
UNIVERSITÀ
DEGLI STUDI
DI PADOVA

*Sede Amministrativa:* Università degli Studi di Padova
*Dipartimento di:* Matematica

*Scuola di dottorato di ricerca in:* Scienze Matematiche
*Indirizzo:* di Matematica Computazionale – XXV ciclo

---

# Aspects of Affine Models in the Pricing of Exotic Options and in Credit Risk

---

**Il direttore della Scuola**
*Ch.mo Prof. Paolo Dai Pra*

**Il coordinatore di indirizzo**
*Ch.mo Prof. Michela Redivo Zaglia*

**Il supervisore**
*Ch.mo Prof. Wolfgang Runggaldier*

**Il dottorando**
*Juan Miguel Montes*

---

# Contents

# Sommario

Le strutture a termine affine con diffusione a salti (AJTSMs) stanno recentemente ricevendo molta attenzione in finanza matematica, perché spesso è semplice analizzare le funzioni di distribuzione ad esse associate. Questa tesi riguarda tre diversi aspetti della finanza matematica, applicati su certe classi di AJTSMs.

Il primo aspetto riguarda il problema del prezzaggio, nel caso particolare in cui il processo sottostante $X_t$ sia una Catena Markoviana a Tempo Continuo (CTMC). Per opzioni esotiche, dove il "claim", cioè il "payoff" del derivato è dipende dal tempo oppure dalle traiettorie, solitamente i prezzi devono essere stimati attraverso simulazioni di tipo Monte Carlo. Mostriamo che, quando si condiziona prima sul numero $N_{t,T}$ dei salti della catena, il calcolo di questa stima si semplifica. Viene proposta una ricorsione per calcolare il valore atteso del "payoff" scontato, dato $N_{t,T} = k$; in seguito si calcola il valore atteso del "payoff" rispetto alla distribuzione di $N_{t,T}$ attraverso un metodo Monte Carlo. Questo condizionamento comporta una riduzione della varianza. Presentiamo i risultati di vari test numerici, che indicano che, per diversi tipi di "claims", il metodo proposto supera spesso un semplice "vanilla" Monte Carlo.

Il secondo aspetto riguarda la calibrazione, cioè la stima dei parametri di un modello finanziario, dove il processo sottostante (una Catena Markoviana finita) è solo parzialmente osservabile tramite i prezzi corrotti del titolo. In questo lavoro, assumiamo che anche i salti del prezzo del titolo corrispondenti ai tempi dei salti della catena Markoviana siano osservabili. Questo è un caso particolare della classe di modelli trattati in [FR10b]. I loro parametri possono essere stimati mediante l'algoritmo "expectation-maximization" (EM), seguendo l'approccio di [EAM08], che, nel caso delle catene a tempo discreto, coinvolge il filtro di Kalman. Estendiamo questo approccio al caso CTMC, usando invece il filtro di Wonham. Il contributo principale di questa parte della tesi è l'approssimazione numerica dei filtri e degli "smoothers" dell'algoritmo EM. Confrontiamo i classici metodi di Eulero e di Milstein con una nuova strategia, simile a [PR10a], che chiamiamo "soluzione quasi-esatta" e che è anche collegata al metodo di "splitting-up" di [BGR90] e [Gla92]. Dimostriamo che tale schema ha un ordine di convergenza forte di almeno 0.5 e che pertanto è almeno tanto efficace quanto lo schema di Eulero. Presentiamo alcuni risultati numerici che indicano che, di fatto, in certi casi il nuovo metodo converge più velocemente di entrambi i metodi di Eulero e di Milstein.

Il terzo aspetto riguarda un quadro unificato per la modellazione del rischio di "equity" e "credit", con applicazioni alla gestione del rischio. Trattiamo un AJTSM di un'azione con un'unica discontinuità ("jump-to-default"), dove il tempo di fallimento dell'azione è un tempo aleatorio doppiamente stocastico con intensità determinata da un sottostante processo affine. Questo approccio permette una piena trattabilità analitica pur lasciando flessibilità nel definire le interazioni

tra il prezzo dell'azione fallibile, la volatilità stocastica e l'intensità del fallimento. Infine caratterizziamo tutte le misure di rischio neutrale che conservano la struttura affine del modello e mostriamo che sia la gestione del rischio che i problemi del prezzaggio possono essere trattati in modo efficiente passando a misure di sopravivenza appropriate. Come esempio, estendiamo il modello di volatilità stocastica di Heston considerando la possibilità di un "jump-to-default".

# Abstract

Affine jump-diffusion term structure models (AJTSMs) are recently receiving much attention in mathematical finance, because they often lead to a tractable analysis of the price distribution functions. This thesis concerns three aspects of mathematical finance, when applied to certain classes of AJTSMs.

The first aspect concerns the pricing problem in the special case when the underlying process $X_t$ is a Continuous-Time Markov Chain. For exotic options, where the claims are time or path dependent, prices can only be estimated by Monte-Carlo simulation, in most cases. We show that this computation is simplified by conditioning first on the number $N_{t,T}$ of the jumps of the chain. A recursion is proposed to compute the expected discounted payoff given $N_{t,T} = k$; Monte Carlo is then used to average out the result over the distribution of $N_{t,T}$. This leads to a variance reduction by conditioning. We present results of numerical tests which indicate that the method often outperforms plain vanilla Monte Carlo for different kinds of claims.

The second aspect concerns the calibration of a financial model by parameter estimation, when the underlying, a finite state Markov chain, is only partially observed through noisy asset prices. Here, we assume that the jumps of the asset price occurring at the jump-times of the Markov chain are observable as well. Such a model is a special case of the class of models treated in [FR10b]. Their parameter estimation can be addressed via the EM algorithm, following the approach by [EAM08] which, in the case of discrete-time chains, involves the Kalman filter. We extend this approach to to the case of CTMCs via the use of the Wonham filter. Our main contribution is the numerical approximation of the filters and smoothers in the EM algorithm. We compare the classical Euler and Milstein schemes to a new scheme, inspired by [PR10a], that we call a quasi-exact solution and is related to the splitting-up method of [BGR90] and [Gla92]. We prove that such a scheme is of strong convergence order at least 0.5, hence it performs at least as well as the Euler scheme. We present numerical evidence indicating that in fact, in certain cases the method outperforms both the Euler and the Milstein scheme.

The third aspect concerns a unified framework for equity and credit risk modeling, with applications to risk management. Here we treat an affine jump-diffusion model with a single jump-to-default, where the default time is a doubly stochastic random time with intensity driven by an underlying affine factor process. This approach allows for flexible interactions between the defaultable underlying asset price, its stochastic volatility and the default intensity, while maintaining full analytical tractability. We characterize all risk-neutral measures which preserve the affine structure of the model and show that risk management as well as pricing problems can be dealt with efficiently by shifting to suitable survival measures. As an example, we consider a jump-to-default extension of the Heston stochastic volatility model.

# Acknowledgments

At the end of this thesis writing and almost at the end of the entire Ph.D. endeavor, I cannot help but reflect on what has taken place and be moved to borrow from the African proverb which says that *"It takes a village to raise a child"*. Indeed child-rearing and villages serve as good metaphors, respectively, to Ph.D. thesis writing and to the host of persons - other than the author himself - who advertently or inadvertently contributed in some way to the conceiving, giving birth, nourishing, sheltering, weening and disciplining of the new born runt, namely, the said thesis. It takes an entire cast of professors, academic advisers, colleagues, students, flatmates, friends, girlfriends, and family members to "raise" a Ph.D. thesis. The following acknowledgements in go to the various "villagers" to whom I owe my most heartfelt "Thanks"!

First and foremost, I am grateful to my thesis supervisor, Wolfgang Runggaldier, for his guidance all throughout my Ph.D. studies. Even at an early stage of my studies he encouraged and supported me to participate and present my preliminary results frequently in various summer schools, special semesters, and academic conferences. I am thankful to him for giving me the opportunity to work together on the joint paper with V. Prezioso which was to become Part 1 of this thesis, and also for advising me to pursue research with C. Fontana that, in turn, has culminated into the article serving as Part 3. Prof. Runggaldier has always been ready to lend his advice, whether in terms of practical concerns when I was a newly-arrived foreigner in Italy, or insightful hints while grappling with and unravelling anomalous numerical "bugs" during our research - these have many times lead to the correct solution and a better understanding of the problem. Last but not the least, I thank Prof. Runggaldier for his patience, for being at once gentle yet firm whenever it was needed to focus my perspective "stick-to-the-problem-at-hand". A true mentor, he has often imparted in me invaluable lessons that come from his many years of experience. For these it has been both an honor and pleasure to be his Ph.D. advisee.

Still from the academics' side, I am very grateful to Claudio Fontana, with whom it was a pleasure to co-write my first publication. Claudio, an exemplary person and researcher, has been also a great inspiration for me not only because of the calibre of his work, but also for his vision in future research, recognizing subtle but important open questions and being able to frame solutions rather elegantly; in a word, he inspires good "taste" in mathematical finance. It is both Prof. Runggaldier and Claudio to whom I owe my own acquired taste for stochastic filtering and credit risk applications.

I am very thankful also to Valentina Prezioso, whose own Ph.D. thesis - which she defended when I was just a freshman - served as the theoretical foundation for mine. While many of our interactions had subsequently been limited to e-mail, due to her busy life as a quant and and as a proud new wife and mother after her studies, nevertheless I have had the opportunity to be

her guest for a couple of research discussions in Milan, for which I remember her warmth and cheerfulness. On the other hand, I have read her Ph.D. thesis back-and-forth $N \to \infty$ times and it is due to her thesis that I have become a confirmed *appassionato* of Monte-Carlo and Markov Chain approximation. Thus, I am very grateful for her work and her encouragement.

Several other key professors at Padua have certainly also contributed to my formation. To Lorenzo Finesso I am grateful not only through the very interesting graduate course he taught on Information Theory, but also for the exploratory research discussions in my first year and introducing me to his work on Hidden Markov Models. While we had been subsequently unable to re-open those discussions for a possible collaborative work, they nevertheless have been an indirect source of the particular "flavors" or "color palette", so to say, that the Part 2 of my thesis came to have. I remain eager to have an opportunity to collaborate with him, hopefully soon. On the other hand, I am also grateful to Tiziano Vargiolou, who had also always facilitated our research groups' participation in conferences and who has taught the most graduate lectures I had attended (three semesters of Special Topics in Finance, covering advanced Arbitrage Pricing Theory, Dynamic Programming and Optimal Control on Markov Processes, and Monte-Carlo Methods in Finance). Thus it is to him that I also owe a lot of my specialized mathematical finance formation. In a special way I mention also Paolo Dai Pra, Markus Fischer, and Paolo Guiotto for the wonderful trilogy on Markov Semigroups, and to Francesco Caravenna for his excellent, essential course on Stochastic Analysis. Still special mention goes to colleagues who had lent a hand in different ways, big or small: Silvia, Gabriella, Alessandro, Cecilia, Giulia, Daniele, Federico, Marco, Francesco (and all from Torre Archimede, Room 712) and Christian, Thilo, Francesca, Sorin, Andreas, Jan, Hannes, Max and Jacopo (from LMU Munich). Special thanks goes to the very friendly and helfpul administrative staff and coordinators: Angela Pucca, Alessandro Lanza, Lorenzo Capanna, Cristina Destro, Loretta Dalla Costa, Katia Milan and Sara Fedel (Padua) Isabel Sousa von Blume, Ingrid Schehrer, Christina Broza, and Anna Warlimont (LMU Munich).

My European formation was not only limited to mathematical finance, but also all the aspects of living for the first time very far away from my home entailed. Here I acknowledge the friends who have let me live life fully while I was in Padua: all the guys of the Math departments' "calcetto-pd" football group 2011-2012 (Giovanni, Luong, Khai, Duong, Vittorio, Stefano, Marco, Giulio, Bien, *e chiunque altro mi dimentico!*) my dear flatmates at Centro Universitario (Francesco, Marianna, don Roberto, don Marco, Elisa, Yuko, Senthil, Barbara, Barbara, Matteo, Rino, Laura, Giovanna, Roberto, Fabrizio) at Galileo (Qin Qong, Kim Ha, Phan, Nataly), don Giorgio of San Gaetano, friends from Centro Riviera (Luigi, Luigi, don Alberto, Andrea), the Frezza family, the Opportuno family and the Bovassi family, and my *kababayans* in Padua (Cecile and Laarni Silva, Irene Casas, Allan Francisco and friends and their families).

My Padua years will always be my fondest thanks to the friends I spent them with: Eleni, Neeraja, Martino, Szabolcs, Marco, Sophie, Jonnabelle, Joel, Maura, Elisa and Vincenza. *Ognuno di voi mi avete dato motivi per sorridere e stare sorridente durante il mio quotidiano padovano. Vi ringrazio di cuore!* In a special way I thank Serena for her kindness, love, strength and support and I also extend my thanks to the Bellon family for their warmth and for welcoming me to their home. In many ways I owe to you and to your family the second-wind that helped me get through my final year.

Last but not the least, my heartfelt gratitude goes to Mommy, Papa, Crissy, Tusa, Meg, Belay and Gushi, for the countless favors they have done, giving their love and support from home, and visiting me and/or travelling together with me whenever they could. This project was possible because of you!

*Juan Miguel Montes*
*Padua, 23 January, 2014*

# Chapter 0

# Introduction

## 0.1 Introduction

This work consists of three parts, each of which relates to one of three important problem domains in mathematical finance, namely the pricing problem, the risk-management problem, and the calibration problem. We do not aim to provide an exhaustive survey of these problem domains, nor to present a unifying framework by which these three problems can be treated in synthesis within a single model. Rather, we have chosen to focus our attention only on one class of models, which is receiving increasing attention lately, namely the class of affine models with jumps. For each problem domain, we tackle different, specific aspects of this class which are, nevertheless interrelated.

The class of affine models is a popular choice for many financial applications, especially in tackling the pricing problem for the term structure of interest rates. The models in this class are widely regarded as highly computationally tractable and are able to capture empirical features of given financial time series. Affine models have also seen applications beyond interest rates, for example in stochastic volatility option pricing and in credit risk models. For comprehensive treatments of affine models, see [DK96], [DFS03], [DF09], as well as the book [Fil09].

Most of the literature on affine term-structure models assume a diffusion model or a jump-diffusion model for the interest rate. To illustrate with a simple example for the case of diffusions, in [DK96], the authors propose a model for the "short rate" process $r_t$ under which $r_t$ is a function $r(X_t)$ of a process $X_t \in D \subset \mathbb{R}^n$ that is a solution to the SDE

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t$$

$$X_0 \in D \subset \mathbb{R}^n, \quad t \geq 0.$$

Here $\mu$ and $\sigma\sigma^\top$ are assumed to be affine functions of $X$, that is $\mu(X) = a + bX$ and $\sigma\sigma^\top = \mathcal{A} + \mathcal{B} \cdot X$, for $a$, $b$, $\mathcal{A}$, and $\mathcal{A}$ constant parameters of the appropriate (scalar / vector) dimensions, satisfying certain conditions. The authors of [DK96] show that the yield then has the form

$$Y(t,T) := \log P(t,T) = \alpha(T-t) + \beta(T-t) \cdot X_t,$$

where $P(t,T)$ denotes the zero coupon bond price at $t$ maturing at $T$.

Within the general class of affine models also belong models for which $r(X_t)$ is a pure-jump process. An example is the case when $X$ is a Finite State, Continuous Time Markov Chain (CTMC). The affine property of $r(X_t)$ is discussed in more detail in [Pre10] and in the paper [BKR97] concerning jump-diffusions, of which CTMC's are a special case.

Concerning the pricing of derivatives in the context of CTMCs, some references are the papers [Nor03], [Nor05], [PR10b], and the original work [MPR13]. In Part I we present [MPR13] where we tackle the pricing problem for generic, possibly exotic derivative payoffs, when the underlying process is a CTMC; this represents an alternative to the techniques presented in [Nor03], [Nor05] and is an extension of [PR10b].

On the other hand the paper [EHJ00] deals with a CTMC term-structure model in the context of model calibration by parameter estimation. In Part II of the present work, we also deal with the calibration problem for derivatives with CTMC payoffs. Our approach is an extension of [EHJ00] to a continuous-time filter setup, where we assume that only continuous noise corrupted time-series of prices are observable and the calibration is performed by combined Wonham filtering and Expectation-Maximization (EM). The contribution of our work is to propose quasi - exact solutions to the Wonham Filters via strong approximation schemes for the corresponding Zakai SDEs. We provide results on the strong order of such schemes, an algorithm for serial filtering due to the presence of jumps in the observation process, and some numerical results on simulated data.

In the third part, we present the original paper [FM14] which deals with a risk-management framework that jointly models equity and credit-type risks for affine-diffusion equity price processes of stocks that may default. Our proposed risk-management framework is able to also incorporate information about derivatives prices in a structure-preserving way, via the risk-premium. While in this third part we depart from the CTMC context of Part I and Part II, nevertheless we are still treating affine processes with jumps in so far as the default event represents a single jump-process added to the state dynamics.

# Part I

# Pricing Aspects

# Chapter 1

# Monte Carlo Variance Reduction by conditioning for pricing when the Underlying Dynamics follow a Continuous-Time Finite State Markov Process

## 1.1 Introduction

Traditional pricing formulae concern market models either in continuous or discrete time. There are however situations, in which it is more natural to model the evolution of the underlying as a continuous time Markov chain (CTMC) that combines features of both, continuous as well as discrete time models; in fact, while being a model in continuous time, it has an embedded Markov chain that corresponds to a discrete time one. A CTMC has the following relevant features:

**i)** it jumps at random points in time;

**ii)** the number of jumps in a fixed time interval is random.

Market models based on CTMC appear to be relevant in the following specific situations:

**a)** One such situation arises in the pricing of bonds and interest rate derivatives, where the "underlying" can be considered as given by the short rate of interest. This covers, similarly, all those situations where the "underlying" is a rate (e.g. exchange rate) or an intensity (e.g. default intensity). The evolution of a rate or intensity is generally modeled as a diffusion process, but the actual evolution may present jumps. Continuous-time term structure models that allow for jumps have already been considered in the literature and we limit ourselves to mention here just a couple of them. For the case of jump-diffusions, the article [BKR97] illustrates how, by assuming an appropriate affine structure, the bond price can be expressed in terms of a system of ODEs. This approach is theoretically interesting, but it does not consider derivative pricing and turns out to be difficult to implement in practice,

especially in the multivariate case. For a more general Levy driven interest rate model, the article [EK06] considers also Caps and Swaptions and the authors obtain explicit analytic solution formulae in the scalar case which require however rather sophisticated mathematical tools; moreover, their numerical results do not concern the prices as such, which is our main goal. While less general than jump-diffusions or Levy models, CTMCs still retain the basic features of the actual evolution of rates and allow to obtain explicit pricing formulae by which to compute numerical values. This has been shown for simple claims and time homogeneous CTMC in [PR10b], where bond prices, Caps and Swaptions are shown to be particular cases or linear combinations of what is called there a "Prototype product" that is an analog here of Arrow-Debreu prices.

**b)** A further more general situation, where CTMCs appear to be natural models, arises whenever one considers small time scales on which asset prices vary by tick sizes at random times in reaction to trading or to the arrival of significant new information. With respect to a), here the jumps are in general much more frequent.

In this chapter we shall consider a generic underlying, of which the evolution is given by a CTMC and that may be multivariate and/or time inhomogeneous. We derive a method that allows for an explicit computation also of path dependent claims. It is a mixture of an analytic expression and a Monte Carlo (MC) simulation and corresponds to an MC method with variance reduction by conditioning. The basic idea, illustrated for the moment for the case of simple claims, is as follows.

Denote by $X_t$ the underlying, evolving as a CTMC with infinitesimal generator $Q$ (below we shall denote by the same symbol also the transition intensity matrix that is equivalent to the generator, but has zeroes on the main diagonal) and let $H = H(X_T)$ be, for the moment, a simple claim with maturity $T$. Let $\tilde{P} \sim P$ be an equivalent (to the physical measure $P$) martingale measure that is used for pricing and that will typically result from a calibration to the market. The arbitrage-free price at $t < T$ of the claim $H$, when $X_t = x^i$ is then

$$\Pi_i(t) = E^{\tilde{P}} \left\{ \exp\left[ -\int_t^T r_s ds \right] H \mid X_t = x^i \right\} \tag{1.1.1}$$

A theory of financial markets, when the underlying follows a time homogeneous CTMC has been developed in [Nor03] according to which the price $\Pi_i(t)$ of a simple claim $H$ can be computed as

$$\Pi_i(t) = \left[ \exp\{(Q - R)(T - t)\} H \right]_i \tag{1.1.2}$$

where $[z]_i$ denotes the $i-$th component of the vector $z$, $R$ is the diagonal $(N \times N)-$matrix with elements $r^i$ $(i = 1, \cdots, N)$ and $Q$ is assumed here to be time homogeneous.

As mentioned above, the main purpose of the present study is to extend the basic theory and to provide an efficient semianalytic method to compute the price $\Pi_i(t)$ for these extensions, for which the explicit formula (1.1.2) cannot be applied or is difficult to apply. The extension concerns basically path dependent claims that in many cases can be reduced to simple claims by augmenting the dimension of the underlying. In some cases this may lead to a time homogeneous, although multivariate, underlying; in many cases however one ends up with a time inhomogeneous underlying. We shall also consider barrier options that lead to an intrinsically time inhomogeneous

situation. In the time inhomogeneous case the explicit formula (1.1.2) cannot be applied and it may become cumbersome to apply it also in the time homogeneous case when the underlying is multivariate. In all these more general cases one can always resort to a Monte Carlo (MC) simulation to compute the expectation in (1.1.1). As it will appear also from the numerical results below, a plain MC approach has various drawbacks: in addition to a possibly large variance, it may also lead to biased results, unless one performs an extremely large number of simulation runs. The MC-based approach proposed in this chapter and that we shall call "conditional MC" approach, results from first conditioning on the number $N_{t,T}$ of transitions of the chain $X_t$ between $t$ and $T$. It results from rewriting the expression in (1.1.1) as

$$\begin{aligned}\Pi_i(t) &= E^{\tilde{P}}\left\{e^{-\int_t^T r_s ds} H(X_T) \mid X_t = i\right\} \\ &= E^{\tilde{P}}\left\{E^{\tilde{P}}\left\{e^{-\int_t^T r_s ds} H(X_T) \mid N_{t,T}, X_t = i\right\} \mid X_t = i\right\}\end{aligned} \tag{1.1.3}$$

We shall show that the inner expression allows for an explicit analytic computation also in the case of claims that are not simple claims and this is by itself one of the other contributions of the present work. In this way there remains to possibly simulate only $N_{t,T}$ (an explicit, but cumbersome formula for computing the distribution of $N_{t,T}$ can be found in [PR10b] for $X_t$ scalar). The fact that a fair portion of what is computed by simulation in plain MC is here computed explicitly, makes it intuitively clear that one can thus obtain more precise results.

The outline of this chapter is as follows. In section 1.2 we describe more specifically our CTMC market model. In section 1.3 we show how to reformulate conveniently path dependent claims with a CTMC underlying so that they can be priced according to what we call "Prototype product" and for which we describe our conditional MC approach in section 1.4. In section 1.5 we then show how to apply the conditional MC approach to the pricing of path dependent and barrier options. Finally, in section 1.6, we present numerical results and comparisons.

## 1.2 The model

Let $X_t$ be a CTMC with values in $\{x^1, \cdots, x^N\}$ and transition intensity matrix $Q$ that we assume for the moment to be time homogeneous. Below we shall occasionally identify $x^i$ with $i$, ($i = 1, \cdots, N$). By a transition intensity matrix $Q$ we mean here the matrix, where the off-diagonal elements $q_{i,j}$ represent the transition rates from state $i$ to state $j$, while the diagonal elements $q_{i,i}$ are equal to zero. This is different from the infinitesimal generator matrix, where $q_{i,i} = -\sum_{j\neq i} q_{i,j}$, while $q_{i,j}$ remain the same for $i \neq j$ (notice that the information content is the same in both types of matrices).

For a given maturity $T$ let there be given a simple claim of the form

$$H = H(X_T) = \left[H(x^1), \cdots, H(x^N)\right]' \quad \text{where } ' \text{ denotes transposition}$$

and which, given the finite state assumption for $X_t$, can be represented as a vector.

Denoting by $\tau_n$ the random time at which the $n-$th transition of the chain $X_t$ takes place, put, for simplicity, $X_n := X_{\tau_n}$ so that $X_s = X_n$ for $s \in [\tau_n, \tau_{n+1})$.

Assume also that the short rate of interest $r_t$ is related to the underlying $X_t$ in the sense that $r_t = r(X_t)$ so that $r_t$ can undergo a change only at the time points $\tau_n$ and we let $r_n := r_{\tau_n} = r(X_n) \in \{r^1, \cdots, r^N\}$.

Starting from the pricing formula (1.1.1) and denoting by $N_t := \sup\{n \mid \tau_n \leq t\}$ the number of transitions of $X_t$ up to a given time $t$ so that $N_{t,T} = N_T - N_t$, one may write

$$\Pi_i(t) = E^{\tilde{P}}\Big\{\exp[r_t(t-\tau_{N_t}) \exp\left[-\sum_{i=N_t}^{N_T-1} r_i(\tau_{i+1}-\tau_i) - r_T(T-\tau_{N_T})\right] H(X_T) \mid X_t = i\Big\}$$

$$= \exp[r_t(t-\tau_{N_t})\, E^{\tilde{P}}\left\{exp\left[-\sum_{i=N_t}^{N_T-1} r_i(\tau_{i+1}-\tau_i) - r_T(T-\tau_{N_T})\right] H(X_T) \mid X_t = i\right\} \tag{1.2.1}$$

where we have used the fact that $\tau_{N_t}$ is known at time $t$ and so $\exp[r_t(t-\tau_{N_t})$ can be taken as being deterministic. As a consequence, we may without loss of generality assume $t = \tau_{N_t}$ and consider the computation of

$$V_{H,t,T}(X_t) = E^{\tilde{P}}\left\{\exp\left[-\sum_{i=N_t}^{N_T-1} r_i(\tau_{i+1}-\tau_i) - r_{N_T}(T-\tau_{N_T})\right] H(X_T) \mid X_t\right\} \tag{1.2.2}$$

where $V_{H,t,T}(X_t)$ denotes the vector with components $V_{H,t,T}(X_t)\mathbf{1}_{\{X_t=x^i\}}$ that occasionally we shall also denote by $V_{H,t,T}(X_t)_{|X_t=x^i}$ and where

$$H(\cdot) := H_0(\cdot) = \sum_{i=1}^{N} w_i^0 \mathbf{1}_{\{\cdot=x^i\}},\ x^i \in E,\ \ w_i^0 \in \mathbb{R} \tag{1.2.3}$$

with $w_i^0$ representing the value $H(x^i)$.

A plain MC approach consists now in simulating the successive transition times $\tau_n$ of the chain $X_t$ and the values $X_n$ of $X_t$ at $\tau_n$ and then averaging over the values obtained in each simulation run for the argument in the expectation of the right hand side of (1.2.2). To be precise, consider for a moment again a time homogeneous chain with transition intensity matrix $Q$. Putting

$$q_i := \sum_{i \neq j} q_{i,j}, \tag{1.2.4}$$

one has that, if $X_n = X_{\tau_n} = x^i$, then the inter-jump times $\tau_{n+1} - \tau_n$ are exponentially distributed with parameter $q_i$ and the transition probability of the embedded chain, namely the probability that $X_{\tau_{n+1}} = x^j \neq x^i$ is given by

$$p_{i,j} = \frac{q_{i,j}}{q_i} \tag{1.2.5}$$

implying, as it should be, that $p_{i,i} = 0$. Given these values for $q_i$ and $p_{i,j}$, one can then simulate the successive values of $\tau_n$ and of the corresponding $X_n$.

Contrary to the plain MC approach, by our semi-analytic approach only the number of jumps is being simulated while the remaining calculations are performed on the basis of an explicit analytic formula.

We close this section by mentioning the notation that we shall use for the time inhomogeneous and multivariate case. In the time inhomogeneous case, instead of a fixed transition intensity

matrix $Q$, we shall consider a sequence $Q(n)$ where $n$ refers to the $n-$th transition time $\tau_n$ of the underlying $X_t$. For the multivariate case let us take the special case of a bivariate CTMC $(X_t, Y_t)$ with $X_t \in \{x^1, \cdots, x^N\}$, $Y_t \in \{y^1, \cdots, y^M\}$. Letting $\tau_n$ denote the $n-$th jump time of the pair $(X_t, Y_t)$, assume again that the short rate $r_t$ changes only at the time points $\tau_n$, i.e. assume that $r_n := r_{\tau_n} = r(X_n, Y_n)$ with $(X_n, Y_n) = (X_{\tau_n}, Y_{\tau_n})$. For the (time inhomogeneous) multivariate transition intensity matrix we shall put

$$Q(n) = \left\{q_{(i,h),(j,k)}(n)\right\}_{\substack{i,j=1,\cdots,N \\ h,k=1,\cdots,M}} \tag{1.2.6}$$

Notice that the multivariate case may arise not only when the underlying itself is multivariate, but also when the underlying is scalar and one has a path dependent claim that can be transformed into a simple claim by increasing the dimension of the underlying as we shall discuss in the next section. Furthermore, the multivariate case may also arise e.g. in defaultable bond pricing, where the underlying is $X_t = r_t$ and the price is given by

$$\Pi(t) = \mathbf{1}_{\{\tau > t\}} E^{\tilde{P}} \left\{ \exp\left[ -\int_t^T (r_s + \lambda_s)\, ds \right] \mid \mathcal{F}_t \right\} \tag{1.2.7}$$

with $\tau$ denoting the default time and $\lambda_t$ the default intensity. The pair $(r_t,\ \lambda_t)$ may then be taken to form a bivariate CTMC, i.e. $X_t = r_t, , Y_t = \lambda_t$. To take correlation into account, one may put $r_t = r(X_t, Z_t)$ and $\lambda_t = \lambda(Y_t, Z_t)$ with $X_t, Y_t, Z_t$ three independent CTMCs.

## 1.3 Path dependent claims

Here we show how, for an underlying evolving as a CTMC, some of the standard path dependent claims can be transformed into simple claims with an underlying evolving as a multivariate (possibly time inhomogeneous) CTMC. In this section we shall limit ourselves to lookback and Asian options but later on we shall also consider barrier options, discussing in particular knock-out options, since for this case formula (1.1.2) can clearly not be applied but our approach turns out to be relatively simple to apply.

### 1.3.1 Lookback options

Given is an underlying CTMC $X_t \in \{x^1, \cdots, x^N\}$ where for simplicity we shall occasionally identify $x^i$ with $i$, $(i = 1, \cdots, N)$. Consider a claim of the form

$$H_T = \left(X_T - g(X_0^T)\right)^+ \tag{1.3.1}$$

where $g(\cdot)$ is a measurable function of the generic trajectory $X_0^t := (X_0, \cdots, X_t)$ for $t \le T$ of the process $X_t$ for which, recalling that $\tau_n$ denote the transition times of $X_t$, we assume

$$g(X_0^{\tau_n}) = G(X_{\tau_n}, g(X_0^{\tau_{n-1}})) \quad \text{for some measurable } G(\cdot,\cdot) \tag{1.3.2}$$

Put $Y_t := g(X_0^t)$ and notice that, for $t \in [0, T]$, the process $Y_t$ also takes a finite number of possible values; denote them by $h = 1, \cdots, M$. More importantly, $Y_t$ can make a transition only at the transition times of $X_t$ and it can be easily seen that $(X_t, Y_t)$ forms a bivariate CTMC.

Furthermore we have that $H_T = (X_T - Y_T)^+$ and we also recall that $(X_T, Y_T) = (X_{N_T}, Y_{N_T})$. To deal with the chain $(X_t, Y_t)$ and to price $H_T$, we need to derive the transition intensity matrix $Q$ for $(X_t, Y_t)$, for which we have

**Proposition 1.3.1.** *Given a chain $X_t$ with transition intensity matrix $Q = \{q_{i,j}\}_{i,j=1,\cdots,N}$, the chain $(X_t, Y_t)$ has*

$$\left\{q_{(i,h),(j,k)}(n)\right\}_{\substack{i,j=1,\cdots,N \\ h,k=1,\cdots,M}}$$

*with*

$$q_{(i,h),(j,k)} = q_{i,j}\mathbf{1}_{\{G(j,h)=k\}} \tag{1.3.3}$$

*Proof.* Recall first that (see (1.2.5)), if for a scalar CTMC $X_t$ the $Q-$matrix is $Q = \{q_{i,j}\}$, then the transition probabilities of the embedded chain $X_n$ are $p_{i,j} = \frac{q_{i,j}}{q_i}$ with $q_i = \sum_{j\neq i} q_{i,j}$ $(q_{i,i} = p_{i,i} = 0)$. Viceversa, given $p_{i,j}$, there are various possible $q_{i,j}$ that lead to the same $p_{i,j}$. They differ by the choice of $q_i$ since we have $q_{i,j} = q_i p_{i,j}$. Given that in our case $Y_t$ jumps exactly when $X_t$ does, we may put

$$q_{(i,h)} = \sum_{j,k} q_{(i,h),(j,k)} = q_i \quad \forall h = 1, \cdots, M \tag{1.3.4}$$

and notice that at a generic $\tau_n$ the process $X_t$ actually leaves the current state, while $Y_t$ may jump to itself. To conclude, it thus suffices to construct $p_{(i,h),(j,k)}$. Recalling that we had put $X_n = X_{\tau_n}, Y_n = Y_{\tau_n}$, we have

$$\begin{aligned}
p_{(i,h),(j,k)} &:= P\{X_{n+1} = j, Y_{n+1} = k \mid X_n = i, Y_n = h\} \\
&= P\{X_{n+1} = j, G(X_{n+1}, Y_n) = k \mid X_n = i, Y_n = h\} \\
&= P\{G(X_{n+1}, Y_n) = k \mid X_{n+1} = j, X_n = i, Y_n = h\} \\
&\qquad \cdot P\{X_{n+1} = j \mid X_n = i, Y_n = h\} \\
&= \mathbf{1}_{\{G(j,h)=k\}} P\{X_{n+1} = j \mid X_n = i\} = \mathbf{1}_{\{G(j,h)=k\}} p_{i,j}
\end{aligned} \tag{1.3.5}$$

from which

$$q_{(i,h),(j,k)} = p_{(i,h),(j,k)} \cdot q_i = q_{i,j}\mathbf{1}_{\{G(j,h)=k\}} \tag{1.3.6}$$

∎

**Example 1.** As an example consider the standard case where

$$Y_t = g(X_0^t) := \min_{s\leq t} X_s \tag{1.3.7}$$

Notice that in this case $Y_t$ has the same finite number of possible values as $X_t$ and the assumptions on $g(\cdot)$ are satisfied since

$$G(X_{\tau_n}, g(X_0^{\tau_{n-1}})) = \min\left[X_{\tau_n}, \min_{s\leq\tau_{n-1}} X_s\right] = g(X_0^{\tau_n}) \tag{1.3.8}$$

The relation (1.3.5) particularizes into

$$p_{(i,h),(j,k)} = \mathbf{1}_{\{G(j,h)=k\}} p_{i,j} = \mathbf{1}_{\{\min\{j,h\}=k\}} p_{i,j} \tag{1.3.9}$$

which, with the states $x^i$ in increasing order of magnitude, implies that

$$p_{(i,h),(j,k)} = \begin{cases} p_{ik} & \text{if} \quad k < h \\ p_{ij} & \text{if} \quad k = h,\, j \geq k \\ 0 & \text{if} \quad k > h \end{cases} = \begin{cases} \frac{q_{ik}}{q_i} & \text{if} \quad k < h \\ \frac{q_{ij}}{q_i} & \text{if} \quad k = h,\, j \geq k \\ 0 & \text{if} \quad k > h \end{cases} \tag{1.3.10}$$

and, consequently,

$$q_{(i,h),(j,k)} = p_{(i,h),(j,k)} \cdot q_i = \begin{cases} q_{ik} & \text{if} \quad k < h \\ q_{ij} & \text{if} \quad k = h,\, j \geq k \\ 0 & \text{if} \quad k < h \end{cases} \tag{1.3.11}$$

Notice that the chain $(X_t, Y_t)$, as described above, is multivariate but still time homogeneous so that the pricing of the claim could be performed by the explicit analytic formula (1.1.2), although it is more complex due to the increased dimensionality. If, however, one would consider a claim of the form $H_T = (X_T - g(X_{T-\sigma}^T))^+$ for a given $0 < \sigma < T$, then our process $Y_t$ may be defined as taking a suitable fixed value for $t \leq T - \sigma$ and thereafter evolves as in the case of $g(X_0^T)$ by letting $t = 0$ correspond to $t = T - \sigma$. In the specific case of Example 1, the process $Y_t$ may then be defined as

$$Y_t = \begin{cases} \max_i x^i & \text{for} \quad t \leq T - \sigma \\ \\ \min_{T-\sigma<s\leq t} X_s & \text{for} \quad t > T - \sigma \end{cases} \tag{1.3.12}$$

It is then quite evident that the chain $(X_t, Y_t)$ is not anymore time homogeneous, even if $X_t$ is. The transition intensity matrix is then a sequence $Q(n)$ that has a certain expression for all $n$ such that $\tau_n \leq T - \sigma$ and another one for those $n$ for which $\tau_n > T - \sigma$. Notice also that the change from one expression to the other one depends on $\tau_n$ and therefore on the individual trajectory of $X_t$. Formula (1.1.2) then does not apply anymore, but by the approach described in the next section 1.4 we shall still be able to compute the price as we shall discuss in subsection 1.5.2.

### 1.3.2 Asian options

Given an underlying CTMC process $X_t$, define the process

$$Y_t := \int_0^t X_s ds = \sum_{\tau_n \leq t} X_{\tau_{n-1}}(\tau_n - \tau_{n-1}) + X_{\tau_n}(t - \tau_n) \tag{1.3.13}$$

and write $X_n$ and $Y_n$ for $X_{\tau_n}$ and $Y_{\tau_n}$ respectively.

The claim of a standard (average price) Asian option can then be represented as

$$H_T = \left( \frac{1}{\sigma} \int_{T-\sigma}^T X_s ds - \kappa \right)^+ = \left( \frac{1}{\sigma}(Y_T - Y_{T-\sigma}) - \kappa \right)^+ \tag{1.3.14}$$

for a given $0 < \sigma < T$ and where $\kappa$ denotes the strike. While $X_t$ is finite-state, $Y_t$ is continuous-valued. If we want the pair $(X_t, Y_t)$ to be finite-state Markov, we have to discretize the values of $Y_t$. While for this purpose one might introduce an optimal quantization (see e.g. [BPP01]), for simplicity we shall use here a standard discretization based on an equipartition of the range of the values of $Y_t$. Assuming that $X_t \in \left\{x^1, \cdots, x^N\right\}$, where the values $x^i$ are in increasing order

of magnitude (as before, we may identify $x^i$ with $i$) the range for the values for $Y_t$, when $t \leq T$, is $[0,\, T\max_{t\leq T} X_t] = \left[0, Tx^N\right]$. Partition now $\left[0, Tx^N\right]$ into intervals of equal length $\Delta > 0$ assuming that $Tx^N = K\Delta$ for a suitable positive integer $K$. The generic $k-$th interval of the partition is then

$$A^k = [a^{k-1}, a^k) = [(k-1)\Delta, k\Delta), \quad k = 1, \cdots, K \tag{1.3.15}$$

Denote by $y^k$ the midpoint of $A^k$ (*other choices are possible*) and let the discretized process $Y_t^\Delta$ for $t \leq T$ be such that $Y_t^\Delta = y^k$ if $Y_t \in A^k$. In what follows we may denote this value simply by $k\,, (k = 1, \cdots, K)$. Since $Y_0 = 0$, we have also to allow for the value $y = 0$ that we may consider as corresponding to $k = 0$.

We introduce next the bivariate finite-state process $(X_t, Y_t^\Delta)$ that below we shall simply denote by $(X_t, Y_t)$ and where the state transitions occur only at the jump times $\tau_n$ of the chain $X_t$ so that occasionally we shall denote this process by $(X_n, Y_n^\Delta) = (X_{\tau_n}, Y_{\tau_n}^\Delta)$. The transitions of the component $X_n$ are those of the given underlying CTMC $X_t$, while for $Y_n$ we have the following

*i*) For $\tau_0 = 0$ put $Y_0 = 0$.

*ii*) For $n > 0$ with $\tau_n \leq T$ let

$$\begin{aligned} Y_n = y^k \quad &\leftrightarrow \quad Y_{n-1} + X_{n-1}(\tau_n - \tau_{n-1}) \in A^k \\ &\leftrightarrow (k-1)\Delta \leq Y_{n-1} + X_{n-1}(\tau_n - \tau_{n-1}) < k\Delta \end{aligned} \tag{1.3.16}$$

*iii*) For $n > 0$ with $\tau_n > T$ let

$$\begin{aligned} Y_T = y^k \quad &\leftrightarrow \quad Y_{n-1} + X_{n-1}(T - \tau_{n-1}) \in A^k \\ &\leftrightarrow (k-1)\Delta \leq Y_{n-1} + X_{n-1}(T - \tau_{n-1}) < k\Delta \end{aligned}$$

Since the distribution of $\tau_n - \tau_{n-1}$ is completely determined by the value of $X_{n-1}$ (it is exponential with parameter $q_i$ if $X_{n-1} = x^i$), we see that the pair is Markov as long as it is restricted to the event $\tau_n \leq T$, afterwards $\tau_n$ has to be added to the state to have Markovianity.

In order to deal with the chain $(X_t, Y_t)$ for $t \leq T$ and to price $H_T$ we need to derive, as for the lookback case, the transition intensity matrix

$$Q = \left\{q_{(i,h),(j,k)}(n)\right\}_{\substack{i,j=1,\cdots,N \\ h,k=1,\cdots,K}}$$

For this purpose notice that, considering the second component of the finite state process $(X_t, Y_t)$ as described above, we may rewrite the claim $H_T$ in (1.3.14) as follows (for simplicity of presentation we put here $\sigma = T$)

$$H_T = \left(\frac{Y_T}{T} - \kappa\right)^+ = \left(\frac{Y_{N_T}}{T} - \kappa\right)^+ = \sum_{n=1}^{\infty} \left(\frac{Y_n}{T} - \kappa\right)^+ \mathbf{1}_{\{n=N_T\}} \tag{1.3.17}$$

so that we need the transition intensities only for those values of $n$ that correspond to the event $\{n \leq N_T\}$. On the other hand, since, as in the lookback case, $Y_t$ makes a transition exactly when $X_t$ does, the jump intensity for the pair $(X_t, Y_t)$ is the same as that for $X_t$, i.e. $q_{(i,h)} = \sum_{j,k} q_{(i,h),(j,k)} = q_i \quad \forall h = 1, \cdots, K$ and so, by $q_{(i,h),(j,k)} = q_i\, p_{(i,h),(j,k)}$, it suffices to construct

$p_{(i,h),(j,k)}$ $(i,j=1,\cdots,N;\, h,k=1,\cdots,K)$, namely the transition probabilities for the embedded chain. Given the comment after (1.3.17), we shall now determine $p_{(i,h),(j,k)}$ restricted to those transition times for which $\tau_n \leq T$, $(n \leq N_T)$. On the event $\tau_n \leq T$ the definition of the discretized process $Y_t$ is limited to that in point ii) of (1.3.16) above and so we obtain

$$\begin{aligned} p_{(i,h),(j,k)} &= \tilde{P}\{X_{n+1}=j,\, Y_{n+1}=k \mid X_n=i,\, Y_n=h\} \\ &= P\{Y_n=y^k \mid X_n=x^j,\, X_{n-1}=x^i,\, Y_{n-1}=y^h\}\, p_{i,j} \\ &= P\{Y_n \in A^k \mid X_{n-1}=x^i,\, Y_{n-1}=y^h\}\, p_{i,j} \\ &= P\left\{\tfrac{(k-1)\,\Delta - y^h}{x^i} \leq \tau_n - \tau_{n-1} < \tfrac{k\,\Delta - y^h}{x^i}\right\} p_{i,j} \\ &= e^{-q_i \frac{(k-1)\,\Delta - y^h}{x^i}} \left[1 - e^{-q_i \frac{\Delta}{x^i}}\right] \frac{q_{ij}}{q_i} \end{aligned} \tag{1.3.18}$$

It follows that

$$q_{(i,h),(j,k)} = q_{i,j}\, e^{-q_i \frac{(k-1)\,\Delta - y^h}{x^i}} \left[1 - e^{-q_i \frac{\Delta}{x^i}}\right] \tag{1.3.19}$$

Notice that the probabilities computed in (1.3.18) are relative to the event $\tau_n \leq T$ and so do not in general sum up to 1. At this point we can determine a $Q-$matrix

$$Q = \left\{q_{(i,h),(j,k)}\right\}_{\substack{i,j=1,\cdots,N \\ h,k=1,\cdots,K}}$$

with the elements according to (1.3.19). Since the $p_{(i,h),(j,k)}$ do not necessarily sum up to 1, this $Q$ is not necessarily a transition intensity matrix but, in view of the above, it suffices to compute the price of $H_T$ in (1.3.14).

## 1.4 Background for the Monte Carlo method by conditioning

We recall formula (1.2.2) for the price of a contingent claim $H(X_T)$, namely

$$V_{H,t,T}(X_t) = E^{\tilde{P}}\left\{\exp\left[-\sum_{i=N_t}^{N_T-1} r_i(\tau_{i+1}-\tau_i) - r_{N_T}(T-\tau_{N_T})\right] H(X_T) \mid X_t\right\} \tag{1.4.1}$$

recalling also that $V_{H,t,T}(X_t)$ stands for the vector with components $V_{H,t,T}(X_t)\mathbf{1}_{\{X_t=x^i\}} = V_{H,t,T}(X_t)_{|X_t=x^i}$. We shall call this *Prototype product* since various more general derivatives can be obtained either as particular cases or as linear combinations of prototype products. This is in particular the case when $X_t = r_t$ and one has to deal with interest rate derivatives where e.g. Caps and Swaptions can be expressed as linear combinations of prototype products (see [PR10b]). Notice also the analogy between the prototype product and Arrow-Debreu prices.

As already mentioned in section 1.2, the plain MC approach consists in simulating $\tau_n$ and the values $X_n$ of $X_t$ at $\tau_n$ and then averaging over the values obtained in each simulation run for the argument in the expectation of the right hand side of (1.2.2) (see also (1.4.1)). In the MC approach based on conditioning on the number of state transitions of $X_t$ that we are going

to present here, we are simulating only the number of transitions $N_{t,T}$ of $X_t$ between $t$ and $T$ and, conditionally on $N_{t,T}$, perform the remaining calculations in an explicit analytic way. From (1.4.1) we now obtain that the price of the claim $H(X_T)$, conditionally on $N_{t,T}$, is given by

$$V_{H,t,T}^{N_{t,T}}(X_t) = E^{\tilde{P}}\left\{exp\left[-\sum_{i=N_t}^{N_T-1} r_i(\tau_{i+1}-\tau_i) - r_{N_T}(T-\tau_{N_T})\right] H(X_T) \mid X_t, N_{t,T}\right\} \tag{1.4.2}$$

and we have

$$V_{H,t,T}(X_t) = E^{\tilde{P}}\left\{V_{H,t,T}^{N_{t,T}}(X_t) \mid X_t\right\} \tag{1.4.3}$$

so that, if we are able to compute in an exact analytic way $V_{H,t,T}^{N_{t,T}}(X_t)$ in (1.4.2), then we may compute the expectation in (1.4.3) by MC simulations of only $N_{t,T}$. It turns out that the exact analytic computation of $V_{H,t,T}^{N_{t,T}}(X_t)$ in (1.4.2) is made difficult by the presence of the random variables $r_{N_T}$ and $\tau_{N_T}$ both in the sum appearing in the exponential as well as in the last term thus preventing the use of a double conditioning to separate the expectation of the exponential of the sum from that of the last term. As we shall show below, it is however possible to compute in an exact analytic way both an upper and a lower bound on $V_{H,t,T}^{N_{t,T}}(X_t)$, which we shall denote by $\bar{V}_{H,t,T}^{N_{t,T}}(X_t)$ and $\underline{V}_{H,t,T}^{N_{t,T}}(X_t)$ respectively and that are defined as follows. First, let for $i = 1, \cdots, N$

$$V_{H,t,T}^{N_{t,T},0}(X_t) = E^{\tilde{P}}\left\{exp\left[-\sum_{i=N_t}^{N_T-1} r_i(\tau_{i+1}-\tau_i)\right] H(X_T) \mid X_t, N_{t,T}\right\} \tag{1.4.4}$$

and

$$V_{H,t,T}^{N_{t,T},1}(X_t) = E^{\tilde{P}}\left\{exp\left[-\sum_{i=N_t}^{N_T} r_i(\tau_{i+1}-\tau_i)\right] H(X_T) \mid X_t, N_{t,T}\right\} \tag{1.4.5}$$

denoting, furthermore, by $V_{H,t,T}^{N_{t,T},0}(\underline{x})$ and $V_{H,t,T}^{N_{t,T},1}(\underline{x})$ the vectors with components $V_{H,t,T}^{N_{t,T},0}(X_t)_{|X_t=x^i}$ and $V_{H,t,T}^{N_{t,T},0}(X_t)_{|X_t=x^i}$ respectively. Then put

$$\bar{V}_{H,t,T}^{N_{t,T}}(X_t); = \max\left[V_{H,t,T}^{N_{t,T},0}(X_t), V_{H,t,T}^{N_{t,T},1}(X_t)\right] \tag{1.4.6}$$

$$\underline{V}_{H,t,T}^{N_{t,T}}(X_t) := \min\left[V_{H,t,T}^{N_{t,T},0}(X_t), V_{H,t,T}^{N_{t,T},1}(X_t)\right] \tag{1.4.7}$$

considering, analogously to $V_{H,t,T}^{N_{t,T},i}(\underline{x})$, $i = 0, 1$, also here the vectors $\bar{V}_{H,t,T}^{N_{t,T}}(\underline{x})$ and $\underline{V}_{H,t,T}^{N_{t,T}}(\underline{x})$. We also recall that $X_T = X_{N_T}$ in (1.4.4) and $X_T = X_{N_T+1}$ in (1.4.5).

**Remark 1.4.1.** *Since for each $X_t = x^i$ the argument of the expectation in the right hand side of (1.4.5) is a.s. smaller than the corresponding one in (1.4.4), one might wonder why we did not define the upper and lower bounds directly as $V_{H,t,T}^{N_{t,T},0}(X_t)$ and $V_{H,t,T}^{N_{t,T},1}(X_t)$ respectively. The reason is that it is the norm of the entire vector $V_{H,t,T}^{N_{t,T},1}(\underline{x})$ determined according to (1.4.5) that is guaranteed to be smaller than or equal to that determined according to (1.4.4), but the individual components of $V_{H,t,T}^{N_{t,T},1}(\underline{x})$ may not necessarily be smaller than those of $V_{H,t,T}^{N_{t,T},0}(\underline{x})$. In fact, in our computations the individual components of $V_{H,t,T}^{N_{t,T},0}(\underline{x})$ and $V_{H,t,T}^{N_{t,T},1}(\underline{x})$ turned out to exhibit an initial oscillatory behavior. It is however always the case that the actual price $V_{H,t,T}(X_t)$ in (1.4.1) belongs to the interval with extreme points given by the corresponding values of $\bar{V}_{H,t,T}^{N_{t,T}}(X_t)$ and $\underline{V}_{H,t,T}^{N_{t,T}}(X_t)$.*

Instead of computing analytically the exact value of $V_{H,t,T}^{N_{t,T}}(X_t)$ in (1.4.2), we shall now determine analytically the approximation given by the midpoint between $\bar{V}_{H,t,T}^{N_{t,T}}(X_t)$ and $\underline{V}_{H,t,T}^{N_{t,T}}(X_t)$, namely

$$V_{H,t,T}^{N_{t,T}}(X_t) \sim \tilde{V}_{H,t,T}^{N_{t,T}}(X_t) := \frac{1}{2}\left(\bar{V}_{H,t,T}^{N_{t,T}}(X_t) + \underline{V}_{H,t,T}^{N_{t,T}}(X_t)\right) \tag{1.4.8}$$

which, combined with (1.4.3), leads to the approximation

$$V_{H,t,T}(X_t) \sim E^{\tilde{P}}\left\{\tilde{V}_{H,t,T}^{N_{t,T}}(X_t)\right\} := \frac{1}{2}E^{\tilde{P}}\left\{\bar{V}_{H,t,T}^{N_{t,T}}(X_t) + \underline{V}_{H,t,T}^{N_{t,T}}(X_t)\right\} \tag{1.4.9}$$

where the expectation is with respect to $N_{t,T}$.

**Remark 1.4.2.** *Concerning the accuracy of the approximation in (1.4.8) and (1.4.9) it will be shown in Corollary 1.4.9 below that the norm of the difference $\bar{V}_{H,t,T}^{n}(\underline{x}) - \underline{V}_{H,t,T}^{n}(\underline{x})$ tends to zero for $n$ tending to infinity so that the approximation can be expected to be rather precise in situations where there are many transitions of $X_t$. This is also in line with the fact that, as can be easily shown, the norm of the difference $\bar{V}_{H,t,T}^{N_{t,T}}(\underline{x}) - \underline{V}_{H,t,T}^{N_{t,T}}(\underline{x})$ is bounded from above by $\bar{\rho} := \max_{i\leq N} \frac{r_i}{q_i}$, which is in general a small value, especially if the transitions of $X_t$ are frequent. In section 1.6 we perform some numerical tests to assess the accuracy of the approximation of $V_{H,t,T}(X_t)$ by $E^{\tilde{P}}\left\{\tilde{V}_{H,t,T}^{N_{t,T}}(X_t)\right\}$.*

The expectation with respect to $N_{t,T}$ in (1.4.9) will be computed by MC simulations. In order to simulate just the number of jumps $N_{t,T}$, for each simulation run we cannot avoid determining also the successive values of $\tau_n$ and $X_n$, but we do not have to record them. As we shall mention in Remark 1.4.10 at the end of the next subsection 1.4.1, instead of determining the value of $E^{\tilde{P}}\left\{\tilde{V}_{H,t,T}^{N_{t,T}}(X_t)\right\}$ by MC simulations of $N_{t,T}$, one might also compute this expectation in a fully analytical way, but the corresponding procedure is rather cumbersome.

In the next subsection we shall now present our approach to determine in an analytically exact way the value of $\tilde{V}_{H,t,T}^{N_{t,T}}(X_t)$ for each given value of $N_{t,T}$.

### 1.4.1 Derivation of the algorithm

Given $t$ and $N_t$ (recall that, see description following (1.2.1), we had assumed without loss of generality that $t = N_t$ and so $X_t = X_{N_t}$) fix an $M \in \mathbb{N}$ and define the sequence of functions $H_n(\cdot)$, $n = 0, \cdots, M$ recursively as follows: $H_0(\cdot)$ is given by the Prototype payoff, namely (see (1.2.3)

$$H_0(\cdot) = \sum_{i=1}^{N} w_i^0 \mathbf{1}_{\{\cdot = x^i\}} \tag{1.4.10}$$

and, for $0 < n \leq M$,

$$H_n(X_{N_t+M-n}) = E^{\tilde{P}}\left\{e^{-r_{N_t+M-n}(\tau_{N_t+M-n+1}-\tau_{N_t+M-n})} H_{n-1}(X_{N_t+M-n+1}) \mid X_{N_t+M-n}\right\} \tag{1.4.11}$$

This definition has been inspired by an approach described in [FZ02], where the authors derive the analog in discrete time of the continuous time affine term structure models, and here we use the fact that $X_t$ is Markov and that, given $X_{N_t}$, the distribution of the inter-arrival time $\tau_{N_t+1} - \tau_{N_t}$ depends only on $X_{N_t}$. Notice also that from (1.4.11) it follows that

$$H_M(X_{N_t}) = E^{\tilde{P}}\left\{\exp\left[-\sum_{m=N_t}^{N_t+M-1} r_m(\tau_{m+1} - \tau_m)\right] H_0(X_{N_t+M}) \mid X_{N_t}\right\} \tag{1.4.12}$$

Since $X_t$ takes a finite number of possible values, just as with $H(\cdot) = H_0(\cdot)$, also each of the $H_n(x)$ can be represented as a vector. More precisely, letting $\underline{x} = [x^1, \cdots, x^N]'$, we have

$$H_n(\underline{x}) = [w_1^n, \cdots, w_N^n]' \tag{1.4.13}$$

with $w_i^n$ representing the value $H_n(x^i)$.

**Remark 1.4.3.** *Notice the immediate relationship between the vectors $H_n(\underline{x})$ defined recursively in (1.4.11) above and the basic quantities of our MC-with-conditioning approach, namely the vectors $V_{H,t,T}^{N_{t,T},0}(\underline{x})$ and $V_{H,t,T}^{N_{t,T},1}(\underline{x})$ with components defined for each value of $X_t = x^i$, $i = 1, \cdots, N$ according to (1.4.4) and (1.4.5) respectively. From (1.4.4), (1.4.5), the equality $H(X_T) = H_0(X_T)$, the representation of $H_M(\cdot)$ in (1.4.12), and recalling that $N_{t,T} = N_T - N_t$, we have in fact that*

$$V_{H,t,T}^{N_{t,T},0}(\underline{x}) = H_{N_{t,T}}(\underline{x}) \quad ; \quad V_{H,t,T}^{N_{t,T},1}(\underline{x}) = H_{N_{t,T}+1}(\underline{x}) \tag{1.4.14}$$

From the definitions of $\bar{V}_{H,t,T}^{N_{t,T}}(\underline{x})$ and $\underline{V}_{H,t,T}^{N_{t,T}}(\underline{x})$ with components defined for each value of $X_t = x^i$, $i = 1, \cdots, N$ according to (1.4.6) and (1.4.7), from that of $\tilde{V}_{H,t,T}^{N_{t,T}}(X_t)$ in (1.4.8), from the equalities (1.4.14), and recalling that $X_T = X_{N_T}$ in (1.4.4) and $X_T = X_{N_T+1}$ in (1.4.5), one obtains immediately the following corollary where, in line with the other analogous vectors defined previously, $\tilde{V}_{H,t,T}^{N_{t,T}}(\underline{x})$ denotes the vector with components $\tilde{V}_{H,t,T}^{N_{t,T}}(X_t)_{|X_t=x^i}$, $i = 1, \cdots, N$.

**Corollary 1.4.4.** *We have*

$$\bar{V}_{H,t,T}^{N_{t,T}}(\underline{x}) = \max\left[H_{N_{t,T}}(\underline{x}), H_{N_{t,T}+1}(\underline{x})\right] \quad ; \quad \underline{V}_{H,t,T}^{N_{t,T}}(\underline{x}) = \min\left[H_{N_{t,T}}(\underline{x}), H_{N_{t,T}+1}(\underline{x})\right]$$

$$\tilde{V}_{H,t,T}^{N_{t,T}}(\underline{x}) = \tfrac{1}{2}\left(\bar{V}_{H,t,T}^{N_{t,T}}(\underline{x}) + \underline{V}_{H,t,T}^{N_{t,T}}(\underline{x})\right) = \tfrac{1}{2}\left(H_{N_{t,T}}(\underline{x}) + H_{N_{t,T}+1}(\underline{x})\right)$$

This Corollary shows clearly the relevance of the functions $H_n(\cdot)$ computed recursively in (1.4.11).

We shall now derive an easily implementable procedure to analytically compute the functions $H_n(\cdot)$. Considering the general case, where the transition intensity of the process $X_t$ is given as a sequence of $Q-$matrices $Q(n) = \{q_{i,j}(n)\}$, define the sequence of matrices $\widetilde{Q}(n)$ as

$$\widetilde{Q}(n) = (\tilde{q}_{i,j}(n))_{1 \le i,j \le N} \quad \text{with} \quad \tilde{q}_{i,j}(n) = \begin{cases} \frac{q_{i,j}(n)}{r^i + q_i(n)} & i \neq j \\ 0 & i = j \end{cases} \tag{1.4.15}$$

where $r^i = r(X)$ when $X = x^i$ and $q_i(n) = \sum_{j \neq i} q_{i,j}(n) = \sum_{i,j} q_{i,j}(n)$.

