A state predictor for continuous-time stochastic systems

Filippo Cacace[a], Valerio Cusimano[b,*], Alfredo Germani[c], Pasquale Palumbo[b]

[a]*Università Campus Bio-Medico di Roma, Via Álvaro del Portillo, 21, 00128 Roma - Italy (e-mail: f.cacace@unicampus.it)*
[b]*IASI-CNR Viale Manzoni 30, 00185 Roma, Italy (e-mail: valerio.cusimano@iasi.cnr.it, pasquale.palumbo@iasi.cnr.it)*
[c]*Dipartimento di Ingegneria Elettrica e dell'Informazione, Università degli Studi dell'Aquila, Via Vetoio, Coppito 67100 L'Aquila - Italy (e-mail: germani@ing.univaq.it)*

Abstract

This work investigates the state prediction problem for nonlinear stochastic differential systems, affected by multiplicative state noise. This problem is relevant in many state-estimation frameworks such as filtering of continuous-discrete systems (i.e. stochastic differential systems with discrete measurements) and time-delay systems. A very common heuristic to achieve the state prediction exploits the numerical integration of the deterministic nonlinear equation associated to the noise-free system. Unfortunately this methods provide the exact solution only for linear systems. Instead here we provide the exact state prediction for nonlinear system in term of the series expansion of the expected value of the state conditioned to the value in a previous time instant, obtained according to the Carleman embedding technique. The truncation of the infinite series allows to compute the prediction at future times with an arbitrary approximation. Simulations support the effectiveness of the proposed state-prediction algorithm in comparison to the aforementioned heuristic method.

Keywords: Nonlinear Filtering, Stochastic Systems, Nonlinear Systems, Kalman Filtering, Carleman Approximation

---

*Corresponding author
  Email address:* `valerio.cusimano@iasi.cnr.it` (Valerio Cusimano)

1. Problem formulation and background

Consider the following nonlinear stochastic differential system in the Itô formulation

$$dx_t = f(x_t)dt + \sum_{j=1}^{p} g_j(x_t)dW_{j,t} \tag{1}$$

defined on a probability space $(\Omega, \mathcal{F}, P)$, where x_t is the state vector, $f, g_j : \mathbb{R}^n \to \mathbb{R}^n$ are nonlinear analytic maps and $\{W_j \in \mathbb{R}, j = 1, \cdots, p\}$ is a set of pairwise independent standard Wiener processes with respect to a family of increasing σ-algebras $\{\mathcal{F}_t, \ t \geq 0\}$. The initial state $x_0 = \bar{x}$ is an $\mathcal{F}_0$-measurable random vector, independent of the state noises $W_{j,t}$.

The problem here investigated is the prediction of x_t given the value x_s at a previous time instant $s < t$, that is, the aim is to compute the conditional expectation $\mathbb{E}(x_t|x_s)$. According to the explicit solution of (1), and to the properties of the Itô integral [1] the expected value is given by

$$\mathbb{E}(x_t|x_s) = x_s + \int_s^t \mathbb{E}(f(x_\xi)|x_s)d\xi. \tag{2}$$

Predictions are broadly exploited in the more general setting of nonlinear filters, according to the usual paradigm suggesting to write the filter equations as a "*prediction*" + a "*correction*" term (see e.g. [1, 2]). The common denominator of such approaches, that include the well known Extended Kalman-Bucy filter for the continuous-time case, is that the prediction step is entrusted to the following coarse simplification:

$$\mathbb{E}(f(x_\xi)|x_s) \simeq f\big(\mathbb{E}(x_\xi|x_s)\big). \tag{3}$$

Clearly, such an approximation is exact only in special cases, like linear systems (see e.g. [3]). Such simplification provides an easy-to-handle heuristics since, by exploiting (3), the prediction $z_t = \mathbb{E}(x_t|x_s)$ of (2) can be computed as the solution of the deterministic differential system

$$\dot{z}_t = f(z_t), \qquad z_s = x_s. \tag{4}$$

Motivation for the present note stems from the need to build up a theory providing the solution to the state prediction problem that, in principle, could

be applied to such a broad range of nonlinear filtering framework. This problem is especially important when designing continuous-discrete (CD) filters, that is, filters for stochastic differential systems with sampled measurements. CD filters are ubiquitous in problems such as tracking [4], finance [1] and systems biology [5], and they are receiving growing attention in recent years [6, 7, 8]. One reason for this interest is that new application areas such as systems biology often employ continuous-time models as in (1), coupled with sampled measurements with large sampling intervals. There can be found many solutions to CD filters in the literature, such as continuous-discrete extended Kalman filter CD-EKF [2, 9], CD unscented Kalman filter, CD-UKF [10], and CD cubature Kalman filter, CD-CKF [11]. In the case of CD filters the crucial problem is the evolution of the state moments during the sampling interval, since the optimal state estimate over the inter-sampling period (e.g. for $t \in [k\Delta, (k + 1)\Delta)$, where $t = k\Delta$, $k = 0, 1, \ldots$ are the time instants when measurements are acquired) is provided by the prediction $\mathbb{E}(x_t|\hat{x}_{k\Delta})$, formally defined by

$$\mathbb{E}(x_t|\hat{x}_{k\Delta}) = \hat{x}_{k\Delta} + \int_{k\Delta}^{t} \mathbb{E}(f(x_\xi)|\hat{x}_{k\Delta})d\xi. \tag{5}$$

This is, clearly, the same problem introduced in (2) with $s = k\Delta$ and $x_s = \hat{x}(k\Delta)$. In the literature, CD filters share the same approach to solve the optimal state prediction problem based on (3), and eq.(4) (written with $s = k\Delta$) endowed with the equation of the covariance of the prediction error are also called *differential moment equations*: in this framework several recent works have been devoted to propose precise and efficient methods to evaluate the solution of (4) on small discretization intervals, see for example [6, 7, 8]. However, these methods provide precise solutions to the approximate equation (4), whereas the exact prediction (2) cannot be obtained by solving an ordinary differential equation (see for example [2], p. 168). The right hand-side of (2) involves an expectation that requires the whole conditional density for its evaluation. Stated differently, the evaluation of the first two moments of the prediction depends on all the other moments. It should be mentioned that when the discretization interval is not negligible the solutions of (2) and (4) can be quite different even in the

3

simple case of scalar systems.

