APPROXIMATE LINEAR MINIMUM VARIANCE FILTERS FOR CONTINUOUS-DISCRETE STATE SPACE MODELS: CONVERGENCE AND PRACTICAL ALGORITHMS

JUAN CARLOS JIMENEZ*

Abstract. In this paper, approximate Linear Minimum Variance (LMV) filters for continuous-discrete state space models are introduced. The filters are obtained by means of a recursive approximation to the predictions for the first two moments of the state equation. It is shown that the approximate filters converge to the exact LMV filter when the error between the predictions and their approximations decreases. As particular instance, the order-β Local Linearization filters are presented and expounded in detail. Practical algorithms are also provided and their performance in simulation is illustrated with various examples. The proposed filters are intended for the recurrent practical situation where a nonlinear stochastic system should be identified from a reduced number of partial and noisy observations distant in time.

Key words. system identification, linear minimum variance filters, local linearization filters, stochastic differential equations, numerical integrators

AMS subject classifications. 93E12, 93E11, 65C30

1. Introduction. The estimation of unobserved states of a continuous stochastic dynamical system from noisy discrete observations is of central importance to solve diverse scientific and technological problems. The major contribution to the solution of this estimation problem is due to Kalman and Bucy [30, 31], who provided a sequential and computationally efficient solution to the optimal filtering and prediction problem for linear state space models with additive noise. However, the optimal estimation of nonlinear state space models is still a subject of active researches. Typically, the solution of optimal filtering problems involves the resolution of evolution equations for conditional probabilistic densities, moments or modes, which in general have explicit solutions in few particular cases. Therefore, a variety approximations have been developed. Examples of such approximate nonlinear filters are the classical ones as the Extended Kalman, the Iterated Extended Kalman, the Gaussian and the Modified Gaussian filters [20]; and other relatively recents as the Local Linearization [43, 28], the Projection [3] and the Particle filters [12] methods.

In a variety of practical situations, the solution of the general optimal filtering problem is dispensable since the solution provided by a suboptimal filter is satisfactory. This is the case of the signal filtering and detection problems, the system stabilization, and the parameter estimation of nonlinear systems, among others. Prominent examples of suboptimal filters are the linear, the quadratic and the polynomial one, which have been widely used for the estimation of the state of both, continuous-continuous [34, 38, 47] and discrete-discrete [10, 46, 47, 11, 7] models. In the case of continuous-discrete models, exact expressions for Linear Minimum Variance filter (LMV) have also been derived [27], but they are restricted to linear models. For nonlinear models, this kind of suboptimal filter has in general no exact solution since the first two conditional moments of the state equation has no explicit solution. Therefore, adequate approximations are required in this situation as well.

In this paper, approximate LMV filters for nonlinear continuous-discrete state space models are introduced. The filters are obtained by means of a recursive approximation to the predictions for the first two moments of the state equation. It is shown that the approximate filters converge to the exact LMV filter when the error between the predictions and their approximations decreases. Based on the well-known Local Linear approximations for the state equation, the order-β Local Linearization filters are presented as a particular instance. Their convergence, practical algorithms and performance in simulations are also considered in detail. The simulations show that these Local Linearization filters provide accurate and computationally efficient estimation of the unobserved states of the stochastic systems given a reduced number of partial and noisy observations, which is a typical situation in practical control engineering.

The paper is organized as follows. In section 2, basic notations and results on LMV filters, Local Linear approximations and Local Linearization filters are presented. The general class of approximate LMV filters is introduced in section 3 and its convergence is stated. In section 4, the order-β Local

*Instituto de Cibernética, Matemática y Física, Departamento de Matemática Interdisciplinaria, Calle 15, no. 551, Vedado, La Habana, Cuba (http://sites.google.com/site/locallinearization).

Linearization filters are presented and their convergence analyzed. In the last two sections, practical algorithms for these filters and their performance in simulations are considered.

2. Notations and Preliminaries. Let $(\Omega, \mathcal{F}, P)$ be a complete probability space, and $\{\mathcal{F}_t, t \geq t_0\}$ be an increasing right continuous family of complete sub σ-algebras of $\mathcal{F}$. Consider the state space model defined by the continuous state equation

$$d\mathbf{x}(t) = \mathbf{f}(t, \mathbf{x}(t))dt + \sum_{i=1}^{m} \mathbf{g}_i(t, \mathbf{x}(t))d\mathbf{w}^i(t), \tag{2.1}$$

for all $t \in [t_0, T]$, and the discrete observation equation

$$\mathbf{z}_{t_k} = \mathbf{C}\mathbf{x}(t_k) + \mathbf{e}_{t_k}, \tag{2.2}$$

for all $k = 0, 1, .., M-1$, where $\mathbf{f}$, $\mathbf{g}_i : [t_0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ are functions, $\mathbf{w} = (\mathbf{w}^1, \dots, \mathbf{w}^m)$ is an m-dimensional $\mathcal{F}_t$-adapted standard Wiener process, $\{\mathbf{e}_{t_k} : \mathbf{e}_{t_k} \sim \mathcal{N}(0, \Sigma_{t_k}),\ k = 0, .., M-1\}$ is a sequence of r-dimensional i.i.d. Gaussian random vectors independent of $\mathbf{w}$, Σ_{t_k} an $r \times r$ positive semi-definite matrix, and $\mathbf{C}$ an $r \times d$ matrix. Here, it is assumed that the M time instants t_k define an increasing sequence $\{t\}_M = \{t_k : t_k < t_{k+1},\ t_{M-1} = T,\ k = 0, 1, .., M-1\}$. Conditions for the existence and uniqueness of a strong solution of (2.1) with bounded moments are assumed.

Let $\mathbf{x}_{t/\rho} = E(\mathbf{x}(t)/Z_\rho)$ and $\mathbf{Q}_{t/\rho} = E(\mathbf{x}(t)\mathbf{x}^\intercal(t)/Z_\rho)$ be the first two conditional moments of $\mathbf{x}$ with $\rho \leq t$, where $E(.)$ denotes the mathematical expectation value, and $Z_\rho = \{\mathbf{z}_{t_k} : t_k \leq \rho,\ t_k \in \{t\}_M\}$ is a time series with observations from (2.2). Further, let us denote by

$$\begin{aligned}\mathbf{U}_{t/\rho} &= E((\mathbf{x}(t) - \mathbf{x}_{t/\rho})(\mathbf{x}(t) - \mathbf{x}_{t/\rho})^\intercal/Z_\rho) \\ &= \mathbf{Q}_{t/\rho} - \mathbf{x}_{t/\rho}\mathbf{x}_{t/\rho}^\intercal\end{aligned}$$

the conditional variance of $\mathbf{x}$.

Denote by $\mathcal{C}_P^l(\mathbb{R}^d, \mathbb{R})$ the space of l time continuously differentiable functions $g : \mathbb{R}^d \to \mathbb{R}$ for which g and all its partial derivatives up to order l have polynomial growth.

2.1. Linear minimum variance filtering problem. According to [2, 51, 56, 20] the linear minimum variance filter $\mathbf{x}_{t_{k+1}/t_{k+1}}$ for a state space model with discrete observation equation (2.2) is defined as

$$\mathbf{x}_{t_{k+1}/t_{k+1}} = \mathbf{x}_{t_{k+1}/t_k} + \mathbf{G}_{t_{k+1}}(\mathbf{z}_{t_{k+1}} - \mathbf{C}\mathbf{x}_{t_{k+1}/t_k}),$$

where the filter gain $\mathbf{G}_{t_{k+1}}$ is to be determined so as to minimize the error variance

$$E((\mathbf{x}(t_{k+1}) - \mathbf{x}_{t_{k+1}/t_{k+1}})(\mathbf{x}(t_{k+1}) - \mathbf{x}_{t_{k+1}/t_{k+1}})^\intercal).$$

This yields to the following definition.

Definition 2.1. *The Linear Minimum Variance filter for the state space model (2.1)-(2.2) is defined, between observations, by*

$$\frac{d\mathbf{x}_{t/t}}{dt} = E(\mathbf{f}(t, \mathbf{x})/Z_t) \tag{2.3}$$

$$\begin{aligned}\frac{d\mathbf{U}_{t/t}}{dt} &= E(\mathbf{x}\mathbf{f}^\intercal(t, \mathbf{x})/Z_t) - \mathbf{x}_{t/t}E(\mathbf{f}^\intercal(t, \mathbf{x})/Z_t) + E(\mathbf{f}(t, \mathbf{x})\mathbf{x}^\intercal/Z_t) \\ &\quad - E(\mathbf{f}(t, \mathbf{x})/Z_t)\mathbf{x}_{t/t}^\intercal - \sum_{i=1}^{m} E(\mathbf{g}_i(t, \mathbf{x})\mathbf{g}_i^\intercal(t, \mathbf{x})/Z_t)\end{aligned} \tag{2.4}$$

for all $t \in (t_k, t_{k+1})$, and by

$$\mathbf{x}_{t_{k+1}/t_{k+1}} = \mathbf{x}_{t_{k+1}/t_k} + \mathbf{G}_{t_{k+1}}(\mathbf{z}_{t_{k+1}} - \mathbf{C}\mathbf{x}_{t_{k+1}/t_k}) \tag{2.5}$$

$$\mathbf{U}_{t_{k+1}/t_{k+1}} = \mathbf{U}_{t_{k+1}/t_k} - \mathbf{G}_{t_{k+1}}\mathbf{C}\mathbf{U}_{t_{k+1}/t_k} \tag{2.6}$$

for each observation at t_{k+1}, with filter gain

$$\mathbf{G}_{t_{k+1}} = \mathbf{U}_{t_{k+1}/t_k}\mathbf{C}^{\intercal}(\mathbf{C}\mathbf{U}_{t_{k+1}/t_k}\mathbf{C}^{\intercal} + \Sigma_{t_{k+1}})^{-1} \tag{2.7}$$

for all $t_k, t_{k+1} \in \{t\}_M$. The predictions $\mathbf{x}_{t/t_k}$, $\mathbf{U}_{t/t_k}$ are accomplished, respectively, via expressions (2.3)-(2.4) with initial conditions $\mathbf{x}_{t_k/t_k}$ and $\mathbf{U}_{t_k/t_k}$ for all $t \in (t_k, t_{k+1}]$ and $t_k, t_{k+1} \in \{t\}_M$.