We have now

**Proposition 1.4.5.** *Starting from the given $H_0(\cdot)$, the functions $H_n(\cdot)$ in (1.4.11) can be computed recursively by the following matrix multiplication*

$$H_n(\underline{x}) = \widetilde{Q}(n)\, H_{n-1}(\underline{x}) \tag{1.4.16}$$

*Proof.* Fixing a generic $n$, by the representation of $H_n(\underline{x})$ in (1.4.13) we have to prove that

$$[w_1^n, \cdots, w_N^n]' = \widetilde{Q}(n)\, [w_1^{n-1}, \cdots, w_N^{n-1}]'$$

For this purpose it suffices to show that for a generic $i \in \{1, \cdots, N\}$ we have

$$w_i^n = \sum_{j=1}^{N} \frac{q_{i,j}(n)}{r^i + q_i(n)} \, w_j^{n-1} \tag{1.4.17}$$

where, due to the fact that $q_{i,i} = 0$, the sum extends actually only over the $j \neq i$.

We next have, see (1.4.11)

$$\begin{aligned} & w_i^n = H_n(X_{N_t+M-n} = x^i) \\ & = E^{\tilde{P}} \left\{ e^{-r_{N_t+M-n}(\tau_{N_t+M-n+1} - \tau_{N_t+M-n})} H_{n-1}(X_{N_t+M-n+1}) \mid X_{N_t+M-n} = x^i \right\} \\ & = E^{\tilde{P}} \left\{ e^{-r^i(\tau_{N_t+M-n+1} - \tau_{N_t+M-n})} \sum_{j=1}^{N} w_j^{n-1} \mathbf{1}_{\{X_{N_t+M-n+1} = x^j\}} \mid X_{N_t+M-n} = x^i \right\} \\ & = \sum_{j=1}^{N} w_j^{n-1} E^{\tilde{P}} \left\{ e^{-r^i(\tau_{N_t+M-n+1} - \tau_{N_t+M-n})} \mathbf{1}_{\{X_{N_t+M-n+1} = x^j\}} \mid X_{N_t+M-n} = x^i \right\} \end{aligned} \tag{1.4.18}$$

where we have used the fact that, at a generic transition time $\tau_n$, we have $r_n = r^i$ if $X_n = x^i$. We have also used the definition of $w_j^{n-1}$ according to (1.4.13) with $n-1$ replacing $n$.

Next, by the general properties of CTMCs we have that, conditional on $X_{N_t+M-n}$, the inter-arrival time $\tau_{N_t+M-n+1} - \tau_{N_t+M-n}$ is independent of $X_{N_t+M-n+1}$. Recalling furthermore that, for $X_{N_t+M-n} = x^i$, the distribution of $\tau_{N_t+M-n+1} - \tau_{N_t+M-n}$ is (negative) exponential with parameter $q_i$, we obtain

$$\begin{aligned} & E^{\tilde{P}} \left\{ e^{-r^i(\tau_{N_t+M-n+1} - \tau_{N_t+M-n})} \mathbf{1}_{\{X_{N_t+M-n+1} = x^j\}} \mid X_{N_t+M-n} = x^i \right\} \\ & = E^{\tilde{P}} \left\{ e^{-r^i(\tau_{N_t+M-n+1} - \tau_{N_t+M-n})} \mid X_{N_t+M-n} = x^i \right\} E^{\tilde{P}} \left\{ \mathbf{1}_{\{X_{N_t+M-n+1} = x^j\}} \mid X_{N_t+M-n} = x^i \right\} \\ & = \int_0^\infty e^{-r^i u} q_i(n) e^{-q_i(n)u} du \; \tilde{P} \left\{ X_{N_t+M-n+1} = x^j \mid X_{N_t+M-n} = x^i \right\} \\ & = \tfrac{q_i(n)}{r^i + q_i(n)} \, p_{i,j}(n) = \tfrac{q_{i,j}(n)}{r^i + q_i(n)} \end{aligned} \tag{1.4.19}$$

where in the last passage we have used the fact that (see (7) for the case of a time homogeneous $Q$) one has that $p_{i,j} = \frac{q_{i,j}}{q_i}$.

Combining (1.4.18) with (1.4.19) we obtain (1.4.17) and thus the statement of the Proposition.

∎

Notice that (1.4.16) implies $H_n(\underline{x}) = \widetilde{Q}(n)\widetilde{Q}(n-1)\cdots\widetilde{Q}(1)H_0(\underline{x})$ which in the time homogeneous case becomes $H_n(\underline{x}) = \widetilde{Q}^n H_0(\underline{x})$. Notice also that in the bivariate (multivariate) case, putting $\underline{z} = (\underline{x}, \underline{y})'$ with $\underline{x} = (x^1, \cdots, x^N)$, $\underline{y} = (y^1, \cdots, y^M)$ we have $H_0(\underline{z}) = [w_1^0, \cdots, w_{N \cdot M}^0]'$ and $H_n(\underline{z}) = \widetilde{Q}(n) H_{n-1}(\underline{z})$ where

$$\widetilde{Q}(n) = \left\{ \frac{q_{(i,h),(j,k)}(n)}{r^{i,h} + q_{i,h}(n)} \right\}_{\substack{i,j=1,\cdots,N \\ h,k=1,\cdots,M}} \tag{1.4.20}$$

whereby $r^{i,h} = r(X, Y)$ when $X = x^i$ and $Y = y^h$ and $q_{i,h}(n) = \sum_{j \neq i, k \neq h} q_{(i,h),(j,k)}(n)$.

The computation of $\tilde{V}^{N_{t,T}}_{H,t,T}(X_t)$ by recursive matrix multiplication and then that of the expectation $E^{\tilde{P}}\left\{\tilde{V}^{N_{t,T}}_{H,t,T}(X_t)\right\}$ by simulating $N_{t,T}$ forms the backbone of our (hybrid) MC method with conditioning. It is based on the rather immediate next Proposition, which in fact follows from (1.4.9), Proposition 1.4.5 and Corollary 1.4.4.

**Proposition 1.4.6.** *The value of $E^{\tilde{P}}\left\{\tilde{V}^{N_{t,T}}_{H,t,T}(X_t)\right\}$, by which (see (1.4.9)) we determine the price $V_{H,t,T}(X_t)$ of the Prototype product, is given by*

$$E^{\tilde{P}}\left\{\tilde{V}^{N_{t,T}}_{H,t,T}(X_t)\right\}_{|X_t=x^i} = \frac{1}{2}\,E^{\tilde{P}}\left\{\left[\left(1+\widetilde{Q}(N_{t,T}+1)\right)\widetilde{Q}(N_{t,T})\cdots\widetilde{Q}(1)H_0(\underline{x})\right]_i\right\} \tag{1.4.21}$$

*where $[z]_i$ denotes the $i-$th component of the vector $z$ and where the expectation is with respect to $N_{t,T}$. In the time homogeneous case the above expression reduces to*

$$E^{\tilde{P}}\left\{\tilde{V}^{N_{t,T}}_{H,t,T}(X_t)\right\}_{|X_t=x^i} = \frac{1}{2}\,E^{\tilde{P}}\left\{\left[\left(1+\widetilde{Q}\right)\widetilde{Q}^{N_{t,T}}H_0(\underline{x})\right]_i\right\} \tag{1.4.22}$$

**Remark 1.4.7.** *The expressions in (1.4.22) can be further simplified if $\widetilde{Q}$ is diagonalizable (see [PR10b]).*

Based on the above Proposition 1.4.6, our conditional MC approach (hybrid MC) can now be synthesized as follows:

**i)** Simulate a sufficiently large number of realizations of the random variable $N_{t,T}$.

**ii)** Record the maximum value, say $M$, of $N_{t,T}$ obtained during the simulations and determine the empirical distribution of $N_{t,T}$ derived from the simulations.

**iii)** Compute recursively the values of $\left(1+\widetilde{Q}(n+1)\right)\widetilde{Q}(n)\cdots\widetilde{Q}(1)H_0(\underline{x})$ for $n=1$ up to $n=M$.

**iv)** Determine the average of the values computed in iii) with respect to the empirical distribution of $N_{t,T}$ determined in ii).

Applications of this approach are discussed in the next section 1.5 and in section 1.6 we then present numerical results and comparisons.

As already mentioned in Remark 1.4.2, the accuracy of the approximation of the exact price $V_{H,t,T}(X_t)$ in (1.4.1) by the $E^{\tilde{P}}\left\{\tilde{V}^{N_{t,T}}_{H,t,T}(X_t)\right\}$, where $\tilde{V}^{N_{t,T}}_{H,t,T}(X_t)$ is the midpoint between the upper and lower bounds $\bar{V}^{N_{t,T}}_{H,t,T}(X_t)$ and $\underline{V}^{N_{t,T}}_{H,t,T}(X_t)$ respectively, can also be seen as a consequence of Corollary 1.4.9 below. For this purpose consider the operator, acting on $\mathbb{R}^N$ with values in $\mathbb{R}^N$, that is associated with the matrix $\widetilde{Q}(n)$ defining the recursions (1.4.16). It is given by the expectation operator in (1.4.11) which, for the generic $i-$th component of $H_n(\underline{x})$ can be expressed as (recall that we had assumed $r^i = r(x^i)$)

$$\widetilde{Q}(n)H_{n-1}(X_{n-1})_{|X_{n-1}=x^i} = E^{\tilde{P}}\left\{e^{-r^i\mathcal{I}}H(X_n)\mid X_{n-1}=x^i\right\} \tag{1.4.23}$$

with $\mathcal{I}$ denoting a (negative) exponential random variable with parameter $q_i$. We have the following

**Proposition 1.4.8.** *The operator $\widetilde{Q}(n)$ in (1.4.23) is a contraction operator in $\mathbb{R}^N$ with contraction constant*

$$\gamma := \max_{i\leq N, n\in\mathbb{N}} \frac{q_i(n)}{r^i + q_i(n)} < 1$$

.

*Proof.* By Jensen's inequality we have

$$\|\widetilde{Q}(n)H(\underline{x}) - \widetilde{Q}(n)\bar{H}(\underline{x})\| \leq \sup_{i,n} E^{\tilde{P}}\left\{e^{-r^i\mathcal{I}}\right\} \|H(\underline{x}) - \bar{H}(\underline{x})\|$$

where

$$\sup_{i,n} E^{\tilde{P}}\left\{e^{-r^i\mathcal{I}}\right\} = \sup_{i,n} \int_0^\infty e^{-r^i s} q_i(n) e^{-q_i(n)s} ds = \max_{i,n} \frac{q_i(n)}{r^i + q_i(n)}$$

∎

This Proposition 1.4.8 as well as Corollary 1.4.4 lead immediately to the following

**Corollary 1.4.9.** *By the contraction property of the operator associated with $\widetilde{Q}(n)$ we have that*

$$\lim_{n\to\infty} \|\bar{V}^n_{H,t,T}(\underline{x}) - \underline{V}^n_{H,t,T}(\underline{x})\| = 0$$

**Remark 1.4.10.** *As already mentioned at the end of the introductory part to this section, instead of determining, on the basis of (1.4.21), the value of $E^{\tilde{P}}\left\{\tilde{V}^{N_{t,T}}_{H,t,T}(X_t)\right\}$ by MC simulations of $N_{t,T}$, one might compute this expectation in a fully analytical way as*

$$E^{\tilde{P}}\left\{\tilde{V}^{N_{t,T}}_{H,t,T}(X_t)\right\}_{|X_t=x^i} = \frac{1}{2}\sum_{n=0}^{\infty}\left[\left(1+\widetilde{Q}(n+1)\right)\widetilde{Q}(n)\cdots\widetilde{Q}(1)H_0(\underline{x})\right]_i \tilde{P}(N_{t,T}=n \mid X_t = x^i)$$

*The difficulties for an actual use of this formula consist in the infinite sum and the probability distribution of $N_{t,T}$. Concerning the infinite sum notice that, since (see Proposition 1.4.8) the operator associated with $\widetilde{Q}(n)$ is contracting, for the actual computations one may truncate the infinite sum thereby introducing an approximation that can be made arbitrarily precise provided the truncation is chosen to be sufficiently large. On the other hand, the probability distribution of $N_{t,T}$ can in fact be determined explicitly, however the corresponding procedure is rather cumbersome. Details for the case of a scalar $X_t$ can again be found in [PR10b].*

## 1.5 Applications of the conditional MC approach to the pricing of path dependent options

Recall that the proposed conditional MC method is described in steps i) to iv) in subsection 1.4.1. As already mentioned, with respect to a plain MC this method allows to reduce the variance ("variance reduction by conditioning") and, as we shall see from the numerical results in the next section 1.6, also a possible bias. The great advantage lies however in the wide applicability, in particular in the time inhomogeneous and multivariate cases, of our method that is based on successive matrix multiplications. In subsection 1.5.1 we shall show this first for barrier options *(knock-out options)* for which formula (1.1.2) cannot be used not even in the time homogeneous case. Successively, in subsections 1.5.2 and 1.5.3 we shall then show it for the lookback and Asian options as we had described them in subsections 1.3.1 and 1.3.2 and for which our method allows to overcome the difficulties alluded to in Section 1.3.

### 1.5.1 Barrier options (knock-out options)

Barrier options include various kinds of derivatives, in particular they include also credit risky derivatives in the context of the structural approach. As an example we consider here a specific case, namely knock-out options.

We consider the case when an option with underlying $X_t$ is knocked out as soon as $X_t$ reaches or falls below a level $L$. Using a notation corresponding to (1.2.3), assume that for the "background" (not knocked out) option we have

$$\bar{H}(\cdot) = \bar{H}_0(\cdot) = \sum_{i=1}^{N} \bar{w}_i^0 \mathbf{1}_{\{\cdot = x^i\}},\ x^i \in E,\ \ w_i^0 \in \mathbb{R} \tag{1.5.1}$$

Assume furthermore that the values $x^i$ are arranged in increasing order of magnitude and put

$$\ell := \min[i \in \{1, \cdots, N\} \mid x^i > L] \tag{1.5.2}$$

For the knock-out option we may then start from

$$H(X_t) = H_0(X_T) = \sum_{i=1}^{N} \bar{w}_i^0 \mathbf{1}_{\{X_T = x^i, i \geq \ell\}} := \sum_{i=1}^{N} w_i^0 \mathbf{1}_{\{X_T = x^i\}} \tag{1.5.3}$$

having put $w_i^0 := \bar{w}_i^0 \mathbf{1}_{\{i \geq \ell\}}$.

In order to be able to apply our conditional MC approach, we want also here to obtain a relation of the form (1.4.16) for a suitable $\widetilde{Q}(n)$. We have now the rather immediate

**Proposition 1.5.1.** *Starting from*

$$H_0(\cdot) = \sum_{i=1}^{N} \bar{w}_i^0 \mathbf{1}_{\{\cdot = x^i, i \geq \ell\}} := \sum_{i=1}^{N} w_i^0 \mathbf{1}_{\{\cdot = x^i\}} \tag{1.5.4}$$

*with $w_i^0 := \bar{w}_i^0 \mathbf{1}_{\{i \geq \ell\}}$ we have, for $n \leq N_T$,*

$$H_n(\cdot) = \sum_{i=1}^{N} w_i^n \mathbf{1}_{\{\cdot = x^i\}} \tag{1.5.5}$$

*where $w^n = [w_1^n, \cdots, w_N^n]'$ are given recursively by*

$$w^n = I_\ell \widetilde{Q}(n) w^{n-1} \tag{1.5.6}$$

*with $I_\ell$ a unit matrix having the first $\ell$ rows equal to zero and, as before,*

$$\widetilde{Q}(n) = \left\{ \frac{q_{i,j}(n)}{r_i + q_i(n)} \right\}_{i,j=1,\cdots,N} \tag{1.5.7}$$

As a consequence of Proposition 1.5.1 we may restrict attention to an $(N - \ell)-$vector $\tilde{w}^n$ for which

$$\tilde{w}_i^0 = w_i^0 := \bar{w}_i^0 \mathbf{1}_{\{i \geq \ell\}} \quad \text{and} \quad \tilde{w}^n = \widetilde{Q}_\ell(n) \tilde{w}^{n-1} \tag{1.5.8}$$

where $\widetilde{Q}_\ell(n)$ is the $(N-\ell)\times(N-\ell)$ sub matrix of $\widetilde{Q}^n$ formed by the last $N-\ell$ rows and columns. We have the equivalent representations

$$H_n(X_{N_T-n}) = \sum_{i=1}^{N} w_i^n \mathbf{1}_{\{X_{N_T-n}=x^i\}} = \sum_{i=1}^{N-\ell} \tilde{w}_i^n \mathbf{1}_{\{X_{N_T-n}=x^i\}} \tag{1.5.9}$$

Notice the importance here of having the recursive relation (1.5.6) even if the underlying is time-homogeneous. Notice also that, in the case of barrier options, the time in-homogeneity arises not only if the underlying $X_t$ is time in-homogeneous, but also if the barrier $L$ is time varying.

### 1.5.2 Lookback options

In the general case of a lookback option of the form $H_T = (X_T - g(X_{T-\sigma}^T))^+$ with $0 < \sigma < T$ our conditional MC approach (see steps i) to iv) in subsection 1.4.1) appears to be particularly appropriate since the expression to be averaged, namely $\left[1+\widetilde{Q}(n+1)\right]\widetilde{Q}(n)\cdots\widetilde{Q}(1)H_0(\underline{x})$ can be computed separately for each individual simulated trajectory.

### 1.5.3 Asian options

In subsection 1.3.2 we had derived the transition intensity matrix $Q$ for the bivariate Markov process $(X_t, Y_t)$ on the event $\{n \le N_T\} = \{\tau_n \le T\}$, which suffices to evaluate the claim $H_T = \left(\frac{Y_{N_T}}{T} - \kappa\right)^+$ (see (1.3.17)). In the MC-with-conditioning the basic quantities are $V_{H,t,T}^{N_{t,T},0}(X_t)$ in (1.4.4) and $V_{H,t,T}^{N_{t,T},1}(X_t)$ in (1.4.5) where, for the case of $V_{H,t,T}^{N_{t,T},1}(X_t)$, we need the transition intensities of $(X_t, Y_t)$ on the larger event $\{n \le N_T + 1\} = \{\tau_{n+1} \le T\}$. However, since the last discounting factor in (1.4.5) is $\exp\left[-r_{N_T}\left(\tau_{N_T+1} - \tau_{N_T}\right)\right]$ rather than $\exp\left[-r_{N_T}\left(T - \tau_{N_T}\right)\right]$ as in the expression (1.2.2) for the exact price (see also (1.4.1) and (1.4.2)), we can still use the definition for the discretization of $Y_t$ as described in point ii) of (1.3.16) to compute the transition intensities also on the larger set $\{n \le N_T+1\}$ so that they keep the same expression as in (1.3.19). Accordingly, we may consider also here a matrix $\widetilde{Q}$ constructed in line with (2.4.12), namely

$$\tilde{q}_{(i,h),(j,k)} = \begin{cases} \frac{q_{(i,h),(j,k)}}{r^i+q_i} & i \ne j \\ 0 & i = j \end{cases} \tag{1.5.10}$$

where, as in (2.4.12), $r^i = r(X)$ when $X = x^i$ and $q_i = \sum_{j\ne i} q_{i,j}$. This matrix can now be used to compute $V_{H,t,T}(X_t)$ according to (1.4.9) and (1.4.21), namely it can be used to implement our conditional MC approach as described in steps i) to iv) in subsection 1.4.1.

For simplicity of presentation, the analysis in subsection 1.3.2 was carried out for the case of a time homogeneous $X_t$ and for $\sigma = T$ and so the $Q-$matrix for $(X_t, Y_t)$ turned out to be time homogeneous. In this time homogeneous case one may compute the price of $H_T$ also via the explicit analytic formula (1.1.2) and, in fact, in the next numerical section 1.6 we perform also this computation for comparison purposes (it will, though, not be the exact price, but a sufficiently good "proxy" since it is based on the discretization of $Y_t$). If, however, $X_t$ is time inhomogeneous and/or $\sigma < T$, then formula (1.1.2) is not applicable anymore, but our MC-with-conditioning can still be used.

## 1.6 Numerical results and comparisons

The purpose of this section is to implement numerically our suggested approach and thereby to show that, with respect to a plain MC, in the MC with conditioning the variance is indeed reduced and the results themselves are more precise. To apply our method, we have first to specify a transition intensity matrix $Q$. For actual applications, this matrix would have to be calibrated to actual market data. Here the purpose is however that of providing some test examples and for this it may be convenient to choose the matrix freely. In fact, even if we would perform a calibration, we would end up with one specific matrix $Q$. To test the behavior of our algorithm it may thus be more significant to try more than one $Q-$matrix which is what we do below at least in one case. While calibration is thus beyond the scope of this work, we still want to make the following

**Remark 1.6.1.** *Concerning* **calibration** *we would like to mention that, for an actual calibration to become feasible, one would have to choose specific patterns of $Q-$matrices, parametrized by a small number of parameters, and so calibrate just these parameters. One such pattern can for example be obtained by taking a classical continuous-time affine term structure model and (see section 1.6.2 below) perform a space discretization according e.g. to [KD01] thus obtaining a corresponding CTMC with a transition matrix that turns out to be tri-diagonal and is characterized by the same parameters as the continuous-time model. We also mention that, for a time homogeneous $X_t$ it is possible to set up a filtering approach to estimate the values of the $Q-$matrix, either by combined filtering and parameter estimation, or by filtering and EM-parameter estimation (see e.g.[EAM08], see also Part 2 of this thesis).*

The section is structured as follows. In subsection 1.6.1, we discuss the performance criteria, which is the convergence of the empirical estimator variance per sample. Then in each of the succeeding subsections we present the performance of MC with conditioning for different types of financial products as described in previous sections. We start in subsection 1.6.2 by discussing the pricing of zero-coupon bonds for which, at least in the time homogeneous case, comparison with theoretically exact values is possible; this may thus serve as an assessment of the quality of our approach which is then applied to situations where a comparison is not possible anymore. Prices of zero-coupon bonds previously computed via MC with conditioning in [PR10b] are here reported and compared against prices computed using other methods, in particular almost exact ones. It is shown that our approach yields comparable values. In subsection 1.6.3 we present the numerically tested performance of MC with conditioning for pricing barrier options (knock-out options); in subsections 1.6.4 and 1.6.5 we present the analogous numerical tests for the performance of our approach in the case of lookback options as well as Asian options prices.

### 1.6.1 Monte Carlo criterion

The book [Gla04] treats comprehensively the MC method and its applications to finance, including techniques to improve the performance of the *vanilla* MC by analyzing various more sophisticated kinds of MC. In order to compare two different MC estimators it is important to establish the relevant performance criterion. Let $\hat{C}_n$ denote the MC estimate for the random variable $C$, computed after $n$ simulations that are each i.i.d. Due to the central limit theorem, it is well

known that under an unbiased estimator $\hat{C}$, the error

$$\hat{C}_n - C \Rightarrow \mathcal{N}(0, \sigma_C^2/n)$$

in distribution, so the convergence rate

$$\left|\mathrm{Var}[\hat{C}_n - C] - \sigma_C^2/n\right| /n$$

is a simple performance criterion in the ideal case. This is however not applicable in our situation. In the first place, given the underlying CTMC $X$, we often do not explicitly have the theoretical distribution of the derivative price, especially when treating complicated payoffs, i.e., there are no closed-form pricing formulas that could provide a basis of comparison for more complicated derivatives. In the second place, estimator bias depends on the parameters of the CTMC $X$. Indeed, in our various experiments for the case when the theoretical distribution of prices is available, it has been empirically observed that certain choices of parameters for $X$ lead to MC estimators that converge to the theoretical mean faster than others.

For methods that improve on a plain MC, additional computational effort is necessary to ensure that it is an asymptotically unbiased estimator. It is therefore observed in [Gla04] that a better criterion must take into account both the convergence rate of the error variance and also the bias reduction rate. A criterion proposed was the mean square error (MSE), which is the sum of the squared bias and the variance (see [Gla04]), more explicitly, if $\hat{\alpha}$ is an estimator of a quantity $\alpha$ then

$$\begin{aligned}\mathrm{MSE}(\hat{\alpha}) =& E[(\hat{\alpha} - \alpha)^2] \\ =& (E[\hat{\alpha}] - \alpha)^2 + E[(\hat{\alpha} - E[\hat{\alpha}])^2] \\ =& \mathrm{Bias}^2(\hat{\alpha}) + \mathrm{Variance}(\hat{\alpha}).\end{aligned}$$

In the absence of explicit expressions for the theoretical mean of the prices, we are in general unable to use the component $\mathrm{Bias}^2(\hat{\alpha})$ of the MSE, and therefore we shall assume the estimator variance as the basis of our principal performance criterion instead of the entire MSE. The performance criterion we will therefore use is

$$\frac{Var\left[\hat{C}_n\right]}{n}$$

which reflects the estimator variance per number of simulations.

We remark that a possible future improvement to our study is therefore to investigate bias estimators for the case of complicated derivative payoffs, allowing hence to evaluate the estimator on the basis of MSE performance, as well as investigating bias reduction techniques. Nevertheless we wish to point out that, since the MC with conditioning approach provides analytic expressions for quantities that are on the other hand left as random variables in plain MC, intuitively there is less bias in the MC with conditioning approach (it is more closely related to the explicit formula).

### 1.6.2 Zero Coupon Bonds

As mentioned at the beginning of this section 1.6, the pricing of zero coupon bonds is simple enough to allow for a comparison also with theoretically almost exact values and corresponds to

our setup by putting $X_t = r_t$ and $H(X_T) \equiv 1$. The CTMC for $X_t = r_t$ can for example be chosen as the approximation, obtained via a space discretization in the spirit of [KD01], of well-known affine diffusion models for the short rate of interest $r_t$, for which the exact price can be computed analytically. We report here results from [PR10b], where the CTMC $X_t = r_t$ is obtained from a space discretization according to [KD01] of the CIR model for a corresponding continuous-time short rate $\tilde{r}_t$, namely

$$\begin{cases} d\tilde{r}_t = k(\theta - \tilde{r}_t)dt + \sigma dW_t \\ \tilde{r}_0 = \bar{r} \end{cases} \tag{1.6.1}$$

The bond prices obtained from the MC method with conditioning are compared not only with the theoretically exact price $\widetilde{P}(t,T)$ for model (1.6.1), but also with the prices obtained by other methods, including plain MC. While in the present work we consider the midpoint between the upper and lower bounds as the value to be computed analytically in our (hybrid) MC method with conditioning (see (1.4.8) and Corollary 1.4.4), in [PR10b] the authors consider just the upper bound that in general does not differ much from the midpoint. In the Tables below, prices for various maturities $T$ and different initial values for the short rate $\tilde{r}_0 = r_0 = \bar{r}$, as well as for different values of the parameters in (1.6.1) are reported from [PR10b]. As already mentioned, $\widetilde{P}(t,T)$ denotes the exact bond price for model (1.6.1), while the other reported prices are: $P_{RBT}(t,T)$ the bond price obtained from a recombining binomial tree model, $P_{FZ}(t,T)$ the price computed for a time discretization of (1.6.1) and computed according to the recursive pricing method described in [FZ02], $P_{Exp}(t,T)$ the price computed by formula (1.1.2) (recall that $Q$ is here time homogeneous), $P_{pl}(t,T)$ the price computed by plain MC and $P_c^{ub}(t,T)$ the upper bound on the price computed by MC with conditioning.

The following parameters were used: $t = 0$ years, $T = 0.5$, 2 and 5 years. The number of MC simulations $M$ and the RBT steps $\bar{M}$ were taken to be 500, which, while it is a low number with respect to what is typical, was enough because a very fine space discretization was chosen. In Table 1 and Table 2 the numerical results are presented when the values of the initial spot rate $\bar{r}$ and the mean-reversion constant $\theta$ are of the order of one hundredth; in Table 3 the values of $\bar{r}$ and $\theta$ are of the order of one tenth.

**Table 1**: Bond prices $\widetilde{P}(t,T)$, $P_{RBT}(t,T)$, $P_{FZ}(t,T)$, $P_{Exp}(t,T)$ and $P_c^{ub}(t,T)$ ($\mathbf{M} = \bar{\mathbf{M}} = 500$)

| $T$(years) | 0.5 | 2 | 5 | 0.5 | 2 | 5 |
|---|---|---|---|---|---|---|
| $\bar{r}(= r^i)$ | 0.01 | 0.01 | 0.01 | 0.02 | 0.02 | 0.02 |
| $k$ | 0.01 | 0.01 | 0.01 | 0.02 | 0.02 | 0.02 |
| $\theta$ | 0.8 | 0.8 | 0.8 | 0.5 | 0.5 | 0.5 |
| $\sigma$ | 0.1 | 0.1 | 0.1 | 0.05 | 0.05 | 0.05 |
| $\widetilde{P}(t,T)$ | 0.995014 | 0.980244 | 0.951462 | 0.990051 | 0.960821 | 0.905046 |
| $P_{RBT}(t,T)$ | 0.995042 | 0.980302 | 0.951556 | 0.99007 | 0.960898 | 0.905226 |
| $P_{FZ}(t,T)$ | 0.995014 | 0.980244 | 0.951463 | 0.990051 | 0.960821 | 0.905046 |
| $P_{Exp}(t,T)$ | 0.995012 | 0.979568 | 0.947174 | 0.990051 | 0.960821 | 0.905047 |
| $P_c^{ub}(t,T)$ | 0.995024 | 0.980276 | 0.951621 | 0.990143 | 0.960734 | 0.905318 |

**Table 2**: Bond prices $\widetilde{P}(t,T)$, $P_{RBT}(t,T)$, $P_{FZ}(t,T)$, $P_{Exp}(t,T)$ and $P_c^{ub}(t,T)$ ($\mathbf{M} = \bar{\mathbf{M}} = 500$)

| $T$(years) | 0.5 | 2 | 5 | 0.5 | 2 | 5 |
|---|---|---|---|---|---|---|
| $\bar{r}(=r^i)$ | 0.03 | 0.03 | 0.03 | 0.02 | 0.02 | 0.02 |
| $k$ | 0.03 | 0.03 | 0.03 | 0.02 | 0.02 | 0.02 |
| $\theta$ | 1.1 | 1.1 | 1.1 | 1.2 | 1.2 | 1.2 |
| $\sigma$ | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| $\widetilde{P}(t,T)$ | 0.985116 | 0.941861 | 0.861094 | 0.990052 | 0.960849 | 0.905072 |
| $P_{RBT}(t,T)$ | 0.985146 | 0.941974 | 0.86135 | 0.990072 | 0.960926 | 0.905251 |
| $P_{FZ}(t,T)$ | 0.985116 | 0.941861 | 0.861094 | 0.990052 | 0.960849 | 0.905072 |
| $P_{Exp}(t,T)$ | 0.985116 | 0.941841 | 0.861042 | 0.990052 | 0.960738 | 0.904656 |
| $P_c^{ub}(t,T)$ | 0.985128 | 0.941968 | 0.861319 | 0.990059 | 0.95647 | 0.90193 |

**Table 3**: Bond prices $\widetilde{P}(t,T)$, $P_{RBT}(t,T)$, $P_{FZ}(t,T)$, $P_{Exp}(t,T)$, $P_{pl}(t,T)$, $P_c(t,T)$ ($\mathbf{M} = \bar{\mathbf{M}} = 500$)

| $T$(years) | 0.5 | 0.5 | 0.5 | 0.5 |
|---|---|---|---|---|
| $\widetilde{r}(=r^i)$ | 0.01 | 0.02 | 0.03 | 0.02 |
| $k$ | 0.8 | 0.5 | 1.1 | 1.2 |
| $\theta$ | 0.01 | 0.02 | 0.03 | 0.02 |
| $\sigma$ | 0.1 | 0.05 | 0.1 | 0.1 |
| $\widetilde{P}(t,T)$ | 0.995014 | 0.990051 | 0.985116 | 0.990052 |
| $P_{RBT}(t,T)$ | 0.995042 | 0.99007 | 0.985146 | 0.990072 |
| $P_{FZ}(t,T)$ | 0.995014 | 0.990051 | 0.985116 | 0.990052 |
| $P_{Exp}(t,T)$ | 0.995012 | 0.990051 | 0.985116 | 0.990052 |
| $P_{pl}(t,T)$ | 0.995012 | 0.990051 | 0.985067 | 0.989930 |
| $P_c(t,T)$ | 0.981921 | 0.990059 | 0.985041 | 0.989885 |

### 1.6.3 Barrier options (knock-out options)

Following the pricing approach in section 1.5.1 for barrier options, we are able to compare the barrier options price estimator variance rate for plain MC and MC with conditioning. With $N = 5$, we define the following $Q$ matrix:

$$Q = \begin{bmatrix} -190 & 30 & 25 & 50 & 85 \\ 5 & -185 & 140 & 25 & 15 \\ 5 & 135 & -230 & 70 & 20 \\ 5 & 40 & 165 & -230 & 20 \\ 5 & 10 & 45 & 60 & -120 \end{bmatrix}$$

For the other parameters, we set the (non-knocked-out) terminal payoff $H(X_T)$ as

$$H(X_T) = [1, 1, 1, 1, 1]^\top,$$

the *Down-and-Out* barrier level $L = x^2 \quad (x \in \{x^1, \cdots, x^5\})$, the interest rate values $r_t$ as

$$r_t = r(X_t) = [0.01, 0.015, 0.02, 0.025, 0.05],$$

the initial state $X_0 = x^4$ again with $x \in \{x^1, \cdots, x^5\}$, initial time $t = 0$ and terminal time $T = 1$ year.

*Down-and-Out* barrier options were hence computed using these specified parameters. Figure 1.1 shows a graph of the estimator with plain MC (dotted line) and MC with conditioning (dashed line), over the number of MC simulations, from 1000 to 10,000.

Figure 1.2 shows the estimator variance rate of plain MC (dotted line) and MC with conditioning (dashed line); the y-axis is given in $\log_{10}$ scale to improve readability. It is clear that

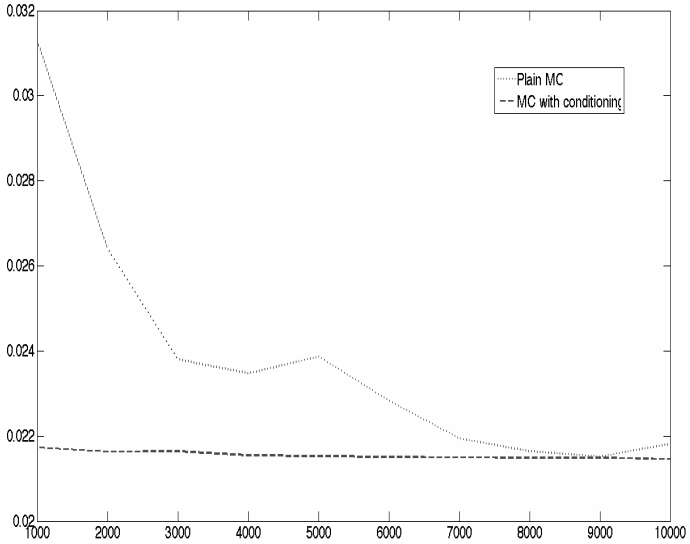


**Figure 1.1:** Estimator for barrier options using Plain MC (dotted line) and MC with conditioning (dashed line) over total number of MC simulations. Final estimated price is $\mathit{Bar}_{pl}(t,T) = 0.021811$, *std. dev*$= 0.14443$, and $\mathit{Bar}_c(t,T) = 0.021461$, *std. dev*$= 0.005472$, $n = 10000$.

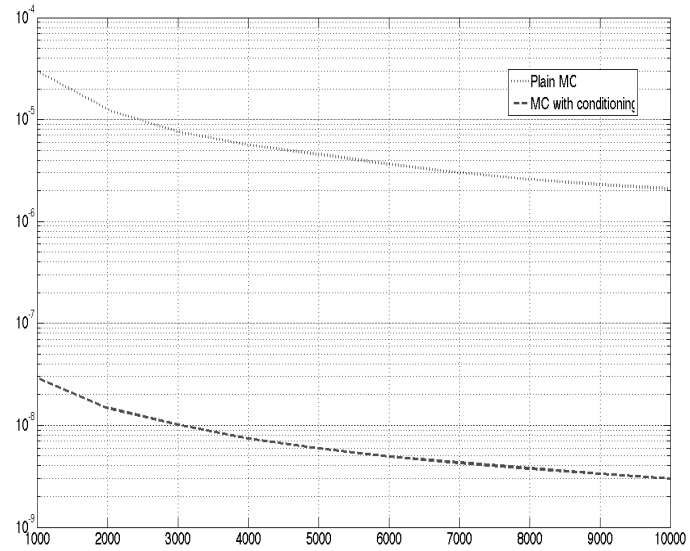


**Figure 1.2:** The estimator variance rate for barrier options using Plain MC (dotted line) and MC with conditioning (dashed line), shown on a $\log_{10}$ scaled y-axis, over the total number of simulations. For $n = 10000$ the Plain MC variance rate is $2.085972 \times 10^{-6}$ while the MC-with-conditioning variance rate is $2.994255 \times 10^{-9}$. The performance gain with conditioning is evident (lower is better).

MC with conditioning performs better for barrier options pricing compared to plain MC using estimator variance rate as our performance criterion.

### 1.6.4 Lookback options

We performed the MC pricing for lookback options as discussed in subsections 1.3.1 and 1.5.2 considering the case where $g(X_0^t) := \min_{s\leq t} X_s$ (Example 1) assuming, furthermore, that $\sigma = T$. This implies that, if the chain $X_t$ is time-homogeneous, also the bivariate chain $(X_t, Y_t)$ is and has thus a time homogeneous transition intensity matrix. Since the claim $H_T = (X_T - Y_T)^+$ is furthermore a simple claim, it is possible to compute the price also according to the analytic formula (1.1.2). We thus computed lookback prices for a plain MC ($\widetilde{LB}(t,T)$), for MC with conditioning ($LB(t,T)$). and according to formula (1.1.2) ($\bar{LB}(t,T)$) for different sets of parameters. We report here just two of the tests that we performed, which however show two quite different situations that may occur.

In both cases, we assumed a state space $E = [0.8, 0.9, 1.0, 1.1, 1.2]$ with $N = 5$; initial state $X_0 = 1.1$, and maturity $T = 2$ years. For Test 1, we used the following Q-matrix:

$$Q = \begin{bmatrix} -1440 & 360 & 360 & 360 & 360 \\ 7.2 & -28.8 & 7.2 & 7.2 & 7.2 \\ 0.72 & 0.72 & -2.88 & 0.72 & 0.72 \\ 2.52 & 2.52 & 2.52 & -10.08 & 2.52 \\ 480 & 480 & 480 & 480 & -1920 \end{bmatrix}$$

and, for Test 2:

$$Q = \begin{bmatrix} -0.8 & 0.2 & 0.2 & 0.2 & 0.2 \\ 0.7\bar{3} & -2.5\bar{3} & 0.\bar{6} & 0.6 & 0.5\bar{3} \\ 0.02 & 0.02 & -0.08 & 0.02 & 0.02 \\ 0.6 & 0.5\bar{3} & 0.\bar{6} & -2.4\bar{6} & 0.\bar{6} \\ 0.4 & 0.4 & 0.\bar{3} & 0.4 & -1.5\bar{3} \end{bmatrix}$$

where the over-bar denotes repeated decimals. Notice that the basic difference in the two cases consists in the fact that, in the first case, we obtain more frequent jumps/transitions of $X_t$.

The results are as follows: for Test 1, the plain MC estimate of the lookback price is $\widetilde{LB}(t,T) = 0.216839$ (std. dev $\tilde{\delta} = 0.128226$) (Figure 1.3), for the MC-with-conditioning the estimated price is $LB(t,T) = 0.194265$ (std. dev $\delta = 0.011603$), while $\bar{LB}(t,T) = 0.194707$. Note that $\widetilde{LB}(t,T) - \tilde{\delta} < LB(t,T) - \delta < LB(t,T) + \delta < \widetilde{LB}(t,T) + \tilde{\delta}$.

Figure 1.3 and Figure 1.4 show the graphs of the empirical mean as a function of the total iteration count of the MC simulation, together with a plot of the 1.5 standard deviations interval for $\widetilde{LB}(t,T)$ and $LB(t,T)$ in Test 1, respectively. For comparison, each of the Figures includes also the level corresponding to $\bar{LB}(t,T)$.

For Test 1, Figure 1.5 is a graph of the estimator variance rate over the total number of MC simulations for the plain MC estimator (solid line) and the MC-with-conditioning estimator (dashed line). Again here we see that the MC-with-conditioning performs better under the variance criterion.

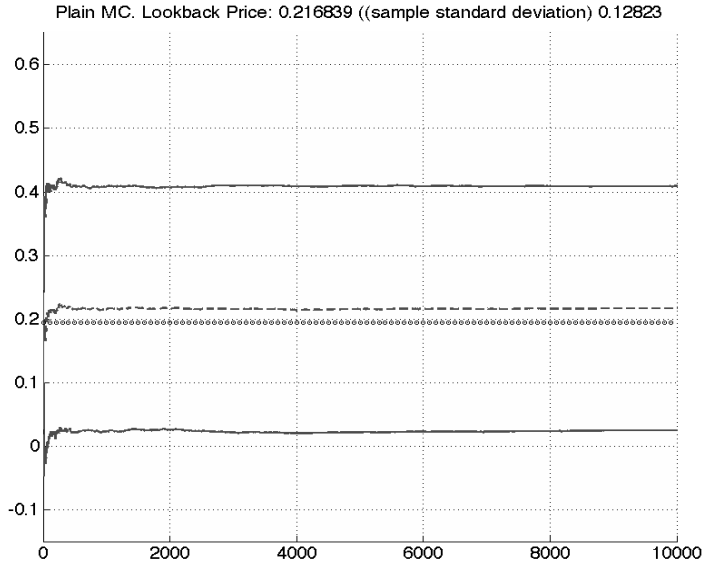


**Figure 1.3:** Plain MC estimator for lookback options showing the 1.5 standard deviation interval (Test 1). *lookback price* $\widetilde{LB}(0,2) = 0.216839$ (dashed line), *sample std dev* $= 0.128226$, $\bar{LB}(t,T) = 0.194707$ (dotted line), $n = 10000$

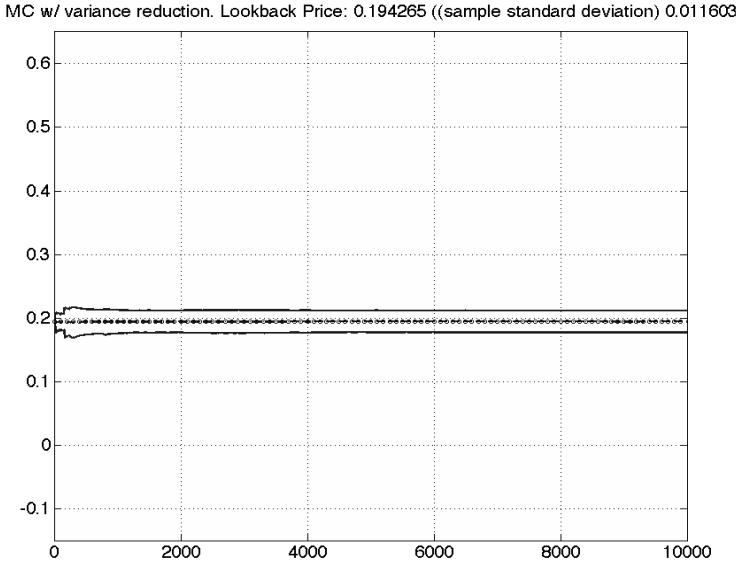


**Figure 1.4:** MC-with-Conditioning estimator for lookback options showing the 1.5 standard deviation interval (Test 1). *lookback price* $LB(0,2) = 0.194265$ (dashed line), *sample std dev* $= 0.011603$, $\bar{LB}(t,T) = 0.194707$ (dotted line), $n = 10000$

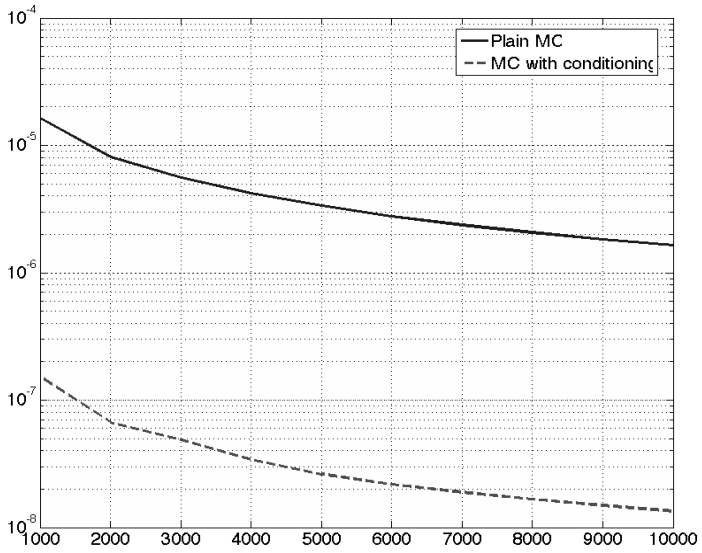


**Figure 1.5:** The estimator variance rate for lookback options using Plain MC (solid line) and MC with conditioning (dashed line), shown on a $\log_{10}$ scaled y-axis, over the total number of MC simulations (Test 1). For $n = 10000$ the Plain MC variance rate is $1.644202 \times 10^{-6}$ while the MC-with-conditioning variance rate is $1.346261 \times 10^{-8}$. The performance gain with conditioning is evident (lower is better).

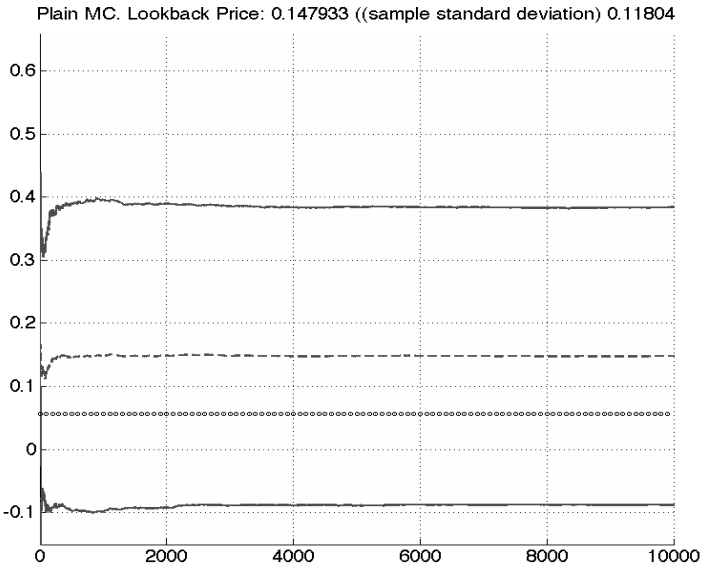


**Figure 1.6:** Plain MC estimator for lookback options showing the 2 standard deviation interval (Test 2). *lookback price* $= 0.147932$ (dashed line), *sample std dev* $= 0.118036$, $\bar{LB}(t,T) = 0.056398$ (dotted line), $n = 10000$

For Test 2, where the jump intensities are considerably lower, the lookback price with the plain MC estimator is $\widetilde{LB}(t,T) = 0.147932$ (std. dev $\tilde{\delta} = 0.118035$), the MC-with-conditioning price has empirical mean $LB(t,T) = 0.118050$ (std. dev $\delta = 0.037996$) and $\bar{LB}(t,T) = 0.056398$. Unlike in the previous test, we now have that $\widetilde{LB}(t,T) + \tilde{\delta} < LB(t,T) - \delta$ (see Figure 1.6 and Figure 1.7) and we need to plot the 2 standard deviation interval so that it includes $\bar{LB}(t,T)$. This may indicate that, since we have less frequent state transitions, it takes more simulations to obtain sufficiently reliable estimates. In fact, clustering problems may arise, i.e., the drawn samples in the simulation belong to just a small subset of the event space. This is a well known phenomenon that may lead to highly biased MC estimators. Indeed we find that in our MC tests, samples are sometimes not drawn over a large enough subset of the event space for the random variable $N_{t,T}$ - not every possible transition count in $[t,T]$ is achieved by the simulated paths (see Figures 1.8 and 1.9 which show the empirical distribution of jump counts for test 1 and test 2). As a result, many of the possible outcomes are weighted too low, causing the empirical mean, that is the weighted average, to be unreliable (see Figures 1.10 and 1.11 which show respectively the sampled price and the theoretical price for each possible jump count).

Nevertheless we still observe better performance for MC-with-conditioning under the empirical variance rate criterion as well as a considerably smaller squared bias with respect to the value $\bar{LB}(t,T)$ calculated from (1.1.2). Figure 1.12 shows in fact the estimator variance rate over the number of simulations, which just as before reflects the improved performance of MC-with-conditioning. This suggests that in order to meaningfully apply MC, in particular plain MC, it must be understood what conditions on the parameters may lead to models that are susceptible to clustering, and in that case, it may be better to apply a Quasi-MC approach.

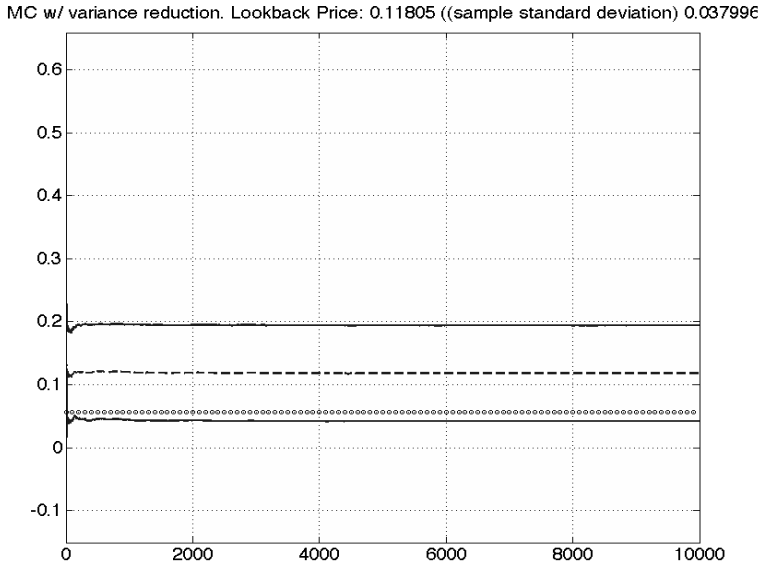


**Figure 1.7:** MC-with-Conditioning estimator for lookback options showing the 2 standard deviation interval (Test 2) *lookback price* $= 0.118050$ (dashed line), *sample std dev* $= 0.037996$, $\bar{LB}(t,T) = 0.056398$ (dotted line), $n = 10000$

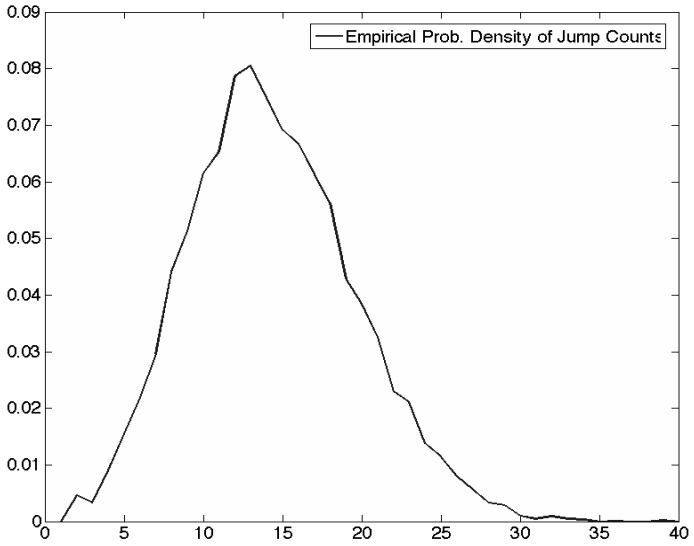


**Figure 1.8:** Empirical Distribution of Jump Counts for Test 1 samples, for $N = 0 \dots 40$ jumps

### 1.6.5 Asian options

The pricing for Asian options, as discussed in Section 1.5.3, can be performed using MC with conditioning by obtaining first a suitable bi-variate Markov process that, for simplicity of presentation, we choose here to be time-homogeneous. Following the procedure outlined in section 1.3.2 and restricting ourselves (see section 1.5.3) to events for which $\tau_n < T$, we first obtain the matrix $P$ of transition probabilities for $(X_t, Y_t)$ for the case when $\sigma = T$. Due to the restriction $\tau_n < T$ and the effect of the space discretization of $Y_t$, the matrix $P$ is not a *true* probability transition matrix of the Markov chain $(X_t, Y_t)$, i.e., its rows will not necessarily sum up to one. Nevertheless we can obtain the pricing formula to compute the right-most expression in (1.4.21) (see also (1.4.9)), using the $Q$-matrix that corresponds to $P$.

In this section, we present the prices obtained by plain MC according to (1.4.1) and by MC with conditioning according to (1.4.21). Recalling that the $Y-$component of the bi-variate Markov process $(X_t, Y_t)$ involves a time discretization (see subsection 1.3.2), we also wanted to investigate the effect of this discretization. For this purpose, thanks to the time homogeneity of $Q$ and the fact that, for $\sigma = T$, we obtain a simple claim, we include also a *proxy* for the theoretical price, by computing (1.1.2) on the basis of this $Q$-matrix. Here we call it "proxy" because it is the theoretical price for discretized $Y_t$ instead of the actual continuous one. We shall see that, indeed, we achieve a reduction in the estimator variance, as well as a reduction in the squared-bias with respect to the *proxy* for the theoretical price. We also tested the effect of varying the discretization level of $Y_t$.

For the CTMC X we have specified the following $Q$-matrix:

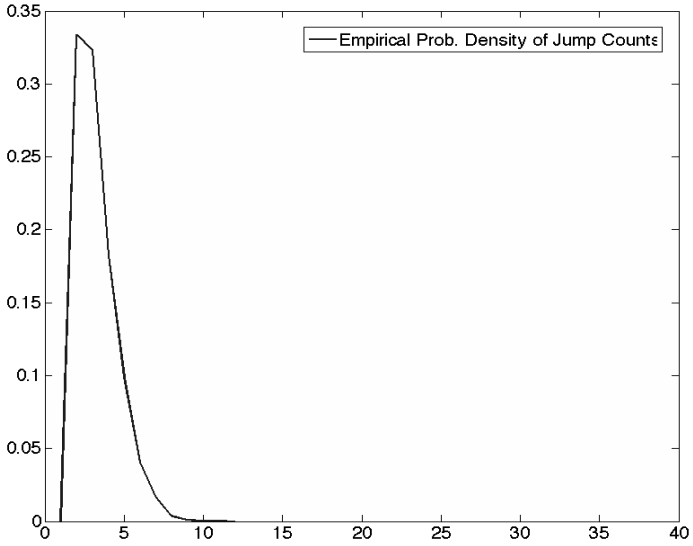


**Figure 1.9:** Empirical Distribution of Jump Counts for Test 2 samples, for $N = 0 \dots 40$ jumps

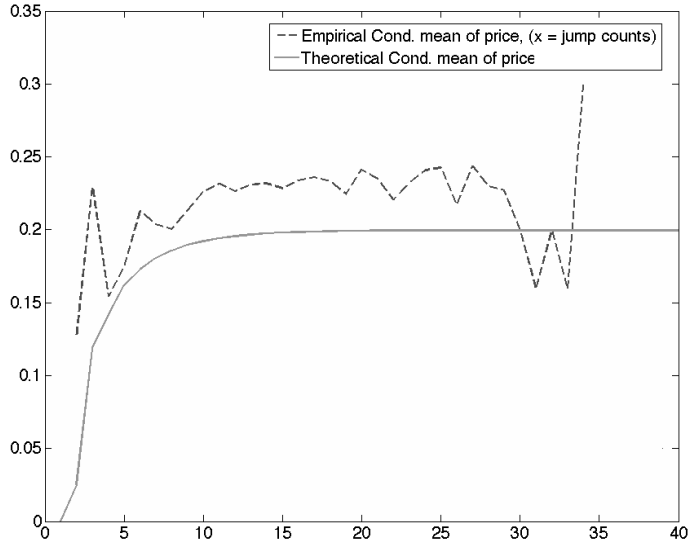


**Figure 1.10:** Conditional mean of prices per jump count (Test 1 samples). Dashed line - sample price (Plain MC); Solid line - theoretical price (MC -with-Conditioning)

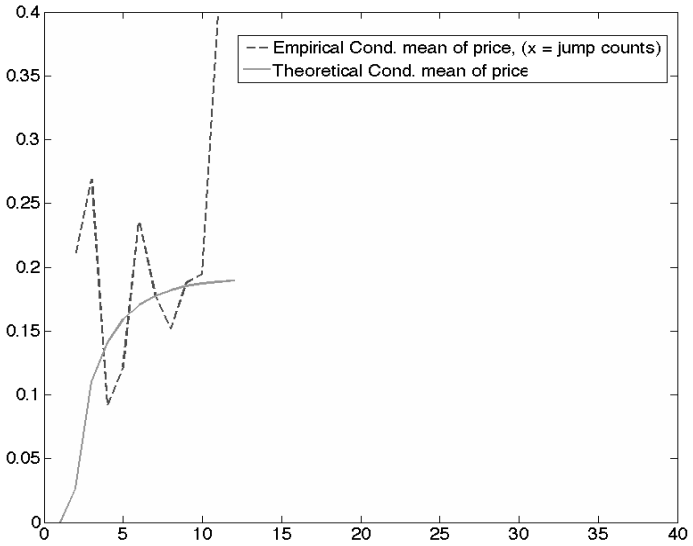


**Figure 1.11:** Conditional mean of prices per jump count (Test 2 samples). Dashed line - sample price (Plain MC); Solid line - theoretical price (MC-with-Conditioning)

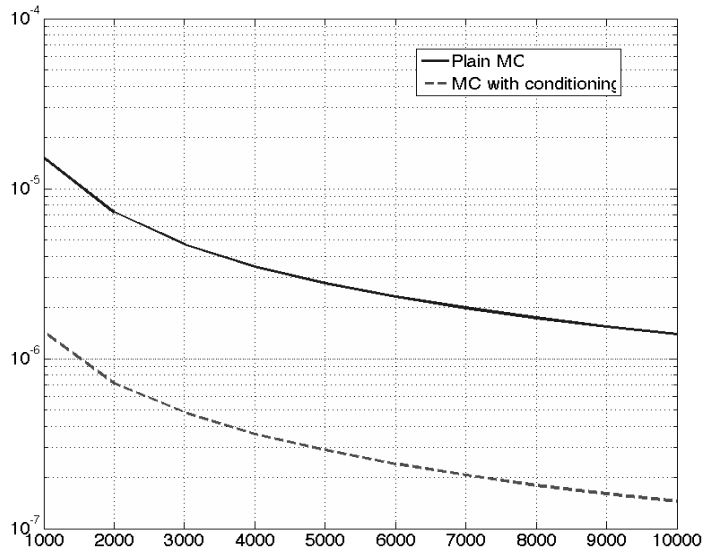


**Figure 1.12:** The estimator variance rate for lookback options using Plain MC (solid line) and MC with conditioning (dashed line), shown on a $\log_{10}$ scaled y-axis, over the total number of MC simulations (Test 2). For $n = 10000$ the Plain MC variance rate is $1.393242 \times 10^{-6}$ while the MC-with-conditioning variance rate is $1.443663 \times 10^{-7}$.The performance gain with conditioning is evident (lower is better).