Furthermore, the application of state predictors for stochastic systems is not limited to CD filters, for example they are useful in the area of stochastic delay equations [12] as well as predictors and filters for stochastic systems with delays in the input and/or the output [13, 14].

From a theoretical viewpoint, the solution to the prediction problem can be pursued by searching for the conditional density $p(x_t, t | x_s, s)$ provided by the solution of the Kolmogorov forward equation, and then use p to compute the conditional expectation (2). Since the solution of the Kolmogorov forward equation can be obtained by analytic means only in few cases, a number of numerical methods have been proposed to this aim, including finite-difference method [15], finite elements [16], adaptive finite-elements ([17], pp. 115–123), quadrature-based methods [18], the adjoint method [19], Galerkin's method [20, 21], particle methods [22, 23, 24] and Markov chain Monte Carlo methods [25]. The computational complexity of solving the Kolmogorov equation increases exponentially with the dimension of the state vector. For this reason, this approach is not well suited for the implementation of real-time predictors (or filters) even for systems of moderate size.

In this work we introduce an approximation scheme for the state prediction equation (2). The proposed solution has several positive features. In the first place we provide the exact solution to the correct problem statement: its analytical form is expressed in terms of a Taylor series expansion, thus the state prediction (2) can be computed with any arbitrary precision. In the second place, and for the same reason, the prediction can be made precise on arbitrary sampling intervals. A final advantage is that, because of the analytic expression provided by the method, the state prediction $\mathbb{E}(x_t | x_s)$ can be used for analysis purposes, and not only as a numerical value. As a drawback, it can be mentioned that our method applies only to systems of the form (1), that is, to time-invariant nonlinear systems, in contrast with other methods that include also the time-varying case.

The approximation scheme is in essence based on the Carleman embedding

4

technique, already exploited in [26] for a slightly different stochastic differential system (nonlinear drift + additive Gaussian noise, instead of the more general nonlinear diffusion term here considered) with the aim of continuous-time filters. The Carleman technique results in the embedding of the original finite-dimensional nonlinear system into an infinite-dimensional bilinear one. Differently from [26], where the state-estimation problem required a further finite-dimensional approximation of the Carleman embedding, here we propose the exact stochastic discretization of the Carleman embedding without any approximation, thus obtaining the optimal prediction as the sum of a series, that we name the *Carleman prediction* in analogy to the deterministic case [27, 28].

Section 2 describes the approach, while Section 3 is devoted to its evaluation. In particular we consider a class of systems widely used in financial mathematics and for which the solution of the prediction problem here described has important applications. In a few cases the exact solution can be found by analytic tools. We consider one such cases, with the aim of comparing the exact solution of (2) with the standard approximation (4) and the one provided by our method.

Notation. I_n denotes the identity matrix in $\mathbb{R}^n$. $0_{n,m}$ denotes a matrix of zeros in $\mathbb{R}^{n \times m}$. The symbol $\otimes$ denotes the Kronecker matrix product, the notation $A^{[i]}$ is used for the Kronecker power of matrix A, that is $A \otimes A \otimes \cdots \otimes A$, repeated i times. The standard Jacobian of $f : \mathbb{R}^n \to \mathbb{R}^n$ can be formally written as $\nabla_x \otimes f$, where ∇_x denotes the operator $[\partial/\partial x_1 \ldots \partial/\partial x_n]$. Higher-order derivatives of f are represented as $\nabla_x^{[i]} \otimes f = \nabla_x \otimes \left(\nabla_x^{[i-1]} \otimes f \right)$, where $\nabla_x^{[i]} \otimes f : \mathbb{R}^n \to \mathbb{R}^{n \times n^i}$.

2. Carleman predictor for stochastic systems

Consider the problem of computing the state prediction $\mathbb{E}(x_t|\bar{x})$, with $x_s = \bar{x}$, $t \geq s$ and x_t that evolves according to (1). Whenever useful, the non negative displacement $t - s$ will be referred to as $\Delta \geq 0$. The proposed prediction algorithm is based on the following steps.

1. Define the displacement

$$\varphi_t = x_t - \bar{x}. \tag{6}$$

2. By using the Carleman embedding technique [27] we transform the nonlinear stochastic differential system for φ_t into an infinite-dimensional bilinear system (linear drift and multiplicative noise).

3. Since the system is bilinear, the exact prediction of the state in the embedded space is obtained by integrating the corresponding linear drift without the noise terms.

4. Finally, we project the solution onto the original finite-dimensional space to obtain $\mathbb{E}(x_t|\bar{x})$.

To exploit the Carleman embedding, we shall make use of the representation of the analytic maps $f(x_t)$ and $g_j(x_t)$ as Taylor expansions around $\bar{x}$, written according to the Kronecker formalism (see [26] for details):

$$f(x_t) = \sum_{i=0}^{\infty} A_i(\bar{x})\varphi_t^{[i]}, \quad g_j(\varphi_t) = \sum_{i=0}^{\infty} G_i^j(\bar{x})\varphi_t^{[i]}, \tag{7}$$

$$A_i(\bar{x}) = \frac{\nabla_x^{[i]} \otimes f(x)}{i!}\Big|_{x=\bar{x}} \in \mathbb{R}^{n \times n^i}, \tag{8}$$

$$G_i^j(\bar{x}) = \frac{\nabla_x^{[i]} \otimes g_j(x)}{i!}\Big|_{x=\bar{x}} \in \mathbb{R}^{n \times n^i}. \tag{9}$$

From (1),(6)-(9) we have the following stochastic differential system for φ_t, with initial value $\varphi_s = 0$.

$$d\varphi_t = \sum_{i=0}^{\infty} A_i(\bar{x})\varphi_t^{[i]}dt + \sum_{j=1}^{p}\sum_{i=0}^{\infty} G_i^j(\bar{x})\varphi_t^{[i]}dW_{j,t}. \tag{10}$$

The Carleman embedding technique requires to express the stochastic differentials $d(\varphi_t^{[h]})$, $h \geq 1$. This expression is provided by the following Theorem.

