4 = (2,0),\ \alpha_5 = (1,1),\ \alpha_6 = (0,2).$$

From here on, we refer to the basis $\beta_i = f_{\alpha_i}$ as the canonical basis of Pol_k. Since the canonical basis of Pol_{k+1} is a basis extension of the canonical basis of Pol_k, we have actually defined a basis of $\mathrm{Pol} = \cup_{n\in\mathbb{N}} \mathrm{Pol}_n$ and refer to that one as the canonical basis of Pol.

2.3 Function spaces

We denote by $C(X, Y)$ the set of continuous functions $f : X \mapsto Y$ for two topological spaces (usually some form of $\mathbb{R}^d$) X and Y. The set of Borel-measurable functions is denoted by $\mathcal{E}(X, Y)$. In case of $Y = \mathbb{R}$, we simply write $C(X)$ (resp. $\mathcal{E}(X)$). The restriction to bounded functions is denoted by $bC(X)$ (resp. $b\mathcal{E}(X)$). With $\mathcal{D}(I, \mathbb{R}^d)$, we denote the set of all functions that are indexed by $I \subset \mathbb{R}$, take values in $\mathbb{R}^d$ and are càdlàg. For a set I equipped with a topology (usually $\mathbb{R}^d$ or some subset), we denote with $\mathcal{B}(I)$ the corresponding Borel σ-field.

Let $(\Omega, \mathscr{F}, P)$ be a probability space and $p \in [1, \infty)$. We denote with $\mathcal{L}^p(\Omega, \mathcal{F}, P)$ the set of all $\mathcal{F}$ measurable random variables X, taking value in $\mathbb{R}^d$ for some $d \in \mathbb{N}$ such that $\mathbb{E}\left[\|X\|^p_{\mathbb{R}^d}\right] < \infty$ where the expectation is taken under the measure P.

Note that under this definition, two random variables X, Y might belong to the set $\mathcal{L}^p(\Omega, \mathcal{F}, P)$ while having different state spaces. For our purposes, this will not pose a problem since we only want to indicate integrability.

Chapter 3

General theory

In this section we want to prepare some of the general theory of stochastic processes for the readers convenience. The first part will state some general results from probability theory as we will need it. We will not introduce all concepts we use, for the main part of the theory of semimartingales, we follow the school of Protter [Pro05] and Jacod and Shiryaev [JS13] and refer the reader to these references for any concept or notation we fail to introduce here.

3.1 General probability

Let us start with a fundamental definition.

Definition 3.1. *We call a collection* $(\Omega, \mathscr{F}, \mathbb{F}, P)$ *a* **stochastic basis***, if* $(\Omega, \mathscr{F}, P)$ *is a probability space and* $\mathbb{F} = (\mathcal{F}_s)_{s\geq 0}$ *a filtration of sub-*σ*-algebras of* $\mathcal{F}$*, satisfying the usual conditions:*

(i) non-decreasing, i.e. for all $s \leq t$, $\mathcal{F}_s \subset \mathcal{F}_t$,

(ii) right continuous, i.e. $\mathcal{F}_t = \cap_{\varepsilon>0}\mathcal{F}_{t+\varepsilon}$,

(iii) completeness, $\mathcal{F}_0$ *is* P *complete and* $\mathcal{F}_0$ *contains all* P*-null sets of* $\mathcal{F}$.

This definition slightly differs from Jacod and Shiryaev [JS13, Definition I.1.2 and I.1.3] as we always assume property *(iii)* above whenever we speak of a stochastic basis in this thesis.

In this thesis, all equalities between random variables is to be understood in an almost sure sense unless otherwise stated. Sometimes we want to emphasize this or want to emphasize under which measure (not always probability measure) an equality is to be understood. Hence for a measure Q, we write

$$X = Y \quad [Q],$$

whenever X and Y agree outside a Q-zero set. Further, for a sequence of random variables (X_n) and a random variable X, we write $X_n \stackrel{\mathcal{L}^p}{\to} X$, if

$$\lim_{n\to\infty} \mathbb{E}\left[\|X_n - X\|_{\mathbb{R}^d}^p\right] = 0, p \in [1,\infty).$$

Consider the probability space $(\Omega, \mathcal{F}, P)$. We will need the concept of uniformly integrable family of random variables defined on a common probability space, compare Meyer [Mey66, Definition D17 on page 16]. Note that this concept exists in a more general measure theoretical framework, see for example Rudin [Rud87].

Definition 3.2. *Let $\mathcal{H}$ be a family of $\mathbb{R}^d$-valued random variables $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$. $\mathcal{H}$ is called uniformly integrable, if for all $\varepsilon > 0$ there is a $c \in \mathbb{R}_+$ such that*

$$\mathbb{E}\left[\|X\|_{\mathbb{R}^d} \mathbb{1}_{\{\|X\|_{\mathbb{R}^d} \geq c\}}\right] < \varepsilon \qquad \forall X \in \mathcal{H}. \tag{3.1}$$

The motivation for the concept of uniform integrability becomes clear with the next theorem which is a generalization of the dominated convergence theorem. Again we refer to Meyer [Mey66, Theorem 21 on page 18] for the proof.

Theorem 3.3. *Let $(X_n)_{n\geq 1}$ be a sequence of $\mathbb{R}^d$-valued random variables such that for all $n \in \mathbb{N}$, $X_n \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ and assume that they converge a.s. to a random variable X. Then $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ and it holds $X_n \stackrel{\mathcal{L}^1}{\to} X$ if and only if the sequence $(X_n)_{n\geq 0}$ is uniformly integrable.*

We will now state a theorem due to La Vallée Poussin (1866 – 1962) that allows for a convenient characterization of uniform integrability. For the proof see Meyer [Mey66, Theorem T22].

Theorem 3.4. *Let $\mathcal{H}$ be a set of $\mathbb{R}^d$-valued random variables which is a subset of $\mathcal{L}^1(\Omega, \mathcal{F}, P)$. The following properties are equivalent:*

(1) $\mathcal{H}$ is uniformly integrable

(2) There exists a positive, convex and increasing function $\varphi(t)$ on $\mathbb{R}_+$ such that

$$\lim_{t\to+\infty} \frac{\varphi(t)}{t} = +\infty, \tag{3.2}$$

and

$$\sup_{X\in\mathcal{H}} \mathbb{E}\left[\varphi(\|X\|_{\mathbb{R}^d})\right] < \infty. \tag{3.3}$$

Let $(\Omega, \mathscr{F}, \mathbb{F}, P)$ be a stochastic basis. We will need the following lemma when dealing with polynomial processes.

Lemma 3.5. *Let X be an $\mathbb{R}$-valued adapted càdlàg process and assume that for every $t > 0$, we have that*

$$\{X_s : 0 \leq s \leq t\} \qquad \text{is uniformly integrable.}$$

Then it holds that

(i)
$$\sup_{0 \leq s \leq t} \mathbb{E}\left[|X_s|\right] < \infty, \qquad \sup_{0 < s \leq t} \mathbb{E}\left[|X_{s-}|\right] < \infty \quad \forall\, t \geq 0.$$

(ii) for every $t_n \downarrow t$ we have for fixed s,
$$\mathbb{E}\left[X_{t_n} \mid \mathcal{F}_s\right] \xrightarrow{\mathcal{L}^1} \mathbb{E}\left[X_t \mid \mathcal{F}_s\right] \tag{3.4}$$
and for $\tilde{t}_n \uparrow t$
$$\mathbb{E}\left[X_{\tilde{t}_n} \mid \mathcal{F}_s\right] \xrightarrow{\mathcal{L}^1} \mathbb{E}\left[X_{t-} \mid \mathcal{F}_s\right]. \tag{3.5}$$

Proof. Choose $c > 0$ (depends on t) large enough such that

$$\sup_{0 \leq s \leq t} \mathbb{E}\left[|X_s| \, \mathbb{1}_{|X_s| \geq c}\right] \leq \varepsilon.$$

This is possible by the definition of uniform integrability. Hence we get

$$\sup_{0 \leq s \leq t} \mathbb{E}\left[|X_s|\right] = \sup_{0 \leq s \leq t} \mathbb{E}\left[|X_s| \left(\mathbb{1}_{|X_s| \geq c} + \mathbb{1}_{|X_s| < c}\right)\right] \leq \varepsilon + c$$

Further, for $\tilde{u}_n \uparrow u$ for $0 < u \leq t$, we have by continuity of $|\cdot|$ and the càdlàg property of X and Fatou's lemma that

$$\mathbb{E}\left[|X_{u-}|\right] = \mathbb{E}\left[\liminf_{n \to \infty} |X_{\tilde{u}_n}|\right] \leq \liminf_{n \to \infty} \mathbb{E}\left[|X_{\tilde{u}_n}|\right] \leq \sup_{0 \leq s \leq t} \mathbb{E}\left[|X_s|\right] < \infty.$$

Hence, since the upper bound holds for all u, we have shown (i). Regarding (ii), note that (i) implies that the conditional expectations in (ii) are well defined. Again by the càdlàg property of X we have for $t_n \uparrow t$ and $\tilde{t}_n \uparrow t$,

$$X_{t_n} \longrightarrow X_t \text{ and } X_{\tilde{t}_n} \longrightarrow X_{t-} \quad [P].$$

Together with the uniform integrability assumption, (ii) follows from theorem 3.3 since

$$\mathbb{E}\left[\left|\mathbb{E}\left[X_{t_n} | \mathcal{F}_s\right] - \mathbb{E}\left[X_t | \mathcal{F}_s\right]\right|\right] \leq \mathbb{E}\left[|X_{t_n} - X_t|\right],$$

which goes goes to zero by $\mathcal{L}^1$ convergence. The same is true for $\tilde{t}_n$. $\square$

We follow the notation in Jacod and Shiryaev [JS13] and denote with $\mathcal{M}$ the set of all uniformly integrable martingales. For a set of processes $\mathcal{A}$, we denote with $\mathcal{A}_{\text{loc}}$ the localized class. To be precise, we call a sequence of stopping times $(T_n)_n$ a localkizing sequence, if $T_n \to \infty$ a.s. for $n \to \infty$. A process M is then an element of $\mathcal{A}_{\text{loc}}$, if the stopped process $M^{T_n} \in \mathcal{A}$ for all n. In our notation we have $M_t^{T_n} = M_{t\wedge T_n}$. Hence, $\mathcal{M}_{\text{loc}}$ denotes the set of all local martingales. We further make the following standard convention, where a $\mathbb{R}^d$ valued process is a (local) martingale (resp. semimartingale), if each component is a (local) martingale (resp. semimartingale).

The following lemma is a well known result and gives a sufficient condition for a local martingale to be a true martingale, see e.g. [CAH17]. It is stated for dimension $d = 1$ which implies the general case.

Lemma 3.6. *Let M be a $\mathbb{R}$-valued local martingale such that for all $t \geq 0$,*

$$\mathbb{E}\left[M_t^*\right] = \mathbb{E}\left[\sup_{0\leq s\leq t} |M_s|\right] < \infty. \tag{3.6}$$

Then M is a true martingale.

Proof. First note that since for all $t \geq 0$, we have $|M_t| \leq \sup_{0\leq s\leq t} |M_s|$, the condition $\mathbb{E}\left[|M_t|\right] < \infty$ is satisfied for all $t \geq 0$. Further, let τ_n be a localizing sequence of M. Hence, for each n, M^{τ_n} is a true martingale. For all stopping times $\tau \in [0, t]$ a.s. we have

$$|M_\tau| \leq \sup_{0\leq s\leq t} |M_s|, \tag{3.7}$$

which due to $u \wedge \tau_n \in [0, t]$ for $0 \leq u \leq t$ justifies the application of the dominated convergence theorem yielding

$$\begin{aligned}\mathbb{E}\left[M_t \mid \mathcal{F}_s\right] &= \mathbb{E}\left[\lim_{n\to\infty} M_{t\wedge\tau_n} \mathcal{F}_s\right] = \lim_{n\to\infty} \mathbb{E}\left[M_{t\wedge\tau_n} \mid \mathcal{F}_s\right] \\ &= \lim_{n\to\infty} M_s^{\tau_n} = M_s.\end{aligned} \tag{3.8}$$

Hence, M is a true martingale. □

3.2 Processes of locally bounded variation

This section intends to give a brief overview regarding processes (including deterministic functions) that are càdlàg and have finite variation on compacts. These processes play an important role in the theory of semimartingales and we do not aim for a complete overview but rather want to use this opportunity to introduce some additional notation. We refer the reader to [JS13] for a comprehensive overview that exhausts all the points we make in this section. In particular we refer the reader to this reference for the proofs of the statements we make in this section.

We denote with $\mathcal{V}^+$ the set of all $\mathbb{R}$-valued processes A that are càdlàg , adapted and with $A_0 = 0$ such that for all paths, $t \mapsto A_t(\omega)$ is nondecreasing. Denote by $\mathcal{V} = \mathcal{V}^+ \ominus \mathcal{V}^+$. In particular we get for each $A \in \mathcal{V}$ that for arbitrary finite $T \geq 0$ we have for all $\omega \in \Omega$, that

$$\bigvee_{[0,T]} (A)(\Omega) < \infty,$$

where $\bigvee_{[0,T]}(A)(\omega)$ is the total variation of the path $t \mapsto A_t(\omega)$ on the interval $[0,T]$, compare [JS13, Proposition I.3.3]. We will sometimes refer to these functions as of finite variation (FV).

Next we introduce the set of functions as above that satisfy an additional integrability condition. Define $\mathcal{A}^+$ to be the set of functions $A \in \mathcal{V}^+$ such that

$$\mathbb{E}\left[A_\infty\right] < \infty.$$

Similarly, define $\mathcal{A}$ to be the set of functions $A \in \mathcal{V}$, such that

$$\mathbb{E}\left[\lim_{T\to\infty} \bigvee_{[0,T]} (A)\right] < \infty.$$

Again by [JS13, Proposition I.3.3], we get $\mathcal{A} = \mathcal{A}^+ \ominus \mathcal{A}^+$. Overall we get the following inclusions:

$$\mathcal{A}_{\text{loc}} = \mathcal{A}^+_{\text{loc}} \ominus \mathcal{A}^+_{\text{loc}}, \quad \mathcal{A}^+ \subset \mathcal{A}^+_{\text{loc}} \subset \mathcal{V}^+, \quad \mathcal{A} \subset \mathcal{A}_{\text{loc}} \subset \mathcal{V}.$$

It is easy to see that for deterministic $A \in \mathcal{V}$ one gets $A \in \mathcal{A}_{\text{loc}}$. A more general version of this result is given by [JS13, Lemma I.3.10] that we state here for the readers convenience.

Lemma 3.7. *For a predictable process $A \in \mathcal{V}$ there exists a localizing sequence (τ_n) such that $(\bigvee_{[0,\cdot]}(A))^{\tau_n} = (\bigvee_{[0,\cdot\wedge\tau_n]}(A)) \leq n$. In particular we get $A \in \mathcal{A}_{loc}$.*

Further, we say a function (process) that takes values in $\mathbb{R}^{n_1\times n_2}$ is in $\mathcal{A}$, if $f(\cdot)_{i,j} \in \mathcal{A}$ holds for all i,j. We do the same for $\mathcal{V}, \mathcal{V}^+$ etc. Finally, let us define $\mathcal{V}^+_{\not\Omega}$ as the set of all deterministic, càdlàg and nondecreasing functions A with $A_0 = 0$. Hence, by identifying A as an element of $\mathcal{V}^+$ that is constant on Ω, we have $\mathcal{V}^+_{\not\Omega} \subset \mathcal{V}^+$. In the same way, define $\mathcal{V}_{\not\Omega} = \mathcal{V}^+_{\not\Omega} \ominus \mathcal{V}^+_{\not\Omega}$.

Since a process $A \in \mathcal{V}$ defines a signed measure for each $\omega \in \Omega$, we can define the integral

$$\int_{(0,t]} f(u)A(du) = (f \bullet A)_t, \tag{3.9}$$

for an appropriate function f. We always define such an integral in a Lebesgue-Stieltjes sense and not as a Riemann-Stieltjes integral. The same notation will be used for the stochastic integral in a semimartingale setting, and we will switch between both notations in (3.9) whenever we believe it to be preferable.

Chapter 4

Matrix evolution systems and measure integral equations

This chapter is intended to give an introduction to the concept of matrix evolution systems (MESs) and their special case in terms of homogeneity in form of matrix semigroups (MSGs). We start by discussing the semigroup case including the concept of infinitesimal generators. As certain regularity properties that are already implied under very mild assumptions for the semigroup case are lost when dealing with the general evolution system setting - making it unclear how the concept of infinitesimal generators can be extended beyond homogeneity - we show how a generalizing notion of ***extended generators*** can be established, that coincides in the homogeneous case with the classical definition and is applicable in situations, where a classical generator does not even exist.

After having established first results for infinitesimal generators of MESs, we will continue with showing how such evolution systems are related to certain integral equations through their extended generators. This will equip us with the tools needed to study a class of stochastic processes we call ***polynomial processes*** and certain types of semimartingales we call ***polynomial semimartingales***. In particular we can then show that extended generators of these stochastic processes are tightly connected to the extended generators of evolution systems induced by these processes.

Before we start, let us first give a motivation for the study of evolution systems and by that semigroups as a special case. This should only be regarded as an informal exposition, we will formalize these concepts in their respective sections.

Assume we have a deterministic dynamical system starting at some configuration $x \in E$ for some state vector space E at time $s \geq 0$ such that the state of the system at time $t \geq s$ is described by the function $\phi(s, x; t)$. If we denote that state at t with y, we have $y = \phi(s, x; t)$. Now imagine we consider the state of the system at $t + u$, $u \geq 0$, again starting in x at time s, we get that the state z satisfies $z = \phi(s, x; t + u)$. Then we can "split" the evolution of the system into two parts, the first being that we start at s in x end evolve until time t. The second that we start in y at time t and evolve until time $t+u$, landing in z. Hence, this flow property leads to the following well established flow equations

$$\phi(s,x;t+u) = \phi(t,\phi(s,x;t),t+u), \quad 0 \leq s \leq t,\, u \geq 0,\, x \in E.$$

When we consider evolution systems, we assume $x \mapsto \phi(s,x;t)$ to be a linear map from E into itself for any $0 \leq s \leq t$. Hence one assumes to have a family of linear operators τ indexed by (s,t) with $s \leq t$ and $\tau_{s,t} : E \to E$ such that $\phi(s,x;t) = \tau_{s,t}x$. This yields the operator equation $\tau_{s,t+u} = \tau_{t,t+u} \circ \tau_{s,t}$ and considering that the system should remain unchanged when no time has passed, one has additionally $\phi(s,x;s) = x$ and by that $\tau_{s,s} = \mathrm{Id}_E$, the identity on E. When we say the system is ***homogeneous***, we mean by that that the evolution of the system only depends on the length of time it evolved, not at what time this evolution started. Hence one would have $\phi(s,x;t) = \phi(0,x;t-s)$. In the linear case, this will lead to the special case of semigroups where one would define a one-parameter family of linear operators by $T(t) := M(0,t)$. The flow property then translates to $T(s+t) = T(s)T(t) = T(t)T(s)$ and one would have $T(0) = \mathrm{Id}_E$.

Note that when studying polynomial processes or polynomial semimartingales, we make use of certain algebraic properties that allow to reduce the discussions to finite dimensions, or matrix evolution systems. Hence one can assume $E = \mathbb{R}^n$ and by fixing a basis of E, $\tau_{s,t}$ can be identified with a family of matrices $M(s,t) \in \mathbb{R}^{n\times n}$. Further we shall deviate from standard convention in the literature on evolution systems by requiring the flow property, or as we will refer to it from here on evolution property, for the transposed of the matrices, as that will be more convenient in our following discussions, see the formal definition of matrix evolution systems in the respective section.

Let us finish this prelude by fixing some notation. Throughout we fix some dimension $n \in \mathbb{N}$ and we denote the ring of $n \times n$ matrices over $\mathbb{R}$ by $M_n(\mathbb{R}) = (\mathbb{R}^{n\times n}, +, \cdot)$, with standard component wise addition " + " and matrix multiplication " $\cdot$ " where the multiplication "$\cdot$" is usually omitted in tradition of standard convention, i.e. unless stated otherwise, we always mean for two elements $L, J \in M_n(\mathbb{R})$ the matrix multiplication when we write $L\ J$. The identity matrix is denoted by Id_n. With $\mathrm{GL}_n \subset M_n(\mathbb{R})$ we denote the general linear group, i.e. all those matrices in $M_n(\mathbb{R})$ that are invertible.

Further, define the index set $\mathbb{T}$ by

$$\mathbb{T} = \left\{(s,t) \in \mathbb{R}^2 : 0 \leq s \leq t\right\}.$$

4.1 Matrix semigroups

This section intends to introduce matrix semigroups (MSGs) and give a first set of results for these. These results, together with proofs and further references can be found in

[EN99, Chapter 1]. We in particular want to name this reference for the general infinite dimensional case. The study of these MSGs and many more is subject of the theory of functional equations. It is worth mentioning that it was the french mathematician A. Cauchy who motivated the study of the scalar case (see [EN99, page 2] for reference and the translation):

> **Déterminer la fonction $\phi(x)$ de manière qu'elle reste continue entre deux limites réelles quelconques de la variable x, et que l'on ait pour toutes les valeurs réelles des variables x et y**
>
> $$\phi(x + s) = \phi(x)\phi(y).$$
>
> **(A. Cauchy, 1921)**

Note that $\phi(0) = 1$ is not required and indeed this would follow already when excluding $\phi(x) \equiv 0$. Analogously, in the matrix case, one would not need to assume $T(0) = \mathrm{Id}_E$, provided one assumes there exists a $t \geq 0$ such that $T(t) \in \mathrm{GL}_n$. This is due to the fact that one has for any $x \in E$

$$T(t)x = T(t)T(0)x,$$

which due to $T(t) \in \mathrm{GL}_n$ implies $T(0)x = x \ \forall x \in E$, hence $T(0) = \mathrm{Id}_n$. As we are primarily interested in evolution systems, where $T(0) = \mathrm{Id}_n$ is a priory satisfied, we shall include this property in our first definition.

Definition 4.1. *We call $T : \mathbb{R}_+ \to M_n(\mathbb{R})$ a matrix semigroup (MSG) (in dimension n or on $M_n(\mathbb{R})$), if*

- $T(0) = \mathrm{Id}_n$,
- $T(s + t) = T(s)T(t)$.

If $T(t) \in \mathrm{GL}_n$ for all $t \geq 0$ and $t \mapsto T(t) \in \mathcal{C}(\mathbb{R}_+, M_n(\mathbb{R}))$, we call T **regular**. *In that case, T can always be extended to a group over $\mathbb{R}$ by setting $T(-t) = T(t)^{-1}$ for $t \geq 0$.*

Let us establish some first preliminary results regarding regularity of the map $t \mapsto T(t)$. The following proposition shows that T is already regular under mild assumptions.

Proposition 4.2. *Let T be a MSG on $M_n(\mathbb{R})$ and assume $t \mapsto T(t)$ is right continuous at $t = 0$. Then T is regular.*

Proof. First note that by right continuity, there is a $\varepsilon^* > 0$ such that $T([0, \varepsilon^*]) \subset \mathrm{GL}_n$, since GL_n is an open subset of $M_n(\mathbb{R})$. But that actually implies that $T(t) \in \mathrm{GL}_n$ for all $t \geq 0$. To see that, assume

$$t^* = \inf \{t > 0 : T(t) \notin \mathrm{GL_n}\} < \infty.$$

Note that it holds $t^* \geq \varepsilon^*$. Otherwise there exists a $t \in (t^*, \varepsilon^*)$ s.t. $T(t) \notin \mathrm{GL}_n$ by the very definition of t^*. As $T(\varepsilon^*) = T(t)T(\varepsilon^* - t)$, this implies $T(\varepsilon^*) \notin \mathrm{GL}_n$ which is a contradiction to the choice of ε^*, hence we have $t^* \geq \varepsilon^*$. Then for $0 < \varepsilon \leq \varepsilon^*$, we have $T(t^* - \varepsilon) \in \mathrm{GL}_n$ and $T(\varepsilon) \in \mathrm{GL}_n$. Hence, because GL_n is a group wrt the matrix multiplication, we get

$$T(t^*) = T(t^* - \varepsilon)T(\varepsilon) \in \mathrm{GL_n} .$$

By the assumption $t^* < \infty$, we have that $T(t^* + \varepsilon) \notin \mathrm{GL_n}$ for $\varepsilon > 0$ small enough. But now we continue with

$$T(t^*)T(\varepsilon) = T(t^* + \varepsilon),$$

which implies $T(t^* + \varepsilon) \in \mathrm{GL_n}$, which is a contradiction to the definition of t^*. Hence we get $\{t \geq 0 : T(t) \notin \mathrm{GL_n}\} = \emptyset$. Therefore, we can extend T to a group by setting $T(-t) = T(t)^{-1}$ for $t \geq 0$. Let $t > 0$. We first show left continuity. Note that we have for $\varepsilon > 0$ small,

$$T(t) = T(t - \varepsilon)T(\varepsilon).$$

By assumption, $T(\varepsilon) \to \mathrm{Id}_n$ as $\varepsilon \downarrow 0$. This yields $T(t) = \lim_{\varepsilon \downarrow 0} T(t - \varepsilon)$, which shows left continuity. Next we show right continuity, where we shall make use of the extension to a group. It holds

$$T(t) = T(t)T(\varepsilon)T(-\varepsilon) = T(t + \varepsilon)T(-\varepsilon) = T(t + \varepsilon)\left(T(\varepsilon)\right)^{-1} .$$

Since the function $A \mapsto A^{-1}$ is continuous on GL_n, right continuity follows by the same arguments we made for left continuity. As $t > 0$ was arbitrary, this concludes the proof. $\square$

The next theorem is a very well known fact from MSG theory. It shows that regular MSGs are of a particular form. In particular, in combination with the previous result, it shows the following chain of implications (compare [EN99, section 1.6]):

$$\text{right continuity at } t = 0 \Rightarrow T \text{ is regular} \Rightarrow \text{ The map } t \mapsto T(t) \text{ is smooth.}$$

As Engel and Nagel noted in [EN99, section 1.6], this can be seen as a contribution to the second part of Hilbert's fifth problem. There, one can also find the following quote by David Hilbert, together with a translation into the English language:

> **Überhaupt werden wir auf das weite und nicht uninteressante Feld der Funktionalgleichungen geführt, die bisher meist nur unter Voraussetzung der Differenzierbarkeit der auftretenden Funktionen untersucht worden ist. Insbesondere die von Abel mit so vielem Scharfsinn behandelten Funktionalgleichungen, die Differenzengleichungen und andere in der Literatur vorkommende Gleichungen weisen an sich nichts auf, was zur Forderung der Differenzierbarkeit der auftretenden Funktionen zwingt. (...) In allen Fällen erhebt**

> **sich daher die Frage, inwieweit etwa die Aussagen, die wir im Falle der Annahme differenzierbarer Funktionen machen können, unter geeigneten Modifikationen ohne diese Voraussetzung gültig sind.**
>
> **(D. Hilbert, 1900, see [Hil70])**

We now state the theorem that yields smoothness under mild assumptions. We note that we do not state the most general form of the following theorem as one can already establish strong results by just assuming $t \mapsto T(t)$ is measurable, see Exercise 1.7 in [EN99]. As we are interested in using MSGs and MESs primarily in the study of càdlàg processes, it suffices for our purposes to assume more regularity.

Theorem 4.3. *Let T be a regular MSG on $M_n(\mathbb{R})$. Then there exists a unique matrix $G \in M_n(\mathbb{R})$ such that we have*

$$T(t) = \exp(tA), \quad \forall t \geq 0.$$

In reverse, any function T that is defined by the above matrix exponential for an arbitrary $A \in M_n(\mathbb{R})$, is a regular MSG.

Proof. The reverse statement is just a consequence of elementary properties of the matrix exponential, for the proof see Proposition 2.3 in [EN99]. The first direction is proved by Theorem 2.9 in [EN99]. To be complete we should note that by that theorem, one can only deduce the existence of a matrix $A \in M_n(\mathbb{C})$. However, as $M_n(\mathbb{R})$ is a closed subset of $M_n(\mathbb{C})$, and since by Proposition 2.8 in [EN99] one has

$$A = \lim_{\varepsilon \downarrow 0} \frac{T(\varepsilon) - T(0)}{\varepsilon}, \tag{4.1}$$

it immediately follows that $A \in M_n(\mathbb{R})$. Regarding uniqueness, assume there are $A_1, A_2 \in M_n(\mathbb{R})$ with $T(t) = \exp(tA_1) = \exp(tA_2)\, \forall t \geq 0$. Hence (4.1) implies $A_1 = A_2$ which finishes the proof. □

The matrix A in the theorem above plays an important role in the study of MSGs as it allows for an infinitesimal description of T. We want to discuss this a little bit further. First, let us make the following definition, compare Definition 2.4 in [EN99].

Definition 4.4. *Given a regular MSG T with $T(t) = \exp(tA)$, we call A the (infinitesimal) generator of T. Conversely, we say T is the regular MSG generated by A.*

To make clear what we mean by infinitesimal description of T by A, let us state the following result we have already cited above.

Proposition 4.5. *Given the a MSG $T(t) = \exp(tA)$, it holds $T \in \mathcal{C}^\infty(\mathbb{R}_+, M_n(\mathbb{R}))$ and*

$$\begin{cases} \frac{d}{dt}T(t) = AT(t) = T(t)A, \ \textit{for } t \geq 0, \\ T(0) = \mathrm{Id}_n . \end{cases}$$

Proof. This is just [EN99, Proposition 2.8]. Note that from that proposition we get $\frac{d}{dt}T(t)|_{t=0} = A$. Hence it holds

$$AT(t) = \left(\lim_{\varepsilon\downarrow 0} \frac{T(\varepsilon) - \mathrm{Id}_n}{\varepsilon}\right) T(t) = \lim_{\varepsilon\downarrow 0} \frac{T(t+\varepsilon) - T(t)}{\varepsilon} = T(t)A,$$

which is a well known fact for MSGs. □

In the next section we shall study the inhomogeneous case in form of evolution systems. There, unlike above, the regularity assumptions we shall make do not imply the evolution system to be differentiable or even continuous. Nonetheless we will be able to to define a concept of infinitesimal description in form of what we will call the ***extended generator*** of an evolution system. The idea is to look at

$$\frac{d}{dt}T(t) = T(t)A \tag{4.2}$$

as the defining property of the extended generator where the time derivative is replaced with a more general Radon-Nikodým derivative where the measure that replaces the Lebesgue measure is connected to the discontinuities of the evolution system. We will formalize this in the following section.

4.2 Matrix evolution systems

This section can be seen as the time inhomogeneous counter part to the previous section. Instead of MSGs, we will consider matrix evolution systems. Evolution systems have been treated in the literature for a long time in an infinite dimensional setting, see in particular [EN99, Chapter 9] where such systems are introduced in the study of non-autonomous Cauchy problems and [Böt14] where generators are defined by left and right limits corresponding to the semigroup case for the study of time inhomogeneous Markov processes. We also refer the reader to the plenty of references given in those two sources. In both cases, strong continuity assumptions are made and a key tool in the analysis is the transformation to the homogeneous setting by means of space-time homogenization of either the evolution system stemming from a partial differential equation to a semigroup, compare [EN99, Lemma 9.10] or a time inhomogeneous Markov process to a homogeneous one, see [EN99, Section 3].

Our situation however is different since we are interested in evolution systems as a tool to study a certain class of stochastic processes we call ***polynomial processes***. First, we want to include cases beyond strongly continuous evolution systems. Second, the stochastic processes considered have a defining algebraic property, that allows it to consider finite dimensional evolution systems. Hence, a space-time homogenization would force us to leave the finite dimensional setting. It is however this finite dimensionality that we are interested in as it allows us to consider (measure) ordinary differential equations rather then (measure) partial differential equations for polynomials as ***initial conditions***, see e.g. the step *(iii)* $\Rightarrow$ *(i)* in the proof of [Cuc+12, Theorem 2.7] where this is done in the homogeneous case with strong time derivatives rather then Radon-Nikodým derivatives.

As mentioned already, we do not want to assume continuity for our evolution systems we will shortly introduce. In the study of càdlàg processes, as we intend to do, it seems more appropriate to consider such evolution systems that only satisfy càdlàg properties as we introduce further down in Definition 4.10. The following subsection will give a short summary of results for the inhomogeneous version of Cauchy's functional equation. There the reader will see a full characterization in the spirit of Theorem 4.3. Unfortunately, these results will require assumptions that we do not wish to incorporate a priory and indeed we will see that they do not hold valid in all cases. Hence, the reader may skip the following subsection as it is not needed for the discussions that follow it.

4.2.1 Cauchy-Sincov-Equations

This subsection only intends to present results to the inhomogeneous version of Cauchy's multiplicative functional equation known as Cauchy-Sincov functional equations. We want to mention that there are even further generalizations in form of Cauchy-Pexider-Sincov equations where the binary operation between two elements of the range of the respective functions (in the above setting this was the matrix multiplication) is substituted with a general function.

For a more detailed exposition, we refer the reader to the excellent book of Aczél [Acz66], where the theory of functional equations is treated to a greater extent with many more functional equations that go beyond the Cauchy types, see [Acz66, Section 2.1].

We consider the following functional equation. Let A be an arbitrary set and G a set on which there is a binary operation $\circ : G \times G \to G$. Consider a function $F : A \times A \to G$. The Cauchy-Sincov functional equation is given by

$$F(x, y) \circ F(y, z) = F(x, z), \text{ for all } x, y, z \in A, \tag{C-S FE}$$

and the goal is to characterize all those functions F that satisfy (C-S FE). In the context of matrix evolution systems, the above setting is different since A can be an arbitrary set. When dealing with evolution systems, we consider $A = \mathbb{R}_+$ and additionally require $x \leq y$ in $F(x, y)$, or $(x, y) \in \mathbb{T}$. On the other hand, recall that we have seen in the MSG case that $T(t) \in \mathrm{GL}_n$ under very mild assumptions. Hence, if we assume G to be a group wrt $\circ$, $A = \mathbb{R}_+$ and denote for $g \in G$ the inverse element by g^{-1}, we can extend F to $\mathbb{R} \times \mathbb{R}$ by setting $F(y, x) := F(x, y)^{-1}$ for $x \leq y$ and using $F(-x, y) = F(-x, 0)F(0, y) = F(0, x)^{-1} \circ F(0, y)$ for $x, y \geq 0$. This extended F still satisfies (C-S FE), and one has

$$F(x, y) = F(x, x) \circ F(x, y) = F(x, y) \circ F(y, y),$$

yielding that $F(x,x) = F(y,y) = e$, where e is the neutral element of G. In particular, one has for $s \leq u \leq t$ that

$$F(s,t) \circ F(t,u) = F(s,t) \circ F(u,t)^{-1} = F(s,u) \circ F(u,t) \circ F(u,t)^{-1} = F(s,u).$$

It is this very assumption that F takes values in a group G, allowing for an ***inverse operation***, that yields a full characterization of solutions to (C-S FE). The following theorem can be found in [Acz66, Theorem 2 in Section 8.1].

Theorem 4.6. *Assume G is a group with group operation $\circ$ and A an arbitrary set such that $\forall x, y \in A$ it holds $F(x,y) \in G$. Then all solutions to*

$$F(x,y) \circ F(y,z) = F(x,z) \text{ for all } x,y,z \in A$$

are of the form

$$F(x,y) = n(x)^{-1} n(y),$$

for some function $n : A \to G$.

Note that the above theorem does not assume any regularity on F, only the group structure is important. Further down, we see that it is desirable to have an evolution system that takes values in a group, namely GL_n as that simplifies the definition of the extended generator. The hope is therefore, that by posing regularity assumptions wrt the domain of F, it is possible to show that F automatically takes values in a group, compare Proposition 4.2. However, we will show that there are examples where despite our regularity assumptions (càdlàg properties of $x \mapsto F(x,y)$ and $y \mapsto F(x,y)$), F will not always lie in a group.

Nonetheless there is a result that guaranties that G is indeed a group. Unfortunately, the conditions are too strong for us so we only state this theorem for completeness. For reference and proof, see [Acz66, Theorem 3 in Section 8.1].

Theorem 4.7. *If a function $F(x,y)$ is defined for $x, y \in A$ and if its values lie in a set G such that there is a binary operation $\circ$ between two elements $F(x,y), F(y,z)$ (i.e. the second argument of the left coincides with the first of the right), for which for any $x, y, z \in A$ the functional equation*

$$F(x,y) \circ F(y,z) = F(x,z)$$

is satisfied, and such that we have

(a) For all $x \in G$ fixed, range$(F(x,\cdot)) = G$,

(b) there exists an element $a \in G$, such that range$(F(\cdot,a)) = G$,

then G is a group wrt the operation $\circ$.

4.2.2 Matrix evolution systems revisited

We will now begin with our detailed discussions of matrix evolution systems (MESs), and introduce the concept of extended generators that will play a central role when considering polynomial processes in Chapter 6. Recall that

$$\mathbb{T} = \{(s,t) \in \mathbb{R}^2 : 0 \leq s \leq t\}.$$

We shall define MESs as functions of two parameters (such that the pair is in $\mathbb{T}$) that take values in $M_n(\mathbb{R})$ such that they are solutions to (C-S FE) together with the property that when both parameters coincide, we get the neutral element of $M_n(\mathbb{R})$, compare with the remarks made right after (C-S FE). Let us make this formal.

Definition 4.8. *We call a function $P : \mathbb{T} \to M_n(\mathbb{R})$ with $(s,t) \mapsto P(s,t)$ a* **matrix evolution system** *(in dimension n), if*

(i) $P(s,s) = \mathrm{Id}_n$ for all $s \in [0,\infty)$,

(ii) For any $0 \leq s \leq u \leq t$, P satisfies

$$P(s,u)P(u,t) = P(s,t).$$

Note that we do not need $P(s,t)$ to be defined if $t < s$ which would yield the existence of inverse elements, implying that P takes values in GL_n. Further down, we will see that even after the additional regularity assumptions we will make, it is possible for P to become singular, see Example 4.24.

The following shows how MESs are a generalization of MSGs.

Definition 4.9. *A matrix evolution system P is called* **homogeneous** *if for any $0 \leq s \leq t$,*

$$P(s,t) = P(0,t-s).$$

In other words, the study of P can be reduced to the study of the one parameter semigroup T defined via

$$T(t) = P(0,t).$$

The semigroup properties hold for obvious reasons. As we have seen in section 4.1, assuming that the map $t \mapsto T(t)$ is right continuous at $t = 0$ implies

$$T(t) = \exp(tA),$$

for a unique $A \in \mathrm{M}_n(\mathbb{R})$ that we referred to as generator. We now want to introduce the regularity assumptions we shall make for MESs. The goal is to define a reasonable concept of generator suited for our purposes when studying polynomial semimartingales.

Definition 4.10. *We call a matrix evolution system P* **càdlàg in the first coordinate**, *if for all $t \geq 0$, $s \mapsto P(s,t)$ is càdlàg on $[0,t]$. We call it* **càdlàg in the second coordinate**, *if for all $s \geq 0$, $t \mapsto P(s,t)$ is càdlàg on $[s,\infty)$.*

We call P **càdlàg**, *if it is càdlàg in the first and in the second coordinate.*

We can immediately give the following relation between these properties.

Proposition 4.11. *Let P be a matrix evolution system which is càdlàg in the second coordinate. Then for $t > 0$ fixed, the function $s \mapsto P(s,t)$ is right continuous on $[0,t)$.*

Proof. We need to show for any $s \in [0,s)$, that

$$\lim_{\varepsilon \downarrow 0} P(s+\varepsilon,t) = P(s,t)$$

holds. By the evolution system property we have for any $\varepsilon > 0$,

$$P(s,t) = P(s,s+\varepsilon)P(s+\varepsilon,t).$$

Since $P(s,s) = \mathrm{Id_n}$, and by the right continuity in the second variable, we have that there is a $\varepsilon > 0$ such that $P(s,s+\tilde{\varepsilon}) \in \mathrm{GL}_n \; \forall \tilde{\varepsilon} \leq \varepsilon$. Hence we get

$$\lim_{\varepsilon \downarrow 0} P(s+\varepsilon,t) = \lim_{\varepsilon \downarrow 0} P(s,s+\varepsilon)^{-1}P(s,t).$$

The claim now follows by the fact that $(\cdot)^{-1}$ is continuous on GL_n and $t \mapsto P(s,t)$ is right continuous by assumption. □

The following example shows that the previous proposition can not be improved in the sense of existence of left limits.

Example 4.12. *Consider the following evolution system ρ in dimension $n = 1$ given by*

$$\rho(s,t) = \frac{\eta(t)}{\eta(s)},$$

where we have $\eta(x) = (1-x)\mathbb{1}_{\{x<1\}} + \mathbb{1}_{\{x \geq 1\}}$. It is then easy to see that this evolution system is càdlàg in the second coordinate but fails to have a left limit in the first coordinate at $s = 1$ for $t = 1$ since $\rho(1-\varepsilon,1) = 1/\varepsilon$ which does not exist as $\varepsilon \downarrow 0$.

Finally, let us note that if an evolution system that is càdlàg in the second coordinate fails to have left limits in a certain point, then this is equivalent to a slightly stronger statement as illustrated by the following proposition that is an immediate consequence of the evolution property.

Proposition 4.13. *Let P be a matrix evolution system càdlàg in the second coordinate. Then the following two statements are equivalent:*

(i) The limit $P(s-, s)$ does not exist,

(ii) There is a $\delta > 0$, such that the limit $P(s-, t)$ does not exist for all $t \in [s, s + \delta)$.

Proof. The direction $(ii) \Rightarrow (i)$ is obviously true. For the other direction, note that

$$\lim_{s \downarrow 0} P(s - \varepsilon, t) = \left(\lim_{\varepsilon \downarrow 0} P(s - \varepsilon, s) \right) P(s, t).$$

Hence, by right continuity in the second coordinate, there exists a $\delta > 0$ such that $P(s, t) \in \mathrm{GL}_n$. This implies $\neg(ii) \Rightarrow \neg(i)$", which finishes the proof. $\square$

We now want to motivate the next definition regarding regularity of P by the following observations. Consider again a regular MSG T. As we have seen in section 4.1, we have

$$\dot{T}(t) = \frac{dT(t)}{dt} = T(t)A. \tag{4.3}$$

Let us collect a few facts about this result. First note that the assumptions we made give us smoothness for $t \mapsto T(t)$. Second, we get that $T(t) \in \mathrm{GL}_n \forall t \geq 0$ which actually allows us to extend T to a group indexed by all of $\mathbb{R}$. Finally, note that we can write (4.3) in integral form as

$$T(t) = \mathrm{Id_n} + \int_0^t T(u)A \, du. \tag{4.4}$$

Further, note that the generator A can be represented by

$$A = T(t)^{-1} \dot{T}(t),$$

which will become useful in a more general setting where (4.3) can not be applied anymore.

Before we introduce the concept of extended generators for MESs, let us first extend the definition of (infinitesimal) generators to MESs. We shall define these under very strong regularity assumptions.

Definition 4.14. *Let P be a MES such that for each $s \geq 0$, we have $t \mapsto P(s, t)$ is right differentiable. Define the function $f : \mathbb{T} \to M_n(\mathbb{R})$ by*

$$f(s, t) := \lim_{\varepsilon \downarrow 0} \frac{P(s, t + \varepsilon) - P(s, t)}{\varepsilon}.$$

We call the function $G : \mathbb{R}_+ \to M_n(\mathbb{R})$ defined by $G(s) := f(s, s)$ the **(infinitesimal) generator** *of P.*

Remark 4.15. *Note that by the evolution property we have $f(s,t) = P(s,t)G(t)$. If we assume G to be continuous, this implies $t \mapsto P(s,t)$ to be continuously right differentiable, meaning that $t \mapsto f(s,t)$ is continuous. Then we would have*

$$P(s,t) - P(s,u) = \int_u^t f(s,u)du,$$

where the integral above is a Riemann integral. If we were to replace G continuous with bounded, we could extend this definition to the absolute continuous setting (wrt the Lebesgue integral). This idea is contained in the next definition as a special case.

Let us revisit the semigroup case. The points we made regarding regularity can also be formulated in the following way. As T is smooth, we have $T \in \mathcal{V}$. By defining $\gamma((s,t]) := T(t) - T(s)$ one can therefore define a signed vector measure γ on $\mathbb{R}_+$. From (4.4) one then gets $\gamma \ll dt$ where dt is the Lebesgue measure. This motivates the next notion of regularity.

Definition 4.16. *We call a matrix evolution system P* **regular in the second coordinate**, *if it is càdlàg in the second coordinate and the function $t \mapsto P(s,t)$ has finite variation on $[s,T]$ for all $T \in (s,\infty)$. If P is even càdlàg in the sense of Definition 4.10, we call it* **regular**.

The finite variation property for regular evolution systems implies that $P(s,\cdot)$ can be associated with a signed (vector) Borel measure γ_s on (s,∞), i.e. for $0 \le s \le u \le t$,

$$\gamma_s((u,t]) = P(s,t) - P(s,u). \tag{4.5}$$

For $i,j \in \{1,\dots,n\}$, we denote by $\gamma_s^{i,j}$ the i,j-component of γ_s, i.e. $\gamma_s^{i,j}((u,t]) = (P(s,t))^{i,j} - (P(s,u))^{i,j}$.

By Equation (4.3), it holds that $T(t)$ is differentiable. Here we are interested in measures which are only absolutely continuous with respect to some fixed measure, hence the following definition.

Definition 4.17. *Let μ be a σ-finite measure on $\mathbb{R}_{>0}$ and $t > 0$. The family of measures $\gamma = (\gamma_s)_{s\ge 0}$ introduced in* (4.5) *is called* **dominated by $\boldsymbol{\mu}$ until $\boldsymbol{t}$**, *if for all $i,j \in \{1,\dots,n\}$, we have*

$$(\gamma_u)_{i,j} \ll \mu|_{(u,\infty)}, \quad \forall 0 \le u \le t.$$

We write $\gamma \overset{t}{\ll} \mu$ if the family of measures γ is dominated by μ until t. If this is the case for all $t > 0$, we simply write $\gamma \overset{\infty}{\ll} \mu$.

Remark 4.18. *If one assumes $P(s,t) \in \mathrm{GL}_n$ for all $s \le t$, then the above definition would simplify due to the following fact. Because,*

$$\gamma_{s'}((u,t]) = P(s',t) - P(s',u) = P(s',s)\left(P(s,t) - P(s,u)\right) = P(s',s)\gamma_s((u,t]),$$

for $0 \le s' \le s \le u \le t$, it would suffice to require $\gamma_0 \ll \mu$ component wise which would then imply $\gamma \overset{\infty}{\ll} \mu$, since $\mu((u,t]) = 0$ implies $\gamma_0((u,t]) = 0$ which is equivalent to $\gamma_s((u,t]) = 0$ since $P(0,s) \in \mathrm{GL_n}$.

Let $A \in \mathcal{V}^+_{\mathfrak{A}}$. Then A defines a measure on $\mathbb{R}_{>0}$ which we denote by dA. Note that with expressions of the form

$$\int f(u)dA_u = \int f(u)A(du)$$

we always mean the Lebesgue-Stieltjes integral of a measurable function f against the measure dA, not the Riemann-Stieltjes integral of a sufficiently regular function f against the right-continuous function A.

Definition 4.19. *We call a collection (P, γ, A) a* **proper matrix evolution system** *if P is a regular matrix evolution system, γ is the family of vector-measures induced by P as above and $A \in \mathcal{V}^+_{\mathfrak{A}}$ such that $\gamma \overset{\infty}{\ll} dA$.*

We are now able to introduce the concept of ***extended generators*** for MESs which in light of (4.4) can be seen as a generalization of generators for MSGs.

Definition 4.20. *Let $G : \mathbb{R}_{>0} \to M_n(\mathbb{R})$ be a a Borel-measurable function. We call G the* **extended generator** *of the proper MES (P, γ, A), if for any $s \leq t$ we have*

$$P(s,t) = \mathrm{Id_n} + \int_{(s,t]} P(s, u-)G(u)A(du).$$

Due to the evolution system property it is clear that the above condition follows if it is satisfied for $s = 0$. We further will say G is the **extended $\boldsymbol{A}$-generator** *of P, if G is the extended generator of (P, γ, A) for the measure γ induced by P.*

The definition above requires G to satisfy the A-a.e. identity

$$\frac{d\gamma_s}{dA}(\cdot) = P(s, \cdot-)G(\cdot) \tag{4.6}$$

on (s, ∞), hence a family $\tilde{G}$ that is A-a.e. equal to G is also a generator. We will deal with existence and uniqueness for G further down.

In the following we will see that unlike in the semigroup case, our càdlàg assumptions do not guaranty that $P(s,t) \in \mathrm{GL}_n$ for all $0 \leq s \leq t$. If we assume however, that $P(s,t) \in \mathrm{GL}_n \forall t \geq s$, then we can compute G on (s, ∞) simply by

$$G(\cdot) = P(s, \cdot-)^{-1}\frac{d\gamma_s}{dA}(\cdot), \tag{4.7}$$

for arbitrary version of $d\gamma_0/dA$. Note that the assumption also implies $P(s, t-) \in \mathrm{GL}_n$ as we will see further down when we deal with existence and uniqueness of G, see Proposition 4.23 and Proposition 4.31 below.

Remark 4.21. *Note that by Remark 4.15, extended generators are strict generalizations of infinitesimal generators. By that we mean that if G is an infinitesimal generator of P, then it is the extended generator of the proper MES (P, γ, A), where γ is the family of measures defined by $P(s, \cdot) \in \mathcal{V}_{\mathfrak{A}}$ for each $s \geq 0$ and $A_t = t$, while the reverse is not true as we will later see with Example 4.24.*

We will now collect a few facts about matrix evolution systems.

Lemma 4.22. *Let P be a càdlàg MES. Then, for $0 \leq s \leq u \leq t$, then*

(i) $P(s,u-)P(u-,t) = P(s,t)$, $P(s-,s)P(s,t) = P(s-,t)$ *and if additionally* $u < t$, *then* $P(s,u)P(u,t-) = P(s,t-)$,

(ii) for $s < u < t$ *we have* $P(s-,t-)$ *is well defined and* $P(s-,t-)P(t-,t) = P(s-,t)$ *as well as* $P(s,u-)P(u-,t-) = P(s,t-)$.

If there is a family of measure γ *and* $A \in \mathcal{V}^+_{\mathfrak{A}}$ *such that* (P,γ,A) *is a proper MES,*

(iii) then we have on (u,∞), *that* $\frac{d\gamma_s}{dA}(\cdot)|_{(u,\infty)} = P(s,u)\frac{d\gamma_u}{dA}(\cdot)$ *where the equality is in an* dA*-a.e. sense.*

Define the matrix valued function $\Gamma_s(\cdot)$ *with domain* (s,∞) *by*

$$\Gamma_s(\cdot) := P(s-,s)\frac{d\gamma_s}{dA}(\cdot). \tag{4.8}$$

Then,

(iv) the function $t \mapsto P(s-,t)$ *is FV over* $[s,T]$ *for any* $T > s$ *with* $\gamma_{s-} \ll A|_{(s,\infty)}$ *and with density* $\frac{d\gamma_{s-}}{dA} = \Gamma_s$. *Further it holds*

$$\frac{d\gamma_s}{dA}|_{(u,\infty)} = P(s,u-)\frac{d\gamma_{u-}}{dA} \text{ and } \frac{d\gamma_{s-}}{dA} = P(s-,s)\frac{d\gamma_s}{dA} \tag{4.9}$$

for $u > s$ *with the equalities are to be understood in a* A*-a.e. sense.*

Proof. Let $(\varepsilon_n)_n$ and $(\delta_n)_n$ be two positive zero sequences.

(i) Since the two limits $P(s,u-)$ and $P(u-,t)$ exist, and we have

$$P(s,t) = P(s,u-\varepsilon_n)P(u-\varepsilon_n,t)\,\forall n, \tag{4.10}$$

we get

$$P(s,t) = \lim_n \left(P(s,u-\varepsilon_n)P(u-\varepsilon_n,t)\right) = P(s,u-)P(u-,t). \tag{4.11}$$

The other statement follows similarly. Again we have

$$P(s-,s)P(s,t) = \lim_{n\to\infty}\left(P(s-\varepsilon_n,s)\right)P(s,t) = \lim_{n\to\infty}\left(P(s-\varepsilon_n,t)\right) = P(s-,t).$$

(ii) We need to show that the two limits satisfy

$$\lim_n \lim_m P(s-\delta_m, t-\varepsilon_n) = \lim_m \lim_n P(s-\delta_m, t-\varepsilon_n) \tag{4.12}$$

and define $P(s-,t-)$. Indeed we have for n,m large enough

$$\lim_n \lim_m P(s-\delta_m, t-\varepsilon_n) = \lim_n (\lim_m P(s-\delta_m, s)) P(s, t-\varepsilon_n) = P(s-,s)P(s,t-) \tag{4.13}$$

Likewise we have

$$\lim_m \lim_n P(s-\delta_m, t-\varepsilon_n) = \lim_m P(s-\delta_m, s) \lim_n P(s, t-\varepsilon_n) = P(s-,s)P(s,t-). \tag{4.14}$$

Using this, we get

$$P(s-,t-)P(t-,t) = P(s-,s)P(s,t-)P(t-,t) \overset{(i)}{=} P(s-,s)P(s,t) \overset{(i)}{=} P(s-,t).$$

Finally, we have using the previous results,

$$P(s,u-)P(u-,t-) = P(s,u-)P(u-,u)P(u,t-) = P(s,t-).$$

(iii) An application of the evolution property yields

$$\int_{(u,t]} P(s,u)\frac{d\gamma_u}{dA}(v)dA_v = P(s,u)(P(u,t) - P(u,u)) = P(s,t) - P(s,u), \tag{4.15}$$

hence we get that $P(s,u)\frac{d\gamma_u}{dA}$ is a version of $\frac{d\gamma_s}{dA}$ restricted to (u,∞).

(iv) We first show $t \mapsto P(s-,t)$ is locally FV. This is clear since we have by (i)

$$P(s-,t) = P(s-,s)P(s,t), \tag{4.16}$$

where $P(s-,s)$ is constant in t and $P(s,t)$ is locally FV by assumption. The claim $\gamma_{s-} \ll A_{(s,\infty)}$ follows from the fact that we have

$$\gamma_{s-}((u,t]) = P(s-,t) - P(s-,u) = P(s-,s)\gamma_s((u,\infty]),$$

for $s \le u \le t$. Hence, since we have $\gamma_s \ll A_{(s,\infty)}$ by assumption we get the claim. Finally, we have

$$\int_{(u,t]} P(s-,s)\frac{d\gamma_s}{dA}(u)dA_u = P(s-,s)\left(P(s,t) - P(s,u)\right) = P(s-,t) - P(s-,u).$$

This shows similarly to (iii) that $P(s-,s)\frac{d\gamma_s}{dA}$ is a version if $\frac{d\gamma_{s-}}{dA}$. $\square$

4.2.3 Existence and uniqueness of extended generators

We now want to discuss existence and uniqueness of the extended generator for a proper MES (P, γ, A). As noted before, if one assumes for fixed $s \geq 0$, that $P(s,t) \in \mathrm{GL_n}\, \forall t \geq s$, G exists on (s, ∞) since for an arbitrary version of $\frac{d\gamma_s}{dA}$ one has a version of G on (s, ∞) given by (4.7). As a first step, let us proof the following proposition that is a direct consequence of the group structure of GL_n.

Proposition 4.23. *Let P be a càdlàg MES. If for given $s \geq 0$, there exists $t^* > s$ such that we have $P(s,t) \in \mathrm{GL}_n \forall t \in (s, t^*)$, then also $P(s, t-) \in \mathrm{GL}_n$ and for $s < u \leq t < t^*$ we have $P(u-, t) \in \mathrm{GL}_n$.*

Proof. Note that we have by Lemma 4.22 (i) that

$$P(s,t) = P(s, u-)P(u-, t).$$

Since $P(s,t) \in \mathrm{GL}_n$, so are $P(s, u-)$ and $P(u-, t)$ (Just apply the determinant function to both sides to see that this must hold true). Hence the proof is complete. □

The following example shows that MESs can leave GL_n even if they are càdlàg and hence "start" in GL_n.

Example 4.24. *Consider the sets $M_i \subset \mathbb{R}^2$ defined for $i \in \mathbb{N}$ via*

$$M_i := \left\{(s,t) \in \mathbb{R}^2 : i - 1 \leq s \leq t < i\right\}.$$

Define $M := \bigcup_{i \in \mathbb{N}} M_i$. The set M is shown in Figure 4.1. Define $\phi(s,t) := \mathbb{1}_M(s,t)$. By construction of M, ϕ is a càdlàg MES on $M_n(\mathbb{R})$ for $n = 1$.[1]

However, for any s fixed, $0 = \phi(s,t) \notin \mathrm{GL}_1 = \{0\}$ if t satisfies $t > s + 1$. Likewise, for any càdlàg MES P with n arbitrary, $\tilde{P}(s,t) := \phi(s,t)P(s,t)$ is again a càdlàg MES for which $\tilde{P} \in \mathrm{GL}_n$ is violated.

The above example shows that it is not clear if for a proper MES (P, γ, A) an extended generator G exists on all $\mathbb{R}_{>0}$. However, by the càdlàg property we know that for any $s > 0$ there is an $\varepsilon_s > 0$ such that $P(s,t) \in \mathrm{GL_n}$ for $s \leq t \leq \varepsilon_s$. Further down, we present sufficient conditions based on this observation that imply the existence of extended generators, see Proposition 4.31 and Theorem 4.35. Regarding "leaving" GL_n, let us proof the following result.

Proposition 4.25. *Let P be a càdlàg MES. Fix some $s \geq 0$ and define t^* by*

$$t^* := \inf\left\{t > s : P(s,t) \notin \mathrm{GL_n}\right\}.$$

Then the following points hold true

[1] The author of this thesis would like to thank Dr. Lukas Steinberger for providing this example.

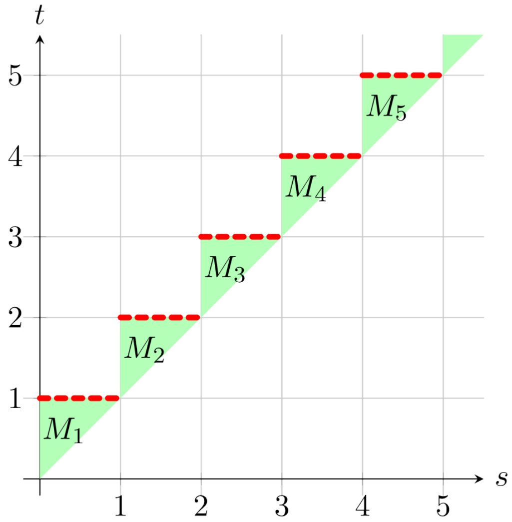


Figure 4.1: Graphical visualization of the set M. The red lines indicate that that line segment does not belong to M, e.g. $\{(s,1) : 0 \leq s < 1\} \cap M = \emptyset$.

(i) $P(s,t^*) \notin \mathrm{GL_n}$,

(ii) $P(s,u), P(s,u-), P(u-,u) \in \mathrm{GL}_n \forall u \in (s,t^*)$,

(iii) $P(u,t^*), P(u-,t^*) \notin \mathrm{GL_n}\ \forall u \in (s,t^*)$ *and* $P(t^*-,t^*) \notin \mathrm{GL_n}$,

(iv) $P(s,t^*+\varepsilon) \notin \mathrm{GL_n}\ \forall \varepsilon \geq 0$.

Proof. (i) Note that we have by the càdlàg property that there is an $\varepsilon > 0$ such that $P(t^*,t^*+\varepsilon) \in \mathrm{GL_n}$. Further, by the evolution property we have

$$P(s,t^*+\varepsilon) = P(s,t^*)P(t^*,t^*+\varepsilon).$$

By definition of t^*, it holds that $P(s,t^*+\varepsilon) \notin \mathrm{GL_n}$, hence $P(s,t^*)$ can not be invertible as claimed.

(ii) The first statement, that is $P(s,u) \in \mathrm{GL}_n$ is obvious by the very definition of t^*. For the next two, note that we can apply Lemma 4.22(i) since it only requires P to be càdlàg . Hence we have

$$P(s,u) = P(s,u-)P(u-,u).$$

As $P(s,u) \in \mathrm{GL_n}$, so are $P(s,u-)$ and $P(u-,u)$ which follows from the fact that we are dealing with square matrices and

$$\det P(s,u) = \det P(s,u-)\, \det P(u-,u).$$

(iii) By *(ii)*, we have $P(s,u) \in \mathrm{GL_n}$, hence

$$P(u,t^*) = P(s,u)^{-1}P(s,t^*).$$

Since $P(s,t^*)$ singular, the first claim follows. For the second, note that

$$P(u-,u)P(u,t^*) = P(u-,t^*),$$

which implies the second statement. For the final statement, note that we have

$$P(t^*-,t^*) = \lim_{\varepsilon \downarrow 0} P(t^* - \varepsilon, t^*).$$

For each $\varepsilon > 0$ small enough, we have seen that $P(t^* - \varepsilon, t^*) \notin \mathrm{GL_n}$, hence by the closedness of the set of singular matrices in $\mathrm{M}_n(\mathbb{R})$, we conclude $P(t^*-,t^*) \notin \mathrm{GL_n}$. □

We will see that when dealing with existence of the extended generator, the set

$$\{t > 0 : P(t-,t) \notin \mathrm{GL_n}\}$$

will play an important role. We will not be able to show existence of the extended generator for any proper MES (P,γ,A). We will however be able to do so under an additional assumption that we will formulate as a condition. To motivate this condition, let us make the following observation.

