Optimal linear filter for a class of nonlinear stochastic differential systems with discrete measurements

F. Cacace, V. Cusimano, A. Germani, P. Palumbo, and M. Papi

Abstract— **Continuous-discrete models refer to systems described by continuous ordinary or stochastic differential equations, with measurements acquired at discrete sampling instants. Here we investigate the state estimation problem in the stochastic framework, for a class of nonlinear systems characterized by a linear drift and a generic nonlinear diffusion term. Motivation stems from a large variety of applications, ranging from systems biology to finance. By using a Carleman linearization approach we show how the original system can be embedded into an infinite dimensional bilinear system, for which it is possible to write the equations of the optimal linear filter, in case of measurements provided by linear state transformations. A finite dimensional approximation of the optimal linear filter is finally derived. Results are applied to a case of interest in financial applications.**

I. INTRODUCTION

This work investigates the state estimation problem for continuous differential systems endowed with discrete sampled measurements, a filtering framework easy to find in tracking [34], finance [29] and systems biology [13], that is receiving growing attention in recent years [17], [24], [28]. Different approaches can be found in the literature, including adaptations of the Extended Kalman Filter (EKF), like the continuous-discrete extended Kalman filter [21], [26], unscented Kalman filter, [31], and cubature Kalman filter, [2]. When exploiting the EKF, the main difficulty is the evolution of the moments, specifically the expectation of the state and the error covariance, across the discretization interval. For this reason, many techniques have been proposed to overcome this problem in the context of both the extended and unscented Kalman filters [17], [24], [28], [33] and a comparison can be found, for example, in [17].

On the other hand, one can search for the filter solution by properly exploiting the conditional expectation. However, this approach requires the solution of the Kolmogorov forward equation, which is not provided analytically except for very special cases; instead, different numerical methods have been proposed to this aim, including finite-difference method [22], finite elements [16], quadrature-based methods [35],

F. Cacace and M. Papi are with Università Campus Bio-Medico di Roma, Via Álvaro del Portillo, 21, 00128, Rome, Italy, (f.cacace@unicampus.it; m.papi@unicampus.it), A. Germani is with Department of Information Engineering, Computer Science and Mathematics, University of L'Aquila, Via Vetoio, 67100 Coppito (L'Aquila), Italy. (e-mail: alfredo.germani@univaq.it), V. Cusimano and P. Palumbo are with CNR-IASI, Italian National Research Council - Institute for Systems Analysis and Computer Science "A. Ruberti", Via dei Taurini 19, 00185, Rome, Italy (e-mail: valerio.cusimano@iasi.cnr.it; pasquale.palumbo@iasi.cnr.it)

Galerkin's method [3], [20], particle methods [6], [9], [30] and Markov chain Monte Carlo methods [4]. Unfortunately, the computational complexity of solving the Kolmogorov equation increases exponentially with the dimension of the state vector, thus preventing an effective real-time implementation even for systems of moderate size.

In this note, the stochastic differential system under investigation is described by a linear drift and a generic nonlinear diffusion term, with sampled measurements provided by a linear output transformation, and affected by an additive Gaussian noise. To address the filtering issue, we embed the nonlinear finite-dimensional system into a bilinear infinite-dimensional system for which the optimal linear filter is achieved. To this end the Carleman embedding technique for stochastic systems is exploited: the idea of transforming a nonlinear system in a bilinear one (with respect to the deterministic input) has been pursued, among others, in [25], [27]. A central point of the proposed algorithm involves the computation of the moment equations, a problem studied in [32], whilst the use of the Carleman embedding (related to different frameworks) can be found both for stochastic [19], [18], [8] and deterministic systems [1], [7], [23].

A finite-dimensional implementation of the infinite-dimensional optimal linear filter is as well provided by the truncation of the higher order terms in the Carleman expansion.

The paper is organized as follows. In Section II we introduce the Carleman embedding for the class of stochastic nonlinear systems under investigation. The resulting optimal linear filter is developed in Section III, with the finite-dimensional approximation detailed in subsection III-A. Section IV discusses an application in the mathematical finance field.

Notation. Throughout the paper we use boldface notation to denote infinite-dimensional vectors and matrices. For any two matrices A, B of any dimension, $A \otimes B$ denotes the Kronecker product between A and B, and $A^{[n]}$ is the n-th order Kronecker power of A, recursively defined by $A^{[n]} = A \otimes A^{[n-1]}$ and $A^{[0]} = 1$. The stack operator of a matrix A is the vector that piles up all its entries and is denoted by $\mathrm{st}(A)$. The inverse operation is denoted by $\mathrm{st}_{r,c}^{-1}$ and transform a vector of $r \cdot c$ entries into an $r \times c$ matrix.

