

Markovian lifts of positive semidefinite affine Volterra-type processes

Cuchiero, Christa; Teichmann, Josef

Published in:
Decisions in Economics and Finance

DOI:
[10.1007/s10203-019-00268-5](https://doi.org/10.1007/s10203-019-00268-5)

Published: 01/01/2019

Document Version
Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for published version (APA):
Cuchiero, C., & Teichmann, J. (2019). Markovian lifts of positive semidefinite affine Volterra-type processes. *Decisions in Economics and Finance*, 42, 407-448. <https://doi.org/10.1007/s10203-019-00268-5>



Markovian lifts of positive semidefinite affine Volterra-type processes

Christa Cuchiero¹ · Josef Teichmann²

Received: 5 May 2019 / Accepted: 7 September 2019
© The Author(s) 2019

Abstract

We consider stochastic partial differential equations appearing as Markovian lifts of matrix-valued (affine) Volterra-type processes from the point of view of the generalized Feller property (see, e.g., Dörsek and Teichmann in *A semigroup point of view on splitting schemes for stochastic (partial) differential equations*, 2010. [arXiv:1011.2651](https://arxiv.org/abs/1011.2651)). We introduce in particular Volterra Wishart processes with fractional kernels and values in the cone of positive semidefinite matrices. They are constructed from matrix products of infinite dimensional Ornstein–Uhlenbeck processes whose state space is the set of matrix-valued measures. Parallel to that we also consider positive definite Volterra pure jump processes, giving rise to multivariate Hawkes-type processes. We apply these affine covariance processes for multivariate (rough) volatility modeling and introduce a (rough) multivariate Volterra Heston-type model.

Keywords Stochastic partial differential equations · Affine processes · Wishart processes · Hawkes processes · Stochastic Volterra processes · Rough volatility models

Mathematics Subject Classification 60H15 · 60J25

JEL Classification C.5 · G.1

The authors are grateful for the support of the ETH Foundation and Erwin Schrödinger Institut Wien. Christa Cuchiero gratefully acknowledges financial support by the Vienna Science and Technology Fund (WWTF) under Grant MA16-021.

✉ Christa Cuchiero
christa.cuchiero@wu.ac.at

¹ Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Vienna, Austria

² ETH Zürich, Rämistrasse 101, 8092 Zurich, Switzerland

1 Introduction

It is the goal of this article to investigate the results of Cuchiero and Teichmann (2018) on infinite dimensional Markovian lifts of stochastic Volterra processes in a multivariate setup: We are mainly interested in the case where the stochastic Volterra processes take values in the cone of positive semidefinite matrices \mathbb{S}_+^d . We shall concentrate on the affine case due to its relevance for tractable *rough covariance modeling*, extending rough volatility (see, e.g., Alòs et al. 2007; Gatheral et al. 2018; Bayer et al. 2016) to a setting of d “roughly correlated” assets.

Viewing stochastic Volterra processes from an infinite dimensional perspective allows to dissolve a generic non-Markovianity of the at first sight naturally low-dimensional volatility process. Indeed, this approach makes it actually possible to go beyond the univariate case considered so far and treat the problem of multivariate rough covariance models for more than one asset. Moreover, the considered Markovian lifts allow to apply the full machinery of affine processes. We refer to the introduction of Cuchiero and Teichmann (2018) for an overview of theoretical and practical advantages of Markovian lifts in the context of Volterra-type processes.

Let us start now by explaining why the matrix-valued positive definite case is actually more involved than the scalar one in \mathbb{R}_+ , where, for instance, the Volterra Cox–Ingersoll–Ross process takes values. The latter appears as variance process in a rough Heston model (see, e.g., El Euch 2019; Abi Jaber and El Euch 2019; Alòs and Yang 2017). Consider now a standard Wishart process on \mathbb{S}_+^d , as defined in Bru (1991), Cuchiero et al. (2011), of the form

$$dX_t = (d - 1) \text{Id}_d dt + \sqrt{X_t} dW_t + dW_t^\top \sqrt{X_t}, \quad X_0 \in \mathbb{S}_+^d. \quad (1.1)$$

Here $\sqrt{\cdot}$ denotes the matrix square root, Id_d the identity matrix and W a $d \times d$ matrix of Brownian motions. The (necessary) presence of the dimension d in the drift is an obvious obstruction to infinite dimensional versions of this equation, which could be projected to obtain Volterra-type equations by the variation of constants formula; see Cuchiero and Teichmann (2018) for such a projection on \mathbb{R}_+ . In order to circumvent this difficulty, we present two approaches in this paper:

- We develop a theory of infinite dimensional affine Markovian lifts of pure jump positive semidefinite Volterra processes.
- We develop a theory of squares of Gaussian processes in a general setting to construct infinite dimensional analogs of Wishart processes. Their finite dimensional projections, however, look different from naively conjectured Volterra Wishart processes following the role model of Volterra Cox–Ingersoll–Ross processes. They are also different in dimension one, as outlined below.

The jump part appears natural and comes without any further probabilistic problem when constrained to finite variation jumps. Note that in the (non-Volterra) case of affine processes on positive semidefinite matrices, quadratic variation jumps are not possible either (see Mayerhofer 2012). With the generalized Feller approach from Dörsek and Teichmann (2010), Cuchiero and Teichmann (2018), we obtain a new class of stochastic Volterra processes taking values in \mathbb{S}_+^d of the form

$$V_t = h(t) + \int_0^t (K(t-s)V_s + V_s K(t-s))ds + \int_0^t K(t-s)dN_s + \int dN_s K(t-s), \tag{1.2}$$

where $h : \mathbb{R}_+ \rightarrow \mathbb{S}_+^d$ is some deterministic function, K a (potentially fractional) kernel in $L^2(\mathbb{R}_+, \mathbb{S}_+^d)$ and N a pure jump process of finite variation with jump sizes in \mathbb{S}_+^d , whose compensator is a linear function in V . This allows, for instance, to define a multivariate Hawkes process \widehat{N}^1 with values in \mathbb{N}_0^d given by the diagonal entries of N , i.e., $\text{diag}(N) = \widehat{N}$, and the compensator of \widehat{N}_i is given by $\int_0^\cdot V_{s,i} ds$ (see Example 4.16). By means of the affine transform formula for the infinite dimensional lift of (1.2), we are able to derive an expression for the Laplace transform of V_t which can be computed by means of matrix Riccati–Volterra equations.

The difficulty of the continuous part arises from geometric constraints, which can, however, be circumvent by building squares of unconstrained processes. Let us illustrate the idea in a finite dimensional setting: Let W be an $n \times d$ matrix of Brownian motions and let v be a matrix in $\mathbb{R}^{d \times dk}$ consisting of k submatrixes $v_i \in \mathbb{R}^{d \times d}$, $i = 1, \dots, k$, i.e., $v = (v_1, \dots, v_k)$.

Define now a Gaussian process with values in $\mathbb{R}^{n \times dk}$ by $\gamma := Wv$. Then, by Itô’s product formula, the $\mathbb{R}^{dk \times dk}$ valued process $\gamma_t^\top \gamma_t$ satisfies the following equation

$$d\gamma_t^\top \gamma_t = nv^\top v dt + v^\top dW_t^\top \gamma_t + \gamma_t^\top dW_t v. \tag{1.3}$$

Following Bru (1991, Subsection 5.2) and setting $\lambda_t := \gamma_t^\top \gamma_t$, this can, however, also be written via a $kd \times kd$ matrix of independent Brownian motions B satisfying

$$\sqrt{\gamma_t^\top \gamma_t} dB_t \sqrt{v^\top v} = \gamma_t^\top dW_t v \tag{1.4}$$

in the more familiar form

$$d\lambda_t = nv^\top v dt + \sqrt{v^\top v} dB_t^\top \sqrt{\lambda_t} + \sqrt{\lambda_t} dB_t \sqrt{v^\top v}. \tag{1.5}$$

Our article is devoted to analyze the situation where the index variable v gets continuous, which is the only possible form of an infinite dimensional Wishart process. We believe that generalized Feller processes are the right arena to achieve this purpose. In this article, we choose measure spaces, but an analogous analysis can be done in the setting of function spaces as, for instance, the Hilbert space setting of Filipović (2001); see Cuchiero and Teichmann (2018, Section 5.2). In the measure-valued setting, we proceed as follows: Let γ be an infinite dimensional Ornstein–Uhlenbeck process taking values in $\mathbb{R}^{n \times d}$ -valued regular Borel measures on \mathbb{R}_+ . Then, Volterra Wishart processes arise as finite dimensional projections of $\gamma^\top(dx_1)\gamma(dx_2)$ on \mathbb{S}_+^d and can be written as

¹ We refer to the original paper of (Hawkes 1971) for the one-dimensional case.

$$V_t = h(t) + n \int_0^t K(t-s)K(t-s)ds + \int_0^t K(t-s)dW_s^\top Y(t,s)ds + \int_0^t Y(t,s)^\top dW_s K(t-s), \quad (1.6)$$

where h and K are as in (1.2), W an $n \times d$ matrix of Brownian motions and $Y(t, s) = \int_0^\infty e^{-x(t-s)} \gamma_s(dx)$. As explained in Remark 5.4, V_t corresponds to the matrix square of a Volterra Ornstein–Uhlenbeck process X_t , obtained as finite dimensional projection of $\gamma(dx)$. The Volterra Wishart process (1.6) can then also be written in terms of the forward process of X_t , i.e., $(\mathbb{E}[X_t|\mathcal{F}_s])_{s \leq t}$, namely

$$V_t = h(t) + n \int_0^t K(t-s)K(t-s)ds + \int_0^t K(t-s)dW_s^\top \mathbb{E}[X_t|\mathcal{F}_s]ds + \int_0^t \mathbb{E}[X_t^\top|\mathcal{F}_s]dW_s K(t-s).$$

Note that this is not of standard Volterra form, as, e.g., in Abi Jaber et al. (2019), since $Y(t, s)$ or $\mathbb{E}[X_t|\mathcal{F}_s]$, respectively, cannot be expressed as a function of V_t . By moving to a Brownian field analogous to (1.4), it could, however, be expressed as a path functional of $(V_s)_{s \leq t}$. For $n = d = 1$, it also gives rise to a different equation than the Volterra CIR process. We explain the connection between (1.6) and (1.3)–(1.5) in detail in Sect. 5.

Note that by choosing K to be a matrix of fractional kernels, the trajectories of (1.6) become rough, whence V qualifies for rough covariance modeling with potentially different roughness regimes for different assets and their covariances. This is in accordance with econometric observations. In Sect. 6, we show how such models can be defined: We introduce a (rough) multivariate Volterra Heston-type model with jumps and show that it can again be cast in the affine framework. This is particularly relevant for pricing basket or spread options using the Fourier pricing approach.

The remainder of the article is organized as follows: In Sect. 1.1, we introduce some notation and review certain functional analytic concepts. In Sects. 2 and 3, we recall and extend results on generalized Feller processes as outlined in Cuchiero and Teichmann (2018). In particular, Theorem 2.8 provides a result on invariant (sub)spaces for generalized Feller processes that is crucial for the square construction as outlined above. In Sect. 4, we apply the presented theory to SPDEs which are lifts of matrix-valued stochastic Volterra jump processes of type (1.2). Section 5 is devoted to present a theory of infinite dimensional Wishart processes which in turn give rise to (rough) Volterra Wishart processes. In Sect. 6, we apply these processes for multivariate (rough) volatility modeling.

1.1 Notation and some functional analytic notions

For the background in functional analysis, we refer to the excellent textbook of Schaefer and Wolff (1999) as main reference and to the equally excellent books of Engel and Nagel (2000) and Pazy (1983) for the background in strongly continuous semigroups.

We shall apply the following notations: Let Y be a Banach space and Y^* its dual space, i.e., the space of linear continuous functionals with the strong dual norm

$$\|\lambda\|_{Y^*} = \sup_{\|y\| \leq 1} |\langle y, \lambda \rangle|,$$

where $\langle y, \lambda \rangle := \lambda(y)$ denotes the evaluation of the linear functional λ at the point $y \in Y$. Since in the case of Eq. (1.2), cones \mathcal{E} of Y^* will be our state spaces, we denote the polar cones in pre-dual notation, i.e.,

$$\mathcal{E}_* = \{y \in Y \mid \langle y, \lambda \rangle \leq 0 \text{ for all } \lambda \in \mathcal{E}\}.$$

We denote spaces of bounded linear operators from Banach spaces Y_1 to Y_2 by $L(Y_1, Y_2)$ with norm

$$\|A\|_{L(Y_1, Y_2)} := \sup_{\|y_1\|_{Y_1} \leq 1} \|Ay_1\|_{Y_2}.$$

If $Y_1 = Y_2$, we only write $\|\cdot\|_{L(Y)}$. On Y^* , we shall usually consider beside the strong topology (induced by the strong dual norm) the weak- $*$ -topology, which is the weakest locally convex topology making *all* linear functionals $\langle y, \cdot \rangle$ on Y^* continuous. Let us recall the following facts:

- The weak- $*$ -topology is metrizable if and only if Y is finite dimensional: This is due to Baire’s category theorem since Y^* can be written as a countable union of closed sets, whence at least one has to contain an open set, which in turn means that compact neighborhoods exist, i.e., a strictly finite dimensional phenomenon.
- Norm balls K_R of any radius R in Y^* are compact with respect to the weak- $*$ -topology, which is the Banach–Alaoglu theorem.
- These balls are metrizable if and only if Y is separable: This is true since Y can be isometrically embedded into $C(K_1)$, where $y \mapsto \langle y, \cdot \rangle$, for $y \in Y$. Since Y is separable, its embedded image is separable, too, which means—by looking at the algebra generated by Y in $C(K_1)$ —that $C(K_1)$ is separable, which is the case if and only if K_1 is metrizable.

Even though some results are more general, in particular, often only compactness of K_R is used, we shall always assume separability in this article.

Finally, a family of linear operators $(P_t)_{t \geq 0}$ on a Banach space Y with $P_t P_s = P_{t+s}$ for $s, t \geq 0$ and with $P_0 = I$ where I denotes the identity is called strongly continuous semigroup if $\lim_{t \rightarrow 0} P_t y = y$ holds true for every $y \in Y$. We denote its generator usually by A which is defined as $\lim_{t \rightarrow 0} \frac{P_t y - y}{t}$ for all $y \in \text{dom}(A)$, i.e., the set of elements where the limit exists. Notice that $\text{dom}(A)$ is left invariant by the semigroup P and that its restriction on the domain equipped with the operator norm

$$\|y\|_{\text{dom}(A)} := \sqrt{\|y\|^2 + \|Ay\|^2}$$

is again a strongly continuous semigroup.

Moreover, as already used in the introduction, \mathbb{S}^d denotes the vector space of symmetric $d \times d$ matrices and \mathbb{S}_+^d the cone of positive semidefinite ones. Furthermore, we denote by $\text{diag}(A)$ the vector consisting of the diagonal elements of a matrix A .

2 Generalized Feller semigroups and processes

In the context of Markovian lifts of stochastic Volterra processes (signed), measure-valued processes appear in a natural way. The generalized Feller framework is tailor-made for such processes, as it allows to consider non-locally compact state spaces, going beyond the standard theory of Feller processes as provided e.g. in Ethier and Kurtz (1986). This is explicitly needed in Sect. 5 for Ornstein-Uhlenbeck processes which take values in the space of matrix-valued measures. Beyond that jump processes with unbounded, but finite activity can be easily constructed in this setting, see Proposition 3.4 and Sect. 4. We shall first collect some results from Cuchiero and Teichmann (2018) and generalize accordingly for the purposes of this article.

2.1 Definitions and results

First, we introduce weighted spaces and state a central Riesz–Markov–Kakutani representation result. The underlying space X here is a completely regular Hausdorff topological space.

Definition 2.1 A function $\varrho: X \rightarrow (0, \infty)$ is called *admissible weight function* if the sets $K_R := \{x \in X: \varrho(x) \leq R\}$ are compact and separable for all $R > 0$.

An admissible weight function ϱ is necessarily lower semicontinuous and bounded from below by a positive constant. We call the pair X together with an admissible weight function ϱ a *weighted space*. A weighted space is σ -compact. In the following remark, we clarify the question of local compactness of convex subsets $\mathcal{E} \subset X$ when X is a locally convex topological space and ϱ convex.

Remark 2.2 Let X be a separable locally convex topological space and \mathcal{E} a convex subset. Moreover, let ϱ be a *convex* admissible weight function. Then, ϱ is continuous on \mathcal{E} if and only if \mathcal{E} is locally compact. Indeed, if ϱ is continuous on \mathcal{E} , then of course, the topology on \mathcal{E} is locally compact since every point has a compact neighborhood of type $\{\varrho \leq R\}$ for some $R > 0$. On the other hand, if the topology on \mathcal{E} is locally compact, then for every point $\lambda_0 \in \mathcal{E}$, there is a convex, compact neighborhood $V \subset \mathcal{E}$ such that $\varrho(\lambda) - \varrho(\lambda_0)$ is bounded on V by a number $k > 0$, whence by convexity $|\varrho(s(\lambda - \lambda_0) + \lambda_0) - \varrho(\lambda_0)| \leq sk$ for $\lambda - \lambda_0 \in s(V - \lambda_0)$ and $s \in]0, 1]$. This in turn means that ϱ is continuous at λ_0 .

From now on, ϱ shall always denote an admissible weight function. For completeness, we start by putting definitions for general Banach space valued functions, although in the sequel, we shall only deal with \mathbb{R} -valued functions: Let Z be a Banach space with norm $\|\cdot\|_Z$. The vector space

$$B^\varrho(X; Z) := \left\{ f: X \rightarrow Z: \sup_{x \in X} \varrho(x)^{-1} \|f(x)\|_Z < \infty \right\} \tag{2.1}$$

of Z -valued functions f equipped with the norm

$$\|f\|_\varrho := \sup_{x \in X} \varrho(x)^{-1} \|f(x)\|_Z, \tag{2.2}$$

is a Banach space itself. It is also clear that for Z -valued bounded continuous functions, the continuous embedding $C_b(X; Z) \subset B^\varrho(X; Z)$ holds true, where we consider the supremum norm on bounded continuous functions, i.e., $\sup_{x \in X} \|f(x)\|$.

Definition 2.3 We define $\mathcal{B}^\varrho(X; Z)$ as the closure of $C_b(X; Z)$ in $B^\varrho(X; Z)$. The normed space $\mathcal{B}^\varrho(X; Z)$ is a Banach space.