Assume there exists an interval (s,t) such that for any $u \in (s,t)$ it holds $P(v,u) \notin \mathrm{GL_n}\, \forall v < u$. Assume further that P is càdlàg . Then, since $\mathrm{M}_n(\mathbb{R}) \setminus \mathrm{GL_n}$ is closed, we get $P(u-,u) \notin \mathrm{GL_n}$ for all $v \in (s,t)$. However, by the càdlàg assumption, there exists an $\varepsilon > 0$ such that $u + \varepsilon \in (s,t)$ and $P(u, u+\varepsilon) \in \mathrm{GL_n}$. But since $s < u < u + \varepsilon < t$, the initial assumption would require $P(v, v+\varepsilon) \notin \mathrm{GL_n}$, hence a contradiction. Based on the observation above where there was a continuum of $u \in (s,t)$ such that $P(u-,u) \notin \mathrm{GL_n}$, we formulate the following two conditions. Note that the first condition implies the second, see Remark 4.27.

Condition 4.26. *Consider a proper MES (P,γ,A).*

(i) The set

$$\{t > 0 : P(t-,t) \notin \mathrm{GL_n}\}$$

is at most countable.

(ii) For dA denoting the measure induced by A, we have

$$dA\left(\{t > 0 : P(t-,t) \notin \mathrm{GL_n}, \Delta A_t = 0\}\right) = 0.$$

Remark 4.27. *Note that in the condition above, (i) implies (ii). Indeed assume (i) holds. Then because A is càdlàg , we have $\Delta A_t = 0 \Rightarrow dA(\{t\}) = 0$. Hence, (i) implies that the set in (ii) can be written as the at most countable union of zero sets. Further down, we formulate Theorem 4.35 under (ii). However, we believe that condition (i) always holds for a càdlàg MES P. So far this is only a conjecture since we can not prove this at this point.*

In the sequel, let (P, γ, A) be a fixed proper MES. Let us briefly present a result regarding limits of the form $P(t-, t)$ as they appear in Condition 4.26 in the case where one assumes invertibility.

Proposition 4.28. *Assume $P(s,t) \in \mathrm{GL_n}$ for some $s < t$ and $\Delta A_t = 0$. Then $P(t-, t) = \mathrm{Id_n}$.*

Proof. By Proposition 4.25(ii) we have $P(t-, t) \in \mathrm{GL_n}$. Further, note that we have

$$P(s,t) = P(s,t-)P(t-,t) = P(s,t-) + \frac{d\gamma_s}{dA}(t)\Delta A_t.$$

Hence, $P(s,t) = P(s,t-)$. That implies

$$P(s,t)P(t-,t)P(t-,t)^{-1} = P(s,t)P(t-,t)P(t-,t).$$

Hence we have $P(t-,t) = P(t-,t)^{-1}$ showing the claim. ☐

Before presenting our most general existence result, let us first proof a special case in the form of Proposition 4.31. Note that the conditions we make in that theorem imply Condition 4.26 (i) and (ii). This proposition will illustrate the basic idea we follow to show existence. The general case is then treated in Theorem 4.35 below.

Let us make a few definitions in order to prepare Proposition 4.31. Given a proper MES (P, γ, A), set $t_0 = 0$ and define for $n \in \mathbb{N}$

$$t_n := \inf \{t > t_{n-1} : P(t_{n-1}, t) \notin \mathrm{GL_n}\}, \tag{4.17}$$

if $t_{n-1} < \infty$ and $t_n = \infty$ else. Note that if $t_{n-1}, t_n < \infty$, then by the càdlàg property $t_n - t_{n-1} > 0$.

The following can be seen as a corollary to Proposition 4.25. It shows that the definition of the sequence $(t_n)_n$ is in some sense universal.

Corollary 4.29. *Let $s \in [t_{n-1}, t_n)$. Then for*

$$t^* = \inf \{t > s : P(s,t) \notin \mathrm{GL_n}\},$$

we have $t^* = t_n$ and $P(s, t^*) \notin \mathrm{GL_n}$.

Proof. Since we have $P(t_{n-1}, s) \in \mathrm{GL_n}$, for any $u < t_n$ we have $P(s, u) \in \mathrm{GL}_n$. Hence $t^* = t_n$ since $P(s, t_n) \notin \mathrm{GL_n}$. □

Remark 4.30. *We have seen that $\forall s \in [t_{n-1}, t_n)$, it holds $P(s, t_n) \notin \mathrm{GL_n}$ and also $P(t_n-, t_n) \notin \mathrm{GL_n}$ which implies Condition 4.26 if $\lim_n t_n = \infty$. We can however not say anything regarding singularity about $P(s, t_n-)$. We can however make the following statement about the kernel. Consider $t_{n-1} \le s \le s' < t_n$. Then*

$$\ker P(s, t_n-) = \ker P(s', t_n-),$$

which follows from $P(s, s') \in \mathrm{GL_n}$.

Proposition 4.31. *Consider a proper MES (P, γ, A) and define the sequence $(t_n)_{n\in\mathbb{N}_0}$ as in (4.17). If $\sum_{n\in\mathbb{N}} \mathbb{1}_{(t_{n-1}<\infty)}(t_n - t_{n-1}) = \lim_{n\to\infty} t_n = \infty$, then an extended generator G exists for (P, γ, A) on all of $\mathbb{R}_{>0}$. Further, G is dA a.e. unique outside a countable set.*

Proof. Define $N \in \mathbb{N} \cup \{\infty\}$ by

$$N = \inf \{n \in \mathbb{N} : t_n = \infty\}.$$

Note that the condition $\lim_{n\to\infty} t_n = \infty$ implies

$$\mathbb{R}_{>0} = \dot{\bigcup_{n=1}^{N}} (t_{n-1}, t_n) \quad \dot\cup \quad \dot{\bigcup_{n=1}^{N-1}} \{t_n\}. \tag{4.18}$$

By construction, for any $n \le N$, we have $P(t_{n-1}, t) \in \mathrm{GL_n}$ for all $t \in (t_{n-1}, t_n)$ and hence by Proposition 4.23 $P(t_{n-1}, t-) \in \mathrm{GL_n}$. Hence we can define

$$G(\cdot)|_{(t_{n-1},t_n)} = P(t_{n-1}, \cdot-)^{-1} \frac{d\gamma_{t_{n-1}}}{dA}(\cdot).$$

We are left to define G on $\dot\bigcup_{n=1}^{N-1}\{t_n\}$ such that G satisfies the definition of the extended generator. Take an arbitrary t_n. For $t_{n-1} \le u < t_n$, G needs to satisfy

$$\begin{aligned} P(t_{n-1}, t_n) - P(t_{n-1}, u) &= \int_{(u,t_n]} P(t_{n-1}, s-)G(s)A(ds) \\ = \int_{(u,t_n)} P(t_{n-1}, s-)G(s)A(ds) &+ P(t_{n-1}, t_n-)G(t_n)\Delta A_{t_n}. \end{aligned} \tag{4.19}$$

If $\Delta A_{t_n} = 0$, we can just set $G(t_n) = 0$ and the extended generator property is satisfied. Now assume A charges t_n, i.e. $\Delta A_{t_n} > 0$. Using the evolution system property of P, we get that (4.19) is equivalent to

$$\begin{aligned} &P(t_{n-1}, t_n) - P(t_{n-1}, u) \\ &= P(t_{n-1}, t_n-) - P(t_{n-1}, u) + P(t_{n-1}, t_n-)G(t_n)\Delta A_{t_n}. \end{aligned} \tag{4.20}$$

canceling out $P(t_{n-1}, u)$ gives equivalently

$$P(t_{n-1}, t_n) = P(t_{n-1}, t_n-) + P(t_{n-1}, t_n-)G(t_n)\Delta A_{t_n}. \tag{4.21}$$

An application of the evolution property yields

$$P(t_{n-1}, t_n-)P(t_n-, t_n) = P(t_{n-1}, t_n-)(\mathrm{Id_n} + G(t_n)\Delta A_{t_n}).$$

Hence, it is sufficient for G to satisfy

$$P(t_n-, t_n) - \mathrm{Id_n} = G(t_n)\Delta A_{t_n}.$$

Since by assumption $\Delta A_{t_n} > 0$, (4.20) is satisfied if we set

$$G(t_n) = \frac{P(t_n-, t_n) - \mathrm{Id_n}}{\Delta A_{t_n}}, \tag{4.22}$$

which is always possible. Hence we have defined G on all of $\mathbb{R}_{>0}$ such that the defining conditions of the extended generator are satisfied. This shows existence. As We have $P(t_{n-1}, u-) \in \mathrm{GL_n}$ for $u < t_n$, we get that G is unique on the intervals (t_{n-1}, t_n) outside dA-zero sets. This is a consequence of Lemma 4.22 (iii), which for $s \leq u$ and

$$\tilde{G}(\cdot) := P(s, \cdot-)^{-1}\frac{d\gamma_s}{dA}(\cdot) \text{ and } \hat{G}(\cdot) := P(u, \cdot-)^{-1}\frac{d\gamma_u}{dA}(\cdot),$$

implies for $l > u$,

$$\tilde{G}(l) = P(s, l-)^{-1}\frac{d\gamma_s}{dA}(l) = (P(s,u)P(u,l-))^{-1}\, P(s,u)\frac{d\gamma_u}{dA}(l) = \hat{G}(l) \quad [dA],$$

since $d\gamma_s/dA$ is dA-a.e. unique. Hence, non uniqueness is only possible on $\dot\bigcup_{n=1}^{N-1}\{t_n\}$, which is a countable set as claimed. $\square$

Remark 4.32. *Let us briefly discuss the condition* $\lim_{n\to\infty} t_n = \infty$ *in Proposition 4.31. The following example shows that this assumption does not always hold. Consider Example 4.24 where* $\lim_{n\to\infty} t_n < \infty$ *holds. We modify this example by defining the set* M *as*

$$M = \dot\bigcup_{i\in\mathbb{N}} \tilde{M}_i \,\dot\cup\, \dot\bigcup_{i\in\mathbb{N}} \widehat{M_i},$$

where we are left to define the sets $\tilde{M}_i$ *and* $\widehat{M_i}$*. First, define the sequence* $(j_i)_{i\in\mathbb{N}_0}$ *by* $j_0 := 0$ *and*

$$j_i := j_{i-1} + \frac{1}{2^{j-1}}.$$

Hence, $\lim_{n\to\infty} j_n = \sum_{i\in\mathbb{N}} 2^{-(j-1)} = 2 < \infty$*. Next, for* $i \in \mathbb{N}$ *we define*

$$\tilde{M}_i = \left\{(s,t) \in \mathbb{R}^2 : j_{i-1} \leq s \leq t \leq j_i\right\}.$$

Further, we define for $i \in \mathbb{N}$,

$$\widehat{M_i} = \left\{(s,t) \in \mathbb{R}^2 : 2 + (i-1) \le s \le t \le 2 + i\right\}.$$

By construction, $\phi(s,t) = \mathbb{1}_M(s,t)$ is a càdlàg MES over $M_n(\mathbb{R})$ for $n = 1$. However, the sequence t_n in (4.17) satisfies $\lim_{n\to\infty} t_n = 2 < \infty$. On the other hand, ϕ can be embedded in a proper MES (P, γ, A) with

$$\phi(s,t) - \phi(s,u) = \int_{(u,t]} \rho(s,u) A(du),$$

where $\rho(s,u) = 2^{i-1}(\phi(s,t) - 1)$ for $0 \le s \le u < 2$ and $A(t) = j_i$ for $t \in [j_{i-1}, j_i)$. Further, on $[2,\infty)$, ρ is given by $\rho(s,u) = \phi(s,u) - 1$ and A is given by $A(t) = 2 + i$ for $t \in [2+(i-1), 2+i)$. Hence γ is given by $\frac{d\gamma_s}{dA}(\cdot) = \rho(s,\cdot)$. Indeed, while Proposition 4.31 can not be applied to this example, we will see that by Theorem 4.35 this proper MES too has an extended generator.

An interesting consequence in case of continuity is the following proposition showing under stronger continuity assumptions the extended generator always exists.

Proposition 4.33. *Consider a càdlàg MES P and assume $s \mapsto P(s,t)$ is even continuous. Then $P(s,t) \in \mathrm{GL_n}$ for all $(s,t) \in \mathbb{T}$.*

Proof. Since $\mathrm{GL_n} \subset \mathrm{M}_n(\mathbb{R})$ is a closed subset and $s \mapsto P(s,t)$ is continuous for arbitrary $t > 0$ on $[0,t]$, we have that there exists an $\varepsilon_t > 0$ such that $P(t - \varepsilon_t, t) \in \mathrm{GL_n}$ where we use $P(t,t) = \mathrm{Id_n} \in \mathrm{GL_n}$. Let $s \ge 0$ be arbitrary. Define

$$t^* := \inf\{t > s : P(s,t) \notin \mathrm{GL_n}\},$$

and assume $t^* < \infty$. By Proposition 4.25 we have $P(s,t^*) \notin \mathrm{GL_n}$. Let ε_{t^*} be like above. Then, we have

$$P(s,t^*) = P(s, t^* - \varepsilon_{t^*}) P(t^* - \varepsilon_{t^*}, t^*).$$

The right hand side is a product of two matrices, that by our assumptions are non-singular. However, that implies $P(s,t^*) \in \mathrm{GL_n}$, which is a contradiction to $t^* < \infty$. This yields $P(s,t) \in \mathrm{GL}_n$ as desired. □

Remark 4.34. *Let us apply Proposition 4.31 to Example 4.24. Recall that we have defined the MES ϕ by*

$$\phi(s,t) = \mathbb{1}_M(s,t), \quad 0 \le s \le t.$$

Let $n \in \mathbb{N}$ be such that $(n-1) \le s < n$. Then for $t \ge s$, we have $\phi(s,t) = \mathbb{1}_{\{t<n\}}$. We get that $t_n = n$, hence the condition is satisfied. Indeed, it is easy to see that $\gamma \ll dA$ for

$$A(t) := n - 1, \text{ for } n - 1 \le t < n.$$

By setting $G(t) = -\mathbb{1}_{\mathbb{N}}(t)$ it is easy to check that G is an A extended generator for ϕ and indeed satisfies (4.22).

Let us now treat the more general case where the set in Condition 4.26(i) can have accumulation points. We will see that this is just a modification of the special case presented above. Recall that in the first part of the condition implies the second so we proof the following theorem under the second condition.

Theorem 4.35. *Given a proper MES (P, γ, A), assume Condition 4.26(ii) holds. Then, there exists an extended generator G that is unique dA-a.e. outside a countable subset of $\mathbb{R}_+$.*

Proof. For $q \in \mathbb{Q}_+ = \mathbb{R}_+ \cap \mathbb{Q}$, let $\varepsilon_q = \inf\{\varepsilon > 0 : P(q, q+\varepsilon) \notin \mathrm{GL_n}\}$ which is bigger then zero by P being càdlàg . By Proposition 4.25 (ii), that implies $P(q, v-) \in \mathrm{GL_n}$ for all $v \in (q, q+\varepsilon_q)$. Define the intervals $I_q = (q, q+\varepsilon_q)$. We can therefor define $G(\cdot)|_{I_q} = P(q, \cdot)^{-1}\frac{d\gamma_q}{dA}(\cdot)$. Hence, G is defined on the set

$$I = \bigcup_{q \in \mathbb{Q}_+} I_q,$$

and by the same arguments as at the end of the proof of Proposition 4.31, G is well defined since for two $q, p \in \mathbb{Q}_+$ with $I_q \cap I_p \neq \emptyset$, the two definitions of G are equal up to a dA-zero set. Since $P(q, \cdot-) \in \mathrm{GL}_n$ on I_q, G is further unique up to dA-zero set due to the dA uniqueness of the Radon-Nikodým derivative.

We are left to deal with $I^c = \mathbb{R}_+ \setminus I$. Let $s \in I^c$. Let $q_l \uparrow s$ and $q'_l \downarrow s$ be two strictly monotonic sequences in $\mathbb{Q}_+$. Then, $P(q_l, q'_j) \notin \mathrm{GL}_n$ for all $l, j \in \mathbb{N}$ because otherwise $s \in \mathcal{I}_{q_l} \subset I$. This holds for all l, j and since GL_n is a closed subspace of $\mathrm{M}_n(\mathbb{R})$, we get that $P(s-, q'_j) \notin \mathrm{GL_n}$ for all $j \geq 1$. Again by the same argument, we can take a limit once again to get $P(s-, s) \notin \mathrm{GL_n}$. Note that we have

$$\begin{aligned}&\{t > 0 : P(t-, t) \notin \mathrm{GL_n}\} \\ &\qquad = \{t > 0 : P(t-, t) \notin \mathrm{GL_n}, \Delta A_t = 0\} \uplus \{t > 0 : P(t-, t) \notin \mathrm{GL_n}, \Delta A_t > 0\}.\end{aligned}$$

By assumption, the set where $\Delta A_t = 0$ has measure zero wrt. dA. Therefore we can define G on that set arbitrarily, hence we just set G to be the zero matrix on that set. Since the set on the right is at most countable as A is càdlàg and hence has at most countably many jumps, we have already shown the uniqueness claim. We are left to argue that G can be defined on the set on the right. But we have already shown in the proof of Proposition 4.31, that G can be defined on that set by

$$G(t) = \frac{P(t-, t) - \mathrm{Id_n}}{\Delta A_t}, \quad \forall\, t \in \{t > 0 : P(t-, t) \notin \mathrm{GL_n}, \Delta A_t > 0\}, \tag{4.23}$$

finishing the proof. □

Remark 4.36. *We will later see that Condition 4.26 (ii) plays an important role when studying MESs with certain invariance properties motivated by the study of polynomial semimartingales.*

4.3 Ordinary measure integral equations

In this section we want to discuss the link between evolution systems previously introduced, their extended generators and a certain class of linear measure integral equations or as they are commonly referred to, measure differential equations. We will use both terminologies interchangeably but since the integral formulation is equivalent to the weak Radon-Nikodým derivative formulation, this should not lead to any confusions.

For the rest of this section, we denote with $\|\cdot\|_{\mathbb{R}^n}$ any fixed norm on $\mathbb{R}^n$. Whenever an operator norm occurs, it is always the one corresponding to the fixed norm $\|\cdot\|_{\mathbb{R}^n}$. When there is no confusion, we will also just write $\|\cdot\|$ if it is clear that this is a norm on $\mathbb{R}^n$.

4.3.1 Existing literature

The classical case, that is where the derivatives are classical derivatives, are among the most common topics in mathematics. In particular, the existence and uniqueness result by Picard-Lindelöf are treated on an introductory level when studying ordinary differential equations, see e.g. [Arn92] or [Wal13] and the references therein.

The continuous version of our problem is given in the context of systems of linear differential equations of the form

$$\frac{dy(t)}{dt} = V(t, y(t)), \quad y(s) = x \in \mathbb{R}^n,$$

for a vector field $V : [0, \infty) \times \mathbb{R}^n \mapsto \mathbb{R}^n$ satisfying Lipschitz and continuity assumptions guaranteeing existence and uniqueness of solutions. It is well understood that when V is linear in y, that a fundamental solution $M(s,t) \in \mathrm{M}_n(\mathbb{R})$ for $(s,t) \in \mathbb{T}$ to these equations exist such that for the flow of the ode above denoted by $\phi(s,x;t)$, one has $\phi(s,x;t) = M(s,t)x$ corresponding to the linearity of the equation. Indeed, by studying the so called Wronski-determinant of these solutions, one can not only establish existence and uniqueness results of the equation but show that solutions to different initial values do not intersect. Since by assumption, everything is sufficiently regular, one can make a time reversal transformation to derive an ode with initial condition $F(t)$, such that the solution "after" time $t-s$ is at x. This shows that $M(s,t)$ is invertible for any $s \leq t$.

In our setup, the assumptions made above do not hold anymore. Further, we do not assume absolute continuity wrt the Lebesgue measure. The problem we shall deal with is of the following form. Let A be a deterministic and increasing càdlàg function, i.e. $A \in \mathcal{V}^+_{\mathfrak{X}}$. Given a function $V : (s,\infty) \times \mathbb{R}^n \to \mathbb{R}^n$, $s \geq 0$, we consider measure integral or measure differential equations (MDE) of the form

$$\begin{cases} F(t) = F(s) + \displaystyle\int_{(s,t]} V(u, F(u-))dA_u \\ F(s) = x, \end{cases} \tag{MDE:A,V}$$

with initial value $x \in \mathbb{R}^n$ at time s. Note that if the above equation permits existence and uniqueness of solutions for any $x \in \mathbb{R}^n$, this defines the associated flow $\phi(s, x; t) = F(t)$ where F_s is the unique solution to (MDE:A, V) with initial data $F_s(s) = x$. It will sometimes be convenient to make the dependency of the solution F with respect to s and x explicit which we shall do with $F_{s,x}(t)$. When showing existence and uniqueness of a global solution, it is enough to restrict to a finite time horizon T provided that T can be chosen arbitrary large. Hence we shall fix $T > s$ and consider the time interval $[s, T]$ instead of $\mathbb{R}_+$. We are interested in the case where V is Lipschitz which we define later. Note that we need V to be defined on $(s, T]$ only since for $t \leq s$, the integral above is over the empty set. Since we do not require V to be continuous in the first variable, the Lipschitz property in the second variable, namely that there exists a $L < \infty$ such that

$$\|V(t, x) - V(t, y)\| \leq L \|x - y\|, \qquad \text{for all } t \in (s, T], \tag{4.24}$$

will not be enough for us to establish existence and uniqueness. Hence we will need an additional integrability assumption in the first variable. We will make this precise later, see Definition 4.43.

Let us briefly discuss the pertinent literature. In [Sha72] the authors consider measure differential equations of type (MDE:A, V), however they only show existence and uniqueness on some interval, hence they only provide a result for a local solution. The same is true for [DS72].

The setting in [Gil07] is slightly different from ours as it would require A to be left continuous. Further, [Gil07, Theorem 19.7.1] poses continuity assumptions on the first variable of V, which is not satisfied in our setting. Hence we will proof a suitable existence and uniqueness result for (MDE:A, V).

Another approach would be too consider (MDE:A, V) as a stochastic differential equation. This is possible since A is a càdlàg process of locally finite variation and hence a Semimartingale under the measure P where P is the Dirac measure on $\mathcal{D}(\mathbb{R}_+, \mathbb{R})$ charging A. Uniqueness would then be a consequence of Protter [Pro05, Theorem 7 in Chapter V]. For that theorem however, it is necessary to make slightly stronger regularity assumptions on V then we would like.

We believe it is preferable to provide a result tailored to our setting, allowing for a more direct treatment of the issues at hand. In particular we believe this to be an advantage for the reader, hence we will do so in the following section.

We will also discuss Gronwall type inequalities for measure differential equations as we will need them in our discussions regarding regularity of solutions wrt. initial data and in the next chapter when studying polynomial processes. These kind of inequalities have been studied among others in [AP97], [DS80], [EK90], [Hor96] and [Gil07]. However, there the assumptions made (such as only isolated points of discontinuity for A) are not given in our case. Nonetheless, a version of a Gronwall inequality almost suitable for us has been essentially given in Ethier and Kurtz [EK09] by means of Theorem 5.1 in the appendix and we will present a slight modification of that proof that covers our situation.

4.3.2 Existence and uniqueness for measure integral equations

We will now state our result on existence and uniqueness of (MDE:A, V), which is a Picard-Lindelöf type theorem. The proof follows the idea of the classical version meaning we want to apply Banach's fixed point theorem to a contraction mapping Φ where a fixed point of Φ is a solution of (MDE:A, V). Since in general the Lipschitz constant L of V is not smaller then one, it is not possible to just use the norm of uniform convergence on the space of all càdlàg functions. In fact, one needs to define an equivalent norm such that Φ becomes a contraction. To the best of our knowledge, such a theorem has not been proven in the literature.

Let us note that assuming V is bounded, the right-continuity of A implies that any function that satisfies (MDE:A, V) is right continuous. Further, in order for the expression $F(u-)$ to make even sense, one needs the existence of left limits for F. Hence it is reasonable to assume that any solution is a càdlàg function.

We begin by making a few short remarks regarding our solution space. As a first step, let us make precise what we mean by a solution to (MDE:A, V).

Definition 4.37. *We call $F \in \mathcal{D}([s,T], \mathbb{R}^n)$ a solution of* (MDE:A, V)*, if F satisfies* (MDE:A, V) *for all $t \in [s,T]$. We say the solution is unique on $[s,T]$, if F is the only element in $\mathcal{D}([s,T], \mathbb{R}^n)$ wrt the norm*

$$\|\kappa\|_\infty := \sup_{u \in [s,T]} \|\kappa(u)\|_{\mathbb{R}^n} ,$$

that satisfies (MDE:A, V).

Let us recall some basic facts about the space of càdlàg functions. First, note that it is a Banach space with respect to the topology of uniform convergence induced by the norm $\|\cdot\|_\infty$, see e.g. [JS13, Chapter VI]. Unlike the space of continuous function however, it fails to be separable wrt this norm. To see that, just consider the family of càdlàg function κ_s defined on $(0,T]$ via $\kappa_s(t) := \mathbb{1}_{(0,s)}(t)$ indexed by s for each $s \in (0,T)$. Then, for $s' \neq s$, we have $\|\kappa_s - \kappa_{s'}\|_\infty = 1$.

Non separability can often proof to be a problem in probability theory which is why in many cases, the topology of uniform convergence is replaced with the Skorochod topology, again see e.g. [JS13, Chapter VI] for more details and results. In our situation however, we do not need separability, completeness is sufficient for Banach's fixed point theorem.

As mentioned before, a key tool in proving existence and uniqueness is the definition of a suitable norm such that the Picard mapping is a contraction (we will define shortly what we mean by that). In the classical setting, for $\kappa \in \mathcal{C}([s,T], \mathbb{R}^n)$ one defines the norm

$$\|\kappa\|_{\rho,(s,T]} := \sup_{u \in (s,T]} \rho(u) \|\kappa(u)\|_{\mathbb{R}^n} \tag{4.25}$$

for $\rho(u) = \exp\left(-\tilde{L}(u-s)\right)$, where $\tilde{L} > L$, with L being the Lipschitz constant of V. Usually, one finds $\tilde{L} = 2L$ in the literature. Notice that ρ^{-1} solves

$$\rho^{-1}(t) = \rho^{-1}(s) + \tilde{L} \int_s^t \rho^{-1}(u)du, \tag{4.26}$$

where the integral is a Riemann integral. In fact, ρ^{-1} is the only solution to (4.26) and the exponential function could even be defined as the unique solution to (4.26) for $\tilde{L} = 1$. The general case corresponding to our setting now coincides precisely with a widely studied object in semimartingale theory, a generalization of the exponential function (where the Riemann integral is replaced with a stochastic integral) referred to as stochastic exponential. We only need the version where the semimartingale is a process of locally finite version (actually increasing is enough) reducing the stochastic integral to the Lebesgue-Stieltjes integral. Hence we can solve pathwise including a deterministic setting and the term ***stochastic*** should not lead to confusions.

Definition 4.38 (Stochastic exponential)**.** *Consider $A \in \mathcal{V}^+_{\mathscr{A}}$. We call a càdlàg function $\kappa \in \mathcal{D}([s,T],\mathbb{R})$ the stochastic exponential of A, if*

$$\kappa(t) = 1 + \int_{(s,t]} \kappa(u-)A(du), \tag{4.27}$$

for all $t \in [s,T]$.

The next result on stochastic exponentials can be found in [JS13, Theorem 4.61 in I.4], together with the general version for semimartingales.

Theorem 4.39. *Consider $A \in \mathcal{V}^+_{\mathscr{A}}$. Then there is a unique càdlàg solution to 4.27 denoted by $\mathcal{E}(A)$ with $\mathcal{E}(A) \in \mathcal{V}^+_{\mathscr{A}}$. Further, it holds*

$$\mathcal{E}(A)_t = \exp\left(A_t - A_s - \sum_{s<u\le t} \Delta A_u\right) \prod_{s<u\le t} (1 + \Delta A_u) \ge 1. \tag{4.28}$$

We can now state the following corollary that will be central for our main result in this section.

Corollary 4.40. *Let $\tilde{L}, \rho^{-1}(s) > 0$. There exists exactly one càdlàg solution ρ^{-1} to the equation*

$$\rho^{-1}(t) = \rho^{-1}(s) + \tilde{L} \int_{(s,t]} \rho^{-1}(u-)A(du), \tag{4.29}$$

where $\rho^{-1}(t) \ge \rho^{-1}(s)$ for all $t \in [s,T]$ and we have for each such t,

$$\frac{\int_{(s,t]} \rho^{-1}(u-)A(du)}{\rho^{-1}(t)} \le \frac{1}{\tilde{L}}.$$

Further, ρ^{-1} is non decreasing which implies $\rho^{-1}(t-) \le \rho^{-1}(t)$.

Proof. The first statement follows by just considering $\tilde{\rho}^{-1} = \rho^{-1}/\rho^{-1}(s)$ as well as the transformation $\tilde{A} = \tilde{L}A$ to which we apply Theorem 4.39. The second follows by basic manipulations of (4.29). Monotonicity is a direct consequence of A being non decreasing. $\square$

Remark 4.41. *As the notation ρ^{-1} suggests, we are interested in the reciprocal which we denote by ρ. Note that by the above corollary, $\rho^{-1}(t) \geq \rho^{-1}(s) > 0$ for all $t \in [s,T]$, which implies ρ is càdlàg on $[s,T]$. Hence for any $T > s$ fixed, we have*

$$0 < \rho(T) = \rho_- = \inf_{u\in[s,T]} \rho(u) \leq \sup_{u\in[s,T]} \rho(u) = \rho_+ = \rho(s) < \infty.$$

Now, if we define the norm $\|\cdot\|_{\rho,[s,T]}$ on $\mathcal{D}((s,T],\mathbb{R}^n)$ by (4.25) where ρ^{-1} is the solution to (4.29) for given $\rho^{-1}(s), \tilde{L} > 0$, we can conclude that

$$\rho_- \|\cdot\|_{\infty,[s,T]} \leq \|\cdot\|_{\rho,[s,T]} \leq \rho_+ \|\cdot\|_{\infty,[s,T]},$$

for $\|\cdot\|_{\infty,[s,t]}$ being the norm of uniform convergence on $[s,T]$ which then implies that $\|\cdot\|_{\infty,[s,T]}$ and $\|\cdot\|_{\rho,[s,T]}$ are two equivalent norms on $\mathcal{D}([s,T],\mathbb{R}^n)$.

Remark 4.42. *Recall that for a Banach space $\mathcal{X}$ with two equivalent norms $\|\cdot\|_1, \|\cdot\|_2$ and a mapping $T: \mathcal{X} \to \mathcal{X}$ such that T is a contraction wrt $\|\cdot\|_1$ with unique fixed point $x^* \in \mathcal{X}$, i.e. for any $x, y \in \mathcal{X}$ one has $\|T(x) - T(y)\|_1 < \|x - y\|_1$ and $T(x^*) = x^*$, the statement of Banach's fixed point theorem holds true for $\|\cdot\|_2$ too, i.e for any $x \in \mathcal{X}$ we have*

$$\lim_{l\to\infty} T^l(x) \overset{\|\cdot\|_2}{\to} x^*.$$

Let us briefly make precise what we meant in (4.24) by Lipschitz and integrability in the first variable. Recall that for $A \in \mathcal{V}^+_{\mathfrak{A}}$ we denote with dA the measure induced by A.

Definition 4.43. *We call a measurable function $V : (s,T] \times \mathbb{R}^n \to \mathbb{R}^n$ A-Lipschitz on $(s,T]$ for an $A \in \mathcal{V}^+_{\mathfrak{A}}$, if there exists a $L > 0$ and a set $N \subset (s,T]$ with $dA(N) = 0$ such that for any $t \in N^c$ we have*

$$\|V(t,x) - V(t,y)\|_{\mathbb{R}^n} \leq L \|x - y\|_{\mathbb{R}^n}.$$

If additionally, there exists a $x^* \in \mathbb{R}^n$ such that the function $t \mapsto \|V(t,x^*)\|$ is A-essentially bounded on $(s,T]$, i.e. it holds

$$\int_{(s,T]} \|V(u,x^*)\|_{\mathbb{R}^n} A(du) \leq \operatorname*{ess\,sup}_{u\in(s,T]} \|V(u,x^*)\|_{\mathbb{R}^n} (A(T) - A(s)) < \infty,$$

we call V A-Lipschitz and integrable on $(s,T]$. We call a measurable function $V : \mathbb{R}_{>0} \times \mathbb{R}^n \to \mathbb{R}^n$ locally A-Lipschitz and integrable, if for all $T > 0$, there exists a constant $L_T > 0$ such that V is locally A-Lipschitz and integrable on $(0,T]$. If there is no ambiguity wrt. the function A, we call V (locally) Lipschitz and integrable.

Remark 4.44. *(i) The condition that a point x^* exists such that $V(t, x^*)$ is integrable is always satisfied for such V where $V(t,\cdot)$ is a linear map since then one can simply choose $x^* = 0$.*

(ii) Note that if V is locally A-Lipschitz and integrable, then it is A-Lipschitz and integrable on all intervals $(s, T]$.

The following definition will play a central role when showing existence and uniqueness of (MDE:A, V) and is widely referred to as Picard map since it defines the Picard iteration by means of composition.

Definition 4.45. *Let A, V be as in* (MDE:A, V) *and $x \in \mathbb{R}^n$. Define the map $T_{A,V,x} : \mathcal{D}([s,T], \mathbb{R}^n) \mapsto \mathcal{D}([s,T], \mathbb{R}^n)$ by*

$$T_{A,V,x}(\kappa)_t = x + \int_{(s,t]} V(u, \kappa(u-))A(du), \quad t \in (s, T].$$

We call $T_{A,V,x}$ the **Picard map** *of* (MDE:A, V) *(or just Picard map if there is no ambiguity).*

Remark 4.46. *Note that this map is not a priori well defined since the integral does not need to exist for arbitrary V while the Picard map must map càdlàg functions again to càdlàg functions. We give a sufficient condition with the next lemma.*

Lemma 4.47. *If V is Lipschitz and integrable with constant $L > 0$, then the Picard map is well defined.*

Proof. We calculate for arbitrary $t \in (s, T]$ and $\kappa \in \mathcal{D}([s,T], \mathbb{R}^n)$,

$$\begin{aligned}\int_{(s,t]} \|V(u, \kappa(u-))\|_{\mathbb{R}^n} A(du) &= \int_{(s,t]} \|V(u, \kappa(u-)) - V(u, x^*) + V(u, x^*)\|_{\mathbb{R}^n} A(du) \\ &\le \int_{(s,t]} \|V(u, \kappa(u-)) - V(u, x^*)\|_{\mathbb{R}^n} A(du) + \int_{(s,t]} \|V(u, x^*)\|_{\mathbb{R}^n} A(du) \\ &\le L \int_{(s,t]} \|\kappa(u-) - x^*\|_{\mathbb{R}^n} A(du) + \int_{(s,t]} \|V(u, x^*)\|_{\mathbb{R}^n} A(du).\end{aligned}$$

Since $\kappa(u) - x^*$ is càdlàg , it is bounded on $[s, T]$. Hence, $\kappa(u-) - x^*$ is bounded on $(s, T]$ which shows the existence of the first integral as $dA((s,T]) < \infty$. The second is finite by the Lipschitz assumption on V. We are left to show that $T(\kappa)$ is indeed càdlàg . Regarding right continuity, we need to show

$$\lim_{\varepsilon \downarrow 0} \int_{(t,t+\varepsilon]} \|V(u, \kappa(u-))\|_{\mathbb{R}^n} = 0.$$

But this already follows from the previous calculations using the fact that A is càdlàg and hence $A(t+\varepsilon) - A(t) \to 0$ as $\varepsilon \downarrow 0$. The existence of left limits follows from

$$\lim_{\varepsilon \downarrow 0} \int_{(s,t-\varepsilon]} V(u, \kappa(u-))A(du) = \int_{(s,t)} V(u, \kappa(u-))A(du),$$

making use of $(s,t) \subset (s,t]$ for the existence of the rhs. $\square$

We can now state our main existence and uniqueness result for measure integral equations.

Theorem 4.48 (Picard-Lindelöf)**.** *Let $V : (s,T] \times \mathbb{R}^n \to \mathbb{R}^n$ be A-Lipschitz and integrable on $(s,T]$ for an $A \in \mathcal{V}^+_{\mathfrak{A}}$. Then, there exists a unique solution to* (MDE:A,V) *for any initial value $x \in \mathbb{R}^n$.*

Proof. Let $T_{A,V,x}$ be the Picard map of (MDE:A,V). By Lemma 4.47, the map is well defined. Hence, a function $\kappa \in \mathcal{D}((s,T],\mathbb{R}^n)$ is a solution if and only if $T(\kappa) = \kappa$, or in another words a fixed point of $T_{A,V,x}$. If we show that there exists a unique fixed point, we are done. By Banach's fixed point theorem it is sufficient to show that $T_{A,V,x}$ is a contraction under a norm that is equivalent to the norm of uniform convergence. We shall pick the norm $\|\cdot\|_{\rho,[s,T]}$ where ρ is defined via (4.29) for some $\rho(s) > 0$ with $\tilde{L} = L(1+\varepsilon)$ for $\varepsilon > 0$ and L being the Lipschitz constant of V. We compute for two arbitrary $\kappa, \nu \in \mathcal{D}((s,T],\mathbb{R}^n)$,

$$
\begin{aligned}
&\|T_{A,V,x}(\kappa) - T_{A,V,x}(\nu)\|_{\rho,[s,T]} = \sup_{u\in[s,T]} \rho(u) \left\|(T_{A,V,x}(\kappa) - T_{A,V,x}(\nu))(u)\right\|_{\mathbb{R}^n} \\
&= \sup_{u\in[s,T]} \rho(u) \left\| \int_{(s,u]} V(\xi,\kappa(\xi-)) - V(\xi,\nu(\xi-)))A(d\xi) \right\|_{\mathbb{R}^n} \\
&\leq \sup_{u\in[s,T]} \rho(u) \int_{(s,u]} \|V(\xi,\kappa(\xi-)) - V(\xi,\nu(\xi-)))\|_{\mathbb{R}^n} A(d\xi) \\
&\leq L \sup_{u\in[s,T]} \rho(u) \int_{(s,u]} \rho(\xi)\rho^{-1}(\xi) \|\kappa(\xi-) - \nu(\xi-)\|_{\mathbb{R}^n} A(d\xi) \qquad (4.30) \\
&\leq L \|\kappa - \nu\|_{\rho[s,T]} \sup_{u\in[s,T]} \rho(u) \int_{(s,u]} \rho^{-1}(\xi) A(d\xi) \\
&\leq L \|\kappa - \nu\|_{\rho[s,T]} \sup_{u\in[s,T]} \frac{\int_{(s,u]} \rho^{-1}(\xi-)A(d\xi)}{\rho(u)} \\
&\leq \frac{1}{1+\varepsilon} \|\kappa - \nu\|_{\rho[s,T]},
\end{aligned}
$$

where we used Corollary 4.40. Hence, by equivalence of the two norms, we have shown existence and uniqueness to (MDE:A,V). □

We can also state the next obvious corollary to the previous theorem.

Corollary 4.49. *Let V be locally A-Lipschitz and integrable. Then there exists a unique solution of* (MDE:A,V) *on the interval $(s,T]$ for all $s < T$ with initial value $x \in \mathbb{R}^n$.*

From here on we always consider V to be defined on $\mathbb{R}_+ \times \mathbb{R}^n$. The next theorem shows that the solution flows of (MDE:A,V) satisfy certain càdlàg properties. Recall that the flow $\phi : \mathbb{T}\times\mathbb{R}^n \to \mathbb{R}^n$ of (MDE:A,V) is defined in case (MDE:A,V) has a unique solution for all $x \in \mathbb{R}^n$ and $s \in \mathbb{R}_+$ on all intervals $[s,T]$ the following way. For $(s,t) \in \mathbb{T}$ and $x \in \mathbb{R}^n$, set

$$\phi(s,x;t) := F_{s,x}(t),$$

where $F_{s,x}$ is the unique solution with initial value x at time s. In particular, it is well known that these flows satisfy the evolution property we have already seen when we dealt with evolution systems.

Theorem 4.50. *Consider* (MDE:A,V) *with locally Lipschitz and integrable* V *and denote by* ϕ *the corresponding flow. Then for any* $x \in \mathbb{R}^n$ *and* $0 \le s \le t$*, the functions*

$$s \mapsto \phi(s,x;t), \quad t \mapsto \phi(s,x;t)$$

are càdlàg on $[0,t]$ *(resp.* $[s,\infty)$*).*

Proof. We have already shown that solutions to (MDE:A,V) are càdlàg , which yields $t \mapsto \phi(s,x;t)$ is càdlàg on $[s,T]$ for all $T > s$, hence on $[s,\infty)$. The more challenging part is the càdlàg property in s. Let us first show right continuity, so we can assume $s < t$. We will apply the Gronwall inequality, see Theorem 4.55, to the function $F_{s+\varepsilon,x}(t)$ where $\varepsilon \ge 0$ such that $s+\varepsilon < t$. Let $K_* := \operatorname{ess\,sup}_{t\in(0,T]} \|V(t,x^*)\|_{\mathbb{R}^n}$. By our existence and uniqueness result we have

$$\begin{aligned}
\|F_{s+\varepsilon,x}(t)\|_{\mathbb{R}^n} &\le \|x\|_{\mathbb{R}^n} + \int_{(s+\varepsilon,t]} \|V(\xi, F_{s+\varepsilon,x}(\xi-))\|_{\mathbb{R}^n} \, A(d\xi) \\
&= \|x\|_{\mathbb{R}^n} + \int_{(s+\varepsilon,t]} \|V(\xi, F_{s+\varepsilon,x}(\xi-)) - V(\xi,x^*) + V(\xi,x^*)\|_{\mathbb{R}^n} \, A(d\xi) \\
&\le \|x\|_{\mathbb{R}^n} + L\int_{(s+\varepsilon,t]} \|F_{s+\varepsilon,x}(\xi-) - x^*\|_{\mathbb{R}^n} \, A(d\xi) + K_* A(T) \\
&\le \Big[\, \|x\|_{\mathbb{R}^n} + A(T)(K_* + \|x^*\|_{\mathbb{R}^n}) \Big] + L\int_{(s+\varepsilon,t]} \|F_{s+\varepsilon,x}(\xi-)\|_{\mathbb{R}^n} \, A(d\xi).
\end{aligned}$$

Define $K' = \|x\|_{\mathbb{R}^n} + A(T)(K_* + \|x^*\|_{\mathbb{R}^n})$. Then, by the Gronwall inequality we have

$$\begin{aligned}
\|F_{s+\varepsilon,x}(t)\|_{\mathbb{R}^n} &\le K' \exp\{L(A(T) - A(s+\varepsilon))\} \\
&\le K' \exp\{LA(T)\} =: K_F.
\end{aligned} \tag{4.31}$$

Hence, we get a uniform bound K_F wrt ε and t. We compute

$$\begin{aligned}
\|F_{s+\varepsilon,x}(t) - F_{s,x}(t)\|_{\mathbb{R}^n} \le &\int_{(s,s+\varepsilon]} \|V(\xi, F_{s,x}(\xi-))\|_{\mathbb{R}^n} \, A(d\xi) \\
+ \int_{(s+\varepsilon,t]} \|V(\xi, F_{s+\varepsilon,x}(\xi-)) - &V(\xi, F_{s,x}(\xi-))\|_{\mathbb{R}^n} \, A(d\xi)
\end{aligned} \tag{4.32}$$

We first show that the first integral on the rhs tends to zero as $\varepsilon \downarrow 0$. We have

$$\begin{aligned}
&\int_{(s,s+\varepsilon]} \|V(\xi, F_{s,x}(\xi-))\|_{\mathbb{R}^n} A(d\xi) \\
&= \int_{(s,s+\varepsilon]} \|V(\xi, F_{s,x}(\xi-)) - V(\xi, x^*) + V(\xi, x^*)\|_{\mathbb{R}^n} A(d\xi) \\
&\leq \int_{(s,s+\varepsilon]} \|V(\xi, F_{s,x}(\xi-)) - V(\xi, x^*)\|_{\mathbb{R}^n} A(d\xi) + \int_{(s,s+\varepsilon]} \|V(\xi, x^*)\|_{\mathbb{R}^n} A(d\xi) \\
&\leq L \int_{(s,s+\varepsilon]} \|F_{s,x}(\xi-) - x^*\|_{\mathbb{R}^n} A(d\xi) + K_*(A(s+\varepsilon) - A(s)) \\
&\leq (L(K_F + \|x^*\|_{\mathbb{R}^n}) + K_*)(A(s+\varepsilon) - A(s)) =: \rho(\varepsilon).
\end{aligned}$$

Since A is càdlàg , this tends to zero as $\varepsilon \downarrow 0$. Now consider again (4.32). We get

$$\|F_{s+\varepsilon,x}(t) - F_{s,x}(t)\|_{\mathbb{R}^n} \leq \rho(\varepsilon) + L \int_{(s+\varepsilon,t]} \|F_{s+\varepsilon,x}(\xi-) - F_{s,x}(\xi-)\|_{\mathbb{R}^n} A(d\xi) \tag{4.33}$$

We can therefore apply the Gronwall inequality again and get

$$\|F_{s+\varepsilon,x}(t) - F_{s,x}(t)\|_{\mathbb{R}^n} \leq \rho(\varepsilon) \exp\{LA(T)\}.$$

As we have already shown that $\rho(\varepsilon)$ tends to zero as $\varepsilon \downarrow 0$, we have shown right continuity.

We will now show that left limits exist, so assume $0 < s \leq t$. We shall do so again by applying the Gronwall inequality in an appropriate way. For the existence of left limits, we need to show that the limit

$$\lim_{\varepsilon \downarrow 0} F_{s-\varepsilon,x}(t)$$

exists for all x and $t \in [s,T]$. It suffices to show that for arbitrary positive zero sequence ε_n with $s_n = s - \varepsilon_n$ we have that $F_{s_n,x}(t)$ forms a Cauchy sequence. Let $m, n \in \mathbb{N}$ be large enough such that $s_n, s_m > 0$. Further, let $\tilde{s}_n = s_n \wedge s_m$ and $\tilde{s}_m = s_n \vee s_m$. We compute

$$\begin{aligned}
&\|F_{s_n,x}(t) - F_{s_m,x}(t)\|_{\mathbb{R}^n} \leq \int_{(\tilde{s}_n,\tilde{s}_m]} \|V(\xi, F_{\tilde{s}_n,x}(\xi-)) - V(\xi, x^*) + V(\xi, x^*)\|_{\mathbb{R}^n} A(d\xi) \\
&+ \int_{(\tilde{s}_m,t]} \|V(\xi, F_{s_n,x}(\xi-)) - V(\xi, F_{s_m,x}(\xi-))\|_{\mathbb{R}^n} A(d\xi) \\
&\leq L \int_{(\tilde{s}_n,\tilde{s}_m]} \|F_{\tilde{s}_n,x}(\xi-) - x^*\|_{\mathbb{R}^n} A(d\xi) + K_* \left(A(\tilde{s}_m) - A(\tilde{s}_n)\right) \\
&+ L \int_{(\tilde{s}_m,t]} \|F_{s_n,x}(\xi-) - F_{s_m,x}(\xi-)\|_{\mathbb{R}^n} A(d\xi).
\end{aligned} \tag{4.34}$$

Note that we have

$$\begin{aligned}
&L \int_{(\tilde{s}_n,\tilde{s}_m]} \|F_{\tilde{s}_n,x}(\xi-) - x^*\|_{\mathbb{R}^n} A(d\xi) + K_* \left(A(\tilde{s}_m) - A(\tilde{s}_n)\right) \\
&\leq \left(LK_F + L\|x^*\|_{\mathbb{R}^n} + K_*\right) \left(A(\tilde{s}_m) - A(\tilde{s}_n)\right) =: K(n,m),
\end{aligned}$$

where we use (4.31) to get

$$\max_{\xi\in(\tilde{s}_n,\tilde{s}_m]} \|F_{\tilde{s}_n,x}(\xi-)\|_{\mathbb{R}^n} \leq \max_{\xi\in[\tilde{s}_n,\tilde{s}_m]} \|F_{\tilde{s}_n,x}(\xi)\|_{\mathbb{R}^n} \leq K_F.$$

Since we have assumed $t \geq s > \tilde{s_m}$, again the Gronwall inequality yields

$$\|F_{s_n,x}(t) - F_{s_m,x}(t)\|_{\mathbb{R}^n} \leq K(n,m)\exp\{LA(T)\}.$$

Hence, since A is càdlàg , we get that $\sup_{h\in\mathbb{N}} K(n,n+h)$ goes to zero as n goes to infinity, proving that indeed we have a Cauchy sequence. By completeness, the limit exists, finishing the proof. □

Linear measure differential equations and MESs

We will now show how our previous discussions relate to MESs. For that, we consider the case where V is linear, that is the exists a measurable function $G : \mathbb{R}_{>0} \to \mathrm{M}_n(\mathbb{R})$, such that $V(t,x) = G(t)^\top x$. Fix a function $A \in \mathcal{V}^+_{\mathfrak{A}}$ and denote the induced measure on $\mathbb{R}_{>0}$ with dA. Then,

$$\|G\|_{dA,t} := \operatorname*{ess\,sup}_{dA}(\|G(\cdot)\|_\infty \mathbb{1}_{\cdot\leq t}),$$

where the norm is the operator norm wrt some fixed norm on $\|\cdot\|_{\mathbb{R}^n}$ on $\mathbb{R}^n$. By finite dimensionality, $\|\cdot\|_{dA,t}$ differs for a different choice of $\|\cdot\|_{\mathbb{R}^n}$ only by a multiplicative constant since all norms in finite dimensions are equivalent. The next Lemma just states the obvious.

Lemma 4.51. *Assume $\|G\|_{dA,t} < \infty$ for all $t \geq 0$. Then, V is locally A-Lipschitz and integrable.*

Proof. Let us note that for the operator norm $\|\cdot\|_\infty$ we have $\left\|G^\top\right\|_\infty \leq C\|G\|_\infty$ for all matrices $G \in \mathrm{M}_n(\mathbb{R})$ where the constant $C > 0$ is independent of G. Just consider the matrix norm of the maximum of all entries of a matrix. Then under this norm, a matrix has the same norm as it's transposed and the rest follows from equivalence of matrix norms due to finite dimensionality.[2] Fix some arbitrary $T > 0$. Then $\|G\|_{dA,T} < \infty$ implies the existence of a set $N \subset (0,T]$ with $dA(N) = 0$ s.t.

$$\begin{aligned}\|V(t,x) - V(t,y)\|_{\mathbb{R}^n} &= \left\|G(t)^\top(x-y)\right\|_{\mathbb{R}^n} \leq C\|G(t)\|_\infty \|x-y\|_{\mathbb{R}^n} \\ &\leq C\|G\|_{dA,T}\|x-y\|_{\mathbb{R}^n} \quad \forall t \in N^c,\end{aligned}$$

hence V is A-Lipschitz on $(0,T]$. Regarding the integrability condition, we choose $x* = 0$ since $G(t)^\top x* = 0$ for all $t \geq 0$. Hence, V is A-Lipschitz and integrable on $(0,T]$. Since T was arbitrary, this finishes the proof. □

[2] In case $\|\cdot\|_{\mathbb{R}^n}$ is the $p = 2$ norm, it is easy to see that for the corresponding operator norm we have $\|G\|_\infty = \left\|G^\top\right\|_\infty$.

Remark 4.52. *Note that the assumption in the lemma above imply existence and uniqueness of* (MDE:A,V)*. The important implication of V being defined by a matrix-function G is that solutions to* (MDE:A,V) *obey the superposition principle. That is under the assumption in Lemma 4.51, we can denote the unique solutions of* (MDE:A,V) *with $F_{s.x}(t)$ and we have for any $x,y\in\mathbb{R}^n$ and $(s,t)\in\mathbb{T}$ the equality*

$$F_{s,x}(t)+F_{s,y}(t)=F_{s,x+y}(t).$$

Linearity now implies that there exists a matrix function $(s,t)\mapsto M(s,t)$, called the fundamental solution, such that

$$F_{s,x}(t)=M(s,t)x \quad \forall x\in\mathbb{R}^n \text{ and } (s,t)\in\mathbb{T}.$$

Proposition 4.53. *Consider the assumptions in Lemma 4.51. Then, the fundamental solution for $V(t,x)=G(t)^\top x$ of* (MDE:A,V) *denoted by M defines a MES P through $P(s,t)=M(s,t)^\top$ and G is the extended generator of the proper MES (P,γ,A).*

Proof. We have already argued the existence of M in Remark 4.52. By Theorem 4.50, we have that for any $x\in\mathbb{R}^n$, the function $t\mapsto M(s,t)x$ is càdlàg on $[s,\infty)$ and $s\mapsto M(s,t)x$ is càdlàg on $[0,t]$. This is equivalent to $(s,t)\mapsto M(s,t)$ being càdlàg in the sense of MESs since we are dealing with matrices. Further, since $F_{s,x}(s)=x$ for all $s\geq 0$, we have $M(s,t)=\mathrm{Id_n}$ Note that the evolution property of solution flows implies for $s\leq u\leq t$,

$$M(u,t)M(s,u)=M(s,t)\Rightarrow P(s,u)P(u,t)=P(s,t).$$

Hence we have that P is a càdlàg MES. Next, note that we have

$$\begin{aligned}
P(s,t)^\top x=M(s,t)x=F_{s,x}(t)&=x+\int_{(s,t]}G(u)^\top F_{s,x}(u-)A(du)\\
&=x+\int_{(s,t]}G(u)^\top M(s,u-)xA(du)\\
&=(\mathrm{Id_n}+\int_{(s,t]}G(u)\top M(s,u-))xA(du)\\
&=(\mathrm{Id_n}+\int_{(s,t]}G(u)\top P(s,u-)^\top A(du))x\\
&=(\mathrm{Id_n}+\int_{(s,t]}P(s,u-)G(u)A(du))^\top x.
\end{aligned}$$

Since this holds for all $x\in\mathbb{R}^n$, we get the equality

$$P(s,t)=\mathrm{Id_n}+\int_{(s,t]}P(s,u-)G(u)A(du).$$

This implies that for all $s\geq 0$, the function $t\mapsto P(s,t)$ is of finite variation on compacts, defining a signed vector measure γ_s such that $P(s,u-)G(u)$ is a version of the Radon-Nikodým derivative wrt dA of γ_s. Since this holds for all $s\geq 0$, we get $\gamma\overset{\infty}{\ll}dA$. Hence, (P,γ,A) is a proper MES with extended generator G as claimed. $\square$

Remark 4.54. *When discussing MESs, we saw that an important property such an MES P with is whether it becomes singular. Assume G satisfies the condition of Lemma 4.51. Let $M = P^\top$ be the fundamental solution of* (MDE:A, V) *for g. Since $P(s,t) \in \mathrm{GL_n} \Leftrightarrow M(s,t) \in \mathrm{GL_n}$, we have a clear interpretation for the time point $t*$ defined by*

$$t^* = \inf\{t > s : P(s,t) \notin \mathrm{GL_n}\}.$$

By Proposition 4.25 (i), $P(s,t^*) \notin \mathrm{GL_n}$ which implies $M(s,t^*) \notin \mathrm{GL_n}$. Hence, for $0 \notin y \in \ker M(s,t^*)$ we have

$$F_{s,y}(t^*) = 0 \text{ or } F_{s,x+y}(t^*) = F_{s,x}(t^*).$$

Hence, t^* is the first time solutions to different initial values at starting time s merge, something that can not happen in the classical, fully regular case.

4.3.3 A Gronwall type inequality for measure differential equations

Here we just state a Gronwall type inequality applicable to our setup. As mentioned before, this inequality has been established in [EK09] in a slightly different way. The proof however, is also valid for our situation and we shall only proof this result here for the readers convenience.

The setup is as follows. As before, let $A \in \mathcal{V}^+_{\not{\mathfrak{A}}}$ and denote the corresponding measure with dA. Then, dA defines a Radon measure on $(\mathbb{R}_{>0}, \mathscr{B}(\mathbb{R}_{>0}))$. Note that we have $dA((0,t]) = A(t) - A(0) = A(t)$.

Theorem 4.55. *Let $u : [s,T] \to \mathbb{R}_+$ for some $T \in \mathbb{R}_{>0} \cup \{\infty\}$ and $s \geq 0$. Assume u to be càdlàg . Assume further, that there exists a non-negative function $u_0(t) \geq 0$ such that u satisfies the following inequality for any $T \geq t \geq s$,*

$$0 \leq u(t) \leq u_0(t) + \int_{(s,t]} u(\xi-)A(d\xi).$$

Then for all $t \leq T$,

$$u(t) \leq u_0(t)\exp\left(A(t) - A(s)\right) \leq u_0(t)\exp\left(A(t)\right) \tag{4.35}$$

Let us first proof the following lemma about the measure of the n-dimensional simplex wrt the product measure of dA, denoted by $dA^{\otimes n}$. We define the n-dimensional simplex over $(s,t]$ to be

$$I^n_{s,t} := \{(s_1,\ldots,s_n) \in (s,t]^n : s < s_1 < s_2 < \cdots < s_n = t\}.$$

The measure of $I^n_{s,t}$ wrt $dA^{\otimes n}$ is then given by

$$dA^{\otimes n}(I^n_{s,t}) = \int_{(s,t]} \int_{(s,\xi_1)} \cdots \int_{(s,\xi_{n-1})} A(d\xi)A(d\xi_{n-1})\cdots A(d\xi_1),$$

in particular it holds

$$dA^{\otimes 1}(I^1_{s,t}) = dA(I^1_{s,t}) = \int_{(s,t]} A(d\xi) = A(t) - A(s).$$

The following is a well established lemma for the measure of the simplex and we only provide a proof for the reader's convenience.

Lemma 4.56. *For $0 \leq s < t$ and $n \in \mathbb{N}$ it holds*

$$dA^{\otimes n}(I^n_{s,t}) \leq \frac{(dA^{\otimes 1}(I^1_{s,t}))^n}{n!} = \frac{(A(t) - A(s))^n}{n!}$$

Proof. Let S_n be the set of permutations on $\{1, \dots, n\}$. Recall that $|S_n| = n!$ and for each $\sigma \in S_n$, define the sets

$$I^{n,\sigma}_{s,t} = \left\{(s_1, \dots, s_n) \in (s,t]^n : s < s_{\sigma(1)} < s_{\sigma(2)} < \cdots < s_{\sigma(n)} = t\right\}.$$

By symmetry we have for two arbitrary $\sigma, \sigma' \in S_n$,

$$dA^{\otimes n}(I^{n,\sigma}_{s,t}) = dA^{\otimes n}(I^{n,\sigma'}_{s,t}).$$

Further, for all two $\sigma \neq \sigma'$, we have

$$I^{n,\sigma}_{s,t} \cap I^{n,\sigma'}_{s,t} = \emptyset.$$

Using $I^{n,\sigma}_{s,t} \subset (s,t]^n$, and $dA^{\otimes n}((s,t]) = (A(t) - A(s))^n$, we get

$$\sum_{\sigma \in S_n} dA^{\otimes n}(I^{n,\sigma}_{s,t}) \leq dA^{\otimes n}((s,t]) = (A(t) - A(s))^n,$$

which finally yields

$$dA^{\otimes n}(I^n_{s,t}) \leq \frac{(A(t) - A(s))^n}{n!}.$$

□

We can now prove Theorem 4.55.

Proof of Theorem 4.55. Let us start by noting that because u is càdlàg , we have

$$u(t-) \leq u_0(t) + \int_{(s,t)} u(\xi-)A(d\xi).$$

Hence we can iterate the assumed inequality to get

$$u(t) \leq u_0(t) + u_0(t)\sum_{k=1}^{n} dA^{\otimes k}(I^k_{s,t}) + \int_{(s,t]} u(\xi-)A(d\xi)(dA^{\otimes n}(I^n_{s,t})).$$

Note that because u is càdlàg and so is A, the integral is finite and does not depend on n. An application of Lemma 4.56 now yields

$$u(t) \leq u_0(t)(1 + \sum_{k=1}^{n} \frac{(A(t)-A(s))^n}{n!}) + R_n \tag{4.36}$$

with

$$R_n = \int_{(s,t]} u(\xi-)A(d\xi)\frac{(A(t)-A(s))^n}{n!}.$$

Hence, $\lim_{n\to\infty} R_n \to 0$. Taking $n \to \infty$ in (4.36) yields the desired inequality (4.35). □

Remark 4.57. *Let us stress that in our version of the Gronwall inequality, it is crucial that the integrand is the left limit of the function u, i.e. it appears $u(\xi-)$ rather then $u(\xi)$. Otherwise, we would need to redefine the simplex to fit our needs to be*

$$I^n_{s,t} := \{(s_1, \dots, s_n) \in (s,t]^n : s < s_1 \leq s_2 \leq \dots \leq s_n = t\}.$$

In this case, we would however have

$$I^{n,\sigma}_{s,t} \cap I^{n,\sigma'}_{s,t} \neq \emptyset,$$

which is why a more combinatorial discussion is necessary to bound the measure of the simplex in dimension n by a summable sequence wrt n. While possible to derive sufficient conditions for such a sequence to exist in the form of $\Delta A_t < 1$ for all t, the general case is unclear.

