Notes on Polynomial and Affine Jump-Diffusions

Martin Larsson[1]

December 25, 2017

[1]Department of Mathematics, ETH Zurich, Rämistrasse 101, CH-8092, Zurich, Switzerland, martin.larsson@math.ethz.ch.

Contents

Chapter 1

Introduction and preliminaries

1.1. Models for dynamic phenomena are frequently described by *local laws of motion.* Such phenomena are often stochastic in nature, or at least sufficiently complex that it makes sense to view some of their motion as "noise". This is in particular true in finance, where randomness or "noise" is a highly prevalent feature that cannot be ignored. In such situations, a local law of motion often takes the form of a stochastic differential equation, such as

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x_0, \tag{1.1}$$

where b and σ are suitably regular functions, W is Brownian motion, and $x_0 \in \mathbb{R}$. For example, the famous Black–Scholes model is of this form with $b(x) = \mu x$ and $\sigma(x) = \nu x$ for some constants μ and ν. There are many ways (1.1) can be generalized. For example, one can increase the dimension of X, allow the coefficients b and σ to depend on the entire history of X (not just its current value), replace the Brownian motion W by a more general process, or allow X to exhibit discontinuous trajectories.

The chosen model is only useful if statements can be made about its probabilistic properties. One would like to be able to compute expectations, variances, average excursion lengths, or other quantities of interest. In finance this corresponds to risk assessment, optimal trading decisions, asset valuation, etc. A standard device for performing such computations is the Feynman–Kac formula, which states that under suitable conditions and for any $T \geq 0$, the solution X of (1.1) satisfies

$$\mathbb{E}[f(X_T)] = u(0, x_0)$$

for a large class of functions f, where $u\colon [0,T]\times\mathbb{R}\to\mathbb{R}$ solves the partial differential equation

$$\begin{aligned} u_t(t,x) + b(x)u_x(t,x) + \frac{1}{2}\sigma(x)^2 u_{xx}(t,x) &= 0, \qquad && (t,x)\in[0,T)\times\mathbb{R},\\ u(T,x) &= f(x), && x\in\mathbb{R}. \end{aligned}$$

(The subscripts denote partial derivatives.) A variety of numerical methods exist for solving such equations. Another common method is Monte-Carlo simulation, which in its basic form consists of generating a large number of independent replications $X_T^{(1)},\ldots,X_T^{(n)}$ of X_T, and then using the law of large numbers to obtain

$$\mathbb{E}[f(X_T)] \approx \frac{1}{n}\sum_{i=1}^n f(X_T^{(i)}).$$

However, if the dimensionality of X grows large, or if $\mathbb{E}[f(X_T)]$ has to be computed a large number of times (e.g. for different functions f or different coefficients b and σ), such methods eventually become computationally taxing. It is therefore of interest to look for classes of processes X and functions f with enhanced tractability. The goal of these notes is to consider two possibilities:

- Affine jump-diffusions with $f(x)=e^{ux}$ where u is constant, and
- Polynomial jump-diffusions with $f(x)$ a polynomial in x.

The scope is surprisingly broad. In finance this leads to models for equities, interest rates, credit risk, optimal investment, economic equilibrium, etc. We will look at some of these applications later on.

1.2. Let us consider a basic example of how special structure of X can be used to simplify the Feynman–Kac formula. Assume X satisfies

$$dX_t = \beta X_t dt + \sqrt{X_t}\, dW_t, \quad X_0 = x_0, \tag{1.2}$$

for some constants β and x_0. We assume that X_t is nonnegative for all t so that the square-root is well-defined. We aim to compute

$$\mathbb{E}[e^{uX_T}]$$

for $T>0$ and $u\in\mathbb{R}$. As an Ansatz, consider the process

$$M_t = e^{\psi(T-t)X_t}, \quad t\geq 0,$$

for a smooth function ψ to be determined later. An application of Itô's formula with the function $f(t,x)=\exp(\psi(T-t)x)$ yields

$$\begin{aligned} dM_t &= M_t\left(-\psi'(T-t)X_t\right)dt + M_t\psi(T-t)dX_t + \frac{1}{2}M_t\psi(T-t)^2 d\langle X\rangle_t \\ &= M_t\left(-\psi'(T-t)+\beta\psi(T-t)+\frac{1}{2}\psi(T-t)^2\right)X_t dt + M_t\psi(T-t)\sqrt{X_t}dW_t. \end{aligned}$$

Therefore, if ψ solves the ODE

$$\psi' = \beta\psi + \frac{1}{2}\psi^2, \qquad \psi(0)=u, \tag{1.3}$$

then M is a local martingale with $M_T = e^{uX_T}$. If it is even a true martingale, we obtain

$$\mathbb{E}[e^{uX_T}] = \mathbb{E}[M_T] = M_0 = e^{\psi(T)x_0}.$$

We see that in order to compute $\mathbb{E}[e^{uX_T}]$ one does not need to solve a Feynman–Kac type PDE, but is left with the much easier task of finding the solution of the ODE (1.3).

Remark 1.3.

- The above argument is very similar in flavor to the proof of the Feynman–Kac formula: Make a suitable Ansatz, apply Itô's formula, identify a deterministic equation acting as (local) martingale condition, and finally take expectations to express the quantity of interest using the solution of the deterministic equation.
- The special form of dX_t played a crucial role. Can you find a more general form of dX_t for which a similar calculation still works?
- We did not verify that (i) X exists, (ii) ψ exists, and (iii) M is a true (not just local) martingale. These points are delicate because the SDE (1.2) for X and the ODE (1.3) for ψ both involve non-Lipschitz coefficients. In fact, without any further conditions, the above computation is not legitimate; try for instance $\beta=0$, $x_0=1$, $u=1$, and $T\geq 2$.

1.4. The computation in 1.2 captures the essence of the idea behind *affine* jump-diffusions. A similar argument based on polynomials rather than exponentials lies at the heart of *polynomial* jump-diffusions. As an exercise, try to compute the second moment $\mathbb{E}[X_T^2]$ of the process in (1.2) by considering the Ansatz $M_t = \phi(T-t) + \psi(T-t)X_t + \pi(T-t)X_t^2$ for suitable functions ϕ, ψ, and π! In these notes we will develop these ideas is much greater generality and look at how they can be brought to bear on applications in finance.

1.1 Notation and preliminaries

1.5. We always work on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ satisfying the *usual conditions*. That is, $\mathcal{F}$ is $\mathbb{P}$-complete, the filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ is right-continuous, and $\mathcal{F}_0$ contains all $\mathbb{P}$-nullsets of $\mathcal{F}$.

1.6. A process $X = (X_t)_{t\geq 0}$ is called *càdlàg* (or *RCLL*) if all its paths are right-continuous with left limits. In this case one defines the two processes X_- and ΔX by

$$X_{0-} = X_0, \quad X_{t-} = \lim_{s\uparrow t} X_s \text{ for } t > 0,$$
$$\Delta X_t = X_t - X_{t-} \text{ for } t \geq 0.$$

The process X is of *finite variation (FV)* if all its paths are of finite variation. If X is càdlàg, adapted, and of finite variation, we write $X \in \mathrm{FV}$, and if in addition $X_0 = 0$, we write $X \in \mathrm{FV}_0$. The total variation process $\mathrm{Var}(X)$ is again càdlàg and adapted ($\longrightarrow$ exercise). The set of càdlàg local martingales is denoted by $\mathcal{M}_{\mathrm{loc}}$, the subspace of continuous local martingales by $\mathcal{M}^c_{\mathrm{loc}}$, and we write $\mathcal{M}_{0,\mathrm{loc}}$ respectively $\mathcal{M}^c_{0,\mathrm{loc}}$ if the value at time zero is zero. We will often deal with multidimensional processes $X = (X^1, \ldots, X^d)$, in which case the above notions are applied component-wise. Thus, for instance, $X \in \mathcal{M}_{\mathrm{loc}}$ then means that $X^i \in \mathcal{M}_{\mathrm{loc}}$ for $i = 1, \ldots, d$.

Exercise 1.7. Every $X \in \mathrm{FV}_0$ admits a unique (up to indistinguishability) decomposition $X = A - B$ with A, B nondecreasing and in FV_0 and $A + B = \mathrm{Var}(X)$.

1.8. Since the usual conditions hold, every local martingale X admits a càdlàg modification, unique up to indistinguishability.[1] We always choose such a modification, which has the advantage that $X_{t-}(\omega)$ and $\Delta X_t(\omega)$ are well-defined for *every* $(t, \omega) \in \mathbb{R}_+ \times \Omega$. One could dispense with the usual conditions at the cost of more involved bookkeeping of nullsets.

[1] See Revuz and Yor (1999, Theorem II.2.9).

Chapter 2

Semimartingales and characteristics

Before discussing *affine* and *polynomial* jump-diffusions, we need to discuss *jump-diffusions* (not necessarily polynomial or affine!). This involves elements of the theory of semimartingales, such as special semimartingales and semimartingale characteristics, eventually leading to the crucial notion of an *(extended) generator*. While a more parsimonious presentation would have been possible, it seems worthwhile to develop some aspects of semimartingale theory, which is interesting and useful in its own right.

2.1 Semimartingales

Definition 2.1. A *semimartingale* is a process of the form $X = X_0 + M + A$, where X_0 is an $\mathcal{F}_0$-measurable random variable, $M \in \mathcal{M}_{0,\mathrm{loc}}$, and $A \in \mathrm{FV}_0$.

Example 2.2. The following are all semimartingales: Any deterministic function of finite variation; Brownian motion; Brownian motion with drift; Solutions of stochastic differential equations; Any Lévy process.

Exercise 2.3. (i) Consider the deterministic process $X_t = \sum_{n=1}^{\infty} \frac{1}{n}(-1)^n \mathbf{1}_{[t_n,\infty)}(t)$, $t \geq 0$, where $t_n = 1 - 1/n$. Is X well-defined? Is it a semimartingale?

(ii) Consider the stochastic process $X_t = \sum_{n=1}^{\infty} \frac{1}{n} B_n \mathbf{1}_{[t_n,\infty)}(t)$, $t \geq 0$, where $t_n = 1 - 1/n$

and $\{B_n : n \in \mathbb{N}\}$ is a sequence of iid Bernoulli random variables with $\mathbb{P}(B_n = 1) = \mathbb{P}(B_n = -1) = 1/2$. Is X well-defined? Is it a semimartingale?

2.4. One reason semimartingales are important is because they allow for a rich stochastic integration theory. In fact, the so-called Bichteler-Dellacherie theorem characterizes semimartingales as *good integrators*. A different way to characterize semimartingales is via a *compensation property*: any suitably regular function f can be assigned a predictable (see below) process $A^f \in \mathrm{FV}$ such that the compensated process $f(X) - f(X_0) - A^f$ is a local martingale. The latter property is the most relevant one for the theory of affine and polynomial jump-diffusions, while stochastic integration plays a rather indirect role.

2.5. The decomposition $X = X_0 + M + A$ is not unique in general. This is because there exist non-constant finite variation martingales, such as the compensated Poisson process $X_t = N_t - \lambda t$, where N is a Poisson process with intensity $\lambda > 0$. For this process, $X = 0 + 0 + X$, $X = 0 + X + 0$, and $X = 0 + \frac{1}{2}X + \frac{1}{2}X$ are some possible decompositions. As we will see, a unique decomposition is obtained if A is in addition required to be *predictable*.

2.2 Predictability and the Doob–Meyer decomposition

In this section we study the notion of *predictability*. We start with basic definitions and properties that follow from measure-theoretic manipulations. We then state (but do not prove) the important *Doob–Meyer decomposition theorem*, which we use to obtain *compensators* of sufficiently integrable increasing processes. Recall that we always work under the usual conditions.

Definition 2.6.

- The *predictable σ-algebra* is the σ-algebra $\mathcal{P}$ on $\mathbb{R}_+ \times \Omega$ generated by all left-continuous adapted processes (viewed as maps from $\mathbb{R}_+ \times \Omega$ to $\mathbb{R}$).
- A process X is *predictable* if the map $(t, \omega) \mapsto X_t(\omega)$ is $\mathcal{P}$-measurable.

2.7. It is often useful to work with *stochastic intervals*. For example,

$$[\![\sigma, \tau]\!] = \{(t, \omega) \in \mathbb{R}_+ \times \Omega : \sigma(\omega) \leq t \leq \tau(\omega)\}$$

for any stopping times σ and τ. Other intervals such as $[\![\sigma,\tau[\![$ and $[\![\sigma,\tau]\!]$ are defined analogously. The *graph* of a stopping time τ is $[\![\tau]\!] = [\![\tau,\tau]\!]$. **Caution:** By definition, all stochastic intervals are disjoint from $\{\infty\} \times \Omega$. In particular, $[\![0,\infty]\!] = \mathbb{R}_+ \times \Omega$.

Lemma 2.8. (i) *$\mathcal{P}$ is generated by all sets of the form $\{0\} \times A$, with $A \in \mathcal{F}_0$, and $[\![0,\tau]\!]$, with τ a stopping time.*

(ii) *$\mathcal{P}$ is generated by all sets of the form $\{0\} \times A$, with $A \in \mathcal{F}_0$, and $(s,t] \times A$, with $s < t$ and $A \in \mathcal{F}_s$.*

(iii) *If X is a predictable process and τ a stopping time, then X^τ is predictable.*

(iv) *If $X \in \mathrm{FV}_0$ is predictable, then $\mathrm{Var}(X)$ is predictable.*

Proof. (i) and (ii): Let $\mathcal{P}'$ respectively $\mathcal{P}''$ be the σ-algebra generated by all sets of the form in (i) respectively (ii). If $G \in \mathcal{P}'$, then the process X defined by $X_t(\omega) = \mathbf{1}_G(t,\omega)$ is left-continuous and adapted. Thus $\mathcal{P}' \subseteq \mathcal{P}$. Moreover, for $s < t$ and $A \in \mathcal{F}_s$ one has $(s,t] \times A = [\![0,\tau]\!] \setminus [\![0,\sigma]\!]$ for the stopping times $\sigma = s\mathbf{1}_A + \infty\mathbf{1}_{A^c}$ and $\tau = t\mathbf{1}_A + \infty\mathbf{1}_{A^c}$. Thus $\mathcal{P}'' \subseteq \mathcal{P}'$. Finally, let X be any left-continuous adapted process, and define for each $n \in \mathbb{N}$ a process X^n by

$$X^n_t(\omega) = X_0(\omega)\mathbf{1}_{\{0\}}(t) + \sum_{k\in\mathbb{N}} X_{k2^{-n}}(\omega)\mathbf{1}_{(k2^{-n},(k+1)2^{-n}]}(t).$$

Each X^n is $\mathcal{P}''$-measurable ($\longrightarrow$ exercise), and $X^n \to X$ pointwise by left-continuity. Thus X is also $\mathcal{P}''$-measurable, and we deduce that $\mathcal{P} \subseteq \mathcal{P}''$. This proves (i) and (ii).

(iii): If $X = \mathbf{1}_G$ is the indicator process of a set G of the form in (i), then stopping X yields a predictable process ($\longrightarrow$ exercise). Now apply the monotone class theorem.

(iv): Write $\mathrm{Var(X)} = \mathrm{Var(X)}_- + \Delta\mathrm{Var(X)}$. This is predictable because $\mathrm{Var(X)}_-$ is left-continuous and adapted, and $\Delta\mathrm{Var(X)} = |\Delta X| = |X - X_-|$. $\square$

Exercise 2.9. Show that every predictable $X \in \mathrm{FV}_0$ admits a unique (up to indistinguishability) decomposition $X = A - B$ with A, B predictable, nondecreasing, and in FV_0.

2.10. The *Doob–Meyer decomposition theorem* is a cornerstone of continuous time martingale theory. We will not prove this rather difficult result here; several different proofs exist in the literature.[1] To state the theorem, we need some terminology:

- A process X is said to be of *Class (D)* if $\{X_\tau : \tau \text{ is a finite stopping time}\}$ is a uniformly integrable family.
- A nondecreasing process $A \in \mathrm{FV}_0$ is called *integrable* if $\mathbb{E}[A_\infty] < \infty$. A process $A \in \mathrm{FV}_0$ is of *integrable variation* if $\mathbb{E}[\mathrm{Var}(A)_\infty] < \infty$, or equivalently, if the processes A^+ and A^- in the decomposition $A = A^+ - A^-$ of Exercise 1.7 are integrable.
- A process A is *locally integrable* or of *locally integrable variation*, respectively, if there is a sequence (τ_n) of stopping times, a.s. tending to infinity, such that A^{τ_n} satisfies the corresponding property for each n. We call (τ_n) a *localizing sequence.*

Theorem 2.11 (Doob–Meyer decomposition)**.** *If X is a càdlàg submartingale of Class (D), then there exists a unique (up to indistinguishability) predictable process $A \in \mathrm{FV}_0$ such that $X - A$ is a uniformly integrable martingale. The process A is nondecreasing and integrable.*

Corollary 2.12. *Let $M \in \mathcal{M}_{\mathrm{loc}} \cap \mathrm{FV}$. If M is predictable, then M is a.s. constant.*

Proof. We may assume that $M_0 = 0$. For any stopping time τ, M^τ is still predictable by Lemma 2.8(iii). By localization we may therefore assume M is a uniformly integrable martingale, hence of Class (D) by the stopping theorem. Since $M = M + 0 = 0 + M$ are two possible Doob–Meyer decompositions of M, the uniqueness assertion in Theorem 2.11 yields $M = 0$. □

Corollary 2.13. *Let $A \in \mathrm{FV}_0$ be of locally integrable variation. Then there exists a unique (up to indistinguishability) predictable process $A^p \in \mathrm{FV}_0$ of locally integrable variation, called the* compensator *of A, such that $A - A^p$ is a local martingale. If A is nondecreasing, then so is A^p.*

Proof. Uniqueness follows from Corollary 2.12, so let us prove existence. Due to Exercise 2.9 we may assume that A is nondecreasing. Let (τ_n) be a localizing sequence so that $\mathbb{E}[A_{\tau_n}] < \infty$ for each n. Then each A^{τ_n} is a submartingale of Class (D), and Theorem 2.11 yields nondecreasing predictable $B^n \in \mathrm{FV}_0$ such that $A^{\tau_n} - B^n$ is a uniformly

[1] Meyer's original proof can be found in Meyer (1962, 1963), while Protter (2005, Chapter III.3) presents a proof based on Bass (1996). See Beiglböck et al. (2012) for a different, more recent approach.

integrable martingale. By uniqueness (and Lemma 2.8(iii)), $(B^{n+1})^{\tau_n} = B^n$. The process $A^p = \sum_{n=1}^\infty B^n \mathbf{1}_{]\!]\tau_{n-1},\tau_n]\!]}$, where we set $\tau_0 = 0$, then satisfies the required properties. □

Example 2.14. Let N be a Poisson process with intensity $\lambda > 0$. Then its compensator is $N_t^p = \lambda t$, because $N_t - \lambda t$ is a martingale. On the other hand, if $A \in \mathrm{FV}_0$ is a continuous process, then A is automatically predictable, and is therefore equal to its compensator: $A^p = A$. This shows why the notion of compensator is only relevant when processes with jumps are involved.

Exercise 2.15. Exhibit a process A, on a suitable filtered probability space, which lies in FV_0 but is not of locally integrable variation.

2.3 Special semimartingales

2.16. If a semimartingale X has a decomposition $X = X_0 + M + A$ where A is predictable, then this decomposition is unique. Indeed, If $X = X_0 + M' + A'$ is another decomposition, then $M - M' = A' - A \in \mathcal{M}_{\mathrm{loc}} \cap \mathrm{FV}_0$, so that Corollary 2.12 yields $A' = A$ and $M' = M$.

Definition 2.17. A semimartingale X is called *special* if it admits a decomposition $X = X_0 + M + A$ where A can be chosen predictable. This decomposition (which is unique) is called the *canonical decomposition* of X.

Exercise 2.18. Let $X = X_0 + M + A$ be a semimartingale. Show that the following conditions are equivalent:

(i) X is special;

(ii) $\sup_{s\le\cdot} |X_s - X_0|$ is locally integrable;

(iii) $\sup_{s\le\cdot} |\Delta A_s|$ is locally integrable; and

(iv) A is of locally integrable variation.

Hints: First prove that $\sup_{s\le\cdot} |M_s|$ *is locally integrable for any* $M \in \mathcal{M}_{\mathrm{loc}}$. *Then prove the implications (i)* $\Rightarrow$ *(ii)* $\Rightarrow$ *(iii)* $\Rightarrow$ *(iv)* $\Rightarrow$ *(i). For the first implications, you may use that any predictable process in* FV_0 *is of locally integrable variation (c.f. Lemma I.3.10 in Jacod and Shiryaev (2003)).*

2.4 The structure of local martingales

Definition 2.19. A local martingale $M \in \mathcal{M}_{\text{loc}}$ is *purely discontinuous* if $MN \in \mathcal{M}_{0,\text{loc}}$ for every $N \in \mathcal{M}^c_{\text{loc}}$. The set of all purely discontinuous local martingales is denoted $\mathcal{M}^d_{\text{loc}}$.

2.20. The definition says $M \in \mathcal{M}_{\text{loc}}$ is purely discontinuous if it is in a certain sense orthogonal to all continuous local martingales. This does not mean that M is "pure jump" in the sense of being constant between jumps; see Example 2.21. Nonetheless, a purely discontinuous local martingale is completely determined by its jumps; see Lemma 2.22. Note that every $M \in \mathcal{M}^d_{\text{loc}}$ starts at zero, so we have no need for the notation "$\mathcal{M}^d_{0,\text{loc}}$".

Example 2.21. The compensated Poisson process $M_t = N_t - \lambda t$ is a purely discontinuous martingale (careful with the notation: here N denotes a Poisson process with intensity $\lambda > 0$!). This can be seen using stochastic integration theory for general local martingales. One can also verify this directly using the properties of the Poisson process ($\longrightarrow$ exercise).

Lemma 2.22. *If $M, N \in \mathcal{M}^d_{\text{loc}}$ satisfy $\Delta M = \Delta N$, then $M = N$.*

Proof. The process $L = M - N$ is a local martingale with $L_0 = 0$ and $\Delta L = 0$, so that $L \in \mathcal{M}^c_{0,\text{loc}}$. It is also orthogonal to every continuous local martingale, in particular to itself. Thus L and L^2 are both in $\mathcal{M}^c_{0,\text{loc}}$ and hence equal to zero. □

Theorem 2.23 (Decomposition of local martingales)**.** *Every local martingale M admits a unique (up to indistinguishability) decomposition $M = M_0 + M^c + M^d$ where $M^c \in \mathcal{M}^c_{0,\text{loc}}$ and $M^d \in \mathcal{M}^d_{0,\text{loc}}$.*

2.24. We will not give a full proof of Theorem 2.23, but rather outline some special cases.

- Recall that $\mathcal{H}^2$ denotes the Hilbert space of all L^2-bounded martingales M with inner product $(M, N)_{\mathcal{H}^2} = \mathbb{E}[M_\infty N_\infty]$. The subspace $\mathcal{H}^{2,c}$ of continuous elements is closed, and its orthogonal complement consists of purely discontinuous martingales; this is a consequence of the theory of stable subspaces.[2] The assertion of Theorem 2.23 for $M \in \mathcal{H}^2$ thus follows from the Hilbert space decomposition $\mathcal{H}^2 = \mathcal{H}^{2,c} \oplus (\mathcal{H}^{2,c})^\perp$.