Theorem 1. *The differential $d(\varphi_t^{[h]})$, for $h \geq 2$, can be written as:*

$$d(\varphi_t^{[h]}) = \overline{H}_{h,h-2}(\bar{x})\varphi_t^{[h-2]}dt$$
$$+ \sum_{k=h-1}^{\infty}\left(\overline{A}_{h,k}(\bar{x}) + \overline{H}_{h,k}(\bar{x})\right)\varphi_t^{[k]}dt$$
$$+ \sum_{j=1}^{p}\sum_{k=h-1}^{\infty}\overline{G}_{h,k}^j(\bar{x})\varphi_t^{[k]}dW_{j,t} \tag{11}$$

6

where, for $k \geq h - 2$:

$$\overline{H}_{h,k}(\bar{x}) = \frac{1}{2} O_n^h \left(\sum_{j=1}^{p} H_{k-h+2}^{j}(\bar{x}) \otimes I_{n^{h-2}} \right) \tag{12}$$

with

$$H_l^j(\bar{x}) = \sum_{i=0}^{l} G_i^j(\bar{x}) \otimes G_{l-i}^j(\bar{x}), \tag{13}$$

and, for $k \geq h - 1$:

$$\overline{A}_{h,k}(\bar{x}) = U_n^h \left(A_{k-h+1}(\bar{x}) \otimes I_{n^{h-1}} \right) \tag{14}$$

$$\overline{G}_{h,k}^{j}(\bar{x}) = U_n^h \left(G_{k-h+1}^{j}(\bar{x}) \otimes I_{n^{h-1}} \right), \tag{15}$$

with matrices O_n^h and U_n^h defined in Lemma 5 in Appendix.

PROOF. By using Theorem 5.2 in [29], differentials $d(\varphi_t^{[h]})$ can be written as:

$$d(\varphi_t^{[h]}) = \left(\nabla_\varphi \otimes \varphi_t^{[h]} \right) \sum_{i=0}^{\infty} A_i(\bar{x}) \varphi_t^{[i]} dt$$

$$+ \frac{1}{2} \left(\nabla_\varphi^{[2]} \otimes \varphi_t^{[h]} \right) \sum_{j=1}^{p} \left(\sum_{i=0}^{\infty} G_i^j(\bar{x}) \varphi_t^{[i]} \right)^{[2]} dt$$

$$+ \left(\nabla_\varphi \otimes \varphi_t^{[h]} \right) \sum_{j=1}^{p} \sum_{i=0}^{\infty} G_i^j(\bar{x}) \varphi_t^{[i]} dW_{j,t}. \tag{16}$$

According to Lemma 5 in Appendix, the first term in the right-hand-side of (16) can be written as follows:

$$\left(\nabla_\varphi \otimes \varphi_t^{[h]} \right) \sum_{i=0}^{\infty} A_i(\bar{x}) \varphi_t^{[i]} dt$$

$$= \sum_{i=0}^{\infty} U_n^h \left(I_n \otimes \varphi_t^{[h-1]} \right) \cdot \left(\left(A_i(\bar{x}) \varphi_t^{[i]} \right) \otimes 1 \right) dt$$

$$= \sum_{i=0}^{\infty} U_n^h \left(\left(A_i(\bar{x}) \varphi_t^{[i]} \right) \otimes \left(I_{n^{h-1}} \varphi_t^{[h-1]} \right) \right) dt$$

$$= \sum_{i=0}^{\infty} U_n^h \left(A_i(\bar{x}) \otimes I_{n^{h-1}} \right) \varphi_t^{[i+h-1]} dt$$

$$= \sum_{k=h-1}^{\infty} U_n^h \left(A_{k-h+1}(\bar{x}) \otimes I_{n^{h-1}} \right) \varphi_t^{[k]} dt$$

$$= \sum_{k=h-1}^{\infty} \overline{A}_{h,k}(\bar{x}) \varphi_t^{[k]} dt \tag{17}$$

7

where the identity

$$(A \otimes B) \cdot (C \otimes D) = (A \cdot C) \otimes (B \cdot D), \tag{18}$$

that holds true for matrices of suitable size, has been repeatedly applied, and matrix $\overline{A}_{h,k}(\bar{x})$ is defined in (14). Analogously, the third term in (16) can be written as:

$$\left(\nabla_\varphi \otimes \varphi_t^{[h]} \right) \sum_{j=1}^{p} \sum_{i=0}^{\infty} G_i^j(\bar{x}) \varphi_t^{[i]} dW_{j,t}$$

$$= \sum_{j=1}^{p} \sum_{k=h-1}^{\infty} U_n^h \left(G_{k-h+1}^j(\bar{x}) \otimes I_{n^{h-1}} \right) \varphi_t^{[k]} dW_{j,t}$$

$$= \sum_{j=1}^{p} \sum_{k=h-1}^{\infty} \overline{G}_{h,k}^j(\bar{x}) \varphi_t^{[k]} dW_{j,t} \tag{19}$$

with $\overline{G}_{h,k}^j(\bar{x})$ defined in (15). For the second term in the right-hand-side of (16) we have, by properly exploiting identity (18),

$$\left(\sum_{i=0}^{\infty} G_i^j(\bar{x}) \varphi_t^{[i]} \right)^{[2]} = \sum_{i=0}^{\infty} \sum_{k=0}^{\infty} \left(G_i^j(\bar{x}) \otimes G_k^j(\bar{x}) \right) \varphi_t^{[i+k]}$$

$$= \sum_{i=0}^{\infty} \sum_{l=i}^{\infty} \left(G_i^j(\bar{x}) \otimes G_{l-i}^j(\bar{x}) \right) \varphi_t^{[l]} = \sum_{l=0}^{\infty} H_l^j(\bar{x}) \varphi_t^{[l]} \tag{20}$$

with $H_l^j(\bar{x})$ defined in (13). Thus, according to Lemma 5 in Appendix, by