Chapter 5

Polynomials and matrix evolution systems

In the previous chapter we have seen how proper MESs (P, γ, A) relate to measure differential equations through their extended generators. These systems will play an important role in the next chapter, where we will introduce polynomial processes and show that these are indeed special semimartingales, motivating the introduction of so called polynomial semimartingales, provided that a MES related to the process satisfies assumptions of the kind we have encountered in the previous chapter. As the name suggests, multivariate polynomials and mappings between spaces of such polynomials will play an important role. An important observation the authors made in Cuchiero et al. [Cuc+12] is that when considering polynomials of at most a certain degree, finite dimensionality of these spaces implies that linear mappings such as conditional expectation operators can be represented by matrices when fixing a basis of Pol_k. Further, the extended generator of a process X, a mapping between function spaces, can be shown to map polynomials to polynomials of at most the same degree and hence permit a representing matrix.

We will start by introducing some notation and preliminaries, relating polynomials with corresponding Euclidean spaces, MSGs and MESs. Following that, we discuss the homogeneous case as considered in [Cuc+12]. This will serve as a motivation for us to extend the observations we make in that section to the two-parameter case in form of MESs. This extension will be presented in the last section of this chapter.

5.1 Preliminaries

As a first step, let us introduce some notation that will add clarity to the subsequent discussion. Throughout the rest of this chapter, let $k \in \mathbb{N}_0$ and $d \in \mathbb{N}$. Recall that we always consider polynomials over the whole $\mathbb{R}^d$, i.e. $\mathrm{Pol}_k = \mathrm{Pol}_k(\mathbb{R}^d)$ with

$$\mathrm{Pol}_k = \left\{ x \mapsto \sum_{|\mathbf{l}|=0}^{k} \alpha_{\mathbf{l}} x^{\mathbf{l}} \mid \alpha_{\mathbf{l}} \in \mathbb{R},^{l} = (l_1, \dots, l_d), x^{\mathbf{l}} = x_1^{l_1} \dots x_d^{l_d} \right\},$$

where we use the multi index notation, i.e the sum is taken over all multi indices $\mathbf{l} = (l_1, \dots, l_d)$ with $|\mathbf{l}| = \sum_i l_i \leq k$. We will usually omit the dependency wrt d since this will not lead to any ambiguity. This includes the dimensions of the respective polynomial spaces. Hence, define $N_k := \dim \mathrm{Pol}_k$. We denote the vector space of all polynomials with arbitrary large but finite degree with $\mathrm{Pol} = \bigcup_{n \in \mathbb{N}} \mathrm{Pol}_n$.

Denote by $\beta^l = (\beta_1^l, \dots, \beta_{N_l}^l)$ a fixed basis of Pol_l. To keep the following discussions simple, we make the following convention that is without loss of generality.

Convention 5.1. *For the set of bases $\{\beta^l : l \in \mathbb{N}_0\}$, it holds*

(i) $\beta^i \subset \beta^{i+1}$ for $0 \leq i \leq i+1 \leq k$, and

(ii) if $p \in \beta^i \cap \beta^{i+1}$, then $p = \beta_l^i = \beta_m^{i+1}$ implies $m = l$. Hence, β^{i+1} is obtained by extending the basis β^i.

Therefore we can define a basis of Pol, denoted with $\beta = \{\beta_j : j \in \mathbb{N}\}$, by $\beta_j = \beta_j^i$, where i is such that $j \leq N_i$. From here on we fix β and all β^l.

Remark 5.2. *Note that the canonical basis of* Pol *and* Pol_k *for $k \in \mathbb{N}_0$ introduced in the preliminaries satisfy this condition.*

To show that these conventions are without loss of generality, it will be sufficient to show that matrix evolution systems are stable under basis transformations.

Proposition 5.3. *Let $T \in \mathrm{GL}_n$.*

(i) Let P be a MES. Then TPT^{-1} is a MES.

(ii) Let (P, γ, A) be a proper MES with generator G. Then $(\tilde{P}, \tilde{\gamma}, A)$ is a proper MES with generator $\tilde{G}$, where

$$\tilde{P} = TPT^{-1}, \tilde{\gamma}_s = T\gamma_s T^{-1}, \tilde{G} = TGT^{-1}.$$

Proof. (i) Since $P(s,s) = \mathrm{Id}_{\mathrm{N_k}}$, we get $\tilde{P}(s,s) = \mathrm{Id}_{\mathrm{N_k}}$. Further, we have for $s \leq u \leq t$,

$$\tilde{P}(s,u)\tilde{P}(u,t) = TP(s,u)T^{-1}TP(u,t)T^{-1} = TP(s,t)T^{-1} = \tilde{P}(s,t).$$

(ii) We have already shown that $\tilde{P}$ is a MES. Further, since T and T^{-1} are constant matrices, $\tilde{P}$ is again càdlàg and of finite variation on compacts. Further, we compute

$$\tilde{P}(s,t) - \tilde{P}(s,u) = T(\int_{(u,t]} \frac{\gamma_s}{dA}(\xi)A(d\xi))T^{-1} = \int_{(u,t]} T\frac{\gamma_s}{dA}(\xi)T^{-1}A(d\xi).$$

Hence, if $\tilde{\gamma}_s$ is the signed vector measure induced by $\tilde{P}$, then $\tilde{\gamma} \overset{\infty}{\ll} dA$ and $\frac{d\tilde{\gamma}_s}{dA} = T\frac{d\gamma_s}{dA}T^{-1}$, hence $\tilde{\gamma}_s = T\gamma_s T^{-1}$. This shows that $(\tilde{P}, \tilde{\gamma}, A)$ is a proper MES. We further have

$$\begin{aligned} &\mathrm{Id}_{\mathrm{N_k}} + \int_{(s,t]} \tilde{P}(s,u-)\tilde{G}(u)A(du) \\ &= T\int_{(s,t]} P(s,u-)G(u)A(du)T^{-1} = TP(s,t)T^{-1} = \tilde{P}(s,t), \end{aligned}$$

hence $\tilde{G}$ is the extended generator of $(\tilde{P}, \tilde{\gamma}, A)$.

□

When dealing with mappings from some polynomial space to another, we are interested in properties such as continuity. Hence we will define a norm on Pol which induces a topology. We can then equip all subspaces Pol_n for $n \in \mathbb{N}$ with the trace topology. It is well known that Pol can be identified with c_{00}, the space of all zero sequences with all but finitely many elements non-equal to zero. Then for $p \in \mathrm{Pol}$, $x \in \mathbb{R}^d$, there exists a $c = (c_1, \ldots) \in c_{00}$ with

$$p(x) = \sum_{i \in \mathbb{N}} c_i \beta_i(x).$$

We define $\|\cdot\|_{\mathrm{Pol}} : \mathrm{Pol} \to \mathbb{R}_+$ for a polynomial p with coefficients $c \in c_{00}$ via

$$\|p\|_{\mathrm{Pol}} = \sum_{i \in \mathbb{N}} |c_i|.$$

Recall that the sum of the absolute values defines a norm $\|\cdot\|_{c_00}$ on c_{00}, so the relationship $p \leftrightarrow c$ defines an isometric isomorphism between Pol and c_{00} wrt. their respective norms. We denote this mapping with $\iota : \mathrm{Pol} \to c_{00}$, $\iota(p) = c$. While Pol and equivalently c_{00} are not complete under these norms, their finite dimensional subspaces are as they can be identified with Euclidean spaces equipped with the $\|\cdot\|_1$ norm. For us, the most interesting subspaces of c_{00} are those, that correspond to Pol_l for some $l \in \mathbb{N}$. We formalize this by introducing two maps.

First, for $i \in \mathbb{N}$, define the map $T_i : c_{00} \to \mathbb{R}^i$ by $(T_i c)_j = c_j$ for $j \leq i$. Second, for $l \in \mathbb{N}_0$, define $\mathrm{pr}_l : \mathrm{Pol} \to \mathrm{Pol}_l$ the following way. Let $p \in \mathrm{Pol}$ with $\iota(p) = c$. Then

$$\mathrm{pr}_l(p) = \sum_{i=1}^{N_l} c_i \beta_i \in \mathrm{Pol}_l.$$

$$
\begin{tikzcd}
\mathrm{Pol} \arrow[r, shift right, "\iota"'] \arrow[d, "\mathrm{pr}_l"'] & c_{00} \arrow[l, shift right, "\iota^{-1}"'] \arrow[d, "T_{N_l}"] \\
\mathrm{Pol}_l \arrow[r, dashed, shift right, "\iota_l"'] & \mathbb{R}^{N_l} \arrow[l, dashed, shift right, "\iota_l^{-1}"']
\end{tikzcd}
$$

(a)

$$
\begin{tikzcd}
\mathrm{Pol} \arrow[r, "\iota"] & c_{00} \\
\mathrm{Pol}_l \arrow[u, "\pi_l"] \arrow[r, "\iota_l"'] & \mathbb{R}^{N_l} \arrow[u, "\lambda_{N_l}"']
\end{tikzcd}
$$

(b)

$$
\begin{tikzcd}
\mathrm{Pol}_l \arrow[r, "\tau"] & \mathrm{Pol}_l \arrow[d, "\iota_l"] \\
\mathbb{R}^{N_l} \arrow[u, "\iota_l^{-1}"] \arrow[r, dashed, "P"'] & \mathbb{R}^{N_l}
\end{tikzcd}
$$

(c)

Figure 5.1: (a): Commutative diagram depicting T_{N_l} as the representing matrix of pr_l and defining the maps ι_l and ι_l^{-1}. (b): Commutative diagram showing that π_l and λ_{N_l} define the same maps under ι. (c): Commutative diagram defining the representing matrix P.

Hence, the mappings T_{N_l} and pr_l define the same mappings under ι. This is shown by the commutative diagram in Figure 5.1 (a) where the maps ι_l and ι_l^{-1} are defined by the rest of the diagram, indicated by the dashed arrows. This definition is possible since pr_l and T_{N_l} are both surjective. Further, let π_l be the inclusion map from $\mathrm{Pol}_l \to \mathrm{Pol}$ and λ_{N_l} the one from $\mathbb{R}^{N_l} \to c_{00}$, i.e. in particular we have $\mathrm{pr}_l \circ \pi_l = \mathrm{id}_{\mathrm{Pol}_l}$ and $T_{N_l} \circ \lambda_{N_l} = \mathrm{id}_{\mathbb{R}^{N_l}}$. We then have $\iota \circ \pi_l = \lambda_{N_l} \circ \iota_l$, compare Figure 5.1 (b).

We equip each Pol_k with the subspace norm, inducing the trace topology wrt. Pol. Since for $v \in \mathbb{R}^{N_l}$, we have $\|v\|_1 = \|\lambda_{N_l}(v)\|_{c_{00}}$, this shows that indeed each Pol_k is a complete normed space. This should be no surprise since one could identify each Pol_k directly with a corresponding Euclidean space. We will however be dealing with a situation where we are confronted with several subspaces of Pol simultaneously, therefore this seems to be the most natural approach.

The following proposition states the obvious.

Proposition 5.4. *The map ι_l is an isometric isomorphism between $(\mathrm{Pol}_l, \|\cdot\|_{\mathrm{Pol}})$ and $(\mathbb{R}^{N_l}, \|\cdot\|_1)$ and ι_l^{-1} is it's inverse.*

Proof. Consider Figure 5.1. Diagram (a) implies $\iota_l \circ \mathrm{pr}_l = T_{N_l} \circ \iota$ and $\iota_l^{-1} \circ T_{N_l} = \mathrm{pr}_l \circ \iota^{-1}$. Using that ι is an isomorphism, we have $\iota_l^{-1} \circ T_{N_l} \circ \iota = \mathrm{pr}_l$ which implies $\mathrm{pr}_l = \iota_l^{-1} \circ \iota_l \circ \mathrm{pr}_l$. Since the range of pr_l is Pol_l, we get $\iota^{-1} \circ \iota = \mathrm{id}_{\mathrm{Pol}_l}$. The same way, we get $\iota_l \circ \mathrm{pr}_l \circ \iota^{-1} = T_{N_l}$, hence $\iota_l \circ \iota_l^{-1} \circ T_{N_l} = T_{N_l}$. Again we argue that since T_{N_l} has range $\mathbb{R}^{N_l}$, we get $\iota_l \circ \iota_l^{-1} = \mathrm{id}_{\mathbb{R}^{N_l}}$. Let $p \in \mathrm{Pol}_l$. Using that ι is isometric, we compute

$$\|p\|_{\mathrm{Pol}} = \|\pi_l p\|_{\mathrm{Pol}} = \|(\iota \circ \pi_l)p\|_{c_{00}} = \|(\lambda_{N_l} \circ \iota_l)p\|_{c_{00}} = \|\iota_l p\|_1 \,,$$

where we made use of $(\iota \circ \pi_l) = (\lambda_{N_l} \circ \iota_l)$ and in the last step the fact that under our choice of norms, the inclusion map λ_{N_l} preserves the norm. $\square$

Let us now consider the finite dimensional spaces Pol_k and $\mathbb{R}^{N_k}$ and for $l \leq k$, consider Pol_l and $\mathbb{R}^{N_l}$. We have that Pol_l is a subspace of Pol_k, so an embedding is clear and given by $\pi_{l,k} := \mathrm{pr}_k \circ \pi_l$. Further, the projection $\mathrm{pr}_{k,l} : \mathrm{Pol}_k \to \mathrm{Pol}_l$ is defined by $\mathrm{pr}_{k,l} = \mathrm{pr}_l \circ \pi_k$.

Further, $\mathbb{R}^{N_l}$ is only isometrically isomorph to N_l-dimensional subspaces of $\mathbb{R}^{N_l}$ and there are several equally reasonable embeddings possible. However, in light of the convention we made for β, we choose to fix the embedding induced by λ_{N_l} and T_{N_k}, hence we define the embedding of $\mathbb{R}^{N_l}$ into $\mathbb{R}^{N_k}$ denoted by $\lambda_{N_l,N_k} : \mathbb{R}^{N_l} \to \mathbb{R}^{N_k}$ and the projections $T_{N_k,N_l} : \mathbb{R}^{N_k} \to \mathbb{R}^{N_l}$ by

$$\lambda_{N_l,N_k} := T_{N_k} \circ \lambda_{N_l}, \quad T_{N_k,N_l} := T_{N_l} \circ \lambda_{N_k}.$$

As a next step, let us recall the following definitions. The function $\mathbf{H}_{\beta^k} : \mathbb{R}^d \to \mathbb{R}^{N_k}$ was defined as

$$x \mapsto \mathbf{H}_{\beta^k}(x) = (\beta_1(x), \ldots, \beta_{N_k}(x))^\top .$$

Hence, for $p \in \mathrm{Pol}_k$, we have

$$p(x) = \iota_k(p)^\top \mathbf{H}_{\beta^k}(x) \quad \forall x \in \mathbb{R}^d.$$

We will sometimes indicate $p = \iota_k(p)$ through bold letters, if it is clear which k is meant, i.e. $\iota_k(p) = \mathbf{p}$. Let $\tau : \mathrm{Pol}_k \to \mathrm{Pol}_k$. Define the map $P : \mathbb{R}^{N_k} \to \mathbb{R}^{N_k}$ via Figure 5.1 (c), i.e. $P = \iota_l \circ \tau \circ \iota_l^{-1}$. Since ι_l is an isomorphism, we get that P is linear if and only if τ is linear. In that case, P can be identified wrt it's representing matrix and in the sequel, we will always do so.

Let $l \leq k$. For $v \in \mathbb{R}^{N_l}$ and $w \in \mathbb{R}^{N_k}$, we write $v \overset{\mathrm{Pol}}{=} w$, if $\lambda_{N_l}(v) = \lambda_{N_k}(w)$. This implies $w \in \mathrm{im}(\lambda_{N_l,N_k})$ with $\lambda_{N_l,N_k} v = w$. Further, we then have $\exists! p \in \mathrm{Pol}_l$, such that

$$p(x) = v^\top \mathbf{H}_{\beta^l}(x) = w^\top \mathbf{H}_{\beta^k}(x), \quad \forall x \in \mathbb{R}^d.$$

Note that the last two implications are equivalent characterizations of $v \overset{\mathrm{Pol}}{=} w$. Let us now define the projection of mappings $\tau : \mathrm{Pol}_k \to \mathrm{Pol}_k$ to subspaces Pol_l for $l \leq k$.

Definition 5.5. *Let $l \leq k$ and consider $\tau : \mathrm{Pol}_k \to \mathrm{Pol}_k$. We define the map $\tau^l : \mathrm{Pol}_l \to \mathrm{Pol}_l$, the* **projection** *to Pol_l by*

$$\tau^l = \mathrm{pr}_{k,l} \circ \tau \circ \pi_{l,k},$$

compare the diagram in Figure 5.2 (a).

We can now define a key property of linear maps on Pol_k that will play a central role in our following discussions.

Definition 5.6. *We call a mapping $\tau : \mathrm{Pol}_k \to \mathrm{Pol}_k$* **degree preserving** *on Pol_k (or just degree preserving if it is clear which k is meant), if for all $l \leq k$, with projections τ^l, we have*

$$\pi_k \circ \tau \circ \pi_{l,k} = \pi_l \circ \tau^l,$$

or equivalently

$$\tau \circ \pi_{l,k} = \pi_{l,k} \circ \tau^l.$$

$$\begin{tikzcd}
\mathrm{Pol}_l \arrow[r, dashed, "\tau^l"] \arrow[d, "\pi_{l,k}"] & \mathrm{Pol}_l \\
\mathrm{Pol}_k \arrow[r, "\tau"] & \mathrm{Pol}_k \arrow[u, "\mathrm{pr}_{k,l}"']
\end{tikzcd}$$

(a)

$$\begin{tikzcd}
\mathrm{Pol}_l \arrow[d, "\tau^l"] & \mathbb{R}^{N_l} \arrow[l, "\iota_l^{-1}"'] \arrow[rr, "\lambda_{N_l,N_k}"] \arrow[d, dashed, "P^l"] & & \mathbb{R}^{N_k} \arrow[r, "\iota_k^{-1}"] \arrow[d, dashed, "P"] & \mathrm{Pol}_k \arrow[d, "\tau"] \\
\mathrm{Pol}_l \arrow[r, "\iota_l"'] & \mathbb{R}^{N_l} \arrow[r, "\lambda_{N_l}"'] & c_{00} & \mathbb{R}^{N_k} \arrow[l, "\lambda_{N_k}"] & \mathrm{Pol}_k \arrow[l, "\iota_k"]
\end{tikzcd}$$

(b)

Figure 5.2: (a): Commutative diagram defining the projection τ^l of a map τ for $l \leq k$. (b): Diagram showing the relation of degree preserving properties between maps and their respective projections.

This property can be formulated as $\tau(\mathrm{Pol}_l) \subset \mathrm{Pol}_l$ for all $l \leq k$, or for all $p \in \mathrm{Pol}_l$, noting that the inclusion map satisfies $\pi_{l,k}p = p$,

$$\tau^l p = \tau p.$$

It is clear that for degree preserving mappings τ, this property is inherited by the representing matrices of τ, denoted by P and the one fore τ^l denoted by P^l, see the next proposition. We call P^l the projection of P on Pol_l. By definition of the projection τ^l, P^l is given by for each

$$P^l = T_{N_k,N_l} P \lambda_{N_l,N_k}.$$

Proposition 5.7. *Consider τ with projection τ^l to Pol_l and P, P^l the respective representing matrices. Then for any $v \in \mathbb{R}^{N_l}$ and $w \in \mathbb{R}^{N_k}$ with $v \overset{\mathrm{Pol}}{=} w$ we have*

$$Pw \overset{\mathrm{Pol}}{=} P^l v,$$

if and only if τ is degree preserving on Pol_k.

Proof. By definition of the representing matrices, P and P^l are defined such that the left and the right subdiagrams in Figure 5.2 (b) commute. If now the whole diagram commutes, then this implies that both, τ preserves the degree and $Pw \overset{\mathrm{Pol}}{=} P^l v$, compare in particular with . We show either condition implies that that diagram commutes. Assume now τ is degree preserving, hence $\pi_k \circ \tau \circ \pi_{l,k} = \pi_l \circ \pi_{l,k} \circ \tau^l$. Since $\pi_i = \iota^{-1} \circ \lambda_{N_i} \circ \iota_i$ for $i \in \{l, k\}$, this yields

$$\iota^{-1} \circ \lambda_{N_k} \circ \iota_k \circ \tau \circ \pi_{l,k} = \iota^{-1} \circ \lambda_{N_l} \circ \iota_l \circ \tau^l \Rightarrow \lambda_{N_k} \circ \iota_k \circ \tau \circ \pi_{l,k} = \lambda_{N_l} \circ \iota_l \circ \tau^l,$$

since ι is isomorph. Further, since $\pi_{l,k} = \iota_k^{-1} \circ \lambda_{N_l,N_k} \circ \iota_l$, this yields using $\iota_l \circ \iota_l^{-1} = \mathrm{id}_{N_l}$,

$$\lambda_{N_k} \circ \iota_k \circ \tau \circ \iota_k^{-1} \circ \lambda_{N_l,N_k} = \lambda_{N_l} \circ \iota_l \circ \tau_l \circ \iota_l^{-1},$$

implying that the diagram commutes. The rest is more direct. Assume $Pw \overset{\text{Pol}}{=} P^l v$ for all $w = \lambda_{N_l,N_k} v$. Then this implies

$$\lambda_{N_l} \circ P^l = \lambda_{N_k} \circ P \circ \lambda_{N_l,N_k},$$

hence the diagram commutes. This finishes the proof. □

In the sequel, we write $P = \iota_k(\tau)$ to indicate that P is the representing matrix of the linear map $\tau : \mathrm{Pol}_k \to \mathrm{Pol}_k$. By linearity we then have for all $p \in \mathrm{Pol}_k$,

$$(\tau p)(x) = \iota_k(\tau p)^\top \mathbf{H}_{\beta^k}(x) = (P\,\iota_k(p))^\top \mathbf{H}_{\beta^k}(x) = (P\,\mathbf{p})^\top \mathbf{H}_{\beta^k}(x),$$

where $P\mathbf{p}$ is the matrix-vector product. The following definition will simplify notation in the future.

Definition 5.8. *For an index set $I \in \{\mathbb{R}_+, \mathbb{R}_{>0}, \mathbb{T}\}$, and a map $\tau : I \times \mathrm{Pol}_k \to \mathrm{Pol}_k$ with $(t,p) \mapsto \tau_t p$, we call τ pointwise linear on Pol_k if each τ_t is a linear map on Pol_k for $t \in I$. For $P : I \to M_{N_k}(\mathbb{R})$, we write $\iota_k(\tau) = P$ if for all $t \in I$, $\iota_k(\tau_t) = P(t)$. Finally, we call τ degree preserving on Pol_k, if it is pointwise degree preserving on Pol_k.*

Remark 5.9. *Note that by bijectivity of ι_k, one can conversely define an operator $\tau : I \times \mathrm{Pol}_k \to \mathrm{Pol}_k$ through a matrix function $P : I \to M_{N_k}(\mathbb{R})$ by $P = \iota_k(\tau)$.*

5.2 Degree preserving properties of time-homogeneous polynomial processes

Let us continue our discussions in the context of the results presented in Cuchiero et al. [Cuc+12], in particular we consider the index set $I = \mathbb{R}_+$. The authors consider a Markov process $(X, (\mathbb{P}_x)_{x\in E})$ with state space E such that there exists a semigroup $\tau = (\tau_t)_{t\in\mathbb{R}_+}$ on Pol_k (which by definition of semigroups is pointwise linear) such that for $f \in \mathrm{Pol}_k$,

$$\mathbb{E}_x\left[f(X_t)\right] = \tau_t f(x),$$

where the expectation is taken under the measure $\mathbb{P}_x$ and $t \mapsto \tau_t f(x)$ is right continuous at $t = 0$ for all polynomials $f \in \mathrm{Pol}_k$ and all $x \in E$. Unlike our case, they consider polynomials over a closed subset of $\mathbb{R}^d$ rather then $\mathbb{R}^d$ itself, namely the state space E. For our following discussions however, we assume polynomials over all of $\mathbb{R}^d$.

Setting $P = \iota_k(\tau)$, the continuity assumption implies that P is a regular MSG, c.f. Proposition 4.2. By Proposition 4.2, P has an infinitesimal generator G with $P(t) = \exp(tG)$. For the readers convenience, let us prove these points.

Proposition 5.10. *Given a pointwise linear τ indexed by $\mathbb{R}_+$, let $P = \iota_k(\tau)$. If τ satisfies the right continuity assumption at $t = 0$ made above, then P is a regular MSG.*

Proof. As mentioned, we only need to show $\lim_{t\downarrow 0} P(t) = P(0) = \mathrm{Id}_{N_k}$. Since τ is a semigroup, $\tau_0 = \mathrm{id}_{\mathrm{Pol}_k}$, hence $P(0) = \mathrm{Id}_{N_k}$. For arbitrary polynomial $p \in \mathrm{Pol}_k$, we have by the semigroup property of τ that $\tau_{s+t}p = \tau_s \circ \tau_t p$, hence $P(s+t)\mathbf{p} = P(s)P(t)\mathbf{p}$. Since p was arbitrary, we get $P(s)P(t) = P(s+t)$. Further,

$$\lim_{t\downarrow 0} \tau_t f(x) = f(x)$$

by assumption. This implies

$$\lim_{t\downarrow 0}(P(t)\mathbf{f})^\top \mathbf{H}_{\beta^k}(x) - \mathbf{f}^\top \mathbf{H}_{\beta^k}(x) = \mathbf{f}^\top \left(\lim_{t\downarrow 0}(P(t) - \mathrm{Id_{N_k}})\right)^\top \mathbf{H}_{\beta^k}(x) = 0,$$

for all $\mathbf{f} \in \mathbb{R}^{N_k}, x \in \mathbb{R}^d$, which implies the claim. Otherwise, there is a vector $\mathbf{g} \neq 0$ such that $\mathbf{g}^\top \mathbf{H}_{\beta^k}(x) = 0$ for all x which is only possible for the zero polynomial that we excluded with $\mathbf{g} \neq 0$. $\square$

The next lemma shows that projections inherit the finite variation on compacts property. We consider the case $I = \mathbb{R}_+$, but the proof is valid for $I = \mathbb{T}$ if one fixes the first coordinate s of $(s,t) \in \mathbb{T}$.

Lemma 5.11. *Assume $P = \iota_k(\tau)$. The following two points are equivalent.*

(i) The function $t \mapsto P(t)$ is of finite variation on compacts in each component,

(ii) for all $l \leq k$, the projections $P^l = \iota_l(\tau^l)$ satisfy $t \mapsto P^l(t)$ is of finite variation on compacts in each component.

Proof. As a first step, let us show that the finite variation property of P is equivalent to $t \mapsto \tau_t f(x)$ is of finite variation for all $f \in \mathrm{Pol}_k$ and $x \in \mathbb{R}^d$ fixed. Note that we have

$$t \mapsto \tau_t p(x) = (P(t)\mathbf{p})^\top \mathbf{H}_{\beta^k}(x) = \sum_{i=1}^{N_k} \alpha_i(t)\beta_i(x) = \sum_{i=1}^{N_k} \tau_t \beta_i(x),$$

for some functions α_i. Hence if P satisfies the finite variation property, than so must $t \mapsto \tau_t f(x)$ since in the second part, only P depends on t. For the converse, assume $t \mapsto \tau_t p(x)$ satisfies the finite variation property for all p and x while P does not. But that implies that there exists a vector $\mathbf{p} \in \mathbb{R}^{N_k}$ such that at least for one i, α_i is not of finite variation on compacts if x is fixed. That implies $t \mapsto \tau_t \beta_i(x)$ is not of finite variation on compacts which is a contradiction since $\beta_i \in \mathrm{Pol}_k$.

Considering $(ii) \Rightarrow (i)$, note that this is clear since we can choose $l = k$, hence $P^k = P$. We now show $(i) \Rightarrow (ii)$. We have by the first arguments we made that $kt \mapsto \tau_f(x)$ is of finite variation on compacts for all $f \in \mathrm{Pol}_k, x \in \mathbb{R}^d$. Further, we have

$$\tau_t^l f = \mathrm{pr}_{k,l} \circ \tau_t \circ \pi_{l,k}.$$

Since $\mathrm{pr}_{k,l}$ and the inclusion $\pi_{l,k}$ do not depend on t, this implies $t \mapsto \tau_t^l f(x)$ is of finite variation on compacts for all $f \in \mathrm{Pol}_l, x \in \mathbb{R}^d$. This in return implies by our first arguments (ii). $\square$

Note that a matrix $G \in \mathrm{M}_{N_k}(\mathbb{R})$ defines an operator $\mathcal{G} : \mathrm{Pol}_k \to \mathrm{Pol}_k$ through (note we do not differentiate between a linear map on $\mathbb{R}^{N_k}$ and the representing matrix)

$$\mathcal{G}f = \iota_k^{-1} \circ G \circ \iota_k, \quad \mathcal{G}f(x) = (G\mathbf{f})^\top \mathbf{H}_{\beta^k}(x).$$

The authors in [Cuc+12] show that the generator G defines a map $\mathcal{G}$, that coincides with the extended generator of the process X, a concept we will present in the next chapter. A key observation is that if τ_t is degree preserving on Pol_k for all $t \geq 0$, an assumption that is part of the definition of polynomial processes, compare [Cuc+12, Remark 2.2 (ii)], then so is $\mathcal{G}$. In light of Proposition 5.7, this is equivalent to the preserving property of the representing matrices. For illustration, we want to rephrase the condition $w \overset{\mathrm{Pol}}{=} v$ for $w \in \mathbb{R}^{N_k}, v \in \mathbb{R}^{N_l}$. Denote by $e^k = \{e_1^k, \dots, e_{N_k}^k\}$ the canonical standard basis of $\mathbb{R}^{N_k}$. Define $E_l = \mathrm{span}\{e_i : i \leq N_l\}$. We then have

$$E_l = \{w \in \mathbb{R}^{N_k} : \exists v \in \mathbb{R}^{N_l} \text{ s.t. } w \overset{\mathrm{Pol}}{=} v\}.$$

The property that $\mathcal{G}$ is degree preserving than is equivalent to $Gw \in E_l$ for all $w \in E_l$ and all $l \leq k$, hence we make the following definition.

Definition 5.12. *We call a matrix $G \in M_{N_k}(\mathbb{R})$ degree preserving on* Pol_k*, if for all $l \leq k$, $Gw \in E_l$ for all $w \in E_l$.*

Hence we have the following corollary to Proposition 5.7.

Corollary 5.13. *Consider a linear mapping $\mathcal{G} : \mathrm{Pol}_k \to \mathrm{Pol}_k$ with $G = \iota_k(\mathcal{G})$. Then $\mathcal{G}$ is degree preserving on* Pol_k *if and only if G is degree preserving on* Pol_k*.*

In the classical homogeneous case where G can be expressed as the limit

$$G = \lim_{t \downarrow 0} \frac{P(t) - \mathrm{Id}_{N_k}}{t},$$

it follows directly that $G(E_l) \subset (E_l)$ and by that $\mathcal{G}$ is preserving since $P(t)(E_l) \subset (E_l)$ for all t and $\mathrm{Id}_{N_k}(E_l) = E_l$ combined with the fact that the finite dimensional subspaces E_l are closed. This extends to the inhomogeneous case in which the infinitesimal generator can be expressed as a limit and it is the approach that is implicitly taken in C. A. Hurtado [CAH17]. This approach will not be available to us when dealing with polynomial semimartingales. There is however another approach to showing that $\mathcal{G}$ inherits the preserving property and it is the one the authors took in [Cuc+12].

Under the assumption that P is degree preserving, each projection P^l defines a MSG. Since the projections τ^l are right continuous if τ is right continuous, Proposition 5.10 implies that each P^l is regular with infinitesimal generator G^l. It is then possible to show that the infinitesimal generator G of P can be represented by a sum of maps, defined by each G^l implying that G is indeed degree preserving. For us the situation presented in the next section is different as it is not clear that a projected MES has an extended generator if the original one does. We will however be able to show that under our most general assumption that implies the existence of an extended generator (c.f. Theorem 4.35), there exists a version of the generator that preserves the degree.

5.3 Degree preserving properties of matrix evolution systems and extended generators

Let us now turn to the inhomogeneous setting as it is relates to our setup, hence we have the index set $I = \mathbb{T}$. Consider $\tau : \mathbb{T} \times \mathrm{Pol}_k \to \mathrm{Pol}_k$. We call τ càdlàg , if $s \mapsto \tau_{s,t} f(x)$ is càdlàg on $[0, t]$ and $t \mapsto \tau_{s,t} f(x)$ is càdlàg on $[t, \infty)$ for all $f \in \mathrm{Pol}_k$ and $x \in \mathbb{R}^d$. Define $P = \iota_k(\tau)$. We call τ an evolution system on Pol_k if P is a MES in dimension N_k. It is clear that τ is càdlàg if and only if P is càdlàg , since

$$\iota_k(\tau_{s,t} f) = P(s,t)\iota_k(f), \tag{5.1}$$

noting that the components of the left hand site are the coefficients of $\tau_{s,t} f$ wrt. the basis β^k. The next definition extends the degree preserving property to matrix functions such as MESs and extended generators. Denote with $I \in \{\mathbb{R}_{>0}, \mathbb{T}\}$ an index set.

Definition 5.14. *We call a function $T : I \to M_{N_k}(\mathbb{R})$ degree preserving on Pol_k, if for any $i \in I$, $T(i)$ is degree preserving. Hence, a mapping $\tau : I \times \mathrm{Pol}_k \to \mathrm{Pol}_k$ with $\iota_k(\tau) = T$ is degree preserving, if and only if T is degree preserving.*

Let us continue our discussions with the following lemma.

Lemma 5.15. *Let $l \le k$ and consider a degree preserving operator $\tau : \mathrm{Pol}_k \to \mathrm{Pol}_k$ with projections $\tau^l : \mathrm{Pol}_l \to \mathrm{Pol}_l$. Then τ is bijective, if and only if τ^l is bijective for all $l \le k$.*

Proof. The one direction, that is the if and only if part, is clear since $\tau^k = \tau$. Assume now τ is bijective. Let $p \in \mathrm{Pol}_l$. Then the degree preserving property implies $\tau_l p = \tau p$, hence $(\tau^l)^{-1} := \mathrm{pr}_{k,l} \circ \tau^{-1} \circ \pi_{l,k}$ satisfies $(\tau^l)^{-1}\tau^l = \tau^l(\tau^l)^{-1} = \mathrm{id}_{\mathrm{Pol}_l}$. □

The next is an important implication as we will need it later.

Corollary 5.16. *Let $l \le k$ and assume $\tau : \mathbb{T} \times \mathrm{Pol}_k \to \mathrm{Pol}_k$ has representing matrices $P = \iota_k(\tau)$. Let $\tau^l : \mathbb{T} \times \mathrm{Pol}_l \to \mathrm{Pol}_l$ be the pointwise defined projections with representing matrices $P^l = \iota_l(\tau^l)$. If τ is degree preserving, then*

$$\{t : P(t-,t) \notin \mathrm{GL}_{N_k}\} = \bigcup_{l \le k} \left\{ t : P^l(t-,t) \notin \mathrm{GL}_{N_l} \right\}.$$

We now show that under the assumption that τ is degree preserving, the projections define evolution systems on the respective subspaces.

Proposition 5.17. *Consider an evolution system on Pol_k with $P = \iota_k(\tau)$. If τ is degree preserving on Pol_k, then we have for the projections τ^l with $l \le k$ that they are an evolution system on Pol_l and $\iota_l(\tau^l) = P^l$ is a MES in dimension N_l. Further, if P belongs to a proper MES (P, γ, A), then there exist families of signed vector measures $\gamma^l \in M_{N_l}(\mathbb{R})$ for each l such that (P^l, γ^l, A) is a proper MES.*

Proof. Fix $s \leq u \leq t$. If τ is degree preserving, then $\tau^l_{s,t}p = \tau_{s,t}p$ for all $p \in \mathrm{Pol}_l$. Since $\tau_{s,s} = \mathrm{id}_{\mathrm{Pol}_k}$, this implies $\tau^l_{s,t} = \mathrm{id}_{\mathrm{Pol}_l}$. This implies $P^l(s,s) = \mathrm{Id}_{N_l}$. Further, we get $\mathrm{pr}_{k,l} \circ \tau^l_{u,t}p = \tau^l_{u,t}p$. Hence, $\tau^l_{s,u} \circ \tau^l_{u,t}p = \tau_{s,u} \circ \tau_{u,t}p = \tau_{s,t}p = \tau^l_{s,t}p$ which implies $P^l(s,u)P^l(u,t) = P^l(s,t)$. Therefore, P^l is a MES in dimension N_l and by that τ^l an evolution system on Pol_l.

Assume now (P, γ, A) is a proper evolution system. By Lemma 5.11, this implies that for all $l \leq k$, the functions $t \mapsto P(s,t)$ are of finite variation on compacts for all fixed $s \leq t$. Since P is a càdlàg MES, so is τ by Equation (5.1). Since $\tau^l = \mathrm{pr}_{k,l} \circ \tau \circ \pi_{l,k}$, and neither $\mathrm{pr}_{k,l}$ nore $\pi_{l,k}$ depend on s or t, τ^l is càdlàg which again by (5.1) implies P^l is càdlàg . Here we stress the point that ι_l is an isomorphism and hence $\{v : \exists f \in \mathrm{Pol}_l \text{ s.t. } \iota_l(f) = v\} = \mathbb{R}^{N_l}$.

Let now $f \in \mathrm{Pol}_k$ and $p \in \mathrm{Pol}_l$ such that $p = \mathrm{pr}_{k,l}f$. Note that $\mathrm{pr}_{k,l}$ is surjective. By Proposition 5.7, we have

$$P\iota_k(\pi_{l,k} \circ \mathrm{pr}_{k,l}f) \overset{\mathrm{Pol}}{=} P^l\iota_l(\mathrm{pr}_{k,l}f). \tag{5.2}$$

Recall that T_{N_k,N_l} is the representing matrix of $\mathrm{pr}_{k,l}$ and λ_{N_l,N_k} is the embedding of $\mathbb{R}^{N_l}$ into $\mathbb{R}^{N_k}$, i.e. we have for all $f \in \mathrm{Pol}_k$ the equality $\iota_k(\pi_{l,k} \circ \mathrm{pr}_{k,l}f) = \lambda_{N_l,N_k}T_{N_k,N_l}\iota_k(f)$. Hence we get

$$P\lambda_{N_l,N_k}T_{N_k,N_l}\iota_k(f) \overset{\mathrm{Pol}}{=} P^l\iota_l(\mathrm{pr}_{k,l}f).$$

This in return yields

$$T_{N_k,N_l}P\lambda_{N_l,N_k}T_{N_k,N_l}\iota_k(f) = P^l\iota_l(\mathrm{pr}_{k,l}f).$$

By surjectivity of $\mathrm{pr}_{k,l}$, we conclude

$$T_{N_k,N_l}P\lambda_{N_l,N_k} = P^l,$$

as one would expect from (5.2). To keep the notation simply, let us define $\tilde{F} = T_{N_k,N_l}$ and $F = \lambda_{N_l,N_k}$. Since (P, γ, A) is a proper MES, we have

$$\begin{aligned} P^l(s,t) - P^l(s,u) &= \tilde{F}P(s,t)F - \tilde{F}P(s,u)F \\ &= \tilde{F}\int_{(u,t]} \frac{d\gamma_s}{dA}(u)A(du)F = \int_{(u,t]} \tilde{F}\frac{d\gamma_s}{dA}(u)FA(du). \end{aligned}$$

Hence, the family γ^l defined by P^l through the already established finite variation property has densities $\frac{d\gamma^l_s}{dA} = \tilde{F}\frac{d\gamma_s}{dA}F$ which implies $\gamma^l \overset{\infty}{\ll} dA$ implying that (P^l, γ^l, A) is a proper MES as claimed. Since $l \leq k$ was arbitrary, this finishes the proof. $\square$

In the following, we call (P^l, γ_l, A) the projection of (P, γ, A) on Pol_l. We can now formulate our main result that gives sufficient conditions such that all projections of (P, γ, A) on Pol_l have an extended generator, allowing the construction of a version of the extended generator G of (P, γ, A) that preserves the degree on Pol_k. The next Theorem can therefore be seen as an improvement to Theorem 4.35.

Theorem 5.18. *Given a proper MES (P, γ, A), assume Condition 4.26 (ii) holds (or (i) which implies (ii)). Assume P preserves the degree of* Pol_k. *Then there exists an extended generator G that is unique dA-a.e. outside a countable subset of $\mathbb{R}_+$ and that preserves the degree on* Pol_k.

Proof. For $k = 0$, the statement is trivial since then there are no true subspaces of lower polynomial degree, hence in the sequel, let us assume $k > 0$. Since P preserves the degree, so does τ by Corollary 5.13. By Corollary 5.16, Condition 4.26 implies that we have for any $l \leq k$, that

$$\Big\{t : P^l(t-, t) \notin \mathrm{GL}_{N_l}, \Delta A_t = 0\Big\}$$

has dA measure zero for the proper MESs (P^l, γ^l, A) that exist by Proposition 5.17. Hence, Theorem 4.35 implies the existence of an extended generator G^l for each $l \leq k$. Define the operator $\mathcal{G}^l : \mathbb{R}_{>0} \times \mathrm{Pol}_l \to \mathrm{Pol}_l$, by $G = \iota_l(\mathcal{G})$, c.f. Definition 5.8, hence

$$(u, p) \mapsto \mathcal{G}^l_u p(\cdot) = \Big(G^l(u)\iota_l(p)\Big)^\top \mathbf{H}_{\beta^l}(\cdot).$$

Note that we use two notations, one where $\mathcal{G}^l$ is a map that takes arguments from $\mathbb{R}_{>0} \times \mathrm{Pol}_k$ and once where $\mathcal{G}^l$ is a family of operators $\mathcal{G}^l_u$ defined pointwise for each u by all $G^l(u)$ for each $l \leq k$. This should not lead to any confusion but will allow us to keep the notation simple while defining the same operators.

Let $p \in \mathrm{Pol}_k$. Define now for $l \leq k$, $\kappa_l : \mathrm{Pol}_k \to \mathrm{Pol}_l$ with $\kappa_0 p = \mathrm{pr}_{k,0}\, p$ and for $0 < l \leq k$, define

$$\kappa_l\, p = \mathrm{pr}_{k,l}\, p - \pi_{l-1,l} \circ \mathrm{pr}_{k,l-1}\, p.$$

Hence for any $p \in \mathrm{Pol}_l$, we have $\kappa_{\tilde{l}} \circ \pi_{l,k}\, p = 0$ for $\tilde{l} > l$. Next, define $\mathcal{G}^\# : \mathbb{R}_{>0} \times \mathrm{Pol}_k \to \mathrm{Pol}_k$ by

$$(u, p) \mapsto \mathcal{G}^\#_u p = \sum_{l=0}^{k} \pi_{l,k} \circ \mathcal{G}^l_u \circ \kappa_l\, p.$$

By construction, $\mathcal{G}^\#$ preserves the degree on Pol_k pointwise for all u. Further, $\mathcal{G}^\#$ is pointwise linear since it is the sum of pointwise linear maps on Pol_k. Hence, for each u it defines a matrix $G^\#(u)$. In other words, we have $G^\# = \iota_k(\mathcal{G}^\#)$ and by Corollary 5.13, $G^\#$ is degree preserving. We now show that $G^\#$ is a version of the extended generator of (P, γ, A). Let τ be defined by $\iota_k(\tau) = P$ and denote with τ^l and P^l the respective projections. Since we know that P, P^l are càdlàg , Equation (5.1) implies that the expression $\tau_{s,u-} f$ and $\tau^l_{s,u-} g$ are well defined for all $f \in \mathrm{Pol}_k$ and $g \in \mathrm{Pol}_l$. Further, note that since P preserves the degree by assumption, hence Corollary 5.13 implies that so does τ.

Note that we want to show

$$P(s, t) = \mathrm{Id}_{\mathrm{N_k}} + \int_{(s,t]} P(s, u-) G^\#(u) A(du),$$

which is equivalent to

$$\begin{aligned} P(s,t)w &= w + \int_{(s,t]} P(s,u-)G^{\#}(u)A(du)w \\ &= w + \int_{(s,t]} P(s,u-)G^{\#}(u)wA(du), \quad \forall w \in \mathbb{R}^{N_k}. \end{aligned} \tag{5.3}$$

Since we have equipped the finite dimensional space Pol_k with a norm under which it is complete, the integral is well defined in a Bochner sense. Further, since ι_k preserves the metric between Pol_k and $\mathbb{R}^{N_k}$, we get for all pointwise linear (and measurable) $c : \mathbb{R}_+ \times \mathrm{Pol}_k \to \mathrm{Pol}_k$, with $\iota_k(c_t) = C(t)$,

$$\int_\gamma c_\xi f A(d\xi) = \iota_k^{-1} \circ \int_\gamma C(\xi)\iota_k(f)A(d\xi), \quad \gamma \in \mathcal{B}(\mathbb{R}_+).$$

Hence, (5.3) is equivalent to showing

$$\tau_{s,t} f = f + \int_{(s,t]} \tau_{s,u-} \circ \mathcal{G}_u^{\#} f A(du), \quad \forall f \in \mathrm{Pol}_k.$$

As a first step, note that since τ is degree preserving, and càdlàg , the operator $\tau_{s,u-} : \mathrm{Pol}_k \to \mathrm{Pol}_k$, given by

$$\tau_{s,u-} f = \lim_{\varepsilon \downarrow 0} \tau_{s,u-\varepsilon} f$$

is also degree preserving. This is due to the fact that for arbitrary sequence $u_n \uparrow u$, the degree preserving property of τ implies $\tau_{s,u_n} f \in \mathrm{Pol}_l$ for all $f \in \mathrm{Pol}_l$. Since $\mathrm{Pol}_l \subset \mathrm{Pol}_k$ is a closed subset, this implies that the limit is in Pol_l too. We continue by computing

$$f + \int_{(s,t]} \tau_{s,u-} \circ \mathcal{G}_u^{\#} f A(du) = f + \sum_{l=0}^{k} \int_{(s,t]} \tau_{s,u-} \circ \pi_{l,k} \circ \mathcal{G}_u^l \circ \kappa_l f A(du). \tag{5.4}$$

Since $\mathcal{G}_u^l \circ \kappa_l f \in \mathrm{Pol}_l$, the degree preserving property of τ implies $\tau_{s,u-} \circ \pi_{l,k} = \pi_{l,k} \circ \tau_{s,u-}^l$ above yielding that (5.4) equals

$$\begin{aligned} & f + \sum_{l=0}^{k} \pi_{l,k} \circ \int_{(s,t]} \tau_{s,u-}^l \circ \mathcal{G}_u^l \circ \kappa_l f A(du) \\ &= f + \sum_{l=0}^{k} \pi_{l,k} \circ \iota_l^{-1} \circ \left(\int_{(s,t]} P^l(s,u-)G^l(u)A(du)\iota_l(\kappa_l f) \right) \\ &= f + \sum_{l=0}^{k} \pi_{l,k} \circ \iota_l^{-1} \circ \left(P^l(s,t)\iota_l(\kappa_l f) - \mathrm{Id}_{\mathrm{N_l}} \, \iota_l(\kappa_l f) \right) \\ &= f + \sum_{l=0}^{k} \pi_{l,k} \circ \iota_l^{-1} \circ \left(P^l(s,t)\iota_l(\kappa_l f) \right) - \sum_{l=0}^{k} \pi_{l,k} \circ \iota_l^{-1} \circ (\mathrm{Id}_{\mathrm{N_l}} \, \iota_l(\kappa_l f)) \\ &= f + \sum_{l=0}^{k} \pi_{l,k} \circ \iota_l^{-1} \circ \left(P^l(s,t)\iota_l(\kappa_l f) \right) - \sum_{l=0}^{k} \pi_{l,k} \circ \kappa_l f \end{aligned} \tag{5.5}$$

Noting that for all $f \in \mathrm{Pol}_k$, the definition of the κ_l implies $\sum_{l=1}^k \pi_{l,k} \circ \kappa_l f = f$, we get that (5.5) equals

$$\sum_{l=0}^{k} \pi_{l,k} \circ \iota_l^{-1} \circ \left(P^l(s,t)\iota_l(\kappa_l f)\right) = \sum_{l=0}^{k} \pi_{l,k} \circ \tau_{s,t}^l \circ \kappa_l f,$$

where we use that $\tau_{s,t}^l$ is represented by $P^l(s,t)$ through ι_l. Using once again that τ is degree preserving, we get

$$\sum_{l=0}^{k} \pi_{l,k} \circ \tau_{s,t}^l \circ \kappa_l f = \sum_{l=0}^{k} \tau_{s,t} \circ \pi_{l,k} \circ \kappa_l f = \tau_{s,t} \circ \sum_{l=0}^{k} \pi_{l,k} \circ \kappa_l f = \tau_{s,t} f.$$

Because τ is represented by P, this finally yields

$$P(s,t) = \mathrm{Id}_{\mathrm{N_k}} + \int_{(s,t]} P(s,u-)G^{\#}(u)A(du),$$

hence $G^{\#}$ is a version of the extended generator of (P, γ, A). As this generator is unique dA-a.e. outside a countable set by Theorem 4.35, we get $G^{\#}(u) = G^k(u)$ dA-a.e. outside a countable set. This finishes the proof. □

Let us now consider the reverse direction, i.e. if $G : \mathbb{R}_{>0} \to \mathrm{M}_{N_k}(\mathbb{R})$ is degree preserving on Pol_k, is then the MES defined through the fundamental solution of (MDE:A, V) also degree preserving? Let us make the convention that for a matrix function G indexed by some set i, AG is defined pointwise by $A(i)G(i), i \in I$, where in case A is a constant matrix, it is interpreted as a constant function over the index set of G. We have seen before that if a MES P was degree preserving, then so were it's projections P^l, which are representing matrices of the projections τ^l wrt ι_l. Further, recall that we have

$$\tau^l = \mathrm{pr}_{k,l} \circ \tau \circ \pi_{l,k},$$

which in terms of the representations is equivalent to

$$P^l = T_{N_k,N_l} P \lambda_{N_l,N_k}.$$

By definition, τ preserves the degree on Pol_k, if for all $l \leq k$,

$$\tau \circ \pi_{l,k} = \pi_{l,k} \circ \tau^l.$$

Since $\mathrm{pr}_{k,l}$ is surjective on Pol_l, this gives the equivalent characterization

$$\tau \circ \pi_{l,k} \circ \mathrm{pr}_{k,l} = \pi_{l,k} \circ \tau^l \circ \mathrm{pr}_{k,l},$$

hence by definition of τ^l, for all $l \leq k$

$$\tau \circ \pi_{l,k} \circ \mathrm{pr}_{k,l} = \pi_{l,k} \circ \mathrm{pr}_{k,l} \circ \tau \circ \pi_{l,k} \circ \mathrm{pr}_{k,l}.$$

By Corollary 5.13, we know that τ is degree preserving if and only if P is degree preserving. Hence, P is degree preserving if and only if

$$P\lambda_{N_l,N_k}T_{N_k,N_l} = \lambda_{N_l,N_k}T_{N_k,N_l}P\lambda_{N_l,N_k}T_{N_k,N_l}. \tag{5.6}$$

Alternatively, P is degree preserving by definition, provided for all $l \leq k$,

$$Pw \in E_l, \quad \forall w \in E_l.$$

Let $L_l \in \mathrm{M}_{N_k}(\mathbb{R})$ be the diagonal matrix with elements $(L_l)_{i,j} = \mathbb{1}_{\{i=j\}}\mathbb{1}_{\{i\leq N_l\}}$. Given that E_l is the subspace of $\mathbb{R}^{N_k}$ that contains precisely those elements that are identified under ι_k with polynomials of at most degree l, a closer look shows $\lambda_{N_l,N_k}T_{N_k,N_l} = L_l$ and since $L_l w = w$ for all $w \in E$, we get P is degree preserving, if and only if for all $l \leq k$,

$$L_l P L_l w = P L_l w, \quad \forall w \in \mathbb{R}^{N_k} \Leftrightarrow L_l P L_l = P L_l$$

which is the same as (5.6). Denote further with L_l^c the diagonal matrix with $L_l^c = \mathrm{Id}_{\mathrm{N_k}} - L_l$. Note that by surjectivity of L_l on E_l and $\mathrm{im} L_l = E_l$, we have for a matrix $M \in \mathrm{M}_{N_k}(\mathbb{R})$,

$$L_l M = 0 \Leftrightarrow f^\top M = 0, \quad \forall f \in E_l. \tag{5.7}$$

Let us now proof the following theorem.

Theorem 5.19. *Let $G : \mathbb{R}_{>0} \to M_{N_k}(\mathbb{R})$ and $A \in \mathcal{V}^+_{\mathfrak{A}}$. Consider* (MDE:$A, V$) *for $V(t,x) = G(t)^\top x$ and assume $\|G\|_{dA,t} < \infty$ for all $t \geq 0$. Then there exists a proper MES (P, γ, A) for which G is an extended generator. If G preserves the degree on* Pol_k*, then so does P.*

Proof. The first statement, that is the existence of a proper MES (P, γ, A) is the statement of Proposition 4.53. Let M be the fundamental solution, i.e. $M = P^\top$ and recall that we have uniqueness of solutions to (MDE:A, V). We first show

$$f^\top M(s,t) L_l = f^\top L_l M(s,t), \quad \forall f \in E_l. \tag{5.8}$$

Let $w \in \mathbb{R}^{N_k}$ be arbitrary. Then, since M is the fundamental solution, we have

$$L_l M(s,t) w = L_l w + \int_{(s,t]} L_l G(u)^\top M(s,u-) w A(du).$$

Since L_l is a diagonal matrix we have $L_l = L_l^\top$. Further, since by assumption G is degree preserving, we have $L_l G L_l = G L_l$ which implies $L_l G(u)^\top = L_l G(u)^\top L_l$, yielding

$$L_l M(s,t) w = L_l w + L_l \int_{(s,t]} G(u)^\top L_l M(s,u-) w A(du).$$

In the same way, $M(s,t)L_l w$ is the unique solution with initial value $L_l w$, hence we have

$$M(s,t)L_l w = L_l w + \int_{(s,t]} G(u)^\top M(s,u-)L_l w A(du).$$

Define $g(s,t) = L_l M(s,t) - M(s,t)L_l$, hence since $M(s,s) = \mathrm{Id}_{\mathrm{N_k}}$, $g(s,s) = 0$. We compute for arbitrary $w \in \mathbb{R}^{N_k}$, using $L_l G(u)^\top L_l = L_l G(u)^\top$ by the degree preserving assumption,

$$\begin{aligned} g(s,t)w &= \int_{(s,t]} L_l G(u)^\top M(s,u-)wA(du) - \int_{(s,t]} G(u)^\top M(s,u-)L_l w A(du) \\ &= L_l \int_{(s,t]} G(u)^\top L_l M(s,u-)wA(du) - \int_{(s,t]} G(u)^\top M(s,u-)L_l w A(du) \\ &= \int_{(s,t]} G(u)^\top g(s,u-)wA(du) - L_l^c \int_{(s,t]} G(u)^\top L_l M(s,u-)wA(du) \end{aligned}$$

Since we have $L_l L_l^c = 0$ and using again $L_l G(u)^\top L_l = L_l G(u)^\top$, this gives us

$$\begin{aligned} L_l g(s,t)w &= L_l \int_{(s,t]} G(u)^\top g(s,u-)wA(du) \\ &= \int_{(s,t]} (G(u)L_l)^\top L_l g(s,u-)wA(du). \end{aligned}$$

Since $\|L_l\| < \infty$, we have $\|GL_l\|_{dA,t} \le \|G\|_{dA,t} \|L_l\| < \infty$ for all $t \ge 0$. This implies that $L_l g(s,t)w$ is the unique solution to (MDE:A,V) for $V(t,x) = (G(u)L_l)^\top x$ with initial value 0, hence $L_l g(s,t)w \equiv 0$ for all $w \in \mathbb{R}^{N_k}$, implying $L_l g(s,t) = 0$. This implies $f^\top g(s,t) = 0$ for all $f \in E_l$, which in turn implies (5.8). Hence we get

$$L_l P(s,t)f = P(s,t)f, \quad \forall f \in E_l,$$

implying P preserves the degree on Pol_k since $l \le k$ was arbitrary. □

Chapter 6

Polynomial Semimartingales

In this chapter we will introduce and discuss the core subject of this part of this thesis, ***polynomial processes*** and ***polynomial semimartingales***. As discussed in the introduction, it is our goal to combine the two approaches taken in [FL16] by Filipović and Larsson and [Cuc+12] by Cuchiero et al. In particular, we want to extend the class of polynomial processes, see Chapter 1, to include those processes that map polynomials of a certain degree to polynomials of at most the same degree while allowing for time inhomogeneity and stochastic discontinuity.

While Cuchiero et al. started with a Markov process and hence with a semigroup that allows for a straight forward definition of the polynomial property, Filipović and Larsson start by considering diffusion processes with coefficients satisfying a polynomial condition which is equivalent to assuming an invariance property of the extended generator of the diffusion. This then is shown to imply a polynomial property in the spirit of [Cuc+12] under natural integrability assumption. Hence the starting point of the discussions is a semimartingale setting. One key observations the authors make in [FL16] is that processes as they introduce them do not necessarily need to be Markov process, or to be precise they do not assume that it is always possible to embed their class of processes in a Markov process.

Our approach relates to the two above in the following way. We start our discussions by considering a càdlàg process and define the polynomial property in a natural way, hence under a fixed stochastic basis similar to [FL16]. This however will lead to technical challenges that would not exist if one would start with a semigroup $(P_t)_{t\geq 0}$ such that for any $x \in E$ with E being the state space and $f \in \mathrm{Pol}_k$ one has

$$P_t f(x) := \mathbb{E}_x \left[f(X_t) \right], \quad \forall x \in E,$$

which defines a function on all of E. In our situation however, we do not want to a priory fix a state space E since we neither consider a Markov-process, nor do we want to deal with time varying supports. Hence we consider processes taking values in all of $\mathbb{R}^d$ instead of a subset E. In that case however, we face the issue that there might be two $g, g' \in \mathrm{Pol}_k$ with $g \neq g'$ but "agree" on the support of the conditioning value, meaning one has the a.s. identity

$$\mathbb{E}\left[f(X_t) \mid X_s\right] = g(X_s) = g'(X_s).$$

In the setting of [Cuc+12], this can not happen since polynomials are considered over E, hence one would have the identification $g = g'$. The goal however is to construct a Markov-semigroup-like structure defined on polynomials that allows for similar discussions as was done in [Cuc+12]. It turns out that the concept needed for these discussions are the matrix evolution systems introduced in Chapter 4. As we will see, the regularity assumptions we made for MESs there are well suited to study our class of polynomial processes allowing for stochastic discontinuity.

We will also see that our processes are semimartingales on our fixed stochastic basis similar to [FL16] and unlike [Cuc+12], where semimartingality is defined for each probability measure $P_x, x \in E$ and hence with $\mathcal{L}^{\mathrm{law}}(X_0) = X_{0\#}P = \delta_x$. Hence, we will introduce our class of processes as polynomial processes, but in light of the semimartingality result refer to them finally as polynomial semimartingales.

Recall that with $\|\cdot\|_{\mathbb{R}^d}$ we denote any norm on $\mathbb{R}^d$ which is reasonable as by equivalence of all norms in $\mathbb{R}^d$, the statements in this chapter will be independent of the concrete choice unless otherwise stated. We will sometimes pick a concrete norm since it will simplify notation without losing us generality. For that, recall that the p-norms on $\mathbb{R}^m$ for $p \geq 1$ are defined by

$$\|x\|_p^p = \sum_{i=1}^{m} |x_i|^p.$$

Also recall, that by the conventions we made in the preliminaries, we define the polynomials $f_i, i \in \{1, \ldots, d\}$ on $\mathbb{R}^d$ as the projections to the i-th coordinate, i.e. for $x = (x_1, \ldots, x_d)^\top \in \mathbb{R}^d$ we have $f_i(x) = x_i$.

We will further use the notation introduced in the previous chapter. Further, Convention 5.1 holds to simplify notation. Note that this is w.l.o.g. since the following discussions hold for arbitrary choice of a basis of the spaces $\mathrm{Pol}_k, k \in \mathbb{N}$ unless explicitly indicated otherwise.

In the sequel, whenever we consider a stochastic basis $(\Omega, \mathscr{F}, P)$, we assume that the usual conditions hold regardless of whether we explicitly say so or not. Further, let us fix the notation $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$.

6.1 Setting and the polynomial property

Consider the stochastic basis $(\Omega, \mathscr{F}, \mathbb{F}, P)$ satisfying the usual conditions and let X be an $\mathbb{R}^d$-valued adapted càdlàg process on this basis. Recall that with Pol_k we always mean $\mathrm{Pol}_k(\mathbb{R}^d)$. Recall that for each $k \in \mathbb{N}_0$, we defined $N_k := \dim \mathrm{Pol}_k$. If there is no ambiguity, we will write N for N_k, hence omit the degree k.

We start with a lemma regarding integrability assumptions we will make later on.

