CrossMark

Tensor Approximation of Generalized Correlated Diffusions and Functional Copula Operators

Antonio Dalessandro[1] · Gareth W. Peters[1]

Abstract In this paper we develop a class of applied probabilistic continuous time but discretized state space decompositions of the characterization of a multivariate generalized diffusion process. This decomposition is novel and, in particular, it allows one to construct families of mimicking classes of processes for such continuous state and continuous time diffusions in the form of a discrete state space but continuous time Markov chain representation. Furthermore, we present this novel decomposition and study its discretization properties from several perspectives. This class of decomposition both brings insight into understanding locally in the state space the induced dependence structures from the generalized diffusion process as well as admitting computationally efficient representations in order to evaluate functionals of generalized multivariate diffusion processes, which is based on a simple rank one tensor approximation of the exact representation. In particular, we investigate aspects of semimartingale decompositions, approximation and the martingale representation for multidimensional correlated Markov processes. A new interpretation of the dependence among processes is given using the martingale approach. We show that it is possible to represent, in both continuous and discrete space, that a multidimensional correlated generalized diffusion is a linear combination of processes originated from the decomposition of the starting multidimensional semimartingale. This result not only reconciles with the existing theory of diffusion approximations and decompositions, but defines the general representation of infinitesimal generators for both multidimensional generalized diffusions and, as we will demonstrate, also for the specification of copula density dependence structures. This new result provides immediate representation of the approximate weak solution for correlated stochastic differential equations. Finally, we demonstrate desirable convergence results for the proposed multidimensional semimartingales decomposition approximations.

✉ Antonio Dalessandro
info@antonio-dalessandro.com

[1]   Department of Statistical Science, University College London, London, UK

🙋 Springer

1 Introduction

We develop a class of applied probabilistic continuous time but discretized state space decompositions of the characterization of a multivariate generalized diffusion process. The general results in this paper are aimed at developing representations of such generalized diffusion processes that are tractable to work with both algebraically and computationally. To achieve this we develop novel mimicking representations of generalized diffusion processes which are based around asymptotically exact continuous time, discrete state-space, Markov chain representations. We derive such representations, prove there exactness in a weak sense and also develop a class of simulation based approximation algorithms for efficient and accurate computations of such classes of mimicking process. The representations we develop are novel extensions of ideas considered in works such as Hackbusch ([2012](#)).

We present and develop different algebraic tensor representations of this novel decomposition approach that are beneficial for different numerical approximation settings and in particular we study the approximation behaviour and discretization properties from several perspectives. This class of decomposition both brings insight into understanding locally in the state space the induced dependence structures from the generalized diffusion process as well as admitting computationally efficient representations in order to evaluate functionals of generalized multivariate diffusion processes.

We consider the multidimensional stochastic differential equation (SDE) of the form

$$dX_t = b(t, X_t)dt + \Sigma(t, X_t)dW_t, \tag{1}$$

where $b(t, X_t) : [0, \infty) \times \mathbb{R}^d \to \mathbb{R}^d$ and $\Sigma(t, X_t) : [0, \infty) \times \mathbb{R}^d \to \mathbb{R}^{d \times d}$ and we assume that $b = (b_j)$ is continuous vector valued function and $\Sigma = ((\Sigma_{ij}))$ is a continuous, symmetric, nonnegative definite, $d \times d$ matrix valued function. Let the infinitesimal generator associated to the SDE of Eq. ([1](#)) be denoted,

$$A = \left\{ \left(f, Gf = \frac{1}{2} \sum_{ij}^{d} \Sigma_{ij} \partial_i \partial_j f + \sum_{i}^{d} b_i \partial_i f \right) : f \in C_c^\infty(\mathbb{R}^d) \right\}, \tag{2}$$

where $C_c^\infty(\mathbb{R}^d)$ is the set of smooth functions with compact support on $\mathbb{R}^d$.

Specifically, the aim of our work is to develop tractable and computationally efficient mimicking representations of the solution for the generalized multivariate correlated diffusion processes of Eq. ([1](#)), *in settings which may involve potentially very high dimensional state spaces displaying non-trivial dependence structures.*

We propose two ways to calculate the weak solution of the SDE in (Eq. [1](#)) in an approximate manner:

1. Direct approximation of the infinitesimal generator A with particular emphasis on the structure of its mixed derivative terms.
2. Decomposition of the infinitesimal generator A into orthogonal components.

The proposed approximation schemes are based on tensor algebra decompositions such as those considered in Hackbusch ([2012](#)). The novelty of our work consists in the introduction

of new concepts like the copula infinitesimal generator, correlated tensor representation, conditional infinitesimal generator and the framework developed to perform a parametric copula function mapping. The proposed results aim to develop a new characterization of the cross space among dimensions using tensor algebra and they have a general validity. Furthermore, our schemes are supported by approximation and convergence results that constitute a very useful perspective to look at solutions of multidimensional SDEs.

Our investigation is focused on aspects of the semimartingale decomposition and martingale representation for multi-dimensional correlated Markov processes. The objective is to construct a continuous time Markov chain (CTMC) that approximates or mimicks such processes and their dependence structures induced throughout the state space, which are only implicitly defined by the joint structure of the marginal process volatility functional forms and the joint coupling of the correlation structures in the driving noise processes. Once such a mimicking process is obtained we may then transform its structure to produce a copula function mapping theorem that allows one to obtain structural characterizations of general dependence frameworks, through the specification of the mimicking multivariate diffusion process.

This work is motivated by the problems of finding the expression of a multi-dimensional CTMC that both closely follows the dynamics of the corresponding correlated Ito-processes and also can effectively deal with the representation and simulation of large dimensional processes that exhibit various correlation structures. Although the literature on Markov processes and Markov chains is very rich and mature, see (Rogers and Williams 2000; Ethier and Kurtz 1985; Karatzas and Shreve 2000; Kushner and Dupuis 2001), we find that there is still room for further investigation and characterization of multi-dimensional chains and the relationship between the correlation structure among marginal Markov chains and dependence concepts like copula functions (Nelsen 1999) and concordance measures of dependence (McNeil et al. 2005; Scarsini 1984). In fact these concepts have always been treated separately, and there is lack in literature of a theory that begins to reconcile them.

Our findings and results show that our approach, based on linear and tensor algebra, is a powerful way to produce accurate solutions of multidimensional correlated SDEs that exhibit a correlation that can be fully modelled through copula functions. Specifically given a multi-dimensional Ito processes whose drift and diffusion terms are adapted processes, we show how to construct the approximated infinitesimal generator and how to characterize the process properties by its associated continuous time Markov chain (CTMC). We construct an approximated weak solution to the stochastic differential equation that weakly converges to the distribution of the multi-dimensional Ito processes.

We develop an interpretation for the correlation among processes using the martingale approach applied to the study of diffusions. The novelty is that it is possible to represent, in both continuous and discrete space, that a multidimensional correlated generalized diffusion is a linear combination of processes that originate from the decomposition of the starting multidimensional semimartingale.

The only assumption required by our approximation approach is that the martingale problem for the associated generator of the multidimensional Markov process is well posed. Stroock and Srinivasa Varadhan (1997) formulated the martingale problem as a means of studying Markov processes, especially multidimensional diffusions. This approach is deemed to be more powerful and more intrinsic than the alternative approaches represented by the Markov process approach and the Ito approach. More recently some authors (Brunick 2013) extended the study of the martingale problem associated to operators of the same type as in (Eq. 2) and showed that is well posed when the covariance matrix takes a particular lower-diagonal block form.

Our result reconciles with the existing theory of diffusion approximations and decompositions existing in the probability literature and is more closely related to the work of Gyöngy (1986) and more recently to Brunick and Shreve (2013). In the seminal manuscript (Gyöngy 1986) considers a multi-dimensional Ito process, and constructs a weak solution to a stochastic differential equation which mimics the marginals of the original Ito process at each fixed time instant. The drift and covariance coefficients for the mimicking process can be interpreted as the expected value of the instantaneous drift and covariance of the original Ito process, conditional on its terminal value. In Brunick and Shreve (2013) the authors extend the result of Gyöngy (1986), proving that they can match the joint distribution at each fixed time for various functionals of the Ito process. The mimicking process takes the form of a stochastic functional differential equation and the diffusion coefficient is given by the so-called Markovian projection. In our framework we further generalize findings from Brunick and Shreve (2013) and the mimicking process takes form of a sequence of conditional continuous time Markov chains with instantaneous drift and diffusion coefficients given by projected instantaneous local moments. We use generalized diffusion approximations, similarly to the work of Ekström et al. (2013), in order to produce desired target multivariate distributions. However the results presented in Ekström et al. (2013) are limited to the univariate case and the authors propose an approach that involves the speed of measure of a diffusion and time-changes of a Brownian motion allowing for target distributions with arbitrary support. The results reported in this manuscript define the general representation of the approximated infinitesimal generators for both multidimensional generalized diffusions and for what we will define as the function copula specification.

The paper is organized as following: Section 2 introduces the martingale problem for correlated Markov processes. In Section 3.2 we introduce and characterize the approximation schemes for the infinitesimal generator of correlated Markov processes while in Section 4 we proposed some desirable convergence results of our approximations and then in Section 5 we develop an applications section which also presents a generic algorithm for implementation of the mimicking process solutions developed.

2 Martingale Problems for Correlated Markov Processes

Our analysis takes place on a measure space $(\Omega, \mathscr{F}, \mu)$, where μ is a non zero σ-finite, positive measure on the measurable space $(\Omega, \mathscr{F})$. We denote by $\|f\|_p$ for $p \in [1, \infty]$, the $L^p(\mu)$ norm of a function f: $\|f\|_p^p = \int |f|^p d\mu$. Furthermore we define the Markov semigroup $(P_t(x, dy))$ on a metric, complete, separable space $\mathbf{E}$, as the family of probability transition kernels on $\mathbf{E}$ depending on the parameter $t \in [0, \infty)$, such that

$$\int_{y \in \mathbf{E}} P_s(x, dy) P_t(y, dz) = P_{s+t}(x, dz), \text{ for } s, t \in [0, \infty), \tag{3}$$

and its action on bounded and positive functions is denoted by

$$P_t f(x) = \int f(y) P_t(x, dy). \tag{4}$$

The family of operators P_t satisfies the following axioms $P_0 = I$, $P_t \circ P_s = P_{t+s}$, $\lim_{t \to 0^+} \|P_t f - f\| = 0$. Markov processes are naturally related to Markov semigroups because the probability measure $P_t(x, dy)$ with $P_t f(x) = \mathbb{E}_x[f(X_t)]$ is the law of $\{X_t\}_{t \geq 0}$

the process itself starting from the value x at time 0. We assume in this paper that the measure μ is an invariant measure for the semigroup P_t, and this means that for all $f \in L^1(\mu)$,

$$\int P_t f(x)\mu(dx) = \int f(x)\mu(dx),$$

and that P_t is a contraction semigroup in $L^1(\mu)$ for all t. We define the infinitesimal generator of a strongly continuous contraction semigroup by the map

$$Af := \lim_{t \to 0^+} \frac{P_t f - f}{t}, \tag{5}$$

for all $f \in D_{\mathbf{E}}$, where the space $D_{\mathbf{E}}$, subspace of $L^p(\mu)$, is the space of right continuous functions $f : \mathbf{E} \to \mathbb{R}$ with left limits, see Ch. 5, pp. 115 of Ethier and Kurtz (1985) for more details. The properties of A and $D_{\mathbf{E}}$ entirely specify the semigroup P_t. In fact given $f \in D_{\mathbf{E}}$ the function $U(x, t) = P_t f(x)$ is the unique solution of the equation

$$\frac{\partial U(x, t)}{\partial t} = A U(x, t), \tag{6}$$

defined in $D_{\mathbf{E}}$ for all $t > 0$ and with $U(x, 0) = \nu$, the initial probability distribution. We restrict ourself to the set of compactly supported functions denoted by $C_c^\infty(\mathbb{R}^d)$ and belonging to the set $\bar{D} \subseteq D_{\mathbf{E}}$, where $\bar{D}$ is defined as the core of A, and we study only generators of local diffusions of the form

$$A_s = \sum_i^d b_i(s, x) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j}^d \Sigma_{i,j}(s, x) \frac{\partial^2}{\partial x_i \partial x_j}, \tag{7}$$

where $b = (b_i(s, x))$, $i = 1, \ldots, d$ is a drift vector and $\Sigma = ((\Sigma_{i,j}))$, $i, j = 1, \ldots, d$ is a dispersion matrix with $\Sigma = CC'$ that characterize locally stochastic differential equations (SDE) with expression

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t C(s, X_s)dW_s, \tag{8}$$

where W is a d-dimensional Brownian motion. We note that the dynamics of X_t are characterized completely by the infinitesimal operator and therefore by the laws of the drift and diffusion coefficients, including the conditional probability law. In particular the infinitesimal operator in Eq. (26) is specified also by such coefficient with an explicit expression. In this respect the connection between the generator of X_t and the solution of the SDE for X_t has a rigorous formulation given by the Martingale problem of Stroock and Srinivasa Varadhan (1997). It is straightforward to show the connection between the differential operator A and the probabilistic interpretation of the solution to the corresponding SDE. If $f \in C_c^2(\mathbb{R}^d)$ then Ito's Lemma yields

$$f(X(t)) = f(X(0)) + \int_0^t Af(X(s))ds + \int_0^t \nabla f(X(s))'\sigma(X(s))dW(s). \tag{9}$$

This means that

$$f(X_t) - f(X_0) - \int_0^t Af(X_s)ds, \tag{10}$$

is a local martingale. In particular the processes

$$M_t = X(t) - X(0) - \int_0^t b(X(s))ds = \int_0^t \sigma(X(s))dW(s), \tag{11}$$

$$M_t^i M_t^j - \int_0^t \sigma_{ij}(X_s)ds, \tag{12}$$

are local martingales.

Definition 1 *Denote the generic probability measure ν on $(C[0,\infty)^d, \mathscr{B}(C[0,\infty)^d))$ under which*

$$M_f(t) = f(X_t) - f(X_0) - \int_0^t A_s f(X_s))ds, \tag{13}$$

is a martingale for every $f \in \bar{D}$, and is called a solution to the martingale problem associated with the operator A_t, where A_t is as defined in (Eq. 7).

In addition, it will also be useful to observe the following representation and uniqueness results regarding the process X_t and the infinitesimal operator A_t, as detailed for instance in Stroock and Srinivasa Varadhan (1997).

Theorem 1 *The process X_t is a weak solution to the SDE of* (Eq. 1) *if and only if it satisfies the martingale problem of* (Eq. 1) *with A_t as the infinitesimal operator of X_t as defined in* (Eq. 7).

Another fundamental result we mention that is directly relevant background to the framework we develop is that given by the following theorem.