[2] See, e.g., Chapter 4.5 in M. Schweizer's lecture notes on Brownian Motion and Stochastic Calculus. Alternatively, see Protter (2005, Chapter IV.3).

- Using this, the general case can be deduced from the following two facts: (i) every local martingale M can be decomposed into $M = M' + M''$, where M' has bounded jumps and M'' is of finite variation; and (ii) any local martingale of finite variation is purely discontinuous. The proofs of both facts are somewhat involved.

- A more intuitive approach is the following: Suppose M has summable and locally integrable jumps in the sense that the process $J_t = \sum_{s \le t} |\Delta M_s|$ is locally integrable. The process $A_t = \sum_{s \le t} \Delta M_s$ is then well-defined and of locally integrable variation, and thus has a compensator A^p by Corollary 2.13. The local martingale $M^d = A - A^p$, which can be understood as a "compensated sum of jumps", is purely discontinuous with the same jumps as M (this requires some effort to prove). Thus $M^c = M - M^d - M_0$ is in $\mathcal{M}^c_{0,\mathrm{loc}}$, which yields the desired decomposition. One still has to work rather hard to eliminate the local integrability assumption on J.

2.5 Jump measures and their compensators

Definition 2.25. The *jump measure* of a d-dimensional càdlàg process X is the random measure μ on $\mathcal{B}(\mathbb{R}_+ \times \mathbb{R}^d)$ given by

$$\mu(dt, d\xi; \omega) = \sum_{t>0} \mathbf{1}_{\{\Delta X_t(\omega) \neq 0\}} \delta_{(t, \Delta X_t(\omega))}(dt, d\xi)$$

2.26. For each ω, one thus has a measure on $\mathbb{R}_+ \times \mathbb{R}^d$ consisting of at most countably many atoms (càdlàg processes have at most countably many jumps), and assigning zero mass to the set $\mathbb{R}_+ \times \{0\}$ (jumps of size zero do not count; a zero jump just means the process did not jump!).

Example 2.27. Let $X = N$ be a Poisson process with intensity $\lambda > 0$. Since all jumps are of size 1, the jump measure μ is concentrated on $\mathbb{R}_+ \times \{1\}$. If instead X is a compound Poisson process[3] with jump distribution F, then μ is concentrated on $\mathbb{R}_+ \times \mathrm{supp}(F)$.

[3] Recall that X is a *compound Poisson process* if $X_t = \sum_{n=1}^{N_t} Y_n$, $t \ge 0$, where N is a Poisson process and $\{Y_n\}_{n \in \mathbb{N}}$ is an iid sequence independent of N. Thus X is in effect a Poisson process with random jump sizes.

Example 2.28. Using the jump measure one can conveniently express a number of quantities related to the jumps of X. For example, the number of jumps over an interval $[0,t]$ is given by $\mu([0,t]\times\mathbb{R}^d)$, while the sum of squared jump sizes is equal to $\int_{[0,t]\times\mathbb{R}}|\xi|^2\mu(ds,d\xi)$. These quantities may of course be infinite. As usual, the dependence on ω is often suppressed.

Remark 2.29. Although we do not use it here, Itô's formula for a general semimartingale X can be expressed via its jump measure μ. Without going into the precise meaning of the terms involved, let us mention that for any C^2 function f one has, in the scalar case $d=1$,

$$\begin{aligned} f(X_t) &= f(X_0) + \int_0^t f'(X_{s-})dX_s + \frac{1}{2}\int_0^t f''(X_{s-})d[X,X]_s \\ &\quad + \int_{[0,t]\times\mathbb{R}} \left(f(X_{s-}+\xi) - f(X_{s-}) - f'(X_{s-})\xi - \frac{1}{2}f''(X_{s-})\xi^2 \right) \mu^X(ds,d\xi). \end{aligned} \tag{2.1}$$

2.30. For some purposes the jump measure μ is too irregular. A smoother object is another random measure known as the *compensator of* μ. This is the random measure μ^p constructed in the next theorem. The following terminology is used:

- A *predictable function* is a map $F\colon \mathbb{R}_+\times\Omega\times\mathbb{R}^d\to\mathbb{R}$ that is $\mathcal{P}\otimes\mathcal{B}(\mathbb{R}^d)$-measurable. An example is $F(s,\omega,\xi)=f(X_{s-}(\omega)+\xi)$, which appears in (2.1) above ($\longrightarrow$ exercise).
- A *predictable random measure* is a collection of measures $\{\mu(dt,d\xi;\omega)\colon \omega\in\Omega\}$ on $\mathcal{B}(\mathbb{R}_+\times\mathbb{R}^d)$ such that $\int_{[0,t]\times\mathbb{R}^d}F(t,\xi)\mu(dt,d\xi)$, $t\geq 0$, is a predictable process for every predictable function F such that the process is well-defined.

Theorem 2.31. *Let μ be the jump measure of a càdlàg adapted process. There exists a unique (up to indistinguishability) predictable random measure μ^p, called the* compensator *of μ, such that*

$$\int_{[0,t]\times\mathbb{R}^d} F(s,\xi)\mu(ds,d\xi) - \int_{[0,t]\times\mathbb{R}^d} F(s,\xi)\mu^p(ds,d\xi), \quad t\geq 0,$$

is a local martingale for any predictable function F such that $\int_{[0,t]\times\mathbb{R}^d}|F(s,\xi)|\mu(ds,d\xi)$ is locally integrable.

2.32. Again we do not give the a proof, only an outline based on the Riesz–Markov–Kakutani theorem. A different and more general proof is given in Jacod and Shiryaev (2003, Theorem II.1.8).

(i) By means of a localization argument one reduces to the case $\mathbb{E}[\mu(\mathbb{R}_+ \times \mathbb{R}^d)] < \infty$, which means that the expected total number of jumps of X is finite.

(ii) For any $f \in C_c(\mathbb{R}_+ \times \mathbb{R}^d)$, define the process A^f by

$$A_t^f = \int_{[0,t]\times\mathbb{R}^d} f(s,\xi)\mu(ds,d\xi) = \sum_{s\leq t} f(s,\Delta X_s)\mathbf{1}_{\{\Delta X_s \neq 0\}}.$$

Due to (i) one has $A^f \in \mathrm{FV}_0$ and $\mathbb{E}[\mathrm{Var}(A)_\infty] \leq \|f\|_\infty \mathbb{E}[\mu(\mathbb{R}_+ \times \mathbb{R}^d)] < \infty$, so that A^f is of integrable variation. In particular, A_∞^f is well-defined and finite (up to nullsets).

(iii) Thanks to (ii) and Corollary 2.13, each A^f has a unique compensator B^f, which is nondecreasing if $f \geq 0$. Furthermore, since A^f is linear in f we have for $a, b \in \mathbb{R}$ and $f, g \in C_c(\mathbb{R}_+ \times \mathbb{R})$ that

$$A^{af+bg} - (aB^f + bB^g) = a(A^f - B^f) + b(A^g - B^g) = \text{local martingale}.$$

By uniqueness of compensators it follows that $aB^f + bB^g = B^{af+bg}$. Thus B^f is also linear in f.

(iv) By (iii), and ignoring issues with nullsets, the map $f \mapsto B_\infty^f(\omega)$ is a positive linear functional on the space $C_c(\mathbb{R}_+ \times \mathbb{R})$ for every ω. The Riesz–Markov–Kakutani theorem therefore gives, for every ω, a measure $\mu^p(dt, d\xi; \omega)$ such that

$$B_\infty^f(\omega) = \int_{\mathbb{R}_+\times\mathbb{R}} f(s,\xi)\mu^p(ds,d\xi;\omega).$$

(v) It remains to show that μ^p has the stated properties, and this is done via the monotone class theorem.

Exercise 2.33. Let μ be the jump measure associated with a Poisson process $X = N$ with intensity $\lambda > 0$. Show that the compensator of μ is given by $\mu^p(dt, d\xi) = \lambda\, \delta_{\{1\}}(d\xi)dt$. In particular μ^p is deterministic.

2.6 Special semimartingale characteristics

Definition 2.34. Consider a d-dimensional special semimartingale with canonical decomposition $X = X_0 + M + B$. Let $M = M^c + M^d$ be the decomposition of M in its continuous and purely discontinuous parts, and let $C = \langle M^c, M^c \rangle$ be the quadratic variation of M^c.[4] Let μ be the jump measure of X and μ^p its compensator. The triplet (B, C, μ^p) is called the *characteristics of* X.

Remark 2.35. Characteristics (B, C, μ^p) can also be defined for non-special semimartingales X, although we will not use this. As before, μ^p is the compensator of the jump measure of X. To obtain B and C, one fixes a truncation function, i.e. a bounded measurable function $h \colon \mathbb{R} \to \mathbb{R}$ with $h(\xi) = \xi$ in a neighborhood of the origin. One then defines $\hat{X}(h)_t = \sum_{s \le t} (\Delta X_s - h(\Delta X_s))$ and $X(h) = X - \hat{X}(h)$. Then $\hat{X}(h)$ is càdlàg, adapted, and of finite variation, hence a semimartingale. Thus $X(h)$ is also a semimartingale, and since its jumps are bounded it is a special semimartingale by Exercise 2.18(ii). One then lets B and C be first two characteristics of $X(h)$. One can show that B, but not C, depends on the choice of truncation function.

Lemma 2.36. *Let X be a special semimartingale with characteristics (B, C, μ^p). Then*

$$\Delta B_t = \int_{\mathbb{R}^d} \xi \mu^p(\{t\}, d\xi), \tag{2.2}$$

$$C_t - C_s \text{ is symmetric positive semidefinite for all } s \le t, \tag{2.3}$$

$$\int_{[0,t] \times \mathbb{R}} |\xi|^2 \wedge |\xi| \, \mu^p(ds, d\xi) \text{ is locally integrable and } \mu^p(\{t\} \times \mathbb{R}^d) \le 1. \tag{2.4}$$

Sketch of proof. We do not prove (2.2), as this requires additional tools. To see (2.3), use bilinearity of the quadratic covariation to get

$$\langle x_1 M^{c,1} + \cdots + x_d M^{c,d} \rangle = \sum_{i,j=1}^{d} x_i x_j \langle M^{c,i}, M^{c,j} \rangle = \sum_{i,j=1}^{d} x_i x_j C^{ij}$$

for any real $x_1, \ldots, x_d$. The left-hand side is nondecreasing, so positive semidefiniteness of the increments of C follows. Symmetry follows from symmetry of the quadratic covariations. Finally, consider (2.4). Since only one jump can occur at a time, we have

[4] Thus C is a predictable matrix-valued process whose (i, j)th component is $C^{ij} = \langle M^{c,i}, M^{c,j} \rangle$.

$\mu(\{t\} \times \mathbb{R}^d) \leq 1$, which can be shown to imply $\mu^p(\{t\} \times \mathbb{R}^d) \leq 1$. The integrability statement follows from the fact that $\sup_{s\leq \cdot} |X_s - X_0|$ is locally integrable (see Exercise 2.18) and that $\sum_{s\leq t}(\Delta X_s)^2 < \infty$ for all $t \geq 0$. The latter is an important result in stochastic integration, which we do not prove here. $\square$

Theorem 2.37. *Let X be a càdlàg adapted process, and let (B, C, μ^p) be a triplet with B, C predictable in FV_0 and μ^p a predictable random measure, satisfying (2.2)–(2.4). Then X is a special semimartingale with characteristics (B, C, μ^p) if and only if*

$$\begin{aligned} f(X_t) - f(X_0) - \int_0^t \nabla f(X_{s-})^\top dB_s - \frac{1}{2}\int_0^t \mathrm{Tr}(\nabla^2 f(X_{s-}) dC_s) \\ - \int_{[0,t]\times\mathbb{R}} \Big(f(X_{s-} + \xi) - f(X_{s-}) - \nabla f(X_{s-})^\top \xi \Big) \mu^p(ds, d\xi), \quad t \geq 0, \end{aligned}$$

is a local martingale for every bounded C^2 function $f\colon \mathbb{R}^d \to \mathbb{R}$.

2.7 Jump-diffusions and generators

2.38. We may now finally define jump-diffusions, which are semimartingales whose characteristics are of a particular form. We only consider the case of special semimartingales, although the non-special case can be treated as well using truncation functions; cf. Remark 2.35.

Definition 2.39. Let X be a d-dimensional special semimartingale. We say that X is a *(time-homogeneous) jump-diffusion* if its characteristics (B, C, μ^p) are of the form

$$B_t = \int_0^t b(X_s)\, ds, \quad C_t = \int_0^t a(X_s)\, ds, \quad \mu^p(dt, d\xi) = \nu(X_{t-}, d\xi)dt$$

for some measurable functions $b\colon \mathbb{R}^d \to \mathbb{R}^d$ and $a\colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$, and a kernel $\nu(x, d\xi)$ from $\mathbb{R}^d$ into $\mathbb{R}^d$ such that $a(x)$ is symmetric positive semidefinite, $\nu(x, \{0\}) = 0$, and $\int_{\mathbb{R}^d} |\xi|^2 \wedge |\xi| \nu(x, d\xi) < \infty$ for all $x \in \mathbb{R}^d$. We refer to (b, a, ν) as the *coefficients* of X.

Remark 2.40.

- The coefficients need not be uniquely determined. For instance, if X never visits a certain point $\overline{x} \in \mathbb{R}^d$, then one can modify $b(\overline{x})$, $a(\overline{x})$, and $\nu(\overline{x}, d\xi)$ and obtain another set of valid coefficients.

- In these notes, every jump-diffusion is a special semimartingale, even if we sometimes may neglect to say so explicitly. **Warning:** In most sources, jump-diffusions are *not* required to be special (though they are always semimartingales).

Example 2.41. Each of the following processes is a jump-diffusion.

(i) Brownian motion. The coefficients are constant and given by $b = 0$, $a = 1$, $\nu = 0$.

(ii) The Poisson process N with intensity $\lambda > 0$. Its canonical decomposition is $N_t = 0 + (N_t - \lambda t) + \lambda t$. Hence N is a jump-diffusion whose coefficients are constant and given by $b = \lambda$, $a = 0$, $\nu(x, d\xi) = \lambda\delta_1(d\xi)$; cf. Exercise 2.33.

(iii) The compound Poisson process $X_t = \sum_{n=1}^{N_t} Y_n$, if the jump sizes Y_n are integrable. Its canonical decomposition is then ($\longrightarrow$ exercise)

$$X_t = 0 + \left(\sum_{n=1}^{N_t} Y_n - \lambda\mathbb{E}[Y_1]t\right) + \lambda\mathbb{E}[Y_1]t.$$

Moreover, its jump measure has compensator $\mu^p(dt, d\xi) = \lambda F(d\xi)dt$, where $F(d\xi)$ is the law of Y_1 ($\longrightarrow$ exercise). Therefore X is a jump-diffusion with coefficients $a = 0$, $b = \lambda\mathbb{E}[Y_1]$, and $\nu(x, d\xi) = \lambda F(d\xi)$.

(iv) Solutions of SDEs of the form $dX_t = b(X_t)dt + \sigma(X_t)dW_t$. Indeed, the canonical decomposition is by construction given by

$$X_t = X_0 + \int_0^t \sigma(X_s)dW_s + \int_0^t b(X_s)ds,$$

which is a jump-diffusion with coefficients $b(x) = b(x)$ (!), $a(x) = \sigma(x)^2$, $\nu = 0$.

2.42. For measurable $A \subseteq \mathbb{R}^d$, the jump coefficient $\nu(x, A)$ can be interpreted as the *instantaneous arrival intensity of jumps whose sizes lie in A, given that $X_{t-} = x$.* Example 2.41(ii)–(iii) make this interpretation more concrete in the case of state-independent jumps, i.e., when $\nu(x, d\xi)$ does not depend on x.

Definition 2.43. Let X be a jump-diffusion with coefficients (b, a, ν). The *generator* of X (also called *extended generator* or *Dynkin operator*) is the operator $\mathcal{G}$ defined by

$$\mathcal{G}f(x) = b(x)^\top \nabla f(x) + \frac{1}{2}\operatorname{Tr}(a(x)\nabla^2 f(x)) + \int_{\mathbb{R}^d} \left(f(x+\xi) - f(x) - \xi^\top \nabla f(x)\right)\nu(x, d\xi) \tag{2.5}$$

for any C^2 function f such that the integral is well-defined.

The following property of the generator will be extremely useful for us.

Lemma 2.44. *Let X be a jump-diffusion with coefficients (b, a, ν) and generator $\mathcal{G}$. Then the process M^f given by*

$$M_t^f = f(X_t) - f(X_0) - \int_0^t \mathcal{G}f(X_s)ds \tag{2.6}$$

is well-defined and a local martingale for any C^2 function $f\colon \mathbb{R}^d \to \mathbb{R}$ such that

$$\int_0^t \int_{\mathbb{R}^d} \left| f(X_s + \xi) - f(X_s) - \xi^\top \nabla f(X_s) \right| \nu(X_s, d\xi)ds < \infty, \quad t \geq 0. \tag{2.7}$$

Proof. Let $X = X_0 + M + B$ be the canonical decomposition of X, and $M = M^c + M^d$ the decomposition of the local martingale part into continuous and purely discontinuous components. Itô's formula[5] for general semimartingales states that

$$\begin{aligned} f(X_t) = f(X_0) &+ \sum_{i=1}^d \int_0^t D_i f(X_{s-})dX_s + \frac{1}{2} \sum_{i,j=1}^d \int_0^t D_{ij}^2 f(X_{s-})d\langle M^{c,i}, M^{c,j}\rangle_s \\ &+ \sum_{0<s\leq t} \Big(f(X_s) - f(X_{s-}) - \Delta X_s^\top \nabla f(X_{s-}) \Big), \quad t \geq 0. \end{aligned}$$

Since X is a jump-diffusion with generator $\mathcal{G}$, the same thing can be written in different notation as

$$\begin{aligned} f(X_t) = f(X_0) &+ \int_0^t \mathcal{G}f(X_s)ds + \sum_{i=1}^d \int_0^t D_i f(X_{s-})dM_s^i \\ &+ \int_{[0,t]\times\mathbb{R}^d} F(s,\xi)\mu(ds, d\xi) - \int_{[0,t]\times\mathbb{R}^d} F(s,\xi)\mu^p(ds, d\xi), \quad t \geq 0, \end{aligned}$$

where we define the predictable function $F(s, \xi) = f(X_{s-} + \xi) - f(X_{s-}) - \xi^\top \nabla f(X_{s-})$. We must show that the last three terms together form a local martingale. Stochastic integration theory tells us that each integral with respect to M^i is a local martingale. Moreover, Theorem 2.31 gives that the last two terms form a local martingale, if we can

[5] See e.g. Jacod and Shiryaev (2003, Theorem I.4.57).

show that $\int_{[0,t]\times\mathbb{R}^d} |F(s,\xi)|\mu(ds,d\xi)$ is locally integrable. This however follows from the assumption (2.7) along with the following fact ($\longrightarrow$ exercise):

> For any predictable function $G \geq 0$, $\int_{[0,t]\times\mathbb{R}^d} G(s,\xi)\mu(ds,d\xi)$ is locally integrable if and only if $\int_{[0,t]\times\mathbb{R}^d} G(s,\xi)\mu^p(ds,d\xi)$ is locally integrable.

This completes the proof. □

Remark 2.45.

- If the function f in Lemma 2.44 were in addition bounded, the result would have followed immediately from the forward implication of Theorem 2.37. This does not really save us any work however, because the argument given above is exactly how the forward implication of Theorem 2.37 is proved (up to replacing $\int_0^t \mathcal{G}f(X_s)ds$ by the appropriate expression involving the characteristics (B, C, μ^p) of X).

- The reverse implication of Theorem 2.37 can be used to show the following: *If X is a càdlàg adapted process and $\mathcal{G}$ an operator of the form (2.5) for some coefficients (b, a, ν) as in Definition 2.39 such that (2.6) is well-defined (in the sense that $\int_0^t |\mathcal{G}f(X_s)|ds < \infty$ for all $t \geq 0$) and a local martingale for every bounded C^2 function $f\colon \mathbb{R}^d \to \mathbb{R}$, then X a special semimartingale and a jump-diffusion with coefficients (b, a, ν) and generator $\mathcal{G}$.*

- Due to Lemma 2.44, jump-diffusions are closely connected to *martingale problems* and *Markov processes*. Why is it intuitive that jump-diffusions are typically Markov processes?

Chapter 3

Polynomial and affine jump-diffusions

We are now in a position to define and study the central objects of these notes: polynomial jump-diffusions (PJDs) and affine jump-diffusions (AJDs). The basic ingredient is a jump-diffusion taking values in a given subset $E \subseteq \mathbb{R}^d$ called the *state space*. Typical examples of state spaces are the

- Euclidean space $E = \mathbb{R}^d$,
- nonnegative orthant $E = \mathbb{R}^d_+$,
- unit cube $E = [0, 1]^d$,
- unit simplex $E = \Delta^{d-1} = \{x \in [0, 1]^d \colon x_1 + \cdots + x_d = 1\}$,
- unit sphere $E = S^{d-1} = \{x \in \mathbb{R}^d \colon |x| = 1\}$,
- unit ball $E = B^d = \{x \in \mathbb{R}^d \colon |x| \le 1\}$,
- symmetric and positive semidefinite $m \times m$ matrices $E = \mathbb{S}^m_+$, viewed as a subset of $\mathbb{R}^d$ with $d = m(m + 1)/2$,

as well as cartesian products of such spaces.

Throughout this chapter we fix a state space $E \subseteq \mathbb{R}^d$ and an E-valued jump-diffusion X with coefficients (b, a, ν) and generator $\mathcal{G}$ given by (2.5). Roughly speaking, X is called *polynomial* if $\mathcal{G}$ maps any polynomial p of degree at most n to another polynomial $\mathcal{G}p$ of degree at most n, for every n. It is called *affine* if $\mathcal{G}$ maps the function $x \mapsto f_u(x) = e^{u^\top x}$ to a function $x \mapsto \mathcal{G}f_u(x) = R(u,x)e^{u^\top x}$, for every imaginary vector $u \in \mathrm{i}\mathbb{R}^d$, where $R(u,x)$ is affine in x. Some care is needed in making these definitions precise, due to subtleties in the notions of polynomial and degree.

3.1 PJDs: Definition and characterization

Definition 3.1.

- For a *multi-index* $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}_0^d$, we write $|\boldsymbol{\alpha}| = \alpha_1 + \cdots + \alpha_d$ and $x^{\boldsymbol{\alpha}} = x_1^{\alpha_1} \cdots x_d^{\alpha_d}$ for $x \in \mathbb{R}^d$.