Lemma 6.1. *Given a random variable $Y \in \mathbb{R}^d$ and a positive integer $k \in \mathbb{N}$, the following integrability conditions are equivalent:*

(i) $\mathbb{E}\left[|f(Y)|\right] < \infty$ *for all* $f \in \mathrm{Pol}_k$

(ii) $\mathbb{E}\left[\|Y\|_{\mathbb{R}^d}^k\right] < \infty$ *for all choices of* $\|\cdot\|_{\mathbb{R}^d}$.

Proof. $(i) \Rightarrow (ii)$: First note that in (ii), by the equivalence of all p-norms on $\mathbb{R}^d$, it is always sufficient to to show the integrability condition for $p = 1$. Recall that from Jensen's inequality for sums, we have for $a_i \in \mathbb{R}$, $i = 1, \dots, l$, $l \in \mathbb{N}$,

$$\left(\sum_{i=1}^{l} a_i\right)^k \leq \left(l \sum_{i=1}^{l} \frac{1}{l} a_i\right)^k = l^k \sum_{i=1}^{l} \frac{1}{l}(a_i)^k = l^{k-1} \sum_{i=1}^{l}(a_i)^k. \tag{6.1}$$

Denote by $f_i^k(x) = x_i^k$. Hence $f_i^k \in \mathrm{Pol}_k$ for all $i = 1, \dots, d$. We have

$$\mathbb{E}\left[\|X\|_1^k\right] = \mathbb{E}\left[\left(\sum_{i=1}^{d} |X_i|\right)^k\right] \overset{(6.1)}{\leq} d^{k-1} \sum_{i=1}^{d} \mathbb{E}\left[|X_i|^k\right]. \tag{6.2}$$

Since $|X_i|^k = \left|f_i^k(X)\right|$, we get (ii) for $p = 1$ and hence for all p. We now show the other direction.

$(ii) \Rightarrow (i)$: First note that we have $|X_i| \leq \|X\|_1$. Further, for any $f \in \mathrm{Pol}_k$, there exists a constant C depending on f such that

$$|f(X)| \leq C(1 + \sum_{i=1}^{d} |X_i|^k) \leq C(1 + d \cdot \|X\|_1^k). \tag{6.3}$$

Hence (i) follows from (ii) for $p = 1$ and we are done. □

A direct consequence of the previous lemma is the following lemma.

Lemma 6.2. *Let X be a stochastic process and $k \in \mathbb{N}$ a positive integer. Then the following two integrability conditions are equivalent:*

(i) for all $f \in \mathrm{Pol}_k$ and $t > 0$ we have

$$\sup_{s \leq t} \mathbb{E}\left[|f(X_s)|\right] < \infty$$

(ii) for all $p \geq 1$ *we have*

$$\sup_{s \leq t} \mathbb{E}\left[\|X_s\|_p^k\right] < \infty$$

Proof. $(i) \Rightarrow (ii)$: We have seen from the proof of Lemma 6.1 that

$$\sup_{s \leq t} \mathbb{E}\left[\|X_s\|_1^k\right] \leq d^{k-1} \sum_{i=1}^{d} \sup_{s \leq t} \mathbb{E}\left[\left|f_i^k(X_s)\right|\right].$$

Since by assumption each summand is finite and we sum over finitely many values, the first direction is proven.

$(ii) \Rightarrow (i)$: Again from the proof of Lemma 6.1, we have

$$\sup_{s \leq t} \mathbb{E}\left[|f(X_s)|\right] \leq \sup_{s \leq t} \mathbb{E}\left[C(1 + d\,\|X_s\|_1^k\right] = C + d \sup_{s \leq t} \mathbb{E}\left[\|X_s\|_1^k\right],$$

which finishes the proof. □

As motivation for Definition 6.4 further down, let us state the following well known result under the assumption of square integrability, see [JP12, Theorem 23.3].

Theorem 6.3. *Let* $(\Omega, \mathscr{F}, P)$ *be a probability space with sub* σ*-algebra* $\mathcal{A} \subset \mathscr{F}$. *Let* $Y \in L^2(\Omega, \mathscr{F}, P)$ *with the usual* P*-a.s. identification. If for some random variable* Z *we have* $\sigma(Z) = \mathcal{A}$, *then there exists a Borel measurable function* f *such that*

$$\mathbb{E}\left[Y \mid \mathcal{A}\right] = f(Z) \quad [P].$$

The idea is now the following. Given a càdlàg process X, set $Y = \varphi(X_t)$ and $Z = X_s$ for $s \leq t$, where $\varphi : \mathbb{R}^d \to \mathbb{R}$ and require that if φ belongs to a certain class of functions, then so does f. In the context of polynomial processes, this class of functions shall be the space of multivariate polynomials up to a given degree.

Definition 6.4. *We say the process* X *has the* ***k*-polynomial** *property, if for all* $l \in \{0, \ldots, k\}$, $f \in \mathrm{Pol}_l$ *we have* $\mathbb{E}\left[\|X_t\|_1^k\right] < \infty\ \forall\, t \geq 0$ *and for any* $0 \leq s \leq t$ *there exists a* $q_{s,t}^f \in \mathrm{Pol}_l$ *such that we have*

$$\mathbb{E}\left[f(X_t) \mid \mathcal{F}_s\right] = q_{s,t}^f(X_s) \quad [P]. \tag{6.4}$$

Finally, if X *is has the* k*-polynomial property for all* $k \in \mathbb{N}$, *we say it has the polynomial property.*

Remark 6.5. *(i) Since $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ is a filtration, we have due to Equation* (6.4) *and the tower property,*

$$\mathbb{E}\left[f(X_t) \mid X_s\right] = \mathbb{E}\left[\mathbb{E}\left[f(X_t) \mid \mathcal{F}_s\right] \mid X_s\right] = \mathbb{E}\left[q^f_{s,t}(X_s) \mid X_s\right] = q^f_{s,t}(X_s). \tag{6.5}$$

Hence, a process with the k-polynomial property satisfies the Markov property on the finite dimensional space Pol_k rather on the space of bounded measurable functions, compare e.g. [EK09, Chapter 4.1].

(ii) Note that in Definition 6.4, property (6.4) *holds for all degrees $l \leq k$, hence the polynomial $q^f_{s,t}$ has at most the same degree as f. This is similar to Cuchiero et al. [Cuc+12] where the same condition is required and the reason given there, c.f. Remark 2.2 (ii). For us too, this degree preserving property is needed for Theorem 6.44 and Theorem 6.57 that we too consider the most important from an application point of view. The reader will notice that this property corresponds to degree preserving properties of maps $\tau : \mathbb{T} \times \mathrm{Pol}_k \to \mathrm{Pol}_k$ and indeed we will see that this precisely why we make the definition in this way.*

Note that in Definition 6.4, we do not require $q^f_{s,t}$ to be unique. This has several consequences which we want to discuss now.

Remark 6.6. *First, note that we consider polynomials over the whole space $\mathbb{R}^d$, so while two polynomials f, f' might satisfy $f(X_s) = f'(X_s)$ a.s., they do not need to be the same. Just consider the following example where Z is standard normal and $X := (Z, Z)$. Hence $d = 2$. However, the support of X is a submanifold of $\mathbb{R}^2$, and for $f(x, y) = x$ and $f'(x, y) = y$ we have $f(X) = f'(X)$ a.s. while $f \neq f'$. This implies that in general, the map $f \mapsto q^f_{s,t}$ above is not well defined as a map from Pol_l to Pol_l.*

Let $\tau : \mathbb{T} \times \mathrm{Pol}_k \to \mathrm{Pol}_k$ with $(s, t, f) \mapsto \tau_{s,t} f$ such that for any $(s, t) \in \mathbb{T}$ and $l = 0, \ldots, k$, $\tau_{s,t}(\mathrm{Pol}_l) \subset \mathrm{Pol}_l$ and

$$\mathbb{E}\left[f(X_t) \mid \mathcal{F}_s\right] = (\tau_{s,t} f)(X_s) \quad [P]. \tag{6.6}$$

Note that by Remark 6.6, there might be more than one such function τ. Further, at least one such function always exists by defining it pointwise for any $(s, t) \in \mathbb{T}$ through the axiom of choice. A construction without invoking the axiom of choice by finite dimensionality of Pol_k is actually possible by the next remark.

Remark 6.7. *One might think that $\forall (s, t) \in \mathbb{T}$, the map $\tau_{s,t}$ is linear* **(or $\boldsymbol{\tau}$ is pointwise linear)** *due to the linearity of the conditional expectation. However, by Remark 6.6, $\tau_{s,t}$ is not unique. This is the reason why a priory we do not have linearity. In Example 6.11 below, we provide an example in which $\tau_{s,t}$ for fixed $(s, t) \in \mathbb{T}$ is not linear. However, by fixing an arbitrary basis β of Pol_k, one can always construct a pointwise linear $\tilde{\tau}$ from τ in the following way. For $f \in \mathrm{Pol}_k$, let $\vec{\alpha} \in \mathbb{R}^{N_k}$ s.t.*

$$f(x) = \sum_{i=1}^{N_k} \alpha_i \beta_i(x) \quad \forall x \in \mathbb{R}^d. \tag{6.7}$$

Then, we define $\tilde{\tau}_{s,t} : \mathrm{Pol}_k \to \mathrm{Pol}_k$ *by*

$$\tilde{\tau}_{s,t} f = \sum_{i=1}^{N_k} \alpha_i \tau_{s,t}(\beta_i). \tag{6.8}$$

Then, the linearity of the conditional expectation operator implies

$$\tilde{\tau}_{s,t} f(X_s) = \tau_{s,t} f(X_s) \quad [P].$$

Further, $\tilde{\tau}$ *is linear by construction, which finishes the overall construction. Note that the construction of* $\tilde{\tau}$ *in* (6.8) *only required* τ *to be defined on a finite set, namely the basis. Hence,* $\tilde{\tau}$ *can be constructed without an explicit invoking of the axiom of choice.*

Remark 6.8. *Let us note that the tower property of the conditional expectation would indicate that a map* τ *would satisfy an evolution property, i.e.*

$$\tau_{s,u} \circ \tau_{u,t} f = \tau_{s,t} f. \tag{6.9}$$

Again, this is in general not true, even if τ *is pointwise linear. We provide a counterexample with Example 6.11. Even the condition* $\tau_{s,s} = \mathrm{id}$ *would not need to be satisfied in light of Remark 6.6.*

Remark 6.9. *As one expects, the above points regarding linearity and the evolution property are rendered mute if one assumes* τ *to be unique. In that case, linearity and the evolution property follow by Proposition 6.13. Unlike the case for linearity however, we can not always assume existence of a* τ *satisfying Equation* (6.9) *in the general case. Let us make the following observation. If we were to consider a discrete time model, i.e* $\{(n,m) \in \mathbb{N}_0^2 : n \leq m\}$ *rather then* $\mathbb{T}$*, we can construct a* τ *that is pointwise linear and satisfies Equation* (6.9) *by first constructing a pointwise linear version, again denoted by* τ *and then defining for each* $n \in \mathbb{N}_0$

$$\tilde{\tau}_{n,n} := \mathrm{id}_{\mathrm{Pol}_k} \;\; \textit{and} \; \tilde{\tau}_{n,n+1} := \tau_{n,n+1}.$$

Finally, we extend $\tilde{\tau}$ *to the whole parameter set by defining*

$$\tilde{\tau}_{n,n+m} := \tilde{\tau}_{n,n+1} \circ \ldots \circ \tilde{\tau}_{n+m-1,n+m}.$$

Then, by construction, $\tilde{\tau}$ *satisfies Equation* (6.9) *and is pointwise linear. This construction would not work in the continuous parameter case. There however, provided an infinitesimal description of the evolution "*$\tau_{s,t} \to \tau_{s,t+\varepsilon}$*" is available in form of what we will introduce as an extended generator, it is possible to show existence of such a* τ*, compare Theorem 6.41.*

The following proposition is just a summary of the previous remarks regarding existence of a linear map τ.

Proposition 6.10. *Let X have the k-polynomial property. Then there exists a map $\tau : \mathbb{T} \times \mathrm{Pol}_k \to \mathrm{Pol}_k$ such that for each $(s,t) \in \mathbb{T}$, Equation* (6.6) *is satisfied and $\tau_{s,t}$ is linear on Pol_k.*

Let us now give the example we promised in the remarks above regarding nonlinearity and violation of the evolution property.

Example 6.11. *Let X be an $\mathbb{R}$-valued Markov process with $X_t \in \{-1,1\} \forall t \geq 0$ defined by $P(X_0 = -1) = h = 1 - P(X_0 = 1)$ for some $h \in [0,1]$ and $X_t = X_{\lfloor t \rfloor}$ such that we only need to specify X_n for $n \in \mathbb{N}$ and $\tau_{n,m}$ for $n, m \in \mathbb{N}$ with $n \leq m$. For that, we define*

$$P(X_{n+1} = -1 \mid X_n) = \begin{cases} p \in [0,1], & \text{if } X_n = -1 \\ \tilde{p} \in [0,1], & \text{if } X_n = 1. \end{cases}$$

For $n \in \mathbb{N}_0$, we define $\tau_{n,n} f_{(0)} = f_{(0)}$ which satisfies Equation (6.6). *Further, we have*

$$\mathbb{E}\left[f_{(1)}(X_{n+1}) \mid X_n\right] = \begin{cases} 1 - 2p, & \text{if } X_n = -1 \\ 1 - 2\tilde{p}, & \text{if } X_n = 1. \end{cases}$$

Hence τ must satisfy

$$\begin{aligned} \tau_{n-1,n} f_{(1)}(-1) &= 1 - 2p \\ \tau_{n-1,n} f_{(1)}(1) &= 1 - 2\tilde{p}, \end{aligned} \tag{6.10}$$

and it is clear that there is always exactly one polynomial $g_1 \in \mathrm{Pol}_1$ that satisfies (6.10) *so we define $\tau_{n-1,n} f_{(1)} = g_1$. Next, note that*

$$\mathbb{E}\left[f_{(2)}(X_{n+1}) \mid X_n\right] = \begin{cases} 1, & \text{if } X_n = -1 \\ 1, & \text{if } X_n = 1. \end{cases}$$

Hence τ must satisfy

$$\begin{aligned} \tau_{n-1,n} f_{(2)}(-1) &= 1 \\ \tau_{n-1,n} f_{(2)}(1) &= 1, \end{aligned} \tag{6.11}$$

and there are infinitely many polynomials $g_2 \in \mathrm{Pol}_2$ such that (6.11) *is satisfied. Let us set $p = \frac{1}{2}$ and $\tilde{p} = \frac{3}{4}$. Then $g_1(x) = -\frac{x}{4} - \frac{1}{4}$ and one possible choice for g_2 is $g_2(x) = f_{(2)}(x) = x^2$. Next, consider the polynomial $f = f_{(2)} - f_{(1)}$. Again, we have*

$$\mathbb{E}\left[f(X_{n+1}) \mid X_n\right] = \begin{cases} 2p = 1, & \text{if } X_n = -1 \\ 2\tilde{p} = \frac{3}{2}, & \text{if } X_n = 1. \end{cases}$$

Hence τ must satisfy

$$\begin{aligned} \tau_{n-1,n} f(-1) &= 1 \\ \tau_{n-1,n} f(1) &= \frac{3}{2}. \end{aligned} \tag{6.12}$$

Hence, for $g(x) := \frac{5}{4} + \frac{x}{4}$, we can set $\tau_{n-1,n} f = g$ and (6.12) is satisfied. Another choice could have been $g_2 - g_1$, compare Equation (6.8), but our choice is valid as it maps to a polynomial that satisfies Equation (6.4). But since $g \neq g_2 - g_1$, we see that τ is not a pointwise linear map as linearity would require $g = g_2 - g_1$.

Example 6.12. *Next we show that the evolution property can be violated. So far, we have defined τ only for the indices $(n-1, n)$. Let us change our example above to $\tau_{n-1,n} f = \tau_{n-1,n} f_{(2)} - \tau_{n-1,n} f_{(1)}$. In fact, since $(f_{(0)}, f_{(1)}, f_{(2)})$ forms a basis of Pol_2, let us extend $\tau_{n-1,n}$ to all of Pol_2 as we did in Equations (6.7) and (6.8). Hence, now for each n, $\tau_{n,n}$ and $\tau_{n-1,n}$ are linear maps. Note that we only need to extend τ to the parameters $n < m$ with $m - n > 1$ in order to have τ fully specified. This could be done as proposed in Remark 6.9 and by construction, the evolution property would be satisfied. This construction however is not the only choice possible as we show in the following. Consider $m = n + 2$. We have $\tau_{n,n+1} f_{(2)} = f_{(2)}$. This fixed point property combined with the tower property yields*

$$\begin{aligned}\mathbb{E}\left[f_{(2)}(X_{n+2}) \mid X_n\right] &= \mathbb{E}\left[\tau_{n+1,n+2} f_{(2)}(X_{n+1}) \mid X_n\right] \\ &= \mathbb{E}\left[f_{(2)}(X_{n+1}) \mid X_n\right] = \mathbb{E}\left[1 \mid X_n\right] = 1 \quad [P],\end{aligned}$$

hence we can set $\tau_{n,n+2} f_{(2)} = f_{(0)}$ together with $\tau_{n,n+2} f_{(i)} = \tau_{n,n+1} \circ \tau_{n+1,n+2} f_{(i)}$ for $i = 0, 1$ and extend to all of Pol_k as we did in Equations (6.7) and (6.8). But then

$$f_{(0)} = \tau_{n,n+2} f_{(2)} \neq \tau_{n,n+1} \circ \tau_{n+1,n+2} f_{(2)} = f_{(2)},$$

so the evolution property (6.9) does not hold, which finishes our example.

In the example above, we have used the non-uniqueness in Equation (6.4) to construct nonlinearity and a violation of the evolution property. The following proposition shows that in case of uniqueness, both properties would be implied.

Proposition 6.13. *Let X have the k-polynomial property and assume for all $0 \leq t$ and $f \in \mathrm{Pol}_k$, the polynomial $q^f_{s,t} \in \mathrm{Pol}_k$ for which Equation (6.4) is satisfied is unique. Then, the mapping τ is uniquely defined, pointwise linear and satisfies the evolution property together with $\tau_{s,s} = \mathrm{id}_{\mathrm{Pol}_k}$ for all $s \geq 0$. Further, for all $(s,t) \in \mathbb{T}$ we have $\tau_{s,t}(\mathrm{Pol}_l) \subset \mathrm{Pol}_l$ for all $l \in \{0, \ldots, k\}$ and $\tau_{s,t}$ is bijective for all $(s,t) \in \mathbb{T}$.*

Proof. Uniqueness of τ is implied by the uniqueness of $q^f_{s,t}$. Note that by the definition of the k-polynomial property we can deduce $\deg(q^f_{s,t}) \leq \deg(f)$. Further, uniqueness implies that for $t \geq 0$ and $p, q \in \mathrm{Pol}_k$, $p(X_t) = q(X_t)$ P-a.s. implies $p = q$. Hence linearity follows from linearity of conditional expectation. The evolution property now comes from the tower property, again using the implication $p = q$ if $p(X_t) = q(X_t)$ P-a.s., yielding

$$\begin{aligned}\tau_{s,u} \circ \tau_{u,t} f(X_s) &= \mathbb{E}\left[\tau_{u,t} f(X_u) \mid \mathcal{F}_s\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[f(X_t) \mid \mathcal{F}_u\right] \mid \mathcal{F}_s\right] = \mathbb{E}\left[f(X_t) \mid \mathcal{F}_s\right] = \tau_{s,t} f(X_s) \quad [P],\end{aligned}$$

and hence $\tau_{s,u} \circ \tau_{u,t} = \tau_{s,t}$. The property that τ is pointwise bijective is just a rephrasing of existence and uniqueness of $q^f_{s,t}$ for all $f \in \mathrm{Pol}_k$. Finally, note that $\tau_{s,s} f(X_s) = f(X_s)$ P-a.s. which implies $\tau_{s,s} = \mathrm{id}_{\mathrm{Pol}_k}$. □

In the previous chapter we have called mappings τ evolution systems if the corresponding representing matrices form a MES. It is clear that this definition can be formulated without invoking representing matrices. Hence we introduce the next equivalent definition. Additionally, we define what it means if a mapping τ belongs to a process X.

Definition 6.14. *We call a mapping τ, that is pointwise linear, satisfies the evolution property with $\tau_{s,s} = \mathrm{id}_{\mathrm{Pol}_k}$ $\forall s \geq 0$ an* **evolution system** *on* Pol_k*. If in addition, it maps for fixed $(s,t) \in \mathbb{T}$ a polynomial $f \in \mathrm{Pol}_k$ to a polynomial of at most the same degree (see Definition 6.4) such that $\tau_{s,t} f$ satisfies* (6.6)*, we call it an* **evolution system of X on** Pol_k*.*

Remark 6.15. *Note that by the definition above, if τ is an evolution system of a k-polynomial process X on* Pol_k*, then it is automatically degree preserving on* Pol_k*.*

From Proposition 6.13 we know that a sufficient condition for τ to be linear and satisfy the evolution property is uniqueness of the polynomials $q^f_{s,t}$ in Equation (6.4). The next condition will be sufficient for such a uniqueness.

Condition 6.16. *We say that X satisfies the* ***full support condition****, if there is a $D \subset \mathbb{R}^d$ such that it holds*

$$D \subset \bigcap_{s \geq 0} supp(X_s) \tag{6.13}$$

with $\mathring{D} \neq \emptyset$.

As noted in Remark 6.5 (i), k-polynomial processes satisfy the Markov-property on Pol_k instead of $b\mathcal{E}(\mathbb{R}^d)$ as one often requires (compare e.g. Blumenthal and Getoor [BG07, Definition 1.1 and Theorem 1.3 (iii)]). Indeed it does not follow that equation (6.4) implies that X can be embedded in a Markov process. Therefore we can not assume the existence of an infinite dimensional evolution system on $b\mathcal{E}(\mathbb{R}^d)$, compare Böttcher [Böt14, Section 2], as is usually the starting point when studying Markov processes.

The next lemma shows that if Condition 6.16 holds, then X has a unique evolution system on Pol_k for all $k \in \mathbb{N}_0$. In particular, by the definition of the polynomial property, the evolution system will be degree preserving.

Lemma 6.17. *Let X satisfy the polynomial property and assume that the full support condition is satisfied. Then there is a mapping $\tau : \mathbb{T} \times \mathrm{Pol} \to \mathrm{Pol}$ such that $\tau|_{\mathrm{Pol}_k}$ is the unique evolution system for X on* Pol_k *for all $k \in \mathbb{N}$.*

Proof. By Proposition 6.13, it is sufficient to show that for all all $s \geq 0$, the implication

$$p(X_s) = q(X_s) \quad [P] \text{ for } p, q \in \mathrm{Pol} \Rightarrow p = q,$$

holds. Hence assume $p(X_s) = q(X_s)$ for fixed $s \geq 0$. Note that it suffices to show that $p = q$ on a dense subset of an open ball in $\mathbb{R}^d$ since that uniquely determines a polynomial on $\mathbb{R}^d$.

Since $p(X_s) = q(X_s)$, there exists a set $\mathcal{J} \subset \Omega$, such that $P(\mathcal{J}) = 1$ and

$$p(X_s(\omega)) = q(X_s(\omega)) \quad \text{ for all } \ \omega \in \mathcal{J}. \tag{6.14}$$

Set $\mathcal{I} = X_s(\mathcal{J})$ so that we have $P(X_s \in \mathcal{I}) = 1$. Hence, since we have

$$D \subset \mathrm{supp}(X_s) = \bigcap_C \left\{ C : C \subset \mathbb{R}^d \text{ closed }, P(X_s \in C) = 1 \right\}, \tag{6.15}$$

we get that $D \subset \overline{\mathcal{I}}$. Note that (6.14) is equivalent to

$$p(x) = q(x) \quad \forall x \in \mathcal{I}. \tag{6.16}$$

By the full support condition, there exists an non-empty open ball $B \subset \mathbb{R}^d$ with $B \subset \mathring{D} \subset \overline{\mathcal{I}}$. As $\mathcal{I}$ is dense in $\overline{\mathcal{I}}$, we have that $\mathcal{I} \cap B$ is dense in B. Combined with (6.16), this finishes the proof for uniqueness of the polynomial in (6.4). Hence, by Proposition 6.13 for each $k \in \mathbb{N}$ there is a unique τ^k which is an evolution system on Pol_k for X and for all $(s,t) \in \mathbb{T}$, $m \in \mathbb{N}_0$,

$$\tau^{k+m}_{s,t}|_{\mathrm{Pol}_k} = \tau^k_{s,t},$$

defining a mapping $\tau : \mathbb{T} \times \mathrm{Pol} \to \mathrm{Pol}$ with the desired properties which finishes the proof of this Lemma. □

Remark 6.18. *The full support condition at $s = 0$ might appear rather restrictive. Just think of the prime example of a process with the polynomial property in $d = 1$, namely Brownian motion (c.f. C. A. Hurtado [CAH17] or Filipović and Larsson [FL16]). Indeed, if one considers that Brownian motion B starting at some $x \in \mathbb{R}$, then the full support condition is violated in $s = 0$. This however is an a priory restriction in our setup. Instead, one should think of Brownian motion as a process on our stochastic basis with initial value B_0, where B_0 is a $\mathcal{F}_0$ measurable random variable. Then, it is very well possible to have B_0 be distributed by an initial distribution that satisfies the full support condition since $\mathcal{F}_0$ does not need to be the trivial σ-algebra. The situation where B starts in x is then recovered by conditioning on $\{B_0 = x\}$ which is a measurable set since by our initial assumptions, the filtration is complete.*

In the following, for a process X with the k-polynomial property, we will indicate that a mapping τ is an evolution system on Pol_k for X by writing

$$X \xrightarrow{\mathrm{Pol}_k} \tau.$$

Definition 6.19. *We call a pair* (X, τ) ***k*-polynomial process** *for an adapted càdlàg process* X *and a mapping* $\tau : \mathbb{T} \times \mathrm{Pol}_k \to \mathrm{Pol}_k$*, if* X *has the k-polynomial property and* $X \overset{\mathrm{Pol}_k}{\longrightarrow} \tau$.

In the sequel we will not assume uniqueness of τ as it is sufficient for our purposes to just have one degree preserving evolution system available. However, the example further down shows that even under perfect regularity conditions for X such as continuity and boundedness, it is then possible that the càdlàg property of X is not inherited by τ. Recall that in the last chapter we equipped Pol with the topology induced by the norm $\|\cdot\|_{\mathrm{Pol}}$ and the subspaces Pol_k are equipped with the subspace norm. In particular we get that the topology induced on Pol_k coincides with the trace topology. Recall the following definition.

Definition 6.20. *We call a mapping* $\tau : \mathbb{T} \times \mathrm{Pol}_k \to \mathrm{Pol}_k$ *càdlàg , if for any* $f \in \mathrm{Pol}_k$ *and* $x \in \mathbb{R}^d$,

(i) the function $t \mapsto \tau_{s,t} f(x)$ *is càdlàg on* $[s, \infty)$ *and*

(ii) the function $s \mapsto \tau_{s,t} f(x)$ *is càdlàg on* $[0, t]$.

This definition is compatible with our topology on Pol_k as demonstrated by the next proposition.

Proposition 6.21. *Consider a mapping* $\tau : \mathbb{R}_+ \times \mathrm{Pol}_k \to \mathrm{Pol}_k$ *and let* t_n *be a sequence in* $\mathbb{R}_+$ *converging to* t*. Then*

$$\lim_{n\to\infty} \|\tau_{t_n} f - \tau_t f\|_{\mathrm{Pol}} = 0 \tag{6.17}$$

if and only if for all $x \in \mathbb{R}^d$,

$$\lim_{n\to\infty} |\tau_{t_n} f(x) - \tau_t f(x)| = 0. \tag{6.18}$$

Proof. Note that since τ_{t_n}, τ_t map polynomials to polynomials of at most degree k, we have

$$\tau_{t_n} f(x) - \tau_t f(x) = ((\tau_{t_n} - \tau_t) f)(x) = \sum_{i=1}^{N_k} \alpha_{n,i} \beta_i(x),$$

where for each i, $(\alpha_{n,i})_n$ is a sequence in $\mathbb{R}$. Note that by the definition of $\|\cdot\|_{\mathrm{Pol}}$, (6.17) is equivalent to $\alpha_{n,i}$ being a zero sequence in n for each n. Hence, if (6.17) holds, then (6.18) follows.

For the other direction, note that since we are dealing with polynomials, we get that the left hand side converges uniformly to zero on all compact sets, including $\overline{B_1(0)}$. But that implies that $(\tau_{t_n} - \tau_f) f$ converges to the zero polynomial with respect to the supremum-norm on $\overline{B_1(0)}$, hence all coefficients $\alpha_{n,i}$ for all i must be zero sequences in n, implying (6.17). □

Let us now give an example, showing that τ can loose the càdlàg property of X.

Example 6.22. *Consider the deterministic process defined below and the basis* $\beta = (\beta_0, \beta_1, \beta_2)$ *of* $\mathrm{Pol}_1(\mathbb{R}^2)$*, also defined below. We shall use the notation* $z = (x, y) \in \mathbb{R}^2$*.*

$$X_t = \begin{pmatrix} 0 \\ t \end{pmatrix}, \ \beta_0(x,y) = 1, \ \beta_1(x,y) = x, \ \beta_2(x,y) = y.$$

Since X *is deterministic, we have for any* $f \in \mathrm{Pol}_1$*,*

$$\mathbb{E}\left[f(X_t) \mid \mathcal{F}_s\right] = f(X_t).$$

Since we require the evolution system τ *of* X *on* Pol_1 *to be pointwise linear, it is enough to specify* $\tau_{s,t}\beta_i$ *for* $i = 0, 1, 2$*. The polynomial property implies*

$$\begin{aligned}
1 &= \beta_0(X_t) = c_{00}(s,t)\beta_0(X_s) \\
0 &= \beta_1(X_t) = c_{10}(s,t)\beta_0(X_s) + c_{11}(s,t)\beta_1(X_s) + c_{12}(s,t)\beta_2(X_s) \\
t &= \beta_2(X_t) = c_{20}(s,t)\beta_0(X_s) + c_{21}(s,t)\beta_1(X_s) + c_{22}(s,t)\beta_2(X_s),
\end{aligned}$$

where the functions $c_{ij} : \mathbb{T} \to \mathbb{R}$ *must be chosen so that* τ *satisfies the evolution property with* $\tau_{s,s} = \mathrm{id}_{\mathrm{Pol}_1}$*. Using* $\beta_1(X_t) \equiv 0$ *and* $\beta_2(X_t) = t$ *for all* $t \geq 0$*, we see that the specification* $c_{00} = c_{11} = c_{22} = 1, c_{10} = c_{12} = c_{21} = 0$ *and* $c_{20}(s,t) = t - s$ *satisfies the above equations and defines* τ *to be an evolution system on* Pol_1 *as one sees by observing* $\tau_{s,t}\beta_i = \beta_i$ *for* $i = 0, 1$ *which yields the evolution property for the first two basis polynomials. For the third, consider for* $0 \leq s \leq u \leq t$

$$\tau_{u,t}\beta_2 = (t-u)\beta_0 + \beta_2 \Rightarrow \tau_{s,u} \circ \tau_{u,t}\beta_2 = (t-u)\tau_{s,u}\beta_0 + \tau_{s,u}\beta_2 = t - s,$$

which shows τ *is indeed an evolution system of* X *on* Pol_1*. However, a closer look at the necessary conditions on the functions* c_{ij} *above shows that due to* $\beta_1(X_t) = 0$ *for all* $t \geq 0$*, one can actually change the definition of* c_{11} *by setting* $c_{11}(s,t) = \rho(s,t)$*, where* ρ *is an arbitrary MES in dimension* $d = 1$*. Hence,*

$$\tau_{s,t}\beta_1(x,y) = \rho(s,t)\beta_1(x,y) = \rho(s,t)x.$$

Then τ *is still an evolution system of* X *on* Pol_1*. Consider now Example 4.12 and define* ρ *as in that example. As we have already seen, this defines an evolution system which is not càdlàg in* $s = 1$ *for* $t = 1$*. Hence,* τ *is not a càdlàg evolution system.*

Motivated by the example above, we will make regularity assumptions in the next section that allow us to consider MESs that satisfy regularity properties suitable for our following discussions.

6.2 Regular polynomial processes and representations in finite dimensions

The goal of this section is to make use of the fact that linear maps from $\mathrm{Pol}_k \mapsto \mathrm{Pol}_k$ can be represented as matrix-vector operations as we have seen in the previous chapter. We will see that the study of an evolution system on Pol_k of a k-polynomial process (X, τ) can be reduced to the study of matrix evolution systems (MESs) as introduced in Chapter 4.

Recall that for all polynomials $f \in \mathrm{Pol}_k$, we have

$$f(x) = \iota_k(f)^\top \mathbf{H}_{\beta^k}(x) = \mathbf{f}^\top \mathbf{H}_{\beta^k}(x), \quad \forall x \in \mathbb{R}^d.$$

Note that we indicate $\mathbf{f} = \iota_k(f)$ by using bold letters. In the sequel we could use any basis of Pol_k, but that would not gain us any generality, hence we keep our notation simple by keeping the conventions and notations we made in the previous chapter regarding the bases $\beta^l, l \in \mathbb{N}_0$, unless otherwise specified.

Consider a k-polynomial process (X, τ). A function $P : \mathbb{T} \to \mathrm{M}_N(\mathbb{R})$ is the unique representation of τ wrt the vector space isomorphism ι_k, denoted by an abuse of notation as $\iota_k(\tau) = P$, if for all $(s,t) \in \mathbb{T}$ and $f \in \mathrm{Pol}_k$ we have

$$(\tau_{s,t} f)(x) = (\iota_k(\tau_{s,t} f))^\top \mathbf{H}_{\beta^k}(x) = (P(s,t)\mathbf{f})^\top \mathbf{H}_{\beta^k}(x), \forall x \in \mathbb{R}^d.$$

We have seen in the previous chapter that since τ is an evolution system on Pol_k, P is a MES on $\mathrm{M}_N(\mathbb{R})$. Note that τ has the additional property $\tau_{s,t}(\mathrm{Pol}_l) \subset \mathrm{Pol}_l$ for $l \in \{0, \ldots, k\}$, which is inherited by P in the sense that it each $P(s,t)$ has upper block triangular form since the basis β^k under the conventions of the previous chapter satisfies

$$\deg(\beta_i) < \deg(\beta_j) \Rightarrow i < j,$$

hence τ is degree preserving on Pol_k in the terminology introduced in the previous chapter and equivalently so is P. In that case the blocks correspond to the indices belonging to those basis functions with the same degree, hence there are $k+1$ blocks. Equivalently to $P = \iota_k(\tau)$ we will sometimes write $\tau = \iota_k^{-1}(P)$ to indicate that τ is defined by P and not the other way around.

Definition 6.23. *We call (X, P)* ***k*-polynomial process***, if P is a MES on $M_N(\mathbb{R})$ and for $\tau = \iota_k^{-1}(P)$ we have that (X, τ) is a k-polynomial process.*

Since we have fixed the basis of Pol_k for all k, it is always clear which isomorphism ι_k is meant. We will now define regularity properties for k-polynomial processes through regularity of P as a MES.

Definition 6.24. *We call a k-polynomial process (X, τ) (resp. (X, P))* **càdlàg** *, if P is a càdlàg MES (which is equivalent to saying τ is càdlàg). If P is even regular, we call the first and the latter* **regular**. *Further, we call (X, P, γ, A) a* **proper k-polynomial process**, *if (X, P) is a k-polynomial process such that (P, γ, A) is a proper MES on $M_{N_k}(\mathbb{R})$.*

Remark 6.25. *Recall that for a proper MES (P, γ, A), by definition we have that P is càdlàg . Hence, if (X, P, γ, A) is a proper k-polynomial process (wrt. a fixed basis β), then (X, P) is in particular a càdlàg k-polynomial process and by that also (X, τ) for $\tau = \iota_k^{-1}(P)$.*

We will later see that under these regularity conditions and additional assumptions, a notion of extended generator for X similar to that for Markov processes can be represented on Pol_k precisely by the generator of a proper MES coming from a k-polynomial process X. This will be the topic of the next section.

6.3 Extended generators for k-polynomial processes

In this section we want to extend the concept of infinitesimal description by means of extended generators of a stochastic process to our setting. As noted in the introduction, the extended generator plays a central role in e.g. Cuchiero et al. [Cuc+12] and Filipović et al. [Fil+17]. So far, the study of finite dimensional polynomial processes has been under regularity assumptions that imply stochastic continuity as a consequence of the Markov inequality.

In the existing literature for polynomial processes, extended generators have been defined as mappings $\mathcal{G}$ (or as a collection of mappings $(\mathcal{G}_s)_{0 \leq s}$, c.f. C. A. Hurtado [CAH17]) that map functions f coming from a certain domain to functions $\mathcal{G}_s f$ such that

$$f(X_t) - f(X_s) - \int_s^t (\mathcal{G}_u f)(X_u) du \in \mathcal{M}_{loc}.$$

Notice that the integral is wrt the Lebesgue measure, which requires that $f(X)$ is quasi left continuous. This implies certain properties of $f(X)$, compare Jacod and Shiryaev [JS13, Propositions 2.26 and 2.9]. Hence, the only polynomials that lie in the domain of the extended generator in case of stochastic discontinuity, or $\Delta X_t \neq 0$, are the constant polynomials. We however want to extend the class of polynomial processes to include processes with deterministic jump-times in the spirit of Keller-Ressel et al. [KR+18] where this was done for the affine case. This requires a more general notion of an extended generator, we will introduce it further down and ask the reader to indulge us, will make it necessary for us to be able to deal with expressions of the form

$$\mathbb{E}\left[f(X_{t-}) \mid \mathcal{F}_s\right] \text{ rather then } \mathbb{E}\left[f(X_t) \mid \mathcal{F}_s\right],$$

where f is a polynomial with $\deg(f) \leq k$ and (X, τ) a k-polynomial process. In case of quasi-left continuity, i.e. $\Delta X_t = 0$, both expressions coincide a.s., but when this is not the case, we can so far only compute expressions as on the right side using τ. assume that (X, τ) is a càdlàg k-polynomial process. Then, if we assume that we can apply the dominated convergence theorem to the limit below, we get for $s < t$ using continuity of polynomials that

$$\tau_{s,t-} f(X_s) = \lim_{\varepsilon \downarrow 0} \tau_{s,t-\varepsilon} f(X_s) = \lim_{\varepsilon \downarrow 0} \mathbb{E}\left[f(X_{t-\varepsilon}) \mid \mathcal{F}_s\right] = \mathbb{E}\left[f(X_{t-}) \mid \mathcal{F}_s\right] \quad [P], \tag{6.19}$$

where the limit exists by our càdlàg assumption and the right hand side implicitly by the dominated convergence theorem. This is precisely what we need. Alternatively, the generalized dominated convergence theorem (see Theorem 3.3) could be used to obtain the same result using the following well known fact about convergence of random variables.

Proposition 6.26. *Let (X_n) be a sequence of random variables such that $X_n \to Y$ a.s. and $X_n \xrightarrow{\mathcal{L}^1} Y'$. Then $Y' = Y \quad [P]$.*

Therefore, the property we consider is uniform integrability for the sets

$$\mathcal{H}_t^f = \{f(X_s) : 0 \leq s \leq t\},$$

for all $t \geq 0$ and $f \in \mathrm{Pol}_k$. This motivates the following definition.

Definition 6.27. *We call a càdlàg k-polynomial process (X, τ)* **uniformly integrable***, if for each $t \geq 0$ and $f \ni \mathrm{Pol}_k$, the set $\mathcal{H}_t^f$ is uniformly integrable. Similarly, we call proper k-polynomial process (X, P, γ, A)* **uniformly integrable***, if (X, τ) is uniformly integrable for $\tau = \iota_\beta^{-1}(P)$.*

Remark 6.28. *Note that for k-polynomial processes, we have*

$$\mathbb{E}\left[\left\|\mathbf{H}_{\beta^k}(X_t)\right\|_{\mathbb{R}^{N_k}}\right] < \infty \quad \forall t \geq 0, \tag{6.20}$$

since by definition $\mathbb{E}\left[|\beta_i(X_t)|\right] < \infty$ for all $i \leq N_k$.

The usual conditions of our stochastic basis guarantee that

$$\mathcal{F}_s = \bigcap_{\varepsilon > 0} \mathcal{F}_{s+\varepsilon}.$$

Unlike intersections however, we do not have that unions of σ-fields are again σ-fields. Hence we define the left limit of our filtration the following way,

$$\mathcal{F}_{s-} := \sigma\Big(\bigcup_{\varepsilon > 0} \mathcal{F}_{s-\varepsilon}\Big). \tag{6.21}$$

Lemma 6.29. *Let (X, τ) be a càdlàg k-polynomial process. Then the following holds for any $f \in \mathrm{Pol}_k$:*

(i) For $s < t$ and (X, τ) uniformly integrable,

$$\mathbb{E}\left[f(X_{t-})|\mathcal{F}_s\right] = \tau_{s,t-}f(X_s) \quad [P]. \tag{6.22}$$

(ii) For $0 < s \leq t$,

$$\mathbb{E}\left[f(X_t)|\mathcal{F}_{s-}\right] = \tau_{s-,t}f(X_{s-}) \quad [P]. \tag{6.23}$$

(iii) For $s < t$ and (X, τ) uniformly integrable, $\tau_{s-,t-}(X_{s-})$ exists and

$$\mathbb{E}\left[f(X_{t-})|\mathcal{F}_{s-}\right] = \tau_{s-,t-}f(X_{s-}) \quad [P]. \tag{6.24}$$

Proof. (i) Note that the limit $\tau_{s,t-}f$ exists by the càdlàg property for all $f \in \mathrm{Pol}_k$, hence the limit below exists

$$\lim_{\varepsilon \downarrow 0} \tau_{s,t-\varepsilon}f(X_s) = \tau_{s,t-}f(X_s) \quad [P].$$

Next, note that by assumption we have that $\mathcal{H}_T^f$ is uniformly integrable for some $T > t$. Since f is continuous and X a càdlàg stochastic process, we have $f(X_{t-\varepsilon}) \overset{[P]}{\to} f(X_{t-})$ as $\varepsilon \downarrow 0$. Hence, by Theorem 3.3, we get $f(X_{t-\varepsilon}) \overset{\mathcal{L}^1}{\to} f(X_{t-})$ as $\varepsilon \downarrow 0$. By Lemma 3.5, we have $\mathbb{E}\left[|f(X_{t-})|\right] < \infty$, hence $\mathbb{E}\left[f(X_{t-}) \mid \mathcal{F}_s\right]$ is well defined. By Jensen's inequality and the tower property we get

$$\begin{aligned}
&\lim_{\varepsilon \downarrow 0} \mathbb{E}\left[\,|\,\mathbb{E}\left[f(X_{t-\varepsilon}) \mid \mathcal{F}_s\right] - \mathbb{E}\left[f(X_{t-}) \mid \mathcal{F}_s\right]|\,\right] \\
&\quad = \lim_{\varepsilon \downarrow 0} \mathbb{E}\left[\,|\,\mathbb{E}\left[f(X_{t-\varepsilon}) - f(X_{t-}) \mid \mathcal{F}_s\right]|\,\right] \\
&\quad \leq \lim_{\varepsilon \downarrow 0} \mathbb{E}\left[\,|\,f(X_{t-\varepsilon}) - f(X_{t-})\,|\,\right] = 0,
\end{aligned} \tag{6.25}$$

which shows

$$\mathbb{E}\left[f(X_{t-\varepsilon}) \mid \mathcal{F}_s\right] \xrightarrow{\mathcal{L}^1} \mathbb{E}\left[f(X_{t-}) \mid \mathcal{F}_s\right] \text{ as } \varepsilon \downarrow 0.$$

Putting these points together, by page 87 we conclude

$$\tau_{s,t-}f(X_s) = \mathbb{E}\left[f(X_{t-}) \mid \mathcal{F}_s\right] \quad [P].$$

(ii) Let (s_n) be an increasing sequence with $s_n \uparrow s$. Define

$$Y_n := \mathbb{E}\left[f(X_t) \mid \mathcal{F}_{s_n}\right].$$

Then, (Y_n) forms a martingale wrt. the filtration $(\mathcal{F}_{s_n})_{n \geq 1}$ and since $f(X_t) \in \mathcal{L}^1(P, \Omega, \mathcal{F})$ by the k-polynomial property, Lévy's upwards convergence theorem (c.f. Kallenberg [Kal06, Theorem 7.23]) yields $\lim_{n\to\infty} Y_n = \mathbb{E}\left[f(X_t) \mid \mathcal{F}_{s-}\right] =: Y_\infty\, [P]$. Further, we have

$$Y_n = \tau_{s_n,t} f(X_{s_n}).$$

Fix an arbitrary basis β of Pol_k and with $P = \iota_\beta^{-1}(\tau)$. Then a.s. we have

$$Y_n = \left(P(s_n,t)\mathbf{f}\right)^\top \mathbf{H}_\beta(X_{s_n}).$$

By our càdlàg assumptions, $P(s_n,t) \to P(s-,t)$ and $\mathbf{H}_\beta(X_{s_n}) \to \mathbf{H}_\beta(X_{s-})\,[P]$. Hence,

$$Y_\infty = \left(P(s-,t)\mathbf{f}\right)^\top \mathbf{H}_\beta(X_{s-}),$$

and since $\iota_\beta(\tau_{s-,t} f) = P(s-,t)\mathbf{f}$ for all $f \in \mathrm{Pol}_k$, we have

$$\tau_{s-,t} f(X_{s-}) = Y_\infty = \mathbb{E}\left[f(X_t) \mid \mathcal{F}_{s-}\right].$$

(iii) We need to show for $s_n \uparrow s$ and $t_m \uparrow t$, that

$$\lim_{n\to\infty}\lim_{m\to\infty} \tau_{s_n,t_m} = \lim_{m\to\infty}\lim_{n\to\infty} \tau_{s_n,t_m}.$$

Fix a basis β of Pol_k. Then for $P = \iota_\beta(\tau)$, this is equivalent to the limits being equal for $P(s_n,t_m)$ instead of τ_{s_n,t_m}. Since P is a càdlàg MES, this follows from Lemma 4.22. We now get by (ii) that

$$\lim_{m\to\infty}\lim_{n\to\infty} \mathbb{E}\left[f(X_{t_m}) \mid \mathcal{F}_{s_n}\right] = \lim_{m\to\infty} \mathbb{E}\left[f(X_{t_m}) \mid \mathcal{F}_{s-}\right],$$

and the rhs limit exist by the same arguments as in (i). Combining this with the previous result, we get just as we did in (ii),

$$\mathbb{E}\left[f(X_{t-}) \mid \mathcal{F}_{s-}\right] = \tau_{s-,t-} f(X_{s-}) \quad [P]. \tag{6.26}$$

Finally, we have

$$\lim_{n\to\infty}\lim_{m\to\infty} \mathbb{E}\left[f(X_{t_m}) \mid \mathcal{F}_{s_n}\right] = \lim_{n\to\infty} \mathbb{E}\left[f(X_{t-}) \mid \mathcal{F}_{s_n}\right],$$

and by the same Lévy upward convergence theorem argument as in (ii), we get that this limit exists and is a.s. equal to the rhs of (6.26), which finishes the proof. □

The next definition will play the central role in our discussions of polynomial processes. In particular, it will allow us to show a tight connection between the extended generator of a stochastic process that we will introduce further down and the extended generator of a MES in the context of k-polynomial processes.

Recall the following definition from Chapter 4. Let μ be a measure on $\mathbb{R}_{>0}$ and $G : \mathbb{R}_{>0} \to \mathrm{M}_n(\mathbb{R})$ for some $n \in \mathbb{N}$. We define the running essential supremum of G wrt μ as

$$\|G\|_{\mu,t} := \operatorname*{ess\,sup}_\mu (\|G(\cdot)\|_\infty \mathbb{1}_{\cdot \le t}),$$

where the norm is the operator norm wrt some norm on $\mathbb{R}^n$. Note that a condition of the form $\|G\|_{\mu,t} < \infty$ is independent of the concrete choice of norm that defines the operator norm.

Definition 6.30 (bounded k-polynomial process)**.** *We call a collection (X, P, A, G) a* **bounded k-polynomial process***, if there exists a family of measures γ such that (X, P, γ, A) is a proper k-polynomial process and G is an extended generator of the MES P such that for all $t \geq 0$ we have $\|G\|_{dA,t} < \infty$, where dA is the measure induced by the increasing càdlàg function A.*

We now introduce the concept of extended generators for stochastic processes as done in Çinlar et al. [Çin+80]. The original introduction in this generality can be found in Kunita [Kun69, Definition 3.1].

Definition 6.31. *Fix a stochastic basis $(\Omega, \mathscr{F}, \mathbb{F}, P)$. Let X be an adapted càdlàg process on this basis and let $A \in \mathcal{V}^+$. Denote by $\mathcal{D}$ the set of all $f \in \mathcal{E}(\mathbb{R}^d)$ for which there exists a function $g \in \mathcal{E}(\mathbb{R}_{>0} \times \mathbb{R}^d)$, such that*

$$M_t^f = f(X_t) - f(X_0) - \int_{(0,t]} g(u, X_{u-}) dA_u \in \mathcal{M}_{loc}. \tag{6.27}$$

We call a mapping $\tilde{\mathcal{G}}$ that maps $f \mapsto g$ a **pre-version** *of the* ***extended A-generator*** *of the process X with domain $\mathcal{D}_{X,A}$.*

By definition, the domain is independent of the choice of a pre-version of the extended A-generator.

Note that in our definition we speak of pre-versions of an extended generator. Let us discuss this. First, we only talk about a version of an extended generator, since for two functions g, g' such that $g(u, \cdot) = g'(u, \cdot)$ outside a dA-zero set we have that if $f \mapsto g$ satisfies (6.27), then so does $f \mapsto g'$. Hence we can at best expect uniqueness outside dA-zero sets. Further, we call it a pre-version, since just as in the discussion of the maps $\tau_{s,t}$, we can not in general expect linearity of a map $\tilde{G}$ since it is well possible that two functions $g \neq g'$ exist and yet the two processes $u \mapsto g(u, X_{u-})$ and $u \mapsto g'(u, X_{u-})$ are equal up to evanescence. As we did with $\tau_{s,t}$ however, we can always construct a linear modification on Pol_k. This is demonstrated in the next proposition.

Proposition 6.32. *Assume the setting in Definition 6.31. Assume for each $f \in \mathrm{Pol}_k$ there exists a function $g \in \mathcal{E}(\mathbb{R}_{>0} \times \mathbb{R}^d)$ such that (6.27) holds. Then, there exists a linear map $\mathcal{G} : \mathrm{Pol}_k \to \mathcal{B}(\mathbb{R}_{>0} \times \mathbb{R}^d)$ such that $\mathcal{G}$ is the restriction on Pol_k of a pre-version of the extended A-generator of X and $\mathrm{Pol}_k \subset \mathcal{D}_{X,A}$.*

Proof. The statement $\mathrm{Pol}_k \subset \mathcal{D}_{X,A}$ is part of the assumption. We write $N = N_k$. Fix the basis $\beta^k = \{\beta_1, \ldots, \beta_N\}$ of Pol_k. By assumption, for each $i = 1, \ldots, N$, there exists a $g_i \in \mathcal{B}(\mathbb{R}_{>0} \times \mathbb{R}^d)$ such that (6.27) holds. Hence, define $\mathcal{G}(\beta_i) := g_i$. We now extend the definition of $\mathcal{G}$ to all of Pol_k. To that extend, let $f \in \mathrm{Pol}_k$ be arbitrary with

$$f = \sum_{i=1}^{N} c_i \beta_i.$$

Then we define

$$\mathcal{G}(f) = \sum_{i=1}^{N} c_i \mathcal{G}(\beta_i).$$

By construction, $\mathcal{G}$ is a linear mapping. It is further a pre-version of the extended A-generator of X restricted to Pol_k due to linearity of the Lebesgue integral and the fact that local martingales are defined as the localization of uniformly integrable martingales which by Protter [Pro05, Chapter I Theorem 18] yields that the finite sum of local martingales is again a local martingale. □

Remark 6.33. *Note that if M_t^f is a local martingale, it follows that*

$$M_{s,t}^f = f(X_t) - f(X_s) - \int_{(s,t]} (\tilde{\mathcal{G}} f)(u, X_{u-}) dA_u \tag{6.28}$$

is a local martingale with respect to the filtration $\mathbb{F}_s = (\mathcal{F}_t)_{t \geq s}$, since $M_{s,t}^f = M_t^f - M_s^f$ and the sum of two local martingales is again a local martingale.

Remark 6.34. *Let us stress that by our definition, the extended A-generator does not need to be unique, even modulo agreeing outside a dA-zero set. This is because we choose as function domains $\mathbb{R}^d$ regardless of the support of the underlying process. If however, one would make the identification of processes that are indistinguishable, one would get uniqueness modulo agreeing outside dA-zero sets since local martingales are special semimartingales. Hence they have a unique (up to indistinguishability) canonical decomposition into local martingale part and finite variation part. This viewpoint is taken e.g. in Çinlar et al. [Çin+80], where the extended A-generator is a mapping between processes instead of deterministic functions. They also consider the case where A is itself a random process in the context of extended random generators. For our discussion however, we will only consider the case $A \in \mathcal{V}^+_{\not{\Omega}}$. We finish this remark by noting that the reason we take our approach is that it will allow for a simplified discussion. Otherwise, we would have to deal with situations where we identify two polynomials at a time point t but not at a different time point t' since the processes we consider can have a time dependent support.*

Remark 6.35. *Note that $\mathcal{D}_{X,A}$ is a linear space since the finite linear combination of local martingales is a local martingale. If one allows for the axiom of choice (and the equivalent lemma of Zorn), one can construct a linear pre-version of the extended A-generator the same way we did for the finite dimensional case in Proposition 6.32. Indeed, by Zorn's lemma, there exists a vector space basis of $\mathcal{D}_{X,A}$. By the axiom of choice, we can define a mapping on the basis as we did in the finite dimensional setting. Then, one extends canonically to the linear hull, noting that an arbitrary element of $\mathcal{D}_{X,A}$ has a finite representation as a linear combination, extending the local martingale property to the extended map. The existence of a linear pre-version leads to the following definition.*

Definition 6.36. *We call a pre-version of the extended A-generator $\mathcal{G}$ a* **version of the extended A-generator** *(or in short a version of $\mathcal{G}_{X,A}$), if it is a linear map between* *domain and range. If it is linear as a restriction on a subspace $\mathcal{S}$, we say $\mathcal{G}|_{\mathcal{S}}$ is a version of the extended A-generator on $\mathcal{S}$ or in short $\mathcal{G}|_{\mathcal{S}}$ is a version of $\mathcal{G}_{X,A}$ on $\mathcal{S}$. If $\mathcal{G}$ is only defined on $\mathcal{S}$, then we say $\mathcal{G}$ is a version of $\mathcal{G}_{X,A}$ on $\mathcal{S}$.*

Recall the notation from the previous chapter, where for a matrix function $G : \mathbb{R}_{>0} \to \mathrm{M}_{N_k}(\mathbb{R})$, we define an operator $\mathcal{G} : \mathbb{R}_{>0} \times \mathrm{Pol}_k \to \mathrm{Pol}_k$ pointwise through

$$(u, f) \mapsto \mathcal{G}_u f(\cdot) = (G(u)\iota_k(f))^\top \mathbf{H}_{\beta^k}(\cdot).$$

This was denoted by $\iota_k(G) = \mathcal{G}$. Since we will be using matrix functions such as G to define versions $\mathcal{G}_{X,A}$ on Pol_k, it will be convenient to interpret a mapping $\mathcal{G}$ defined by G as a mapping $\mathcal{G} : \mathrm{Pol}_k \to \mathcal{E}(\mathbb{R}_{>0} \times \mathbb{R}^d)$, by

$$\mathcal{G}f(u, x) = \mathcal{G}_u f(x), \quad u \in \mathbb{R}_{>0}, x \in \mathbb{R}^d.$$

In particular, we get that for each $u \in \mathbb{R}_{>0}$ fixed, we have

$$\mathcal{G}f(u, \cdot) = \mathcal{G}_u f(\cdot) = (G(u)\iota_k(f))^\top \mathbf{H}_{\beta^k}(\cdot) \in \mathrm{Pol}_k.$$

We define the function space $\mathrm{Pol}_k(\mathbb{R}_{>0} \times \mathbb{R}^d)$ as the set of all functions $g : \mathbb{R}_{>0} \times \mathbb{R}^d \to \mathbb{R}$ such that for fixed $u \in \mathbb{R}_{>0}$, $g(u, \cdot) \in \mathrm{Pol}_k$ and for all $x \in \mathbb{R}^d$, $u \mapsto g(u, x)$ is Borel measurable. In the sequel, we write $G \overset{\iota_k}{\to} \mathcal{G}$ to indicate that $\mathcal{G}$ is defined by G as a mapping $\mathcal{G} : \mathrm{Pol}_k \to \mathrm{Pol}_k(\mathbb{R}_{>0} \times \mathbb{R}^d)$. An important observation is that by the definition of $\mathcal{G}$, if G is degree preserving on Pol_k, then for any $l \leq k, u \in \mathbb{R}_{>0}$ we have $\mathcal{G}f(u, \cdot) \in \mathrm{Pol}_l$ whenever $f \in \mathrm{Pol}_l$.

Note that for all functions $\mathcal{G} : \mathrm{Pol}_k \to \mathrm{Pol}_k(\mathbb{R}_{>0} \times \mathbb{R}^d)$, which are linear for the pointwise sum on Pol_k and $\mathrm{Pol}_k(\mathbb{R}_{>0} \times \mathbb{R}^d)$, there exists a measurable function $G : \mathbb{R}_{>0} \to \mathrm{M}_{N_k}(\mathbb{R})$ such that G defines $\mathcal{G}$ as above, i.e. $G \overset{\iota_k}{\to} \mathcal{G}$. To see that, note that by linearity we only need to look at the basis polynomials of Pol_k. Hence, there exist functions $a_{ij} : \mathbb{R}_{>0} \to \mathbb{R}$ for $i, j \in \{1, \ldots, N_k\}$ with

$$(\mathcal{G}\beta_i)(u, x) = \sum_{j=1}^{N_k} a_{ij}(u)\beta_j(x), \quad u \in \mathbb{R}_{>0}, \ x \in \mathbb{R}^d.$$

Note that $\iota_k(\beta_i) \in \mathbb{R}^{N_k}$ is the vector with the i-th coordinate being one and the rest zero. Hence the elements if $G(u)$ are given by $a_{ij}(u) = (G(u))_{ij}$. The measurability claim is now obvious since all β_i are linear independent continuous and hence measurable functions and the product Borel σ-algebra is the Borel σ algebra wrt the product topology of $\mathbb{R}_{>0} \times \mathbb{R}^d$. Hence, $\mathcal{G}$ is Borel measurable if and only if G is Borel measurable.

The next lemma shows that we get $\mathcal{L}^1$ boundedness for the sets $\mathcal{H}_t^f$ of a stochastic process if the extended A-generator has a certain polynomial structure. The idea of the proof is to use the Gronwall inequality in the spirit of the proof of Cuchiero et al. [Cuc+12, Theorem 2.10].

Lemma 6.37. *Let $k \geq 0$ be an even number, $\beta^k = \{\beta_1, \dots, \beta_{N_k}\}$ our fixed basis of* Pol_k *and $A \in \mathcal{V}^+_{\mathfrak{A}}$. Let X be a càdlàg adapted process such that $\mathrm{Pol}_k \subset \mathcal{D}_{X,A}$. Assume there exists a measurable function*

$$G : \mathbb{R}_{>0} \to M_{N_k}(\mathbb{R}) \text{ with } \|G\|_{dA,t} < \infty \quad \forall t \geq 0,$$

such that the map $\mathcal{G} : \mathrm{Pol}_k \to \mathrm{Pol}_k(\mathbb{R}_{>0} \times \mathbb{R}^d)$ defined by $G \overset{\iota_k}{\to} \mathcal{G}$, i.e.

$$f \mapsto \mathcal{G}f(u,x) = (G(u)\mathbf{f})^\top \mathbf{H}_{\beta^k}(x), \quad (u,x) \in \mathbb{R}_{>0} \times \mathbb{R}^d,$$

is a version of $\mathcal{G}_{X,A}$ on Pol_k. Then for any $f \in \mathrm{Pol}_k$ and $T > 0$,

(i) there exists a constant K_T^f, depending on f and T such that

$$\operatorname*{ess\,sup}_{u \in [0,T],\, dA} |\mathcal{G}f(u,x)| \leq K_T^f F(x),$$

for a polynomial $F \in \mathrm{Pol}_k$ that does not depend on the choice f and T.

(ii) If $\mathbb{E}\left[\|X_0\|^k_{\mathbb{R}^d}\right] < \infty$, the sets $\mathcal{H}_T^f$ are $\mathcal{L}^1$-bounded, i.e. for all $T \geq 0$,

$$\sup_{0 \leq t \leq T} \mathbb{E}\left[|f(X_t)|\right] < \infty \quad \forall f \in \mathrm{Pol}_k.$$

(iii) For all $f \in \mathrm{Pol}_k$, the process g defined via $g_t = \int_{(0,t]} \mathcal{G}f(u, X_{u-})A(du)$ is càdlàg, adapted, predictable and belongs to $\mathcal{V}$.