8

repeatedly applying identity (18), it follows that

$$
\frac{1}{2}\left(\nabla_\varphi^{[2]} \otimes \varphi_t^{[h]}\right) \sum_{j=1}^{p}\left(\sum_{i=0}^{\infty} G_i^j(\bar{x})\varphi_t^{[i]}\right)^{[2]}
$$

$$
=\frac{1}{2}\sum_{l=0}^{\infty} O_n^h \sum_{j=1}^{p}\left(I_{n^2} \otimes \varphi_t^{[h-2]}\right) \cdot \left(\left(H_l^j(\bar{x})\varphi_t^{[l]}\right) \otimes 1\right)
$$

$$
=\frac{1}{2}\sum_{l=0}^{\infty} O_n^h \sum_{j=1}^{p}\left(H_l^j(\bar{x})\varphi_t^{[l]}\right) \otimes \left(I_{n^{h-2}}\varphi_t^{[h-2]}\right)
$$

$$
=\frac{1}{2}\sum_{l=0}^{\infty} O_n^h \left(\sum_{j=1}^{p} H_l^j(\bar{x}) \otimes I_{n^{h-2}}\right)\varphi_t^{[l+h-2]}
$$

$$
=\frac{1}{2}\sum_{k=h-2}^{\infty} O_n^h \left(\sum_{j=1}^{p} H_{k-h+2}^j(\bar{x}) \otimes I_{n^{h-2}}\right)\varphi_t^{[k]}
$$

$$
=\sum_{k=h-2}^{\infty} \overline{H}_{h,k}(\bar{x})\varphi_t^{[k]} \tag{21}
$$

with $\overline{H}_{h,k}(\bar{x})$ defined in (12). In particular, notice that

$$
\overline{H}_{2,0}(\bar{x}) = \frac{1}{2}O_n^2 \sum_{j=1}^{p} H_0^j(\bar{x}) = \frac{1}{2}O_n^2 \sum_{j=1}^{p}(G_0^j(\bar{x}))^{[2]}. \tag{22}
$$

The theorem is proved by substituting (17), (19) and (21) in (16).

$$\square$$

We can now proceed to embed (1) into an infinite dimensional bilinear system, with linear drift.

Theorem 2. *Define the infinite-dimensional vector composed by the Kronecker powers of the displacement φ_t defined in (6) as follows:*

$$
\Phi_t = \begin{bmatrix} \varphi_t^T & \varphi_t^{[2]T} & \cdots \end{bmatrix}^T. \tag{23}
$$

Then, Φ_t is the solution of the following stochastic differential system (linear drift and multiplicative noise) with initial condition $\Phi_s = 0$

$$
\begin{aligned}
d\Phi_t = {}& \widetilde{A}(\bar{x})\Phi_t dt + \widetilde{L}(\bar{x})dt \\
& + \sum_{j=1}^{p}\left(\widetilde{B}^j(\bar{x})\Phi_t + \widetilde{F}^j(\bar{x})\right) dW_{j,t}
\end{aligned} \tag{24}
$$

where the building blocks $\widetilde{A}_{il}, \widetilde{B}_{il}^j \in \mathbb{R}^{n^i \times n^l}$, $\widetilde{L}_i, \widetilde{F}_i^j \in \mathbb{R}^{n^i \times 1}$, $i, l = 1, 2, \dots$ providing the infinite-dimensional block matrices $\widetilde{A}$, $\widetilde{L}$, $\widetilde{B}^j$, $\widetilde{F}^j$ are defined as

$$
\widetilde{A}_{i,l}(\bar{x}) = \begin{cases} A_l(\bar{x}), & \text{if } i = 1, \\ \overline{H}_{i,i-2}(\bar{x}), & \text{if } l = i - 2, \\ \overline{A}_{i,l}(\bar{x}) + \overline{H}_{i,l}(\bar{x}) & \text{if } i > 1, \ l > i - 2 \\ 0_{n^i \times n^l}, & \text{otherwise} \end{cases} \tag{25}
$$

$$
\widetilde{L}_i(\bar{x}) = \begin{cases} A_0(\bar{x}), & \text{if } i = 1 \\ \overline{H}_{2,0}(\bar{x}), & \text{if } i = 2 \\ 0_{n^i \times 1}, & \text{otherwise} \end{cases} \tag{26}
$$

$$
\widetilde{B}_{i,l}^j(\bar{x}) = \begin{cases} \overline{G}_{i,l}^j(\bar{x}), & \text{if } l \geq i - 1 \\ 0_{n^i \times n^l}, & \text{otherwise} \end{cases} \tag{27}
$$

$$
\widetilde{F}_i^j(\bar{x}) = \begin{cases} G_0^j(\bar{x}), & i = 1 \\ 0_{n^i \times 1}, & \text{otherwise} \end{cases} \tag{28}
$$

95 PROOF. The proof is immediately obtained by aggregating the differentials $d(\varphi_t^{[h]})$, $h \geq 1$, defined in (10) and in Theorem 1.