Theorem 2 (**Uniqueness of SDE solution.**) (Stroock and Srinivasa Varadhan 1997, Th. 3.2.1.) *The SDE*

$$dX_t = b(X_t)dt + C(X_t)dW_t, 0 \le t \le s, X_0 \sim \nu(dx), \tag{14}$$

with X_0 a random variable independent from W_t, $t \ge 0$ and $\mathbb{E}[|X_0|^2] < \infty$, has unique solution adapted to the filtration generated by W_t and X_0 if the measurable functions $b(x)$ and $C(x)$, for $t \le s$, satisfy the following two conditions for a positive real constant K such that:

1. *Lipschitz continuity*

$$|b(x) - b(y)| + |C(x) - C(y)| \le K(|x - y|), for\, all x, y \in \mathbb{R}^d, \tag{15}$$

2. *Linear Growth*

$$|b(x)| + |C(x)| \le K(1 + |x|), for\, all x \in \mathbb{R}^d. \tag{16}$$

For completeness we give the definition of generalized diffusion in Definition 2, also see the reference (Portenko 1982), pag. 121. For a homogeneous Markov process in $\mathbb{R}^n$ with transition probability $P(t, x, \Gamma)$ where Γ is a Borel subset of $\mathbb{R}^d$, let

$$a_\varepsilon(t, x) = \frac{1}{t} \int_{|y-x| \leq \varepsilon} (x - y) P(t, x, dy), \tag{17}$$

$$(b_\varepsilon(t, x)\theta, \theta) = \frac{1}{t} \int_{|y-x| \leq \varepsilon} (x - y, \theta)^2 P(t, x, dy), \tag{18}$$

$$c_\varepsilon(t, x) = \frac{1}{t} \int_{|y-x| \leq \varepsilon} P(t, x, dy), \tag{19}$$

where $t > 0$, $\varepsilon > 0$, and $x, \theta \in \mathbb{R}^d$. Let $C_0(\mathbb{R}^d)$ be the collection of all real valued continuous function with compact support of $\mathbb{R}^d$, and $L^s(\mathbb{R}^d)$ the collection of symmetric linear operators acting on $\mathbb{R}^n$.

Definition 2 *[Generalized diffusion process] A Markov process with values in $\mathbb{R}^d$ and with transition probability $P(t, x, \Gamma)$ is called a generalized diffusion process if the following two conditions holds:*

1. *For all $\varepsilon > 0$ and $\varphi \in C_0(\mathbb{R}^n)$*

$$\lim_{t \to 0} \int_{\mathbb{R}^n} \varphi(x) c_\varepsilon(t, x) dx = 0 \tag{20}$$

2. *There exist linear functionals $A(\varphi)$ and $B(\varphi)$, $\varphi \in C_0(\mathbb{R}^n)$, with values in $\mathbb{R}^n$ and $L^s(\mathbb{R}^n)$, respectively, such that for some $\varepsilon > 0$ and all $\varphi \in C_0(\mathbb{R}^n)$*

$$\lim_{t \to 0} \int_{\mathbb{R}^n} \varphi(x) a_\varepsilon(t, x) dx = A(\varphi) \tag{21}$$

$$\lim_{t \to 0} \int_{\mathbb{R}^n} \varphi(x) b_\varepsilon(t, x) dx = B(\varphi) \tag{22}$$

Remark 1 *Observe that if $\varphi \geq 0$ for $x \in \mathbb{R}^d$, then $(B(\varphi)\theta, \theta) \geq 0$ for all $\theta \in \mathbb{R}^d$. The limit in condition 2) exist for all $\varepsilon > 0$ and do not depend on ε. The vector $A(\varphi)$ and the matrix $B(\varphi)$ are called the drift vector and the diffusion matrix.*

Having developed the general theoretical construct for the processes we will be working with, we may now proceed with the specification of our framework.

3 Approximation of Correlated Markov Processes

In this section we illustrate two new approximation schemes based on tensor algebra for correlated Markov processes where dynamics are expressed by the SDE in (Eq. 8). The novelty we introduce in these approximations involves the characterization of the cross space among dimensions using tensor algebra and in the introduction of new concepts like the copula infinitesimal generator, correlated tensor representation, conditional infinitesimal generator and parametric copula functions mapping. The proposed schemes we develop are based on tensor algebra which we will demonstrate makes them highly amenable to address problems in high dimensional state spaces for correlated processes. As an overview, the approximations we develop involve two aspects:

1. Direct approximation of the infinitesimal generator A with particular emphasis on its mixed derivatives terms;
2. Decomposition of the infinitesimal generator A into orthogonal components.

In what follows we first illustrate the SDE approximation for $d = 1$, in order to establish some useful notation and the building blocks of the approximation schemes. We then present the details of approximations of correlated processes when $d \geq 2$.

3.1 Univariate Diffusion Approximation by CMTC

In order to approximate a one dimensional process, we construct a state space $\mathbf{X} \subset \mathbf{E}$ with $n \in \mathbb{N}$ elements and define the sets of such stencils by

$$\mathbf{X} := \{\underline{x} := -\frac{2^{2n}}{2^n} = x_1, -\frac{2^{2n}}{2^n} + \frac{1}{2^n} := x_2, \ldots, \frac{1}{2^n}, \frac{2}{2^n}, \ldots, x_m = \frac{2^{2n}}{2^n} := \overline{x}\}, \quad (23)$$

with $\mathbf{X}^o := \mathbf{X} \backslash \partial \mathbf{X}$ and $h = \frac{1}{2^n}$ a positive constant that represents the discretization unit, and where the boundary $\partial \mathbf{X}$ consist of the smallest (i.e. $\underline{x}$) and largest (i.e. $\overline{x}$) elements in $\mathbf{X}$, possibly a countably infinite set, and the interior $\mathbf{X}^o$ is the complement of the boundary. We denote by $\pi_n : E \to \mathbf{X}$ the bounded transformation from the continuous state space to the discretized one.

It is possible to construct the continuous time Markov chain $X^{(n)} := \{X_t^n\}_{t \geq 0}$ as the discrete approximation of X on $\mathbf{X}$ by building a matrix $A^{(n)} = \{a(x_i, x_j)\}$ for all $i, j = 1, \ldots, 2^{2n+1} + 1 = m$, that is the discretized counterpart of A in (Eq. 7) and each entry can be calculated by solving the following system of local moment matching equations:

$$\begin{cases} a(x_1, x_2) = a(x_m, x_{m-1}) = 0, \\ a(x_i, x_{i+1}) = \frac{1}{2} \left(\frac{b(x_i)}{h} + \frac{\sigma^2(x_i)}{h^2} \right), \\ a(x_i, x_{i-1}) = \frac{1}{2} \left(\frac{\sigma^2(x_i)}{h^2} - \frac{b(x_i)}{h} \right), \\ a(x_i, x_i) = -(a(x_i, x_{i-1}) + a(x_i, x_{i+1})), \end{cases} \quad (24)$$

for all $i = 2, \ldots, 2^{2n+1}$, with $-\frac{\sigma^2(x_i)}{h} \leq b(x_i) \leq \frac{\sigma^2(x_i)}{h}$. However, the discrete state space $\mathbf{X}$ does not need to be uniform and alternative discretization routines are presented in Tavella and Randall (2000).

Remark 2 *In the following we will denote by $A^{(n)} = A_X^{(n)} := A_X^{(n)}(b(x_i), \sigma(x_i))$, for $i = 1, \ldots, n$, the approximated infinitesimal generator for the Markov process X with local parameters $\mu(\cdot)$ and $\sigma(\cdot)$. In particular $A_{X_k}^{(n)}$ is the approximated infinitesimal generator for the Markov process X_k. Each operator $A_{X_k}^{(n)}$ can be thought of as an element of a Hilbert space H_{X_k}, for all k. Furthermore, given a process X_k we will denote by $\mathbf{X}_k$ the vector corresponding to its discrete state space, in a similar fashion as $\mathbf{X}$ in Eq. 23.*

Assumption 1 *Note that for $x \in \partial \mathbf{X}$, for computational aspects, we impose an absorbing boundary condition. However, it is important to choose the boundary states sufficiently in the extreme of the state space that the laws of the processes $X^{(n)}$ and X are close to each other during the finite time interval of interest in the approximation. This can always be trivially achieved and so is not overly restrictive.*

The resulting matrix $A^{(n)}$ is a tridiagonal matrix in $\mathbb{R}^{m \times m}$, $m = 2^{2n+1} + 1$, with always positive extra-diagonal elements. The previous system calculates the entries of this generator by specifying the first and second instantaneous moments of the process $X^{(n)}$ that have to coincide with those of X on the set $\mathbf{X}^o$. This is equivalent to satisfy the following conditions

$$\mathbb{E}_{X_t}[(X_{t+\Delta t} - X_t)^z] = \mathbb{E}_{X_t}[(X^n_{t+\Delta t} - X^n_t)^z] + o(\Delta t), \qquad z \in \{1, 2\} \text{and} X^{(n)} \in \mathbf{X}^o. \quad (25)$$

Furthermore, one could in principle produce more accurate results by matching higher instantaneous moments of the process and in general there will be a trade off between the number of local moments matched and the coarsity of the grid/stencil h. We also observe that the numerical problem we face is of the same type as in (Eq. 6) and its analytic solution is $U_t = e^{tA}v$ which represents the transient probability of a Markov chain with n states.

3.2 Multivariate Diffusion Approximation by CMTC

We are now in a position to introduce the approximation schemes when $d \geq 2$ for multivariate generalized correlated diffusions as introduced previously. Let $A^{(n_k)}_{X_k} \in \mathbb{R}^{n_k \times n_k}$ denote the approximated infinitesimal generator for the continuous Markov process X_k, with $k = 1, 2, \ldots, d$. This notation is useful when describing d correlated processes and the unique approximated generator for the multidimensional process. Each matrix $A^{(n_k)}_{X_k}$ is tridiagonal and its entries calculated using instantaneous local moment matching as described in (Eq. 127).

Under the local moment matching formulation the representation of the infinitesimal generator for correlated Markov processes given in (Eq. 7) can be rewritten as follows:

$$L_t = \sum_i^d b_i(t, x) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{\substack{i,j \\ i=j}}^d a_{i,j}(t, x) \frac{\partial^2}{\partial x_i \partial x_j} + \frac{1}{2} \sum_{\substack{i,j \\ i \neq j}}^d a_{i,j}(t, x) \frac{\partial^2}{\partial x_i \partial x_j}. \quad (26)$$

Now denote this operator in two components,

$$A_{X_1,\ldots,X_d} := \sum_i^d b_i(t, x) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{\substack{i,j \\ i=j}}^d a_{i,j}(t, x) \frac{\partial^2}{\partial x_i \partial x_j}, \quad (27)$$

and

$$A^{(c)}_{X_1,\ldots,X_d} = \frac{1}{2} \sum_{\substack{i,j \\ i \neq j}}^d a_{i,j}(t, x) \frac{\partial^2}{\partial x_i \partial x_j}, \quad (28)$$

then we can rewrite L_t in (Eq. 26) as the sum of two linear operators

$$L_t = A_{X_1,\ldots,X_d} + A^{(c)}_{X_1,\ldots,X_d}. \quad (29)$$

In particular $A_{X_1,\ldots,X_d}$ is the continuous operator for the independent Markov processes $X_1, \ldots, X_d$, while $A^{(c)}_{X_1,\ldots,X_d}$ is the continuous operator just for the dependence structure of such processes.

Next we develop two ways to approximate (Eq. 26) on a discrete multidimensional space $\bigoplus_{i=1}^n \mathbf{X}_i$, namely:

1. With direct approximation of the operators $A^{(n_1)}_{X_1}, \ldots, A^{(n_d)}_{X_d}$ within the orthogonal dimensions and the operator $A^{(c)(n_1,\ldots,n_d)}_{X_1,\ldots,X_d}$ defined on the cross spaces. This approach

directly approximates the cross derivatives operators and their action over the cross state spaces;

2. The operator is approximated only over the orthogonal spaces. This is possible through the introduction of the notion of conditional operator $A_{X_i|X_j=a}^{(n_i)}$ with $a \in \mathbf{X}_j$ and $i \neq j$, $j = 1, \ldots, d$.

Definition 3 (**Multidimensional approximated generator (independent processes)**) *The multidimensional approximated generator for n independent Markov process $\{X_i(t)\}_{t \geq 0}$, for $i = 1, \ldots, d$ with approximated generators $A_{x_k}^{(n_k)} \in \mathbb{R}^{n_k \times n_k}$, $k = 1, 2, \ldots, n$ is the $(n_1 n_2 \cdots n_d) \times (n_1 n_2 \cdots n_d)$ matrix*

$$L_{X_1, \ldots, X_d}^{(n_1 \cdots n_d)} = A_{x_1}^{(n_1)} \oplus A_{x_2}^{(n_2)} \oplus \cdots \oplus A_{x_d}^{(n_d)}, \tag{30}$$

where $\oplus$ denotes the standard Kronecker sum.