II. CARLEMAN EMBEDDING

Consider the following nonlinear stochastic differential system in the Itô formulation endowed with a linear discrete-

time measurement equation

$$dx_t = Ax_tdt + u_tdt + \sum_{l=1}^{p} g^l(x_t)dW_t^l, \qquad t \geq 0 \qquad (1)$$

$$Y_k = C_kx_{k\Delta} + \gamma_k + D_kN_k, \qquad\qquad k \in \mathbb{N} \qquad (2)$$

defined on a filtered probability space $(\Omega, \{\mathcal{F}_t\}, P)$, where $x_t \in \mathbb{R}^n$ is the state vector, $\{W_t, \ t \geq 0\}$ is an $\mathcal{F}_t$-adapted $\mathbb{R}^p$-valued standard Wiener process, $Y_k \in \mathbb{R}^q$ are the sampled measurements, acquired at sampling interval Δ, $u_t \in \mathbb{R}^n$ and $\gamma_k \in \mathbb{R}^q$ are known deterministic biases and $\{N_k \in \mathbb{R}^q, \ k \in \mathbb{N}\}$ is a sequence of zero-mean independent gaussian vectors, with identity covariance matrix, independent of W_t. $g^l : \mathbb{R}^n \to \mathbb{R}^n$ are Lipschitz continuous analytic maps, and A, C_k, D_k are matrices in $\mathbb{R}^{n \times n}$, $\mathbb{R}^{q \times n}$, $\mathbb{R}^{q \times q}$, respectively. The initial state x_0 is a random variable independent of the state noise W_t, with finite moments of any order. In order to avoid singular filtering problems, see [5], we use the standard assumption that $\mathrm{rank}(D_kD_k^T) = q$. These assumptions ensure that the solution to (1) is unique and exists for all $t \geq 0$.

Denoting $X_k = x_{k\Delta}$, the exact discretization applied to (1) is described by the equation:

$$X_{k+1} = e^{A\Delta}X_k + U_k + Z_k, \qquad k \in \mathbb{N}, \qquad (3)$$

with

$$U_k = \int_{k\Delta}^{(k+1)\Delta} e^{A((k+1)\Delta - \tau)} u_\tau d\tau, \qquad (4)$$

$$Z_k = \sum_{l=1}^{p} \int_{k\Delta}^{(k+1)\Delta} e^{A((k+1)\Delta - \tau)} g^l(x_\tau)dW_\tau^l. \qquad (5)$$

The noise affecting the discretized system (3) $\{Z_k, \ k \in \mathbb{N}\}$ is a white sequence of zero-mean random vectors, whose covariance (denoted in the following by $\Psi_{Z,k}$) computation requires the knowledge of the probability density function for $x_t, t \in \mathcal{I}_k = [k\Delta, (k+1)\Delta]$. This lack of knowledge prevents the straightforward application of the optimal linear filter equations to (3)-(5) endowed with the output linear equation (2). Such information is recovered by the knowledge of the moments of x_t by exploiting a problem reformulation in an infinite-dimensional setting: this task is achieved by means of the Carleman embedding, which suitably exploits the Taylor expansion of the nonlinear functions $g^l(x_\tau)$ in the neighborhood of the actual estimate $\widehat{X}_k$. To this end, consider the displacement

$$\varepsilon_{k,t} = x_t - \widehat{X}_k, \qquad t \in \mathcal{I}_k. \qquad (6)$$

and the infinite-dimensional vector $\boldsymbol{\Theta}_{k,t}$ composed by the aggregate of the Kronecker powers of the displacement $\varepsilon_{k,t}$, namely $\Theta_{k,t}^i = \varepsilon_{k,t}^{[i]} \in \mathbb{R}^{n^i}$, $i = 1, 2, \dots$ (see the Appendix of [10] for an exhaustive description of the Kronecker product and of its main properties). According to a little abuse of notation, when $t = k\Delta$ or $t = (k+1)\Delta$ then $\boldsymbol{\Theta}_{k,k\Delta}$ and $\boldsymbol{\Theta}_{k,(k+1)\Delta}$ will be denoted by $\boldsymbol{\Theta}_{k,k}$ and $\boldsymbol{\Theta}_{k,k+1}$ for short.

Since $d\varepsilon_{k,t} = dx_t$, by properly exploiting the Taylor expansion of $g^l(x_t)$ around $\widehat{X}_k$, we get

$$d\Theta_{k,t}^1 = (A\Theta_{k,t}^1 + A\widehat{X}_k + u_t)dt + \sum_{l=0}^{p}\sum_{j=0}^{\infty} G_j^l(\widehat{X}_k)\Theta_{k,t}^j dW_t^l, \qquad (7)$$

where $G_j^l(\widehat{X}_k)$ stand for the Taylor expansion coefficients, see e.g. [19] for the details. Higher order differentials of the type $d\Theta_{k,t}^i$ are computed by properly exploiting Theorem 5.2 in [11], so that by using the expression of first and second order differentials of the Kronecker powers, [19], and the properties of Kronecker product in [8], we obtain that $\boldsymbol{\Theta}_{k,t}$ is the solution for $t \in \mathcal{I}_k$ of the following infinite-dimensional bilinear stochastic system (linear drift and multiplicative noise)

$$d\boldsymbol{\Theta}_{k,t} = \mathbf{A}_{k,t}\boldsymbol{\Theta}_{k,t}dt + \mathbf{L}_{k,t}dt + \sum_{l=1}^{p} \left(\mathbf{B}_k^l\boldsymbol{\Theta}_{k,t} + \mathbf{F}_k^l\right)dW_t^l, \qquad (8)$$

with initial condition

$$\boldsymbol{\Theta}_{k,k} = \begin{bmatrix} X_k - \widehat{X}_k \\ (X_k - \widehat{X}_k)^{[2]} \\ \vdots \end{bmatrix}, \qquad (9)$$

where the infinite-dimensional matrices $\mathbf{A}_{k,t}, \mathbf{L}_{k,t}, \mathbf{B}_k^l, \mathbf{F}_k^l$ (the index k recalls that they depend on $\widehat{X}_k$), have the structure reported in Appendix.