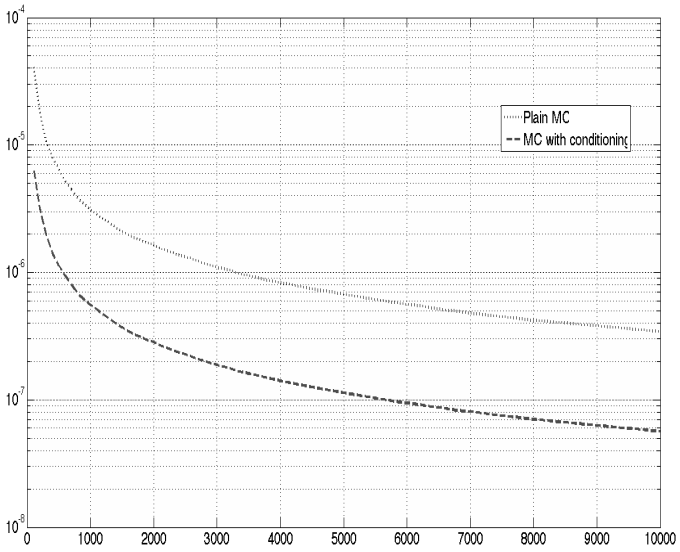


**Figure 1.13:** The variance rate for Asian options using Plain MC (dotted line) and MC with conditioning (dashed line), shown on a $\log_{10}$ scaled y-axis, over the total number of MC simulations. The discretization level is $K = 100$ for $Y_t$. For $n = 10000$ the Plain MC variance rate is $3.431721 \times 10^{-7}$ while the MC-with-conditioning variance rate is $5.65751 - \times 10^{-8}$. The performance gain with conditioning is evident (lower is better).

$$Q = \begin{bmatrix} -1.6 & 0.8 & 0.0\bar{3} & 0.2\bar{6} & 0.5 \\ 0.7\bar{6} & -2.0 & 0.2\bar{3} & 0.4\bar{6} & 0.5\bar{3} \\ 0.1\bar{3} & 0.2 & -1.7\bar{3} & 0.\bar{6} & 0.7\bar{3} \\ 0.\bar{3} & 0.4 & 0.6\bar{3} & -1.4\bar{6} & 0.1 \\ 0.3\bar{6} & 0.6 & 0.8\bar{3} & 0.0\bar{6} & -1.8\bar{6} \end{bmatrix}.$$

The other parameters are: $t = 0$, $T = 1.5$ years; $E = [0.8, 0.9, 1.0, 1.1, 1.2]^\top$, $R = [0.1, 0.12, 0.13, 0.14]^\top$, $X_0 = 1.1$, strike price $= 0.9$.

We start in Figure 1.13 by giving numerical evidence for the fact that by MC with conditioning we achieve a lower variance. We do this only for one discretized level of $K$ (for the definition of this level $K$ see subsection 1.3.2), since $K$ does not affect the frequency of jumps and thus the variance.

We also want to give some numerical evidence for the fact that MC with conditioning allows to reduce a possible bias. To this effect we compare the prices for MC with conditioning with those of the proxy for the theoretical price. We present two figures, Fig. 1.14 and Fig. 1.15 in order to furthermore show that the bias decreases with increasing level of $K$.

### 1.6.6 Computation Times

The Table 4 shows the example computation times in running each of the various numerical experiments. They represent the computational cost to obtain the performance corresponding to

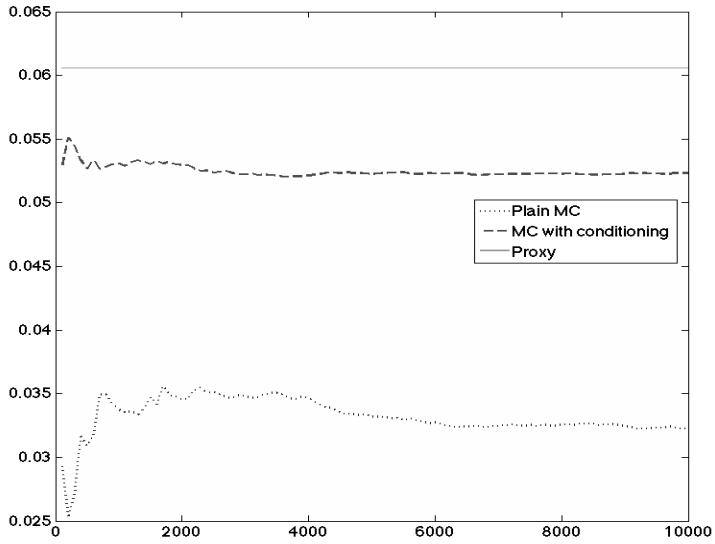


**Figure 1.14:** Estimator for Asian options using Plain MC (dotted line) and MC with conditioning (dashed line), over total number of MC simulations. We take $K = 25$ states for $Y_t$. The proxy for the theoretical price is also shown (solid line). Final estimated price is $AsianOpt_{pl}(t,T) = 0.032274$, *std. dev*$= 0.064392$, and $AsianOpt_c(t,T) = 0.052327$, *std. dev*$= 0.024023$, $n = 10000$. The value of the *proxy* for the theoretical price is $\tilde{AsianOpt}(t,T) = 0.060617$

each test as reported in Sections 1.6.2-1.6.5. The computations were performed using the software package *MATLAB Release 2012b* on a 2.6 GHz Intel Core i7 processor, with 16GB Memory, under an OS X Version 10.9.1 operating system.

For the MC pricing, we distinguish the time spent on the simulation of paths and the actual time for computing prices. It can be observed that whenever an explicit, analytic formula is available, that is, for simple claims such as zero-coupon bonds, the total computation time by analytic formula is lower compared to the MC, which is expected, in general. For the case of Asian Options, the times reported under the column *Analytic Formula* are indicated with an (*) since they refer to the *proxy* and not an actual analytic formula (see the discussion at the end of Section 1.5.3).

Comparing the computation times of plain MC to those of the MC with conditioning, it can be seen that the latter is slightly faster than the former in the case of Zero-Coupon Bonds, Barrier Options and Lookback Options. However, it is significantly slower than the former in the case of Asian Options (space discretization size $K = 25$, and $K = 100$). This can be explained by the fact that scientific computing packages such as MATLAB are generally optimized for matrix computations. However this computation advantage with matrices holds only up to a certain dimension. Indeed, pricing Asian Options under MC with conditioning involves high-dimensional matrices (namely, the matrices $\widetilde{Q}(n)$ ) representing such a case.

In summary, the MC with conditioning still presents a comparable, if not slightly better,

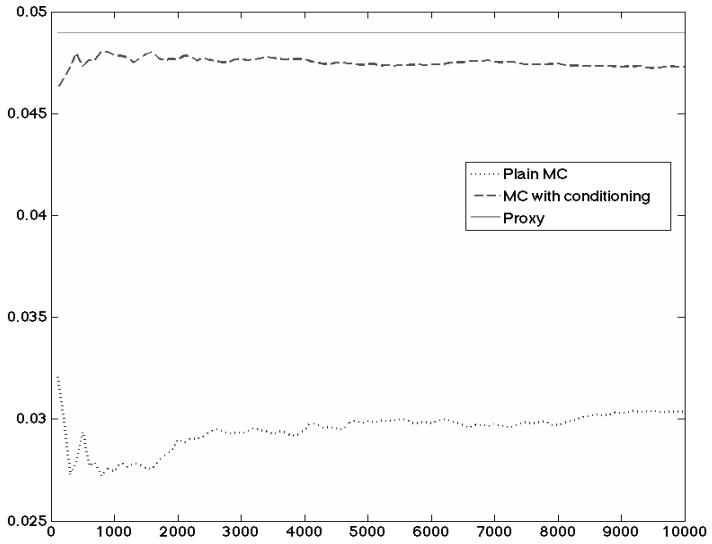


**Figure 1.15:** Estimator for Asian options using Plain MC (dotted line) and MC with conditioning (dashed line), over total number of MC simulations. We take $K = 100$ states for $Y_t$. The proxy for the theoretical price is also shown (solid line). Final estimated price is $\mathit{AsianOpt}_{pl}(t,T) = 0.030362$, *std. dev*$= 0.058581$, and $\mathit{AsianOpt}_{c}(t,T) = 0.047291$, *std. dev*$= 0.023786$, $n = 10000$. The value of the *proxy* for the theoretical price is $\tilde{AsianOpt}(t,T) = 0.048981$

method over plain MC, not only for variance reduction but also with respect to computational cost, except in cases such as Asian Options, that lead to very large $\widetilde{Q}(n)$ matrices. The added computational cost in that case is certainly undesirable but is the trade-off for the variance- and, possibly, bias - reduction. One usually applies pre-conditioning techniques on large matrices (see for example [Ben02]), however we did not apply this in the implementation our computer algorithms.

| Test | Simulation Time | Plain MC | MC with Conditioning | Analytic Formula |
|---|---|---|---|---|
| Zero-Coupon Bonds | 59.57 s | 0.602 s | 0.503 s | 0.00635 s |
| Barrier Options | 505.4 s | 1.803 s | 1.721 s | N.A. |
| Lookback Options | 15.774 s | 5.611 s | 2.321s | N.A. |
| Asian Options (K = 25) | 38.51 s | 3.212 s | 5.909 s | 0.0622* s |
| Asian Options (K = 100) | 36.82s | 3.182 s | 141.8 s | 0.293* s |

**Table 4**: Computation Time for the Numerical Tests

# Part II

# Calibration Aspects

# Chapter 2

# Calibrating Market Models when the underlying factor dynamics is a Continuous-Time Markov Chain

## 2.1 Introduction

Financial models are often specified under complete information (i.e. the investor filtration is the model filtration itself) and perfect knowledge of the true model parameters. These are, of course, idealized assumptions that are no longer true when implementing a model. Two related problems have therefore been tackled in order to arrive at more practical models. The first is the stochastic filtering problem that arises from specifying the model under incomplete information, and the second is the calibration problem, also known as system identification or parameter estimation in the control theory literature.

In the former, one typically assumes an incomplete investor filtration $\mathcal{F}_t$, such that the model values are measurable only under some enlarged filtration $\mathcal{G}_t \supseteq \mathcal{F}_t$. The investor is then constrained to perform computations such as those necessary for derivatives pricing, under the $\mathcal{F}_t$-conditional expectation. Hence there is a need to obtain the $\mathcal{F}_t$-conditional density $\pi_t(\cdot) := P[\cdot \mid \mathcal{F}_t]$. Given a state process $X$ and an observation process $Y$, that is a noise-corrupted function of $X$, if we define the investor filtration as $\mathcal{F}_t = \mathcal{F}_t^Y$, i.e. the observation process represents the investors' direct source of information, then the computation of optimal estimates $\widehat{f(X_t)} := E[f(X_t) \mid \mathcal{F}_t^Y]$ for any measurable function $f(\cdot)$, amounts to the *stochastic filtering problem.*

In the parameter estimation problem, on the other hand, one considers a parametrized family of probability measures $(P_\theta)_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}$ where the parameters $\boldsymbol{\theta}$ belong to some set of admissible parameters. Let us denote by $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}$ the true model parameters, namely, we suppose that the true or "physical" system induces the probability measure $P_{\theta^*}$. Next we assume that the investor *a priori* can only make an initial guess $\boldsymbol{\theta}_0$ of the model parameters (or equivalently, an initial guess $P_{\theta_0}$ of the model induced probability measure). The goal of parameter estimation is to obtain a parameter estimate $\tilde{\boldsymbol{\theta}}$ that induces $P_{\tilde{\theta}}$, an estimate of the probability measure $P_{\theta^*}$. We will adopt in this work a well-known recursive algorithm to obtain a sequence $(\tilde{\boldsymbol{\theta}}_k)_k$ of parameter estimates,

is the Expectation-Maximization (EM) algorithm, so named because in order to obtain $\tilde{\boldsymbol{\theta}}_k$ at each iteration of the algorithm, one must compute the expectation of the *log-likelihood* function $\log \frac{dP_{\theta_k}}{dP_{\tilde{\theta}}}$ and then perform the maximization of this expectation ([Wu83]).

In fact, the parameter estimation problem by EM is itself connected to the stochastic filtering problem. This is a central topic throughout the book [EAM08]. One approach that applies filter-based EM to the case of pricing payoffs $F(t, H, \xi)$ that depend on a CTMC state process $\xi$ with values in a finite set $E$ is due to [EHJ00]. The authors of [EHJ00] specify $\xi$ in continuous time and suppose that markets can only indirectly observe the price $F(t, H, \xi)$ of future claims $H$ depending on $\xi$, since the price is corrupted by multiplicative noise. Specifically, they assume a discrete-time i.i.d. Gaussian noise sequence $v_k = v(t_k)$ defined over the $L+1$ fixed observation times $0 = t_0 < t_1 < \dots t_L = T$. For each generic observation time $t_l$, the authors of [EHJ00] then define the time discretized observations $Y$ as the noise-corrupted *yield*, that is the instantaneous spot rate, of

$$F(t_l, H, \xi) e^{\eta(\xi_{t_l}) v_l} =: e^{-Y_l (T - t_l)}$$

for some function $\eta(\cdot)$. Since the observations are not defined for $t$ outside of $\{t_0, t_1, \dots, t_L\}$, in effect this approach only considers the time discretized $\xi_t$, i.e. the sequence $(\xi_l)_l = (\xi_{t_l})_l$ . [1]

Having defined the observation process $Y$ in this way, and denoting by $\boldsymbol{X}_n = \boldsymbol{e}_{\xi_n}$ where $\boldsymbol{e}_i$ is the $i$-th basis vector in $\boldsymbol{R}^N$ (i.e. $\boldsymbol{e}_1 = (1, 0, \dots)^\top$, $\boldsymbol{e}_2 = (0, 1, \dots 0)^\top$, etc.), [EHJ00] perform parameter estimation by filter-based EM on the system

$$\begin{cases} \boldsymbol{X}_n = \boldsymbol{Q} \boldsymbol{X}_{n-1} + \boldsymbol{M}_n \\ Y_n = -\frac{1}{T-t} \log F(t_n, H, \xi_n) - \frac{\sigma(\xi_n)}{T-t} v_n. \end{cases} \tag{2.1.1}$$

where $\boldsymbol{Q}$ is the generator matrix of $\xi_t$, and the dynamics for $\boldsymbol{X}$ come from its discrete-time martingale representation. The filter-based EM technique (see [EAM08]), is essentially an application of the Kalman filter.

Due to the fixed time-step discretization, the randomness in the jump times of the original (non-discretized) CTMC $\xi_t$ is effectively lost. We believe that this discrete-time set-up could lead to inefficient estimators for the CTMC $\xi$; indeed, to capture the random dynamics of a continuous trajectory via a time discretization with a deterministic choice of time points, a fine time step must be chosen. The true jump times of $\xi$ now occur very rarely, with respect to this fine time discretization, and hence the estimators for the model parameters in $\boldsymbol{Q}$ may become quite biased. This would create difficulties especially in Monte-Carlo based pricing.

One way to retain the randomness in the jump-times of the CTMC $X$ is to define the observation noise as a Brownian Motion instead of an i.i.d. Gaussian sequence. Then, to perform an EM-based parameter estimation algorithm, in the spirit of [EHJ00], we must use the *Wonham* filter (see [Won65]). In continuous time, the filter equations are given by a *Zakai* SDE. The filter also forms the basis of an EM-algorithm, analogous to the filter recursions in discrete time. In general a numerical scheme must be chosen to obtain the solutions of these SDEs. Here one

[1] The choice to define the observations process from the yield of a bond instead of its price is not an arbitrary choice - in practice, bonds are indeed quoted in terms of their yield. [EHJ00] note that a disadvantage of this set-up is the possibility to observe prices larger than 1 (negative yields). Here we remark that, while in the past, negative yields were considered to be an impossible phenomenon, more recently this phenomenon has also been observed, albeit still in rare cases. Hence this may now actually be considered an advantage of the model.

typically applies a fixed-time discretization such as the Euler-Maruyama or the Milstein scheme to discretize the SDE into recursive difference equations. These are so-called Ito Taylor approximation schemes. This is different from the approach in [EHJ00] where the filter model is itself discrete; here instead one has a continuous-time model on which a discretization is applied.

The standard numerical schemes approximate the solution to the SDE via the truncated Ito-Taylor expansion of the SDE's corresponding differential operator. This is analogous to the numerical solution of deterministic ODEs. These schemes suffer from a cumulative discretization error due to the fact that, like in the deterministic ODE case, they behave as *slope estimates*, in the sense that they estimate the next value by projecting it along the approximate tangent curve starting from the previous estimated value.Therefore to achieve sufficient accuracy, the fixed-time discretization must be fine, especially near the jump times. However, using a fixed-time discretization still causes information about the randomness of jumps to be lost. By using a *jump-time adapted discretization* one recovers the randomness due to the jumps. Such a random time discretization scheme first takes a fixed-time discretization, and then joins to its time points the set of random jump-times of $X$.

We then propose the following extension to [EHJ00], which is an alternative approach inspired by [PR10a]. Here we apply jump-adapted time discretization and investigate the use of standard numerical schemes as well as quasi-exact approximation schemes. The quasi-exact approximation is essentially a proxy for a highly accurate, exact explicit scheme, that incurs no discretization errors. This is due to the fact that an exact scheme approximates the process itself instead of the SDE. The term quasi-exact refers to the use of the exact - explicit solution if this is available or a quasi - exact solution otherwise. [PR10a] numerically solve the *Zakai* equation using such a jump - adpated quasi-exact approximation. They observed that such a scheme converges faster than an Euler-Maruyama or Milstein discretization scheme even for coarse time-discretization steps. [PR10a] treated the quasi-exact solutions of the Zakai SDE the state filters; however they did not treat the problem of the quasi-exact solutions to the Zakai SDE's for the EM-based estimation of parameters, which is the main contribution of this work.

The rest of Part II is structured as follows: Section 2.2 contains the background for the concepts and tools we will use. Section 2.3 focuses on the quasi-exact scheme, while Section 2.4 describes our proposed calibration model. Section 2.5 presents some results of a numerical test to check the effectiveness of our proposed algorithm for a simple case.

## 2.2 Basic Concepts and Tools: CTMCs, Filtering, Parameter Estimation, and Strong Approximations

In this Section we review the basic concepts and probabilistic tools we will use, such as CTMCs, Wonham filtering, parameter estimation by EM, and we will briefly recall the Euler-Maruyama and the Milstein scheme.

### 2.2.1 Basic facts concerning Continuous-Time Markov Chains

Let $(\Omega, \mathcal{F}, P)$ be a probability space with a Continuous-Time Markov Chain (CTMC) $\xi = (\xi_t)_{t\geq 0}$ with values in $E := \{x^1, x^2, \ldots, x^N\} \subset \mathbb{R}$ that we assume for the moment to be time-

homogeneous. We assume that $P$ is the risk-neutral measure under which the no-arbitrage condition is guaranteed, hence prices are computed under $P$.

The state values $x^1, \ldots, x^N$ are real scalars, but in general may be in $\mathbb{R}^N$ or in any arbitrary space. As we will see, it is convenient to consider an $N$-dimensional CTMC $\boldsymbol{X} = (\boldsymbol{X}_t)_{0\leq t\leq T}$ that is identical to $\xi$ under a state space transformation (see Section 2.2 of [EAM08]). We define the state space for $\boldsymbol{X}$ to be the set $\boldsymbol{E} := \{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_N\}$ of unit vectors in $\mathbb{R}^N$, with $\boldsymbol{e}_1 = (1, 0, 0, \ldots, 0)^\top$, $\boldsymbol{e}_2 = (0, 1, 0, \ldots, 0)^T$ and so on. Then we define the transformation from $\boldsymbol{X}$ to $\xi$ as

$$\xi_t = \boldsymbol{X}_t^\top \underline{\boldsymbol{x}} =: \langle \boldsymbol{X}_t, \underline{\boldsymbol{x}} \rangle$$

with $\underline{\boldsymbol{x}} = (x^1, x^2, \ldots, x^N)^\top \in \mathbb{R}^N$, and where we have denoted the scalar product of vectors by $\langle \cdot, \cdot \rangle$. On the other hand the transformation from $\xi$ to $\boldsymbol{X}$ is

$$\boldsymbol{X}_t = \boldsymbol{e}_{\xi_t}.$$

Under the probability measure $P$, $\boldsymbol{X}$ is characterized by a given $Q$-matrix $\boldsymbol{Q} = (q_{ij})_{1\geq i,j\geq N}$, where $q_{ii} = -\sum_{j\neq i} q_{ij}$ and the jump intensity of state $i$ is $q_i = -q_{ii}$. The transition probabilities $p_{ij} := P(\boldsymbol{X}_{t+s} = \boldsymbol{e}_j \mid \boldsymbol{X}_t = \boldsymbol{e}_i)$ from state $i$ to state $j$ for $i, j \in E$ and $\forall t, s \geq 0$ are defined as

$$p_{ij} := \begin{cases} \frac{q_{ij}}{q_i} & \text{if} \quad i \neq j \\ 0 & \text{if} \quad i = j. \end{cases} \tag{2.2.1}$$

We denote the jump times as $\tau_1 < \tau_2, \ldots \tau_\nu < T$ where the distribution of the jump time intervals is $(\tau_{k+1} - \tau_k \mid \boldsymbol{X}_k = \boldsymbol{e}_i) \sim \mathcal{E}xp(q_i)\ \forall k \geq 0$. We also denote the sequences $\boldsymbol{X}_n$ and $\xi_n$ of the values of $\boldsymbol{X}$ and $\xi$ respectively, i.e. $X_n = X_t$ if $\tau_n \leq t < t$ and analogously for $\xi$.

Associated to the jump process $\boldsymbol{X}_t$ (and $\xi_t$) are the state-wise jump counts $N_t^j$, $j = 1, \ldots, N$ defined by

$$N_t^j = \sum_{k=0}^{\infty} \mathbf{1}_{\{\tau_k \leq t\}} \mathbf{1}_{\{\xi_k = j\}} = \sum_{k=0}^{\infty} \mathbf{1}_{\{\tau_k \leq t\}} \mathbf{1}_{\{\boldsymbol{X}_k = \boldsymbol{e}_j\}},$$

and the counting process $N_t = \sum_{i=1}^N N_t(i)$.

Functions of the type $f(t, \xi_t) = \sum_{i=1}^N f^i(t) \mathbf{1}_{\{\xi_t = i\}}$ are equivalent to $f(t, \boldsymbol{X}_t)$ with the vector notation

$$f(t, \boldsymbol{X}_t) = \langle \boldsymbol{f}(t), \boldsymbol{X}_t \rangle,$$

where $\boldsymbol{f}(t) = (f^1(t), f^2(t), \ldots, f^N(t))^\top$.

**Martingale Representation** We state two known results concerning martingale representations for CTMCs.

**Theorem 2.2.1** (Martingale Representation Theorem for Markov Chains)**.** *The process $M_t$ defined as*

$$\boldsymbol{M}_t = \boldsymbol{X}_t - \boldsymbol{X}_0 - \int_0^t \boldsymbol{Q}\boldsymbol{X}_s ds \tag{2.2.2}$$

*is a martingale.*

*Proof.* This is a standard result in Markov Chain theory due to Dynkin (see Appendix B of the book [EAM08] for an example of a proof). □

Let us also recall the following

**Proposition 2.2.2** (Martingale Representation Theorem for functions of a Markov Chain)**.** *A function $f(t, \boldsymbol{X}_t)$ that is differentiable in $t$ and which has the vector form*

$$f(t, \boldsymbol{X}_t) = \boldsymbol{f}(t)^\top \boldsymbol{X}_t = \langle \boldsymbol{f}(t), \boldsymbol{X}_t \rangle$$

*where $\boldsymbol{f}(t) = (f_1(t), \ldots, f_N(t))^\top$, satisfies the following differentiation rule and representation result:*

$$f(t, \boldsymbol{X}_t) = f(0, \boldsymbol{X}_0) + \int_0^t \langle \dot{\boldsymbol{f}}(s), \boldsymbol{X}_s \rangle ds + \int_0^t \langle \boldsymbol{f}(s), \boldsymbol{Q}\boldsymbol{X}_{s-} \rangle ds + \int_0^t \langle \boldsymbol{f}(s), d\boldsymbol{M}_s \rangle. \tag{2.2.3}$$

*Here $\int_0^t \langle \boldsymbol{f}(s), d\boldsymbol{M}_s \rangle$ is an $\mathcal{F}_t$-martingale.*

*Proof.* See Appendix B of the book [EAM08]. This representation result is just a corollary of the generalized Ito formula for semi-martingales and an application of Theorem 2.2.1. □

**Remark 2.2.3.** *Note that the preceding differentiation rule can be remembered easily by substituting the differential $d\boldsymbol{X}_t$. Hence, equation (2.2.3) is just*

$$f(t, \boldsymbol{X}_t) = f(0, \boldsymbol{X}_0) + \int_0^t \langle \dot{\boldsymbol{f}}(s), \boldsymbol{X}_s \rangle ds + \int_0^t \langle \boldsymbol{f}(s), d\boldsymbol{X}_s \rangle \tag{2.2.4}$$

### 2.2.2 Wonham Filters

Let us now suppose that the CTMC $\boldsymbol{X}_t$ represents the hidden state that one observes the process $y_t$. We shall assume that $\boldsymbol{X}_t$ and $y_t$ satisfy a stochastic dynamical system of the type

$$\begin{aligned} \boldsymbol{X}_t &= \boldsymbol{X}_0 + \int_0^t \boldsymbol{Q}\boldsymbol{X}_s ds + M_t \\ y(t) &= \int_0^t h(s, \boldsymbol{X}_s) ds + \int_0^t \eta dW_s. \end{aligned} \tag{2.2.5}$$

We suppose throughout this work that the parameters that enter into the dynamics of $\boldsymbol{X}$ and $y$ satisfy sufficient conditions that guarantee a well-defined solution to (2.2.5). In particular we shall assume that $h(t, \boldsymbol{X}_t)$ is a bounded (and hence also square integrable) process.

Systems of the type (2.2.5) form the basis for the *Wonham filter problem*, namely, compute the estimates $\widehat{\boldsymbol{X}_t} = E[\boldsymbol{X}_t \mid \mathcal{F}_t^y]$. The derivation of Wonham filters can be obtained through the change-of-measure technique, which we present in the Appendix 2.A.1; here we shall only state the filter equations[2].

#### 2.2.2.1 Kallianpur-Striebel Formula

Define a probability measure $\bar{P}$ by setting

$$d\bar{P} = \Lambda_T^{-1} dP$$

[2]For more details on deriving the filters see the original paper [Won65], Chapter 8 of the textbook [EAM08] or Chapter 7 of the lecture notes [Han07b].

with

$$\Lambda_T = \exp\left\{-\frac{1}{2}\int_0^T \eta^{-2}\|h(s,\boldsymbol{X}_s)\|^2 + \int_0^T \eta^{-1}h(s,\boldsymbol{X}_s)dy_s\right\}.$$

Note that by the boundedness of $h(t,\boldsymbol{X}_t)$, we have that $E[\Lambda_T]=1$. By the Girsanov Theorem (see Theorem 2.2.9 below or Theorem 8.6.5 of the book [Øks10]) the process $\bar{y} := \eta^{-1}y$ is a $\bar{P}$-Wiener process. Here $\Lambda_t = \frac{dP}{d\bar{P}}\big|_{\mathcal{F}_t}$ is the corresponding Radon-Nikodym derivative from $\bar{P}\sim P$ restricted to $\mathcal{F}_t$. By applying the Ito formula under $\bar{P}$ we have that $\Lambda_t$ satisfies

$$\Lambda_t = 1 + \int_0^t \Lambda_s \cdot \eta^{-1}\cdot h(s,\boldsymbol{X}_s)dy_s, \tag{2.2.6}$$

and hence $\Lambda$ is a $\bar{P}$-martingale.

For each $i\in E$ let us define the *unnormalized $P$-conditional probability* $\sigma^i(\xi_t)$ for the CTMC to be in state $x^i\in E$ at time $t$, by the $\bar{P}$-conditional expectation[3]

$$\sigma^i(\xi_t) := \bar{E}[\mathbf{1}_{\{\xi_t=i\}}\Lambda_t \mid \mathcal{F}_t^y],$$

for $t\in[0,T]$. It follows from the *Kallianpur-Striebel formula*, see [FKK72], that the $P$-conditional probabilities of $\xi_t$ given $\mathcal{F}_t^y$ are

$$P(\xi_t = i \mid \mathcal{F}_t^y) = E[\mathbf{1}_{\{\xi_t=i\}} \mid \mathcal{F}_t^y] = \frac{\sigma^i(\xi_t)}{\sum_{j=1}^N \sigma(\xi_t)^j},$$

for $t\in[0,T]$. We can write the vector process $\boldsymbol{\sigma}(\xi_t) = (\sigma^1(\xi_t),\dots,\sigma^N(\xi_t))^\top$ and define $\sigma(\boldsymbol{X}_t)$ by

$$\sigma(\boldsymbol{X}_t) = \langle \boldsymbol{\sigma}, \boldsymbol{X}_t\rangle = \boldsymbol{\sigma}(\xi).$$

The least-squares estimate at time $t$ for $f(\xi_t)$ with respect to the observations available at time $t$, i.e. with respect to $\mathcal{F}_t^y$, is then the *Wonham filter*, which is given by the $P$-conditional expectation

$$\pi_t(f) = E[f(\xi_t)\mid \mathcal{F}_t^y] = \frac{\sum_{i=1}^N f(i)\sigma(\xi_t)^i}{\sum_{i=1}^N \sigma(\xi_t)^i} \tag{2.2.7}$$

for $t\in[0,T]$.

#### 2.2.2.2 The Zakai Equations

From equation (2.2.7) we see that for a given measurable function $f$, the Wonham filter for $f(\xi_t)$ can be computed by knowing the unnormalized conditional expectation $\sigma(\xi_t)$. This unnormalized conditional expectation satisfies an SDE, the Zakai equation, whose derivation we present in Appendix 2.A.1, following the proof of Theorem 3.2 in Chapter 8 of [EAM08].

[3]The $P$-conditional probability (normalized) of $\xi_t = i$ is

$$P[\xi_t = i \mid \mathcal{F}_t^y] = \frac{\bar{E}[\mathbf{1}_{\{\xi_t=i\}}\Lambda_t \mid \mathcal{F}_t^y]}{\bar{E}[\Lambda_t \mid \mathcal{F}_t^y]}$$

while the unnormalized $P$-conditional probability is

$$\bar{E}[\mathbf{1}_{\{\xi_t=i\}}\Lambda_t \mid \mathcal{F}_t^y] := \sigma^i(\xi_t).$$

**Proposition 2.2.4** (Zakai Equation for the state $\boldsymbol{X}_t$)**.** *The unnormalized conditional expectation* $\sigma(\boldsymbol{X}_t) = \bar{E}\left[\Lambda_t \boldsymbol{X}_t \mid \mathcal{F}_t^y\right]$ *satisfies the following SDE*

$$\begin{cases} d\boldsymbol{\sigma}(\xi_t) = & \boldsymbol{Q}\boldsymbol{\sigma}(\xi_t)dt + \bar{\boldsymbol{D}}(t)\boldsymbol{\sigma}(\xi_t)d\bar{y}_t, \\ \sigma(\boldsymbol{X}_0) = & P(\boldsymbol{X}_0), \end{cases} \tag{2.2.8}$$

*for* $t \in [0,T]$*. This is a homogeneous linear Ito SDE in* $\bar{P}$*.*

*Proof.* See Appendix 2.A.1.1 □

For the parameter estimation problem that we will discuss in Section 2.2.3, we shall need some more Zakai equations.

#### 2.2.2.3 The Zakai Equation for the Transition Count Process $N_t^{i,j}$

Define $N_t^{i,j}$ the counting process at time $t$ for transitions of $\xi_s$ from $i$ to $j$ for $s \in [0,t)$. Define the martingale $M_t^{i,j}$ as

$$M_t^{i,j} := \int_0^t \langle \boldsymbol{X}_{s-}, \boldsymbol{e}_i \rangle \boldsymbol{e}_j^\top dM_s \tag{2.2.9}$$

We then have, from the martingale representation of $\boldsymbol{X}$, that

$$M_t^{i,j} = \int_0^t \langle \boldsymbol{X}_{s-}, \boldsymbol{e}_i \rangle \boldsymbol{e}_j^\top d\boldsymbol{X}_s - \int_0^t \langle \boldsymbol{X}_{s-}, \boldsymbol{e}_i \rangle \boldsymbol{e}_j^\top \boldsymbol{Q}\boldsymbol{X}_{s-} ds.$$

Note that the first term is

$$\int_0^t \langle \boldsymbol{X}_{s-}, \boldsymbol{e}_i \rangle \boldsymbol{e}_j^\top d\boldsymbol{X}_s = \int_0^t \mathbf{1}_{\{\xi_{s-}=i\}} \boldsymbol{e}_j^\top d\boldsymbol{X}_s = N_t^{i,j}$$

and the integrand in the second term simplifies to

$$\langle \boldsymbol{X}_{s-}, \boldsymbol{e}_i \rangle \boldsymbol{e}_j^\top \boldsymbol{Q}\boldsymbol{X}_{s-} = \langle \boldsymbol{X}_{s-}, \boldsymbol{e}_i \rangle q_{ij}.$$

Therefore,

$$N_t^{i,j} = \int_0^t \langle \boldsymbol{X}_{s-}, \boldsymbol{e}_i \rangle q_{ij} ds + \int_0^t \langle \boldsymbol{X}_{s-}, \boldsymbol{e}_i \rangle \boldsymbol{e}_j^\top dM_s. \tag{2.2.10}$$

observe that $N_t^{i,j} = \langle N_t^{i,j}\boldsymbol{X}_t, \underline{\mathbf{1}} \rangle$, and hence

$$\sigma(N_t^{i,j}) = \sigma(\langle N_t^{i,j}\boldsymbol{X}_t, \underline{\mathbf{1}} \rangle) = \langle \sigma(N_t^{i,j}\boldsymbol{X}_t), \underline{\mathbf{1}} \rangle,$$

so that it is enough to compute $\sigma(N_t^{i,j}\boldsymbol{X}_t)$ which is given in the following

**Theorem 2.2.5.** *The counting process for transitions from* $i$ *to* $j$*,* $\sigma(N_t^{i,j}\boldsymbol{X}_t)$*, satisfies the following Zakai equation:*

$$\begin{cases} d\sigma(N_t^{i,j}\boldsymbol{X}_t) = & \left\{ \langle \sigma(\boldsymbol{X}_t), \boldsymbol{e}_i \rangle q_{i,j} \boldsymbol{e}_j + \boldsymbol{Q}\sigma(N_t^{i,j}\boldsymbol{X}_t) \right\} dt + \bar{\boldsymbol{D}}(t)\sigma(N_t^{i,j}\boldsymbol{X}_t)d\bar{y}_t \\ \sigma(N_0^{i,j}\boldsymbol{X}_0) = & 0. \end{cases} \tag{2.2.11}$$

*Proof.* See Appendix 2.A.1.2.

□

#### 2.2.2.4 The Zakai Equation for the State Occupation - Time $\mathcal{T}_t(i)$

Next, define the occupation time $\mathcal{T}_t(i)$ of state $i$ in $[0,t)$,that is

$$\mathcal{T}_t(i) := \int_0^t \langle \boldsymbol{X}_s, \boldsymbol{e}_i \rangle ds, \quad 1 \leq i \leq N.$$

We have analogously that

$$\mathcal{T}_t(i) = \langle \mathcal{T}_t(i)\boldsymbol{X}_t, \underline{\mathbf{1}} \rangle,$$

hence also

$$\sigma(\mathcal{T}_t(i)) = \langle \sigma(\mathcal{T}_t(i)\boldsymbol{X}_t), \underline{\mathbf{1}} \rangle.$$

**Theorem 2.2.6.** *The Zakai equation for the unnormalized conditional expectation $\sigma(\mathcal{T}_t(i)\boldsymbol{X}_t)$ is*

$$\begin{cases} d\sigma(\mathcal{T}_t(i)\boldsymbol{X}_t) = & \{\langle \sigma(\boldsymbol{X}_t), \boldsymbol{e}_i \rangle \boldsymbol{e}_i + \boldsymbol{Q}\sigma(\mathcal{T}_t(i)\boldsymbol{X}_t)\}\, dt + \bar{\boldsymbol{D}}(t)\sigma(\mathcal{T}_t(i)\boldsymbol{X}_t) d\bar{y}_t \\ \sigma(\mathcal{T}_0(i)\boldsymbol{X}_0) = & 0. \end{cases} \tag{2.2.12}$$

*Proof.* See Appendix 2.A.1.3. □

#### 2.2.2.5 The Zakai Equation for the Drift Estimator $\Gamma_t(i)$

Lastly, we define the process $\Gamma_t(i)$ under $P$ as

$$\Gamma_t(i) := \int_0^t \langle \boldsymbol{X}_s, \boldsymbol{e}_i \rangle dy_s = \int_0^t h^i \langle \boldsymbol{X}_s, \boldsymbol{e}_i \rangle ds + \int_0^t \langle \boldsymbol{X}_s, \boldsymbol{e}_i \rangle \sigma dW_s,$$

which we shall use in the estimator for the drift components $h^i(s)$ of the vector $\boldsymbol{h}(s)$ that enters in the drift of $y_t$. Hence we shall call $\sigma(\Gamma_t(i))$ the *unnormalized drift component estimator* or simply *estimator for the drift*. Again we have that

$$\Gamma_t(i) = \langle \Gamma_t(i)\boldsymbol{X}_t, \underline{\mathbf{1}} \rangle$$

hence in the same manner as with $\sigma(N_t^{i,j})$ and $\sigma(\mathcal{T}_t(i))$ the following holds:

$$\sigma(\Gamma_t(i)) = \langle \sigma(\Gamma_t(i)\boldsymbol{X}_t), \underline{\mathbf{1}} \rangle.$$

**Theorem 2.2.7.** *The unnormalized conditional expectation $\sigma(\Gamma_t(i)\boldsymbol{X}_t)$ satisfies the following Zakai equation*

$$\begin{cases} d\sigma(\Gamma_t(i)\boldsymbol{X}_t) = & \left\{h^i \langle \sigma(\boldsymbol{X}_s), \boldsymbol{e}_i \rangle \boldsymbol{e}_i + \boldsymbol{Q}\sigma(\Gamma_t^i \boldsymbol{X}_r)\right\} dt + \left\{\bar{\boldsymbol{D}}(t)\sigma(\Gamma_t(i)\boldsymbol{X}_t) + \langle \sigma(\boldsymbol{X}_t), \boldsymbol{e}_i \rangle \boldsymbol{e}_i\right\} d\bar{y}_t \\ \sigma(\Gamma_0(i)\boldsymbol{X}_0) = & 0. \end{cases} \tag{2.2.13}$$

*Proof.* See Appendix 2.A.1.4. □

We have seen that computing the Wonham filter amounts to computing the solutions of certain Zakai SDEs that, in general, have to be performed numerically. In the next section, we show an application of Wonham filters in treating the problem of parameter estimation of a model of the type (2.2.5).

### 2.2.3 Parameter Estimation

In this section we will show that the parameter estimation of systems of the type (2.4.22) can be done on the basis of the Wonham filter, which enters in the context of a recursive, EM algorithm. This is discussed in Ch. 8 of [EAM08] and is based on the work of [DZ86] and [ZD88]. Here we give a review of this procedure.

To solve the parameter estimation problem, one must specify an objective function $\boldsymbol{Q}(\boldsymbol{\theta}^*, \boldsymbol{\theta}_k)$ and a sequence of maximizing parameter estimates $(\tilde{\boldsymbol{\theta}}_k)_k \in \boldsymbol{\Theta}$ such that

$$\tilde{\boldsymbol{\theta}}_k := \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \boldsymbol{Q}(\boldsymbol{\theta}_{k-1}, \boldsymbol{\theta})$$

converges to $\boldsymbol{\theta}^*$, i.e.

$$\left|\tilde{\boldsymbol{\theta}}_k - \boldsymbol{\theta}^*\right| \to 0$$

for some norm $|\cdot|$ on $\boldsymbol{\Theta}$. A milder criterion that typically leads to an easier choice of an approximating sequence $(\tilde{\boldsymbol{\theta}}_k)_k$ is that

$$\left|\tilde{\boldsymbol{\theta}}_k - \tilde{\boldsymbol{\theta}}_{k-1}\right| \to 0. \tag{2.2.14}$$

This leads to a recursive algorithm to estimate $\boldsymbol{\theta}^*$. Usually a stopping criterion $\epsilon$ is chosen so that the algorithm stops when

$$\left|\tilde{\boldsymbol{\theta}}_k - \tilde{\boldsymbol{\theta}}_{k-1}\right| < \epsilon.$$

It has been shown in [Wu83] that assuming $\boldsymbol{\Theta}$ to be a compact Euclidean subset and that specifying $\boldsymbol{Q}(\tilde{\boldsymbol{\theta}}_{k-1}, \boldsymbol{\theta})$ to be the log-likelihood, i.e.

$$\boldsymbol{Q}(\tilde{\boldsymbol{\theta}}_{k-1}, \boldsymbol{\theta}) = E\left[\log \left.\frac{dP_{\theta_{k-1}}}{dP_\theta}\right|_{\mathcal{F}_t} \mid \mathcal{F}_t^y\right] \tag{2.2.15}$$

leads to a converging algorithm under the criteria (2.2.14). It is then our interest in this section to compute and maximize (2.2.15) with respect to each of our parameters.

Let us define the parameter vector $\boldsymbol{\theta}$ as

$$\boldsymbol{\theta} := (q_{i,j}, 1 \le i, j \le N, h^i, 1 \le i \le N)^\top.$$

Suppose we have the latest estimate

$$\boldsymbol{\theta} = (q_{i,j}, h^i, 1 \le i, j \le N)$$

and we wish to determine the updated estimates

$$\tilde{\boldsymbol{\theta}} = (\tilde{q}_{i,j}, \tilde{h}^i, 1 \le i, j \le N)$$

which maximizes the log-likelihood in (2.2.15). To compute such maximum, first we observe that the process $dP_{\theta_{k-1}}/dP_\theta|_{\mathcal{F}_t}$ appearing in the likelihood in (2.2.15) is the Radon-Nikodym derivative that enters in an absolutely continuous measure transformation from $P_\theta$ to $P_{\theta_{k-1}}$. There are two components to such a measure transformation, the first is the change in the intensity of the CTMC and the second is the Girsanov Transformation to translate the drift term.

#### 2.2.3.1 Change Of Measure for CTMCs

Suppose the given $E$-valued, Time-Homogeneous CTMC $\xi_t$, defined on the finite time interval $[0,T]$, is specified under $P$ by a given intensity matrix $\boldsymbol{Q} = (q_{i,j})_{1\leq i,j\leq N}$. Let us suppose that the given $q'_{i,j}s$ correspond to the elements of $\boldsymbol{\theta}$, and hence we define our given measure $P$ itself to be $P_\theta$.

**Theorem 2.2.8.** *Define $L_t^{i,j}$, for $i,j = 1\dots N$ as*

$$L_t^{i,j} =: \left(\frac{\tilde{q}_{i,j}}{q_{i,j}}\right)^{N_t^{i,j}} \exp\left\{\int_0^t (q_{i,j} - \tilde{q}_{i,j})\langle \boldsymbol{X}_{s-}, \boldsymbol{e}_i\rangle ds\right\}, \tag{2.2.16}$$

*and let*

$$L_t := \prod_{i\neq j} L_t^{i,j} \tag{2.2.17}$$

*Let $\overline{P}$ be the measure defined by the Radon-Nikodym derivative*

$$\frac{d\overline{P}}{dP}|_{\mathcal{F}_t} := L_t.$$

*Then under $\overline{P}$, $\xi_t$ is a CTMC with intensity matrix $\widetilde{\boldsymbol{Q}} = (\tilde{q}_{i,j})_{1\leq i,j\leq N}$. The transition probability matrix $\widetilde{\boldsymbol{P}} = (\tilde{p}_{i,j})_{1\leq i,j\leq N}$ under $\overline{P}$ is determined from $\widetilde{\boldsymbol{Q}}$ in the usual manner (see (2.2.1)).*

*Proof.* See Proposition 11.2.3 of [BR02]. □

According to Theorem 2.2.8, the probability measure $\overline{P}(\cdot) := \dfrac{E\left[L_t \mathbf{1}_{\{\cdot\}}\right]}{E\left[L_t\right]}$ is the probability measure equivalent to $P$ under which the process $\xi_t$ is a CTMC with the desired new state transition intensities $\tilde{q}_{i,j}$. We still need the change of measure to translate the drift parameter.

#### 2.2.3.2 Change Of Measure to Translate the Drift

We recall the Girsanov Theorem for Ito processes allows us to apply an absolutely continuous measure change $\frac{dP}{d\overline{P}}\Big|_{\mathcal{F}_t}$ that transforms the drift from $\boldsymbol{h}(t) = (h^1(t),\dots,h^N)^\top$ (under $\overline{P}$) to $\tilde{\boldsymbol{h}}(t) = (\tilde{h}^1(t),\dots,\tilde{h}^N(t))^\top$ (under $P$). Indeed, let us suppose that the $\overline{P}$-dynamics of $y_t$ satisfy

$$dy_t = \langle \boldsymbol{h}(t), \boldsymbol{X}_t\rangle dt + \eta dW_t^\theta$$

Let us further recall the Girsanov Theorem (see Theorem 8.6.5 of [Øks10]),

**Theorem 2.2.9.** *If one defines $\frac{d\widetilde{P}}{d\overline{P}}\Big|_{\mathcal{F}_t} := \Lambda_t$ where*

$$\Lambda_t := \exp\left\{\int_0^t \langle \boldsymbol{X}_s, \eta^{-1}(\tilde{\boldsymbol{h}} - \boldsymbol{h})\rangle dW_s^\theta - \frac{1}{2}\int_0^t (\langle \boldsymbol{X}_s, \eta^{-1}\tilde{\boldsymbol{h}}\rangle^2 - \langle \boldsymbol{X}_s, \eta^{-1}\boldsymbol{h}\rangle^2) ds\right\},$$

*i.e.*

$$d\Lambda_t = \Lambda_t \langle \boldsymbol{X}_s, \eta^{-1}(\tilde{\boldsymbol{h}} - \boldsymbol{h})\rangle dW_t^P, \qquad \Lambda_0 = 1, \tag{2.2.18}$$

*then the dynamics of $y_t$ under the probability measure $\widetilde{P}$ defined by $\frac{d\widetilde{P}}{d\overline{P}}\Big|_{\mathcal{F}_t}$ is*

$$dy_t = \langle \tilde{\boldsymbol{h}}(t), \boldsymbol{X}_t\rangle dt + \eta dW_t^{\tilde{\theta}},$$

*where*

$$dW_t^{\tilde{\theta}} = \frac{\langle \boldsymbol{X}_t, \boldsymbol{h}(t)\rangle - \langle \boldsymbol{X}_t, \tilde{\boldsymbol{h}}(t)\rangle}{\eta} + dW_t^{\theta}$$

*is a Brownian Motion under* $\widetilde{P}$.

From Theorem 2.2.9, we have that the probability measure defined as $\widetilde{P}(\cdot) := \dfrac{\overline{E}\left[\Lambda_t \mathbf{1}_{\{\cdot\}}\right]}{\overline{E}\left[\Lambda_t\right]}$ is the probability measure equivalent to $\overline{P}$ under which $y$ has the drift parameter $\tilde{\boldsymbol{h}}$ with respect to the Brownian Motion $W_t^{\tilde{\theta}}$. Recalling the definition of $\overline{P}$ in terms of the measure $P$, the new measure $\widetilde{P}$ in terms of $P$ is

$$\widetilde{P}(\cdot) = \frac{\overline{E}\left[\Lambda_t \mathbf{1}_{\{\cdot\}}\right]}{\overline{E}\left[\Lambda_t\right]} \propto E\left[L_t \cdot \Lambda_t \mathbf{1}_{\{\cdot\}}\right].$$

We define $P_{\tilde{\theta}}$ to be the new probability measure $\widetilde{P}$. Hence, the combined measure transformation to change the entire parameter vector from $\theta$ to $\tilde{\theta}$ is given by[4]:

$$\left.\frac{dP_{\tilde{\boldsymbol{\theta}}}}{dP_{\boldsymbol{\theta}}}\right|_{\mathcal{F}_t} := L_t{\cdot}\Lambda_t = \prod_{i\neq j} L_t^{i,j} \exp\left\{\int_0^t \langle \boldsymbol{X}_s, \eta^{-1}(\tilde{\boldsymbol{h}} - \boldsymbol{h})\rangle dW_s^{\theta} - \frac{1}{2}\int_0^t (\langle \boldsymbol{X}_s, \eta^{-1}\tilde{\boldsymbol{h}}\rangle^2 - \langle \boldsymbol{X}_s, \eta^{-1}\boldsymbol{h}\rangle^2) ds\right\} \tag{2.2.19}$$

where $L_t^{i,j}$ are defined as in (2.2.16).

#### 2.2.3.3 Maximizing the Log-Likelihood

Taking the log of both sides of (2.2.19), we obtain,

$$\begin{aligned}
\log \left.\frac{dP_{\tilde{\boldsymbol{\theta}}}}{dP_{\boldsymbol{\theta}}}\right|_{\mathcal{F}_t} &= \sum_{i\neq j} \log\left\{L_t^{i,j}\right\} + \int_0^t \langle \boldsymbol{X}_s, \eta^{-1}(\tilde{\boldsymbol{h}} - \boldsymbol{h})\rangle dW_s^{\theta} - \frac{1}{2}\int_0^t (\langle \boldsymbol{X}_s, \eta^{-1}\tilde{\boldsymbol{h}}\rangle^2 - \langle \boldsymbol{X}_s, \eta^{-1}\boldsymbol{h}\rangle^2) ds \\
&= \sum_{i\neq j} N_t^{i,j}(\log \tilde{q}_{i,j} - \log q_{i,j}) - \int_0^t (\tilde{q}_{i,j} - q_{i,j})\langle \boldsymbol{X}_s, \boldsymbol{e}_i\rangle ds \\
&\quad + \int_0^t \langle \boldsymbol{X}_s, \eta^{-1}\tilde{\boldsymbol{h}}\rangle dW_s^{\theta} - \frac{1}{2}\int_0^t (\langle \boldsymbol{X}_s, \eta^{-1}\tilde{\boldsymbol{h}}\rangle^2) ds + \left\{\text{terms independent of } \tilde{\theta}\right\} \\
&= \sum_{i\neq j} N_t^{i,j} \log \tilde{q}_{i,j} - \int_0^t \tilde{q}_{i,j}\langle \boldsymbol{X}_s, \boldsymbol{e}_i\rangle ds \\
&\quad + \int_0^t \langle \boldsymbol{X}_s, \eta^{-1}\tilde{\boldsymbol{h}}\rangle dW_s^{\theta} - \frac{1}{2}\int_0^t (\langle \boldsymbol{X}_s, \eta^{-1}\tilde{\boldsymbol{h}}\rangle^2) ds + \left\{\text{terms independent of } \tilde{\theta}\right\}.
\end{aligned} \tag{2.2.20}$$

Recalling from the definition of the processes $\mathcal{T}_t(i)$ and $\Gamma_t(i)$ at the end of the Section 2.2.2.2 that

$$\mathcal{T}_t(i) := \int_0^t \langle \boldsymbol{X}_s, \boldsymbol{e}_i\rangle ds$$

and

$$\Gamma_t(i) := \int_0^t \langle \boldsymbol{X}_s, \boldsymbol{e}_i\rangle dW_s^{\theta}$$

[4]See Equation 5.1 of Section 8.5 in [EAM08] or the analogous Equation 10.3.18 in [PBL10].

hence

$$\begin{aligned}\frac{1}{2}\int_0^t (\langle \boldsymbol{X}_s, \eta^{-1}\tilde{\boldsymbol{h}}\rangle^2) ds =& \frac{1}{2}\sum_{i=1}^N \eta^{-2}(\tilde{h}^i)^2 \int_0^t \langle \boldsymbol{X}_s, \boldsymbol{e}_i\rangle ds \\ =& \frac{1}{2}\sum_{i=1}^N \eta^{-2}(\tilde{h}^i)^2 \mathcal{T}_t(i)\end{aligned}$$

and

$$\begin{aligned}\int_0^t \langle \boldsymbol{X}_s, \eta^{-1}\tilde{\boldsymbol{h}}\rangle dW_s^\theta =& \sum_{i=1}^N \eta^{-1}\tilde{h}^i \int_0^t \langle \boldsymbol{X}_s, \boldsymbol{e}_i\rangle dW_s^\theta \\ =& \sum_{i=1}^N \eta^{-1}\tilde{h}^i \Gamma_t(i).\end{aligned}$$

Continuing from (2.2.20)

$$= \sum_{i\neq j} N_t^{i,j} \log \tilde{q}_{i,j} - \tilde{q}_{i,j}\mathcal{T}_t(i)) + \sum_{i=1}^N \eta^{-1}\tilde{h}^i\Gamma_t(i) - \frac{1}{2}\eta^{-2}(\tilde{h}^i)^2\mathcal{T}_t(i) + \left\{\text{terms independent of } \tilde{\theta}\right\}$$

We then take the $\mathcal{F}_t^y$-conditional expectation on both sides of (2.2.20). Applying the $\widehat{\cdot} := E[\cdot \mid \mathcal{F}_t^y]$ notation, we obtain the following expression:

$$\begin{aligned}E\left[\log \frac{dP_{\tilde{\boldsymbol{\theta}}}}{dP_{\boldsymbol{\theta}}} \mid \mathcal{F}_t^y\right] =& \sum_{\substack{i,j=1\\ i\neq j}}^N (\widehat{N}_t(i,j)\log\tilde{q}_{i,j} - \tilde{q}_{i,j}\widehat{\mathcal{T}}_t(i)) + \\ &+ \sum_{i=1}^N (\eta^{-1}\tilde{h}^i\widehat{\Gamma}_t(i) - \frac{1}{2}\eta^{-2}(\tilde{h}^i)^2\widehat{\mathcal{T}}_t(i)) + \left\{\text{terms independent of } \tilde{\theta}\right\}.\end{aligned} \tag{2.2.21}$$

We have the following partial derivatives $\frac{\partial}{\partial\tilde{\theta}_l}$ in each of the components of the parameter vector $\tilde{\theta}$:

$$\frac{\partial}{\partial\tilde{q}_{i,j}} E\left\{\log \frac{dP_{\tilde{\boldsymbol{\theta}}}}{dP_{\boldsymbol{\theta}}}\bigg|_{\mathcal{F}_t} \mid \mathcal{F}_t^y\right\} = \frac{1}{\tilde{q}_{i,j}}\widehat{N}_t^{i,j} - \widehat{\mathcal{T}_t(i)} \tag{2.2.22}$$

and

$$\frac{\partial}{\partial\tilde{h}^i} E\left\{\log \frac{dP_{\tilde{\boldsymbol{\theta}}}}{dP_{\boldsymbol{\theta}}}\bigg|_{\mathcal{F}_t} \mid \mathcal{F}_t^y\right\} = \eta^{-1}\widehat{\Gamma}_t(i) - \eta^{-2}\tilde{h}^i\widehat{T}_t(i). \tag{2.2.23}$$

The unique maximum of (2.2.21) over $\tilde{\boldsymbol{\theta}}$, obtained by equating to zero the partial derivatives (2.2.22) and (2.2.23), is therefore given by

$$\tilde{q}_{i,j}^* = \frac{\sigma(N_t^{i,j})}{\sigma(\mathcal{T}_t(i))}, \tag{2.2.24}$$

$$\tilde{h}_i^* = \eta\frac{\sigma(\Gamma_t(i))}{\sigma(\mathcal{T}_t(i))}. \tag{2.2.25}$$

The parameter $\tilde{\boldsymbol{\theta}}^*$ formed by taking (2.2.24) and (2.2.25) gives $P_{\tilde{\boldsymbol{\theta}}}$, the next probability measure in the sequence of steps in the EM procedure. The sequence of log-likelihoods constructed this way is increasing and so converges. The convergence of the sequence of $\theta$ is discussed in [DZ86] and [ZD88].

### 2.2.4 Standard Numerical Schemes

We have seen in the previous sections that parameter-estimation by EM can be achieved through the use of filters. In turn, these are computed by solving certain Zakai SDEs. In practice such computation must be performed numerically. We may use either an Euler- Maruayama scheme or a Milstein scheme (see Section 2.2.4). We will propose a different approach to numerically solve the Zakai SDEs, following Section 2 of the paper [PR10a], that is an approximation based on quasi-exact solutions (to be discussed in Section 2.3.1).

**The Euler and Milstein Approximation Schemes**
The standard procedure in numerical solution of SDE's over a time horizon $[t_0, t)$ is first to chose a fixed time discretization $\left\{t_0 := t_0^{(\Delta)} \leq t_1^{(\Delta)} \ldots \leq t_n^{(\Delta)} = t\right\} =: \mathcal{T}^\Delta$ parametrized by a chosen maximum step size $\Delta = \max_l \{\Delta_l\}$ where $\Delta_l := t_{l+1}^{(\Delta)} - t_l^{(\Delta)}$. We shall drop the superscript "$(\Delta)$" and write $t_l$ to mean $t_l^{(\Delta)}$ in the following discussion. One can then define a time-discretization of the SDE, which takes the form of recursive equations. Two popular time-discretizatons are the Euler - Maruyama scheme and the Milstein scheme:

**Definition 2.2.10.** *Let $W_t$ be a standard Brownian-motion and let $\boldsymbol{S}_t$ be the unique solution of a general SDE*

$$d\boldsymbol{S}_t = \alpha(t, \boldsymbol{S}_t)dt + \beta(t, \boldsymbol{S}_t)dW_t$$

*then the discrete-time stochastic process $\boldsymbol{S}^\Delta = (\boldsymbol{S}_k^\Delta)_{k=0}^n$ defined by*

$$\boldsymbol{S}_{k+1}^\Delta = \boldsymbol{S}_k^\Delta + \alpha(t_k, \boldsymbol{S}_k^\Delta)\Delta t_k + \beta(t_k, \boldsymbol{S}_k^\Delta)\Delta W_k, \qquad \boldsymbol{S}_0^\Delta = \boldsymbol{S}_{t_0}$$

*is called an Euler-Maruyama scheme of the process $\boldsymbol{S}$, where $\Delta t_k := t_{k+1} - t_k$ and $\Delta W_k := W_{t_{k+1}} - W_{t_k}$.*

**Definition 2.2.11.** *Let $W_t$ be a standard Brownian-motion and let $\boldsymbol{S}_t$ be the unique solution of the SDE*

$$d\boldsymbol{S}_t = \alpha(t, \boldsymbol{S}_t)dt + \beta(t, \boldsymbol{S}_t)dW_t$$

*then the discrete-time stochastic process $\boldsymbol{S}^\Delta = (\boldsymbol{S}_k^\Delta)_{k=0}^n$ defined by*

$$\boldsymbol{S}_{k+1}^\Delta = \boldsymbol{S}_k^\Delta + \alpha(t_k, \boldsymbol{S}_k^\Delta)\Delta t_k + \beta(t_k, \boldsymbol{S}_k^\Delta)\Delta W_k + \frac{1}{2}\beta(t_k, \boldsymbol{S}_k^\Delta)\beta_v(t_k, \boldsymbol{S}_k^\Delta)(\Delta W_k^2 - \Delta t_k), \qquad \boldsymbol{S}_0^\Delta = \boldsymbol{S}_{t_0}$$

*is called a Milstein scheme of the process $\boldsymbol{S}$, where $\Delta t_k := t_{k+1} - t_k$, $\Delta W_k := W_{t_{k+1}} - W_{t_k}$ and $\beta_v := \dfrac{\partial}{\partial V}\beta$.*

The Euler- Maruyama and Milstein schemes are very simple to implement and they have the property that the values of the discretized process $\boldsymbol{S}^\Delta(t_k)$ at each time point $t_k$ can be made as close as desired to the value of the continuous process $\boldsymbol{S}_{t_k}$ by taking $\Delta$ small enough. This is formalized in the concept of the strong-order of an approximation.