Definition 4 *Let the matrix $I^{(n)} \in \mathbb{R}^{n \times n}$ be the identity matrix of size n, $I_m^{(n)} \in \mathbb{R}^{n \times n}$ a matrix of all zeros and with lower diagonal equal to ones, and $I_p^{(n)} \in \mathbb{R}^{n \times n}$ a matrix of all zeros and with upper diagonal equal to ones, namely:*

$$I_m^{(n)} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} I_p^{(n)} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \tag{31}$$

Proposition 1 (**Joint generator I**) *Let X_1 and X_2 be two Markov processes with properties as described in Section 2 and with associated approximated infinitesimal generators $A_{X_1}^{(n_1)}$ and $A_{X_2}^{(n_2)}$ respectively. It is possible to define a Gaussian factor Z which is a Markov process acting instantaneously on the spaces $(\mathbf{X}_1, \mathbf{X}_2)$ with approximated generator the triadiagonal matrix $A_Z^{(n_1 n_2)}$, such that the infinitesimal approximated generator of the correlated processes (X_1, X_2) with local correlation parameter $\rho := \rho(\mathbf{X}_1, \mathbf{X}_2)$ can be written as*

$$A_{X_1, X_2}^{(n_1 n_2)} = A_{X_1}^{(n_1)} \oplus A_{X_2}^{(n_2)} + A_{X_1, X_2}^{(c)(n_1 n_2)} \tag{32}$$

with

$$\begin{aligned} A_{X_1, X_2}^{(c)(n_1 n_2)} = &- \left(L_Z^{(n_1)} \oplus L_Z^{(n_2)} \right) - diag \left(I^{(n_1)} \otimes diag(L_Z^{(n_2)}) \right) \\ &+ 1_{\{\rho > 0\}} \left(I_m^{(n_2)} \otimes L_Z^{(m)(n_1)} + L_Z^{(p)(n_2)} \otimes I_p^{(n_1)} \right) \\ &+ 1_{\{\rho < 0\}} \left(I_m^{(n_2)} \oplus L_Z^{(p)(n_1)} + L_Z^{(p)(n_2)} \otimes I_p^{(n_1)} \right) \end{aligned} \tag{33}$$

and where $L_Z^{(n_1)}$ and $L_Z^{(n_2)}$ denote the component generator matrices of $A_Z^{(n_1 n_2)}$ that act on X_1 and X_2 respectively. The superscript (p) means that the $L_Z^{(p)(n_1)}$ has upper triangular entries as the matrix $L_Z^{(n_1)}$, while $L_Z^{(m)(n_1)}$ has lower triangular entries as $L_Z^{(n_1)}$. Furthermore (Eq. 32) can be rewritten as a conditional structure given by

$$A_{X_1, X_2}^{(n_1 n_2)} = A_{X_1|Z}^{(n_1)} \oplus A_{X_2|Z}^{(n_2)} + A_Z^{(n_1 n_2)}. \tag{34}$$

Proof First we consider the mixed derivative terms in (Eq. 28) when $d = 2$. For positive correlation the following approximations hold for finite element approximations for the partial derivative operators,

$$
\begin{aligned}
\frac{\partial f}{\partial x_1 \partial x_2} &= \frac{\left(\frac{\partial f}{\partial x_2}\right)_{i+1,j} - \left(\frac{\partial f}{\partial x_2}\right)_{i,j}}{\Delta x_1} \\
&= \frac{f_{i+1,j+1} - f_{i+1,j} - f_{i,j+1} + f_{i,j}}{\Delta x_1 \Delta x_2} + O(\Delta x_1) + O(\Delta x_2), \quad (35)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial f}{\partial x_1 \partial x_2} &= \frac{\left(\frac{\partial f}{\partial x_2}\right)_{i,j-1} - \left(\frac{\partial f}{\partial x_2}\right)_{i-1,j-1}}{\Delta x_1} \\
&= \frac{f_{i,j} - f_{i,j-1} - f_{i-1,j} + f_{i-1,j-1}}{\Delta x_1 \Delta x_2} + O(\Delta x_1) + O(\Delta x_2). \quad (36)
\end{aligned}
$$

Equation (35) can be combined to yield,

$$
\frac{\partial f_{ij}}{\partial x_1 \partial x_2} = \frac{f_{i+1,j+1} - f_{i+1,j} - (f_{i,j-1} - 2f_{i,j} + f_{i,j+1}) - f_{i-1,j} + f_{i-1,j-1}}{2\Delta x_1 \Delta x_2} \\
+ O(\Delta x_1^2) + O(\Delta x_2^2) + O(\Delta x_1 \Delta x_2). \quad (37)
$$

Note the same scheme applies for negative correlation. However (Eq. 37) can be decomposed into the following three terms $T3 - (T1 + T2)$, where

$$
\begin{aligned}
T1 &= \frac{f_{i+1,j} - 2f_{i,j} + f_{i-1,j}}{2\Delta x_1 \Delta x_2}, \\
T2 &= \frac{f_{i,j+1} - 2f_{i,j} + f_{i,j-1}}{2\Delta x_1 \Delta x_2}, \\
T3 &= \frac{f_{i+1,j+1} - 2f_{i,j} + f_{i-1,j-1}}{2\Delta x_1 \Delta x_2}. \quad (38)
\end{aligned}
$$

Next we observe that the operator $A_{X_1,X_2}^{(c),(n_1 n_2)}$ is a 'correlation' operator acting on the joint discretized product space $\mathbf{X}_1 \times \mathbf{X}_2$. The term T1 acts only along the discretized support of X_1, the term T2 acts only along the discretized support of X_2, while T3 acts only along the discretized cross support for both X_1 and X_2. We then use these finite difference operators to calculate the entries of the operators in (Eq. 33). In particular, we use the scheme T1 for $L_Z^{(n_1)}$, the scheme T2 for $L_Z^{(n_2)}$ and T3 for $1_{\{\rho>0\}}\left(I_m^{(n_2)} \otimes L_Z^{(m)(n_1)} + L_Z^{(p)(n_2)} \otimes I_p^{(n_1)}\right)$. The magnitude of the local instantaneous intensities is $\rho(\mathbf{X}_1, \mathbf{X}_2)\sigma\mathbf{X}_1, \sigma\mathbf{X}_2$. We can therefore rewrite (Eq. 32) as

$$
\begin{aligned}
A_{X_1,X_2}^{(n_1 n_2)} = & (A_{X_1}^{(n_1)} - L_Z^{(n_1)}) \oplus (A_{X_2}^{(n_2)} - L_Z^{(n_2)}) \\
& - \Big[\text{diag}(I^{(n_1)} \otimes diag(L_Z^{(n_2)})) \\
& + 1_{\{\rho>0\}}\left(I_m^{(n_2)} \otimes L_Z^{(m)(n_1)} + L_Z^{(p)(n_2)} \otimes I_p^{(n_1)}\right) \\
& + 1_{\{\rho<0\}}\left(I_m^{(n_2)} \oplus L_Z^{(p)(n_1)} + L_Z^{(p)(n_2)} \otimes I_p^{(n_1)}\right)\Big]. \quad (39)
\end{aligned}
$$

If we define the following operators,

$$
\begin{aligned}
A_{X_1|Z}^{(n_1)} &= (A_{X_1}^{(n_1)} - L_Z^{(n_1)}), \\
A_{X_2|Z}^{(n_2)} &= (A_{X_2}^{(n_2)} - L_Z^{(n_2)}), \\
A_Z^{(n_1 n_2)} &= -\operatorname{diag}(I^{(n_1)} \otimes diag(L_Z^{(n_2)})) + 1_{\{\rho>0\}} \left(I_m^{(n_2)} \otimes L_Z^{(m)(n_1)} + L_Z^{(p)(n_2)} \otimes I_p^{(n_1)} \right) \\
&\quad + 1_{\{\rho<0\}} \left(I_m^{(n_2)} \oplus L_Z^{(p)(n_1)} + L_Z^{(p)(n_2)} \otimes I_p^{(n_1)} \right).
\end{aligned}
\tag{40}
$$

This proves (Eq. 34). □

Proposition 2 (Multivariate Uniform Distribution Infinitesimal Operator) *The operator $A_{X_1,X_2}^{(c)(n_1 n_2)}$ is the infinitesimal generator associated to a bivariate correlated distribution with uniform marginals.*

Proof Let's consider the operator $A_{X_1,X_2}^{(c)(n_1 n_2)}$ in (Eq. 33) acting on the joint Hilbert space $H_{X_1 X_2} = H_{X_1} \oplus H_{X_2}$. Without loss of generality we consider the structure of the approximated infinitesimal operator when $\rho > 0$, and we note that the case with $\rho < 0$ is identical. In this case the operator is given by,

$$
\begin{aligned}
A_{X_1,X_2}^{(c)(n_1 n_2)} &= -\left(L_Z^{(n_1)} \oplus L_Z^{(n_2)} \right) - \operatorname{diag}\left(I^{(n_1)} \otimes diag(L_Z^{(n_2)}) \right) \\
&\quad + \left(I_m^{(n_2)} \otimes L_Z^{(m)(n_1)} + L_Z^{(p)(n_2)} \otimes I_p^{(n_1)} \right),
\end{aligned}
\tag{41}
$$

and it is a linear function of the operators $L_Z^{(n_1)}$ and $L_Z^{(n_2)}$ that act on the spaces H_{X_1} and H_{X_2} respectively. □

Theorem 3 (Kronecker Product of Matrix Sequence, $\otimes_S$) *If A is an $\mathbb{R}^{m \times n}$ matrix and $\{B_j\}$, $j = 1, \ldots, m$ is a sequence of m $\mathbb{R}^{p \times q}$ matrices, then the Kronecker product of a matrix sequence $A \otimes_S \{B_j\}$ is the $nm \times pq$ block matrix:*

$$
A \otimes_S \{B_j\} =
\begin{bmatrix}
a_{11} B_1 & \cdots & a_{1n} B_1 \\
\vdots & \ddots & \vdots \\
a_{m1} B_m & \cdots & a_{mn} B_m
\end{bmatrix}
\tag{42}
$$

We have that:

$$
A \otimes_S \{B_j\}^T = \left(A \otimes I^{(p)} \right) \left(I^{(n)} \otimes_S \{B_j\} \right)^T = \left(I^{(n)} \otimes_S \{B_j\} \right) \left(A \otimes I^{(p)} \right)^T
\tag{43}
$$

and also

$$
A \oplus_S \{B_j\}^T = \left(A \otimes I^{(p)} \right) + \left(I^{(n)} \otimes_S \{B_j\} \right)^T = \left(I^{(n)} \otimes_S \{B_j\} \right) + \left(A \otimes I^{(p)} \right)^T
\tag{44}
$$

Proof Results in Eqs. (43) and (44) follow from standard tensor algebra, see the reference (Zhang and Ding 2013). □

Theorem 4 *If A is an $\mathbb{R}^{m \times n}$ matrix and $\{B_j\}$, $j = 1, \ldots, m$ is a sequence of m $\mathbb{R}^{p \times q}$ matrices, then*

$$
\exp\left(A \oplus_S \{B_j\} \right) \approx \exp(A) \otimes_S \exp(\{B_j\}) + O([A, \{B_j\}])
\tag{45}
$$

Proof See (Zhang and Ding 2013). □

For completeness we report the following property of the exponential of the Kronecker sum of two matrices, which is extensively used in all the paper.

Theorem 5 *Let $A \in \mathbb{R}^{m \times m}$ and $B \in \mathbb{R}^{n \times n}$; then*

$$e^{A \oplus B} = e^A \otimes e^B \tag{46}$$

Proof The Kronecker sum of the two matrices is a matrix $C \in \mathbb{R}^{mn \times mn}$ given by:

$$M = A \oplus B = A \otimes I_n + I_m \otimes B \tag{47}$$

The exponential of the matrix M is:

$$e^C = e^{A \oplus B} = e^{(A \otimes I_n + I_m \otimes B)} = e^{A \otimes I_n} e^{I_m \otimes B} \tag{48}$$

Equation (48) is true if $(A \otimes I_n)(I_m \otimes B) = (I_m \otimes B)(A \otimes I_n)$. This a property of the matrix exponential where

$$e^{(A+B)t} = e^{At} e^{Bt} \text{if} AB = BA. \tag{49}$$

We have by theorem 1 of Zhang and Ding (2013) that

$$A \otimes B = (A \otimes I_n)(I_m \otimes B) = (I_m \otimes B)(A \otimes I_n) \tag{50}$$

Therefore the matrices $(A \otimes I_n)$ and $(I_m \otimes B)$ in Eq. (48) always commute. Furthermore by theorem 3 of Zhang and Ding (2013) we have that, if $A, C \in \mathbb{R}^{m \times m}$ and $B, D \in \mathbb{R}^{n \times n}$, then

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD) \tag{51}$$

By theorem 18 of Zhang and Ding (2013) we have that, if $A \in \mathbb{R}^{m \times m}$ and $f(z)$ is analytic and $f(A)$ exists, then

$$f(I_n \otimes A) = I_n \otimes f(A) \text{and} f(A \otimes I_n) = f(A) \otimes I_n \tag{52}$$

By combining Eqs. (47), (48) and (52), we have

$$e^{A \oplus B} = e^{(A \otimes I_n + I_m \otimes B)} = e^{A \otimes I_n} e^{I_m \otimes B} \tag{53}$$

$$= (e^A I_m) \otimes (I_n e^B) = e^A \otimes e^B \tag{54}$$

□

Theorem 6 (**Conditional Infinitesimal Generator**) *Let the infinitesimal generators of generalized diffusions $\{X_i(t)\}_{t \geq 0}, i = 1, 2, \ldots$ be*

$$A_{X_i}(s) = \mu_i(x, s)\frac{\partial}{\partial x} + \frac{1}{2}\sigma_i^2(x, s)\frac{\partial^2}{\partial x^2}, i = 1, 2, \ldots$$

and with general and local properties as described in Section 3.2. Without loss of generality let us drop the time $s \geq 0$ dependency and assume that the diffusions $\{X_i(t)\}_{t \geq 0}, i = 1, 2, \ldots$ are locally correlated with instantaneous local covariation given by

$$\left\langle \sigma_i(x)dW_i(t), \sigma_j(y)dW_j(t) \right\rangle = \sigma_i(x)\sigma_j(x)\rho_{ij}(x, y), \quad \text{for all } i, j = 1, 2, \ldots \tag{55}$$

In order to introduce the concept of conditional infinitesimal generator, without loss of generality, we consider the 2-dimensional operator,

$$L_t = A_{X_1, X_2} + A_{X_1, X_2}^{(c)} \tag{56}$$

being the d-dimensional case just an algebraic extension. We will make all the necessary dimensionality considerations for the d-dimensional representation. Therefore let's consider

the processes pair $X_1 = X$ and $X_2 = Y$. The conditional approximated infinitesimal generator $A_{X|Y}^{(n_1)}$ is defined by the sequence of operator matrices $\left\{ A_{X|Y=y_j}^{(n_1)} \right\} \in \mathbb{R}^{n_1 \times n_1}$, $y_j \in \mathbf{Y}$ each of whose entries are obtained according to local moment matching by:

$$A_{X_1|X_2=y_j}^{(n_1)} = \begin{cases} a(x_1, x_2) = a(x_m, x_{m-1}) = 0, \\[4pt] a(x_i, x_{i+1}) = \frac{1}{2} \left(\frac{\mu_1(x_i) + \rho_{12}(x_i,y_j)\frac{\sigma_1(x_i)}{\sigma_2(y_j)}(y_j - \mu_2(y_j))}{h} + \frac{\sigma_1^2(x_i)(1-\rho_{12}^2(x_i,y_j))}{h^2} \right), \\[4pt] a(x_i, x_{i-1}) = \frac{1}{2} \left(\frac{\sigma_1^2(x_i)(1-\rho_{12}^2(x_i,y_j))}{h^2} - \frac{\mu_1(x_i) + \rho_{12}(x_i,y_j)\frac{\sigma_1(x_i)}{\sigma_2(y_j)}(y_j - \mu_2(y_j))}{h} \right), \\[4pt] a(x_i, x_i) = -(a(x_i, x_{i-1}) + a(x_i, x_{i+1})), \end{cases}$$

(57)

for all $y_j \in \mathbf{Y}$, $x_i \in \mathbf{X}$, with $-\frac{\sigma_k^2(x_i)}{h} \leq \mu_k(x_i) \leq \frac{\sigma_k^2(x_i)}{h}$ for $k = 1, 2$, $-1 \leq \rho_{12}(x_i, y_j) \leq 1$, and $a(x_i, x_j) \geq 0$ for all $i \neq j$.