If the range space $Z = \mathbb{R}$, which from now on will be the case, we shall write $\mathcal{B}^\varrho(X)$ for $\mathcal{B}^\varrho(X; \mathbb{R})$ and analogously $B^\varrho(X)$.

We consider elements of $\mathcal{B}^\varrho(X)$ as continuous functions whose growth is controlled by ϱ . More precisely, we have by Dörsek and Teichmann (2010, Theorem 2.7) that $f \in \mathcal{B}^\varrho(X)$ if and only if $f|_{K_R} \in C(K_R)$ for all $R > 0$ and

$$\lim_{R \rightarrow \infty} \sup_{x \in X \setminus K_R} \varrho(x)^{-1} \|f(x)\| = 0. \tag{2.3}$$

Additionally, by Dörsek and Teichmann (2010, Theorem 2.8), it holds that for every $f \in \mathcal{B}^\varrho(X)$ with $\sup_{x \in X} f(x) > 0$, there exists $z \in X$ such that

$$\varrho(x)^{-1} f(x) \leq \varrho(z)^{-1} f(z) \quad \text{for all } x \in X, \tag{2.4}$$

which emphasizes the analogy with spaces of continuous functions vanishing at ∞ on locally compact spaces.

Let us now state the following crucial representation theorem of Riesz type:

Theorem 2.4 (Riesz representation for $\mathcal{B}^\varrho(X)$) *For every continuous linear functional $\ell: \mathcal{B}^\varrho(X) \rightarrow \mathbb{R}$ there exists a finite signed Radon measure μ on X such that*

$$\ell(f) = \int_X f(x) \mu(dx) \quad \text{for all } f \in \mathcal{B}^\varrho(X). \tag{2.5}$$

Additionally,

$$\int_X \varrho(x) |\mu|(dx) = \|\ell\|_{L(\mathcal{B}^\varrho(X), \mathbb{R})}, \tag{2.6}$$

where $|\mu|$ denotes the total variation measure of μ .

We shall next consider strongly continuous semigroups on $\mathcal{B}^\varrho(X)$ spaces and recover very similar structures as well known for Feller semigroups on the space of continuous functions vanishing at ∞ on locally compact spaces.

Definition 2.5 A family of bounded linear operators $P_t: \mathcal{B}^\varrho(X) \rightarrow \mathcal{B}^\varrho(X)$ for $t \geq 0$ is called *generalized Feller semigroup* if

- (i) $P_0 = I$, the identity on $\mathcal{B}^q(X)$,
- (ii) $P_{t+s} = P_t P_s$ for all $t, s \geq 0$,
- (iii) for all $f \in \mathcal{B}^q(X)$ and $x \in X$, $\lim_{t \rightarrow 0} P_t f(x) = f(x)$,
- (iv) there exist a constant $C \in \mathbb{R}$ and $\varepsilon > 0$ such that for all $t \in [0, \varepsilon]$, $\|P_t\|_{L(\mathcal{B}^q(X))} \leq C$.
- (v) P_t is positive for all $t \geq 0$, that is, for $f \in \mathcal{B}^q(X)$, $f \geq 0$, we have $P_t f \geq 0$.

We obtain due to the Riesz representation property the following key theorem:

Theorem 2.6 *Let $(P_t)_{t \geq 0}$ satisfy (i) to (iv) of Definition 2.5. Then, $(P_t)_{t \geq 0}$ is strongly continuous on $\mathcal{B}^q(X)$, that is,*

$$\lim_{t \rightarrow 0} \|P_t f - f\|_q = 0 \quad \text{for all } f \in \mathcal{B}^q(X). \tag{2.7}$$

One can also establish a positive maximum principle in case that the semigroup P_t grows around 0 like $\exp(\omega t)$ for some $\omega \in \mathbb{R}$ with respect to the operator norm on $\mathcal{B}^q(X)$. Indeed, the following theorem proved in Dörsek and Teichmann (2010, Theorem 3.3) is a reformulation of the Lumer–Phillips theorem for pseudo-contraction semigroups using a *generalized positive maximum principle* which is formulated in the sequel.

Theorem 2.7 *Let A be an operator on $\mathcal{B}^q(X)$ with domain D , and $\omega \in \mathbb{R}$. A is closable with its closure \bar{A} generating a generalized Feller semigroup $(P_t)_{t \geq 0}$ with $\|P_t\|_{L(\mathcal{B}^q(X))} \leq \exp(\omega t)$ for all $t \geq 0$ if and only if*

- (i) D is dense,
- (ii) $A - \omega_0$ has dense image for some $\omega_0 > \omega$, and
- (iii) A satisfies the generalized positive maximum principle, that is, for $f \in D$ with $(\varrho^{-1} f) \vee 0 \leq \varrho(z)^{-1} f(z)$ for some $z \in X$, $Af(z) \leq \omega f(z)$.

As a new contribution to the general theorems, we shall work out a statement on invariant subspaces which will be crucial for constructing squares of infinite dimensional OU processes.

Theorem 2.8 *Let X be a weighted space with weight ϱ_1 and $q : X \rightarrow q(X)$ be a (surjective) continuous map from (X, ϱ_1) to the weighted space $(q(X), \varrho_2)$. Let $P^{(1)}$ be a generalized Feller semigroup acting on $\mathcal{B}^{q_1}(X)$. Assume that $\varrho_2 \circ q \leq \varrho_1$ on X . Let D be a dense subspace of $\mathcal{B}^{q_2}(q(X))$. Furthermore, for every $f \in D \subset \mathcal{B}^{q_2}(q(X))$ and for every $t \geq 0$, there is some $g \in \mathcal{B}^{q_2}(q(X))$ such that*

$$P_t^{(1)}(f \circ q) = g \circ q, \tag{2.8}$$

and additionally, there is a constant $C \geq 1$ such that

$$P_t^{(1)}(\varrho_2 \circ q) \leq C \varrho_2 \circ q. \tag{2.9}$$

Then, there is a generalized Feller semigroup $P^{(2)}$ acting on $\mathcal{B}^{q_2}(q(X))$ such that

$$P_t^{(1)}(f \circ q) = (P_t^{(2)} f) \circ q. \tag{2.10}$$

Proof The continuous map q defines a linear operator M from $\mathcal{B}^{\varrho_2}(q(X))$ to $\mathcal{B}^{\varrho_1}(X)$ via $f \mapsto f \circ q$. Notice that M is bounded, since

$$\|Mf\|_{\varrho_1} \leq \|f\|_{\varrho_2}, \quad f \in \mathcal{B}^{\varrho_2}(q(X))$$

due to the assumption $\varrho_2 \circ q \leq \varrho_1$. It is also injective, but its image is not necessarily closed. Assumptions (2.8) and (2.9) now mean that

$$P_t^{(1)}Mf \in \text{rg}(M)$$

for every $f \in \mathcal{B}^{\varrho_2}(q(X))$ and not only for $f \in D$. Hence, we can define

$$P_t^{(2)}f := M^{-1}P_t^{(1)}Mf,$$

which is by the very construction a semigroup of linear operators on $\mathcal{B}^{\varrho_2}(q(X))$. Since M is continuous, its graph is closed, whence $P_t^{(2)}$ is a bounded linear operator by the closed graph theorem. Moreover, property (iv) of Definition 2.5 holds true due to Assumption (2.9). Positivity is also preserved, since for $f \geq 0$, we have due to Assumption (2.8) and the fact that $P^{(1)}$ is a generalized Feller semigroup,

$$P_t^{(2)}f = M^{-1}P_t^{(1)}Mf = M^{-1}\underbrace{P_t^{(1)}(f \circ q)}_{\geq 0} = M^{-1}(g \circ q) = g \geq 0.$$

Here, g is nonnegative due the positivity of $P_t^{(1)}(f \circ q)$. By (2.8) and the definition of $P^{(2)}$, (2.10) clearly holds true. Hence,

$$\lim_{t \rightarrow 0} P_t^{(2)}f(q(x)) = \lim_{t \rightarrow 0} P_t^{(1)}f(q(x)) = f(q(x))$$

for $x \in X$ and thus property (iii) of Definition 2.5. Hence, all conditions of Definition 2.5 are satisfied and we can conclude that the operators $(P_t^{(2)})$ form a generalized Feller semigroup.

□

Remark 2.9 In the setting of general semigroups, it is not clear that restrictions of semigroups to (not even closed) subspaces preserve strong continuity.

Remark 2.10 There are several methods to show that (2.8) is satisfied. In general, it is not sufficient to assume that the generator of $P^{(1)}$ has this property.

Corollary 2.11 *Let the assumptions of Theorem 2.8 except Assumption (2.9) hold true and suppose additionally that*

$$\varrho_2 \circ q = \varrho_1.$$

Then, the same conclusions hold true. In particular, the range of the operator $M : \mathcal{B}^{\varrho_2}(q(X)) \rightarrow \mathcal{B}^{\varrho_1}(X)$, $f \mapsto f \circ q$ is closed.

We restate from Cuchiero and Teichmann (2018) assertions on existence of generalized Feller processes and path properties. It is remarkable that in this very general context, càg versions exist for countably many test functions.

Theorem 2.12 *Let $(P_t)_{t \geq 0}$ be a generalized Feller semigroup with $P_t 1 = 1$ for $t \geq 0$. Then, there exists a filtered measurable space $(\Omega, (\mathcal{F}_t)_{t \geq 0})$ with right continuous filtration, and an adapted family of random variables $(\lambda_t)_{t \geq 0}$ such that for any initial value $\lambda_0 \in X$ there exists a probability measure \mathbb{P}^{λ_0} with*

$$\mathbb{E}_{\lambda_0}[f(\lambda_t)] := \mathbb{E}_{\mathbb{P}^{\lambda_0}}[f(\lambda_t)] = P_t f(\lambda_0)$$

for $t \geq 0$ and every $f \in \mathcal{B}^q(X)$. The Markov property holds true, i.e.,

$$\mathbb{E}_{\mathbb{P}^{\lambda_0}}[f(\lambda_t) \mid \mathcal{F}_s] = P_{t-s} f(\lambda_s)$$

almost surely with respect to \mathbb{P}^{λ_0} .

Theorem 2.13 *Let $(P_t)_{t \geq 0}$ be a generalized Feller semigroup, and let $(\lambda_t)_{t \geq 0}$ be a generalized Feller process on a filtered probability space. Then, for every countable family $(f_n)_{n \geq 0}$ of functions in $\mathcal{B}^q(X)$, we can choose a version of the processes $\left(\frac{f_n(\lambda_t)}{\varrho(\lambda_t)}\right)_{t \geq 0}$, such that the trajectories are càglàd for all $n \geq 0$. If additionally $P_t \varrho \leq \exp(\omega t) \varrho$ holds true, then $(\exp(-\omega t) \varrho(\lambda_t))_{t \geq 0}$ is a super-martingale and can be chosen to have càglàd trajectories. In this case, we obtain that the processes $(f_n(\lambda_t))_{t \geq 0}$ can be chosen to have càglàd trajectories.*

Remark 2.14 In the general case, when $P_t \varrho \leq M \exp(\omega t) \varrho$ for $M > 1$, we obtain for $(f_n(\lambda_t))_{t \geq 0}$ only càg trajectories. To see this, consider the measurable set of sample events $\{\sup_{0 \leq t \leq 1} \varrho(\lambda_t) \leq R\}$. Then, we can construct on the metrizable compact set $\{\varrho \leq R\}$ a càglàd version of the processes $\left(\frac{f_n(\lambda_t)}{\varrho(\lambda_t)}\right)_{t \leq 1}$ and $\left(\frac{1}{\varrho(\lambda_t)}\right)_{t \leq 1}$ and in turn also of $(f_n(\lambda_t))_{t \geq 0}$. The limit $R \rightarrow \infty$, however, only leads to a càg version since we cannot control the right limits.

2.2 Dual spaces of Banach spaces

The most important playground for our theory will be closed subsets of duals of Banach spaces, where the weak- $*$ -topology appears to be σ -compact due to the Banach–Alaoglu theorem. Assume that $\mathcal{E} \subset Y^*$ is a closed subset of the dual space Y^* of some Banach space Y where Y^* is equipped with its weak- $*$ -topology. Consider a lower semicontinuous function $\varrho: \mathcal{E} \rightarrow (0, \infty)$ and denote by (\mathcal{E}, ϱ) the corresponding weighted space. We have the following approximation result (see Döörsek and Teichmann (2010, Theorem 4.2)) for functions in $\mathcal{B}^q(\mathcal{E})$ by cylindrical functions. Set

$$\begin{aligned} \text{Cyl}_N := \{ & g(\langle \cdot, y_1 \rangle, \dots, \langle \cdot, y_N \rangle) : g \in C_b^\infty(\mathbb{R}^N) \\ & \text{and } y_j \in Y, j = 1, \dots, N\}, \end{aligned} \quad (2.11)$$

where $\langle \cdot, \cdot \rangle$ denotes the pairing between Y^* and Y . We denote by $\text{Cyl} := \bigcup_{N \in \mathbb{N}} \text{Cyl}_N$ the set of bounded smooth continuous cylinder functions on \mathcal{E} .

Theorem 2.15 *The closure of Cyl in $B^q(\mathcal{E})$ coincides with $\mathcal{B}^q(\mathcal{E})$, whose elements appear to be precisely the functions $f \in \mathcal{B}^q(\mathcal{E})$ which satisfy (2.3) and that $f|_{K_R}$ is weak- $*$ -continuous for any $R > 0$.*

Proof See Cuchiero and Teichmann (2018). □

Assumption 2.16 Let $(\lambda_t)_{t \geq 0}$ denote a time homogeneous Markov process on some stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P}^{\lambda_0})$ with values in \mathcal{E} .

Then, we assume that

- (i) there are constants C and $\varepsilon > 0$ such that

$$\mathbb{E}_{\lambda_0}[\varrho(\lambda_t)] \leq C\varrho(\lambda_0) \quad \text{for all } \lambda_0 \in \mathcal{E} \text{ and } t \in [0, \varepsilon]; \tag{2.12}$$

- (ii)

$$\lim_{t \rightarrow 0} \mathbb{E}_{\lambda_0}[f(\lambda_t)] = f(\lambda_0) \quad \text{for any } f \in \mathcal{B}^q(\mathcal{E}) \text{ and } \lambda_0 \in \mathcal{E}; \tag{2.13}$$

- (iii) for all f in a dense subset of $\mathcal{B}^q(\mathcal{E})$, the map $\lambda_0 \mapsto \mathbb{E}_{\lambda_0}[f(\lambda_t)]$ lies in $\mathcal{B}^q(\mathcal{E})$.

Remark 2.17 Of course inequality (2.12) implies that $|\mathbb{E}_{\lambda_0}[f(\lambda_t)]| \leq C\varrho(\lambda_0)$ for all $f \in \mathcal{B}^q(\mathcal{E})$, $\lambda_0 \in \mathcal{E}$ and $t \in [0, \varepsilon]$.

Theorem 2.18 *Suppose Assumptions 2.16 hold true. Then, $P_t f(\lambda_0) := \mathbb{E}_{\lambda_0}[f(\lambda_t)]$ satisfies the generalized Feller property and is therefore a strongly continuous semigroup on $\mathcal{B}^q(\mathcal{E})$.*

Proof This follows from the arguments of Dörsek and Teichmann (2010, Section 5). □

3 Approximation theorems

In order to establish existence of Markovian solutions for general generators A , we could at least in the pseudo-contractive case either directly apply Theorem 2.7, where we have to assume that the generator A satisfies on a dense domain D a generalized positive maximum principle and that for at least one $\omega_0 > \omega$ the range of $A - \omega_0$ is dense, or we approximate a general generator A by (finite activity pure jump) generators A^n and apply the following (well known) approximation theorems. They also work in the general context when the constant $M > 1$.

Theorem 3.1 *Let $(P_t^n)_{n \in \mathbb{N}, t \geq 0}$ be a sequence of strongly continuous semigroups on a Banach space Z with generators $(A^n)_{n \in \mathbb{N}}$ such that there are uniform (in n) growth bounds $M \geq 1$ and $\omega \in \mathbb{R}$ with*

$$\|P_t^n\|_{L(Z)} \leq M \exp(\omega t) \tag{3.1}$$

for $t \geq 0$. Let furthermore $D \subset \bigcap_n \text{dom}(A^n)$ be a dense subspace with the following three properties:

- (i) D is an invariant subspace for all P^n , i.e., for all $f \in D$, we have $P_t^n f \in D$, for $n \geq 0$ and $t \geq 0$.
- (ii) There is a norm $\|\cdot\|_D$ on D such that there are uniform growth bounds with respect to $\|\cdot\|_D$, i.e., there are $M_D \geq 1$ and $\omega_D \in \mathbb{R}$ with

$$\|P_t^n f\|_D \leq M_D \exp(\omega_D t) \|f\|_D$$

for $t \geq 0$ and for $n \geq 0$.

- (iii) The sequence $A^n f$ converges as $n \rightarrow \infty$ for each $f \in D$, in the following sense: There exists a sequence of numbers $a_{nm} \rightarrow 0$ as $n, m \rightarrow \infty$ such that

$$\|A^n f - A^m f\| \leq a_{nm} \|f\|_D$$

holds true for every $f \in D$ and for all n, m .

Then, there exists a strongly continuous semigroup $(P_t^\infty)_{t \geq 0}$ with the same growth bound on Z such that $\lim_{n \rightarrow \infty} P_t^n f = P_t^\infty f$ for all $f \in Z$ uniformly on compacts in time and on bounded sets in D . Furthermore on D , the convergence is of order $O(a_{nm})$. If in addition for each $n \in \mathbb{N}$, $(P_t^n)_{t \geq 0}$ is a generalized Feller semigroup, then this property transfers also to the limiting semigroup.

Proof See Cuchiero and Teichmann (2018). □

For the purposes of affine processes, a slightly more general version of the approximation theorem is needed, which we state in the sequel:

Theorem 3.2 Let $(P_t^n)_{n \in \mathbb{N}, t \geq 0}$ be a sequence of strongly continuous semigroups on a Banach space Z with generators $(A^n)_{n \in \mathbb{N}}$ such that there are uniform (in n) growth bounds $M \geq 1$ and $\omega \in \mathbb{R}$ with

$$\|P_t^n\|_{L(Z)} \leq M \exp(\omega t)$$

for $t \geq 0$. Let furthermore $D \subset \bigcap_n \text{dom}(A^n)$ be a subset with the following two properties:

- (i) The linear span $\text{span}(D)$ is dense.
- (ii) There is a norm $\|\cdot\|_D$ on $\text{span}(D)$ such that for each $f \in D$ and for $t > 0$, there exists a sequence $a_{nm}^{f,t}$, possibly depending on f and t ,

$$\|A^n P_u^m f - A^m P_u^m f\| \leq a_{nm}^{f,t} \|f\|_D$$

holds true for n, m and for $0 \leq u \leq t$, with $a_{nm}^{f,t} \rightarrow 0$ as $n, m \rightarrow \infty$.