Notice that the knowledge of the conditional expectation $\mathbb{E}[\boldsymbol{\Theta}_{k,t}|\mathcal{F}_{k\Delta}]$, $t \in \mathcal{I}_k$ is equivalent to the knowledge of the conditional moments of any order of $\varepsilon_{k,t}$ and, hence, to the conditional probability distribution of x_t.

Let us denote these moments with $\overline{\Theta}_{k,t}^i = \mathbb{E}[\Theta_{k,t}^i|\widehat{X}_k]$. From (8), we have that the vector $\overline{\boldsymbol{\Theta}}_{k,t}$ (assembled by properly aggregating the blocks $\overline{\Theta}_{k,t}^i$ $i = 1, 2, \dots$) obeys the following infinite-dimensional linear equation,

$$\dot{\overline{\boldsymbol{\Theta}}}_{k,t} = \mathbf{A}_{k,t}\overline{\boldsymbol{\Theta}}_{k,t} + \mathbf{L}_{k,t}, \qquad t \in \mathcal{I}_k, \qquad (10)$$

$$\overline{\boldsymbol{\Theta}}_{k,k} = \mathbb{E}[\boldsymbol{\Theta}_{k,k}|\widehat{X}_k], \qquad\qquad (11)$$

It will be shown in the next Section that the integration of (10)-(11) is instrumental for the computation of the covariance matrix $\Psi_{Z,k}$ related to Z_k in (5).

III. THE OPTIMAL LINEAR FILTER

The optimal linear filter refers to the discrete system (3), (2) with the initial estimate $\widehat{X}_0 = \mathbb{E}[X_0]$ and conditional error covariance matrix $P(0) = \mathbb{E}[(X_0 - \widehat{X}_0)(X_0 - \widehat{X}_0)^T|\widehat{X}_0] = \Psi_{X_0}$, where Ψ_{X_0} is the covariance matrix of the initial state X_0. $\mathbb{E}[X_0]$ and Ψ_{X_0} are supposed to be known.

Because of the linear drift, the state prediction is:

$$\widehat{X}_{k+1|k} = e^{A\Delta}\widehat{X}_k + U_k, \qquad (12)$$

with the prediction error covariance matrix $P(k+1|k)$

$$P(k+1|k) = e^{A\Delta}P(k)e^{A^T\Delta} + \Psi_{Z,k}. \qquad (13)$$

The computation of $\Psi_{Z,k}$ will be later addressed. The Kalman gain $K(k+1)$ and the updated estimation error covariance $P(k+1)$ are:

$$
\begin{aligned}
K(k+1) =&P(k+1|k)C_{k+1}^T \\
&\left(C_{k+1}P(k+1|k)C_{k+1}^T + D_{k+1}D_{k+1}^T\right)^{-1},
\end{aligned}
$$
(14)

$$
P(k+1) =\Xi_{k+1}P(k+1|k), \qquad \Xi_k = I_n - K(k)C_k
$$
(15)

with the updated estimate $\widehat{X}_{k+1}$,

$$
\widehat{X}_{k+1} = \widehat{X}_{k+1|k}+K(k+1)\left(Y_{k+1} - C_{k+1}\widehat{X}_{k+1|k} - \gamma_{k+1}\right).
$$
(16)

The Kalman filter (12)–(16) is actually infinite-dimensional, since the computation of $\Psi_{Z,k}$ requires all the moments of the error $\Theta_{k,k}^1 = x_{k\Delta} - \widehat{X}_k$. To this end, we need to propagate $\overline{\Theta}_{k,k}$ to $\overline{\Theta}_{k+1,k+1}$. For these reasons it is essential to prove that all the moments can be computed, though according to a possibly infinite dimensional procedure. We prove this property constructively through a sequence of steps.

Step 1. *Computation of* $\overline{\Theta}_{k,k+1}$: It is obtained by integrating (10) in $\mathcal{I}_k$ with initial condition (11).

Step 2. *Computation of the moments of* Z_k: Denote $\zeta_k^i = \mathbb{E}[Z_k^{[i]}|\widehat{X}_k] \in \mathbb{R}^{n^i}$ and compute X_{k+1} from (6)

$$
X_{k+1} = \widehat{X}_k + \Pi_n\boldsymbol{\Theta}_{k,k+1} = \widehat{X}_k + \Theta_{k,k+1}^1,
$$
(17)

where Π_n denotes the projector onto $\mathbb{R}^n$. Consequently,

$$
\mathbb{E}[X_{k+1}^{[i]}|\widehat{X}_k] = \sum_{j=0}^{i} M_j^i\left(\widehat{X}_k^{[j]} \otimes \overline{\Theta}_{k,k+1}^{i-j}\right)
$$
(18)

where M_j^i, $j = 1,\ldots,i-1$ are the coefficients for the binomial expansion of an i-th order Kronecker power (see e.g. [10]). On the other hand, from (3) and the fact that X_k and Z_k are independent random vectors it follows