- A *polynomial* (or more accurately, a *polynomial function*) *on* $\mathbb{R}^d$ is a function $p\colon \mathbb{R}^d \to \mathbb{R}$ of the form
$$p(x) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}_0^d} c_{\boldsymbol{\alpha}} x^{\boldsymbol{\alpha}},$$
where only finitely many of the constants $c_{\boldsymbol{\alpha}} \in \mathbb{R}$ are nonzero. Such a representation is unique. The *degree* of p, denoted $\deg(p)$, is the largest $|\boldsymbol{\alpha}|$ such that $c_{\boldsymbol{\alpha}} \neq 0$. If p is the zero polynomial, then $\deg(p) = -\infty$. We define
$$\mathrm{Pol}(\mathbb{R}^d) = \{\text{all polynomials on } \mathbb{R}^d\},$$
$$\mathrm{Pol}_n(\mathbb{R}^d) = \{p \in \mathrm{Pol}(\mathbb{R}^d)\colon\ \deg(p) \le n\}, \qquad n \in \mathbb{N}_0.$$

- A *polynomial on* E is the restriction $p = q|_E$ to E of some $q \in \mathrm{Pol}(\mathbb{R}^d)$. Its *degree* is $\deg(p) = \min\{\deg q\colon p = q|_E,\ q \in \mathrm{Pol}(\mathbb{R}^d)\}$. We define
$$\mathrm{Pol}(E) = \{\text{all polynomials on } E\},$$
$$\mathrm{Pol}_n(\mathbb{R}^d) = \{p \in \mathrm{Pol}(E)\colon\ \deg(p) \le n\}, \qquad n \in \mathbb{N}_0.$$
Although slightly inaccurate, it is very convenient to write $f \in \mathrm{Pol}_n(E)$ for any function $f = (f_1, \dots, f_k)\colon \mathbb{R}^d \to \mathbb{R}^k$ such that $f_i|_E \in \mathrm{Pol}_n(E)$ for $i = 1, \dots, k$.

- The operator $\mathcal{G}$ is said to be *well-defined on* $\mathrm{Pol}(E)$ if
$$\int_{\mathbb{R}^d} |\xi|^n \, \nu(x, d\xi) < \infty \text{ for all } x \in E \text{ and all } n \ge 2, \text{ and} \tag{3.1}$$
$$\mathcal{G}f = 0 \text{ on } E \text{ for any } f \in \mathrm{Pol}(\mathbb{R}^d) \text{ with } f = 0 \text{ on } E. \tag{3.2}$$

Exercise 3.2. When and why do we need to worry about the distinction between $\mathrm{Pol}(E)$ and $\mathrm{Pol}(\mathbb{R}^d)$? The following exercises illustrate the potential issues.

- Assume E contains a nonempty open set. Show that for every $p \in \mathrm{Pol}(E)$ there exists exactly one $q \in \mathrm{Pol}(\mathbb{R}^d)$ such that $p = q|_E$. In particular, $\deg(p) = \deg(q)$. Deduce that $\mathrm{Pol}_n(E)$ and $\mathrm{Pol}_n(\mathbb{R}^d)$ are isomorphic as vector spaces for each $n \in \mathbb{N}_0$.[1]

- Let $E = \{0\}$ be a singleton. Show that every $p \in \mathrm{Pol}(E)$ is either the zero polynomial or has degree zero, $\deg(p) = 0$. Show that $\mathrm{Pol}(E)$ is isomorphic to $\mathbb{R}$.

- Let $E = \mathcal{S}^1 \subset \mathbb{R}^2$ be the unit circle in the plane. What is the degree $\deg(p)$ of the polynomial $p \in \mathrm{Pol}(E)$ given by $p(x_1, x_2) = x_1^4 + x_2^4 + 2x_1^2 x_2^2$?

- Let $E = \mathbb{R} \times \{0\} \subset \mathbb{R}^2$ and define $\mathcal{G}f(x_1, x_2) = D_2 f(x_1, x_2) + \frac{1}{2} D_{11}^2 f(x_1, x_2)$. Is this operator well-defined on $\mathrm{Pol}(E)$?

- Find an example of a non-trivial second order differential operator $\mathcal{G}$ that is well-defined on $\mathrm{Pol}(E)$, where $E = \mathcal{S}^1 \subset \mathbb{R}^2$.

Exercise 3.3. As noted above, $\mathrm{Pol}_n(\mathbb{R}^d)$ is a vector space for each $n \in \mathbb{N}_0$. Show that its dimension is given by $\binom{n+d}{n}$. *Hint: As a basis for* $\mathrm{Pol}_n(\mathbb{R}^d)$*, take the monomials* $x^{\boldsymbol{\alpha}}$ *with* $|\boldsymbol{\alpha}| \le n$*. Writing this as* $1^{\alpha_0} x_1^{\alpha_1} \cdots x_d^{\alpha_d}$*, the question boils down to counting the number of multi-indices* $(\alpha_0, \ldots, \alpha_d) \in \mathbb{N}_0^{d+1}$ *with* $\alpha_0 + \cdots + \alpha_d = n$*.*

3.4. If the generator $\mathcal{G}$ is well-defined on $\mathrm{Pol}(E)$, it is possible to let it act on polynomials $p \in \mathrm{Pol}(E)$. Indeed, if $p = q|_E$ for some $q \in \mathrm{Pol}(\mathbb{R}^d)$, condition (3.1) ensures that $\mathcal{G}q(x)$ makes sense for all $x \in E$. Moreover, if $r \in \mathrm{Pol}(\mathbb{R}^d)$ is another representative of p in the sense that $p = r|_E$, then condition (3.2) ensures that $\mathcal{G}r(x) = \mathcal{G}q(x)$ for all $x \in E$. We may therefore define $\mathcal{G}p(x) = \mathcal{G}q(x)$ for all $x \in E$ without ambiguity.

Definition 3.5. The operator $\mathcal{G}$ is called *polynomial on* E if it is well-defined on $\mathrm{Pol}(E)$ and maps $\mathrm{Pol}_n(E)$ to itself for each $n \in \mathbb{N}$. In this case, we call X a *polynomial jump-diffusion on* E.

Proposition 3.6. *Assume* $\mathcal{G}$ *is well-defined on* $\mathrm{Pol}(E)$*. Then the following are equivalent:*

[1] Recall that two vector spaces V and W are isomorphic if there exists a linear bijection $T\colon V \to W$.

(i) *$\mathcal{G}$ is polynomial on E;*

(ii) *The coefficients (b, a, ν) satisfy*

$$b \in \mathrm{Pol}_1(E), \tag{3.3}$$

$$a + \int_{\mathbb{R}^d} \xi\xi^\top \nu(\,\cdot\,, d\xi) \in \mathrm{Pol}_2(E), \tag{3.4}$$

$$\int_{\mathbb{R}^d} \xi^{\boldsymbol{\alpha}} \nu(\,\cdot\,, d\xi) \in \mathrm{Pol}_{|\boldsymbol{\alpha}|}(E), \tag{3.5}$$

for all $|\boldsymbol{\alpha}| \geq 3$.

In this case, the polynomials on E in (3.3)–(3.5) are uniquely determined by the action of $\mathcal{G}$ on $\mathrm{Pol}(E)$. Moreover, a, b, and $\int_{\mathbb{R}^d} \xi^{\boldsymbol{\alpha}} \nu(\,\cdot\,, d\xi)$ are locally bounded on E for all $|\boldsymbol{\alpha}| \geq 2$.

Notice the notational shortcut: a more pedantic statement of (3.3) would be "$b_i|_E \in \mathrm{Pol}_1(E)$ for $i = 1 \ldots, d$", and similarly for (3.4)–(3.5). This would soon become tedious, which is why the simplified but slightly inaccurate notation is used. Nonetheless, the proof of Proposition 3.6 will take the pedantic approach to emphasize the subtleties involved. First, we need some basic facts about degrees of polynomials on E.

Lemma 3.7. *For any $p, q \in \mathrm{Pol}(E)$ we have $\deg(p + q) \leq \max\{\deg(p), \deg(q)\}$ and $\deg(pq) \leq \deg(p) + \deg(q)$.*

Proof. By definition there exist $r, s \in \mathrm{Pol}(\mathbb{R}^d)$ such that $p = r|_E$, $\deg(p) = \deg(r)$, and $q = s|_E$, $\deg(q) = \deg(s)$. We then get

$$\deg(p + q) \leq \deg(r + s) \leq \max\{\deg(r), \deg(s)\} = \max\{\deg(p), \deg(q)\},$$

where the first inequality uses the definition of $\deg(p + q)$, the second inequality is a basic fact about polynomials on $\mathbb{R}^d$, and the equality follows from the choice of r and s. We similarly get

$$\deg(pq) \leq \deg(rs) = \deg(r) + \deg(s) = \deg(p) + \deg(q),$$

as claimed. □

Proof of Proposition 3.6. It is useful to express $\mathcal{G}$, given in (2.5), in the equivalent form

$$\begin{aligned} \mathcal{G}f(x) &= b(x)^\top \nabla f(x) + \frac{1}{2} \operatorname{Tr}\left(\left(a(x) + \int_{\mathbb{R}^d} \xi\xi^\top \nu(x, d\xi) \right) \nabla^2 f(x) \right) \\ &\quad + \int_{\mathbb{R}^d} \left(f(x + \xi) - f(x) - \xi^\top \nabla f(x) - \frac{1}{2} \xi^\top \nabla^2 f(x) \xi \right) \nu(x, d\xi), \end{aligned} \tag{3.6}$$

for any $f \in \mathrm{Pol}(\mathbb{R}^d)$ and $x \in E$, which is permissible since $\mathcal{G}$ is well-defined on E and in particular satisfies (3.1).

We first prove (ii) $\Rightarrow$ (i). Fix any $n \in \mathbb{N}$ and $p \in \mathrm{Pol}_n(E)$, and choose $f \in \mathrm{Pol}_n(\mathbb{R}^d)$ such that $p = f|_E$. We need to show that $(\mathcal{G}f)|_E \in \mathrm{Pol}_n(E)$. Since polynomials lose degree when differentiated, we have $D_i f|_E \in \mathrm{Pol}_{n-1}(E)$ and $D^2_{ij} f|_E \in \mathrm{Pol}_{n-2}(E)$ for $i,j = 1,\dots,d$. Therefore, using (3.3)–(3.4) and Lemma 3.7, the map

$$E \to \mathbb{R}, \qquad x \mapsto b(x)^\top \nabla f(x) + \frac{1}{2}\operatorname{Tr}\left(\left(a(x) + \int_{\mathbb{R}^d} \xi\xi^\top \nu(x,d\xi)\right)\nabla^2 f(x)\right),$$

is in $\mathrm{Pol}_n(E)$. Next, note that f is a linear combination of monomials on $\mathbb{R}^d$ of the form $x^{\boldsymbol{\alpha}}$ with $|\boldsymbol{\alpha}| \le n$. The multi-binomial theorem[2] thus implies that the expression

$$f(x+\xi) - f(x) - \xi^\top \nabla f(x) - \frac{1}{2}\xi^\top \nabla^2 f(x)\xi$$

is a linear combination of terms of the form $x^{\boldsymbol{\alpha}}\xi^{\boldsymbol{\beta}}$ with $|\boldsymbol{\beta}| \ge 3$ and $|\boldsymbol{\alpha}| + |\boldsymbol{\beta}| \le n$. Due to (3.5) and Lemma 3.7, we have that the maps

$$E \to \mathbb{R}, \qquad x \mapsto x^{\boldsymbol{\alpha}} \int_{\mathbb{R}^d} \xi^{\boldsymbol{\beta}} \nu(x,d\xi)$$

lie in $\mathrm{Pol}_{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}(E) \subseteq \mathrm{Pol}_n(E)$. Summing up all the terms and using Lemma 3.7, we finally deduce that $(\mathcal{G}f)|_E \in \mathrm{Pol}_n(E)$, as required.

The reverse implication (i) $\Rightarrow$ (ii) is proved by applying $\mathcal{G}$ to monomials. Fix $i \in \{1,\dots,d\}$ and define $f \in \mathrm{Pol}_1(\mathbb{R}^d)$ by $f(x) = x_i$. Then $\mathcal{G}f = b_i$. On the other hand, since $f|_E \in \mathrm{Pol}_1(E)$, (i) gives $(\mathcal{G}f)|_E \in \mathrm{Pol}_1(E)$. We deduce (3.3). Similarly, with $f(x) = x_i x_j$, we have

$$\mathcal{G}f(x) = x_i b_j(x) + x_j b_i(x) + a_{ij}(x) + \int_{\mathbb{R}^d} \xi_i \xi_j \nu(x,d\xi),$$

using also that $a(x)$ is a symmetric matrix. As before this yields (3.4). Next, set $f(x) = x^{\boldsymbol{\gamma}}$ with $|\boldsymbol{\gamma}| \ge 3$. A slightly more careful use of the multi-binomial theorem gives

$$f(x+\xi) - f(x) - \xi^\top \nabla f(x) - \frac{1}{2}\xi^\top \nabla^2 f(x)\xi = \xi^{\boldsymbol{\gamma}} + \sum_{\substack{3 \le |\boldsymbol{\alpha}| < |\boldsymbol{\gamma}| \\ |\boldsymbol{\beta}|+|\boldsymbol{\alpha}| \le |\boldsymbol{\gamma}|}} c_{\boldsymbol{\beta},\boldsymbol{\alpha}}\, x^{\boldsymbol{\beta}} \xi^{\boldsymbol{\alpha}} \tag{3.7}$$

[2]The multi-binomial theorem states that $(x+\xi)^{\boldsymbol{\alpha}} = \sum_{\boldsymbol{\beta} \le \boldsymbol{\alpha}} \binom{\boldsymbol{\alpha}}{\boldsymbol{\beta}} x^{\boldsymbol{\beta}} \xi^{\boldsymbol{\alpha}-\boldsymbol{\beta}}$, where the multi-binomial coefficients are defined by $\binom{\boldsymbol{\alpha}}{\boldsymbol{\beta}} = \prod_{i=1}^d \binom{\alpha_i}{\beta_i}$.

for some coefficients $c_{\boldsymbol{\beta},\boldsymbol{\alpha}}$. Thus, if (3.5) has been proved for all $\boldsymbol{\alpha}$ with $3 \le |\boldsymbol{\alpha}| < |\boldsymbol{\gamma}|$, we rearrange (3.6) and use (3.3)–(3.4) and (3.7) to obtain

$$\int_{\mathbb{R}^d} \xi^{\boldsymbol{\gamma}} \nu(x, d\xi) = \mathcal{G}f(x) + p(x), \quad x \in E,$$

for some $p \in \mathrm{Pol}_{|\boldsymbol{\gamma}|}(E)$. Thus (3.5) holds also for $\boldsymbol{\alpha} = \boldsymbol{\gamma}$. The base case $|\boldsymbol{\alpha}| = 3$ is clear since (3.7) is then simply equal to $\xi^{\boldsymbol{\alpha}}$. It follows by induction that (3.5) holds for all $|\boldsymbol{\alpha}| \ge 3$, as claimed.

As a by-product of the above proof, we see that the polynomials on E listed in property (ii) are uniquely determined by the action of $\mathcal{G}$ on $\mathrm{Pol}(E)$. All that remains to show is that a and $\int_{\mathbb{R}^d} |\xi|^2 \nu(\,\cdot\,, d\xi)$ are locally bounded on E. But $a(x) \in \mathbb{S}^d_+$ for all $x \in E$, and hence has nonnegative diagonal entries. Since $a_{ii} + \int_{\mathbb{R}^d} \xi_i^2 \nu(\cdot, d\xi) \in \mathrm{Pol}_2(E)$ for $i = 1, \ldots, d$, it follows that both $\mathrm{Tr}(a)$ and $\int_{\mathbb{R}^d} |\xi|^2 \nu(\,\cdot\,, d\xi)$ are locally bounded on E. Then a is also locally bounded on E, due to the inequality $\|A\|_F^2 \le \mathrm{Tr}(A)^2$ for any $A \in \mathbb{S}^d_+$, where $\|\cdot\|_F$ denotes the Frobenius norm ($\longrightarrow$ exercise). $\square$

Example 3.8. Let $d = 1$ and consider the SDE

$$dX_t = (b + \beta X_t)dt + \sqrt{a + \alpha X_t + A X_t^2}\, dW_t$$

for some real parameters b, β, a, α, A. The state space E depends on the parameters, and must be such that $a + \alpha X + AX^2 \ge 0$. Any E-valued solution of this equation is a polynomial diffusion (i.e., a PJD without jumps) by Proposition 3.6. Conversely, one can show that every one-dimensional polynomial diffusion is the solution of an equation of this form. Up to affine transformations of X, there are three main cases:

(i) $E = \mathbb{R}$ with b, β unconstrained, $a > 0$, $\alpha = 0$, and $A \ge 0$. This covers in particular Brownian motion with drift ($\beta = A = 0$) as well as the Ornstein–Uhlenbeck process ($A = 0$).

(ii) $E = \mathbb{R}_+$ with $b \ge 0$, β unconstrained, $a = 0$, $\alpha \ge 0$, and $A \ge 0$. This covers in particular the square-root process ($a = A = 0$, also known as the Cox–Ingersoll–Ross, or CIR, process), geometric Brownian motion ($b = a = \alpha = 0$), and GARCH diffusion ($a = \alpha = 0$).

(iii) $E = [0, 1]$ with $b \ge 0$, $b + \beta \le 0$, $a = 0$, and $A = -\alpha$, so that $a + \alpha x + Ax^2 = \alpha x(1 - x)$. This is known as the Jacobi process.

3.2 The moment formula

3.9. Our next goal is to establish the *moment formula*, which describes how to calculate conditional expectations of the form $\mathbb{E}[p(X_T) \mid \mathcal{F}_t]$ where X is a PJD and p is a polynomial. This is the most important result about PJDs. The following example illustrates the basic idea. Suppose, for this example only, that X solves the SDE

$$dX_t = \beta X_t dt + \sqrt{X_t} dW_t, \quad X_0 = x_0.$$

This is the SDE we looked at in the introduction. By Itô's formula, the squared process satisfies $dX_t^2 = (X_t + 2\beta X_t^2)dt + 2X_t^{3/2} dW_t$. Writing this in integral form, we get

$$\begin{pmatrix} X_t \\ X_t^2 \end{pmatrix} = \begin{pmatrix} x_0 \\ x_0^2 \end{pmatrix} + \int_0^t \begin{pmatrix} \beta & 0 \\ 1 & 2\beta \end{pmatrix} \begin{pmatrix} X_s \\ X_s^2 \end{pmatrix} ds + \int_0^t \begin{pmatrix} \sqrt{X_s} \\ 2X_s^{3/2} \end{pmatrix} dW_s.$$

Now take expectations on both sides. Assuming that the local martingale term is a true martingale, and that we may interchange time integral and expectation, this gives

$$\begin{pmatrix} \mathbb{E}[X_t] \\ \mathbb{E}[X_t^2] \end{pmatrix} = \begin{pmatrix} x_0 \\ x_0^2 \end{pmatrix} + \int_0^t \begin{pmatrix} \beta & 0 \\ 1 & 2\beta \end{pmatrix} \begin{pmatrix} \mathbb{E}[X_s] \\ \mathbb{E}[X_s^2] \end{pmatrix} ds.$$

This is a *deterministic linear ODE*, whose solution can be written down in closed form,

$$\begin{pmatrix} \mathbb{E}[X_t] \\ \mathbb{E}[X_t^2] \end{pmatrix} = \exp\left(t \begin{pmatrix} \beta & 0 \\ 1 & 2\beta \end{pmatrix} \right) \begin{pmatrix} x_0 \\ x_0^2 \end{pmatrix}.$$

We have thus computed all expectations $\mathbb{E}[p(X_t)]$ for $p \in \mathrm{Pol}_2(\mathbb{R})$. Note that we ignored several points: (i) we only looked at a specific one-dimensional diffusion; (ii) we only considered deterministic initial conditions $X_0 = x_0$; (iii) only the unconditional moments were computed; (iv) we never verified the true martingale property; (v) we only considered moments up to order two.

3.10. We now return to the general situation, where we fix a state space $E \subseteq \mathbb{R}^d$ and an E-valued jump-diffusion X with coefficients (b, a, ν) and generator $\mathcal{G}$ given by (2.5). Assume that $\mathcal{G}$ is polynomial on E, so that X is a PJD on E. Fix $n \in \mathbb{N}$ and set $N = \dim \mathrm{Pol}_n(E) - 1$. By Exercise 3.3 we have $1 + N = \binom{n+d}{d}$ if $E = \mathbb{R}^d$, but N may be smaller in general. Choose N polynomials

$$h_1, \ldots, h_N \in \mathrm{Pol}_n(\mathbb{R}^d)$$

such that

$$1, h_1|_E, \dots, h_N|_E \text{ form a basis for } \mathrm{Pol}_n(E). \tag{3.8}$$

Define the (row) vector valued function

$$H : \mathbb{R}^d \to \mathbb{R}^N, \quad H(x) = (h_1(x), \dots, h_N(x)). \tag{3.9}$$

Every $p \in \mathrm{Pol}_n(E)$ has a coordinate representation with respect to the basis in (3.8), and we denote its coordinate (column) vector by $\vec{p} \in \mathbb{R}^{1+N}$. Thus

$$p(x) = (1, H(x))\vec{p}, \quad x \in E. \tag{3.10}$$

Since $\mathcal{G}$ is polynomial, it maps $\mathrm{Pol}_n(E)$ linearly to itself. This linear map has a matrix representation $G \in \mathbb{R}^{(1+N)\times(1+N)}$ with respect to the basis in (3.8), determined by

$$\mathcal{G}(1, H)(x) = (1, H(x))G, \quad x \in E,$$

where $\mathcal{G}$ is understood to act componentwise on the function $(1, H)$. In particular, we have

$$\mathcal{G}p(x) = (1, H(x))G\vec{p}, \quad x \in E. \tag{3.11}$$

3.11. The moment formula uses the notion of *generalized conditional expectation*, which is defined for any σ-field $\mathcal{F}' \subseteq \mathcal{F}$ and all random variables Y (not just the integrable or nonnegative ones) by

$$\mathbb{E}[Y \mid \mathcal{F}'] = \begin{cases} \mathbb{E}[Y^+ \mid \mathcal{F}'] - \mathbb{E}[Y^- \mid \mathcal{F}'], & \text{on } \{\mathbb{E}[|Y| \mid \mathcal{F}'] < \infty\}, \\ +\infty, & \text{elsewhere.} \end{cases}$$

In particular, $\mathbb{E}[Y \mid \sigma(Y)] = Y$ for any random variable Y.

Theorem 3.12. *Assume $\mathcal{G}$ is polynomial on E. Then for any $p \in \mathrm{Pol}_n(E)$ with coordinate vector $\vec{p} \in \mathbb{R}^{1+N}$, the moment formula holds,*

$$\mathbb{E}[p(X_T) \mid \mathcal{F}_t] = (1, H(X_t))e^{(T-t)G}\,\vec{p}, \quad \textit{for } t \leq T.$$

3.13. According to Theorem 3.12, the $\mathcal{F}_t$-conditional expectation of $p(X_T)$ is a polynomial $q(X_t)$ in X_t with coefficient vector $\vec{q} = e^{(T-t)G}\,\vec{p}$. This vector can be computed numerically either by directly evaluating the matrix exponential, or by solving the linear ODE in $\mathbb{R}^{1+N}$,

$$f' = Gf, \quad f(0) = \vec{p},$$

and setting $\vec{q} = f(T-t)$.