**Definition 2.2.12** (Strong Order)**.** *Let us consider on the interval $[t_0, t] \subseteq \mathbb{R}^+$ a family of time-discretizations $\mathcal{T}^\Delta := \left\{t_l^{(\Delta)}\right\}_{l=0}^{n\Delta}$ parametrized by a decreasing sequence of the maximum step size $\Delta$. Following Section 5.3 of [PBL10] we shall say that a discrete time approximation scheme $\boldsymbol{S}^\Delta$*

*of a stochastic process* $\boldsymbol{S}_t$ *is of order* $\gamma$ *if* $(\boldsymbol{S}_k^\Delta)_k$ *is an approximating sequence for* $\boldsymbol{S}_t$ *such that its global error satisfies*

$$\bar{E}\left\{\left|\boldsymbol{S}_t - \boldsymbol{S}_{t_n}^\Delta\right|\right\} < K \cdot \Delta^\gamma$$

*where* $|\cdot|$ *denotes the Euclidean norm for vectors and* $K$ *is a constant independent of the maximum step-size* $\Delta$*. We shall refer to an approximating sequence* $(\boldsymbol{S}_k^\Delta)_k$ *with strong order* $\gamma$ *as a strong approximation of order* $\gamma$ *to the stochastic process* $\boldsymbol{S}_t$*.*

The Euler - Maruyama and Milstein schemes are known to have strong order of at least 0.5 and 1.0, respectively (see [PBL10] Section 5.2). They belong to the class of Ito-Taylor approximations, which is a stochastic analogue of numerical methods for ODEs where, for some $k$, one takes only the first $k$ terms of the Taylor expansion of the differential operator, and then replaces each derivative by a difference ratio. For example, when $k = 1$, the resulting approximation, in the deterministic ODE case, is just the (deterministic) Euler scheme. In the case of Ito processes, instead of a Taylor expansion one must apply a stochastic analogue such as a Wagner-Platen expansion (Chapter 4 of [PBL10]). The Euler - Maruyama scheme is an approximation that includes the first order (drift and diffusion) terms of a Wagner-Platen expansion, while the Milstein scheme is an approximation that includes the additional second order term due to the Quadratic Variation.

In such schemes, the successive values are approximated by projecting an imaginary tangent curve starting from the previous value, and whose "slope" is due to the differential operator. In other words, while these schemes capture the direction of a path well they do not capture the point-by-point values of the path. This inevitably leads to a discretization error. Starting from a known true value, the one-step error, so-called *local error*, may still be small, but the accumulated error or *global error* grows with each successive time point. The concept of strong order of a scheme provides us with a lower bound on the rate by which the global error can be reduced by taking a smaller step-size $\Delta$.

#### 2.2.4.1 Numerical Implementation of the filter-based EM Algorithm

We now apply the standard numerical schemes to the filter-based parameter estimation by EM, as discussed in Section 2.2.3. Namely, we must numerically compute, the unnormalized conditional expectations $\sigma(\boldsymbol{X}_t)$, $\sigma(N_t^{i,j})$, $\sigma(\mathcal{T}_t(i))$ and $\sigma(\Gamma_t(i))$.

The Zakai equation for $\sigma(\boldsymbol{X}_t)$, see (2.2.8), has the form

$$d\boldsymbol{Z}_t = \boldsymbol{A}\boldsymbol{Z}_t dt + \boldsymbol{D}(t)\boldsymbol{Z}_t dW_t, \tag{2.2.26}$$

where $\boldsymbol{D}$ is a diagonal matrix. On the other hand, the Zakai equation for $\sigma(\boldsymbol{N_t^{i,j}})$ (see 2.2.11) and $\sigma(\mathcal{T}_t(i))$ (see 2.2.12) has the form

$$d\boldsymbol{S}_t = \{\bar{\boldsymbol{a}}(\sigma(\boldsymbol{X}_t)) + \boldsymbol{A}\boldsymbol{S}_t\}\, dt + \boldsymbol{D}(t)\boldsymbol{S}_t dW_t, \tag{2.2.27}$$

a so-called *linear SDE in the narrow sense*. Finally the Zakai equation for $\sigma(\Gamma_t(i))$ (see 2.2.13) has the form

$$d\boldsymbol{S}_t = \{\boldsymbol{a}(\sigma(\boldsymbol{X}_t)) + \boldsymbol{A}\boldsymbol{S}_t\}\, dt + \{\boldsymbol{b}(\sigma(\boldsymbol{X}_t)) + \boldsymbol{D}(t)\boldsymbol{S}_t\}\, dW_t, \tag{2.2.28}$$

which is called a *general linear SDE*.

For a practical computation of the solutions of these SDEs, one considers a given time-discretization $\mathcal{T}^\Delta$ which, for instance, corresponds to the observation times or some sparse subset thereof. Then, the standard procedure is to apply an Euler-Maruyama or Milstein scheme. For example the Euler- Maruyama scheme for an SDE of the form (2.2.26) is

$$\boldsymbol{Z}_{k+1}^\Delta = (\boldsymbol{I} + \boldsymbol{A}\Delta_k + \boldsymbol{D}(t_{k+1})\Delta W_k) \cdot \boldsymbol{Z}_k^\Delta$$

while for an SDE of the form (2.2.28) it is

$$\boldsymbol{S}_{k+1}^\Delta = (\boldsymbol{I} + \boldsymbol{A}\Delta_k + \boldsymbol{D}(t_{k+1})\Delta W_k) \cdot \boldsymbol{S}_k^\Delta + \boldsymbol{a}(\boldsymbol{Z}_k^\Delta)\Delta_k + \boldsymbol{b}(\boldsymbol{Z}_k^\Delta)\Delta W_k. \tag{2.2.29}$$

The Euler-Maruyama scheme for an SDE of the form (2.2.27) is (2.2.29) with $\boldsymbol{b}(\cdot) = \underline{0}$, the zero vector.

On the other hand, the Milstein scheme for an SDE of the form (2.2.26) is

$$\boldsymbol{Z}_{k+1}^\Delta = (\boldsymbol{I} + \boldsymbol{A}\Delta_k + \boldsymbol{D}(t_{k+1})\Delta W_k + \frac{1}{2}\boldsymbol{D}(t_k)^2(\Delta W_k^2 - \Delta t_k)) \cdot \boldsymbol{Z}_k^\Delta$$

while for an SDE of the form (2.2.28) it is

$$\boldsymbol{S}_{k+1}^\Delta = (\boldsymbol{I}+\boldsymbol{A}\Delta_k+\boldsymbol{D}(t_{k+1})\Delta W_k+\frac{1}{2}\boldsymbol{D}(t_k)^2(\Delta W_k^2-\Delta t_k))\cdot\boldsymbol{S}_k^\Delta+\boldsymbol{a}(\boldsymbol{Z}_k^\Delta)\Delta_k+\boldsymbol{b}(\boldsymbol{Z}_k^\Delta)\Delta W_k. \tag{2.2.30}$$

and the analogous scheme for (2.2.27) given by (2.2.30) with $\boldsymbol{b}(\cdot) = \underline{0}$.

#### 2.2.4.2 Including the Jump Times

For an SDE with jumps, analogous strong approximations of a given strong order are very challenging to compute because the Wagner - Platen expansion must then also include the terms of the Poisson random measure due to the jumps (see Chapter 8 of [PBL10]). Higher terms of such expansions become rather complicated. In this regard, [Pla82] observes that within each inter-jump interval, the process is just a continuous diffusion. Hence he proposed the class of *jump-adapted schemes*. Here, one joins the jump-times to the time points $\mathcal{T}^\Delta$ (fixed grid) and then applies a standard strong approximation scheme within the inter-jump intervals. At the jump times, the approximated value is incremented by the value of the jump-measure at that time. [PBL10] observes that even for mark-dependent jumps, the strong order from such jump-adapted schemes are in general, equal to the strong order of the scheme used for the diffusive part.

It is important to qualify that the trade-off for a scheme to have a higher strong order is a higher computational complexity, so that one must either take a very fine step-size $\Delta$ or use a complicated scheme, in order to attain the desired accuracy. In Section 8.6 of [PBL10], they discuss *jump-adapted exact* schemes, which, on the other hand, results in highly accurate approximations and whose formulae still relatively simple expressions. It is an approach that applies to the case when explicit solutions to SDEs can be obtained. Unlike the standard schemes mentioned above, it does not apply a discretization of the differential of the process. Rather, this scheme computes directly the exact values of the process at the discretization points, and hence there is no time-discretization error. In [PR10a], a *quasi-exact jump-adapted* scheme is treated, in which a related *quasi-exact* solution is applied in the case when the exact explicit solution to the SDE cannot be obtained. In Section 2.3 we explore the use of this method for the Zakai SDEs that enter in filter-based EM.

## 2.3 Quasi - Exact Schemes

In this section, we discuss an alternative to the standard numerical schemes reviewed in Section 2.2.4. We shall refer to these as Quasi - Exact Approximation schemes. It is inspired by the jump - adapted quasi - exact solution proposed in [PR10a] (see also Section 8.6 of [PBL10]). We proceed as follows: first, in sub-section 2.3.1 we treat the exact solutions of Zakai SDEs, from which the quasi-exact solutions are defined. Then in Section 2.3.2 we derive the discretization scheme utilizing the quasi - exact solutions defined in Section 2.3.1.

### 2.3.1 Exact Solutions of the Zakai Equations

Let $\boldsymbol{A} : \mathbb{R} \to \mathbb{R}^{N\times N}$ and $\boldsymbol{D} : \mathbb{R} \to \mathbb{R}^{N\times N}$ be deterministic matrix functions, where $\boldsymbol{A}(t)$ and $\boldsymbol{D}(t)$ are both defined for all $t \in [0,T]$. We restrict $\boldsymbol{D}(t)$ to be a diagonal matrix, and further impose that $\boldsymbol{A}(t)$ and $\boldsymbol{D}(t)$ satisfy the condition $\sup_{0\leq t\leq T}\{|\boldsymbol{A}(t)| + |\boldsymbol{D}(t)|\} < \infty$, where $|\cdot|$ is some matrix norm. Consider under some probability measure $P$ and a time horizon $[0,T]$, the scalar $P$ - Brownian Motion $dW_t$ and the $\mathbb{R}^N$ - vector stochastic process $\boldsymbol{Z}_t$, $t \in [0,T)$ that is a solution to the $N$-dimensional SDE

$$d\boldsymbol{Z}_t = \boldsymbol{A}(t)\boldsymbol{Z}_t dt + \boldsymbol{D}(t)\boldsymbol{Z}_t dW_t \tag{2.3.1}$$

for a given deterministic initial value $\boldsymbol{Z}_{t_0} = \boldsymbol{Z}_0$.

The Zakai equation (2.2.8) for the unnormalized conditional expectation of the state $\sigma(\boldsymbol{X}_t)$ and which is defined under $\bar{P}$, is an equation of the type (2.3.1). Indeed we can see this if we substitute the process $\bar{y}_t$ (which is a $\bar{P}$-Brownian motion) from (2.2.8) to the process $W_t$ in (2.3.1). We have the following

**Proposition 2.3.1.** *Suppose that for every fixed $t \in [0,T]$, the matrices $\boldsymbol{A}(t)$ and $\boldsymbol{D}(t)$ commute, i.e.*

$$\boldsymbol{A}(t)\boldsymbol{D}(t) = \boldsymbol{D}(t)\boldsymbol{A}(t),$$

*then an explicit solution of the SDE (2.3.1) can be expressed by*

$$\boldsymbol{Z}_t = \boldsymbol{\Phi}_t \boldsymbol{Z}_0, \tag{2.3.2}$$

*for $t \in [0,\infty)$, where*

$$\boldsymbol{\Phi}_t = \mathbf{exp}\left\{\int_0^t \left(\boldsymbol{A}(s) - \frac{1}{2}(\boldsymbol{D})^2(s)\right) ds + \int_0^t \boldsymbol{D}(s) dW_s\right\}, \tag{2.3.3}$$

*for $t \in [0,T)$, and where $\mathbf{exp}\{\cdot\}$ is the matrix exponential defined for any real or complex valued square matrices $\boldsymbol{M}$ as*

$$\mathbf{exp}\{\boldsymbol{M}\} := \sum_{k=0}^{\infty} \frac{\boldsymbol{M}^k}{k!}.$$

*(Recall that this sum always converges for any real or complex matrix $\boldsymbol{M}$ and is bounded if the matrix-norm $\|\boldsymbol{M}\|$ is bounded).*

*Proof.* A proof for the more general case when $\boldsymbol{Z}$ is driven by $k$ multiple noise factors (i.e., $k$ independent Brownian Motions) is given in [PR10a]. In Appendix 2.A.2 we show the proof, following the approach in [PR10a] simplified to the present case with a scalar Brownian Motion. The proof entails computing $d(\boldsymbol{\Phi}_t \boldsymbol{Z}_0) = d(\boldsymbol{\Phi}_t)\boldsymbol{Z}_0$ by using the Ito product formula to obtain the SDE (2.3.1) with $\boldsymbol{Z}_t = \boldsymbol{\Phi}_t \boldsymbol{Z}_0$. □

Thus a sufficient condition for an SDE of type (2.3.1) to have an explicit solution is if $\boldsymbol{A}$ and $\boldsymbol{D}$ commute. In the SDE (2.2.8) we need to check $\boldsymbol{Q}$, and $\boldsymbol{D}(t)$; in that equation the matrices in general are not commuting due to the general structure of $\boldsymbol{Q}$.

**Remark 2.3.2.** *Since $\boldsymbol{D}$ is diagonal, it commutes with other diagonal matrices, otherwise it may not commute. In the case when $\boldsymbol{X}_t$ is a CTMC whose $\boldsymbol{Q}$-matrix possesses a special structure, such that it commutes with $\boldsymbol{D}$, the explicit solution (2.3.2) holds, and the filtering and parameter estimation procedure outlined in Sections 2.2.3 and 2.2.2 can be performed easily.*

*The authors in [PR10a] observe that, even in the case when $\boldsymbol{A}$ does not commute with the $\boldsymbol{D}_k$'s, if nevertheless one takes the matrix exponential (2.3.3) formally into (2.3.2) then one obtains a "proxy" of the solution of the equation (2.3.1). It is noted in [PR10a] that this so-called quasi-exact solution provides an excellent approximation of the exact solution. They numerically tested the performance of such an approximation against time-discretization schemes such as the Euler-Maruyama and Milstein schemes, and showed that their quasi-exact approximation is able to produce good results - provided they use a jump-adapted time discretization - even for large time steps of $\Delta$ for which the other schemes fail to provide good approximations.*

**Definition 2.3.3** (Quasi-Exact Solution)**.** *When $\boldsymbol{A}(t)$ and $\boldsymbol{D}(t)$ do not commute, we shall call the vector stochastic process $\boldsymbol{Z}_t$ defined by*

$$\boldsymbol{Z}_t = \boldsymbol{\Phi}_t \boldsymbol{Z}_0, \tag{2.3.4}$$

*where $\boldsymbol{Z}_0$ is a given initial vector value $\boldsymbol{Z}_{t_0} = \boldsymbol{Z}_0$ and where*

$$\boldsymbol{\Phi}_t = \mathbf{exp}\left\{ \int_0^t \left( \boldsymbol{A}(s) - \frac{1}{2}(\boldsymbol{D})^2(s) \right) ds + \int_0^t \boldsymbol{D}(s) dW_s \right\}, \tag{2.3.5}$$

*for $t \in [0, T)$, a quasi-exact solution to the SDE*

$$d\boldsymbol{Z}_t = \boldsymbol{A}(t)\boldsymbol{Z}_t dt + \boldsymbol{D}(t)\boldsymbol{Z}_t dW_t \tag{2.3.6}$$

**Remark 2.3.4** (On the relationship between the true solution and the quasi - exact solution for the non-commutative case)**.** *Applying such quasi - exact solutions in solving SDEs requires an understanding of the precise relationship between the quasi-exact and the true solution. A study of this relationship is not pursued in the present work; we observe only that the commutativity of $\boldsymbol{A}(t)$ and $\boldsymbol{D}(t)$ is necessary and sufficient for the following equality to hold*

$$\begin{aligned} &\mathbf{exp}\left\{ \int_0^t \boldsymbol{A}(s) ds \right\} \mathbf{exp}\left\{ -\frac{1}{2}\int_0^t (\boldsymbol{D})^2(s) ds + \int_0^t \boldsymbol{D}(s) dW_s \right\} \\ &\quad = \mathbf{exp}\left\{ \int_0^t \left( \boldsymbol{A}(s) - \frac{1}{2}(\boldsymbol{D})^2(s) \right) ds + \int_0^t \boldsymbol{D}(s) dW_s \right\}, \end{aligned} \tag{2.3.7}$$

*while in the general case, this equality may not be true. In the proof for Proposition2.3.1, the commutativity is needed only for such an identity. If one follows the proof without applying*

*substitutions that use (2.3.7), then one can show that if we define* $\mathbf{\Phi}_t$ *itself to be the left hand side of (2.3.7) and if we do not assume commutativity of* $\boldsymbol{A}(t)$ *and* $\boldsymbol{D}(t)$*, it turns out that in this case* $\mathbf{\Phi}_t$ *is the solution to the SDE*

$$d\mathbf{\Phi}_t = \boldsymbol{A}(t)\mathbf{\Phi}_t dt + \mathbf{\Phi}_t \boldsymbol{D}(t) dW_t. \tag{2.3.8}$$

*Hence, understanding the relationship between (2.3.6) and the SDE (2.3.8) may suggest what link exists between the quasi-exact and the true solution.*

*One may also look at* $\mathbf{log}(\boldsymbol{e}^{\boldsymbol{A}}\boldsymbol{e}^{\boldsymbol{D}})$ *when* $\boldsymbol{A}$ *and* $\boldsymbol{D}$ *do not commute*[5]*. We do not pursue this question further in this study, instead as in [PR10a], we will present a simulation study for a simple case using the quasi - exact approximation scheme to be discussed in sub-section 2.3.2.*

#### 2.3.1.1 Quasi-Exact Solutions for the Counting Process, the Occupation Time, and the Drift Estimator

To estimate parameters according to Section 2.2.3, we also need to solve the Zakai equations (2.2.11), (2.2.12), and (2.2.13) for $\sigma(N_t^{i,j}\boldsymbol{X}_t)$, $\sigma(\mathcal{T}_t(i)\boldsymbol{X}_t)$ and $\sigma(\Gamma_t(i)\boldsymbol{X}_t)$ respectively. Recall that the equations (2.2.11) and (2.2.12) for $\sigma(N_t^{i,j}\boldsymbol{X}_t)$ and $\sigma(\mathcal{T}_t(i)\boldsymbol{X}_t)$ are each vector SDEs of the narrow-sense linear form. That is if we denote by $\boldsymbol{A}(t) : \mathbb{R} \to \mathbb{R}^{N\times N}$ and $\boldsymbol{D}(t) : \mathbb{R} \to \mathbb{R}^{N\times N}$ matrix valued maps that are deterministic, and $\boldsymbol{a}(\cdot) : \mathbb{R}^{N\times N} \to \mathbb{R}^N$ is a deterministic vector functional defined on $\boldsymbol{R}^{N\times N}$. Then (2.2.11) and (2.2.12) are SDEs of the form

$$d\boldsymbol{S}_t = \{\boldsymbol{A}(t)\boldsymbol{S}_t + \boldsymbol{a}\left(\mathbf{\Phi}(t,t_0)\right)\}\, dt + \boldsymbol{D}(t)\boldsymbol{S}_t dY_t. \tag{2.3.9}$$

On the other hand, denoting $\boldsymbol{b}(\cdot) : \mathbb{R}^{N\times N} \to \mathbb{R}^N$ a deterministic vector functional defined on $\mathbb{R}^{N\times N}$, then the equation (2.2.13) for $\sigma(\Gamma_t(i)\boldsymbol{X}_t)$ is an SDE of the general linear vector form

$$d\boldsymbol{S}_t = \{\boldsymbol{A}(t)\boldsymbol{S}_t + \boldsymbol{a}(\mathbf{\Phi}(t,t_0))\}\, dt + \{\boldsymbol{D}(t)\boldsymbol{S}_t + \boldsymbol{b}\left(\mathbf{\Phi}(t,t_0)\right)\}\, dY_t \tag{2.3.10}$$

where $Y_t$ is the observations process, that is a $\bar{P}$-Brownian Motion, and the matrix $\mathbf{\Phi}(t,t_0)$ is the solution to the SDE

$$d\mathbf{\Phi}(t,t_0) = \boldsymbol{A}(t)\mathbf{\Phi}(t,t_0)dt + \boldsymbol{D}(t)\mathbf{\Phi}(t,t_0)dY_t, \qquad \mathbf{\Phi}(t_0,t_0) = \boldsymbol{I}.$$

**Remark 2.3.5.** *The form of the drift translation term* $\boldsymbol{a}(\mathbf{\Phi}(t,t_0))$ *is that of a function* $\boldsymbol{a}(\cdot)$ *evaluated at* $\mathbf{\Phi}(t,t_0)$ *because any function* $f(\sigma(\boldsymbol{X}_t))$ *can be represented as* $f(\Phi(t,t_0)\cdot\sigma(\boldsymbol{X}_0))$*, noting that* $\sigma(\boldsymbol{X}_t) = \mathbf{\Phi}(t,t_0)\cdot\sigma(\boldsymbol{X}_0)$ *(see Proposition 2.3.1).*

We now obtain the quasi-exact solutions for these SDEs. First we have the following proposition concerning the existence of exact, explicit solutions

**Proposition 2.3.6.** *If* $\boldsymbol{A}(t)$ *and* $\boldsymbol{D}(t)$ *are commuting, then the matrix* $\mathbf{\Phi}(t,t_0)$ *is given by the exact solution*

$$\mathbf{\Phi}(t,t_0) = \mathbf{exp}\left\{\int_0^t \left(\boldsymbol{A}(s) - \frac{1}{2}(\boldsymbol{D})^2(s)\right) ds + \int_0^t \boldsymbol{D}(s)dY_s\right\} \tag{2.3.11}$$

[5] see the remarks in Section 4.2 about the *Baker-Campbell-Hausdorff* formula that is used to compute such logarithms

*Proof.* This is immediate, following the proof of Proposition 2.3.1, in particular see equation (2.3.3). □

**Proposition 2.3.7.** *If $\boldsymbol{A}(t)$ and $\boldsymbol{D}(t)$ are commuting $\mathbb{R}^{N\times N}$ matrices, and given the initial value $\boldsymbol{S}_{t_0} = \boldsymbol{S}_0$ at $t_0$, then the vector linear SDE (2.3.9) has the solution*

$$\boldsymbol{S}_t = \boldsymbol{\Phi}(t,t_0)\left(\boldsymbol{S}_0 + \int_{t_0}^{t} \boldsymbol{\Phi}(s,t_0)^{-1}\left[\boldsymbol{a}(\boldsymbol{\Phi}(s,t_0))\right] ds\right), \tag{2.3.12}$$

*where the matrix stochastic process $\boldsymbol{\Phi}(t,t_0)$ is given by (2.3.11) with the initial condition $\boldsymbol{\Phi}(t_0,t_0) = \boldsymbol{I}$, the identity matrix.*

**Proposition 2.3.8.** *If $\boldsymbol{A}(t)$ and $\boldsymbol{D}(t)$ are commuting $\mathbb{R}^{N\times N}$ matrices, and given the initial value $\boldsymbol{S}_{t_0} = \boldsymbol{S}_0$ at $t_0$, then the vector linear SDE (2.3.10) has the solution*

$$\boldsymbol{S}_t = \boldsymbol{\Phi}(t,t_0)\left(\boldsymbol{S}_0 + \int_{t_0}^{t} \boldsymbol{\Phi}(s,t_0)^{-1}\left[\boldsymbol{a}(\boldsymbol{\Phi}(s,t_0)) - \boldsymbol{D}(s)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))\right] ds + \int_{t_0}^{t} \boldsymbol{\Phi}^{-1}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))dY_s\right), \tag{2.3.13}$$

*where the matrix stochastic process $\boldsymbol{\Phi}(t,t_0)$ is given by equation (2.3.11).*

*Proof.* See Appendix 2.A.3. Note that Proposition 2.3.7 follows from Proposition 2.3.8 as the special case when $\boldsymbol{b}(\boldsymbol{\Phi}) = \boldsymbol{0}$, the zero vector, for any $R^{N\times N}$ matrix $\boldsymbol{\Phi}$. Hence in Appendix 2.A.3 we only prove Proposition 2.3.8. □

In light of the fact that the coefficient matrices in equations (2.2.11), (2.2.12) and (2.2.13) do not commute in general, then the process defined in (2.3.13) represents a quasi-exact approximation to the true solution.

**Definition 2.3.9.** *When the coefficient matrices $\boldsymbol{A}(t)$ and $\boldsymbol{D}(t)$ do not commute, we shall call the process $\boldsymbol{S}_t$ defined in (2.3.12) and (2.3.13) the quasi-exact solutions to the SDEs (2.3.9) and (2.3.10), respectively.*

### 2.3.2 Discrete Approximations

In this section we derive the time-discretization schemes for SDEs that we shall call quasi-exact approximations. We shall derive two such approximations, one for the Zakai SDE of the state (sub-section 2.3.1), and another for general linear SDEs (sub-section 2.3.1.1).

#### 2.3.2.1 Quasi - Exact Approximation for the $\sigma(\boldsymbol{X}_t)$

Let us recall that the Zakai equation (2.2.8) to solve the unnormalized conditional probabilities $\boldsymbol{Z}_t := \boldsymbol{\sigma}(\boldsymbol{X}_t)$ for the Wonham filter of (2.2.5), noting that $\bar{\boldsymbol{D}}(t)\boldsymbol{Z}_t d\bar{y}_t = \boldsymbol{D}(t)\boldsymbol{Z}_t dy_t$ (see 2.A.4):

$$\begin{cases} d\boldsymbol{Z}_t = & \boldsymbol{Q}\boldsymbol{Z}_t dt + \boldsymbol{D}(t)\boldsymbol{Z}_t dy_t, \\ \boldsymbol{Z}_0 \sim & P(\boldsymbol{Z}_0), \end{cases} \tag{2.3.14}$$

defined under $\bar{P}$. Let us further recall that in this case, the process

$$\boldsymbol{Z}_t = \boldsymbol{\Phi}_t \boldsymbol{Z}_0,$$

with

$$\boldsymbol{\Phi}_t = \mathbf{exp}\left\{\boldsymbol{Q}t - \frac{1}{2}\int_0^t \boldsymbol{D}^2(s)ds + \int_0^t \boldsymbol{D}(s)dy_s\right\}$$

is either the exact solution, if $\boldsymbol{Q}$ and $\boldsymbol{D}(t)$ commute, or is a quasi-exact solution if they do not commute.

**Remark 2.3.10.** *We will permit ourselves the slight abuse of terminology in the sequel, whereby we always refer to the process $\boldsymbol{Z}_t = \boldsymbol{\Phi}_t \boldsymbol{Z}_0$ as the quasi-exact solution of (2.3.14), even if such a process is in fact the exact solution in the case when $\boldsymbol{Q}$ and $\boldsymbol{D}(t)$ commute.*

Let us now propose a strong approximation to $\boldsymbol{Z}_t$. First, notice that at a generic time point $t = t_{k+1}$ we can write $\boldsymbol{Z}_{t_{k+1}}$ as

$$\boldsymbol{Z}_{t_{k+1}} = \boldsymbol{Z}_{t_0} + \int_{t_0}^{t_{k+1}} d\boldsymbol{Z}_s.$$

The standard approach here would be to apply an Euler-Maruyama (or Milstein) discretization scheme $\Delta\tilde{\boldsymbol{Z}}_k$ to the differential $d\boldsymbol{Z}_t$ and replace the integral with a summation in terms of discretized increments of $\boldsymbol{Z}_t$, that is

$$\boldsymbol{Z}_{t_0} + \int_{t_0}^{t_{k+1}} d\boldsymbol{Z}_s \approx \boldsymbol{Z}_{t_0} + \sum_{l=0}^{k} \Delta\tilde{\boldsymbol{Z}}_k^{\Delta} = \boldsymbol{Z}_{t_0} + \sum_{l=0}^{k} (\tilde{\boldsymbol{Z}}_{l+1}^{\Delta} - \tilde{\boldsymbol{Z}}_l^{\Delta}) = \tilde{\boldsymbol{Z}}_{k+1}^{\Delta},$$

where the sequence $(\tilde{\boldsymbol{Z}}_k^{\Delta})_k$ is defined by the Euler-Maruyama (or equivalently, Milstein) discretion of $d\boldsymbol{Z}_t$, i.e. $\Delta\tilde{\boldsymbol{Z}}_k^{\Delta} \approx d\boldsymbol{Z}_t$ .

However, we wish to approximate the quasi-exact solution $\boldsymbol{Z}_t$ to (2.3.14) itself, and so, we shall follow a different approach. We wish to *define* each $\boldsymbol{Z}_k^{\Delta}$ such that

$$\boldsymbol{Z}_k^{\Delta} \approx \boldsymbol{Z}_{t_k}.$$

Now notice that

$$\boldsymbol{Z}_t = \mathbf{exp}\left\{\boldsymbol{Q}t - \frac{1}{2}\int_0^t \boldsymbol{D}^2(s)ds + \int_0^t \boldsymbol{D}(s)dy_s\right\} \cdot \boldsymbol{Z}_0, \tag{2.3.15}$$

In order to write the right-hand expression in (2.3.15) as a recursion, let us first approximate the integrals with a summation of increments.

**Lemma 2.3.11.** *Suppose that $\boldsymbol{Q}$ commutes with $\boldsymbol{D}(t)$. If we define a matrix stochastic process $\boldsymbol{\zeta}_t$ by the SDE*

$$d\boldsymbol{\zeta}_t = \left\{\boldsymbol{Q} - \frac{1}{2}\boldsymbol{D}^2(t)\right\}dt + \boldsymbol{D}(t)dy_t \quad \text{with } \boldsymbol{\zeta}_0 = \mathbf{0} \tag{2.3.16}$$

*then $\boldsymbol{Z}_t = \mathbf{exp}\{\boldsymbol{\zeta}_t\}\boldsymbol{Z}_0$, where $\mathbf{exp}\{\cdot\}$ denotes the matrix exponential. If $\boldsymbol{Q}$ and $\boldsymbol{D}(t)$ do not commute, then $\boldsymbol{Z}_t = \exp\{\boldsymbol{\zeta}_t\}\boldsymbol{Z}_0$ where $\exp\{\cdot\}$ denotes the matrix of component-wise exponentials.*

*Proof.* The proof for the commuting case is analogous to the proof for Proposition 2.3.1. The proof for the non-commuting case entails applying the Ito formula component-wise to compute the matrix of component-wise differentials $(d(e^{\boldsymbol{\zeta}_{i,j}}))_{1\leq i,j\leq N}$. □

**Remark 2.3.12.** *In the sequel we shall define $\mathbf{exp}\{\boldsymbol{\zeta}_t\}\boldsymbol{Z}_0$ to be the quasi-exact solution $\boldsymbol{Z}_t$ instead of $\exp\{\zeta_t\}\boldsymbol{Z}_0$, when $\boldsymbol{Q}$ and $\boldsymbol{D}(t)$ do not commute.*

Next we take a fixed time-discretization $t_0 = s_0 < s_1 < \cdots < s_n = T$, such that $s_k = k\Delta$, for $k = 0, 1, \ldots, \frac{T}{\Delta}$, i.e. $\Delta = \Delta_l$, $\forall l = 0 \ldots n$. We shall define the matrix sequence denoted by $(\boldsymbol{\zeta}_k^\Delta)_k$ - with $\boldsymbol{\zeta}_k^\Delta = \boldsymbol{\zeta}^\Delta(s_k)$ - to be an Euler-Maruayma approximation of $\boldsymbol{\zeta}_t$ arising from a discretization of $d\boldsymbol{\zeta}_t$. Hence the sequence $(\boldsymbol{\zeta}_k^\Delta)_k$ takes the form of the following recursive equation

$$\boldsymbol{\zeta}_{k+1}^\Delta = \boldsymbol{\zeta}_k^\Delta + \boldsymbol{Q}\Delta s_k - \frac{1}{2}\boldsymbol{D}^2(s_k)\Delta + \boldsymbol{D}(s_k)\Delta Y_k, \quad \text{with } \boldsymbol{\zeta}_0^\Delta = \boldsymbol{\zeta}_{t_0} = \mathbf{0}. \tag{2.3.17}$$

Notice that, since $\mathbf{exp}\{\boldsymbol{\zeta}_t\}\boldsymbol{Z}_0 = \boldsymbol{Z}_t$, the equation (2.3.17) suggests that we approximate (2.3.15) by writing the integrals as summations of increments as follows

$$\begin{aligned}\boldsymbol{Z}_t &= \mathbf{exp}\left\{\boldsymbol{Q}t - \frac{1}{2}\int_0^t \boldsymbol{D}^2(s)ds + \int_0^t \boldsymbol{D}(s)dy_s\right\}\cdot \boldsymbol{Z}_0 \\ &\approx \mathbf{exp}\left\{\boldsymbol{Q}t - \frac{1}{2}\sum_{l=1}^k \boldsymbol{D}^2(s_l)\Delta s_l + \sum_{l=1}^k \boldsymbol{D}(s_l)\Delta y_l\right\}\cdot \boldsymbol{Z}_0,\end{aligned}$$

which we can write in a recursive form as follows

$$\begin{aligned}&= \mathbf{exp}\left\{\boldsymbol{Q}\cdot k\cdot \Delta s_k - \frac{1}{2}\sum_{l=1}^k \boldsymbol{D}^2(s_l)\Delta s_l + \sum_{l=1}^k \boldsymbol{D}(s_l)\Delta y_l\right\}\cdot \boldsymbol{Z}_0 \\ &= \prod_{l=1}^k \mathbf{exp}\left\{\boldsymbol{Q}\cdot \Delta s_l - \frac{1}{2}\boldsymbol{D}^2(s_l)\Delta s_l + \boldsymbol{D}(s_l)\Delta y_l\right\}\boldsymbol{Z}_0 \\ &= \prod_{l=1}^k \mathbf{exp}\left\{\boldsymbol{\zeta}_l^\Delta\right\} \\ &=: \boldsymbol{Z}_k^\Delta\end{aligned} \tag{2.3.18}$$

Notice that we have essentially defined our approximating sequence $(\boldsymbol{Z}_k^\Delta)_k$ for $\boldsymbol{Z}_t$ by equating $\boldsymbol{Z}_k^\Delta/\boldsymbol{Z}_{k-1}^\Delta$ to $\mathbf{exp}\{\boldsymbol{\zeta}_k^\Delta\}$ at each step $k$. The sequence $\boldsymbol{Z}_{s_k}^\Delta$ is then expressed recursively by

$$\boldsymbol{Z}_{k+1}^\Delta = \mathbf{exp}\left\{\boldsymbol{Q}\Delta s_k - \frac{1}{2}\boldsymbol{D}^2(s_k)\Delta + \boldsymbol{D}(s_k)\Delta y_k\right\}\boldsymbol{Z}_k^\Delta \quad \text{with } \boldsymbol{Z}_0^\Delta = \boldsymbol{Z}_{s_0}, \tag{2.3.19}$$

for $s_k = s_0, s_1, \ldots, T$.

**Remark 2.3.13.** *Defining $\boldsymbol{Z}_k^\Delta$ as the exponential of the Euler-Maruyama approximating sequence arising from the discretization of the differential of the logarithm of $\boldsymbol{Z}_t$ may seem indirect; indeed one may ask why not instead compute the Euler-Maruyama approximating sequence arising from the differential of $\boldsymbol{Z}_t$ directly. In fact, the indirect approach is a standard procedure in the numerical simulation of Geometric Brownian Motion SDEs (see for example Section 5.2 of [PBL10]) that ensures non-negative and log-normally distributed approximations $\boldsymbol{Z}_k^\Delta$. Indeed, note that taking the Euler-Maruyama discretization of $d\boldsymbol{Z}_t = \boldsymbol{Q}\boldsymbol{Z}_t dt + \boldsymbol{D}(t)\boldsymbol{Z}_t dy_t$, one obtains the sequence*

$$\tilde{\boldsymbol{Z}}_{k+1}^\Delta = \tilde{\boldsymbol{Z}}_k^\Delta + \boldsymbol{Q}\boldsymbol{Z}_k\Delta s_k + \boldsymbol{D}(s_k)\boldsymbol{Z}_k\Delta y_k,$$

*which is different from $\boldsymbol{Z}_k^\Delta$ - observe that $\tilde{\boldsymbol{Z}}_k^\Delta$ may take negative values and it is $\mathcal{F}_{s_{k-1}}$-conditionally $\mathcal{N}(\tilde{\mu}_k, \tilde{\Sigma}_k^2)$-distributed, with $\tilde{\mu}_k = \tilde{\boldsymbol{Z}}_{k-1}^\Delta + \boldsymbol{Q}\boldsymbol{Z}_k\Delta s_k$ and $\tilde{\Sigma}_k = diag\{\boldsymbol{D}(s_k)\boldsymbol{Z}_k\}$. Instead, each $\boldsymbol{Z}_k^\Delta$ is non-negative for each $k$ and it is $\mathcal{F}_{s_k}$- conditionally log-normally distributed, i.e.*

$$(\log \boldsymbol{Z}_k^\Delta \mid \mathcal{F}_{s_k}) \sim \mathcal{N}(\mu_k, \Sigma_k)$$

*where* $\mu_k = \log \boldsymbol{Z}_{k-1} + \boldsymbol{Q}\Delta s_k - \frac{1}{2}\boldsymbol{D}^2(t_k)\Delta$ *and* $\Sigma_k = diag\{\boldsymbol{D}(s_k)\}$.

*Moreover, discretizing the differential of* $\boldsymbol{Z}_t$ *is more in line with the use of a standard scheme as discussed in Section 2.2.4, whereas our approach corresponds to the exact approximation schemes as we explained sub-section 2.2.4.2.*

**Remark 2.3.14** (Similarity to Splitting-up Schemes [Gla92], [BGR90])**.** *If* $\boldsymbol{Q}$ *and* $\boldsymbol{D}$ *are commuting, then (2.3.19) can be expressed as*

$$\boldsymbol{Z}^{\Delta}_{k+1} = \mathbf{exp}\{\boldsymbol{Q}\Delta s_k\}\,\mathbf{exp}\left\{-\frac{1}{2}\boldsymbol{D}^2(s_k)\Delta + \boldsymbol{D}(s_k)\Delta y_k\right\}\boldsymbol{Z}^{\Delta}_k \quad \text{with } \boldsymbol{Z}^{\Delta}_0 = \boldsymbol{Z}_{s_0}, \tag{2.3.20}$$

*for* $s_k = s_0, s_1, \ldots, T$.

*In this way, each step in the recursion can be seen as the product of two matrix operators*

$$\mathcal{Q}^{\Delta}_k := \mathbf{exp}\{\boldsymbol{Q}\Delta s_k\}$$

*and*

$$\mathcal{P}^{\Delta}_k := \mathbf{exp}\left\{-\frac{1}{2}\boldsymbol{D}^2(s_k)\Delta + \boldsymbol{D}(s_k)\Delta y_k\right\}$$

*applied to the value estimated in the previous time-step:*

$$\boldsymbol{Z}^{\Delta}_{k+1} = \mathcal{Q}^{\Delta}_k \cdot \mathcal{P}^{\Delta}_k \boldsymbol{Z}^{\Delta}_k.$$

*The matrix operators* $\mathcal{Q}^{\Delta}_k$ *and* $\mathcal{P}^{\Delta}_k$ *can be understood as approximating certain matrices* $\mathcal{Q}_{[t_k,t)}$ *and* $\mathcal{P}_{[t_k,t)}$ *that are the solution operators to the ODE*

$$d\widetilde{\boldsymbol{Z}}^1_t = \boldsymbol{Q}\widetilde{\boldsymbol{Z}}^{(1)}_t dt, \qquad \widetilde{\boldsymbol{Z}}^{(1)}_{t_k} = \mathcal{P}_{[t_k,t)}\widetilde{Z}^{(2)}(t_k)$$

*and the SDE*

$$d\widetilde{\boldsymbol{Z}}^{(2)}_t = \boldsymbol{D}(t)\widetilde{\boldsymbol{Z}}^{(2)}_t dy_t, \qquad \widetilde{\boldsymbol{Z}}^{(2)}(t_k) = \boldsymbol{Z}_{t_k},$$

$t \in [t_k, t_{k+1})$*, respectively. Thus, the approximation is analogous to the splitting-up approximation scheme proposed in [Gla92] for solving SDEs in filtering, see also [BGR90]. Under such schemes, the SDE to be solved is split-up into two equations involving simpler (possibly deterministic) differential operators. The split-up recursions are then defined in-terms of the semigroups associated to each of the differential operators. The semigroups are, in this case, analogous to the matrix operators* $\mathcal{Q}^{\Delta}_k$ *and* $\mathcal{P}^{\Delta}_k$.

We have the following result about the convergence of the approximation scheme $(\boldsymbol{Z}^{\Delta}_k)_k$.

**Proposition 2.3.15.** *Supose that sufficient boundedness and regularity conditions on the coefficients of the SDE (2.3.14) hold to guarantee the existence and uniqueness of the solution. Define the sequence* $\left\{\boldsymbol{Z}^{\Delta}_k\right\}_{k=0}^{m(t)}$ *according to (2.3.19). Then this sequence forms a strong approximation scheme of order* 0.5 *to the quasi-exact solution* $(\boldsymbol{Z}_s)_{s\in[t_0,t)}$ *of SDE (2.3.14).*

*Proof.* See Appendix 2.A.4. □

#### 2.3.2.2 Quasi - Exact Approximation for $\sigma(N_t^{i,j})$, $\sigma(\mathcal{T}_t(i))$ and $\sigma(\Gamma_t(i))$

Analogously to (2.3.19), we are then able to compute quasi-exact approximations also of the solutions for $\boldsymbol{\sigma}(N_t^{i,j})$, $\boldsymbol{\sigma}(\mathcal{T}_t(i))$, and $\boldsymbol{\sigma}(\Gamma_t(i))$ on the basis of (2.3.13). To do this, we will now define suitable approximating sequences $\boldsymbol{S}_k^\Delta$ for the solution (2.3.13) to equation (2.3.10), for each step-size $\Delta$. We shall derive such an approximation following the lines of our derivation for $\boldsymbol{Z}_k^\Delta$, which is to approximate the solution $\boldsymbol{S}_t$ directly rather than the increment $d\boldsymbol{S}_t$. Indeed, we note that at time $t_{k+1}$ we can express $\boldsymbol{S}_{t_{k+1}}$

$$\boldsymbol{S}_{t_{k+1}} = \boldsymbol{\Phi}(t_{k+1}, t_0)\left(\boldsymbol{S}_{t_0} + \int_{t_0}^{t_{k+1}} \boldsymbol{\Phi}(s,t_0)^{-1}\left[\boldsymbol{a}(\boldsymbol{\Phi}(s,t_0)) - \boldsymbol{D}(s)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))\right]ds + \int_{t_0}^{t_{k+1}} \boldsymbol{\Phi}^{-1}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))dY_s\right). \tag{2.3.21}$$

First, we observe that $\boldsymbol{\Phi}(t,t_0)$ is a process that we can approximate by the quasi-exact approximation

$$\boldsymbol{\Phi}(t,t_0) \approx \mathbf{exp}\left\{\boldsymbol{\zeta}_k^\Delta\right\} =: \boldsymbol{\Phi}_k^\Delta \tag{2.3.22}$$

given in (2.3.17), for $\boldsymbol{\Phi}_0^\Delta = \boldsymbol{\Phi}(t_0,t_0) = \boldsymbol{I}$.

Next, in order to write the right-hand expression in (2.3.21) as a summation of increments, let us replace the integral in $dt$ with a trapezoidal approximation since it is just a Riemann integral. In other words

$$\begin{aligned}
&\int_{t_0}^{t_{k+1}} \boldsymbol{\Phi}(s,t_0)^{-1}\left[\boldsymbol{a}(\boldsymbol{\Phi}(s,t_0)) - \boldsymbol{D}(s)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))\right]ds \\
&\quad\approx \sum_{l=0}^{k}\left\{(\boldsymbol{\Phi}_{l+1}^\Delta)^{-1}\left[\boldsymbol{a}(\boldsymbol{\Phi}_{l+1}^\Delta) - \boldsymbol{D}(t_{l+1})\boldsymbol{b}(\boldsymbol{\Phi}_{l+1}^\Delta)\right] + (\boldsymbol{\Phi}_l^\Delta)^{-1}\left[\boldsymbol{a}(\boldsymbol{\Phi}_l^\Delta) - \boldsymbol{D}(t_l)\boldsymbol{b}(\boldsymbol{\Phi}_l^\Delta)\right]\right\}\frac{\Delta}{2} \\
&\quad=: \sum_{l=0}^{k}\left(\bar{\psi}_{l+1} + \bar{\psi}_l\right)\frac{\Delta}{2} = \sum_{l=0}^{k}\psi_l\frac{\Delta}{2}
\end{aligned}$$

where we have defined

$$\bar{\psi}_l := (\boldsymbol{\Phi}_l^\Delta)^{-1}\left[\boldsymbol{a}(\boldsymbol{\Phi}_l^\Delta) - \boldsymbol{D}(t_l)\boldsymbol{b}(\boldsymbol{\Phi}_l^\Delta)\right]$$

and

$$\psi_l = \bar{\psi}_{l+1} + \bar{\psi}_l.$$

Furthermore, we shall replace the stochastic integral in $dY$ with the approximating sequence that is the sums of increments $\Delta Y_l$ multiplied by the values at the left-endpoints of the integrand, that is

$$\int_{t_0}^{t_{k+1}} \boldsymbol{\Phi}^{-1}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))dY_s \approx \sum_{l=0}^{k}(\boldsymbol{\Phi}_l^\Delta)^{-1}\boldsymbol{b}(\boldsymbol{\Phi}_l^\Delta)\Delta Y_l.$$

Recalling that $Y$ is a $\bar{P}$-Brownian Motion, then by Donsker's invariance principle, the summation on the right hand side converges to the stochastic integral on the left-hand side, if the integrands satisfy appropriate boundedness and measurability conditions . (In our case, boundedness of the elements of the $\boldsymbol{Q}$-matrix is sufficient).

Applying the foregoing approximations for the integral terms in the right hand side of (2.3.21),

$$\begin{aligned}\boldsymbol{S}_{t_{k+1}} &= \boldsymbol{\Phi}(t_{k+1},t_0)\left(\boldsymbol{S}_{t_0} + \int_{t_0}^{t_{k+1}} \boldsymbol{\Phi}(s,t_0)^{-1}\left[\boldsymbol{a}(\boldsymbol{\Phi}(s,t_0)) - \boldsymbol{D}(s)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))\right]ds\right.\\ &\qquad \left. + \int_{t_0}^{t_{k+1}} \boldsymbol{\Phi}^{-1}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))dY_s\right)\\ &\approx \boldsymbol{\Phi}^{\Delta}_{k+1}\left(\boldsymbol{S}^{\Delta}_k + \sum_{l=0}^{k}\psi_l\frac{\Delta t_l}{2} + \sum_{l=0}^{k}(\boldsymbol{\Phi}^{\Delta}_l)^{-1}\boldsymbol{b}(\boldsymbol{\Phi}^{\Delta}_l)\Delta Y_l\right). \end{aligned} \tag{2.3.23}$$

This motivates the definition of $\boldsymbol{S}^{\Delta}_{k+1}$ as the right hand expression in (2.3.23). We finally have

**Definition 2.3.16.** *On the same jump-adapted time discretization for $\boldsymbol{\Phi}^{\Delta}_k$, we define the sequence $(\boldsymbol{S}^{\Delta}_k)_{k=0}^{m(t)}$ as follows*

$$\begin{aligned}\boldsymbol{S}^{\Delta}_{k+1} &= \boldsymbol{\Phi}^{\Delta}_{k+1}\left(\boldsymbol{S}^{\Delta}_0 + \sum_{l=0}^{k}\psi_l\frac{\Delta t_l}{2} + \sum_{l=0}^{k}(\boldsymbol{\Phi}^{\Delta}_l)^{-1}\boldsymbol{b}(\boldsymbol{\Phi}^{\Delta}_l)\Delta Y_l\right)\\ \boldsymbol{S}^{\Delta}_0 &= \boldsymbol{S}_{t_0}\end{aligned} \tag{2.3.24}$$

*where*

$$\begin{aligned}\psi_k &= \bar{\psi}_{k+1} + \bar{\psi}_k\\ \bar{\psi}_k &= (\boldsymbol{\Phi}^{\Delta}_k)^{-1}\left[\boldsymbol{a}(\boldsymbol{\Phi}^{\Delta}_k) - \boldsymbol{D}(t_k)\boldsymbol{b}(\boldsymbol{\Phi}^{\Delta}_k)\right]\\ \bar{\psi}_0 &= 0.\end{aligned} \tag{2.3.25}$$

Each $\boldsymbol{S}^{\Delta}_k$ in the sequence can also be written in the following recursive form

$$\begin{aligned}\boldsymbol{S}^{\Delta}_{k+1} &= \boldsymbol{\Phi}^{\Delta}_{k+1}\left(\frac{\boldsymbol{S}^{\Delta}_k}{\boldsymbol{\Phi}^{\Delta}_k} + \psi_k\frac{\Delta t_k}{2} + (\boldsymbol{\Phi}^{\Delta}_k)^{-1}\boldsymbol{b}(\boldsymbol{\Phi}^{\Delta}_k)\Delta Y_k\right)\\ \boldsymbol{S}^{\Delta}_0 &= \boldsymbol{S}_{t_0}.\end{aligned} \tag{2.3.26}$$

In Proposition 2.3.17 we shall state another result, namely, that the scheme $\boldsymbol{S}^{\Delta}_k$ approximates $\boldsymbol{S}_t$ with a strong order of at least 0.5. We have the following result concerning the convergence of the approximation scheme $\boldsymbol{S}^{\Delta}_k$:

**Proposition 2.3.17.** *Supose that sufficient boundedness and regularity conditions on the coefficients of the SDE (2.3.10) hold to guarantee the existence and uniqueness of the solution. Define the sequence $\left\{\boldsymbol{S}^{\Delta}_k\right\}_{k=0}^{m(t)}$ according to (2.3.26), with $\boldsymbol{\Phi}^{\Delta}_k$ as given in (2.3.22). Then this sequence forms a strong approximation scheme of order at least 0.5 to the process $(\boldsymbol{S}_s)_{s\in[t_0,t]}$ given in (2.3.13) that is the quasi-exact solution of the general linear SDE (2.3.10).*

*Proof.* See Appendix 2.A.5. □

## 2.4 The Calibration Model

In this Section we describe our proposed calibration method for a market model based on CTMCs. It extends the approach proposed by [EHJ00] to the continuous-time case, retaining the random jump-times of the CTMC $\boldsymbol{X}$. First we formulate the state-observations system to which the

Wonham filter can be applied, in conjunction with *Bayesian updating* at jump-times. Calibration is performed by applying the filter-based EM algorithm. In order to numerically implement this algorithm, we propose the use of Quasi-Exact approximations to compute the Wonham filters. In Section 2.4.1 we review the pricing of payoffs that depend on CTMCs. In Section 2.4.2 we formulate the Filtering Problem applied to noisy-observations of the log-price. Finally in Section 2.4.3 we present the Calibration Algorithm by Filter-based EM.

### 2.4.1 Review of Pricing for CTMCs

Let us first consider payoffs that are simple functions of the terminal value of $\xi$ (equivalently, $\boldsymbol{X}$), i.e. $H(\xi_T) = \sum_i h^i \mathbf{1}_{\{\xi_T = i\}} = \langle \boldsymbol{H}, \boldsymbol{X}_T \rangle$ with $\boldsymbol{H} = (h_1, h_2, \dots, h_N)^\top$. In the general case, payoffs may depend on the entire path. Let $\mathcal{F}_t$ be the natural filtration of $\xi$. Let $r(t) = \sum_{i=1}^N \rho_i \mathbf{1}_{\{\xi_t = i\}}$, $\rho^i \in \mathbb{R}^+ \ \forall i = 1 \dots N$ be the risk-free short rate. We then have $r(t) = \langle \boldsymbol{R}, \boldsymbol{X}_t \rangle$ where $\boldsymbol{R} := (\rho_1, \rho_2, \dots, \rho_N)^\top$. The price at time $t$ of a payoff $H(\xi_T) = \langle \boldsymbol{H}, \boldsymbol{X}_T \rangle$ is given by

$$E\left[e^{-\int_t^T r(s)ds} H(\xi_T) \mid \mathcal{F}_t\right] = E\left[e^{-\int_t^T \langle \boldsymbol{R}, \boldsymbol{X}_s \rangle ds} \langle \boldsymbol{H}, \boldsymbol{X}_T \rangle \mid \mathcal{F}_t\right].$$

By the Markov Property

$$E\left[e^{-\int_t^T \langle \boldsymbol{R}, \boldsymbol{X}_s \rangle ds} \langle \boldsymbol{H}, \boldsymbol{X}_T \rangle \mid \mathcal{F}_t\right] = E\left[e^{-\int_t^T \langle \boldsymbol{R}, \boldsymbol{X}_s \rangle ds} \langle \boldsymbol{H}, \boldsymbol{X}_T \rangle \mid \xi_t\right] =: F(t, H_T, \xi_t).$$

Since $\xi_t \in E$, we can compute this as follows

$$\begin{aligned} F(t, H_T, \xi_t) &= \sum_{i=1}^N E\left[e^{-\int_t^T r(s)ds} H(\xi_T) \mid \xi_t = i\right] \mathbf{1}_{\{\xi_t = i\}} && (2.4.1) \\ &=: \sum_{i=1}^N F^i(t, H_T) \mathbf{1}_{\{\xi_t = i\}} && (2.4.2) \\ &= \langle \boldsymbol{F}(t, H_T), \boldsymbol{X}_t \rangle, && (2.4.3) \end{aligned}$$

with

$$\boldsymbol{F}(t, H) = (F^1, F^2, \dots, F^N)^\top (t, H).$$

As we have seen in Part I, there are two approaches to compute $F(t, H_T, \xi_t) = \langle \boldsymbol{F}(t, H_T), \boldsymbol{X}_t \rangle$. The first approach which applies to simple claims $H(\cdot)$ is treated in [Nor03] (see also [EHJ00] for the case of bonds), while the second approach, discussed in [PR10b] and extended in [MPR13], applies to path-dependent claims and to the time-inhomogeneous case as well.

#### 2.4.1.1 Direct ODE approach for simple claims

In the first approach one considers an ODE in $\boldsymbol{F}(t, H_T)$ using martingale arguments related to the *no-arbitrage* condition. The solution is then available in explicit form as a matrix exponential. It can be shown that

**Proposition 2.4.1.** *The vector $\boldsymbol{F}(t, H_T)$ whose components are the $\{\xi_t = x^i\}$-conditional prices $F^i(t, H)$ satisfies*

$$\frac{d\boldsymbol{F}}{dt}(t, H_T) = (\boldsymbol{R} - \boldsymbol{Q})\boldsymbol{F}(t, H_T), \qquad F(T, H_T) = \boldsymbol{H} \tag{2.4.4}$$

*for all $t \in [0, T]$.*

*Proof.* Since this is a well-known result, we do not give the proof, but instead cite various references where this pricing equation is derived. For example, in Section D Equation 3.14 of [Nor03], which treats the general setting of markets with CTMC-driven stocks, this equation is obtained by applying Ito's lemma to the value function of portfolios and then applying the martingale property that is due to the *no-arbitrage* condition; equation (2.4.4) is just the special case of a portfolio consisting only of one bond. Other references are Section 3 of [EHJ00]. In Proposition 3.1 in [Lan00] where it is obtained for more general Markovian jump-diffusion processes and in Corollary 6.4 [BKR97], which treats general Levy processes. □

In component form, (2.4.4) is given by the equation

$$\frac{d}{dt}F^i(t,H_t) = \rho^i F^i(t,H_t)^i - q_i F^i(t,H_t) - \sum_{\substack{j=1,\dots,N \\ j\neq i}} F^j(t,H_t)q_{ij}, \tag{2.4.5}$$

for $i = 1,\dots,N$. Equation (2.4.5) will be needed in Section 2.4.2.4.

**Remark 2.4.2.** *In general one may be dealing with a payoff function $H$ such that the function $\boldsymbol{F}(t,H)$ is not continuously differentiable and hence Proposition 2.4.1 does not hold. Sufficient conditions are discussed in detail in [Nor05] for the continuous differentiability of $F(t,H,\xi)$. In that paper it was also noted that most derivative payoffs in practice satisfy these conditions, hence we shall simply assume this in what follows.*

We immediately get that

**Proposition 2.4.3.** *The vector $\boldsymbol{F}(t,H)$ is given by*

$$\boldsymbol{F}(t,H_T) = \mathbf{exp}\left\{(\boldsymbol{Q}-\boldsymbol{R})(T-t)\right\}\boldsymbol{H}, \tag{2.4.6}$$

*where $\mathbf{exp}(\cdot)$ denotes the matrix exponential*

$$\mathbf{exp}(\boldsymbol{A}) = \sum_{k=0}^{\infty}\frac{\boldsymbol{A}^k}{k!}.$$

For the case of zero-coupon bonds, with $\boldsymbol{H} = (1,1,\dots,1)^\top := \underline{\mathbf{1}}$ and denoting $\boldsymbol{V}(t,T) = \boldsymbol{F}(t,\underline{\mathbf{1}})$ we have

**Corollary 2.4.4.** *The price vector $\boldsymbol{V}(t,T)$ for zero-coupon bonds is given by*

$$\boldsymbol{V}(t,T) = \mathbf{exp}\left\{(\boldsymbol{Q}-\boldsymbol{R})(T-t)\right\}\underline{\mathbf{1}}. \tag{2.4.7}$$

#### 2.4.1.2 Prototype Product approach for simple and complex claims

The main topic of Part I, see also [MPR13] is the so-called *Prototype Product* approach, that is an alternative pricing approach applicable also to complex claims, such as path-dependent payoffs. For general time-inhomogeneous CTMCs, the elements of the Q-matrix vary in time, i.e., $\boldsymbol{Q} = \boldsymbol{Q}(t)$. By restricting our attention to those time-inhomogeneous CTMCs whose $Q$-matrices change only at the jump times $\tau_n$, we have that $\boldsymbol{Q}(t) = \boldsymbol{Q}(n)$ for each $t = t_n$. If we further restrict our attention to the case when

$$(\tau_{n+1} - \tau_n \mid \boldsymbol{X}_n = \boldsymbol{e}_i) \sim \mathcal{E}xp(q_i(n))$$

holds, where $q_i(n) = \sum_{j\neq i} q_{i,j}(n) = \sum_{i,j} q_{i,j}(n)$, then for such time-inhomogeneous CTMCs, the *Prototype Product* is also applicable. We shall recall here this approach, omitting details and proofs since they have already been discussed in Part I.

Let us suppose, without loss of generality, that $t = \tau_0$. We have that

$$F(t, H_T, \xi_t) = E\left\{\exp\left[-\sum_{i=N_t}^{N_T-1} r_i(\tau_{i+1}-\tau_i) - r_{N_T}(T-\tau_{N_T})\right] H(\xi_T) \mid \xi_t\right\}. \tag{2.4.8}$$

Let us define $N_{t,T} := N_T - N_t$ and consider the $N_{t,T}$ conditional expectation

$$F(t, H_T, \xi_t \mid N_{t,T}) = E\left\{\exp\left[-\sum_{i=N_t}^{N_T-1} r_i(\tau_{i+1}-\tau_i) - r_{N_T}(T-\tau_{N_T})\right] H(\xi_T) \mid \xi_t, N_{t,T}\right\}, \tag{2.4.9}$$

where we have

$$F(t, H_T, \xi_t) = E\left\{F(t, H, \xi_t \mid N_{t,T}) \mid X_t\right\}. \tag{2.4.10}$$

The expectation in (2.4.10) is just

$$E\left\{F(t, H, N_{t,T}) \mid X_t\right\} = \sum_{k=0}^{\infty} F(t, H, k) P(N_{t,T} = k \mid X_t). \tag{2.4.11}$$

The difficulty in computing this expression is due to the fact that it involves a quite cumbersome calculation of the weights $P\left[N_{t,T} \mid X_t\right]$, for very general Markov Chains. Instead, as we have seen in Part I Section 1.4.1, an approximation to the expectation can be computed more efficiently by Monte Carlo. We recall here how such an approximation is computed, for details see Part I Section 1.4.1.