$$
\begin{aligned}
\mathbb{E}[X_{k+1}^{[i]}|\widehat{X}_k] =&\zeta_k^i + \sum_{j_1+j_2+j_3=i,\ j_3<i} M_{j_1,j_2,j_3}^i \\
&\cdot \left(\left((e^{A\Delta})^{[j_1]}\mathbb{E}[X_k^{[j_1]}|\widehat{X}_k]\right) \otimes U_k^{[j_2]} \otimes \zeta_k^{j_3}\right)
\end{aligned}
$$
(19)

where M_{j_1,j_2,j_3}^i are the coefficients for the trinomial expansion of an i-th order Kronecker power (see e.g. [10]) and $\mathbb{E}[X_k^{[j_1]}|\widehat{X}_k]$ can be computed from (6) at $t = k\Delta$ as

$$
\mathbb{E}[X_k^{[j_1]}|\widehat{X}_k] = \sum_{l=0}^{j_1} M_l^{j_1}\left(\widehat{X}_k^{[l]} \otimes \overline{\Theta}_{k,k}^{j_1-l}\right).
$$
(20)

Therefore, by comparing (18) to (19) we obtain the i-th order moment of Z_k, i.e. ζ_k^i, as a function of its lower order moments $\zeta_k^{j_3}$, $j_3 < i$.

Step 3. *Computation of* $\Psi_{Z,k}$: $\Psi_{Z,k}$ can be computed from ζ_k^2 by exploiting the column stack property of the Kronecker product (see [10] for the definition of the stack operator and of its inverse):

$$
\Psi_{Z,k} = \mathrm{st}_{n,n}^{-1}(\zeta_k^2).
$$
(21)

According to the previous **Step 2**, this computation requires the final-time solution (first $n+n^2$ components) of (10)-(11). That is, by denoting with $\boldsymbol{\Pi}_n^2$ the projector onto $\mathbb{R}^{n+n^2}$ we obtain from (10)

$$
\begin{aligned}
\begin{bmatrix} \overline{\Theta}_{k,k+1}^1 \\ \overline{\Theta}_{k,k+1}^2 \end{bmatrix} =&\boldsymbol{\Pi}_n^2\boldsymbol{\Phi}((k+1)\Delta, k\Delta)\overline{\boldsymbol{\Theta}}_{k,k} \\
&+ \int_{k\Delta}^{(k+1)\Delta} \boldsymbol{\Pi}_n^2\boldsymbol{\Phi}((k+1)\Delta, \tau)\mathbf{L}_{k,\tau}d\tau,
\end{aligned}
$$
(22)

where $\boldsymbol{\Phi}(t,\tau)$ is the semigroup generated by $\mathbf{A}_{k,t}$. In case of piecewise constant inputs $u_t = u_k$, $\forall t \in \mathcal{I}_k$, then we can denote $\mathbf{A}_k = \mathbf{A}_{k,t}$ and $\mathbf{L}_k = \mathbf{L}_{k,t}$, $\forall t \in \mathcal{I}_k$, and (22) simplifies into

$$
\begin{aligned}
\begin{bmatrix} \overline{\Theta}_{k,k+1}^1 \\ \overline{\Theta}_{k,k+1}^2 \end{bmatrix} =&\boldsymbol{\Pi}_n^2 e^{\mathbf{A}_k\Delta}\overline{\boldsymbol{\Theta}}_{k,k}+\int_{k\Delta}^{(k+1)\Delta} \boldsymbol{\Pi}_n^2 e^{\mathbf{A}_k((k+1)\Delta-\tau)}d\tau\mathbf{L}_k \\
=&\sum_{i=0}^{\infty} \boldsymbol{\Pi}_n^2\mathbf{A}_k^i\overline{\boldsymbol{\Theta}}_{k,k}\frac{\Delta^i}{i!} + \sum_{i=1}^{\infty} \boldsymbol{\Pi}_n^2\mathbf{A}_k^{i-1}\mathbf{L}_k\frac{\Delta^i}{i!}
\end{aligned}
$$
(23)

The two sums in (23) can computed according to a different level of accuracy. Indeed, each term of the first sum is made of infinite sums itself, whilst the terms of the second sum are made each of finite sums: this is because of the sparsity of the infinite-dimensional matrix $\mathbf{L}_k$ (see also [8]).

Step 4. *Computation of* $\overline{\boldsymbol{\Theta}}_{k+1,k+1}$: In order to have a recursively computable estimate we need $\overline{\boldsymbol{\Theta}}_{k,k}$ to be available at each step. This can be achieved by expressing $\overline{\boldsymbol{\Theta}}_{k+1,k+1}$, that is, all the moments of the estimation error of the updated estimate, as a function of $\overline{\boldsymbol{\Theta}}_{k,k}$, the moments ζ_k^i of Z_k and the measurement noise. To this end we represent the estimation error as

$$
\begin{aligned}
\Theta_{k+1,k+1}^1 = \Xi_{k+1}e^{A\Delta}\Theta_{k,k}^1 + \Xi_{k+1}Z_k \\
- K(k+1)C_{k+1}D_{k+1}N_{k+1},
\end{aligned}
$$
(24)

therefore

$$
\begin{aligned}
\overline{\Theta}_{k+1,k+1}^i =&\sum_{m_1+m_2+m_3=i} (-1)^{m_3}M_{m_1,m_2,m_3}^i \\
&\cdot \left(\left((\Xi_{k+1}e^{A\Delta})^{[m_1]}\overline{\Theta}_{k,k}^{m_1}\right) \otimes (\Xi_{k+1}^{[m_2]}\zeta_k^{m_2})\right. \\
&\left.\otimes \left((K(k+1)C_{k+1}D_{k+1})^{[m_3]}\mathbb{E}[N_{k+1}^{[m_3]}]\right)\right),
\end{aligned}
$$
(25)