Example 3.14. Theorem 3.12 implies that X_T has finite $\mathcal{F}_t$-conditional moments of all orders. This does not imply that X_T has finite unconditional moments. For example, it can be shown that the GARCH diffusion $dX_t = \kappa(\theta - X_t)\,dt + \sqrt{2\kappa}X_t\,dW_t$ with $\kappa, \theta > 0$ has a unique stationary solution, which is a polynomial diffusion on $E = \mathbb{R}_+$. The distribution of X_T is an inverse Gamma distribution with shape parameter 2 and scale parameter $1/\theta$. We then have $\mathbb{E}[X_T] = \theta$ and $\mathbb{E}[X_T^2] = +\infty$.[3] This example motivates the use of generalized conditional expectations.

Example 3.15. Consider the one-dimensional polynomial diffusion from Example 3.8,

$$dX_t = (b + \beta X_t)dt + \sqrt{a + \alpha X_t + A X_t^2}\,dW_t$$

for some real parameters b, β, a, α, A. Its generator is

$$\mathcal{G}f(x) = (b + \beta x)f'(x) + \frac{1}{2}(a + \alpha x + Ax^2)f''(x).$$

Except in trivial cases, E contains a nonempty open interval and we may work with the monomial basis $1, x, \dots, x^n$. The columns of the matrix G are then the coordinate vectors of $\mathcal{G}(x^k)$ for $k = 0, \dots, n$. Here $\mathcal{G}(x^k)$ is shorthand for $\mathcal{G}p$ with $p(x) = x^k$. Since

$$\mathcal{G}(x^k) = k(k-1)\frac{a}{2}x^{k-2} + k\left(b + (k-1)\frac{\alpha}{2}\right)x^{k-1} + k\left(\beta + (k-1)\frac{A}{2}\right)x^k$$

for $k \geq 0$, we infer that

$$G = \begin{pmatrix} 0 & b & 2\frac{a}{2} & 0 & \cdots & 0 \\ 0 & \beta & 2\left(b + \frac{\alpha}{2}\right) & 3 \cdot 2\frac{a}{2} & 0 & \vdots \\ 0 & 0 & 2\left(\beta + \frac{A}{2}\right) & 3\left(b + 2\frac{\alpha}{2}\right) & \ddots & 0 \\ 0 & 0 & 0 & 3\left(\beta + 2\frac{A}{2}\right) & \ddots & n(n-1)\frac{a}{2} \\ \vdots & & & 0 & \ddots & n\left(b + (n-1)\frac{\alpha}{2}\right) \\ 0 & & \cdots & & 0 & n\left(\beta + (n-1)\frac{A}{2}\right) \end{pmatrix}.$$

This illustrates how the matrix G can be computed.

[3] See Forman and Sørensen (2008, Case 4) for further details.

The rest of this section is devoted to the proof of Theorem 3.12, which in principle is straightforward but requires some technical developments. For the rest of this section, we assume that $\mathcal{G}$ is polynomial on E. Then the process

$$M_t^f = f(X_t) - f(X_0) - \int_0^t \mathcal{G}f(X_s)\,ds, \quad t \geq 0,$$

is well defined for any $f \in \mathrm{Pol}(E)$.

Lemma 3.16. *M_t^f is a local martingale for any $f \in \mathrm{Pol}(E)$.*

Proof. Note that $W(x,\xi) = f(x+\xi) - f(x) - \xi^\top \nabla f(x)$ is a linear combination of monomials $x^{\boldsymbol{\beta}}\xi^{\boldsymbol{\gamma}}$ with $2 \leq |\boldsymbol{\gamma}| \leq n$. Hence $|W(x,\xi)| \leq C(x)\left(|\xi|^2 + |\xi|^{2n}\right)$ for some polynomial $C(x)$. It follows that $\int_0^t \int_{\mathbb{R}^d} |W(X_s,\xi)|\nu(X_s,d\xi)ds < \infty$ thanks to Proposition 3.6. Lemma 2.44 then gives the result. $\square$

The quadratic variation of a semimartingale Y is a nondecreasing process given by $[Y,Y] = Y^2 - Y_0^2 - 2\int Y_- dY$. If $Y = M \in \mathcal{M}_{0,\mathrm{loc}}$ then $M^2 - [M,M] = 2\int M_- dM \in \mathcal{M}_{0,\mathrm{loc}}$ as well. This is used in the proof of the following lemma, which will let us work with the compensator of $[M,M]$ rather than $[M,M]$ itself.

Lemma 3.17. *If $M \in \mathcal{M}_{0,\mathrm{loc}}$, $A \in \mathrm{FV}_0$ is nondecreasing, and $M^2 - A \in \mathcal{M}_{0,\mathrm{loc}}$, then*

$$\mathbb{E}[\sup_{t \leq T} M_t^2] \leq 4\,\mathbb{E}[A_T], \quad T \geq 0.$$

Proof. Fix any $T \geq 0$. Let (τ_m) be a localizing sequence for M and $M^2 - [M,M]$. Doob's inequality gives

$$\mathbb{E}[\sup_{t \leq T\wedge\tau_m} M_t^2] \leq 4\,\mathbb{E}[M_{T\wedge\tau_m}^2] = 4\,\mathbb{E}[[M,M]_{T\wedge\tau_m}].$$

Sending m to infinity and using the monotone convergence theorem we get

$$\mathbb{E}[\sup_{t \leq T} M_t^2] \leq 4\,\mathbb{E}[[M,M]_T].$$

Next, we have $[M,M] - A = (M^2 - A) - (M^2 - [M,M]) \in \mathcal{M}_{\mathrm{loc}}$. Let (τ_m) be a localizing sequence. Then $\mathbb{E}[[M,M]_{T\wedge\tau_m}] = \mathbb{E}[A_{T\wedge\tau_n}]$. $\square$

We next identify a suitable such process A for the local martingale M^f.

Lemma 3.18. *Let $f \in \mathrm{Pol}(E)$ and define $\Gamma(f,f) = \mathcal{G}(f^2) - 2f\mathcal{G}f$. Then $\Gamma(f,f)(x) \geq 0$ for all $x \in E$, and*

$$(M^f)^2 - \int_0^\cdot \Gamma(f,f)(X_s)ds \in \mathcal{M}_{0,\mathrm{loc}}.$$

Proof. A calculation gives $\Gamma(f,f)(x) = \nabla f(x)^\top a(x)\nabla f(x) + \int_{\mathbb{R}^d}(f(x+\xi)-f(x))^2\nu(x,d\xi)$, which is nonnegative for all $x \in E$. Next, for simplicity we suppose that $f(X_0) = 0$, so that

$$M_t^f = f(X_t) - \int_0^t \mathcal{G}f(X_s)ds \quad \text{and} \quad M_t^{f^2} = f(X_t)^2 - \int_0^t \mathcal{G}(f^2)(X_s)ds. \tag{3.12}$$

The general case is similar ($\longrightarrow$ exercise). We will also use the identities

$$\left(\int_0^t H_s ds\right)^2 = 2\int_0^t H_s \int_0^s H_r dr\, ds \tag{3.13}$$

and

$$M_t^f \int_0^t H_s ds = \int_0^t M_s^f H_s ds + \int_0^t \int_0^s H_r dr\, dM_s^f, \tag{3.14}$$

valid for any adapted process H with integrable trajectories. We get

$$\begin{aligned}
(M_t^f)^2 &= f(X_t)^2 + \left(\int_0^t \mathcal{G}f(X_s)ds\right)^2 - 2f(X_t)\int_0^t \mathcal{G}f(X_s)ds \\
&= f(X_t)^2 - \left(\int_0^t \mathcal{G}f(X_s)ds\right)^2 - 2M_t^f \int_0^t \mathcal{G}f(X_s)ds && \text{by (3.12)} \\
&= f(X_t)^2 - 2\int_0^t \mathcal{G}f(X_s)\Big(\int_0^s \mathcal{G}f(X_r)dr + M_s^f\Big)ds + (\text{loc. mg.}) && \text{by (3.13)–(3.14)} \\
&= \int_0^t \big(\mathcal{G}(f^2)(X_s) - 2f(X_s)\mathcal{G}f(X_s)\big)\, ds + (\text{loc. mg.}) && \text{by (3.12)}
\end{aligned}$$

In view of the definition of $\Gamma(f,f)$, this proves the lemma. □

Remark 3.19. Define $\Gamma(f,g) = \mathcal{G}(fg) - f\mathcal{G}g - g\mathcal{G}f$. Then $[M^f, M^g] - \int_0^\cdot \Gamma(f,g)(X_s)ds$ is in $\mathcal{M}_{0,\mathrm{loc}}$ ($\longrightarrow$ exercise). The bilinear operator Γ is called the *carré-du-champ* (or *square-field*) operator of $\mathcal{G}$.

The next lemma uses the generalized conditional expectation. In particular, it is not assumed that $\mathbb{E}[|X_0|] < \infty$.

Lemma 3.20. *For any $k \in \mathbb{N}$ there exists a constant $C \in \mathbb{R}_+$ such that*

$$\mathbb{E}[1 + |X_t|^{2k} \mid \mathcal{F}_0] \leq (1 + |X_0|^{2k})\, e^{Ct}, \quad t \geq 0.$$

Proof. Set $f(x) = 1 + |x|^{2k}$. Then $\mathcal{G}f \in \mathrm{Pol}_{2k}(E)$, so there exists a constant $C \in \mathbb{R}_+$ such that $|\mathcal{G}f(x)| \leq Cf(x)$ for all $x \in E$. Let (τ_m) be a localizing sequence for M^f such that $|X_t| \leq m$ for all $t < \tau_m$ (note the strict inequality!). Then

$$\begin{aligned}
\mathbb{E}[f(X_{t\wedge\tau_m}) \mid \mathcal{F}_0] &= f(X_0) + \mathbb{E}\left[\int_0^{t\wedge\tau_m} \mathcal{G}f(X_s)ds \mid \mathcal{F}_0\right] \\
&\leq f(X_0) + C\int_0^t \mathbb{E}\left[f(X_s)\mathbf{1}_{\{s<\tau_m\}} \mid \mathcal{F}_0\right] ds \\
&\leq f(X_0) + C\int_0^t \mathbb{E}\left[f(X_{s\wedge\tau_m}) \mid \mathcal{F}_0\right] ds.
\end{aligned}$$

By Gronwall's lemma, $\mathbb{E}[f(X_{t\wedge\tau_m}) \mid \mathcal{F}_0] \leq f(X_0)e^{Ct}$.[4] Sending m to infinity and using Fatou's lemma,

$$\mathbb{E}[f(X_t) \mid \mathcal{F}_0] \leq \liminf_{m\to\infty} \mathbb{E}[f(X_{t\wedge\tau_m}) \mid \mathcal{F}_0] \leq f(X_0)e^{Ct},$$

as desired. □

Proposition 3.21. *For any $c \in \mathbb{R}_+$ and $f \in \mathrm{Pol}(E)$, the process $N_t = M^f_t \mathbf{1}_{\{|X_0|\leq c\}}$, $t \geq 0$, is a true martingale.*

Proof. Since $M^f \in \mathcal{M}_{0,\mathrm{loc}}$ by Lemma 3.16 and X_0 is $\mathcal{F}_0$-measurable, $N \in \mathcal{M}_{0,\mathrm{loc}}$ as well. Moreover, since $(M^f)^2 - \int_0^\cdot \Gamma(f,f)(X_t)dt \in \mathcal{M}_{0,\mathrm{loc}}$ by Lemma 3.18, we have

$$N^2 - \int_0^\cdot \Gamma(f,f)(X_t)\mathbf{1}_{\{|X_0|\leq c\}}dt \in \mathcal{M}_{0,\mathrm{loc}}.$$

Thus, by Lemma 3.17,

$$\mathbb{E}[\sup_{t\leq T} N_t^2] \leq 4\int_0^T \mathbb{E}[\Gamma(f,f)(X_t)\mathbf{1}_{\{|X_0|\leq c\}}]dt.$$

[4] We use the following form of Gronwall's lemma: let $g\colon \mathbb{R}_+ \to \mathbb{R}$ be continuous and let $\kappa_1, \kappa_2 \in \mathbb{R}_+$. If $g(t) \leq \kappa_1 + \kappa_2 \int_0^t g(s)ds$ for all $t \geq 0$, then $g(t) \leq \kappa_1 e^{\kappa_2 t}$ for all $t \geq 0$.

Since $\mathcal{G}$ is polynomial on E, $\Gamma(f,f)(x) \leq \lambda(1+|x|^{2k})$ for some $\lambda \in \mathbb{R}_+$, $k \in \mathbb{N}$, and all $x \in E$. Thus, by Lemma 3.20, there is a constant $C \in \mathbb{R}_+$ such that

$$\mathbb{E}[\sup_{t\leq T} N_t^2] \leq 4\lambda \int_0^T \mathbb{E}[(1+|X_0|^{2k})\mathbf{1}_{\{|X_0|\leq c\}}]e^{Ct}dt \leq 4\lambda(1+c^{2k})Te^{CT} < \infty.$$

This implies that N is a true martingale ($\longrightarrow$ exercise). $\square$

Proof of Theorem 3.12. Fix $c, t \in \mathbb{R}_+$ and define the row vector valued $\mathcal{F}_t$-measurable function $F(T) = \mathbb{E}[(1,\ H(X_T))\mathbf{1}_{\{|X_0|\leq c\}} \mid \mathcal{F}_t]$, $T \geq 0$. Proposition 3.21 implies that the process M given by

$$M_T = (1,\ H(X_T))\mathbf{1}_{\{|X_0|\leq c\}} - (1,\ H(X_0))\mathbf{1}_{\{|X_0|\leq c\}} - \int_0^T \mathcal{G}(1,\ H)(X_s)\mathbf{1}_{\{|X_0|\leq c\}}ds, \quad T \geq 0,$$

is an $(1+N)$-dimensional row vector valued martingale. Therefore, using also the identity $\mathcal{G}(1,\ H)(x) = (1,\ H(x))G$ for $x \in E$, we get for all $T \geq t$,

$$\begin{aligned}
0 &= \mathbb{E}[M_T - M_t \mid \mathcal{F}_t] \\
&= F(T) - F(t) - \mathbb{E}\left[\int_t^T \mathcal{G}(1,\ H)(X_s)\mathbf{1}_{\{|X_0|\leq c\}}ds \,\middle|\, \mathcal{F}_t\right] \\
&= F(T) - F(t) - \int_t^T \mathbb{E}\left[(1,\ H)(X_s)\mathbf{1}_{\{|X_0|\leq c\}} \mid \mathcal{F}_t\right] Gds \\
&= F(T) - F(t) - \int_t^T F(s)Gds.
\end{aligned}$$

This linear $(1+N)$-dimensional ODE has the unique solution

$$F(T) = F(t)e^{(T-t)G}, \quad T \geq t.$$

Post-multiplying by $\vec{p}$ gives

$$\mathbb{E}[p(X_T) \mid \mathcal{F}_t]\mathbf{1}_{\{|X_0|\leq c\}} = (1,\ H(X_T))e^{(T-t)G}\vec{p}\,\mathbf{1}_{\{|X_0|\leq c\}},$$

from which the result follows on sending c to infinity. $\square$

3.3 AJDs: Definition and characterization

We now turn to affine jump-diffusions (AJDs). Recall that we have fixed a state space $E \subseteq \mathbb{R}^d$ and an E-valued jump-diffusion X with coefficients (b, a, ν) and generator $\mathcal{G}$ given by (2.5).

Definition 3.22. The operator $\mathcal{G}$ is called *affine on E* if there exist functions $R_0, \ldots, R_d$ from $\mathrm{i}\mathbb{R}^d$ to $\mathbb{C}$ such that

$$\mathcal{G}\mathrm{e}^{u^\top x} = \left(R_0(u) + \sum_{i=1}^d R_i(u) x_i \right) \mathrm{e}^{u^\top x} \tag{3.15}$$

holds for all $x \in E$ and $u \in \mathrm{i}\mathbb{R}^d$. In this case, we call X an *affine jump-diffusion on E*.

The following result is analogous to Proposition 3.6.

Proposition 3.23. *The following are equivalent:*

(i) *$\mathcal{G}$ is affine on E;*

(ii) *The coefficients (b, a, ν) are affine of the form*

$$b(x) = b_0 + x_1 b_1 + \cdots + x_d b_d, \tag{3.16}$$

$$a(x) = a_0 + x_1 a_1 + \cdots + x_d a_d, \tag{3.17}$$

$$\nu(x, \cdot\,) = \nu_0 + x_1 \nu_1 + \cdots + x_d \nu_d \tag{3.18}$$

for all $x \in E$, for some matrices $a_i \in \mathbb{S}^d$, vectors $b_i \in \mathbb{R}^d$, and signed measures ν_i on $\mathbb{R}^d$ such that $\nu_i(\{0\}) = 0$ and $\int_{\mathbb{R}^d} |\xi| \wedge |\xi|^2 \, |\nu_i|(d\xi) < \infty$, $i = 0, \ldots, d$.

In this case, the functions $R_0, \ldots, R_d$ in (3.15) can be taken to be given by

$$R_i(u) = b_i^\top u + \frac{1}{2} u^\top a_i u + \int_{\mathbb{R}^d} \left(e^{u^\top \xi} - 1 - u^\top \xi \right) \nu_i(d\xi). \tag{3.19}$$

Proof. We first assume that $0 \in E$ and the affine span of E is all of $\mathbb{R}^d$.[5]

We first prove (i) $\Rightarrow$ (ii) and assume $\mathcal{G}$ is affine on E. Observe that

$$\mathcal{G}\mathrm{e}^{u^\top x} = \left(b(x)^\top u + \frac{1}{2} u^\top a(x) u + \int_{\mathbb{R}^d} \left(e^{u^\top \xi} - 1 - u^\top \xi \right) \nu(x, d\xi) \right) \mathrm{e}^{u^\top x}$$

so that, by virtue of the assumed relation (3.15), we obtain

$$\begin{aligned} R_0(u) + \sum_{i=1}^d R_i(u) x_i &= b(x)^\top u + \frac{1}{2} u^\top a(x) u \\ &\quad + \int_{\mathbb{R}^d} \left(e^{u^\top \xi} - 1 - u^\top \xi \right) \nu(x, d\xi) \quad \text{for all } x \in E,\, u \in \mathrm{i}\mathbb{R}^d. \end{aligned} \tag{3.20}$$

[5] The *affine span* of a set $V \subseteq \mathbb{R}^d$ is the set of points $\lambda_1 x^1 + \cdots + \lambda_d x^d$ with $\lambda_1 + \cdots + \lambda_d = 1$ and $x^1, \ldots, x^d \in V$.

We claim that R_i is of the form (3.19) for all i. Since $0 \in E$, this clear for $R_0(u)$, setting $a_0 = a(0)$, $b_0 = b(0)$, $\nu_0(d\xi) = \nu(0, d\xi)$. Next, since the affine span of E is all of $\mathbb{R}^d$, there exist numbers $\lambda_1, \ldots, \lambda_d$ with $\sum_{k=1}^d \lambda_k = 1$ and points $x^1, \ldots, x^d \in E$ such that $\lambda_1 x^1 + \cdots + \lambda_d x^d = e_1$, the first canonical unit vector. Evaluating both sides of (3.20) at $x = x^k$, multiplying by λ_k, summing over k, and using the form of $R_0(u)$, it follows that $R_1(u)$ is of the form (3.19) with

$$a_1 = \sum_{k=1}^d \lambda_k a(x^k) - a_0, \qquad b_1 = \sum_{k=1}^d \lambda_k b(x^k) - b_0, \qquad \nu_1(d\xi) = \sum_{k=1}^d \lambda_k \nu(x^k, d\xi) - \nu_0(d\xi).$$

The same argument shows that $R_2, \ldots, R_d$ are also of the form (3.19).

We must still prove (3.16)–(3.18). Given the R_i just obtained, it is clear that taking (b, a, ν) as in (3.16)–(3.18) is consistent with (3.20). Furthermore, for each fixed $x \in E$, knowing the right-hand side of (3.20) for all $u \in \mathrm{i}\mathbb{R}^d$ uniquely determines $a(x)$, $b(x)$, $\nu(x, d\xi)$; see Jacod and Shiryaev (2003, Lemma II.2.44). Thus (3.16)–(3.18) is in fact the only possibility. This completes the proof of (i) $\Rightarrow$ (ii).

To prove (ii) $\Rightarrow$ (i), assume (b, a, ν) are given by (3.16)–(3.18). A calculation then shows that $\mathcal{G}$ satisfies (3.15) with the R_i given by (3.19). Thus $\mathcal{G}$ is affine on E, as claimed.

Finally, in the general case, where either $0 \notin E$ or the affine span of E is not $\mathbb{R}^d$, we apply an invertible affine transformation $T \colon \mathbb{R}^d \to \mathbb{R}^d$ such that $0 \in T(E)$ and the affine span of $T(E)$ is $\mathbb{R}^{d'} \times \{0\}$ for some $d' \leq d$. In these new coordinates we set the corresponding a_i, b_i, and $\nu_i(d\xi)$ to zero for $i > d'$, and then transform back by T^{-1}. □

Corollary 3.24. *If X is an affine jump-diffusion on E and $\mathcal{G}$ is well-defined on* Pol(E)*, then X is a polynomial jump-diffusion on E.*

Proof. This follows directly from Proposition 3.6 and Proposition 3.23. □

Example 3.25. The CIR process X with values in $E = \mathbb{R}_+$ is given by

$$dX_t = (b + \beta X_t)dt + \sigma\sqrt{X_t}dW_t$$

for some parameters $b \geq 0$, $\beta \in \mathbb{R}$, $\sigma \geq 0$. Often the drift part is written $\kappa(\theta - X_t)$ instead of $b + \beta X_t$ where $b = \kappa\theta$ and $\kappa = -\beta$ (for $b \neq 0$ this is only possible if $\beta \neq 0$). This process is affine on E, and in the notation of Proposition 3.23 we have

$$b_0 = b,\ b_1 = \beta,\ a_0 = 0,\ a_1 = \sigma^2,\ \nu = 0.$$

In mathematical finance, the CIR process is used as a model for the short-term interest rate; this model was introduced by Cox et al. (1985) and is known as the *CIR short rate model*. The CIR process is also used as a component of the *Heston stochastic volatility model*, introduced by Heston (1993). This model consists of the two-dimensional process $X = (Y, V)$ with values in $E = \mathbb{R} \times \mathbb{R}_+$ given by

$$\begin{aligned} dY_t &= -\frac{1}{2} V_t dt + \sqrt{V_t} dB_t \\ dV_t &= (b + \beta V_t) dt + \sigma \sqrt{V_t} d\widetilde{B}_t, \end{aligned}$$

where $(B, \widetilde{B})$ is a correlated two-dimensional Brownian motion with correlation $\rho \in [0, 1]$.[6] The process X is affine on E ($\longrightarrow$ exercise). In finance, Y_t represents the logarithm of the time t price of an asset, and V_t represents its squared volatility. The Heston model is one of the earliest and most popular stochastic volatility models.

3.4 The affine transform formula

Affine jump-diffusions on E not only satisfy the moment formula in Theorem (3.12), subject to the generator being well-defined on $\mathrm{Pol}(E)$. Their characteristic functions are also analytically tractable.