Proof Let's consider the Markov processes pair (X, Y) and how to derive its corresponding infinitesimal generator approximation in matrix form. The local instantaneous intensities are calculated in the same way as described in Section 3.2 for the one dimensional case, namely using local moment matching as reported in (Eq. 25). For the 2-d case the instantaneous intensities are calculated using the local transition kernel

$$p(X_{t+\Delta t}, Y_{t+\Delta t}|X_t, Y_t), \text{ as} \Delta t \to 0. \tag{58}$$

In particular considering the local states $x_i \in \mathbf{X}$ and $y_j \in \mathbf{Y}$ the transition probability in (Eq. 58) can be rewritten in local form as

$$p(x_{i+1}, y_{j+1}|x_i, y_j) = p(x_{i+1}|y_{j+1}, x_i)p(y_{j+1}|y_j), \tag{59}$$

where

$$\left(Y_{t+} = y_{j+1}|Y_t = y_j \right) \sim p(y_{j+1}|y_j) = N\left(\mu(y_j), \sigma^2(y_j) \right) \tag{60}$$

$$\left(X_{t+} = x_{i+1}|Y_t = y_{j+1}, X_t = x_i \right) \sim p(x_{i+1}|y_{j+1}, x_i)$$
$$= N\left(\mu(x_i) + \frac{\sigma(x_i)}{\sigma(y_j)}\rho(x_i, y_j)(y_j - \mu(y_j)), (1 - \rho^2(x_i, y_j)\sigma^2(x_i)) \right)$$

The calculation of the matrix entries for $A_{X_1|X_2=y_j}^{n_1 n_2}$ is done in identical way as described in Section 3.2, imposing the instantaneous moments of (Eq. 61) for all $y_j \in \mathbf{Y}$ and this completes our proof. Extension to dimensions larger than two is straightforward due to independence of each conditional operator. $\qquad \square$

Remark 3 *We would like to remark that the matrices $\left(A \otimes I^{(p)} \right)$ and $\left(I^{(n)} \otimes_S \{B_j\} \right)$ in Eq. (44) do not commute. However the key result in our proposed multidimensional diffusion approximation is based on theorem 4, namely*

$$\exp\left(A \oplus_S \{B_j\} \right) \approx \exp(A) \otimes_S \exp(\{B_j\}) + O([A, \{B_j\}]) \tag{61}$$

While the left side of Eq. (61) represents the expression of the diffusion approximation that will be used in the study of weak convergence in Section 4, the right end side is instead a further algebraic approximation that is key to a fast computation of the proposed numerical scheme illustrated in Section 5. We have that, while the left side of Eq. (61) can be interpreted as a multidimensional approximation of the transition probability at time 1, the right

end side represents the same multidimensional approximation made by one dimensional approximations. The error term of order $O([A, \{B_j\}])$, is due to the fact that the matrices A and the matrix sequence $\{B_j\}$ do not commute.

Example 1 We saw in Eq. (44) that the expression of the Kronecker sum of a matrix A with a matrix sequence $\{B_j\}$ can be written as:

$$A \oplus_S \{B_j\}^T = \left(A \otimes I^{(p)}\right) + \left(I^{(n)} \otimes_S \{B_j\}\right)^T = \left(I^{(n)} \otimes_S \{B_j\}\right) + \left(A \otimes I^{(p)}\right)^T \quad (62)$$

More specifically, if we consider a coupled diffusion process (X, Y), the expression of the approximated infinitesimal generator reads

$$\begin{aligned} A_X \oplus_S \{A_{Y|X}\} &= \left(A_X \otimes I^{(p)}\right) + \left(I^{(n)} \otimes_S \{A_{Y|X}\}\right) \\ &= \left(A_X \otimes I^{(p)}\right) + \left(I^{(n)} \otimes \{A_Y\}\right) + \left(I^{(n)} \otimes_S \{A_{Y|X}^{(c)}\}\right) \end{aligned} \quad (63)$$

In Eq. (63) we introduce the matrix $\{A_{Y|X}^{(c)}\}$ spanning the X-Y space. Such a matrix is a rotation matrix which corresponds to a specific operator, a copula one, with approximated infinitesimal generator equal to the difference between the joint operator and the independent one. In fact with some algebra we have that:

$$\{A_{Y|X}^{(c)}\} = A_X \oplus_S \{A_{Y|X}\} - (A_X \oplus A_Y) \quad (64)$$

$$= \left(A_X \otimes I^{(p)}\right) + \left(I^{(n)} \otimes_S \{A_{Y|X}\}\right) - \left[\left(A_X \otimes I^{(p)}\right) + \left(I^{(n)} \otimes A_Y\right)\right] \quad (65)$$

$$= \left(I^{(n)} \otimes_S \{A_{Y|X}\}\right) - \left(I^{(n)} \otimes A_Y\right) \quad (66)$$

with generator matrix entries given by:

$$\{A_{Y|X=x_j}^{(c)}\} = \begin{cases} a(y_1, y_2) = a(y_m, y_{m-1}) = 0, \\ a(y_i, y_{i+1}) = \frac{1}{2}\left(\frac{\sigma_1^2(y_i)(1-\rho_{12}^2(x_i,y_j))}{h^2} + \frac{\rho_{12}(x_i,y_j)\frac{\sigma_1(y_i)}{\sigma_2(x_j)}(x_j-\mu_2(x_j))}{h}\right), \\ a(y_i, y_{i-1}) = \frac{1}{2}\left(\frac{\sigma_1^2(y_i)(1-\rho_{12}^2(x_i,y_j))}{h^2} - \frac{\rho_{12}(x_i,y_j)\frac{\sigma_1(y_i)}{\sigma_2(x_j)}(x_j-\mu_2(x_j))}{h}\right), \\ a(y_i, y_i) = -(a(y_i, y_{i-1}) + a(y_i, y_{i+1})), \end{cases} \quad (67)$$

The representation of the copula transition probability in tensor form is:

$$P_y^{(c)} = \alpha \frac{P(X, Y)}{P(X)P(Y)} = \alpha \frac{P(Y|X)}{P(Y)} = \alpha \frac{I^{(n)} \otimes_S e^{\{A_{Y|X}\}}}{I^{(n)} \otimes A_Y} \quad (68)$$

where α is just a scaling factor. Analogously,

$$P_x^{(c)} = \alpha' \frac{P(X, Y)}{P(X)P(Y)} = \alpha' \frac{P(X|Y)}{P(X)} = \alpha' \frac{e^{\{A_{Y|X}\}} \otimes_S I^{(m)}}{A_Y \otimes I^{(m)}} \quad (69)$$

with α' just a scaling factor. Note that $P_y^{(c)} = P_x^{(c)}$ for standardized uniform marginals.

Corollary 1 (Multidimensional approximated generator (correlated processes)) *The multidimensional approximated generator for n independent Markov process $\{X_i(t)\}_{t\geq0}$, for $i = 1, \ldots, d$ with approximated generators $A_{X_k}^{(n_k)} \in \mathbb{R}^{n_k \times n_k}$, $k = 1, 2, \ldots, n$ is the $n_1 n_2 \cdots n_d \times n_1 n_2 \cdots n_d$ matrix*

$$A_{X_1,\ldots,X_d}^{(n_1 \cdots n_d)} = A_{X_1|X_2,\ldots,X_d}^{(n_1)} \oplus_S \cdots \oplus_S A_{x_d}^{(n_d)}, \quad (70)$$

where $\oplus_S$ denotes the Kronecker sum over a matrix sequence.

From Theorem 6 it is clear that we can represent the approximated multidimensional generator of (Eq. 30) as a decomposition of independent conditional generators. Furthermore it is also possible to exploit standard results of conditional probability partitioning in order to facilitate the local characterization of the independent multidimensional conditional generators. Given a multivariate gaussian variable $X \sim N(\mu, \Sigma)$ and consider the partition of X and equivalently of μ and Σ into

$$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{bmatrix} \tag{71}$$

Then $(x_1|x_2) = y = x_1 + Cx_2$, where $C = -\Sigma_{1,2}\Sigma_{2,2}^{-1}$, the conditional distribution of the first partition given the second, is $N(\overline{\mu}, \overline{\Sigma})$, with mean

$$\overline{\mu} = \mu_1 + \Sigma_{1,2}\Sigma_{2,2}^{-1}(x_2 - \mu_2), \tag{72}$$

and covariance matrix

$$\overline{\Sigma} = \Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1}. \tag{73}$$

More generally if we denote by Σ the covariance matrix introduced within the definition of the operator in (Eq. 7), and by $X \sim N(\mu, \Sigma)$ a multivariate normal vector, and fixed $t \geq 0$ the covariance matrix $\Sigma = \{a_{i,j}(t, x)\} \in \mathbb{R}^{d \times d}$ of Eqs. (7), (26) is

$$\Sigma = \begin{pmatrix} \Sigma_{1,1} & \Sigma_{1,2} & \cdots & \Sigma_{1,d} \\ \Sigma_{2,1} & \Sigma_{2,2} & \cdots & \Sigma_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{d,1} & \Sigma_{d,2} & \cdots & \Sigma_{d,d} \end{pmatrix}. \tag{74}$$

Conditional probability partitioning is a very important property when creating the sequence of conditional approximated generators as in (Eq. 57) because large multivariate Gaussian vectors can be easily partitioned as the combination of sets of independent sub-multivariate Gaussian vectors. Each of the sub-multivariate Gaussian vectors can be further characterized and locally approximated through a principal component analysis (PCA). Therefore it is possible to construct a reduced dimensionality infinitesimal generator of a large dimension process without an aggregate PCA of the global process covariance structure.

Example 2 (3-D approximated generator) Given the results in Theorem (6), and applying corollary 1, we show how to calculate $A_{X_1X_2X_3}^{(n_1n_2n_3)}$. In fact we can express the approximated 3-D generator under a conditional decomposition according to

$$A_{X_1X_2X_3}^{(n_1n_2n_3)} = A_{X_1|X_2X_3}^{(n_1)} \oplus_S A_{X_2|X_3}^{(n_2)} \oplus_S A_{X_3}^{(n_3)}. \tag{75}$$

Note that $_{X_1|X_2X_3}^{(n_1)} = \left\{ A_{X_1|X_2=x_2,X_3=x_3}^{(n_1)} \right\} \in \mathbb{R}^{n_1 \times n_1}$ with $x_2 \in X_2$ and $x_3 \in X_3$ and $A_{X_2|X_3}^{(n_2)} = \left\{ A_{X_2|X_3=x_3}^{(n_2)} \right\} \in \mathbb{R}^{n_2 \times n_2}$ with $x_3 \in X_3$.

Remark 4 *It is important to note that the joint infinitesimal generator approximation provided in the main Theorem 6 is not equivalent to the joint representation of Proposition 1. The approximations do not produce same instantaneous local correlations in the cross space, but they produce equivalent 'terminal' joint distribution. We would like to introduce the explicit expressions of the quadratic variation and quadratic covariation among CTMC with generators approximated using the method introduced in Theorem (6). This is formalized in the following Lemmas.*

Lemma 1 (**Orthogonal CTMCs and zero instantaneous covariation**) *Let $\overline{X} = [X_1^{(n_1)},$ $\ldots, X_d^{(n_d)}]'$ be a d-dimensional CTMC and define its partition as $\overline{X} = [x_1, x_2]'$. Let's denote define $y = x_1 + Cx_2$, where $C = -\Sigma_{1,2}\Sigma_{2,2}{}^{-1}$. Then the chains x_2 and the conditional chains y are orthogonal and $var(x_1|x_2) = var(y)$.*

Proof We want to show that the instantaneous covariation of the chains x_2 and y is zero. Therefore we compute

$$
\begin{aligned}
\text{cov}(x_2, y) &= \text{cov}(x_2, x_1) + \text{cov}(x_2, Cx_2) \\
&= \Sigma_{1,2} + C\text{cov}(x_2, x_2) \\
&= \Sigma_{1,2} - \Sigma_{1,2}\Sigma_{2,2}{}^{-1}\Sigma_{2,2} = 0
\end{aligned}
\tag{76}
$$

Furthermore we have

$$
\begin{aligned}
\text{var}(x_1|x_2) &= \text{var}(x_1 + Cx_2) \\
&= \text{var}(x_1) + C\text{var}(x_2)C' + C\text{cov}(x_1, x_2) + \text{cov}(x_2, x_1)C' \\
&= \text{var}(y|x_2) = \text{var}(y).
\end{aligned}
\tag{77}
$$

$\square$

Lemma 2 (**'Terminal' covariance of orthogonal CTMCs**) *Given the settings and results from Proposition 6 and Lemma 1 with some algebra it is straightforward to obtain the covariance matrix Σ. We have*

$$
[y, x_2]\begin{bmatrix} \overline{\Sigma} & 0 \\ 0 & \Sigma_{2,2} \end{bmatrix}\begin{bmatrix} y \\ x_2 \end{bmatrix} = [x_1, x_2]\begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.
\tag{78}
$$

Theorem 7 (**Equivalence of the joint 'terminal' representations**) *Given two correlated Markov processes X_1 and X_2, with properties as described in Section 2, the approximated 'terminal' joint transition probability kernel $P_{X_1,X_2}^{(n_1n_2)}(\mathbf{X}_1, \mathbf{X}_2)$ is the solution of the Cauchy problem of (Eq. 6) where the infinitesimal operator can be expressed in either cross space and marginals decomposition or as conditional decomposition given by*

$$
A_{X_1,X_2}^{(n_1n_2)} = A_{X_1}^{(n_1)} \oplus A_{X_2}^{(n_2)} + A_{X_1,X_2}^{(c)(n_1n_2)},
\tag{79}
$$

or

$$
A_{X_1,X_2}^{(n_1n_2)} = A_{X_1|X_2}^{(n_1)} \oplus_S A_{X_2}^{(n_2)}.
\tag{80}
$$

The approximations in Eqs. (79) and (80) produce equivalent 'terminal' joint transition probabilities.