Note that, in continuous-discrete filtering problem, the filters $E(\mathbf{x}(t)/Z_t)$ and $E(\mathbf{x}(t)\mathbf{x}^{\intercal}(t)/Z_t)$ reduce to the predictions $E(\mathbf{x}(t)/Z_{t_k})$ and $E(\mathbf{x}(t)\mathbf{x}^{\intercal}(t)/Z_{t_k})$ for all t between two consecutive observations t_k and t_{k+1}, that is for all $t \in (t_k, t_{k+1})$. This is because there is not more observations between t_k and t_{k+1}. This implies that, in the above definition, $\mathbf{x}_{t_{k+1}-\varepsilon/t_{k+1}-\varepsilon} \equiv \mathbf{x}_{t_{k+1}-\varepsilon/t_k}$ for all $\varepsilon > 0$ and so $\mathbf{x}_{t_{k+1}-\varepsilon/t_{k+1}-\varepsilon}$ tends to $\mathbf{x}_{t_{k+1}/t_k}$ when ε goes to zero.

Clearly, for linear state equation with additive noise, the LMV filter (2.3)-(2.7) reduces to the classical continuous-discrete Kalman filter. For linear state equation with multiplicative noise, explicit formulas for the LMV filter can be found in [27]. In general, since the integro-differential equations (2.3)-(2.4) of the LMV filter have explicit solution for a few simple state equations, approximations to them are needed. In principle, for this type of suboptimal filter, the same conventional approximations to the general optimal minimum variance filter may be used as well. For instance, those for the solution of (2.3)-(2.4) provided by the conventional Extended Kalman, the Iterated Extended Kalman, the Gaussian, the Modified Gaussian and the Local Linearization filters. However, in all these approximations, once the data Z_{t_M} are given on a time partition $\{t\}_M$ the error between the exact and the approximate predictions for the mean and variance of (2.1) at t_k is completely settled by $t_k - t_{k-1}$ and can not be reduced. Therefore, small enough time distance between consecutive observations would be typically necessary to obtain an adequate approximation to the LMV filter. Undoubtedly, this imposes undesirable restrictions to the time distance between observations that can not be accomplished in many practical situations. This drawback can be overcome by means of the particle filter introduced in [12], but at expense of a very high computation cost. Note that this filter performs, by means of intensive simulations, an estimation of the whole probabilistic distribution of the processes $\mathbf{x}$ solution of (2.1) from which the first two conditional moments of $\mathbf{x}$ can then be computed. Obviously, this general solution to the filtering problem is not practical when an expedited computation of the LMV filter (2.3)-(2.7) is required, which is typically demanded in many applications. For example, the LMV filter and its approximations are a key component of the innovation method for the parameter estimation of diffusion processes from a time series of partial and noisy observations [44, 54, 40, 41, 42, 55, 29]. For this purpose, accurate and computationally efficient approximations to the LMV filter will be certainly usefull.

2.2. Local Linearization filters. A key component for constructing the Local Linearization (LL) filters is the concept of Weak Local Linear (WLL) approximation for Stochastic Differential Equations (SDEs) [23, 28].

Let us consider the SDE (2.1) on the time interval $[a, b] \subset [t_0, T]$, and the time discretization $(\tau)_h = \{\tau_n : n = 0, 1, \ldots, N\}$ of $[a, b]$ with maximum stepsize h defined as a sequence of times that satisfy the conditions $a = \tau_0 < \tau_1 < \cdots < \tau_N = b$, and $\max\limits_n(\tau_{n+1} - \tau_n) \leq h < 1$ for $n = 0, \ldots, N-1$. Further, let

$$n_t = \max\{n = 0, 1, \ldots, N : \tau_n \leq t \text{ and } \tau_n \in (\tau)_h\}$$

for all $t \in [a, b]$.

Definition 2.2. *For a given time discretization $(\tau)_h$ of $[a, b]$, the stochastic process $\mathbf{y} = \{\mathbf{y}(t), t \in [a, b]\}$ is called order-β $(= 1, 2)$ Weak Local Linear approximation of the solution of (2.1) on $[a, b]$ if it is the weak solution of the piecewise linear equation*

$$d\mathbf{y}(t) = (\mathbf{A}(\tau_{n_t})\mathbf{y}(t) + \mathbf{a}^{\beta}(t;\tau_{n_t}))dt + \sum_{i=1}^{m}(\mathbf{B}_i(\tau_{n_t})\mathbf{y}(t) + \mathbf{b}_i^{\beta}(t;\tau_{n_t}))d\mathbf{w}^i(t) \tag{2.8}$$

for all $t \in (\tau_n, \tau_{n+1}]$ and initial value $\mathbf{y}(a) = \mathbf{y}_0$, where the matrices functions $\mathbf{A}, \mathbf{B}_i$ are defined as

$$\mathbf{A}(s) = \frac{\partial \mathbf{f}(s, \mathbf{y}(s))}{\partial \mathbf{y}} \qquad \text{and} \qquad \mathbf{B}_i(s) = \frac{\partial \mathbf{g}_i(s, \mathbf{y}(s))}{\partial \mathbf{y}},$$

and the vectors functions $\mathbf{a}^{\beta}$, $\mathbf{b}_i^{\beta}$ *as*

$$\mathbf{a}^{\beta}(t;s)=\begin{cases}\mathbf{f}(s,\mathbf{y}(s))-\frac{\partial\mathbf{f}(s,\mathbf{y}(s))}{\partial\mathbf{y}}\mathbf{y}(s)+\frac{\partial\mathbf{f}(s,\mathbf{y}(s))}{\partial s}(t-s) & \text{for } \beta=1\\ \mathbf{a}^{1}(t;s)+\frac{1}{2}\sum\limits_{j,l=1}^{d}[\mathbf{G}(s,\mathbf{y}(s))\mathbf{G}^{\intercal}(s,\mathbf{y}(s))]^{j,l}\ \frac{\partial^2\mathbf{f}(s,\mathbf{y}(s))}{\partial\mathbf{y}^j\partial\mathbf{y}^l}(t-s) & \text{for } \beta=2\end{cases}$$

and

$$\mathbf{b}_i^{\beta}(t;s)=\begin{cases}\mathbf{g}_i(s,\mathbf{y}(s))-\frac{\partial\mathbf{g}_i(s,\mathbf{y}(s))}{\partial\mathbf{y}}\mathbf{y}(s)+\frac{\partial\mathbf{g}_i(s,\mathbf{y}(s))}{\partial s}(t-s) & \text{for } \beta=1\\ \mathbf{b}_i^{1}(t;s)+\frac{1}{2}\sum\limits_{j,l=1}^{d}[\mathbf{G}(s,\mathbf{y}(s))\mathbf{G}^{\intercal}(s,\mathbf{y}(s))]^{j,l}\ \frac{\partial^2\mathbf{g}_i(s,\mathbf{y}(s))}{\partial\mathbf{y}^j\partial\mathbf{y}^l}(t-s) & \text{for } \beta=2\end{cases}$$

for all $s\leq t$. *Here,* $\mathbf{G}=[\mathbf{g}_1,\ldots,\mathbf{g}_m]$ *is an* $d\times m$ *matrix function.*

The drift and diffusion coefficients of the equation (2.8) are, respectively, weak approximations of order β to the drift and diffusion coefficients of the equation (2.1) obtained from the Ito-Taylor expansion of order β. That is [33],

$$\sup_{s\leq t\leq s+h}\left|E(g(\mathbf{f}(t,\mathbf{y}(t)))-E(g(\mathbf{A}(s)\mathbf{y}(t)+\mathbf{a}^{\beta}(t;s)))\right|\leq Ch^{\beta}$$

and

$$\sup_{s\leq t\leq s+h}\left|E(g(\mathbf{g}_i(t,\mathbf{y}(t)))-E(g(\mathbf{B}_i(s)\mathbf{y}(t)+\mathbf{b}_i^{\beta}(t;s)))\right|\leq Ch^{\beta}$$

for all $h>0$ and $s\in[a,b-h]$, where $g\in\mathcal{C}_P^{2(\beta+1)}(\mathbb{R}^d,\mathbb{R})$ and C is a positive constant.

Explicit formulas for the conditional mean $\mathbf{y}_{t/\rho}$ and variance $\mathbf{V}_{t/\rho}$ of $\mathbf{y}$ were initially given in [27, 28] and simplified later in [21].

The conventional Local Linearization filters for the model (2.1)-(2.2) are obtained in two steps [28]: 1) by approximating the solution of the nonlinear state equation on each time subinterval $[t_k,t_{k+1}]$ by the Local Linear approximation (2.8) on $[t_k,t_{k+1}]$ with time discretization $(\tau)_h\equiv\{t_k,t_{k+1}\}$ for all $t_k,t_{k+1}\in\{t\}_M$; and 2) by the recursive application of the linear minimum variance filter [27] to the resulting piecewise linear continuous-discrete model. This yields to the following.

Definition 2.3. *Given a time discretization* $(\tau)_h\equiv\{t\}_M$, *the Local Linearization filter for the state space model (2.1)-(2.2) is defined, between observations, by the linear equations*

$$\frac{d\mathbf{y}_{t/t}}{dt}=\mathbf{A}(t_{n_t})\mathbf{y}_{t/t}+\mathbf{a}^{\beta}(t;t_{n_t}) \tag{2.9}$$

$$\frac{d\mathbf{V}_{t/t}}{dt}=\mathbf{A}(t_{n_t})\mathbf{V}_{t/t}+\mathbf{V}_{t/t}\mathbf{A}^{\intercal}(t_{n_t})+\sum_{i=1}^{m}\mathbf{B}_i(t_{n_t})\mathbf{V}_{t/t}\mathbf{B}_i^{\intercal}(t_{n_t})+\mathcal{B}(t;t_{n_t}) \tag{2.10}$$

for all $t\in(t_k,t_{k+1})$, *and by*

$$\mathbf{y}_{t_{k+1}/t_{k+1}}=\mathbf{y}_{t_{k+1}/t_k}+\mathbf{K}_{t_{k+1}}(\mathbf{z}_{t_{k+1}}-\mathbf{C}\mathbf{y}_{t_{k+1}/t_k}) \tag{2.11}$$

$$\mathbf{V}_{t_{k+1}/t_{k+1}}=\mathbf{V}_{t_{k+1}/t_k}-\mathbf{K}_{t_{k+1}}\mathbf{C}\mathbf{V}_{t_{k+1}/t_k} \tag{2.12}$$

for each observation at t_{k+1}, *with filter gain*

$$\mathbf{K}_{t_{k+1}}=\mathbf{V}_{t_{k+1}/t_k}\mathbf{C}^{\intercal}(\mathbf{C}\mathbf{V}_{t_{k+1}/t_k}\mathbf{C}^{\intercal}+\Sigma_{t_{k+1}})^{-1} \tag{2.13}$$

for all $t_k,t_{k+1}\in\{t\}_M$. *The predictions* $\mathbf{y}_{t/t_k}$ *and* $\mathbf{V}_{t/t_k}$ *are accomplished, respectively, via expressions (2.9)-(2.10) with initial conditions* $\mathbf{y}_{t_k/t_k}$ *and* $\mathbf{V}_{t_k/t_k}$ *for* $t\in(t_k,t_{k+1}]$. *Here,*

$$\mathcal{B}(t;s)=\sum_{i=1}^{m}\mathbf{B}_i(s)\mathbf{y}_{t/t}\mathbf{y}_{t/t}^{\intercal}\mathbf{B}_i^{\intercal}(s)+\mathbf{B}_i(s)\mathbf{y}_{t/t}(\mathbf{b}_i^{\beta}(t;s))^{\intercal}+\mathbf{b}_i^{\beta}(t;s)\mathbf{y}_{t/t}^{\intercal}\mathbf{B}_i^{\intercal}(t)+\mathbf{b}_i(t;s)(\mathbf{b}_i^{\beta}(t;s))^{\intercal},$$

and the matrices $\mathbf{A}, \mathbf{B}_i$ *and the vectors* $\mathbf{a}, \mathbf{b}_i^{\beta}$ *are defined as in the WLL approximation (2.8) but, replacing* $\mathbf{y}(s)$ *by* $\mathbf{y}_{s/s}$.