Eq. (25) benefits of the fact that random variables $\Theta_{k,k}^1$, Z_k and N_{k+1} are mutually independent.

A. Optimal linear filter: finite-dimensional approximation

Here we provide the ν-degree finite-dimensional approximation of the optimal linear filter, required for the computation of $\overline{\Theta}_{k,k+1}^1$ and $\overline{\Theta}_{k,k+1}^2$ in (22). Define the finite-dimensional vector $\overline{\boldsymbol{\Theta}}_{k,t}^{\nu} = \left[\overline{\Theta}_{k,t}^{\nu,1\,T} \cdots \overline{\Theta}_{k,t}^{\nu,\nu\,T}\right]^T$, such that:

$$\dot{\overline{\boldsymbol{\Theta}}}_{k,t}^{\nu} = \mathbf{A}_{k,t}^{\nu}\overline{\boldsymbol{\Theta}}_{k,t}^{\nu} + \mathbf{L}_{k,t}^{\nu}, \qquad t \in \mathcal{I}_k, \qquad (26)$$

with initial condition $\overline{\boldsymbol{\Theta}}_{k,k}^{\nu}$ inherited from the previous step of the filtering algorithm. Finite-dimensional matrices $\mathbf{A}_{k,t}^{\nu}$, $\mathbf{L}_{k,t}^{\nu}$ are easily obtained from (40)-(41) by properly neglecting the blocks $(\mathbf{A}_{k,t})_{ij}$ and $(\mathbf{L}_{k,t})_i$ with $i,j > \nu$. Hence, the implementable (approximate) version of the optimal linear filtering algorithm is below resumed:

0) The initialization requires the moments of the initial states up to degree ν to compute $\widehat{X}_{\nu,0} = \mathbb{E}[x_0]$, $P^{\nu}(0) = \mathrm{Cov}(x_0)$ and $\Theta_{0,0}^{\nu,i} = \mathbb{E}[(x_0 - \widehat{X}_{\nu,0})^{[i]}]$, $i = 1, \ldots, \nu$.
 Given $\widehat{X}_{\nu,k}$, $P^{\nu}(k)$ and $\overline{\boldsymbol{\Theta}}_{k,k}^{\nu}$ at time $t = k\Delta$:

1) Compute the state prediction $\widehat{X}_{\nu,k+1|k}$ by means of

$$\widehat{X}_{\nu,k+1|k} = e^{A\Delta}\widehat{X}_{\nu,k} + U_k. \qquad (27)$$

2) Compute $\overline{\boldsymbol{\Theta}}_{k,k+1}^{\nu}$ as the final-value ($t = (k+1)\Delta$) solution of (26).

3) Compute $\widehat{\Psi}_{Z,k}^{\nu} = \mathrm{st}_{n,n}^{-1}(\zeta_k^{\nu,2})$, with $\zeta_k^{\nu,2}$ achieved by properly exploiting the first $n + n^2$ components of $\overline{\boldsymbol{\Theta}}_{k,k+1}^{\nu}$

4) Compute $P^{\nu}(k+1|k)$ from

$$P^{\nu}(k+1|k) = e^{A\Delta}P^{\nu}(k)e^{A^T\Delta} + \widehat{\Psi}_{Z,k}^{\nu} \qquad (28)$$

5) Compute the gain matrix $K^{\nu}(k+1)$ from

$$\begin{aligned} K^{\nu}(k+1) =& P^{\nu}(k+1|k)C_{k+1}^T \\ & \cdot \left(C_{k+1}P^{\nu}(k+1|k)C_{k+1}^T + D_{k+1}D_{k+1}^T\right)^{-1} \end{aligned} \qquad (29)$$

6) Update the state estimate $\widehat{X}_{\nu,k+1}$ from

$$\begin{aligned} \widehat{X}_{\nu,k+1} =& \widehat{X}_{\nu,k+1|k} + K^{\nu}(k+1) \\ & \cdot \left(Y_{k+1} - C_{k+1}\widehat{X}_{\nu,k+1|k} - \gamma_{k+1}\right) \end{aligned} \qquad (30)$$

7) Update the error covariance matrix $P^{\nu}(k+1)$ from

$$P^{\nu}(k+1) = \Xi_{\nu,k+1}P^{\nu}(k+1|k), \qquad (31)$$
$$\text{with} \qquad \Xi_{\nu,k} = I_n - K^{\nu}(k)C_k$$

8) Iteratively compute $\zeta_k^{\nu,i}$, $i = 3, \ldots, \nu$ by properly exploiting all the components of $\overline{\boldsymbol{\Theta}}_{k,k+1}^{\nu}$.