Proof Refer to Lemma 1 and Lemma 2. $\square$

Theorem 8 (**Correlated kernel tensor representation**) *Given m correlated Markov processes, with properties as described in Section* 2, *the approximated solution of the martingale problem for* (Eq. 1) *is given by the product measure representation involving the generator decomposition according to,*

$$P_{X_1,\ldots,X_m}^{(n_1\cdots n_m)}(t) = e^{tA_{X_1|X_2,\ldots,X_m}^{(n_1)}} \otimes_S e^{tA_{X_2|X_3,\ldots,X_m}^{(n_2)}} \otimes_S \cdots \otimes_S e^{tA_{X_m}^{(n_m)}} \tag{81}$$

Proof In order to calculate the m-dimensional solution of (Eq. 81) we exploit the orthogonality of the conditional approximated infinitesimal operators $A_{X_i|\ldots,X_m}^{(n_i)}$ for all i. This follows from Theorem 6. Then the transition density of (Eq. 81) is computed along same lines as in example 2. □

3.3 Functional Copula Constructions

We are interested in generating joint distributions with a variety of dependence structures and to achieve this purpose we use copula function specifications. In what follows we explore how copulas and families of copulas are generated and introduce a general approach to construct copulas in a tensor product space. In particular we show how copulas are related to other methods of generating joint distributions in a Hilbert space based on specified marginal tensors and decomposition properties of the product space. In doing this we exploit some desirable convergence results of the proposed transformed tensor representation of multivariate correlated processes to the continuous copula functions.

After the definition of a copula function we recall Sklar's theorem, a fundamental result about the relationship between marginals and joint distribution for multivariate correlated random variables. This is important to recall as we will develop a new characterization of Sklar theorem as a function generator representation.

Definition 5 (**m-Copula Function**) *An m-dimensional copula (or m-copula) is a function C from the unit m-cube* $[0, 1]^m$ *to the unit interval* $[0, 1]$ *which satisfies the following conditions:*

1. *$C(1, \ldots, 1, a_n, 1, \ldots, 1) = a_n$ for every $n \leq m$ all a_n in $[0, 1]$;*
2. *$C(a_1, \ldots, a_m) = 0$ if $a_n = 0$ for $a_n \leq a_m$;*
3. *C is m-increasing.*

Remark 5 *Property 1 says that if the realizations of m-1 variables are known each with marginal probability one, then the joint probability of the outcomes is the same as the probability of the remaining uncertain outcome. Property 2 is sometimes referred to as the grounded property of a copula. It says that the joint probability of all outcomes is zero if the marginal probability of any outcome is zero. Property 3 says that the C-volume of any m-dimensional interval is non-negative. Properties 2 and 3 are general properties of multivariate cdfs. It follows that an m-copula can be defined as an m-dimensional cdf whose support is contained in $[0, 1]^m$ and whose one-dimensional margins are uniform on $[0, 1]$. In other words, an m-copula is an m-dimensional distribution function with all m univariate margins being $U(0, 1)$.*

The relationship between distribution functions and copulas is given by the following result, see an overview in Sklar (1996).

Theorem 9 (**Sklar's Theorem**) *Let X and Y be random variables with distribution functions F and G respectively and joint distribution function H. Then there exists a copula C such that for all $(x, y) \in \mathbb{R} \times \mathbb{R}$*

$$H(x, y) = C(F(x), G(y)) \tag{82}$$

If F and G are continuous, then C is unique; otherwise, C is uniquely determined on Ran(F) × Ran(G). Conversely, if C is a copula and F and G are distribution functions, then the function H defined by (82) *is a joint distribution function with margins F and G.*

By Sklar's theorem, given continuous margins F_1 and F_2 and the joint continuous distribution function $F(x_1, x_2) = C(F_1(x_1), F_2(y_2))$, the corresponding copula is generated using the unique inverse transformation

$$C(u_1, u_2) = C(F(x_1), F_2(x_2)) = F(x_1, x_2) = F(F_1^{-1}(u_1), F_2^{-1}(u_2)), \tag{83}$$

where u_1 and u_2 are standard uniform variates. Given the result in theorem 8 it is possible to extend it to the representation of joint distribution function using tensor algebra. Without loss of generality we present the result in two dimensions. Extension to higher dimensional case is straightforward by induction.

Proposition 3 (**Sklar's Theorem in Generator Space**) *Let $P_{X_1, X_2}^{(n_1 n_2)}(\mathbf{X}_1, \mathbf{X}_2)(t), t \geq 0$, be the approximated transition probability kernel solution of the martingale problem in* (Eq. 1) *for the infinitesimal generator*

$$A_s = \sum_i^2 b_i(s, x) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j}^2 \Sigma_{i,j}(s, x) \frac{\partial^2}{\partial x_i \partial x_j}. \tag{84}$$

Then the approximated joint distribution function for the process starting at $(\underline{x}_1, \underline{x}_2)$ is given by

$$F_{X_1, X_2}^{(n_1 n_2)}(\mathbf{X}_1, \mathbf{X}_2 | X_1 = \underline{x}_1, X_2 = \underline{x}_2) = F_{X_1}^{(n_1)}(\mathbf{X}_1 | X_1 = \underline{x}_1) \otimes_S F_{X_2}^{(n_2)}(\mathbf{X}_2 | \mathbf{X}_1, X_2 = \underline{x}_2) \tag{85}$$

Proof Due to the main results of theorem 6 and theorem 81 the following expression holds:

$$P_{X_1 X_2}^{(n_1 n_2)} = P_{X_1}^{(n_1)} \otimes_S P_{X_2 | X_1}^{(n_2)} \tag{86}$$

Then we can calculate:

$$\sum_{i=\underline{x}_1}^{x_1} P^{(n_1)}(x_i^{(1)} | X_1 = \underline{x}_1) \otimes_S \sum_{j=\underline{x}_2}^{x_2} P(x_j^{(2)} | X_2 = \underline{x}_2, X_1 = x_i^{(1)})$$

$$= F_{X_1, X_2}^{(n_1 n_2)}(\mathbf{X}_1, \mathbf{X}_2 | X_1 = \underline{x}_1, X_2 = \underline{x}_2). \tag{87}$$

$\square$

Due to the general validity of our result we can proceed to formulate it in the following theorem.

Theorem 10 (**Joint Distribution Function Convergence**) *Let $X_1(t)$ and $X_2(t)$ be correlated Markov processes with marginal distributions F_{X_1} and F_{X_2} and approximated joint*

distribution function $F_{X_1,X_2}^{(n_1 n_2)}(\mathbf{X}_1, \mathbf{X}_2)$ *as in* (Eq. 85). *Let* $C(u_1, u_2) : [0, 1]^2 \mapsto [0, 1]$ *a continuous copula function. Then the following convergence result holds:*

$$\lim_{n_1, n_2 \to \infty} F_{X_1, X_2}^{(n_1 n_2)}(\mathbf{X}_1, \mathbf{X}_2) = C(F_{X_1}(X_1), F_{X_2}(X_2)) \tag{88}$$

Corollary 2 (**General Copula Mapping**) *Let* $D^{(n_1, n_2, \dots)} := [\frac{i}{n_1}] \times [\frac{j}{n_2}] \times \dots$ *where* $i = 0, \dots, n_1$ *and* $j = 0, \dots, n_2$ *and so on, denote the discretization of the unit hypercube. Let* $C^{n_1, n_2}(u_i, v_j)$ *be a copula distribution function defined for all* $(u_i, v_j) \in D^{(n_1, n_2)}$. *Then according to theorem 10, the following equalities hold:*

$$C^{(n_1, n_2)}(u_i, v_j) = C^{(n_1, n_2)} \left(F_{X_1}^{(n_1)}(x_i^{(1)}), F_{X_2}^{(n_2)}(x_j^{(2)}) \right) = F(x_i^{(1)}, x_j^{(2)}) \tag{89}$$

Furthermore, if we denote by θ *the set of the copula parameters, and by* $\boldsymbol{p}$ *the set of local cross space parameters for the approximated tensor representation, then given* $C^{(n_1, n_2)} \left(F_{X_1}^{(n_1)}(x_i^{(1)}), F_{X_2}^{(n_2)}(x_j^{(2)}); \theta \right)$, *its equivalent representation in a tensor space is the solution of the following minimization problem:*

$$\min_{\boldsymbol{p}} \left\| C^{(n_1, n_2)}(u_i, v_j; \theta) - F_{X_1, X_2}^{(n_1 n_2)}(\mathbf{X}_1, \mathbf{X}_2; \boldsymbol{p}) \right\|_2^2 \tag{90}$$

(Equation 90) means that for a given parametric copula distribution function C belonging to any copula family it is possible to find a set of local parameters $\mathbf{p}$ that would produce a joint distribution function F that has minimal Euclidean distance from C in the tensor space. A practical way to solve (Eq. 90) is to compute the local likelihood with respect to the set of parameters $\mathbf{p} = \{p_{ij}\}$ such that

$$\min_{p_{ij}} \left\{ \log \left(C_{\text{target}}(u_i, u_j; \theta) - F_{X_1, X_2}^{(n_1 n_2)} \left(F_{X_1}^{(n_1)}(x_i^{(1)}), F_{X_2}^{(n_2)}(x_j^{(2)}); p_{ij} \right) \right) \right\}, \text{ for all } i, j. \tag{91}$$

where C_{target} can be any target copula function.

Theorem 11 (**Copula Infinitesimal Operator**) *Let* $c^{(n_1, n_2)}(u_i, v_j)$ *be a copula density function defined for all* $(u_i, v_j) \in D^{(n_1, n_2)}$. *Given the result in theorem 10, we have that*

$$c^{(n_1, n_2)} \left(F_{X_1}^{(n_1)}(x_i^{(1)}), F_{X_2}^{(n_2)}(x_j^{(2)}) \right) = P^{(c)}(x_i^{(1)}, x_j^{(2)}), \text{ for all } i, j. \tag{92}$$

with infinitesimal operator given by (Eq. 33).

Theorem 12 (**Copula Tensor Representation**) *Given the tensor product basis* $Z = B_X \oplus B_Y$ *associated to* $A_{X,Y}^{(n_x, n_y)}$ *as in* (Eq. 79), *where* B_X *and* B_Y *denote the basis of the operators* $A_X^{(n_x)}$ *and* $A_X^{(n_x)}$ *respectively, it is possible to specify a point* $(x_i, y_j) \in (\mathbf{X}, \mathbf{Y})$ *with corresponding associated subspace* $(M \oplus M^\perp) \subset Z$ *of the direct sum of vectors* $x \in M$ *and* $y \in M^\perp$ *with origin* (x_i, y_j), *such that the following relations hold,*

$$z = x \oplus y \tag{93}$$

and

$$z = z' + z'' = (x' \oplus y') + (-a \oplus -a^\perp), \tag{94}$$

where the vector a is the shared component among the vectors x and y or the instantaneous local covariance part of the joint process (X_t, Y_t) and represents the copula. Equations (93) and (94) are represented in Fig. 1.

4 Convergence of the Approximated Generator

In Section 3.2 we defined the Markov chain $X_t^{(n)}$ approximating the multidimensional generalized diffusion X_t. Here we present weak convergence results for $X_t^{(n)}$ to the solution of the SDE for X_t, introduced in (Eq. 1). The convergence is studied from different perspectives: first from a semigroup point of view, secondly in a spectral way through a Fourier unitary transformation of the approximated generator providing desired rate of convergence results, and lastly through the martingale problem for the associated infinitesimal generator.
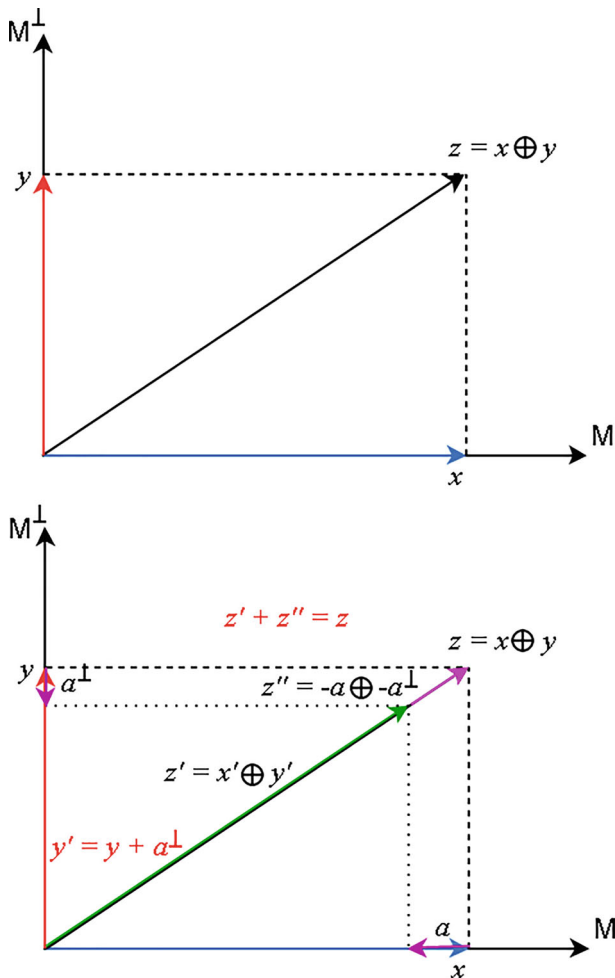
Fig. 1 Representation of $z = z' + z'' = (x' \oplus y') + (-a \oplus -a^\perp)$

It is important to remark that convergence is studied within the multidimensional space, and not onto the projected subspaces introduced in our approximation.

We want that the continuous Markov chain $X^{(n)} := \{X_t^n\}_{t \geq 0}$ has a dynamics as close as possible to the corresponding approximated process $\{X_t\}_{t \geq 0}$. At this purpose we can define an error

$$\epsilon_n(f) := \sup_{x \in \mathbf{H}} \|A^{(n)} f(x) - Af(x)\|. \tag{95}$$

Using the semigroup approach to weak convergence we can state that, if $\epsilon_n(f)$ tends to zero as n tends to infinity for f in $\overline{\mathscr{D}}$, then the sequences of processes $X^{(n)}$, converges weakly to X, in the space $D_\mathbf{E}$. The following theorem states that if the error $\epsilon_n(f)$ goes to zero as n tends to infinity, implying norm convergence of the approximated generator $A^{(n)} f(x)$ to $Af(x)$, this would imply also convergence of the corresponding approximated semigroup $P_t^{(n)}$ to $P_t f$ and of the chain $X^{(n)}$ to X for $t \geq 0$.

Theorem 13 *Let X be a Feller process with state space $\mathbf{E}$ and infinitesimal generator A satisfying the properties and assumptions defined in Section 2, and $X^{(n)}$ be a sequence of Markov chains with generator matrices $A^{(n)}$. Then it holds that*

$$\lim_{n \to \infty} \epsilon_n(f) = 0 \tag{96}$$

for every function f in the core of A, which is equivalent to the statement that

$$P_t^{(n)} \pi_n f \to P_t f, \qquad \text{for all } f \in C_0(\mathbf{E}), t \geq 0.$$

Then the sequences of processes $X^{(n)}$, converges weakly to X, in the space $D_\mathbf{E}$.