Both, the Local Linear approximations and the Local Linearization filters have had a number of important applications. The first ones, in addition to the filtering problems, have been used for the derivation of effective integration [23, 6, 5, 26] and inference [52, 53, 13, 55, 19] methods for SDEs, in the estimation of distribution functions in Monte Carlo Markov Chain methods [57, 50, 15] and the simulation of likelihood functions [39]. The second ones have played a crucial role in the practical implementation of innovation estimators for the identification of continuous-discrete state space models [44, 54, 45, 29]. In a variety of applications, these approximate innovation methods have shown high effectiveness and efficiency for the estimation of unobserved components and unknown parameters of SDEs given a set of discrete observations. Remarkable is the identification, from actual data, of neurophysiological, financial and molecular models, among others (see, e.g., [4, 32, 8, 9, 25, 48, 49]).

3. Approximate Linear Minimum Variance filters. Let $(\tau)_h$ be a time discretization of $[t_0, T]$ such that $(\tau)_h \supset \{t\}_M$, and $\mathbf{y}_n$ the approximate value of $\mathbf{x}(\tau_n)$ obtained from a discretization of the equation (2.1) for all $\tau_n \in (\tau)_h$. Let us consider the continuous time approximation $\mathbf{y} = \{\mathbf{y}(t),\ t \in [t_0, T] : \mathbf{y}(\tau_n) = \mathbf{y}_n$ for all $\tau_n \in (\tau)_h\}$ of $\mathbf{x}$ with initial conditions

$$E\left(\mathbf{y}(t_0)\middle|\mathcal{F}_{t_0}\right) = E\left(\mathbf{x}(t_0)\middle|\mathcal{F}_{t_0}\right) \quad \text{and} \quad E\left(\mathbf{y}(t_0)\mathbf{y}^{\intercal}(t_0)\middle|\mathcal{F}_{t_0}\right) = E\left(\mathbf{x}(t_0)\mathbf{x}^{\intercal}(t_0)\middle|\mathcal{F}_{t_0}\right);$$

satisfying the bound condition

$$E\left(\left|\mathbf{y}(t)\right|^{2q}\middle|\mathcal{F}_{t_k}\right) \leq L \tag{3.1}$$

for all $t \in [t_k, t_{k+1}]$; and the weak convergence criteria

$$\sup_{t_k \leq t \leq t_{k+1}} \left| E\left(g(\mathbf{x}(t))\middle|\mathcal{F}_{t_k}\right) - E\left(g(\mathbf{y}(t))\middle|\mathcal{F}_{t_k}\right)\right| \leq L_k h^{\beta} \tag{3.2}$$

for all $t_k, t_{k+1} \in \{t\}_M$, where $g \in \mathcal{C}_P^{2(\beta+1)}(\mathbb{R}^d, \mathbb{R})$, L and L_k are positive constants, $\beta \in \mathbb{N}_+$, and $q = 1, 2 \ldots$. The process $\mathbf{y}$ defined in this way is typically called order-β approximation to $\mathbf{x}$ in weak sense [33].

When an order-β approximation to the solution of the state equation (2.1) is chosen, the following approximate filter can be naturally defined.

Definition 3.1. *Given a time discretization* $(\tau)_h \supset \{t\}_M$, *the order-*β *Linear Minimum Variance filter for the state space model (2.1)-(2.2) is defined, between observations, by*

$$\mathbf{y}_{t/t} = E(\mathbf{y}(t)/Z_t) \qquad \text{and} \qquad \mathbf{V}_{t/t} = E(\mathbf{y}(t)\mathbf{y}^{\intercal}(t)/Z_t) - \mathbf{y}_{t/t}\mathbf{y}_{t/t}^{\intercal} \tag{3.3}$$

for all $t \in (t_k, t_{k+1})$, *and by*

$$\mathbf{y}_{t_{k+1}/t_{k+1}} = \mathbf{y}_{t_{k+1}/t_k} + \mathbf{K}_{t_{k+1}}(\mathbf{z}_{t_{k+1}} - \mathbf{C}\mathbf{y}_{t_{k+1}/t_k}), \tag{3.4}$$

$$\mathbf{V}_{t_{k+1}/t_{k+1}} = \mathbf{V}_{t_{k+1}/t_k} - \mathbf{K}_{t_{k+1}}\mathbf{C}\mathbf{V}_{t_{k+1}/t_k}, \tag{3.5}$$

for each observation at t_{k+1}, *with filter gain*

$$\mathbf{K}_{t_{k+1}} = \mathbf{V}_{t_{k+1}/t_k}\mathbf{C}^{\intercal}(\mathbf{C}\mathbf{V}_{t_{k+1}/t_k}\mathbf{C}^{\intercal} + \Sigma_{t_{k+1}})^{-1} \tag{3.6}$$

for all $t_k, t_{k+1} \in \{t\}_M$, *where* $\mathbf{y}$ *is an order-*β *approximation to the solution of (2.1) in weak sense. The predictions* $\mathbf{y}_{t/t_k} = E(\mathbf{y}(t)/Z_{t_k})$ *and* $\mathbf{V}_{t/t_k} = E(\mathbf{y}(t)\mathbf{y}^{\intercal}(t)/Z_{t_k}) - \mathbf{y}_{t/t_k}\mathbf{y}_{t/t_k}^{\intercal}$, *with initial conditions* $\mathbf{y}_{t_k/t_k}$ *and* $\mathbf{V}_{t_k/t_k}$, *are defined for all* $t \in (t_k, t_{k+1}]$ *and* $t_k, t_{k+1} \in \{t\}_M$.

Note that the goodness of the approximation $\mathbf{y}$ to $\mathbf{x}$ is measured (in weak sense) by the left hand side of (3.2). Thus, the inequality (3.2) gives a bound for the errors of the approximation $\mathbf{y}$ to $\mathbf{x}$, for all $t \in [t_k, t_{k+1}]$ and all pair of consecutive observations $t_k, t_{k+1} \in \{t\}_M$. Moreover, this inequality states the convergence (in weak sense and with rate β) of the approximation $\mathbf{y}$ to $\mathbf{x}$ as the maximum stepsize h of the time discretization $(\tau)_h \supset \{t\}_M$ goes to zero. Clearly this includes, as particular case, the convergence of the first two conditional moments of $\mathbf{y}$ to those of $\mathbf{x}$. Since the approximate filter in Definition 3.1 is designed in terms of the first two conditional moments of the approximation $\mathbf{y}$, the weak convergence of $\mathbf{y}$ to $\mathbf{x}$ should imply the convergence of the approximate filter to the exact one. Next result deals with this matter.

THEOREM 3.2. *Let $\mathbf{x}_{t/\rho}$ and $\mathbf{U}_{t/\rho}$ be the conditional mean and variance corresponding to the LMV filter (2.3)-(2.7) for the model (2.1)-(2.2), and $\mathbf{y}_{t/\rho}$ and $\mathbf{V}_{t/\rho}$ their respective approximations given by the order-β LMV filter (3.3)-(3.6). Then, between observations, the filters satisfy*

$$\left|\mathbf{x}_{t/t} - \mathbf{y}_{t/t}\right| \leq K_1 h^{\beta} \qquad \text{and} \qquad \left|\mathbf{U}_{t/t} - \mathbf{V}_{t/t}\right| \leq K_1 h^{\beta} \tag{3.7}$$

for all $t \in (t_k, t_{k+1})$ and, at each observation t_{k+1},

$$\left|\mathbf{x}_{t_{k+1}/t_{k+1}} - \mathbf{y}_{t_{k+1}/t_{k+1}}\right| \leq K_1 h^{\beta} \qquad \text{and} \qquad \left|\mathbf{U}_{t_{k+1}/t_{k+1}} - \mathbf{V}_{t_{k+1}/t_{k+1}}\right| \leq K_1 h^{\beta} \tag{3.8}$$

for all $t_k, t_{k+1} \in \{t\}_M$, where K_1 is a positive constant. For the predictions,

$$\left|\mathbf{x}_{t/t_k} - \mathbf{y}_{t/t_k}\right| \leq K_2 h^{\beta} \qquad \text{and} \qquad \left|\mathbf{U}_{t/t_k} - \mathbf{V}_{t/t_k}\right| \leq K_2 h^{\beta} \tag{3.9}$$

hold for all $t \in (t_k, t_{k+1}]$ and $t_k, t_{k+1} \in \{t\}_M$, where K_2 is a positive constant.