$\square$

For the ease of the reader we report below the structure of the block matrices previously defined:

$$
\widetilde{A} = \begin{bmatrix} A_1 & A_2 & A_3 & \cdots \\ \overline{A}_{2,1} + \overline{H}_{2,1} & \overline{A}_{2,2} + \overline{H}_{2,2} & \overline{A}_{2,3} + \overline{H}_{2,3} & \cdots \\ \overline{H}_{3,1} & \overline{A}_{3,2} + \overline{H}_{3,2} & \overline{A}_{3,3} + \overline{H}_{3,3} & \cdots \\ 0 & \overline{H}_{4,2} & \overline{A}_{4,3} + \overline{H}_{4,3} & \cdots \\ 0 & 0 & \overline{H}_{5,3} & \ddots \\ \vdots & \vdots & \ddots & \ddots \end{bmatrix}, \tag{29}
$$

$$\widetilde{L} = \begin{bmatrix} A_0 \\ \overline{H}_{2,0} \\ 0 \\ \vdots \end{bmatrix}, \quad \widetilde{B}^j = \begin{bmatrix} G_1^j & G_2^j & \cdots \\ \overline{G}_{2,1}^j & \overline{G}_{2,2}^j & \cdots \\ 0 & \overline{G}_{3,2}^j & \cdots \\ 0 & 0 & \ddots \\ \vdots & \vdots & \ddots \end{bmatrix}, \quad \widetilde{F}^j = \begin{bmatrix} G_0^j \\ 0 \\ 0 \\ \vdots \end{bmatrix}. \tag{30}$$

Theorem 2 provides an infinite-dimensional bilinear systems (i.e. linear drift + multiplicative noise) with respect to the extended state accounting for the displacement $\varphi_t = x_t - \bar{x}$ and for its powers. By defining $\zeta_t = \mathbb{E}(\Phi_t|\bar{x})$ for $t \geq s$, it is readily shown that it obeys the following ordinary differential equation system:

$$\dot{\zeta}_t = \widetilde{A}(\bar{x})\zeta_t + \widetilde{L}(\bar{x}), \qquad \zeta_s = \mathbb{E}(\Phi_s|\bar{x}) = 0, \tag{31}$$

whose explicit solution is given by

$$\zeta_t = \int_s^t e^{\widetilde{A}(\bar{x})(t-\tau)}\widetilde{L}(\bar{x})d\tau = \sum_{i=1}^{\infty} \widetilde{A}^{i-1}(\bar{x})\widetilde{L}(\bar{x})\frac{\Delta^i}{i!}, \tag{32}$$

The following Theorem allows to extract the finite-dimensional information from the sequence of infinite-dimensional vectors provided by (32).

Theorem 3. *Consider a projection operator Π_n providing the first n components of an infinite-dimensional vector. Then:*

$$\mathbb{E}(x_t|\bar{x}) = \bar{x} + \sum_{i=1}^{\infty} Q_{i-1}(\bar{x})\frac{\Delta^i}{i!}, \tag{33}$$

where for $i \geq 0$,

$$Q_i(\bar{x}) = \Pi_n \widetilde{A}^i(\bar{x})\widetilde{L}(\bar{x}). \tag{34}$$

PROOF. According to the definition of φ_t in (6) it is:

$$\mathbb{E}(x_t|\bar{x}) = \mathbb{E}(\varphi_t + \bar{x}|\bar{x}) = \bar{x} + \mathbb{E}(\varphi_t|\bar{x}) = \bar{x} + \mathbb{E}(\Pi_n\Phi_t|\bar{x})$$
$$= \bar{x} + \Pi_n\mathbb{E}(\Phi_t|\bar{x}) = \bar{x} + \Pi_n\zeta_t \tag{35}$$

with ζ_t given by (32). The proof is readily completed by substituting (32) in (35).

$\square$

11

The essential property here is that in spite of the infinite dimension of the matrices involved in (34), each term of the sum in (33) has finite size. This allows to compute $\mathbb{E}(x_t|\bar{x})$ with a desired precision by truncating the series at an appropriate index $\nu < \infty$. In fact, consider the following block-diagonal decomposition for $\widetilde{A}^i \widetilde{L}$ (we neglect $\bar{x}$ in the matrices),

$$\widetilde{A}^i \widetilde{L} = \begin{bmatrix} [\widetilde{A}^i \widetilde{L}]_1 \\ \vdots \\ [\widetilde{A}^i \widetilde{L}]_j \\ \vdots \end{bmatrix}, \qquad [\widetilde{A}^i \widetilde{L}]_j \in \mathbb{R}^{n^j \times 1}. \tag{36}$$

Definitions (25)–(26) imply that $[\widetilde{A}^i \widetilde{L}]_j$ is a zero block for $j > 2i+2$. Moreover, from (34) it is clear that that $Q_i = [\widetilde{A}^i \widetilde{L}]_1$. The following lemma allows to recursively build the nontrivial blocks of $\widetilde{A}^i \widetilde{L}$.

Lemma 4. *Define*

$$\mu^i = \begin{bmatrix} \mu_1^i \\ \vdots \\ \mu_{2i+2}^i \end{bmatrix} \qquad \text{with} \qquad \mu_j^i = [\widetilde{A}^i \widetilde{L}]_j \in \mathbb{R}^{n^j}. \tag{37}$$

Then $\mu^{i+1} = \mathcal{A}(i)\mu^i$, with

$$\mathcal{A}(i) = \begin{bmatrix} \widetilde{A}_{1,1} & \cdots & \widetilde{A}_{1,2i+2} \\ \vdots & \ddots & \vdots \\ \widetilde{A}_{2i+4,1} & \cdots & \widetilde{A}_{2i+4,2i+2} \end{bmatrix} \tag{38}$$

PROOF. The proof is obtained by computing μ_j^{i+1}, for $j = 1, \ldots, 2i+4$. Indeed, by definition, it is:

$$\mu_j^{i+1} = [\widetilde{A}^{i+1} \widetilde{L}]_j = \sum_{l=1}^{2i+2} \widetilde{A}_{j,l} [\widetilde{A}^i \widetilde{L}]_l, \qquad j = 1, \ldots, 2i+4. \tag{39}$$

Thus, according to μ^i and $\mathcal{A}(i)$ definitions, μ_j^{i+1} in (39) becomes:

$$\mu_j^{i+1} = \sum_{l=1}^{2i+2} [\mathcal{A}(i)]_{j,l} \cdot \mu_l^i \tag{40}$$

12

that completes the proof.