Let us define the sequence of matrices $\widetilde{\boldsymbol{Q}}(n)$ as

$$\overline{\boldsymbol{Q}}(n) = (\bar{q}_{i,j}(n))_{1\leq i,j\leq N} \quad \text{with} \quad \bar{q}_{i,j}(n) = \begin{cases} \frac{q_{i,j}(n)}{r^i + q_i(n)} & i \neq j \\ 0 & i = j \end{cases} \tag{2.4.12}$$

where $r^i = r(\xi)$ when $\xi = x^i$. Let $\xi$ be any $E$-valued random variable and let us define

$$H_0(\cdot) = H(\cdot) = \sum_{i=1}^{N} h^i \mathbf{1}_{\{\cdot = i\}}$$

and define the vectors $H_n$, representing the sequence of *Prototype Payoffs* according to (1.4.11) in Part I, Section 1.4. These vectors can be computed recursively as follows (see Proposition 1.4.5, Part I, Section 1.4.1):

$$\boldsymbol{H}_n(\xi) = \widetilde{\boldsymbol{Q}}(n)\, \boldsymbol{H}_{n-1}(\xi). \tag{2.4.13}$$

Recall that $F(t, H_T, X_t) = \langle \boldsymbol{F}(t, H_T), \boldsymbol{X}_t \rangle$ with $\boldsymbol{F}(t, H_T) = (F^1(t, H_T), \ldots, F^N(t, H_t))^\top$. Finally, the components $F^i(t, H_T)$ can be approximated as follows (see Proposition 1.4.6 from Part I, Section 1.4.1)

**Proposition 2.4.5.**

$$\begin{aligned} F^i(t, H_T) \approx & E\left\{\tilde{F}(t, H, N_{t,T}) \mid \boldsymbol{X} = \boldsymbol{e}_i\right\} \\ = & \frac{1}{2} E\left\{\boldsymbol{e}_i^\top \left(1 + \widetilde{\boldsymbol{Q}}(N_{t,T}+1)\right) \widetilde{\boldsymbol{Q}}(N_{t,T}) \cdots \widetilde{\boldsymbol{Q}}(1) H_0(\xi)\right\}. \end{aligned} \tag{2.4.14}$$

The expectation in Proposition 2.4.5 may be efficiently computed by Monte - Carlo simulation. In fact, applying a Monte - Carlo simulation to compute the expectations in (2.4.14) converges faster than applying Monte-Carlo directly to compute (2.4.8). That is due to the Variance Reduction effect by conditioning on $N_t$ for each *Prototype Product*. Hence the *Prototype Product* approach also leads to an improved Monte-Carlo scheme over a *plain vanilla* Monte - Carlo.

We have shown how to apply the *Prototype Product* approach only to the case of path-independent claims. In Part I, it was shown that this approach also extends to the case of path-dependent claims. Hence the *Prototype Product* approach, while seemingly indirect, nevertheless offers the important advantage of being easily generalized to the case of path-dependent derivative payoffs and time-inhomogeneous CTMCs.

### 2.4.2 Filtering in Continuous - Time with Bayesian Updating

We now present a calibration approach which extends the approach in [EHJ00] to the continuous-time case, retaining the random jump-times of the CTMC $\boldsymbol{X}$. To this effect, we must first define our incomplete-information problem with continuous-time observation noise, in the form of a scaled Brownian Motion $\eta W_t$. In [EHJ00], the observations are the yield of noise-corrupted prices, where the noise is multiplicative on the price; in order to simplify the presentation we shall instead define as our observations process the noise-corrupted log-price[6], that is, the log of $F(t, H, \boldsymbol{X})$ multiplied by a scaled lognormal process:

$$F(t, H, \boldsymbol{X})e^{\eta W_t} =: e^{-y_t},$$

or equivalently

$$y_t = \log F(t, H, \boldsymbol{X}_t) + \eta W_t. \tag{2.4.16}$$

Let us now derive the state-observations system for $\boldsymbol{X}$ and $y_t$ in a form suitable for the application of filtering. Re-writing (2.4.16) gives

$$\begin{aligned} y_t = &\quad \log F(t, H, \boldsymbol{X}_t) + \eta W_t \\ = &\quad \log\{\langle \boldsymbol{F}(t, H_T), \boldsymbol{X}_t)\rangle\} + \eta W_t \end{aligned}$$

Let the vector function $\boldsymbol{C}(t)$ be defined such that

$$\langle \boldsymbol{C}(t), \boldsymbol{X}_t\rangle = \log\{\langle \boldsymbol{F}(t, H_T), \boldsymbol{X}_t\rangle\} = \langle \log(\boldsymbol{F}(t, H_T)), \boldsymbol{X}_t\rangle,$$

hence

$$y_t = \langle \boldsymbol{C}(t), \boldsymbol{X}_t\rangle + \eta W_t, \tag{2.4.17}$$

[6]Defining the observations process to be the yield of noise-corrupted prices leads to the equation

$$\bar{y}_t = -\frac{1}{T-t}\log F(t, H, \boldsymbol{X}_t) - \frac{1}{T-t}\eta W_t \tag{2.4.15}$$

instead of (2.4.16). Since no generality is lost by working with the log-price instead of the yield, we choose to work on the log-price in order to avoid unwieldy formulas involving the time-derivative due to the $\frac{1}{T-t}$ coefficient. In practice, our results can be applied to the case of noisily observed yields by a simple transformation of the observations process $\bar{y}_t = -\frac{1}{T-t}y_t$.

Let $c(t, \boldsymbol{X}_t) := \langle \boldsymbol{C}(t), \boldsymbol{X}_t \rangle$. We shall further re-write equation (2.4.17) in a jump-diffusion form. To this end, let us recall the Theorem 2.2.1 and also Proposition 2.2.2. Applying Proposition 2.2.2 to $c(t, \boldsymbol{X}_t)$ then substituting in equation (2.4.17), we obtain

$$\begin{aligned} y_t &= \langle \boldsymbol{C}, \boldsymbol{X}_t \rangle + \int_0^t \eta dW_s \\ &= \langle \boldsymbol{C}(0), \boldsymbol{X}_0 \rangle + \int_0^t \langle \dot{\boldsymbol{C}}(s), \boldsymbol{X}_s \rangle ds + \int_0^t \langle \boldsymbol{C}(s), \boldsymbol{Q}\boldsymbol{X}_{s-} \rangle ds + \int_0^t \langle \boldsymbol{C}(s), d\boldsymbol{M}_s \rangle + \int_0^t \eta dW_s \end{aligned} \tag{2.4.18}$$

We shall therefore work with the following state-observations system:

$$\begin{cases} \boldsymbol{X}_t = \boldsymbol{X}_0 + \int_0^t \boldsymbol{Q}\boldsymbol{X}_s ds + M_t \\ y_t = \langle \boldsymbol{C}(0), \boldsymbol{X}_0 \rangle + \int_0^t \langle \dot{\boldsymbol{C}}(s), \boldsymbol{X}_s \rangle ds + \int_0^t \langle \boldsymbol{C}(s), \boldsymbol{Q}\boldsymbol{X}_{s-} \rangle ds + \int_0^t \langle \boldsymbol{C}(s), d\boldsymbol{M}_s \rangle + \int_0^t \eta dW_s \end{cases} \tag{2.4.19}$$

We wish now to compute the filtered estimates $E[f(\xi_t) \mid \mathcal{F}_t^y] := \pi_t(f)$. Note that to apply filtering to system (2.4.19) it is not enough to apply a *Wonham filter* because the observations process includes also the martingale term

$$\int_0^t \langle \boldsymbol{C}(s), d\boldsymbol{M}_s \rangle$$

for the jumps, whereas in the *Wonham filter* problem, the jump process enters only in the jump-modulated drift coefficient. Therefore, a direct filtering approach would require the use of filters for Jump-Diffusions, for example, using a combination of the techniques in Chapter 7 and Chapter 8 of [EAM08]. Instead we shall take a different approach that will still allow us to use the *Wonham filter*, inspired by [FR10b] (see also [CC12]) whereby we utilize *Bayesian updating* at the jump times. First, note that if we write $c(t, \boldsymbol{X}_t) := \langle \boldsymbol{C}(t), \boldsymbol{X}_t \rangle = c(0, \boldsymbol{X}_0) + \int_0^t dc(s, \boldsymbol{X}_s)$ and apply the differentiation rule from Proposition 2.2.2, then (2.4.19) is simply the following

$$\begin{cases} \boldsymbol{X}_t = \boldsymbol{X}_0 + \int_0^t \boldsymbol{Q}\boldsymbol{X}_s ds + M_t \\ y_t = c(0, \boldsymbol{X}_0) + \int_0^t dc(s, \boldsymbol{X}_s) + \int_0^t \eta dW_s \end{cases} \tag{2.4.20}$$

The key is then to re-write $\boldsymbol{X}_t$ and $c(t, \boldsymbol{X}_t)$ initialized at the time $\tau_{N(t^-)}$ instead of $t_0$, that is

$$\begin{cases} \boldsymbol{X}_t = \boldsymbol{X}_{N(t^-)} + \int_{\tau_{N(t^-)}}^t \boldsymbol{Q}\boldsymbol{X}_s ds + \int_{\tau_{N(t^-)}}^t M_s \\ y_t = c(\tau_{N(t^-)}, \boldsymbol{X}_{N(t^-)}) + \int_{\tau_{N(t^-)}}^t dc(s, \boldsymbol{X}_s) + \int_0^t \eta dW_s \end{cases} \tag{2.4.21}$$

Supposing that $t \in (\tau_{N(t^-)}, \tau_{N(t^-)+1}]$, this leads to a much simpler analysis, distinguishing only two possible cases: either $t = \tau_{N(t^-)+1}$, i.e. $t$ falls on the next jump time, or $t \in (\tau_{N(t^-)}, \tau_{N(t^-)+1})$, i.e. $t$ is in-between jumps.

#### 2.4.2.1 The case when $t$ falls in between jump times

Denoting by $y_t^1$ the observation process $y_t$ in between jump times, the system (2.4.19) reduces to

$$\begin{cases} \boldsymbol{X}_t = \boldsymbol{X}_{\tau_{N(t^-)}} + \int_{\tau_{N(t^-)}}^t \boldsymbol{Q}\boldsymbol{X}_s ds + \int_{\tau_{N(t^-)}}^t dM_s \\ y_t^1 = \langle \boldsymbol{C}(\tau_{N(t^-)}), \boldsymbol{X}_{N(t^-)} \rangle + \int_{\tau_{N(t^-)}}^t \langle \dot{\boldsymbol{C}}(s), \boldsymbol{X}_s \rangle ds + \int_0^t \eta dW_s \end{cases} \tag{2.4.22}$$

To this system one can directly apply the *Wonham Filter* for $t \in (\tau_{N(t^-)}, \tau_{N(t^-)+1})$. Note that the equation for $y_t^1$ in (2.4.22) arises from the differentiation rule (2.2.4) and the fact that $d\boldsymbol{X}$ is zero in between jump times. By translating the system to the initial time $\tau_{N(t)}$ instead of $t = 0$, and then substituting $y^1(t)$ into $y(t)$ and $\langle \dot{\boldsymbol{C}}(t), \boldsymbol{X}_t \rangle$ into $h(t, \xi_t)$, we obtain exactly the system (2.2.5) in Section 2.2.2.

#### 2.4.2.2 The case when $t$ falls on the next jump time

In the case when $t$ falls on the next jump time, that is, when $t = \tau_{N(t^-)+1}$ then the increment $d\boldsymbol{X}$ is non-zero at $t$. Let us suppose that at $t$ a transition from $\boldsymbol{X}_{t^-} = \boldsymbol{e}_i$ to $\boldsymbol{X}_t = \boldsymbol{e}_j$ occurs, so at time $t$, $d\boldsymbol{X}_t = \boldsymbol{e}_j - \boldsymbol{e}_i$. Then

$$\begin{aligned} c(t, \boldsymbol{X}_t) &= \langle \boldsymbol{C}(\tau_{N(t)}), \boldsymbol{X}_{N(t)} \rangle + \int_{\tau_{N(t^-)}}^{\tau_{N(t^-)+1}} \langle \dot{\boldsymbol{C}}(s), \boldsymbol{X}_s \rangle ds + \int_{\tau_{N(t^-)}}^{\tau_{N(t^-)+1}} \langle \boldsymbol{C}(s), d\boldsymbol{X}_s \rangle \\ &= \langle \boldsymbol{C}(\tau_{N(t^-)}), \boldsymbol{e}_i \rangle + \int_{\tau_{N(t^-)}}^{\tau_{N(t^-)+1}} \langle \dot{\boldsymbol{C}}(s), \boldsymbol{e}_i \rangle ds + \langle \boldsymbol{C}(\tau_{N(t^-)+1}), \boldsymbol{e}_j \rangle - \langle \boldsymbol{C}(\tau_{N(t^-)+1}^-), \boldsymbol{e}_i \rangle \end{aligned}$$

Therefore

$$\begin{aligned} y_t &= \langle \boldsymbol{C}(\tau_{N(t^-)}), \boldsymbol{X}_{N(t^-)} \rangle + \int_{\tau_{N(t^-)}}^t \langle \dot{\boldsymbol{C}}(s), \boldsymbol{X}_s \rangle ds + \int_0^t \sigma dW_s + c(t, \boldsymbol{X}_t) - c(t^-, \boldsymbol{X}_{t^-}) \\ &= \langle \boldsymbol{C}(\tau_{N(t^-)}), \boldsymbol{X}_{N(t^-)} \rangle + \int_{\tau_{N(t^-)}}^t \langle \dot{\boldsymbol{C}}(s), \boldsymbol{X}_s \rangle ds + \int_0^t \sigma dW_s + \langle \boldsymbol{C}(t), \boldsymbol{X}_{N(t^-)+1} \rangle - \langle \boldsymbol{C}(t^-), \boldsymbol{X}_{N(t^-)} \rangle \\ &= y_t^1 + \langle \boldsymbol{C}(t), \boldsymbol{X}_{N(t^-)+1} \rangle - \langle \boldsymbol{C}(t^-), \boldsymbol{X}_{N(t^-)} \rangle. \end{aligned}$$

the system (2.4.19) becomes

$$\begin{cases} \boldsymbol{X}_t = \boldsymbol{X}_{\tau_{N(t^-)}} + \int_{\tau_{N(t^-)}}^t \boldsymbol{Q}\boldsymbol{X}_s ds + \int_{\tau_{N(t^-)}}^t M_s \\ y_t = y_t^1 + \langle \boldsymbol{C}(t), \boldsymbol{X}_{N(t^-)+1} \rangle - \langle \boldsymbol{C}(t^-), \boldsymbol{X}_{N(t^-)} \rangle. \end{cases} \tag{2.4.23}$$

In order to deal with this system, we note that, by defining $y_t^2$ as the increments of $c(t, \boldsymbol{X}_t)$, that is

$$y_t^2 := c(t, \boldsymbol{X}_t) - c(t^-, \boldsymbol{X}_{t^-}),$$

then system (2.4.23) is equivalent to a system with two observation processes

$$\begin{cases} \boldsymbol{X}_t = \boldsymbol{X}_0 + \int_0^t \boldsymbol{Q}\boldsymbol{X}_s ds + M_t \\ y_t^1 = \langle \boldsymbol{C}(\tau_{N(t)}), \boldsymbol{X}_{N(t)} \rangle + \int_{\tau_{N(t)}}^t \langle \dot{\boldsymbol{C}}(s), \boldsymbol{X}_s \rangle ds + \int_0^t \sigma dW_s \\ y_t^2 = \langle \boldsymbol{C}(t, \boldsymbol{X}_{N(t^-)+1} \rangle - \langle \boldsymbol{C}(t^-), \boldsymbol{X}_{N(t^-)} \rangle \end{cases} \tag{2.4.24}$$

of which, the subsystem $\left(\boldsymbol{X}_t, y_t^1\right)^\top$ is a state-observations system to which we can apply the Wonham filter up to time $t^- = \tau_{N(t)}^-$; the resulting filter must then be updated to time $t = \tau_{N(t)}$ using the information from $y_t^2$ at the jump times.

At first glance, it may not seem obvious that the observations process

$$y_t = y_t^1 + y_t^2$$

can indeed be decomposed and represented as the vector $\boldsymbol{Y}_t = (y_t^1, y_t^2)^\top$ so that $y_t = \langle \boldsymbol{Y}_t, \underline{\mathbf{1}} \rangle$. Such a decomposition is indeed possible - it stems from the fact that $y$ and $c(t, \boldsymbol{X}_t)$ have common jumps, hence the jump times $\tau_k$ are fully observed, and moreover, all the jump increments $y_t^2 = c(t, \boldsymbol{X}_t) - c(t^-, \boldsymbol{X}_{t^-})$ can be directly observed from the jump increments $y_t - y_{t^-}$. In fact they coincide and are zero in-between jumps:

$$y_t^2 = \begin{cases} y_t - y_{t^-} & \text{if } t \text{ is a jump time,} \\ 0 & \text{otherwise.} \end{cases} \tag{2.4.25}$$

We express this in the following

**Remark 2.4.6.** *Since $y_t$ and $\boldsymbol{X}_t$ share the same jump times, then every $\tau_k$ for $k = 0, 1, \ldots, N(t)$ is $\mathcal{F}_t^y$-measurable. From (2.4.25),*

$$y_t^2 = c(t, \boldsymbol{X}_t) - c(t^-, \boldsymbol{X}_{t^-}) = \sum_{k=0}^{N(t)} (y_t - y_{t^-}) \mathbf{1}_{\{t=\tau_k\}},$$

*which is $\mathcal{F}_t^y$ measurablefor all $t$.*

The component $y^1$ is simply the continuous part of $y$ in between jump times, and the component $y^2$ simply comes from observing the jump increments of $y$ at each point of discontinuity, that is, at each jump time of $y$.

On the basis of these foregoing observations, we propose a serial filtering algorithm where we iterate between the following steps:

1. We observe $y_t^2 = y_t - y_{t^-}$ continuously. At each jump time $t = \tau_k$ it is non-zero and equal to $c(\tau_k, \xi_k) - c(\tau_k -, \xi_{k-1})$, hence we can fully observe the history of $c(\tau_k, \xi_k)$ from $y_t^2$.

2. For $t$ in between jump times, that is $t \in (\tau_{N(t)}, \tau_{N(t)+1})$ apply the *Wonham Filter* to the system (2.4.22) to compute $\pi_{t+\tau_{N(t)}}(\boldsymbol{X}_t)$ for all $t \in [\tau_{N(t)}, \tau_{N(t)+1})$, using the initial condition $\pi_{\tau_{N(t)}}(\boldsymbol{X}_{\tau_{N(t)}})$ previously obtained at $t = \tau_{N(t)}$ (Note that $\pi_{\tau_{N(t)}}(\boldsymbol{X}_{\tau_{N(t)}}) \neq \pi_{\tau_{N(t)}}(\boldsymbol{X}_{\tau_{N(t)}-})$)

3. For $t$ falling on a jump time $\tau_{N(t)} := \tau_n$, we shall update the $\mathcal{F}_t^y$-conditional expectation $\pi_{t^-}(\boldsymbol{X}_{t^-})$ to obtain $\pi_t(\boldsymbol{X}_t)$ in the following manner:

#### 2.4.2.3 Procedure for updating the filter at jump times

First, observe that $\pi_t(\boldsymbol{X}_t) = (\pi_t^1, \pi_t^2, \ldots, \pi_t^N)^\top \boldsymbol{X}_t$, where $\pi_t^i(\boldsymbol{X}_t) = \pi_t(\mathbf{1}_{\{\xi_t = i\}})$. Hence

$$\begin{aligned}\pi_t^i(\boldsymbol{X}_t) =& P[\xi_t = i \mid \mathcal{F}_t^y] \\ =& E[\mathbf{1}_{\{\xi_t=i\}} \mid \mathcal{F}_t^y] \\ =& E\left[E[\mathbf{1}_{\{\xi_t=i\}} \mid \xi_{t^-}] \mid \mathcal{F}_t^y\right]\end{aligned}$$

$$\begin{aligned}=& E\left[\sum_{h=1, h\neq i}^{N} E[\mathbf{1}_{\{\xi_t=i\}} \mid \xi_{t^-} = h]\mathbf{1}_{\{\xi_{t^-}=h\}} \mid \mathcal{F}_t^y\right] \\ =& \sum_{h=1, h\neq i}^{N} P[\xi_t = i \mid \xi_{t^-} = h] \cdot P[\xi_{t^-} = h \mid \mathcal{F}_t^y] \\ =& \sum_{h=1, h\neq i}^{N} p_{h,i} \cdot P[\xi_{t^-} = h \mid \mathcal{F}_t^y]\end{aligned}$$

Next, for each $h = 1 \ldots, N$, $h \neq i$, we obtain an expression for $P[\xi_{t^-} = h \mid \mathcal{F}_t^y]$ in terms of $\pi_{t^-}(\boldsymbol{X}_{t^-})$ as follows:

$$\begin{aligned}P[\xi_{t^-} = h \mid \mathcal{F}_t^y] &= P[\xi_{t^-} = h \mid \mathcal{F}_{t^-}^y, y_t - y_{t^-}] \\ &= \sum_{k=1}^{N} P[\xi_{t^-} = h \mid \mathcal{F}_{t^-}^y, y_t - y_{t^-} = k]\mathbf{1}_{\{y_t - y_{t^-}=k\}} \\ &\propto \sum_{k=1}^{N} P[y_t - y_{t^-} = k \mid \mathcal{F}_{t^-}^y, \xi_{t^-} = h] \cdot P[\xi_{t^-} = h \mid \mathcal{F}_{t^-}^y]\mathbf{1}_{\{y_t - y_{t^-}=k\}} \\ &= \pi_{t^-}^h(\xi_{t^-}) \cdot \sum_{k=1}^{N} P[y_t - y_{t^-} = k \mid \mathcal{F}_{t^-}^y, \xi_{t^-} = h]\mathbf{1}_{\{y_t - y_{t^-}=k\}},\end{aligned}$$

where the form of the *emmission probabilities* $P[y_t - y_{t^-} = k \mid \mathcal{F}_{t^-}^y, \xi_{t^-} = h]$ is specified according to the particular dependence between $y_t$ and $\xi_t$ as defined in $c(t, \xi_t)$.

**Remark 2.4.7.** *This updating procedure is analogous to Corollary 4.2 (see also Algorithm 4.3) of [FR10b] . In our case the counting process is itself the factor process so that all jumps are common and, in particular, the jumps of the counting process are identical to the jumps of the factor process. Furthermore, unlike in [FR10b], we are considering only a single issuer of a financial derivative, that is, our counting process consists of only one component. Since, evidently, we are considering a much simpler model, the derivation of the updating procedure for the filter is much simpler in our case compared with the proof of Corollary 4.2 in [FR10b].*

**Remark 2.4.8** (Procedure for updating the filter for generic functions $f(\boldsymbol{X}_t)$ at jump times)**.** *The update of the filter $\pi_t(f(\boldsymbol{X}_t)) = \pi_t(H_t\boldsymbol{X}_t)$ from $\pi_{t^-}(f(\boldsymbol{X}_{t^-}))$ at jump times follows a procedure completely analogous to the update of $\pi_t(\boldsymbol{X}_t)$, by simply applying the state transformation $x^i \mapsto f(x^i) := f^i$, for $i = 1, \ldots, N$. In this case, the update is applied to $\pi_{t^-}(H_{t^-}\boldsymbol{X}_{t^-})$ instead of $\pi_{t^-}(\boldsymbol{X}_{t^-})$.*

**Remark 2.4.9.** *A subtle consequence that arises, however, from the Remark 2.4.6 must be mentioned, namely, if the map $(t,x) \mapsto c(t,x)$ is bijective (here we assume perfect knowledge of the true parameters that enter in $c(\cdot,\cdot)$) then for every jump time $t = \tau_k$, the inverse map $c(t,x) \mapsto x$ allows us to identify the value for $\xi_t$ at a jump time $t = \tau_k$ on the basis of the value of $c(t,\xi_t) - c(t-,\xi_{t^-})$. Hence, it follows from the bijectivity $(t,x) \mapsto c(t,x)$ and the fact that $y^2$ is $\mathcal{F}_t^y$-measurabale, that the value of $\xi_t$ can be backed-out from $c(t,\xi_t)$ and from observations of $y^2$ in this case. In other words (2.4.21) reduces to a degenerate filtering problem.*

*Nevertheless, in the case of a bijective map $(t,x) \mapsto c(t,x)$, in a practical model the parameters are unknown and $y$ (hence also $c(t,\xi_t)$ is not continuously observed, therefore the value of $\xi_t$ cannot be backed-out from observations of $y^2$, and instead one must still rely on the conditional distributions of $\xi_t$ via the filters.*

#### 2.4.2.4 Example: The case of zero-coupon bonds

In the particular case when $F(t,H,\boldsymbol{X}_t)$ represents the payoff function of a zero-coupon bond, recall that Proposition 2.4.3 holds with $\boldsymbol{H}_T = \underline{\mathbf{1}}$. Substituting the formula in Proposition 2.4.3 for $\boldsymbol{F}(t,H_T)$ in $F(t,H,\boldsymbol{X}) = \langle \boldsymbol{F}(t,\boldsymbol{H}_T), \boldsymbol{X}_t\rangle$ we have the following particular expression for $y_t$:

$$\begin{aligned} y_t &= \log\left\{\langle \boldsymbol{F}(t,\boldsymbol{H}_T), \boldsymbol{X}_t)\rangle\right\} + \sigma W_t \\ &= \log\left\{\langle \mathbf{exp}\left\{(\boldsymbol{Q}-\boldsymbol{R})(T-t)\right\}\boldsymbol{H}_T, \boldsymbol{X}_t)\rangle\right\} + \sigma W_t \end{aligned} \tag{2.4.26}$$

Now observe that $\boldsymbol{G}(t) := \mathbf{exp}\left\{(\boldsymbol{Q}-\boldsymbol{R})(T-t)\right\}\boldsymbol{H}$ is a product of a matrix (i.e. the matrix exponential $\mathbf{exp}\left\{(\boldsymbol{Q}-\boldsymbol{R})(T-t)\right\}$) with the vector $\boldsymbol{H}$, hence $\boldsymbol{G}(t)$ is a vector,

$$\boldsymbol{G}(t) = (g_1(t), g_2(t), \ldots, g_N(t))^\top.$$

The value of the scalar product $\langle \mathbf{exp}\left\{(\boldsymbol{Q}-\boldsymbol{R})(T-t)\right\}\boldsymbol{H}, \boldsymbol{X}_t\rangle$ can be written as

$$\begin{aligned} \langle \mathbf{exp}\left\{(\boldsymbol{Q}-\boldsymbol{R})(T-t)\right\}\boldsymbol{H}, \boldsymbol{X}_t\rangle &= \langle \boldsymbol{G}(t), \boldsymbol{X}_t\rangle \\ &= \sum_{i=1}^N g_i(t)\mathbf{1}_{\{\xi_t=i\}}, \end{aligned}$$

hence

$$\begin{aligned} \log\left\{\langle \boldsymbol{G}(t), \boldsymbol{X}_t\rangle\right\} &= \log\left\{\sum_{i=1}^N g_i(t)\mathbf{1}_{\{\xi_t=i\}}\right\} \\ &= \sum_{i=1}^N \log(g_i(t))\mathbf{1}_{\{\xi_t=i\}}. \end{aligned}$$

Hence $\bar{\boldsymbol{C}}(t) = \left(\log\left\{g_1(t)\right\}, \log\left\{g_2(t)\right\}, \ldots, \log\left\{g_N(t)\right\}\right)^\top$. We then have that

$$\log\left\{\langle \boldsymbol{G}(t), \boldsymbol{X}_t\rangle\right\} = \langle \bar{\boldsymbol{C}}(t), \boldsymbol{X}_t\rangle.$$

Let us suppose that $\xi_{N(t^-)} = i$. Then, for $t \in [\tau_{N(t^-)}, \tau_{N(t^-)+1})$

$$c(t,\boldsymbol{X}_t) = \log F(t,H,\xi_{N(t^-)}) = \log F(t,H,i).$$

Note that $F(t,H,i) = \langle \boldsymbol{F}(t,H), \boldsymbol{e}_i \rangle = \boldsymbol{F}(t,H)^i$. Let us recall from (2.4.5) that

$$\frac{d}{dt}F(t,H,i) = \rho^i F(t,H,i) - q_i F(t,H,i) - \sum_{\substack{j=1,\dots,N \\ j\neq i}} F(t,H,j) q_{ij}$$

and hence

$$\begin{aligned}\frac{d}{dt}\log F(t,H,i) &= \frac{1}{F(t,H,i)}\frac{d}{dt}F(t,H,i) \\ &= \rho^i - q_i - \sum_{\substack{j=1,\dots,N \\ j\neq i}} \frac{F(t,H,j)}{F(t,H,i)} q_{ij}.\end{aligned}$$

Since $\boldsymbol{X}_t = \boldsymbol{e}_i$ is fixed for all $t \in [\tau_{N(t)}, t)$, we have

$$\begin{aligned}c(t,\boldsymbol{X}_t) =& c(\tau_{N(t)}, \boldsymbol{X}_{N(t)}) + \int_{\tau_{N(t)}}^t dc(s, \boldsymbol{X}_s) \\ =& c(\tau_{N(t)}, \boldsymbol{e}_i) + \int_{\tau_{N(t)}}^t dc(s, \boldsymbol{e}_i) \\ =& c(\tau_{N(t)}, \boldsymbol{X}_{N(t)}) + \int_{\tau_{N(t)}}^t \dot{c}(s,i) ds\end{aligned}$$

where

$$\dot{c}(t,i) = \rho^i - q_i - \sum_{\substack{j=1,\dots,N \\ j\neq i}} \frac{F(t,H,j)}{F(t,H,i)} q_{ij},$$

due to the fact that $c(s, \boldsymbol{X}_t) = c(s,i)$ is a deterministic function of $t$ in between jump times. Therefore if we consider a generic value $\boldsymbol{X}_t$ for $t \in (\tau_{N(t^-)}, \tau_{N(t^-)+1})$, and define $\dot{\boldsymbol{C}}(s) = (\dot{c}(s,1), \dots, \dot{c}(s,N))^\top$, then we have

$$c(t,\boldsymbol{X}_t) = c(\tau_{N(t^-)}, \boldsymbol{X}_{N(t^-)}) + \int_{\tau_{N(t^-)}}^t \langle \dot{\boldsymbol{C}}(s), \boldsymbol{X}_s \rangle ds$$

**Remark 2.4.10** (Using the Prototype Product approach for Path-dependent Claims)**.** *For a path dependent claim, we have seen that Proposition 2.4.3 does not hold and so we do not have the equation (2.4.26). However* $\boldsymbol{F}(t,\boldsymbol{H}_T)$ *and hence* $\log\{\langle \boldsymbol{F}(t,\boldsymbol{H}_T), \boldsymbol{X}_t \rangle\}$ *can still be computed using the techniques in [Pre10] and [MPR13] that are summarized in Section 2.4.1.2. While cumbersome, the subsequent computations can be modified accordingly, replacing the computations involving* $\boldsymbol{F}(t,\boldsymbol{H}_T)$ *with a Monte-Carlo average; in particular, the computation of* $\dot{\boldsymbol{C}}(t) = \frac{d}{dt}\boldsymbol{F}(t,\boldsymbol{H}_T)$ *must be performed approximately. In such a way, the techniques in this chapter can be extended to path dependent claims up to approximations. We will, however, not show this in the sequel.*

### 2.4.3 The Calibration Algorithm

Here we synthesize the results of Sections 2.2 - 2.4 into an algorithm for estimating the parameters of (2.4.19).

We specify the of calibration parameters $\boldsymbol{\theta}$ of the model to include the elements of the Q-matrix, $(q_{i,j})_{1\leq i,j\leq N}$and the the parameters, previously denoted $h^i$, that are those which enter in the function $c(t,\xi_t)$. Each of the $h^i$'s are determined by the corresponding value of $\rho^i$ of the

interest rate $r(\xi_t)$. That is, $\boldsymbol{\theta} = \left\{\rho^i, q_{i,j} \mid 1 \leq i, j \leq N\right\}$. We may then assume that either a subset of $\boldsymbol{\theta}$ is already known or that a good guess is known. For example, in the numerical results presented in 2.5, we assume that the true values for $\rho^i$ are known $\forall i = 1, \ldots, N$. We have also assumed that the scaling coefficient of the noise, that is, the coefficient $\eta$, is already known. For example, it can be taken as the average empirical variance of the time series $y(t)$ taken over inter-jump intervals.

We fix $\boldsymbol{Q}^0$, the initial guess for the matrix $\boldsymbol{Q}$ by taking $q^0_{i,j}$ to be the empirical transition counts for transitions from state $i$ to $j$, divided by the average occupation time in state $i$.

The parameter estimation algorithm is the following:

1. Make an initial guess $\boldsymbol{\theta}_0$ of the unknown parameters ($\boldsymbol{Q}^0$ as defined above).
2. Observe $y_t$ over a fixed time discretization $s_1 < s_2 < \ldots < s_m = T$of $[0, T]$ as well as over the jump times $\tau_1, \tau_2, ...$
3. Form the jump adapted time discretization as follows: Take the fixed discretization $s_1 < s_2 < \ldots < s_m = T$ and join the jump times $\tau_1, \tau_2, ..., \tau_\nu < T$ to form the jump-adapted time points $t_1 < t_2 < \ldots < t_{n(T)} < T$.
4. Iterate the following steps:
   (a) Observe $y^1_t = y_t$ for all $t$ before the first jump time.

   **Serial Filtering with Bayesian Updating at the Jumps**

   i. On the basis of the process $y^1(t)$ and the initial parameters $\boldsymbol{\theta}_0$, also initialize $\boldsymbol{\sigma}(\boldsymbol{X}_t)$, $\boldsymbol{\sigma}(N^{i,j}_t)$, $\boldsymbol{\sigma}(\mathcal{T}_t(i))$ and $\boldsymbol{\sigma}(\Gamma_t(i))$ using the *quasi-exact* approximation scheme on the *jump-adapted* time discretization, see equations (2.3.19) and (2.3.26).

   ii. At a jump-time $\tau_k$, apply Bayesian updating to re-initialize the filters starting from $\tau_k$ and compute $\boldsymbol{\sigma}(\boldsymbol{X}_t)$, $\boldsymbol{\sigma}(N^{i,j}_t)$, $\boldsymbol{\sigma}(\mathcal{T}_t(i))$ and $\boldsymbol{\sigma}(\Gamma_t(i))$ until the next jump-time $\tau^-_{k+1}$.

   iii. Iterate between re-initializing and filtering over the inter-jump intervals $(\tau_k, \tau_{k+1}]$ until the terminal time $T$ is reached.

   **Expectation Step** Compute $\hat{\boldsymbol{\sigma}}(\boldsymbol{X}_t)$ using the Quasi - Exact Approximation (2.3.19) and then also compute $\hat{\boldsymbol{\sigma}}(N^{i,j}_t)$, $\hat{\boldsymbol{\sigma}}(\mathcal{T}_t(i))$ and $\hat{\boldsymbol{\sigma}}(\Gamma_t(i))$ by the corresponding Quasi - Exact Approximation (2.3.26).

   **Maximization Step** Obtain $\tilde{\boldsymbol{\theta}}$ for the next iteration of the EM (see (2.2.24) and (2.2.25)).

   (b) Repeat the step (a) with $\boldsymbol{\theta}_0$ replaced by $\hat{\boldsymbol{\theta}}$ successively from each E - M step, until a stopping criterion is reached, for example $|\hat{\boldsymbol{\theta}}_k - \hat{\boldsymbol{\theta}}_{k-1}| < \epsilon$ for some threshold $\epsilon$.

## 2.5 Numerical Results

In this section we present the numerical results from implementing our proposed calibration algorithm for a simple test case, where the time series for $y_t$ is generated by simulating the

CTMC $\xi_t$, instead of taking market data. We chose as our CTMC model the same used by [PR10a] in their numerical tests, that is, the CTMC has $\boldsymbol{Q}$-matrix

$$\boldsymbol{Q} = \begin{bmatrix} -1 & 1 & 0 \\ 0.5 & -1 & 0.5 \\ 0 & 1 & -1 \end{bmatrix},$$

the diffusion coefficient of the noise term $W$ of the observations process $y$ is $\eta = 1.0$ and the diffusion matrix of the Zakai SDE for the Wonham filter of the state is $\boldsymbol{D}$ is

$$\boldsymbol{D} = \begin{bmatrix} 0.05 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -0.05 \end{bmatrix}. \tag{2.5.1}$$

We note that in this set-up, the drift and diffusion matrix coefficients of the Zakai SDE are not-commuting, and hence this is not a trivial set-up, i.e. we are able to test the performance of the quasi-exact scheme in a way that takes into account errors arising from the use of the quasi-exact solution as *proxy*. **Financial Interpretation** In [PR10a] this set-up served as a simple numerical test to show that the quasi-exact scheme works when approximating the Zakai SDE for the state-filter. Here we also attempt to give these choice of parameters a financial interpretation. First we observe that they arise from taking the price $F(t, H, \xi_t)$ defined as

$$F(t, H, \xi_t) = \begin{cases} e^{0.05(T-t)} & \text{if } \xi_t = x^1 \\ 1 & \text{if } \xi_t = x^2 \\ e^{-0.05(T-t)} & \text{if } \xi_t = x^3. \end{cases} \tag{2.5.2}$$

Indeed in that case, $c(t, \xi_t) = \log\{F(t, H, \xi_t)\}$ is

$$c(t, \xi_t) = \begin{cases} 0.05(T-t) & \text{if } \xi_t = x^1 \\ 0 & \text{if } \xi_t = x^2 \\ -0.05(T-t) & \text{if } \xi_t = x^3 \end{cases},$$

$h(t, \xi_t) = \dot{c}(t, \xi_t)$ equals

$$h(t, \xi_t) = \begin{cases} 0.05 & \text{if } \xi_t = x^1 \\ 0 & \text{if } \xi_t = x^2 \\ -0.05 & \text{if } \xi_t = x^3 \end{cases},$$

and finally $\boldsymbol{D}(t) = \text{diag}\{\boldsymbol{h}\}$, which is exactly (2.5.1).

**Remark 2.5.1.** *A price $F(t, H, \xi_t)$ that is given as (2.5.2) arises as follows: Let $V(t, H, \xi_t)$ be the price of a zero-coupon bond whose value depends on $\xi_t$, as in the Example 2.4.2.4. Then $F(t, H, \xi_t)$ is the value at time $t$ of a dynamically re-balanced portfolio consisting of $e^{0.5(T-t)}/V(t, H, \xi_t)$ units of the bond if $\xi_t = x^1$, $e^{-0.5(T-t)}/V(t, H, \xi_t)$ units of the bond if $\xi_t = x^3$ and $1/V(t, H, \xi_t)$ units of the bond if $\xi_t = x^2$. A possible interpretation of this is that it replicates a double-touch digital caplet on the short rate $r(\xi_t)$. This simple example can serve as a basis for applications with other types of payoffs. Indeed, under a general Markov functional model (see the remarks of Section 4.2), the calibration of digital caplets is the basis for completely determining all the price functionals of market-traded derivatives.*

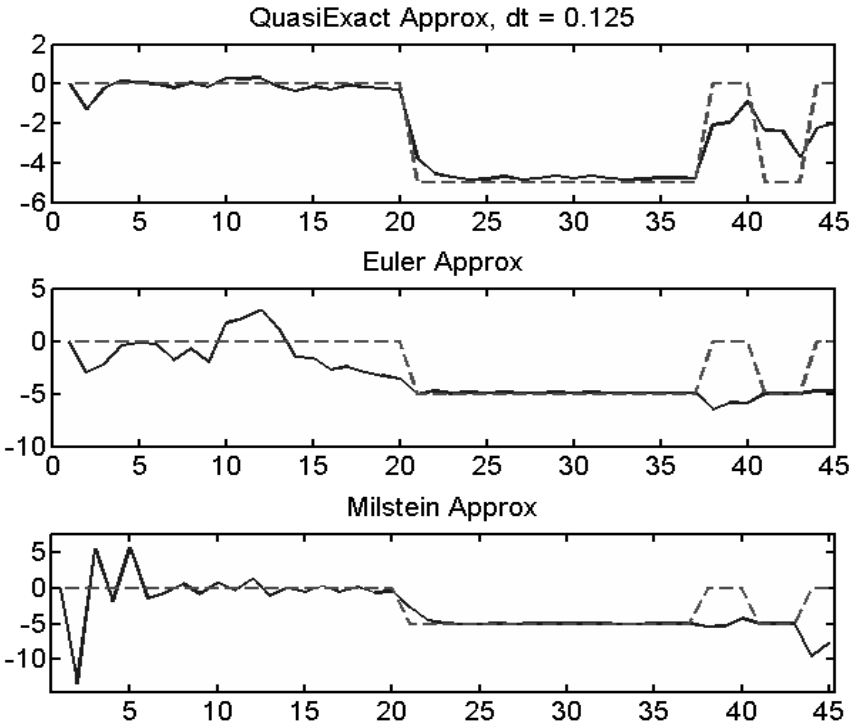


**Figure 2.1:** A sample path of $\hat{\xi}_k^{\Delta}$, $\Delta = 2^{-3}$. The dashed line is the true path.

Returning to the numerical test itself, we first computed the discrete time approximation $\hat{\xi}_k^{\Delta}$ of the Wonham filter for the state, $\hat{\xi}_t$, using the Euler-Maruyama, Milstein, and Quasi-Exact Schemes, and a jump-adapted time grid, where the fixed time points had time step sizes of $\Delta = 2^{-3}, 2^{-4}, \ldots, 2^{-7}$. As a first experiment, we use the true parameters in the computation of the filters. We used an ultra-fine discretization of $2^{-10}$ to obtain a *proxy* for the true solution $\hat{\xi}_t$. Figures 2.1 and 2.2 show graphs of the sample paths obtained using each scheme, at a discretization level of $2^{-3}$ and $2^{-4}$, respectively. The implementation was performed using textitMatlab R2012b for OS X 10.8.3 on a Macbook Pro with a 2.6 GHZ Intel Core i7 processor and 16 GB of memory.

At each discretization level we then computed the global error at the final time point $t_n(T)$. We also computed the (path wise) $L^2$-error, that is

$$\sqrt{\sum_k \left(\hat{\xi}_k^{\Delta} - \hat{\xi}_{t_k}\right)^2}.$$

In order to compare the quality of each scheme, we plotted the logarithm (base 2) of the global error versus the logarithm (base 2) of the discretization size $\Delta$. Figure 2.3 shows the $\log_2$-$\log_2$ plots for the three schemes. We can see that, while an error reduction is achieved in all three schemes by decreasing the step-size $\Delta$, the Quasi-Exact Scheme has a relatively quite low error at the coarsest discretization level. We see a similar comparison in the $\log_2$-$\log_2$ plot of the $L^2$ error versus $\Delta$ (Figure 2.4).

We then implemented the Wonham filters for the occupation time $\widehat{\mathcal{T}}_t(i)$ and the counting process for transitions $\widehat{\mathcal{N}}_t^{i,j}$. As before, we varied the step-size $\Delta$ and compared the global error

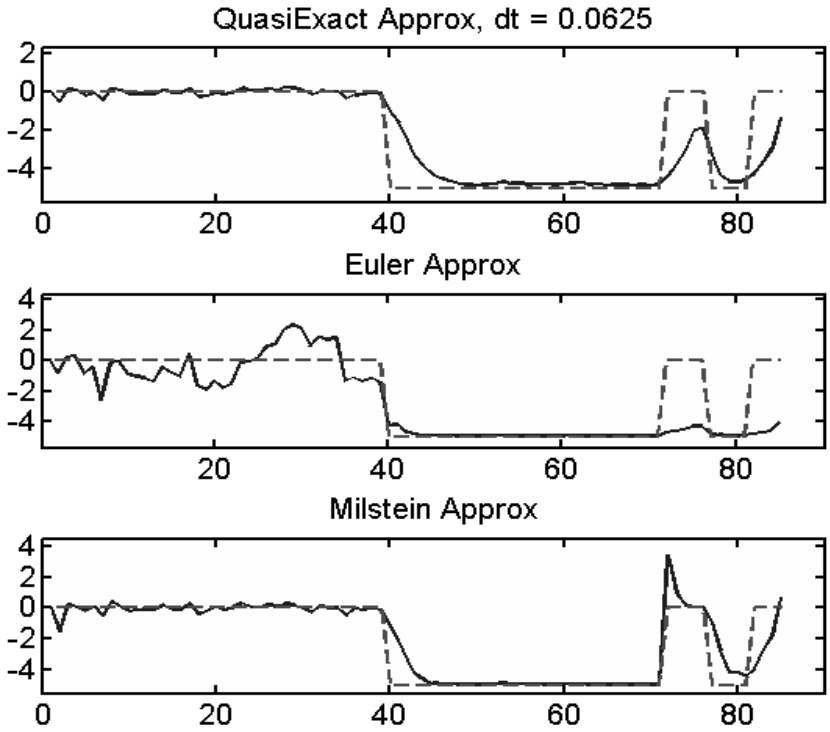


**Figure 2.2:** A sample path of $\hat{\xi}_k^{\Delta}$, $\Delta = 2^{-4}$. The dashed line is the true path.

and the $L^2$-error. For brevity, we show here only the $\log_2$-$\log_2$ plots of the global error (Figure 2.6) and the $L^2$ error (Figure 2.7) for the filter $\widehat{\mathcal{T}}_t(2)$ (Figure 2.5). In this case we see again that in the coarsest time-discretization, the Quasi-Exact Scheme attains the lowest error, while for a fine time-discretization it has a similar accuracy to the Milstein Scheme. The Euler Scheme performs relatively poorly for this filter, due to the more complicated Zakai SDEs. Similar results were observed for the filter $\widehat{N}_t^{i,j}$.

Finally we also implemented the filter-based EM algorithm, when the true parameters in the $\boldsymbol{Q}$ matrix of $\xi$ are

$$\boldsymbol{Q} = \begin{bmatrix} -1 & 1 & 0 \\ 0.4 & -1 & 0.6 \\ 0 & 1 & -1 \end{bmatrix},$$

and we estimate only $q_{2,3}$, assuming that $q_i = 1$ for $i = 1, 2, 3$. We take as our initial guess the matrix $\boldsymbol{Q}_0$

$$\boldsymbol{Q}_0 = \begin{bmatrix} -1 & 1 & 0 \\ 0.5 & -1 & 0.5 \\ 0 & 1 & -1 \end{bmatrix}.$$

The estimate for $q_{2,3}$ is computed by taking $\widehat{N}_T^{2,3}/\widehat{\mathcal{T}}_T(2)$ (see sub-section 2.2.3) at each iteration of the EM algorithm. We compare the results for $\widehat{N}_T^{2,3}/\widehat{\mathcal{T}}_T(2)$ computed using each of the three discretization schemes (Quasi-Exact, Euler-Maruyama and Milstein).

We noted, over several different simulated time series for $\xi_t$ and observations $y_t$, that the

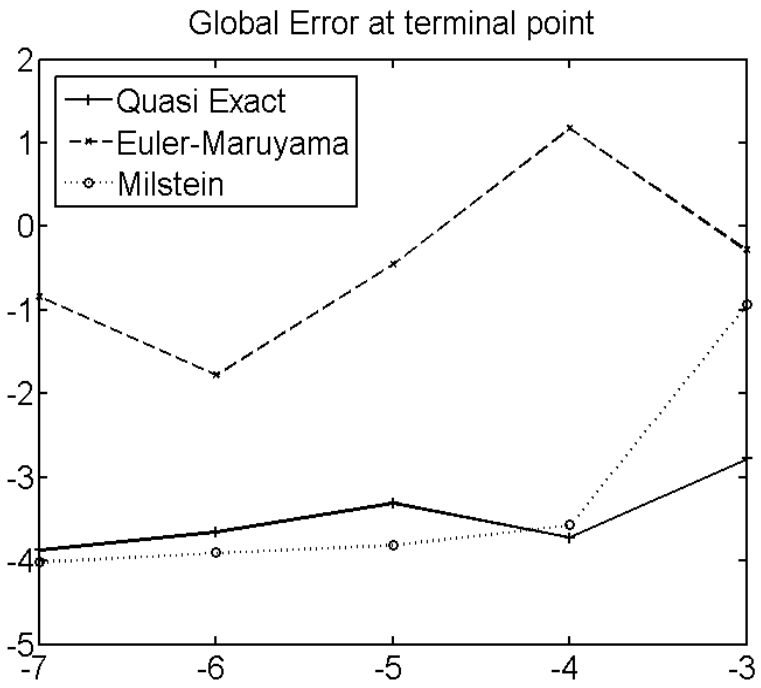


**Figure 2.3:** $\log_2$-$\log_2$ plot of the global error of $\hat{\xi}^{\Delta}$ at the final time point $T$, versus step-size $\Delta$. The solid, dashed, and dotted lines correspond to the Quasi-Exact, Euler-Maruyama and Milstein scheme, respectively. At the coarsest discretization level, $\Delta = 2^{-3}$, the respective global errors are $0.145$, $0.823$, and $0.523$.

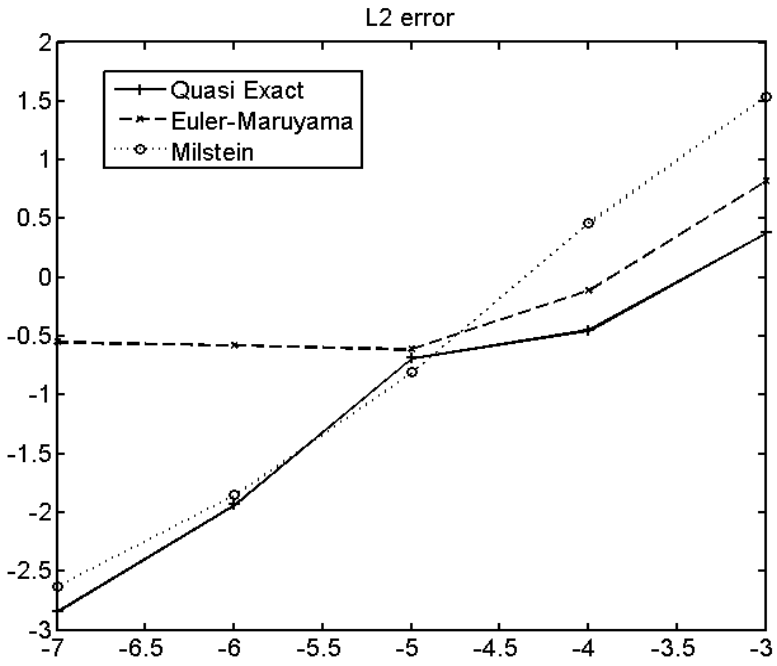


**Figure 2.4:** $\log_2$-$\log_2$ plot of the $L^2$-error of $\hat{\xi}^{\Delta}$, versus step-size $\Delta$. The solid, dashed, and dotted lines correspond to the Quasi-Exact, Euler-Maruyama and Milstein scheme, respectively. At the coarsest discretization level, $\Delta = 2^{-3}$, the respective $L^2$-errors are $1.344$, $1.859$, and $1.586$.

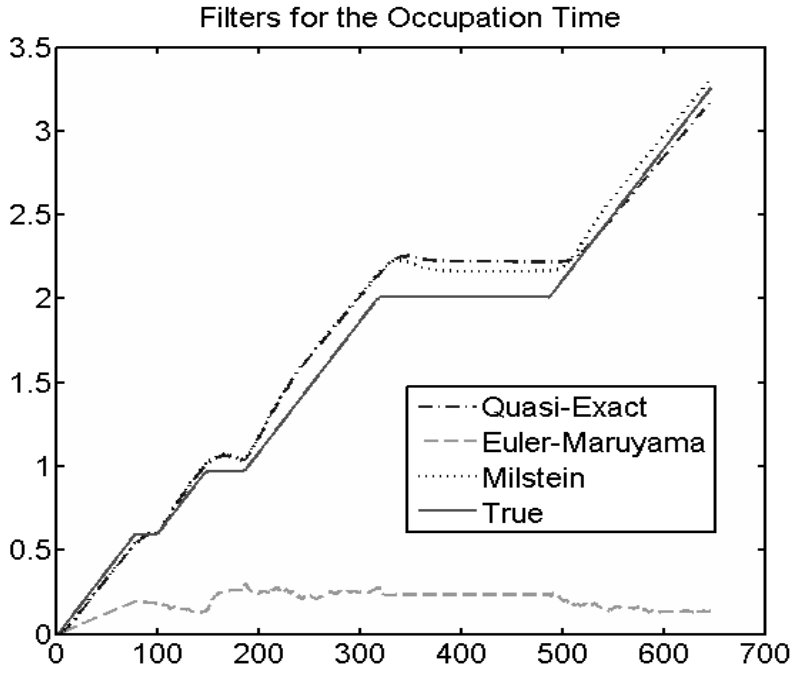


**Figure 2.5:** Sample path of $\widehat{\mathcal{T}}(2)_k^{\Delta}$ by the Quasi-Exact, Euler-Maruyama and Milstein Scheme at $\Delta = 2^{-10}$.

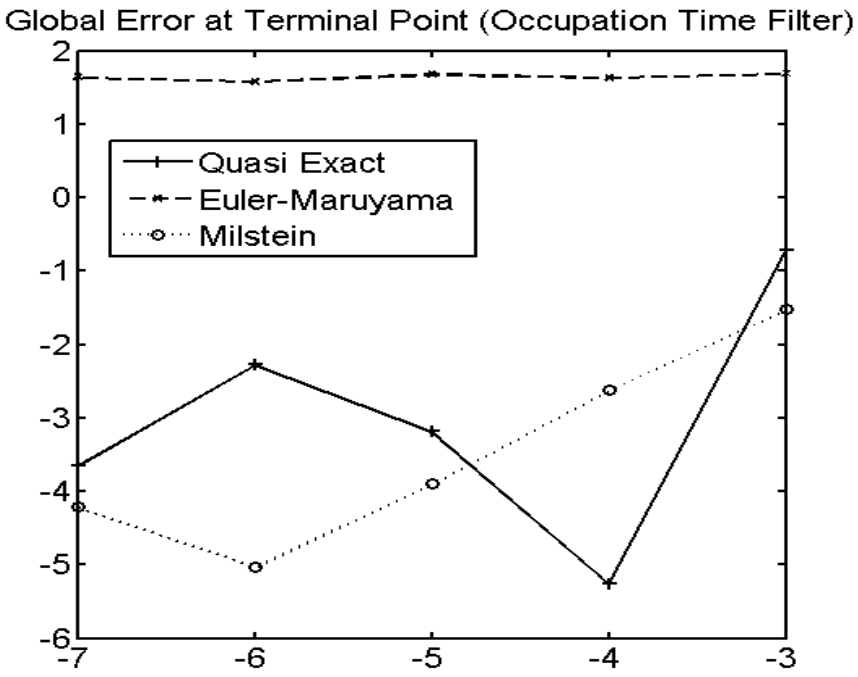


**Figure 2.6:** $\log_2$-$\log_2$ plot of the global error of $\widehat{\mathcal{T}}(2)^{\Delta}$ at the final time point $T$ versus step-size $\Delta$. The solid, dashed, and dotted lines correspond to the Quasi-Exact, Euler-Maruyama and Milstein scheme, respectively. At the coarsest discretization level, $\Delta = 2^{-3}$, the respective global errors are $0.6084$, $3.2337$, and $0.3473$.

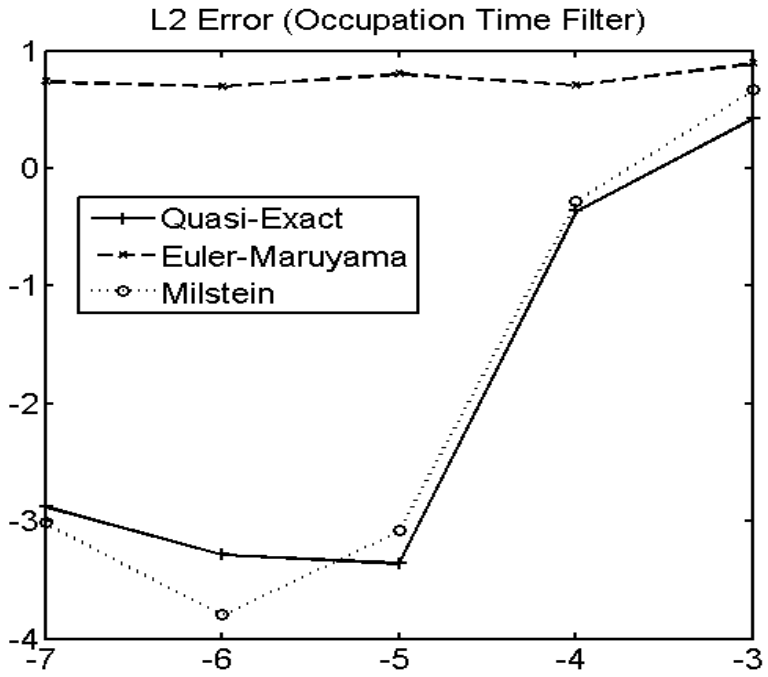


**Figure 2.7:** $\log_2$-$\log_2$ plot of the $L^2$-error of $\widehat{\mathcal{T}}(2)^{\Delta}$ versus step-size $\Delta$. The solid, dashed, and dotted lines correspond to the Quasi-Exact, Euler-Maruyama and Milstein scheme, respectively. At the coarsest discretization level, $\Delta = 2^{-3}$, the respective $L^2$-errors are $1.30$, $1.76$, and $2.91$.

| scheme | series 1 | error | series 2 | error |
|---|---|---|---|---|
| Euler-Maruyama | 0.6133 | −0.0133 | 0.8821 | −0.2821 |
| Milstein | 0.4486 | 0.1514 | 0.4551 | 0.1449 |
| Quasi-Exact | 0.4841 | 0.1159 | 0.5796 | 0.0204 |

**Table 2.1:** EM Parameter Estimates for the true model parameter $q_{2,3} = 0.6$, using an initial guess $q^0_{2,3} = 0.5$

results of parameter estimation are quite variable from one time series to another and, while in general the algorithm yields parameter estimates approaching the true parameters, there are several instances when the estimate exceeds the true value, or oscillates near the true value, i.e., the convergence is not always observed and is not always monotonic.

In Table 2.1 we show the results of one of the simulation studies we performed. It lists some of the estimated values for $q_{2,3}$ computed using each of the three discretization schemes. The series' correspond to each EM step (we show only 2 EM iterations in the table). The fixed step-size $\Delta = 2^{-7}$ was used, and each time series was simulated from $t = 0$ to $T = 10$ (1280 points in the fixed time-discretization).

#### 2.5.0.1 Final comments on the numerical results

From the parameter estimates attained in Table 2.1 (which reflect the also nature of the results over various simulations performed), these results perhaps do not definitively confirm whether the EM algorithm using our proposed discretization scheme is a good choice. The Euler - Maruyama estimate is the most accurate in Series 1, while the Quasi-Exact estimate is the most accurate in Series 2. This may be explained by the presence of oscillations when using an Euler-Maruyama scheme, which results in certain fluctuating estimates over each iterations. Perhaps more puzzling is why the Milstein scheme underperforms both schemes. A more thorough numerical test would need to also investigate the issue of stability of this algorithm. It is well known that the EM-based parameter estimation may lead to convergence of the objective function to local maxima, and therefore several simulation studies using several different initial values is necessary to fully ascertain by experiments, the effectiveness of this approach. Unfortunately due to constraints in time and resources, we are unable to do such a thorough study in the present work and must leave it to a possible future investigation. Nevertheless, we have shown that the core discretization scheme performs relatively as expected.

## 2.A Technical Proofs

### 2.A.1 Derivation of the Wonham Filter Equations

In order to derive the filter equations for the state $\boldsymbol{X}_t$, the transition counts $N_t^{i,j}$, the occupation time $\mathcal{T}_t(i)$ and the drift estimator $\Gamma_t(i)$ given respectively in Proposition 2.2.4, Theorem 2.2.5, Theorem 2.2.6 and Theorem 2.2.7, we must first derive a general equation for the filter of a generic measurable function of $\boldsymbol{X}_t$, and then particularize such an equation to obtain each of the specific filter equations.

Let us then denote by $H_t := f(\boldsymbol{X}_t)$ the process defined by $f$ that is a generic, measurable function of $\boldsymbol{X}_t$. From the differentiation rule (2.2.3) we may assume that $H_t$ takes the general semi-martingale form:

$$H_t := H_0 + \int_0^t \alpha_s ds + \int_0^t \beta_s^\top dM_t + \int_0^t \delta_s dW_t,$$

$\alpha_t, \delta_t \in \mathbb{R}$ and $\beta_t \in \mathbb{R}^N$.