Proof. Let us start proving inequalities (3.9) and (3.7). For the functions $g(\mathbf{x}(t)) = \mathbf{x}^i(t)$ and $g(\mathbf{x}(t)) = \mathbf{x}^i(t)\mathbf{x}^j(t)$ belonging to the function space $\mathcal{C}_P^{2(\beta+1)}(\mathbb{R}^d, \mathbb{R})$, for all $i, j = 1..d$, condition (3.2) directly implies that

$$\left|\mathbf{x}_{t/t_k} - \mathbf{y}_{t/t_k}\right| \leq \sqrt{d} L_k h^{\beta} \tag{3.10}$$

and

$$\left|\mathbf{Q}_{t/t_k} - \mathbf{P}_{t/t_k}\right| \leq d L_k h^{\beta}$$

for all $t \in (t_k, t_{k+1}]$, where $\mathbf{P}_{t/t_k} = E(\mathbf{y}(t)\mathbf{y}^{\intercal}(t)/Z_{t_k})$. Since the solution of (2.1) has bounded moments, there exists a positive contant Λ such that of $\left|\mathbf{x}_{t/t_k}\right| \leq \Lambda$ for all $t \in [t_k, t_{k+1}]$. Condition (3.1) implies that $\left|\mathbf{y}_{t/t_k}\right| \leq L$ for all $t \in [t_k, t_{k+1}]$. From the formula of the variance in terms of the first two moments, it follows that

$$\left|\mathbf{U}_{t/t_k} - \mathbf{V}_{t/t_k}\right| \leq \left|\mathbf{Q}_{t/t_k} - \mathbf{P}_{t/t_k}\right| + \left|\mathbf{x}_{t/t_k}\mathbf{x}^{\intercal}_{t/t_k} - \mathbf{y}_{t/t_k}\mathbf{y}^{\intercal}_{t/t_k}\right|.$$

Since

$$\begin{aligned}
\left|\mathbf{x}_{t/t_k}\mathbf{x}^{\intercal}_{t/t_k} - \mathbf{y}_{t/t_k}\mathbf{y}^{\intercal}_{t/t_k}\right| &= \left|\mathbf{x}_{t/t_k}\mathbf{x}^{\intercal}_{t/t_k} - \mathbf{x}_{t/t_k}\mathbf{y}^{\intercal}_{t/t_k} + \mathbf{x}_{t/t_k}\mathbf{y}^{\intercal}_{t/t_k} - \mathbf{y}_{t/t_k}\mathbf{y}^{\intercal}_{t/t_k}\right| \\
&\leq \left|\mathbf{x}_{t/t_k}(\mathbf{x}^{\intercal}_{t/t_k} - \mathbf{y}^{\intercal}_{t/t_k})\right| + \left|(\mathbf{x}_{t/t_k} - \mathbf{y}_{t/t_k})\mathbf{y}^{\intercal}_{t/t_k}\right| \\
&\leq (\left|\mathbf{x}_{t/t_k}\right| + \left|\mathbf{y}_{t/t_k}\right|)\left|\mathbf{x}_{t/t_k} - \mathbf{y}_{t/t_k}\right|,
\end{aligned}$$

$$\left|\mathbf{U}_{t/t_k} - \mathbf{V}_{t/t_k}\right| \leq \alpha_k h^{\beta} \tag{3.11}$$

for all $t \in (t_k, t_{k+1}]$, where $\alpha_k = (\sqrt{d} + L + \Lambda)\sqrt{d}L_k$. Hence, inequalities (3.9) are obtained from (3.10) and (3.11) with $K_1 = \max_k\{\alpha_k\}$. Inequalities (3.7) can be derived in the same way.

For the remainder inequalities follow this. From (2.5) and (3.4), it is obtained

$$\begin{aligned}
\left|\mathbf{x}_{t_{k+1}/t_{k+1}} - \mathbf{y}_{t_{k+1}/t_{k+1}}\right| &\leq \left|\mathbf{x}_{t_{k+1}/t_k} - \mathbf{y}_{t_{k+1}/t_k}\right| \\
&\quad + \left|\mathbf{G}_{t_{k+1}}(\mathbf{z}_{t_{k+1}} - \mathbf{C}\mathbf{x}_{t_{k+1}/t_k}) - \mathbf{K}_{t_{k+1}}(\mathbf{z}_{t_{k+1}} - \mathbf{C}\mathbf{y}_{t_{k+1}/t_k})\right| \\
&\leq (1 + \left|\mathbf{G}_{t_{k+1}}\mathbf{C}\right|)\left|\mathbf{x}_{t_{k+1}/t_k} - \mathbf{y}_{t_{k+1}/t_k}\right| \\
&\quad + (\left|\mathbf{z}_{t_{k+1}}\right| + \left|\mathbf{C}\mathbf{y}_{t_{k+1}/t_k}\right|)\left|\mathbf{G}_{t_{k+1}} - \mathbf{K}_{t_{k+1}}\right|.
\end{aligned}$$

From (2.6) and (3.5),

$$\begin{aligned}\left|\mathbf{U}_{t_{k+1}/t_{k+1}}-\mathbf{V}_{t_{k+1}/t_{k+1}}\right| &\leq \left|\mathbf{U}_{t_{k+1}/t_k}-\mathbf{V}_{t_{k+1}/t_k}\right|+\left|\mathbf{G}_{t_{k+1}}\mathbf{C}\mathbf{U}_{t_{k+1}/t_k}-\mathbf{K}_{t_{k+1}}\mathbf{C}\mathbf{V}_{t_{k+1}/t_k}\right| \\ &\leq (1+\left|\mathbf{G}_{t_{k+1}}\mathbf{C}\right|)\left|\mathbf{U}_{t_{k+1}/t_k}-\mathbf{V}_{t_{k+1}/t_k}\right|+\left|\mathbf{C}\mathbf{V}_{t_{k+1}/t_k}\right|\left|\mathbf{G}_{t_{k+1}}-\mathbf{K}_{t_{k+1}}\right|.\end{aligned}$$

By rewriting (3.6) and (2.7) as

$$\mathbf{K}_{t_{k+1}}(\mathbf{C}\mathbf{V}_{t_{k+1}/t_k}\mathbf{C}^\intercal+\Sigma_{t_{k+1}})=\mathbf{V}_{t_{k+1}/t_k}\mathbf{C}^\intercal$$

and

$$\mathbf{G}_{t_{k+1}}(\mathbf{C}\mathbf{U}_{t_{k+1}/t_k}\mathbf{C}^\intercal+\Sigma_{t_{k+1}})-\mathbf{G}_{t_{k+1}}(\mathbf{C}\mathbf{V}_{t_{k+1}/t_k}\mathbf{C}^\intercal+\Sigma_{t_{k+1}})+\mathbf{G}_{t_{k+1}}(\mathbf{C}\mathbf{V}_{t_{k+1}/t_k}\mathbf{C}^\intercal+\Sigma_{t_{k+1}})=\mathbf{U}_{t_{k+1}/t_k}\mathbf{C}^\intercal,$$

and subtracting the first expression to the second one, it follows that

$$(\mathbf{G}_{t_{k+1}}-\mathbf{K}_{t_{k+1}})(\mathbf{C}\mathbf{V}_{t_{k+1}/t_k}\mathbf{C}^\intercal+\Sigma_{t_{k+1}})=\mathbf{G}_{t_{k+1}}\mathbf{C}(\mathbf{V}_{t_{k+1}/t_k}-\mathbf{U}_{t_{k+1}/t_k})\mathbf{C}^\intercal+(\mathbf{U}_{t_{k+1}/t_k}-\mathbf{V}_{t_{k+1}/t_k})\mathbf{C}^\intercal.$$

Thus,

$$(\mathbf{G}_{t_{k+1}}-\mathbf{K}_{t_{k+1}})=(\mathbf{I}-\mathbf{G}_{t_{k+1}}\mathbf{C})(\mathbf{U}_{t_{k+1}/t_k}-\mathbf{V}_{t_{k+1}/t_k})\mathbf{C}^\intercal(\mathbf{C}\mathbf{V}_{t_{k+1}/t_k}\mathbf{C}^\intercal+\Sigma_{t_{k+1}})^{-1}$$

and

$$\left|\mathbf{G}_{t_{k+1}}-\mathbf{K}_{t_{k+1}}\right|\leq\left|(\mathbf{I}-\mathbf{G}_{t_{k+1}}\mathbf{C})\right|\left|\mathbf{C}^\intercal(\mathbf{C}\mathbf{V}_{t_{k+1}/t_k}\mathbf{C}^\intercal+\Sigma_{t_{k+1}})^{-1}\right|\left|\mathbf{U}_{t_{k+1}/t_k}-\mathbf{V}_{t_{k+1}/t_k}\right|.$$

From the above inequalities, and taking into account that $\left|\mathbf{V}_{t_{k+1}/t_k}\right|$, $\left|\mathbf{G}_{t_{k+1}}\right|$, $\left|\Sigma_{t_{k+1}}\right|$ and $|\mathbf{C}|$ are also bound, it is obtained that

$$\left|\mathbf{x}_{t_{k+1}/t_{k+1}}-\mathbf{y}_{t_{k+1}/t_{k+1}}\right|\leq\beta_k h^\beta \qquad \text{and} \qquad \left|\mathbf{U}_{t_{k+1}/t_{k+1}}-\mathbf{V}_{t_{k+1}/t_{k+1}}\right|\leq\beta_k h^\beta,$$

where β_k is a positive constant. This implies (3.8) with $K_2=\max_k\{\beta_k\}$. □

Theorem 3.2 states that, given a set of M partial and noisy observations of the states $\mathbf{x}$ on $\{t\}_M$, the approximate LMV filter of Definition 3.1 converges with rate β to the exact LMV filter of Definition 2.1 as h goes to zero, where h is the maximum stepsize of the time discretization $(\tau)_h\supset\{t\}_M$ on which the approximation $\mathbf{y}$ to $\mathbf{x}$ is defined. This means that the approximate filter inherits the convergence rate of the approximation employed for its design. Note that, the convergence results of Theorem 3.2 can be easily extended for noisy observations of any realization of $\mathbf{x}$ just by taking expectation value in the inequalities (3.7)-(3.9). Further note that in both, Definition 3.1 and Theorem 3.2, no restriction on the time partition $\{t\}_M$ for the data has been assumed. Thus, there are not specific constraints about the time distance between two consecutive observations, which allows the application of the approximate filter in a variety of practical problems (see, e.g., [49, 17, 18]) with a reduced number of not close observations in time, with sequential random measurements, or with multiple missing data. Neither there are restrictions on the time discretization $(\tau)_h\supset\{t\}_M$ on which the approximate filter is defined. Thus, $(\tau)_h$ can be set by the user by taking into account some specifications or previous knowledge on the filtering problem under consideration, or automatically designed by an adaptive strategy as it will be shown in the section concerning the numerical simulations.

The order-β LMV filter of Definition 3.1 has been proposed for models with linear observation equation. However, by following the procedure proposed in [29], it can be easily applied as well to models with nonlinear observation equation.