Proof From Theorem 6.1 in Chapter 1 pag. 28 of Ethier and Kurtz (1985), we have that if $P_t^{(n)}, n = 1, 2, \ldots$, and P_t are strongly continuous contraction semigroups on $\mathbf{H} \in \mathbf{E}$ with generators $A^{(n)}$ and A respectively, and let $\overline{\mathscr{D}}$ be a core for A, then the following are equivalent:

(i)   For each $f \in \mathbf{E}$

$$P_t^{(n)} \pi_n f \to P_t f$$

uniformly on bounded intervals.

(ii) For each $f \in \mathbf{E}$

$$P_t^{(n)} \pi_n f \to P_t f$$

for all $t \geq 0$.

(iii) For each $f \in \overline{\mathscr{D}}$, there exists $f_n \in D(A^{(n)})$ ($D(A^{(n)})$ is the domain of $A^{(n)}$) for each $n \geq 1$ such that $f_n \to f$ and $A^{(n)} f_n \to Af$.

Furthermore, following (Ethier and Kurtz 1985), Chapter 4 pag. 172 Theorem 2.11, we have that if P_t is a Feller Semigroup on $C_0(\mathbf{E})$ and that for each $t \geq 0$ and $f \in C_0(\mathbf{E})$

$$P_t^{(n)} \pi_n f \to P_t f.$$

If $X_0^{(n)}$ has limiting distribution ν, then there is a Markov Process X corresponding to P_t with initial distribution ν and sample paths in D_p, and

$$X^{(n)} \to X \quad \text{in} D_p.$$

$\square$

Therefore, in order to be able to use the result of theorem 13, we just need to prove the convergence of the proposed approximated generator $A^{(n)}$ and this is done in the following theorem.

Theorem 14 *For all $f \in \overline{\mathscr{D}}$, with $\overline{\mathscr{D}}$ core of the generator A as previously defined,*

$$\lim_{n \to \infty} \sup_{x \in \mathbf{H}} |A^{(n)} f(x) - Af(x)| = 0$$

This mode of convergence is called strong convergence.

Proof Let $f \in \overline{\mathscr{D}}$ and $x \in \mathbf{H}$, and $h := \Delta x \in \mathbb{R}$ the discretization unit. We can locally write the generator of the semigroup $P_t^{(n)}$ as

$$A^{(n)} f(x) = a^{(n)}(x, x + h) (f(x + h) - f(x)) + a^{(n)}(x, x - h) (f(x - h) - f(x))$$

By Taylor approximation

$$f_n(x) \approx f(x + \Delta x) \approx f(x) + f'(x)\Delta x + \frac{1}{2} f''(x)(\Delta x)^2 + o((\Delta x)^2)$$

we obtain for all $i = 1, 2, \ldots, x_i \in \mathbf{H}$,

$$A^{(n)} f_n(x_i) = a(x_i, x_{i+1}) \left(f(x_i) + f'(x_i)\Delta x + \frac{1}{2} f''(x_i)(\Delta x)^2 + o((\Delta x)^2) - f(x_i) \right)$$

$$+ a(x_i, x_{i-1}) \left(f(x_i) - f'(x_i)\Delta x + \frac{1}{2} f''(x_i)(\Delta x)^2 + o((\Delta x)^2) - f(x_i) \right)$$

$$= f'(x_i)\Delta x \, (a(x_i, x_{i+1}) - a(x_i, x_{i-1}))$$

$$+ \frac{1}{2} f''(x_i)(\Delta x)^2 \, (a(x_i, x_{i+1}) - a(x_i, x_{i-1}))$$

$$+ (a(x_i, x_{i+1}) - a(x_i, x_{i-1})) \, o((\Delta x)^2).$$

Due to the fact that $f \in \overline{\mathscr{D}}$, the error term $o((\Delta x)^2)$ is uniform in x. We have that

$$a(x_i, x_{i+1}) + a(x_i, x_{i-1}) = \frac{\sigma^2(x_i)}{h^2},$$

$$a(x_i, x_{i+1}) - a(x_i, x_{i-1}) = \frac{\mu(x_i)}{h}.$$

In Section 3.2 we made precise assumptions, see the assumption 1 on the behaviour at the boundary of the process. Without loss of generality we can assume that the boundary is at a point of infinity, i.e. not attainable in finite time and it is furthermore absorbing. It is possible to express $A^{(n)} f_n(x)$ for all $x \in \mathbf{H}$, as

$$A^{(n)} f_n(x) = f'(x)\Delta x \left(\frac{\mu_i}{\Delta x} \right) + \frac{1}{2} f''(x)(\Delta x)^2 \left(\frac{\sigma^2(x_i)}{(\Delta x)^2} \right) + \left(\frac{\mu(x_i)}{\Delta x} \right) o((\Delta x)^2)$$

$$= Af(x) + \left(\frac{\mu(x_i)}{\Delta x} \right) o((\Delta x)^2).$$

We obtain

$$\sup_{x \in \mathbf{H}} |A_n f(x) - Af(x)| = \sup_{x \in \mathbf{H}} |\frac{\mu(x_i)}{\Delta x} o((\Delta x)^2)| = C_1 o((\Delta x)^2) \xrightarrow{n \to \infty} 0. \quad (97)$$

In case the elements of $\partial \mathbf{H}$, consisting of the smallest (i.e. $\underline{x}$) and largest (i.e. $\overline{x}$) elements in $\mathbf{H}$, are absorbing states of the Markov chain we can continue the above analysis with a further investigation of the weak convergence. Furthermore the behavior on the boundary of the functions $f \in \overline{D}$ is $f'(\underline{x}) = f'(\overline{x}) = 0$. For $x_i = \underline{x}$ we have,

$$
\begin{aligned}
|A^{(n)} f(\underline{x}) - Af(\underline{x})| \leq\ & |A^{(n)} f(\underline{x}) - A^{(n)} f(\underline{x} + h)| \\
& + |A^{(n)} f(\underline{x} + h) - Af(\underline{x} + h)| \\
& + |Af(\underline{x} + h) - Af(\underline{x})|
\end{aligned}
$$

The second term tends to 0 as shown above, the third term by continuity of Af in $\underline{x}$. For the first term we have

$$
\begin{aligned}
|A^{(n)} f(\underline{x}) - A^{(n)} f(\underline{x} + h)| =\ & |a(\underline{x}, \underline{x} + h)(f(\underline{x} + h) - f(\underline{x})) \\
& - a(\underline{x} + h, \underline{x})(f(\underline{x} + 2h) - f(\underline{x} + h)) \\
& - a(\underline{x} + h, \underline{x} + 2h)(f(\underline{x} + h) - f(\underline{x}))| \\
=\ & O(n)o(h) \xrightarrow{n \to \infty} 0.
\end{aligned}
$$

since $a(\underline{x}, \underline{x} + h)$, $a(\underline{x} + h, \underline{x})$, $a(\underline{x} + h, \underline{x} + 2h)$ are on order n and $(f(\underline{x} + h) - f(\underline{x}))$, $(f(\underline{x} + 2h) - f(\underline{x} + h))$, $(f(\underline{x} + h) - f(\underline{x}))$ are of order $o(h)$ because $f'(\underline{x}) = 0$ for $f \in \overline{D}$. The result for $\overline{x}$ follows in the same way. $\qquad\square$

4.1 Correlated Diffusions Approximation Convergence

For this purpose we need to introduce multiplication operators that are considered as an infinite-dimensional generalization of diagonal matrices and they are extremely simple to construct. Furthermore, they appear naturally in the context of the Fourier transform or when one applies the spectral theorem and deals with spectral representation of operators on Hilbert spaces. It is an equivalent way to represent the same operator and can be useful for calculations and further analysis.

We recall that the definitions of discrete Fourier transform matrix and its inverse as a unitary operator are given by,

$$
F_{s,k} = \frac{1}{\sqrt{n}} e^{-i \frac{2\pi}{n} sk},
$$

$$
F_{k,s}^{-1} = \frac{1}{\sqrt{n}} e^{i \frac{2\pi}{n} sk},
$$

$$
F_{l,k} F_{k,j}^{-1} = \delta_{l,j},
$$

where $\delta_{l,j}$ is the Kronecker delta, so these matrices give the resolution of the identity matrix and define a unitary transformation. Also, if $f(x)$ is a function belonging to the space L^2 and B_n be the Brillouin zone defined as $B_n = \left\{ -\frac{\pi}{h} + kh, \quad k = 0, \ldots, \frac{2\pi}{h^2} = \frac{n}{h} \right\}$ the transformation $\mathscr{F}_n : L^2(\mathbf{H}) \mapsto L^2(B_n)$

$$
\mathscr{F}_n(f)(s) = \sum_{x \in \mathbf{H}} F_{s,k} f(x), \tag{98}
$$

is the discrete Fourier transform of f, with $\mathbf{H} := h\mathbb{Z}$ [1] . In fact,

$$\mathscr{F}(f)(s) = \frac{1}{\sqrt{n}} \int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} e^{-\mathrm{i}sx} f(x)dx \tag{99}$$

$$\approx \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} e^{\mathrm{i}s(-\frac{\pi}{h}+kh)} \underbrace{f(-\frac{\pi}{h} + kh)}_{f_k}, \tag{100}$$

where $h = \frac{2\pi}{n}$. We now extend the above transformation to the d-dimensional case, and the following results set the notation for our subsequent theorems and proofs.

Lemma 3 *For any nonnegative-definite symmetric matrix Λ the function*

$$\varphi_{\mathbf{X}} = \exp\left(-\mathrm{i}\mu s - \frac{1}{2}s'\Lambda s\right)t \tag{101}$$

is the characteristic function at time $t > 0$ of the random vector $\mathbf{X}$ with $\mathbb{E}[\mathbf{X}] = \mu$ and $Cov[\mathbf{X}] = \blacksquare$.

Theorem 15 *(cf.* (Lukacs 1958)*, Th. 3.2.3) Let $f(s)$ be an arbitrary characteristic function. For every real x the limit*

$$p(x) = \lim_{T\downarrow 0} \frac{1}{2T} \int_{-T}^{T} e^{-\mathrm{i}sx} f(s)dx \tag{102}$$

exists and is equal to the saltus of the distribution function of $f(s)$ at the point x.

We can obtain a spectral representation of the operator $L_{X_1,\ldots,X_d}^{n_1\cdots n_d}$ by applying the above unitary transformation leading to the following diagonal operator,

$$q_{n_1\cdots n_d}(s) = \mathscr{F}(L_{X_1,\ldots,X_d}^{(n_1\cdots n_d)}(f))\mathscr{F}^{-1}(s,s) \tag{103}$$

where the approximated operator of (Eq. 30) can be written as

$$L_{X_1,\ldots,X_d}^{(n_1\cdots n_d)} = \mu'\nabla + \frac{1}{2}\sigma'\tilde{H}\sigma \tag{104}$$

with ∇ the discrete d-dimensional operator gradient and $\tilde{H}$ the discrete d-dimensional Hessian operator.

In order to derive some converge properties of the operator (30) we do this by comparing the spectral representations of the probability density functions for the continuous infinitesimal generator and its approximated counterpart, namely

$$P_t^{(n_1\cdots n_d)}(x,y) = \frac{1}{(2\pi)^d} \int_{[-\frac{\pi}{h},\frac{\pi}{h}]^d} e^{q_{n_1\cdots n_d}(s)t} e^{\mathrm{i}s(y-x)}ds,$$

and

$$p_t(x,y) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \varphi_{\mathbf{X}} e^{\mathrm{i}s(y-x)}ds.$$

---

[1]$\mathbf{H}$ is possibly unbounded but in all our practical applications we consider $\mathbf{H} \subset D := [-K, K] \subset \mathbb{R}$, $K \in (0, \infty)$, with D the operator domain and assuming for simplicity also periodic boundary conditions.

In this way we are able to assess the order of convergence of the error

$$\epsilon_n := \left| p_t(x, y) - P_t^{(n_1 \cdots n_d)}(x, y) \right|. \tag{105}$$

To assess the rate of convergence of (Eq. 105), we exploit the relationship between the distribution function and its corresponding characteristic function and in particular we refer to the Continuity Theorem.

Theorem 16 (**Continuity Theorem, cf.** (Lukacs 1958), **Th. 3.6.1.**) *Let $\{F_n(x)\}$ be a sequence of distribution functions and denote by $\{f_n(s)\}$ the sequence of the corresponding characteristic functions. The sequence $\{F_n(x)\}$ converges weakly to a distribution function $F(x)$ if, and only if, the sequence $\{f_n(s)\}$ converges for every s to a function $f(s)$ which is continuous at $s = 0$. The limiting function is then the characteristic function of $F(x)$.*

We can therefore focus on the analysis of the passage to the limit $h \to 0$ of the following spectral representation, conditional on a time t,

$$\lim_{h \downarrow 0} \frac{1}{(2\pi)^d} \int_{[-\frac{\pi}{h}, \frac{\pi}{h}]^d} e^{q_{n_1 \cdots n_d}(s)t} e^{is(y-x)} ds = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \varphi_{\mathbf{X}} e^{is(y-x)} ds. \tag{106}$$

Theorem 17 (**Convergence of the d-dimensional approximated operator**) *For all x we consider the sequence of distribution functions $P_t^{(n_1 \cdots n_d)}(x, y)$ and by $\{e^{q_{n_1 \cdots n_d}(s)t}\}$ the sequence of the corresponding characteristic functions. The sequence $\{F_n(x)\}$ converges weakly to a distribution function $F(x)$ if, and only if, the sequence $\{e^{q_{n_1 \cdots n_d}(s)t}\}$ converges for every s to a function $\varphi_{\mathbf{X}}$ which is continuous at $s = 0$. The limiting function is then the characteristic function of $F(x)$.*

Proof We prove convergence and characterization of the rate of convergence for $d = 1$ and $d = 2$ being the proof in higher dimensions just an algebraic extension of the case of $d = 2$. The calculation of $q_{n_1 \cdots n_d}(s)t$ is straightforward and it is just an application of the shift theorem. For $d = 1$

$$\mathscr{F}_n(A_{X_1}^{(n_1)})(f)(s) = \sum_{x \in \mathbf{H}} F_{s,k} \left(\mu \nabla_{h_1}(f)(x) + \frac{\sigma^2}{2} \Delta_{h_1}(f)(x) \right)$$

$$\mathscr{F}_n(\mu \nabla_{h_1})(f)(s) = \sum_{x \in \mathbf{H}} F_{s,k} \mu \nabla_{h_1}(f)(x)$$

$$= \mu \sum_{k=0}^{n-1} F_{s,k} \frac{f(kh + h_1) - f(kh - h_1)}{2h_1}$$

$$= \frac{\mu}{2h_1} \left(e^{ih_1 s} \sum_{k=0}^{n-1} F_{s,k} f(kh) - e^{-ih_1 s} \sum_{k=0}^{n-1} F_{s,k} f(kh) \right)$$

$$= \mu \frac{e^{ih_1 s} - e^{-ih_1 s}}{2h_1} \mathscr{F}(f)(s) = -i\mu \frac{\sin h_1 s}{h_1} \mathscr{F}(f)(s).$$