Theorem 3.26. *Assume X is an affine jump-diffusion on E. For $u \in \mathrm{i}\mathbb{R}^d$ and $T > 0$ let $\phi\colon [0, T] \to \mathbb{C}$ and $\psi = (\psi_1, \dots, \psi_d)\colon [0, T] \to \mathbb{C}^d$ be functions that solve the generalized Riccati equations*

$$\begin{aligned} \phi'(\tau) &= R_0(\psi(\tau)), \qquad \phi(0) = 0, \\ \psi_i'(\tau) &= R_i(\psi(\tau)), \qquad \psi_i(0) = u_i, \quad i = 0, \dots, d, \end{aligned} \tag{3.21}$$

for $\tau \in [0, T]$, where $R_i(u)$ are the functions in (3.19). If

$$\operatorname{Re} \phi(\tau) + \operatorname{Re} \psi(\tau)^\top x \le 0 \quad \textit{for all } (\tau, x) \in [0, T] \times E, \tag{3.22}$$

then the affine transform formula holds,

$$\mathbb{E}[\mathrm{e}^{u^\top X_T} \mid \mathcal{F}_t] = \mathrm{e}^{\psi_0(T-t) + \psi(T-t)^\top X_t}, \quad t \le T.$$

[6] This means that B and $\widetilde{B}$ are scalar Brownian motions with $\langle B, \widetilde{B} \rangle_t \equiv \rho t$. One can then find a bivariate standard Brownian motion $W = (W^1, W^2)$ such that $B = W^1$ and $\widetilde{B} = \rho W^1 + \sqrt{1 - \rho^2} W^2$ ($\longrightarrow$ exercise).

Proof. Define $F(t,x) = \exp(\phi(T-t)+\psi(T-t)^\top x)$ and consider the complex-valued process $M = (M_t)_{t\in[0,T]}$ given by $M_t = F(t, X_t)$. A calculation using (3.21) yields

$$\partial_t F(t,x) + \mathcal{G}F(t,x) = 0, \qquad 0 \le t \le T, \quad x \in E,$$

where $\mathcal{G}$ acts on the real and imaginary parts of $F(t,\cdot)$ separately. Therefore, since the process $Y_t := (t, X_t)$ is a jump-diffusion with generator $\partial_t + \mathcal{G}$, we may apply Lemma 2.44 (with X replaced by Y and $f(x)$ replaced by $F(t,x)$) to see that M is a local martingale. Note that the condition (2.7) needs to be verified, which can be done as follows. Let $c := 1/\max_{\tau\in[0,T]} |\psi(\tau)|$, where we suppose the denominator is non-zero (otherwise the situation is rather trivial $\longrightarrow$ exercise). Whenever $|\xi| \le c$, we have $|\psi(T-t)^\top \xi| \le 1$, and thus the general inequality

$$|e^z - 1 - z| \le \frac{3}{2}|z|^2 \quad \text{for all } z \in \mathbb{C} \text{ with } |z| \le 1$$

implies

$$\begin{aligned}
|F(t,x+\xi) - F(t,x) - \xi^\top \nabla_x F(t,x)| &= |F(t,x)|\,|e^{\psi(T-t)^\top \xi} - 1 - \xi^\top \psi(T-t)| \\
&\le |F(t,x)|\,\frac{3}{2}|\psi(T-t)^\top \xi|^2 \\
&\le \frac{3}{2}(|\xi|/c)^2
\end{aligned}$$

for all $(t,x) \in [0,T]\times E$, where in the last step we also used that (3.22) implies that $|F(t,x)| \le 1$ for all $(t,x) \in [0,T]\times E$. On the other hand, whenever $|\xi| > c$, we have

$$\begin{aligned}
|F(t,x+\xi) - F(t,x) - \xi^\top \nabla_x F(t,x)| &\le |F(t,x+\xi)| + |F(t,x)| + |F(t,x)|\,|\psi(T-t)|\,|\xi| \\
&\le 1 + 1 + 1 \times |\xi|/c \\
&\le 3|\xi|/c.
\end{aligned}$$

Combining these bounds gives

$$|F(t,x+\xi) - F(t,x) - \xi^\top \nabla_x F(t,x)| \le \left(\frac{3}{2c^2} \vee \frac{3}{c}\right) |\xi| \wedge |\xi|^2,$$

which in view of (3.18) implies that (2.7) holds for the process Y and the function F. This justifies the above application of Lemma 2.44, showing that M is a local martingale. Since $|M_t| \le 1$ due to (3.22), M is in fact a true martingale. Moreover, $M_T = \exp(u^\top X_T)$. The affine transform formula is therefore just the equality $M_t = \mathbb{E}[M_T \mid \mathcal{F}_t]$. $\square$

Remark 3.27. Compared to the moment formula, the proof of the affine transform formula looks rather short and simple. This is deceptive, because several questions are left unanswered by Theorem 3.26: (i) nothing is said about existence and uniqueness of solutions of the generalized Riccati equations (3.21); (ii) even if existence and uniqueness is established abstractly, one still has to verify (3.22) which can be difficult if the solution is not explicitly given; and (iii) it is often of interest to obtain the affine transform formula for u with non-zero real part. The martingale property of M then becomes more difficult to verify.

Exercise 3.28. Consider again the CIR process X with values in $E = \mathbb{R}_+$ given by

$$dX_t = (b + \beta X_t)dt + \sigma\sqrt{X_t}dW_t.$$

- Verify that the Riccati equations (3.21) become

$$\begin{aligned} \phi' &= b\psi, & \phi(0) &= 0, \\ \psi' &= \beta\psi + \frac{\sigma^2}{2}\psi^2, & \psi(0) &= u. \end{aligned}$$

  Show that the solution is given by

$$\phi(\tau) = b\int_0^\tau \psi(s)ds, \quad \psi(\tau) = \frac{e^{\beta\tau}}{u^{-1} - \frac{\sigma^2}{2}(e^{\beta\tau} - 1)/\beta}$$

  for any purely imaginary $u \neq 0$, provided $\beta \neq 0$. If $u = 0$ then clearly $\psi \equiv 0$. If $\beta = 0$, then passing to limit $\beta \to 0$ in the above expression suggests that $\psi(\tau) = (u^{-1} - \frac{\sigma^2}{2}\tau)^{-1}$. Verify that this indeed gives the correct solution.

- Show that $\operatorname{Re}\psi(\tau) \leq 0$ for all $\tau \geq 0$ if u is purely imaginary.

- What happens if u is not purely imaginary? Does the Riccati equations still have a global solution?

Example 3.29. One typically expects (and can often prove) that the Riccati equations will have global solutions whenever u is purely imaginary and the parameters b_i, a_i, ν_i are such that the process X exists. However, there are exceptions. For example, consider the two-point state space $E = \{0, 1\} \subseteq \mathbb{R}$, and the process X that jumps from 1 to 0 with intensity λ and is absorbed once it reaches 0. This is probably the simplest example of a non-trivial continuous time Markov chain. It is also a jump-diffusion with generator

$$\mathcal{G}f(x) = \lambda x(f(x-1) - f(x)),$$

whose coefficients are $a(x) = 0$, $b(x) = \lambda x$, $\nu(x, d\xi) = \lambda x \delta_{-1}$. Thus X is an affine jump-diffusion with $R_0(u) = 0$ and $R_1(u) = \lambda(e^{-u} - 1)$. The associated generalized Riccati equation is

$$\psi'(\tau) = \lambda(e^{-\psi(\tau)} - 1). \tag{3.23}$$

We claim that this equation does not have a global solution for the initial condition $u = \mathrm{i}\pi$. We argue by contradiction and assume that $\psi(\tau)$ is a global solution of (3.23). Then $\Psi(\tau) = e^{\psi(\tau)}$ satisfies the linear equation

$$\Psi'(\tau) = -\lambda\Psi(\tau) + \lambda, \qquad \Psi(0) = -1,$$

whose unique solution is $\Psi(\tau) = 1 - 2e^{-\lambda\tau}$. This becomes zero for $\tau = \lambda^{-1}\log 2$, a contradiction. The deeper reason behind this fact is that the characteristic function of X_T for $u = \mathrm{i}\pi$ and $T = \lambda^{-1}\log 2$ given $X_0 = 1$ becomes zero, and hence cannot be written as exponential as in the affine transform formula. Indeed, we have

$$\mathbb{E}[e^{uX_T}] = \mathbb{P}(X_T = 0) + e^u\,\mathbb{P}(X_T = 1) = 1 - (1 - e^u)e^{-\lambda T}.$$

Chapter 4

Existence and uniqueness of jump-diffusions

In previous chapters we have studied properties of *given* jump-diffusions, without worrying about questions of existence and uniqueness. Of course, in some cases, existence and uniqueness are obtained by other means, such as for Brownian motion and the Ornstein–Uhlenbeck process. In this chapter we outline a general method for obtaining existence of jump-diffusions via simple analytical criteria. Uniqueness is a more difficult problem, which we will only touch on briefly.

4.1 The existence theorem

Fix a closed state space $E \subseteq \mathbb{R}^d$ and a candidate generator $\mathcal{G}$ with coefficients (b, a, ν), that is,

$$\mathcal{G}f(x) = b(x)^\top \nabla f(x) + \frac{1}{2}\operatorname{Tr}(a(x)\nabla^2 f(x)) + \int_{\mathbb{R}^d} \Big(f(x+\xi) - f(x) - \xi^\top \nabla f(x) \Big)\, \nu(x, d\xi)$$

with $a(x) \in \mathbb{S}^d_+$, $\nu(x,\{0\}) = 0$, $\nu(x,(E-x)^c) = 0$, and $\int_{\mathbb{R}^d} |\xi|^2 \wedge |\xi| \nu(x,d\xi) < \infty$ for all $x \in E$. We assume that the coefficients are locally bounded[1], and that

$$\mathcal{G}f \text{ is continuous on } E \text{ and vanishes at infinity for any } f \in C^2_c(\mathbb{R}^d), \tag{4.1}$$

where $C^2_c(\mathbb{R}^d)$ denotes the space of compactly supported C^2 functions on $\mathbb{R}^d$. For any given law μ on E, the goal is to find an E-valued jump-diffusion X with $X_0 \sim \mu$, coefficients (b, a, ν), and generator $\mathcal{G}$.

4.1. One important method in the diffusion case ($\nu = 0$) is to try to obtain X as the solution of an SDE,

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x_0,$$

where $\sigma(x)\sigma(x)^\top = a(x)$ for all $x \in E$. The main tool for this is Itô's theorem on existence and uniqueness of solutions of SDEs. However:

- One needs b, σ Lipschitz on $\mathbb{R}^d$ (or at least *locally* Lipschitz on $\mathbb{R}^d$). This is too restrictive in practice; just consider the CIR process $dX_t = \kappa(\theta - X_t)dt + \sigma\sqrt{X_t}dW_t$.
- It is not clear how to guarantee that X stays in E, unless $E = \mathbb{R}^d$.
- Generalization to processes with jumps is not straightforward.

4.2. A different method, which is the one we will focus on below, is to proceed via the *martingale problem*. This method is very general and, importantly, easy to use. Throughout, we fix E, $\mathcal{G}$, (b, a, ν).

Definition 4.3. An E-valued càdlàg process $X = (X_t)_{t\geq 0}$ defined on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F})_{t\geq 0}, \mathbb{P})$ satisfying the usual conditions is called a *solution of MP*$(E, \mathcal{G}, \mu)$ ("the martingale problem for $(E, \mathcal{G})$ with initial law μ"), where μ is a probability measure on E, if $\mathbb{P}(X_0 \in \cdot\,) = \mu$ and

$$f(X_t) - f(X_0) - \int_0^t \mathcal{G}f(X_s)ds, \quad t \geq 0,$$

[1] More precisely, we assume that the function

$$E \to \mathbb{R}, \quad x \mapsto |a(x)| + |b(x)| + \int_{\mathbb{R}^d} |\xi|^2 \wedge |\xi| \nu(x, d\xi)$$

is locally bounded. By Proposition 3.6, this holds if $\mathcal{G}$ is polynomial on E.

is a local martingale for every $f \in C_b^2(\mathbb{R}^d)$.

Remark 4.4. Note that, when constructing a solution of a martingale problem, one is free to choose the underlying filtered probability space at one's convenience. In this sense, solutions of martingale problems are similar to weak solutions of SDEs. For us, a solution of a martingale problem is always a càdlàg process X. However, in many sources it is the *law* of X that is referred to as the solution of the martingale problem.

Remark 4.5. Other classes of test functions than $C_b^2(\mathbb{R}^d)$ are sometimes used.

Due to the following result, solving the martingale problem is enough to get existence of jump-diffusions.

Lemma 4.6. *Let X be a solution of MP$(E, \mathcal{G}, \mu)$. Then X is a jump-diffusion with coefficients (b, a, ν) and generator $\mathcal{G}$.*

Proof. Define $B_t = \int_0^t b(X_s)ds$, $C_t = \int_0^t a(X_s)ds$, and $\mu^p(dt, d\xi) = \nu(X_{t-}, d\xi)dt$ for $t \geq 0$. By Theorem 2.37, X is a special semimartingale with characteristics (B, C, μ^p). Thus, by the definition of jump-diffusion and generator (see Definitions 2.39 and 2.43), the result follows. □

We cannot expect to get existence of solutions of MP$(E, \mathcal{G}, \mu)$ without any further conditions. For instance, if $E = \mathbb{R}_+$ and $\mathcal{G}f(x) = \frac{1}{2}f''(x)$, no solution will exist. This is because any solution must be a Brownian motion, and will therefore not remain in $\mathbb{R}_+$. The following lemma gives an important necessary condition for existence.

Lemma 4.7. *Fix $x \in E$ and let X be a solution of MP$(E, \mathcal{G}, \delta_x)$. If $f \in C_b^2(\mathbb{R}^d)$ and x maximizes f over E, i.e. $f(x) = \max_{y \in E} f(y)$, then $\mathcal{G}f(x) \leq 0$.*

Proof. Assume for contradiction that $f(x) = \max_{y \in E} f(y)$ and $\mathcal{G}f(x) > 0$ for some $f \in C_b^2(\mathbb{R}^d)$. Define

$$\tau := \inf\{t \geq 0 : \mathcal{G}f(X_t) \leq 0\}$$

and note that this is a strictly positive stopping time. The assumptions on $\mathcal{G}$ and the maximality of $f(x)$ then yields

$$f(X_{t\wedge\tau}) - f(x) - \int_0^{t\wedge\tau} \mathcal{G}f(X_s)ds < 0 \qquad \forall t \geq 0.$$

But since X solves $\mathrm{MP}(E, \mathcal{G}, \delta_x)$, this is also a local martingale starting from 0. Since any nonpositive local martingale starting from zero is constant, we obtain a contradiction. $\square$

This motivates the following definition.

Definition 4.8. $\mathcal{G}$ satisfies $\mathrm{PMP}(E)$ ("the *positive maximum principle on* E") if the following property holds:

$$x \in E,\ f \in C_b^2(\mathbb{R}^d),\ f(x) = \max_{y \in E} f(y) \geq 0 \qquad \Longrightarrow \qquad \mathcal{G}f(x) \leq 0.$$

Due to Lemma 4.7, if there exists a solution of $\mathrm{MP}(E, \mathcal{G}, \delta_x)$ for every $x \in E$, then $\mathcal{G}$ satisfies $\mathrm{PMP}(E)$. It is a remarkable fact that the converse statement is true as well, up to one further condition.

Theorem 4.9. *Assume*

(i) $\mathcal{G}$ *satisfies PMP*(E),

(ii) *there exist functions* $f_n \in C_c^2(\mathbb{R}^d)$ *with* $\sup_{y \in E}(|f_n(y)| + |\mathcal{G}f_n(y)|) < \infty$ *as well as* $f_n(x) \to 1$ *and* $\mathcal{G}f_n(x) \to 0$ *for all* $x \in E$.

Then MP$(E, \mathcal{G}, \mu)$ *has a solution for every probability measure* μ *on* E.

Remark 4.10.
- If E is compact, then (ii) automatically holds: just take $f_n = 1$ on a neighborhood of E.

- If $\nu = 0$ (no jumps) and b and a satisfy the growth condition $|b(x)| + |a(x)|^{1/2} \leq c(1 + |x|)$ for all $x \in E$ and some constant c, then (ii) holds ($\longrightarrow$ exercise). In particular, this holds for polynomial diffusions.

- **Exercise.** Find a similar condition (involving ν as well as b and a) that works for polynomial jump-diffusions.

- If $\mathcal{G}$ satisfies $\mathrm{PMP}(E)$, then $\mathcal{G}f = \mathcal{G}g$ whenever $f(x) = g(x)$ for all $x \in E$: just note that $f - g$ has a minimum and a maximum over E in each $x \in E$. This in particular means that $\mathcal{G}f$ is independent of the behavior of f outside E. It also means that $\mathcal{G}$ is well-defined on $\mathrm{Pol}(E)$, provided (3.1) holds.

We outline the proof of this theorem in the case where E is compact. The idea is to approximate (in a suitable sense) the operator $\mathcal{G}$ by a sequence of simpler operators $\mathcal{G}^{(n)}$ for which the martingale problem is easy to solve. The solutions $X^{(n)}$ are then shown to converge (in a suitable sense, at least along a subsequence) to a process X that solves $\mathrm{MP}(E, \mathcal{G}, \mu)$. The positive maximum principle plays a central role for the construction of the sequence of operators $\mathcal{G}^{(n)}$ (see the proof of Lemma 4.12).

Notation: $\|f\|_E = \sup_{x \in E} |f(x)|$ denotes the supremum norm.

Lemma 4.11. *Assume E is compact and fix a law μ on E. Let $\mathcal{G}^{(n)} \colon C_b^2(\mathbb{R}^d) \to C(E)$ be linear operators such that $\|\mathcal{G}^{(n)} f - \mathcal{G} f\|_E \to 0$ for every $f \in C_b^2(\mathbb{R}^d)$. Let $X^{(n)}$ be a solution of MP($E, \mathcal{G}^{(n)}, \mu$), let X be an E-valued càdlàg process, and assume that $X^{(n)}$ converges to X in the sense of FDMDs, i.e.,*

$$(X_{t_1}^{(n)}, \ldots, X_{t_m}^{(n)}) \Rightarrow (X_{t_1}, \ldots, X_{t_m}) \text{ for all } m \in \mathbb{N},\ 0 \le t_1 < \cdots < t_m.$$

Then X is a solution of MP($E, \mathcal{G}, \mu$), if we take $(\mathcal{F}_t)_{t \ge 0}$ to be the right-continuous completion of $(\mathcal{F}_t^0)_{t \ge 0}$, where $\mathcal{F}_t^0 = \sigma(X_s \colon s \le t)$.

Proof. First consider the processes

$$\begin{aligned} N_t^f &:= f(X_t) - f(X_0) - \int_0^t \mathcal{G} f(X_s) ds, \qquad f \in C_b^2(\mathbb{R}^d), \\ N_t^{f,n} &:= f(X_t^{(n)}) - f(X_0^{(n)}) - \int_0^t \mathcal{G}^{(n)} f(X_s^{(n)}) ds, \qquad f \in C_b^2(\mathbb{R}^d), \end{aligned}$$

and note that since $X^{(n)}$ solves $\mathrm{MP}(E, \mathcal{G}^{(n)}, \mu)$, the processes $N^{f,n}$ are local martingales. More precisely, since f and $\mathcal{G}^{(n)} f$ are continuous and $X^{(n)}$ take values in E which is compact, $N^{f,n}$ are bounded local martingales and thus true martingales. Fix then $h_i \in C_b^2(\mathbb{R}^d)$ for $i = 1, \ldots, k$ and $0 \le t_1 \le \ldots \le t_k \le s < t$. Setting $\mathcal{F}_t^{0,n} := \sigma(X_s^{(n)} \colon s \le t)$ and noting that $X_{t_1}^{(n)}, \ldots, X_{t_k}^{(n)}$ are $\mathcal{F}_s^{0,n}$-measurable we then get

$$0 = \mathbb{E}\Big[\mathbb{E}[N_t^{f,n} - N_s^{f,n} | \mathcal{F}_s^{0,n}] \prod_{i=1}^k h_i(X_{t_i}^{(n)})\Big] = \mathbb{E}\Big[\Big(N_t^{f,n} - N_s^{f,n}\Big) \prod_{i=1}^k h_i(X_{t_i}^{(n)})\Big]. \tag{4.2}$$

Fix $\varepsilon > 0$ and note that for n large enough we have that $\|\mathcal{G}^{(n)} f - \mathcal{G} f\|_E < \varepsilon$. Thus (4.2) can be bounded from above by

$$\mathbb{E}\Big[f(X_t^{(n)})\prod_{i=1}^k h_i(X_{t_i}^{(n)})\Big]-\mathbb{E}\Big[f(X_s^{(n)})\prod_{i=1}^k h_i(X_{t_i}^{(n)})\Big]\\ -\int_s^t \mathbb{E}\Big[\mathcal{G}f(X_u^{(n)})\prod_{i=1}^k h_i(X_{t_i}^{(n)})\Big]du+\varepsilon C,$$

where $C=\mathbb{E}\big[\prod_{i=1}^k h_i(X_{t_i}^{(n)})\big](t-s)$. Letting n go to infinity, the dominated convergence theorem yields

$$0\le \mathbb{E}\Big[f(X_t)\prod_{i=1}^k h_i(X_{t_i})\Big]-\mathbb{E}\Big[f(X_s)\prod_{i=1}^k h_i(X_{t_i})\Big]-\int_s^t \mathbb{E}\Big[\mathcal{G}f(X_u)\prod_{i=1}^k h_i(X_{t_i})\Big]du+\varepsilon C.$$

Since ε was arbitrary, we conclude that $0\le\mathbb{E}[(N_t^f-N_s^f)\prod_{i=1}^k h_i(X_{t_i})]$. By a similar argument, we also get the converse inequality. The monotone class theorem and the definition of conditional expectation then yield $\mathbb{E}\big[N_t^f-N_s^f|\mathcal{F}_s]=0$, proving that N^f is a martingale. We deduce that X is a solution of $\mathrm{MP}(E,\mathcal{G},\mu)$. □

Lemma 4.12. *Assume that E is compact and $\mathcal{G}$ satisfies PMP(E). Then there exist kernels $\kappa^{(n)}$ on $\mathbb{R}^d$ such that*

- *$\kappa^{(n)}(x,E^c)=0$ and $\kappa^{(n)}(x,E)=1$ for all $x\in E$ and $n\in\mathbb{N}$.*
- *The operators $\mathcal{G}^{(n)}$ defined by*

$$\mathcal{G}^{(n)}f(x):=n\int_{\mathbb{R}^d}(f(y)-f(x))\kappa^{(n)}(x,dy),\qquad f\in C_b^2(\mathbb{R}^d)\tag{4.3}$$

satisfy $\|\mathcal{G}^{(n)}f-\mathcal{G}f\|_E\to0$ for every $f\in C_b^2(\mathbb{R}^d)$.

Proof. Observe that PMP(E) implies that the operator $\mathcal{G}$ is *dissipative*, i.e.

$$\|\lambda f-\mathcal{G}f\|_E\ge\lambda\|f\|_E\qquad\forall\lambda>0,\ f\in C_b^2(\mathbb{R}^d).\tag{4.4}$$

Indeed, choosing $x\in E$ such that $f(x)=\|f\|_E$ (without loss of generality $f(x)\ge0$, otherwise consider $-f$), by PMP(E) we have $\mathcal{G}f(x)\le0$ and hence

$$\|\lambda f-\mathcal{G}f\|_E\ge\lambda f(x)-\mathcal{G}f(x)\ge\lambda f(x)=\lambda\|f\|_E.$$

The dissipativity (4.4) of $\mathcal{G}$ is the key property for the construction of the kernels $\kappa^{(n)}$.