To illustrate this, let us consider the state space model defined by the continuous state equation (2.1) and the discrete observation equation

$$\mathbf{z}_{t_k}=\mathbf{h}(t_k,\mathbf{x}(t_k))+\mathbf{e}_{t_k},\ \text{for } k=0,1,..,M-1, \tag{3.12}$$

where $\mathbf{e}_{t_k}$ is defined as in (2.2) and $\mathbf{h}:\mathbb{R}\times\mathbb{R}^d\to\mathbb{R}^r$ is a twice differentiable function. By using the Ito formula,

$$\begin{aligned} d\mathbf{h}^j &= \{\frac{\partial\mathbf{h}^j}{\partial t}+\sum_{k=1}^{d}f^k\frac{\partial\mathbf{h}^j}{\partial\mathbf{x}^k}+\frac{1}{2}\sum_{s=1}^{m}\sum_{k,l=1}^{d}\mathbf{g}_s^l\mathbf{g}_s^k\frac{\partial^2\mathbf{h}^j}{\partial\mathbf{x}^l\partial\mathbf{x}^k}\}dt+\sum_{s=1}^{m}\sum_{l=1}^{d}\mathbf{g}_s^l\frac{\partial\mathbf{h}^j}{\partial\mathbf{x}^l}d\mathbf{w}^s \\ &= \boldsymbol{\rho}^j dt+\sum_{s=1}^{m}\boldsymbol{\sigma}_s^j d\mathbf{w}^s\end{aligned}$$

with $j = 1,..,r$. Hence, the state space model (2.1) and (3.12) is transformed to the following higher-dimensional state space model with linear observation

$$d\mathbf{v}(t) = \mathbf{a}(t, \mathbf{v}(t))dt + \sum_{i=1}^{m} \mathbf{b}_i(t, \mathbf{v}(t))d\mathbf{w}^i(t),$$

$$\mathbf{z}_{t_k} = \mathbf{C}\mathbf{v}(t_k) + \mathbf{e}_{t_k}, \text{ for } k = 0, 1, .., M-1,$$

where

$$\mathbf{v} = \begin{bmatrix} \mathbf{x} \\ \mathbf{h} \end{bmatrix}, \ \mathbf{a} = \begin{bmatrix} \mathbf{f} \\ \boldsymbol{\rho} \end{bmatrix}, \ \mathbf{b}_i = \begin{bmatrix} \mathbf{g}_i \\ \boldsymbol{\sigma}_i \end{bmatrix}$$

and the matrix $\mathbf{C}$ is such that $\mathbf{h}(t_k, \mathbf{x}(t_k)) = \mathbf{C}\mathbf{v}(t_k)$.

In this way, the state space model (2.1) and (3.12) is transformed to the form of the state space model (2.1)-(2.2), and so the order-β LMV filter of Definition 3.1 and the convergence result of Theorem 3.2 can be applied.

4. Order-β Local Linearization filters. In principle, according to Theorem 3.2, any kind of approximation $\mathbf{y}$ converging to $\mathbf{x}$ in a weak sense can be used to construct approximate LMV filters (e.g., those in [33]). Therefore, additional selection criterions could be taking into account for this purpose. For instance, high order of convergence, efficient algorithm for the computation of the moments, and so on. In this paper, we elected the Local Linear approximation (2.8) for the following reasons: 1) its first two conditional moments have simple explicit formulas that can be computed by means of efficient algorithm (including high dimensional state equations) [27, 28, 21]; 2) its first two conditional moments are exact for linear state equations in all the possible variants (with additive and/or multiplicative noise, autonomous or not) [27]; 3) it has an adequate order of weak convergence for state equations with additive noise [6]; and 4) the high effectiveness of the conventional LL filters for the identification of complex nonlinear models in a variety of applications (see, e.g., [4, 8, 25, 48, 49]).

Once the order-β Local Linear approximation (2.8) is chosen for approximating the state equation (2.1), the well know ordinary differential equations for the first two moments of linear SDEs [1] can be directly used to define the following filter.

Definition 4.1. *Given a time discretization $(\tau)_h \supset \{t\}_M$, the order-β Local Linearization filter for the state space model (2.1)-(2.2) is defined, between observations, by the piecewise linear equations*

$$\frac{d\mathbf{y}_{t/t}}{dt} = \mathbf{A}(\tau_{n_t})\mathbf{y}_{t/t} + \mathbf{a}^{\beta}(t; \tau_{n_t}) \tag{4.1}$$

$$\frac{d\mathbf{P}_{t/t}}{dt} = \mathbf{A}(\tau_{n_t})\mathbf{P}_{t/t} + \mathbf{P}_{t/t}\mathbf{A}^{\intercal}(\tau_{n_t}) + \sum_{i=1}^{m} \mathbf{B}_i(\tau_{n_t})\mathbf{P}_{t/t}\mathbf{B}_i^{\intercal}(\tau_{n_t}) + \mathcal{B}(t; \tau_{n_t}) \tag{4.2}$$

$$\mathbf{V}_{t/t} = \mathbf{P}_{t/t} - \mathbf{y}_{t/t}\mathbf{y}_{t/t}^{\intercal} \tag{4.3}$$

for all $t \in (t_k, t_{k+1})$, and by

$$\mathbf{y}_{t_{k+1}/t_{k+1}} = \mathbf{y}_{t_{k+1}/t_k} + \mathbf{K}_{t_{k+1}}(\mathbf{z}_{t_{k+1}} - \mathbf{C}\mathbf{y}_{t_{k+1}/t_k}) \tag{4.4}$$

$$\mathbf{V}_{t_{k+1}/t_{k+1}} = \mathbf{V}_{t_{k+1}/t_k} - \mathbf{K}_{t_{k+1}}\mathbf{C}\mathbf{V}_{t_{k+1}/t_k} \tag{4.5}$$

for each observation at t_{k+1}, with filter gain

$$\mathbf{K}_{t_{k+1}} = \mathbf{V}_{t_{k+1}/t_k}\mathbf{C}^{\intercal}(\mathbf{C}\mathbf{V}_{t_{k+1}/t_k}\mathbf{C}^{\intercal} + \Sigma_{t_{k+1}})^{-1} \tag{4.6}$$

for all $t_k, t_{k+1} \in \{t\}_M$. *Here,*

$$\mathcal{B}(t;s) = \mathbf{a}^{\beta}(t;s)\mathbf{y}_{t/t}^{\intercal} + \mathbf{y}_{t/t}(\mathbf{a}^{\beta}(t;s))^{\intercal} + \sum_{i=1}^{m} \mathbf{B}_i(s)\mathbf{y}_{t/t}(\mathbf{b}_i^{\beta}(t;s))^{\intercal} + \mathbf{b}_i^{\beta}(t;s)\mathbf{y}_{t/t}^{\intercal}\mathbf{B}_i^{\intercal}(s) + \mathbf{b}_i^{\beta}(t;s)(\mathbf{b}_i^{\beta}(t;s))^{\intercal} \tag{4.7}$$

with matrix functions $\mathbf{A}, \mathbf{B}_i$ *and vector functions* $\mathbf{a}^{\beta}, \mathbf{b}_i^{\beta}$ *defined as in the WLL approximation (2.8) but, replacing* $\mathbf{y}(s)$ *by* $\mathbf{y}_{s/s}$. *The predictions* $\mathbf{y}_{t/t_k}$, $\mathbf{P}_{t/t_k}$ *and* $\mathbf{V}_{t/t_k}$ *are accomplished, respectively, via expressions (4.1)-(4.3) with initial conditions* $\mathbf{y}_{t_k/t_k}$ *and* $\mathbf{P}_{t_k/t_k}$ *for* $t \in (t_k, t_{k+1}]$ *and* $t_k, t_{k+1} \in \{t\}_M$, *and with* $\mathbf{A}, \mathbf{B}_i, \mathbf{a}^{\beta}, \mathbf{b}_i^{\beta}$ *also defined as in (2.8) but, replacing* $\mathbf{y}(s)$ *by* $\mathbf{y}_{s/t_k}$.

The approximate LL filter (4.1)-(4.6) reduces to the conventional LL filter (2.9)-(2.13) when $(\tau)_h \equiv \{t\}_M$. For linear state equations with multiplicative noise, the LL filter (4.1)-(4.6) reduces to the LMV filter proposed in [27], whereas for linear state equations with additive noise, the LL filter (4.1)-(4.6) reduces to the classical Kalman filter.

According with Theorem 3.2, the approximate LL filter (4.1)-(4.6) will inherit the order of convergence of the WLL approximation (2.8). As it was mention before, the weak convergence rate of that approximation was early stated in [6] for SDEs with additive noise. For equations with multiplicative noise, this subject will be considered in what follows.

Lemma 4.2. *Suppose that the drift and diffusion coefficients of the SDE (2.1) satisfy the following conditions*

$$\mathbf{f}^k, \mathbf{g}_i^k \in \mathcal{C}_P^{2(\beta+1)}([a,b] \times \mathbb{R}^d, \mathbb{R}) \tag{4.8}$$

$$|\mathbf{f}(s,\mathbf{u})| + \sum_{i=1}^{m}(|\mathbf{g}_i(s,\mathbf{u})| + \sum_{k,l=1}^{d} \left|\mathbf{g}_i^k(s,\mathbf{u})\mathbf{g}_i^l(s,\mathbf{u})\right| \delta_\beta^2) \leq K(1+|\mathbf{u}|), \tag{4.9}$$

$$\left|\frac{\partial \mathbf{f}(s,\mathbf{u})}{\partial t}\right| + \left|\frac{\partial \mathbf{f}(s,\mathbf{u})}{\partial \mathbf{x}}\right| + \left|\frac{\partial^2 \mathbf{f}(s,\mathbf{u})}{\partial \mathbf{x}^2}\right| \delta_\beta^2 \leq K \tag{4.10}$$

and

$$\left|\frac{\partial \mathbf{g}_i(s,\mathbf{u})}{\partial t}\right| + \left|\frac{\partial \mathbf{g}_i(s,\mathbf{u})}{\partial \mathbf{x}}\right| + \left|\frac{\partial^2 \mathbf{g}_i(s,\mathbf{u})}{\partial \mathbf{x}^2}\right| \delta_\beta^2 \leq K \tag{4.11}$$

for all $s \in [a,b]$, $\mathbf{u} \in \mathbb{R}^d$, *and* $i = 1,..,m$, *where* K *is a positive constant. Then the order-*β *WLL approximation (2.8) satisfies*

$$E\left(\sup_{a\leq t\leq b} |\mathbf{y}(t)|^{2q} \,\middle|\, \mathcal{F}_a\right) \leq C(1+|\mathbf{y}(a)|^{2q}) \tag{4.12}$$

for each $q = 1, 2, \ldots,$ *where* C *is positive constant.*

Proof. Let us denote the drift and diffusion coefficients of the SDE (2.8) by

$$\mathbf{p}(t,\mathbf{y}(t);\tau_{n_t}) = \mathbf{A}(\tau_{n_t})\mathbf{y}(t) + \mathbf{a}^{\beta}(t;\tau_{n_t})$$

and

$$\mathbf{q}_i(t,\mathbf{y}(t);\tau_{n_t}) = \mathbf{B}_i(\tau_{n_t})\mathbf{y}(t) + \mathbf{b}_i^{\beta}(t;\tau_{n_t}),$$

respectively.