□

Remark 1. *In absence of noise (i.e. when $g_j(x) = 0$ in (1)), the matrix $\widetilde{A}$ is composed only by blocks A_l and $\overline{A}_{i,l}$, since $\overline{H}_{i,l} = 0$. Moreover, all matrices $\widetilde{B}^j$ vanish, yielding a linear (instead of a bilinear) extended system for Φ_t. In this case eq.(33) is the Taylor series of the solution to $\dot{x}(t) = f(x(t))$ with initial condition $\bar{x}$, as shown in [28].*

Remark 2. *In presence of noise terms, (33) is the Taylor series with respect to time of $\mathbb{E}(x_t|\bar{x})$, where x_t is the solution of (1) with initial condition $\bar{x} = x_s$. The approximation error when the series is truncated at index ν can be estimated with the usual methods for Taylor series. For example, if there exists a compact set $D \subset \mathbb{R}^n$ containing the evolution of $\mathbb{E}(x_t|\bar{x})$, we can define*

$$M = \max_{x \in D} Q_\nu(x), \tag{41}$$

and the estimation error for a finite Carleman approximation $\nu < \infty$ satisfies the following uniform (i.e. independent from $\bar{x}$) bound

$$r_{\bar{x},\nu}(t-s) = \mathbb{E}(x_t|\bar{x}) - \left(\bar{x} + \sum_{i=1}^{\nu} Q_{i-1}(\bar{x}) \frac{(t-s)^i}{i!} \right) \leq M \frac{(t-s)^{\nu+1}}{(\nu+1)!}. \tag{42}$$

3. Simulations

In order to validate the proposed approach we consider a class of nonlinear stochastic models in the area of financial mathematics for which the prediction problem is of great practical interest and that admit a closed-form expression of the conditional expectation in some special cases. We are therefore in the condition to evaluate the accuracy of the Carleman predictor in terms of displacements from the real solution, and to compare our results to the ones obtained by applying the deterministic prediction (3)-(4) so broadly exploited in the literature [6, 7, 8].

13

Consider the $\mathbb{R}^m$-valued process y which satisfies the following stochastic differential equation with initial condition $y_0 \in \mathbb{R}^m$:

$$dy_t = b(t, y_t)dt + \sigma(t, y_t)dW_t, \tag{43}$$

where $b : [0, T] \times \mathbb{R}^m \to \mathbb{R}^m$ and $\sigma : [0, T] \times \mathbb{R}^m \to \mathbb{R}^{m \times m}$ are continuous functions, such that $b(t, \cdot)$, $\sigma(t, \cdot)$ are Lipschitz continuous on $\mathbb{R}^m$, uniformly in time, for $t \in [0, T]$, $T > 0$. Given $z_0 \in \mathbb{R}$, let us introduce the $\{\mathcal{F}_t\}$-adapted stochastic process z, defined for $t \geq 0$,

$$z_t = z_0 \exp\left(-\int_0^t c(s, y_s)ds\right), \tag{44}$$

where $c : [0, T] \times \mathbb{R}^m \to \mathbb{R}$ is also a continuous functions. The extended process $x := [y^T, z]^T \in \mathbb{R}^{m+1}$, with $x_0 := (y_0, z_0)$ satisfies

$$dx_t = \begin{bmatrix} b(t, y_t) \\ -c(t, y_t)z_t \end{bmatrix} dt + \begin{bmatrix} \sigma(t, y_t) \\ 0_{1 \times m} \end{bmatrix} \cdot dW_t. \tag{45}$$

The computation of the expectation value of processes like (45) is one of the main problems faced in financial mathematics [30], [31], [32]. In particular, for a given real valued function ϕ, there is interest in the conditional expectation

$$\mathbb{E}\left[z_T \phi(y_T)|\mathcal{F}_t\right], \tag{46}$$

for $0 \leq t \leq T$, which, under a suitable specification of functions c and ϕ, represent the no-arbitrage price of a financial derivative with maturity T at time t, in several pricing situations.

In some particular cases (46) admits a closed-form solution. For example, this happens if process y is scalar and satisfies

$$dy_t = k(\theta - y_t)dt + \sigma\sqrt{y_t}dW_t, \tag{47}$$

where $y_0 > 0$ and W_t is a 1-dimensional Brownian motion. The coefficient $k > 0$ is the *speed* of this mean reversion, $\theta > 0$ is the *long run average intensity* and $\sigma > 0$ is called *volatility* and when they satisfy Feller's condition $2k\theta \geq \sigma^2$ [33] the trajectories of the process are positive. The coefficient values used in

14

the simulations are reported in Table 1. This model exhibits mean reversion of the intensity, causing the intensity to be pulled downward when it is above the long run average intensity and upward in the opposite case. Model (47) was developed in [34] and has been extensively applied in financial mathematics to describe the term structure of interest rates.

k	θ	σ
0.1209	0.0423	0.1642

Table 1: Parameters of process (47).

130

When (47) is used in (45) with $\phi(y) \equiv 1$, $c(t, y_t) = y_t$, it is possible to derive the analytical expression of the prediction for x_T [34]

$$\mathbb{E}\left[x_T | \mathcal{F}_t\right] = \begin{bmatrix} y_t e^{-k(T-t)} + \theta(1 - e^{-k(T-t)}) \\ z u(y_t, t) \end{bmatrix} \tag{48}$$

for any $0 \leq t \leq T$. The closed form solution for the function u is:

$$u(y, t) = A(T - t)e^{-B(T-t)y}, \tag{49}$$

where

$$A(T - t) = \left[\frac{2\gamma e^{(k+\gamma)(T-t)/2}}{(\gamma + k)(e^{\gamma(T-t)} - 1) + 2\gamma}\right]^{2k\theta/\sigma^2}, \tag{50}$$

$$B(T - t) = \frac{2(e^{\gamma(T-t)} - 1)}{(\gamma + k)(e^{\gamma(T-t)} - 1) + 2\gamma}, \tag{51}$$

with $\gamma = \sqrt{k^2 + 2\sigma^2}$.