Then, there exists a strongly continuous semigroup $(P_t^\infty)_{t \geq 0}$ with the same growth bound on Z such that $\lim_{n \rightarrow \infty} P_t^n f = P_t^\infty f$ for all $f \in Z$ uniformly on compacts in time. If in addition for each $n \in \mathbb{N}$, $(P_t^n)_{t \geq 0}$ is a generalized Feller semigroup, then this property transfers also to the limiting semigroup.

Proof See Cuchiero and Teichmann (2018). □

Our first application of Theorem 3.1 is the next proposition that extends well-known results on bounded generators toward unbounded limits.

We repeat here a remark from Cuchiero and Teichmann (2018) since it helps to understand the fourth condition on the measures:

Remark 3.3 Let $(P_t)_{t \geq 0}$ be a generalized Feller semigroup with $\|P_t\|_{L(\mathcal{B}^q(X))} \leq M \exp(\omega t)$ for some $M \geq 1$ and some ω . Additionally, it is assumed to be of transport type, i.e.,

$$P_t f(x) = f(\psi_t(x)) \tag{3.2}$$

for some continuous map $\psi_t : X \rightarrow X$. Define now a new function

$$\tilde{\varrho}(x) := \sup_{t \geq 0} \exp(-\omega t) P_t \varrho(x)$$

for $x \in X$. Notice that $\tilde{\varrho}$ is an admissible weight function, since

$$\{\tilde{\varrho} \leq R\} = \bigcap_{t \geq 0} \{P_t \varrho \leq \exp(\omega t) R\} \subseteq \{\varrho \leq R\}$$

is compact by the definition of ϱ and the continuity of $x \mapsto \psi_t(x)$ which leads to an intersection of closed subsets of compacts. Additionally, we have that

$$\varrho \leq \tilde{\varrho} \leq M \varrho$$

by the growth bound, and therefore, the norm on $\mathcal{B}^q(X)$ is equivalent to

$$\|f\|_{\tilde{\varrho}} = \sup_{x \in X} \frac{|f(x)|}{\tilde{\varrho}(x)}.$$

Furthermore,

$$\|P_t f\|_{\tilde{\varrho}} \leq \exp(\omega t) \|f\|_{\tilde{\varrho}}$$

holds for all $t \geq 0$ and $f \in \mathcal{B}^q(X)$. Indeed, this is a consequence of the following estimate

$$\begin{aligned} \|P_t f\|_{\tilde{\varrho}} &= \sup_x \left| \frac{f(\psi_t(x))}{\sup_s \exp(-\omega s) \varrho(\psi_s(x))} \right| \leq \sup_x \left| \frac{f(\psi_t(x))}{\sup_s \exp(-\omega(t+s)) \varrho(\psi_{t+s}(x))} \right| \\ &\leq \exp(\omega t) \sup_x \left| \frac{f(\psi_t(x))}{\sup_s \exp(-\omega s) \varrho(\psi_s(\psi_t(x)))} \right| \leq \exp(\omega t) \|f\|_{\tilde{\varrho}}. \end{aligned}$$

Hence,

$$|P_t f(x)| \leq \exp(\omega t) \tilde{\varrho}(x) \|f\|_{\tilde{\varrho}},$$

which implies

$$P_t \tilde{Q} \leq \exp(\omega t) \tilde{Q}, \quad t \geq 0.$$

Proposition 3.4 *Let (X, ϱ) be a weighted space with weight function $\varrho \geq 1$. Consider an operator A on $\mathcal{B}^\varrho(X)$ with dense domain $\text{dom}(A)$ generating on $\mathcal{B}^\varrho(X)$ a generalized Feller semigroup $(P_t)_{t \geq 0}$ of transport type as in (3.2), such that for all $t \geq 0$, we have $\|P_t\|_{L(\mathcal{B}^\varrho(X))} \leq M_1 \exp(\omega t)$ for some M_1 and ω and such that $\mathcal{B}^{\sqrt{\varrho}}(X) \subset \mathcal{B}^\varrho(X)$ is left invariant.*

Consider furthermore a family of finite measures $\mu(x, \cdot)$ for $x \in X$ on X such that the operator B acts on $\mathcal{B}^\varrho(X)$ by

$$Bf(x) := \int (f(y) - f(x))\mu(x, dy)$$

for $x \in X$ yielding continuous functions on $\{\varrho \leq R\}$ for $R \geq 0$, and such that the following properties hold true:

- For all $x \in X$

$$\int \varrho(y)\mu(x, dy) \leq M\varrho^2(x), \quad (3.3)$$

as well as

$$\int \sqrt{\varrho(y)}\mu(x, dy) \leq M\varrho(x), \quad (3.4)$$

and

$$\int \mu(x, dy) \leq M\sqrt{\varrho(x)}, \quad (3.5)$$

hold true for some constant M .

- For some constant $\tilde{\omega} \in \mathbb{R}$,

$$\int \left| \frac{\sup_{t \geq 0} \exp(-\omega t) P_t \varrho(y) - \sup_{t \geq 0} \exp(-\omega t) P_t \varrho(x)}{\sup_{t \geq 0} \exp(-\omega t) P_t \varrho(x)} \right| \mu(x, dy) \leq \tilde{\omega}, \quad (3.6)$$

for all $x \in X$. In particular, $y \mapsto \sup_{t \geq 0} \exp(-\omega t) P_t \varrho(y)$ should be integrable with respect to $\mu(x, \cdot)$

Then, $A + B$ generates a generalized Feller semigroup $(P_t^\infty)_{t \geq 0}$ on $\mathcal{B}^\varrho(X)$ satisfying $\|P_t^\infty\|_{L(\mathcal{B}^\varrho(X))} \leq M_1 \exp((\omega + \tilde{\omega})t)$.

Proof See Cuchiero and Teichmann (2018). □

Remark 3.5 In contrast to classical Feller theory, also processes with unbounded jump intensities can be constructed easily if ϱ is unbounded on X . The general character of the proposition allows to build general processes from simple ones by perturbation.

4 Lifting stochastic Volterra jump processes with values in \mathbb{S}_+^d

Building on the theory of generalized Feller processes from the above, we shall now treat the following type of matrix measure-valued SPDEs

$$\begin{aligned} d\lambda_t(dx) &= \mathcal{A}^* \lambda_t(dx) dt + \nu(dx) dX_t + dX_t \nu(dx), \\ \lambda_0 &\in \mathcal{E}. \end{aligned} \tag{4.1}$$

As shown below, this equation corresponds to a Markovian lift of the Volterra jump process in (1.2).

We consider here the setting of Sect. 2.2. The underlying Banach space Y^* is here the space of finite \mathbb{S}^d -valued regular Borel measures on the extended half real line $\overline{\mathbb{R}}_+ := \mathbb{R}_+ \cup \{\infty\}$, and \mathcal{E} denotes a (positive definite) subset of Y^* . Moreover, \mathcal{A}^* is the generator of a strongly continuous semigroup \mathcal{S}^* on Y^* , $\nu \in Y^*$ (or in a slightly larger space denoted by Z^* in the sequel). The pre-dual space Y is given by $C_b(\overline{\mathbb{R}}_+, \mathbb{S}^d)$ functions. Note that since $\overline{\mathbb{R}}_+$ is compact, $Y = C_b(\overline{\mathbb{R}}_+, \mathbb{S}^d)$ is separable. The driving process X is an \mathbb{S}^d -valued pure jump Itô-semimartingale, whose differential characteristics depend linearly on λ , precisely specified below. Let us remark that other forms of differential characteristics of X , in particular beyond the linear case, can be easily incorporated in this setting.

The pairing between Y and Y^* , denoted by $\langle \cdot, \cdot \rangle$, is specified via:

$$\langle \cdot, \cdot \rangle : Y \times Y^* \rightarrow \mathbb{R}, \quad (y, \lambda) \mapsto \langle y, \lambda \rangle = \text{Tr} \left(\int_0^\infty y(x) \lambda(dx) \right),$$

where Tr denotes the trace. We also define another bilinear map via

$$\langle \langle \cdot, \cdot \rangle \rangle : Y \times Y^* \rightarrow \mathbb{S}^d, \quad (y, \lambda) \mapsto \langle \langle y, \lambda \rangle \rangle = \int_0^\infty y(x) \lambda(dx) + \int_0^\infty \lambda(dx) y(x). \tag{4.2}$$

In the following, we summarize the main ingredients of our setting. For the norm on \mathbb{S}^d , we write $\| \cdot \|$, which is given by $\|u\| = \sqrt{\text{Tr}(u^2)}$ for $u \in \mathbb{S}^d$.

Assumption 4.1 Throughout this section, we shall work under the following conditions:

- (i) We are given an admissible weight function ϱ on Y^* (in the sense of Sect. 2) such that

$$\varrho(\lambda) = 1 + \|\lambda\|_{Y^*}^2, \quad \lambda \in Y^*,$$

where $\| \cdot \|_{Y^*}$ denotes the norm on Y^* , which is the total variation norm of λ .

- (ii) We are given a closed convex cone $\mathcal{E} \subset Y^*$ (in the sequel the cone of \mathbb{S}_+^d valued measures) such that (\mathcal{E}, ϱ) is a weighted space in the sense of Sect. 2. This will serve as state space of (4.1).

- (iii) Let $Z \subset Y$ be a continuously embedded subspace.
- (iv) We assume that a semigroup \mathcal{S}^* with generator \mathcal{A}^* acts in a strongly continuous way on Y^* and Z^* , with respect to the respective norm topologies. Moreover, we suppose that for any matrix $A \in \mathbb{S}^d$, it holds that

$$\mathcal{S}_t^*(\lambda(\cdot)A + A\lambda(\cdot)) = (\mathcal{S}_t^*\lambda(\cdot))A + A(\mathcal{S}_t^*\lambda(\cdot)). \tag{4.3}$$

- (v) We assume that $\lambda \mapsto \mathcal{S}_t^*\lambda$ is weak- $*$ -continuous on Y^* and on Z^* for every $t \geq 0$ (considering the weak- $*$ -topology on both the domain and the image space).
- (vi) We suppose that the (pre-) adjoint operator of \mathcal{A}^* , denoted by \mathcal{A} and domain $\text{dom}(\mathcal{A}) \subset Z \subset Y$, generates a strongly continuous semigroup on Z with respect to the respective norm topology (but *not* necessarily on Y).

To analyze solvability of (4.1), we first consider the following linear deterministic equation

$$d\lambda_t(dx) = \mathcal{A}^*\lambda_t(dx)dt + \nu(dx)\beta(\lambda_t(\cdot))dt + \beta(\lambda_t(\cdot))\nu(dx)dt \tag{4.4}$$

for $\lambda_0 \in Y^*$, $\nu \in Z^*$ and β a bounded linear operator from $Y^* \rightarrow \mathbb{S}^d$ which satisfies for $A \in \mathbb{S}^d$ and $\lambda \in Y^*$

$$\beta(\lambda(\cdot)A + A\lambda(\cdot)) = \beta(\lambda(\cdot))A + A\beta(\lambda(\cdot)). \tag{4.5}$$

We denote by $\beta_* : \mathbb{S}^d \rightarrow Y$ the adjoint operator defined via

$$\text{Tr}(u\beta(\lambda)) = \text{Tr}\left(\int_0^\infty \beta_*(u)(x)\lambda(dx)\right) = \langle \beta_*(u), \lambda \rangle, \quad u \in \mathbb{S}^d, \lambda \in Y^*.$$

Remark 4.2 Notice that drift specifications could be more general here, but for the sake or readability, we leave this direction for the interested reader.

For notational convenience, we shall often leave the dx argument away when writing an (S)PDE of type (4.4) subsequently. Under the following assumptions on \mathcal{S}^* and $\nu \in Z^*$, we can guarantee that (4.4) can be solved on the space Y^* for all times in the mild sense with respect to the dual norm $\|\cdot\|_{Y^*}$ by a standard Picard iteration method.

Assumption 4.3 We assume that

- (i) $\mathcal{S}_t^*\nu \in Y^*$ for all $t > 0$ even though ν does not necessarily lie in Y^* itself, but only in Z^* ;
- (ii) $\int_0^t \|\mathcal{S}_s^*\nu\|_{Y^*}^2 ds < \infty$ for all $t > 0$.

For the linear operator β as of (4.5), we define

$$K(t) := \beta(\mathcal{S}_t^*\nu), \tag{4.6}$$

which will correspond to a kernel in $L^2_{\text{loc}}(\mathbb{R}_+, \mathbb{S}^d)$ of a Volterra equation. Define furthermore $R_K \in L^2_{\text{loc}}(\mathbb{R}_+, \mathbb{S}^d)$ as a symmetrized version of the resolvent of the second kind [(see, e.g., Gripenberg et al. (1990, Theorem 3.1)] that solves

$$K * R_K + R_K * K = K - R_K, \tag{4.7}$$

where $K * R_K$ denotes the convolution, i.e., $K * R_K = \int_0^\cdot K(\cdot - s)R_K(s)ds$.

Example 4.4 The main examples that we have in mind for β and for \mathcal{S}^* , and thus in turn for the kernel K , are the following specifications:

$$\beta(\lambda) = \int_0^\infty \lambda(dx), \quad \mathcal{S}_t^* \nu(dx) = e^{-xt} \nu(dx).$$

In this case, $K = \int_0^\infty e^{-xt} \nu(dx)$ and the adjoint operator β_* is given by the constant function

$$(\beta_*(u))(x) = u, \quad \text{for all } x \in \mathbb{R}_+.$$

Remark 4.5 To the semigroup $\mathcal{S}_t^* = e^{-xt}$ of the above example, we associate our (main) specification of the space Z : Let $Z \subset Y$ such that for all $y \in Y$ the map

$$h_y : \overline{\mathbb{R}}_+ \rightarrow \mathbb{S}^d, \quad x \mapsto xy(x)$$

lies in Z equipped with the operator norm, i.e.,

$$\|h_y\|_Z = \sqrt{\sup_{x \geq 0} \|y(x)\| + \sup_{x \geq 0} \|xy(x)\|} \text{ for } h_y \in Z.$$

The corresponding dual space $Z^* \supset Y^*$ is the space of regular \mathbb{S}^d -valued Borel measures ν on $\overline{\mathbb{R}}_+$ that satisfy

$$\left\| \int_0^\infty \left(\frac{1}{x} \wedge 1 \right) \nu(dx) \right\| < \infty.$$

Note that we can specify the components of ν to be measures of the form

$$\nu_{ij}(dx) = x^{-\frac{1}{2} - H_{ij}}, \quad H_{ij} \in \left(0, \frac{1}{2} \right),$$

which gives rise to fractional kernels $K_{ij}(t) = \int_0^\infty e^{-xt} \nu_{ij}(dx) \approx t^{H_{ij} - \frac{1}{2}}$. These are in turn main ingredients of rough covariance modeling.

Remark 4.6 In this article, we choose to work with state spaces of matrix-valued measures using the representation of the kernel K as Laplace transform of a matrix-valued measure ν as specified in Example 4.4. We could, however, perform the same analysis

on a Hilbert space of forward covariance curves. This corresponds then to a multivariate analogon of Cuchiero and Teichmann (2018, Section 5.2).

Proposition 4.7 *Under Assumption 4.3, there exists a unique mild solution of (4.4) with values in Y^* . Additionally, the solution operator is a weak- $*$ -continuous map $\lambda_0 \mapsto \lambda_t$, for each $t > 0$, and the solution satisfies*

$$\varrho(\lambda_t) \leq C\varrho(\lambda_0), \quad \text{for all } \lambda_0 \in Y^* \text{ and } t \in [0, \varepsilon]$$

for some positive constants C and ε .

Remark 4.8 The unique mild solution of Equation (4.4) satisfies by means of (4.3) the variation of constants equation

$$\lambda_t = \mathcal{S}_t^* \lambda_0 + \int_0^t (\mathcal{S}_{t-s}^* \nu \beta(\lambda_s) + \beta(\lambda_s) \mathcal{S}_{t-s}^* \nu) ds,$$

for all $t \geq 0$. Applying the linear operator β and using property (4.5), we obtain a deterministic linear Volterra equation of the form

$$\begin{aligned} \beta(\lambda_t) &= \beta(\mathcal{S}_t^* \lambda_0) + \int_0^t \beta(\mathcal{S}_{t-s}^* \nu \beta(\lambda_s) + \beta(\lambda_s) \mathcal{S}_{t-s}^* \nu) ds \\ &= \beta(\mathcal{S}_t^* \lambda_0) + \int_0^t (K(t-s)\beta(\lambda_s) + \beta(\lambda_s)K(t-s)) ds \end{aligned} \quad (4.8)$$

where we have used (4.6).

Proof We follow the arguments of Cuchiero and Teichmann (2018) and translate the proof to the matrix-valued setting. We show first the completely standard convergence of the Picard iteration scheme with respect to the dual norm on Y^* . Define

$$\begin{aligned} \lambda_t^0 &= \lambda_0, \\ \lambda_t^{n+1} &= \mathcal{S}_t^* \lambda_0 + \int_0^t (\mathcal{S}_{t-s}^* \nu) \beta(\lambda_s^n) ds + \int_0^t \beta(\lambda_s^n) (\mathcal{S}_{t-s}^* \nu) ds, \quad n \geq 0. \end{aligned}$$

Then, by Assumption 4.3, (i) each λ_t^n lies Y^* . Consider now

$$\begin{aligned} \|\lambda_t^{n+1} - \lambda_t^n\|_{Y^*} &= \left\| \int_0^t (\mathcal{S}_{t-s}^* \nu) (\beta(\lambda_s^n) - \beta(\lambda_s^{n-1})) ds \right. \\ &\quad \left. + \int_0^t (\beta(\lambda_s^n) - \beta(\lambda_s^{n-1})) (\mathcal{S}_{t-s}^* \nu) ds \right\|_{Y^*} \\ &\leq 2\|\beta\|_{\text{op}} \int_0^t \|\mathcal{S}_{t-s}^* \nu\|_{Y^*} \|\lambda_s^n - \lambda_s^{n-1}\|_{Y^*} ds, \end{aligned}$$

where $\|\beta\|_{\text{op}}$ denotes the operator norm of β . Assumption 4.3 (ii) and an extended version of Gronwall's inequality see Dalang (1999, Lemma 15) then yield convergence

of $(\lambda_t^n)_{n \in \mathbb{N}}$ to some λ_t with respect to the dual norm $\|\cdot\|_{Y^*}$ uniformly in t on compact intervals. For details on strongly continuous semigroups and mild solutions, see Pazy (1983).