For each q, the Ito formula applied to $|\mathbf{y}(t)|^{2q}$ implies that

$$\begin{aligned}|\mathbf{y}(t)|^{2q} &= |\mathbf{y}(\tau_{n_t})|^{2q} + \int_{\tau_{n_t}}^{t} 2q\,|\mathbf{y}(s)|^{2q-2}\,\mathbf{y}^{\intercal}(s)\mathbf{p}(s,\mathbf{y}(s);\tau_{n_t})ds \\ &+ \sum_{i=1}^{m}\int_{\tau_{n_t}}^{t} 2q\,|\mathbf{y}(s)|^{2q-2}\,\mathbf{y}^{\intercal}(s)\mathbf{q}_i(s,\mathbf{y}(s);\tau_{n_t})d\mathbf{w}^i(s) \\ &+ \sum_{i=1}^{m}\int_{\tau_{n_t}}^{t} q\,|\mathbf{y}(s)|^{2q-2}\,|\mathbf{q}_i(s,\mathbf{y}(s);\tau_{n_t})|^2\,ds \\ &+ \sum_{i=1}^{m}\int_{\tau_{n_t}}^{t} 2q(q-1)\,|\mathbf{y}(s)|^{2q-4}\,|\mathbf{y}^{\intercal}(s)\mathbf{q}_i(s,\mathbf{y}(s);\tau_{n_t})|^2\,ds\end{aligned}$$

for all $t \in [\tau_{n_t}, \tau_{n_t+1}]$.

By recursive application of the expression above it is obtained that

$$\begin{aligned}|\mathbf{y}(t)|^{2q} &= |\mathbf{y}(a)|^{2q} + \int_{a}^{t} 2q\,|\mathbf{y}(s)|^{2q-2}\,\mathbf{y}^{\intercal}(s)\mathbf{p}(s,\mathbf{y}(s);\tau_{n_s})ds \\ &+ \sum_{i=1}^{m}\int_{a}^{t} 2q\,|\mathbf{y}(s)|^{2q-2}\,\mathbf{y}^{\intercal}(s)\mathbf{q}_i(s,\mathbf{y}(s);\tau_{n_s})d\mathbf{w}^i(s) \\ &+ \sum_{i=1}^{m}\int_{a}^{t} q\,|\mathbf{y}(s)|^{2q-2}\,|\mathbf{q}_i(s,\mathbf{y}(s);\tau_{n_s})|^2\,ds \\ &+ \sum_{i=1}^{m}\int_{a}^{t} 2q(q-1)\,|\mathbf{y}(s)|^{2q-4}\,|\mathbf{y}^{\intercal}(s)\mathbf{q}_i(s,\mathbf{y}(s);\tau_{n_s})|^2\,ds\end{aligned}$$

for all $t \in [a,b]$.

Theorem 4.5.4 in [33] implies that $E\left(|\mathbf{y}(t)|^{2q}\right) < \infty$ for $a \leq t \leq b$. Hence, the function $\mathbf{r}$ defined as $\mathbf{r}(t) = \mathbf{0}$ for $0 \leq t < a$ and as $\mathbf{r}(t) = |\mathbf{y}(t)|^{2q-2}\,\mathbf{y}^{\intercal}(t)\mathbf{q}_i(t,\mathbf{y}(t);\tau_{n_t})$ for $a \leq t \leq b$ belongs to the class $\mathcal{L}_b^2$ of function $\mathcal{L}\times\mathcal{F}-$ measurable. Then, Lemma 3.2.2 in [33] implies that

$$E\left(\int_{a}^{t} |\mathbf{y}(s)|^{2q-2}\,\mathbf{y}^{\intercal}(s)\mathbf{q}_i(s,\mathbf{y}(s);\tau_{n_s})d\mathbf{w}^i(s)\right) = 0$$

for all $i = 1,..,m$. From this and the previous expression for $|\mathbf{y}(t)|^{2q}$ follows that

$$\begin{aligned}E\left(\sup_{a\leq u\leq t}|\mathbf{y}(u)|^{2q}\,\Big|\mathcal{F}_a\right) &\leq |\mathbf{y}(a)|^{2q} + 2q\int_{a}^{t} E\left(|\mathbf{y}(s)|^{2q-2}\,|\mathbf{y}^{\intercal}(s)\mathbf{p}(s,\mathbf{y}(s);\tau_{n_s})|\,\Big|\mathcal{F}_a\right)ds \\ &+ q\sum_{i=1}^{m}\int_{a}^{t} E\left(|\mathbf{y}(s)|^{2q-2}\,|\mathbf{q}_i(s,\mathbf{y}(s);\tau_{n_s})|^2\,\Big|\mathcal{F}_a\right)ds \\ &+ 2q(q-1)\sum_{i=1}^{m}\int_{a}^{t} E\left(|\mathbf{y}(s)|^{2q-4}\,|\mathbf{y}^{\intercal}(s)\mathbf{q}_i(s,\mathbf{y}(s);\tau_{n_s})|^2\,\Big|\mathcal{F}_a\right)ds.\end{aligned}$$

From conditions (4.9)-(4.11) follows that

$$|\mathbf{p}(s,\mathbf{y}(s);\tau_{n_s})| \leq K(|\mathbf{y}(s)| + |\mathbf{y}(\tau_{n_s})|) + K_\beta(1 + |\mathbf{y}(s)|) + K$$

and

$$|\mathbf{q}_i(s,\mathbf{y}(s);\tau_{n_s})| \leq K(|\mathbf{y}(s)| + |\mathbf{y}(\tau_{n_s})|) + K_\beta(1 + |\mathbf{y}(s)|) + K,$$

where

$$K_\beta = \begin{cases} K & \text{for } \beta = 1 \\ K(1 + \frac{1}{2}K) & \text{for } \beta = 2 \end{cases}.$$

Thus, there exists a positive constant C such that

$$|\mathbf{y}^\intercal(s)\mathbf{p}(s,\mathbf{y}(s);\tau_{n_s})| \leq C(1 + |\mathbf{y}(s)|^2) + C(1 + |\mathbf{y}(\tau_{n_s})|^2),$$

$$|\mathbf{q}_i(s,\mathbf{y}(s);\tau_{n_s})|^2 \leq C(1 + |\mathbf{y}(s)|^2) + C(1 + |\mathbf{y}(\tau_{n_s})|^2),$$

$$|\mathbf{y}^\intercal(s)\mathbf{q}_i(s,\mathbf{y}(s);\tau_{n_s})|^2 \leq C|\mathbf{y}(s)|^2(1 + |\mathbf{y}(s)|^2) + C|\mathbf{y}(s)|^2(1 + |\mathbf{y}(\tau_{n_s})|^2),$$

and so

$$E\left(\sup_{a\leq u\leq t} |\mathbf{y}(u)|^{2q} \,\Big|\mathcal{F}_a\right) \leq |\mathbf{y}_0|^{2q} + L\int_a^t E\left(\sup_{a\leq u\leq s}(1 + |\mathbf{y}(u)|^2)\,|\mathbf{y}(u)|^{2q-2}\,\Big|\mathcal{F}_a\right) ds,$$

where $L = 2qC(2 + 2qm - m)$. From the inequality $(1 + z^2)z^{2q-2} \leq 1 + 2z^{2q}$,

$$E\left(\sup_{a\leq u\leq t} |\mathbf{y}(u)|^{2q} \,\Big|\mathcal{F}_a\right) \leq |\mathbf{y}_0|^{2q} + L(t-a) + 2L\int_a^t E\left(\sup_{a\leq u\leq s} |\mathbf{y}(u)|^{2q} \,\Big|\mathcal{F}_a\right) ds.$$

From this and the Gronwall Lemma, the assertion of the Theorem is obtained. □

In what follows, additional notations and results of [33] will be used. Briefly recall us that $\mathcal{M}$ denotes the set of all the multi-indexes $\alpha = (j_1, \ldots, j_{l(\alpha)})$ with $j_i \in \{0, 1, \ldots, m\}$ and $i = 1, \ldots, l(\alpha)$, where m is the dimension of $\mathbf{w}$ in (2.1). $l(\alpha)$ denotes the length of the multi-index α and $n(\alpha)$ the number of its zero components. $-\alpha$ and $\alpha-$ are the multi-indexes in $\mathcal{M}$ obtained by deleting the first and the last component of α, respectively. The multi-index of length zero will be denoted by v. Further,

$$L^0 = \frac{\partial}{\partial t} + \sum_{k=1}^{d} \mathbf{f}^k \frac{\partial}{\partial \mathbf{x}^k} + \frac{1}{2}\sum_{k,l=1}^{d}\sum_{j=1}^{m} \mathbf{g}_j^k \mathbf{g}_j^l \,\frac{\partial^2}{\partial \mathbf{x}^k \partial \mathbf{x}^l}$$

denotes the diffusion operator for the SDE (2.1), and

$$L^j = \sum_{k=1}^{d} \mathbf{g}_j^k \frac{\partial}{\partial \mathbf{x}^k},$$

for $j = 1, \ldots, m$.

Let us consider the hierarchical set

$$\Gamma_\beta = \{\alpha \in \mathcal{M} : l(\alpha) \leq \beta\}$$

with $\beta = 1, 2$; and $\mathcal{B}(\Gamma_\beta) = \{\alpha \in \mathcal{M}\backslash\Gamma_\beta : -\alpha \in \Gamma_\beta\}$ the remainder set of Γ_β.