Fix $x \in E$ and $n \in \mathbb{N}$ and consider the operator $\mathcal{A}_n^x : (\mathcal{R}(n-\mathcal{G})|_E, \|\cdot\|_E) \to \mathbb{R}$ given by

$$\mathcal{A}_n^x f = n(n-\mathcal{G})^{-1} f(x),$$

where $n-\mathcal{G}$ denotes the linear operator mapping g to $ng-\mathcal{G}g$, $\mathcal{R}(n-\mathcal{G})|_E$ denotes the restriction to E of its range, i.e.

$$\mathcal{R}(n-\mathcal{G})|_E = \{f|_E : \ f = ng - \mathcal{G}g, \ g \in C_b^2(\mathbb{R}^d)\},$$

and $(n-\mathcal{G})^{-1}$ its inverse. This in particular means that $\mathcal{A}_n^x f = g$ for some g such that $\frac{ng-\mathcal{G}g}{n} = f$ on E. Observe then that

(i) $\mathcal{A}_n^x$ is well defined: suppose that $\frac{ng_1-\mathcal{G}g_1}{n} = \frac{ng_2-\mathcal{G}g_2}{n} = f$ on E, then

$$0 = \frac{1}{n}\|(n-\mathcal{G})(g_2-g_1)\|_E \overset{(4.4)}{\geq} \|g_2-g_1\|_E.$$

This implies that $g_2 = g_1$ on E and thus that $\mathcal{A}_n^x f = g_1(x)$ is well defined.

(ii) $\mathcal{A}_n^x$ is linear ($\longrightarrow$ exercise).

(iii) $\mathcal{A}_n^x 1 = 1$ ($\longrightarrow$ exercise).

(iv) $\mathcal{A}_n^x$ is a bounded functional of norm 1, i.e. $|\mathcal{A}_n^x f| \leq \|f\|_E$:

$$\|f\|_E = \frac{1}{n}\|(n-\mathcal{G})\big(n(n-\mathcal{G})^{-1}f\big)\|_E \overset{(4.4)}{\geq} \|n(n-\mathcal{G})^{-1}f\|_E \geq |n(n-\mathcal{G})^{-1}f(x)| = |\mathcal{A}_n^x f|.$$

(v) $\mathcal{A}_n^x$ is a positive functional, i.e. $\mathcal{A}_n^x f \geq 0$ whenever $f \geq 0$ on E: by (iii), (iv), and the nonnegativity of f on E we can compute

$$\|f\|_E - \mathcal{A}_n^x f = \mathcal{A}_n^x(\|f\|_E - f) \leq \big\|\|f\|_E - f\big\|_E = \|f\|_E.$$

Since $\mathcal{A}_n^x$ is a bounded linear operator, by the Hahn-Banach extension theorem it can be extended to a functional on $C(E)$ still satisfying (ii)-(iv). As before, we can also show for this extension that (iii) and (iv) imply (v). Due to (ii) and (v), the Riesz–Markov representation theorem is applicable, and we obtain that

$$\mathcal{A}_n^x f = \int_E f(y)\kappa^{(n)}(x,dy), \qquad f \in \mathcal{R}(n-\mathcal{G})|_E$$

for some (well chosen) measure $\kappa^{(n)}(x,\,\cdot\,)$ on E. By (iii), we also know that $\kappa^{(n)}(x,E)=1$.

For all $f\in C_b^2(\mathbb{R}^d)$ define now $\mathcal{A}_n^x f=\int_E f(y)\kappa^{(n)}(x,dy)$. In order to show that the linear operator $\mathcal{G}^{(n)}$ given by (4.3) satisfies the last item of the lemma, observe that since $\mathcal{A}_n^x\big(\frac{n-\mathcal{G}}{n}f\big)=n(n-\mathcal{G})^{-1}\big(\frac{n-\mathcal{G}}{n}f\big)(x)=f(x)$ for all $f\in C_b^2(\mathbb{R}^d)$ we get

$$\mathcal{A}_n^x f=\mathcal{A}_n^x f-\frac{1}{n}\mathcal{A}_n^x(\mathcal{G}f)+\frac{1}{n}\mathcal{A}_n^x(\mathcal{G}f)=f(x)+\frac{1}{n}\mathcal{A}_n^x(\mathcal{G}f) \tag{4.5}$$

and hence $\mathcal{G}^{(n)}f(x)=n\int f(y)-f(x)\kappa^{(n)}(x,dy)=n(\mathcal{A}_n^x f-f(x))=\mathcal{A}_n^x(\mathcal{G}f)$. Since by (4.5) and (iv) we also know that $\sup_{x\in E}|\mathcal{A}_n^x f-f(x)|$ converges to 0 for n going to infinity, we can conclude that

$$\|\mathcal{G}^{(n)}f-\mathcal{G}f\|_E=\sup_{x\in E}|\mathcal{A}_n^x(\mathcal{G}f)-\mathcal{G}f(x)|_E\to 0.$$

as $n\to\infty$. This completes the proof of the lemma. $\square$

Lemma 4.13. *Assume that E is compact, fix $n\in\mathbb{N}$, and let $\mathcal{G}^{(n)}$ be the operator (4.3) given by Lemma 4.12. Then there exists a solution of MP$(E,\mathcal{G}^{(n)},\mu)$ for any law μ on E.*

Proof. Fix a probability space $(\Omega,\mathcal{F},\mathbb{P})$ and let N be a poisson process with intensity n. Let $(Y_k)_{k\in\mathbb{N}_0}$ be the Markov chain given by $\mathbb{P}(Y_0\in A)=\mu(A)$ and

$$\mathbb{P}(Y_{k+1}\in A|Y_0,\dots,Y_k)=\kappa^{(n)}(Y_k,A),$$

for all $A\in\mathcal{B}(E)$. Define a process X by $X_t:=Y_0+\sum_{n=1}^{N_t}Y_n-Y_{n-1}$, $t\geq 0$. Then X along with (the right-continuous completion of) the filtration it generates solves MP$(E,\mathcal{G}^{(n)},\mu)$ ($\longrightarrow$ exercise). $\square$

Sketch of proof of Theorem 4.9. We only consider the case where E is compact. Fix a law μ on E. By Lemma 4.12 we obtain operators $\mathcal{G}^{(n)}$ that approximate $\mathcal{G}$, and by Lemma 4.13 we obtain solutions $X^{(n)}$ of MP$(E,\mathcal{G}^{(n)},\mu)$. The main technical point is now to argue that the sequence $\{X^{(n)}\}_{n\in\mathbb{N}}$ admits a limit point, in the sense that there exists a càdlàg process X such that a subsequence $\{X^{(n_k)}\}_{k\in\mathbb{N}}$ converges to X in the sense of FDMDs. This uses the theory of weak convergence in the Polish space $D([0,\infty);E)$ of càdlàg trajectories with values in E. We will not pursue this point here, though we do mention that it does not rely on further use of the positive maximum principle. Once X has been obtained and the convergence established, an application of Lemma 4.11 completes the proof of the theorem. $\square$

Remark 4.14. If E is not compact, then one performs essentially the same steps with E replaced by its one-point compactification E^Δ (and $\mathcal{G}$ extended in a suitable manner). One then gets a solution X of $\mathrm{MP}(E^\Delta, \mathcal{G}, \mu)$, which *a priori* may reach the cemetery state Δ ("explode in finite time"). The role of condition (ii) is to prevent this from happening, so that X remains in E and is a solution of $\mathrm{MP}(E, \mathcal{G}, \mu)$.

4.2 Applying the existence theorem to PJDs

Fix a closed state space $E \subseteq \mathbb{R}^d$ and the candidate generator $\mathcal{G}$ given by

$$\mathcal{G}f(x) = b(x)^\top \nabla f(x) + \frac{1}{2}\operatorname{Tr}(a(x)\nabla^2 f(x)) + \int_{\mathbb{R}^d} \Big(f(x+\xi) - f(x) - \xi^\top \nabla f(x) \Big)\, \nu(x, d\xi)$$

with $a(x) \in \mathbb{S}^d_+$, $\nu(x, \{0\}) = 0$, $\nu(x, (E-x)^c) = 0$, and $\int_{\mathbb{R}^d} |\xi|^2 \wedge |\xi| \nu(x, d\xi) < \infty$ for all $x \in E$. From the previous section we know that if $\mathcal{G}$ satisfies (4.1) and the conditions of Theorem 4.9, then $\mathrm{MP}(E, \mathcal{G}, \mu)$ has a solution for every probability measure μ on E. Since we are interested in PJDs, we assume that Proposition 3.6(ii) is satisfied. In this case, (4.1) automatically holds whenever $\nu = 0$ or E is compact. Moreover, Remark 4.10 shows that Theorem 4.9(ii) holds. We thus only need to check $\mathrm{PMP}(E)$. We illustrate now by means of some examples how $\mathrm{PMP}(E)$ can be verified.

Example 4.15 ($E = \mathbb{R}$)**.** Fix $g \in C^2_b(\mathbb{R})$ such that $g(x) = \max_{\mathbb{R}} g \geq 0$ for some $x \in \mathbb{R}$. Then one has the first- and second-order optimality conditions

$$g'(x) = 0 \qquad \text{and} \qquad g''(x) \leq 0. \tag{4.6}$$

This directly implies that the following operators satisfy $\mathrm{PMP}(\mathbb{R})$.

- Brownian motion: Set $\mathcal{G}f(x) = \frac{1}{2}f''(x)$. Then (4.6) yields $\mathcal{G}g(x) \leq 0$.

- Geometric Brownian motion: $\mathcal{G}f(x) = \beta x f'(x) + \frac{A}{2}x^2 f''(x)$ with $\beta \in \mathbb{R}$ and $A \geq 0$. Then (4.6) yields $\mathcal{G}g(x) = \frac{A}{2}x^2 g''(x) \leq 0$.

- PJD with bounded jumps:

$$\mathcal{G}f(x) = (b + \beta x) f'(x) + \frac{a + \alpha x + A x^2}{2} f''(x) + \int_{\mathbb{R}} \big(f(x+y) - f(x) - y f'(x) \big)\, \mu(dy),$$

for $b, \beta \in \mathbb{R}$, $a, \alpha, A \in \mathbb{R}$ such that $a + \alpha x + Ax^2 \geq 0$ for all $x \in \mathbb{R}$, and μ being a probability measure supported on $[-C, C] \setminus \{0\}$ for some $C > 0$. Then (4.6) yields $\mathcal{G}g(x) \leq \int_{\mathbb{R}} g(x+y) - g(x))\mu(dy)$, which is non-positive since $g(x) = \max_{\mathbb{R}} g$.

Since those operators also satisfy (4.1) ($\longrightarrow$ exercise), by Theorem 4.9 the corresponding martingale problems have an $\mathbb{R}_+$-valued solution for every initial condition.

Example 4.16 ($E = \mathbb{R}_+$)**.** Fix $g \in C_b^2(\mathbb{R})$ such that $g(x) = \max_{\mathbb{R}_+} g \geq 0$ for some $x \in \mathbb{R}_+$. Then one has the first- and second-order optimality conditions, which are special cases of the so-called *Karush–Kuhn–Tucker* conditions:

$$\text{If } x > 0, \text{ then (4.6) holds. If } x = 0, \text{ then } g'(0) \leq 0.$$

Indeed, the result for $x > 0$ is basic calculus, while for $x = 0$ one simply observes that $g'(0) = \lim_{y \to 0} \frac{g(y)-g(0)}{y}$, which is nonnegative since $g(0) = \max_{\mathbb{R}_+} g$. This condition automatically implies that the following operators satisfy PMP($\mathbb{R}_+$).

- Geometric Brownian motion: $\mathcal{G}f(x) = \beta x f'(x) + \frac{A}{2}x^2 f''(x)$ with $\beta \in \mathbb{R}$ and $A \geq 0$. Note that in Example 4.15 we obtained existence of an $\mathbb{R}$-valued solution of the corresponding martingale problem. What we are proving now is existence of an $\mathbb{R}_+$-valued solution.

- CIR process: $\mathcal{G}f(x) = \kappa(\theta - x)f'(x) + \frac{\sigma^2}{2} x f''(x)$ with $\kappa, \theta, \sigma \in \mathbb{R}$ such that $\kappa\theta \geq 0$.

Since those operators also satisfy $\nu = 0$ and thus condition (4.1), by Theorem 4.9 the corresponding martingale problems have $\mathbb{R}_+$-valued solutions for every initial condition.

Remark 4.17. Example 4.16 show that, at the boundary of E, the drift needs to be "inward-pointing" and the diffusion needs to vanish. This is the case for $d = 1$, but what about $d \geq 2$? For instance, what do the drift and diffusion coefficients look like for the polynomial diffusion given by (W, W^2), where W a one-dimensional Brownian motion?

Example 4.18. ($E = [0, 1]$.) Fix $g \in C_b^2(\mathbb{R})$ such that $g(x) = \max_{[0,1]} g \geq 0$ for some $x \in [0, 1]$. Then one has the first- and second-order optimality conditions ($\longrightarrow$ exercise):

$$\text{If } x \in (0, 1), \text{ then (4.6) holds. If } x = 0, \text{ then } g'(0) \leq 0, \text{ and if } x = 1, \text{ then } g'(1) \geq 0.$$

This condition automatically implies that the following operators satisfy PMP($[0, 1]$).

- Jacobi Diffusion: $\mathcal{G}f(x) = \kappa(\theta - x)f'(x) + \frac{\sigma^2}{2}x(1-x)f''(x)$ with $\kappa, \theta, \sigma \in \mathbb{R}$ such that $\kappa\theta \geq 0$, $\kappa(\theta - 1) \leq 0$. For $x \in (0,1)$ Lemma **??** yields $\mathcal{G}g(x) \leq 0$. For $x = 0, 1$ Lemma **??** yields $\mathcal{G}g(0) = \kappa\theta g'(0) \leq 0$ and $\mathcal{G}g(1) = \kappa(\theta - 1)g'(1) \leq 0$, respectively.

- Polynomial Jump Diffusion with negative jumps:

$$\mathcal{G}f(x) = \kappa(\theta - x)f'(x) + \frac{\sigma^2}{2}x(1-x)f''(x) + \int_{[0,1]} \big(f(x - xy) - f(x) + xyf'(x)\big)\,\mu(dy)$$

  with $\kappa, \theta, \sigma \in \mathbb{R}$ and μ being a probability measure on $(0,1]$ such that $\kappa\theta \geq 0$ and $\kappa(\theta - 1) + \int y\mu(dy) \geq 0$. With this specification, if a jump occurs when the process is at level x, then the size of the jump is $-yx$, where y is μ-distributed.

Since these operators also satisfy condition (4.1) ($\longrightarrow$ exercise), by Theorem 4.9 the corresponding martingale problems have $[0,1]$-valued solution for every initial condition.

Remark 4.19. Example 4.18 shows that if $\nu \neq 0$, the intuition given in Remark 4.17 needs to be adjusted: It is now the *effective drift*, namely $b(x)$ adjusted by the jump compensation $\int -\xi\nu(x, d\xi)$, that needs to be inward-pointing at the boundary.

4.20. The approach illustrated in the examples above can be used for (much) more general state spaces. Checking the positive maximum principle is always central, and this is done by applying optimality conditions like the Karush–Kuhn–Tucker conditions.

4.3 Uniqueness of PJDs and AJDs

While the existence problem has a satisfactory solution for a very general class of jump-diffusions, uniqueness (in law) is much more delicate. For PJDs and AJDs, it is sometimes possible to obtain uniqueness by appealing to the moment formula and the affine transform formula, respectively.

- Let X be a PJD on $E \subseteq \mathbb{R}^d$ with generator $\mathcal{G}$. By iterating the moment formula, it follows that all joint moments of all FDMDs, namely the quantities

$$\mathbb{E}[X_{t_1}^{\boldsymbol{\alpha}^{(1)}} \cdots X_{t_m}^{\boldsymbol{\alpha}^{(m)}}], \quad m \in \mathbb{N},\ 0 \leq t_1 < \cdots < t_m,\ \boldsymbol{\alpha}^{(1)}, \ldots, \boldsymbol{\alpha}^{(m)} \in \mathbb{N}_0^d,$$

are uniquely determined by $\mathcal{G}$ and the law of X_0. In some cases it can be shown that this uniquely determines all FDMDs, which proves uniqueness in law of solutions of the martingale problem. However, there are cases where the moments do *not* uniquely determine the distribution. A famous such example is the lognormal distribution: there exist distributions different from the lognormal, which nonetheless have the same moments as the lognormal. In such cases, uniqueness in law must be proved by other means. For example, this is the case with geometric Brownian motion, whose one-dimensional marginal distributions are lognormal!

- Let X be an AJD on $E \subseteq \mathbb{R}^d$ with generator $\mathcal{G}$. By iterating the affine transform formula, and assuming that existence of solutions can be proved for all involved Riccati equations, it follows that the characteristic functions of all FDMDs, namely the quantities

$$\mathbb{E}[\exp(\mathrm{i}\boldsymbol{\lambda}^{(1)\top} X_{t_1} + \cdots + \mathrm{i}\boldsymbol{\lambda}^{(1)\top} X_{t_1})], \quad m \in \mathbb{N},\ 0 \leq t_1 < \cdots < t_m,\ \boldsymbol{\lambda}^{(1)}, \ldots, \boldsymbol{\lambda}^{(m)} \in \mathbb{R}^d,$$

are uniquely determined by $\mathcal{G}$ and the law of X_0. This uniquely determines the FDMDs themselves, and proves uniqueness in law of solutions of the martingale problem.

Chapter 5

Applications in Finance

The goal of this chapter is to illustrate how polynomial and affine jump-diffusions can be used to construct financial models. There is a large and growing literature where this is done. In this course, we will focus on a few specific but important examples. Specifically, we will consider affine and polynomial stochastic volatility models, as well as affine and polynomial interest rate models.

Beyond stochastic volatility and interest rates, polynomial jump-diffusions have been used to model exchange rates, life insurance liabilities, variance swaps, credit risk, dividend futures, commodities and electricity, stochastic portfolio theory, among other things. For references, see Filipović and Larsson (2017). Affine jump-diffusions have been studied longer, and has rich history in the finance literature, where they have long been used to address a large number of problems in asset pricing, optimal investment, equilibrium analysis, etc.

5.1 Derivatives pricing via absence of arbitrage

In this section we briefly review the fundamental theorem of asset pricing, which provides a convenient framework for constructing asset pricing models. Fix a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ satisfying the usual conditions. Let $X = (X^1, \ldots, X^d)$ be a semimartingale that models the (discounted[1]) prices of d liquidly traded assets.

[1] That is, $X^i_t = e^{-\int_0^t r_s ds} P^i_t$, where P^i_t models the nominal price and r_t models the short interest rate.

Informal definition. X is *arbitrage-free* if it is impossible generate riskless profits by trading in X using $(\mathcal{F}_t)$-predictable strategies.

5.1. Of course, work is needed to make this definition mathematically rigorous. Slightly different definitions may be appropriate depending on the circumstances, and one of the achievements of mathematical finance has been to clarify the relations between different possible definitions.

5.2. Absence of arbitrage is a desirable consistency property that a pricing model ought to satisfy. Without it, the analysis may yield questionable outcomes. For example, if one tries to numerically calculate optimal trading strategies in a model with arbitrage, these trading strategies are likely to be nonsensical.

5.3. Given an arbitrage-free model, one can obtain prices of *derivative securities* (or at least a range of possible prices) by the principle of *arbitrage-free pricing.* This principle states that the price process of the derivative should be chosen so that the joint model remains arbitrage-free. We illustrate this with an example.

Example 5.4. Suppose $S = (S_t)_{t\geq 0}$ is an arbitrage-free model for the price of a stock. Let us take $r_t = 0$ for simplicity, so that S is already discounted. A *European call option* written on the stock with maturity T and strike price K can be viewed as a security that pays the amount $C_T := (S_T - K)^+$ to the holder at time T. This payoff is stochastic and unknown at times $t < T$, but still only depends on the behavior of the *underlying*, that is the stock. We wish to identify the price C_t of the call option at times $t < T$. The principle of arbitrage-free pricing states that $C = (C_t)_{t\in[0,T]}$ should be selected to that the joint model $X = (S, C)$ is arbitrage-free. The key tool for doing so is the following theorem.

Fundamental theorem of asset pricing. Let $X = (X^1, \ldots, X^d)$ model the (discounted) prices of d liquidly traded assets. Then, up to technical conditions,

$$X \text{ is arbitrage-free} \quad \Longleftrightarrow \quad \exists \mathbb{Q} \sim \mathbb{P} \text{ such that } X \text{ is a local } \mathbb{Q}\text{-martingale.}$$

The probability measure $\mathbb{Q}$ is called a *risk-neutral measure* or *equivalent local martingale measure.*

5.5. The validity of this theorem depends on the details of the definition of "arbitrage-free" as well as the precise assumptions made on X. However, the sufficiency direction "$\Leftarrow$" holds under any reasonable definition of "arbitrage-free". This has two important practical consequences:

(i) By directly specifying the (discounted) model price $S = (S_t)_{t\geq 0}$ of an underlying asset as a local martingale under a probability measure $\mathbb{Q}$, we are guaranteed to obtain an arbitrage-free model.

(ii) By then specifying the price of the derivative as $C_t := \mathbb{E}_{\mathbb{Q}}[C_T \mid \mathcal{F}_t]$, we are additionally guaranteed to obtain an arbitrage-free joint model (S, C). One then often speak about $\mathbb{Q}$ as the *pricing measure.*

(In the case of non-zero interest rate, we specify $(e^{-\int_0^t r_s ds} S_t)_{t\geq 0}$ to be a local martingale, and price the derivative by $C_t = \mathbb{E}_{\mathbb{Q}}[e^{-\int_t^T r_s ds} C_T \mid \mathcal{F}_t]$.)

This general procedure presents us with two competing objectives: on one hand, the model S of the underlying has to be rich enough to capture empirically observed features; on the other hand, S has to be simple enough that quantities like $\mathbb{E}_Q[(S_T - K)^+ \mid \mathcal{F}_t]$ can be computed. Polynomial and affine jump-diffusions provide an excellent trade-off.

Example 5.6. The classical Black–Scholes stock price model is $S_t = S_0 \exp(\sigma W_t - \frac{\sigma^2}{2} t)$ under the risk-neutral measure $\mathbb{Q}$ assuming zero interest rates, where W is a standard Brownian motion under $\mathbb{Q}$ and $\sigma > 0$ is a volatility parameter. The time-zero price of a call option,

$$C_0 = \mathbb{E}_{\mathbb{Q}}[(S_T - K)^+] = \mathbb{E}_{\mathbb{Q}}[(e^{\sigma W_t - \frac{\sigma^2}{2} t} - K)^+],$$

can then be calculated explicitly.

5.2 Affine stochastic volatility models

5.7. Fix a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{Q})$ satisfying the usual conditions, where $\mathbb{Q}$ will play the role of risk-neutral measure. We will consider some affine stochastic volatility models, focusing on the diffusion case (no jumps). We take the interest rate to be zero, $r_t \equiv 0$.