Having established the existence of a mild solution of (4.4) in Y^* , consider now the \mathbb{S}^d -valued process $\beta(\lambda_t)$:

$$\begin{aligned} \beta(\lambda_t) &= \beta(\mathcal{S}_t^* \lambda_0) + \int_0^t \beta(\mathcal{S}_{t-s}^* \nu \beta(\lambda_s) + \beta(\lambda_s) \mathcal{S}_{t-s}^* \nu) \, ds, \\ &= \beta(\mathcal{S}_t^* \lambda_0) + \int_0^t (\beta(\mathcal{S}_{t-s}^* \nu) \beta(\lambda_s) + \beta(\lambda_s) \beta(\mathcal{S}_{t-s}^* \nu)) \, ds \\ &= \beta(\mathcal{S}_t^* \lambda_0) + \int_0^t (R_K(t-s) \beta(\mathcal{S}_s^* \lambda_0) + \beta(\mathcal{S}_s^* \lambda_0) R_K(t-s)) \, ds \quad (4.9) \end{aligned}$$

where we applied property (4.5). Remember that R_K denotes the resolvent of the second kind of $K(t) = \beta(\mathcal{S}_t^* \nu)$ as introduced in (4.7) by means of which we can solve the above equation in terms of integrals of $t \mapsto \beta(\mathcal{S}_t^* \lambda_0)$. Since by assumption, \mathcal{S}^* is a weak- $*$ -continuous solution operator, the map $\lambda_0 \mapsto (t \mapsto \beta(\mathcal{S}_t^* \lambda_0))$ is weak- $*$ -continuous as a map from Y^* to $C(\mathbb{R}_+, \mathbb{S}^d)$ (with the topology of uniform convergence on compacts on $C(\mathbb{R}_+, \mathbb{S}^d)$). From (4.9), we thus infer that $\beta(\lambda_t)$ is weak- $*$ -continuous for every $t \geq 0$, which clearly translates to the solution map of Equation (4.4).

Finally, we have to show that the stated inequality for $\varrho(\lambda_t)$ holds true on small time intervals $[0, \varepsilon]$. Observe first that for $t \in [0, \varepsilon]$

$$\|\mathcal{S}_t^* \lambda\|_{Y^*}^2 \leq C \|\lambda\|_{Y^*}^2$$

for all $\lambda \in Y^*$ just by the assumption that \mathcal{S}_t^* is strongly continuous, for some constant $C \geq 1$. Furthermore for $t \in [0, \varepsilon]$,

$$\begin{aligned} \|\lambda_t\|_{Y^*}^2 &\leq 3 \left(C \|\lambda_0\|_{Y^*}^2 + t \int_0^t \|\mathcal{S}_{t-s}^* \nu \beta(\lambda_s)\|_{Y^*}^2 + t \int_0^t \|\beta(\lambda_s) \mathcal{S}_{t-s}^* \nu\|_{Y^*}^2 \right) \\ &\leq 3 \left(C \|\lambda_0\|_{Y^*}^2 + 2\varepsilon \|\beta\|_{\text{op}}^2 \int_0^t \|\mathcal{S}_{t-s}^* \nu\|_{Y^*}^2 \|\lambda_s\|_{Y^*}^2 \, ds \right). \end{aligned}$$

Consider now the kernel $K'(t, s) = 6\varepsilon \|\beta\|_{\text{op}}^2 \|\mathcal{S}_{t-s}^* \nu\|_{Y^*}^2 1_{\{s \leq t\}}$ and denote by R' the resolvent of $-K'$, which is non-positive. By exactly the same arguments as in Cuchiero and Teichmann (2018), we then have for $t \in [0, \varepsilon]$

$$\|\lambda_t\|_{Y^*}^2 \leq \tilde{C} \|\lambda_0\|_{Y^*}^2 \left(1 - \int_0^\varepsilon R'(s) \, ds \right),$$

for some constant \tilde{C} . This leads to the desired assertion due to the definition of ϱ . From this inequality, also uniqueness follows in a standard way. \square

As our goal is to consider \mathbb{S}_+^d -measure-valued processes, we denote by \mathcal{E} the following weak- $*$ -closed convex cone

$$\mathcal{E} = \{\lambda_0 \in Y^* \mid \lambda_0 \text{ is an } \mathbb{S}_+^d \text{-valued measure on } \overline{\mathbb{R}}_+\}.$$

The next proposition establishes that the solution of (4.4) leaves \mathcal{E} invariant, if the following assumption holds true:

Assumption 4.9 We assume that

- (i) $\mathcal{S}_t^*(\mathcal{E}) \subseteq \mathcal{E}, \quad \forall t \geq 0;$
- (ii) ν is an \mathbb{S}_+^d -valued measure;
- (iii) $\beta(\mathcal{E}) \subseteq \mathbb{S}_+^d.$

Proposition 4.10 *Let Assumptions 4.3 and 4.9 be in force. Then, the solution of (4.4) leaves \mathcal{E} invariant and it defines a generalized Feller semigroup on (\mathcal{E}, ϱ) by $P_t f(\lambda_0) := f(\lambda_t)$ for all $f \in \mathcal{B}^{\varrho}(\mathcal{E})$ and $t \geq 0$.*

Proof Consider first the slightly modified equation

$$d\lambda_t(dx) = \mathcal{A}^* \lambda_t(dx) dt + \mathcal{S}_\varepsilon^* \nu(dx) \beta(\lambda_t(\cdot)) dt + \beta(\lambda_t(\cdot)) \mathcal{S}_\varepsilon^* \nu(dx) dt \tag{4.10}$$

for some $\varepsilon > 0$. Then, the operator $B = \mathcal{S}_\varepsilon^* \nu(dx) \beta(\cdot) + \beta(\cdot) \mathcal{S}_\varepsilon^* \nu(dx)$ is bounded and the associated semigroup is given by $P_t^\varepsilon = e^{Bt}$. Due to the assumptions on \mathcal{S}^* , ν and β , we have $B(\mathcal{E}) \subseteq \mathcal{E}$ implying that $P_t^\varepsilon(\mathcal{E}) \subseteq \mathcal{E}$ for all $t \geq 0$. The Trotter-Kato theorem, see, e.g., Engel and Nagel (2000, Theorem III.5.8), then yields that the semigroup associated with (4.10) maps \mathcal{E} to itself. This then also holds true for the limit when $\varepsilon = 0$ by Theorem 3.1.

Since by Proposition 4.7, the solution operator is weak- $*$ -continuous, we can conclude that $\lambda_0 \mapsto f(\lambda_t)$ lies in $\mathcal{B}^{\varrho}(\mathcal{E})$ for a dense set of $\mathcal{B}^{\varrho}(\mathcal{E})$ by Theorem 2.15. Moreover, it satisfies the necessary bound (2.12) for ϱ and (2.13) is satisfied by (norm)-continuity of $t \mapsto \lambda_t$. Hence, all the conditions of Assumption 2.16 are satisfied and the solution operator therefore defines a generalized Feller semigroup (P_t) on $\mathcal{B}^{\varrho}(\mathcal{E})$ by Theorem 2.18. This generalized Feller semigroup of course coincides with the previously constructed limit. □

By the previous results, we can now construct a generalized Feller process on \mathcal{E} which jumps up by multiples of $\mathcal{S}_\varepsilon^* \nu$ for some $\varepsilon \geq 0$ and with an instantaneous intensity of size $\beta(\lambda_t)$. Recall that $\mathcal{E}_* \subset Y$ denotes the (pre-)polar cone of \mathcal{E} , that is,

$$\mathcal{E}_* = \{y \in Y \mid y \in C_b(\overline{\mathbb{R}}_+, \mathbb{S}_+^d)\}.$$

Recall the notation from (4.2) and define the following set

$$\mathcal{D} = \{y \in Y \mid y \in \text{dom}(\mathcal{A}) \text{ s.t. } \langle \cdot, y \rangle \text{ is well defined}\}. \tag{4.11}$$

Proposition 4.11 *Let Assumptions 4.3 and 4.9 be in force. Moreover, let μ be a finite \mathbb{S}_+^d -valued measure on \mathbb{S}_+^d such that $\int_{\|\xi\| \geq 1} \|\xi\|^2 \|\mu(d\xi)\| < \infty$. Consider the SPDE*

$$d\lambda_t = \mathcal{A}^* \lambda_t dt + v\beta(\lambda_t)dt + \beta(\lambda_t)vdt + \mathcal{S}_\varepsilon^* v dN_t + dN_t \mathcal{S}_\varepsilon^* v, \tag{4.12}$$

where $(N_t)_{t \geq 0}$ is a pure jump process with jump sizes in \mathbb{S}_+^d and compensator

$$\int_0^\cdot \int_{\mathbb{S}_+^d} \xi \operatorname{Tr}(\beta(\lambda_s) \mu(d\xi)) ds.$$

- (i) Then, for every $\lambda_0 \in \mathcal{E}$ and $\varepsilon > 0$, the SPDE (4.12) has a solution in \mathcal{E} given by a generalized Feller process associated with the generator of (4.12).
- (ii) This generalized Feller process is also a probabilistically weak and analytically mild solution of (4.12), i.e.,

$$\begin{aligned} \lambda_t = & \mathcal{S}_t^* \lambda_0 ds + \int_0^t \mathcal{S}_{t-s}^* v \beta(\lambda_s) ds + \int_0^t \beta(\lambda_s) \mathcal{S}_{t-s}^* v ds + \\ & + \int_0^t \mathcal{S}_{t-s+\varepsilon}^* v dN_s + \int_0^t dN_s \mathcal{S}_{t-s+\varepsilon}^* v, \end{aligned}$$

which justifies Eq. (4.12). In particular for every initial value the process N can be constructed on an appropriate probabilistic basis. The stochastic integral is defined in a pathwise way along finite variation paths. Moreover, for every family $(f_n)_n \in \mathcal{B}^q(\mathcal{E})$, $t \mapsto f_n(\lambda_t)$ can be chosen to be càglàd for all n .

- (iii) For every $\varepsilon > 0$, the corresponding Riccati equation $\partial_t y_t = R(y_t)$ with $R : \mathcal{D} \cap \mathcal{E}_* \rightarrow Y$ given by

$$\begin{aligned} R(y) = & \mathcal{A}y + \beta_* \left(\int_0^\infty y(x)v(dx) + v(dx)y(x) \right) \\ & + \beta_* \left(\int_{\mathbb{S}_+^d} (\exp(\langle y, \mathcal{S}_\varepsilon^* v \xi + \xi \mathcal{S}_\varepsilon^* v \rangle) - 1) \mu(d\xi) \right), \end{aligned} \tag{4.13}$$

admits a unique global solution in the mild sense for all initial values $y_0 \in \mathcal{E}_*$.

- (iv) The affine transform formula holds true, i.e.,

$$\mathbb{E}_{\lambda_0} [\exp(\langle y_0, \lambda_t \rangle)] = \exp(\langle y_t, \lambda_0 \rangle),$$

where y_t solves $\partial_t y_t = R(y_t)$ for all $y_0 \in \mathcal{E}_*$ in the mild sense with R given by (4.13). Moreover, $y_t \in \mathcal{E}_*$ for all $t \geq 0$.

Proof We assume that $v \neq 0$, otherwise there is nothing to prove. To prove the first assertion, we apply Proposition 3.4. By Propositions 4.7 and 4.10, the deterministic equation (4.4) has a mild solution on \mathcal{E} which—by Assumption 4.3—defines a generalized Feller semigroup $(P_t)_{t \geq 0}$ on $\mathcal{B}^q(\mathcal{E})$. The operator A in Proposition 3.4 then corresponds to the generator of $(P_t)_{t \geq 0}$, i.e., the semigroup associated with the purely

deterministic part of (4.12). This is a transport semigroup, and in view of Remark 3.3, we can have an equivalent norm with respect to a new weight function $\tilde{\varrho}$ on $\mathcal{B}^{\varrho}(\mathcal{E})$, such that $\|P_t\|_{L(\mathcal{B}^{\tilde{\varrho}}(\mathcal{E}))} \leq \exp(\omega t)$. Therefore, we find ourselves in the conditions of Proposition 3.4.

Note that by the same arguments as in Proposition 4.10 and by applying Theorem 2.18, we can prove that $(P_t)_{t \geq 0}$ also defines a generalized Feller semigroup on $\mathcal{B}^{\sqrt{\varrho}}(\mathcal{E})$. For the detailed proof which translates literally to the present setting, we refer to Cuchiero and Teichmann (2018).

Finally, we need to verify (3.3)–(3.5), which read as follows

$$\begin{aligned} \int \varrho(\lambda + \mathcal{S}_\varepsilon^* \nu \xi + \xi \mathcal{S}_\varepsilon^* \nu) \operatorname{Tr}(\beta(\lambda) \mu(d\xi)) &\leq M \varrho(\lambda)^2, \\ \int \sqrt{\varrho}(\lambda + \mathcal{S}_\varepsilon^* \nu \xi + \xi \mathcal{S}_\varepsilon^* \nu) \operatorname{Tr}(\beta(\lambda) \mu(d\xi)) &\leq M \varrho(\lambda), \\ \int \operatorname{Tr}(\beta(\lambda) \mu(d\xi)) &\leq M \sqrt{\varrho(\lambda)}, \end{aligned}$$

which hold true by the second-moment condition on μ . Concerning (3.6), denote as in Remark 3.3

$$\tilde{\varrho}(\lambda) = \sup_{t \geq 0} \exp(-\omega t) P_t \varrho(\lambda).$$

In particular, we know that $\varrho \leq \tilde{\varrho}$ and it holds that $P_t f(x) = f(\psi_t(x))$ where ψ is the solution of (4.4) which is linear. Using this together with $|\sup_t c(t) - \sup_t d(t)| \leq \sup_t |c(t) - d(t)|$, we obtain for some $\tilde{\omega}$

$$\begin{aligned} &\int \left| \frac{\tilde{\varrho}(\lambda + \mathcal{S}_\varepsilon^* \nu \xi + \xi \mathcal{S}_\varepsilon^* \nu) - \tilde{\varrho}(\lambda)}{\tilde{\varrho}(\lambda)} \right| \operatorname{Tr}(\beta(\lambda) \mu(d\xi)) \\ &\leq \int \left| \frac{\sup_{t \geq 0} \exp(-\omega t) |P_t \varrho(\lambda + \mathcal{S}_\varepsilon^* \nu \xi + \xi \mathcal{S}_\varepsilon^* \nu) - P_t \varrho(\lambda)|}{\tilde{\varrho}(\lambda)} \right| \operatorname{Tr}(\beta(\lambda) \mu(d\xi)) \\ &\leq \int \left| \frac{\sup_{t \geq 0} \exp(-\omega t) |\varrho(\psi_t(\lambda + \mathcal{S}_\varepsilon^* \nu \xi + \xi \mathcal{S}_\varepsilon^* \nu)) - \varrho(\psi_t(\lambda))|}{\tilde{\varrho}(\lambda)} \right| \operatorname{Tr}(\beta(\lambda) \mu(d\xi)) \\ &= \int \left| \frac{\sup_{t \geq 0} \exp(-\omega t) (2 \|\psi_t(\lambda)\|_{Y^*} \|\psi_t(\mathcal{S}_\varepsilon^* \nu \xi + \xi \mathcal{S}_\varepsilon^* \nu)\|_{Y^*} + \|\psi_t(\mathcal{S}_\varepsilon^* \nu \xi + \xi \mathcal{S}_\varepsilon^* \nu)\|_{Y^*}^2)}{\varrho(\lambda)} \right| \\ &\quad \times \operatorname{Tr}(\beta(\lambda) \mu(d\xi)) \leq \tilde{\omega}. \end{aligned}$$

The last inequality holds by the linearity of ψ and the second-moment condition on μ . Proposition 3.4 now allows to conclude that $A + B$, where B is given by

$$Bf(\lambda) = \int (f(\lambda + \mathcal{S}_\varepsilon^* \nu \xi + \xi \mathcal{S}_\varepsilon^* \nu) - f(\lambda)) \operatorname{Tr}(\beta(\lambda) \mu(d\xi)),$$

generates a generalized Feller semigroup \tilde{P} as asserted.

For (ii), we now construct the probabilistically weak and analytically mild solution directly from the properties of the generalized Feller process: take $y \in \mathcal{D}$ where \mathcal{D} is defined in (4.11) and consider the \mathbb{S}^d -valued martingale

$$M_t^y := \langle\langle y, \lambda_t \rangle\rangle - \langle\langle y, \lambda_0 \rangle\rangle - \int_0^t \langle\langle \mathcal{A}y, \lambda_s \rangle\rangle + \langle\langle y, \nu\beta(\lambda_s) + \beta(\lambda_s)\nu \rangle\rangle ds - \int_0^t \int \langle\langle y, \mathcal{S}_\varepsilon^* \nu \xi + \xi \mathcal{S}_\varepsilon^* \nu \rangle\rangle \text{Tr}(\beta(\lambda_s)\mu(d\xi)) ds \tag{4.14}$$

for $t \geq 0$ (after an appropriate and possible regularization according to Theorem 2.13).

Let now y be as above with the additional property that $\langle\langle y, \mathcal{S}_\varepsilon^* \nu \xi + \xi \mathcal{S}_\varepsilon^* \nu \rangle\rangle = \pi \xi + \xi \pi$ for all $\xi \in \mathbb{S}_+^d$ and some fixed $\pi \in \mathbb{S}_+^d$. For such y , define

$$N_t^\pi = \pi N_t + N_t \pi := M_t^y + \int_0^t \int \langle\langle y, \mathcal{S}_\varepsilon^* \nu \xi + \xi \mathcal{S}_\varepsilon^* \nu \rangle\rangle \text{Tr}(\beta(\lambda_s)\mu(d\xi)) ds \tag{4.15}$$

for $t \geq 0$, which is a càglàd semimartingale. Notice that the left-hand side only defines N^π and not the more suggestive $\pi N + N\pi$. Then, N^π does not depend on y by construction. Indeed, for all y_i with $\langle\langle y_i, \mathcal{S}_\varepsilon^* \nu \xi + \xi \mathcal{S}_\varepsilon^* \nu \rangle\rangle = \pi \xi + \xi \pi$ for all ξ , $i = 1, 2$, we clearly have

$$\int_0^t \int \langle\langle y_1 - y_2, \mathcal{S}_\varepsilon^* \nu \xi + \xi \mathcal{S}_\varepsilon^* \nu \rangle\rangle \text{Tr}(\beta(\lambda_s)\mu(d\xi)) ds = 0$$

and $M^{y_1} - M^{y_2} = M^{y_1 - y_2} = 0$ as well. The latter follows from the fact that the martingale M^y is constant if $\langle\langle y, \mathcal{S}_\varepsilon^* \nu \xi + \xi \mathcal{S}_\varepsilon^* \nu \rangle\rangle = 0$ for all ξ , since its quadratic variation vanishes in this case.