Let us observe that

$$H_t = \langle H_t \boldsymbol{X}_t, \underline{\mathbf{1}} \rangle,$$

hence

$$\sigma(H_t) = \sigma(\langle H_t \boldsymbol{X}_t, \underline{\mathbf{1}} \rangle) = \langle \sigma(H_t \boldsymbol{X}_t), \underline{\mathbf{1}} \rangle,$$

so that, in order to compute $\sigma(H_t)$ it suffices to know $\sigma(H_t \boldsymbol{X}_t)$.

We shall now derive equations for [7] $\sigma(H_t \boldsymbol{X}_t)$. For the following formulas we define $\boldsymbol{h}(t)$ to be the vector such that

$$h(t, \boldsymbol{X}_t) = \langle \boldsymbol{h}(t), \boldsymbol{X}_t \rangle,$$

and we shall also define

$$\boldsymbol{D}(t) := \operatorname{diag}\left\{\eta^{-1} \cdot \boldsymbol{h}(t)\right\}, \tag{2.A.1}$$

the diagonal matrix having in its main diagonal the elements of the vector $\eta^{-1} \cdot \boldsymbol{h}(t)$. For convenience of notation, we further define

$$\bar{\boldsymbol{D}}(t) := \operatorname{diag}\left\{\boldsymbol{h}(t)\right\}. \tag{2.A.2}$$

We now have the following

**Theorem 2.A.1.** *Let $H_t := H_0 + \int_0^t \alpha_s ds + \int_0^t \beta_s^\top dM_t + \int_0^t \delta_s dW_t$ for a $P$-Wiener process $W$. Then the equation for the unnormalized estimate $\sigma(H_t \boldsymbol{X}_t)$ is given by the following linear SDE:*

$$\begin{cases} d\sigma(H_t \boldsymbol{X}_t) & = \left\{\sigma(\alpha_t \boldsymbol{X}_t) + \boldsymbol{Q}\sigma(H_t \boldsymbol{X}_t)\right\} dt \\ & + \sum_{i,j=1}^N \langle \sigma(\beta_t^j \boldsymbol{X}_t - \beta_t^i \boldsymbol{X}_t), \boldsymbol{e}_i \rangle q_{i,j} (\boldsymbol{e}_j - \boldsymbol{e}_i) dt \\ & + \left\{\sigma(\delta_t \boldsymbol{X}_t) + \bar{\boldsymbol{D}}(t)\sigma(H_t \boldsymbol{X}_t)\right\} d\bar{y}_t \end{cases} \tag{2.A.3}$$

*where $\bar{y}_t = \eta^{-1} y_t$.*

Note that, by the definition of $\bar{y}$ together with equations (2.A.1) and (2.A.2), we have

$$\bar{\boldsymbol{D}}(t)\bar{y}_t = \eta^{-1}\bar{\boldsymbol{D}}(t) y_t = \boldsymbol{D}(t) y_t. \tag{2.A.4}$$

**Remark 2.A.2.** *This theorem is analogous to Theorem 3.2 in Chapter 8 of [EAM08], the difference being that in [EAM08], the coefficient matrix $\boldsymbol{D}(s) \equiv \boldsymbol{D}$ is constant, whereas the present Theorem 2.A.1 considers a time varying $\boldsymbol{D}(s)$. This is due to the fact that we are now considering a time varying drift coefficient $h(s, \boldsymbol{X}_s)$ in the observations process. While the proof is completely analogous, we provide it here in detail for completeness of our discussion.*

[7] We may also directly derive equations in $\sigma(H_t)$, however, as explained in Remark 2 of [Ell93], this results in filter equations that are quite cumbersome to deal with, whereas equations in $\sigma(H_t \boldsymbol{X}_t)$ turn out to be more convenient.

*Proof.* Analogous to the proof of Theorem 3.2 from Chapter 8 of [EAM08], our approach is to first obtain the differential equation for $\Lambda_t H_t \boldsymbol{X}_t$, and then apply the $\bar{E}[\cdot \mid \mathcal{F}_t^y]$-conditional expectation.

From the Ito product rule we have

$$\begin{aligned} H_t\boldsymbol{X}_t = H_0\boldsymbol{X}_0 &+ \int_0^t \alpha_s \boldsymbol{X}_{s-} ds + \int_0^t \beta_s \boldsymbol{X}_{s-} dM_s + \int_0^t \delta_s \boldsymbol{X}_{s-} dW_s \\ &+ \int_0^t H_{s-}\boldsymbol{Q}\boldsymbol{X}_{s-} ds + \int_0^t H_{s-} dM_s + \sum_{0<k\leq N(t)} (\beta_k^\top \Delta \boldsymbol{X}_k)\Delta \boldsymbol{X}_k. \end{aligned} \tag{2.A.5}$$

where $\beta_k = \beta_{\tau_k}$ as before. Let us observe that if $\xi_t$ jumps from state $i$ to state $j$ at time $t$, then $\Delta\xi_t = \xi_t - \xi_{t^-} = j - i$ and

$$\Delta \boldsymbol{X}_t = \boldsymbol{X}_t - \boldsymbol{X}_{t^-} = (\boldsymbol{e}_j - \boldsymbol{e}_i). \tag{2.A.6}$$

Note further that

$$\mathbf{1}_{\{\xi_t=j\}}\mathbf{1}_{\{\xi_{t^-}=i\}} = \langle \boldsymbol{X}_t, \boldsymbol{e}_j\rangle\langle \boldsymbol{X}_{t^-}, \boldsymbol{e}_i\rangle,$$

by definition. Hence

$$\begin{aligned} \Delta \boldsymbol{X}_t &= \sum_{i,j=1}^N (\boldsymbol{e}_j - \boldsymbol{e}_i)\mathbf{1}_{\{\xi_t=j\}}\mathbf{1}_{\{\xi_{t^-}=i\}} \\ &= \sum_{i,j=1}^N (\boldsymbol{e}_j - \boldsymbol{e}_i)\langle \boldsymbol{X}_t, \boldsymbol{e}_j\rangle\langle \boldsymbol{X}_{t^-}, \boldsymbol{e}_i\rangle. \end{aligned}$$

Let us denote by $\beta_t^i$ the $i$-th component of the vector $\beta_t$. By a similar argument as before, one can easily show that

$$\begin{aligned} \beta_t^\top \Delta \boldsymbol{X}_t =& \beta_t^\top \sum_{i,j=1}^N (\boldsymbol{e}_j - \boldsymbol{e}_i)\langle \boldsymbol{X}_t, \boldsymbol{e}_j\rangle\langle \boldsymbol{X}_{t^-}, \boldsymbol{e}_i\rangle \\ &= \sum_{i,j=1}^N (\beta_t^\top \boldsymbol{e}_j - \beta_t^\top \boldsymbol{e}_i)\langle \boldsymbol{X}_t, \boldsymbol{e}_j\rangle\langle \boldsymbol{X}_{t^-}, \boldsymbol{e}_i\rangle \\ &= \sum_{i,j=1}^N (\beta_t^j - \beta_t^i)\langle \boldsymbol{X}_t, \boldsymbol{e}_j\rangle\langle \boldsymbol{X}_{t^-}, \boldsymbol{e}_i\rangle. \end{aligned}$$

We further have that

$$\begin{aligned} (\beta_t^\top \Delta \boldsymbol{X}_t)\Delta \boldsymbol{X}_t &= \left(\sum_{i,j=1}^N (\beta_t^j - \beta_t^i)\langle \boldsymbol{X}_t, \boldsymbol{e}_j\rangle\langle \boldsymbol{X}_{t^-}, \boldsymbol{e}_i\rangle\right) \cdot \sum_{i,j=1}^N (\boldsymbol{e}_j - \boldsymbol{e}_i)\langle \boldsymbol{X}_t, \boldsymbol{e}_j\rangle\langle \boldsymbol{X}_{t^-}, \boldsymbol{e}_i\rangle \\ &= \sum_{i,j=1}^N (\beta_t^j - \beta_t^i)\langle \boldsymbol{X}_t, \boldsymbol{e}_j\rangle\langle \boldsymbol{X}_{t^-}, \boldsymbol{e}_i\rangle(\boldsymbol{e}_j - \boldsymbol{e}_i) \\ &= \sum_{i,j=1}^N (\beta_t^j - \beta_t^i)\langle \boldsymbol{X}_t - \boldsymbol{X}_{t^-}, \boldsymbol{e}_j\rangle\langle \boldsymbol{X}_{t^-}, \boldsymbol{e}_i\rangle(\boldsymbol{e}_j - \boldsymbol{e}_i), \end{aligned}$$

where the third equation is due to the fact that when there is a transition from $\boldsymbol{X}_{t^-} = \boldsymbol{e}_i$ to $\boldsymbol{X}_t = \boldsymbol{e}_j$, then $\langle \boldsymbol{X}_{t^-}, \boldsymbol{e}_j\rangle = 0$. The last term in (2.A.5) can then be rewritten as

$$\sum_{0<k\leq N(t)} (\beta_k^\top \Delta \boldsymbol{X}_k)\Delta \boldsymbol{X}_k = \sum_{i,j=1}^N \int_0^t (\beta_s^j - \beta_s^i)\langle \boldsymbol{X}_{s-}, \boldsymbol{e}_i\rangle(\boldsymbol{e}_j - \boldsymbol{e}_i)\langle \boldsymbol{e}_j, d\boldsymbol{X}_s\rangle. \tag{2.A.7}$$

Substituting $d\boldsymbol{X}_s$ from the martingale representation

$$d\boldsymbol{X}_s = \boldsymbol{Q}\boldsymbol{X}_{s-}ds + dM_s$$

into equation (2.A.7) gives

$$\begin{aligned}\sum_{0<k\leq N(t)} (\beta_k^\top \Delta\boldsymbol{X}_k)\Delta\boldsymbol{X}_k &= \sum_{i,j=1}^N \int_0^t (\beta_s^j - \beta_s^i)\langle \boldsymbol{X}_{s-}, \boldsymbol{e}_i\rangle(\boldsymbol{e}_j - \boldsymbol{e}_i)\langle \boldsymbol{e}_j, dM_s\rangle \\ &\quad + \sum_{i,j=1}^N \int_0^t (\beta_s^j - \beta_s^i)\langle \boldsymbol{X}_{s-}, \boldsymbol{e}_i\rangle(\boldsymbol{e}_j - \boldsymbol{e}_i)\langle \boldsymbol{e}_j, \boldsymbol{Q}\boldsymbol{X}_{s-}ds\rangle \\ &= \sum_{i,j=1}^N \int_0^t (\beta_s^j - \beta_s^i)\langle \boldsymbol{X}_{s-}, \boldsymbol{e}_i\rangle(\boldsymbol{e}_j - \boldsymbol{e}_i)\langle \boldsymbol{e}_j, dM_s\rangle \\ &\quad + \sum_{i,j=1}^N \int_0^t (\beta_s^j - \beta_s^i)\langle \boldsymbol{X}_{s-}, \boldsymbol{e}_i\rangle q_{i,j}(\boldsymbol{e}_j - \boldsymbol{e}_i)ds. \end{aligned} \tag{2.A.8}$$

The second term in the right-hand side of equation (2.A.8) is obtained by observing that, in the event of a transition from $\boldsymbol{X}_{s-} = \boldsymbol{e}_i$ to $\boldsymbol{X}_s = \boldsymbol{e}_j$, the expression $\langle \boldsymbol{Q}\boldsymbol{X}_{s-}ds, \boldsymbol{e}_j\rangle$ equals

$$\langle \boldsymbol{Q}\boldsymbol{X}_{s-}ds, \boldsymbol{e}_j\rangle = q_{i,j}ds.$$

Continuing,

$$\begin{aligned} &= \sum_{i,j=1}^N \int_0^t (\beta_s^j - \beta_s^i)\langle \boldsymbol{X}_{s-}, \boldsymbol{e}_i\rangle(\boldsymbol{e}_j - \boldsymbol{e}_i)\langle \boldsymbol{e}_j, dM_s\rangle \\ &\quad + \sum_{i,j=1}^N \int_0^t \langle \beta_s^j \boldsymbol{X}_{s-} - \beta_s^i \boldsymbol{X}_{s-}, \boldsymbol{e}_i\rangle q_{i,j}(\boldsymbol{e}_j - \boldsymbol{e}_i)ds \end{aligned} \tag{2.A.9}$$

Substituting the expression from (2.A.9) into (2.A.5) we get

$$\begin{aligned} H_t\boldsymbol{X}_t &= H_0\boldsymbol{X}_0 + \int_0^t \alpha_s\boldsymbol{X}_{s-}ds + \int_0^t \beta_s^\top \boldsymbol{X}_{s-}dM_s + \int_0^t \delta_s\boldsymbol{X}_{s-}dW_s \\ &\quad + \int_0^t H_{s-}\boldsymbol{Q}\boldsymbol{X}_{s-}ds + \int_0^t H_{s-}dM_s \\ &\quad + \sum_{i,j=1}^N \int_0^t (\beta_s^j - \beta_s^i)\langle \boldsymbol{X}_{s-}, \boldsymbol{e}_i\rangle(\boldsymbol{e}_j - \boldsymbol{e}_i)\langle \boldsymbol{e}_j, dM_s\rangle \\ &\quad + \sum_{i,j=1}^N \int_0^t \langle \beta_s^j \boldsymbol{X}_{s-} - \beta_s^i \boldsymbol{X}_{s-}, \boldsymbol{e}_i\rangle q_{ij}(\boldsymbol{e}_j - \boldsymbol{e}_i)ds. \end{aligned} \tag{2.A.10}$$

The term

$$\sum_{i,j=1}^N \int_0^t (\beta_s^j - \beta_s^i)\langle \boldsymbol{X}_{s-}, \boldsymbol{e}_i\rangle(\boldsymbol{e}_j - \boldsymbol{e}_i)\langle \boldsymbol{e}_j, dM_s\rangle$$

can be written as

$$\sum_{i,j=1}^N \int_0^t (\beta_s^j - \beta_s^i)\langle \boldsymbol{X}_{s-}, \boldsymbol{e}_i\rangle \cdot (\boldsymbol{e}_j - \boldsymbol{e}_i) \cdot \boldsymbol{e}_j^\top dM_s.$$

Let us define $\bar{\alpha}_s \in \mathbb{R}$, $\bar{\beta}_s \in \mathbb{R}^N$, and $\bar{\delta}_s \in \mathbb{R}$ as follows

$$\begin{aligned}
\bar{\alpha}_s &=: \alpha_s \boldsymbol{X}_{s-} + H_{s-}\boldsymbol{Q}\boldsymbol{X}_{s-} + \sum_{i,j=1}^{N} \langle \beta_s^j \boldsymbol{X}_{s-} - \beta_s^i \boldsymbol{X}_{s-}, \boldsymbol{e}_i \rangle q_{ij}(\boldsymbol{e}_j - \boldsymbol{e}_i) \\
\bar{\beta}_s &=: \beta_s^\top \boldsymbol{X}_{s-} + H_{s-} + \sum_{i,j=1}^{N} (\beta_s^j - \beta_s^i)\langle \boldsymbol{X}_{s-}, \boldsymbol{e}_i \rangle \cdot (\boldsymbol{e}_j - \boldsymbol{e}_i) \cdot \boldsymbol{e}_j^\top \\
\bar{\delta}_s &=: \delta_s \boldsymbol{X}_{s-}.
\end{aligned}$$

Hence equation (2.A.10) is

$$H_t \boldsymbol{X}_t = H_0 \boldsymbol{X}_0 + \int_0^t \bar{\alpha}_s ds + \int_0^t (\bar{\beta}_s)^\top dM_s + \int_0^t \bar{\delta}_s dW_s \tag{2.A.11}$$

Once again applying the Ito product rule, and recalling that the differential $d\Lambda_t$ is given in (2.2.6) as

$$d\Lambda_t = \Lambda_t \cdot \eta^{-1} \cdot h(t, \boldsymbol{X}_t) dy_t, \quad \text{with } \Lambda_0 = 1$$

we obtain the following expression

$$\begin{aligned}
d(\Lambda_t H_t \boldsymbol{X}_t) =& \Lambda_t \bar{\alpha}_t dt + \Lambda_t (\bar{\beta}_t)^\top dM_t + \Lambda_t \bar{\delta}_t dW_t \\
&+ \Lambda_t \cdot \eta^{-1} \cdot h(t, \boldsymbol{X}_t) H_{t^-} \boldsymbol{X}_{t^-} dy_t + d\left[\Lambda, H\boldsymbol{X}\right]_t \\
=& \Lambda_t \bar{\alpha}_t dt + \Lambda_t (\bar{\beta}_t)^\top dM_t + \Lambda_t \bar{\delta}_t dW_t \\
&+ \Lambda_t \cdot \eta^{-1} \cdot h(t, \boldsymbol{X}_t) H_{t^-} \boldsymbol{X}_{t^-} dy_t + \text{“}\Lambda_t \cdot \eta^{-1} \cdot h(t, \boldsymbol{X}_t) \bar{\alpha}_t dt dy_t\text{”} \\
&+ \Lambda_t \cdot \eta^{-1} \cdot h(t, \boldsymbol{X}_t) \bar{\beta}_t^\top d\left[y, M\right]_t + \Lambda_t \cdot \eta^{-1} \cdot h(t, \boldsymbol{X}_t) \bar{\delta}_t d\left[y, W\right]_t .
\end{aligned} \tag{2.A.12}$$

The expression inside the quotation marks “· ” represents higher order terms which can be neglected due to the fact that in computing the quadratic variation $[\Lambda, H\boldsymbol{X}]_t$ we are performing a limiting operation[8]. Note also that

$$\left[y, W\right]_t = \left[\int_0^t h(s, \xi_s) ds + \eta W_t, W_t\right] = \eta \langle W \rangle_t = \eta \cdot t,$$

following standard arguments from the calculus of Ito processes. Hence $d[y, W]_s = \eta \cdot ds$.

We then write the integrated expansion of $\Lambda_t H_t \boldsymbol{X}_t$,

$$\Lambda_t H_t \boldsymbol{X}_t = H_0 \boldsymbol{X}_0 + \int_0^t d(\Lambda_s H_s \boldsymbol{X}_s),$$

[8]The quadratic variation $[\Lambda, H\boldsymbol{X}]_t$ is the limit in probability of the product of increments, i.e.

$$[\Lambda, H\boldsymbol{X}]_t = \underset{\Delta \downarrow 0}{\text{P-lim}} \sum_{k=0}^{n(t)} (\Lambda_{k+1}^\Delta - \Lambda_k^\Delta)((H\boldsymbol{X})_{k+1}^\Delta - (H\boldsymbol{X})_k^\Delta),$$

as the time discretization step $\Delta = \frac{t}{n(t)}$ is taken to be infinitesimally small, and where $\Lambda_k^\Delta$ and $(H\boldsymbol{X})_k^\Delta$ are families of sequences, having a time step size of $\Delta$, that are approximating sequences for $\Lambda_t$ and $H_t \boldsymbol{X}_t$, respectively. See for example [Pro05].

by integrating both sides of (2.A.12).

$$\begin{aligned}\Lambda_t H_t \boldsymbol{X}_t =& H_0\boldsymbol{X}_0 + \int_0^t \Lambda_s\bar{\alpha}_s ds + \int_0^t \Lambda_s(\bar{\beta}_s)^\top dM_s + \int_0^t \Lambda_s\bar{\delta}_s dW_s \\ &+ \int_0^t \Lambda_s\cdot\eta^{-1}\cdot h(s,\boldsymbol{X}_s)H_{s-}\boldsymbol{X}_{s-}dy_s \\ &+ \int_0^t \Lambda_s\cdot\eta^{-1}\cdot h(s,\boldsymbol{X}_s)\bar{\beta}_s^\top d\left[y,M\right]_s + \int_0^t \Lambda_s\cdot\eta^{-1}\cdot h(s,\boldsymbol{X}_s)\bar{\delta}_s\eta ds \\ =& H_0\boldsymbol{X}_0 + \int_0^t \Lambda_s\bar{\alpha}_s ds + \int_0^t \Lambda_s(\bar{\beta}_s)^\top dM_s + \int_0^t \Lambda_s\bar{\delta}_s dW_s \\ &+ \int_0^t \Lambda_s\cdot\eta^{-1}\cdot h(s,\boldsymbol{X}_s)H_{s-}\boldsymbol{X}_{s-}dy_s \\ &+ \int_0^t \Lambda_s\cdot\eta^{-1}\cdot h(s,\boldsymbol{X}_s)\bar{\beta}_s^\top d\left[y,M\right]_s + \int_0^t \Lambda_s h(s,\boldsymbol{X}_s)\bar{\delta}_s ds. \end{aligned} \tag{2.A.13}$$

Next we take the $\bar{E}\left[\cdot \mid \mathcal{F}_t^y\right]$-conditional expectation of both sides of (2.A.13), applying Fubini's theorem to interchange the order of expectations and integrals. This has the following consequences: first, the $\mathcal{F}_t^y$-conditional $\bar{P}$-expectation of the quadratic variation term in $d\left[y,M\right]_s$ vanishes because $y\cdot M$ is a martingale under $\bar{P}$ (see the proof of Proposition 1.1.10 in [Han07a]). Secondly, we recall that the $\mathcal{F}_t^y$ -optional projections under $\bar{P}$ of the $dy_s$ (as well as the $dW_s$) integrals are $dy_s$ (respectively $dW_s$) integrals of the optional projections, and thirdly we note that the $\bar{E}\left[\cdot \mid \mathcal{F}_t^y\right]$-conditional expectations of the $dM_s$ terms vanish (see [Han07b], Lemma 7.2.7 for both properties). Define the $\widehat{\cdot}$ notation as

$$\widehat{w} = \bar{E}\left[w \mid \mathcal{F}_t^y\right], \qquad w \text{ a random variable.}$$

We are left with

$$\widehat{\Lambda_t H_t \boldsymbol{X}_t} = H_0\boldsymbol{X}_0 + \int_0^t \widehat{\Lambda_s\bar{\alpha}_s} ds + \int_0^t \widehat{\Lambda_s\bar{\delta}_s dW_s} + \int_0^t \widehat{\Lambda_s\eta^{-1}h(s,\boldsymbol{X}_s)H_s\boldsymbol{X}_s} dy_s + \int_0^t \widehat{\Lambda_s h(s,\boldsymbol{X}_s)\bar{\delta}_s} ds. \tag{2.A.14}$$

From the definitions of $\bar{\alpha}_s$ and $\bar{\delta}_s$ and the equality

$$h(s,\boldsymbol{X}_s)\boldsymbol{e}_i = \langle \boldsymbol{h}(s),\boldsymbol{X}_s\rangle \boldsymbol{e}_i = h^i(s)\boldsymbol{e}_i$$

the right hand side of (2.A.14) equals

$$\begin{aligned}\widehat{\Lambda_t H_t \boldsymbol{X}_t} =& H_0\boldsymbol{X}_0 + \int_0^t \widehat{\Lambda_s\alpha_s\boldsymbol{X}_{s-}} ds + \int_0^t \widehat{\delta_s\Lambda_s\boldsymbol{X}_{s-}dW_s} \\ &+ \int_0^t \widehat{\Lambda_s H_{s-}\boldsymbol{Q}\boldsymbol{X}_{s-}} ds \\ &+ \sum_{i,j=1}^N \int_0^t \widehat{\langle \beta_s^j\Lambda_s\boldsymbol{X}_{s-} - \beta^i\Lambda_s\boldsymbol{X}_{s-}, \boldsymbol{e}_i\rangle q_{i,j}(\boldsymbol{e}_j - \boldsymbol{e}_i)} ds \\ &+ \sum_{i=1}^N \int_0^t \widehat{\langle \Lambda_s\boldsymbol{X}_s, \boldsymbol{e}_i\rangle\cdot\eta^{-1}\cdot h_i(s)H_s\boldsymbol{e}_i} dy_s + \sum_{i=1}^N \int_0^t \widehat{\langle \Lambda_s\boldsymbol{X}_{s-}, \boldsymbol{e}_i\rangle\delta_s h_i(s)\boldsymbol{e}_i} ds. \end{aligned} \tag{2.A.15}$$

We can re-write the terms

$$\begin{aligned}
&\int_0^t \delta_s \Lambda_{s-} \boldsymbol{X}_{s-} dW_s + \sum_{i=1}^N \int_0^t \langle \Lambda_s \boldsymbol{X}_{s-}, \boldsymbol{e}_i \rangle \delta_s h_i(s) \boldsymbol{e}_i ds \\
&= \int_0^t \Lambda_{s-} \delta_s \boldsymbol{X}_{s-} dW_s + \int_0^t \Lambda_s \delta_s \langle \boldsymbol{h}(s), \boldsymbol{X}_{s-} \rangle \boldsymbol{X}_{s-} ds \\
&= \int_0^t \Lambda_{s-} \delta_s \boldsymbol{X}_{s-} dW_s + \int_0^t \Lambda_s \delta_s \boldsymbol{X}_{s-} h(s, \boldsymbol{X}_{s-}) ds \\
&= \int_0^t \Lambda_s \delta_s \boldsymbol{X}_{s-} d\bar{y}_s.
\end{aligned}$$

and also

$$\sum_{i=1}^N \int_0^t \langle \Lambda_s \boldsymbol{X}_{s-}, \boldsymbol{e}_i \rangle \cdot \eta^{-1} \cdot h_i(s) H_{s-} \boldsymbol{e}_i dy_s = \int_0^t \Lambda_s \bar{\boldsymbol{D}}(s) H_{s-} \boldsymbol{X}_{s-} d\bar{y}_s$$

where we recall that we have defined $\bar{y}_t = \eta^{-1} y_t$.

Hence (2.A.15) simplifies to

$$\begin{aligned}
\widehat{\Lambda_t H_t \boldsymbol{X}_t} &= H_0 \boldsymbol{X}_0 + \int_0^t \widehat{\Lambda_s \alpha_s \boldsymbol{X}_{s-}} ds + \int_0^t \widehat{\Lambda_s H_{s-} \boldsymbol{Q} \boldsymbol{X}_{s-}} ds \\
&+ \sum_{i,j=1}^N \int_0^t \widehat{\langle \beta_s^j \Lambda_s \boldsymbol{X}_{s-} - \beta^i \Lambda_s \boldsymbol{X}_{s-}, \boldsymbol{e}_i \rangle q_{i,j} (\boldsymbol{e}_j - \boldsymbol{e}_i)} ds \\
&+ \int_0^t \bar{\boldsymbol{D}}(s) \widehat{\Lambda_s H_{s-} \boldsymbol{X}_{s-}} d\bar{y}_s + \int_0^t \widehat{\Lambda_s \delta_s \boldsymbol{X}_{s-}} d\bar{y}_s. \qquad \text{(2.A.16)}
\end{aligned}$$

Recalling the definition of $\sigma(f(\boldsymbol{X}_t))$ for any function $f(\boldsymbol{X}_t)$, we observe that

$$\sigma(f(\boldsymbol{X}_t)) = \bar{E}[\Lambda_t f(\boldsymbol{X}_t) \mid \mathcal{F}_t^y] = \widehat{\Lambda_t f(\boldsymbol{X}_t)}.$$

Applying this observation to (2.A.16) gives us the result. □

Now we particularize the filter equations to obtain the filter equations from Proposition 2.2.4, Theorem 2.2.5, Theorem 2.2.6 and Theorem 2.2.7.

### 2.A.1.1 Proof of Proposition 2.2.4

*Proof.* Taking $H_t = H_0 = 1$, $\alpha_s = 0$, $\beta_s = \underline{\mathbf{0}} \in \mathbb{R}^N$, and $\delta_s = 0$, and applying Theorem (2.A.1), we obtain the result. □

### 2.A.1.2 Proof of Theorem 2.2.5

*Proof.* We would like to obtain the Zakai equation for the unnormalized conditional expectation $\sigma(N_t^{i,j})$. First observe that $N_t^{i,j} = \langle N_t^{i,j} \boldsymbol{X}_t, \underline{\mathbf{1}} \rangle$, and hence

$$\sigma(N_t^{i,j}) = \sigma(\langle N_t^{i,j} \boldsymbol{X}_t, \underline{\mathbf{1}} \rangle) = \langle \sigma(N_t^{i,j} \boldsymbol{X}_t), \underline{\mathbf{1}} \rangle.$$

This allows us to apply Theorem 2.A.1 to $\sigma(N_t^{i,j}\boldsymbol{X}_t)$. Take $H_t = N_t^{i,j}$, $H_0 = 0$, $\alpha_t = \langle \boldsymbol{X}_{t^-}, \boldsymbol{e}_i\rangle q_{i,j}$, $\beta_t = \langle \boldsymbol{X}_{t^-}, \boldsymbol{e}_i\rangle \boldsymbol{e}_j$, and $\delta_s = 0$, to obtain

$$\begin{aligned}\sigma(N_t^{i,j}\boldsymbol{X}_t) = &\left\{\langle \sigma(\boldsymbol{X}_{t^-}), \boldsymbol{e}_i\rangle q_{i,j}\boldsymbol{X}_{t^-} + \boldsymbol{Q}\sigma(N_t^{i,j}\boldsymbol{X}_t)\right\} dt \\ &+ \sum_{k,l=1}^{N} \langle \sigma(\beta_t^l \boldsymbol{X}_t - \beta_t^k \boldsymbol{X}_t), \boldsymbol{e}_k\rangle q_{k,l}(\boldsymbol{e}_l - \boldsymbol{e}_k)dt \\ &+ \bar{\boldsymbol{D}}(t)\sigma(N_t^{i,j}\boldsymbol{X}_t)d\bar{y}_t.\end{aligned}$$

Since $\langle \boldsymbol{X}_{t^-}, \boldsymbol{e}_i\rangle \boldsymbol{X}_{t^-} = \langle \boldsymbol{X}_{t^-}, \boldsymbol{e}_i\rangle \boldsymbol{e}_i$ we find that

$$\begin{aligned}\sum_{k,l=1}^{N} \langle \beta_t^l \boldsymbol{X}_t - \beta_t^k \boldsymbol{X}_t, \boldsymbol{e}_k\rangle q_{k,l}(\boldsymbol{e}_l - \boldsymbol{e}_k) =&(\langle \boldsymbol{X}_{t^-}, \boldsymbol{e}_i\rangle \boldsymbol{e}_i^\top \boldsymbol{e}_i) q_{i,j}(\boldsymbol{e}_j - \boldsymbol{e}_i \\ =&\langle \boldsymbol{X}_{t^-}, \boldsymbol{e}_i\rangle q_{i,j}(\boldsymbol{e}_j - \boldsymbol{e}_i).\end{aligned}$$

Hence

$$\begin{aligned}\sigma(N_t^{i,j}\boldsymbol{X}_t) &= \left\{\langle \sigma(\boldsymbol{X}_{t^-}), \boldsymbol{e}_i\rangle q_{i,j}\boldsymbol{e}_i + \boldsymbol{Q}\sigma(N_t^{i,j}\boldsymbol{X}_t)\right\} dt + \langle \sigma(\boldsymbol{X}_{t^-}), \boldsymbol{e}_i\rangle q_{i,j}(\boldsymbol{e}_j - \boldsymbol{e}_i)dt \\ &\quad + \bar{\boldsymbol{D}}(t)\sigma(N_t^{i,j}\boldsymbol{X}_t)d\bar{y}_t \\ &= \left\{\langle \sigma(\boldsymbol{X}_{t^-}), \boldsymbol{e}_i\rangle q_{i,j}\boldsymbol{e}_j + \boldsymbol{Q}\sigma(N_t^{i,j}\boldsymbol{X}_t)\right\} dt + \bar{\boldsymbol{D}}(t)\sigma(N_t^{i,j}\boldsymbol{X}_t)d\bar{y}_t.\end{aligned}$$

□

### 2.A.1.3 Proof of Theorem 2.2.6

*Proof.* To prove Theorem 2.2.6, we can apply Theorem 2.A.1 with $H_t = \mathcal{T}_t(i)$, $H_0 = 0$, $\alpha_s = \langle \boldsymbol{X}_s, \boldsymbol{e}_i\rangle$, $\beta_s = \underline{\boldsymbol{0}} \in \mathbb{R}^N$, and $\delta_s = 0$. □

### 2.A.1.4 Proof of Theorem 2.2.7

*Proof.* Again, Theorem 2.2.7 is derived by applying Theorem 2.A.1, with $H_t = \Gamma_t(i)$, $H_0 = 0$, $\alpha_s = h_i(s)\langle \boldsymbol{X}_s, \boldsymbol{e}_i\rangle$, $\beta_s = \underline{\boldsymbol{0}}$, and $\delta_s = \langle \boldsymbol{X}_s, \boldsymbol{e}_i\rangle$. Note that $\boldsymbol{X}_s\alpha_s = h_i(s)\langle \boldsymbol{X}_s, \boldsymbol{e}_i\rangle$ and $\boldsymbol{X}_s\delta_s = \boldsymbol{X}_s\langle \boldsymbol{X}_s, \boldsymbol{e}_i\rangle = \langle \boldsymbol{X}_s, \boldsymbol{e}_i\rangle \boldsymbol{e}_i$. □

## 2.A.2 Proof of Proposition 2.3.1

*Proof.* We recall that for $\mathbb{R}^{k\times k}$ - matrix stochastic processes $\boldsymbol{X}$ and $\boldsymbol{Y}$, the Ito product formula,

$$d(\boldsymbol{XY}) = \boldsymbol{X}d\boldsymbol{Y} + d\boldsymbol{X}\boldsymbol{Y} + d\left[\boldsymbol{X}, \boldsymbol{Y}\right]_t,$$

is analogous to the scalar case with the obvious difference that the matrix product is not, in general, commutative.

We also recall two properties of the matrix exponential (see [PP12]): the first is that

$$\mathbf{exp}\left\{\boldsymbol{M}_1 + \boldsymbol{M}_2\right\} = \mathbf{exp}\left\{\boldsymbol{M}_1\right\}\mathbf{exp}\left\{\boldsymbol{M}_2\right\} = \mathbf{exp}\left\{\boldsymbol{M}_2\right\}\mathbf{exp}\left\{\boldsymbol{M}_1\right\}$$

if and only if $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ commute. The second is that if $\int_0^t \boldsymbol{M}(s)ds$ is deterministic then (see [PP12])

$$d\left(\mathbf{exp}\left\{\int_0^t \boldsymbol{M}(s)ds\right\}\right) = \mathbf{exp}\left\{\int_0^t \boldsymbol{M}(s)ds\right\}\boldsymbol{M}(t)dt = \boldsymbol{M}(t)\mathbf{exp}\left\{\int_0^t \boldsymbol{M}(s)ds\right\}dt. \tag{2.A.17}$$

In particular if $\boldsymbol{M}$ is a constant matrix and $t \in \mathbb{R}$, then

$$\frac{d}{dt}\mathbf{exp}\left\{\boldsymbol{M}\cdot t\right\} = \boldsymbol{M}\cdot\mathbf{exp}\left\{\boldsymbol{M}\cdot t\right\} = \mathbf{exp}\left\{\boldsymbol{M}\cdot t\right\}\cdot\boldsymbol{M}.$$

We then have

$$\begin{aligned}
d(\boldsymbol{\Phi}_t) &= d\left(\mathbf{exp}\left\{\int_0^t\left(\boldsymbol{A}(s)-\frac{1}{2}(\boldsymbol{D})^2(s)\right)ds+\int_0^t \boldsymbol{D}(s)dW_s\right\}\right)\\
&\qquad\text{(by commutativity of } A(t) \text{ and } D(t))\\
&= d\left(\mathbf{exp}\left\{\int_0^t \boldsymbol{A}(s)ds\right\}\mathbf{exp}\left\{-\frac{1}{2}\int_0^t(\boldsymbol{D})^2(s)ds+\int_0^t \boldsymbol{D}(s)dW_s\right\}\right)\\
&= \mathbf{exp}\left\{\int_0^t \boldsymbol{A}(s)ds\right\}d\left(\mathbf{exp}\left\{-\frac{1}{2}\int_0^t(\boldsymbol{D})^2(s)ds+\int_0^t \boldsymbol{D}(s)dW_s\right\}\right)\\
&\quad + d\left(\mathbf{exp}\left\{\int_0^t \boldsymbol{A}(s)ds\right\}\right)\mathbf{exp}\left\{-\frac{1}{2}\int_0^t(\boldsymbol{D})^2(s)ds+\int_0^t \boldsymbol{D}(s)dW_s\right\}\\
&\quad + d\left(\mathbf{exp}\left\{\int_0^t \boldsymbol{A}(s)ds\right\}\right)d\left(\mathbf{exp}\left\{-\frac{1}{2}\int_0^t(\boldsymbol{D})^2(s)ds+\int_0^t \boldsymbol{D}(s)dW_s\right\}\right).
\end{aligned} \tag{2.A.18}$$

We can apply the property (2.A.17) to compute the $d(\mathbf{exp}\left\{\boldsymbol{A}(s)\right\})$ terms in (2.A.18). Next we claim that the term

$$\mathbf{exp}\left\{-\frac{1}{2}\int_0^t(\boldsymbol{D})^2(s)ds+\int_0^t \boldsymbol{D}(s)dW_s\right\} =: \boldsymbol{Y}(t), \tag{2.A.19}$$

where $\boldsymbol{Y}(t)$ is a matrix stochastic process, has the Ito differential

$$d\boldsymbol{Y}_t = \boldsymbol{D}(t)\boldsymbol{Y}_t dW_t = \boldsymbol{Y}_t\boldsymbol{D}(t)dW_t. \tag{2.A.20}$$

Indeed, we can easily see this by first observing that $-\frac{1}{2}\int_0^t(\boldsymbol{D})^2(s)ds+\int_0^t \boldsymbol{D}(s)dW_s$ is a diagonal matrix, and hence $\boldsymbol{Y}_t$, which is the left hand side of (2.A.19), is also a diagonal matrix:

$$\boldsymbol{Y}^{(i,j)}(t) = \begin{cases} \exp\left\{-\frac{1}{2}\int_0^t(\boldsymbol{D})^2_{i,i}(s)ds+\int_0^t \boldsymbol{D}_{i,i}(s)dW_s\right\} & \text{if } i=j\\ 0 & \text{otherwise}\end{cases}$$

Hence each entry on the main diagonal satisfies the scalar SDE

$$d\boldsymbol{Y}^{(i,i)}(t) = \boldsymbol{D}_{i,i}(t)\boldsymbol{Y}^{(i,i)}(t)dW_t = \boldsymbol{Y}^{(i,i)}(t)\boldsymbol{D}_{i,i}(t)dW_t.$$

Let us now compute $d\boldsymbol{Y}(t) = (d(\boldsymbol{Y}(t))^{(i,j)})_{1\le i,j\le N}$:

$$d\boldsymbol{Y}_t^{(i,j)} = \begin{cases} \boldsymbol{D}_{i,i}(t)\boldsymbol{Y}^{(i,i)}(t)dW_t = \boldsymbol{Y}^{(i,i)}(t)\boldsymbol{D}(t)dW_t, & \text{if } i=j\\ 0 & \text{otherwise}\end{cases},$$

hence we immediately obtain (2.A.20).

From (2.A.17) and (2.A.20) we can re-write equation (2.A.18) as follows:

$$\begin{aligned} d(\mathbf{\Phi}_t) &= \mathbf{exp}\left\{\int_0^t \boldsymbol{A}(s)ds\right\}\mathbf{exp}\left\{-\frac{1}{2}\int_0^t (\boldsymbol{D})^2(s)ds + \int_0^t \boldsymbol{D}(s)dW_s\right\}\boldsymbol{D}(t)dW_t \\ &\quad + \left(\boldsymbol{A}(t)\mathbf{exp}\left\{\int_0^t \boldsymbol{A}(s)ds\right\}dt\right)\mathbf{exp}\left\{-\frac{1}{2}\int_0^t (\boldsymbol{D})^2(s)ds + \int_0^t \boldsymbol{D}(s)dW_s\right\} \qquad (2.A.21) \\ &\quad + \left(\mathbf{exp}\left\{\int_0^t \boldsymbol{A}(s)ds\right\}\boldsymbol{A}(t)dt\right)\mathbf{exp}\left\{-\frac{1}{2}\int_0^t (\boldsymbol{D})^2(s)ds + \int_0^t \boldsymbol{D}(s)dW_s\right\}\boldsymbol{D}(t)dW_t \\ &= \boldsymbol{D}(t)\mathbf{exp}\left\{\int_0^t \boldsymbol{A}(s)ds - \frac{1}{2}\int_0^t (\boldsymbol{D})^2(s)ds + \int_0^t \boldsymbol{D}(s)dW_s\right\}dW_t \\ &\quad + \ \boldsymbol{A}(t)\mathbf{exp}\left\{\int_0^t \boldsymbol{A}(s)ds - \frac{1}{2}\int_0^t (\boldsymbol{D})^2(s)ds + \int_0^t \boldsymbol{D}(s)dW_s\right\}dt \qquad (2.A.22) \\ &= \boldsymbol{A}(t)\mathbf{\Phi}_t dt + \boldsymbol{D}(t)\mathbf{\Phi}_t dW_t. \qquad (2.A.23) \end{aligned}$$

where (2.A.22) is due to the commutativity of $\boldsymbol{A}(t)$ and $\boldsymbol{D}(t)$, and since terms in "$dt \cdot dW$" represent higher order terms that can be neglected. □

### 2.A.3 Proof of Proposition 2.3.8

*Proof.* From Proposition 2.3.6 we know that $\mathbf{\Phi}(t, t_0)$ is the solution of

$$d\mathbf{\Phi}(t, t_0) = \boldsymbol{A}(t)\mathbf{\Phi}(t, t_0)dt + \boldsymbol{D}(t)\mathbf{\Phi}(t, t_0)dY_t, \qquad \mathbf{\Phi}(t_0, t_0) = \boldsymbol{I}.$$

By applying Ito's formula and since $\boldsymbol{A}(t)$, $\boldsymbol{D}(t)$ and $\mathbf{\Phi}(t, t_0)$ commute, we can compute the differential $d\mathbf{\Phi}(t, t_0)^{-1}$ (note that $\mathbf{\Phi}(t, t_0) > 0 \ \forall t$)

$$d\mathbf{\Phi}(t, t_0)^{-1} = -\mathbf{\Phi}(t, t_0)^{-1}\left\{\boldsymbol{A}(t) - \boldsymbol{D}^2(t)\right\}dt - \mathbf{\Phi}(t, t_0)^{-1}\boldsymbol{D}(t)dY_t \qquad (2.A.24)$$

Let us define $\mathbf{\Psi}_t := \mathbf{\Phi}(t, t_0)^{-1} \cdot \boldsymbol{S}_t$ and compute $d\mathbf{\Psi}_t = d(\mathbf{\Phi}^{-1}\boldsymbol{S})_t$ by the Ito product formula

$$d(\mathbf{\Phi}^{-1}\boldsymbol{S})_t = (\mathbf{\Phi}_t^{-1})d\boldsymbol{S}_t + d(\mathbf{\Phi}_t^{-1})\boldsymbol{S}_t + d\left[\mathbf{\Phi}^{-1}, \boldsymbol{S}\right]_t$$

$$\begin{aligned} d(\mathbf{\Phi}^{-1}\boldsymbol{S})_t &= \mathbf{\Phi}(t, t_0)^{-1}\left\{\boldsymbol{A}(t)\boldsymbol{S}_t + \boldsymbol{a}\left(\mathbf{\Phi}(t, t_0)\right)\right\}dt + \mathbf{\Phi}(t, t_0)^{-1}\left\{\boldsymbol{D}(t)\boldsymbol{S}_t + \boldsymbol{b}\left(\mathbf{\Phi}(t, t_0)\right)\right\}dY_t \\ &\quad - \mathbf{\Phi}(t, t_0)^{-1}\left\{\boldsymbol{A}(t) - \boldsymbol{D}^2(t)\right\}\boldsymbol{S}_t dt - \mathbf{\Phi}(t, t_0)^{-1}\boldsymbol{D}(t)\boldsymbol{S}_t dY_t \\ &\quad - \mathbf{\Phi}(t, t_0)^{-1}\boldsymbol{D}(t) \cdot \left\{\boldsymbol{D}(t)\boldsymbol{S}_t + \boldsymbol{b}\left(\mathbf{\Phi}(t, t_0)\right)\right\}dt. \qquad (2.A.25) \end{aligned}$$

After cancellations we have

$$d(\boldsymbol{S}\Phi^{-1})_t = d\mathbf{\Psi}_t = \mathbf{\Phi}(t, t_0)^{-1}\left\{\boldsymbol{a}(\mathbf{\Phi}(t, t_0)) + \boldsymbol{D}(t)\boldsymbol{b}(\mathbf{\Phi}(t, t_0))\right\}dt + \mathbf{\Phi}(t, t_0)^{-1}\boldsymbol{b}(\mathbf{\Phi}(t, t_0))dY_t$$

which has the integrated form (note that $\mathbf{\Psi}_0 = \boldsymbol{S}_0 \cdot \boldsymbol{I}$):

$$\mathbf{\Psi}_t = \boldsymbol{S}_0 + \int_0^t \mathbf{\Phi}(s, t_0)^{-1}\left[\boldsymbol{a}(\mathbf{\Phi}(s, t_0)) - \boldsymbol{D}(t)\boldsymbol{b}(\mathbf{\Phi}(s, t_0))\right]ds + \int_0^t \mathbf{\Phi}(s, t_0)^{-1}\boldsymbol{b}(\mathbf{\Phi}(s, t_0))dY_s \qquad (2.A.26)$$

Finally, the equality

$$\boldsymbol{S}_t = \mathbf{\Phi}_t\mathbf{\Psi}_t$$

gives us the result. □

### 2.A.4 Proof of Proposition 2.3.15

We shall need a probabilistic analogue of the mean value theorem in order to obtain the strong order of the approximating sequence $(\boldsymbol{Z}_k^{\Delta})_k$.

**Definition 2.A.3.** *We denote by the symbol $\leq_{st}$, the inequality operator in the usual stochastic sense, to mean the following relation:*
*Two real random variables $A$ and $B$ are said to satisfy the inequality $A \leq_{st} B$ if and only if $P(A > x) \leq P(B > x)$ for all $x \in \mathbb{R}$.*

**Lemma 2.A.4.** *If $E[A] < E[B] < \infty$ and $A \leq_{st} B$, then there exists an absolutely continuous non-negative random variable $C$ having probability density function*

$$f_C(x) = \frac{P[B > x] - P[A > x]}{E[B] - E[A]}, \quad x \geq 0.$$

*Let $g$ be a measurable and differentiable function such that $E[g(A)]$, $E[g(B)] < \infty$, and let its derivative $g'$ be measurable and Riemann-integrable on the interval $[x, y]$ for all $y \geq x \geq 0$. Then $E[g'(Z)]$ is finite and*

$$E[g(B)] - E[g(A)] = E[g'(C)] \cdot (E[B] - E[A]) .$$

*Proof.* See the paper [Cre99]. □

**Lemma 2.A.5.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{B}$, the Borel $\sigma$-algebra on $\mathbb{R}$. Let $A$, $B$ be real random variables in $(\Omega, \mathcal{F}, P)$ with $E[A], E[B] < \infty$. Define $\overline{C} := \max\{A, B\}$ and $\underline{C} := \min\{A, B, \}$, and $C := A - B$. Then $\overline{C}$ and $\underline{C}$ and $|C|$ are real random variables with $\underline{C} \leq_{st} \overline{C}$, $E[\underline{C}] \leq E[\overline{C}]$, and moreover*

$$|C| = \overline{C} - \underline{C}.$$

*Proof.* The $(\mathcal{F}, \mathcal{B})$-measurability of $\overline{C}$, $\underline{C}$ and $|C|$ are immediate, due to the measurability of $\min\{\cdot\}$, $\max\{\cdot\}$ and $|\cdot|$ (see any book on measure theory).To prove the first part of the claim, let us take a generic $x \in \mathbb{R}$ and define the subset $\mathcal{A} := \{\omega \in \Omega \mid A(\omega) \geq B(\omega)\}$. Since $\mathcal{A} = \cup_{x \in \mathbb{R}} \{A(\omega) \geq x\} \cap \{x \geq B(\omega)\}$ and the sets $\{A(\omega) \geq x\}$, and $\{B(\omega) \geq x\}$ are $\mathcal{B}$-measurable, then $\mathcal{A}$ is $(\mathcal{F}, \mathcal{B})$-measurable.

Note that $\mathcal{A} \cap \mathcal{A}^c = \Omega$ and for any $\omega \in \Omega$, we can write

$$\overline{C}(\omega) = A(\omega)\mathbf{1}_{\{\mathcal{A}\}} + B(\omega)\mathbf{1}_{\{\mathcal{A}^c\}},$$

and

$$\underline{C}(\omega) = A(\omega)\mathbf{1}_{\{\mathcal{A}^c\}} + B(\omega)\mathbf{1}_{\{\mathcal{A}\}}.$$

Let us then prove that the claim holds for all $\omega \in \mathcal{A}$ and also for all $\omega \in \mathcal{A}^c$. Take $\omega \in \mathcal{A}$. Then

$$\overline{C}(\omega) = A(\omega) \qquad \underline{C}(\omega) = B(\omega)$$

and for any $x \in \mathbb{R}$,

$$P\left\{\overline{C}(\omega) > x\right\} = P\left\{A(\omega) > x\right\} \qquad P\left\{\underline{C}(\omega) > x\right\} = P\left\{B(\omega) > x\right\}.$$

We can write

$$\{A(\omega) > x\} = \{A(\omega) - x > 0\} = \{A(\omega) - x \geq B(\omega) + x > 0\} \cup \{A(\omega) - x > 0 \geq B(\omega) - x\},$$

and on the other hand

$$\{B(\omega) > x\} = \{B(\omega) - x > 0\} = \{A(\omega) - x \geq B(\omega) - x > 0\}$$

since $\omega \in \mathcal{A}$. Hence

$$P\left\{\overline{C}(\omega) > x\right\} = P\{A(\omega) > x\} > P\{B(\omega) > x\} = P\{\underline{C}(\omega) > x\}.$$

If we then take $\omega \in \mathcal{A}^c$ and interchange the role of $A$ and $B$, we obtain analogously that

$$P\left\{\overline{C}(\omega) > x\right\} = P\{B(\omega) > x\} > P\{A(\omega) > x\} = P\{\underline{C}(\omega) > x\}.$$

That proves

$$\underline{\mathrm{C}} \leq_{st} \overline{C}.$$

Since $\underline{\mathrm{C}} \leq_{st} \overline{C}$ if and only if $E[\varphi(\underline{\mathrm{C}})] \leq E[\varphi(\overline{C})]$ for any non-decreasing functions $\varphi$ for which the expectations exist (see [SS07], Section 1.A.1), then $E[\underline{\mathrm{C}}] \leq E[\overline{C}]$ immediately follows. To prove the second part of the claim, note that for any real numbers $a, b < \infty$, then $|a - b| = \max\{a, b\} - \min\{a, b\}$ by definition of $|\cdot|$. Thus the equality

$$|C(\omega)| = \overline{C}(\omega) - \underline{C}(\omega)$$

holds for any arbitrary fixed $\omega \in \Omega$, and hence

$$|C| = \overline{C} - \underline{C}.$$

□

We shall make use of the following

**Definition 2.A.6.** *The induced operator norm $\|\cdot\|$ for matrices $\boldsymbol{A} \in \mathbb{R}^{k\times k}$ is defined as*

$$\|\boldsymbol{A}\| := \max\left\{\frac{\|\boldsymbol{A}x\|}{\|x\|} \mid x \in \mathbb{R}^k\right\},$$

*where $\|\|$ are any vector p-norms. In the case when the operator norm $\|\cdot\|$ is induced by the Euclidean ($p = 2$) norm, we shall distinguish it by the symbol $\|\cdot\|_2$.*

As a consequence of the definition, it is immediate that

$$|\boldsymbol{A}x| \leq \|\boldsymbol{A}\|\,|x|.$$

Note also that this norm is a so-called *sub-multiplicative* norm, that is

$$\|\boldsymbol{A}\boldsymbol{B}\| \leq \|\boldsymbol{A}\| \cdot \|\boldsymbol{B}\|.$$

**Proposition 2.A.7.** *Given a matrix $\boldsymbol{A} \in \mathbb{R}^{k\times k}$, then $\|\boldsymbol{A}\|_2 = \sqrt{\lambda_{\max}(\boldsymbol{A}^\top \boldsymbol{A})}$ where $\lambda_{\max}(\boldsymbol{A}^\top \boldsymbol{A})$ is the largest eigenvalue of $\boldsymbol{A}^\top \boldsymbol{A}$.*

*Proof.* See [PP12]. □

We now prove Proposition 2.3.15

*Proof.* We shall prove the case when the initial vector $\boldsymbol{Z}_0$ may itself be any random vector, in particular we will not have independence of $\boldsymbol{Z}_0$ and $\boldsymbol{Z}_t$ nor $\boldsymbol{Z}_0$ and $\boldsymbol{Z}_k^{\Delta}$.

Let $t = t_k$. Let us recall that for any Euler-Maruyama scheme, the strong order is $\gamma = 0.5$ (see [PBL10] section 5.2), that is, for any initial vector $\boldsymbol{z}_0$,

$$\bar{E}\left\{\left|\boldsymbol{\zeta}_t \boldsymbol{z}_0 - \boldsymbol{\zeta}_k^{\Delta} \boldsymbol{z}_0\right|\right\} \leq K \cdot \Delta^{0.5} \tag{2.A.27}$$

for some constant matrix $K$.

Now we consider the following global error:

$$\bar{E}\left\{\left|\mathbf{exp}\left\{\boldsymbol{\zeta}_t\right\} \boldsymbol{Z}_0 - \mathbf{exp}\left\{\boldsymbol{\zeta}_k^{\Delta}\right\} \boldsymbol{Z}_0\right|\right\}$$

Observe that $t = t_k$ and we can factor out $\mathbf{exp}\left\{\boldsymbol{Q} \cdot t\right\} = \mathbf{exp}\left\{\boldsymbol{Q} \cdot t_k\right\}$, which is deterministic, from $\mathbf{exp}\left\{\boldsymbol{\zeta}_t\right\} \boldsymbol{Z}_0 - \mathbf{exp}\left\{\boldsymbol{\zeta}_k^{\Delta}\right\} \boldsymbol{Z}_0$. Let

$$\bar{\boldsymbol{\zeta}}_t := \boldsymbol{Q} \cdot t - \boldsymbol{\zeta}_t$$

and

$$\bar{\boldsymbol{\zeta}}_k^{\Delta} := \boldsymbol{Q} \cdot t - \boldsymbol{\zeta}_k^{\Delta}.$$

$$\begin{aligned}
\bar{E}\left\{\left|\left(\mathbf{exp}\left\{\boldsymbol{\zeta}_t\right\} - \mathbf{exp}\left\{\boldsymbol{\zeta}_k^{\Delta}\right\}\right) \cdot \boldsymbol{Z}_0\right|\right\} =& \bar{E}\left\{\left|e^{\boldsymbol{Q} \cdot t_k} \cdot \left(\mathbf{exp}\left\{\bar{\boldsymbol{\zeta}}_t\right\} - \mathbf{exp}\left\{\bar{\boldsymbol{\zeta}}_k^{\Delta}\right\}\right) \boldsymbol{Z}_0\right|\right\} \\
\leq& \bar{E}\left\{\left\|e^{\boldsymbol{Q} \cdot t_k}\right\|_2 \cdot \left|\left(\mathbf{exp}\left\{\bar{\boldsymbol{\zeta}}_t\right\} - \mathbf{exp}\left\{\bar{\boldsymbol{\zeta}}_k^{\Delta}\right\}\right) \boldsymbol{Z}_0\right|\right\} \\
=& \left\|e^{\boldsymbol{Q} \cdot t_k}\right\|_2 \cdot \bar{E}\left\{\left|\left(\mathbf{exp}\left\{\bar{\boldsymbol{\zeta}}_t\right\} - \mathbf{exp}\left\{\bar{\boldsymbol{\zeta}}_k^{\Delta}\right\}\right) \cdot \boldsymbol{Z}_0\right|\right\}
\end{aligned}$$

Let us focus on the following global error

$$\bar{E}\left\{\left|\left(\mathbf{exp}\left\{\bar{\boldsymbol{\zeta}}_t\right\} - \mathbf{exp}\left\{\bar{\boldsymbol{\zeta}}_k^{\Delta}\right\}\right) \boldsymbol{Z}_0\right|\right\},$$

since the factor $\left\|e^{\boldsymbol{Q} \cdot t_k}\right\|_2$ is a scalar constant which does not depend on $\Delta$. We have that

$$\begin{aligned}
&\bar{E}\left\{\left|\left(\mathbf{exp}\left\{\bar{\boldsymbol{\zeta}}_t\right\} - \mathbf{exp}\left\{\bar{\boldsymbol{\zeta}}_k^{\Delta}\right\}\right) \boldsymbol{Z}_0\right|\right\} \\
&= \bar{E}\left\{\sqrt{\boldsymbol{Z}_0^{\top} \left(\mathbf{exp}\left\{\bar{\boldsymbol{\zeta}}_t\right\} - \mathbf{exp}\left\{\bar{\boldsymbol{\zeta}}_k^{\Delta}\right\}\right)^{\top} \left(\mathbf{exp}\left\{\bar{\boldsymbol{\zeta}}_t\right\} - \mathbf{exp}\left\{\bar{\boldsymbol{\zeta}}_k^{\Delta}\right\}\right) \boldsymbol{Z}_0}\right\}
\end{aligned}$$

Note that, since $\boldsymbol{D}(t)$ is a diagonal matrix, then also $\bar{\boldsymbol{\zeta}}_t$ and $\bar{\boldsymbol{\zeta}}_k^{\Delta}$ are just diagonal matrices with the component-wise exponential functions on each element of the main diagonal. Hence

$$\left(\mathbf{exp}\left\{\bar{\boldsymbol{\zeta}}_t\right\} - \mathbf{exp}\left\{\bar{\boldsymbol{\zeta}}_k^{\Delta}\right\}\right)^{\top} \left(\mathbf{exp}\left\{\bar{\boldsymbol{\zeta}}_t\right\} - \mathbf{exp}\left\{\bar{\boldsymbol{\zeta}}_k^{\Delta}\right\}\right)$$

is also diagonal. We can then write

$$\begin{aligned}
&= \bar{E}\left\{\sqrt{\sum_{j=1}^{N} (\boldsymbol{Z}_0^{(j)})^2 \cdot \left(e^{\bar{\boldsymbol{\zeta}}_t^{(j)}} - e^{\bar{\boldsymbol{\zeta}}_k^{\Delta,j}}\right)^2}\right\} \\
&\leq \bar{E}\left\{\sqrt{N \cdot \max_j \left\{(\boldsymbol{Z}_0^{(j)})^2 \cdot \left(e^{\bar{\boldsymbol{\zeta}}_t^{(j)}} - e^{\bar{\boldsymbol{\zeta}}_k^{\Delta,j}}\right)^2\right\}}\right\}
\end{aligned}$$

Let $i = \arg\max_j \left\{ \left| e^{\bar{\zeta}_t^{(j)}} - e^{\bar{\zeta}_k^{\Delta,j}} \right| \right\}$. Then

$$
\begin{aligned}
&= \bar{E}\left\{ \sqrt{N \cdot (\boldsymbol{Z}_0^{(i)})^2 \cdot \left( e^{\bar{\zeta}_t^{(i)}} - e^{\bar{\zeta}_k^{\Delta,i}} \right)^2} \right\} \\
&= \sqrt{N}\bar{E}\left\{ \left| (\boldsymbol{Z}_0^{(i)}) \cdot \left( e^{\bar{\zeta}_t^{(i)}} - e^{\bar{\zeta}_k^{\Delta,i}} \right) \right| \right\}
\end{aligned}
$$

Let us define the scalar real random variable $A(i) = \min\left\{ \bar{\zeta}_t^{(i)}, \bar{\zeta}_k^{\Delta,i} \right\} + \log(\boldsymbol{Z}_0^{(i)})$, and $B(i) = \max\left\{ \bar{\zeta}_t^{(i)}, \bar{\zeta}_k^{\Delta,i} \right\} + \log(\boldsymbol{Z}_0^{(i)})$. Then, by Lemma 2.A.5 $A(i) \leq_{st} B(i)$ and $B(i) - A(i) = \left| \bar{\zeta}_t^{(i)} - \bar{\zeta}_k^{\Delta,i} \right|$, and by (2.A.27)

$$
\bar{E}\left\{ B(i) - A(i) \right\} \leq K^i \cdot \Delta^{0.5}. \tag{2.A.28}
$$

Note that if $x \leq_{st} y$, then $e^x \leq_{st} e^y$ since $\exp\{\cdot\}$ is an increasing function. Hence

$$
\begin{aligned}
\bar{E}\left\{ \left| (\boldsymbol{Z}_0^{(i)}) \cdot \left( e^{\bar{\zeta}_t{}^{(i)}} - e^{\zeta_k^{\Delta,i}} \right) \right| \right\} & \\
&= \bar{E}\left\{ \left| e^{\bar{\zeta}_t{}^{(i)} + \log(\boldsymbol{Z}_0^{(i)})} - e^{\zeta_k^{\Delta,i} + \log(\boldsymbol{Z}_0^{(i)})} \right| \right\} \\
&= \bar{E}\left\{ e^{B(i)} - e^{A(i)} \right\}
\end{aligned}
$$

by Lemma 2.A.5.