LEMMA 4.3. *Let* $\mathbf{y}$ *be the order-*β *WLL approximation (2.8), and* $\mathbf{z} = \{\mathbf{z}(t),\ t \in [a,b]\}$ *be the stochastic process defined by*

$$\mathbf{z}(t) = \mathbf{y}_{n_t} + \sum_{\alpha\in\Gamma_\beta/\{\nu\}} I_\alpha[\Lambda_\alpha(\tau_{n_t}, \mathbf{y}_{n_t}; \tau_{n_t})]_{\tau_{n_t},t} + \sum_{\alpha\in\mathcal{B}(\Gamma_\beta)} I_\alpha[\Lambda_\alpha(.,\mathbf{y}_.;\tau_{n_t})]_{\tau_{n_t},t}, \tag{4.13}$$

where $I_\alpha[.]_{\tau_{n_t},t}$ *denotes the multiple Ito integral and, for any given* $(\tau_{n_t}, \mathbf{y}_{n_t})$,

$$\Lambda_\alpha(s,\mathbf{v};\tau_{n_t}) = \begin{cases} L^{j_1}\dots L^{j_{l(\alpha)-1}}\mathbf{p}^\beta(s,\mathbf{v};\tau_{n_t}) & if\ j_{l(\alpha)} = 0 \\ L^{j_1}\dots L^{j_{l(\alpha)-1}}\mathbf{q}^\beta_{j_{l(\alpha)}}(s,\mathbf{v};\tau_{n_t}) & if\ j_{l(\alpha)} \neq 0 \end{cases}$$

is a function of s *and* $\mathbf{v}$*, with*

$$\mathbf{p}^\beta(s,\mathbf{v};\tau_{n_t}) = \mathbf{A}(\tau_{n_t})\mathbf{v} + \mathbf{a}^\beta(s;\tau_{n_t}) \qquad \textit{and} \qquad \mathbf{q}^\beta_i(s,\mathbf{v};\tau_{n_t}) = \mathbf{B}_i(\tau_{n_t})\mathbf{v} + \mathbf{b}^\beta_i(s;\tau_{n_t}),$$

for all $s \in [a,b]$ *and* $\mathbf{v} \in \mathbb{R}^d$*, and matrix functions* $\mathbf{A}, \mathbf{B}_i$ *and vector functions* $\mathbf{a}^\beta, \mathbf{b}^\beta_i$ *defined as in the WLL approximation (2.8). Then*

$$E\left(g(\mathbf{y}(t))\right) = E\left(g(\mathbf{z}(t))\right), \tag{4.14}$$

$$E\left(g(\mathbf{y}(t) - \mathbf{y}(\tau_{n_t}))\right) = E\left(g(\mathbf{z}(t) - \mathbf{z}(\tau_{n_t}))\right) \tag{4.15}$$

for all $t \in [a,b]$ *and* $g \in \mathcal{C}_P^{2(\beta+1)}(\mathbb{R}^d,\mathbb{R})$*; and*

$$I_\alpha[\Lambda_\alpha(\tau_{n_t}, \mathbf{y}_{n_t}; \tau_{n_t})]_{\tau_{n_t},t} = I_\alpha[\lambda_\alpha(\tau_{n_t}, \mathbf{y}_{n_t})]_{\tau_{n_t},t}, \tag{4.16}$$

for all $\alpha \in \Gamma_\beta/\{\nu\}$ *and* $t \in [a,b]$*, where* λ_α *denotes the Ito coefficient function corresponding to the SDE (2.1).*

Proof. The identities (4.14)-(4.15) trivially hold, since (4.13) is the order-β weak Ito-Taylor expansion of the solution of the piecewise linear equation (2.8) with initial value $\mathbf{y}(a) = \mathbf{y}_0$.

By simple calculations it is obtained that Ito coefficient functions λ_α corresponding to the SDE (2.1) are

$$\begin{aligned}
\lambda^k_{(0)} &= \mathbf{f}^k, \\
\lambda^k_{(j)} &= \mathbf{g}^k_j, \\
\lambda^k_{(0,j)} &= \frac{\partial \mathbf{g}^k_j}{\partial t} + \sum_{i=1}^d \mathbf{f}^i \frac{\partial \mathbf{g}^k_j}{\partial \mathbf{x}^i} + \frac{1}{2}\sum_{i,l=1}^d \sum_{j=1}^m \mathbf{g}^i_j \mathbf{g}^l_j\ \frac{\partial^2 \mathbf{g}^k_j}{\partial \mathbf{x}^i \partial \mathbf{x}^l}, \\
\lambda^k_{(j,0)} &= \sum_{i=1}^d \mathbf{g}^i_j\ \frac{\partial \mathbf{f}^k}{\partial \mathbf{x}^i}, \\
\lambda^k_{(0,0)} &= \frac{\partial \mathbf{f}^k}{\partial t} + \sum_{i=1}^d \mathbf{f}^i \frac{\partial \mathbf{f}^k}{\partial \mathbf{x}^i} + \frac{1}{2}\sum_{i,l=1}^d \sum_{j=1}^m \mathbf{g}^i_j \mathbf{g}^l_j\ \frac{\partial^2 \mathbf{f}^k}{\partial \mathbf{x}^i \partial \mathbf{x}^l}, \\
\lambda^k_{(i,j)} &= \sum_{l=1}^d \mathbf{g}^l_i \frac{\partial \mathbf{g}^k_j}{\partial \mathbf{x}^l}
\end{aligned}$$

for $\alpha \in \Gamma_2$. By taking into account that $\mathbf{p}^\beta(s,\mathbf{v};\tau_n)$ and $\mathbf{q}^\beta_i(s,\mathbf{v};\tau_n)$ are linear functions of s and $\mathbf{v}$, it is not difficult to obtain that $\Lambda_\alpha(\tau_{n_s}, \mathbf{y}_{n_s}; \tau_{n_s}) = \lambda_\alpha(\tau_{n_s}, \mathbf{y}_{n_s})_{t_{n_s},s}$ for all $\alpha \in \Gamma_2$, which implies (4.16). ☐

Note that, the stochastic process $\mathbf{z}$ defined in the previous lemma is solution of the piecewise linear SDE (2.8) and Λ_α denotes the Ito coefficient functions corresponding to that equation. Therefore, (4.13) is the Ito-Taylor expansion of the Local Linear approximation (2.8).

The main convergence result for the WLL approximations is them stated in the following theorem.

THEOREM 4.4. *Let* $\mathbf{x}$ *be the solution of the SDE (2.1) on* $[a,b]$*, and* $\mathbf{y}$ *the order-*β *weak Local Linear approximation of* $\mathbf{x}$ *defined by (2.8). Suppose that the drift and diffusion coefficients of the SDE (2.1) satisfy the conditions (4.8)-(4.11). Further, suppose that the initial values of* $\mathbf{x}$ *and* $\mathbf{y}$ *satisfy the conditions*

$$E(|\mathbf{x}(a)|^q) < \infty$$

and

$$|E\left(g(\mathbf{x}(a))\right) - E\left(g(\mathbf{y}(a))\right)| \leq C_0 h^\beta$$

for $q = 1, 2, \ldots,$ *some constant* $C_0 > 0$ *and all* $g \in \mathcal{C}_P^{2(\beta+1)}(\mathbb{R}^d, \mathbb{R})$*. Then there exits a positive constant* C *such that*

$$\left| E\left(g(\mathbf{x}(b))\middle|\mathcal{F}_a\right) - E\left(g(\mathbf{y}(b))\middle|\mathcal{F}_a\right)\right| \leq C(b-a)h^\beta. \tag{4.17}$$

Proof. For $l = 1, 2, \ldots$, let $P_l = \{\mathbf{p} \in \{1, \ldots, d\}^l\}$, and let $\mathbf{F_p} : \mathbb{R}^d \to \mathbb{R}$ be the function defined as

$$\mathbf{F_p}(\mathbf{x}) = \prod_{i=1}^{l} \mathbf{x}^{p_i},$$

where $\mathbf{p} = (p_1, \ldots, p_l) \in P_l$.

By applying Lemma 5.11.7 in [33] to (2.8) and taking into account that (4.13) is the order-β weak Ito-Taylor expansion of the solution of (2.8), it is obtained

$$\left| E\left(\mathbf{F_p}(\mathbf{y}_{n+1} - \mathbf{y}_n) - \mathbf{F_p}(\sum_{\alpha \in \Gamma_\beta / \{\nu\}} I_\alpha[\Lambda_\alpha(\tau_n, \mathbf{y}_n; \tau_n)]_{\tau_n, \tau_{n+1}}) \middle| \mathcal{F}_{\tau_n}\right)\right| \leq K(1 + |\mathbf{y}_n|^{2r}) \cdot (\tau_{n+1} - \tau_n) h_n^\beta,$$

for all $\mathbf{p} \in P_l$ and $l = 1, \ldots, 2\beta + 1$, some $K > 0$ and $r \in \{1, 2, \ldots\}$, where Λ_α denotes the Ito coefficient function corresponding to (2.8), and $h_n = \tau_{n+1} - \tau_n$. Further, Lemma 4.3 implies that

$$E\left(\mathbf{F_p}(\sum_{\alpha \in \Gamma_\beta / \{\nu\}} I_\alpha[\lambda_\alpha(\tau_n, \mathbf{y}_n)]_{\tau_n, \tau_{n+1}}) \middle| \mathcal{F}_{\tau_n})\right) = E\left(\mathbf{F_p}(\sum_{\alpha \in \Gamma_\beta / \{\nu\}} I_\alpha[\Lambda_\alpha(\tau_n, \mathbf{y}_n; \tau_n)]_{\tau_n, \tau_{n+1}}) \middle| \mathcal{F}_{\tau_n}\right),$$

where λ_α denotes the Ito coefficient function corresponding to (2.1). Hence,

$$\begin{aligned}\left| E\left(\mathbf{F_p}(\mathbf{y}_{n+1} - \mathbf{y}_n) - \mathbf{F_p}(\sum_{\alpha \in \Gamma_\beta / \{\nu\}} I_\alpha[\lambda_\alpha(\tau_n, \mathbf{y}_n)]_{\tau_n, \tau_{n+1}}) \middle| \mathcal{F}_{\tau_n}\right)\right| &\leq K(1 + |\mathbf{y}_n|^{2r})(\tau_{n+1} - \tau_n) h_n^\beta \\ &\leq K(1 + \max_{0 \leq k \leq n} |\mathbf{y}_k|^{2r}) \\ &\quad \cdot (\tau_{n+1} - \tau_n) h_n^\beta.\end{aligned}$$

On the other hand, Theorem 4.5.4 in [33] applied to (2.8) and Lemma 4.2 imply

$$E\left(|\mathbf{y}_{n+1} - \mathbf{y}_n|^{2q} \middle| \mathcal{F}_{\tau_n}\right) \leq L(1 + \max_{0 \leq k \leq n} |\mathbf{y}_k|^{2q})(\tau_{n+1} - \tau_n)^q$$

for all $0 \leq n \leq N - 1$, and

$$E\left(\max_{0 \leq k \leq n_b} |\mathbf{y}_k|^{2q} \middle| \mathcal{F}_a\right) \leq C(1 + |\mathbf{y}_0|^{2q}),$$

respectively, where C and L are positive constants. The proof concludes by using Theorem 14.5.2 in [33] with the last three inequalities. $\square$

For state equations with additive noise, the order of weak convergence of the WLL approximations provided by this Theorem matches with that early obtained in [6].