Moreover, by the definition of N^π in (4.15), its compensator is given by $\int_0^t \int (\pi \xi + \xi \pi) \text{Tr}(\beta(\lambda_s)\mu(d\xi)) ds$. Since it is sufficient to perform the previous construction for finitely many π to obtain all necessary projections, a process N can be defined such that $N^\pi = \pi N + N\pi$, as suggested by the notation.

By (4.14) and the very definition of (4.15), we obtain that

$$\begin{aligned} \langle\langle y, \lambda_t \rangle\rangle &= \langle\langle y, \lambda_0 \rangle\rangle + \int_0^t \langle\langle \mathcal{A}y, \lambda_s \rangle\rangle ds + \int_0^t \langle\langle y, \nu\beta(\lambda_s) + \beta(\lambda_s)\nu \rangle\rangle ds \\ &\quad + \langle\langle y, \mathcal{S}_\varepsilon^* \nu N_t \rangle\rangle + \langle\langle y, N_t \mathcal{S}_\varepsilon^* \nu \rangle\rangle \end{aligned}$$

for $y \in \mathcal{D}$. This analytically weak form can be translated into a mild form by standard methods. Indeed, notice that the integral is just along a finite variation path, and therefore, we can readily apply variation of constants. The last assertion about the càglàd property is a consequence of Theorem 2.13 by noting that $\varrho(\lambda)$ does not explode. This proves (ii).

Concerning (iii), note first that we have a unique mild solution to

$$\partial_t y_t = \mathcal{A}y_t + \beta_* \left(\int_0^\infty y(x)\nu(dx) + \int_0^\infty \nu(dx)y(x) \right), \quad y_0 \in Y, \quad (4.16)$$

since this is the adjoint equation of (4.4). For the equation with jumps, we proceed as in Proposition 4.7 via Picard iteration. Denote the semigroup associated with (4.16) by \mathcal{S}^{β_*} and define

$$y_t^0 = y_0, \\ y_t^n = \mathcal{S}_t^{\beta_*} y_0 + \int_0^t \mathcal{S}_{t-s}^{\beta_*} \beta_* \left(\int_{\mathbb{S}_+^d} \left(\exp(\langle y_s^{n-1}, \mathcal{S}_\varepsilon^* \nu \xi + \xi \mathcal{S}_\varepsilon^* \nu \rangle) - 1 \right) \mu(d\xi) \right) ds.$$

Moreover, for $t \in [0, \delta]$ for some $\delta > 0$, we have by local Lipschitz continuity of $x \mapsto \exp(x)$

$$\|y_t^{n+1} - y_t^n\|_Y \leq \int_0^t \mathcal{S}_{t-s}^{\beta_*} \beta_* \left(\int_{\mathbb{S}_+^d} \left(\exp(\langle y_s^n, \mathcal{S}_\varepsilon^* \nu \xi \rangle) - \exp(\langle y_s^{n-1}, \mathcal{S}_\varepsilon^* \nu \xi \rangle) \right) \mu(d\xi) \right) ds \|y_t^n\|_Y \\ \leq \int_0^t C \|\mathcal{S}_{t-s}^{\beta_*} \beta_*\|_{\text{op}} \|y_s^n - y_s^{n-1}\|_Y \left(\int_{\mathbb{S}_+^d} \|\mathcal{S}_\varepsilon^* \nu \xi\|_{Y^*} \mu(d\xi) \right) ds.$$

By an extension of Gronwall’s inequality, see Dalang (1999, Lemma 15), this yields convergence of $(y_t^n)_{n \in \mathbb{N}}$ with respect to $\|\cdot\|_Y$ and hence the existence of a unique local mild solution to (4.13) up to some maximal life time $t_+(y_0)$. That $t_+(y_0) = \infty$ for all $y_0 \in \mathcal{E}_*$ follows from the subsequent estimate

$$\|y_t\|_Y = \|\mathcal{S}_t^{\beta_*} y_0 + \int_0^t \mathcal{S}_{t-s}^{\beta_*} \beta_* \left(\int_{\mathbb{S}_+^d} \left(\exp(\langle y_s, \mathcal{S}_\varepsilon^* \nu \xi + \xi \mathcal{S}_\varepsilon^* \nu \rangle) - 1 \right) \mu(d\xi) \right) ds\|_Y \\ \leq \|\mathcal{S}_t^{\beta_*} y_0\|_Y + \int_0^t \|\mathcal{S}_{t-s}^{\beta_*} \beta_*\|_{\text{op}} \left(\int_{\mathbb{S}_+^d} |\exp(\langle y_s, \mathcal{S}_\varepsilon^* \nu \xi + \xi \mathcal{S}_\varepsilon^* \nu \rangle) - 1| \mu(dx) \right) ds \\ \leq \|\mathcal{S}_t^{\beta_*} y_0\|_Y + t \sup_{s \leq t} \|\mathcal{S}_s^{\beta_*} \beta_*\|_{\text{op}} \mu(\mathbb{S}_+^d),$$

where we used $|\exp(\langle y, \mathcal{S}_\varepsilon^* \nu \xi + \xi \mathcal{S}_\varepsilon^* \nu \rangle) - 1| \leq 1$ for all $y \in \mathcal{E}_*$ in the last estimate.

To prove (iv), just note that by the existence of a generalized Feller semigroup, the abstract Cauchy problem for the initial value $\exp(\langle y_0, \cdot \rangle)$ can be solved uniquely for $y_0 \in \mathcal{E}_*$. Indeed, $\mathbb{E}_\lambda[\exp(\langle y_0, \lambda_t \rangle)]$ uniquely solves

$$\partial_t u(t, \lambda) = Au(t, \lambda), \quad u(0, \lambda) = \exp(\langle y_0, \lambda \rangle),$$

where A denotes the generator associated with (4.12). Setting $u(t, \lambda) = \exp(\langle y_t, \lambda \rangle)$, we have

$$\partial_t u(t, \lambda) = \exp(\langle y_t, \lambda \rangle) R(y_t),$$

where the right-hand side is nothing else than $A \exp(\langle y_t, \lambda \rangle)$; hence, the affine transform formula holds true. This also implies that $y_t \in \mathcal{E}_*$ for all $t \geq 0$, simply because $\mathbb{E}_\lambda[\exp(\langle y_0, \lambda_t \rangle)] \leq 1$ for all $\lambda \in \mathcal{E}$. \square

We are now ready to state the main theorem of this section, namely an existence and uniqueness result for equations of the type

$$d\lambda_t = \mathcal{A}^* \lambda_t dt + \nu dX_t + dX_t \nu, \tag{4.17}$$

where $(X_t)_{t \geq 0}$ is a \mathbb{S}_+^d -valued pure jump Itô semimartingale of the form

$$X_t = \int_0^t \beta(\lambda_s) ds + \int_0^t \int_{\mathbb{S}_+^d} \xi \mu^X(d\xi, ds), \tag{4.18}$$

with β specified in (4.5) satisfying Assumption 4.9 and random measure of the jumps μ^X . Its compensator satisfies the following condition:

Assumption 4.12 The compensator of μ^X is given by

$$\text{Tr} \left(\beta(\lambda_t) \frac{\mu(d\xi)}{\|\xi\| \wedge 1} \right)$$

where μ is a \mathbb{S}_+^d -valued finite measure on \mathbb{S}_+^d satisfying $\int_{\|\xi\| \geq 1} \|\xi\|^2 \mu(d\xi) < \infty$.

For the formulation of the subsequent theorem, we shall need the following set of Fourier basis elements

$$\mathcal{D} = \{f_y : \mathcal{E} \rightarrow [0, 1]; \lambda \mapsto \exp(\langle y, \lambda \rangle) \mid y \in \mathcal{E}_* \cap \text{dom}(\mathcal{A}) \text{ s.t. } \langle \langle y, \nu \rangle \rangle \text{ is well defined}\}. \tag{4.19}$$

Theorem 4.13 *Let Assumptions 4.3, 4.9 and 4.12 be in force.*

(i) *Then, the stochastic partial differential equation (4.17) admits a unique Markovian solution $(\lambda_t)_{t \geq 0}$ in \mathcal{E} given by a generalized Feller semigroup on $\mathcal{B}^e(\mathcal{E})$ whose generator takes on the set of Fourier elements*

$$f_y : \mathcal{E} \rightarrow [0, 1]; \lambda \mapsto \exp(\langle y, \lambda \rangle)$$

for $y \in \mathcal{D} \cap \mathcal{E}_*$ where \mathcal{D} is defined in (4.11) the form

$$\mathcal{A}f_y(\lambda) = f_y(\lambda)(\langle \mathcal{A}y, \lambda \rangle + \langle \mathcal{R}(\langle y, \nu \rangle), \lambda \rangle), \tag{4.20}$$

with $\mathcal{R} : \mathbb{S}_+^d \rightarrow Y$ given by

$$\mathcal{R}(u) = \beta_*(u) + \beta_* \left(\int_{\mathbb{S}_+^d} (\exp(\text{Tr}(u\xi)) - 1) \frac{\mu(d\xi)}{\|\xi\| \wedge 1} \right). \tag{4.21}$$

(ii) This generalized Feller process is also a probabilistically weak and analytically mild solution of (4.17), i.e.,

$$\lambda_t = \mathcal{S}_t^* \lambda_0 ds + \int_0^t \mathcal{S}_{t-s}^* v dX_s + \int_0^t dX_s \mathcal{S}_{t-s}^* v,$$

This justifies Eq. (4.17); in particular, for every initial value, the process X can be constructed on an appropriate probabilistic basis. The stochastic integral is defined in a pathwise way along finite variation paths. Moreover, for every family $(f_n)_n \in \mathcal{B}^0(\mathcal{E})$, $t \mapsto f_n(\lambda_t)$ can be chosen to be càg for all n .

(iii) The affine transform formula is satisfied, i.e.,

$$\mathbb{E}_{\lambda_0} [\exp(\langle y_0, \lambda_t \rangle)] = \exp(\langle y_t, \lambda_0 \rangle),$$

where y_t solves $\partial_t y_t = R(y_t)$ for all $y_0 \in \mathcal{E}_*$ and $t > 0$ in the mild sense with $R : \mathcal{D} \cap \mathcal{E}_* \rightarrow Y$ given by

$$R(y) = \mathcal{A}y + \mathcal{R}(\langle y, v \rangle) \tag{4.22}$$

with \mathcal{R} defined in (4.21). Furthermore, $y_t \in \mathcal{E}_*$ for all $t \geq 0$.

(iv) For all $\lambda_0 \in \mathcal{E}$, the corresponding stochastic Volterra equation, $V_t := \beta(\lambda_t)$, given by

$$\begin{aligned} V_t = \beta(\lambda_t) &= \beta(\mathcal{S}_t^* \lambda_0) + \int_0^t \beta(\mathcal{S}_{t-s}^* v) dX_s + \int_0^t dX_s \beta(\mathcal{S}_{t-s}^* v) \\ &= h(t) + \int_0^t K(t-s) dX_s + \int_0^t dX_s K(t-s) \end{aligned} \tag{4.23}$$

admits a probabilistically weak solution with càg trajectories. Here, $h(t) := \beta(\mathcal{S}_t^* \lambda_0)$.

(v) The Laplace transform of the Volterra equation V_t is given by

$$\mathbb{E}_{\lambda_0} [\exp(\text{Tr}(u V_t))] = \exp\left(\text{Tr}(u h(t)) + \int_0^t \text{Tr}(\mathfrak{R}(\psi_s) h(t-s)) ds\right), \tag{4.24}$$

where $h(t) = \beta(\mathcal{S}_t^* \lambda_0)$, $\mathfrak{R} : \mathbb{S}_-^d \rightarrow \mathbb{S}_-^d$, $u \mapsto \mathfrak{R}(u) = u + \int_{\mathbb{S}_+^d} (e^{\text{Tr}(u\xi)} - 1) \frac{\mu(d\xi)}{\|\xi\| \wedge 1}$ and ψ_t solves the matrix Riccati–Volterra equation

$$\psi_t = uK(t) + \int \mathfrak{R}(\psi_s) K(t-s) ds, \quad t > 0.$$

Hence, the solution of the stochastic Volterra equation in (4.23) is unique in law.

Remark 4.14 One essential point here is that we loose the càglàd property as stated in Proposition 4.11 (ii) when we let ε of \mathcal{S}_ε tend to zero. As long as the kernel K has a singularity at $t = 0$, it is impossible to preserve finite growth bounds with $M = 1$,

as $\varepsilon \rightarrow 0$, but we get càg versions (compare with the second conclusion in Theorem 2.13 and Remark 2.14).

Remark 4.15 Note that for β as of Example 4.4, the above equations simplify considerably. In particular, β_* in (4.21) is simply the identity.

Proof We apply Theorem 3.2 and consider a sequence of generalized Feller semigroups $(P^n)_{n \in \mathbb{N}}$ with generators A^n corresponding to the solution λ^n of (4.12) for $\varepsilon = \frac{1}{n}$, and compensator

$$\text{Tr} \left(\beta(\lambda_t^n) \frac{1_{\{\|\xi\| > \frac{1}{n}\}} \mu(d\xi)}{\|\xi\| \wedge 1} \right), \quad n \in \mathbb{N}.$$

Let us first establish a uniform growth bound for this sequence. To this end, denote

$$F^n(d\xi) := \frac{1_{\{\|\xi\| > \frac{1}{n}\}} \mu(d\xi)}{\|\xi\| \wedge 1}.$$

Note that for the solution of (4.12), we have due to Proposition 4.11 (ii) the following estimate for $t \in [0, T]$ for some fixed $T > 0$

$$\begin{aligned} \mathbb{E}[\|\lambda_t^n\|_{Y^*}^2] &\leq 5\|\mathcal{S}_t^* \lambda_0\|_{Y^*}^2 + 10t \int_0^t \|\mathcal{S}_{t-s}^* \nu\|_{Y^*}^2 \|\beta\|_{\text{op}}^2 \mathbb{E}[\|\lambda_s^n\|_{Y^*}^2] ds \\ &\quad + 10\mathbb{E} \left[\left\| \int_0^t \mathcal{S}_{t-s+\frac{1}{n}}^* \nu dN_s - \int_0^t \int \mathcal{S}_{t-s+\frac{1}{n}}^* \nu \xi \text{Tr}(\beta(\lambda_s^n) F^n(d\xi)) ds \right\|_{Y^*}^2 \right] \\ &\quad + 10\mathbb{E} \left[\left\| \int_0^t dN_s \mathcal{S}_{t-s+\frac{1}{n}}^* \nu - \int_0^t \int \xi \mathcal{S}_{t-s+\frac{1}{n}}^* \nu \text{Tr}(\beta(\lambda_s^n) F^n(d\xi)) ds \right\|_{Y^*}^2 \right] \\ &\quad + 10\mathbb{E} \left[\left\| \int_0^t \int \mathcal{S}_{t-s+\frac{1}{n}}^* \nu \xi \text{Tr}(\beta(\lambda_s^n) F^n(d\xi)) ds \right\|_{Y^*}^2 \right] \\ &\quad + 10\mathbb{E} \left[\left\| \int_0^t \int \xi \mathcal{S}_{t-s+\frac{1}{n}}^* \nu \text{Tr}(\beta(\lambda_s^n) F^n(d\xi)) ds \right\|_{Y^*}^2 \right]. \end{aligned}$$