Now we compare the exact prediction with the values obtained from (33) at different levels of accuracy corresponding to $\nu = 1, \cdots, 7$, and with the "deterministic" approximation (3)-(4). In particular, fixed the initial state $x_0^T = [\theta, 1]$, the state prediction of z_t for $T \in [0, 8]$ computed according to the three methods is reported in Figure 1.

Figure 2 shows the percentage error for different order of the Carleman approximation in comparison with the deterministic approximation at $T = 4$. As
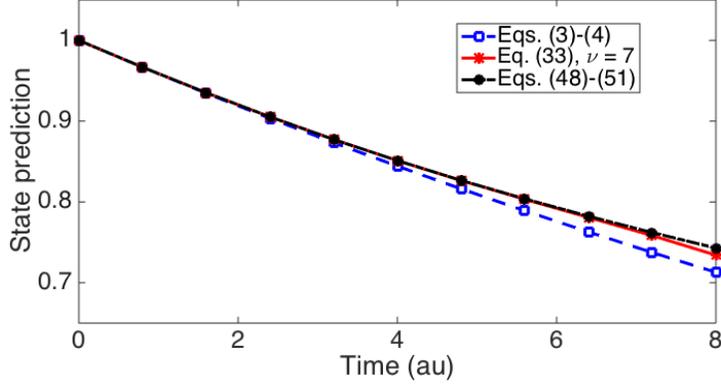
15

Figure 1: Time evolution of the state prediction.
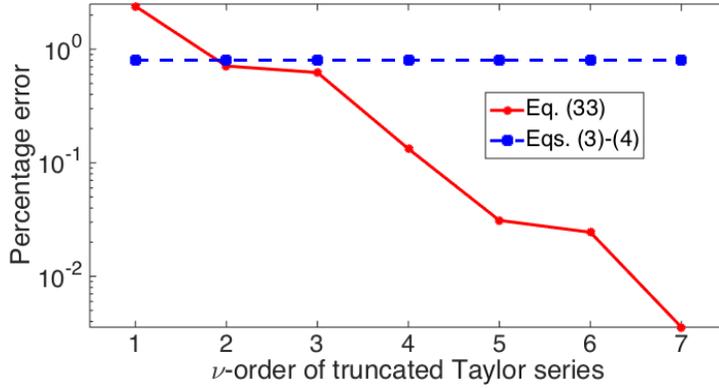

Figure 2: Percentage error

expected the accuracy of the Carleman predictor increases when adding more
terms in the sum (33) and in fact only in the linear case the deterministic ap-
proximation has a smaller error. The same results are reported in Table 2. As
for the computational cost, for a prediction interval $T = 5$ the implementa-
tion of the Carleman predictor used in these simulation with $\nu \leq 6$ requires
about $5 \cdot 10^{-2}$s of computation compared with about $5 \cdot 10^{-3}$s for the standard
MATLAB® ODE solver ode45(). Notice that the computation time for the
ODE solver depends on the prediction interval, whereas the Carleman predictor
performs only one step and its computational time is independent from T.

16

		$T = 2$	$T = 4$	$T = 6$
Eqs. (4)-(5)		0.12	0.80	2.15
Eq. (33)	$\nu = 1$	0.50	2.39	5.89
	$\nu = 3$	0.04	0.62	3.00
	$\nu = 5$	$6 \cdot 10^{-4}$	0.03	0.27
	$\nu = 7$	10^{-5}	$3 \cdot 10^{-3}$	0.11

Table 2: Percentage error for $T \in \{2,\ 4,\ 6\}$.

4. Conclusion

The method described in this paper provides a series expansion for the prediction problem of nonlinear stochastic systems. The appropriate balance between precision and computational cost can be achieved by tuning the number of terms. Since the prediction problem is ubiquitous in the area of continuous-discrete systems the method proposed here can be useful in a number of applications. Further work will be devoted to the derivation of a filtering algorithm for continuous-time systems with discrete measurements based on the Carleman linearization technique that we have presented.

Acknowledgement

We gratefully acknowledge the help of Prof. Marco Papi who drew our attention to the relevance of the prediction problem for stochastic systems in the area of financial mathematics, and introduced us to the class of systems described in Section 3.

Appendix

Properties of the stochastic differential of Kronecker powers

The following Lemma shows a useful property shared by first and second order differentials of Kronecker powers.

17

Lemma 5. *([26])*

$$\nabla_x \otimes x^{[h]} = U_n^h \left(I_n \otimes x^{[h-1]} \right), \quad U_n^h \in \mathbb{R}^{n^h \times n^h} \tag{52}$$

$$\nabla_x^{[2]} \otimes x^{[h]} = O_n^h \left(I_{n^2} \otimes x^{[h-2]} \right), \quad O_n^h \in \mathbb{R}^{n^h \times n^h} \tag{53}$$

where

$$U_n^h = I_{n^h} + C_{n^{h-1},n}^T (U_n^{h-1} \otimes I_n), \tag{54}$$

$$O_n^h = U_n^h C_{n^{h-1},n}^T ((U_n^{h-1} C_{n^{h-2},n}^T) \otimes I_n) C_{n^2,n^{h-2}}^T \tag{55}$$

with $U_n^1 = I_n$, and C_{n_1,n_2}^T stands for the commutation matrix of the Kronecker product between vectors of size n_1, n_2.

4.1. Commutation matrices

For any given pair of matrices $A \in \mathbb{R}^{r \times s}$, $B \in \mathbb{R}^{n \times m}$, we have

$$B \otimes A = C_{r,n}^T (A \otimes B) C_{s,m}, \tag{56}$$

where the commutation matrix $C_{u,v}$ is the $(u \cdot v) \times (u \cdot v)$ matrix such that its (h,l) entry is given by

$$\{C_{u,v}\} = \begin{cases} 1 & \text{if } l = (|h-1|_v)u + \left(\left[\frac{h-1}{v}\right] + 1\right) \\ 0 & \text{otherwise} \end{cases} \tag{57}$$

$$\{C_{1,1}\} = 1$$

More details on commutation matrices and the Kronecker algebra properties can be found in [35].