5.8. A generic stochastic volatility model has the following structure. The stock price is modeled by a strictly positive process $S = (S_t)_{t\geq 0}$ whose log-price $Y_t = \log S_t$ is given by

$$dY_t = -\frac{1}{2}\sigma_t^2 dt + \sigma_t dB_t,$$

where B is a standard Brownian motion, and $(\sigma_t)_{t\geq 0}$ is a predictable process such that $\int_0^t \sigma_s^2 ds < \infty$ for all $t \geq 0$ (in other words, $(\sigma_t)_{t\geq 0}$ should be B-integrable). By Itô's formula ($\longrightarrow$ exercise), S then satisfies

$$dS_t = S_t \sigma_t dB_t,$$

so S is a local martingale. Thus the model is arbitrage-free.

5.9. We already saw the Heston (1993) model in Example 3.25. It consists of the two-dimensional process (Y, V) with values in $E = \mathbb{R} \times \mathbb{R}_+$ given by

$$\begin{aligned} dY_t &= -\frac{1}{2}V_t dt + \sqrt{V_t} dB_t \\ dV_t &= (b + \beta V_t)dt + \sigma\sqrt{V_t} d\widetilde{B}_t, \end{aligned}$$

where $(B, \widetilde{B})$ is a correlated two-dimensional Brownian motion with correlation $\rho \in [0, 1]$. This is a stochastic volatility model in the sense of 5.8 with $\sigma_t = \sqrt{V_t}$. (Be careful with the potentially confusing notation: here $\sigma_t = \sqrt{V_t}$ stands for the volatility process as above, but there is also a constant parameter $\sigma \geq 0$ that enters in the dynamics of V.)

The Heston model can be extended with several factors. This means that the volatility of the stock depends on more than one stochastic process. Here is a simple two-factor extension of the Heston model, which preserves the affine structure.

$$dY_t = -\frac{1}{2}(V_t^1 + V_t^2)dt + \sqrt{V_t^1} dB_t^1 + \sqrt{V_t^2} dB_t^2, \tag{5.1}$$

$$dV_t^1 = (b_1^V + \beta_{11}^V V_t^1 + \beta_{12}^V V_t^2)dt + \sigma_1\sqrt{V_t^1} d\widetilde{B}_t^1, \tag{5.2}$$

$$dV_t^2 = (b_2^V + \beta_{21}^V V_t^1 + \beta_{22}^V V_t^2)dt + \sigma_2\sqrt{V_t^2} d\widetilde{B}_t^2, \tag{5.3}$$

where $B^1, B^2, \widetilde{B}^1, \widetilde{B}^2$ are Brownian motions with $\langle B^i, \widetilde{B}^i\rangle_t = \rho_i t$ $(i = 1, 2)$, and all other pairs uncorrelated.

Remark 5.10. Note that (5.1)–(5.3) indeed specify a stochastic volatility model in the sense of 5.8, because we have

$$dY_t = -\frac{1}{2}\sigma_t^2 + \sigma_t dW_t,$$

where $\sigma_t := \sqrt{V_t^1 + V_t^2}$ and W is defined by

$$dW_t := \frac{1}{\sigma_t}\left(\sqrt{V_t^1}dB_t^1 + \sqrt{V_t^2}dB_t^2\right).$$

Note that W is indeed a Brownian motion by Lévy's characterization theorem ($\rightarrow$ exercise).

Proposition 5.11. *Suppose $b_i^V \geq 0$ and $\beta_{ij}^V \geq 0$ for all i and $j \neq i$. Then for any starting point in $E := \mathbb{R} \times \mathbb{R}_+^2$ the process $X := (Y, V^1, V^2)$, is E-valued, and is affine on E.*

Proof. Existence follows from Theorem 4.9 ($\rightarrow$ exercise). To verify the affine property, note that X is an E-valued diffusion with coefficients $(b, a, 0)$, where

$$b(x) = b(y, v_1, v_2) = \begin{pmatrix} 0 \\ b_1^V \\ b_2^V \end{pmatrix} + \begin{pmatrix} 0 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \beta_{11}^V & \beta_{12}^V \\ 0 & \beta_{21}^V & \beta_{22}^V \end{pmatrix} \begin{pmatrix} y \\ v_1 \\ v_2 \end{pmatrix}.$$

To compute $a(x)$, consider the continuous local martingale part of X,

$$dM_t^c = \Sigma(X_t)dB_t,$$

where

$$\Sigma(x) = \Sigma(y, v_1, v_2) = \begin{pmatrix} \sqrt{v_1} & \sqrt{v_2} & 0 & 0 \\ 0 & 0 & \sigma_1\sqrt{v_1} & 0 \\ 0 & 0 & 0 & \sigma_2\sqrt{v_2} \end{pmatrix}, \qquad dB_t = \begin{pmatrix} dB_t^1 \\ dB_t^2 \\ d\widetilde{B}_t^1 \\ d\widetilde{B}_t^2 \end{pmatrix}.$$

Letting as usual $\langle M^c, M^c \rangle$ be the matrix-valued process with components $\langle M^{c,i}, M^{c,j} \rangle$, and similarly for $\langle B, B \rangle$, one verifies that

$$a(X_t)dt = d\langle M^c, M^c \rangle_t = \Sigma(X_t)d\langle B, B \rangle_t \Sigma(X_t)^\top = \Sigma(X_t) \begin{pmatrix} 1 & 0 & \rho_1 & 0 \\ * & 1 & 0 & \rho_2 \\ * & * & 1 & 0 \\ * & * & * & 1 \end{pmatrix} \Sigma(X_t)^\top$$

$$= \begin{pmatrix} V_t^1 + V_t^2 & \rho_1\sigma_1 V_t^1 & \rho_2\sigma_2 V_t^2 \\ * & \sigma_1^2 V_t^1 & 0 \\ * & * & \sigma_2^2 V_t^2 \end{pmatrix} dt,$$

where "$*$" is to be filled in by symmetry. Thus X is affine on E by Proposition 3.23. $\square$

Remark 5.12. One can clearly consider generalizations to more than two volatility factors and more than one stock. However, the affine requirement imposes restrictions on the possible dynamics. For example, in the two-factor model (5.1)–(5.3), try to modify the correlation structure of dB_t without destroying the affine property.

5.13. Next, we consider a model that does not look affine at first glance, but turns out to have an affine structure after *augmenting* the factor process. This augmentation technique is sometimes very useful. The Stein and Stein (1991) stochastic volatility model is specified by

$$\begin{aligned} dY_t &= -\frac{1}{2}\sigma_t^2 dt + \sigma_t dB_t, \\ d\sigma_t &= (b^\sigma + \beta^\sigma \sigma_t)dt + \alpha d\widetilde{B}_t, \end{aligned}$$

where b^σ, β^σ, and α are real parameters, and B and $\widetilde{B}$ are correlated Brownian motions with correlation parameter ρ. We see from Proposition 3.23 that the joint process (Y, σ) is *not* affine, since for example the drift of Y is quadratic in σ. Nonetheless, we have the following result:

Proposition 5.14. *The process $X = (Y, \sigma, \sigma^2)$ is affine on $E := \mathbb{R} \times \{(u, v) \in \mathbb{R}^2 \colon v = u^2\}$.*

Proof. First observe that Itô's formula gives

$$d\sigma_t^2 = 2\sigma_t d\sigma_t + d\langle\sigma\rangle_t = (\alpha^2 + 2b^\sigma\sigma_t + 2\beta^\sigma\sigma_t^2)dt + 2\alpha\sigma_t d\widetilde{B}_t.$$

From this we infer that X is an E-valued diffusion whose coefficients $(b, a, 0)$ satisfy

$$b(x) = \begin{pmatrix} 0 \\ b^\sigma \\ \alpha^2 \end{pmatrix} + \begin{pmatrix} 0 & 0 & -\frac{1}{2} \\ 0 & \beta^\sigma & 0 \\ 0 & 2b^\sigma & 2\beta^\sigma \end{pmatrix} \begin{pmatrix} y \\ \sigma \\ \sigma^2 \end{pmatrix}, \quad a(x) = \begin{pmatrix} \sigma^2 & \rho\alpha\sigma & 2\rho\alpha\sigma^2 \\ * & \alpha^2 & 2\alpha^2\sigma \\ * & * & 4\alpha^2\sigma^2 \end{pmatrix}.$$

Since both depend on (y, σ, σ^2) in an affine way, we deduce that X is affine on E. □

5.3 Option pricing in affine stochastic volatility models

We continue to work on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{Q})$ satisfying the usual conditions, and consider an affine jump-diffusion X on $E \subseteq \mathbb{R}^d$. In an affine stochastic volatility model such as those we have seen above, the log-price is a component of the process X, say $\log S_t = Y_t := X_t^1$. To price European puts (and similarly for other options), we need to compute

$$C_t := \mathbb{E}_{\mathbb{Q}}[(K - S_T)^+ \mid \mathcal{F}_t] = \mathbb{E}_{\mathbb{Q}}[(K - e^{Y_T})^+ \mid \mathcal{F}_t].$$

However, the affine transform formula only allows us to compute quantities like $\mathbb{E}_{\mathbb{Q}}[e^{u^\top X_T} \mid \mathcal{F}_t]$ in a tractable way. How can we leverage this tractability to price options? A key method for doing so is *Fourier pricing*, which is based on the following identity from Fourier analysis.

Lemma 5.15. *Let $K > 0$ and $w > 0$. Then*

$$(K - e^y)^+ = \frac{1}{2\pi} \int_{\mathbb{R}} e^{(\mathrm{i}\lambda - w)y} \frac{K^{w+1-\mathrm{i}\lambda}}{(i\lambda - w)(\mathrm{i}\lambda - w - 1)} d\lambda$$

for all $y \in \mathbb{R}$.

Proof. Observe that the function $h(y) := e^{wy}(K - e^y)^+$ is in $L^1(\mathbb{R})$. Its Fourier transform is thus well-defined and given by

$$\hat{h}(\lambda) := \int_{\mathbb{R}} e^{-\mathrm{i}\lambda y} h(y) dy = \frac{K^{w+1-\mathrm{i}\lambda}}{(i\lambda - w)(\mathrm{i}\lambda - w - 1)}.$$

Through a computation ($\longrightarrow$ exercise) one finds

$$|\hat{h}(\lambda)| = \frac{K^{w+1}}{\lambda^2 + w(w+1)},$$

which shows that $\hat{h} \in L^1(\mathbb{R})$. We may thus apply the inverse Fourier transform to obtain

$$h(y) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{\mathrm{i}\lambda y} \hat{h}(y) dy.$$

This is the claimed identity. $\square$

Using Lemma 5.15 together with Fubini's theorem, we are able to derive the following formula for the put price in a model where the log-price is $Y_t := \log S_t$:

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}}[(K - S_T)^+ \mid \mathcal{F}_t] &= \mathbb{E}_{\mathbb{Q}}[(K - e^{Y_T})^+ \mid \mathcal{F}_t] \\
&= \mathbb{E}_{\mathbb{Q}}\left[\frac{1}{2\pi} \int_{\mathbb{R}} e^{(\mathrm{i}\lambda - w)Y_T} \frac{K^{w+1-\mathrm{i}\lambda}}{(i\lambda - w)(\mathrm{i}\lambda - w - 1)} d\lambda \mid \mathcal{F}_t\right] \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} \mathbb{E}_{\mathbb{Q}}[e^{(\mathrm{i}\lambda - w)Y_T} \mid \mathcal{F}_t] \frac{K^{w+1-\mathrm{i}\lambda}}{(i\lambda - w)(\mathrm{i}\lambda - w - 1)} d\lambda.
\end{aligned}$$

The application of Fubini's theorem is valid provided that $\mathbb{E}_{\mathbb{Q}}[e^{-wY_T} \mid \mathcal{F}_t] < \infty$. This has to be verified on a case by case basis, and typically only holds for sufficiently small values of w. The formula that we obtain can now be used, since the conditional expectations $\mathbb{E}_{\mathbb{Q}}[e^{(\mathrm{i}\lambda - w)Y_T} \mid \mathcal{F}_t]$ are accessible via the affine transform formula.

5.4 The Jacobi stochastic volatility model

The following is an example of stochastic volatility model based on a polynomial diffusion, introduced by Ackerer et al. (2016). We work on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{Q})$ satisfying the usual conditions, on which a two-dimensional standar Brownian motion $W = (W^1, W^2)$ is defined. Fix two parameters $0 \leq v_{\min} < v_{\max} < \infty$, and define

$$Q(v) := \frac{(v - v_{\min})(v_{\max} - v)}{(\sqrt{v_{\max}} - \sqrt{v_{\min}})^2}.$$

The squared volatility is specified as

$$dV_t = \kappa(\theta - V_t)dt + \sigma\sqrt{Q(V_t)}dW_t^1, \tag{5.4}$$

for some parameters $\kappa \geq 0$, $\theta \in [v_{\min}, v_{\max}]$, $\sigma > 0$. The process V exists and is valued in $[v_{\min}, v_{\max}]$ due to Theorem 4.9 ($\longrightarrow$ exercise). Next, the log-price is specified as

$$dY_t = -\frac{1}{2}V_t dt + \rho\sqrt{Q(V_t)}dW_t^1 + \sqrt{V_t - \rho^2 Q(V_t)}dW_t^2, \tag{5.5}$$

where $\rho \in [-1, 1]$. Observe that $Q(v) \leq v$, so that $\sqrt{V_t - \rho^2 Q(V_t)}$ makes sense. Moreover, $d\langle Y\rangle_t = V_t dt$, so that V_t is the squared volatility and we indeed have a stochastic volatility model in the sense of 5.8.

Exercise 5.16. Show that $X := (Y, V)$ is a polynomial diffusion on $E := \mathbb{R} \times [v_{\min}, v_{\max}]$ with coefficients $(b, a, 0)$ given by

$$b(x) = b(y, v) = \begin{pmatrix} -v/2 \\ \kappa(\theta - v) \end{pmatrix}, \qquad a(x) = a(y, v) = \begin{pmatrix} v & \rho\sigma Q(v) \\ x & \sigma^2 Q(v) \end{pmatrix}$$

5.5 Option pricing in polynomial stochastic volatility models

5.17. For stochastic volatility models based on polynomial jump-diffusions the Fourier methodology used in the affine case is no longer appropriate. Rather than expressing option prices in terms of characteristics functions, we wish to express them in terms of moments. We now describe a general method for doing so. We consider the problem of computing expectations of the form

$$\mathbb{E}_{\mathbb{Q}}[f(Y_T)],$$

where Y is a component of a polynomial jump-diffusion, and $f\colon \mathbb{R} \to \mathbb{R}$ is a function determined by the option payoff. The running example we have in mind is to price European call options in the Jacobi stochastic volatility model (5.4)–(5.5), in which case Y is the log-price and $f(y) = (e^y - K)^+$.

5.18. We assume that Y_T has a density function g_T. In the Jacobi model, a sufficient condition for this is that $\int_0^T (V_s - \rho^2 Q(V_s))ds > 0$ a.s. We also fix an auxiliary density function w on $\mathbb{R}$ such that

$$\int_{\mathbb{R}} \frac{g_T(x)^2}{w(x)} dx < \infty. \tag{5.6}$$

We may then define the weighted Hilbert space

$$L^2_w := \{\text{all measurable } g\colon \mathbb{R} \to \mathbb{R} \text{ such that } \int_{\mathbb{R}} g(x)^2 w(x) dx < \infty\},$$

where as usual we identify functions that are w-a.e. equal. The inner product on this space is

$$\langle g, h \rangle_w := \int_{\mathbb{R}} g(x)h(x)w(x)ds.$$

The basic assumption is now that the payoff function f satisfies

$$f \in L^2_w.$$

Note that (5.6) expresses that the *likelihood ratio* function is in L^2_w,

$$\ell := \frac{g_T}{w} \in L^2_w$$

The basic idea is now the following. If $\{h_n\}_{n\geq 0}$ is an orthonormal basis (ONB) for L^2_w, then, since both f and ℓ lie in L^2_w, one has

$$\mathbb{E}_{\mathbb{Q}}[f(Y_T)] = \int_{\mathbb{R}} f(x) g_T(x) dx = \langle f, \ell \rangle_w = \sum_{n\geq 0} \langle f, h_n \rangle_w \langle \ell, h_n \rangle_w. \tag{5.7}$$

By suitably choosing the objects involved, it turns out that the moment formula can be used to evaluate the right-hand side (of course, one has to truncate the infinite sum). Let us look at how this works in the Jacobi model.

5.19. We put ourselves in the setup of the Jacobi stochastic volatility model. Let w be the Gaussian density function with parameters (μ_w, σ_w) with $\sigma_w > v_{\max}T/2$. One can then show that (5.6) holds. Next, set

$$h_n(x) := \frac{1}{\sqrt{n!}}\mathcal{H}_n\left(\frac{x-\mu_n}{\sigma_w}\right)$$

where $\mathcal{H}_n$ is the nth Hermite polynomial,

$$\mathcal{H}_n(x) = (-1)^n e^{x^2/2}\frac{d^n}{dx^n}e^{-x^2/2}.$$

(This is a polynomial of degree n!) It is then known that $\{h_n\}_{n\geq 0}$ is an ONB for L^2_w. We now consider how to compute the right-hand side of (5.7).

- $\langle f, h_n\rangle_w = \int_{\mathbb{R}} f(x)h_n(x)w(x)dx$ is the expectation of fh_n under a Normal(μ_w, σ_w) distribution, which can be computed efficiently. In fact, one can even obtain extremely efficient recursive formulas.
- $\langle \ell, h_n\rangle_w = \int_{\mathbb{R}} \frac{g_T(x)}{w(x)}h_n(x)w(x)dx = \int_{\mathbb{R}} h_n(x)g_T(x)dx = \mathbb{E}_{\mathbb{Q}}[h_n(Y_T)]$ can be computed efficiently using the moment formula, since Y_T is a component of a polynomial (jump-)diffusion, and h_n is a polynomial.

Further description of the general method can be found in Filipović and Larsson (2017), and for details in the case of the Jacobi model, see Ackerer et al. (2016).

5.6 Overview of interest rate and fixed income models

5.20. A major application of polynomial and affine jump-diffusions is in the area of interest rates and fixed income. The basic security is now the *zero-coupon bond* (ZCB). The holder of a ZCB with maturity T is guaranteed a payoff of one currency unit (e.g. USD or CHF) at time T. Thus the value of a ZCB at maturity is known in advance: it is equal to one. The question is what its value is at times $t < T$. Letting $P(t,T)$ denote the (model) price, the general arbitrage-free pricing framework thus postulates

$$P(t,T) = \mathbb{E}_Q[e^{-\int_t^T r_s ds} \times 1 \mid \mathcal{F}_t], \tag{5.8}$$

where $\mathbb{Q}$ is a pricing measure, and r_t is the short rate. Note that $P(T,T) = 1$, as it should. Note also that, unlike the stock price models considered in previous sections, it is crucial to work with stochastic interest rates. If the interest rate is deterministic (let alone zero), the pricing problem is trivial.

The collection $\{P(t,T), T \geq t\}$ of ZCB prices at each time t is called the *term structure of bond prices*. Consequently, models for ZCBs are often called *term structure models*.[2]

In these notes we focus on ZCBs. There are many other important fixed income instruments, such as forward rate agreements, caps, floors, coupon bonds, swaps, swaptions, etc., but they are all in some way constructed from simple ZCBs.

5.21. There are several possible modeling approaches. The most important ones are:

- **Short-rate models:** One directly models the short rate r_t in such a way that (5.8) can be computed directly. Affine jump-diffusions form a natural basis for such models.

- **Heath–Jarrow–Morton (HJM) models:** Define the *forward rates* $f(t,s)$ via the equation $P(t,T) = e^{-\int_t^T f(t,s)ds}$. The function $s \mapsto f(t,s)$, which is observed at time t, is called the *forward curve* at time t. As t increases, the forward curve evolves stochastically, and an HJM model seeks to describe this evolution.

- **State price density models:** Consider the real-world measure $\mathbb{P}$ and the pricing measure $\mathbb{Q} \sim \mathbb{P}$, and let $Z_t := \frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{F}_t}$ be the corresponding Radon–Nikodym density process. Bayes' rule gives
$$P(t,T) = \mathbb{E}_Q[e^{-\int_t^T r_s ds} \mid \mathcal{F}_t] = \frac{1}{Z_t e^{-\int_0^t r_s ds}} \mathbb{E}_{\mathbb{P}}[Z_T e^{-\int_0^T r_s ds} \mid \mathcal{F}_t] = \frac{1}{\zeta_t} \mathbb{E}_{\mathbb{P}}[\zeta_T \mid \mathcal{F}_t],$$
where we define $\zeta_t := Z_t e^{-\int_0^t r_s ds}$ to be the discounted density process. The process ζ is called *state price density* (or *pricing kernel* or *stochastic discount factor*). In a state price density model, one directly models the dynamics of ζ (under $\mathbb{P}$!). Polynomial jump-diffusions form a natural basis for such models.

- **Market models:** In reality, ZCBs exist only for finitely many maturities $T \in \{T_1, \ldots, T_m\}$. A *market model* (or sometimes LIBOR market model) specifies dy-

[2] The notion of a term structure is more general however. For instance, if $C(t,T,K)$ denotes the price at time t of a call option with maturity T and strike price K, one often refers to $\{C(t,T,K)\colon T \geq t\}$ as the term structure of option prices.

namics for the *δ-period forward LIBOR rates* defined by

$$L(t,T_k) := \frac{1}{\delta}\left(\frac{P(t,T_k)}{P(t,T_k+\delta)} - 1\right), \quad t \le T_k, \quad k = 1,\dots,m,$$

for a given $\delta > 0$.

The above approaches have all been studied rather extensively. In this course we will look at short-rate models based on affine jump-diffusions (often called *affine term structure models*), and state price density models based on polynomial jump-diffusions (often called *polynomial term structure models*).

5.7 Affine short-rate models

Fix a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge 0}, \mathbb{Q})$ satisfying the usual conditions. Let X be an affine jump-diffusion on $E \subseteq \mathbb{R}^d$ with coefficients (b, a, ν) and generator $\mathcal{G}$. Recall from Proposition 3.23 that the coefficients are affine in x,

$$b(x) = b_0 + \sum_{i=1}^d x_i b_i, \qquad a(x) = a_0 + \sum_{i=1}^d x_i a_i, \qquad \nu(x, \,\cdot\,) = \nu_0 + \sum_{i=1}^d x_i \nu_i.$$

An *affine short rate model* is now obtained by taking $\mathbb{Q}$ to be the pricing measure, and define the short rate process by

$$r_t := \gamma^\top X_t$$

for some $\gamma \in \mathbb{R}^d$. For later use, we also introduce the integrated short rate process,

$$R_t := \int_0^t r_s ds.$$

The following result shows how to calculate bond prices in an affine short-rate model.