Proof. Note that the case $k = 0$ is trivial since then all polynomial in Pol_k are constant. We set $N = N_k$. Recall $f_i(x) = x_i$. Hence, we have $f_i^k \in \mathrm{Pol}_k$. Let us first show (i). As in [Cuc+12], define the polynomial $F \in \mathrm{Pol}_k$ by

$$0 \leq F(x) := 1 + \sum_{i=1}^d f_i(x)^k.$$

For a polynomial $f \in \mathrm{Pol}_k$ we define the function $|f|$ by

$$f(x) = \sum_{i=1}^N c_i \beta_i(x) \Rightarrow |f|(x) := \sum_{i=1}^N |c_i| \, |\beta_i(x)| \, .$$

Note that for any $f \in \mathrm{Pol}_k$, there is a $K < \infty$ such that

$$|f(x)| \leq |f|(x) \leq KF(x). \tag{6.29}$$

Define the polynomial $\lambda \in \mathrm{Pol}_k$ by $\iota_k(\lambda) = (1, \dots, 1)^\top \in \mathbb{R}^N$. Define

$$c_\infty = \max_{i \in \{1,\dots,N\}} |c_i| \, , \text{ hence } c_\infty = \|\iota_\beta(f)\|_\infty = \|\mathbf{f}\|_\infty \, .$$

Further, let K_λ be the constant in (6.29) for the polynomial λ. By equivalence of norms on $\mathbb{R}^N$, we can pick any norm, so we pick the norm $\|\cdot\|_1$ and the associated operator norm. It then holds

$$\begin{aligned}\|\mathbf{f}\|_1 \left\|\mathbf{H}_{\beta^k} k(x)\right\|_1 &= \left(\sum_{i=1}^N |c_i|\right)\left(\sum_{j=1}^N |\beta_i(x)|\right) = \left(\sum_{i,j=1}^N |c_i|\,|\beta_i(x)|\right)\\ &\le \left(\sum_{i,j=1}^N c_\infty\,|\beta_i(x)|\right) = Nc_\infty|\lambda|(x) \le N\|\mathbf{f}\|_\infty K_\lambda F(x).\end{aligned}$$

Noting that $\|G\|_{dA,t}$ is the running essential supremum wrt the measure dA, we can compute for dA-almost any $u \in (0,t]$,

$$\begin{aligned}\operatorname*{ess\,sup}_{u\in[0,T],\,dA} |\ \mathcal{G}f(u,x)\ | &= \operatorname*{ess\,sup}_{u\in[0,T],\,dA}\left|(G(u)\mathbf{f})^\top \mathbf{H}_{\beta^k}(x)\right| \le \|G\|_{dA,T}\|\mathbf{f}\|_1\left\|\mathbf{H}_{\beta^k}(x)\right\|_1\\ &\le \|G\|_{dA,T}\, N\,\|\mathbf{f}\|_\infty K_\lambda F(x).\end{aligned}$$

Hence, we have shown (i) for $K_T^f = \|G\|_{dA,T}\, N\,\|\mathbf{f}\|_\infty K_\lambda$. Regarding (ii), note that we can choose $f = F$ so that $c_\infty = \|\iota_\beta(F)\|_\infty = \|\mathbf{F}\|_\infty$. Recall that by assumption, $\mathrm{Pol}_k \subset \mathcal{D}_{X,A}$. Since $F \in \mathrm{Pol}_k$, we get that M^F is a local martingale with $M_0^F = 0$. Let $\tilde{T}_n$ be a localizing sequence of M^F. Define the stopping times

$$\hat{T}_n = \{t \ge 0 : F(X_t) > n\}.$$

Since X is càdlàg and F continuous, we get $F(X_{(t\wedge\hat{T}_n)-}) \le n$. Because $(M^F)^{T_n}$ is a uniformly integrable martingale, by Doob's optional stopping theorem we have that for $T_n := \tilde{T}_n \wedge \hat{T}_n$, $(M^F)^{T_n}$ is a uniformly integrable martingale, implying T_n is a localizing sequence of M^F. Let $T > t$. Then, since $\mathbb{E}\left[M^F_{t\wedge T_n}\right] = \mathbb{E}\left[M^F_0\right] = 0$, we get

$$\begin{aligned}\mathbb{E}\left[F(X_{t\wedge T_n})\right] &= \mathbb{E}\left[F(X_0)\right] + \mathbb{E}\left[\int_{(0,t\wedge T_n]} \mathcal{G}F(u,X_{u-})A(du)\right]\\ &\le \mathbb{E}\left[F(X_0)\right] + \mathbb{E}\left[\int_{(0,t\wedge T_n]} |\ \mathcal{G}F(u,X_{u-})|A(du)\right]\\ &\le \mathbb{E}\left[F(X_0)\right] + \|G\|_{dA,t}\, Nc_\infty K_\lambda \mathbb{E}\left[\int_{(0,t\wedge T_n]} F(X_{u-})A(du)\right]\\ &\le \mathbb{E}\left[F(X_0)\right] + K_T^F\,\mathbb{E}\left[\int_{(0,t]} F(X_{(u\wedge T_n)-})A(du)\right],\end{aligned}$$

where in the last step we use that $F \ge 0$ and K_T^F is the constant from (i). Since $dA((0,t]) < \infty$ and $T_n = \tilde{T}_n \wedge \hat{T}_n$, we have for all $u \in (0,t]$ that $0 \le F(X_{(u\wedge T_n)-}) \le n$, which justifies an application of the Fubini theorem, yielding

$$\mathbb{E}\left[F(X_{t\wedge T_n})\right] \le \mathbb{E}\left[F(X_0)\right] + K_T^F \int_{(0,t]} \mathbb{E}\left[F(X_{(u\wedge T_n)-})\right] A(du).$$

By assumption, $\mathbb{E}\left[\|X_0\|_{\mathbb{R}^d}^k\right] < \infty$ which by Lemma 6.1 yields $\mathbb{E}[F(X_0)] < \infty$. We can therefore apply Gronwall's inequality (Theorem 4.55) which gives us for any $0 \leq t \leq T$,

$$\mathbb{E}[F(X_{t\wedge T_n})] \leq \mathbb{E}[F(X_0)] \exp\left(K_T^F\, A(t)\ \right).$$

Let now $g \in \mathrm{Pol}_k$ be arbitrary and denote by K_g the constant in Equation (6.29) for g. By the lemma of Fatou we now have

$$\begin{aligned}
\mathbb{E}[|g(X_t)|] &\leq K_g\, \mathbb{E}[F(X_t)] = K_g\, \mathbb{E}\left[\lim_{n\to\infty} F(X_{t\wedge T_n})\right] \\
&\leq K_g\, \liminf_{n\to\infty} \mathbb{E}[F(X_{t\wedge T_n})] \\
&\leq K_g\, \mathbb{E}[F(X_0)] \exp\left(K_T^F\ A(T)\ \right) < \infty.
\end{aligned} \tag{6.30}$$

Since the upper bound does not depend on t and g was arbitrary, this yields

$$\sup_{0\leq t\leq T} \mathbb{E}[|f(X_t)|] < \infty \text{ for all } f \in \mathrm{Pol}_k, \tag{6.31}$$

as claimed. Let us now show (iii). Define the process $g(u) = (\tilde{\mathcal{G}}f)(u, X_{u-})$. By (i), there is a constant K_t such that we have

$$\operatorname*{ess\,sup}_{u\in[0,t],dA} |\mathcal{G}f(u,x)| \leq K_t F(x).$$

Hence we have

$$\begin{aligned}
(|g| \bullet A)_t &= \int_{(0,t]} |g(u)| A(du) = \int_{(0,t]} |\mathcal{G}f(u, X_{u-})| A(du) \\
&\leq K_t \int_{(0,t]} F(X_{u-}) A(du) \leq K_t \left(\sup_{u\in[0,t]} F(X_u)\right) A(t).
\end{aligned}$$

Since F is continuous and X is a càdlàg process, we get that the supremum above is finite for almost all $\omega \in \Omega$. Since A is càdlàg , $A(t)$ is finite for all $t \geq 0$. Hence we have shown $(|g| \bullet A)_t(\omega) < \infty$ for all $t \geq 0$ for all ω outside a null set. By Proposition I.3.5 in [JS13] to positive and negative part of g separately we now get

$$g^+ \bullet A \in \mathcal{V}^+ \text{ and } g^- \bullet A \in \mathcal{V}^+, \tag{6.32}$$

hence $g \bullet A \in \mathcal{V}$ and by that càdlàg , where we have that g^+ and g^- are adapted because X is adapted.

We are left to show that $g \bullet A$ is predictable. Because $g(u) = (\mathcal{G}f_i)(u, X_{u-})$ where $(\mathcal{G}f_i)(u,\cdot)$ is deterministic and hence predictable, and X_{u-} is càglad and therefore predictable, we get that g is a predictable process. Similarly, positive and negative parts of g are also predictable since $(\mathcal{G}f)(u,\cdot)^{\pm}$ are deterministic. This shows that again by the proposition cited above, that

$$(g \bullet A)_\cdot = \int_{(0,\cdot]} \mathcal{G}f(u, X_{u-}) dA_u \tag{6.33}$$

is a predictable finite variation process. By definition, this process starts at zero which implies special semimartingality. $\square$

We will in the sequel require to be able to integrate over paths of a process given by conditional expectations, i.e. for $t \geq s$,

$$t \mapsto \mathbb{E}\left[\mathbf{H}_{\beta^k}(X_t) \mid \mathcal{F}_s\right].$$

By linearity this is equivalent to being able to integrate over paths $t \mapsto \mathbb{E}\left[f(X_t) \mid \mathcal{F}_s\right]$ for all $f \in \mathrm{Pol}_k$. Since conditional expectations are defined up to a null set and the uncountable union of null sets can have any measure between zero and one (it can even be non-measurable), the process above is a priory not well defined.[1] If however, two modifications of the process above exist and they are both càd, then they are indistinguishable, c.f. [Pro05, Theorem I.2]. Hence, if we can show existence of a càd modification of the process, then we can always pick such a version and integration over paths are well defined. In case of càdlàg k-polynomial processes, the situation is by assumption already easy since a closer look reveals that the existence of a càdlàg modification is part of the definition of the process. This leads to the following lemma.

Lemma 6.38. *Let (X, τ) be a càdlàg k-polynomial process. Then, there exists a unique (up to indistinguishability) càdlàg modification of the process*

$$[s, \infty) \ni t \mapsto \mathbb{E}\left[\mathbf{H}_{\beta^k}(X_t) \mid \mathcal{F}_s\right].$$

Proof. Let $f \in \beta$ be any basis polynomial of Pol_k. Hence, for all $t \geq s$, we have

$$\mathbb{E}\left[f(X_t) \mid \mathcal{F}_s\right] = \tau_{s,t} f(X_s) \quad [P].$$

Hence, $\tau_{s,\cdot} f(X_s)$ has càdlàg paths on $[s, \infty)$ for ***all*** $\omega \in \Omega$ since τ is càdlàg and deterministic. Hence it is a modification of $\mathbb{E}\left[f(X_\cdot) \mid \mathcal{F}_s\right]$, which is càdlàg and therefore unique up to indistinguishability, compare Theorem A.3 in the appendix. $\square$

We can now establish a link between the matrix-evolution-system induced by a regular k-polynomial process and the extended generator of the process. We will see that the extended generator restricted to Pol_k is in fact given by the extended generator of the matrix-evolution-system. That will allow us to relate the mapping τ of a k-polynomial process (X, τ) to the extended generator of the process X.

The next theorem is the first part of our version of [Cuc+12, Theorem 2.7].

Theorem 6.39. *Let (X, P, A, G) be a bounded and uniformly integrable k-polynomial process for $k \geq 0$. Then $\mathrm{Pol}_k \subset \mathcal{D}_{X,A}$ and $\mathcal{G}$ defined by $G \overset{\iota_k}{\to} \mathcal{G}$ is a version of $\mathcal{G}_{\mathcal{X},A}$ on Pol_k. Further, for all $f \in \mathrm{Pol}_k$, the process M^f defined by*

$$M_t^f = f(X_t) - f(X_0) - \int_{(0,t]} \mathcal{G} f(u, X_{u-}) A(du)$$

is a true martingale. If $k > 0$, then X is a special semimartingale. If G preserves the degree on Pol_k, then $\mathcal{G}(\mathrm{Pol}_l) \subset \mathrm{Pol}_l$ for all $l \leq k$.

[1] Compare the remarks made in the very beginning of Protter [Pro05, Chapter I.1]

Proof. For $k = 0$, all $f \in \mathrm{Pol}_k$ are constant polynomials and the statement is obviously true. Hence, assume $k \geq 1$. We first show for arbitrary $f \in \mathrm{Pol}_k$

$$f(X_t) - f(X_0) - \int_{(0,t]} \mathcal{G}f(u, X_{u-})A(du) \tag{6.34}$$

is a true martingale, which implies $\mathrm{Pol}_k \subset \mathcal{D}_{X,A}$ and $\mathcal{G}$ is a version of $\mathcal{G}_{X,A}$ on Pol_k. Note that since (X, P, A, G) is uniformly integrable, by definition the sets $\mathcal{H}_t^f$ are uniformly integrable for all $f \in Pol_k$ and $t \geq 0$ and $\|G\|_{dA,t} < \infty$ for all $t \geq 0$ holds by the boundedness assumption. By Lemma 6.38, there exists a unique càdlàg modification of $[s, \infty) \ni t \mapsto \mathbb{E}[\mathbf{H}_\beta(X_t) \mid \mathcal{F}_s]$, hence we shall always pick this modification. Further, by uniform integrability of the sets $\mathcal{H}^{\beta_i}$ for all $i \in \{1, \dots, N\}$, Lemma 6.29 (i) implies that the left limit in each point of this modification is given by $\mathbb{E}[\mathbf{H}_\beta(X_{u-}) \mid \mathcal{F}_s]$, i.e in particular we have for all $f \in \mathrm{Pol}_k$,

$$\mathbb{E}[f(X_{t-}) \mid \mathcal{F}_s] = (P(s, t-)\mathbf{f})^\top \mathbf{H}_{\beta^k}(X_s).$$

We start with integrability. We compute

$$\begin{aligned}
\int_{(s,t]} \mathbb{E}\left[\, |\mathcal{G}f(u, X_{u-})| \,\right] A(du) &= \int_{(s,t]} \mathbb{E}\left[\, \left|(G(u)\mathbf{f})^\top \mathbf{H}_\beta(X_{u-})\right| \,\right] A(du) \\
&\leq \|G\|_{dA,t} \|\mathbf{f}\|_1 \int_{(s,t]} \mathbb{E}\left[\, \|\mathbf{H}_\beta(X_{u-})\|_1 \,\right] A(du) \\
&= \|G\|_{dA,t} \|\mathbf{f}\|_1 \int_{(s,t]} \mathbb{E}\left[\sum_{i=1}^N |\beta_i(X_{u-})| \right] A(du) \\
&= \|G\|_{dA,t} \|\mathbf{f}\|_1 \sum_{i=1}^N \int_{(s,t]} \mathbb{E}\left[\, |\beta_i(X_{u-})| \,\right] A(du) \\
&\leq \|G\|_{dA,t} \|\mathbf{f}\|_1 A(t) \sum_{i=1}^N \sup_{0<u\leq t} \mathbb{E}\left[\, |\beta_i(X_{u-})| \,\right].
\end{aligned} \tag{6.35}$$

Since $\beta_i \in \mathrm{Pol}_k$, we have that $\mathcal{H}_t^{\beta_i}$ is uniformly integrable. Hence, by Lemma 3.5, the sum is finite. By Fubini-Tonelli, we get

$$\int_{(s,t]} \mathbb{E}\left[\, |\mathcal{G}f(u, X_{u-})| \,\right] A(du) = \mathbb{E}\left[\int_{(s,t]} |\mathcal{G}f(u, X_{u-})| A(du) \right] < \infty.$$

Now, since $f \in \mathrm{Pol}_k$, we get again by Lemma 3.5 that $\mathbb{E}[|f(X_t)|]$ and $\mathbb{E}[|f(X_0)|]$ are finite. Hence,

$$\mathbb{E}\left[|M_t^f|\right] \leq \mathbb{E}[|f(X_t)|] + \mathbb{E}[|f(X_0)|] + \mathbb{E}\left[\int_{(s,t]} |\mathcal{G}f(u, X_{u-})| A(du) \right] < \infty.$$

Next we show the martingale property. Let $0 \leq s \leq t$. Then,

$$\begin{aligned}\mathbb{E}\left[M_t^f \mid \mathcal{F}_s\right] &= \mathbb{E}\left[f(X_t) - f(X_0) - \int_{(0,t]} (G(u)\mathbf{f})^\top \mathbf{H}_\beta(X_{u-})A(du)|\mathcal{F}_s\right] = \\ &= \mathbb{E}\left[f(X_t)|\mathcal{F}_s\right] - f(X_0) - \mathbb{E}\left[\int_{(0,t]} (G(u)\mathbf{f})^\top \mathbf{H}_\beta(X_{u-})A(du)|\mathcal{F}_s\right].\end{aligned} \tag{6.36}$$

Recalling, that the càdlàg MES P represents τ wrt. the basis β, and an application of the Fubini-Tonelli theorem for conditional expectations (c.f. Schilling [Sch17, Theorem 27.17]) that we already justified above yields

$$\begin{aligned}\mathbb{E}\left[M_t^f \mid \mathcal{F}_s\right] - M_s^f &= (P(s,t)\mathbf{f})^\top \mathbf{H}_\beta(X_s) - (P(s,s)\mathbf{f})^\top \mathbf{H}_\beta(X_s) \\ &\quad - \int_{(s,t]} \mathbb{E}\left[(G(u)\mathbf{f})^\top \mathbf{H}_\beta(X_{u-})|\mathcal{F}_s\right] A(du),\end{aligned} \tag{6.37}$$

where P is a MES and hence $P(s,s) = \mathrm{Id}_N$. By linearity we have that (6.37) is equal to

$$\begin{aligned}\mathbf{f}^\top \Big(& P(s,t)^\top \mathbf{H}_\beta(X_s) - P(s,s)^\top \mathbf{H}_\beta(X_s) \\ & - \int_{(s,t]} G(u)^\top \mathbb{E}\left[\mathbf{H}_\beta(X_{u-})|\mathcal{F}_s\right] A(du)\Big).\end{aligned} \tag{6.38}$$

Recalling, that for the i-th element we have $[\mathbf{H}_\beta(x)]_i = \beta_i(x)$, and hence $[\iota_\beta(\beta_i)]_j = \mathbb{1}_{\{i=j\}}$, we have

$$\mathbb{E}\left[\beta_i(X_{u-}) \mid \mathcal{F}_s\right] = (P_i)^\top \mathbf{H}_\beta(X_s),$$

where $P_i \in \mathbb{R}^N$ is the i-th column of the matrix $P(s,u-)$. This then gives us

$$\mathbb{E}\left[\mathbf{H}_\beta(X_{u-}) \mid \mathcal{F}_s\right] = P(s,u-)^\top \mathbf{H}_\beta(X_s).$$

Using this, we get that Equation (6.38) is equal to

$$\begin{aligned}&\mathbf{f}^\top \left(P(s,t)^\top \mathbf{H}_\beta(X_s) - P(s,s)^\top \mathbf{H}_\beta(X_s) - \int_{(s,t]} G(u)^\top P(s,u-)^\top \mathbf{H}_\beta(X_s) A(du)\right) \\ &= \mathbf{f}^\top \left(P(s,t)^\top - P(s,s)^\top - \int_{(s,t]} G(u)^\top P(s,u-)^\top A(du)\right) \mathbf{H}_\beta(X_s) \\ &= \mathbf{f}^\top \left(P(s,t)^\top - P(s,s)^\top - \int_{(s,t]} P(s,u-) G(u) A(du)\right)^\top \mathbf{H}_\beta(X_s) = 0,\end{aligned} \tag{6.39}$$

where we use that G is an extended generator of the MES P and hence the bracket vanishes. Hence, $t \mapsto M_t^f$ is a true martingale with respect to the filtration $\mathbb{F}$. Note that this also implies that for all $s \geq 0$, we have that $t \mapsto M^f_{s+t} - M^f_s$ is a martingale wrt. the filtration $\mathbb{F}_s = (\mathcal{F}_{t+s \geq s})_{t \geq 0}$. Indeed we have already shown that

$$\mathbb{E}\left[\left|M^f_{s+t} - M^f_s\right|\right] < \infty$$

by applying the triangle inequality. We are left to show that X is a special semimartingale.

To see that X is indeed a special semimartingale, define $f_i(x) = x_i$, the projection to the i-th coordinate. Since X is a special semimartingale on $\mathbb{R}^d$ if and only if each component is a special semimartingale, we need to show that for all $i \in \{1, \dots, d\}$ we have that $f_i(X)$ is special. Since $f_i \in \mathrm{Pol}_k$ we have $f_i \in \mathcal{D}_{X,A}$. This implies that $f_i(X)$ has the decomposition

$$f_i(X_0) + \left(f_i(X) - f_i(X_0) - \int_{(0,\cdot]} (\tilde{\mathcal{G}} f_i)(u, X_{u-}) A(du) \right) + \int_{(0,\cdot]} (\tilde{\mathcal{G}} f_i)(u, X_{u-}) A(du). \tag{6.40}$$

and we are left to show (for the semimartingale property) that the integral on the right defines a process belonging to $\mathcal{V}$. That means we need to show that it is a càdlàg process starting from zero which is adapted and of finite variation. If we then show that the process is also predictable, we get that $f_i(X)$ is a special semimartingale for all $i = 1, \dots, d$. But that is precisely the statement of Lemma 6.37 (iii) for $f = f_i$. Finally, if now G is degree preserving on Pol_k, then $\mathcal{G}(\mathrm{Pol}_l) \subset \mathrm{Pol}_l$ for all $l \leq k$ by the way $\mathcal{G}$ was defined via G. □

Remark 6.40. *In the proof above, we can see that the uniform integrability assumption for the sets $\mathcal{H}^{\beta_i}_t$ for all $t \geq 0$ would imply by Equation* (6.35)*, that*

$$\mathbb{E}\left[\mathrm{Var}\left((g \bullet A)_t\right)\right] = \mathbb{E}\left[(|g| \bullet A)_t\right] < \infty,$$

implying $(g \bullet A) \in \mathcal{A}_{loc}$. The special semimartingale property would then alternatively follow from Jacod and Shiryaev [JS13, Proposition I.4.23].

We will next give the reverse statement of the previous theorem, finishing our version of [Cuc+12, Theorem 2.7]. Note that in [Cuc+12], the authors prove an intermediary result. Since they consider a homogeneous setting, they only need to assume continuity of the semigroup to deduce that the representing MSG has an infinitesimal generator, c.f. Theorem 4.3. In our setting, for the one direction, we needed to a priory assume the existence of an extended generator since we do not have a general existence result under our regularity assumptions on τ, even though the existence follows already under very mild assumptions, c.f. Theorem 4.35. In particular, since by definition the MES P belonging to a càdlàg k-polynomial process (X, P) preserves the degree, Theorem 5.18 yields that this generator can be chosen to be degree preserving. The situation

Theorem 6.41. *Consider a càdlàg process X on our stochastic basis and let $k \in \mathbb{N}_0$. Let $A \in \mathcal{V}^+_{\mathscr{A}}$. Assume*

(i) $\mathrm{Pol}_k \subset \mathcal{D}_{X,A}$ *and there exists an on* Pol_k *degree preserving measurable function* $G : \mathbb{R}_{>0} \to M_{N_k}(\mathbb{R})$*, such that* $\mathcal{G}$ *defined by* $G \overset{\iota_k}{\to} \mathcal{G}$ *is a version of* $\mathcal{G}_{X,A}$ *on* Pol_k *with* $\|G\|_{dA,t} < \infty$ *for all* $t \geq 0$*,*

(ii) for any $f \in \mathrm{Pol}_k$*, the sets* $\mathcal{H}_t^f$ *are uniformly integrable for all* $t \geq 0$ *and the process* M^f *defined by*

$$M_t^f = f(X_t) - f(X_0) - \int_{(0,t]} \mathcal{G}f(u, X_{u-})A(du)$$

is a true martingale.

Then, there exists a regular MES P such that G is the extended A-generator of P, and (X, P, A, G) is a bounded and uniformly integrable k-polynomial process. In particular, if $k > 0$, then X is a special semimartingale.

Proof. Let us first argue that there exists a càdlàg modification of the process

$$[s, \infty) \ni t \mapsto \mathbb{E}\left[\mathbf{H}_{\beta^k}(X_t) \mid \mathcal{F}_s\right] =: F_s(t).$$

By assumption (ii), for all $i \leq N_k$, the process $[s, \infty) \ni t \mapsto \beta_i(X_t)$ has the decomposition

$$\beta_i(X_t) = \beta_i(X_0) + M_t^f + Z_t,$$

where M_t^f is a true martingale and $Z_t = (g \bullet A)_t$ for $g(u) = \mathcal{G}\beta_i(u, X_{u-})$. Since by assumption, all sets $\mathcal{H}_t^{\beta_i}$ are uniformly integrable, we get

$$\mathbb{E}\left[\mathrm{Var}\left((g \bullet A)_t\right)\right] = \mathbb{E}\left[(|g| \bullet A)_t\right] < \infty$$

for all t, implying $Z \in \mathcal{A}_{\mathrm{loc}}$, compare the computations in Equation (6.35) that are valid here, including the application of the theorem of Fubini-Tonelli. Hence, the process

$$[s, \infty) \ni t \mapsto \mathbb{E}\left[\beta_i(X_t) \mid \mathcal{F}_s\right]$$

has a càdlàg modification. Hence, $(F_s(t))_i$ has a càdlàg modification for all $i \leq N_k$, yielding the existence of a càdlàg modification of $F_s(t)$. For the rest of this proof, we always pick such a càdlàg modification without further mentioning it. Let $f \in \mathrm{Pol}_k$ be arbitrary. Hence, $\mathcal{H}_t^f$ is uniformly integrable for all $t \geq 0$ by assumption and by Lemma 3.5, we get $\mathbb{E}\left[|f(X_t)|\right] < \infty$ for all $t \geq 0$. We next show the polynomial property. Since we picked the càdlàg version of $F_s(t)$, we have that $F_s(t-)$ exists a.s. and by the uniform integrability assumption, Lemma 3.5 yields $\lim_{\varepsilon \downarrow 0} F_s(t - \varepsilon)$ converges as a $\mathcal{L}^1$-limit to $\mathbb{E}\left[\mathbf{H}_{\beta^k}(X_{t-}) \mid \mathcal{F}_s\right]$. With Proposition 6.26, this gives us $F_s(t-) = \mathbb{E}\left[\mathbf{H}_{\beta^k}(X_{t-}) \mid \mathcal{F}_s\right]$.

Note that we have

$$\mathbb{E}\left[f(X_t) \mid \mathcal{F}_s\right] = \mathbb{E}\left[\mathbf{f}^\top \mathbf{H}_{\beta^k}(X_t) \mid \mathcal{F}_s\right] = \mathbf{f}^\top F_s(t).$$

The true martingale assumption yields

$$\mathbb{E}\left[f(X_t) \mid \mathcal{F}_s\right] = f(X_s) + \mathbb{E}\left[\int_{(s,t]} \mathcal{G}f(u, X_{u-})A(du)\right].$$

Again the same computations in Equation (6.35), combined with the uniform integrability assumption justifies an application of the Fubini-Tonelli theorem for the conditional expectation, yielding

$$\begin{aligned}
\mathbf{f}^\top F_s(t) = \mathbb{E}\left[f(X_t) \mid \mathcal{F}_s\right] &= f(X_s) + \int_{(s,t]} \mathbb{E}\left[\mathcal{G}f(u, X_{u-}) \mid \mathcal{F}_s\right] A(du) \\
&= \mathbf{f}^\top \mathbf{H}_{\beta^k}(X_s) + \int_{(s,t]} \mathbb{E}\left[(G(u)\mathbf{f})^\top \mathbf{H}_{\beta^k}(X_{u-}) \mid \mathcal{F}_s\right] A(du) \\
&= \mathbf{f}^\top F_s(s) + \int_{(s,t]} (G(u)\mathbf{f})^\top F_s(u-)A(du) \\
&= \mathbf{f}^\top \left(F_s(s) + \int_{(s,t]} G(u)^\top F_s(u-)A(du)\right),
\end{aligned}$$

where we used $F_s(s) = \mathbf{H}_{\beta^k}(X_s)$. All these equalities are to be understood in an a.s. way. Since the above has to hold for all $\mathbf{f} \in \mathbb{R}^{N_k}$, we get that $F_s(t)$ solves (MDE:A, V) for $V(u, x) = G(u)^\top x$. By Proposition 4.53, solutions exist for all initial values $w \in \mathbb{R}^{N_k}$ and define a fundamental solution $M(s,t)$ for $s \le t$. Hence we have $F_s(t) = M(s,t)\mathbf{H}_{\beta^k}(X_s)$. Define $P = M^\top$. Then by the same theorem, we know that P is a regular MES such that (P, γ, A) is a proper MES with extended generator G. Since G is degree preserving on Pol_k by assumption, Theorem 5.19 yields that P preserves the degree on Pol_k. Hence we get

$$\mathbb{E}\left[f(X_t) \mid \mathcal{F}_s\right] = (P(s,t)\mathbf{f})^\top \mathbf{H}_{\beta^k}(X_s),$$

with $\iota_k^{-1}(P(s,t)\mathbf{f}) \in \mathrm{Pol}_l$ for $f \in \mathrm{Pol}_l$. Since this holds for all $l \le k$, and since $\|G\|_{dA,t} < \infty$ for all $t \ge 0$ and $\mathcal{H}_t^f$ uniformly integrable for all $f \in \mathrm{Pol}_k$ and $t \ge 0$, was part of the assumption, we get that (X, P, A, G) is a bounded and uniformly integrable k-polynomial process. If $k > 0$, Theorem 6.39 yields that X is a special semimartingale, finishing the proof. $\square$

Remark 6.42. *In the theorem above, assumption* (ii) *requires* M_t^f *to be a true martingale and the sets* $\mathcal{H}_t^f$ *need to be uniformly integrable for all* $f \in \mathrm{Pol}_k$*. In light of Lemma 6.37, in the case where k is an even number, one could "sacrifice" one degree to achieve uniform integrability of all* $\mathcal{H}_t^f$ *for* $f \in \mathrm{Pol}_{k-1}$ *by means of Theorem 3.4. This is shown for illustration in the appendix, see Proposition B.1. However, a stronger result is possible by adapting the ideas presented in Cuchiero et al. [Cuc+12, Theorem 2.10]. We will see* *that the assumption (ii) above is implied by the first, in case k is an even number. Just as in [Cuc+12], it will be convenient to formulate the following condition.*

Condition 6.43. *Let $(\Omega, \mathscr{F}, \mathbb{F}, P)$ be our stochastic basis and X a càdlàg and adapted process. Let $A \in \mathcal{V}^+_{\mathfrak{A}}$. Then $\mathrm{Pol}_k \subset \mathcal{D}_{X,A}$ and there exists a measurable function $G : \mathbb{R}_{>0} \to M_{N_k}(\mathbb{R})$ that is degree preserving on Pol_k such that $\mathcal{G}$ defined by $G \overset{\iota_k}{\to} \mathcal{G}$ is a version of $\mathcal{G}_{X,A}$ on Pol_k and $\|G\|_{dA,t} < \infty$ for all $t \geq 0$.*

The goal is now to proof the next theorem that is our version of [Cuc+12, Theorem 2.10].

Theorem 6.44. *Let $k \geq 0$ be an even number and X a càdlàg and adapted process with $\mathbb{E}\left[\|X_0\|^k_{\mathbb{R}^d}\right] < \infty$. Then there exists a proper MES (P, γ, A) in dimension N_k with extended generator G that preserves the degree on Pol_k, such that (X, P, A, G) is a bounded and uniformly integrable k-polynomial process if and only if X satisfies Condition 6.43 for k and $A \in \mathcal{V}^+_{\mathfrak{A}}$.*

The condition $k \neq 1$ is necessary as demonstrated in [Cuc+12, Remark 2.11 (*ii*)]. There a counterexample for $k = 1$ is provided. The case $k = 0$ is a trivial statement. As an intermediary result, let us first proof the following lemma.

Lemma 6.45. *Let $k \geq 2$ be an even number and assume X satisfies Condition 6.43 for k and $A \in \mathcal{V}^+_{\mathfrak{A}}$. Assume further $\mathbb{E}\left[\|X_0\|^k_{\mathbb{R}^d}\right] < \infty$. Then $f(X)^*_t \in \mathcal{L}^1$ for all $f \in \mathrm{Pol}_k$, $t \geq 0$ and M^f_t is a true martingale for all $f \in \mathrm{Pol}_k$.*

Proof. We follow Cuchiero et al. [Cuc+12] and show $\mathbb{E}\left[(M^f)^*_t\right] < \infty$ for all $t \geq 0$. Then by Lemma 3.6 we get that M^f is a true martingale as claimed. In the same way, we show for $F \in \mathrm{Pol}_k$ as defined in Lemma 6.37, that $\mathbb{E}\left[F(X)^*_t\right] < \infty$. This will then imply that $\mathbb{E}\left[f(X)^*_t\right] < \infty$ for all $f \in \mathrm{Pol}_k$, implying that the sets $\mathcal{H}^f_t$ are all uniformly integrable. Note that all conditions of Lemma 6.37 are satisfied. As in that lemma, define $F \in \mathrm{Pol}_k$ with

$$1 \leq F(x) = 1 + \sum_{i=1}^{d} f_i(x)^k.$$

Further, recall the definition of a function $|f|(x)$ for a polynomial $f \in \mathrm{Pol}_k$, compare Equation (6.29). We first consider the case $f \in \mathrm{Pol}_l$ for $l < k$. Let $p := \frac{k}{l} > 1$. Then for all $f \in \mathrm{Pol}_l$, there exists a $g \in \mathrm{Pol}_k$ such that

$$|f(x)|^p \leq |f|(x)^p = |g|(x),$$

and by (6.29) we get

$$|f(x)|^p \leq CF(x),$$

for some constant $C > 0$ depending on f. In the same way, we compute

$$\operatorname*{ess\,sup}_{u\in[0,T],\,dA} |\mathcal{G}f(u,x)|^p = \operatorname*{ess\,sup}_{u\in[0,T],\,dA} \left|(G(u)\mathbf{f})^\top \mathbf{H}_{\beta^k}(x)\right|^p.$$

Since G is degree preserving, we have recalling the definition of the matrix L_l from the previous chapter,

$$\begin{aligned}\operatorname*{ess\,sup}_{u\in[0,T],\,dA} \left|(G(u)\mathbf{f})^\top \mathbf{H}_{\beta^k}(x)\right|^p &= \operatorname*{ess\,sup}_{u\in[0,T],\,dA} \left|(G(u)\mathbf{f})^\top L_l \mathbf{H}_{\beta^k}(x)\right|^p \\ &= \|G\|^p_{dA,T}\, \|\mathbf{f}\|_1^p\, N_l^{p-1} \sum_{i=1}^{N_l} |\beta_i(x)|^p,\end{aligned}$$

where in the last step we used Jensen's inequality for sums and that

$$(L_l \mathbf{H}_{\beta^k}(x))_i = \beta_i(x) \mathbb{1}_{\{i\le N_l\}}.$$

Since for $i \le N_l$, $\beta_i \in \mathrm{Pol}_l$, we again get that there exists a constant $C > 0$, depending on f such that

$$\operatorname*{ess\,sup}_{u\in[0,T],\,dA} |\mathcal{G}f(u,x)|^p \le CF(x). \tag{6.41}$$

Recall from the proof in Lemma 6.37, that we can choose a localizing sequence $(T_n)_n$ of M^f such that $0 \le F(X_{(t\wedge T_n)-}) \le n$. We compute using the previous estimates

$$\begin{aligned}\left|M^f_{t\wedge T_n}\right|^p &= \left|f(X_{t\wedge T_n}) - f(X_0) - \int_{(0,t\wedge T_n]} \mathcal{G}f(u, X_{u-})A(du)\right|^p \\ &\le C\left(F(X_{t\wedge T_n}) + F(X_0) + \int_{(0,t\wedge T_n]} F(X_{u-})A(du)\right) \\ &\le C\left(F(X_{t\wedge T_n}) + F(X_0) + \int_{(0,t]} F(X_{(u\wedge T_n)-})A(du)\right)\end{aligned}$$

For some constant $C > 0$ and $t \le T$. In the last step we used that $F > 0$. From the proof of Lemma 6.37, compare Equation (6.30), we have with $g = F$ that there exists a constant $K > 0$ such that

$$\sup_{t\le T} \mathbb{E}\left[F(X_{(t\wedge T_n)-})\right] \le \sup_{t\le T} \mathbb{E}\left[F(X_t)\right] \le K\mathbb{E}\left[F(X_0)\right] \exp\left[KA(T)\right]. \tag{6.42}$$

By assumption, $\mathbb{E}\left[\|X_0\|^k_{\mathbb{R}^d}\right] < \infty$, hence Lemma 6.1 implies $\mathbb{E}\left[F(X_0)\right] < \infty$. We further have

$$\begin{aligned}&\int_{(0,t]} \mathbb{E}\left[|F(X_{(u\wedge T_n)-})|\right] A(du) = \int_{(0,t]} \mathbb{E}\left[F(X_{(u\wedge T_n)-})\right] A(du) \\ &\le K\mathbb{E}\left[F(X_0)\right] \exp\left[KA(T)\right] A(T) < \infty.\end{aligned} \tag{6.43}$$

Note that in the last step we used (6.42) in combination with the standard estimate for the dA integral. This justifies an application of the theorem of Fubini-Tonelli, overall yielding

$$\mathbb{E}\left[\left|M^f_{t\wedge T_n}\right|^p\right] \le C, \tag{6.44}$$

where the constant $C > 0$ does depend on f but not on n. We continue with Doob's maximal L^p-inequality for $p = \frac{k}{l} > 1$, yielding for all n, that there exists a constant $\tilde{K} < \infty$ independent of n such that

$$\mathbb{E}\left[\sup_{t\le T}\left|M^f_{t\wedge T_n}\right|\right] \le \tilde{K}\mathbb{E}\left[\left|M^f_{T\wedge T_n}\right|^p\right] \le \tilde{K}C,$$

where C is from (6.44). The lhs is increasing in n, hence by monotone convergence we get $(M^f)^*_T \in \mathcal{L}^p$ for all $f \in \mathrm{Pol}_{k-1}$ and $T \ge 0$. Since $p > 1$, this then implies the desired

$$\mathbb{E}\left[\sup_{t\le T}\left|M^f_t\right|\right] < \infty.$$

Let us now consider the case $l = k$. Let $q = k/2$, which is an integer by assumption. Let $f \in \mathrm{Pol}_q$ be defined by $f(x) = f_i(x)^q$ for $i \in \{1, \dots, d\}$. We compute, using Jensen's inequality for sums and for integrals, where with $C < \infty$ we denote an appropriate constant, possibly changing from line to line,

$$\begin{aligned}
f(X_t)^2 &= \left(M^f_t + f(X_0) + \int_{(0,t]} \mathcal{G}f(u, X_{u-})A(du)\right)^2 \\
&\le C\left((M^f_t)^2 + f(X_0)^2 + \int_{(0,t]} |\mathcal{G}f(u, X_{u-})|^2 A(du)\right) \\
&\le C\left((M^f_t)^2 + f(X_0)^2 + \int_{(0,t]} F(X_{u-})A(du)\right),
\end{aligned}$$

where for the last step we used (6.41). Since $F \ge 0$, the integral is increasing in t, yielding

$$\sup_{t\le T} f(X_t)^2 \le C\left(\sup_{t\le T}(M^f_t)^2 + f(X_0)^2 + \int_{(0,T]} F(X_{u-})A(du)\right).$$

Since $f^2 \in \mathrm{Pol}_k$, the assumption $\mathbb{E}\left[\|X_0\|^k_{\mathbb{R}^d}\right] < \infty$ implies $\mathbb{E}\left[f(X_0)^2\right] < \infty$. We have already shown above that from (6.43) and Fubini-Tonelli we get

$$\mathbb{E}\left[\int_{(0,T]} F(X_{u-})A(du)\right] < \infty.$$

Since $f \in \mathrm{Pol}_q$, we have $p = k/q = 2$ and by the previous result, we get $(M^f)_t^* \in \mathcal{L}^2$, hence we have

$$\mathbb{E}\left[\sup_{t\le T} f(X_t)^2\right] < \infty.$$

Noting that $F(x) = 1 + \sum_{i=1}^d f_i(x)^k$, and $f(x)^2 = f_i(x)^k$, summing over $i \in \{1,\dots,d\}$ yields

$$\mathbb{E}\left[F(X)_T^*\right] = \mathbb{E}\left[\sup_{t\le T} F(X_t)\right] < \infty. \tag{6.45}$$

From $|f(x)| \le CF(x)$ for some constant C we conclude

$$\mathbb{E}\left[f(X)_T^*\right] = \mathbb{E}\left[\sup_{t\le T} |f(X_t)|\right] \le C\mathbb{E}\left[F(X)_T^*\right] < \infty,$$

hence $f(X)_t^* \in \mathcal{L}^1$ for all $f \in \mathrm{Pol}_k$, $t \ge 0$ as claimed. For $f \in \mathrm{Pol}_k$ arbitrary, we have by Lemma 6.37 (i) and (ii) that

$$\begin{aligned}\sup_{t\le T}\int_{(0,t]} |\mathcal{G}f(u,X_{u-})|A(du) &\le \int_{(0,T]} |\mathcal{G}f(u,X_{u-})|A(du)\\ &\le K_T^f\, A(T)\, \sup_{t\le T} F(X_{u-}),\end{aligned}$$

and we have already shown with (6.45) that the rhs is integrable. Hence we get from

$$\sup_{t\le T}|M_t^f| \le \sup_{t\le T} F(X_t) + F(X_0) + \sup_{t\le T}\int_{(0,t]} |\mathcal{G}f(u,X_{u-})|A(du),$$

that $(M^f)_T^* \in \mathcal{L}^1$ for all $f \in \mathrm{Pol}_k$ and $T \ge 0$ as claimed. By Lemma 3.6, we get M^f is a true martingale finishing the proof. □

Proof of Theorem 6.44. The one direction, that is (X,P,A,G) is a bounded and uniformly integrable k-polynomial process with G preserving the degree on Pol_k, then Condition 6.43 holds is the statement of Theorem 6.39 for all $k \in \mathbb{N}_0$, not just for even ones. For the reverse direction, the case $k = 0$ is a trivial statement since all càdlàg adapted processes have the 0-polynomial property with $\tau_{s,t} \equiv 1$ and $\mathcal{H}_t^f$ for $f \in \mathrm{Pol}_0$ contains only one real number, hence is uniformly integrable. For the case $k \ge 2$, we have by Lemma 6.45 that $f(X)_t^* \in \mathcal{L}^1$ for all $f \in \mathrm{Pol}_k$. Further, X is càdlàg and since polynomials are continuous, we get that for almost all $\omega \in \Omega$,

$$\sup_{s\le t} |f(X_s(\omega))| < \infty.$$

Fix $\mathcal{H}_T^f$. The goal is to show uniform integrability. We have for $n \in \mathbb{N}_0$,

$$\sup_{t\le T}\mathbb{E}\left[f(X_t)\mathbb{1}_{\{|f(X_t)|>n+1\}}\right] \le \mathbb{E}\left[\sup_{t\le T} f(X_t)\mathbb{1}_{\{|f(X_t)|>n\}}\right] \le \mathbb{E}\left[f(X)_T^*\right] < \infty.$$

By the dominated convergence theorem with dominating r.v. $f(X)^*_T$, we have for all $f \leq T$,

$$\begin{aligned} 0 &= \mathbb{E}\left[\lim_{n\to\infty} f(X)^*_T \mathbb{1}_{\{f(X)^*_T > n\}}\right] = \lim_{n\to\infty} \mathbb{E}\left[f(X)^*_T \mathbb{1}_{\{|f(X)^*_T| > n\}}\right] \\ &\geq \lim_{n\to\infty} \mathbb{E}\left[f(X_t)\mathbb{1}_{\{|f(X_t)|>n\}}\right]. \end{aligned}$$

Since the upper bound goes to zero and is independent of $t \leq T$, we get that $\mathcal{H}^f_T$ is indeed uniformly integrable. Further, by Lemma 6.45 we have that M^f is a true martingale for all $f \in \mathrm{Pol}_k$. Hence, with the rest of Condition 6.43, we get that the assumptions of Theorem 6.41 hold, which finishes the proof. □

Remark 6.46. *Note that when defining the k-polynomial property, we follow the authors Cuchiero et al. [Cuc+12] and require P to preserve the degree. We have seen that this makes it reasonable to assume that the extended generator also preserves the degree, compare in particular Theorem 5.18 and Theorem 5.19. This property was crucial in Lemma 6.45 and indeed, by means of [Cuc+12, Remark 2.11 (iii)], there is a counterexample where a process X has an extended generator whose representing matrix G, while existing, does not preserve the degree and despite all other assumptions being satisfied, the process X fails to have the polynomial property.*

6.4 Polynomial Processes as Semimartingales

We have seen that k-polynomial processes for $k \geq 1$, under mild regularity assumptions are special semimartingales. In this section, we will deal with the characterization of k-polynomial processes in the context of semimartingality, i.e. we want to characterize their semimartingale characteristics. As we do not a priory assume a k-polynomial process to be Markov, the results we give are weaker then those in [Cuc+12], compare Remark 6.50.

The results we present in this section are all based on the ideas presented in Cuchiero et al. [Cuc+12] and the proofs we present are all straight forward modifications of the proofs therein. Since we consider a time inhomogeneous setting, we will need to introduce the following function space. Recall that a function $g \in \mathrm{Pol}_k(\mathbb{R}_{>0} \times \mathbb{R}^d)$ is of the form

$$\sum_{i=1}^{N_k} a_i(u)\beta_i(x),$$

where each function $a_i : \mathbb{R}_{>0} \to \mathbb{R}$ is measurable. Hence, $u \mapsto \iota_k(g(u,\cdot))$ is a measurable function. We further have

$$\|\iota_k(g(u,\cdot))\|_1 = \sum_{i=1}^{N_k} |a_i(u)|.$$

Let μ be a locally finite measure on $\mathbb{R}_{>0}$. We define

$$\mathrm{Pol}_k^{\mu}(\mathbb{R}_{>0} \times \mathbb{R}^d) := \left\{ g \in \mathrm{Pol}_k(\mathbb{R}_{>0} \times \mathbb{R}^d) : \operatorname*{ess\,sup}_{u \le t, \mu} \|\iota_k(g(u,\cdot))\|_1 < \infty, \ \forall t \ge 0 \right\}.$$

Let us motivate the introduction of $\mathrm{Pol}_k^{\mu}(\mathbb{R}_{>0} \times \mathbb{R}^d)$ with the next lemma.

Lemma 6.47. *Let $A \in \mathcal{V}^+_{\mathfrak{A}}$ and $G : \mathbb{R}_{>0} \to M_{N_k}(\mathbb{R})$ for $k \in \mathbb{N}_0$. Assume $\|G\|_{dA,t} < \infty$ for all $t \ge 0$ and that G preserves the degree on Pol_k. Define $\mathcal{G} \overset{\iota_k}{\leftarrow} G$. Then for any $l \in \mathbb{N}_0$ with $l \le k$,*

(i) if $f \in \mathrm{Pol}_l$, then

$$\mathcal{G}f \in \mathrm{Pol}_l^{dA}(\mathbb{R}_{>0} \times \mathbb{R}^d).$$

(ii) Further, for any $f \in \mathrm{Pol}_l^{dA}(\mathbb{R}_{>0} \times \mathbb{R}^d)$, there exists a constant $K > 0$ such that

$$\operatorname*{ess\,sup}_{u \le t, dA} |f(u,x)| \le K \operatorname*{ess\,sup}_{u \le t} \|\iota_l(f(u,\cdot))\|_1 \,(1 + \|x\|^l_{\mathbb{R}^d}) < \infty,$$

where K depends on f and the concrete choice of the norm $\|\cdot\|_{\mathbb{R}^d}$ but not on t .

Proof. Let us start with (i). Note that since G preserves the degree on Pol_k, we have $\mathcal{G}f \in \mathrm{Pol}_l(\mathbb{R}_{>0} \times \mathbb{R}^d)$. Further, we compute

$$\begin{aligned}\operatorname*{ess\,sup}_{u \le t, dA} \|\mathcal{G}f(u,\cdot)\|_1 &= \operatorname*{ess\,sup}_{u \le t, dA} \|G(u)\mathbf{f}\| = \operatorname*{ess\,sup}_{u \le t, dA} \|G(u)\| \, \|\mathbf{f}\|_1 \\ &\le C \, \|G\|_{dA,t} \, \|\mathbf{f}\|_1 < \infty,\end{aligned}$$

where $\tilde{C}$ is a finite constant coming from the equivalence of norms in $\mathbb{R}^{N^k}$. Hence (i) is shown and we next turn to (ii). We compute

$$\operatorname*{ess\,sup}_{u \le t, dA} |f(u,x)| = \operatorname*{ess\,sup}_{u \le t, dA} |\iota_l(f(u,\cdot))^\top \mathbf{H}_{\beta^l}(x)| \le \tilde{C} \operatorname*{ess\,sup}_{u \le t} \|\iota_l(f(u,\cdot))\|_1 \left\|\mathbf{H}_{\beta^k}(x)\right\|_1.$$

By assumption, the essential supremum on the right is finite. Further, we have

$$\left\|\mathbf{H}_{\beta^k}(x)\right\|_1 = \sum_{i=1}^{N_l} |\beta_i(x)|.$$

Since each summand has degree of at most l, there is a constant $C_i > 0$ such that $|\beta_i(x)| \le C_i(1 + \|x\|_1^l)$. Hence the claim follows by summing over finitely many sums for the $\|\cdot\|_1$ norm. The general case then follows by equivalence of norms. $\square$

Note that any $f \in \mathrm{Pol}_{k_1}$ is an element of $\mathrm{Pol}_{k_1}^{dA}(\mathbb{R}_{>0} \times \mathbb{R}^d)$, since it is constant in u. In particular, that implies for any $g \in \mathrm{Pol}_{k_2}^{dA}(\mathbb{R}_{>0} \times \mathbb{R}^d)$, that $fg \in \mathrm{Pol}_{k_1+k_2}^{dA}(\mathbb{R}_{>0} \times \mathbb{R}^d)$, where the product is the pointwise product.

Throughout the rest of this section, we denote with μ^X the random measure associated with the jumps of X, see Jacod and Shiryaev [JS13, Proposition II.1.16], given by

$$\mu^X(\omega; dt, d\xi) = \sum_{s \in \mathbb{R}_+} \mathbb{1}_{\{\Delta X_s(\omega) \neq 0\}} \varepsilon_{\{s, \Delta X_s(\omega)\}}(dt, d\xi),$$

where ε is the Dirac-distribution on $\mathbb{R}_+ \times \mathbb{R}^d$. Note that the sum is well defined since $X(\omega)$ is càdlàg for all $\omega \in \Omega$ and hence for only countably many s we have $\Delta X_s(\omega) \neq 0$.

In the sequel, for a function $f : \mathbb{R}^d \to \mathbb{R}$ with $f \in \mathcal{C}^2(\mathbb{R}^d)$ and $i, j \in \{1, \dots, d\}$, we denote with D_i and D_{ij} the partial derivatives wrt the coordinates i and j, i.e.

$$D_i f(x) = \frac{\partial}{\partial x_i} f(x), \quad D_{ij} f(x) = \frac{\partial^2}{\partial x_i \partial x_j} f(x).$$

From Lemma 6.37 (*iii*), we know that a càdlàg and adapted process X for which Condition 6.43 holds for a given $A \in \mathcal{V}^+_{\mathscr{A}}$ and $k \geq 0$ is a special semimartingale (provided $k > 0$). In the next proposition we show that the characteristics wrt the "truncation function" $h(x) = x$ satisfy certain polynomial properties. Combined with Proposition 6.54, this can be seen as our version of [Cuc+12, Proposition 2.12]. All the statements we shall make regarding characteristics are to be understood up to evanescence. Further, we assume that our stochastic basis satisfies the usual conditions, hence an arbitrary version of the characteristics is predictable, c.f. Jacod and Shiryaev [JS13, Remark II.2.8].

Set $\tilde{\Omega} = \Omega \times \mathbb{R}_+ \times \mathbb{R}^d$. We denote the predictable σ-algebra with $\mathcal{P}$ and set $\tilde{\mathcal{P}} = \mathcal{P} \otimes \mathcal{B}(\mathbb{R}^d)$, compare [JS13, p. II.1.4] where we have $E = \mathbb{R}^d$ and $\mathcal{E} = \mathcal{B}(\mathbb{R}^d)$. We call a random measure μ predictable, if for all $\tilde{\mathcal{P}}$-measurable functions W, $W * \mu$ is $\mathcal{P}$-measurable (c.f. [JS13, Definition II.1.6]).

Proposition 6.48. *Let X be a càdlàg and adapted process, $A \in \mathcal{V}^+_{\mathscr{A}}$ and $k \in \mathbb{N}$ with $k \geq 2$. Assume that Condition 6.43 holds. Then, X is a special semimartingale with semimartingale characteristics (B, C, ν) wrt the "truncation" function $h(x) = x$, such that for $i, j \in \{1, \dots, d\}$, the characteristics can be written as*

$$B_{t,i} = \int_{(0,t]} b_i(u, X_{u-}) A(du), \quad b_i \in \mathrm{Pol}_1^{dA}(\mathbb{R}_{>0} \times \mathbb{R}^d), \tag{6.46}$$

$$C_{t,ij} + \int_{(0,t]} \int_{\mathbb{R}^d} f_i(\xi) f_j(\xi) \nu(du, d\xi) = \int_{(0,t]} a_{ij}(u, X_{u-}) A(du), \qquad a_{ij} \in \mathrm{Pol}_2^{dA}(\mathbb{R}_{>0} \times \mathbb{R}^d). \tag{6.47}$$

Further, the characteristics C and ν can be written as

$$C_{t,ij} = \int_{(0,t]} c_{u,ij} A(du), \ \nu(\omega; dt, d\xi) = K(\omega, t; d\xi) A(dt),$$

for some predictable processes $c_{\cdot,ij}$ and $\mathcal{P}$-measurable transition kernel K from $(\Omega \times \mathbb{R}_+, \mathcal{P})$ to $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Finally, if $k \geq 3$, then for arbitrary multi-index $\mathbf{l}$ with $3 \leq |\mathbf{l}| \leq k$, there exists a polynomial $g \in \mathrm{Pol}^{dA}_{|l|}(\mathbb{R}_{>0} \times \mathbb{R}^d)$, such that,

$$\int_{\mathbb{R}^d} \xi^{\mathbf{l}} K(\omega, u; d\xi) = g(u, X_{u-}) \quad [dA \otimes P].$$

Remark 6.49. *(i) Note that by the definition of the characteristic C, we have that $t \mapsto C_t$ is continuous. Hence we get for the process c that if $\Delta A_t > 0$, then $c_{t,ij} = 0$ for all $i, j \in \{1, \dots, d\}$.*

(ii) Let us make the remark that our characterization of the characteristics is different from the affine case (c.f. [KR+18, Theorem 3.2]). There, the authors consider the decomposition of A into continuous part and pure jump part. One can therefore ask for conditions under which a semimartingale with characteristics of the form above is an affine semimartingale in the sense of [KR+18]. This however would go beyond the scope of this thesis and we leave this question for future work.