Doing a similar calculation for $\mathscr{F}(\frac{\sigma^2}{2} \Delta_{h_1})(f)(s)$ we obtain

$$q_{n_1}(s) = \left(-i\mu \frac{\sin h_1 s}{h_1} + \sigma^2 \frac{\cos(h_1 s) - 1}{h_1^2} \right)(s). \tag{107}$$

We have

$$P_t^{(n_1)}(x, y) = p^{(n_1)}(y \leq X_t \leq y + h_1 | X_0 = x)$$
$$= \frac{1}{n_1} \sum_{s \in B_n} e^{q_{n_1}(s)t} e^{is(y-x)}.$$

Without loss of generality, this result is a particular case of the continuity theorem (16), and the convergence error of (Eq. 105) is measured in correspondence of saltus point of the distribution, see Th.(15). Lets consider the integral on the left end side of (Eq. 106) for $d = 1$ and it can be rewritten as

$$\frac{1}{2\pi} \int_{-\frac{\pi}{h_1}}^{\frac{\pi}{h_1}} e^{\left(-i\mu \frac{\sin h_1 s}{h_1} + \sigma^2 \frac{\cos(h_1 s)-1}{h_1^2}\right)(s)t} e^{is(y-x)} ds$$

$$= \frac{1}{2\pi} \int_{-\frac{\pi}{h_1}}^{-K} e^{\left(-i\mu \frac{\sin h_1 s}{h_1} + \sigma^2 \frac{\cos(h_1 s)-1}{h_1^2}\right)(s)t} e^{is(y-x)} ds$$

$$+ \frac{1}{2\pi} \int_{-K}^{K} e^{\left(-i\mu \frac{\sin h_1 s}{h_1} + \sigma^2 \frac{\cos(h_1 s)-1}{h_1^2}\right)(s)t} e^{is(y-x)} ds$$

$$+ \frac{1}{2\pi} \int_{K}^{\frac{\pi}{h_1}} e^{\left(-i\mu \frac{\sin h_1 s}{h_1} + \sigma^2 \frac{\cos(h_1 s)-1}{h_1^2}\right)(s)t} e^{is(y-x)} ds$$

and it is possible to make the first and the third integral on the right end side arbitrary small by choosing a large number $K > 0$ and by selecting $h_1 > 0$ sufficiently small. If we consider the second integral we can analyze the behaviour as $h_1 \to 0$. We notice that the function

$$\lim_{h_1 \to 0} \left(i\mu \frac{\sin h_1 s}{h_1}\right) = \lim_{h_1 \to 0} i\mu \frac{s}{s} \frac{\sin h_1 s}{h_1}$$
$$= \lim_{h_1 \to 0} is\mu \frac{\sin h_1 s}{h_1 s} = is\mu$$

and

$$\lim_{h_1 \to 0} \left(\sigma^2 \frac{\cos h_1 s - 1}{h_1^2}\right) = \lim_{h_1 \to 0} \sigma^2 \frac{s^2}{s^2} \frac{\cos h_1 s - 1}{h_1^2}$$
$$= \lim_{h_1 \to 0} \sigma^2 s^2 \frac{\cos h_1 s - 1}{(h_1 s)^2} = -\frac{1}{2}\sigma^2 s^2.$$

We would like to examine in more details the order of convergence of the above functions as $h_1 \to 0$. For the limit

$$\lim_{h_1 \to 0} \frac{\sin h_1}{h_1} = 1,$$

using $\sin h_1 = h_1 - \frac{h_1^3}{6} + \ldots$, we get

$$\frac{\sin h_1}{h_1} - 1 = \frac{\sin h_1 - h_1}{h_1} = -\frac{h_1^3}{6h} + \ldots = -\frac{h_1^2}{6} + \ldots$$

we find an order of $O(h_1^2)$. In the same way using $\cos h_1 = 1 - \frac{h_1^2}{2} + \ldots$, we can assess the order of convergence of the limit

$$\lim_{h_1 \to 0} \frac{1 - \cos h_1}{h_1^2} = \frac{1}{2},$$

$$\frac{1 - \cos h}{h_1^2} = \frac{1 - \left(1 - \frac{h_1^2}{2} + \ldots\right)}{h_1^2} = \frac{1}{2} + \ldots$$

and convergence order of $O(1)$. Therefore the order of convergence is at most $O(h_1^2)$. This results can be extended to all the marginals of a d-dimensional approximated operator in case of independent marginals. In the presence of correlation we have the presence of mixed derivative terms. For $d = 2$ the calculation of $q_{n_1 n_2}$ is as follows:

$$\mathscr{F}_n\left(A_{X_1, X_2}^{(n_1 n_2)}\right)(f)(s)$$
$$= \mathscr{F}_n\left(A_{X_1}^{(n_1)} \oplus A_{X_2}^{(n_2)} + A_{X_1, X_2}^{(c)(n_1 n_2)}\right)(f)(s)$$
$$= \sum_{x_1 \in \mathbf{X}_1} \sum_{x_2 \in \mathbf{X}_2} F_{s,k}\left(\left(\mu_1 \nabla_{h_1} + \frac{\sigma_2^2}{2}\Delta_{h_1} + \mu_2 \nabla_{h_2} + \frac{\sigma_2^2}{2}\Delta_{h_2} + \rho\sigma_1\sigma_2\nabla_{h_1}\nabla_{h_2}\right)(f)(x)\right).$$

(Eq. 108) is equivalent to

$$\mathscr{F}_n\left(A_{X_1}^{(n_1)} \oplus A_{X_2}^{(n_2)} + A_{X_1, X_2}^{(c)(n_1 n_2)}\right)(f)(s)$$
$$= \mathscr{F}_n\left(A_{X_1}^{(n_1)}\right)(f)(s) \oplus \mathscr{F}_n\left(A_{X_2}^{(n_2)}\right)(f)(s) + \mathscr{F}_n\left(A_{X_1, X_2}^{(c)(n_1 n_2)}\right)(f)(s).$$

Therefore it is sufficient to analyze the term

$$\mathscr{F}_n\left(A_{X_1, X_2}^{(c)(n_1 n_2)}\right)(f)(s), \tag{108}$$

where the operator $A_{X_1, X_2}^{(c)(n_1 n_2)}$ was introduced in proposition 1 as the correlation operator.

$$\mathscr{F}_n\left(A_{X_1, X_2}^{(c)(n_1 n_2)}\right)(f)(s)$$
$$= \sum_{x_1 \in \mathbf{X}_1} \sum_{x_2 \in \mathbf{X}_2} F_{s,k}\left(\rho\sigma_1\sigma_2\nabla_{h_1}\nabla_{h_2}\right)(f)(x))$$
$$= \rho_{12}\sigma_1\sigma_2 \frac{1}{h_1 h_2}\left(\cos(h_1 s_1 + h_2 s_2) - \cos(h_1 s_1) - \cos(h_2 s_2) + 1\right),$$

because of the mixed derivative term approximation of proposition 1. The following approximations hold:

$$\cos(h_1 s_1 + h_2 s_2) - 1 = -\frac{h_2^2 s_2^2}{2} - s_1 s_2 h_1 h_2 - \frac{h_1^2 s_1^2}{2} + \frac{h_1^2 s_1^2 h_2^2 s_2^2}{4}$$
$$+ (1 + h_1 + h_1^2)O(h_2)^3 + O(h)^3,$$

$$\cos(h_1 s_1) - 1 = -\frac{s_1^2 h_1^2}{2} + O(h_1)^3,$$

$$\cos(h_2 s_2) - 1 = -\frac{s_2^2 h_2^2}{2} + O(h_2)^3.$$

Therefore,

$$\lim_{h_1,h_2\to 0} \rho_{12}\sigma_1\sigma_2 \frac{1}{h_1 h_2}\left(\cos(h_1 s_1 + h_2 s_2) - \cos(h_1 s_1) - \cos(h_2 s_2) + 1\right) = -\rho_{12}\sigma_1\sigma_2 s_1 s_2,$$

(109)

that is exactly the characteristic function of the covariance term for the bivariate normal distribution at time t. Given this proof, extension to higher dimensions is algebraically straightforward. □

4.2 Weak Convergence of the Approximating Chain Using the Martingale Central Limit Theorem

Here we present a weak convergence result for the multivariate diffusion approximation introduced in Section 3.2, along the lines of the general diffusion convergence Theorem 4.1 at pag. 354 of Ethier and Kurtz (1985). The convergence result we propose is based on the arguments belonging to the formulation of diffusion theory in terms of the martingale problem, see (Stroock and Srinivasa Varadhan 1997), which requires minimal assumptions about the smoothness of the coefficients of the SDE and can be seen as an extension of the martingale central limit theorem [Martingale CLT, Th. 1.4, (Ethier and Kurtz 1985), pag. 339]. The results are mainly obtained by compactness arguments which do not require a priori regularity. These arguments are the same as those devised to provide an existence theory for the multidimensional SDE under consideration, and they just refer to properties of the SDE coefficients, that are the prerequisites for our analysis as per Section 2.

Theorem 18 *Let $\Sigma = ((\Sigma_{ij}))$ be a continuous, symmetric, nonnegative definite, $d \times d$ valued function on $\mathbb{R}^d$ and let $b : \mathbb{R}^d \to \mathbb{R}^d$ be continuous. Let*

$$A = \left\{ \left(f, Gf = \frac{1}{2}\sum \Sigma_{ij}\partial_i\partial_j f + \sum b_i \partial_i f \right) : f \in C_c^\infty(\mathbb{R}^d) \right\}$$

(110)

and suppose that the $C_{\mathbb{R}^d}[0,\infty)$ martingale problem for A is well posed. For $n = 1, 2, \ldots$ let X_n and B_n be processes with sample paths in $D_{\mathbb{R}^d}[0,\infty)$ and let $A_n = ((A_n^{ij}))$ be a symmetric $d \times d$ matrix-valued process such that A_n^{ij} has sample paths in $D_R[0,\infty)$ and $A_n(t) - A_n(s)$ is non negative definite for $t > s \geq 0$. Set $\mathscr{F}_t^n = \sigma(X_n(s), B_n(s), A_n(s) : s \leq t)$. Let $\tau_n^r = \inf\{t : |X_n(t)| \geq r \text{ or } |X_n(t-)| \geq r\}$, and suppose that

$$M_n = X_n - B_n$$

(111)

and

$$M_n^i M_n^j - A_n^{ij}$$

(112)

are $\mathscr{F}_t^n$-local martingales, and for each $r > 0$, $T > 0$, and $i, j = 1, 2, \ldots, d$

$$\lim_{n\to\infty} \mathbb{E}\left[\sup_{t \leq T \wedge \tau_n^r} \left| X_n(t) - X_n(t-) \right|^2 \right] = 0,$$

(113)

$$\lim_{n\to\infty} \mathbb{E}\left[\sup_{t \leq T \wedge \tau_n^r} \left| B_n(t) - B_n(t-) \right|^2 \right] = 0,$$

(114)

$$\lim_{n\to\infty} \mathbb{E}\left[\sup_{t \leq T \wedge \tau_n^r} \left| A_n^{ij}(t) - A_n^{ij}(t-) \right|^2 \right] = 0,$$

(115)

$$\sup_{t \leq T \wedge \theta_n^r} \left| B_n^i(t) - \int_0^t b_i(X_n(s))ds \right| \xrightarrow{p} 0$$

(116)

and

$$\sup_{t \leq T \wedge \theta_n^r} \left| A_n^{ij}(t) - \int_0^t \Sigma_{ij}(X_n(s))ds \right| \xrightarrow{p} 0. \tag{117}$$

Suppose that $P(X^{(n)}(0)^{-1}) \Rightarrow \nu$, with ν the starting distribution. Then $X^{(n)} \Rightarrow X$, with X the solution of the martingale problem for (A, ν).

Proof Due to the fact that the martingale problem for A is well posed, the process $M(t) = X(t) - B(t)$ is a martingale. By the optional stopping theorem, see (Rogers and Williams 2000), if τ is a stopping time also $M(\tau)$ is a martingale. In particular this is valid also for the stopping time $\tau_n^r = \inf\{t : |X_n(t)| \geq r \text{ or } |X_n(t-)| \geq r\}$. By (Eq. 111) the process $M_n(t) = X_n(t) - B_n(t)$ is a martingale. Relatively compactness properties of $M_n(t)$ implies relatively compactness of $X_n(t)$ and $B_n(t)$ and therefore for $X_n(t \wedge \tau_n)$ and $B_n(t \wedge \tau_n)$ as in Eqs. (113) and (114) respectively. This extends also to the stopped martingale $M_n(t \wedge \tau_n \wedge \tau_a)$ where $\tau_a = \inf\{t : A_n^{ii}(t) > t \sup_{|x| \leq r} a_{ii}(x) + 1, \text{ for some } i\}$. Furthermore relative compactness in a set C is a condition equivalent to the condition that each sequence in C contains a convergent subsequence, see for example (Billingsley 1999). This means that every subsequence $X_{n_k}(t \wedge \tau_n^r) \Rightarrow X^{\bar{r}}(t \wedge \tau^r)$, where $\tau^r = \inf\{t : |X^{\bar{r}}(t)| \geq r\}$, for all $\bar{r} \geq r > 0$ Therefore the stopped processes

$$M^{\bar{r}}(t \wedge \tau^r) = X^{\bar{r}}(t \wedge \tau^r) - \int_0^{t \wedge \tau^r} b(X^{\bar{r}}(s))ds \tag{118}$$

$$M_i^{\bar{r}}(t \wedge \tau^r) M_j^{\bar{r}}(t \wedge \tau^r) - \int_0^{t \wedge \tau^r} a_{ij}(X^{\bar{r}}(s))ds \tag{119}$$

are martingales and by Ito's lemma

$$f(X^{\bar{r}}(t \wedge \tau^r)) - \int_0^{t \wedge \tau^r} Af(X^{\bar{r}}(s))ds \tag{120}$$

is a martingale for each $f \in C_c^\infty(\mathbb{R}^d)$, and $Af(X^{\bar{r}}(s))$ is the approximated infinitesimal generator applied to the function f. In particular if the martingale problem is well posed uniqueness argument for the solution hold and hold also for the solution for the stopped problem, hence

$$X_n(t \wedge \tau^r) \Rightarrow X(t \wedge \tau^r) \tag{121}$$

for all r. Also $r \to \infty$ implies $\tau^r \to \infty$. Therefore $X_n \to X$. $\square$

5 Application to the Solution of Multi-Dimensional SDEs

In this section we are going to apply the results developed in the previous sections to calculate the approximate solution of a multi-dimensional SDE, as per Eq. 1. To achieve this in a manner that can then be trivially generalized to more complex higher dimensional settings it suffices to illustrate the process on a simple representative toy example.