9) Update the $\overline{\Theta}_{k+1,k+1}^{\nu,i}$, $i = 1, \ldots, \nu$:

IV. MATHEMATICAL FINANCE APPLICATION

To demonstrate the good performances of the proposed algorithm, we decided to focus our attention on a classical problem in the mathematical finance described by the following stochastic differential system

$$dX_t = (V - KX_t)dt + g(X_t)dW_t, \qquad (32)$$

where the state vector $X_t = [r_t, \lambda_t]^T \in \mathbb{R}^2$ represents the short-rate process r and the intensity of default process λ with

$$V = \begin{bmatrix} k_r\theta_r \\ k_\lambda\theta_\lambda \end{bmatrix}, \qquad K = \begin{bmatrix} k_r & 0 \\ 0 & k_\lambda \end{bmatrix}, \qquad (33)$$

$$g(X) = \begin{bmatrix} c_r\sqrt{X_1} & 0 \\ 0 & c_\lambda\sqrt{X_2} \end{bmatrix}, \qquad (34)$$

where k_r, k_λ, θ_r, θ_λ, c_r and c_λ are positive parameters and $W_t^T = [W_t^r \; W_t^\lambda]$ is a two-dimensional Brownian motion.

The goal is to estimate X_t by using the observed zero rates $Y_t \in \mathbb{R}$ associated to a zero-coupon bond price with respect to the date maturity T:

$$Y_t = F_t + H_tX_t + N_t, \qquad t = k\Delta, \qquad k = 1, \ldots, M \qquad (35)$$

with

$$F_t = -\left(\frac{\log[A_r] + \log[A_\lambda]}{(T-t)}\right), \qquad (36)$$

$$(H_t)_1 = \frac{B_r}{(T-t)}, \qquad (H_t)_2 = \frac{B_\lambda}{(T-t)}, \qquad (37)$$

and

$$\begin{aligned} A_h &= \left[\frac{2\gamma e^{(k_h+\gamma_h)(T-t)/2}}{(k_h+\gamma_h)(e^{\gamma_h(T-t)}-1)+2\gamma_h}\right]^{2k_h\theta_h/c_h^2} \\ B_h &= \frac{2\left(e^{\gamma_h(T-t)}-1\right)}{(k_h+\gamma_h)(e^{\gamma_h(T-t)}-1)+2\gamma_h} \end{aligned} \qquad (38)$$

where $\gamma_h = \sqrt{k_h^2 + 2c_h^2}$, $h = r, \lambda$ (see [12], [14] for more details). Note that if the measurement is acquired at very distant times from the value of maturity, the informative contribution is less than that in the case of nearest maturity.

Given a generic process h, the coefficients k_h and θ_h are the speed of adjustment and the long run mean of the process, respectively, while $c_h\sqrt{h_t}$ is the standard deviation factor. When the process is close to zero, also the rate becomes very small, and the drift factor becomes dominate for the system evolution. Moreover if the Feller condition is satisfied ($2k_h\theta_h \geq c^2$ [15]), the process is strictly positive. In Table 1, the parameter values are reported.

TABLE I

MODEL PARAMETERS

k_r	k_λ	θ_r	θ_λ	c_r	c_λ
0.50239	0.15	$6 \cdot 10^{-3}$	$1 \cdot 10^{-3}$	$5 \cdot 10^{-3}$	$4 \cdot 10^{-2}$

The resulting system (32)-(35) is a general formulation of the problem of estimating continuous-time bond pricing models from bond prices with correlated pricing errors.

To evaluate the performance of the algorithm, the sampling interval Δ varies form 0.2 to 1.2 with step of 0.2. For each Δ, a simulation final time $t_f = M \cdot \Delta$, with M the number of measurements samples, and a maturity $T = \alpha \cdot t_f$, with $\alpha = 4$, 10, are defined.

Moreover, for each triplet (Δ, t_f, α), N random realizations were run and the performance was evaluated by using

the Mean Square Error (MSE):

$$\text{MSE} = \frac{1}{N} \sum_{\nu=1}^{N} \frac{1}{M} \sum_{k=1}^{M} \sqrt{||X_k - \widehat{X}_k||^2} \qquad (39)$$

where X_k and $\widehat{X}_k$ are the the real and estimated values at time $t = k\Delta$, respectively.

The variance of the measurement noise is chosen equal to $1 \cdot 10^{-4}$ for $\alpha = 4$ and equal to $1 \cdot 10^{-6}$ for $\alpha = 10$, and the filtering algorithm has been tested with a Carleman approximation order equal to $\nu = 4$.

As an example, Figure 1 shows the evolution of the real and estimated systems for a generic realization.
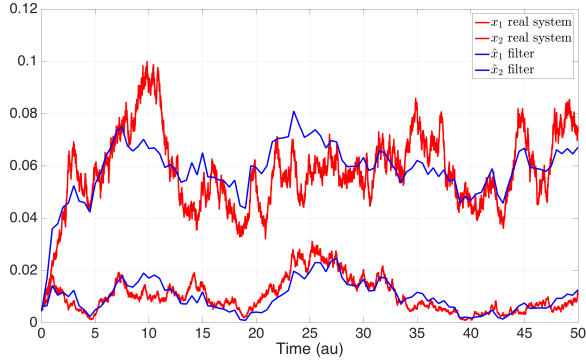

Fig. 1.   Evolution of real system x and filter algorithms $\hat{x}$

In Table 2 and Figure 2, the numerical results for different values of Δ and α are reported, with $N = 100$ and $M = 150$.