Proof of Proposition 6.48. Since $k \geq 2$, we have $f_i \in \mathrm{Pol}_k$ for all $i \in \{1, \dots, d\}$. Hence, Lemma 6.37 *(iii)* yields that X is a special semimartingale. Let $X_{t,i} = M_{t,i} + B_{t,i}$ be the canonical decomposition of $f_i(X)$ and $\mathbf{l}$ any multi-index with $|\mathbf{l}| \leq k$. Note that we have $M_{t,i} = M^{f_i}_t$. By Itô's formula we have

$$\begin{aligned} f_{\mathbf{l}}(X_t) = f_{\mathbf{l}}(X_0) &+ \int_{(0,t]} \sum_{i=1}^d D_i f_{\mathbf{l}}(X_{u-}) dX_{u,i} + \frac{1}{2} \int_{(0,t]} \sum_{i,j=1}^d D_{ij} f_{\mathbf{l}}(X_{u-}) dC_{u,ij} \\ &+ \sum_{s \leq t} \left[f_{\mathbf{l}}(X_s) - f_{\mathbf{l}}(X_{s-}) - \sum_{i=1}^d D_i f_{\mathbf{l}}(X_{s-}) \Delta X_{i,s} \right]. \end{aligned} \tag{6.48}$$

Next, note that since by convention we have $X_{0-} = X_0$ and by that $\Delta X_{0-} = 0$, we can write (we suppress the ω-dependency of μ^X)

$$\begin{aligned} &\sum_{s \leq t} \left[f_{\mathbf{l}}(X_s) - f_{\mathbf{l}}(X_{s-}) - \sum_{i=1}^d D_i f_{\mathbf{l}}(X_{s-}) \Delta X_{i,s} \right] \\ &= \int_{(0,t]} \int_{\mathbb{R}^d} \left(f_{\mathbf{l}}(X_{u-} + \xi) - f_{\mathbf{l}}(X_{u-}) - \sum_{i=1}^d D_i f_{\mathbf{l}}(X_{u-}) \xi_i \right) \mu^X(du, d\xi). \end{aligned}$$

Recall that $\xi_i = f_i(\xi)$ and that $f_{\mathbf{l}}(x) = x^{\mathbf{l}}$. Then Lemma 2.1 yields

$$\begin{aligned}
&\sum_{s \le t} \left[f_{\mathbf{l}}(X_s) - f_{\mathbf{l}}(X_{s-}) - \sum_{i=1}^{d} D_i f_{\mathbf{l}}(X_{s-}) \Delta X_{i,s} \right] \\
&= \int_{(0,t]} \int_{\mathbb{R}^d} W(u,\xi) \mu^X(du, d\xi),
\end{aligned} \tag{6.49}$$

where

$$W(u,\xi) := \sum_{|\mathbf{m}|=2}^{|\mathbf{l}|} \binom{\mathbf{l}}{\mathbf{m}} f_{\mathbf{l}-\mathbf{m}}(X_{u-}) \xi^{\mathbf{m}}.$$

Using this, combined with the canonical decomposition of X, (6.48) becomes

$$\begin{aligned}
f_{\mathbf{l}}(X_t) = f_{\mathbf{l}}(X_0) &+ \int_{(0,t]} \sum_{i=1}^{d} D_i f_{\mathbf{l}}(X_{u-}) dM_u^{f_i} + \int_{(0,t]} \sum_{i=1}^{d} D_i f_{\mathbf{l}}(X_{u-}) dB_{u,i} \\
&+ \frac{1}{2} \int_{(0,t]} \sum_{i,j=1}^{d} D_{ij} f_{\mathbf{l}}(X_{u-}) dC_{u,ij} + \int_{(0,t]} \int_{\mathbb{R}^d} W(u,\xi) \mu^X(du, d\xi).
\end{aligned} \tag{6.50}$$

Since polynomials are smooth and $X_{t,i}$ is càglàd, $M^{f_i} \in \mathcal{M}_{\mathrm{loc}}$ yields

$$D_i f_{\mathbf{l}}(X_-) \bullet M^{f_i} \in \mathcal{M}_{\mathrm{loc}}.$$

By [JS13, Proposition I.3.5] we have $(D_i f_{\mathbf{l}}(X_-) \bullet B_i) \in \mathcal{V}$ and $(D_{ij} f_{\mathbf{l}}(X_-) \bullet C_{ij}) \in \mathcal{V}$ and by the same proposition, both are predictable. By [JS13, Lemma I.3.10], both belong to $\mathcal{A}_{\mathrm{loc}}$. From (6.49), we have

$$\int_{(0,t]} \int_{\mathbb{R}^d} W(u,\xi) \mu^X(du, d\xi) \in \mathcal{V}.$$

Hence, from [JS13, Proposition I.4.23 (*iii*)] it follows

$$W * \mu^X = \int_{(0,t]} \int_{\mathbb{R}^d} W(u,\xi) \mu^X(du, d\xi) \in \mathcal{A}_{\mathrm{loc}},$$

since $f_{\mathbf{l}}(X)$ is a special semimartingale. Hence $|W| * \mu^X \in \mathcal{A}^+_{\mathrm{loc}}$, and by [JS13, Proposition II.1.28] we have $W * \mu^X - W * \nu = W * (\mu^X - \nu)$, implying $W * \mu^X - W * \nu$ is a local martingale (compare [JS13, Definition II.1.27]). Since $f_{\mathbf{l}}(X)$ is special, it's canonical decomposition is unique. Therefore, by combining the definition of $M^{f_{\mathbf{l}}}$ with (6.50), we get

$$\begin{aligned}
M^{f_{\mathbf{l}}} &= \sum_{i=1}^{d} D_i f_{\mathbf{l}}(X_-) \bullet M^{f_i} + W * (\mu^X - \nu) \\
&= f_{\mathbf{l}}(X_t) - f_{\mathbf{l}}(X_0) - \sum_{i=1}^{d} D_i f_{\mathbf{l}}(X_-) \bullet B_i - \frac{1}{2} \sum_{i,j=1}^{d} D_{ij} f_{\mathbf{l}}(X_-) \bullet C_{ij} - W * \nu.
\end{aligned}$$

This, again by definition of $M^{f_{\mathbf{l}}}$, yields

$$
\begin{aligned}
\int_{(0,t]} \mathcal{G} f_{\mathbf{l}}(u, X_{u-}) A(du) = &\sum_{i=1}^{d} (D_i f_{\mathbf{l}}(X_-) \bullet B_i)_t \\
&+ \frac{1}{2} \sum_{i,j=1}^{d} (D_{ij} f_{\mathbf{l}}(X_-) \bullet C_{ij})_t + W * \nu_t.
\end{aligned}
\tag{6.51}
$$

Let $\mathbf{l} = \mathbf{e}_i$, i.e. in particular $|\mathbf{l}| = 1$ and $f_{\mathbf{l}}(x) = f_i(x)$. Then $D_j f_{\mathbf{l}}(x) = \mathbb{1}_{\{j=i\}}$ and $D_{ij} f_{\mathbf{l}}(x) = 0$. Therefore (6.51) implies

$$
\int_{(0,t]} \mathcal{G} f_i(u, X_{u-}) A(du) = B_{i,t}, \quad i \in \{1, \dots, d\}.
$$

Hence we can set $b_i := \mathcal{G} f_i$, and since G is degree preserving on Pol_k for $k \geq 2$ and $\|G\|_{dA,t} < \infty$ for all $t \geq 0$, we have by Lemma 6.47 that $b_i \in \mathrm{Pol}_1^{dA}(\mathbb{R}_{>0} \times \mathbb{R}^d)$. Next, consider the polynomials of degree two, i.e. $f_{ij}(x) = f_{\mathbf{e}_i + \mathbf{e}_j}(x)$ for $i, j \in \{1, \dots, d\}$. In this case, we have $\mathcal{G} f_{ij} \in \mathrm{Pol}_2$. Further, since $D_i f_{ij} = f_j \in \mathrm{Pol}_1$, we get $D_i f_{ij} b_i \in \mathrm{Pol}_2^{dA}$. Since $|\mathbf{e}_i + \mathbf{e}_j| = 2$, we have

$$
W(u, \xi) = \sum_{i,j=1}^{d} f_{ij}(\xi) = \sum_{i,j=1}^{d} \xi_i \xi_j.
$$

Hence, (6.51) yields

$$
C_{t,ij} + W * \nu_t = C_{t,ij} + \int_{(0,t]} \int_{\mathbb{R}^d} \xi_i \xi_j \nu(du, d\xi) = \int_{(0,t]} a_{ij}(u, X_{u-}) A(du),
$$

for some $a_{ij} \in \mathrm{Pol}_2^{dA}(\mathbb{R}_{>0} \times \mathbb{R}^d)$. Hence we have shown (6.46) and (6.47). We continue by defining

$$
A'_t(\omega) := \int_{(0,t]} \int_{\mathbb{R}^d} \|\xi\|_2^2 \, \nu(\omega; du, d\xi).
$$

Note that we have $\sum_{i=1}^{d} \xi_i \xi_i = \|\xi\|_2^2$, hence

$$
\sum_{i=1}^{d} C_{t,ii}(\omega) + A'_t(\omega) = \sum_{i=1}^{d} \int_{(0,t]} a_{ii}(u, X_{u-}) A(du).
\tag{6.52}
$$

By the definition of the characteristic C, the diagonal elements are non-negative increasing and predictable processes. Again by the definition of the characteristic ν, we get for the $\tilde{\mathcal{P}}$ measurable function $\hat{V}(\omega, t, \xi) = \|\xi\|_2^2$ that $A' = \hat{V} * \nu$ is predictable, hence A' is a predictable, non-negative and increasing process. By (6.52), we have for almost all $\omega \in \Omega$,

$$
d\left(\sum_{i=1}^{d} C_{\cdot,ii}(\omega) + A'(\omega) \right) \ll dA.
$$

Then, [JS13, Proposition I.3.13] implies that there exists a non-negative and predictable process H^1 such that

$$\sum_{i=1}^{d} C_{t,ii} + A'_t = (H^1 \bullet A)_t.$$

By non-negativity and the increasing property, we then have for a.a. $\omega \in \Omega$ and all $i \in \{1, \dots, d\}$,

$$dC_{\cdot,ii}(\omega) \ll dA \text{ and } dA'(\omega) \ll dA.$$

Hence, again by the same proposition, there are non-negative and predictable processes $c_{\cdot,ii}, H^2$ such that

$$C_{t,ii} = (c_{\cdot,ii} \bullet A)_t, \; A'_t = (H^2 \bullet A)_t.$$

Define $\tilde{V}(\omega, t, \xi) := \hat{V}(\omega, t, \xi) + \mathbb{1}_{\{0\}}(\xi)$. Then $\tilde{V}$ is a strictly positive, $\tilde{P}$-measurable function on $\tilde{\Omega}$. Further, since $\mu^X(\omega; \mathbb{R}_+ \times \{0\}) = 0$, by construction of the compensator ν we can always assume $\nu(\omega; \mathbb{R}_+ \times \{0\}) = 0$. That implies $A' = \hat{V} * \nu = \tilde{V} * \nu$. Hence by [JS13, Theorem II.1.8], there exists a transition kernel $K'(\omega, t; d\xi)$ from $(\Omega, \mathbb{R}_+, \mathcal{P})$ to $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ with

$$\nu(\omega; du, d\xi) = K'(\omega, u; d\xi) dA'_u.$$

Combining this with the existence of H^2 above, we get

$$\nu(\omega; du, d\xi) = H^2_u(\omega) K'(\omega, u; d\xi) dA_u.$$

Defining $K(\omega, u; d\xi) := H^2_u(\omega) K'(\omega, u; d\xi)$, we get that K is again a $\mathcal{P}$-measurable transition kernel with the desired property

$$\nu(\omega; du, d\xi) = K(\omega, u; d\xi) dA_u,$$

for almost all ω. This implies that (6.52) can be written as

$$C_{t,ij} = \int_{(0,t]} \left(a_{ij}(u, X_{u-}) - \int_{\mathbb{R}^d} \xi_i \xi_j K(u; d\xi) \right) A(du) \quad [P].$$

This implies for $i \neq j$, we also have $dC_{\cdot,ij} \ll dA$ and since a_{ij} are deterministic and X_{u-} is càglàd, we get $a_{ij}(u, X_{u-})$ is predictable. Further K is $\mathcal{P}$ measurable, implying that $\int_{\mathbb{R}^d} \xi_i \xi_j K(\omega, \cdot; d\xi)$ is predictable. Hence we get that the densities $c_{\cdot,ij}$ with $C_{t,ij} = (c_{\cdot,ij} \bullet A)_t$ are predictable processes. Let us continue. A closer look shows that we have the identity

$$\frac{1}{2} \sum_{i,j=1}^{d} D_{ij} f_{\mathbf{l}} \xi_i \xi_j = \sum_{|\mathbf{m}|=2}^{2} \binom{\mathbf{l}}{\mathbf{m}} f_{\mathbf{l}-\mathbf{m}} \xi^{\mathbf{m}},$$

where the important observation is that all multi-indices $\mathbf{m}$ with $|\mathbf{m}| = 2$ are of the form

$\mathbf{m} = \mathbf{e}_i + \mathbf{e}_j$. Hence we can write (6.51) as

$$\begin{aligned}
&\int_{(0,t]} \mathcal{G} f_{\mathbf{l}}(u, X_{u-}) A(du) = \int_{(0,t]} \sum_{i=1}^{d} D_i f_{\mathbf{l}}(X_{u-}) b_i(u, X_{u-}) A(du) \\
&+ \int_{(0,t]} \frac{1}{2} \sum_{i,j=1}^{d} D_{ij} f_{\mathbf{l}}(X_{u-}) \left(c_{t,ij} + \int_{\mathbb{R}^d} \xi_i \xi_j K(u; d\xi) \right) A(du) \\
&+ \int_{(0,t]} \int_{\mathbb{R}^d} \left(\sum_{|\mathbf{m}|=3}^{|\mathbf{l}|} \binom{\mathbf{l}}{\mathbf{m}} f_{\mathbf{l}-\mathbf{m}}(X_{u-}) \xi^{\mathbf{m}} \right) K(u; d\xi) A(du) \\
&= \int_{(0,t]} \sum_{i=1}^{d} D_i f_{\mathbf{l}}(X_{u-}) b_i(u, X_{u-}) A(du) + \int_{(0,t]} \frac{1}{2} \sum_{i,j=1}^{d} D_{ij} f_{\mathbf{l}}(X_{u-}) a_{ij}(u, X_{u-}) A(du) \\
&+ \int_{(0,t]} \int_{\mathbb{R}^d} \left(\sum_{|\mathbf{m}|=3}^{|\mathbf{l}|} \binom{\mathbf{l}}{\mathbf{m}} f_{\mathbf{l}-\mathbf{m}}(X_{u-}) \xi^{\mathbf{m}} \right) K(u; d\xi) A(du).
\end{aligned} \tag{6.53}$$

Consider the case where $\mathbf{l}$ is such that $|\mathbf{l}| = 3$. Then the summands in the last term above are unequal zero only for $\mathbf{m} = \mathbf{l}$. Hence we have

$$\sum_{|\mathbf{m}|=3}^{|\mathbf{l}|} \binom{\mathbf{l}}{\mathbf{m}} f_{\mathbf{l}-\mathbf{m}}(X_{u-}) \xi^{\mathbf{m}} = \xi^{\mathbf{l}}.$$

As all other terms belong to $\mathrm{Pol}_{\mathbf{l}}^{dA}$ for $|\mathbf{l}| = l$, we get that there exists a $g \in \mathrm{Pol}_l^{dA}(\mathbb{R}_{>0} \times \mathbb{R}^d)$ such that

$$\int_{(0,t]} \int_{\mathbb{R}^d} \xi^{\mathbf{l}} K(u; d\xi) A(du) = \int_{(0,t]} g(u, X_{u-}) A(du) \quad [P],$$

and since this holds for all $t \geq 0$, we get $\int_{\mathbb{R}^d} \xi^{\mathbf{l}} K(u; d\xi) = g(u, X_{u-})$ outside $dA \otimes P$ zero sets for all $\mathbf{l}$ with $l = 3$. We extend this result to all $\mathbf{l}$ with $|\mathbf{l}| \leq k$ by induction. Assume for all $\mathbf{m}$ with $|\mathbf{m}| \leq l - 1$ there exists a polynomial $g_{\mathbf{m}} \in \mathrm{Pol}_{|\mathbf{m}|}^{dA}(\mathbb{R}_{>0} \times \mathbb{R}^d)$ such that

$$\int_{\mathbb{R}^d} \xi^{\mathbf{m}} K(u; d\xi) = g_{\mathbf{m}}(u, X_{u-}) \quad [dA \otimes P].$$

Note that we can assume $f_{\mathbf{l}-\mathbf{m}} \in \mathrm{Pol}_{|\mathbf{l}|-|\mathbf{m}|}$ since for all $\mathbf{m} \not\leq \mathbf{l}$ the binomial coefficients are zero. Hence for $\mathbf{m} \leq \mathbf{l}$ we get $f_{\mathbf{l}-\mathbf{m}} g_{\mathbf{m}} \in \mathrm{Pol}_l^{dA}(\mathbb{R}_{>0} \times \mathbb{R}^d)$. Further, we have

$$\begin{aligned}
&\int_{(0,t]} \int_{\mathbb{R}^d} \left(\sum_{|\mathbf{m}|=3}^{|\mathbf{l}|} \binom{\mathbf{l}}{\mathbf{m}} f_{\mathbf{l}-\mathbf{m}}(X_{u-}) \xi^{\mathbf{m}} \right) K(u; d\xi) A(du) \\
&= \int_{(0,t]} \sum_{|\mathbf{m}|=3}^{|\mathbf{l}|} \binom{\mathbf{l}}{\mathbf{m}} f_{\mathbf{l}-\mathbf{m}}(X_{u-}) \left(\int_{\mathbb{R}^d} \xi^{\mathbf{m}} K(u; d\xi) \right) A(du) \\
&= \int_{(0,t]} \sum_{|\mathbf{m}|=3}^{l-1} \binom{\mathbf{l}}{\mathbf{m}} f_{\mathbf{l}-\mathbf{m}}(X_{u-}) \left(\int_{\mathbb{R}^d} \xi^{\mathbf{m}} K(u; d\xi) \right) A(du) \\
&+ \int_{(0,t]} \sum_{|\mathbf{m}|=l-1}^{l} \binom{\mathbf{l}}{\mathbf{m}} f_{\mathbf{l}-\mathbf{m}}(X_{u-}) \left(\int_{\mathbb{R}^d} \xi^{\mathbf{m}} K(u; d\xi) \right) A(du) \\
&= \int_{(0,t]} \sum_{|\mathbf{m}|=3}^{l-1} \binom{\mathbf{l}}{\mathbf{m}} f_{\mathbf{l}-\mathbf{m}}(X_{u-}) g_{\mathbf{m}}(u, X_{u-}) A(du) \\
&+ \int_{(0,t]} \sum_{|\mathbf{m}|=l-1}^{l} \binom{\mathbf{l}}{\mathbf{m}} f_{\mathbf{l}-\mathbf{m}}(X_{u-}) \left(\int_{\mathbb{R}^d} \xi^{\mathbf{m}} K(u; d\xi) \right) A(du) \\
&= \int_{(0,t]} \sum_{|\mathbf{m}|=3}^{l-1} \binom{\mathbf{l}}{\mathbf{m}} f_{\mathbf{l}-\mathbf{m}}(X_{u-}) g_{\mathbf{m}}(u, X_{u-}) A(du) + \int_{(0,t]} \int_{\mathbb{R}^d} \xi^{\mathbf{l}} K(u; d\xi) A(du).
\end{aligned}$$

Hence by the same arguments as before, we get that there exists a $g_{\mathbf{l}} \in \mathrm{Pol}_l^{dA}(\mathbb{R}_{>0} \times \mathbb{R}^d)$ such that

$$\int_{\mathbb{R}^d} \xi^{\mathbf{l}} K(u; d\xi) = g_{\mathbf{l}}(u, X_{u-}) \quad [dA \otimes P],$$

for all $\mathbf{l}$ with $|\mathbf{l}| \leq k$, which finishes the proof. □

Remark 6.50. *In Cuchiero et al. [Cuc+12], the authors show in their version a stronger result regarding the processes $c_{\cdot,ij}$ and the random measure $K(u; d\xi)$ above. A result in our situation analogous to theirs (they proved a time homogeneous version) would be that one can pick autonomous versions of $c_{\cdot,ij}$ and $K(u; d\xi)$, that is one can pick version of the form*

$$c_{t,ij}(\omega) = c(t, X_{u-}(\omega)) \text{ and } K(\omega, u; d\xi) = K(u, X_{u-}(\omega); d\xi).$$

The authors took advantage of the Markov structure they assumed so that they could apply results from Çinlar et al. [Çin+80]. This is not possible in our situation, explaining the motivation for the next definition and the subsequent proposition that can be seen as the second part of our version of [Cuc+12, Proposition 2.12].

Let us first recall the following proposition from Jacod and Shiryaev [JS13, Proposition II.2.9], where we omit the proof. We restrict ourself to the case of special semimartingales.

Proposition 6.51. *Let X be a special semimartingale. Then one can find a version of the characteristics (B, C, ν), called "good versions", wrt the "truncation" function $h(x) = x$ which are of the form*

$$\begin{aligned} B_i &= b_i \bullet A \\ C_{ij} &= c_{ij} \bullet A \\ \nu(\omega; dt, d\xi) &= K(\omega, t; d\xi) dA_t, \end{aligned}$$

where $A \in \mathcal{A}^+_{loc}$ is predictable, for all $i, j \in \{1, \dots, d\}$, the $\mathbb{R}$-valued processes b_i and c_{ij} are predictable and $c = (c_{ij})^d_{i,j=1}$ takes values in the set of all symmetric non-negative (positive definite) matrices and $K(\omega, t; d\xi)$ is a transition kernel from $(\Omega \times \mathbb{R}_+, \mathcal{P})$ into $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

Note that K is indexed with time $t \in \mathbb{R}_+$. Since $\Delta X_0 = 0$ by convention, we can always assume $K(0, x; d\xi) = 0 d\xi$. Further, by construction, for the canonical decomposition $X = X_0 + M + V$ one has $M_0 = 0$ and $V_0 = 0$. Hence, it is sufficient to define the characteristics on $\mathbb{R}_{>0}$ rather then $\mathbb{R}_+$ as we will do from here on. This can also be seen when considering the A integrals above. Since A is càdlàg , b and c do not need to be defined at time $t = 0$.

Definition 6.52. *Let X be a special semimartingale. We call X a* ***k*-polynomial semimartingale** *for $k \geq 2$, if one can choose a representation in Proposition 6.51, such that for $i, j \in \{1, \dots, d\}$,*

(i) *$A \in \mathcal{A}^+_{loc}$ can be chosen as an element of $\mathcal{V}^+_{\mathfrak{A}} \subset \mathcal{A}^+_{loc}$,*

(ii) *$b_{t,i} = b_i(t, X_{t-})$ for some $b_i \in \mathrm{Pol}^{dA}_1(\mathbb{R}_{>0} \times \mathbb{R}^d)$,*

(iii) *$c_{t,ij} = c_{ij}(t, X_{t-})$ for some measurable function c_{ij},*

(iv) *$K(\omega, t; d\xi) = K(t, X_{t-}, d\xi)$, where K is some transition kernel from*

$$(\mathbb{R}_+ \times \mathbb{R}^d, \mathcal{B}(\mathbb{R}_+) \otimes \mathcal{B}(\mathbb{R}^d)) \text{ into } (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)),$$

(v) *the functions c_{ij} and K satisfy for all $i, j \in \{1, \dots, d\}$ and $x \in \mathbb{R}^d$,*

$$c_{ij}(u, x) + \int_{\mathbb{R}^d} \xi_i \xi_j K(u, x; d\xi) = a_{ij}(u, x),$$

where all $a_{ij} \in \mathrm{Pol}^{dA}_2(\mathbb{R}_{>0} \times \mathbb{R}^d)$,

(vi) *if $k \geq 3$, then for all multi-indices $\mathbf{l}$ with $3 \leq |\mathbf{l}| \leq k$, it holds*

$$\int_{\mathbb{R}^d} \xi^{\mathbf{l}} K(u, x; d\xi) \in \mathrm{Pol}^{dA}_{|\mathbf{l}|}(\mathbb{R}_{>0} \times \mathbb{R}^d).$$

If X is a k-polynomial semimartingale for all $k \geq 2$, then we simply say X is a polynomial semimartingale.

Remark 6.53. *(i) Just as in Remark 6.49, note that since we require the above functions to define characteristics, we implicitly get that c vanishes for those times where $\Delta A_t > 0$. Again, this is different from the characterization results for affine semimartingales in [KR+18].*

(ii) Note that we implicitly assume that all the expressions in the definition above are well defined. In particular, we have $\int_{\mathbb{R}^d} \|\xi\|^k K(u, x; d\xi) < \infty$ for all $(u, x) \in \mathbb{R}_{>0} \times \mathbb{R}^d$.

In the sequel, whenever we talk about a polynomial semimartingale, we assume the characteristics are given in the form above.

Proposition 6.54. *Let X be a k-polynomial semimartingale, hence $k \geq 2$. Then for $A \in \mathcal{V}^+_{\mathfrak{A}}$ from the definition of polynomial semimartingales and k, X satisfies Condition 6.43 for said A and k.*

Proof. Note that $\{f_{\mathbf{l}} : |\mathbf{l}| \leq k\}$ forms a basis of Pol_k. Hence for any $f \in \mathrm{Pol}_l$ with $l \leq k$, there are coefficients $\alpha_{\mathbf{m}}$ indexed by multi-indices $\mathbf{m}$ with $|\mathbf{m}| \leq k$, such that

$$f(x) = \sum_{|\mathbf{m}|=0}^{k} \alpha_{\mathbf{m}} f_{\mathbf{m}}(x), \quad \alpha_m = (0, \dots, 0) \quad \forall\, \mathbf{m} \text{ with } l < |\mathbf{m}| \leq k.$$

Consider now (6.53). It follows that a version of $\mathcal{G}_{X,A}$ on Pol_k, denoted by $\mathcal{G}$, is given by

$$\mathcal{G} f(u, x) = \sum_{|\mathbf{m}|=0}^{k} \alpha_{\mathbf{m}} \mathcal{G} f_{\mathbf{m}}(u, x) = \sum_{|\mathbf{m}|=0}^{l} \alpha_{\mathbf{m}} \mathcal{G} f_{\mathbf{m}}(u, x),$$

implying $\mathcal{G} f \in \mathrm{Pol}_l^{dA}$. Hence $\mathcal{G}$ is a linear map on Pol_k that preserves the degree. Hence there exists a matrix function G such that $\mathcal{G} \overset{\iota k}{\leftarrow} G$ where G is degree preserving on Pol_k by Corollary 5.13. To be precise, each matrix $G(u)$ is defined the following way. The i-th column with $i \leq N_k$ is given by the coefficients of the polynomial $\mathcal{G}\beta_i(u, \cdot)$ wrt the basis β^k.

We are left to show $\|G\|_{dA,t} < \infty$ for all $t \geq 0$. Since this property is preserved under basis transformations as these are matrices that are constant in u, we choose to work under the canonical basis of Pol_k introduced in the preliminaries. In particular, we have $\beta_i = f_{\mathbf{l}_i}$ for all $\beta_i \in \beta^k$ where $\mathbf{l}_i$ is the i-th element in Λ_k under the graded reverse-lexicographic order. Note that this basis satisfies Convention 5.1.

Define the operators $\mathcal{G}_i : \mathrm{Pol}_k \to \mathrm{Pol}_k^{dA}$ for $i = 1, 2, 3$ by their action on the basis polynomials β_λ for $\lambda \in \{1, \dots, N_k\}$ via

$$
\begin{aligned}
\mathcal{G}_1\beta_\lambda(u,x) &= \sum_{i=1}^{d} D_i\beta_\lambda(x) b_i(u,x),\\
\mathcal{G}_2\beta_\lambda(u,x) &= \frac{1}{2}\sum_{i,j=1}^{d} D_{ij}\beta_\lambda(x) a_{ij}(u,x),\\
\mathcal{G}_3\beta_\lambda(u,x) &= \int_{\mathbb{R}^d} \left(\sum_{|\mathbf{m}|=3}^{k} \binom{\mathbf{l}_\lambda}{\mathbf{m}} f_{\mathbf{l}_\lambda - \mathbf{m}}(X_{u-})\xi^{\mathbf{m}} \right) K(u; d\xi)\\
&= \int_{\mathbb{R}^d} \left(\sum_{|\mathbf{m}|=3}^{|\mathbf{l}_\lambda|} \binom{\mathbf{l}_\lambda}{\mathbf{m}} f_{\mathbf{l}_\lambda - \mathbf{m}}(X_{u-})\xi^{\mathbf{m}} \right) K(u; d\xi).
\end{aligned}
$$

Note that just as before, the sum is well defined since the multi-binomial coefficients are zero if $\mathbf{m} \le \mathbf{l}_\lambda$ does not hold. Hence we have for all Pol_k the identity $\mathcal{G}f = \sum_{i=1}^3 \mathcal{G}_i f$ and for $G_i \overset{\iota_k}{\to} \mathcal{G}_i$ we have $G = \sum_{i=1}^3 G_i$ for all $u \in \mathbb{R}_{>0}$.

Consider $\mathcal{G}_1$ and define for $1 \in \{1, \dots, d\}$ the operators $\mathcal{T}_i f(u,x) = D_i f(x) b_i(u,x)$ such that we have $\mathcal{G}_1 = \sum_{i=1}^d \mathcal{T}_i$. Since D_i reduces the degree of a non-constant polynomial by one, we have

$$
D_i\beta_\lambda(x) = \sum_{z=1}^{N_k} \tilde{\alpha}_z^\lambda \beta_z(x) = \sum_{z=1}^{N_{k-1}} \tilde{\alpha}_z^\lambda \beta_z(x),
$$

for some finite constants $\tilde{\alpha}_z^\lambda$. Hence, since $b_i \in \mathrm{Pol}_1^{dA}(\mathbb{R}_{>0} \times \mathbb{R}^d)$, there are measurable functions $\alpha_z^\lambda(u)$, such that

$$
D_i\beta_\lambda(x) b_i(u,x) = \sum_{z=1}^{N_k} \alpha_z^\lambda(u)\beta_z(x),
$$

and since $\|\iota_k(b_i(u,\cdot))\|_1$ is essentially locally bounded, we get

$$
\operatorname*{ess\,sup}_{u \le t, dA} |\alpha_z^\lambda(u)| < \infty, \quad \text{for all } t \ge 0 \text{ and } \lambda, z \le N_k.
$$

Hence the representing matrix of $\mathcal{T}_i \overset{\iota_k}{\leftarrow} T_i$ is given by these coefficients, that is

$$
(T_i(u))_{z,\lambda} = \alpha_z^\lambda(u).
$$

Denote with $\|\cdot\|_\Sigma$ the matrix norm given by

$$
\|M\|_\Sigma := \sum_{\lambda,z=1}^{N_k} |M_{z,\lambda}| \;\text{ for } M = (M_{z,\lambda})_{z,\lambda=1}^{N_k}.
$$

We get

$$\operatorname*{ess\,sup}_{u\le t,dA} \|T_i(u)\|_\Sigma \le \sum_{z,\lambda=1}^{N_k} \operatorname*{ess\,sup}_{u\le t,dA} |\alpha_z^\lambda(u)| < \infty.$$

Since all norms in finite dimension are equivalent, this implies $\|T_i\|_{dA,t} < \infty$ for all $t \ge 0$. This implies $\|G_1\|_{dA,t} < \infty$ for all $t \ge 0$ by means of the triangle inequality applied to $G_1(u) = \sum_{i=1}^d T_i(u)$.

By the same arguments we get $\|G_i\|_{dA,t} < \infty$ for all $t \ge 0$ where $i \in \{2, 3\}$. Hence, again by the triangle inequality we get $\|G\|_{dA,t} < \infty$ for all $t \ge 0$, which finishes the proof. □

Remark 6.55. *If X is a polynomial semimartingale, the form of the characteristics might suggest that X has the Markov property. However, since we do not assume the corresponding semimartingale problem to be well posed, even in the pure diffusion case, this does not need to hold. This observation is analogous to the one made by Filipović and Larsson [FL16] and Filipović and Larsson [FL17] in the respective introductions.*

We now want to address the question of conditions under which a k-polynomial semimartingale X can be embedded in bounded and uniformly integrable k-polynomial process (X, P, A, G). By means of Theorem 6.44, if X is a polynomial semimartingale with $\mathbb{E}\left[\|X_0\|_{\mathbb{R}^d}^k\right] < \infty$, then this is always possible for all $k \ge 2$ since by definition, P and G preserve the degree and X has the l-polynomial property for all $l \le k$. Hence if X is a $(2k)$-polynomial semimartingale, then it can be embedded in a bounded and uniformly integrable $2k$-polynomial process (X, P, A, G). Since P and G can be projected to $\mathbb{R}^{N_k}$, one can then embed X into the bounded and uniformly integrable k-polynomial process $(X, \tilde{P}, A, \tilde{G})$ with projections $\tilde{P}, \tilde{G}$ that are MESs and corresponding extended generators due to the degree preserving properties, see Proposition 5.17, Theorem 5.19 combined with the uniqueness result Proposition 4.53.

Let us continue with the following lemma that is our version of [Cuc+12, Lemma 2.17]. Again the proof we present is a modification of the one presented there.

Lemma 6.56. *Let X be a special semimartingale where for $k \ge 2$, the characteristics (B, C, ν) wrt the "truncation" function $h(x) = x$ satisfy properties $(i) - (v)$ in Definition 6.52 for $A \in \mathcal{V}_{\mathfrak{A}}^+$. Then there exists a constant K that depends on t such that*

$$\begin{aligned}\mathbb{E}\left[\sup_{s\le t} \|X_s\|_{\mathbb{R}^d}^k\right] \le K\Big(1 + \mathbb{E}\left[\|X_0\|_{\mathbb{R}^d}^k\right] + \int_{(0,t]} \mathbb{E}\left[\int_{\mathbb{R}^d} \|\xi\|_{\mathbb{R}^d}^k K(u, X_{u-}; d\xi)\right] A(du) \\ + \int_{(0,t]} \mathbb{E}\left[\|X_{u-}\|_{\mathbb{R}^d}^k\right] A(du)\Big).\end{aligned} \tag{6.54}$$

If additionally, we have $\mathbb{E}\left[\|X_0\|_{\mathbb{R}^d}^k\right] < \infty$ and either

$$\int_{(0,t]} \mathbb{E}\left[\int_{\mathbb{R}^d} \|\xi\|_{\mathbb{R}^d}^k K(u, X_{u-}; d\xi)\right] A(du) < \infty, \quad \forall t \ge 0, \tag{6.55}$$

or if for all $t \geq 0$, there exists some finite constant $\widehat{K}$ such that for all $t \leq T$,

$$\int_{\mathbb{R}^d} \|\xi\|_{\mathbb{R}^d}^k K(t,x;d\xi) \leq \widehat{C}\,(1+\|x\|_{\mathbb{R}^d}^k), \quad x \in \mathbb{R}^d, \tag{6.56}$$

then for all $t \geq 0$,

$$\mathbb{E}\left[\sup_{s\leq t} \|X_s\|_{\mathbb{R}^d}^k\right] < \infty.$$

Proof. As a first step we show that $\mathbb{E}\left[\sup_{(0,t]} |f_i(X_s)|^k\right] = \mathbb{E}\left[\sup_{(0,t]} |X_{s,i}|^k\right]$ satisfies the inequality. Let $X_i = X_{0,i} + M + F$ be the canonical decomposition of the special semimartingale X_i. Hence we have $M = M^{f_i}$ and $F = B_i$ which gives us

$$F_t = \int_{(0,t]} b_i(u, X_{u-})A(du).$$

By [JS13, Theorem I.4.18] we can decompose M into continuous and purely discontinuous part $M = M_0 + M^c + M^d$ where in our case $M_0 = M_0^c = M_0^d = 0$. Since a continuous local martingale has no jumps, hence bounded jumps, $M_0^c = 0$ yields that M^c is a locally square integrable martingale and we denote with $\langle\cdot,\cdot\rangle$ the predictable covariation. Hence we have $C_{t,ii} = \langle M^c\rangle_t = \langle M^c, M^c\rangle_t$. By [JS13, Theorem I.4.52], the increasing and non-negative process Z defined by

$$Z_t := \sum_{s\leq t} (\Delta X_{s,i})^2 = \int_{(0,t]}\int_{\mathbb{R}^d} \xi_i^2 \mu^X(ds, d\xi),$$

gives us

$$[M] = [M,M] = C + Z,$$

where $[\cdot,\cdot]$ denotes the quadratic covariation of two semimartingales. Next, define the stopping times

$$\begin{aligned} T_j^X &:= \inf\{t \geq 0 : |X_t| \geq j\} \\ T_j^Z &:= \inf\{t \geq 0 : |Z_t| \geq j\} \\ T_j &:= T_j^X \wedge T_j^Z. \end{aligned}$$

For the rest of the proof, recall that we denote with $\tilde{C}$ a finite constant (that is independent of t) that can change from estimate to estimate. For notational convenience, let us define the function

$$\rho_1(t) := \operatorname*{ess\,sup}_{u\leq t, dA} \|\iota_1(b_i(u,\cdot))\|_1\,,$$

which is finite for all $t \geq 0$ since $b_i \in \mathrm{Pol}_1^{dA}(\mathbb{R}_{>0}\times\mathbb{R}^d)$. A standard estimate yields

$$\sup_{s\leq t}\left|X_{i,s}^{T_j}\right|^k \leq \tilde{C}\left(|X_{0,i}|^k + \sup_{s\leq t}\left|M_s^{T_j}\right|^k + \sup_{s\leq t}\left|\int_{(0,s\wedge T_j]} b_i(u,X_{u-})A(du)\right|^k\right). \tag{6.57}$$

Note that we have $|X_{0,i}|^k \leq \|X_0\|_1^k \leq \tilde{C}\, \|X_0\|_{\mathbb{R}^d}^k$. Further, we have $b_i \in \mathrm{Pol}_1^{dA}(\mathbb{R}_{>0} \times \mathbb{R}^d)$ by assumption. With Lemma 6.47 (*ii*), we estimate

$$\begin{aligned}
\sup_{s\leq t}\left|\int_{(0,s\wedge T_j]} b_i(u, X_{u-})A(du)\right|^k &\leq \sup_{s\leq t}\int_{(0,s\wedge T_j]} |b_i(u, X_{u-})|^k\, A(du) \\
&\leq \sup_{s\leq t}\int_{(0,s\wedge T_j]} \left|\tilde{C}\rho_1(t)(1 + \|X_{u-}\|_{\mathbb{R}^d})\right|^k A(du) \\
&\leq \tilde{C}(\rho_1(t) + A(t))\Big(1 + \int_{(0,t]} \left\|X_{(u\wedge T_j)-}\right\|_{\mathbb{R}^d}^k A(du)\Big).
\end{aligned} \tag{6.58}$$

Let us now consider $\sup_{s\leq t} |M_{s\wedge T_j}|^k$. An application of the Burkholder-Davis-Gundy inequality yields

$$\mathbb{E}\left[\sup_{s\leq t} |M_{s\wedge T_j}|^k\right] \leq \tilde{C}\mathbb{E}\left[[M]_{t\wedge T_j}^{\frac{k}{2}}\right] \leq \tilde{C}\mathbb{E}\left[C_{t\wedge T_j}^{\frac{k}{2}} + Z_{t\wedge T_j}^{\frac{k}{2}}\right], \tag{6.59}$$

where in the last step we used Jensen's inequality for sums. Recall that by our assumptions, we have

$$C_{t,ii} + \int_{(0,t]} \xi_i^2 K(u, X_{u-}; d\xi)A(du) = \int_{(0,t]} a_{ii}(u, X_{u-})A(du),$$

for $a_{ii} \in \mathrm{Pol}_2^{dA}(\mathbb{R}_{>0} \times \mathbb{R}^d)$. Define

$$\rho_2(t) := \operatorname*{ess\,sup}_{u\leq t, dA} \|\iota_2(a_{ii}(u, \cdot))\|_1\,,$$

which is finite for all $t \geq 0$ since $a_{ii} \in \mathrm{Pol}_2^{dA}(\mathbb{R}_{>0} \times \mathbb{R}^d)$. Since $\xi_i^2 \geq 0$, this yields again with Lemma 6.47 (*ii*),

$$\begin{aligned}
\mathbb{E}\left[C_{t\wedge T_j}^{\frac{k}{2}}\right] &\leq \tilde{C}(\rho_2(t) + A(t))\left(1 + \mathbb{E}\left[\int_{(0,t\wedge T_j]} \|X_{u-}\|_{\mathbb{R}^d}^k\, A(du)\right]\right) \\
&\leq \tilde{C}(\rho_2(t) + A(t))\left(1 + \int_{(0,t]} \mathbb{E}\left[\left\|X_{(u\wedge T_j)-}\right\|_{\mathbb{R}^d}^k\right] A(du)\right).
\end{aligned}$$

In the last step we use Fubini-Tonelli due to non-negativity of the integrand. We now consider $Z_{t\wedge T_j}^{\frac{k}{2}}$. We continue following the ideas in [Cuc+12] where the authors base their arguments on ideas presented in Jacod et al. [Jac+05]. Note that we have

$$\Delta Z_t = \sum_{s\leq t}(\Delta X_{i,s})^2 - \sum_{s<t}(\Delta X_{i,s})^2 = (\Delta X_{i,t})^2.$$

Since Z is increasing and piecewise constant, we get that for any $p \geq 1$, we have that Z^p is increasing and piecewise constant with $Z_0^p = 0$. This implies

$$Z_t^{\frac{k}{2}} = \sum_{s\leq t} \Delta(Z_s^{\frac{k}{2}}) = \sum_{s\leq t} Z_s^{\frac{k}{2}} - Z_{s-}^{\frac{k}{2}} = \sum_{s\leq t} (Z_{s-} + \Delta Z_s)^{\frac{k}{2}} - Z_{s-}^{\frac{k}{2}}.$$

Hence we get

$$Z_{t\wedge T_j}^{\frac{k}{2}} = \int_{(0,t\wedge T_j]}\int_{\mathbb{R}^d}\left[\left(Z_{s-}+\xi_i^2\right)^{\frac{k}{2}} - Z_{s-}^{\frac{k}{2}}\right]\mu^X(ds,d\xi).$$

Noting that ν is the predictable compensator of μ^X and the integrand above is $\tilde{\mathcal{P}}$ measurable, we have by [JS13, Theorem II.1.8]

$$\mathbb{E}\left[Z_{t\wedge T_j}^{\frac{k}{2}}\right] = \mathbb{E}\left[\int_{(0,t\wedge T_j]}\int_{\mathbb{R}^d}\left[\left(Z_{s-}+\xi_i^2\right)^{\frac{k}{2}} - Z_{s-}^{\frac{k}{2}}\right]\nu(ds,d\xi)\right]. \tag{6.60}$$

The following fundamental inequalities will play an important role in the sequel. They were proposed in [Jac+05] and we shall use them in the same way they were used in [Cuc+12]. For $x, z \geq 0$, $\varepsilon > 0$ and $\frac{k}{2} = p \geq 1$,

$$(z+x)^p - z^p \leq 2^{p-1}(z^{p-1}x + x^p), \tag{6.61}$$

$$z^{p-1}x \leq \varepsilon z^p + \frac{x^p}{\varepsilon^{p-1}}. \tag{6.62}$$

We apply (6.61) to (6.60), yielding

$$\begin{aligned}\mathbb{E}\left[Z_{t\wedge T_j}^{\frac{k}{2}}\right] &\leq \mathbb{E}\left[\int_{(0,t\wedge T_j]}\int_{\mathbb{R}^d}2^{\frac{k}{2}-1}\left[Z_{s-}^{\frac{k}{2}-1}\xi_i^2 + (\xi_i^2)^{\frac{k}{2}}\right]\nu(ds,d\xi)\right]\\ &= \mathbb{E}\left[\int_{(0,t\wedge T_j]}\int_{\mathbb{R}^d}2^{\frac{k}{2}-1}\left[Z_{s-}^{\frac{k}{2}-1}\xi_i^2 + |\xi_i|^k\right]\nu(ds,d\xi)\right].\end{aligned} \tag{6.63}$$

By assumption, property (iii) in Definition 6.52 holds for ν. Further, since $C_{t,ii}$ is increasing and non-negative, we have $c_{ii}(u, X_{u-}) \geq 0$ outside a $dA\otimes P$-zero set. This implies $\int_{\mathbb{R}^d}\xi_i^2 K(u, X_{u-}; d\xi) \leq a_{ii}(u, X_{u-})$ outside a $dA\otimes P$-zero set. Hence we compute for the first part on the rhs above,

$$\begin{aligned}&\mathbb{E}\left[\int_{(0,t\wedge T_j]}\int_{\mathbb{R}^d}2^{\frac{k}{2}-1}Z_{s-}^{\frac{k}{2}-1}\xi_i^2\nu(ds,d\xi)\right]\\ &= \mathbb{E}\left[\int_{(0,t\wedge T_j]}\int_{\mathbb{R}^d}2^{\frac{k}{2}-1}Z_{s-}^{\frac{k}{2}-1}\xi_i^2 K(s,X_{s-})A(ds)\right]\\ &\leq \mathbb{E}\left[\int_{(0,t\wedge T_j]}2^{\frac{k}{2}-1}Z_{s-}^{\frac{k}{2}-1}a_{ii}(s,X_{s-})A(ds)\right]\\ &\leq \tilde{C}(1+A(t))\,\mathbb{E}\left[\int_{(0,t\wedge T_j]}\varepsilon Z_{s-}^{\frac{k}{2}} + \frac{1+\|X_{s-}\|_{\mathbb{R}^d}^{2\frac{k}{2}}}{\varepsilon^{\frac{k}{2}-1}}A(ds)\right].\end{aligned}$$

In the last step, we used $|Z_s| = Z_s$, Lemma 6.47 (ii) for $a_{ii} \in \mathrm{Pol}_2^{dA}(\mathbb{R}_{>0}\times\mathbb{R}^d)$ and (6.62). Since Z is non-decreasing, we can estimate using $Z_-^{T_j} \leq j$,

$$\int_{(0,t\wedge T_j]}Z_{s-}^{\frac{k}{2}}A(ds) \leq A(t)(j^{\frac{k}{2}} \wedge Z_{t\wedge T_j}^{\frac{k}{2}}).$$

Applying these estimates to (6.63), we get

$$
\begin{aligned}
\mathbb{E}\left[Z^{\frac{k}{2}}_{t\wedge T_j}\right] &\le \tilde{C}(1+A(t))\varepsilon\,\mathbb{E}\left[\int_{(0,t\wedge T_j]} Z^{\frac{k}{2}}_{s-}A(ds)\right] \\
&\quad + \tilde{C}(1+A(t))\,\mathbb{E}\left[\int_{(0,t\wedge T_j]} \frac{1+\|X_{s-}\|^k_{\mathbb{R}^d}}{\varepsilon^{\frac{k}{2}-1}}A(ds)\right] \\
&\quad + \mathbb{E}\left[\int_{(0,t\wedge T_j]}\int_{\mathbb{R}^d} 2^{\frac{k}{2}-1}|\xi_i|^k\nu(ds,d\xi)\right] \\
&\le \tilde{C}(1+A(t)+A(t)^2)\varepsilon\,\mathbb{E}\left[j^{\frac{k}{2}}\wedge Z^{\frac{k}{2}}_{t\wedge T_j}\right] \\
&\quad + \tilde{C}(1+A(t)+A(t)^2)\,\mathbb{E}\left[\int_{(0,t\wedge T_j]} \frac{1+\|X_{s-}\|^k_{\mathbb{R}^d}}{\varepsilon^{\frac{k}{2}-1}}A(ds)\right] \\
&\quad + \tilde{C}\mathbb{E}\left[\int_{(0,t\wedge T_j]}\int_{\mathbb{R}^d} |\xi_i|^k K(s,X_{s-};d\xi)A(ds)\right].
\end{aligned} \tag{6.64}
$$

Since $Z^{\frac{k}{2}}_{t\wedge T_j} \ge (j^{\frac{k}{2}}\wedge Z^{\frac{k}{2}}_{t\wedge T_j})$, we have $\frac{1}{2}Z^{\frac{k}{2}}_{t\wedge T_j} \le Z^{\frac{k}{2}}_{t\wedge T_j} - \frac{1}{2}(j^{\frac{k}{2}}\wedge Z^{\frac{k}{2}}_{t\wedge T_j})$. Hence, choosing $\varepsilon = (2\tilde{C}(1+A(t)+A(t)^2))^{-1}$ and using $|\xi_i| \le \|\xi\|_1$, we infer from (6.64)

$$
\begin{aligned}
\frac{1}{2}\mathbb{E}\left[Z^{\frac{k}{2}}_{t\wedge T_j}\right] &\le \mathbb{E}\left[Z^{\frac{k}{2}}_{t\wedge T_j}\right] - \frac{1}{2}\mathbb{E}\left[j^{\frac{k}{2}}\wedge Z^{\frac{k}{2}}_{t\wedge T_j}\right] \\
&\le \tilde{C}(1+A(t)+A(t)^2)\Big(\mathbb{E}\left[\int_{(0,t]} 1+\left\|X_{(s\wedge T_j)-}\right\|^k_{\mathbb{R}^d}A(ds)\right] \\
&\quad + \mathbb{E}\left[\int_{(0,t\wedge T_j]}\int_{\mathbb{R}^d}\|\xi\|^k_{\mathbb{R}^d}K(s,X_{s-};d\xi)A(ds)\right]\Big),
\end{aligned}
$$

where $\tilde{C}$ contains all constants coming from the equivalence between $\|\cdot\|_{\mathbb{R}^d}$ and $\|\cdot\|_1$. Define

$$
\rho_3(t) = (1+A(t)+A(t)^2+\rho_1(t)+\rho_2(t)).
$$

Applying the above, together with (6.58) and (6.59), to (6.57) and changing order of integration due to non-negativity, yields

$$
\begin{aligned}
\mathbb{E}\left[\sup_{s\le t}\left|X^{T_j}_{i,s}\right|^k\right] &\le \tilde{C}\rho_3(t)\Big(1+\mathbb{E}\left[\|X_0\|^k_{\mathbb{R}^d}\right]+\mathbb{E}\left[C^{\frac{k}{2}}_{t\wedge T_j}\right]+\mathbb{E}\left[Z^{\frac{k}{2}}_{t\wedge T_j}\right] \\
&\quad + \int_{(0,t]}\mathbb{E}\left[\left\|X_{(s\wedge T_j)-}\right\|^k_{\mathbb{R}^d}\right]A(ds)\Big) \\
&\le \tilde{C}\rho_3(t)\Big(1+\mathbb{E}\left[\|X_0\|^k_{\mathbb{R}^d}\right]+\int_{(0,t]}\mathbb{E}\left[\left\|X_{(s\wedge T_j)-}\right\|^k_{\mathbb{R}^d}\right]A(ds) \\
&\quad + \mathbb{E}\left[\int_{(0,t\wedge T_j]}\int_{\mathbb{R}^d}\|\xi\|^k_{\mathbb{R}^d}K(s,X_{s-};d\xi)A(ds)\right]\Big).
\end{aligned} \tag{6.65}
$$

Note that if either term on the rhs of (6.54) is infinite, then the claim is always true. Hence, we can assume they are all finite. Since X and Z are càdlàg , they can not explode in finite time. That implies that for almost all $\omega \in \Omega$ we have $T_j(\omega) \to \infty$ as j goes to infinity. Hence, by monotone convergence we get

$$\lim_{j\to\infty} \mathbb{E}\left[\sup_{s\le t} |X_{i,s}^{T_j}|^k\right] = \mathbb{E}\left[\sup_{s\le t} |X_{i,s}|^k\right].$$

Since $j \mapsto \left\|X_{(s\wedge T_j)-}\right\|_{\mathbb{R}^d}^k$ is monotone, again by monotone convergence we have

$$\lim_{j\to\infty} \int_{(0,t]} \mathbb{E}\left[\left\|X_{(s\wedge T_j)-}\right\|_{\mathbb{R}^d}^k\right] A(ds) = \int_{(0,t]} \mathbb{E}\left[\|X_{s-}\|_{\mathbb{R}^d}^k\right] A(ds),$$

where we use that $\|\cdot\|_{\mathbb{R}^d}$ is continuous, $s\wedge T_j$ is a monotone increasing sequence converging to s and X_{s-} is left continuous. Finally, we have monotonicity of

$$j \mapsto \int_{(0,t\wedge T_j]} \int_{\mathbb{R}^d} \|\xi\|_{\mathbb{R}^d}^k K(s, X_{s-}; d\xi) A(ds).$$

Hence we have

$$\begin{aligned}
&\lim_{j\to\infty} \mathbb{E}\left[\int_{(0,t\wedge T_j]} \int_{\mathbb{R}^d} \|\xi\|_{\mathbb{R}^d}^k K(s, X_{s-}; d\xi) A(ds)\right] \\
&= \mathbb{E}\left[\int_{(0,t]} \int_{\mathbb{R}^d} \|\xi\|_{\mathbb{R}^d}^k K(s, X_{s-}; d\xi) A(ds)\right] \\
&= \int_{(0,t]} \mathbb{E}\left[\int_{\mathbb{R}^d} \|\xi\|_{\mathbb{R}^d}^k K(s, X_{s-}; d\xi)\right] A(ds),
\end{aligned}$$

where for the last step we use non-negativity of the integrand. By summing over the index $i \in \{1,\dots,d\}$, we now get with equivalence of norms and Jensen's inequality for sums,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{s\le t} \left\|X_s^{T_j}\right\|_{\mathbb{R}^d}^k\right] \le \tilde{C} \sum_{i=1}^d \mathbb{E}\left[\sup_{s\le t} \left|X_{i,s}^{T_j}\right|^k\right] \\
&\le \tilde{C}\rho_3(t)\Big(1 + \mathbb{E}\left[\|X_0\|_{\mathbb{R}^d}^k\right] + \int_{(0,t]} \mathbb{E}\left[\left\|X_{(s\wedge T_j)-}\right\|_{\mathbb{R}^d}^k\right] A(ds) \\
&+ \mathbb{E}\left[\int_{(0,t\wedge T_j]} \int_{\mathbb{R}^d} \|\xi\|_{\mathbb{R}^d}^k K(s, X_{s-}; d\xi) A(ds)\right]\Big).
\end{aligned} \tag{6.66}$$

Hence, (6.54) follows by applying monotone convergence to (6.66).

We continue with showing the integrability of the running supremum. Assume we have $\mathbb{E}\left[\|X_0\|_{\mathbb{R}^d}^k\right] < \infty$ and consider (6.66). We can bound the last term and apply Fubini-Tonelli to get

$$\begin{aligned}
\mathbb{E}\left[\sup_{s\le t}\left\|X_s^{T_j}\right\|_{\mathbb{R}^d}^k\right] &\le \tilde{C}\rho_3(t)\Big(1+\mathbb{E}\left[\|X_0\|_{\mathbb{R}^d}^k\right]+\int_{(0,t]}\mathbb{E}\left[\left\|X_{(s\wedge T_j)-}\right\|_{\mathbb{R}^d}^k\right]A(ds) \\
&\quad +\mathbb{E}\left[\int_{(0,t]}\int_{\mathbb{R}^d}\|\xi\|_{\mathbb{R}^d}^k\,K(s,X_{s-};d\xi)A(ds)\right]\Big) \\
&= \tilde{C}\rho_3(t)\Big(1+\mathbb{E}\left[\|X_0\|_{\mathbb{R}^d}^k\right]+\int_{(0,t]}\mathbb{E}\left[\left\|X_{(s\wedge T_j)-}\right\|_{\mathbb{R}^d}^k\right]A(ds) \\
&\quad +\int_{(0,t]}\mathbb{E}\left[\int_{\mathbb{R}^d}\|\xi\|_{\mathbb{R}^d}^k\,K(s,X_{s-};d\xi)\right]A(ds)\Big).
\end{aligned} \tag{6.67}$$

Assume (6.55) holds. Then the function ρ_4, defined by

$$\rho_4(t)=\int_{(0,t]}\mathbb{E}\left[\int_{\mathbb{R}^d}\|\xi\|_{\mathbb{R}^d}^k\,K(u,X_{u-};d\xi)\right]A(du),$$

is finite for all $t\ge 0$. Note that $\rho_j, j=3,4$ are non-decreasing. Hence we get the estimate for all $0\le t\le T$,

$$\begin{aligned}
\mathbb{E}\left[\sup_{s\le t}\left\|X_s^{T_j}\right\|_{\mathbb{R}^d}^k\right] &\le \tilde{C}\rho_3(t)\Big(1+\mathbb{E}\left[\|X_0\|_{\mathbb{R}^d}^k\right]+\rho_4(t)\Big) \\
&\quad +\tilde{C}\rho_3(t)\int_{(0,t]}\mathbb{E}\left[\left\|X_{(s\wedge T_j)-}\right\|_{\mathbb{R}^d}^k\right]A(ds) \\
&\le K+K\int_{(0,t]}\mathbb{E}\left[\sup_{u\le s}\left\|X_{(u\wedge T_j)-}\right\|_{\mathbb{R}^d}^k\right]A(ds),
\end{aligned}$$

where $K=\tilde{C}\rho_3(T)(1+\mathbb{E}\left[\|X_0\|_{\mathbb{R}^d}^k\right]+\rho_4(T))$. Since $\left\|X_{(s\wedge T_j)-}\right\|\le j$ for all $s\ge 0$, we have due to $A(t)<\infty$ that the rhs is finite. By monotone convergence we get that the running supremum is càdlàg since X is càdlàg . Hence, by Gronwall's inequality, we get for all $0\le t\le T$,

$$\mathbb{E}\left[\sup_{s\le t}\left\|X_s^{T_j}\right\|_{\mathbb{R}^d}^k\right]\le K\exp\left(KA(T)\right)<\infty. \tag{6.68}$$

The upper bound is independent for all j and hence with monotone convergence we get for all $t\le T$,

$$\mathbb{E}\left[\sup_{s\le t}\|X_s\|_{\mathbb{R}^d}^k\right]\le K\exp\left(KA(T)\right)<\infty. \tag{6.69}$$

Since T was arbitrary, the claim is shown. We are left with assuming (6.56) instead of

(6.55). Let $\widehat{C}$ be the constant from (6.56) for T. In that case, (6.66) becomes

$$\begin{aligned}
&\mathbb{E}\left[\sup_{s\leq t}\left\|X_s^{T_j}\right\|_{\mathbb{R}^d}^k\right] \leq \tilde{C}\rho_3(T)\Big(1+\mathbb{E}\left[\|X_0\|_{\mathbb{R}^d}^k\right]\Big)\\
&+\tilde{C}\rho_3(T)\int_{(0,t]}\mathbb{E}\left[\left\|X_{(s\wedge T_j)-}\right\|_{\mathbb{R}^d}^k\right]A(ds)+\widehat{C}\mathbb{E}\left[\int_{(0,t\wedge T_j]}(1+\|X_{s-}\|_{\mathbb{R}^d}^k)A(ds)\right]\\
&\leq \tilde{C}\rho_3(T)\Big(1+\mathbb{E}\left[\|X_0\|_{\mathbb{R}^d}^k\right]\Big)+\tilde{C}\rho_3(T)\int_{(0,t]}\mathbb{E}\left[\sup_{u\leq s}\left\|X_u^{T_j}\right\|_{\mathbb{R}^d}^k\right]A(ds)\\
&+\widehat{C}\mathbb{E}\left[\int_{(0,t]}(1+\left\|X_{(s\wedge T_j)-}\right\|_{\mathbb{R}^d}^k)A(ds)\right]\leq K+K\int_{(0,t]}\mathbb{E}\left[\sup_{u\leq s}\left\|X_u^{T_j}\right\|_{\mathbb{R}^d}^k\right]A(ds),
\end{aligned}$$

where $K=(\tilde{C}\rho_3(T)+\widehat{C})(1+\mathbb{E}\left[\|X_0\|_{\mathbb{R}^d}^k\right])$ By the exact same arguments as before, the rhs is finite and since $\widehat{C}$ is only depending on T and not on t, Gronwall's inequality yields (6.68). Again by monotone convergence we get (6.69), which finishes the proof since T was arbitrary. □

This lemma now allows us to state the main theorem from an application point of view. It shows that polynomial semimartingales can be embedded in a polynomial process if the jump measure satisfies a certain integrability condition. In particular condition (6.56) is useful, as it is a statement about a deterministic kernel.

We say a k-polynomial semimartingale can be embedded in a k-polynomial process, if there are matrix functions P, G and a $A\in\mathcal{V}_{\mathfrak{N}}^+$ such that (X,P,A,G) is a bounded and uniformly integrable k-polynomial process.

Theorem 6.57. *Let X be a k-polynomial semimartingale and assume $\mathbb{E}\left[\|X_0\|_{\mathbb{R}^d}^k\right]<\infty$. If either* (6.55) *or* (6.56) *hold, then X can be embedded in a k-polynomial process.*

Proof. Let $A\in\mathcal{V}_{\mathfrak{N}}^+$ be the function that exists by the definition of k-polynomial semimartingales. By Proposition 6.54, we get that Condition 6.43 holds for A and k where $k\geq 2$ by the definition of k-polynomial semimartingales. By Lemma 6.56, we get that the running supremum is integrable. Hence, just as in the proofs of Theorem 6.44 and Lemma 6.45, we get that the sets $\mathcal{H}_t^f$ are uniformly integrable for all $f\in\mathrm{Pol}_k$ and $t\geq 0$ and that for all $f\in\mathrm{Pol}_k$, M^f is a true martingale. The rest of the claim now follows from Theorem 6.41. □

Remark 6.58. *By [JS13, Proposition II.2.9] we have that for a k-polynomial semimartingale X with function $A\in\mathcal{V}_{\mathfrak{N}}^+$ from the definition of k-polynomial semimartingales, that the points of discontinuity of A correspond to the points where X fails to be quasi-left continuous. In particular this implies stochastic discontinuity*

Let us finish this section with an example of a polynomial process X, that is a polynomial semimartingale, where X is not stochastically continuous.

Example 6.59. *Let Z, W be two independent Brownian motions wrt their respective natural filtrations on the probability space $(\Omega, \mathscr{F}, P)$, both starting in zero. Let X_0 be a on $[0,1]$ uniformly distributed random variable on this probability space independent of Z and W. For each $t \geq 0$, we define the σ-algebras*

$$\mathcal{F}_t = \sigma(X_0, (Z)_{s\in[0,t]}) \text{ if } t < 1 \text{ and } \mathcal{F}_t = \sigma(\mathcal{F}_1, (W)_{s\in[1,t)}) \text{ for } t \geq 1.$$

Then $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ forms a right-continuous filtration. After completion, we get that $(\Omega, \mathscr{F}, \mathbb{F}, P)$ forms a stochastic basis.

Define the process X with initial value X_0 via

$$X_t := X_0 + \mathbb{1}_{\{t<1\}} Z_t + \mathbb{1}_{\{t\geq 1\}} W_t.$$

Noting that X is adapted and for $s < 1 \leq t$, we have $\mathcal{F}_s$ is independent of X_t, we get that X is a polynomial process since Z and W are Brownian motions and hence satisfy the polynomial property in Definition 6.4. Further, we have $P(\Delta X_1 \neq 0) = 1$, hence X fails to be stochastically continuous at $t = 1$.

Next, note that $\mathbb{1}_{\{t<1\}}, \mathbb{1}_{\{t\geq 1\}}, Z\mathbb{1}_{\{t<1\}}$ and $W\mathbb{1}_{\{t\geq 1\}}$ are semimartingales on our stochastic basis, which implies that X is a semimartingale. Since the indicator functions are of finite variation and predictable, we compute with the integration by parts formula ([He+92, Corollary 9.34]),

$$\begin{aligned}
\mathbb{1}_{\{\cdot<1\}} Z = \mathbb{1}^2_{\{\cdot<1\}} Z &= (\mathbb{1}_{\{\cdot<1\}} \bullet Z\mathbb{1}_{\{\cdot<1\}}) + \int_{(0,\cdot]} Z_u \mathbb{1}_{\{u<1\}} d(\mathbb{1}_{\{u<1\}}) \\
&= (\mathbb{1}_{\{\cdot<1\}} \bullet Z) + \int_{(0,\cdot]} Z_u d(\mathbb{1}_{\{u<1\}}),
\end{aligned}$$

and similarly

$$\mathbb{1}_{\{\cdot\geq 1\}} W = (\mathbb{1}_{\{\cdot\geq 1\}} \bullet W) + \int_{(0,\cdot]} W_u d(\mathbb{1}_{\{u\geq 1\}}).$$

Hence we get that X has a semimartingale decomposition $X = X_0 + M + F$ for local martingale part

$$M = (\mathbb{1}_{\{\cdot<1\}} \bullet Z) + (\mathbb{1}_{\{\cdot\geq 1\}} \bullet W)$$

and finite variation part

$$F = \int_{(0,\cdot]} Z_u d(\mathbb{1}_{\{u<1\}}) + \int_{(0,\cdot]} W_u d(\mathbb{1}_{\{u\geq 1\}}) = (W_1 - Z_1)\mathbb{1}_{\{\cdot\geq 1\}}.$$

In particular, this implies $F \in \mathcal{A}_{loc}$, which by [JS13, Proposition I.4.23 (ii)] yields that X is a special semimartingale. Let us compute it's canonical decomposition. We define F^1 and F^2 by

$$F = W_1 \mathbb{1}_{\{\cdot\geq 1\}} + \left(-Z_1 \mathbb{1}_{\{\cdot\geq 1\}}\right) =: F^1 + F^2.$$

Then we get that F^2 is predictable, since Z is predictable. Note that this is not true for F^1 since W_s is independent of $\mathcal{F}_s$ for $s < 1$. A direct computation shows that

$$\mathbb{E}\left[F_t^1 \mid \mathcal{F}_s\right] = \begin{cases} 0 & = F_s^1 \text{ , for } s \leq t < 1 \\ \mathbb{E}\left[W_t \mid \mathcal{F}_s\right] \mathbb{1}_{\{t \geq 1\}} = \mathbb{E}\left[W_t\right] \mathbb{1}_{\{t \geq 1\}} = 0 & = F_s^1 \text{ , for } s < 1 \leq t \\ \mathbb{E}\left[W_t \mid \mathcal{F}_s\right] \mathbb{1}_{\{t \geq 1\}} = W_s \mathbb{1}_{\{t \geq 1\}} = W_s \mathbb{1}_{\{s \geq 1\}} & = F_s^1 \text{ , for } 1 \leq s \leq t. \end{cases}$$

Hence we have that F^1 is a martingale and since it is piecewise constant, it is purely discontinuous. This gives us the canonical decomposition

$$X = X_0 + M^c + M^d + V,$$

where $M^c = M, M^d = F^1$ and $V = F^2$. Note that $M^c = M$ is continuous. Hence we get using the independence of the two Brownian motions and the fact that the Lebesgue measure is without atoms,

$$C := \langle M, M \rangle = \int_{(0,\cdot]} (\mathbb{1}_{\{u<1\}} + \mathbb{1}_{\{u \geq 1\}}) du = \int_{(0,\cdot]} (\mathbb{1}_{\{u<1\}} + \mathbb{1}_{\{u>1\}}) du.$$

Define the increasing function $A(t) := t + \mathbb{1}_{\{t \geq 1\}}$. Defining $\tilde{c}_t := \mathbb{1}_{\{u<1\}} + \mathbb{1}_{\{u>1\}}$, we then get using $\tilde{c}_1 = 0$,

$$C = \int_{(0,\cdot]} \tilde{c}_u A(du).$$

In particular we have for $c(u,x) := \tilde{c}_u$ that $c \in \mathrm{Pol}_0^{dA}(\mathbb{R}_{>0} \times \mathbb{R}^d) \subset \mathrm{Pol}_2^{dA}(\mathbb{R}_{>0} \times \mathbb{R}^d)$. Next, note that Z is continuous. Therefore we have $X_{1-} = Z_{1-} = Z_1$. Define $b \in \mathrm{Pol}_1^{dA}(\mathbb{R}_{>0} \times \mathbb{R}^d)$ as

$$b(u,x) = -x \mathbb{1}_{\{1\}}(u).$$

Hence we get

$$F_t^2 = -Z_1 \mathbb{1}_{\{t \geq 1\}} = \int_{(0,t]} b(u, X_{u-}) A(du).$$

We continue with

$$\mu^X(dt, d\xi) = \varepsilon_{\{1, \Delta X_1\}}(dt, d\xi).$$

Since we have $\Delta X_t = \mathbb{1}_{\{1\}}(W_1 - Z_1)$, we get the decomposition

$$\mu^X(dt, d\xi) = \left(\varepsilon_{\{1, W_1\}} - \varepsilon_{\{1, Z_1\}}\right)(dt, d\xi).$$

Hence, as Z is predictable and W is independent of $\mathcal{F}_{1-}$, see (6.21), we get that the predictable compensator ν of μ^X is given by

$$\nu(dt, d\xi) = -\varepsilon_{\{1, Z_1\}}(dt, d\xi),$$

compare [JS13, Theorem II.1.8], and note that the function W in that theorem (not the Brownian motion in our example) has to be $\tilde{\mathcal{P}}$ measurable. Define $K(u, x; d\xi) := -\mathbb{1}_{\{1\}(u)} \varepsilon_{\{x\}} d\xi$. Then, since $X_{1-} = Z_1$, we get

$$\nu(dt, d\xi) = K(u, X_{u-}; d\xi) A(du).$$

In particular this gives us

$$\int_{\mathbb{R}} \xi^k K(u,x;d\xi) = -x^k \mathbb{1}_{\{1\}}(u) \in \mathrm{Pol}_k^{dA}(\mathbb{R}_{>0} \times \mathbb{R}^d).$$

Since this implies for $j \in \{0,1,2\}$,

$$a(u,x) = c(u,x) + \int_{\mathbb{R}} \xi^j K(u,x;d\xi) \in \mathrm{Pol}_2^{dA}(\mathbb{R}_{>0} \times \mathbb{R}^d),$$

we get that X *is a polynomial semimartingale with characteristics* (B,C,ν) *for* $B = F^2$. *In particular, since* (6.56) *holds, Theorem 6.57 yields that* X *has the polynomial property, an observation we made at the beginning of this example since this can be seen from elementary computations.*

6.5 The discrete time case

In this section we want to briefly discuss the setting for polynomial processes in discrete time. It turns out that the discrete time setting can be seen as a special case of polynomial processes. For the rest of this section, We fix a filtered probability space $(\Omega, \mathscr{F}, \mathbb{F}, P)$, where the filtration $\mathbb{F}$ is in discrete time. Hence we have $\mathbb{F} = (\mathcal{F}_n)_{n\in\mathbb{N}_0}$ and $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ for all $n \in \mathbb{N}_0$. Whenever we say a process X is in discrete time, we mean that it is indexed by $\mathbb{N}_0$. Hence $X = (X_n)_{n\in\mathbb{N}_0}$.