As a consequence of Itô’s isometry, the martingale part can be estimated by

$$\begin{aligned} &\mathbb{E} \left[\left\| \int_0^t \mathcal{S}_{t-s+\frac{1}{n}}^* \nu dN_s - \int_0^t \int \mathcal{S}_{t-s+\frac{1}{n}}^* \nu \xi \text{Tr}(\beta(\lambda_s^n) F^n(d\xi)) ds \right\|_{Y^*}^2 \right] \\ &\leq \mathbb{E} \left[\left\| \int_0^t \int \|\mathcal{S}_{t-s+\frac{1}{n}}^* \nu\|_{Y^*}^2 \|\xi\|^2 \text{Tr}(\beta(\lambda_s^n) F^n(d\xi)) ds \right\| \right] \\ &\leq \int \|\xi\|^2 \|F^n(d\xi)\| \int_0^t \|\mathcal{S}_{t-s+\frac{1}{n}}^* \nu\|_{Y^*}^2 \|\beta\|_{\text{op}} \mathbb{E}[\|\lambda_s^n\|_{Y^*}] ds \\ &\leq \left(\int_{\|\xi\| \leq 1} \|\mu(d\xi)\| + \int_{\|\xi\| > 1} \|\xi\|^2 \|\mu(d\xi)\| \right) \int_0^t \|\mathcal{S}_{t-s+\frac{1}{n}}^* \nu\|_{Y^*}^2 \|\beta\|_{\text{op}} \mathbb{E}[\|\lambda_s^n\|_{Y^*}] ds \end{aligned}$$

$$\begin{aligned} &\leq \tilde{C} \int_0^t \|\mathcal{S}_{t-s+\frac{1}{n}}^* \nu\|_{Y^*}^2 \|\beta\|_{\text{op}} \mathbb{E}[\|\lambda_s^n\|_{Y^*}] ds \\ &\leq \tilde{C} K \int_0^t \|\mathcal{S}_{t-s+\frac{1}{n}}^* \nu\|_{Y^*}^2 (1 + \|\beta\|_{\text{op}}^2 \mathbb{E}[\|\lambda_s^n\|_{Y^*}^2]) ds \end{aligned}$$

where $\tilde{C} = \left(\int_{\|\xi\| \leq 1} \|\mu(d\xi)\| + \int_{\|\xi\| > 1} \|\xi\|^2 \|\mu(d\xi)\| \right)$ and K some other constant. Moreover, for the last terms, we have

$$\begin{aligned} &\mathbb{E} \left[\left\| \int_0^t \int \mathcal{S}_{t-s+\frac{1}{n}}^* \nu \xi \text{Tr}(\beta(\lambda_s^n) F^n(d\xi)) ds \right\|_{Y^*}^2 \right] \\ &\leq t \int_0^t \|\mathcal{S}_{t-s+\frac{1}{n}}^* \nu\|_{Y^*}^2 \mathbb{E} \left[\left\| \int \xi \text{Tr}(\beta(\lambda_s^n) F^n(d\xi)) \right\|^2 \right] ds \\ &\leq 2t \int_0^t \|\mathcal{S}_{t-s+\frac{1}{n}}^* \nu\|_{Y^*}^2 \mathbb{E} \left[\left\| \int_{\|\xi\| \leq 1} \xi \text{Tr}(\beta(\lambda_s^n) F^n(d\xi)) \right\|^2 \right. \\ &\quad \left. + \left\| \int_{\|\xi\| \geq 1} \xi \text{Tr}(\beta(\lambda_s^n) F^n(d\xi)) \right\|^2 \right] ds \\ &\leq 2t \int_0^t \|\mathcal{S}_{t-s+\frac{1}{n}}^* \nu\|_{Y^*}^2 \|\beta\|_{\text{op}}^2 \mathbb{E}[\|\lambda_s^n\|_{Y^*}^2] \int \|\mu(d\xi)\| \\ &\quad \times \left(\int_{\|\xi\| \leq 1} \|\mu(d\xi)\| + \int_{\|\xi\| > 1} \|\xi\|^2 \|\mu(d\xi)\| \right) \\ &\leq 2t \hat{C} \int_0^t \|\mathcal{S}_{t-s+\frac{1}{n}}^* \nu\|_{Y^*}^2 \|\beta\|_{\text{op}}^2 \mathbb{E}[\|\lambda_s^n\|_{Y^*}^2] \end{aligned}$$

where $\hat{C} = \int \|\mu(d\xi)\| \tilde{C}$. Putting this together, we obtain

$$\begin{aligned} \mathbb{E}[\|\lambda_t^n\|_{Y^*}^2] &\leq C_0 \|\lambda_0\|_{Y^*}^2 + 10t \int_0^t \|\mathcal{S}_{t-s}^* \nu\|_{Y^*}^2 \|\beta\|_{\text{op}}^2 \mathbb{E}[\|\lambda_s^n\|_{Y^*}^2] ds \\ &\quad + 20\tilde{C} K \int_0^t \|\mathcal{S}_{t-s+\frac{1}{n}}^* \nu\|_{Y^*}^2 ds \\ &\quad + 20(\tilde{C} K + 2t\hat{C}) \int_0^t \|\mathcal{S}_{t-s+\frac{1}{n}}^* \nu\|_{Y^*}^2 \|\beta\|_{\text{op}}^2 \mathbb{E}[\|\lambda_s^n\|_{Y^*}^2] \\ &\leq C_0 \|\lambda_0\|_{Y^*}^2 + C_1 \int_0^t \|\mathcal{S}_{t-s}^* \nu\|_{Y^*}^2 ds + C_2 \int_0^t \|\mathcal{S}_{t-s}^* \nu\|_{Y^*}^2 \mathbb{E}[\|\lambda_s^n\|_{Y^*}^2] ds \end{aligned}$$

where C_0 and C_2 depend on T . We use $\|\mathcal{S}_t^* \lambda_0\|^2 \leq C_0 \|\lambda_0\|^2$ for $t \in [0, T]$, as well as $\|\mathcal{S}_{t-s+\frac{1}{n}}^* \nu\|_{Y^*} \leq C \|\mathcal{S}_{t-s}^* \nu\|_{Y^*}$ for some constant C and all $n \in \mathbb{N}$ due to strong continuity. Exactly by the same arguments as in the proof of Proposition 4.7, we thus obtain for $t \in [0, T]$ for some fixed T

$$\mathbb{E}[\|\lambda_t\|_{Y^*}^2] \leq \tilde{C}(\|\lambda_0\|_{Y^*}^2 + 1) \left(1 - \int_0^t R'(s), ds\right),$$

where R' denotes the resolvent of $-C_2\|\mathcal{S}_{t-s}^*v\|_{Y^*}$. Hence, $\mathbb{E}[\varrho(\lambda_t)] \leq C\varrho(\lambda_0)$ for $t \in [0, T]$. From this, the desired uniform growth bound $\|P_t\|_{L(\mathcal{B}^e(\mathcal{E}))} \leq M \exp(\omega t)$ for some $M \geq 1$ and $\omega \in \mathbb{R}$ follows.

For the set D as of Theorem 3.2, we here choose Fourier basis elements of the form

$$f_y : \mathcal{E} \rightarrow [0, 1]; \lambda \mapsto \exp(\langle y, \lambda \rangle) \tag{4.25}$$

such that $y \in \mathcal{E}_*$ and $\lambda \mapsto \exp(\langle y, \lambda \rangle)$ lies in $\cap_{n \geq 1} \text{dom}(A^n)$, whose span is dense, whence (i) of Theorem 3.2. Here, A^n denotes the generator corresponding to (4.12) with $\varepsilon = \frac{1}{n}$ and μ replaced by F^n . We now equip $\text{span}(D)$ with the uniform norm $\|\cdot\|_\infty$ and verify Condition (ii), i.e., we check

$$\|A^n P_u^m f_y - A^m P_u^m f_y\|_\varrho \leq \|f_y\|_\infty a_{nm} \tag{4.26}$$

for all $0 \leq u \leq t$ with $a_{nm} \rightarrow 0$ as $n, m \rightarrow \infty$, and possibly depending on y . Note that

$$A^n f_y(\lambda) = \langle R^n(y), \lambda \rangle f_y(\lambda),$$

where R^n corresponds to (4.13) for $\varepsilon = \frac{1}{n}$ and μ replaced by F^n . As P^n leaves D invariant for all $n \in \mathbb{N}$ by Proposition 4.11 (iv), we have

$$\begin{aligned} & \frac{|A^n P_u^m f_y(\lambda) - A^m P_u^m f_y(\lambda)|}{\varrho(\lambda)} \\ & \leq \frac{f_{y_u^m}(\lambda)}{\varrho(\lambda)} \left(\beta_* \left(\underbrace{\int_{\mathbb{S}_+^d} \exp(\langle y_u^m, \mathcal{S}_{\frac{1}{m}}^* v \xi + \xi \mathcal{S}_{\frac{1}{m}}^* v \rangle) 1_{\{\|\xi\| \geq \frac{1}{n}\}}}_{:=b_{nm}(\xi)} \right) \right. \\ & \quad \times \left. \underbrace{|\exp(\langle y_u^m, (\mathcal{S}_{\frac{1}{n}}^* v - \mathcal{S}_{\frac{1}{m}}^* v) \xi + \xi (\mathcal{S}_{\frac{1}{n}}^* v - \mathcal{S}_{\frac{1}{m}}^* v) \rangle) - 1|}_{\tilde{a}_{nm}^1(\xi)} \frac{\mu(d\xi)}{\|\xi\| \wedge 1} \right) \\ & \quad + \beta_* \left(\underbrace{\int_{\mathbb{S}_+^d} \exp(\langle y_u^m, \mathcal{S}_{\frac{1}{m}}^* v \xi + \xi \mathcal{S}_{\frac{1}{m}}^* v \rangle) - 1}_{\tilde{a}_{nm}^2} 1_{\{\|\xi\| \geq \frac{1}{n}\}} - 1_{\{\|\xi\| \geq \frac{1}{m}\}} \frac{\mu(d\xi)}{\|\xi\| \wedge 1} \right). \end{aligned}$$

Here, y_u^m denotes the solution of $\partial_t y_u^m = R^m(y_u^m)$ at time u with $y_0 = y$. Moreover, $\tilde{a}_{nm}^1(\xi)$ and \tilde{a}_{nm}^2 can be chosen uniformly for all $u \leq t$ and tend to 0 as $n, m \rightarrow \infty$. This is possible since for the chosen initial values y we obtain that y_u^m is bounded on compact intervals in time uniformly in m (see Cuchiero and Teichmann 2018 for details). This together with dominated convergence for the first term (note that $b_{nm}(\xi) \tilde{a}_{nm}^1(\xi)$ can be bounded by $\|\xi\| \wedge 1$) we thus infer (4.26). The conditions of Theorem 3.2 are therefore

satisfied, and we obtain a generalized Feller semigroup whose generator is given by (4.20).

For the second assertion, we proceed as in the proof of Proposition 4.11, the proof of the existence of X can be transferred verbatim. However, one loses the existence of càglàd paths of $f_n(\lambda)$ due to the possible lack of finite mass of ν . Here, we only obtain càg trajectories (compare with Remarks 2.14 and 4.14).

Concerning the third assertion, the affine transform formula follows simply from the convergence of the semigroups P^n as asserted in Theorem 3.2 by setting $y_t = \lim_{n \rightarrow \infty} y_t^n$, where y_t^n solves $\partial_t y_t^n = R^n(y_t^n)$ in the mild sense with R^n given again by (4.13) with $\varepsilon = \frac{1}{n}$ and μ replaced by F^n . Since $\exp(\langle y_t, \lambda \rangle)$ is then also the unique solution of the abstract Cauchy problem for initial value $\exp(\langle y_0, \lambda \rangle)$, i.e., it solves

$$\partial_t u(t, \lambda) = Au(t, \lambda), \quad u(0, \lambda) = \exp(\langle y_0, \lambda \rangle),$$

where A denotes the generator (4.20), we infer that y_t satisfies $\partial_t y_t = R(y_t)$ with R given by (4.22). This is because $A \exp(\langle y_t, \lambda \rangle) = \exp(\langle y_t, \lambda \rangle)R(y_t)$.

The fourth claim follows from statement (ii), property (4.5) and the definition of K in (4.6).

Finally to prove (v), note that due to (iv) and the definition of the adjoint operator β_* , we have

$$\text{Tr}(uV_t) = \text{Tr}(u\beta(\lambda_t)) = \langle \beta_*(u), \lambda_t \rangle.$$

Statement (iii) therefore implies that

$$\mathbb{E}[e^{\text{Tr}(uV_t)}] = e^{\langle y_t, \lambda_0 \rangle},$$

where the mild solution of y_t can be expressed by

$$y_t = \mathcal{S}_t \beta_*(u) + \int_0^t \mathcal{S}_{t-s} \mathcal{R}(\langle y_s, \nu \rangle) ds. \tag{4.27}$$

Hence, by definition of \mathcal{R} , \mathfrak{A} and h , we find

$$\begin{aligned} \langle y_t, \lambda_0 \rangle &= \langle \mathcal{S}_t \beta_*(u) + \int_0^t \mathcal{S}_{t-s} \mathcal{R}(\langle y_s, \nu \rangle) ds, \lambda_0 \rangle \\ &= \text{Tr}(u\beta(\mathcal{S}_t^* \lambda_0)) + \int_0^t \text{Tr}(\mathfrak{A}(\langle y_s, \nu \rangle) \beta(\mathcal{S}_{t-s}^* \lambda_0)) ds \\ &= \text{Tr}(uh(t)) + \int_0^t \text{Tr}(\mathfrak{A}(\langle y_s, \nu \rangle) h(t-s)) ds \end{aligned} \tag{4.28}$$

From this and (4.27), it is easily seen that we can replace $\langle y_s, \nu \rangle$ in (4.28) by a solution of the following Volterra–Riccati equation

$$\psi_t = uK(t) + \int_0^t \mathfrak{A}(\psi_s)K(t-s).$$

Note that we do not need to symmetrize here since we apply the trace and h is symmetric. This proves the assertion. \square

The following example illustrates how a multivariate Hawkes process can easily be defined by means of (4.18).

Example 4.16 Let β and \mathcal{S}^* be as of Example 4.4. Define $\mu_{ii}(d\xi) = \delta_{e_{ii}}(d\xi)$ and $\mu_{ij} = 0$ for $i \neq j$. Then, the Volterra equation as of (4.23) is given by

$$V_t = \int_0^\infty e^{-xt} \lambda_0(dx) + \int_0^t (K(t-s)V_s + V_s K(t-s))ds + \int_0^t K(t-s)dN_s + \int_0^t dN_s K(t-s).$$

Only the diagonal components of the matrix-valued process N jump, and we can define $\widehat{N} := \text{diag}(N)$ which is a process with values in \mathbb{N}_0^d . Its components jump by one, and the compensator of $N_{ii} = \widehat{N}_i$ is given by $\int_0^\cdot V_{s,ii}ds$, which justifies the name multivariate Hawkes process. Note that the components of V are not independent if v and in turn K are not diagonal.

5 Squares of matrix-valued Volterra OU processes

As in the finite dimensional setting, squares of Gaussian processes provide us with important process classes for financial and statistical modeling. In this section, we outline this program in utmost generality from a stochastic and analytic point of view. In particular, we consider *continuous affine Volterra-type processes* on \mathbb{S}_+^d , which we construct as squares of matrix-valued Volterra Ornstein–Uhlenbeck (OU) processes (see Remark 5.4). Following the finite dimensional analogon (Bru 1991), we start by considering matrix measure-valued OU processes of the form

$$d\gamma_t(dx) = \mathcal{A}^* \gamma_t(dx)dt + dW_t \nu(dx), \quad \gamma_0 \in Y^*(\mathbb{R}^{n \times d}). \tag{5.1}$$

The underlying Banach space, denoted by $Y^*(\mathbb{R}^{n \times d})$, is the space of finite $\mathbb{R}^{n \times d}$ -valued regular Borel measures on the extended half real line $\overline{\mathbb{R}}_+ := \mathbb{R}_+ \cup \{\infty\}$. Together with

$$\varrho(\gamma) = 1 + \|\gamma\|_{Y^*(\mathbb{R}^{n \times d})}^2, \quad \gamma \in Y^*(\mathbb{R}^{n \times d}),$$

where $\|\cdot\|_{Y^*(\mathbb{R}^{n \times d})}$ denotes the total variation norm, this becomes a weighted space. Moreover, \mathcal{A}^* is the generator of a strongly continuous semigroup \mathcal{S}^* on $Y^*(\mathbb{R}^{n \times d})$, which satisfies a property analogous to (4.3), i.e., for elements $A \in \mathbb{R}^{n \times d}$, and it holds that

$$\mathcal{S}_t^*(\gamma(\cdot)A^\top) = (\mathcal{S}_t^* \gamma(\cdot))A^\top \quad \text{and} \quad \mathcal{S}_t^*(A\gamma^\top(\cdot)) = A(\mathcal{S}_t^* \gamma(\cdot))^\top. \tag{5.2}$$

The process W is a $n \times d$ matrix of Brownian motions and $\nu \in Y^* =: Y^*(\mathbb{S}^d)$ or Z^* , as defined in Sect. 4 such that Assumption 4.3 holds true. The pre-dual space denoted

by $Y(\mathbb{R}^{n \times d})$ is given by $C_b(\overline{\mathbb{R}_+}, \mathbb{R}^{n \times d})$ functions, where we fix the pairing $\langle \cdot, \cdot \rangle$ as follows

$$\langle \cdot, \cdot \rangle : Y(\mathbb{R}^{n \times d}) \times Y^*(\mathbb{R}^{n \times d}) \rightarrow \mathbb{R}, \quad (y, \gamma) \mapsto \langle y, \gamma \rangle = \text{Tr} \left(\int_0^\infty y^\top(x) \gamma(dx) \right).$$

Again Tr denotes the trace. We assume that all relevant properties from Assumption 4.1 are translated to the current setting.

Remark 5.1 Observe the analogy to the process γ defined in the introduction. If $\mathcal{A}^* = 0$ and ν is supported on a finite space with k points, then (5.1) is exactly the process from the introduction.

Proposition 5.2 *For every $\gamma_0 \in Y^*(\mathbb{R}^{n \times d})$, the SPDE (5.1) has a solution given by a generalized Feller semigroup on $\mathcal{B}^q(Y^*(\mathbb{R}^{n \times d}))$ associated with the generator of (5.1). The mild formulation directly yields a stochastically strong solution*

$$\gamma_t(dx) = S_t^* \gamma_0(dx) + \int_0^t dW_s S_{t-s}^* \nu(dx)$$

where order matters, i.e., the matrix Brownian increment is applied to $S_{t-s}^* \nu(dx)$ on the left. The integral is understood in the weak sense, i.e., after pairing with $y \in Y(\mathbb{R}^{n \times d})$.

Proof The construction of the generalized Feller process can be done by jump approximation of the Brownian motion similarly as in Cuchiero and Teichmann (2018, Theorem 4.16). Notice here that we consider the process on the whole space $Y^*(\mathbb{R}^{n \times d})$. So no issues with state space constraints occur.

The right-hand side of the stochastically strong formulation defines —after pairing with $y \in Y(\mathbb{R}^{n \times d})$ — almost surely a continuous linear functional with value

$$\langle y, S_t^* \gamma_0 \rangle + \int_0^t \langle y, dW_s S_{t-s}^* \nu \rangle,$$

since the integrand of the stochastic integral is deterministic and in L^2 for each $t \geq 0$. □

In order to define the actual process of interest, we need to introduce some further notations: For elements in $\gamma \in Y^*(\mathbb{R}^{n \times d})$, we define

$$(\gamma \widehat{\otimes} \gamma)(\cdot, \cdot) := \gamma^\top(\cdot) \gamma(\cdot).$$

The corresponding *contracted*, i.e., one matrix multiplication is performed, algebraic tensor product is denoted by $Y^*(\mathbb{R}^{n \times d}) \widehat{\otimes} Y^*(\mathbb{R}^{n \times d})$, and we set

$$\widehat{\mathcal{E}} := \{ \gamma \widehat{\otimes} \gamma \in Y^*(\mathbb{R}^{n \times d}) \widehat{\otimes} Y^*(\mathbb{R}^{n \times d}) \}. \tag{5.3}$$

This corresponds to the space of finite \mathbb{S}_+^d -valued, rank n , product measures on $\overline{\mathbb{R}}_+ \times \overline{\mathbb{R}}_+$. We shall introduce a particular dual topology on $\widehat{\mathcal{E}}$, namely $\sigma(\widehat{\mathcal{E}}, Y \otimes Y)$, where the corresponding pairing is given by

$$\begin{aligned} (y_1 \otimes y_2, \gamma_1 \widehat{\otimes} \gamma_2) &\mapsto \langle y_1 \widehat{\otimes} y_2, \gamma_1 \widehat{\otimes} \gamma_2 \rangle \\ &= \text{Tr} \left(\int_0^\infty y_1^\top(x_1) y_2(x_2) \gamma_1^\top(dx_1) \gamma_2(dx_2) \right). \end{aligned}$$

We denote the pre-dual cone by

$$-\widehat{\mathcal{E}}_* = \{y \widehat{\otimes} y \in Y(\mathbb{R}^{n \times d}) \widehat{\otimes} Y(\mathbb{R}^{n \times d})\}, \tag{5.4}$$

where we use again the contracted algebraic tensor product corresponding to the following matrix multiplication of $\mathbb{R}^{n \times d}$ valued functions

$$(y \widehat{\otimes} y)(\cdot, \cdot) = y^\top(\cdot) y(\cdot), \quad y \in Y(\mathbb{R}^{n \times d}).$$

The minus on the left-hand side of (5.4) is to obtain elements in the polar cone.