Proposition 5.22. *Assume (A, B) is a solution of the following system of ODEs:*

$$A'(\tau) = b_0^\top B - \frac{1}{2} B(\tau)^\top a_0 B(\tau) - \int_{\mathbb{R}^d} (e^{-B(\tau)^\top \xi} - 1 - B(\tau)^\top \xi)\nu_0(d\xi), \qquad A(0) = 0,$$

$$B_i'(\tau) = b_i^\top B - \frac{1}{2} B(\tau)^\top a_i B(\tau) - \int_{\mathbb{R}^d} (e^{-B(\tau)^\top \xi} - 1 - B(\tau)^\top \xi)\nu_i(d\xi) + \gamma_i, \quad B_i(0) = 0.$$

Then the bond prices are given by

$$P(t,T) = e^{-A(T-t) - B(T-t)^\top X_t}, \quad t \le T,$$

provided the right-hand side is bounded by e^{R_t} for all $t \le T$.

Proof. One can prove this result by repeating the argument leading to the affine transform formula (Theorem 3.26). Alternatively, one can reduce the problem to a direct application of that theorem, which is what we will do here. It constitutes another example of the augmentation technique used in connection with the Stein and Stein model; see Proposition 5.14.

We now start the proof. Define the augmented process $\widetilde{X} = (X, R)$. Note that $dR_t = \gamma^\top X_t dt$. It follows that $\widetilde{X}$ is an AJD on $E \times \mathbb{R}$ with coefficients $(\widetilde{b}, \widetilde{a}, \widetilde{\nu})$ given as follows, where we write $\widetilde{x} = (x, x_{d+1})$ for the column vector representing a generic point in $\mathbb{R}^{d+1}$.

$$\widetilde{b}(\widetilde{x}) = \underbrace{\begin{pmatrix} b_0 \\ 0 \end{pmatrix}}_{=:\widetilde{b}_0} + \sum_{i=1}^d x_i \underbrace{\begin{pmatrix} b_i \\ \gamma_i \end{pmatrix}}_{=:\widetilde{b}_i} + x_{d+1} \times \underbrace{0}_{=:\widetilde{b}_{d+1}}$$

and

$$\widetilde{a}(\widetilde{x}) = \underbrace{\begin{pmatrix} a_0 & 0 \\ 0 & 0 \end{pmatrix}}_{=:\widetilde{a}_0} + \sum_{i=1}^d x_i \underbrace{\begin{pmatrix} a_i & 0 \\ 0 & 0 \end{pmatrix}}_{=:\widetilde{a}_i} + x_{d+1} \times \underbrace{0}_{=:\widetilde{a}_{d+1}} .$$

To derive the jump coefficient $\widetilde{\nu}(\widetilde{x}, d\widetilde{\xi})$, observe first that R cannot jump. Thus $\widetilde{X}$ has a jump of size $\widetilde{\xi} = (\xi, \xi_{d+1})$ at a time t if and only if $\xi_{d+1} = 0$ and X has a jump of size ξ at the time t. From this we infer that

$$\widetilde{\nu}(\widetilde{x}, d\widetilde{\xi}) = \nu(x, d\xi)\mathbf{1}_{\{\xi_{d+1}=0\}} = \underbrace{\nu_0(d\xi)\mathbf{1}_{\{\xi_{d+1}=0\}}}_{=:\widetilde{\nu}_0(d\widetilde{\xi})} + \sum_{i=1}^d x_i \underbrace{\nu_i(d\xi)\mathbf{1}_{\{\xi_{d+1}=0\}}}_{=:\widetilde{\nu}_i(d\widetilde{\xi})} + x_{d+1} \times \underbrace{0}_{=:\widetilde{\nu}_{d+1}(d\widetilde{\xi})} .$$

This establishes that $\widetilde{X}$ is an AJD on $E \times \mathbb{R}$, along with the form its coefficients.

Now, the right-hand side of the Riccati equations appearing in the affine transform formula for $\widetilde{X}$ are given by functions $\widetilde{R}_i(\widetilde{u})$, $\widetilde{u} = (u, u_{d+1})$, where

$$\begin{aligned} \widetilde{R}_i(\widetilde{u}) &= \widetilde{b}_i^\top \widetilde{u} + \widetilde{u}^\top \widetilde{a}_i \widetilde{u} + \int_{\mathbb{R}^{d+1}} (e^{\widetilde{u}^\top \widetilde{\xi}} - 1 - \widetilde{u}^\top \widetilde{\xi}) \widetilde{\nu}_i(d\widetilde{\xi}) \\ &= \begin{cases} 0, & i = d+1, \\ \gamma_i u_{d+1} + b_i^\top u + \frac{1}{2} u^\top a_i u + \int_{\mathbb{R}^d} (e^{u^\top \xi} - 1 - u^\top \xi) \nu_i(d\xi) & \\ \qquad = \gamma_i u_{d+1} + R_i(u), & i = 1, \dots, d, \\ R_0(u), & i = 0. \end{cases} \end{aligned}$$

Here $R_i(u)$, $i = 0, \dots, d$, are the corresponding functions for X. (Warning: do not confuse these functions with the integrated short-rate process R!).

Consider the solution (A, B) of the given system of ODEs. Define functions

$$\widetilde{\phi} := -A, \qquad \widetilde{\psi} := \begin{pmatrix} -B \\ -1 \end{pmatrix}.$$

Then

$$\begin{aligned} \widetilde{\phi}' &= -A' = R_0(-B) = \widetilde{R}_0(\widetilde{\psi}), \\ \widetilde{\psi}_i' &= -B_i' = R_i(-B) - \underbrace{\gamma_i}_{=-\gamma_i \widetilde{\psi}_{d+1}} = \widetilde{R}_i(\widetilde{\psi}), \qquad\qquad i = 1, \dots, d, \\ \widetilde{\psi}_{d+1}' &= 0 = \widetilde{R}_{d+1}(\widetilde{\psi}). \end{aligned}$$

Furthermore, $\widetilde{\phi}(0) = 0$ and $\widetilde{\psi}(0) = (0, -1)$. Finally,

$$\widetilde{\phi}(T-t) + \widetilde{\psi}(T-t)^\top \widetilde{X}_t = -A(T-t) - B(T-t)^\top X_t - R_t,$$

which is nonpositive by assumption. By the affine transform formula (Theorem 3.26) for $\widetilde{X}$ with $\widetilde{u} = (0, -1)$, we thus get

$$\mathbb{E}_{\mathbb{Q}}[e^{-R_T} \mid \mathcal{F}_t] = \mathbb{E}_{\mathbb{Q}}[e^{\widetilde{u}^\top \widetilde{X}_T} \mid \mathcal{F}_t] = e^{\widetilde{\phi}(T-t) + \widetilde{\psi}(T-t)^\top \widetilde{X}_t} = e^{-A(T-t) - B(T-t)^\top X_t - R_t}.$$

Since $R_T - R_t = \int_t^T r_s ds$, it follows that

$$P(t,T) = \mathbb{E}_{\mathbb{Q}}[e^{-\int_t^T r_s ds} \mid \mathcal{F}_t] = e^{-A(T-t) - B(T-t)^\top X_t},$$

as desired. $\square$

Remark 5.23. The "$\leq e^{R_t}$" assumption was imposed to make Theorem 3.26 directly applicable. With more work it can be eliminated. This assumption basically corresponds to having $r_t \geq 0$, which is not always in agreement with reality: it is common to see negative interest rates! (Although this is a rather recent phenomenon.)

We now consider three classical examples of affine short-rate models. These are all one-factor models, where $d = 1$ and $\gamma = 1$, and thus $r_t = X_t$.

Example 5.24 (Vasiček model)**.** The short-rate is $\mathbb{R}$-valued and given by

$$dr_t = (b + \beta r_r)dt + \sigma dW_t$$

for some real parameters b, β, σ. The Riccati equations for (A, B) are

$$A' = bB - \frac{\sigma^2}{2}B^2, \qquad A(0) = 0,$$
$$B' = \beta B + 1, \qquad B(0) = 0.$$

These equations can be solved explicitly ($\longrightarrow$ exercise).

Example 5.25 (CIR model)**.** The short-rate is $\mathbb{R}_+$-valued and given by

$$dr_t = (b + \beta r_r)dt + \sigma\sqrt{r_t}dW_t$$

for some real parameters b, β, σ with $b \geq 0$. The Riccati equations for (A, B) are

$$A' = bB, \qquad A(0) = 0,$$
$$B' = \beta B - \frac{\sigma^2}{2}B^2 + 1, \qquad B(0) = 0.$$

Again, these equations can be solved explicitly. However, unlike the Vasiček model the equation for B is now quadratic, which makes solving it a little bit less trivial. It can however still be done (c.f. Exercise 3.28).

Example 5.26 (Hull–White model)**.** The short-rate is $\mathbb{R}$-valued and given by

$$dr_t = (b(t) + \beta r_r)dt + \sigma dW_t$$

for some real parameters β, σ and some deterministic function $b : \mathbb{R}_+ \to \mathbb{R}$ with suitable regularity properties. Note that this model falls outside the time-homogeneous framework developed in this course. Still, the same techniques can be used: Define

$$M_t := \mathrm{e}^{-A(t,T) - B(t,T)r_t - \int_0^t r_s ds}$$

for fixed T and all $t \in [0, T]$. Itô's formula yields

$$\begin{aligned}
\frac{dM_t}{M_t} &= (-\partial_t A(t,T) - \partial_t B(t,T)r_t)dt - B(t,T)dr_t - r_t dt + \frac{1}{2}B(t,T)^2 d\langle r\rangle_t \\
&= \Big((-\partial_t A(t,T) - b(t)B(t,T) + \frac{1}{2}\sigma^2 B(t,T)^2) + (-\partial_t B(t,T) - \beta B(t,T) - 1)r_t\Big)dt \\
&\quad - B(t,T)\sigma dW_t.
\end{aligned}$$

It follows that M is a local martingale if A and B satisfy the following *time-inhomogeneous Riccati equations*:

$$\begin{aligned}
\partial_t A(t,T) &= -b(t)B(t,T) + \frac{\sigma^2}{2}B(t,T)^2, && A(T,T)=0,\\
\partial_t B(t,T) &= -\beta B(t,T) - 1, && B(T,T)=0.
\end{aligned}$$

(Note that these equations have terminal conditions rather than initial conditions, and the time variable is calendar time rather than time-to-maturity. This also explains the discrepancy in sign compared to the Vasiček and CIR models.) If M is in fact a true martingale, we obtain as before that

$$\mathbb{E}_{\mathbb{Q}}[e^{-\int_0^T r_s ds} \mid \mathcal{F}_t] = \mathbb{E}_{\mathbb{Q}}[M_T \mid \mathcal{F}_t] = M_t = e^{-A(t,T)-B(t,T)r_t-\int_0^t r_s ds},$$

which gives the bond price formula

$$P(t,T) = e^{-A(t,T)-B(t,T)r_t}.$$

Why can it be useful to allow for time-dependent coefficients? The reason is that the forward curve at time $t=0$ can be matched exactly. More precisely, defining the time zero forward curve

$$f_0(T) := -\partial_T \log P(0,T),$$

as well as the function $g(T) := \frac{\sigma^2}{2\beta}(e^{\beta T}-1)^2$, one can show that

$$b(T) = f_0'(T) + g'(T) - \beta(f_0(T)+g(T)).$$

Since $f_0(\,\cdot\,)$ is in principle observed at time zero, $b(\,\cdot\,)$ pinned down once σ and β have been chosen. Put differently, it is always possible to choose $b(\,\cdot\,)$ so that the time zero forward curve produced by the model exactly matches the one observed in the market.

5.8 Polynomial term structure models

Fix a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ satisfying the usual conditions. The probability measure $\mathbb{P}$ is *not* supposed to be a risk-neutral measure. Instead, think of $\mathbb{P}$ as the real-world measure.[3] In this section we will consider state price density models.

[3] It is not necessary that $\mathbb{P}$ be the real-world measure. In principle $\mathbb{P}$ can represent any measure equivalent to the real-world measure, and this flexibility is sometimes important in application. For the purposes of this course, you can always think of $\mathbb{P}$ as being the real-world measure.

Definition 5.27. A *state price density* is a strictly positive semimartingale $\zeta = (\zeta_t)_{t\geq 0}$.

In a state price density model where ζ is given, the time t (model) price of a payoff C_T maturing at time T with $\zeta_T C_T \in L^1(\mathbb{P})$ is given by

$$C_t := \frac{1}{\zeta_t}\mathbb{E}[\zeta_T C_T \mid \mathcal{F}_t], \qquad t \leq T.$$

5.28. The key point with the above construction is that the resulting model is arbitrage-free in a suitable sense. To explain this, recall that the fundamental theorem of asset pricing states that a model is arbitrage-free if the *discounted prices* become (local) martingales under some equivalent measure. A version of the theorem gives the same conclusion in a situation where discounting is done not with the bank account process $e^{\int_0^t r_s ds}$, but with a general strictly positive semimartingale. In a state price density model, all price processes become martingales after discounting by $1/\zeta$; indeed, $C/(1/\zeta) = \zeta C$ is a martingale by construction.[4] In a model where a short rate r_t and pricing measure $\mathbb{Q}$ are given, the state price density is $\zeta_t = e^{-\int_0^t r_s ds}\frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{F}_t}$, as can be seen by an application of Bayes' rule.

5.29. As we are interested in ZCBs, we always require that $\zeta_t \in L^1(\mathbb{P})$ for all $t \geq 0$. We then have

$$P(t,T) = \frac{1}{\zeta_t}\mathbb{E}[\zeta_T \mid \mathcal{F}_t].$$

5.30. Let now X be a polynomial jump-diffusion on $E \subseteq \mathbb{R}^d$ with coefficients (b, a, ν) and generator $\mathcal{G}$. Fix $n \in \mathbb{N}$ and a polynomial $p_\zeta \in \mathrm{Pol}_n(E)$ with $p_\zeta > 0$ on E, as well as a constant $\alpha \in \mathbb{R}$. A *polynomial term structure model* is obtained by specifying the state price density as

$$\zeta_t := e^{-\alpha t} p_\zeta(X_t).$$

Several key quantities can be directly computed in this model. To do so, recall the setup from the moment formula: we fix a basis $1, h_1, \ldots, h_N$ for $\mathrm{Pol}_n(E)$ and define

$$H(x) := (h_1(x), \ldots, h_N(x)).$$

We let $G \in \mathbb{R}^{(1+N)\times(1+N)}$ be the corresponding matrix representation of $\mathcal{G}$. The moment formula states that

$$\mathbb{E}[p(X_T) \mid \mathcal{F}_t] = (1, H(X_t))e^{(T-t)G}\vec{p}$$

[4] Technically, one has that the NUPBR (No Unbounded Profit with Bounded Risk) property holds. This property is also known as NA1 (No Arbitrage of the First Kind).

holds for any polynomial $p(x) = (1, H(x))\vec{p}$ in $\mathrm{Pol}_n(E)$ with coordinate representation $\vec{p} \in \mathbb{R}^{1+N}$. From this we immediately get:

- **Bond prices:**

$$\begin{aligned} P(t,T) &= \frac{1}{\zeta_t}\mathbb{E}[\zeta_T \mid \mathcal{F}_t] \\ &= e^{-\alpha(T-t)}\frac{1}{p_\zeta(X_t)}\mathbb{E}[p_\zeta(X_T) \mid \mathcal{F}_t] \\ &= e^{-\alpha(T-t)}\frac{(1, H(X_t))e^{(T-t)G}\vec{p}_\zeta}{p_\zeta(X_t)} \end{aligned}$$

  Thus bond prices are fully explicit in a polynomial term structure model; there is not even a need to solve any Riccati equations.

- **Short rate:** The short rate is defined by $r_t := -\partial_T \log P(t,T)|_{T=t}$.[5] We then get

$$r_t = \alpha - \frac{(1, H(X_t))G\vec{p}_\zeta}{p_\zeta(X_t)} = \alpha - \frac{\mathcal{G}p_\zeta(X_t)}{p_\zeta(X_t)}.$$

  Thus the role of α is to adjust the level of short rate.

Remark 5.31. Note that both bond prices and short rate are rational functions of X_t, where the degree of the numerators is at most the degree of the denominators. Moreover, the denominators are strictly positive. In most cases, this leads to a bounded short rate. In practice, the interval of possible short rates,

$$\alpha - \sup_{x\in E}\frac{\mathcal{G}p_\zeta(x)}{p_\zeta(x)} \le r_t \le \alpha - \inf_{x\in E}\frac{\mathcal{G}p_\zeta(x)}{p_\zeta(x)},$$

is often bounded but rather wide.

5.32. Let us next briefly outline how option pricing is approached in a polynomial term structure model.

- **Bond options:** Consider the payoff $C_T := (P(T,T') - K)^+$. This is a European call option with maturity T and strike price K, written on a ZCB maturing at time

[5] As an exercise to motivate this definition, suppose a short rate r_t is given along with a pricing measure $\mathbb{Q}$. Thus $P(t,T) = \mathbb{E}_Q[e^{-\int_t^T r_s ds} \mid \mathcal{F}_t]$. Assuming the short rate is bounded, show that $r_t = -\partial_T P(t,T)|_{T=t}$.

$T' > T$. The model price at time t of such an option is

$$\begin{aligned} C_t &:= \frac{1}{\zeta_t}\mathbb{E}[\zeta_T C_T \mid \mathcal{F}_t] \\ &= \frac{1}{\zeta_t}\mathbb{E}[(\zeta_T P(T,T') - \zeta_T K)^+ \mid \mathcal{F}_t] \\ &= \frac{1}{\zeta_t}\mathbb{E}[q(X_T)^+ \mid \mathcal{F}_t], \end{aligned}$$

where $q(X_T)$ is polynomial in X_T. This is because $\zeta_T = e^{-\alpha T} p_\zeta(X_T)$ is polynomial in X_T by construction, and $\zeta_T P(T,T') = \mathbb{E}[\zeta_{T'} \mid \mathcal{F}_T]$ is polynomial in X_T due to the moment formula. Moreover, the degree of q is at most the degree of p_ζ.

- **Options on bond portfolios:** The payoff is now of the form

$$C_T := \Big(\sum_{i=1}^m c_i P(T,T_i) - K\Big)^+$$

for some coefficients $c_1, \ldots, c_m$ and maturities $T_1, \ldots, T_m$ of the underlying bonds, where $T_i > T$ for $i = 1, \ldots, m$. Again we obtain

$$\begin{aligned} C_t &:= \frac{1}{\zeta_t}\mathbb{E}[\zeta_T C_T \mid \mathcal{F}_t] \\ &= \frac{1}{\zeta_t}\mathbb{E}\Big[\Big(\sum_{i=1}^m c_i \zeta_T P(T,T_i) - K\Big)^+ \mid \mathcal{F}_t\Big] \\ &= \frac{1}{\zeta_t}\mathbb{E}[q(X_T)^+ \mid \mathcal{F}_t], \end{aligned}$$

where q is a polynomial (not the same one as before of course!) with degree at most that of p_ζ. The pricing problem for options on bond portfolios is therefore of the same complexity as in the single bond case. This stands in stark contrast to affine short-rate models.

- We have thus reduced the option pricing problem to the question of how to compute $\mathbb{E}[q(X_T)^+ \mid \mathcal{F}_t]$ for q a polynomial. This is done on a case-by-case basis. For example, if X happens to be not only a PJD, but even an AJD, then Fourier pricing can be used similarly as in Section 5.3. If X is not an AJD, one can instead use the polynomial expansion method outlined in Section 5.5.

We end with two simple examples of one-factor polynomial term structure models.

Exercise 5.33. Let X be a geometric Brownian motion and let the state price density be a first degree polynomial in X,

$$\begin{aligned}\frac{dX_t}{X_t} &= \mu dt + \sigma dW_t, \qquad X_0 > 0,\\ \zeta_t &= e^{-\alpha t}(1 + X_t),\end{aligned}$$

where $\mu \in \mathbb{R}$, $\sigma > 0$, W is a standard Brownian motion, and $\alpha \in \mathbb{R}$. This specifies a polynomial term structure model. Show that the ZCB prices and short rate are given by

$$\begin{aligned}P(t,T) &= e^{-\alpha(T-t)}\frac{1 + X_t e^{\mu(T-t)}}{1 + X_t},\\ r_t &= \alpha - \mu\frac{X_t}{1 + X_t}.\end{aligned}$$

Exercise 5.34. Consider instead the model

$$\begin{aligned}dX_t &= \kappa(\theta - X_t)dt + \sigma dW_t,\\ \zeta_t &= e^{-\alpha t}(1 + X_t^2),\end{aligned}$$

where $\kappa, \theta \in \mathbb{R}$, $\sigma > 0$, W is a standard Brownian motion, and $\alpha \in \mathbb{R}$. Again this specifies a polynomial term structure model. Compute the bond prices and short rate.

Bibliography

Damien Ackerer, Damir Filipović, and Sergio Pulido. The Jacobi Stochastic Volatility Model. *Social Science Research Network Working Paper Series*, May 2016. URL http://ssrn.com/abstract=2782486.

Richard F. Bass. The Doob-Meyer decomposition revisited. *Canad. Math. Bull.*, 39(2): 138–150, 1996. ISSN 0008-4395. doi: 10.4153/CMB-1996-018-8. URL http://dx.doi.org/10.4153/CMB-1996-018-8.

Mathias Beiglböck, Walter Schachermayer, and Bezirgen Veliyev. A short proof of the Doob-Meyer theorem. *Stochastic Process. Appl.*, 122(4):1204–1209, 2012. ISSN 0304-4149. doi: 10.1016/j.spa.2011.12.001. URL http://dx.doi.org/10.1016/j.spa.2011.12.001.

John C. Cox, Jonathan E. Ingersoll, and Stephen A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407, 1985.

Damir Filipović and Martin Larsson. Polynomial jump-diffusion models. *arXiv:1711.08043*, 2017.

Julie Lyng Forman and Michael Sørensen. The Pearson diffusions: a class of statistically tractable diffusion processes. *Scand. J. Statist.*, 35(3):438–465, 2008. ISSN 0303-6898. doi: 10.1111/j.1467-9469.2007.00592.x. URL http://dx.doi.org/10.1111/j.1467-9469.2007.00592.x.

Steven L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343, 1993. ISSN 08939454, 14657368. URL http://www.jstor.org/stable/2962057.

Jean Jacod and Albert N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 2003. ISBN 3-540-43932-3. doi: 10.1007/978-3-662-05265-5. URL http://dx.doi.org/10.1007/978-3-662-05265-5.

P. A. Meyer. A decomposition theorem for supermartingales. *Illinois J. Math.*, 6:193–205, 1962. ISSN 0019-2082. URL http://projecteuclid.org/euclid.ijm/1255632318.

P.-A. Meyer. Decomposition of supermartingales: the uniqueness theorem. *Illinois J. Math.*, 7:1–17, 1963. ISSN 0019-2082. URL http://projecteuclid.org/euclid.ijm/1255637477.

Philip E. Protter. *Stochastic integration and differential equations*, volume 21 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2005. ISBN 3-540-00313-4. doi: 10.1007/978-3-662-10061-5. URL http://dx.doi.org/10.1007/978-3-662-10061-5. Second edition. Version 2.1, Corrected third printing.

Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999. ISBN 3-540-64325-7. doi: 10.1007/978-3-662-06400-9. URL http://dx.doi.org/10.1007/978-3-662-06400-9.

Elias M. Stein and Jeremy C. Stein. Stock price distributions with stochastic volatility: An analytic approach. *Review of Financial Studies*, 4:727–752, 1991. URL http://rfs.oxfordjournals.org/.