Definition 6.60. *Let* X *be a process in discrete time. Define the piecewise constant process* Y*, indexed by* $\mathbb{R}_+$*, via*

$$Y_t := X_n, \text{ for } t \in [n, n+1).$$

We call the process Y *the* **natural embedding** *of* X *in continuous time. In the same way, for the filtered probability space* $(\Omega, \mathscr{F}, \mathbb{F}, P)$ *in descrete time, define the filtration* $\tilde{\mathbb{F}} = (\tilde{\mathcal{F}}_t)_{t\in\mathbb{R}_+}$ *via*

$$\tilde{\mathcal{F}}_t := \mathcal{F}_n, \text{ for } t \in [n, n+1).$$

We call the filtered probability space $(\Omega, \mathscr{F}, \tilde{\mathbb{F}}, P)$ *the* **natural embedding** *of the discrete time filtered probability space* $(\Omega, \mathscr{F}, \mathbb{F}, P)$ *in continuous time.*

By construction, a discrete time process X is adapted wrt a discrete time filtered probability space, if and only if it's natural embedding in continuous time is adapted wrt the natural embedding of the filtered probability space. Further, the natural embedding of the filtration is right-continuous by construction. Hence, if the discrete time filtered probability space is complete, then the natural embedding forms a stochastic basis.[2]

To make the following discussions simpler, fix the natural embeddings $(\Omega, \mathscr{F}, \tilde{\mathbb{F}}, P)$ and Y of $(\Omega, \mathscr{F}, \mathbb{F}, P)$ and X.

[2] Recall that in this thesis we defined a stochastic basis to always satisfy the usual conditions, compare Definition 3.1.

Definition 6.61. *We say that an adapted process X has the k-polynomial property, if Y has the k-polynomial property wrt $(\Omega, \mathscr{F}, \tilde{\mathbb{F}}, P)$.*

The next proposition shows that in the discrete time case, all embeddings have a very nice structure, allowing for a complete discussion of X by considering the respective embeddings.

Proposition 6.62. *Let X be a discrete time and adapted process. If X has the k-polynomial property, then there are functions $A \in \mathcal{V}^{+}_{\mathscr{A}}$, $G : \mathbb{R}_{>0} \to M_{N_k}(\mathbb{R})$, $P : \mathbb{T} \to M_{N_k}(\mathbb{R})$ such that (Y, P, A, G) is a bounded and uniformly integrable k-polynomial process.*

Proof. We start by constructing the map $\tau : \mathbb{T} \times \mathrm{Pol}_k \to \mathrm{Pol}_k$ such that (Y, τ) is a càdlàg k-polynomial process. By the properties of the conditional expectation, it is always possible to set $\tau_{s,t} = \mathrm{id}_{\mathrm{Pol}_k}$ for all $s \geq 0$ and $t \in [s, \lfloor s \rfloor + 1)$, since $Y_{\lfloor t \rfloor} = Y_t$. By Proposition 6.10 there exists a linear mapping $\tau_{s,\lfloor s \rfloor+1}$ on Pol_k such that Equation (6.4) holds for $q^f_{s,\lfloor s \rfloor+1} = \tau_{s,\lfloor s \rfloor+1} f$. Hence we have defined τ for indices s, t with $n \leq s \leq t \leq n+1$ for all $n \in \mathbb{N}_0$. Further, since τ is piecewise constant, we have that the evolution property is satisfied. We follow Remark 6.9 and extend τ to all of $\mathbb{T}$. For that, let $t > \lfloor s \rfloor + 1$. Then there is a unique $l \in \mathbb{N}$ such that $\lfloor s \rfloor + l < t \leq \lfloor s \rfloor + l + 1$. Define

$$\tau_{s,t} := \tau_{s,\lfloor s \rfloor+1} \circ \tau_{\lfloor s \rfloor+1,\lfloor s \rfloor+2} \circ \cdots \circ \tau_{\lfloor s \rfloor+l,t}.$$

Hence, by definition τ is piecewise constant and càdlàg . Further, again by definition, τ satisfies the evolution property. We get that τ is an evolution system for X on Pol_k. Define the representing MES $P = \iota_k(\tau)$. Then (X, P) is a càdlàg k-polynomial process. Next, define $A \in \mathcal{V}^{+}_{\mathscr{A}}$ via

$$A(t) := \lfloor t \rfloor.$$

Since P is piecewise constant, it is of finite variation and defines a family of measures γ such that $\gamma \overset{\infty}{\ll} dA$. And we get that (P, γ, A) is a proper MES in dimension N_k. Further, note that the sequence t_n in Proposition 4.31 is given by $t_n = n$. As $dA(\mathbb{R}_+ \setminus \mathbb{N}_0) = 0$, we get that there exists an extended generator G of the proper matrix evolution system. Further, we can always set G to be zero on $\mathbb{R}_+ \setminus \mathbb{N}_0$. Hence, G is only non zero in $\mathbb{N}_0$ or equivalently $\Delta A_t \neq 0$. As there are only finitely many such values on all compact intervals and G is given by Equation (4.22), we get that $\|G\|_{dA,t} < \infty$ for all $t \geq 0$. Next, let $f \in \mathrm{Pol}_k$ be arbitrary. Note that by definition of the embedding Y, we have for all $t \geq 0$,

$$\mathcal{H}^f_t = \{f(X_n) : 0 \leq n \leq \lfloor t \rfloor\}.$$

Since each element of $\mathcal{H}^f_t$ is integrable by the polynomial property and we are dealing with a finite set, we get that the supremum over this set is integrable, yielding uniform integrability of $\mathcal{H}^f_t$. Overall, this yields that (Y, P, A, G) is a bounded and uniformly integrable k-polynomial process, finishing the proof. □

Appendix A

Existence of càdlàg modifications

A.1 Modifications and indistinguishability

We shortly present results on càdlàg modifications of stochastic processes as we will be needing them in the following. We fix a stochastic basis $(\Omega, \mathscr{F}, \mathbb{F}, P)$, in particular this means by our definition from the preliminaries, that property (iii) in Definition 3.1 holds.

Definition A.1. *Two stochastic processes X and Y are called* **versions** *of each other (or* **modification***), if for all $t \in I$ we have*

$$X_t = Y_t \quad [P]. \tag{A.1}$$

We say they are **indistinguishable** *or equal up to evanescence if they are equal, simultaneously for all t outside an evanescent set, i.e*

$$\{X \neq Y\} = \{(\omega, t) : X_t(\omega) \neq Y_t(\omega)\}$$

is an evanescent set. This means they have almost surely the same sample paths.

Remark A.2. *Note that if X and Y are versions of each other, this does not imply that they are indistinguishable. However, if X and Y are càdlàg processes, then it is a well known result that they are indistinguishable as the càdlàg version of a process is unique up to evanescence.*

Theorem A.3. *Let X and Y be two stochastic processes such that X is a modification of Y. If X and Y are right continuous, then X and Y are indistinguishable.*

Proof. See Protter [Pro05, Theorem I.2]. We will give a proof of the case where the process is a two parameter random field. □

Remark A.4. *The crucial observation made above is that if one poses sufficient regularity on the paths of a process, such that the process is fully determined by a countable subset of* $\mathbb{R}_+$*, then one can construct a set of measure zero outside which two modifications agree. This concept plays an important role in different areas in the theory of stochastic processes, e.g. in the study of Gaussian processes or more generally Gaussian random fields where it is known as separability, compare e.g. [Adl85; Bil74]. For us, it is sufficient to assume càdlàg regularity which makes statements regarding indistinguishability easy to prove.*

We will later need the following definition regarding the càdlàg property of a two parameter random field. We define this property in the spirit of how we did for matrix evolution systems. Recall that we define the set $\mathbb{T}$ by

$$\mathbb{T} = \left\{(s,t) \in \mathbb{R}^2 : 0 \leq s \leq t\right\}.$$

Definition A.5. *Consider a two parameter random field* $X : \mathbb{T}\times\Omega \Rightarrow \mathbb{R}^d$ *on a probability space* $(\Omega, \mathscr{F}, P)$*. We call* X *càdlàg (right continuous), if there exists a measurable set* $N \subset \Omega$ *with* $P(N) = 1$*, s.t. for all* $\omega \in N$*,*

(i) and all fixed $s \geq 0$*, the functions*

$$t \mapsto X_{(s,t)}(\omega),$$

are càdlàg (right continuous) in t *on* $[s, \infty)$*,*

(ii) and for all $t > 0$*,*

$$s \mapsto X_{(s,t)}(\omega),$$

are càdlàg (right continuous) in s *on* $[0, t]$*.*

For completeness, we shall define the concept of modification and indistinguishability for two parameter processes. We shall use the notation $(s,t) = \mathbf{t} \in \mathbb{T}$, then $X_{\mathbf{t}} = X_{(s,t)}$.

Definition A.6. *Given two random fields* X *and* Y *indexed by* $\mathbb{T}$*, we say* X *and* Y *are* **versions** *(or* **modifications***) of each other, if for any* $\mathbf{t} \in \mathbb{T}$*,*

$$X_{\mathbf{t}} = Y_{\mathbf{t}} \quad [P].$$

If,

$$P(\omega \in \Omega : X_{\mathbf{t}}(\omega) = Y_{\mathbf{t}}(\omega)\ \forall \mathbf{t} \in \mathbb{T}) = 1,$$

we say they are **indistinguishable***.*

The next proposition is a two-parameter random field version of Theorem A.3. The proof is essentially the same as for the one-parameter case but we will give it for the readers convenience.

Proposition A.7. *Given a complete probability space $(\Omega, \mathscr{F}, P)$ and two $\mathbb{T}$ indexed random fields X and Y, assume they are both right continuous. If X is a version of Y, then they are indistinguishable.*

Proof. Note that by the completeness assumption, all sets of probability zero are measurable. We follow the idea in Protter [Pro05, Theorem I.2] to the letter. Let M_1 be the set of probability zero that X is right-continuous for all $\omega \in M_1^c$ and M_2 the corresponding one for Y. Then, $M := M_1 \cup M_2$ is again a set of probability zero and for each $\omega \in M^c$, $X(\omega)$ and $Y(\omega)$ are right continuous. Define $\mathbb{T}^{\mathbb{Q}} = \mathbb{T} \cap \mathbb{Q}$. Then, because X is a version of Y, we have for any $\mathbf{q} \in \mathbb{T}^{\mathbb{Q}}$, that there exists a set $N_{\mathbf{q}} \subset \Omega$ with $P(N_{\mathbf{q}}) = 0$ such that

$$X_{\mathbf{q}}(\omega) = Y_{\mathbf{q}}(\omega) \quad \forall \omega \in N_{\mathbf{q}}^c .$$

Define $N = \bigcup_{q \in \mathbb{T}^{\mathbb{Q}}} N_{\mathbf{q}}$. Then $P(N) = 0$. Finally, define now $K = N \cup M$. Once again, $P(K) = 0$. Further, for $\omega \in K^c$, it holds that $X(\omega)$ and $Y(\omega)$ are right continuous and for all $q \in \mathbb{T}^{\mathbb{Q}}$,

$$X_{\mathbf{q}}(\omega) = Y_{\mathbf{q}}(\omega).$$

Since $P(K) = 1$, if we can show $X_{\mathbf{t}}(\omega) = Y_{\mathbf{t}}(\omega)$ for all $\omega \in K^c$ and $\mathbf{t} \in \mathbb{T}$, we are done. Let now $(s,t) = \mathbf{t} \in \mathbb{T}$. Then there are decreasing sequences $(s_n) \downarrow s$ and $(t_m) \downarrow t$ with $(s_n), (t_m) \subset \mathbb{T}^{\mathbb{Q}}$. If s is rational, we can simply choose $s_n = s$ and the same goes for t. If s is rational, but t is not, we get by right continuity $X_{\mathbf{t}}(\omega) = Y_{\mathbf{t}}(\omega)$ by just taking the limit in (s_n). The same is true if t is rational but s is not. Assume now that both are not rational. Then for any $m \in \mathbb{N}$, there exists a l_m such that $s_{l_m+n} \leq t_m$ for all $n \geq 0$. Hence we can compute the limit for any $m \in \mathbb{N}$, $\omega \in K$,

$$X_{s,t_m}(\omega) = \lim_{n\to\infty} X_{(s_n,t_m)}(\omega) = \lim_{n\to\infty} Y_{(s_n,t_m)}(\omega) = Y_{(s,t_m)}(\omega).$$

This in return now implies

$$X_{s,t}(\omega) = \lim_{m\to\infty} X_{(s,t_m)}(\omega) = \lim_{m\to\infty} Y_{(s,t_m)}(\omega) = Y_{(s,t)}(\omega).$$

Hence we have shown for all $\omega \in K$, that

$$X_{\mathbf{t}}(\omega) = Y_{\mathbf{t}}(\omega) \quad \forall \mathbf{t} \in \mathbb{T}.$$

Since $P(K) = 1$, X is indistinguishable from Y as claimed. □

A.2 Existence of càdlàg modifications

We will now deal with the situation where a process is defined via conditional expectations for each time point. The problem is then that the conditional expectation is only defined up to a null set. As the index set is in general uncountable, we face the issue that for a given measure P, the union of uncountable P-null sets is not necessarily a P-null set again. A similar problem arises in filtering theory where one considers a process X that is not adapted to the filtration $\mathbb{F}$. One then considers a process Y that is defined by

$$Y_t = \mathbb{E}\left[X_t|\mathcal{F}_t\right].$$

The problem is now that this definition does not specify the whole process outside a P-null set. In particular, for a stopping time τ, the random variable Y_τ is not well defined almost surely. The same holds true for functionals of the whole process such as the quadratic variation or integrals alongside a path. In filtering theory this issue is addresses by considering the optional projection of X, which guarantees uniqueness of the chosen version. The interested reader might want to consider Bain and Crisan [BC08].

Our setting is different however, e.g. we are considering a process that is adapted. To be concrete, given a stochastic basis $(\Omega, \mathscr{F}, \mathbb{F}, P)$, and an adapted process X, we define for each $s \geq 0$ the process $Y^s : [s, \infty) \times \Omega \to \mathbb{R}^d$ pointwise via

$$Y^s_t = \mathbb{E}\left[X_t \mid \mathcal{F}_s\right].$$

As mentioned above, since there are uncountable many $t \geq s$, this process is not well defined. If however, one can show that there exists a càdlàg modification, then we have seen that the above definition restricted to the càdlàg modification is well defined up to indistinguishability.

One way to treat the general setting would be by means of Kolmogorow's continuity criterion (actually the càdlàg version which is a straight forward modification). In our case however, we can not assume Hölder regularity type bounds on the expectation making this approach unavailable to us. Fortunately, it will be sufficient for our purposes to assume significantly more on the process X, namely semimartingality with a semimartingale decomposition where the local martingale is actually a true martingale, allowing us to show existence of a càdlàg modification with elementary tools for each $s \geq 0$ fixed. In the context of polynomial processes, this will actually then imply the existence of a unique (up to indistinguishability) modification of the random field

$$\mathbb{T} \ni (s, t) \mapsto Y^s_t = \mathbb{E}\left[X_t \mid \mathcal{F}_s\right],$$

motivating the discussions in the previous section.

The goal of this section is to prove the following theorem.

Theorem A.8. *Given a stochastic basis* $(\Omega, \mathscr{F}, \mathbb{F}, P)$*, consider a semimartingale* X *on this basis with decomposition* $X = X_0 + M + A$ *for a local martingale* M *and and process of finite variation* $A \in \mathcal{V}$ *with* $\mathbb{E}[\mathrm{Var}(A)_t] < \infty$ *for all* $t \geq 0$*. Assume* M *is a true martingale. Then, for all fixed* $s \geq 0$*, the process*

$$t \mapsto \mathbb{E}[X_t \mid \mathcal{F}_s], \quad t \geq s,$$

has a càdlàg modification that is of finite variation.

Remark A.9. *Let us note that it is possible to proof a more general result of the theorem above where the càdlàg process only needs to be uniformly integrable. This can be done by means of disintegrating the conditional expectation and considering the conditional distribution on appropriate path spaces, see Kallenberg [Kal06, Theorem 6.3]. Since for our purposes, the assumptions made above are sufficient, we opt to take the more direct route and prove the above theorem.*

Let us recall the following fundamental property of the conditional expectation. For the proof, we refer e.g. to Kallenberg [Kal06, Theorem 6.1].

Theorem A.10. *Given a probability space* $(\Omega, \mathscr{F}, P)$*, let* $\mathcal{A} \subset \mathcal{F}$ *be a sub-*σ*-algebra. Then, for all random variables* $\xi \in \mathcal{L}^1(P)$ *with* $\xi \geq 0$ *a.s. we have*

$$\mathbb{E}[\xi \mid \mathcal{A}] \geq 0 \quad [P].$$

We will now prove an intermediary lemma to Theorem A.8.

Lemma A.11. *Given a probability space* $(\Omega, \mathscr{F}, P)$*, let* A *be a càdlàg and nondecreasing* $\mathbb{R}_+$*-valued stochastic process with* $A_0 = 0$ *and* $\mathbb{E}[A_t] < \infty$ *for all* $t \geq 0$*. Then, for any sub-*σ*-algebra* $\mathcal{A} \subset \mathcal{F}$*, there exists a càdlàg modification of the process*

$$t \mapsto \mathbb{E}[A_t \mid \mathcal{A}], \tag{A.2}$$

that is nondecreasing with probability one.

Proof. Clearly, it is sufficient to show existence for arbitrary finite time horizon $T > 0$. Indeed, Since for $T' > T$, a càdlàg modification restricted to $[0, T]$ is indistinguishable from the càdlàg modification with time horizon T. Hence, one can "glue" these modifications to a unique modification on the whole real half-line $\mathbb{R}_+$. Fix $T > 0$. Let Y be any version of (A.2). We will show in a first step that there exists a set $\Delta \in \mathcal{A}$ with $P(\Delta^c) = 0$ such that the process $Y_t(\omega)\mathbb{1}_\Delta(\omega)$ is a modification on a countable dense subset of $[0, T]$ such that $Y\mathbb{1}_\Delta$ is nondecreasing on that dense subset. Note that since Δ^c is a zero-set, we get that $Y\mathbb{1}_\Delta$ is indistinguishable from Y, hence still a modification of (A.2) for any $t \geq 0$.

For $n \in \mathbb{N}$, denote by $\mathcal{D}_n$ the dyadic subset of $[0, T]$ defined by

$$\mathcal{D}_n = \left\{0, \frac{T}{2^n}, \frac{2T}{2^n}, \dots, \frac{(2^n-1)T}{2^n}, T\right\}.$$

Recall that $|\mathcal{D}_n| = 2^n + 1$ and $\mathcal{D}_n \subset \mathcal{D}_{n+1}$. Further, the set $\mathcal{D} := \bigcup_{n\in\mathbb{N}} \mathcal{D}_n$ is dense in $[0, T]$. Further, for any $p, q \in \mathcal{D}$, there exists a finite n such that $p, q \in \mathcal{D}_m$ for all $m \geq n$.

Note that for any $t \geq 0$, we have $Y_t \in \mathcal{A}$. Since by convention, we equip $\mathbb{R}_+$ with the corresponding Borel-σ-algebra, we get for any s, t,

$$\{\omega \in \Omega : Y_t(\omega) - Y_s(\omega) \geq 0\} \in \mathcal{A}.$$

By linearity of the conditional expectation, we have that $Y_t - Y_s$ is a version of

$$\mathbb{E}\left[A_t - A_s \mid \mathcal{A}\right].$$

Further, since $A_t - A_s$ a.s. by assumption, we get that

$$P(\omega \in \Omega : Y_t(\omega) - Y_s(\omega) \geq 0) = 1.$$

Consider now for each $n \in \mathbb{N}$ the set

$$\mathcal{D}_n^{\leq} = \left\{(m, n) \in \mathcal{D}_n^2 : m \leq n\right\}.$$

Note that we have $|\mathcal{D}_n^{\leq}| < \infty$ for all $n \in \mathbb{N}$. For each $(m, l) \in \mathcal{D}_n^{\leq}$, define the sets

$$\Delta_{m,l} := \{\omega \in \Omega : Y_l(\omega) - Y_m(\omega) \geq 0\}.$$

As noted before we have $\Delta_{m,l} \in \mathcal{A}$ and $P(\Delta_{m,l}) = 1$ or equivalently $P(\Delta_{m,l}^c) = 0$. Next, define

$$\Delta_n := \bigcap_{(m,l)\in\mathcal{D}_n^{\leq}} \Delta_{m,l} = \left(\bigcup_{(m,l)\in\mathcal{D}_n^{\leq}} \Delta_{m,l}^c\right)^c.$$

Hence, $\Delta_n \in \mathcal{A}$ for all $n \in \mathbb{N}$ and $P(\Delta_n^c) = 0$. Finally, define $\Delta := \bigcap_{n\geq 1} \Delta_n$. Again, we have $\Delta \in \mathcal{A}$ and $P(\Delta^c) = 0$. By construction, we have for all $\omega \in \Delta$,

$$Y_p(\omega) - Y_q(\omega) \geq 0 \quad \forall p, q \in \mathcal{D}, q \leq p.$$

Hence, Δ is the desired zero set on which $Y\mathbb{1}_\Delta$ is nondecreasing on $\mathcal{D}$. We will now define the desired càdlàg and nondecreasing modification of (A.2), denoted by $\tilde{Y}$ on $[0, T]$. For $q \in \mathcal{D}$, define $\tilde{Y}_q(\omega) := Y_q(\omega)\mathbb{1}_\Delta(\omega)$. Let now $r \in [0, T] \setminus \mathcal{D}$. Note that $r < T$. Hence, there is a sequence $(r_n) \subset \mathcal{D}$ with $r_n \downarrow r$ and we define

$$\tilde{Y}_r(\omega) = \lim_{n\to\infty} \tilde{Y}_{r_n}(\omega). \tag{A.3}$$

Since $\tilde{Y}$ is nondecreasing on $\mathcal{D}$, the above limit exists as a monotone sequence bounded below by $\tilde{Y}_0(\omega)$. Further, the limit is independent by the choice of the sequence by monotonicity. We have that $\tilde{Y}_r(\omega)$ is right-continuous for all $\omega \in \Omega$ by construction (outside Δ

it is constant zero). Further, left limits exists for all $\omega \in \Omega$ again by monotonicity with upper bound $Y_T(\omega)$, hence $\tilde{Y}$ is càdlàg . Note that $\tilde{Y}$ is nondecreasing for all $\omega \in \Omega$. Further, since $Y_t \in \mathcal{A}$ for all $t \geq 0$ and $\mathbb{1}_\Delta \in \mathcal{A}$, we have $\tilde{Y}_t \in \mathcal{A}$ for all $t \geq 0$. By construction, $\tilde{Y}$ is a version of (A.2) for each $t \in \mathcal{D}$ and we are left to show that for $r \in [0,T] \setminus \mathcal{D}$, $\tilde{Y}_r$ is also a version of (A.2).

Recall that by assumption A, is nondecreasing and càdlàg with probability one and $\mathbb{E}[A_T] < \infty$. For the sequence (r_n) above, we therefore get by the dominated convergence theorem for the conditional expectation that

$$\mathbb{E}[A_r \mid \mathcal{A}] = \mathbb{E}\left[\lim_{n\to\infty} A_{r_n} \mid \mathcal{A}\right] = \lim_{n\to\infty} \mathbb{E}[A_{r_n} \mid \mathcal{A}] \quad [P].$$

Combining this with (A.3), yields $\tilde{Y}_r = \mathbb{E}[A_r \mid \mathcal{A}]$ a.s. which finishes the proof. □

We can now give the proof of Theorem A.8.

Proof of Theorem A.8. Let $s \geq 0$ be fixed. Note that by assumption, we have M is a true martingale on our stochastic basis. Hence we can compute for any $t \geq s$,

$$\mathbb{E}[X_t \mid \mathcal{F}_s] = X_0 + \mathbb{E}[M_t \mid \mathcal{F}_s] + \mathbb{E}[A_t \mid \mathcal{F}_s] = X_0 + M_s + \mathbb{E}[A_t \mid \mathcal{F}_s].$$

Hence, we only need to show that the conditional expectation of the finite variation process A has a càdlàg modification. We have $A = A^i - A^d$ for increasing càdlàg processes A^i, A^d, both starting in zero. Further, $\mathbb{E}\left[A^i_t\right] + \mathbb{E}\left[A^d_t\right] \leq \mathbb{E}\left[\mathrm{Var}(A)_t\right] < \infty$ for all $t \geq 0$. Denote by Y^i the càdlàg and increasing modification of $t \mapsto \mathbb{E}\left[A^i_t \mid \mathcal{F}_s\right]$ and in the same way by Y^d the one for A^d. These modifications exist by Lemma A.11. By linearity, $Y^i - Y^d$ is now a càdlàg modification of $\mathbb{E}[A_\cdot \mid \mathcal{F}_s]$ that is of finite variation and we are done. □

It is clear, that in Theorem A.8, the existence of such a modification exists on all of $\mathbb{R}_+$, not just $[s,\infty)$, if one drops the finite variation property of the modification. Hence the following corollary.

Corollary A.12. *Consider the situation in Theorem A.8. Then there exists a càdlàg modification of*

$$t \mapsto \mathbb{E}[X_t \mid \mathcal{F}_s], \quad \forall t \geq 0.$$

Proof. For $s = 0$, this is just the theorem itself. Hence consider $s > 0$. For $t \in [0,s)$, since X is adapted and càdlàg , the version is given by X itself. For $t \in [s,\infty)$, take the modification from Theorem A.8. The combined process is again càdlàg , remains a version of the conditional expectation for each $t \geq 0$ and we are done. □

Appendix B

Additional notes

B.1 Proof of a naive uniform integrability result

We now give the previously mentioned naive uniform integrability result. Since we provided a strictly stronger result, this is just for illustration. We also want to note that the following proof can be significantly shortened in Equation (B.1) by applying Jensen's inequality for sums. Since however this is just for illustration, we opt to give the most elementary proof we have.

Proposition B.1. *Let $k \geq 2$ be an even number, β a basis of Pol_k and $A \in \mathcal{V}^{+}_{\mathfrak{A}}$. Let X be a càdlàg adapted process such that $\mathrm{Pol}_k \subset \mathcal{D}_X$ and $\mathbb{E}\left[\|X_0\|^k_{\mathbb{R}^d}\right] < \infty$. Assume there exists a measurable function*

$$G : \mathbb{R}_{>0} \to M_N(\mathbb{R}) \text{ with } \|G\|_{dA,t} < \infty \quad \forall t \geq 0,$$

such that the map $\mathcal{G} : \mathrm{Pol}_k \to \mathrm{Pol}_k(\mathbb{R}_{>0} \times \mathbb{R}^d))$, defined by

$$(\mathcal{G}f)(u)(x) = \mathcal{G}f(u,x) = (G(u)\mathbf{f})^\top \mathbf{H}_\beta(x), \; \forall x \in \mathbb{R}^d,$$

is a version of $\mathcal{G}_{X,A}$ on Pol_k. Then for any $f \in \mathrm{Pol}_{k-1}$ and $T \geq 0$, the set $\mathcal{H}^f_T$ is uniformly integrable.

Proof. The idea is to apply Theorem 3.4. Let $\varepsilon > 0$ such that $(1+\varepsilon)(k-1) = k$. Define the function $\varphi(x) = x^{1+\varepsilon}$. Note that φ is convex on $\mathbb{R}_+$ and

$$\lim_{t \to +\infty} \frac{\varphi(t)}{t} = +\infty.$$

By Lemma 6.37, we have (6.31). By Lemma 6.2, this is equivalent to

$$\sup_{0 \leq t \leq T} \mathbb{E}\left[\|X_t\|^k_1\right].$$

Let $f \in \mathrm{Pol}_{k-1}$ be arbitrary. Then there exists a constant $C > 0$ such that

$$|f(x)| \leq C(1 + \|x\|_1^{k-1}) \Leftrightarrow |\, f(x)| - C \leq C \, \|x\|_1^{k-1} \,. \tag{B.1}$$

This in return implies

$$|\, |f(x)| - C| \mathbb{1}_{\{|f(x)| \geq C\}} \leq C \, \|x\|_1^{k-1} \,.$$

Using this we get

$$\varphi(\, |\, |f(x)| - C| \, \mathbb{1}_{\{|f(x)| \geq C\}}\,) \leq C^{1+\varepsilon} \, \|x\|_1^{k} \,.$$

Therefore we can conclude

$$\sup_{0 \leq t \leq T} \mathbb{E}\left[\varphi(\, |\, |f(X_t)| - C| \, \mathbb{1}_{\{|f(X_t)| \geq C\}}\,)\right] \leq C^{1+\varepsilon} \sup_{0 \leq t \leq T} \mathbb{E}\left[\|X_t\|_1^k\right] < \infty.$$

By Theorem 3.4, we have that the set

$$\left\{|\, |f(X_t)| - C| \, \mathbb{1}_{\{|f(X_t)| \geq C\}} : 0 \leq t \leq T\right\}$$

is uniformly integrable. From the definition of uniform integrability, this means that for any $\varepsilon > 0$, there is a $L > 0$ such that for all $t \in [0, T]$ simultaneously we have

$$\begin{aligned}
\varepsilon &> \mathbb{E}\left[|\, |f(X_t)| - C| \, \mathbb{1}_{\{|f(X_t)| \geq C\}} \mathbb{1}_{\{|\, |f(X_t)| - C| \, \mathbb{1}_{\{|f(X_t)| \geq C\}} \geq L\}}\right] \\
&= \mathbb{E}\left[|\, |f(X_t)| - C| \, \mathbb{1}_{\{|f(X_t)| - C \geq 0\}} \mathbb{1}_{\{|\, |f(X_t)| - C| \, \mathbb{1}_{\{|f(X_t)| - C \geq 0\}} \geq L\}}\right] \\
&\geq \mathbb{E}\left[|\, |f(X_t)| - C| \, \mathbb{1}_{\{|\, |f(X_t)| - C| \, \mathbb{1}_{\{|f(X_t)| - C \geq 0\}} \geq L\}}\right]
\end{aligned} \tag{B.2}$$

where we use in the last step

$$\mathbb{1}_{\{|f(X_t)| - C \geq 0\}} = 0 \Rightarrow \mathbb{1}_{\{|\, |f(X_t)| - C| \, \mathbb{1}_{\{|f(X_t)| - C \geq 0\}} \geq L\}} = 0$$

Using

$$|\, |f(X_t)| - C\, | \geq L + 2C \Rightarrow |f(X_t)| - C \geq 0 \text{ and } |\, |f(X_t)| - C\, | \geq L,$$

we get

$$\begin{aligned}
\varepsilon &> \mathbb{E}\left[|\, |f(X_t)| - C| \, \mathbb{1}_{\{|\, |f(X_t)| - C| \, \mathbb{1}_{\{|f(X_t)| - C \geq 0\}} \geq L\}}\right] \\
&\geq \mathbb{E}\left[|\, |f(X_t)| - C| \, \mathbb{1}_{\{|\, |f(X_t)| - C| \geq L + 2C\}}\right].
\end{aligned} \tag{B.3}$$

Since this holds for all t simultaneously, we have shown that the set

$$\{|\, |f(X_t)| - C| : 0 \leq t \leq T\}$$

is uniformly integrable. This once again, gives us that

$$\{f(X_t) : 0 \leq t \leq T\} = \mathcal{H}_T^f$$

is uniformly integrable. Since $f \in \mathrm{Pol}_{k-1}$ was arbitrary and this holds for all $T \geq 0$, this finishes the proof. $\square$

Part II

A Deep Learning Approach to Local Stochastic Volatility Calibration

Chapter 7

Preliminaries and deep learning

This part of the thesis is based on joint work with Christa Cuchiero[1] and Josef Teichmann[2], where we only give a partial presentation of the work together with a first set of numerical examples that will serve as a proof of concept.

We start by giving a short exposition to the problem we address with this work and continue with a brief summary of the most basic concepts from the area of machine learning pertinent to our discussions. We continue in Section 7.3, with introducing a variance reduction technique that will make our approach actually feasible. Following that, we present in Chapter 8 the central problem formulation we are addressing, namely the calibration of local stochastic volatility models to market data. In particular, we will specify numerical experiments regarding the variance reduction technique we propose and a toy-model example of calibration. In the interest of readability, we move all plots regarding these experiments to Chapter 9.

For the rest of this thesis, we will assume a zero risk interest rate if $r = 0$ to keep notation simple. An extension to the general case however is straight forward.

7.1 Introduction

A central task that has to be performed each day in banks and financial institutions is to calibrate stochastic models to market and historical data. So far the model choice was not only driven by the capacity of capturing empirically observed market features well, but rather by computational tractability. This is now undergoing a big change since ***neural network approaches*** offer a completely new perspective on model calibration,

[1]Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Vienna
[2]ETH Zurich, Rämistrasse 101, 8092 Zurich

see e.g. Hernandez [Her17], Bayer and Stemper [BS18], Horvath et al. [Hor+19] and [Cuc+18a].

This part of the thesis focuses on the calibration of ***local stochastic volatility (LSV) models***, which is still an intricate task, in particular in multivariate situations. LSV models, going back to [Lip02; Ren+07], combine classical stochastic volatility with local volatility to achieve both, a good fit to time series data and in principle a perfect calibration to the implied volatility smiles and skews. In these models, the discounted price process $(S_t)_{t\geq 0}$ of an asset satisfies

$$dS_t = S_t L(t, S_t)\alpha_t dW_t, \tag{7.1}$$

where $(\alpha_t)_{t\geq 0}$ is some stochastic process taking values in $\mathbb{R}$, $L(t, s)$ the so-called ***leverage function*** depending on time and the current value of the asset and W a one-dimensional Brownian motion.

For notational simplicity we consider here the one-dimensional case, but the setup easily translates to a multivariate situation with several assets and general multivariate variance process.

The function L is the crucial part in this model. It allows in principle to perfectly calibrate the implied volatility surface seen on the market. In order to achieve this goal, it has to satisfy

$$L^2(t, s) = \frac{\sigma^2_{\mathrm{Dup}}(t, x)}{\mathbb{E}[\alpha_t^2 | S_t = s]}, \tag{7.2}$$

where σ_{Dup} denotes Dupire's local volatility function (see [Dup96]). Note that (7.2) is an implicit equation for L as it is needed for the computation of $\mathbb{E}[\alpha_t^2 | S_t = s]$. This in turn means that the SDE for the price process $(S_t)_{t\geq 0}$ is actually a McKean-Vlasov SDE, since the law of S_t enters in the equation. Existence and uniqueness results for this equation are not at all obvious since the coefficients do not satisfy any kind of standard conditions like for instance Lipschitz continuity in the Wasserstein space. Existence of a short-time solution of the associated non-linear Fokker-Planck equation for the density of $(S_t)_{t\geq 0}$ was shown in Abergel and Tachet [AT10] under certain regularity assumptions on the initial distribution. As stated in Guyon and Henry-Labordere [GHL11] (see also Jourdain and Zhou [JZ16], where existence of a simplified version of an LSV model is proved) a very challenging and still open problem is to derive the set of stochastic volatility parameters for which LSV models exist uniquely for a given market smile.

Despite these existence issues, LSV models have attracted – due to their appealing feature of a potentially perfect smile calibration and their econometric properties – a lot of attention from the calibration and implementation point of view. We refer to Guyon and Henry-Labordere [GHL11], Guyon and Henry-Labordère [GHL13], and Cozma et al. [Coz+17]) for Monte Carlo methods, to Ren et al. [Ren+07] and Tian et al. [Tia+15] for PDE methods based on non-linear Fokker-Planck equations and to Saporito et al. [Sap+17] for inverse problem techniques for PDEs.

7.2 Preliminaries on deep learning

We will now briefly introduce two core concepts in deep learning, namely ***artificial neural networks*** and ***stochastic gradient descent***, where the latter is a widely used optimization method for optimization problems involving the first. In standard machine learning terminology, the optimization problem is usually referred to as "training" and in the sequel we will use both terminologies interchangeably. In the context of standard machine learning terminology, let us also mention that in all of our following discussions, we only consider neural networks of feed forward type (we define this below) and we do not make any structural assumptions regarding the topology of the network, hence we consider densely connected layers instead of convolutional type ones.

7.2.1 Artificial neural networks

We start with the definition of ***feed forward neural networks***. These are functions of particular structure, indexed by parameters such that this class of functions can serve as an approximation class for an unknown function under certain assumptions, compare Theorem 7.2 below. In particular, it will be clear from the definition below that derivatives of these functions can be efficiently expressed iteratively with a small set of functions (see e.g. Hecht-Nielsen [HN92]), which is a desirable feature from an optimization point of view.

Definition 7.1. *Let $L, N_0, N_1, \dots N_L \in \mathbb{N}$, $\sigma : \mathbb{R} \to \mathbb{R}$ and for any $\ell \in \{1, \dots, L\}$, let $w_\ell : \mathbb{R}^{N_{\ell-1}} \to \mathbb{R}^{N_\ell}$, $x \mapsto A_\ell x + b_\ell$ be an affine function with $A_\ell \in \mathbb{R}^{N_\ell \times N_{\ell-1}}$ and $b_\ell \in \mathbb{R}^{N_\ell}$ and additionally $b_L = 0$. A function $\mathbb{R}^{N_0} \to \mathbb{R}^{N_L}$ defined as*

$$F(x) = w_L \circ F_{L-1} \circ \cdots \circ F_1, \quad \text{with } F_\ell = \sigma \circ w_\ell \quad \text{for } \ell \in \{1, \dots, L-1\}$$

is called a **feed forward neural network**. *Here the* **activation function** *σ is applied componentwise. L denotes the number of layers and $N_1, \dots, N_{L-1}$ denote the dimensions of the hidden layers and N_0 and N_L the dimension of the input and output layers.*

The following version of the so-called ***universal approximation theorem*** is due to K. Hornik [Hor91]. For its formulation we denote the set of all feed forward neural networks with activation function σ, input dimension N_0 and output dimension N_L by $\mathcal{NN}^\sigma_{\infty,N_0,N_L}$.

Theorem 7.2 (Hornik (1991))**.** *Suppose σ is bounded and non-constant. Then the following statements hold:*

(i) For any finite measure μ on $(\mathbb{R}^{N_0}, \mathcal{B}(\mathbb{R}^{N_0}))$ and $1 \le p < \infty$, the set $\mathcal{NN}^\sigma_{\infty,N_0,1}$ is dense in $L^p(\mathbb{R}^{N_0}, \mathcal{B}(\mathbb{R}^{N_0}), \mu)$.

(ii) If in addition $\sigma \in C(\mathbb{R}, \mathbb{R})$, then $\mathcal{NN}^{\sigma}_{\infty,N_0,1}$ is dense in $C(\mathbb{R}^{N_0}, \mathbb{R})$ for the topology of uniform convergence on compact sets.

Since each component of an $\mathbb{R}^{N_L}$ valued neural network is an $\mathbb{R}$-valued neural network, this result easily generalizes to $\mathcal{NN}^{\sigma}_{\infty,N_0,N_L}$ with $N_L > 1$.

We denote by $\mathcal{NN}_{N_0,N_L}$ the set of all neural networks in $\mathcal{NN}^{\sigma}_{\infty,N_0,N_L}$ with a ***fixed architecture,*** i.e. a fixed number of layers L, fixed input and output dimensions N_ℓ for each hidden layer $\ell \in \{1, \dots, L-1\}$ and a fixed activation function σ. This set can be described by

$$\mathcal{NN}_{N_0,N_L} = \{F(\cdot|\theta) \,|\, F \text{ feed forward neural network and } \theta \in \Theta\},$$

with parameter space $\Theta \in \mathbb{R}^q$ for some $q \in \mathbb{N}$ and $\theta \in \Theta$ corresponding to the entries of the matrices A_ℓ and the vector b_ℓ for $\ell \in \{1, \dots, L\}$.

7.2.2 Stochastic gradient descent

In light of Theorem 7.2, it is clear that neural networks can serve as function approximators and the goal is to find the "correct" parameters. Usually, the situation is such that the unknown function is expressed as an expectation. Probably the most prolific training method for such a setup is stochastic gradient descent, and we will shortly review the most basis facts about this optimization/training method.

The structural properties of neural networks allow to solve minimization problems of the type

$$\min_{\theta \in \Theta} f(\theta) \quad \text{with} \quad f(\theta) = \mathbb{E}[Q(\theta)] \tag{7.3}$$

for some stochastic objective function[3] $Q : \Omega \times \Theta \to \mathbb{R}$, $(\omega, \theta) \mapsto Q(\theta)(\omega)$ that depends on parameters θ in some space Θ very efficiently via ***stochastic gradient descent*** and ***back propagation.***

The classical method how to solve generic optimization problems for some differentiable objective function f (not necessarily of the expected value form as in (7.3)) is to apply a ***gradient descent*** algorithm: starting with an initial guess $\theta^{(0)}$, one iteratively defines

$$\theta^{(k+1)} = \theta^{(k)} - \eta_k \nabla f^{(k)}(\theta^{(k)}) \tag{7.4}$$

for some learning rate η_k and $f^{(k)} = f$. Under suitable assumptions, $\theta^{(k)}$ converges for $k \to \infty$ to a local minimum of the function f.

[3]We shall often omit the dependence on ω.

One of the key insights of deep learning is that ***stochastic gradient descent*** methods, going back to stochastic approximation algorithms proposed by Robbins and Monro [RM51], are much more efficient. To apply this, it is crucial that the objective function f is linear in the sampling probabilities. In other words, f needs to be of the expected value form as in (7.3). In the simplest form of stochastic gradient descent, under the assumption that

$$\nabla f(\theta) = \mathbb{E}[\nabla Q(\theta)],$$

the true gradient of f is approximated by a gradient at a single sample $Q(\theta)(\omega)$ which reduces the computational cost considerably. In the updating step for the parameters θ as in (7.5), ∇f is then replaced by $\nabla Q(\theta)(\omega)$, hence

$$\theta^{(k+1)} = \theta^{(k)} - \eta_k \nabla Q^{(k)}(\theta^{(k)})(\omega), \tag{7.5}$$

with $Q^{(k)} = Q$. The algorithm passes through all samples ω of the so-called training data set and performs the update for each element, several times until an approximate minimum is reached.

A compromise between computing the true gradient of f and the gradient at a single $Q(\theta)(\omega)$ is to compute the gradient of a subsample of size N_{batch}, called (mini)-batch, so that $Q^{(k)}$ used in the update (7.5) is now given by

$$Q^{(k)}(\theta) = \frac{1}{N_{\text{batch}}} \sum_{n=1}^{N_{\text{batch}}} Q(\theta)(\omega_{n+kN_{\text{batch}}}), \quad k \in \{0, 1, ..., \lfloor N/N_{\text{batch}}, \rfloor - 1\}. \tag{7.6}$$

where N is the size of the whole training data set. Any other unbiased estimators of $\nabla f(\theta)$ can of course also be applied in (7.5).

In typical applications of machine learning, the data set available is limited, i.e. $N < \infty$. In our situation, we apply machine learning in a simulated environment, allowing us to generate data at will. This corresponds to $N = \infty$, meaning that we can generate completely new data in each step k.

7.3 Variance reduction for pricing and calibration

This section is dedicated to introduce a generic variance reduction technique for Monte Carlo pricing and calibration by using hedges as control variates. This method will be crucial in our LSV calibration presented in Chapter 8, since without such a variance reduction method applied, our numerical experiments showed that a sufficient calibration result is not possible in a reasonable amount of time.

Consider a financial market in discounted terms with r traded instruments $(Z_t)_{t\geq 0}$ following an $\mathbb{R}^r$-valued stochastic process on some filtered probability space $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathcal{F}, \mathbb{Q})$.

Here, $\mathbb{Q}$ is a risk neutral measure and $(\mathcal{F}_t)_{t\geq 0}$ is supposed to be right continuous. In particular, $(Z_t)_{t\geq 0}$ is a σ-martingale with càdlàg paths.

Let C be an $\mathcal{F}_T$-measurable random variable describing the payoff of some European option at maturity $T > 0$. Then the usual Monte Carlo estimator for the price of this option is given by

$$\pi = \frac{1}{N}\sum_{n=1}^{N} C(\omega_n) \tag{7.7}$$

for $n \in \mathbb{N}$. This estimator can easily be modified by adding a stochastic integral with respect to Z. Indeed, consider a locally bounded predictable strategy $(h_s)_{s\in[0,t]}$ and some constant c. Define the following estimator

$$\widehat{\pi} = \frac{1}{N}\sum_{n=1}^{N}(C(\omega_n) - c(h \bullet Z)_T(\omega_n)). \tag{7.8}$$

Then, for any $(h_t)_{t\in[0,T]}$ and c, this estimator is still an unbiased estimator for the price of the option with payoff C since the expected value of the stochastic integral vanishes, at least when Z is a true martingale. Denoting by

$$I = \frac{1}{N}\sum_{n=1}^{N}(h \bullet Z)_T(\omega_n),$$

the variance of $\widehat{\pi}$ is given by

$$\mathrm{Var}(\widehat{\pi}) = \mathrm{Var}(\pi) + c^2\mathrm{Var}(I) - 2c\mathrm{Cov}(\pi, I).$$

This becomes minimal by choosing

$$c = \frac{\mathrm{Cov}(\pi, I)}{\mathrm{Var}\,(I)}.$$

With this choice, we have

$$\mathrm{Var}(\widehat{\pi}) = (1 - \mathrm{Corr}^2(\pi, I))\mathrm{Var}(\pi).$$

In particular, in the case of a perfect (pathwise) hedge given by $(h_t)_{t\in[0,T]}$ this ratio is equal to 1 and the estimator $\widehat{\pi}$ has 0 variance since in this case

$$\mathrm{Var}(\pi) = \mathrm{Var}(I) = \mathrm{Cov}(\pi, I).$$

Therefore it is crucial to find a good approximate hedge such that $\mathrm{Corr}^2(\pi, I)$ becomes large. This is subject of Section 7.3.1 and 7.3.2 below.

7.3.1 Black Scholes hedge

In many cases of local stochastic volatility models as of form (7.1) and options depending only on the terminal value of the price process, a Delta hedge of the Black-Scholes model works well. Indeed, let $C = g(S_T)$ and let $\pi^g_{\mathrm{BS}}(t, x, \sigma)$ be the price at time t of this claim in the Black-Scholes model. Here, x stands for the price variable and σ for the Black-Scholes volatility. Then choosing as hedging instrument only the price S itself and as approximate hedging strategy

$$h_t = \partial_x \pi^g_{\mathrm{BS}}(t, S_t, L(t, S_t)\alpha_t) \tag{7.9}$$

usually already yields a high variance reduction. In fact it is even sufficient to consider α_t alone to achieve satisfying results, i.e. one has

$$h_t = \partial_x \pi^g_{\mathrm{BS}}(t, S_t, \alpha_t). \tag{7.10}$$

7.3.2 Neural networks hedge

Alternatively, in particular when the space of hedging instruments becomes higher dimensional, one can learn the hedging strategy by parameterizing it via neural networks. Indeed, let the payoff be again a function on the terminal values of the hedging instruments, i.e., $C = g(Z_T)$. Then in Markovian models it makes sense to specify the hedging strategy via a function $h : \mathbb{R}_+ \times \mathbb{R}^r \to \mathbb{R}^r$

$$h_t = h(t, z),$$

which in turn will correspond to a neural network $(t, z) \mapsto h(t, z|\delta) \in \mathcal{NN}_{r+1,r}$ with weights denoted by δ in some parameter space Δ. Following the approach in Buehler et al. [Bue+18], an optimal hedge for the claim C can be computed through

$$\inf_{\delta \in \Delta} \rho\left(-C + (h(\cdot, Z.|\delta) \bullet Z.)_T\right)$$

for some risk measure ρ. If the risk measure is chosen of the form

$$\rho(C) = \inf_{w \in \mathbb{R}} \{w + \mathbb{E}\left[\ell(C - w)\right]\}$$

for some continuous, non-decreasing and convex loss function $\ell : \mathbb{R} \to \mathbb{R}$, we can apply stochastic gradient descent, because we fall in the realm of problem (7.3). Indeed, the objective function $f(\delta, w)$ that should be minimized with respect to δ and w is given by

$$f(\delta, w) = \mathbb{E}\left[w + \ell(-C + (h(\cdot, Z.|\delta) \bullet Z.)_T - w)\right].$$

The optimal hedge $h(\cdot|\delta^*)$ for an optimizer δ^* can then be plugged into (7.8).

Our situation is different from [Bue+18], as we do not want to hedge against a risk expressed by a risk measure, but directly aim to reduce variance. Hence the corresponding optimization problem for the hedge simplifies to

$$\inf_{\delta \in \Delta} \mathbb{E}\left[(C - \pi^{\mathrm{mkt}} - (h(\cdot, Z. \mid \delta) \bullet Z)_T)^2\right],$$

where $\pi^{\mathrm{mkt}} = \mathbb{E}[C]$ is the market price of claim C.

In the context of calibration, we will face the situation where the terminal value $C = g(Z_T)$ is a function of $\omega \in \Omega$ and $\theta \in \Theta$, where θ are those parameters that specify the model. Hence, in general we deal with two different sets of parameter, where δ only serves auxiliary purposes for the calibration problem.

Chapter 8

Calibration of LSV models

We will now make a precise formalization of the considered calibration problem and the method we apply.

Consider a LSV model as of (7.1) defined on $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathcal{F}, \mathbb{Q})$, some filtered probability space, where $\mathbb{Q}$ is a risk neutral measure. We assume the stochastic process α to be fixed. This can for instance be achieved by first calibrating the pure stochastic volatility model with $L \equiv 1$ (e.g. SABR) and by fixing the corresponding parameters.

Our main goal is to determine the leverage function L in perfect accordance with market data. Due to the universal approximation properties outlined in Section 7.2 (Theorem 7.2), we choose to parametrize L via neural networks. More precisely, let $0 = T_0 < T_1 \cdots < T_n = T$ denote the maturities of the available European call options to which we aim to calibrate the LSV model. We then specify the leverage function $L(t, s)$ via a family of neural networks, i.e.,

$$L(t, s) = 1 + F_i(s) \quad t \in [T_{i-1}, T_i), \quad i \in \{1, \ldots, n\}, \tag{8.1}$$

where $F^i \in \mathcal{NN}_{1,1}$. We denote the parameters of F_i by θ_i and the corresponding parameter space by Θ_i. For each maturity T_i, we assume to have J_i options with strikes K_{ij}, $j \in \{1, \ldots, J_i\}$. The calibration functional for the i-th maturity is then of the form

$$\operatorname{argmin}_{\theta_i \in \Theta_i} \sum_{j=1}^{J_i} w_{ij} u(\pi_{ij}^{\mathrm{mod}}(\theta_i) - \pi_{ij}^{\mathrm{mkt}}), \quad i \in \{1, \ldots, n\}, \tag{8.2}$$

where $\pi_{ij}^{\mathrm{mod}}(\theta_i)$ (π_{ij}^{mkt} respectively) denotes the model (market resp.) price of an option with maturity T_i and Strike K_{ij}, $u : \mathbb{R} \to \mathbb{R}_+$ is some (positive, nonlinear, convex) function (e.g. square or absolute value) measuring the distance between market and model prices and w_{ij} are some weights, e.g. of vega type, which allows to match implied volatility data rather then pure prices, our actual goal, very well.

We solve the minimization problems (8.2) iteratively: we start with maturity T_1 and fix θ_1. This then enters in the computation of $\pi_{2j}^{\mathrm{mod}}(\theta_2)$ and thus in (8.2) for maturity T_2, etc. To simplify the notation in the sequel, we shall therefore leave the index i away so that for a generic maturity $T > 0$, (8.2) becomes

$$\operatorname{argmin}_{\theta\in\Theta} \sum_{j=1}^{J} w_j u(\pi_j^{\mathrm{mod}}(\theta) - \pi_j^{\mathrm{mkt}}).$$

Since the model prices are given by

$$\pi_j^{\mathrm{mod}}(\theta) = \mathbb{E}[(S_T(\theta) - K_j)^+], \tag{8.3}$$

we have $\pi_j^{\mathrm{mod}}(\theta) - \pi_j^{\mathrm{mkt}} = \mathbb{E}\left[Q_j(\theta)\right]$ where

$$Q_j(\theta)(\omega) := (S_T(\theta)(\omega) - K_j)^+ - \pi_j^{\mathrm{mkt}}. \tag{8.4}$$

Note that S_T depends via (8.1) on θ. The calibration task then amounts to finding a minimum of

$$f(\theta) := \sum_{j=1}^{J} w_j u(\mathbb{E}\left[Q_j(\theta)\right]). \tag{8.5}$$

If u is a general non-linear function, this is clearly not of the expected value form of problem (7.3). In the following section we illustrate two possibilities how to deal with this non-linearity and the fact that stochastic gradient descent as outlined in Section 7.2.2 is not directly applicable.

8.1 Minimizing the calibration functional

The goal of this section is to specify two methods for minimizing (8.5). It is possible to consider linearized versions of (8.5) such that classical stochastic gradient descent with potentially small batch-size is possible. In this thesis however, we only consider two approaches that both amount to use classical gradient descent.

8.1.1 Standard gradient descent

The most obvious choice (which however does not work in practice) is to use a standard Monte Carlo estimator for $\mathbb{E}[Q_j(\theta)]$ so that (8.5) is estimated by

$$\widehat{f}(\theta) = \sum_{j=1}^{J} w_j u\left(\frac{1}{m}\sum_{l=1}^{m} Q_j(\theta)(\omega_l)\right), \tag{8.6}$$

for i.i.d samples $\{\omega_1, \dots, \omega_m\} \in \Omega$. Since the Monte Carlo error decreases as $\frac{1}{\sqrt{m}}$, the number of simulation m has to be chosen large ($\approx 10^8$) in order to approximate well the true model prices in (8.3). Note that implied volatility to which we actually aim to calibrate is even more sensitive. As it is not obvious how to apply stochastic gradient descent due to the non-linearity of u, it seems necessary at first sight to compute the gradient of the whole function $\widehat{f}(\theta)$. As $m \approx 10^8$, this is however computationally very expensive and does not allow to find a minimum in the usually high dimensional parameter space Θ in a reasonable amount of time.

8.1.2 Standard gradient descent with control variates

One possible remedy is to apply hedging control variates as introduced in Section 7.3 as variance reduction technique. This allows to reduce the number of samples m in the Monte Carlo estimator drastically so that usual (non-stochastic) gradient descent is enough to achieve accurate calibration results.

Assume that we have r hedging instruments (including the price process S) denoted by $(Z_t)_{t\geq 0}$ which are σ-martingale under $\mathbb{Q}$ and take values in $\mathbb{R}^r$. Consider strategies $h_j : [0,T] \times \mathbb{R}^r \to \mathbb{R}^r$ and some constant c. Define

$$X_j(\theta)(\omega) := (S_t(\theta)(\omega) - K_j)^+ - c(h_j(\cdot, Z_\cdot(\theta)(\omega)) \bullet Z_\cdot(\theta)(\omega))_t - \pi_j^{\text{mkt}} \tag{8.7}$$

where $(h_j \bullet Z)_t$ denotes (a discretized version of) the stochastic integral with respect to Z. The calibration functionals (8.5) and (8.6), can then simply be defined by replacing $Q_j(\theta)(\omega)$ by $X_j(\theta)(\omega)$.

Analogously as in Section 7.3.2, we can parametrize the hedging strategies via neural networks and find the optimal weight δ by computing

$$\text{argmin}_{\delta\in\Delta}\frac{1}{N}\sum_{n=1}^{N}\ell(-X_j(\theta,\delta)(\omega_n)).$$

for i.i.d samples $\{\omega_1, \dots, \omega_N\} \in \Omega$ and some loss function ℓ when θ is fixed. Here,

$$X_j(\theta,\delta)(\omega) = (S_T(\theta)(\omega) - K_j)^+ - (h_j(\cdot, Z_\cdot(\theta)(\omega)|\delta) \bullet Z_\cdot(\theta)(\omega))_T - \pi_j^{\text{mkt}}.$$

This means to iterate the two optimization procedures, one for θ and the other one for δ. Clearly the Black-Scholes hedge approach of Section 7.3.1 works as well, in this case without additional optimization with respect to the hedging strategies.

8.2 Numerical Implementation

In this section, we present numerical examples in the context of a toy model we now wont to specify. This toy-model will play the role of the "real world", i.e. we consider data generated by this model and the task is then to calibrate to these prices. Let us first summarize the computational framework we work in. We implement out approach via tensorflow, taking advantage of gpu-accelerated computing. All computations are performed on desktop-pc with AMD® Ryzen 2700x cpu, 16 GB of RAM and one GeForce GTX 1070 gpu. The underlying software libraries are tensorflow 1.12 with CUDA 10. We use the python 3.6 API for tensorflow and for the implied volatility computations, we rely on the python `py_vollib` library.[1]

8.2.1 Toy-model specification

We consider a SABR type model with one dimensional price-process and one dimensional variance process α for which the dynamics are given by

$$\begin{aligned} dS_t &= S_t L^d(t, S_t)\alpha_t dW_t, \\ d\alpha_t &= \nu \alpha_t dB_t, \\ dW_t dB_t &= \rho dt, \end{aligned}$$

for two Brownian motions W, B with correlation $\rho \in [-1, 1]$. Since we will compute model prices with a Monte Carlo method via an Euler-discretization of the model, it will be preferable to work in log-price coordinates for S. In particular, we can then parametrize L^d with $X := \log S$ rather then S. By denoting this parametrization again with L^d, where d stands for "data", we therefore have $L^d(t, X)$ instead of $L^d(t, S)$ and the model dynamics read

$$\begin{aligned} dX_t &= \alpha_t L^d(t, X_t) dW_t - \frac{1}{2}\alpha_t^2 L^d(t, X_t)^2 dt, \\ d\alpha_t &= \nu \alpha_t dB_t, \\ dW_t dB_t &= \rho dt. \end{aligned}$$

Note that α is a geometric Brownian motion, in particular, the closed form solution for α is available and given by

$$\alpha_t = \alpha_0 \exp\left(-\frac{\nu^2}{2}t + \nu B_t\right).$$

We specify the leverage function L^d in our toy-model by

$$L^d(t, x) = 1 + 0.3 \sin(13x + 20t) \cos(1.1t). \tag{8.8}$$

For the other parameters, we choose the ones given in Figure 8.1(a). This toy-model of ours will be used to generate data, i.e. option prices. Hence we separate these parameters from our model parameters θ and δ as we keep them fixed and do not optimize (train)

[1] See http://vollib.org/

ν	ρ	X_0	α_0
0.2	-0.7	1	0.2

(a)

T_1	T_2	T_3	T_4
0.15	0.25	0.5	1.0

(b)

K_1	K_2	K_3	K_4
0.1	0.2	0.4	0.5

(c)

Figure 8.1: (a) Parameters for the toy-model. (b) Maturities to generate data for calibration. (c) Parameters that define the strikes of the call options to which we calibrate.

over them. We generate data using this model consisting of European call prices for $n = 4$ different maturities $T_i, i \in \{1, 2, 3, 4\}$ given in Figure 8.1(b). For each maturity T_i, we consider $J_i = 20$ strikes $K_{i,j}, j \in \{1, \dots, J_i\}$, given by

$$K_{i,j} = \exp(-K_i) + (j-1)\Delta_{K_i}, \text{ for } \Delta_{K_i} = \frac{\exp(K_i) - \exp(-K_i)}{19}.$$

The values for K_i are given in Figure 8.1(c).