Let us now define the actual process of interest, namely

$$\lambda_t(dx_1, dx_2) := \gamma_t^\top(dx_1) \gamma_t(dx_2) = \gamma_t(dx_1) \widehat{\otimes} \gamma_t(dx_2). \tag{5.5}$$

Note again the analogy to the Wishart process λ defined in the introduction. The process (5.5) clearly takes values in $\widehat{\mathcal{E}}$ as defined in (5.3). We will now show that we can define a Volterra-type process by considering projections on \mathbb{S}_+^d . Applying Itô's formula, we see that $\lambda_t(dx_1, dx_2)$ satisfies the following equation

$$\begin{aligned} d\lambda_t(dx_1, dx_2) &= (\mathcal{A}_1^* \lambda_t(dx_1, dx_2) + \mathcal{A}_2^* \lambda_t(dx_1, dx_2) + n\nu(dx_1)\nu(dx_2)) dt \\ &\quad + \nu(dx_1) dW_t^\top \gamma_t(dx_2) + \gamma_t(dx_1)^\top dW_t \nu(dx_2), \end{aligned} \tag{5.6}$$

where $\mathcal{A}_1^* \lambda_t(dx_1, dx_2) = \mathcal{A}^* \lambda_t(\cdot, dx_2)(dx_1)$ and analogously for \mathcal{A}_2^* . Note that for $\mathcal{A}^* = 0$, this is completely analogous to (1.3).

By a lot of abuse of notation, but parallel with Bru (1991) and Eqs. (1.4)–(1.5), we can also write

$$\begin{aligned} d\lambda_t(dx_1, dx_2) &= (\mathcal{A}_1^* \lambda_t(dx_1, dx_2) + \mathcal{A}_2^* \lambda_t(dx_1, dx_2) + n\nu(dx_1)\nu(dx_2)) dt \\ &\quad + \int_0^\infty \int_0^\infty \sqrt{\nu \widehat{\otimes} \nu}(dx_1, dx) dB_t^\top(dy, dx) \sqrt{\lambda_t}(dy, dx_2) \\ &\quad + \int_0^\infty \int_0^\infty \sqrt{\lambda_t}(dx_1, dx) dB_t(dx, dy) \sqrt{\nu \widehat{\otimes} \nu}(dy, dx_2), \end{aligned} \tag{5.7}$$

where heuristically $B(dx, dy)$ is $d \times d$ matrix of Brownian fields. We shall not develop a framework where this notation makes sense, but continue with proving that λ is actually a generalized Feller process, which should be considered the correct infinite dimensional version of a Wishart process.

By only a slight abuse of notation, we understand \mathcal{A}^* , and in the sequel also \mathcal{S}^* and other linear operators, as operators acting on both \mathbb{S}^d -valued measures as well as $\mathbb{R}^{d \times n}$ -valued or $\mathbb{R}^{n \times d}$ -valued ones as in (5.1). The mild formulation of (5.6), denoting the semigroup generated by $\mathcal{A}_1^* + \mathcal{A}_2^*$ by $\mathcal{S}_t^{*, \widehat{\otimes}}$, then reads as

$$\begin{aligned} \lambda_t(dx_1, dx_2) &= \mathcal{S}_t^{*, \widehat{\otimes}} \lambda_0(dx_1, dx_2) + n \int_0^t \mathcal{S}_{t-s}^{*, \widehat{\otimes}} \nu(dx_1) \nu(dx_2) ds \\ &\quad + \int_0^t \mathcal{S}_{t-s}^{*, \widehat{\otimes}} (\nu(dx_1) dW_s^\top \gamma_s(dx_2) + \gamma_s(dx_1)^\top dW_t \nu(dx_2)) \\ &= \mathcal{S}_t^{*, \widehat{\otimes}} \lambda_0(dx_1, dx_2) + n \int_0^t (\mathcal{S}_{t-s}^* \nu(dx_1)) (\mathcal{S}_{t-s}^* \nu(dx_2)) ds \\ &\quad + \int_0^t (\mathcal{S}_{t-s}^* \nu(dx_1)) dW_s^\top (\mathcal{S}_{t-s}^* \gamma_s(dx_2)) \\ &\quad + \int_0^t (\mathcal{S}_{t-s}^* \gamma_s(dx_1))^\top dW_s (\mathcal{S}_{t-s}^* \nu(dx_2)), \end{aligned}$$

where the second equality follows from property (5.2).

Let now β be a linear operator from $Y^*(F)$ to F where F stands here for $\mathbb{R}^{n \times d}$, or \mathbb{S}^d with the property that for a constant matrix A with appropriate matrix dimensions, we have

$$\beta(A\gamma(\cdot)) = A\beta(\gamma(\cdot)), \quad \beta(\gamma(\cdot)A) = \beta(\gamma(\cdot))A. \tag{5.8}$$

By means of β , define now an operator $\widehat{\beta}$ acting on $\mathbb{R}^{d \times d}$ valued product measures as follows

$$\widehat{\beta}(\gamma_1^\top(\cdot)\gamma_2(\cdot)) = \beta(\gamma_1(\cdot))^\top \beta(\gamma_2(\cdot)), \tag{5.9}$$

where γ_1 and γ_2 are either in $Y^*(\mathbb{R}^{n \times d})$ or in $Y^*(\mathbb{S}^d)$. In the latter case, the transpose is not needed. Note that (5.9) implies that $\widehat{\beta}(\gamma^\top(\cdot)\gamma(\cdot))$ is \mathbb{S}_+^d -valued. Applying $\widehat{\beta}$ to λ , we find

$$\begin{aligned} \widehat{\beta}(\lambda_t) &= \widehat{\beta}(\mathcal{S}_t^{*, \widehat{\otimes}} \lambda_0) + n \int_0^t \beta(\mathcal{S}_{t-s}^* \nu) \beta(\mathcal{S}_{t-s}^* \nu) ds \\ &\quad + \int_0^t \beta(\mathcal{S}_{t-s}^* \nu) dW_s^\top \beta(\mathcal{S}_{t-s}^* \gamma_s) + \int_0^t \beta(\mathcal{S}_{t-s}^* \gamma_s)^\top dW_s \beta(\mathcal{S}_{t-s}^* \nu). \end{aligned}$$

Defining as in Eq. (4.6) an \mathbb{S}^d -valued kernel via

$$K(t) = \beta(\mathcal{S}_t^* \nu),$$

we obtain the following generalized \mathbb{S}_+^d -valued Volterra equation

$$V_t := \widehat{\beta}(\lambda_t) = \widehat{\beta}(\mathcal{S}_t^{*,\widehat{\otimes}}\lambda_0) + n \int_0^t K(t-s)K(t-s)ds + \int_0^t K(t-s)dW_s^\top \beta(\mathcal{S}_{t-s}^*\gamma_s) + \int_0^t \beta(\mathcal{S}_{t-s}^*\gamma_s)^\top dW_s K(t-s), \quad (5.10)$$

which we call *Volterra Wishart process* in the following definition.

Definition 5.3 For $\beta, \widehat{\beta}$ as given in (5.8)–(5.9) and an \mathbb{S}^d -valued kernel $K(t)$ defined by $K(t) = \beta(\mathcal{S}_t^*v)$, we call the process defined in (5.10), *Volterra Wishart process*.

Remark 5.4 (i) Note that $\beta(\gamma_t)$ defines an $\mathbb{R}^{n \times d}$ -valued Volterra OU process, that is,

$$X_t := \beta(\gamma_t) = \beta(\mathcal{S}_t^*\gamma_0) + \int_0^t dW_s K(t-s). \quad (5.11)$$

By the definition of $\widehat{\beta}$, the Volterra Wishart process

$$V_t = \widehat{\beta}(\lambda_t) = \beta(\gamma_t(\cdot))^\top \beta(\gamma_t(\cdot)) = X_t^\top X_t$$

is thus the matrix square of a Volterra OU process, which justifies the terminology.

(ii) Note that different lifts of the Volterra OU process given in (5.11) are possible, e.g., the forward process lift $f_t(x) := \mathbb{E}[X_{t+x}|\mathcal{F}_t]$. Then, $f_t(0) = X_t$, and similarly as in Cuchiero and Teichmann (2018, Section 5.2), it can be shown that f is an infinite dimensional OU process that solves the following SPDE (in the mild sense)

$$df_t(x) = \frac{d}{dx} f_t(x)dt + dW_t K(x), \quad f_0(x) = \beta(\mathcal{S}_x^*\gamma_0),$$

on a Hilbert space H of absolutely continuous functions (AC) with values in $\mathbb{R}^{n \times d}$, precisely $H = \{f \in AC(\mathbb{R}_+, \mathbb{R}^{n \times d}) \mid \int_0^\infty \|f'(x)\|^2 \alpha(x)dx < \infty\}$ where $\alpha > 0$ denotes a weight function (compare Filipović 2001). We can then set $\lambda_t(x, y) = f_t^\top(x)f_t(y)$ and define the same Volterra Wishart process as in (5.10) by $V_t := \lambda_t(0, 0) = X_t^\top X_t$. By Itô’s formula and variation of constants, its dynamics can then equivalently be expressed via

$$V_t := \lambda_t(0, 0) = f_0^\top(t)f_0(t) + n \int_0^t K(t-s)K(t-s)ds + \int_0^t K(t-s)dW_s^\top f_s(t-s) + \int_0^t f_s^\top(t-s)dW_s K(t-s). \quad (5.12)$$

Comparing (5.12) and (5.10) yields

$$\beta(\mathcal{S}_x^* \gamma_t) = f_t(x) = \mathbb{E}[X_{t+x} | \mathcal{F}_t], \quad x, t \geq 0. \quad (5.13)$$

(iii) In the case when β and \mathcal{S}^* are as in Example 4.4, (5.10) reads as

$$\begin{aligned} \int_{\mathbb{R}^2} \lambda(dx_1, dx_2) &= \int_{\mathbb{R}^2} e^{-(x_1+x_2)t} \lambda_0(dx_1, dx_2) + n \int_0^t K(t-s)K(t-s)ds \\ &\quad + \int_0^t \int_0^\infty K(t-s) dW_s^\top e^{-x(t-s)} \gamma_s(dx) \\ &\quad + \int_0^t \int_0^\infty e^{-x(t-s)} \gamma_s^\top(dx) dW_s K(t-s). \end{aligned}$$

Hence by (5.13), $\int_0^\infty e^{-x(t-s)} \gamma_s(dx) = \mathbb{E}[X_t | \mathcal{F}_s]$. This yields exactly Eq. (1.6) considered in the introduction. Note that if ν and in turn K are chosen as in Remark 4.5, this Volterra Wishart process has exactly the roughness properties desired in rough covariance modeling.

In the following remark, we list several properties of Volterra Wishart processes.

- Remark 5.5** (i) Note that the marginals of V are Wishart distributed as they arise from squares of Gaussians.
- (ii) In order to bring (5.6) in a “standard” Wishart form (with the matrix square root) as in (1.1) by replacing $\gamma(dx)$ by $\sqrt{\lambda}(dx, dy)$, new notation has to be introduced, compare with (5.7).
- (iii) Nevertheless, both the drift and the diffusion characteristics of λ depend linearly only on λ , e.g.,

$$\begin{aligned} \frac{d[\lambda_{ij}(dx_1, dx_2), \lambda_{kl}(dy_1, dy_2)]_t}{dt} &= (K(x_1)K(y_1))_{ik} \lambda_{t,jl}(dx_2, dy_2) \\ &\quad + (K(x_1)K(y_2))_{il} \lambda_{t,jk}(dx_2, dy_1) \\ &\quad + (K(x_2)K(y_1))_{jk} \lambda_{t,il}(dx_1, dy_2) \\ &\quad + (K(x_2)K(y_2))_{jl} \lambda_{t,ik}(dx_1, dy_1), \end{aligned}$$

which indicates that $(\lambda_t)_{t \geq 0}$ is Markovian on its own. This is shown rigorously below.

Using Theorem 2.8, we now show that λ is a generalized Feller process on $(\widehat{\mathcal{E}}, \widehat{\varrho})$ with weight function $\widehat{\varrho}$ satisfying

$$\widehat{\varrho}(\gamma \widehat{\otimes} \gamma) = \varrho(\gamma). \quad (5.14)$$

We also prove that this generalized Feller process is *affine*, in the sense that its Laplace transform is exponentially affine in the initial value. The process λ can therefore be viewed as an *infinite dimensional Wishart process* on $\widehat{\mathcal{E}}$ analogously to Bru (1991), Cuchiero et al. (2011).

Theorem 5.6 *The process λ defined in (5.5) is Markovian on $\widehat{\mathcal{E}}$. The corresponding semigroup is a generalized Feller semigroup on $\mathcal{B}^{\widehat{\varrho}}(\widehat{\mathcal{E}})$, where $\widehat{\varrho}$ satisfies (5.14). Moreover, for $y \in Y(\mathbb{R}^{n \times d})$,*

$$\mathbb{E}_{\lambda_0} [\exp(-\langle y \widehat{\otimes} y, \lambda_t \rangle)] = \exp(-\phi_t - \langle \psi_t, \lambda_0 \rangle), \tag{5.15}$$

where ψ and ϕ satisfy the following Riccati differential equations, namely $\psi_0 = y \widehat{\otimes} y$ and $\partial_t \psi_t = R(\psi_t)$ in the mild sense with $R : \widehat{\mathcal{E}}_* \rightarrow \widehat{\mathcal{E}}_*$ given by

$$R(y \widehat{\otimes} y)(x_1, x_2) = \mathcal{A}y(x_1) \widehat{\otimes} y(x_2) + y(x_1) \widehat{\otimes} \mathcal{A}y(x_2) - 2 \int_0^\infty \int_0^\infty y(dx_1) \widehat{\otimes} y(dx) \nu \widehat{\otimes} \nu(dx, dy) y(dy) \widehat{\otimes} y(dx_2)$$

and $\phi_0 = 0$ and $\partial_t \phi_t = F(\psi_t)$ with $F : \widehat{\mathcal{E}}_* \rightarrow \mathbb{R}$ given by

$$F(y \widehat{\otimes} y) = n \langle y \widehat{\otimes} y, \nu \widehat{\otimes} \nu \rangle.$$

Proof We apply Theorem 2.8 and Corollary 2.11 with

$$q : \mathcal{Y}^*(\mathbb{R}^{n \times d}) \rightarrow \widehat{\mathcal{E}}, \gamma \mapsto \gamma \widehat{\otimes} \gamma = \gamma(\cdot)^\top \gamma(\cdot).$$

Observe that this is a continuous map, since we use the dual topology $\sigma(\widehat{\mathcal{E}}, Y \otimes Y)$ on $\widehat{\mathcal{E}}$ and the respective polar $\widehat{\mathcal{E}}_*$ defined by (5.4). Consider now the following set of Fourier basis elements

$$\widehat{\mathcal{D}} = \{f_y : \widehat{\mathcal{E}} \rightarrow [0, 1]; \lambda \mapsto \exp(-\langle y \widehat{\otimes} y, \lambda \rangle) \mid y \in Y(\mathbb{R}^{n \times d})\}$$

which is dense in $\mathcal{B}^{\widehat{\varrho}}(\widehat{\mathcal{E}})$ by the very definition of the dual topology. We check now that the generalized Feller semigroup $P^{(OU)}$ corresponding to (5.1) satisfies Assumption (2.8) for $f \in \widehat{\mathcal{D}}$, i.e., for every $f \in \widehat{\mathcal{D}}$, there exists some g such that

$$P_t^{(OU)}(f \circ q) = g \circ q. \tag{5.16}$$

Hence, we need to compute $\mathbb{E}_{\gamma_0} [\exp(-\langle y \widehat{\otimes} y, \gamma_t \widehat{\otimes} \gamma_t \rangle)]$. By Lemma 5.7, this expression is given by (5.17). Therefore, (5.16) is clearly satisfied. This proves the first assertion. Concerning the affine property, we can deduce from Lemma 5.7 that ψ and ϕ are given by

$$\begin{aligned} \psi_t &= (2q_t(y \widehat{\otimes} y) + \text{Id}_d)^{-1} (\mathcal{S}_t y \widehat{\otimes} \mathcal{S}_t y), \\ \phi_t &= \frac{n}{2} \log \det(2q_t(y \widehat{\otimes} y) + \text{Id}_d), \end{aligned}$$

with q_t given in Lemma 5.7. Taking derivatives then leads to the form of the Riccati differential equations. □

The following lemma provides an explicit expression for the Laplace transform of $\gamma_t \widehat{\otimes} \gamma_t$. This resembles not surprisingly the Laplace transform of a non-central Wishart distribution with n degrees of freedom.

Lemma 5.7 *Let γ be an Ornstein–Uhlenbeck process as defined in (5.1). Then for $y \in Y(\mathbb{R}^{n \times d})$, the Laplace transform of $\gamma_t \widehat{\otimes} \gamma_t$ is given by*

$$\begin{aligned} \mathbb{E}_{\gamma_0} [\exp(-\langle y \widehat{\otimes} y, \gamma_t \widehat{\otimes} \gamma_t \rangle)] &= \det(2q_t(y \widehat{\otimes} y) + \text{Id}_d)^{-\frac{n}{2}} \\ &\quad \times \exp(-\langle (2q_t(y \widehat{\otimes} y) + \text{Id}_d)^{-1} (\mathcal{S}_t y \widehat{\otimes} \mathcal{S}_t y), \gamma_0 \widehat{\otimes} \gamma_0 \rangle), \end{aligned} \tag{5.17}$$

where $q_t(y \widehat{\otimes} y) = \int_0^t \int_0^\infty \int_0^\infty \mathcal{S}_s^* \nu(dx_1) y^\top(x_1) y(x_2) \mathcal{S}_s^* \nu(dx_2) ds$.