TABLE II

MSE FOR DIFFERENT Δ

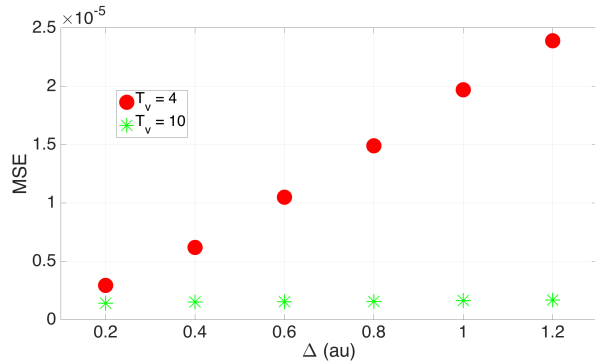| | | Δ | |
		0.2	0.4	0.6	0.8	1.0	1.2	
α	4	2.93	6.19	10.5	14.9	19.7	23.9	10^{-6}
	10	1.42	1.52	1.53	1.56	1.63	1.67	

Fig. 2. Δ vs MSE

In particular in Table 2 it can be noted that the algorithm works better for smaller intervals, differently from the other case where the performances are substantially constant (it can

be better appreciated in Figure 2). This is a consequence of the maturity value more than the filter.

V. CONCLUDING REMARKS AND FUTURE WORK

In summary, in this paper we present an optimal linear filter for Continuous-Discrete (CD) models. This task is achieved by using a Carleman linearization approach through which the original stochastic nonlinear system is embedded into an infinite dimensional bilinear system, for which it is possible to write the equations of the optimal linear filter.

The resulting algorithm has the features:

- to evaluate explicitly, in each sampling interval, the higher order moments of the state evolution up to a designed order. In this way , the obtained representation is more precise than under Gaussian approximations;
- to return a system with a linear structure which allows to use a standard Kalman Filter and the relative implementation advantages;
- that the order of the truncated system is a design parameter that allows a simple trade-off between computational burden and precision. This makes the method well suited when the sampling interval is not negligible.

Future research will involve the study of the existence of the solution for the moment equations in the Carleman approach, i.e. the existence of the solution for the infinite-dimensional system resulting from the linearization procedure, associated to the original nonlinear finite-dimensional one. This is a work in progress by the authors.

APPENDIX

The structure of the infinite-dimensional matrices in (8) is the following.

$$\mathbf{A}_{k,t} = \begin{bmatrix} A & 0_{n \times n^2} & 0_{n \times n^3} & 0_{n \times n^4} & \cdots \\ \overline{A}_{2,1} + \overline{H}_{2,1} & \overline{A}_{2,2} + \overline{H}_{2,2} & \overline{H}_{2,3} & \overline{H}_{2,4} & \cdots \\ \overline{H}_{3,1} & \overline{A}_{3,2} + \overline{H}_{3,2} & \overline{A}_{3,3} + \overline{H}_{3,3} & \overline{H}_{3,4} & \cdots \\ 0_{n^4 \times n} & \overline{H}_{4,2} & \overline{A}_{4,3} + \overline{H}_{4,3} & \overline{A}_{4,4} + \overline{H}_{4,4} & \cdots \\ 0_{n^5 \times n} & 0_{n^5 \times n^2} & \overline{H}_{5,3} & \overline{A}_{5,4} + \overline{H}_{5,4} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \qquad (40)$$

$$\mathbf{L}_{k,t} = \begin{bmatrix} A\widehat{X}_k + u_t \\ \overline{H}_{2,0} \\ 0_{n^3 \times 1} \\ \vdots \end{bmatrix}, \quad \mathbf{B}_k^l = \begin{bmatrix} G_1^l & G_2^l & \cdots \\ \overline{G}_{2,1}^l & \overline{G}_{2,2}^l & \cdots \\ 0_{n^3 \times n} & \overline{G}_{3,2}^l & \cdots \\ 0_{n^4 \times n} & 0_{n^4 \times n^2} & \ddots \\ \vdots & \vdots & \ddots \end{bmatrix}, \quad (41)$$

$$\mathbf{F}_k^l = \begin{bmatrix} G_0^l \\ 0_{n^2 \times 1} \\ 0_{n^3 \times 1} \\ \vdots \end{bmatrix}. \qquad (42)$$

The building blocks $(\mathbf{A}_{k,t})_{ij}, (\mathbf{B}_k^l)_{ij} \in \mathbb{R}^{n^i \times n^j}$,