References

[1] B. Øksendal, Stochastic differential equations, Springer, 2003.

[2] A. H. Jazwinski, Stochastic Processes and Filtering Theory, Academic Press, 1970.

[3] J. Jimenez, T. Ozaki, Linear estimation of continuous-discrete linear state space models with multiplicative noise, Systems & Control Letters 47 (2002) 91–101.

[4] B. Teixeira, M. Santillo, R. Erwin, D. S. Bernstein, Spacecraft tracking using sampled-data Kalman filters, IEEE Contr. Syst. Mag. 28(4) (2008) 78–94.

[5] D. Dochain, State and parameter estimation in chemical and biochemical processes: a tutorial, Journal Process Control 13 (2003) 801–818.

[6] T. Mazzoni, Computational aspects of continuous-discrete extended Kalman-filtering, Computational Statistics 23 (4) (2007) 519–539.

[7] P. Frogerais, J.-J. Bellanger, L. Senhadji, Various ways to compute the continuous-discrete extended Kalman filter, IEEE Trans. Autom. Control 57 (4) (2012) 1000–1004.

[8] G. Kulikov, M. Kulikova, Accurate numerical implementation of the continuous-discrete extended Kalman filter, Automatic Control, IEEE Transactions on 59 (1) (2014) 273–279.

[9] F. Lewis, Optimal Estimation: with an Introduction to Stochastic Control Theory, New York, NY, USA: Wiley, 1986.

[10] S. Särkkä, On unscented Kalman filtering for state estimation of continuous-time nonlinear systems, Automatic Control, IEEE Transactions on 52 (9) (2007) 1631–1641.

[11] I. Arasaratnam, S. Haykin, T. R. Hurd, Cubature Kalman filtering for continuous-discrete systems: theory and simulations, Signal Processing, IEEE Transactions on 58 (10) (2010) 4977–4993.

[12] A. Germani, C. Manes, P. Pepe, A twofold spline approximation for finite horizon LQG control of hereditary systems, SIAM Journal on Control and Optimization 39 (4) (2000) 1233–1295.

[13] F. Cacace, F. Conte, A. Germani, Memoryless approach to the LQ and LQG problems with variable input delay, Automatic Control, IEEE Transactions on 61 (1) (2016) 216–221.

[14] F. Cacace, F. Conte, A. Germani, Filtering continuous-time linear systems with time-varying measurement delay, Automatic Control, IEEE Transactions on 60 (5) (2015) 1368–1373.

[15] K. Kastella, Finite difference methods for nonlinear filtering and automatic target recognition, Multitarget-multisensor tracking: Applications and advances. 3 (2000) 233–258.

[16] C. Floris, Numeric solution of the Fokker-Planck-Kolmogorov equation, Engineering 5 (12) (2013) 975.

[17] W. Bangerth, R. Rannacher, Adaptive finite element methods for differential equations, Birkhäuser, 2013.

[18] Y. Xu, P. Vedula, A quadrature-based method of moments for nonlinear filtering, Automatica 45 (5) (2009) 1291–1298.

[19] F. Daum, M. Krichman, Meshfree adjoint methods for nonlinear filtering, in: Aerospace Conference, 2006 IEEE, IEEE, 2006, pp. 16–pp.

[20] R. Beard, J. Kenney, J. Gunther, J. Lawton, W. Stirling, Nonlinear projection filter based on Galerkin approximation, Journal of guidance, control, and dynamics 22 (2) (1999) 258–266.

[21] A. Germani, M. Piccioni, Finite-dimensional approximations for the equation of nonlinear filtering derived in mild form, Applied mathematics and Optimization 16 (1) (1987) 51–72.

[22] A. Budhiraja, L. Chen, C. Lee, A survey of numerical methods for nonlinear filtering problems, Physica D: Nonlinear Phenomena 230 (1) (2007) 27–36.

[23] F. Campillo, F. Cerou, F. Le Gland, R. Rakotozafy, Particle and cell approximations for nonlinear filtering, Rapports de recherche- INRIA.

[24] P.-A. Raviart, An analysis of particle methods, Springer, 1985.

[25] C. Berzuini, N. G. Best, W. R. Gilks, C. Larizza, Dynamic conditional independence models and Markov chain Monte Carlo methods, Journal of the American Statistical Association 92 (440) (1997) 1403–1412.

[26] A. Germani, C. Manes, P. Palumbo, Filtering of stochastic differential systems via a Carleman approximation approach., IEEE Trans. on Automatic Control 52 (2007) 2166–2172.

[27] K. Kowalski, W. Steeb, Dynamical systems and Carleman linearization, World Scientific, 1991.

[28] F. Cacace, V. Cusimano, A. Germani, An efficient approach to the design of observers for continuous-time systems with discrete-time measurements, in: Proceedings of the 52nd IEEE Conference on Decision and Control, 2011, pp. 7549–7554.

[29] F. Carravetta, A. Germani, M. Raimondi, M. Shuakayev, A new suboptimal approach to the filtering problem for bilinear stochastic differential systems, SIAM Journal on Control and Optimization 38 (4) (2000) 1171–1203.

[30] Y. Ait-Sahalia, R. Kimmel, Estimating affine multifactor term structure models using closed-form likelihood expansions, Journal of Financial Economics 98 (2010) 113–144.

[31] D. Duffie, Dynamic Asset Pricing Theory, Princeton University Press, 2001.

[32] D. Duffie, D. Filipovic, W. Schachermayer, Affine processes and applications in finance. the annals of applied probability, The Annals of Applied Probability 13 (2003) 984–1053.

[33] W. Feller, Two singular diffusion problems, Annals of Mathematics 54 (1951) 173–182.

[34] J. Cox, J. J. Ingersoll, S. A. Ross, A theory of the term structure of interest rates, Econometrica 53 (1985) 385–408.

21

[35] F. Carravetta, A. Germani, M. Raimondi, Polynomial filtering for linear discrete time non-gaussian systems, SIAM Control and Optimization 34 (5) (1996) 1666–1690.