Proof Assume for simplicity first that \mathcal{A}^* is equal to 0. Then, (5.1) becomes

$$\gamma_t(dx) = \gamma_0(dx) + W_t \nu(dx).$$

Fix $y \in Y(\mathbb{R}^{n \times d})$ such that $\int_0^\infty y(x) \nu(dx)$ is well defined. We then have

$$\begin{aligned} \langle y \widehat{\otimes} y, \gamma_t \widehat{\otimes} \gamma_t \rangle &= \langle y \widehat{\otimes} y, (\gamma_0 + W_t \nu) \widehat{\otimes} (\gamma_0 + W_t \nu) \rangle \\ &= \langle y \widehat{\otimes} y, \gamma_0 \widehat{\otimes} \gamma_0 \rangle + \langle y \widehat{\otimes} y, \gamma_0 \widehat{\otimes} W_t \nu \rangle + \langle y \widehat{\otimes} y, W_t \nu \widehat{\otimes} \gamma_0 \rangle \\ &\quad + \langle y \widehat{\otimes} y, W_t \nu \widehat{\otimes} W_t \nu \rangle. \end{aligned}$$

Note now that

$$\begin{aligned} \langle y \widehat{\otimes} y, \gamma_0 \widehat{\otimes} W_t \nu \rangle &= \text{Tr} \left(\left(W_t \int_0^\infty \int_0^\infty \nu(dx_2) y^\top(x_1) y(x_2) \gamma_0^\top(dx_1) \right) \right) \\ &=: \text{Tr}(W_t a), \\ \langle y \widehat{\otimes} y, W_t \nu \widehat{\otimes} \gamma_0 \rangle &= \text{Tr} \left(\left(\int_0^\infty \int_0^\infty \gamma_0(dx_2) y^\top(x_1) y(x_2) \nu(dx_1) \right) W_t^\top \right) \\ &=: \text{Tr}(a_1 W_t^\top) = \text{Tr}(W_t a_1^\top) = \text{Tr}(W_t a), \\ \langle y \widehat{\otimes} y, W_t \nu \widehat{\otimes} W_t \nu \rangle &= \text{Tr} \left(\left(\int_0^\infty \int_0^\infty \nu(dx_2) y^\top(x_1) y(x_2) \nu(dx_1) \right) W_t^\top W_t \right) \\ &=: \text{Tr}(b W_t^\top W_t), \end{aligned}$$

where $a \in \mathbb{R}^{d \times n}$, $a_1 \in \mathbb{R}^{n \times d}$, $b \in \mathbb{R}^{d \times d}$ and $a = a_1^\top$.

For the following calculation, let $n = 1$. Then, using these expressions, we find

$$\begin{aligned} &\mathbb{E} [\exp(-\langle y \widehat{\otimes} y, \gamma_t \widehat{\otimes} \gamma_t \rangle)] \\ &= \exp(-\langle y \widehat{\otimes} y, \gamma_0 \widehat{\otimes} \gamma_0 \rangle) \mathbb{E} \left[\exp(-2 \text{Tr}(W_t a) - \text{Tr}(b W_t^\top W_t)) \right] \\ &= \exp(-\langle y \widehat{\otimes} y, \gamma_0 \widehat{\otimes} \gamma_0 \rangle) \frac{1}{(2\pi)^{\frac{d}{2}} t^{\frac{d}{2}}} \int_{\mathbb{R}^{1 \times d}} e^{-2 \text{Tr}(xa) - \text{Tr}(bx^\top x) - \frac{1}{2t} xx^\top} dx \\ &= \exp(-\langle y \widehat{\otimes} y, \gamma_0 \widehat{\otimes} \gamma_0 \rangle) \end{aligned}$$

$$\begin{aligned} & \times \frac{1}{\det(2b + \frac{1}{t} \text{Id}_d)^{\frac{1}{2}} t^{\frac{d}{2}}} \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}^{1 \times d}} e^{-2xa - \frac{1}{2}x(2b + \frac{1}{t} \text{Id}_d)x^\top} \det(2b + \frac{1}{t} \text{Id}_d)^{\frac{1}{2}} dx \\ & = \frac{1}{\det(2b + \frac{1}{t} \text{Id}_d)^{\frac{1}{2}} t^{\frac{d}{2}}} \exp(-\langle y \widehat{\otimes} y, \gamma_0 \widehat{\otimes} \gamma_0 \rangle) \exp(2a^\top (2b + \frac{1}{t} \text{Id}_d)^{-1} a), \end{aligned}$$

where in the last line we used the formula for the moment generating function of a Gaussian random variable with covariance $(2b + \frac{1}{t} \text{Id}_d)^{-1}$. Simplifying further yields

$$\begin{aligned} & \mathbb{E} [\exp(-\langle y \widehat{\otimes} y, \gamma_t \widehat{\otimes} \gamma_t \rangle)] \\ & = \frac{1}{\det(2b + \frac{1}{t} \text{Id}_d)^{\frac{1}{2}} t^{\frac{d}{2}}} \exp(\langle (2b(2b + \frac{1}{t} \text{Id}_d)^{-1} - \text{Id}_d)(y \widehat{\otimes} y), \gamma_0 \widehat{\otimes} \gamma_0 \rangle) \\ & = \frac{1}{\det(2bt + \text{Id}_d)^{\frac{1}{2}}} \exp(\langle -(\text{Id}_d + 2bt)^{-1} (y \widehat{\otimes} y), \gamma_0 \widehat{\otimes} \gamma_0 \rangle). \end{aligned} \tag{5.18}$$

For general n , note that we can write

$$W_t^\top W_t = \sum_{j=1}^n W_{j,t}^\top W_{j,t},$$

where the W_j are the rows of W and thus take values in $\mathbb{R}^{1 \times d}$. Similarly,

$$\text{Tr}(W_t a) = \text{Tr} \left(\sum_{j=1}^n W_{j,t} \left(\int_0^\infty \int_0^\infty v(dx_2) y^\top(x_1) y(x_2) \gamma_{0,j}^\top(dx_1) \right) \right) =: \sum_{j=1}^n W_{j,t} a_j,$$

where $\gamma_{0,j}$ are the rows of γ_0 . Using the independence of all W_j and applying (5.18) then lead to

$$\mathbb{E} [\exp(-\langle y \widehat{\otimes} y, \gamma_t \widehat{\otimes} \gamma_t \rangle)] = \frac{1}{\det(2bt + \text{Id}_d)^{\frac{n}{2}}} \exp(-\langle (\text{Id}_d + 2bt)^{-1} (y \widehat{\otimes} y), \gamma_0 \widehat{\otimes} \gamma_0 \rangle).$$

The general case for $\mathcal{A}^* \neq 0$ can now be traced back to this situation. Indeed, by the variation of constants formula, γ_t is given by

$$\gamma_t = \mathcal{S}_t^* \gamma_0 + \int_0^t dW_s \mathcal{S}_{t-s}^* \nu(dx).$$

Therefore, we need to replace bt by

$$q_t = \int_0^t \int_0^\infty \int_0^\infty \mathcal{S}_{t-s}^* \nu(dx_1) y^\top(x_1) y(x_2) \mathcal{S}_{t-s}^* \nu(dx_2) ds$$

and γ_0 by $\mathcal{S}_t^* \gamma_0$. This then yields (5.17). Note that this now holds for general $y \in Y(\mathbb{R}^{n \times d})$ even if $\int_0^\infty y(x) \nu(dx)$ is not necessarily well defined. \square

6 (Rough) Volterra-type affine covariance models

The goal of this section is to apply the above constructed affine covariance models for multivariate stochastic volatility models with d assets. We exemplify this with the Volterra Wishart process of Sect. 5 and define a (rough) multivariate Volterra Heston-type model with possible jumps in the price process. Roughness can be achieved by specifying ν and in turn the kernel of the Volterra Wishart process as in Remark 4.5. The log-price process denoted by P and taking values in \mathbb{R}^d evolves according to

$$dP_t = -\frac{1}{2} \text{diag}(V_t)dt - \int_{\mathbb{R}^d} (e^\xi - \mathbf{1} - \xi) \text{Tr}(V_t m(d\xi)) + X_t^\top dB_t + \int_{\mathbb{R}^d} \xi(\mu^P(d\xi) - \text{Tr}(V_t m(d\xi))), \tag{6.1}$$

where X_t denotes the Volterra OU process defined in Remark 5.4, $\mathbf{1}$ the vector in \mathbb{R}^d with all entries being 1 and e^ξ has to be understood componentwise. Moreover, B_t is an \mathbb{R}^n -valued Brownian motion, which can be correlated with the matrix Brownian motion W appearing in (5.1) as follows

$$B_t = W_t \varrho + \sqrt{(1 - \varrho^\top \varrho)} \tilde{B}_t,$$

where \tilde{B}_t is an \mathbb{R}^n -valued Brownian motion independent of W and $\varrho \in \mathbb{R}^d$. Moreover, μ^P denotes the random measure of the jumps with compensator $\text{Tr}(V m(d\xi))$, where V is the Volterra Wishart process of (5.10) and m a positive semidefinite measure supported on \mathbb{R}^d .

As a corollary of Section 5 and Cuchiero (2011, Section 5), we obtain the following result, namely that the log-price process together with the infinite dimensional Wishart process λ given in (5.5) is an affine Markov process.

Before formulating the precise statement, note that the continuous covariation² $\langle P_i, \lambda_{kl}(dx_1, dx_2) \rangle_t$ is given by

$$\frac{\langle P_i, \lambda_{kl}(dx_1, dx_2) \rangle_t}{dt} = (\beta^\top(\gamma_t) \gamma_t(dx_1))_{il} (\nu(dx_2) \varrho)_k + (\beta^\top(\gamma_t) \gamma_t(dx_1))_{ik} (\nu(dx_2) \varrho)_l,$$

where γ is the infinite dimensional OU process of (5.1). Note that $\beta^\top(\gamma_t) \gamma_t(dx_1)$ can also be written as linear map from $\widehat{\mathcal{E}} \rightarrow Y^*(\mathbb{S}^d)$ which we denote by $\tilde{\beta}$, i.e.,

$$\tilde{\beta}(\lambda_t)(dx_1) = \beta^\top(\gamma_t) \gamma_t(dx_1). \tag{6.2}$$

In the standard example of 4.4, we have $\tilde{\beta}(\lambda)(dx_1) = \int_{x_2} \lambda(dx_1, dx_2)$. The adjoint operator of $\tilde{\beta}$ from $Y(\mathbb{S}^d)$ to $Y(\mathbb{R}^{n \times d}) \widehat{\otimes} Y(\mathbb{R}^{n \times d})$ is denoted by $\tilde{\beta}_*$ and given by

$$\langle \tilde{\beta}(\lambda), y \rangle = \langle \lambda, \tilde{\beta}_*(y) \rangle, \quad y \in Y(\mathbb{S}^d),$$

² Here, the brackets stand for the covariation and not for the pairing.

where the brackets are the pairings in the respective spaces. With this notation, we are now ready to state the result. Its proof is a combination of the results of Section 5 and Cuchiero (2011, Section 5).

Corollary 6.1 *The joint process (λ, P) with λ defined in (5.5) and P defined in (6.1) is Markovian with state space $(\widehat{\mathcal{E}}, \mathbb{R}^d)$. It is affine in the sense that for $(y, v) \in Y(\mathbb{R}^{n \times d}) \times \mathbb{R}^d$, we have*

$$\mathbb{E}_{\lambda_0, P_0} \left[\exp \left(-\langle y \widehat{\otimes} y, \lambda_t \rangle + iv^\top P_t \right) \right] = \exp(-\phi_t - \langle \psi_t, \lambda_0 \rangle + iv^\top P_0). \tag{6.3}$$

The function ψ satisfies the following Riccati differential equations, namely $\psi_0 = y \widehat{\otimes} y$ and $\partial_t \psi_t = R(\psi_t, iv)$, in the mild sense with $R : \widehat{\mathcal{E}}_* \times i\mathbb{R}^d \rightarrow \widehat{\mathcal{E}}_*$ given by

$$\begin{aligned} R(y \widehat{\otimes} y, iv)(x_1, x_2) &= \mathcal{A}y(x_1) \widehat{\otimes} y(x_2) + y(x_1) \widehat{\otimes} \mathcal{A}y(x_2) \\ &\quad - 2 \int_0^\infty \int_0^\infty y(dx_1) \widehat{\otimes} y(dx) v \widehat{\otimes} v(dx, dy) y(dy) \widehat{\otimes} y(dx_2) \\ &\quad + \frac{1}{2} \sum_{i=1}^d iv_i \widehat{\beta}_*(e_i e_i^\top)(x_1, x_2) \\ &\quad + \widehat{\beta}_* \left(\int_{\mathbb{R}^d} (iv^\top (e^\xi - \mathbf{1} - \xi)) m(d\xi) \right)(x_1, x_2) \\ &\quad + \frac{1}{2} \widehat{\beta}_*(vv^\top)(x_1, x_2) \\ &\quad + \widetilde{\beta}_* \left(\int_0^\infty y(\cdot) \widehat{\otimes} y(x) v(dx) \right)(x_1, x_2) \varrho iv^\top \\ &\quad + iv \varrho^\top \widetilde{\beta}_* \left(\int_0^\infty v(dx) y(\cdot) \widehat{\otimes} y(x) \right)(x_1, x_2) \\ &\quad - \widehat{\beta}_* \left(\int_{\mathbb{R}^d} (\exp(iv^\top \xi) - 1 - iv^\top \xi) m(d\xi) \right)(x_1, x_2), \end{aligned}$$

where $\widehat{\beta}_*$ and $\widetilde{\beta}_*$ are the adjoint operators of $\widehat{\beta}$ given in (5.9) and $\widetilde{\beta}$ given in (6.2), respectively. The function ϕ satisfies $\phi_0 = 0$ and $\partial_t \phi_t = F(\psi_t)$ with $F : \widehat{\mathcal{E}}_* \rightarrow \mathbb{R}$ given by

$$F(y \widehat{\otimes} y) = n \langle y \widehat{\otimes} y, v \widehat{\otimes} v \rangle.$$

Remark 6.2 In a similar spirit, one can define multivariate affine covariance models with the affine Volterra jump process V given in (4.23). The log-price process (under some risk neutral measure) evolves then according to

$$\begin{aligned} dP_t &= -\frac{1}{2} \text{diag}(V_t) dt - \int_{\mathbb{R}^d} (e^\xi - \mathbf{1} - \xi) \text{Tr}(V_t m(d\xi)) + \sqrt{V_t} dB_t \\ &\quad + \int_{\mathbb{R}^d} \xi (\mu^P(d\xi) - \text{Tr}(V_t m(d\xi))), \end{aligned}$$

where B is a d -dimensional Brownian motion and the jump measure m of P and μ of the Markovian lift λ as given in (4.17) can be the marginals of some common measure supported on $\mathbb{S}_+^d \times \mathbb{R}^d$.

Acknowledgements Open access funding provided by Vienna University of Economics and Business (WU).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Abi Jaber, E., El Euch, O.: Markovian structure of the Volterra Heston model. *Stat. Probab. Lett.* **149**, 63–72 (2019)
- Abi Jaber, E., Larsson, M., Pulido, S.: Affine Volterra processes. *Ann. Appl. Probab.* (2019) (to appear)
- Alòs, E., León, J., Vives, J.: On the short-time behavior of the implied volatility for jump-diffusion models with stochastic volatility. *Finance Stoch.* **11**(4), 571–589 (2007)
- Alòs, E., Yang, Y.: A fractional Heston model with $H > 1/2$. *Stochastics* **89**(1), 384–399 (2017)
- Bayer, C., Friz, P., Gatheral, J.: Pricing under rough volatility. *Quant. Finance* **16**(6), 887–904 (2016)
- Bru, M.F.: Wishart processes. *J. Theor. Probab.* **4**(4), 725–751 (1991)
- Cuchiero, C.: Affine and polynomial processes. Ph.D. thesis, ETH Zürich (2011)
- Cuchiero, C., Filipović, D., Mayerhofer, E., Teichmann, J.: Affine processes on positive semidefinite matrices. *Ann. Appl. Probab.* **21**(2), 397–463 (2011)
- Cuchiero, C., Teichmann, J.: Generalized Feller processes and Markovian lifts of stochastic Volterra processes: the affine case. [arXiv:1804.10450](https://arxiv.org/abs/1804.10450) (2018)
- Dalang, R.C.: Extending the martingale measure stochastic integral with applications to spatially homogeneous SPDEs. *Electron. J. Probab.* **4**(6), 29 (1999)
- Dörsek, P., Teichmann, J.: A semigroup point of view on splitting schemes for stochastic (partial) differential equations. [arXiv:1011.2651](https://arxiv.org/abs/1011.2651) (2010)
- El Euch, O., Rosenbaum, M.: The characteristic function of rough Heston models. *Math. Finance* **29**(1), 3–38 (2019)
- Engel, K.J., Nagel, R.: One-Parameter Semigroups for Linear Evolution Equation. Volume 194 of Graduate Texts in Mathematics. Springer, New York (2000). (With contributions by S. Brendle, M. Campiti, T. Hahn, G. Metafune, G. Nickel, D. Pallara, C. Perazzoli, A. Rhandi, S. Romanelli and R. Schnaubelt)
- Ethier, S.N., Kurtz, T.G.: Markov Processes: Characterization and Convergence. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. Wiley, New York (1986)
- Filipović, D.: Consistency Problems for Heath–Jarrow–Morton Interest Rate Models. Lecture Notes in Mathematics, vol. 1760. Springer, Berlin (2001)
- Gatheral, J., Jaisson, T., Rosenbaum, M.: Volatility is rough. *Quant. Finance* **18**(6), 933–949 (2018)
- Gripenberg, G., Londen, S.-O., Staffans, O.: Volterra Integral and Functional Equations. Volume 34 of Encyclopedia of Mathematics and Its Applications. Cambridge University Press, Cambridge (1990)
- Hawkes, A.G.: Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58**(1), 83–90 (1971)
- Mayerhofer, E.: Affine processes on positive semidefinite dd matrices have jumps of finite variation in dimension $d > 1$. *Stoch. Process. Appl.* **122**(10), 3445–3459 (2012)
- Pazy, A.: Semigroups of Linear Operators and Applications to Partial Differential Equations. Volume 44 of Applied Mathematical Sciences. Springer, New York (1983)
- Schaefer, H.H., Wolff, M.P.: Topological Vector Spaces. Volume 3 of Graduate Texts in Mathematics, 2nd edn. Springer, New York (1999)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.