Let us consider the following pair of stochastic differential equations,

$$dX_t = (\theta - X_t)dt + \sigma_1 dW_t^{(1)} \tag{122}$$

$$dY_t = \mu dt + \sigma_2 dW_t^{(2)} \tag{123}$$

where $\theta \geq 0$, $\sigma_1 > 0$, $\mu \geq 0$, $\sigma_2 > 0$ are real-valued parameters, the initial conditions $X_0 = x_0$, $Y_0 = y_0$ hold, and $W^{(1)}$ and $W^{(2)}$ are correlated Wiener processes with correlation coefficient ρ, i.e. $\mathbb{E}[W_t^{(1)} W_t^{(2)}] = \rho t$.

In this case of time-hogeneous correlated diffusion processes of Eq. 122 and function $f = f(x, y)$ that are twice differentiable with compact support, the Markov generator A is given by

$$Af = (\theta - x)\frac{\partial f}{\partial x} + \sigma_1^2 \frac{\partial^2 f}{\partial x^2} + \mu \frac{\partial f}{\partial y} + \sigma_2^2 \frac{\partial^2 f}{\partial y^2} + \rho \frac{\partial^2 f}{\partial x \partial y} \tag{124}$$

We are interested in calculating the transition probability p that satisfies the partial differential equation (FKE)

$$\frac{\partial p}{\partial t} + Ap = 0 \tag{125}$$

with starting conditions $p(z, t_0; z', t_0) = \delta(z - z')$, with $z = (x, y)$, where δ is the Dirac delta function and the operator A acts on the coordinates z.

In order to approximate the multi-dimensional diffusion process $Z_t = (X_t, Y_t)$ with a Markov chain $Z^{(n)}$, we follow the algorithm below based on the theory and results of the previous sections.

Algorithm: Computational Implementation of Mimicking CTMC Process

STEP 1   Define the state-space for the CTMC $Z^{(n)}$.
This implies that we need to define a discrete state space for each dimension. A grid $X := x_1, \ldots, x_{n_x} \in \mathbb{R}$ for the chain $X^{(n_x)}$ with $n_x \in \mathbb{N}$ elements, such that $x_i < x_j$ for any integers $0 \leq i < j \leq n_x$. Likewise, we generate a grid $Y := y_1, \ldots, y_{n_y} \in \mathbb{R}$ for the chain $Y^{(n_y)}$ with $n_y \in \mathbb{N}$ elements.

STEP 2   Approximate the infinitesimal generator A of Eq. (124) as follows:

$$A \approx A^{n_x n_y} = A^{n_y} \oplus_S \{A_{X|Y}^{n_x}\}^T \tag{126}$$

where:

2.1   The entries of the generator matrix A^{n_y} are computed by locally solving the system of Eq. (127) for the second equation in Eq. (122)

$$\begin{cases} a(y_1, y_2) = a(y_{n_y}, y_{n_y-1}) = 0, \\ a(y_i, y_{i+1}) = \frac{1}{2}\left(\frac{\mu}{h_y} + \frac{\sigma_2^2}{h_y^2}\right), \\ a(y_i, y_{i-1}) = \frac{1}{2}\left(\frac{\sigma_2^2}{h_y^2} - \frac{\mu}{h_y}\right), \\ a(y_i, y_i) = -(a(y_i, y_{i-1}) + a(y_i, y_{i+1})), \end{cases} \tag{127}$$

with $h_y = \frac{y_{n_y} - y_1}{n_y}$ and $-\frac{\sigma_2^2}{h} \leq \mu \leq \frac{\sigma_2^2}{h}$. However, the discrete state space $\mathbf{Y}$ does not need to be uniform.

2.2   Each conditional infinitesimal generator matrix of the sequence $\{A_{X|Y}^{n_x}\}$ can be computed by locally solving the system of Eq. (57) The conditional approximated infinitesimal generator $A_{X|Y}^{(n_1)}$ is defined by the sequence of operator matrices $\left\{A_{X|Y=y_j}^{(n_1)}\right\} \in \mathbb{R}^{n_1 \times n_1}$,

$y_j \in \mathbf{Y}$ each of whose entries are obtained according to local moment matching by:

$$A_{X_1|X_2=y_j}^{(n_1)} = \begin{cases} a(x_1, x_2) = a(x_m, x_{m-1}) = 0, \\[4pt] a(x_i, x_{i+1}) = \frac{1}{2} \left(\frac{\mu_1(x_i)+\rho_{12}(x_i,y_j)\frac{\sigma_1(x_i)}{\sigma_2(y_j)}(y_j-\mu_2(y_j))}{h} + \frac{\sigma_1^2(x_i)(1-\rho_{12}^2(x_i,y_j))}{h^2} \right), \\[4pt] a(x_i, x_{i-1}) = \frac{1}{2} \left(\frac{\sigma_1^2(x_i)(1-\rho_{12}^2(x_i,y_j))}{h^2} - \frac{\mu_1(x_i)+\rho_{12}(x_i,y_j)\frac{\sigma_1(x_i)}{\sigma_2(y_j)}(y_j-\mu_2(y_j))}{h} \right), \\[4pt] a(x_i, x_i) = -(a(x_i, x_{i-1}) + a(x_i, x_{i+1})), \end{cases} \tag{128}$$

for all $y_j \in \mathbf{Y}$, $x_i \in \mathbf{X}$, with $-\frac{\sigma_k^2(x_i)}{h} \leq \mu_k(x_i) \leq \frac{\sigma_k^2(x_i)}{h}$ for $k = 1, 2, -1 \leq \rho_{12}(x_i, y_j) \leq 1$, and $a(x_i, x_j) \geq 0$ for all $i \neq j$.

STEP 3   Compute the transition probability using the generator matrix of Eq. (126) via the discretization as follows:

3.1   Compute the discretized form exactly as follows (exact in asymptotic weak convergence sense)

$$P_t = e^{tA} \approx e^{tA^{n_x n_y}} = e^{tA^{n_y} \oplus_S \{A_{X|Y}^{n_x}\}^T} \tag{129}$$

3.2   Alternatively, make the second stage approximation for a the most efficient approximation using the following additional rank 1 approximation. Where we can approximate the computation of the transtion probability of Eq. (129) further

$$P_t \approx e^{tA^{n_y}} \otimes_S e^{t\{A_{X|Y}^{n_x}\}^T} \tag{130}$$

Remark 6 *Although the transition probability obtained from* Eq. (130) *is approximated, its computation is faster than the one calculated using* Eq. (129)
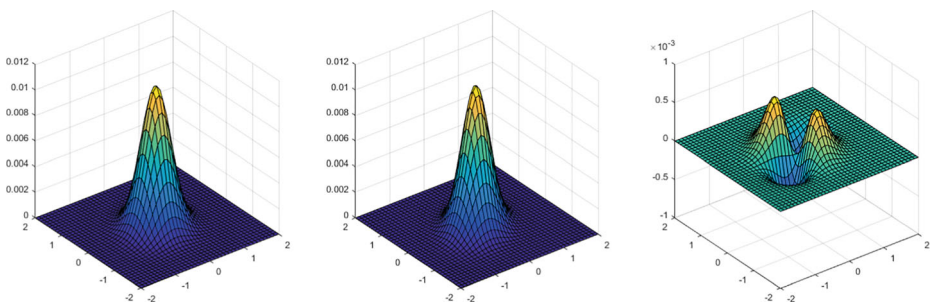
Fig. 2 Approximated transition probabilities computed using the above algorithms with the suggested diffusion parameters with time $t = 1$. On the left hand side the transition probability corresponding to $e^{tA^{n_y} \oplus_S \{A_{X|Y}^{n_x}\}^T}$. In the middle we plot the transition probability corresponding to $e^{tA^{n_y}} \otimes_S e^{t\{A_{X|Y}^{n_x}\}^T}$, therefore simplifying the matrix exponential problem from 2-D into two 1-D problems, and such approximation is numerically faster than the computation of the exponential of the corresponding 2-D matrix. On the right end side we plot the difference of the two obtained transition probability kernels

because the problem of matrix exponentiation is one dimentional instead of bidimensional.

5.1 2D SDE Case Study and Analysis of Results

We compute the joint transition densitiy for the coupled process $(X_1(t), X_2(t))$ using the following parameters for the first marginal process $(\mu_1, \sigma_1) = (0.02, 0.4)$, and these parameters $(\mu_2, \sigma_2) = (0.03, 0.4)$ for the second process, with $\rho_{1,2} = 0.6$, $t = 1$, and $(X_1(0), X_2(0)) = (0, 0)$. Computational results are shown above in (Fig. 2).

We propose in (Fig. 3) the MATLAB code that closely follow the steps of the above algorithm:

```matlab
1   clear all
2   close all
3   T = 1;
4   mu1 = .02; v1 = .4; % marginal 1 drift and vol
5   mu2 = 0.03; v2 = .4; % marginal 2 drift and vol
6   n1 = 40; S1 = linspace(-2,2,n1); % discretization state space 1
7   n2 = 40; S2 = linspace(-2,2,n2); % discretization state space 2
8
9   corrI = -.6;
10
11  corrMtx = ones(n1,n2)*corrI; % correlation
12  h1 = S1(3) - S1(2);
13  h2 = S2(3) - S2(2);
14  G1 = zeros(n1);
15
16  % identity matrices
17  I1 = eye(n1); I2 = eye(n2);
18
19  I2_G1 = zeros(n1*n2);
20
21  P1cP2 = zeros(n1,n1,n2);
22  for j = 1:n2
23  for i = 2:n1-1
24  mcc = mu1 + corrMtx(i,j)*(v1/v2)*(S2(j) - mu2);
25  cv = sqrt((1-corrMtx(i,j)*corrMtx(i,j))*v1*v1);
26  G1(i,i+1) = (cv^2)/(2*h1^2) + mcc/(2*h1);
27  G1(i,i-1) = (cv^2)/(2*h1^2) - mcc/(2*h1);
28  G1(i,i) = -(G1(i,i+1) + G1(i,i-1));
29  end
30  I2_G1((j-1)*n1 + 1:(j-1)*n1 + n1,(j-1)*n1 + 1:(j-1)*n1 + n1) = G1;
31  P1cP2(:,:,j) = expm(sparse(G1)*T);
32  end
33  G2 = zeros(n2);
34  for i = 2:n2-1
35  G2(i,i+1) = (v2^2)/(2*h2^2) + mu2/(2*h2);
36  G2(i,i-1) = (v2^2)/(2*h2^2) - mu2/(2*h2);
37  G2(i,i) = -(G2(i,i+1) + G2(i,i-1));
38  end
39  P2 = expm(G2*T);
40  G2_I1 = kron(G2, I1);
41
42  G = G2_I1 + I2_G1;
43  P = expm(sparse(G));
44  P_split = expm(sparse(G2_I1))*expm(sparse(I2_G1));
45
46  %initial conditions
47  X1_0 = n1/2;
48  X2_0 = n2/2;
49  xx = n1*(X2_0 - 1) + X1_0 ;
50
51  Pm = reshape(P(xx,:),n1,n2);
52  Ps2 = reshape(P_split(xx,:),n1,n2);
53
54  A = P1cP2(X1_0,:,:);
55  B = P2(X2_0,:);
56
57  Ps = repmat(B,n1,1).*reshape(A,n1,n2);
58  [X1, X2] = meshgrid(S1,S2);
59  figure
60  surf(X1,X2,Ps2');
```

Fig. 3 Computation of the approximated transition density for two coupled Diffusions by tensor approximation

6 Conclusion

In this paper we investigated aspects of semimartingale decompositions, approximation and the martingale representation for multidimensional correlated Markov processes. We formulated a new interpretation of the dependence among processes using the martingale approach. We showed that it is possible to represent, in both continuous and discrete space, that a multidimensional correlated generalized diffusion is a linear combination of processes that originate from the decomposition of the starting multidimensional semimartingale. This result not only reconciles with the existing theory of diffusion approximations and decompositions, but defines the general representation of infinitesimal generators for both multidimensional generalized diffusions and, as we demonstrated, also for the specification of copula density dependence structures. This new result provides immediate representation of the approximate solution for correlated stochastic differential equations. We showed desirable convergence results for the proposed multidimensional semimartingales decomposition approximations.

References

Billingsley P (1999) Convergence of Probability Measures, 2nd Edn. Wiley Series in Probability and Statistics

Brunick G, Shreve S (2013) Mimicking an Ito process by a solution of a stochastic differential equation. Ann Appl Probab 23(4):1584–1628

Brunick G (2013) Uniqueness in law for a class of degenerate diffusions with continuous covariance. Probab Theory Related Fields 155:265–302. MR3010399

Ekström E, Hobson D, Janson S, Tysk J (2013) Can time-homogeneous diffusions produce any distribution? Probab Theory Related Fields 155:493–520. MR3034785

Ethier ST, Kurtz TG (1985) Markov Processes, Characterization and Convergence. Wiley

Gyöngy I (1986) Mimicking the one-dimensional marginal distributions of processes having an ito differential. Probab Theory Relat Fields 71(4):501-516

Hackbusch W (2012) Tensor Spaces and Numerical Tensor Calculus. Springer Series in Computational Mathematics

Karatzas I, Shreve S (2000) Brownian Motion and Stochastic Calculus - 2nd Edn. Springer

Kushner HJ, Dupuis P (2001) Numerical Methods for Stochastic Control Problems in Continuous Time - 2nd Edn, Springer Science & Business Media

Lukacs E (1958) Characteristic Functions. Charles Griffin & Company Limited London

McNeil AJ, Frey R, Embrechts P (2005) Quantitative Risk Management: Concepts, Techniques, and Tools. Princeton Series in Finance

Nelsen RB (1999) An Introduction to Copulas. Springer, New York. ISBN 0-387-98623-5

Portenko NI (1982) Generalized Diffusion Processes. American Mathematical Society

Rogers LCG, Williams D (2000) Diffusions, Markov Processes and Martingales. Cambridge University Press

Scarsini M (1984) On measures of concordance. Stochastica 8(3):201-218

Sklar A (1996) Random Variables, Distribution Functions, and Copulas: a personal look backward and forward. Lecture notes-monograph series, JSTOR

Stroock DW, Srinivasa Varadhan SR (1997) Multidimensional Diffusion Processes. Springer

Tavella D, Randall C (2000) Pricing Financial Instrument with the Finite Difference Method. Whiley Financial Engneering

Zhang H, Ding F (2013) On the Kronecker Products and Their Applications. J Appl Math 2013:8. Article ID 296185