We can then apply the Lemma 2.A.4 to $g(A(i))$ and $g(B(i))$with $g(\cdot) = \exp\{\cdot\}$ (the scalar exponential function). Let $C(i)$ denote the random variable $C$ from Lemma 2.A.4. We have

$$
\begin{aligned}
\bar{E}\left\{ e^{B(i)} - e^{A(i)} \right\} &\leq \bar{E}\left[ \mathbf{exp}\left\{ C(i) \right\} \right] \cdot \left( \bar{E}[B(i)] - \bar{E}[A(i)] \right) \\
&= \bar{E}\left[ \mathbf{exp}\left\{ C(i) \right\} \right] \cdot \left( \bar{E}[B(i) - A(i)] \right) \\
&= \bar{E}\left[ \mathbf{exp}\left\{ C(i) \right\} \right] \cdot \left( \bar{E}\left\{ \left| \bar{\zeta}_t^{(i)} - \bar{\zeta}_k^{\Delta,i} \right| \right\} \right) \\
&= \mu^{C(i)} \cdot K_2 \cdot \Delta^{0.5}
\end{aligned}
$$

where $\mu^{C(i)} := \bar{E}\left\{ \mathbf{exp}\left\{ C(i) \right\} \right\}$ and where in the last equation we have used the fact that, since $\boldsymbol{\zeta}_k^{\Delta,i}$ can be considered as an Euler-Maruyama discretization which is a strong approximation scheme of order 0.5 for $\boldsymbol{\zeta}_t^{(i)}$, the global error $\bar{E}\left\{ \left| \bar{\zeta}_t^{(i)} - \bar{\zeta}_k^{\Delta,i} \right| \right\}$ can be bounded as in (2.A.28). Hence

$$
\bar{E}\left\{ \left| \mathbf{exp}\left\{ \boldsymbol{\zeta}_t \right\} - \mathbf{exp}\left\{ \boldsymbol{\zeta}_k^{\Delta} \right\} \right| \right\} \leq \left\| \boldsymbol{e}^{\boldsymbol{Q} \cdot t_k} \right\| \cdot \mu^{C(i)} \cdot K_2 \cdot \Delta^{0.5}
$$

□

### 2.A.5 Proof of Proposition 2.3.17

*Proof.* Let us compute the expected global error that is

$$
\bar{E}\left\{ \left| \boldsymbol{S}_{t_{k+1}} - \boldsymbol{S}_{k+1}^{\Delta} \right| \right\}.
$$

To prove that the approximation is of strong order 0.5, we must show that the expected global error has an upper bound of the form $M \cdot \Delta^{0.5}$, where $M$ is a constant independent of $\Delta$. We

have

$$
\begin{aligned}
&\bar{E}\left\{\left|\boldsymbol{S}_{t_{k+1}}-\boldsymbol{S}_{k+1}^{\Delta}\right|\right\}\\
&=\bar{E}\left\{\left|\boldsymbol{\Phi}(t_{k+1},t_0)\left(\boldsymbol{S}_{t_0}+\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}(s,t_0)^{-1}\left[\boldsymbol{a}(\boldsymbol{\Phi}(s,t_0))-\boldsymbol{D}(s)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))\right]ds\right.\right.\right.\\
&\quad\left.\left.\left.+\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}^{-1}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))dY_s\right)-\boldsymbol{\Phi}_{k+1}^{\Delta}\left(\boldsymbol{S}_0^{\Delta}+\sum_{l=0}^{k}\psi_l\frac{\Delta t_l}{2}+\sum_{l=0}^{k}(\boldsymbol{\Phi}_l^{\Delta})^{-1}\boldsymbol{b}(\boldsymbol{\Phi}_l^{\Delta})\Delta Y_l\right)\right|\right\}\\
&=\bar{E}\left\{\left|\boldsymbol{\Phi}(t_{k+1},t_0)\boldsymbol{S}_{t_0}+\boldsymbol{\Phi}(t_{k+1},t_0)\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}(s,t_0)^{-1}\left[\boldsymbol{a}(\boldsymbol{\Phi}(s,t_0))-\boldsymbol{D}(s)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))\right]ds\right.\right.\\
&\quad+\boldsymbol{\Phi}(t_{k+1},t_0)\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}^{-1}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))dY_s-\boldsymbol{\Phi}_{k+1}^{\Delta}\boldsymbol{S}_0^{\Delta}-\boldsymbol{\Phi}_{k+1}^{\Delta}\sum_{l=0}^{k}\psi_l\frac{\Delta t_l}{2}\\
&\quad\left.\left.-\boldsymbol{\Phi}_{k+1}^{\Delta}\sum_{l=0}^{k}(\boldsymbol{\Phi}_l^{\Delta})^{-1}\boldsymbol{b}(\boldsymbol{\Phi}_l^{\Delta})\Delta Y_l\right|\right\}
\end{aligned}
$$

$$
\begin{aligned}
&=\bar{E}\left\{\left|\boldsymbol{\Phi}(t_{k+1},t_0)\boldsymbol{S}_{t_0}-\boldsymbol{\Phi}_{k+1}^{\Delta}\boldsymbol{S}_0^{\Delta}+\boldsymbol{\Phi}(t_{k+1},t_0)\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}(s,t_0)^{-1}\left[\boldsymbol{a}(\boldsymbol{\Phi}(s,t_0))-\boldsymbol{D}(s)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))\right]ds\right.\right.\\
&\quad-\boldsymbol{\Phi}_{k+1}^{\Delta}\sum_{l=0}^{k}\psi_l\frac{\Delta t_l}{2}+\boldsymbol{\Phi}(t_{k+1},t_0)\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}^{-1}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))dY_s\\
&\quad\left.\left.-\boldsymbol{\Phi}_{k+1}^{\Delta}\sum_{l=0}^{k}(\boldsymbol{\Phi}_l^{\Delta})^{-1}\boldsymbol{b}(\boldsymbol{\Phi}_l^{\Delta})\Delta Y_l\right|\right\}\\
&\leq\bar{E}\left\{\left|\boldsymbol{\Phi}(t_{k+1},t_0)\boldsymbol{S}_{t_0}-\boldsymbol{\Phi}_{k+1}^{\Delta}\boldsymbol{S}_0^{\Delta}\right|\right\}\\
&\quad+\bar{E}\left\{\left|\boldsymbol{\Phi}(t_{k+1},t_0)\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}(s,t_0)^{-1}\left[\boldsymbol{a}(\boldsymbol{\Phi}(s,t_0))-\boldsymbol{D}(s)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))\right]ds-\boldsymbol{\Phi}_{k+1}^{\Delta}\sum_{l=0}^{k}\psi_l\frac{\Delta t_l}{2}\right.\right.\\
&\quad\left.\left.+\boldsymbol{\Phi}(t_{k+1},t_0)\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}^{-1}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))dY_s-\boldsymbol{\Phi}_{k+1}^{\Delta}\sum_{l=0}^{k}(\boldsymbol{\Phi}_l^{\Delta})^{-1}\boldsymbol{b}(\boldsymbol{\Phi}_l^{\Delta})\Delta Y_l\right|\right\}
\end{aligned}
$$

Since $\boldsymbol{S}_{t_0}=\boldsymbol{S}_0^{\Delta}$ and by $\bar{E}\left\{\left|(\boldsymbol{\Phi}(t_{k+1},t_0)-\boldsymbol{\Phi}_{k+1}^{\Delta})\cdot\boldsymbol{S}_{t_0}\right|\right\}<K\Delta^{0.5}$ according to Proposition 2.3.15, replacing $\boldsymbol{S}_0^{\Delta}$ with $\boldsymbol{S}_{t_0}$, we have

$$
\begin{aligned}
&\leq K_3\cdot\Delta^{0.5}\\
&\quad+\bar{E}\left\{\left|\boldsymbol{\Phi}(t_{k+1},t_0)\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}(s,t_0)^{-1}\left[\boldsymbol{a}(\boldsymbol{\Phi}(s,t_0))-\boldsymbol{D}(s)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))\right]ds-\boldsymbol{\Phi}_{k+1}^{\Delta}\sum_{l=0}^{k}\psi_l\frac{\Delta t_l}{2}\right.\right.\\
&\quad\left.\left.+\boldsymbol{\Phi}(t_{k+1},t_0)\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}^{-1}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))dY_s-\boldsymbol{\Phi}_{k+1}^{\Delta}\sum_{l=0}^{k}(\boldsymbol{\Phi}_l^{\Delta})^{-1}\boldsymbol{b}(\boldsymbol{\Phi}_l^{\Delta})\Delta Y_l\right|\right\}\\
&\leq K_3\cdot\Delta^{0.5}\\
&\quad+\bar{E}\left\{\left|\boldsymbol{\Phi}(t_{k+1},t_0)\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}(s,t_0)^{-1}\left[\boldsymbol{a}(\boldsymbol{\Phi}(s,t_0))-\boldsymbol{D}(s)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))\right]ds-\boldsymbol{\Phi}_{k+1}^{\Delta}\sum_{l=0}^{k}\psi_l\frac{\Delta t_l}{2}\right|\right\}\\
&\quad+\bar{E}\left\{\left|\boldsymbol{\Phi}(t_{k+1},t_0)\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}^{-1}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))dY_s-\boldsymbol{\Phi}_{k+1}^{\Delta}\sum_{l=0}^{k}(\boldsymbol{\Phi}_l^{\Delta})^{-1}\boldsymbol{b}(\boldsymbol{\Phi}_l^{\Delta})\Delta Y_l\right|\right\}
\end{aligned}
\tag{2.A.29}
$$

We can re-write and bound the second expectation in (2.A.29) as follows:

$$\begin{aligned}
&\bar{E}\left\{\left|\boldsymbol{\Phi}(t_{k+1},t_0)\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}(s,t_0)^{-1}\left[\boldsymbol{a}(\boldsymbol{\Phi}(s,t_0))-\boldsymbol{D}(s)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))\right]ds-\boldsymbol{\Phi}^{\Delta}_{k+1}\sum_{l=0}^{k}\psi_l\frac{\Delta t_l}{2}\right|\right\}\\
&=\bar{E}\left\{\left|\boldsymbol{\Phi}(t_{k+1},t_0)\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}(s,t_0)^{-1}\left[\boldsymbol{a}(\boldsymbol{\Phi}(s,t_0))-\boldsymbol{D}(s)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))\right]ds-\boldsymbol{\Phi}(t_{k+1},t_0)\sum_{l=0}^{k}\psi_l\frac{\Delta t_l}{2}\right.\right.\\
&\qquad\left.\left.+\boldsymbol{\Phi}(t_{k+1},t_0)\sum_{l=0}^{k}\psi_l\frac{\Delta t_l}{2}-\boldsymbol{\Phi}^{\Delta}_{k+1}\sum_{l=0}^{k}\psi_l\frac{\Delta t_l}{2}\right|\right\}\\
&\leq\bar{E}\left\{\left|\boldsymbol{\Phi}(t_{k+1},t_0)\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}(s,t_0)^{-1}\left[\boldsymbol{a}(\boldsymbol{\Phi}(s,t_0))-\boldsymbol{D}(s)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))\right]ds-\boldsymbol{\Phi}(t_{k+1},t_0)\sum_{l=0}^{k}\psi_l\frac{\Delta t_l}{2}\right|\right\}\\
&\quad+\bar{E}\left\{\left|\boldsymbol{\Phi}(t_{k+1},t_0)\sum_{l=0}^{k}\psi_l\frac{\Delta t_l}{2}-\boldsymbol{\Phi}^{\Delta}_{k+1}\sum_{l=0}^{k}\psi_l\frac{\Delta t_l}{2}\right|\right\}
\end{aligned}$$

The first expectation is just the global error of a trapezoidal approximation to the integral in $dt$. Such a trapezoidal approximation is of strong order at least one (see Section 5 of [PBL10]). Hence

$$\begin{aligned}
&\leq K_4\cdot\Delta+\bar{E}\left\{\left|(\boldsymbol{\Phi}(t_{k+1},t_0)-\boldsymbol{\Phi}^{\Delta}_{k+1})\cdot\sum_{l=0}^{k}\psi_l\frac{\Delta t_l}{2}\right|\right\}\\
&\leq K_4\cdot\Delta+\frac{(k+1)\cdot\Delta}{2}\bar{E}\left\{\left|(\boldsymbol{\Phi}(t_{k+1},t_0)-\boldsymbol{\Phi}^{\Delta}_{k+1})\cdot\max_l\psi_l\right|\right\}\\
&\leq K_4\cdot\Delta+\frac{(k+1)\Delta}{2}\cdot K_5\cdot\Delta^{0.5}\qquad(2.A.30)\\
&=K_4\cdot\Delta+K_6\Delta^{1.5}.
\end{aligned}$$

The last inequality follows again from Proposition 2.3.15 letting the initial vector $\boldsymbol{Z}_0:=\max_l\psi_l$. (We recall that in the proof for Proposition 2.3.15 we did not require the initial vector $\boldsymbol{Z}_0$ to be known. Indeed, we proved it for the general case when it may be random and there may be dependence between the initial vector and the other terms of the approximation.)

It remains to bound the expectation in the third term on the right hand side of (2.A.29), which is

$$\begin{aligned}
&\bar{E}\left\{\left|\boldsymbol{\Phi}(t_{k+1},t_0)\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}^{-1}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))dY_s-\boldsymbol{\Phi}^{\Delta}_{k+1}\sum_{l=0}^{k}(\boldsymbol{\Phi}^{\Delta}_l)^{-1}\boldsymbol{b}(\boldsymbol{\Phi}^{\Delta}_l)\Delta Y_l\right|\right\}\\
&\leq\bar{E}\left\{\left|\boldsymbol{\Phi}(t_{k+1},t_0)\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}^{-1}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))dY_s-\boldsymbol{\Phi}(t_{k+1},t_0)\sum_{l=0}^{k}(\boldsymbol{\Phi}^{\Delta}_l)^{-1}\boldsymbol{b}(\boldsymbol{\Phi}^{\Delta}_l)\Delta Y_l\right|\right\}\\
&\quad+\bar{E}\left\{\left|\boldsymbol{\Phi}(t_{k+1},t_0)\sum_{l=0}^{k}(\boldsymbol{\Phi}^{\Delta}_l)^{-1}\boldsymbol{b}(\boldsymbol{\Phi}^{\Delta}_l)\Delta Y_l-\boldsymbol{\Phi}^{\Delta}_{k+1}\sum_{l=0}^{k}(\boldsymbol{\Phi}^{\Delta}_l)^{-1}\boldsymbol{b}(\boldsymbol{\Phi}^{\Delta}_l)\Delta Y_l\right|\right\}\\
&\leq\bar{E}\left\{\left|\boldsymbol{\Phi}(t_{k+1},t_0)\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}^{-1}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))dY_s-\boldsymbol{\Phi}(t_{k+1},t_0)\sum_{l=0}^{k}(\boldsymbol{\Phi}^{\Delta}_l)^{-1}\boldsymbol{b}(\boldsymbol{\Phi}^{\Delta}_l)\Delta Y_l\right|\right\}\\
&\quad+K_7\cdot\Delta^{0.5},
\end{aligned}$$

where the last term in the right -hand side is a bound that can be obtained following the same

argument as with inequality (2.A.30). It remains then to bound the term

$$\bar{E}\left\{\left|\boldsymbol{\Phi}(t_{k+1},t_0)\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}^{-1}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))dY_s-\boldsymbol{\Phi}(t_{k+1},t_0)\sum_{l=0}^{k}(\boldsymbol{\Phi}_l^{\Delta})^{-1}\boldsymbol{b}(\boldsymbol{\Phi}_l)\Delta Y_l\right|\right\}$$

Let us recall the following Lyapunov inequality:

$$\bar{E}\left[|A-B|\right]\leq\sqrt{\bar{E}\left[|A-B|^2\right]}. \tag{2.A.31}$$

Hence, if we can find a constant $M$ such that

$$\bar{E}\left\{\left|\boldsymbol{\Phi}(t_{k+1},t_0)\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}^{-1}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))dY_s-\boldsymbol{\Phi}(t_{k+1},t_0)\sum_{l=0}^{k}(\boldsymbol{\Phi}_l^{\Delta})^{-1}\boldsymbol{b}(\boldsymbol{\Phi}_l)\Delta Y_l\right|^2\right\}\leq M\Delta, \tag{2.A.32}$$

then by Lyapunov's inequality (2.A.31) it would follow that

$$\bar{E}\left\{\left|\boldsymbol{\Phi}(t_{k+1},t_0)\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}^{-1}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))dY_s-\boldsymbol{\Phi}(t_{k+1},t_0)\sum_{l=0}^{k}(\boldsymbol{\Phi}_l^{\Delta})^{-1}\boldsymbol{b}(\boldsymbol{\Phi}_l^{\Delta})\Delta Y_l\right|\right\}\leq\sqrt{M}\Delta^{0.5}$$

We shall now obtain a constant $M$ for which the inequality (2.A.32) holds. We have

$$\begin{aligned}
&\bar{E}\left\{\left|\boldsymbol{\Phi}(t_{k+1},t_0)\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}^{-1}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))dY_s-\boldsymbol{\Phi}(t_{k+1},t_0)\sum_{l=0}^{k}(\boldsymbol{\Phi}_l^{\Delta})^{-1}\boldsymbol{b}(\boldsymbol{\Phi}_l^{\Delta})\Delta Y_l\right|^2\right\}\\
&=\bar{E}\left\{\left|\left(\boldsymbol{\Phi}(t_{k+1},t_0)\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}^{-1}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))dY_s-\boldsymbol{\Phi}(t_{k+1},t_0)\sum_{l=0}^{k}(\boldsymbol{\Phi}_l^{\Delta})^{-1}\boldsymbol{b}(\boldsymbol{\Phi}_l^{\Delta})\Delta Y_l\right)^2\right|\right\}\\
&=\bar{E}\left\{\left|\left(\boldsymbol{\Phi}(t_{k+1},t_0)\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}^{-1}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))dY_s\right)^2\right.\right.\\
&\quad-2\boldsymbol{\Phi}(t_{k+1},t_0)^2\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}^{-1}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))dY_s\cdot\sum_{l=0}^{k}(\boldsymbol{\Phi}_l^{\Delta})^{-1}\boldsymbol{b}(\boldsymbol{\Phi}_l^{\Delta})\Delta Y_l\\
&\quad\left.\left.+\left(\boldsymbol{\Phi}(t_{k+1},t_0)\sum_{l=0}^{k}(\boldsymbol{\Phi}_l^{\Delta})^{-1}\boldsymbol{b}(\boldsymbol{\Phi}_l^{\Delta})\Delta Y_l\right)^2\right|\right\}\\
&=\bar{E}\left\{\left(\boldsymbol{\Phi}(t_{k+1},t_0)\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}^{-1}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))dY_s\right)^2\right\}\\
&\quad+2\bar{E}\left\{\left|\boldsymbol{\Phi}(t_{k+1},t_0)^2\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}^{-1}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))dY_s\cdot\sum_{l=0}^{k}(\boldsymbol{\Phi}_l^{\Delta})^{-1}\boldsymbol{b}(\boldsymbol{\Phi}_l^{\Delta})\Delta Y_l\right|\right\}\\
&\quad+\bar{E}\left\{\left(\boldsymbol{\Phi}(t_{k+1},t_0)\sum_{l=0}^{k}(\boldsymbol{\Phi}_l^{\Delta})^{-1}\boldsymbol{b}(\boldsymbol{\Phi}_l^{\Delta})\Delta Y_l\right)^2\right\}.
\end{aligned}\tag{2.A.33}$$

The expectation of the squared integral in $dY$ is the variance of an $\mathcal{F}_t^y$-martingale (recall that $Y$ is a $\bar{P}$-Brownian Motion). By the Ito isometry, it is

$$\bar{E}\left\{\left(\boldsymbol{\Phi}(t_{k+1},t_0)\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}^{-1}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))dY_s\right)^2\right\}=\bar{E}\left\{\boldsymbol{\Phi}(t_{k+1},t_0)^2\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}^{-2}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))^2ds\right\}$$

Due to the smoothness and local boundedness assumptions, we can apply Fubini's theorem and hence

$$\begin{aligned}\bar{E}\left\{\boldsymbol{\Phi}(t_{k+1},t_0)^2\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}^{-2}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))^2ds\right\}&=\int_{t_0}^{t_{k+1}}\bar{E}\left\{\boldsymbol{\Phi}(t_{k+1},t_0)^2\boldsymbol{\Phi}^{-2}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))^2\right\}ds\\&\leq\max_s\left\{\bar{E}\left\{\boldsymbol{\Phi}(t_{k+1},t_0)^2\boldsymbol{\Phi}^{-2}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))^2\right\}\mid s\in[t_0,t_{k+1}]\right\}\cdot(t_{k+1}-t_0)\\&=\max_s\left\{\bar{E}\left\{\boldsymbol{\Phi}(t_{k+1},t_0)^2\boldsymbol{\Phi}^{-2}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))^2\right\}\mid s\in[t_0,t_{k+1}]\right\}\cdot(k+1)\cdot\Delta.\end{aligned}$$

The expectation of $\boldsymbol{\Phi}(t_{k+1},t_0)^2\sum_{l=0}^k(\boldsymbol{\Phi}_l^\Delta)^{-2}(\boldsymbol{b}(\boldsymbol{\Phi}_l^\Delta)\Delta Y_l)^2$ is just the sum of the variances of each term, that is

$$\begin{aligned}&\bar{E}\left\{\boldsymbol{\Phi}(t_{k+1},t_0)^2\sum_{l=0}^k(\boldsymbol{\Phi}_l^\Delta)^{-2}(\boldsymbol{b}(\boldsymbol{\Phi}_l^\Delta)\Delta Y_l)^2\right\}=\sum_{l=0}^k\bar{E}\left\{\boldsymbol{\Phi}(t_{k+1},t_0)^2(\boldsymbol{\Phi}_l^\Delta)^{-2}(\boldsymbol{b}(\boldsymbol{\Phi}_l^\Delta)\Delta Y_l)^2\right\}\\&=\sum_{l=0}^k\bar{E}\left\{\boldsymbol{\Phi}(t_{k+1},t_0)^2(\boldsymbol{\Phi}_l^\Delta)^{-2}\boldsymbol{b}(\boldsymbol{\Phi}_l^\Delta)^2\right\}\Delta t_l\\&\leq\max_l\left\{\bar{E}\left\{\boldsymbol{\Phi}(t_{k+1},t_0)^2(\boldsymbol{\Phi}_l^\Delta)^{-2}\boldsymbol{b}(\boldsymbol{\Phi}_l^\Delta)^2\right\}\right\}\cdot(k+1)\cdot\Delta\end{aligned}$$

Define the number

$$\mu=\max_s\left\{E\left\{\boldsymbol{\Phi}(t_{k+1},t_0)^2\Phi(s,t_0)^{-2}\boldsymbol{b}(\Phi(s,t_0))^2\right\}\mid s\in[t_0,t_{k+1}]\right\}.$$

Analogously, define the number

$$\bar{\mu}=\max_l\left\{E\left\{\boldsymbol{\Phi}(t_{k+1},t_0)^2\left|(\boldsymbol{\Phi}_l^\Delta)^{-2}\boldsymbol{b}(\boldsymbol{\Phi}_l^\Delta)^2\right|\right\}\mid l=0,\dots k\right\}.$$

Hence continuing from (2.A.33) and applying the previous two passages,

$$\begin{aligned}&=\int_{t_0}^{t_{k+1}}\bar{E}\left\{\boldsymbol{\Phi}(t_{k+1},t_0)^2\boldsymbol{\Phi}^{-2}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))^2\right\}ds\\&\quad+2\bar{E}\left\{\left|\boldsymbol{\Phi}(t_{k+1},t_0)^2\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}^{-1}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))dY_s\cdot\sum_{l=0}^k(\boldsymbol{\Phi}_l^\Delta)^{-1}\boldsymbol{b}(\boldsymbol{\Phi}_l^\Delta)\Delta Y_l\right|\right\}\\&\quad+\sum_{l=0}^k\bar{E}\left\{\boldsymbol{\Phi}(t_{k+1},t_0)^2(\boldsymbol{\Phi}_l^\Delta)^{-2}\boldsymbol{b}(\boldsymbol{\Phi}_l^\Delta)^2\right\}\Delta t_l\\&\leq\mu\cdot(k+1)\cdot\Delta+2\bar{E}\left\{\left|\boldsymbol{\Phi}(t_{k+1},t_0)^2\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}^{-1}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))dY_s\cdot\sum_{l=0}^k(\boldsymbol{\Phi}_l^\Delta)^{-1}\boldsymbol{b}(\boldsymbol{\Phi}_l^\Delta)\Delta Y_l\right|\right\}\\&\quad+\bar{\mu}\cdot(k+1)\cdot\Delta.\end{aligned}\tag{2.A.34}$$

We can write the integral in the expectation in (2.A.34) as follows

$$\boldsymbol{\Phi}(t_{k+1},t_0)^2\int_{t_0}^{t_{k+1}}\boldsymbol{\Phi}^{-1}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))dY_s=\boldsymbol{\Phi}(t_{k+1},t_0)^2\sum_{l=0}^k\int_{t_l}^{t_{l+1}}\boldsymbol{\Phi}^{-1}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))dY_s$$

and each term in the sum can be bounded in the following way

$$\begin{aligned}&\boldsymbol{\Phi}(t_{k+1},t_0)^2\int_{t_l}^{t_{l+1}}\boldsymbol{\Phi}^{-1}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))dY_s\\&\leq\max_s\left\{\boldsymbol{\Phi}(t_{k+1},t_0)^2\cdot\boldsymbol{\Phi}^{-1}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))\mid s\in[t_k,t_{k+1}]\right\}\cdot\int_{t_l}^{t_{l+1}}dY_s\\&=\max_s\left\{\boldsymbol{\Phi}(t_{k+1},t_0)^2\cdot\boldsymbol{\Phi}^{-1}(s,t_0)\boldsymbol{b}(\boldsymbol{\Phi}(s,t_0))\mid s\in[t_k,t_{k+1}]\right\}\cdot\Delta Y_l.\end{aligned}$$

Let us define the random variables

$$\mu(k) = \sup_s \left\{ \mathbf{\Phi}(t_{k+1}, t_0)^2 \Phi(s, t_0)^{-1} \boldsymbol{b}(\Phi(s, t_0)) \mid s \in [t_k, t_{k+1}] \right\}.$$

Note that

$$\mu = \max \left\{ \bar{E}\left\{\mu(k)\right\} \mid k = 0, \ldots, \frac{T}{\Delta} \right\}.$$

Analogously, we define

$$\bar{\mu}(k) = \max_l \left\{ \left| (\mathbf{\Phi}_l^\Delta)^{-1} \boldsymbol{b}(\mathbf{\Phi}_l^\Delta) \right| \mid l = 0, \ldots k \right\}$$

and also observe that

$$\bar{\mu} = \max \left\{ \bar{E}\left\{\bar{\mu}(k)\right\} \mid k = 0, \ldots, \frac{T}{\Delta} \right\}.$$

Then, continuing from (2.A.34) we have

$$\begin{aligned}
&(\mu + \bar{\mu}) \cdot (k+1) \cdot \Delta + 2\bar{E}\left\{ \left| \mathbf{\Phi}(t_{k+1}, t_0)^2 \int_{t_0}^{t_{k+1}} \mathbf{\Phi}^{-1}(s, t_0) \boldsymbol{b}(\mathbf{\Phi}(s, t_0)) dY_s \cdot \sum_{l=0}^{k} (\mathbf{\Phi}_l^\Delta)^{-1} \boldsymbol{b}(\mathbf{\Phi}_l^\Delta) \Delta Y_l \right| \right\} \\
&= K_8 \cdot \Delta + 2\bar{E}\left\{ \left| \mathbf{\Phi}(t_{k+1}, t_0)^2 \sum_{l=0}^{k} \int_{t_l}^{t_{l+1}} \mathbf{\Phi}^{-1}(s, t_0) \boldsymbol{b}(\mathbf{\Phi}(s, t_0)) dY_s \cdot \sum_{l=0}^{k} (\mathbf{\Phi}_l^\Delta)^{-1} \boldsymbol{b}(\mathbf{\Phi}_l^\Delta) \Delta Y_l \right| \right\} \\
&\leq K_8 \cdot \Delta + 2\bar{E}\left\{ \left| \sum_{l=0}^{k} \left( \max_s \left\{ \mathbf{\Phi}(t_{k+1}, t_0)^2 \cdot \mathbf{\Phi}^{-1}(s, t_0) \boldsymbol{b}(\mathbf{\Phi}(s, t_0)) \right\} \cdot \Delta Y_l \right) \cdot \sum_{l=0}^{k} (\mathbf{\Phi}_l^\Delta)^{-1} \boldsymbol{b}(\mathbf{\Phi}_l^\Delta) \Delta Y_l \right| \right\} \\
&\leq K_8 \cdot \Delta + 2\bar{E}\left\{ \left| \sum_{i=0}^{k} \mu(i) \Delta Y_i \cdot \sum_{l=0}^{k} \bar{\mu}(k) \Delta Y_l \right| \right\} \\
&\leq K_8 \cdot \Delta + 2\bar{E}\left\{ \left| \sum_{i=l=0}^{k} \mu(i) \bar{\mu}(l) \Delta Y_i^2 \right| \right\}
\end{aligned}$$

where we have used the independence property of the Brownian increments $\Delta Y_j \sim \mathcal{N}(0, \Delta t_j)$ and $\Delta Y_i \sim \mathcal{N}(0, \Delta t_i)$ whenever $i > j$, hence $E[\Delta Y_i \Delta Y_j] = 0$. Finally, by the Ito isometry

$$\begin{aligned}
&\leq K_8 \cdot \Delta + 2 \sum_{i=0}^{k} \bar{E}\left\{ \mu(i) \bar{\mu}(i) \Delta Y_i^2 \right\} \\
&\leq K_8 \cdot \Delta + 2 \sum_{i=0}^{k} \bar{E}\left\{ \mu(i) \bar{\mu}(i) \right\} \Delta t_i \\
&\leq K_8 \cdot \Delta + 2\mu\bar{\mu} \cdot (k+1)\Delta \\
&= (K_8 + 2\mu\bar{\mu} \cdot (k+1)) \cdot \Delta = M \cdot \Delta.
\end{aligned}$$

This proves (2.A.32) and the claim. □

# Part III

# Risk-management Aspects

# Chapter 3

# A unified approach to pricing and risk management of equity and credit risk

## 3.1 Introduction

The last few years have witnessed an increasing popularity of hybrid equity/credit risk models, as documented by the recent papers [Bay08, CPS09, CL06, CM10, CS08, CW10, CW12, CK12]. One of the most appealing features of such models is their capability to link the stochastic behavior of the stock price (and of its volatility) with the randomness of the default event and, hence, with the level of credit spreads. The relation between equity and credit risk is supported by strong empirical evidence (we refer the reader to [CL06, CK12] for good overviews of the related literature) and several studies document significant relationships between stock volatility and credit spreads of corporate bonds and Credit Default Swaps ([CT03, CDMW08]).

In Part III of this work, we propose a general framework for the joint modeling of equity and credit risk which allows for a flexible dependence between stock price, stochastic volatility, default intensity and interest rate. The proposed framework is fully analytically tractable, since it relies on the powerful technology of affine processes (see e.g. [DFS03, KR09] for financial applications of affine processes), and nests several stochastic volatility models proposed in the literature, thereby extending their scope to a defaultable setting. Affine models have been successfully employed in credit risk models, as documented by the papers [CW12, Duf05, GS09]. A distinguishing feature of our approach is that, unlike the models proposed in [Bay08, CL06, CM10, CS08, CW12, CK12], we jointly consider both physical and risk-neutral probability measures, ensuring that the analytical tractability is preserved under a change of measure, while at the same time avoiding unnecessarily restrictive specifications of the risk premia. This aspect is of particular importance in credit risk modeling, where one is typically faced with the two problems of computing survival probabilities or related risk measures and of computing arbitrage-free prices of credit derivatives. In this paper, we provide a complete characterisation of the set of risk-neutral measures which preserve the affine structure of the model, thus enabling us to efficiently compute several quantities which are of interest in view of both risk management and pricing applications.

The third part of this work is structured as follows. Section 3.2 introduces the modeling framework, while Section 3.3 gives a characterisation of the family of risk-neutral measures which

preserve the affine structure of the model. In Sections 3.4-3.5, we show how most quantities of interest for risk management and pricing applications, respectively, can be efficiently computed under suitable (risk-neutral) survival measures (we refer the reader to Sect. 2.5 of [Fon12a] for more detailed proofs of the results of Sections 3.4-3.5). Section 3.6 illustrates the main features of the proposed approach within a simple example, which corresponds to a defaultable extension of the Heston [Hes93] model. Finally, Section 4.3 concludes.

## 3.2 The modeling framework

This section presents the mathematical structure of the modeling framework. Let $(\Omega, \mathcal{G}, P)$ be a reference probability space, with $P$ denoting the physical/statistical probability measure (we want to emphasise that our framework will be entirely formulated with respect to the physical measure $P$). Let $T \in (0, \infty)$ be a fixed time horizon and $W = (W_t)_{0 \leq t \leq T}$ an $\mathbb{R}^d$-valued Brownian motion on $(\Omega, \mathcal{G}, P)$, with $d \geq 2$, and denote by $\mathbb{F} = (\mathcal{F}_t)_{0 \leq t \leq T}$ its $P$-augmented natural filtration.

We focus our attention on a single defaultable firm, whose *default time* $\tau : \Omega \to [0, T] \cup \{+\infty\}$ is supposed to be a $(P, \mathbb{F})$-*doubly stochastic* random time, in the sense of Def. 9.11 of [MFE05]. This means that there exists a strictly positive $\mathbb{F}$-adapted process $\lambda^P = (\lambda^P_t)_{0 \leq t \leq T}$ such that

$$P(\tau > t \,|\, \mathcal{F}_T) = P(\tau > t \,|\, \mathcal{F}_t) = \exp\left( - \int_0^t \lambda^P_u du \right), \qquad \text{for all } t \in [0, T].$$

In order to emphasize the role of the reference measure $P$, we call the process $\lambda^P$ the $P$-*intensity* of $\tau$. Let the filtration $\mathbb{G} = (\mathcal{G}_t)_{0 \leq t \leq T}$ be the *progressive enlargement*[1] of $\mathbb{F}$ with respect to $\tau$, i.e., $\mathcal{G}_t := \bigcap_{s > t} \big\{ \mathcal{F}_s \vee \sigma(\tau \wedge s) \big\}$, for all $t \in [0, T]$, and let $\mathcal{G} = \mathcal{G}_T$. It is well-known that $\mathbb{G}$ is the smallest filtration (satisfying the usual conditions) which makes $\tau$ a $\mathbb{G}$-stopping time and contains $\mathbb{F}$, in the sense that $\mathcal{F}_t \subset \mathcal{G}_t$ for all $t \in [0, T]$.

The price at time $t \in [0, T]$ of one share issued by the defaultable firm is denoted by $S_t$. We assume that the $\mathbb{G}$-adapted process $S = (S_t)_{0 \leq t \leq T}$ is continuous and strictly positive on the stochastic interval $[\![ 0, \tau [\![$ and satisfies $S \mathbf{1}_{[\![ \tau, T ]\!]} = 0$. This means that $S$ drops to zero as soon as the default event occurs and remains thereafter frozen at that level. By relying on the Sect. 5.1 of [BR02] together with the fact that all $\mathbb{F}$-martingales are continuous, it can be proved that there exists a continuous strictly positive $\mathbb{F}$-adapted process $\tilde{S} = (\tilde{S}_t)_{0 \leq t \leq T}$ such that $S_t = \mathbf{1}_{\{\tau > t\}} \tilde{S}_t$ holds for all $t \in [0, T]$. We shall refer to the process $\tilde{S}$ as the *pre-default* value of $S$.

The pre-default value $\tilde{S}$ is assumed to be influenced by the $\mathbb{F}$-adapted *stochastic volatility* process $v = (v_t)_{0 \leq t \leq T}$ and by an $\mathbb{R}^{d-2}$-valued $\mathbb{F}$-adapted *factor process* $Y = (Y_t)_{0 \leq t \leq T}$. The process $Y$ can include macro-economic covariates describing the state of the economy as well as firm-specific and latent variables, as considered e.g. in [Fon12b, FR10a]. Let us define the process $L = (L_t)_{0 \leq t \leq T}$ by $L_t := \log \tilde{S}_t$ and the $\mathbb{R}^d$-valued $\mathbb{F}$-adapted process $X = (X_t)_{0 \leq t \leq T}$ by $X_t := (v_t, Y_t^\top, L_t)^\top$, with $^\top$ denoting transposition.

The processes $v$, $Y$ and $L$ are jointly specified through the following square-root-type SDE for the process $X$ on the state space $\mathbb{R}^m_{++} \times \mathbb{R}^{d-m}$, where we let $\mathbb{R}^m_{++} := \{x \in \mathbb{R}^m : x_i > 0, \forall i =$

[1] Due to Lemma 6.1.1 and Lemma 6.1.2 of [BR02], the fact that $P(\tau > t \,|\, \mathcal{F}_T) = P(\tau > t \,|\, \mathcal{F}_t)$, for all $t \in [0, T]$, implies that all $(P, \mathbb{F})$-martingales are also $(P, \mathbb{G})$-martingales. In particular, $W = (W_t)_{0 \leq t \leq T}$ is a Brownian motion with respect to both $\mathbb{F}$ and $\mathbb{G}$. This important fact will be used in the following without further mention.

$1, \dots, m\}$, for some fixed $m \in \{1, \dots, d-1\}$:

$$dX_t = (AX_t + b)\, dt + \Sigma\sqrt{R_t}\, dW_t\,, \qquad X_0 = \left(v_0, Y_0^\top, \log S_0\right)^\top = \bar{x} \in \mathbb{R}^m_{++} \times \mathbb{R}^{d-m}\,, \tag{3.2.1}$$

where $(A, b, \Sigma) \in \mathbb{R}^{d\times d} \times \mathbb{R}^d \times \mathbb{R}^{d\times d}$ and $R_t$ is a diagonal $(d \times d)$-matrix with elements $R_t^{i,i} = \alpha_i + \beta_i^\top X_t$, for all $t \in [0, T]$, with $\alpha := (\alpha_1, \dots, \alpha_d)^\top \in \mathbb{R}^d_+$ and $\beta := (\beta_1, \dots, \beta_d) \in \mathbb{R}^{d\times d}_+$.

Following the notation adopted in Chapt. 10 of [Fil09], for a given $m \in \{1, \dots, d-1\}$, we define the sets $I := \{1, \dots, m\}$, $J := \{m+1, \dots, d\}$ and $D := I \cup J = \{1, \dots, d\}$. Intuitively, the set $I$ collects the indices of the first $m$ elements of the $\mathbb{R}^d$-valued process $X$, while the set $J$ collects the remaining ones. In order to guarantee the existence of a strong solution to the SDE (3.2.1), we introduce the following assumption.

**Assumption 3.2.1.** *The parameters $A, b, \Sigma, \alpha, \beta$ satisfy the following conditions:*

1. *$b_i \geq (\Sigma_{i,i})^2 \beta_{i,i}/2$ for all $i \in I$;*
2. *$A_{i,j} = 0$ for all $i \in I$ and $j \in J$ and $A_{i,j} \geq 0$ for all $i, j \in I$ with $i \neq j$;*
3. *$\Sigma_{i,j} = 0$ for all $i \in I$ and $j \in D$ with $j \neq i$;*
4. *$\beta_{j,i} = 0$ for all $i \in D$ and $j \in J$, $\beta_{i,i} > 0$ for all $i \in I$ and $\beta_{i,j} = 0$ for all $i, j \in I$ with $i \neq j$;*
5. *$\alpha_i = 0$ for all $i \in I$ and $\alpha_j > -\sum_{i\in I} \beta_{i,j}$ for all $j \in J$.*

For any $\bar{x} \in \mathbb{R}^m_{++} \times \mathbb{R}^{d-m}$, Assumption 3.2.1 ensures the existence of a unique strong solution $X = (X_t)_{0\leq t\leq T}$ to the SDE (3.2.1) on the filtered probability space $(\Omega, \mathcal{G}, \mathbb{F}, P)$ such that $X_0 = \bar{x}$ and $X_t \in \mathbb{R}^m_{++} \times \mathbb{R}^{d-m}$ $P$-a.s. for all $t \in [0, T]$. Indeed, the same arguments used in the proof of Lemma 10.6 of [Fil09] give the existence of a unique strong solution $X = (X_t)_{0\leq t\leq T}$ on $\mathbb{R}^m_+ \times \mathbb{R}^{d-m}$, while Lemma A.3 of [DK96] together with Ex. 10.12 of [Fil09] implies that $X$ actually takes values in $\mathbb{R}^m_{++} \times \mathbb{R}^{d-m}$. Due to conditions (iv)-(v) of Assumption 3.2.1, this also implies that the matrix $R_t$ is positive definite for all $t \in [0, T]$. In the remaining part of the paper, we shall always assume that Assumption 3.2.1 is satisfied without further mention.

**Remark 3.2.2.** *The parameter restrictions imposed by Assumption 3.2.1 bear resemblance to the canonical representation of [DS00]. However, we do not require the matrix $\Sigma$ to be diagonal, since this may lead to unnecessary restrictions on the model if $2 \leq m \leq d-2$, as pointed out in [CFK10].*

The following proposition describes the dynamics of the defaultable stock price process $S$.

**Proposition 3.2.3.** *The process $S = (S_t)_{0\leq t\leq T}$ satisfies the following SDE on $(\Omega, \mathcal{G}, \mathbb{G}, P)$:*

$$dS_t = S_{t-}\left(\bar{s} + \mu_1 \log S_{t-} + \mu_2 v_t + \sum_{i=1}^{d-2} \eta_i Y_t^i + \sum_{i=1}^{m-1} \bar{\eta}_i Y_t^i\right) dt + S_{t-}\, \sigma\sqrt{v_t}\, dW_t^1 + S_{t-} \sum_{i=2}^{d} \Sigma_{d,i} \sqrt{R_t^{i,i}}\, dW_t^i - S_{t-}\, d\mathbf{1}_{\{\tau\leq t\}} \tag{3.2.2}$$

*with the convention $S_{t-} \log S_{t-} = 0$ on $\{\tau \leq t\}$ and where*

$$\bar{s} := b_d + \frac{1}{2} \sum_{k=m+1}^{d} (\Sigma_{d,k})^2\, \alpha_k\,, \qquad \mu_1 := A_{d,d}\,, \qquad \mu_2 := A_{d,1} + \frac{1}{2}\,(\Sigma_{d,1})^2\, \beta_{11} + \frac{1}{2} \sum_{k=m+1}^{d} (\Sigma_{d,k})^2\, \beta_{1,k}\,,$$

$$\eta_i := A_{d,i+1}, \qquad \sigma := \Sigma_{d,1}\sqrt{\beta_{1,1}}\,, \qquad \bar{\eta}_i := \frac{1}{2}\,(\Sigma_{d,i+1})^2\, \beta_{i+1,i+1} + \frac{1}{2} \sum_{k=m+1}^{d} (\Sigma_{d,k})^2\, \beta_{i+1,k}\,.$$

*Proof.* Observe first that $dS_t = \mathbf{1}_{\{\tau>t-\}}\tilde{S}_{t-}\left(dL_t + d\langle L\rangle_t/2\right) - \tilde{S}_{t-}\, d\mathbf{1}_{\{\tau\leq t\}}$, due to Itô's formula and integration by parts. Equation (3.2.2) then follows from (3.2.1) together with Assumption 3.2.1 by means of simple computations. □

**Remark 3.2.4.** *The defaultable price process $S$ has a rich structure, influenced by the factor process $Y$ in both the drift and diffusion terms. Furthermore, there are three levels of dependence between $S$ and the stochastic volatility $v$:* (1) *a direct interaction, since $v$ explicitly appears in the dynamics of $S$;* (2) *a semi-direct interaction, since the Brownian motion $W^1$ driving the process $v$ is also one of the drivers of $S$;* (3) *an indirect interaction, since $S$ and $v$ both depend on the factor process $Y$.*

To complete the description of the modeling framework, we specify as follows the $P$-intensity process $\lambda^P = (\lambda^P_t)_{0\leq t\leq T}$ and the risk-free interest rate process $r = (r_t)_{0\leq t\leq T}$:

$$\lambda^P_t := \bar{\lambda}^P + (\Lambda^P)^\top X_t\,, \qquad r_t := \bar{r} + \Upsilon^\top X_t\,, \qquad \text{for all } t\in[0,T]\,, \tag{3.2.3}$$

where $\bar{\lambda}^P, \bar{r} \in \mathbb{R}_+$ and $\Lambda^P, \Upsilon \in \mathbb{R}^m_+ \times \{0\}^{d-m}$, with $\bar{\lambda}^P + \sum_{i=1}^m \Lambda^P_i > 0$ and $\bar{r} + \sum_{i=1}^m \Upsilon_i > 0$. This ensures that the $P$-intensity and the risk-free rate are correlated and strictly positive, since 0 is an unattainable boundary for $X^i$, $\forall i \in I$. Furthermore, the linear structure (3.2.3) permits to obtain analytically tractable formulae for several quantities of interest, as shown in Sections 3.4-3.5. The specification (3.2.3) allows for a direct dependence of $\lambda^P$ on the stochastic volatility $v$, this feature being consistent with several empirical observations (see e.g. [CT03, CDMW08]). Furthermore, the defaultable price process $S$ and the $P$-intensity $\lambda^P$ are linked through the common factor process $Y$. Finally, we want to remark that the proposed modeling framework generalises to a defaultable setting several stochastic volatility models considered in the literature. For instance, defaultable versions of the models considered in [AR09, CS08] and Sect. 4.3 of [DPS00] can be easily recovered within our general setting. As an example, in Section 3.6 we shall study in detail an extended defaultable version of the Heston [Hes93] stochastic volatility model.

**Remark 3.2.5.** *We want to point out that multifactor stochastic volatility models are naturally embedded within our modeling framework. Indeed, the first $m-1$ components of the factor process $Y$ are strictly positive processes and can be interpreted as additional stochastic volatility factors, as can also be seen from equation* (3.2.2)*. For instance, in the case $d = 3$ and $m = 2$, we can easily obtain (a defaultable version of) the two-factor stochastic volatility model proposed by [CHJ09].*

**Remark 3.2.6.** *The modeling framework described in this section can be easily extended to the case of $M > 1$ defaultable firms if we suppose that their random default times $\{\tau_1, \ldots, \tau_M\}$ are $\mathbb{F}$-conditionally independent (see [MFE05], Sect. 9.6). In that case, the process $L$ is an $\mathbb{R}^M$-valued process representing the logarithm of the pre-default values of the $M$ stock prices (and, similarly, the process $v$ representing the stochastic volatilities of the $M$ stocks is also $\mathbb{R}^M$-valued). If the processes $L$, $v$ and the factor process $Y$ are jointly modeled as an affine diffusion of the type* (3.2.1) *and if the $P$-intensity processes $\lambda^{P,\ell} = (\lambda^{P,\ell}_t)_{0\leq t\leq T}$, for $\ell = 1, \ldots, M$, are of the form* (3.2.3)*, then the multi-firm extension of the model is still fully analytically tractable. This generalization can be of particular interest in view of portfolio credit risk modeling.*

## 3.3 Equivalent changes of measure which preserve the affine structure

The modeling framework introduced in Section 3.2 has been formulated entirely with respect to the physical probability measure $P$. However, since we aim at dealing with pricing as well as risk management applications, we need to study the structure of the model under a suitable *risk-neutral* probability measure, formally defined as a probability measure $Q \sim P$ on $(\Omega, \mathcal{G})$ such that the discounted defaultable price process $\exp\left(-\int_0^{\cdot} r_u du\right) S$ is a $(Q, \mathbb{G})$-local martingale[2].

It is important to be aware of the fact that most of the appealing features of the framework described in Section 3.2 may be lost after a change of measure. Aiming at a model which is analytically tractable under both the physical and a risk-neutral measure, we shall consider all risk-neutral measures $Q$ which *preserve the affine structure of* $(X, \tau)$, in the sense of the following Definition.

**Definition 3.3.1.** *Let $Q$ be a probability measure on $(\Omega, \mathcal{G})$ with $Q \sim P$. We say that $Q$* preserves the affine structure of $(X, \tau)$ *if the following hold:*

1. *the process $X = (X_t)_{0 \le t \le T}$ satisfies an SDE of the type (3.2.1) on $(\Omega, \mathcal{G}, \mathbb{F}, Q)$ with respect to an $\mathbb{R}^d$-valued $(Q, \mathbb{F})$-Brownian motion $W^Q = (W_t^Q)_{0 \le t \le T}$ and for some parameters $A^Q, b^Q, \Sigma, \alpha$, $\beta$ satisfying Assumption 3.2.1;*

2. *the default time $\tau$ is a $(Q, \mathbb{F})$-doubly stochastic random time with $Q$-intensity $\lambda^Q = (\lambda_t^Q)_{0 \le t \le T}$ of the form $\lambda_t^Q = \bar{\lambda}^Q + (\Lambda^Q)^\top X_t$, for $\bar{\lambda}^Q \in \mathbb{R}_+$ and $\Lambda^Q \in \mathbb{R}_+^m \times \{0\}^{d-m}$ with $\bar{\lambda}^Q + \sum_{i=1}^m \Lambda_i^Q > 0$.*

We denote by $\mathcal{Q}$ the family of all risk-neutral measures which preserve the affine structure of $(X, \tau)$, in the sense of Definition 3.3.1. The next theorem gives a complete characterisation of the family $\mathcal{Q}$. This result follows from a more general one in Chapt. 2 of [Fon12a], but we outline a self-contained proof for the convenience of the reader. We denote by $\mathcal{E}$ the stochastic exponential and by $M = (M_t)_{0 \le t \le T}$ the $(P, \mathbb{G})$-martingale defined by $M_t := \mathbf{1}_{\{\tau \le t\}} - \int_0^{t \wedge \tau} \lambda_u^P du$ (see [BR02], Prop. 5.1.3).

**Theorem 3.3.2.** *Let $Q$ be a probability measure on $(\Omega, \mathcal{G})$. Then we have $Q \in \mathcal{Q}$ if and only if*

$$\frac{dQ}{dP} = \mathcal{E}\left(\int \theta \, dW + \int \gamma \, dM\right)_T = \exp\left(\sum_{i=1}^d \int_0^T \theta_t^i \, dW_t^i - \frac{1}{2}\sum_{i=1}^d \int_0^T (\theta_t^i)^2 dt - \int_0^{\tau \wedge T} \gamma_t \, \lambda_t^P dt\right)\left(1 + \mathbf{1}_{\{\tau \le T\}} \gamma_\tau\right) \tag{3.3.1}$$

*where $\theta = (\theta_t)_{0 \le t \le T}$ and $\gamma = (\gamma_t)_{0 \le t \le T}$ are $\mathbb{F}$-adapted processes of the following form:*

$$\theta_t = \theta(V_t) := R_t^{-1/2}\left(\hat{\theta} + \Theta X_t\right), \qquad \gamma_t = \gamma(V_t) := \frac{\left(\bar{\lambda}^Q - \bar{\lambda}^P\right) + \left(\Lambda^Q - \Lambda^P\right)^\top X_t}{\bar{\lambda}^P + \left(\Lambda^P\right)^\top X_t}, \tag{3.3.2}$$

*for some $\hat{\theta} \in \mathbb{R}^d$ and $\Theta \in \mathbb{R}^{d \times d}$ such that:*

1. *$\sum_{k=1}^d \Sigma_{i,k} \hat{\theta}_k \ge (\Sigma_{i,i})^2 \beta_{i,i}/2 - b_i$ for all $i \in I$;*

[2] Due to the fundamental result of [DS94], this is equivalent to the validity of *No Free Lunch with Vanishing Risk* (NFLVR) condition for the financial market $(S, \mathbb{G})$, being the process $\exp\left(-\int_0^{\cdot} r_u du\right) S$ locally bounded.

2. $\sum_{k=1}^{d} \Sigma_{i,k}\Theta_{k,j} = 0$, *for all* $i \in I$ *and* $j \in J$, *and* $\sum_{k=1}^{d} \Sigma_{i,k}\Theta_{k,j} \geq -A_{i,j}$, *for all* $i, j \in I$ *with* $i \neq j$;

*for some* $\bar{\lambda}^Q \in \mathbb{R}_+$ *and* $\Lambda^Q \in \mathbb{R}_+^m \times \{0\}^{d-m}$ *with* $\bar{\lambda}^Q + \sum_{i=1}^{m} \Lambda_i^Q > 0$ *and if the following equality holds* $P$*-a.s. on* $\{\tau > t\}$, *using the notation introduced in Proposition 3.2.3:*

$$\bar{s} + \mu_1 \log S_{t-} + \left(\mu_2 + \sigma \frac{\theta_t^1}{\sqrt{v_t}}\right) v_t + \sum_{i=1}^{d-2} \eta_i Y_t^i + \sum_{i=1}^{m-1} \bar{\eta}_i Y_t^i + \sum_{i=2}^{d} \Sigma_{d,i} \sqrt{R_t^{i,i}}\, \theta_t^i = r_t + \lambda_t^P (1 + \gamma_t)\,. \quad (3.3.3)$$

*Proof.* Let $\theta = (\theta_t)_{0 \leq t \leq T}$ and $\gamma = (\gamma_t)_{0 \leq t \leq T}$ be two $\mathbb{F}$-adapted processes satisfying (3.3.2). Since $\theta$ and $\gamma$ are continuous functions of $X$ and the process $X$ is continuous, hence locally bounded, the process $Z := \mathcal{E}\left(\int \theta\, dW + \int \gamma\, dM\right)$ is well-defined as a strictly positive $(P, \mathbb{G})$-local martingale and, as a consequence of Fatou's lemma, it is also a $(P, \mathbb{G})$-supermartingale. Moreover, Thm. 2.4 and Remark 2.5 of [CFY05] allow to conclude that $E[Z_T] = 1$, thus implying that $Z$ is a uniformly integrable $(P, \mathbb{G})$-martingale. So, we can define a probability measure $Q$ on $(\Omega, \mathcal{G})$ via (3.3.1). Part (i) of Definition 3.3.1 then follows from Girsanov's theorem together with (3.3.2), while part (ii) follows from Thm. 6.3 of [CJN12], Girsanov's theorem together with (3.3.2) and Prop. 6.2.2 of [BR02]. Finally, the $(Q, \mathbb{G})$-local martingale property of $\exp\left(-\int_0^{\cdot} r_u du\right) S$ easily follows from Girsanov's theorem together with Proposition 3.2.3 and (3.3.3). Conversely, suppose that $Q \in \mathcal{Q}$. The existence of a representation of the form (3.3.1) follows from Cor. 5.2.4 of [BR02], while (3.3.2) and (3.3.3) follow from Girsanov's theorem together with Definition 3.3.1 and Proposition 3.2.3, respectively. □

Note that the process $\gamma = (\gamma_t)_{0 \leq t \leq T}$ introduced in (3.3.2) satisfies $\gamma_t > -1$ $P$-a.s. for all $t \in [0, T]$, due to the restrictions imposed on the parameters $\bar{\lambda}^P$, $\bar{\lambda}^Q$, $\Lambda^P$ and $\Lambda^Q$. In particular, this ensures that, for every probability measure $Q \in \mathcal{Q}$, both the $P$-intensity process $\lambda^P = (\lambda_t^P)_{0 \leq t \leq T}$ and the $Q$-intensity process $\lambda^Q = (\lambda_t^Q)_{0 \leq t \leq T}$ are $P$-a.s. strictly positive.

Due to Theorem 3.3.2, the preservation of the affine structure of $(X, \tau)$ does not prevent the default intensity to change significantly from the physical to a risk-neutral probability measure $Q \in \mathcal{Q}$, due to the presence of the risk premium $\gamma$ (see also the comments below). From the practical perspective, this is an important aspect of our modeling approach, especially in view of the possibility of valuing credit/equity financial derivatives whose payoff also depends on the $P$-intensity of default through, for instance, the rating score attached to a defaultable firm or the corresponding statistical survival/default probability.

**Remark 3.3.3.** *The processes* $\theta = (\theta_t)_{0 \leq t \leq T}$ *and* $\gamma = (\gamma_t)_{0 \leq t \leq T}$ *admit the financial interpretation of* risk premia *(or* market prices of risk*) associated to the randomness generated by the Brownian motion* $W$ *and by the random default time* $\tau$, *respectively. More specifically:*

1. *The process* $\theta = (\theta_t)_{0 \leq t \leq T}$ *represents the risk premium associated to the diffusive risk generated by the Brownian motion* $W$. *Since the stock price, its stochastic volatility, the default intensity and the interest rate all depend on* $W$ *through* $X$, *the risk premium* $\theta$ *can be considered as a market-wide non-diversifiable risk premium*[3].

[3] In the context of default-free term structure modeling, in [CFK07] the authors demonstrate that the specification (3.3.2) has a considerably better fit to market data than the simpler market price of risk specifications traditionally considered in the literature (see e.g. [CW10, DS00, Duf99, Duf02, Hes93]).

2. *The process $\gamma = (\gamma_t)_{0\le t\le T}$ represents the risk premium associated to the default event or, more precisely, the risk premium associated to the idiosyncratic component of the risk generated by the occurrence of the default event (to this effect, see also [CPS09, NEK01] and Sect. 9.3 of [MFE05]).*

*The importance of explicitly distinguishing between $\theta$ and $\gamma$ has been demonstrated in [Dri05]. Assuming $\gamma \equiv 0$ means that the idiosyncratic component of default risk can be diversified away in the market, as explained in [JLY05], and, therefore, market participants do not require a compensation for it. However, the jump-type risk premium can be significant when it is difficult to hedge the risk associated with the timing of the default event of a given firm. Note that, as can be seen from* (3.3.2)*, the risk premia $\theta$ and $\gamma$ both depend on the common driving process $X$.*

Due to Theorem 3.3.2, our modeling framework enjoys full analytical tractability under both the physical measure $P$ and any risk-neutral measure $Q \in \mathcal{Q}$, thus enabling us to efficiently solve risk management as well as a pricing problems, as we are going to show in Sections 3.4-3.5. We close this section with the following fundamental result, which follows from Thm. 10.4 of [Fil09] together with part (i) of Definition 3.3.1, (3.2.1) and Assumption 3.2.1. For $z \in \mathbb{C}^d$ we denote by $\Re(z)$ and $\Im(z)$ the real and imaginary parts of $z$, respectively, and $\mathbb{C}^m_- := \{z \in \mathbb{C}^m : \Re(z) \in \mathbb{R}^m_-\}$. For $Q \in \mathcal{Q} \cup \{P\}$, we denote by $E^Q$ the (conditional) expectation operator under the measure $Q$.