We compute prices in our toy-model using an explicit Euler discretization. For the time-grid, we discretize the interval $[0, T_4] = [0, 1]$ with step size $1/100$, hence we consider the time grid $\mathcal{T} = \{t_0 = 0, \dots, t_{N_{100}} = 1\}$. Since α is known in closed form, we simulate the Brownian motion B on $\mathcal{T}$ and compute α accordingly. For the 4×20 prices that we consider as "data" in our deep-learning approach, we simulate $N_d = 10^9$ pairs of correlated Brownian paths on $\mathcal{T}$ and compute prices via (7.7). Note that we use the same set of paths for all maturities and strikes. We do not employ the variance reduction method proposed above for the computation of these prices. We implement the Monte Carlo scheme via tensorflow, allowing us to use gpu-accelerated computations without having to write native CUDA[2] code.

8.2.2 A first numerical experiment for the variance reduction

We briefly present a numerical experiment to demonstrate the effect of variance reduction via hedges. Consider the toy model we specified above. It is sufficient to demonstrate the variance reduction for one slice, hence we consider one maturity $T_1 = 1$ and three strikes $K_{1,1} = \exp(-0.3), K_{1,2} = 1$ and $K_{1,3} = \exp(0.3)$. We generate "true" prices via 10^9 simulations with the Monte Carlo scheme we specified above. We compare the variance reduction for each strike with the following experiment. Recall that all plots we refer to are to be found in the next chapter.

Experiment 8.1. *Consider the traded assets α and $\exp(X)$, i.e. $Z = (\exp(X), \alpha)^\top$. Hence, as we consider only one maturity, we can specify all hedges via*

$$h(t, z \mid \delta) = \begin{bmatrix} h_X(t, z \mid \delta) \\ h_\alpha(t, z \mid \delta) \end{bmatrix} = \begin{bmatrix} h_{X,1}(t, z \mid \delta) & h_{X,2}(t, z \mid \delta) & h_{X,3}(t, z \mid \delta) \\ h_{\alpha,1}(t, z \mid \delta) & h_{\alpha,2}(t, z \mid \delta) & h_{\alpha,3}(t, z \mid \delta) \end{bmatrix},$$

[2]For more information regarding the CUDA framework we refer to the framework provider's website https://developer.nvidia.com/cuda-zone.

where each column

$$h(t, z \mid \delta)_i = (h_{X,i}(t, z \mid \delta), h_{\alpha,i}(t, z \mid \delta))^\top$$

is the hedge to the contingent claim with strike $K_{1,i}$. *We model* h *with two neural networks, each representing* h_X *or* h_α *respectively. Hence we have* $h_X, h_\alpha \in \mathcal{NN}_{3,3}$. *We use the same network architecture for both neural networks,*

- *each having* 4 *hidden layers with dimensions* $N_1 = N_2 = N_3 = N_4 = 64$,
- *and a single activation function for all nodes* $\sigma(x) = \tanh(x)$.

All weights are initialized by randomly drawing independent truncated-normal distributed values with zero mean and standard deviation 0.05 *for the matrices* A_l *and setting the vectors* $b_l = 0$ *for initial values. For comparison, we consider four different scenarios, given by*

(i) *no variance reduction is used which can be achieved by fixing the final layer weights of the two neural networks to be zero, hence no training,*

(ii) *only the Black-Scholes hedge is applied, i.e.* $h_\alpha \equiv 0$ *and* h_X *is given by Equation* (7.10), *hence no training is involved,*

(iii) *for* h_X, *we use the Black-Scholes hedge as in* (*ii*) *and* h_α *is given by a neural network that needs to be trained,*

(iv) *both functions* h_X *and* h_α *are given by neural networks of the previously specified architecture such that both sets of parameters need to be trained.*

For each contingent claim and each scenario $(i) - (iv)$, *we simulate* 10^5 *trajectories and compute a normalized histogram of hedge errors, in case of* (*iii*) *and* (*iv*) *after training has finished. In those cases, we train the weights by minimizing*

$$\inf_{\delta \in \Delta} \sum_{i=1}^{3} w_i \, \mathbb{E}\left[\left((\exp(X_T) - K_{1,i})^+ - \pi_{1,i} - (h(\cdot, Z. \mid \delta)_i \bullet Z)_T\right)^2\right],$$

where $\pi_{1,i}$ *is the price of the contingent claim with strike* $K_{1,i}$. *The weights* w_i *are of Vega type, given by* $w_i = \min(700, \frac{1}{v_i})$, *where* v_i *is the Black-Scholes Vega for corresponding strike* $K_{1,i}$, *maturity* $T_1 = 1$ *and implied volatility* α_0. *We bound this weights to ensure that that no contingent claim is overweighted (compare Cont and Ben Hamida [CBH04]). We approximate the expectation in each training/ optimization step with the standard Monte Carlo estimator, compare* (7.7), *with mini-batch size* $N_{batch} = 100$. *Further, we estimate in each training step the mean (which needs to be close to zero since we subtract pre-computed prices) and standard deviation by simulating* 10^5 *trajectories. We make overall* 100 *training steps, using stochastic gradient descent with adaptive step size in the form of the Adam optimizer and implemented in tensorflow, see Kingma and Ba [KB14].*

	$K_{1,1}$	$K_{1,2}$	$K_{1,3}$
scenario (i)	1.88×10^{-1}	1.08×10^{-1}	2.74×10^{-2}
scenario (ii)	7.79×10^{-3}	1.05×10^{-2}	8.03×10^{-3}
scenario (iii)	6.43×10^{-3}	8.71×10^{-3}	6.71×10^{-3}
scenario (iv)	2.40×10^{-2}	5.73×10^{-2}	2.56×10^{-2}

Figure 8.2: Estimated standard deviations per strike and scenario.

We show results of this experiment in Figure 9.1 for strike $K_{1,1}$, Figure 9.1 for strike $K_{1,2}$ and Figure 9.3 for strike $K_{1,3}$. For scenarios (iii) and (iv), we plot the estimated mean and standard deviation as a function of the training steps in Figure 9.4, where (a) and (b) correspond to scenario (iii), (c) and (d) to scenario (iv). Note that the value for training step zero corresponds to not using any variance reduction, not the estimated values wrt the hedge generated by randomly drawing the weights. The estimated values for the standard deviation for each scenario and each strike is given in Figure 8.2.

Let us make the following observations:

- *In each case, an improvement regarding minimizing the variance takes place.*
- *Based on the presented histograms and the estimated standard deviations, scenario (iii) performs best in terms of variance reduction.*
- *Scenario (ii), where only the Black-Scholes hedge is used and no training is taking place, is only slightly worse then scenario (iii). This suggests that the main part of the variance reduction is due to the Black-Scholes hedge.*
- *Scenario (iv) performs significantly worse then (ii) and (iii), suggesting that it is so far not possible to beat the Black-Scholes hedge.*

8.2.3 Numerical results for the calibration problem

We will now turn to the calibration problem and present numerical results. Let us start with the specifications regarding the neural networks that define the parametrization in (8.1) and the one for the variance reduction hedge h. Based on the results of the previous section, we choose to employ the pure Black-Scholes hedge, i.e. scenario (ii) in Experiment 8.1, since this allows us to reduce the dimension of the parameter space over which we optimize. Hence, there is no training wrt h. We are left to specify the network architecture of the neural networks for the leverage function of our parametrized model. As we only consider finitely many prices on finitely many slices, we do not expect to recover L^d. The goal is to calibrate implied volatility, hence we will measure the

	training	impl. vola comp.
Experiment 8.2	00 : 27 : 45	00 : 44 : 56
Experiment 8.3	00 : 36 : 29	00 : 44 : 48

Figure 8.3: Computational time table for Experiment 8.2 and Experiment 8.3, with time given in hours, minutes and seconds.

performance of our approach accordingly. Note that our model leverage function L in (8.1), takes arguments t and $x = \log(s)$. Since we consider 4 different maturities, we need to specify $n = 4$ neural networks $F_i \in \mathcal{NN}_{1,1}$, where each F_i has 4 hidden layers with architecture defined by

- $N_1 = N_2 = N_3 = N_4 = 10$,
- $\sigma(x) = \tanh(x)$.

We use the same weight initialization strategy as in Experiment 8.1. Again, we use stochastic gradient based optimization in form of the Adam method where we compute each optimization step with mini-batch size N_{batch} that needs to be specified. For the loss function u, we choose $u(x) = |x|$. We start by training F_1 using the data from the first slice and after completion, we freeze the weights of F_1 and continue with the next slice and so on. For each slice, we make N_{train} training steps. After termination, we use the trained weights that define L to compute model prices using a sample size of 10^9 simulations. We use the same Vega type weights as in Experiment 8.1 to calibrate implied volatility rather then prices.

We consider the following two concrete numerical experiments.

Experiment 8.2. *For each training step, set $N_{batch} = 500$ and make $N_{train} = 15000$ iterations. The plots consist of implied volatility of data and model, the error of implied volatility and a plot of $x \mapsto L^2(T_{i-1}, x)$ where, $T_0 = 0$ and $x \in [-K_i, K_i]$.*

Results for Experiment 8.2 are given in Figures 9.5 and 9.6, where each column corresponds to one maturity. As we compute model prices by simulating 10^9 pairs of correlated Brownian motions for the Euler discretization, the main computational time is due to the model implied volatility computation. Hence we present the computational time accordingly, by separating pure training time from implied volatility time computations. These results are given in Figure 8.3.

It is clear that in practice, one would apply the variance reduction technique for the model price computation to speed up the overall process. In particular, this would allow for an abort criterion in the training steps to avoid unnecessarily long training.

With the next experiment, we show that there can be a situation where a very high sensitivity between implied volatility and corresponding leverage function exists. The increase in accuracy needed for examining this is achieved by augmenting N_{batch}.

Experiment 8.3. *For each training step, set $N_{batch} = 3000$ and make $N_{train} = 15000$ iterations. For each slice, plot the intermediary training results after 7500 training steps. The plots consist of implied volatility of data and model, the error of implied volatility and a plot of $x \mapsto L^2(T_{i-1}, x)$ where, $T_0 = 0$ and $x \in [-K_i, K_i]$.*

We plot the intermediary and final results wrt training steps per slice in Figures 9.7 to 9.10. Each figure is showing the results wrt one of the four maturities at respective half time $N_{\text{train}}/2$ (left column) and finial training step $N_{\text{train}} = 15000$ (right column).

We can see from from both experiments, that a exact calibration is achieved up to a small error relaxation. However, by comparing the trained leverage functions at the final maturity in Figure 9.6 (bottom right) and Figure 9.10 (last row), one sees that little improvement on the implied volatility matching can lead to significant changes in the trained leverage function.

8.2.4 Conclusion

We have demonstrated how the parametrization by means of neural networks can be used to calibrate local stochastic volatility models to implied volatility data. We make the following remarks:

(i) The method we presented does not require any form of interpolation for the implied volatility surface since we do not calibrate via the Dupire formula. As the calibration result should not depend on the choice of the interpolation result, this is a desirable feature of our method.

(ii) It is possible to “plug in” any stochastic variance process such as rough volatility processes as long as an efficient simulation of trajectories is possible.

(iii) The multivariate extension is straight forward.

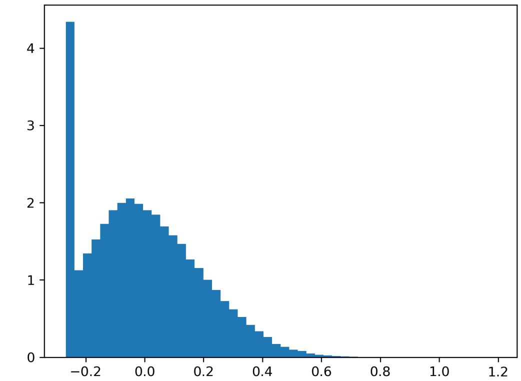
(iv) The level of accuracy of the calibration result is of a very high degree, making the presented method already of interest by this feature alone.

(v) The method can be significantly accelerated by applying distributed computation methods in the context of multi-gpu computational concepts.

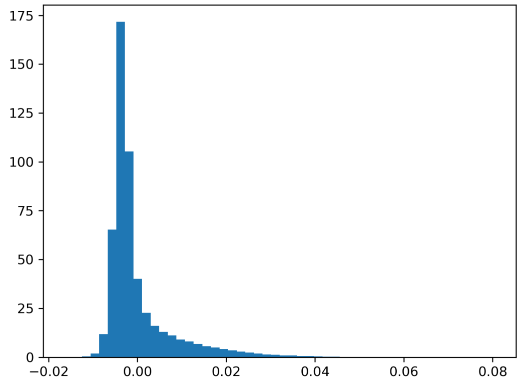
The presented method is - at least conceptionally - further able to deal with non-vanilla type European options since all computations are done by means of Monte Carlo simulations. We are further able to consider options on the variance process, e.g. VIX options, for the calibration. This will be investigated in a separate work as it goes beyond the scope of this thesis.

Chapter 9

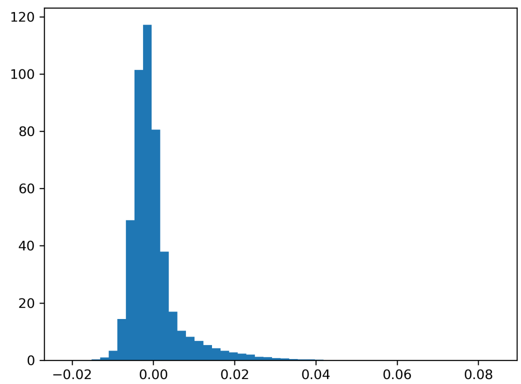
Plots and Histograms



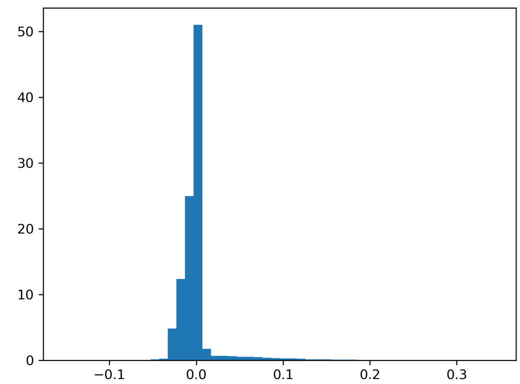
(a) *Scenario (i): Histogram without any variance reduction.*



(b) *Scenario (ii): Variance reduction using only the Black-scholes Δ-hedge.*

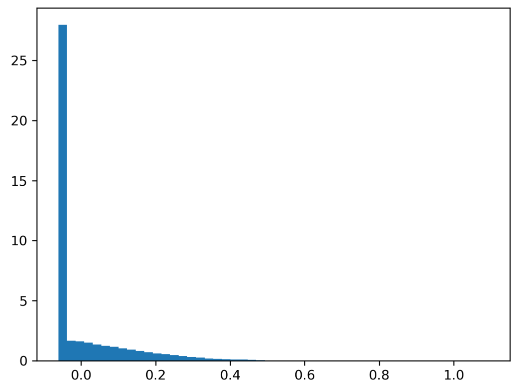


(c) *Scenario (iii): Variance reduction using the Black-Scholes Δ-hedge summed together with a trained neural network.*

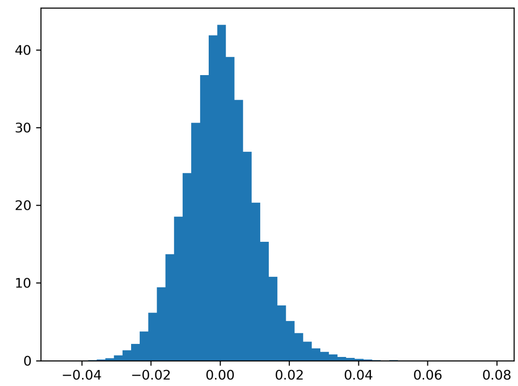


(d) *Scenario (iv): Variance reduction where the trading strategy is model only with a trained neural network.*

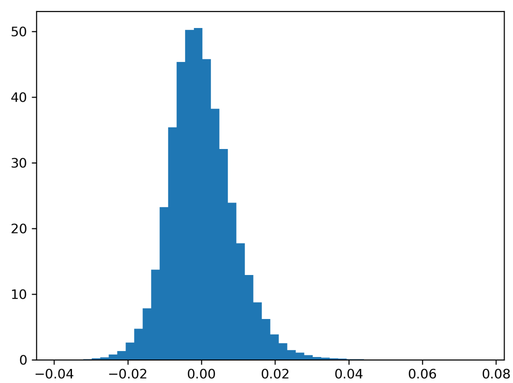
Figure 9.1: Normalized Monte Carlo histograms for Experiment 8.1 with strike $\boldsymbol{K_{1,1}}$ where we compute pricing errors applying different methods for variance reduction via hedging (scenarios (i) to (iv)). Note that the histograms have different scales as a result of the variance reduction. Normalization here means that the area of the blue bars sum up to one. For estimated standard deviations see Figure 8.2.



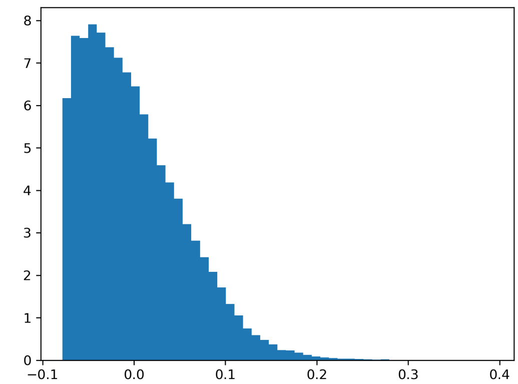
(a) *Scenario (i): Histogram without any variance reduction.*



(b) *Scenario (ii): Variance reduction using only the Black-scholes* Δ*-hedge.*

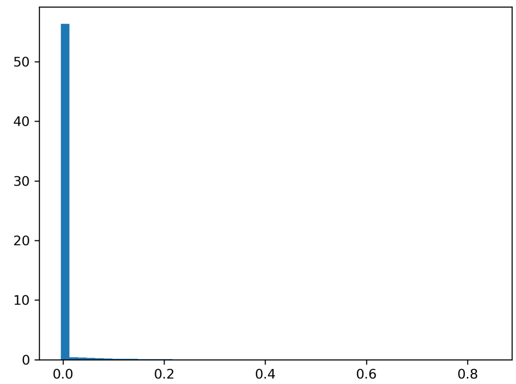


(c) *Scenario (iii): Variance reduction using the Black-Scholes* Δ*-hedge summed together with a trained neural network.*

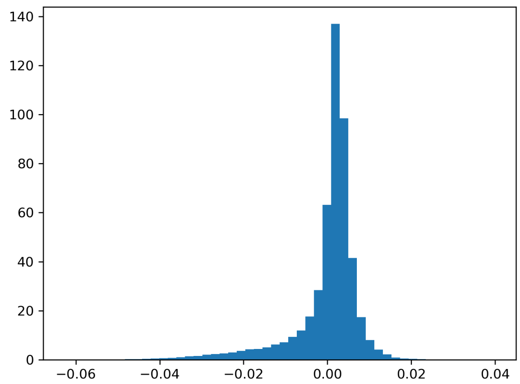


(d) *Scenario (iv): Variance reduction where the trading strategy is model only with a trained neural network.*

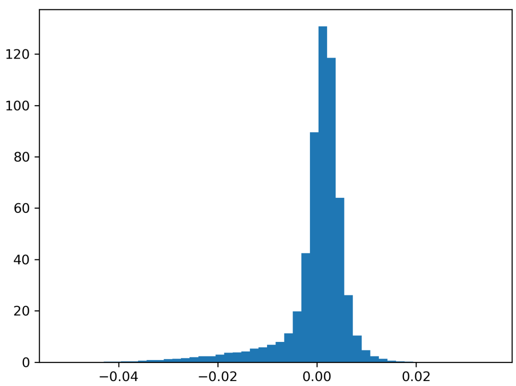
Figure 9.2: Normalized Monte Carlo histograms for Experiment 8.1 with strike $\boldsymbol{K_{1,2}}$ where we compute pricing errors applying different methods for variance reduction via hedging (scenarios (i) to (iv)). Note that the histograms have different scales as a result of the variance reduction. Normalization here means that the area of the blue bars sum up to one. For estimated standard deviations see Figure 8.2.



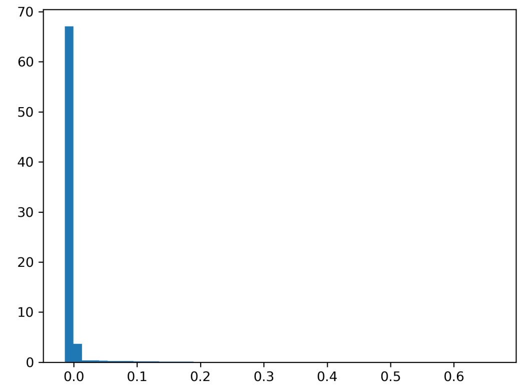
(a)



(b)



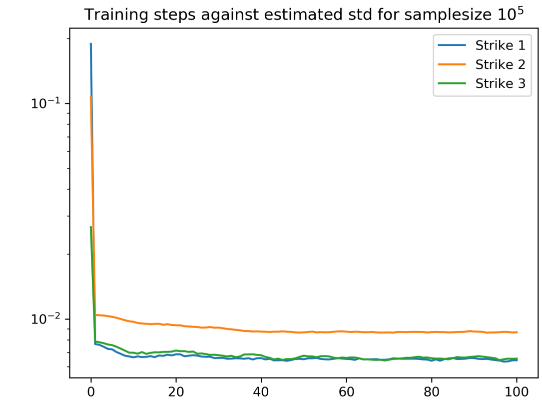
(c)



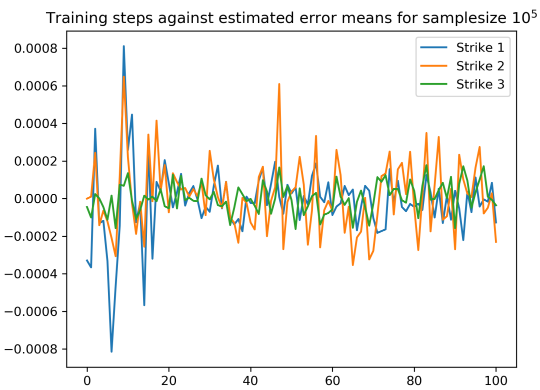
(d)

Figure 9.3: Normalized Monte Carlo histograms for Experiment 8.1 with strike $\boldsymbol{K_{1,3}}$ where we compute pricing errors applying different methods for variance reduction via hedging (scenarios (i) to (iv)). Note that the histograms have different scales as a result of the variance reduction. Normalization here means that the area of the blue bars sum up to one. For estimated standard deviations see Figure 8.2.

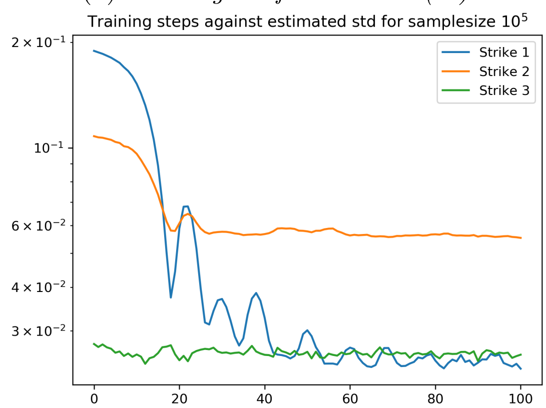
(a) *Running mean for scenario (iii)*



(b) *Running std for scenario (iii)*



(c) *Running mean for scenario (iv)*



(d) *Running std for scenario (iv)*

Figure 9.4: Estimated running mean (see figures (a) and (c)) and standard deviation (see figures (b) and (d)) for scenarios (*iii*) and (*iv*) of Experiment 8.1 against training steps. Training step zero corresponds to no variance reduction at all, not the variance reduction due to randomly drawn weights of the untrained neural networks.

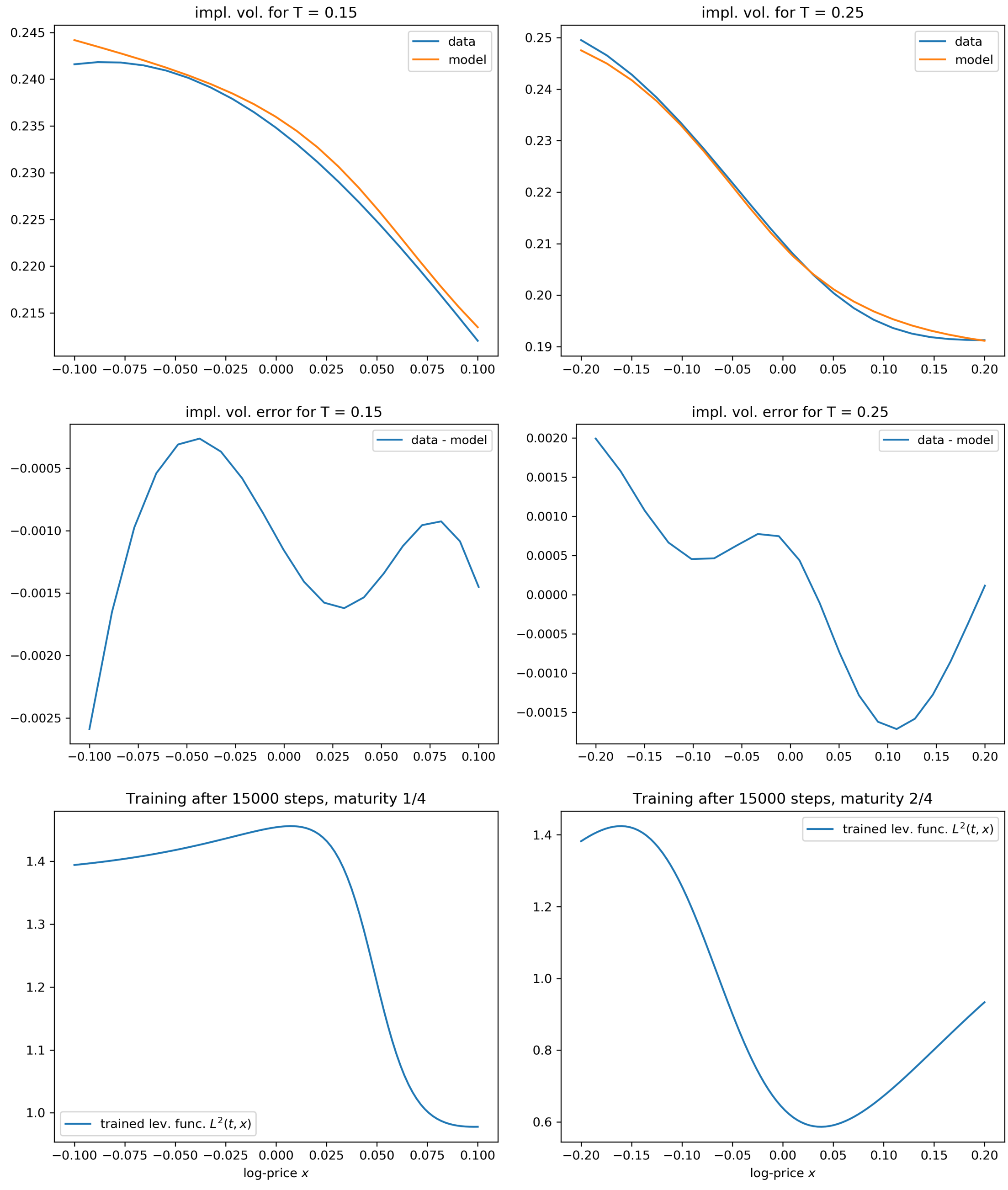


Figure 9.5: Plots for Experiment 8.2 for maturity $T = 0.15$ in the first column and $T = 0.25$ in the second column after 15000 training steps with $N_{\text{batch}} = 500$. Note that implied volatility and leverage function are given in log-strike coordinates $x = \log \frac{K}{S_0}$ for strike K.

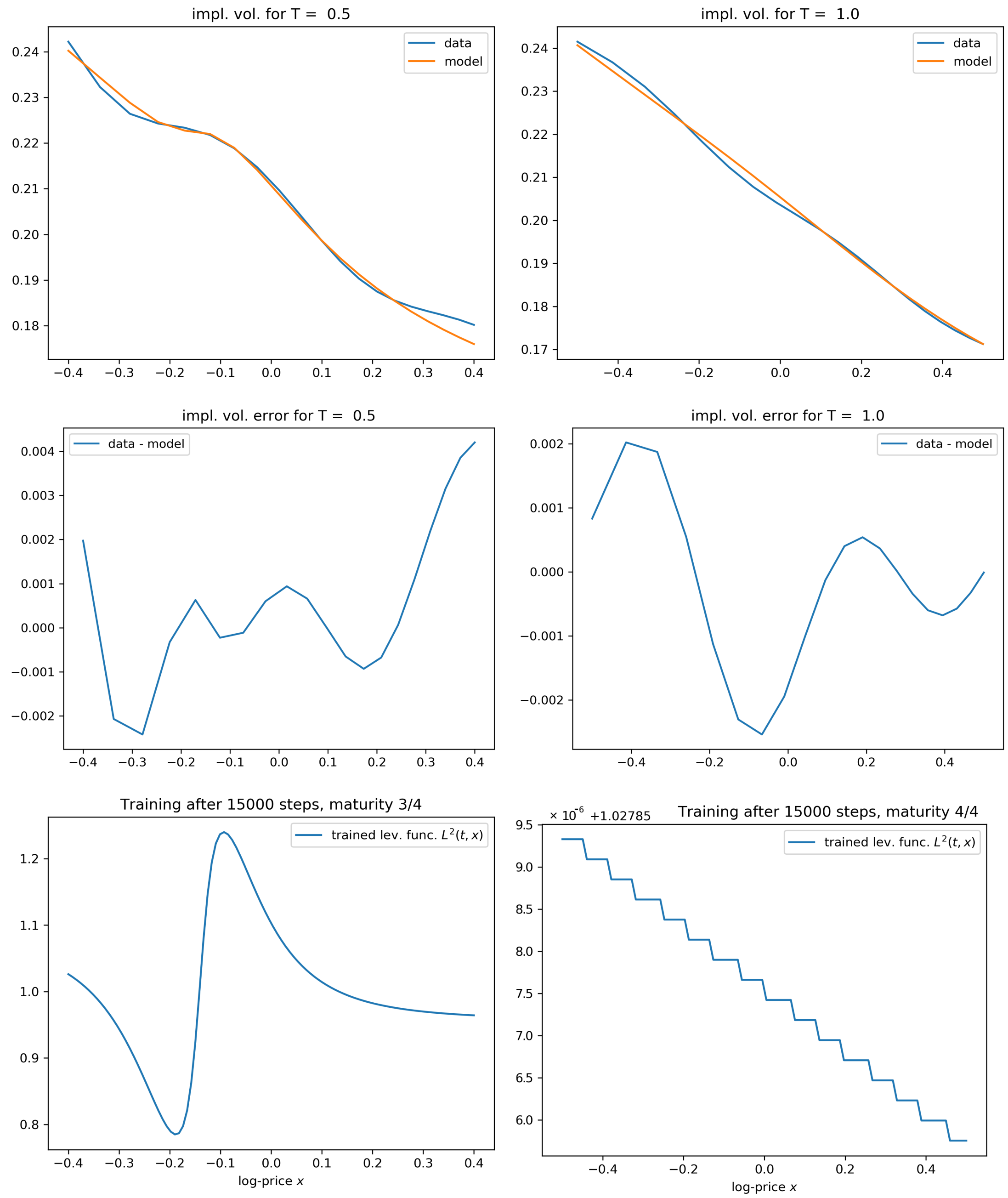


Figure 9.6: Plots for Experiment 8.2 for maturity $T = 0.5$ in the first column and $T = 1.0$ in the second column after 15000 training steps with $N_{\text{batch}} = 500$. Note that implied volatility and leverage function are given in log-strike coordinates $x = \log \frac{K}{S_0}$ for strike K. Mind the scale on the bottom right. The plot shows that the trained leverage function for the last maturity is basically constant.

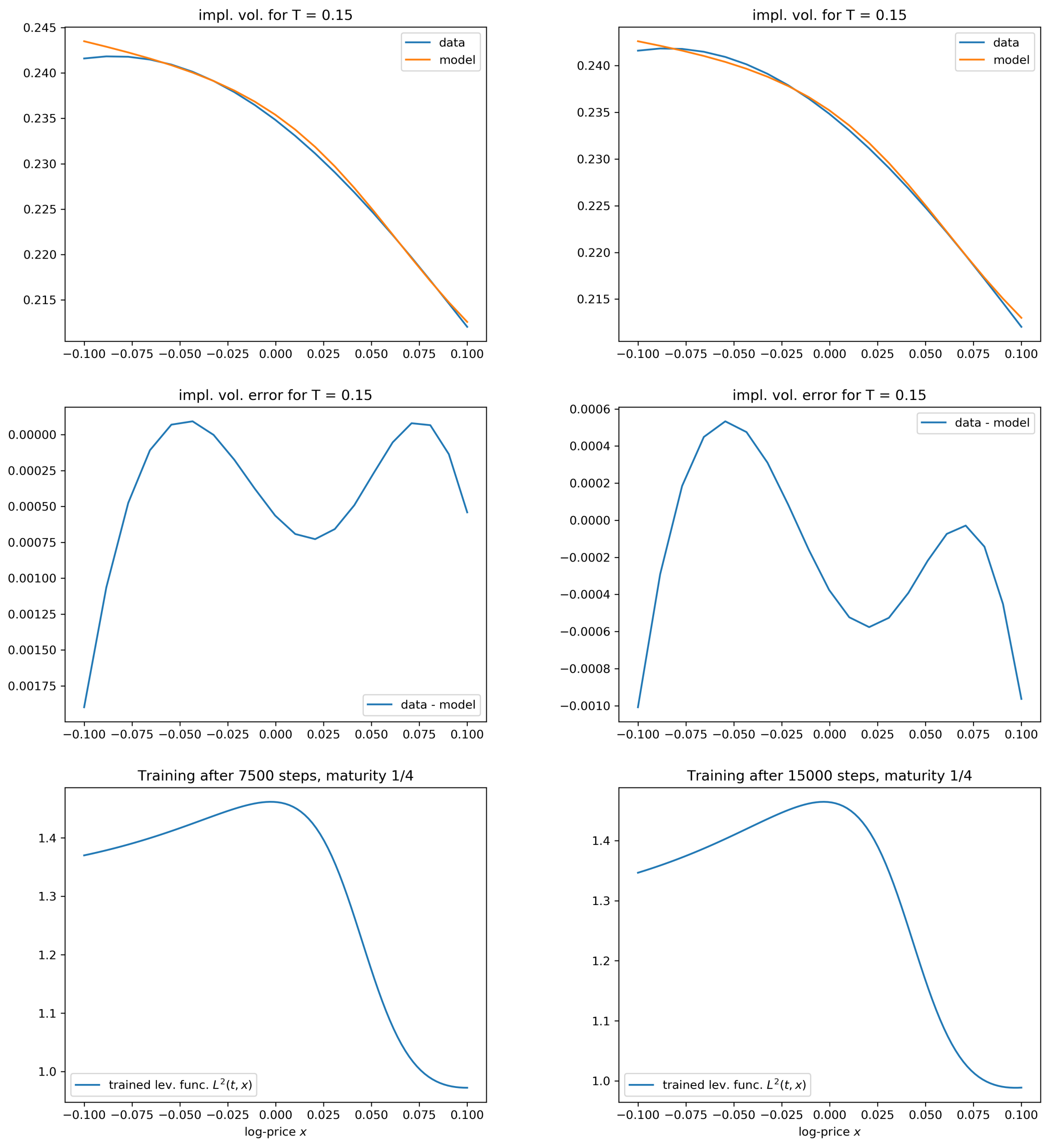


Figure 9.7: Plots for Experiment 8.3 for maturity $T = 0.15$ after 7500 (left column) and 15000 (right column) training steps with $N_{\text{batch}} = 3000$. Note that implied volatility and leverage function are given in log-strike coordinates $x = \log \frac{K}{S_0}$ for strike K.

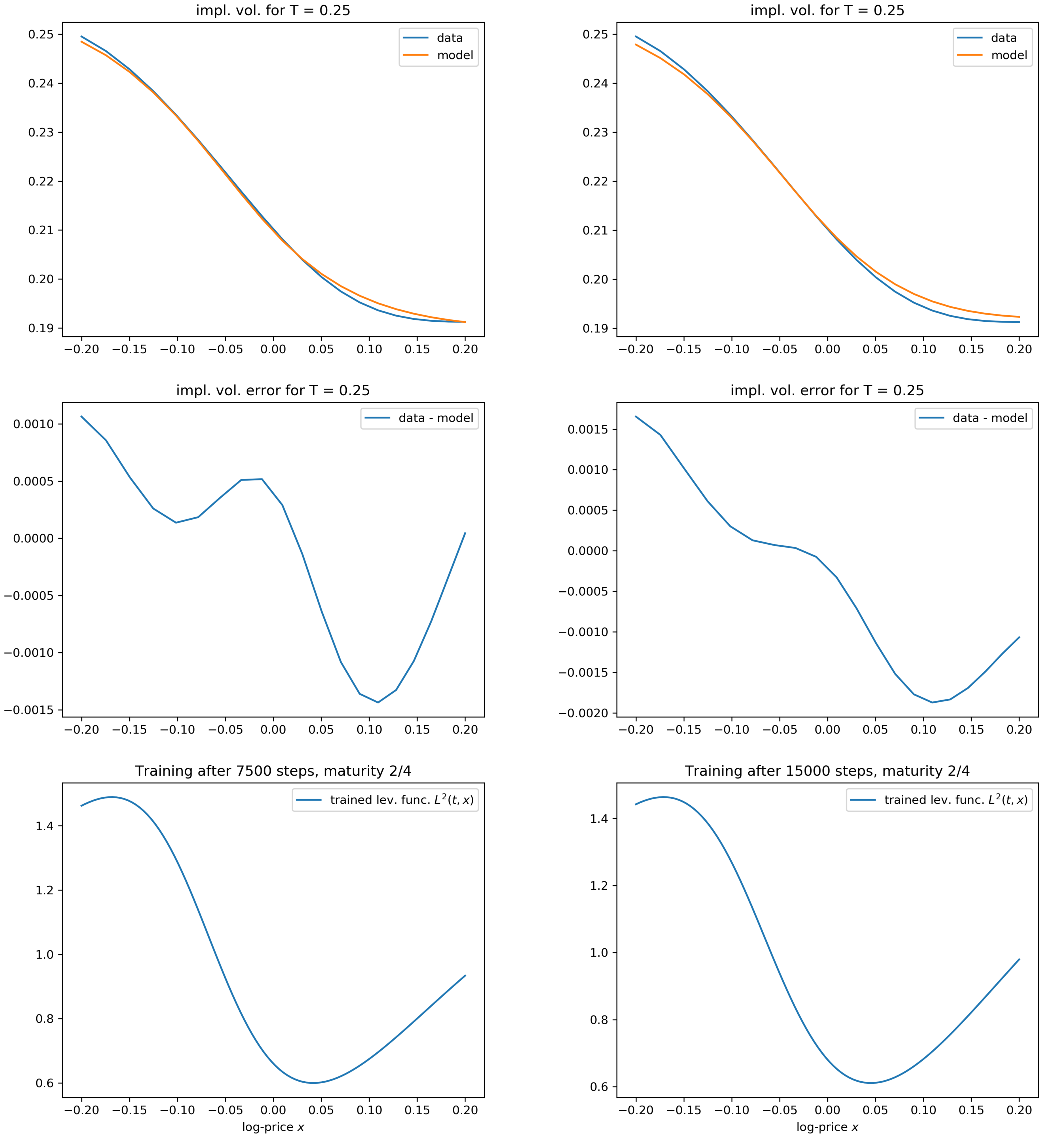


Figure 9.8: Plots for Experiment 8.3 for maturity $T = 0.25$ after 7500 (left column) and 15000 (right column) training steps with $N_{\text{batch}} = 3000$. Note that implied volatility and leverage function are given in log-strike coordinates $x = \log \frac{K}{S_0}$ for strike K.

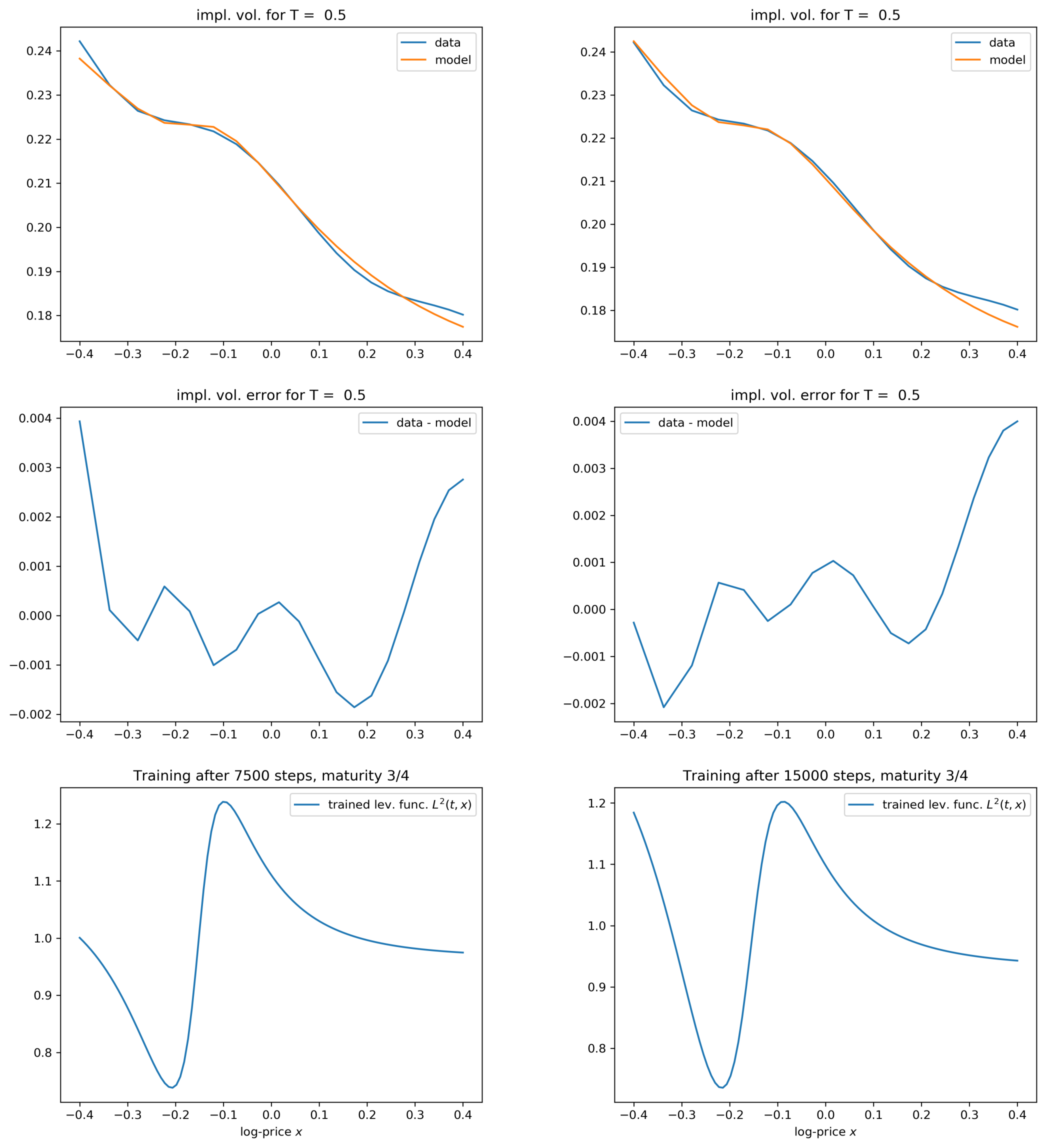


Figure 9.9: Plots for Experiment 8.3 for maturity $T = 0.5$ after 7500 (left column) and 15000 (right column) training steps with $N_{\text{batch}} = 3000$. Note that implied volatility and leverage function are given in log-strike coordinates $x = \log \frac{K}{S_0}$ for strike K.

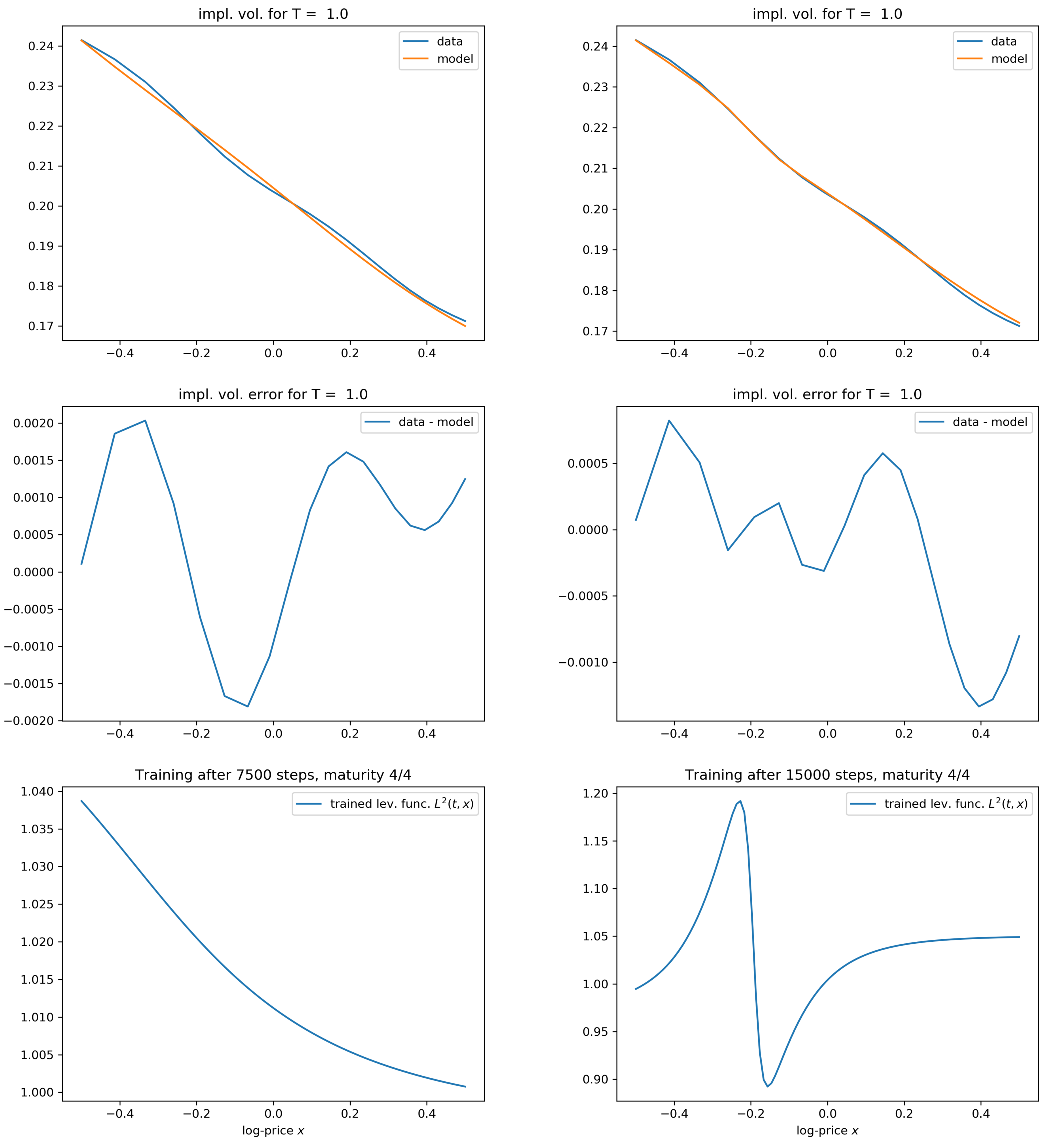


Figure 9.10: Plots for Experiment 8.3 for maturity $T = 1.0$ after 7500 (left column) and 15000 (right column) training steps with $N_{\text{batch}} = 3000$. Note that implied volatility and leverage function are given in log-strike coordinates $x = \log \frac{K}{S_0}$ for strike K. Also, note the difference between the trained leverage functions in the last row, despite the fact that the implied volatility error in the second row is already small after 7500 training steps.

Bibliography

[AP97] William F Ames and BG Pachpatte. ***Inequalities for differential and integral equations***. Vol. 197. Elsevier, 1997.

[AT10] Frédéric Abergel and Rémi Tachet. "A nonlinear partial integro-differential equation from mathematical finance". In: ***Discrete and Continuous Dynamical Systems-Series A*** 27.3 (2010), pp. 907–917.

[Acz66] János Aczél. ***Lectures on functional equations and their applications***. Vol. 19. Academic press, 1966.

[Adl85] Robert J Adler. "Random fields". In: ***Encyclopedia of Statistical Sciences*** (1985).

[Arn92] Vladmimir I. Arnold. ***Ordinary Differential Equations***. Springer Textbook. Springer Berlin Heidelberg, 1992. ISBN: 9783540548133.

[BC08] Alan Bain and Dan Crisan. ***Fundamentals of stochastic filtering***. Vol. 60. Springer Science & Business Media, 2008.

[BG07] Robert McCallum Blumenthal and Ronald Kay Getoor. ***Markov processes and potential theory***. Courier Corporation, 2007.

[BS18] Christian Bayer and Benjamin Stemper. "Deep calibration of rough stochastic volatility models". In: ***arXiv preprint arXiv:1810.03399*** (2018).

[BS73] Fischer Black and Myron Scholes. "The pricing of options and corporate liabilities". In: ***Journal of political economy*** 81.3 (1973), pp. 637–654.

[Ben+18] Fred Espen Benth, Nils Detering, and Paul Kruhner. "Multilinear processes in Banach space". In: ***arXiv preprint*** (2018).

[Bil74] Patrick Billingsley. "A note on separable stochastic processes". In: ***The Annals of Probability*** 2.3 (1974), pp. 476–479.

[Bue+18] Hans Buehler, Lukas Gonon, Josef Teichmann, and Ben Wood. "Deep hedging". In: (2018).

[Böt14] Björn Böttcher. "Feller evolution systems: Generators and approximation". In: ***Stochastics and Dynamics*** 14.03 (2014), p. 1350025.

[CAH17] Maria Fernanda del C. A. Hurtado. “Time-inhomogeneous polynomial processes in electricity spot price models”. PhD thesis. Albert-Ludwigs-Universität Freiburg, June 2017.

[CBH04] Rama Cont and Sana Ben Hamida. “Recovering volatility from option prices by evolutionary optimization”. In: (2004).

[CR98] Fabienne Comte and Eric Renault. “Long memory in continuous-time stochastic volatility models”. In: ***Mathematical finance*** 8.4 (1998), pp. 291–323.

[CT18] Christa Cuchiero and Josef Teichmann. “Generalized Feller processes and Markovian lifts of stochastic Volterra processes: the affine case”. In: ***arXiv preprint*** (2018).

[Coz+17] Andrei Cozma, Matthieu Mariapragassam, and Christoph Reisinger. “Calibration of a Hybrid Local-Stochastic Volatility Stochastic Rates Model with a Control Variate Particle Method”. In: ***Preprint, available at arXiv*** (2017).

[Cuc+12] Christa Cuchiero, Martin Keller-Ressel, and Josef Teichmann. “Polynomial processes and their applications to mathematical finance”. In: ***Finance and Stochastics*** 16.4 (2012), pp. 711–740.

[Cuc+18a] Christa Cuchiero, A. Marr, Mavuso M., A. Mitoulis, S. Singh, and J. Teichmann. “Calibration of mixture interest rate models with neural networks”. In: ***Working paper*** (2018).

[Cuc+18b] Christa Cuchiero, Martin Larsson, Sara Svaluto-Ferro, et al. “Polynomial jump-diffusions on the unit simplex”. In: ***The Annals of Applied Probability*** 28.4 (2018), pp. 2451–2500.

[Cuc+18c] Christa Cuchiero, Martin Larsson, and Sara Svaluto-Ferro. “Probability measure-valued polynomial diffusions”. In: ***arXiv preprint*** (2018).

[Cuc11] Christa Cuchiero. “Affine and polynomial processes”. PhD thesis. ETH-Zurich, 2011.

[Cuc18] Christa Cuchiero. “Polynomial processes in stochastic portfolio theory”. In: ***Stochastic Processes and their Applications*** (2018).

[DS72] Purna Candra Das and Rishi Ram Sharma. “Existence and stability of measure differential equations”. In: ***Czechoslovak Mathematical Journal*** 22.1 (1972), pp. 145–158.

[DS80] PC Das and RR Sharma. “Some stieltjes integral inequalities”. In: ***Journal of Mathematical Analysis and Applications*** 73.2 (1980), pp. 423–433.

[Dup96] Bruno Dupire. “A unified theory of volatility”. In: ***Derivatives pricing: The classic collection*** (1996), pp. 185–196.

[EER19] Omar El Euch and Mathieu Rosenbaum. “The characteristic function of rough Heston models”. In: ***Mathematical Finance*** 29.1 (2019), pp. 3–38.

[EK09] Stewart N Ethier and Thomas G Kurtz. ***Markov processes: characterization and convergence***. Vol. 282. John Wiley & Sons, 2009.

[EK90] Lynn Erbe and QingKai Kong. "Stieltjes integral inequalities of Gronwall type and applications". In: ***Annali di Matematica pura ed applicata*** 157.1 (1990), pp. 77–97.

[EN99] Klaus-Jochen Engel and Rainer Nagel. ***One-parameter semigroups for linear evolution equations***. Vol. 194. Springer Science & Business Media, 1999.

[FL16] Damir Filipović and Martin Larsson. "Polynomial diffusions and applications in finance". In: ***Finance and Stochastics*** 20.4 (2016), pp. 931–972.

[FL17] Damir Filipović and Martin Larsson. "Polynomial jump-diffusion models". In: ***Swiss Finance Institute Research Paper*** 17-60 (2017).

[Fil+17] Damir Filipović, Martin Larsson, and Sergio Pulido. "Markov cubature rules for polynomial processes". In: ***arXiv preprint*** (2017).

[GHL11] Julien Guyon and Pierre Henry-Labordere. "The smile calibration problem solved". In: ***Preprint, https://ssrn.com/abstract=1885032*** (2011).

[GHL13] Julien Guyon and Pierre Henry-Labordère. ***Nonlinear option pricing***. CRC Press, 2013.

[GKR18] Jim Gatheral and Martin Keller-Ressel. "Affine forward variance models". In: ***arXiv preprint*** (2018).

[Gat+14] Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum. "Volatility is rough". In: ***arXiv preprint*** (2014).

[Gil07] Michael Gil. ***Difference equations in normed spaces: stability and oscillations***. Vol. 206. Elsevier, 2007.

[HN92] Robert Hecht-Nielsen. "Theory of the backpropagation neural network". In: ***Neural networks for perception***. Elsevier, 1992, pp. 65–93.

[HW87] John Hull and Alan White. "The pricing of options on assets with stochastic volatilities". In: ***The journal of finance*** 42.2 (1987), pp. 281–300.

[He+92] Sheng-wu He, Jia-gang Wang, and Jia-an Yan. ***Semimartingale theory and stochastic calculus***. Routledge, 1992.

[Her17] Andres Hernandez. "Model Calibration with Neural Networks". In: ***Risk*** (2017).

[Hes93] Steven L Heston. "A closed-form solution for options with stochastic volatility with applications to bond and currency options". In: ***The review of financial studies*** 6.2 (1993), pp. 327–343.

[Hil70] David Hilbert. ***David Hilbert: Gesammelte Abhandlungen***. Springer, 1970.

[Hor+19] Blanka Horvath, Aitor Muguruza, and Mehdi Tomas. “Deep Learning Volatility”. In: ***Available at SSRN 3322085*** (2019).

[Hor91] Kurt Hornik. “Approximation capabilities of multilayer feedforward networks”. In: ***Neural networks*** 4.2 (1991), pp. 251–257.

[Hor96] Laszlo Horvath. “Gronwall–Bellman type integral inequalities in measure spaces”. In: ***Journal of mathematical analysis and applications*** 202.1 (1996), pp. 183–193.

[JP12] Jean Jacod and Philip Protter. ***Probability essentials***. Springer Science & Business Media, 2012.

[JS13] Jean Jacod and Albert N Shiryaev. ***Limit theorems for stochastic processes***. Vol. 288. Springer Science & Business Media, 2013.

[JZ16] Benjamin Jourdain and Alexandre Zhou. “Existence of a calibrated regime switching local volatility model and new fake Brownian motions”. In: ***Preprint available at arXiv*** (2016).

[Jab+17] Eduardo Abi Jaber, Martin Larsson, and Sergio Pulido. “Affine Volterra processes”. In: ***arXiv preprint*** (2017).

[Jac+05] Jean Jacod, Thomas G Kurtz, Sylvie Méléard, and Philip Protter. “The approximate Euler method for Lévy driven stochastic differential equations”. In: ***Annales de l’IHP Probabilités et statistiques***. Vol. 41. 3. 2005, pp. 523–558.

[KB14] Diederik P Kingma and Jimmy Ba. “Adam: A method for stochastic optimization”. In: ***arXiv preprint arXiv:1412.6980*** (2014).

[KR+18] Martin Keller-Ressel, Thorsten Schmidt, and Robert Wardenga. “Affine processes beyond stochastic continuity”. In: ***arXiv preprint*** (2018).

[Kal06] Olav Kallenberg. ***Foundations of modern probability***. Springer Science & Business Media, 2006.

[Kun69] Hiroshi Kunita. “Absolute continuity of Markov processes and generators”. In: ***Nagoya Mathematical Journal*** 36 (1969), pp. 1–26.

[Lip02] A. Lipton. “The vol smile problem.” In: ***Risk Magazine*** 15 (2002), 61–65.

[Mer74] Robert C. Merton. “On the pricing of corporate debt: The risk structure of interest rates”. In: ***The Journal of finance*** 29.2 (1974), pp. 449–470.

[Mey66] Paul André Meyer. ***Probability and potentials***. Vol. 1318. Blaisdell Pub. Co., 1966.

[Pro05] Philip Protter. ***Stochastic integration and differential equations***. Springer, 2005.

[RM51] Herbert Robbins and Sutton Monro. “A stochastic approximation method”. In: ***The annals of mathematical statistics*** (1951), pp. 400–407.

[Ren+07] Yong Ren, Dilip Madan, and M Qian Qian. “Calibrating and pricing with embedded local volatility models”. In: ***RISK-LONDON-RISK MAGAZINE LIMITED-*** 20.9 (2007), p. 138.

[Rud87] Walter Rudin. ***Real and complex analysis***. Tata McGraw-Hill Education, 1987.

[Sap+17] Yuri F Saporito, Xu Yang, and Jorge P Zubelli. “The Calibration of Stochastic-Local Volatility Models-An Inverse Problem Perspective”. In: ***Preprint, available at arXiv:1711.03023*** (2017).

[Sch17] René L Schilling. ***Measures, integrals and martingales***. 2nd edition. Cambridge University Press, 2017.

[Sch98] Philipp J. Schönbucher. “Term structure modelling of defaultable bonds”. In: ***Review of Derivatives Research*** 2.2-3 (1998), pp. 161–192.

[Sha72] RR Sharma. “An abstract measure differential equation”. In: ***Proceedings of the American Mathematical Society*** 32.2 (1972), pp. 503–510.

[Tia+15] Yu Tian, Zili Zhu, Geoffrey Lee, Fima Klebaner, and Kais Hamza. “Calibrating and pricing with a stochastic-local volatility model”. In: ***Journal of Derivatives*** 22.3 (2015), p. 21.

[Wal13] Wolfgang Walter. ***Gewöhnliche Differentialgleichungen: eine Einführung***. Springer-Verlag, 2013.

[Çin+80] Erhan Çinlar, Jean Jacod, Philip Protter, and MJ Sharpe. “Semimartingales and Markov processes”. In: ***Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*** 54.2 (1980), pp. 161–219.