$(\mathbf{L}_{k,t})_i, (\mathbf{F}_k^l)_i \in \mathbb{R}^{n^i \times 1}$, $i, l = 1, 2, \ldots$ are defined as

$$(\mathbf{A}_{k,t})_{ij} = \begin{cases} A, & \text{if } i = 1, \ j = 1, \\ \overline{H}_{i,i-2}(\widehat{X}_k), & \text{if } j = i - 2, \\ \overline{A}_{i,j}(\widehat{X}_k) + \overline{H}_{i,j}(\widehat{X}_k) & \text{if } i > 1, \ j > i - 2 \\ 0_{n^i \times n^j}, & \text{otherwise} \end{cases} \tag{43}$$

$$(\mathbf{L}_{k,t})_i = \begin{cases} A\widehat{X}_k + u_t, & \text{if } i = 1 \\ \overline{H}_{2,0}(\widehat{X}_k), & \text{if } i = 2 \\ 0_{n^i \times 1}, & \text{otherwise} \end{cases} \tag{44}$$

$$(\mathbf{B}_k^l)_{i,l} = \begin{cases} \overline{G}_{i,l}^j(\widehat{X}_k), & \text{if } l \geq i - 1 \\ 0_{n^i \times n^l}, & \text{otherwise} \end{cases} \tag{45}$$

$$(\mathbf{F}_k^l)_i = \begin{cases} G_0^j(\widehat{X}_k), & i = 1 \\ 0_{n^i \times 1}, & \text{otherwise}, \end{cases} \tag{46}$$

and

$$\overline{H}_{i,j}(\widehat{X}_k) = \frac{1}{2} O_n^i \left(\sum_{l=1}^p H_{j-i+2}^l(\widehat{X}_k) \otimes I_{n^{i-2}} \right), \\ j \geq i - 2 \tag{47}$$

$$H_h^l(\widehat{X}_k) = \sum_{r=0}^h G_r^l(\widehat{X}_k) \otimes G_{h-r}^l(\widehat{X}_k) \tag{48}$$

$$\overline{A}_{i,i-1}(\widehat{X}_k) = U_n^i \left((A\widehat{X}_k + u_t) \otimes I_{n^{i-1}} \right), \tag{49}$$

$$\overline{A}_{i,i}(\widehat{X}_k) = U_n^i \left(A \otimes I_{n^{i-1}} \right), \tag{50}$$

$$\overline{A}_{i,j}(\widehat{X}_k) = 0_{n^i \times n^j}, \\ j > i \tag{51}$$

$$\overline{G}_{i,j}^l(\widehat{X}_k) = U_n^i \left(G_{j-i+1}^l(\widehat{X}_k) \otimes I_{n^{i-1}} \right), \quad j \geq i - 1 \tag{52}$$

with matrices O_n^i and U_n^i defined in [19].

REFERENCES

[1] M.S. Al-Tuwaim, O.D. Crisalle, and S.A. Svoronos. Discretization of nonlinear models using a modified carleman linearization technique. In *Proc. of the American Control Conference*, volume 5, pages 3084–3088, 1998.

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

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

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

[5] R. Bucy and E. Jonckheere. Singular filtering problems. *Syst. Contr. Lett.*, 13:339–344, 1989.

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

[7] F. Cacace, V. Cusimano, and 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*, pages 7549–7554, 2011.

[8] F. Cacace, V. Cusimano, A. Germani, and P. Palumbo. A state predictor for continuous-time stochastic systems. *Systems & Control Letters*, 98:37–43, 2016.

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

[10] F. Carravetta, A. Germani, and M. Raimondi. Polynomial filtering of discrete-time stochastic linear systems with multiplicative state noise. *IEEE Trans. Autom. Control*, 42(8):1106–1126, 1997.

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

[12] J.C. Cox, J.E.Jr. Ingersoll, and S.A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53(2), 1985.

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

[14] D. Duffie. Credit risk modeling with affine processes. *Journal of Banking & Finace*, 29:2751–2802, 2005.

[15] W. Feller. Two singular diffusion problems. *Annals of Mathematics*, 54(1):173–182, 1951.

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

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

[18] A. Germani, C. Manes, and P. Palumbo. Polynomial extended Kalman filter. *IEEE Trans. Autom. Control*, 50:2059–2064, 2005.

[19] A. Germani, C. Manes, and P. Palumbo. Filtering of stochastic differential systems via a Carleman approximation approach. *IEEE Trans. Autom. Control*, 52:2166–2172, 2007.

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

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

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

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

[24] G.Y. Kulikov and M.V. Kulikova. Accurate numerical implementation of the continuous-discrete extended Kalman filter. *IEEE Trans. Autom. Control*, 59(1):273–279, 2014.

[25] C.-H. Lee and J.-N. Juang. Nonlinear system identificationa continuous-time bilinear state space approach. *The Journal of the Astronautical Sciences*, 59(1-2):398–420, 2012.

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

[27] J. T.-H. Lo. Global bilinearization of systems with control appearing linearly. *SIAM Journal on Control*, 13(4):879–885, 1975.

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

[29] B. Øksendal. *Stochastic differential equations*. Springer, 2003.

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

[31] S. Särkkä. On unscented Kalman filtering for state estimation of continuous-time nonlinear systems. *IEEE Trans. Autom. Control*, 52(9):1631–1641, 2007.

[32] W. Shing Wong. Carleman linearization and moment equations of nonlinear stochastic equations. *Stochastics: An International Journal of Probability and Stochastic Processes*, 9(1-2):77–101, 1983.

[33] H. Singer. Continuous-discrete unscented kalman filtering. 2006.

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

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