**Proposition 3.3.4.** *For every $Q \in \mathcal{Q} \cup \{P\}$ and for all $z \in \mathbb{C}^m_- \times i\mathbb{R}^{d-m}$, there exists a unique solution $\big(\Phi^Q(\cdot,z), \Psi^Q(\cdot,z)\big) : [0,T] \to \mathbb{C} \times \mathbb{C}^d$ to the following system of Riccati ODEs:*

$$
\begin{aligned}
&\partial_t \Phi^Q(t,z) = (b^Q)^\top \Psi^Q(t,z) + \frac{1}{2}\sum_{k=m+1}^{d} [\Sigma^\top \Psi^Q(t,z)]_k^2\, \alpha_k - \bar\lambda^Q - \bar r\, \mathbf{1}_{Q\neq P}\,,\\
&\Phi^Q(0,z) = 0\,,\\
&\partial_t \Psi_i^Q(t,z) = \sum_{k=1}^{d} A^Q_{k,i}\, \Psi_k^Q(t,z) + \frac{1}{2}[\Sigma^\top \Psi^Q(t,z)]_i^2 \beta_{i,i} + \frac{1}{2}\sum_{k=m+1}^{d} [\Sigma^\top \Psi^Q(t,z)]_k^2\, \beta_{i,k} - \Lambda_i^Q - \Upsilon_i \mathbf{1}_{Q\neq P}\,, \quad \forall i \in I\,,\\
&\partial_t \Psi_j^Q(t,z) = \sum_{k=m+1}^{d} A^Q_{k,j}\, \Psi_k^Q(t,z)\,, \qquad \forall j \in J\,,\\
&\Psi^Q(0,z) = z\,.
\end{aligned}
\tag{3.3.4}
$$

*Furthermore, for any $Q \in \mathcal{Q} \cup \{P\}$, the following holds for all $0 \le t \le u \le T$ and for all $z \in \mathbb{C}^m_- \times i\mathbb{R}^{d-m}$:*

$$
E^Q\left[\exp\left(-\int_t^u (\lambda_s^Q + r_s \mathbf{1}_{Q\neq P})\, ds + z^\top X_u\right)\middle|\, \mathcal{F}_t\right] = \exp\left(\Phi^Q(u-t,z) + \Psi^Q(u-t,z)^\top X_t\right). \tag{3.3.5}
$$

## 3.4 Risk management applications

Many quantities of interest in view of risk management applications can be computed as conditional expectations under the physical measure $P$. As a first and basic application, let us compute the $\mathcal{G}_t$-conditional survival probability of the defaultable firm up to the final horizon $T$. We denote by $\Phi^P(\cdot,\cdot)$ and $\Psi^P(\cdot,\cdot)$ the solutions to the Riccati ODEs (3.3.4) with $Q = P$.

**Proposition 3.4.1.** *For any $t \in [0,T]$, the following holds:*

$$P(\tau > T \,|\, \mathcal{G}_t) = \mathbf{1}_{\{\tau>t\}} \exp\Big(\Phi^P(T-t,0) + \Psi^P(T-t,0)^\top X_t\Big). \tag{3.4.1}$$

*Proof.* Cor. 5.1.1 of [BR02] implies that $P(\tau > T \,|\, \mathcal{G}_t) = \mathbf{1}_{\{\tau>t\}} E[\exp(-\int_t^T \lambda_s^P ds) \,|\, \mathcal{F}_t]$. The result then follows by applying (3.3.5) with $Q = P$, $z = 0$ and $u = T$. □

As can be easily checked from (3.3.4), the right-hand side of (3.4.1) only depends on $\{X^i : i \in I\}$, i.e., on the components of the process $X$ on which the $P$-intensity $\lambda^P$ depends. For computing conditional expectations (under the measure $P$) of more general quantities needed for risk management purposes, it turns out to be convenient to introduce the *$T$-survival measure* $P^T \sim P$ on $(\Omega, \mathcal{G})$ defined by $dP^T/dP := \exp\big(-\int_0^T \lambda_t^P dt\big)/E\big[\exp\big(-\int_0^T \lambda_t^P dt\big)\big]$.

**Lemma 3.4.2.** *For any random variable $F \in L^1(P, \mathcal{F}_T)$ and for any $t \in [0,T]$ the following holds:*

$$E\left[F\,\mathbf{1}_{\{\tau>T\}} \,|\, \mathcal{G}_t\right] = P(\tau > T \,|\, \mathcal{G}_t)\, E^{P^T}[F \,|\, \mathcal{F}_t]. \tag{3.4.2}$$

*Proof.* Cor. 5.1.1 of [BR02] implies that $E\left[F\,\mathbf{1}_{\{\tau>T\}} \,|\, \mathcal{G}_t\right] = \mathbf{1}_{\{\tau>t\}} E\big[F \exp\big(-\int_t^T \lambda_s^P ds\big) \,|\, \mathcal{F}_t\big]$. Equation (3.4.2) then follows by using the definition of the measure $P^T$ together with the conditional version of Bayes' formula (see e.g. [Fil09], Ex. 4.9). □

Lemma 3.4.2 shows that the computation of the $\mathcal{G}_t$-conditional expectation of an $\mathcal{F}_T$-measurable random variable $F$ in the case of survival up to time $T$ reduces to the computation of the $\mathcal{F}_t$-conditional expectation of $F$ under the $T$-survival measure $P^T$, the term $P(\tau > T \,|\, \mathcal{G}_t)$ being given as in (3.4.1). As can be seen from equation (3.4.2), the $T$-survival measure $P^T$ allows to decompose the conditional expectation of the product $F\,\mathbf{1}_{\{\tau>T\}}$ into the product of two conditional expectations. Note also that, from the point of view of practical applications, the term $P(\tau > T \,|\, \mathcal{G}_t)$ does not necessarily have to be computed, since it can often be deduced from publicly available data, notably from rating transition matrices published by rating agencies. Furthermore, as shown in the next lemma, the $\mathcal{F}_t$-conditional characteristic function of the vector $X_T$ under the $T$-survival measure $P^T$ can be computed in closed form.

**Lemma 3.4.3.** *For any $z \in i\mathbb{R}^d$ and for any $t \in [0,T]$ the following holds:*

$$\varphi_t^{P^T}(z) := E^{P^T}\left[e^{z^\top X_T} \,|\, \mathcal{F}_t\right] = \exp\Big(\Phi^P(T-t,z) - \Phi^P(T-t,0) + \big(\Psi^P(T-t,z) - \Psi^P(T-t,0)\big)^\top X_t\Big). \tag{3.4.3}$$

*Proof.* The definition of the measure $P^T$ together with the conditional version of Bayes' formula gives $E^{P^T}\left[e^{z^\top X_T} \,|\, \mathcal{F}_t\right] = E\big[\exp\big(-\int_t^T \lambda_s^P ds + z^\top X_T\big) \,|\, \mathcal{F}_t\big]/E\big[\exp\big(-\int_t^T \lambda_s^P ds\big) \,|\, \mathcal{F}_t\big]$. By applying (3.3.5) with $Q = P$, $u = T$ and $z \in i\mathbb{R}^d$ ($z = 0$, resp.) to the numerator (to the denominator, resp.), we then obtain equation (3.4.3). □

Due to Lemma 3.4.2 and Lemma 3.4.3, we can compute the $\mathcal{G}_t$-conditional expectation (under the physical probability measure $P$) of arbitrary functions of the random vector $X_T$ in the case of survival by relying on well-known Fourier inversion techniques. As an example, we can explicitly compute quantiles of the $\mathcal{G}_t$-conditional distribution of the defaultable price $S_T$ in the case of survival. This is crucial for the computation of *Value-at-Risk* and related risk measures.

**Proposition 3.4.4.** *For any $x \in (0,\infty)$ and for any $t \in [0,T]$ the following holds:*

$$P(S_T \le x, \tau > T \,|\, \mathcal{G}_t) = P\left(\tau > T \,|\, \mathcal{G}_t\right)\left(\frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{\Im\big(e^{-iy\log x}\,\varphi_t^{P^T}(0,\dots,0,iy)\big)}{y}\,dy\right) \tag{3.4.4}$$

*where $P(\tau > T \,|\, \mathcal{G}_t)$ and $\varphi_t^{P^T}(\cdot)$ are explicitly given in* (3.4.1) *and Lemma 3.4.3, respectively.*

*Proof.* Note that $P\left(S_T \le x, \tau > T|\mathcal{G}_t\right) = P\left(L_T \le \log x, \tau > T|\mathcal{G}_t\right) = P\left(\tau > T|\mathcal{G}_t\right) P^T(L_T \le \log x|\mathcal{F}_t)$, where the second equality follows from Lemma 3.4.2. Equation (3.4.4) then follows from standard Fourier inversion techniques (see e.g. [Fon12a], Prop. 2.5.12, and [Pao07], Sect. 1.2.6). □

## 3.5 Valuation of default-sensitive payoffs and defaultable options

Throughout this section, we fix an element $Q \in \mathcal{Q}$. For the purpose of valuing default-sensitive payoffs, the *u-survival risk-neutral measure* $Q^u$, for $u \in [0,T]$, turns out to be quite useful. The measure $Q^u$ is defined by $dQ^u/dQ = \exp\big(-\int_0^u (r_s + \lambda_s^Q)\,ds\big)/E^Q\big[\exp\big(-\int_0^u (r_s+\lambda_s^Q)\,ds\big)\big]$. For $u = T$, the measure $Q^T$ bears resemblance to the *T-survival measure* $P^T$ introduced in Section 3.4, except that $Q^T$ is defined with respect to some $Q \in \mathcal{Q}$ and the density $dQ^T/dQ$ also involves the risk-free interest rate besides the $Q$-intensity $\lambda^Q$ (compare also with [BR02], Def. 15.2.2). Following the same logic of Section 3.4, we show that many pricing problems can be simplified by shifting to the measure $Q^u$, for some $u \in [0,T]$. As a preliminary, let us compute the arbitrage-free price $\Pi(t,T)$ of a *zero-coupon defaultable bond*. We denote by $\Phi^Q(\cdot,\cdot)$ and $\Psi^Q(\cdot,\cdot)$ the solutions to the Riccati ODEs (3.3.4). The proof of the following lemma is completely analogous to that of Proposition 3.4.1 but we include it for the convenience of the reader.

**Lemma 3.5.1.** *For any $t \in [0,T]$ the following holds:*

$$\Pi(t,T) = \mathbf{1}_{\{\tau>t\}} \exp\Big(\Phi^Q(T-t,0) + \Psi^Q(T-t,0)^\top X_t\Big). \tag{3.5.1}$$

*Proof.* Note first that $\Pi(t,T) = E^Q\big[\exp(-\int_t^T r_s ds)\mathbf{1}_{\{\tau>T\}}|\mathcal{G}_t\big] = \mathbf{1}_{\{\tau>t\}} E^Q\big[\exp(-\int_t^T (r_s+\lambda_s^Q)\,ds)|\mathcal{F}_t\big]$, where the second equality follows from Thm. 9.23 of [MFE05]. Equation (3.5.1) then follows from Proposition 3.3.4 with $u = T$ and $z = 0$. □

Of course, coupon-bearing corporate bonds can be valued as linear combinations of zero-coupon defaultable bonds (see [BR02], Sect. 1.1.5). More generally, most default-sensitive payoffs can be decomposed into linear combinations of *zero-recovery* and *pure recovery* payments, the latter being paid only in the case of default, see e.g. Sect. 9.4 of [MFE05]. The next proposition provides general valuation formulas for zero-recovery and pure recovery payments.

**Proposition 3.5.2.** *For any $t \in [0,T]$ and for any measurable function $G : \mathbb{R}^m_{++} \times \mathbb{R}^{d-m} \to \mathbb{R}_+$ the following hold:*

$$E^Q\Big[e^{-\int_t^T r_s ds} G(X_T)\mathbf{1}_{\{\tau>T\}}\big|\,\mathcal{G}_t\Big] = \Pi(t,T)\,E^{Q^T}[G(X_T)\,|\,\mathcal{F}_t\,], \tag{3.5.2}$$

$$\mathbf{1}_{\{\tau>t\}} E^Q\Big[e^{-\int_t^\tau r_s ds} G(X_\tau)\mathbf{1}_{\{\tau\le T\}}\big|\,\mathcal{G}_t\Big] = \int_t^T \Pi(t,u)\,E^{Q^u}[\lambda_u^Q\,G(X_u)\,|\,\mathcal{F}_t\,]\,du\,. \tag{3.5.3}$$

*Proof.* Note first that, due to Thm. 9.23 of [MFE05], we can write:

$$E^Q\left[e^{-\int_t^T r_s ds} G(X_T)\mathbf{1}_{\{\tau>T\}}\big|\,\mathcal{G}_t\right] = \mathbf{1}_{\{\tau>t\}} E^Q\left[e^{-\int_t^T (r_s+\lambda_s^Q)\,ds} G(X_T)\,\big|\,\mathcal{F}_t\right],$$

$$\mathbf{1}_{\{\tau>t\}} E^Q\left[e^{-\int_t^\tau r_s ds} G(X_\tau)\mathbf{1}_{\{\tau\le T\}}\big|\,\mathcal{G}_t\right] = \mathbf{1}_{\{\tau>t\}} E^Q\left[\int_t^T e^{-\int_t^u (r_s+\lambda_s^Q)\,ds}\lambda_u^Q\, G(X_u)\,du\,\Big|\,\mathcal{F}_t\right].$$

Equations (3.5.2)-(3.5.3) then follow by using the definition of the measure $Q^u$, for $u \in [t,T]$, together with the conditional version of Bayes' formula and also, for (3.5.3), with Tonelli's theorem. □

We want to point out that, in view of practical applications, the quantities $\Pi(t,u)$, for $u \in [t,T]$, appearing in equations (3.5.2)-(3.5.3) do not necessarily have to computed, since they can be directly observed on the corporate bond market. This fact represents one of the main advantages of using risk-neutral survival measures for the valuation of defaultable claims (see also [Sch03] for a related discussion and other applications of survival measures to credit risk modeling).

As an application of Lemma 3.5.1 and Proposition 3.5.2, we compute the fair spread $\pi^{\,\mathrm{CDS}}(t,T)$, at time $t \in [0,T]$, of a *Credit Default Swap (CDS)* which exchanges a fixed stream of payments in arrears equal to $\pi^{\,\mathrm{CDS}}(t,T)$ at the dates $\{t_1,\ldots,t_N\}$, with $t \le t_1 < \ldots < t_N \le T$, (*premium payment leg*) against the payment at the default time $\tau$ (if the latter happens before the maturity $T$) of a default protection term equal to a fraction $\delta \in (0,1)$ of the unitary nominal value (*default payment leg*), see e.g. Sect. 9.3 of [MFE05].

**Corollary 3.5.3.** *For any $t \in [0,T]$ and $t_0 := t \le t_1 < \ldots < t_N \le T$, the following holds on $\{\tau > t\}$:*

$$\pi^{\,CDS}(t,T) = \delta\,\frac{\int_t^T \Pi(t,u)\,E^{Q^u}[\lambda_u^Q\,|\,\mathcal{F}_t]\,du}{\sum_{k=1}^N (t_k - t_{k-1})\,\Pi(t,t_k)}. \tag{3.5.4}$$

*Proof.* Due to Lemma 3.5.1, the arbitrage-free price of the premium payment leg is given by:

$$\pi^{\,\mathrm{CDS}}(t,T)\sum_{k=1}^N (t_k - t_{k-1}) E^Q\left[e^{-\int_t^{t_k} r_s ds}\mathbf{1}_{\{\tau>t_k\}}\big|\,\mathcal{G}_t\right] = \pi^{\,\mathrm{CDS}}(t,T)\sum_{k=1}^N (t_k - t_{k-1})\,\Pi(t,t_k).$$

On the other hand, due to equation (3.5.3), the arbitrage-free price of the default payment leg is equal to:

$$\mathbf{1}_{\{\tau>t\}} E^Q\left[e^{-\int_t^\tau r_s ds}\delta\,\mathbf{1}_{\{\tau\le T\}}\big|\,\mathcal{G}_t\right] = \delta\int_t^T \Pi(t,u)\,E^{Q^u}[\lambda_u^Q\,|\,\mathcal{F}_t]\,du.$$

Equation (3.5.4) then follows by recalling that, by definition, the fair spread $\pi^{\,\mathrm{CDS}}(t,T)$ is the premium payment which equates the values of the two legs of the CDS (see [MFE05], Sect. 9.3). □

For $0 \le t \le u \le T$, the next lemma gives the explicit expression of the $\mathcal{F}_t$-conditional characteristic function $\varphi_t^{Q^u}$ of the random vector $X_u$ under the $u$-survival risk-neutral measure $Q^u$. Its proof follows from (3.3.5) and, being analogous to that of Lemma 3.4.3, is omitted.

**Lemma 3.5.4.** *For any $0 \le t \le u \le T$ and for any $z \in i\mathbb{R}^d$ the following holds:*

$$\varphi_t^{Q^u}(z) := E^{Q^u}\left[e^{z^\top X_u}\,|\,\mathcal{F}_t\right] = \exp\Big(\Phi^Q(u-t,z) - \Phi^Q(u-t,0) + \big(\Psi^Q(u-t,z) - \Psi^Q(u-t,0)\big)^\top X_t\Big). \tag{3.5.5}$$

By combining Proposition 3.5.2 with Lemma 3.5.1 and Lemma 3.5.4 and using well-known Fourier inversion techniques, we can obtain semi-explicit formulas for a wide range of default-sensitive as well as equity/credit hybrid products. In particular, we now derive valuation formulas for *Call* and *Put* options (issued by a default-free third party) written on the defaultable stock $S$. We denote by $\Pi_{rf}(t,T) := E^Q\big[\exp(-\int_t^T r_s ds)\,|\,\mathcal{G}_t\big] = E^Q\big[\exp(-\int_t^T r_s ds)\,|\,\mathcal{F}_t\big]$ the arbitrage-free price at time $t \in [0,T]$ of a zero-coupon default-free bond.

**Corollary 3.5.5.** *For any $t \in [0,T]$ and for any strike price $K > 0$ the following hold:*

$$C_K(t,T) := E^Q\Big[e^{-\int_t^T r_s ds}(S_T - K)^+\big|\,\mathcal{G}_t\Big] = \frac{\Pi(t,T)}{2\pi}\int_{-\infty}^{+\infty}\varphi_t^{Q^T}(0,\ldots,0,w+iu)\frac{K^{-(w-1+iu)}}{(w+iu)(w-1+iu)}\,du\,, \tag{3.5.6}$$

$$\begin{aligned} P_K(t,T) &:= E^Q\Big[e^{-\int_t^T r_s ds}(K - S_T)^+\big|\,\mathcal{G}_t\Big] \\ &= K\left(\Pi_{rf}(t,T) - \Pi(t,T)\right) + \frac{\Pi(t,T)}{2\pi}\int_{-\infty}^{+\infty}\varphi_t^{Q^T}(0,\ldots,0,y+iu)\frac{K^{-(y-1+iu)}}{(y+iu)(y-1+iu)}\,du\,. \end{aligned} \tag{3.5.7}$$

*for some $w > 1$ and $y < 0$ such that the system of Riccati ODEs* (3.3.4) *has a unique solution for the initial conditions $z = (0,\ldots,0,w)^\top$ and $z = (0,\ldots,0,y)^\top$.*

*Proof.* Observe first that:

$$E^Q\Big[e^{-\int_t^T r_s ds}(S_T - K)^+\big|\,\mathcal{G}_t\Big] = E^Q\Big[e^{-\int_t^T r_s ds}\big(\tilde{S}_T - K\big)^+\mathbf{1}_{\{\tau>T\}}\big|\,\mathcal{G}_t\Big] = \Pi(t,T)E^{Q^T}\Big[\big(e^{L_T} - K\big)^+\,\big|\,\mathcal{F}_t\Big]$$

where the second equality follows from (3.5.2). As in [CM99] and [Fil09], Lemma 10.2, it can be shown that:

$$(e^x - K)^+ = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{(w+iu)x}\frac{K^{-(w-1+iu)}}{(w+iu)(w-1+iu)}\,du$$

for some $w > 1$. Equation (3.5.6) then follows by Fubini's theorem (see Cor. 2.5.21 of [Fon12a] for more details). Equation (3.5.7) follows by an analogous computation once we observe that:

$$E^Q\Big[e^{-\int_t^T r_s ds}(K - S_T)^+\big|\,\mathcal{G}_t\Big] = E^Q\Big[e^{-\int_t^T r_s ds}\big(K - \tilde{S}_T\big)^+\mathbf{1}_{\{\tau>T\}}\big|\,\mathcal{G}_t\Big] + KE^Q\Big[e^{-\int_t^T r_s ds}\left(1 - \mathbf{1}_{\{\tau>T\}}\right)\big|\,\mathcal{G}_t\Big].$$

□

If the discounted defaultable price process $\exp(-\int_0^\cdot r_u du)\,S$ is not only a $(Q,\mathbb{G})$-local martingale but also a true $(Q,\mathbb{G})$-martingale (this is for instance the case for the Heston with jump-to-default model considered in Section 3.6; see [Fon12a], Prop. 2.4.7), then the classical *put-call parity* relation holds between the arbitrage-free prices of Call and Put options (issued by a default-free third party) with the same maturity $T$ and strike price $K$, written on the defaultable stock $S$:

$$C_K(t,T) - P_K(t,T) = E^Q\Big[e^{-\int_t^T r_s ds}S_T\big|\,\mathcal{G}_t\Big] - KE^Q\Big[e^{-\int_t^T r_s ds}\big|\,\mathcal{G}_t\Big] = S_t - K\,\Pi_{rf}(t,T), \quad \forall t \in [0,T]. \tag{3.5.8}$$

Note that, if the options are issued by an entity defaulting at $\tau$ (for instance, the defaultable firm itself), then the put-call parity relation (3.5.8) still holds if the default-free bond $\Pi_{rf}(t,T)$ is replaced with the defaultable bond $\Pi(t,T)$.

## 3.6 An example: the Heston with jump-to-default model

In this section, we illustrate some of the essential features of the proposed modeling framework within a simple example, which corresponds to a generalisation of the stochastic volatility model introduced by Heston [Hes93], here extended by allowing the stock price process to be killed by a *jump-to-default* event, in the spirit of [CS08].

### 3.6.1 The model

Using the notations introduced in Section 3.2, we let $d = 3$ and consider the following specification:

$$A = \begin{pmatrix} -k & 0 & 0 \\ 0 & -k_0 & 0 \\ -1/2 & 0 & 0 \end{pmatrix} \qquad b = \begin{pmatrix} k\hat{v} \\ k_0\hat{y} \\ \mu \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \bar{\sigma} & 0 & 0 \\ 0 & \sigma_0 & 0 \\ \rho & 0 & \sqrt{1-\rho^2} \end{pmatrix} \qquad R_t = \begin{pmatrix} v_t & 0 & 0 \\ 0 & Y_t & 0 \\ 0 & 0 & v_t \end{pmatrix} \tag{3.6.1}$$

with $k\hat{v} \geq \bar{\sigma}^2/2$, $k_0\hat{y} \geq \sigma_0^2/2$ and $\rho \in [-1, 1]$. The $P$-intensity $(\lambda_t^P)_{0\leq t\leq T}$ is specified as in (3.2.3), i.e., we have $\lambda_t^P = \bar{\lambda}^P + \Lambda_1^P v_t + \Lambda_2^P Y_t$, for some $\bar{\lambda}^P, \Lambda_1^P, \Lambda_2^P \in \mathbb{R}_+$ with $\bar{\lambda}^P + \Lambda_1^P + \Lambda_2^P > 0$. For simplicity, we assume that $r_t = \bar{r} \in \mathbb{R}_+$ for all $t \in [0, T]$. Note that this specification extends the Heston jump-to-default model considered in [CS08] by allowing $\lambda_t^P$ to depend on $v_t$ and on the additional stochastic factor $Y_t$. It can be easily checked that the specification (3.6.1) satisfies Assumption 3.2.1 and, due to Proposition 3.2.3, the defaultable stock price process $S = (S_t)_{0\leq t\leq T}$ has the following dynamics:

$$dS_t = S_{t-}\left(\mu - \lambda_t^P\right)dt + S_{t-}\sqrt{v_t}\left(\rho\, dW_t^1 + \sqrt{1-\rho^2}\, dW_t^3\right) - S_{t-}\, dM_t \tag{3.6.2}$$

where $M = (M_t)_{0\leq t\leq T}$ is the $(P, \mathbb{G})$-martingale defined by $M_t := \mathbf{1}_{\{\tau\leq t\}} - \int_0^{t\wedge\tau} \lambda_u^P du$. We also have:

$$\begin{aligned} dv_t &= k(\hat{v} - v_t)dt + \bar{\sigma}\sqrt{v_t}\, dW_t^1\,, \\ dY_t &= k_0(\hat{y} - Y_t)dt + \sigma_0\sqrt{Y_t}\, dW_t^2\,. \end{aligned} \tag{3.6.3}$$

### 3.6.2 Risk-neutral measures which preserve the Heston with jump-to-default structure

By relying on Theorem 3.3.2, we now characterise the family of all risk-neutral measures $Q \in \mathcal{Q}$ which *preserve the Heston with jump-to-default structure*, namely all risk-neutral measures $Q \in \mathcal{Q}$ which leave unchanged the structure of the SDEs (3.6.2)-(3.6.3) (compare also with [Fon12a], Sect. 2.4.1).

**Lemma 3.6.1.** *A risk-neutral measure $Q \in \mathcal{Q}$ preserves the Heston with jump-to-default structure if and only if $dQ/dP$ admits the representation* (3.3.1) *for some $\mathbb{F}$-adapted processes $\theta = (\theta_t)_{0\leq t\leq T}$ and $\gamma = (\gamma_t)_{0\leq t\leq T}$ of the form* (3.3.2) *with $\bar{\theta} \in \mathbb{R}^3$ and $\Theta \in \mathbb{R}^{3\times 3}$ satisfying the following restrictions:*

$$\hat{\theta} = \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \frac{\bar{r}+\bar{\lambda}^Q-\mu-\rho\,\hat{\theta}_1}{\sqrt{1-\rho^2}} \end{pmatrix} \qquad \Theta = \begin{pmatrix} \Theta_{1,1} & 0 & 0 \\ 0 & \Theta_{2,2} & 0 \\ \frac{\Lambda_1^Q-\rho\,\Theta_{1,1}}{\sqrt{1-\rho^2}} & \frac{\Lambda_2^Q}{\sqrt{1-\rho^2}} & 0 \end{pmatrix} \tag{3.6.4}$$

*with $\hat{\theta}_1 \geq \bar{\sigma}/2 - k\hat{v}/\bar{\sigma}$ and $\hat{\theta}_2 \geq \sigma_0/2 - k_0\hat{y}/\sigma_0$.*

*Proof.* The result follows from conditions (3.3.2)-(3.3.3) of Theorem 3.3.2, noting that the preservation of the Heston with jump-to-default structure consists in the additional restriction $\Theta_{1,2} = \Theta_{2,1} = 0$. □

**Remark 3.6.2.** *The parameter restrictions of Lemma 3.6.1 are significantly weaker than typical parameter restrictions found in the literature. For instance, let us consider the simpler default-free case (i.e., $\tau = +\infty$ P-a.s.) without the additional stochastic factor $Y$. In that case, the model* (3.6.1)-(3.6.3) *reduces to the classical (default-free) Heston [Hes93] stochastic volatility model. In their analysis of the existence of risk-neutral measures in stochastic volatility models, [WH06] show that there exists a risk-neutral measure $Q$ (preserving the Heston structure) if $\hat{\theta}_1 = 0$ and $\Theta_{1,1}$ satisfies $\Theta_{1,1} \geq -k/\bar{\sigma}$ (see [WH06], Thm. 3.5). In Lemma 3.6.1, we show that such a risk-neutral measure exists without any restriction on $\Theta_{1,1}$ and also for non-trivial values of $\hat{\theta}_1$.*

The main benefit of working with risk-neutral measures which preserve the Heston with jump-to-default structure consists in the possibility of obtaining closed-form solutions to the system of Riccati ODEs (3.3.4), as shown in the next lemma (see also Remark 3.6.4), which follows from Lemma 10.12 of [Fil09] by means of simple (but tedious and, hence, omitted) computations.

**Lemma 3.6.3.** *Let $Q \in \mathcal{Q}$ be a risk-neutral measure which preserves the Heston with jump-to-default structure. Then the system of Riccati ODEs* (3.3.4) *admits the following solution, for all $z \in \mathbb{C}^2_- \times i\mathbb{R}$:*

$$\Psi_1^Q(t,z) = -\frac{\left(z_3 - z_3^2 + 2\Lambda_1^Q(1-z_3)\right)\left(e^{\sqrt{\Delta_1}t}-1\right) - \left(\sqrt{\Delta_1}\left(e^{\sqrt{\Delta_1}t}+1\right) + \left(\bar{\sigma}\left(\Theta_{1,1}+\rho z_3\right)-k\right)\left(e^{\sqrt{\Delta_1}t}-1\right)\right) z_1}{\sqrt{\Delta_1}\left(e^{\sqrt{\Delta_1}t}+1\right) - \left(\bar{\sigma}\left(\Theta_{1,1}+\rho z_3\right)-k\right)\left(e^{\sqrt{\Delta_1}t}-1\right) - \bar{\sigma}^2\left(e^{\sqrt{\Delta_1}t}-1\right) z_1}$$

$$\Psi_2^Q(t,z) = -\frac{2\Lambda_2^Q(1-z_3)\left(e^{\sqrt{\Delta_2}t}-1\right) - \left(\sqrt{\Delta_2}\left(e^{\sqrt{\Delta_2}t}+1\right) + \left(\sigma_0\,\Theta_{2,2}-k_0\right)\left(e^{\sqrt{\Delta_2}t}-1\right)\right) z_2}{\sqrt{\Delta_2}\left(e^{\sqrt{\Delta_2}t}+1\right) - \left(\sigma_0\Theta_{2,2}-k_0\right)\left(e^{\sqrt{\Delta_2}t}-1\right) - \sigma_0^2\left(e^{\sqrt{\Delta_2}t}-1\right) z_2}$$

$$\Psi_3^Q(t,z) = z_3$$

$$\Phi^Q(t,z) = \frac{2(k\hat{v}+\bar{\sigma}\hat{\theta}_1)}{\bar{\sigma}^2}\log\left(\frac{2\sqrt{\Delta_1}\exp\left(\frac{\sqrt{\Delta_1}-(\bar{\sigma}(\Theta_{1,1}+\rho z_3)-k)}{2}t\right)}{\sqrt{\Delta_1}(e^{\sqrt{\Delta_1}t}+1)-(\bar{\sigma}(\Theta_{1,1}+\rho z_3)-k)(e^{\sqrt{\Delta_1}t}-1)-\bar{\sigma}^2(e^{\sqrt{\Delta_1}t}-1)z_1}\right)$$
$$+\frac{2(k_0\hat{y}+\sigma_0\hat{\theta}_2)}{\sigma_0^2}\log\left(\frac{2\sqrt{\Delta_2}\exp\left(\frac{\sqrt{\Delta_2}-(\sigma_0\Theta_{2,2}-k_0)}{2}t\right)}{\sqrt{\Delta_2}(e^{\sqrt{\Delta_2}t}+1)-(\sigma_0\Theta_{2,2}-k_0)(e^{\sqrt{\Delta_2}t}-1)-\sigma_0^2(e^{\sqrt{\Delta_2}t}-1)z_2}\right)+\left(r+\bar{\lambda}^Q\right)(z_3-1)t$$

*where* $\Delta_1 := \left(\bar{\sigma}\left(\Theta_{1,1}+\rho z_3\right)-k\right)^2 + \bar{\sigma}^2\left(z_3 - z_3^2 + 2\Lambda_1^Q(1-z_3)\right)$ *and* $\Delta_2 := \left(\sigma_0\,\Theta_{2,2}-k_0\right)^2 + 2\,\sigma_0^2\,\Lambda_2^Q(1-z_3)$.

By combining the above lemma with the results of Sections 3.4-3.5, we can efficiently solve risk management problems and compute arbitrage-free prices of general default-sensitive payoffs.

**Remark 3.6.4.** *In the context of the model* (3.6.1)-(3.6.3)*, it may seem simplistic to restrict the attention to the set of risk-neutral measures which preserve the Heston with jump-to-default structure, i.e., to the set of risk premia processes $\theta = (\theta_t)_{0\leq t\leq T}$ which satisfy the restriction $\Theta_{1,2} = \Theta_{2,1} = 0$ (see the proof of Lemma 3.6.1). However, due to Theorem 4.1 of [GT08], the system of Riccati ODEs* (3.3.4) *for the model* (3.6.1)-(3.6.3) *admits an explicit solution* if and only if $\Theta_{1,2} = \Theta_{2,1} = 0$. *In other words, the set of risk-neutral measures which preserve the Heston with jump-to-default structure characterised in Lemma 3.6.1 coincides with the set of all measures under which system* (3.3.4) *admits a closed-form solution, which is given in Lemma 3.6.3. Of course, by relying on Theorem 3.3.2, we can relax the requirement of the preservation of*

*the Heston with jump-to-default structure with the weaker requirement of the preservation of the affine structure of* $(X, \tau)$ *but, in that case, one has to rely on numerical techniques for solving the Riccati system* (3.3.4).

### 3.6.3 Numerical results

This section reports the results of some numerical experiments for the Heston with jump-to-default model (3.6.1)-(3.6.3). We adopt the following parameters' specification: $k = 0.565$, $\hat{v} = 0.07$, $\bar{\sigma} = 0.281$, $k_0 = 0.325$, $\hat{y} = 0.003$, $\sigma_0 = 0.036$, $\mu = 0.1$, $\rho = -0.558$. These values have been obtained in [CW10] by calibrating (via filtering and maximum likelihood techniques) an analogous stochastic volatility jump-to-default model to market quotes of equity options and CDS spreads on the Citigroup company (period: 5/2002 - 5/2006). The remaining parameters appearing in (3.6.4) are specified as $\bar{r} = 0$, $\Theta_{1,1} = \Theta_{2,2} = 0.002$, $\hat{\theta}_1 = \hat{\theta}_2 = 0.001$ and $\Lambda_1^P = \Lambda_2^P = \bar{\lambda}^P = 0.1225$.

As a first application, we compute the distribution function of the defaultable stock price $S_T$ in the case of survival. More specifically, we consider the model (3.6.1)-(3.6.3) under the physical probability measure $P$ and, by relying on formula (3.4.4) together with Lemma 3.6.3, we compute the surface $(T, x) \mapsto P(S_T \leq x, \tau > T)$, for $T \in [0.5, 3.0]$ and $x \in [0.7, 1.3]$, for $S_0 = 1$. Note that, from the computational point of view, this is an easy task in our modeling framework, since it only requires a one-dimensional numerical integration. As can be observed from Figure 3.1, the shape of the distribution function strongly depends on the time horizon $T$, with a distinct behavior for small and large values of $x$, due to the combined effects of diffusive and jump-type risks. Figure 3.2 shows that the distribution function of the defaultable stock price can be quite different under the physical and a risk-neutral probability measure, even in the case where the overall default probability is kept at the same level (i.e., we have $P(\tau \leq T) = Q(\tau \leq T) = 0.4$), thus accounting for risk-aversion and providing an evidence of the flexibility induced by the possibility of changing the default intensity from the physical to a risk-neutral probability measure (to this regard, compare also the discussion preceding Remark 3.3.3).

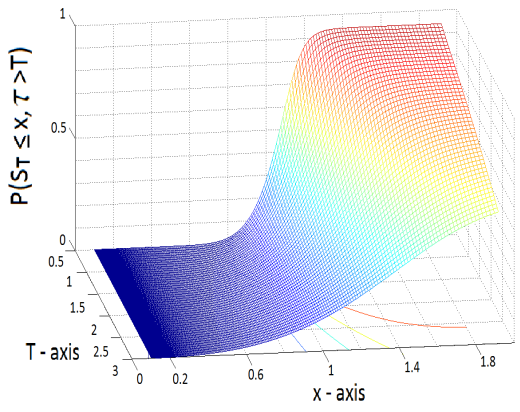


**Figure 3.1:** Surface $(T, x) \mapsto P(S_T \leq x, \tau > T)$ for the Heston with Jump-to-Default model.

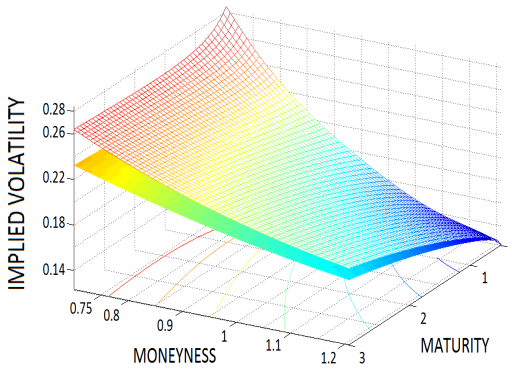


**Figure 3.2:** Distribution functions under physical ($P$) and risk-neutral ($Q$) probability measures.

As a second application, we show the implied volatility surface generated by the model (3.6.1)-(3.6.3). To this effect, we first compute a matrix of prices $P_K(0, T)$ of Put options on the default-

able stock $S_T$, issued by a default-free third party, with maturity $T \in [0.5, 3.0]$ and moneyness $K/S_0 \in [0.7, 1.3]$, letting $\bar{\lambda}^Q = 0.001$ and $\Lambda_i^Q = \Lambda_i^P$, for $i = 1, 2$. The computation is performed via the Fast Fourier Transform method of [CM99], by relying on Corollary 3.5.5 and Lemma 3.6.3. The corresponding implied volatilities are then computed by using the *blsimpv* function in Matlab© (R2012a 64-bit version).

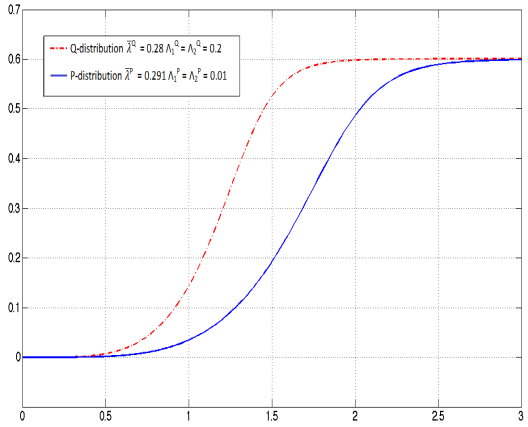


**Figure 3.3:** Implied volatility surfaces: standard Heston (solid) and Heston + Jump-to-Default (mesh).

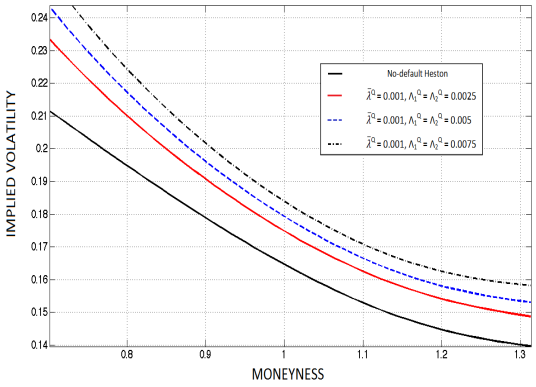


**Figure 3.4:** Implied volatility skew for different $Q$-intensities (for $T = 1.75$).

Figure 3.3 compares the implied volatility surface generated by the model (3.6.1)-(3.6.3) with the implied volatility surface obtained from a standard (default-free) Heston [Hes93] model, i.e., by letting $\bar{\lambda}^Q = \Lambda_1^Q = \Lambda_2^Q = 0$. It is evident that the introduction of default risk (through a jump-to-default) increases the implied volatility along all maturities and strikes. The increase is more pronounced for deep out-of-the-money options, due to the possibility of obtaining $K$ in the case of default (compare also with equation (3.5.7)), thus confirming the fact that default risk is the main responsible for the value of out-of-the-money put options with short maturities. There is also a strong skew effect, which tends to flatten as the maturity increases but is always more significant than in the default-free case. The impact of default risk is also shown in Figure 3.4, which depicts the implied volatility skew for different specifications of the parameters which determine the default intensity $\lambda^Q$ together with the skew generated by a standard default-free Heston model, for the fixed maturity $T = 1.75$. As expected, the implied volatility skew is more pronounced for a higher risk of default as measured by larger values of the default intensity parameters.

# Chapter 4

# Further Developments

In this final chapter we discuss some further developments of interest for Part I, Part II and Part III.

## 4.1 Further Developments of the Pricing Model

In the context of pricing, we have seen that our approach based on CTMCs is versatile enough to cover many different types of exotic payoffs. One possible shortcoming of the model, however, is the fact that interest rates are modeled as the short rate, which is, in fact, an idealized quantity. It would then be of interest to further investigate the extension of the pricing approach to the case, for example, of forward LIBOR rate models. A possible step towards this direction is to study how our proposed pricing approach may be translated to a generalized Markov functional model under a forward LIBOR rate term-structure setting, in the case when the Markovian driver is a CTMC. LIBOR Markov functional models (see [HKP00] and Chapter 27 of [Fri07]) are a class of interest rate models that are gaining attention due to their ability to fit perfectly any market-derived interest rate curves. Under this framework one supposes only that the price processes of the elementary traded financial instruments are each functionals of a Markovian driving process; in particular, the interest rates' functionals may be implied directly from a specification of the functionals for the equity prices. However most work in this area have focused only on the case when the Markovian driver is a Brownian Motion. To briefly illustrate this case, suppose we denote $T_k$, $k = 1 \dots m$ to be the tenors of forward yields quoted in the market. Denote by $P(t, T_k, T_{k+1})$ the value of a discount bond at time $t$ with expiry date $T_k$ and maturity $T_{k+1}$. Under a LIBOR Markov functional model with a Brownian Motion driver, one assumes that the forward LIBOR rate $L(t, T_k)$, defined as (see Section 27.3 of [Fri07])

$$L(T_k) = \frac{1 - P(T_k, T_{k+1})}{P(T_k, T_{k+1})(T_{k+1} - T_k)}$$

is a deterministic function $L(T_k) = L(T_k; x(T_k))$ of $x(T_k)$, where $x$ is a Markov process of the form

$$dx = \sigma(t)dW \quad \text{under } Q^N, x(0) = x_0,$$

and where the choice of numeraire may, for example, be the $T_m$-bond, i.e. $N(T_i) := P(T_i, T_m)$ (this leads to a so-called model under *terminal measure*). In this set-up, it can be shown (Lemma

233 of [Fri07]) that the numeraire is also given by a functional of $x(T_i)$, and as a consequence, also the prices of market-quoted derivatives such as Caplets and Swaps. In particular, if the functional of $L(T_k)$ is specified in some parametrized form, then the price functionals of all market-quoted derivatives are automatically determined. The entire model can be calibrated efficiently on the basis of the prices of digital caplets alone (Section 27.3.1.2 [Fri07]). An extension to the case when the Markovian driver is a CTMC could then provide a connection to our work and an extension to forward rates under a LIBOR Market Model.

Concerning the MC -based pricing under our approach, it was noted that while a variance reduction is attained and a possible bias reduction is also numerically observed, these effects are still quite dependent on the parameters of the CTMC. Numerical experiments suggest that the MC sample distributions of jump transitions tend to be highly clustered, that is, trajectories that lead to certain jump-count numbers are a priori more likely and hence other types of trajectories are under-represented in the sample. This clustering problem is very well known in the field of Monte-Carlo simulation. A possible solution that may also be of interest to investigate is the application of the so-called Quasi-Monte Carlo (QMC). In QMC, the sampling is no-longer performed in a fully random manner, but rather, a semi-deterministic sampling is performed in order to even-out the number of observations of the different possible trajectories.

Finally, we note that the numerical tests reported were only on the basis of two MC techniques and the analytic formula in the case of simple claims. It would be important to also subsequently compare the numerical tests on techniques based on solving the backward ODE for the price. One such example in this regard is [Nor05], which, however, due to time-constraints could, could not be included in this thesis.

## 4.2 Further Developments of the Calibration Method

As with the preceding remarks, the calibration method we propose has the shortcoming that it is defined under a short rate model, whereas financial practitioners more often work directly with market-derived rates. Just as with pricing, it would then be of interest to also extend the calibration method to the case of a forward LIBOR rate model, again possibly by means of the LIBOR Markov functional model framework. An interesting extension would be to look at generalized Markov functional models under incomplete information.

An interesting possible extension of our calibration method is to treat the case of credit ratings migrations. Indeed, several credit rating migrations models, as described in Chapter 12 of [BR02] are based on the assumption that the credit rating transitions follow a CTMC. By combining ideas from Part III, it seems plausible to define a suitable change-of-measure that would allow calibration of the historical credit rating transition probabilities on the basis of time-series of price data of defaultable bonds or equities and other derivatives issued by defaultable entities.

From the theoretical perspective, an important further study is to understand better the role of quasi - exact solutions and their relationship to the true solution. The study of the logarithm of $\boldsymbol{e^A e^B}$ when $\boldsymbol{A}$ and $\boldsymbol{B}$ do not commute, goes beyond the scope of the present thesis, and would involve results from Lie Algebras. Indeed, one computes such a logarithm using the

Baker-Hausdorff-Campbell Formula[1]. It is connected to numerical methods for ODE through the *splitting-up schemes* ([Gla92], [BGR90]). Such schemes were inspired by the application of the Lie-Trotter formula to numerically approximating the solutions of PDEs. The Lie-Trotter formula, a direct result of the Baker-Hausdorff-Campbell formula, states that, for matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ that are possibly non-commuting, then

$$\boldsymbol{e}^{\boldsymbol{A}+\boldsymbol{B}} = \lim_{N\to\infty} \left(\boldsymbol{e}^{\frac{\boldsymbol{A}}{N}} \boldsymbol{e}^{\frac{\boldsymbol{B}}{N}}\right)^N.$$

Roughly, splitting-up schemes generalizes the Lie-Trotter formula to the semigroup operators associated with the split-up differential operators, instead of exponentials of matrices. In the commuting case, the proposed quasi-exact scheme is perfectly analogous to a splitting-up scheme (see the Remark 2.3.14). This suggests that the splitting-up scheme convergence may suggest a way to conjecture whether the quasi-exact scheme also approximates the true solution in the non-commuting case.

Finally, it is evident that sharper bounds for the Quasi - Exact Approximations be found - indeed, the numerical results suggest that they are able to attain a higher strong order than the Euler scheme (0.5) but perform at least as good as the Milstein scheme (1.0).

## 4.3 Extensions of the unified affine model for pricing and risk-management

We have proposed a general framework based on an affine process $X$ and on a doubly stochastic random time $\tau$ for the modeling of a defaultable stock. This approach allows to jointly model equity and credit risk, together with stochastic volatility and stochastic interest rate. Moreover, analytical tractability is ensured under both the physical and a set of risk-neutral probability measures, thanks to a flexible characterisation of all risk-neutral measures which preserve the affine structure of $(X, \tau)$.

In the present paper, we have chosen to specify the driving process $X$ as an affine diffusion on $\mathbb{R}^m_{++} \times \mathbb{R}^{d-m}$, for some $m \in \{1, \ldots, d-1\}$. However, our techniques can be easily adapted to the more general case where $X$ is a continuous matrix-valued affine process (e.g., a Wishart process), as recently considered e.g. in [CFMT11]. We also want to mention that the characterisation of risk-neutral measures which preserve the affine structure of $(X, \tau)$ provided in Theorem 3.3.2 (or in Lemma 3.6.1 for the more specific case of the Heston with jump-to-default model) can also be useful in insurance mathematics for the valuation of mortality-indexed insurance contracts in the context of intensity-based mortality models (see e.g. [Bif05]).

[1]See *Yu.A. Bakhturin (2001), "Campbell - Hausdorff formula", in Hazewinkel, Michael, Encyclopedia of Mathematics, Springer.*

# Bibliography


[AR09] R. Ahlip and M. Rutkowski, *Forward start options under stochastic volatility and stochastic interest rate*, International Journal of Theoretical and Applied Finance **12** (2009), 209–225.

[Bay08] E. Bayraktar, *Pricing options on defaultable stocks.*, Appl. Math. Financ. **15** (2008), 277–304.

[Ben02] Michele Benzi, *Preconditioning techniques for large linear systems: A survey*, Journal of Computational Physics **182** (2002), 418–477.

[BGR90] A. Bensoussan, R. Glowinski, and A. Rascanu, *Approximation of the zakai equation by the splitting up method*, SIAM Journal of Control and Optimization **28** (1990), no. 6, 1420–1431.

[Bif05] E. Biffis, *Affine processes for dynamic mortality and actuarial valuations,*, Insur. Math. Econ. **37** (2005), 443–468.

[BKR97] T. Björk, Y. Kabanov, and W. J. Runggaldier, *Bond market structure in the presence of marked point processes*, Mathematical Finance **7** (1997), no. 2, 211–223.

[BPP01] V. Bally, G. Pages, and J. Printems, *A stochastic quantization method for nonlinear problems.*, Monte Carlo Methods and Applications **7** (2001), 21–34.

[BR02] T. R. Bielecki and M. Rutkowski, *Credit risk: Modeling, valuation and hedging*, Springer, 2002.

[CC12] C. Ceci and K. Colaneri, *Nonlinear filtering for jump diffusion observations*, Adv. in Appl. Probab. **44** (2012), no. 3, 678–701.

[CDMW08] M. Cremers, J. Driessen, P. Maenhout, and D. Weinbaum, *Individual stock-option prices and credit spreads*, J. Bank. Financ. **32** (2008), 2706–2715.

[CFK07] P. Cheridito, D. Filipović, and R.L. Kimmel, *Market price of risk specifications for affine models: Theory and evidence*, J. Financ. Econ. **83** (2007), 123–170.

[CFK10] ———, *A note on the dai-singleton canonical representation of affine term structure models*, Math. Financ. **20** (2010), 509 – 519.

[CFMT11] C. Cuchiero, D. Filipović, E. Mayerhofer, and J. Teichmann, *Affine processes on positive semidefinite matrices.*, Ann. Appl. Probab. **21** (2011), 397 – 463.

[CFY05] P. Cheridito, D. Filipović, and M. Yor, *Equivalent and absolutely continuous measure changes for jump-diffusion processes*, Ann. Appl. Probab. **15** (2005), 1713–1732.

[CHJ09] P. Christoffersen, S. Heston, and K. Jacobs, *The shape and the term structure of the index option smirk: why multifactor stochastic volatility models work so well.*, Manage. Sci. **55** (2009), 1914–1932.

[CJN12] D. Coculescu, M. Jeanblanc, and A. Nikeghbali, *Default ttime, no-arbitrage conditions and changes of probability measures,*, Financ. Stoch. **16** (2012), 513–535.

[CK12] T.K. Chung and Y.K. Kwok, *Handbook on computational economics and finance*, ch. Equity-credit modeling under affine jump-diffusion models with jump-to-default, p. forthcoming, Oxford University Press, 2012.

[CL06] P. Carr and V. Linetsky, *A jump to default extended cev model: An application of bessel processes.*, Financ. Stoch. **10** (2006), 303–330.

[CM99] P. Carr and D.B. Madan, *Option valuation using the fast fourier transform.*, J. Comput. Financ. **2** (1999), 61–73.

[CM10] ———, *Local volatility enhanced by a jump to default.*, SIAM J. Financial Math. **1** (2010), 2–15.

[CPS09] L. Campi, S. Polbennikov, and A. Sbuelz, *Systematic equity-based credit risk: A cev model with jump to default.*, J. Econ. Dyn. Control **33** (2009), 93–108.

[Cre99] A. Di Crescenzo, *A probabilistic analogue of the mean value theorem and its applications to reliability theory.*, J. Appl. Prob. **36** (1999), 706–719.

[CS08] P. Carr and W. Schoutens, *Hedging under the heston model with jump-to-default*, Int. J. Theor. Appl. Financ. **11** (2008), 403–414.

[CT03] J. Campbell and G. Taksler, *Equity volatility and corporate bond yields.*, J. Financ. **LVIII** (2003), 2321–2349.

[CW10] P. Carr and L. Wu, *Stock options and credit default swaps: A joint framework for valuation and estimation.*, J. Financ. Economet. **8** (2010), 409–449.

[CW12] P. Cheridito and A. Wugalter, *Pricing and hedging in affine models with possibility of default.*, SIAM J. Financial Math. **3** (2012), 328–350.

[DF09] E. Mayerhofer D. Filipovic, *Affine diffusion processes: Theory and applications*, Radon Series in Computational and Applied Mathematics **8** (2009), 1–40.

[DFS03] D. Duffie, D. Filipovic, and W. Schachermayer, *Affine processes and applications in finance*, Annals of Applied Probability **13** (2003), 984–1053.

[DK96] D. Duffie and R. Kan, *A yield-factor model of interest rates*, Mathematical Finance **6** (1996), 379–406.

[DPS00] D. Duffie, J. Pan, and K.J. Singleton, *Transform analysis and asset pricing for affine jump-diffusions*, Econometrica **68** (2000), no. 6, 1343–1376.

[Dri05] J. Driessen, *Is default event risk priced in corporate bonds?*, Rev. Financ. Stud. **18** (2005), 165–195.

[DS94] F. Delbaen and W. Schachermayer, *A general version of the fundamental theorem of asset pricing*, Math. Ann. **300** (1994), 463–520.

[DS00] Q. Dai and K. Singleton, *Specification analysis of affine term structure models*, Journal of Finance **55** (2000), 1943–1978.

[Duf99] G. R. Duffee, *Estimating the price of default risk*, The Review of Financial Studies **12** (1999), no. 1, 197–226.

[Duf02] G.R. Duffee, *Term premia and inerest rate forecasts in affine models*, Journal of Finance **LVII** (2002), no. 1, 405–443.

[Duf05] D. Duffie, *Credit risk modeling with affine processes.*, J. Bank. Financ. **29** (2005), 2751–2802.

[DZ86] A. Dembo and O. Zeitouni, *Parameter estimation of partially observed continuous time stochastic processes*, Stochastic Processes and Their Applications **23** (1986), 91–113.

[EAM08] R. J. Elliott, L. Aggoun, and J. B. Moore, *Hidden markov models: Estimation and control*, 2nd ed. ed., Stochastic Modelling and Applied Probability, Springer, 2008.

[EHJ00] R. Elliott, W. Hunter, and B. Jamieson, *Financial signal processing: A self calibrating model*, Working Paper, Federal Reserve Bank of Chicago (2000).

[EK06] E. Eberlein and W. Kluge, *Exact pricing formulae for caps and swaptions in a levy term structure model.*, Journal of Computational Finance **9** (2006), 99–125.

[Ell93] R. Elliott, *New finite-dimensional filters and smoothers for noisily observed markov chains*, IEEE Transactions On Information Theory **39** (1993), no. 1.

[Fil09] D. Filipovic, *Term - structure models, a graduate couse*, Springer, 2009.

[FKK72] M. Fujisaki, G. Kallianpur, and H. Kunita, *Stochastic differential equations for the non-linear filtering problem*, Osaka J. Math **9** (1972), 19–40.

[FM14] C. Fontana and J.M. Montes, *A unified approach to pricing and risk management of equity and credit risk*, Journal of Computational and Applied Mathematics **259** (2014), 350–361.

[Fon12a] C. Fontana, *Four essays in financial mathematics*, Ph.D. thesis, University of Padua, 2012.

[Fon12b] ———, *Mathematical and statistical methods for actuarial sciences and finance.*, ch. Credit risk and incomplete information: a filtering framework for pricing and risk management., pp. 193–201, Springer, Milan, 2012.

[FR10a] C. Fontana and W. J. Runggaldier, *Credit risk and incomplete information: Filtering and EM parameter estimation*, International Journal of Theoretical and Applied Finance **13** (2010), 683–715.

[FR10b] R. Frey and W. J. Runggaldier, *Pricing credit derivatives under incomplete information: a nonlinear filtering approach.*, Finance and Stochastics **14** (2010), no. 4, 495–526.

[Fri07] C. Fries, *Mathematical finance: Theory, modelling and implementation*, Wiley, 2007.

[FZ02] D. Filipovic and J. Zabczyk, *Markovian term structure models in discrete time.*, Annals of Applied Probability **7** (2002), no. 2, 710–729.

[Gla92] F. Le Gland, *Splitting-up approximation for spde's and sde's with application to nonlinear filtering*, in: Stochastic Partial Differential Equations and Their Applications, Charlotte 1991, B. L. Rozovskii and R. B. Sowers, editors, Lecture Notes in Control and Information Sciences **176** (1992), 177–187.

[Gla04] P. Glasserman, *Monte carlo methods in financial engineering.*, Springer-Verlag, 2004.

[GS09] R.M. Gaspar and T. Schmidt, *Financial risks: New developments in structured product and credit derivatives*, ch. CDOs in the light of the current crisis, pp. 33–48, Economica, Paris, 2009.

[GT08] M. Grasselli and C. Tebaldi, *Solvable affine term structure models,*, Math. Financ. **18** (2008), 135 – 153.

[Han07a] R. Van Handel, *Filtering, stability, and robustness.*, Ph.D. thesis, California Institute of Technology, 2007.

[Han07b] ———, *Lecture notes on stochastic calculus, filtering and stochastic control.*, 2007.

[Hes93] S.L. Heston, *A closed-form solution for options with stochastic volatility, with applications to bond and currency options.*, Rev. Financ. Stud. **6** (1993), 327–343.

[HKP00] P. Hunt, J. Kennedy, and A. Pelsser, *Markov-functional interest rate models*, Finance and Stochastics **4** (2000), no. 4, 391–408.

[JLY05] R.A. Jarrow, D. Lando, and F. Yu, *Default risk and diversification: Theory and empirical implications*, Mathematical Finance **15** (2005), no. 1, 1–26.

[KD01] H. J. Kushner and P. G. Dupuis, *Numerical methods for stochastic control problems in continuous time*, Springer-Verlag, 2001.

[KR09] M. Keller-Ressel, *Affine processes - theory and applications in finance*, Ph.D. thesis, TU Wien, 2009.

[Lan00] C. Landen, *Bond pricing in a hidden Markov Model of the short rate.*, Finance and Stochastics **4** (2000), 371–389.

[MFE05] A.J. McNeil, R. Frey, and P. Embrechts, *Quantitative risk management - concepts, techniques and tools,*, Princeton University Press, 2005.

[MPR13] J. Montes, V. Prezioso, and W. Runggaldier, *Monte carlo variance reduction by conditioning for pricing with underlying a continuous-time finite state markov process.*, submitted preprint, University of Padua (2013).

[NEK01] L. Martellini N. El Karoui, *A theoretical inspection of the market price for default risk.*, Working Paper, University of Southern California (2001).

[Nor03] R. Norberg, *The markov chain market*, ASTIN Bulletin (2003).

[Nor05] ———, *Anomalous pdes in markov chains: Domains of validity and numerical solutions.*, Finance and Stochastics **9** (2005), 519–537.

[Øks10] B. Øksendal, *Stochastic differential equations*, Springer-Verlag, 2010.

[Pao07] M.S. Paolella, *Intermediate probability: A computational approach*, Wiley, Chichester, 2007.

[PBL10] E. Platen and N. Bruti-Liberati, *Numerical solution of stochastic differential equations with jumps in finance.*, Springer, 2010.

[Pla82] E. Platen, *An approximation method for a class of ito processes with jump component*, Liet. Mat. Rink. **22** (1982), no. 2, 124–136.

[PP12] K. B. Petersen and M.S. Pedersen, *The matrix cookbook*, 2012, Version 20121115.

[PR10a] E. Platen and R. Rendek, *Quasi-exact approximation of hidden markov chain filters.*, Communications on Stochastic Analysis **4** (2010), 129–142.

[PR10b] V. Prezioso and W. J. Runggaldier, *Interest rate derivatives pricing when the short rate is a continuous time finite state markov process*, Preprint (2010).

[Pre10] V. Prezioso, *Interest rate derivatives pricing when the short rate is a continuous time finite state markov process*, Ph.D. thesis, University of Padua, 2010.

[Pro05] P. E. Protter, *Stochastic integration and differential equations*, 2nd, ver 2.1 ed., Stochastic Modelling and Applied Probability, Springer, 2005.

[Sch03] P.J. Schönbucher, *A note on survival measures and the pricing of options on credit default swaps.*, working paper, Department of Mathematics, ETH Zürich (2003).

[SS07] M. Shaked and G. Shanthikhumar, *Stochastic orders*, Springer Series in Statistics, 2007.

[WH06] B. Wong and C.C. Heyde, *On changes of measure in stochastic volatility models,*, J. Appl. Math. Stoch. Anal. (2006), 1–13.

[Won65] W.M. Wonham., *Some applications of stochastic differential equations to optimal nonlinear filtering*, SIAM J. Control **2** (1965), 347–369.

[Wu83] C.F.J. Wu, *On the convergence properties of the EM algorithm*, The Annals of Statistics **11(1)** (1983), 95–103.

[ZD88] O. Zeitouni and A. Dembo., *Exact filters for the estimation of the number of transitions of finite-state continuous time markov processes.*, IEEE Transactions on Information Theory **34** (1988), 890–893.