Theorem 4.4 provides the global order of weak convergence for the WLL approximations at the time $t = b$. Notice further that inequality (4.17) implies that the uniform bound

$$\sup_{t\in[a,b]} \left| E\left(g(\mathbf{x}(t))\middle|\mathcal{F}_a\right) - E\left(g(\mathbf{y}(t))\middle|\mathcal{F}_a\right)\right| \leq C(b-a)h^{\beta} \tag{4.18}$$

holds as well for the order-β WLL approximation $\mathbf{y}$ since, in general, the global order of weak convergence of a numerical integrator implies the uniform one (see Theorem 14.5.1 and Exercise 14.5.3 in [33] for details).

Convergence in Theorem 4.4 has been proved under the assumption of continuity for $\mathbf{f}$ and $\mathbf{g}_i$. If that is not the case, the consistency of the WLL discretization has been proved in [58]. In other practical situations, it is important to integrate SDEs with nonglobal Lipschitz coefficients on $\mathbb{R}^d$ [35]. Typically, for such type of equations, the conventional weak integrators display explosive values for some realizations. In such a case, if each numerical realization of an order-β scheme leaving a sufficient large sphere $\mathcal{R} \subset \mathbb{R}^d$ is rejected, then Theorem 2.3 in [35] ensures that the accuracy of the scheme is $\varepsilon + O(h^{\beta})$, where ε can be made arbitrary small with increasing the sphere radius. This Theorem could be applied to the WLL approximations as well.

Finally, the rate of convergence of the approximate Local Linearization filter is states as follows.

THEOREM 4.5. *Given a set of M partial and noisy observations of the state equation (2.1) on $\{t\}_M$, and under the assumption that conditions (4.8)-(4.11) hold on $[t_0, T]$, the approximate order-β LL filter (4.1)-(4.6) defined on $(\tau)_h \supset \{t\}_M$ converges with order β to the exact LMV filter (2.3)-(2.7) as h goes to zero.*

Proof. Lemma 4.2 and Theorem 4.4 imply that the order-β LL approximation $\mathbf{y}$ of $\mathbf{x}$ defined by (2.8) satisfies the inequalities (4.12) and (4.18) for any integration interval $[a, b] \subset [t_0, T]$. Thus, by applying that lemma and theorem in each interval $[t_k, t_{k+1}]$ with $\mathbf{y}(t_k) \equiv \mathbf{y}_{t_k/t_k}$ (and $\mathbf{y}_{t_0/t_0} \equiv \mathbf{x}_{t_0/t_0}$), for all $t_k, t_{k+1} \in \{t\}_M$, the bound and convergence conditions (3.1) and (3.2) required by Theorem 3.2 for the convergence of the filter designed from $\mathbf{y}$ are satisfied. Therefore, the inequalities (3.7)-(3.9) hold for the approximate LL filter of the Definition 4.1, and so it has rate of convergence β when h goes to zero. $\square$

5. Practical Algorithms.

This section deals with practical implementation of the order-β LL filter (4.1)-(4.6). Explicit formulas for the predictions $\mathbf{y}_{t/t_k}$ and $\mathbf{P}_{t/t_k}$, an adaptive strategy for the construction of an adequate time discretization $(\tau)_h$, and the resulting adaptive LL filter algorithm are given.

5.1. Formulas for the predictions.

Let us define the vectors $\mathbf{a}_0(\tau)$, $\mathbf{a}_1(\tau)$, $\mathbf{b}_{i,0}(\tau)$ and $\mathbf{b}_{i,1}(\tau)$ satisfying the expressions

$$\mathbf{a}^{\beta}(t;\tau_{n_t}) = \mathbf{a}_0(\tau_{n_t}) + \mathbf{a}_1(\tau_{n_t})(t-\tau_{n_t}) \qquad \text{and} \qquad \mathbf{b}_i^{\beta}(t;\tau_{n_t}) = \mathbf{b}_{i,0}(\tau_{n_t}) + \mathbf{b}_{i,1}(\tau_{n_t})(t-\tau_{n_t})$$

for all $t \in [t_k, t_{k+1}]$, where the vector functions $\mathbf{a}^{\beta}$ and $\mathbf{b}_i^{\beta}$ are defined as in the WLL approximation (2.8) but, replacing $\mathbf{y}(s)$ by $\mathbf{y}_{s/t_k}$. By simplicity, the supraindex β is omitted in the right hand side of the above expressions.

According Theorem 3.1 in [21], the solution of the piecewise linear differential equations (4.1)-(4.2) for the predictions can be computed as

$$\mathbf{y}_{t/t_k} = \mathbf{y}_{t_k/t_k} + \sum_{n=n_{t_k}}^{n_t-1} \mathbf{L}_2 e^{\mathbf{M}(\tau_n)(\tau_{n+1}-\tau_n)}\mathbf{u}_{\tau_n,t_k} + \mathbf{L}_2 e^{\mathbf{M}(\tau_{n_t})(t-\tau_{n_t})}\mathbf{u}_{\tau_{n_t},t_k} \tag{5.1}$$

and

$$vec(\mathbf{P}_{t/t_k}) = \mathbf{L}_1 e^{\mathbf{M}(\tau_{n_t})(t-\tau_{n_t})}\mathbf{u}_{\tau_{n_t},t_k} \tag{5.2}$$

for all $t \in (t_k, t_{k+1}]$ and $t_k, t_{k+1} \in \{t\}_M$, where the vector $\mathbf{u}_{\tau,t_k}$ and the matrices $\mathbf{M}(\tau)$, $\mathbf{L}_1$, $\mathbf{L}_2$ are defined as

$$\mathbf{M}(\tau) = \begin{bmatrix} \mathcal{A}(\tau) & \mathcal{B}_5(\tau) & \mathcal{B}_4(\tau) & \mathcal{B}_3(\tau) & \mathcal{B}_2(\tau) & \mathcal{B}_1(\tau) \\ \mathbf{0} & \mathbf{C}(\tau) & \mathbf{I}_{d+2} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{C}(\tau) & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & 0 & 2 & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & 0 & 0 & 1 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{u}_{\tau,t_k} = \begin{bmatrix} vec(\mathbf{P}_{\tau/t_k}) \\ \mathbf{0} \\ \mathbf{r} \\ 0 \\ 0 \\ 1 \end{bmatrix} \in \mathbb{R}^{(d^2+2d+7)}$$

and

$$\mathbf{L}_1 = \begin{bmatrix} \mathbf{I}_{d^2} & \mathbf{0}_{d^2\times(2d+7)} \end{bmatrix}, \qquad \mathbf{L}_2 = \begin{bmatrix} \mathbf{0}_{d\times(d^2+d+2)} & \mathbf{I}_d & \mathbf{0}_{d\times 5} \end{bmatrix}$$

in terms of the matrices and vectors

$$\mathcal{A}(\tau) = \mathbf{A}(\tau)\oplus\mathbf{A}(\tau) + \sum_{i=1}^{m} \mathbf{B}_i(\tau)\otimes\mathbf{B}_i^{\intercal}(\tau),$$

$$\mathbf{C}(\tau) = \begin{bmatrix} \mathbf{A}(\tau) & \mathbf{a}_1(\tau) & \mathbf{A}(\tau)\mathbf{y}_{\tau/t_k} + \mathbf{a}_0(\tau) \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \in \mathbb{R}^{(d+2)\times(d+2)},$$

$$\mathbf{r}^{\intercal} = \begin{bmatrix} \mathbf{0}_{1\times(d+1)} & 1 \end{bmatrix}$$

$\mathcal{B}_1(\tau) = vec(\boldsymbol{\beta}_1(\tau)) + \boldsymbol{\beta}_4(\tau)\mathbf{y}_{\tau/t_k}$, $\mathcal{B}_2(\tau) = vec(\boldsymbol{\beta}_2(\tau)) + \boldsymbol{\beta}_5(\tau)\mathbf{y}_{\tau/t_k}$, $\mathcal{B}_3(\tau) = vec(\boldsymbol{\beta}_3(\tau))$, $\mathcal{B}_4(\tau) = \boldsymbol{\beta}_4(\tau)\mathbf{L}$ and $\mathcal{B}_5(\tau) = \boldsymbol{\beta}_5(\tau)\mathbf{L}$ with

$$\boldsymbol{\beta}_1(\tau) = \sum_{i=1}^{m} \mathbf{b}_{i,0}(\tau)\mathbf{b}_{i,0}^{\intercal}(\tau)$$

$$\boldsymbol{\beta}_2(\tau) = \sum_{i=1}^{m} \mathbf{b}_{i,0}(\tau)\mathbf{b}_{i,1}^{\intercal}(\tau) + \mathbf{b}_{i,1}(\tau)\mathbf{b}_{i,0}^{\intercal}(\tau)$$

$$\boldsymbol{\beta}_3(\tau) = \sum_{i=1}^{m} \mathbf{b}_{i,1}(\tau)\mathbf{b}_{i,1}^{\intercal}(\tau)$$

$$\boldsymbol{\beta}_4(\tau) = \mathbf{a}_0(\tau) \oplus \mathbf{a}_0(\tau) + \sum_{i=1}^{m} \mathbf{b}_{i,0}(\tau) \otimes \mathbf{B}_i(\tau) + \mathbf{B}_i(\tau) \otimes \mathbf{b}_{i,0}(\tau)$$

$$\boldsymbol{\beta}_5(\tau) = \mathbf{a}_1(\tau) \oplus \mathbf{a}_1(\tau) + \sum_{i=1}^{m} \mathbf{b}_{i,1}(\tau) \otimes \mathbf{B}_i(\tau) + \mathbf{B}_i(\tau) \otimes \mathbf{b}_{i,1}(\tau),$$

$\mathbf{L} = \begin{bmatrix} \mathbf{I}_d & \mathbf{0}_{d\times 2} \end{bmatrix}$, and the d-dimensional identity matrix $\mathbf{I}_d$. The matrix functions $\mathbf{A}, \mathbf{B}_i$ are defined as in the WLL approximation (2.8) but, replacing $\mathbf{y}(s)$ by $\mathbf{y}_{s/t_k}$. The symbols vec, $\oplus$ and $\otimes$ denote the vectorization operator, the Kronecker sum and product, respectively.

Alternatively, see Theorems 3.2 and 3.3 in [21] for simplified formulas in the case autonomous state equations or with additive noise.

5.2. Adaptive selection of a time discretization. In order to write a code that automatically determines a suitable time discretization $(\tau)_h$ for achieving a prescribed accuracy in the computation of the predictions $\mathbf{y}_{t_{k+1}/t_k}$ and $\mathbf{P}_{t_{k+1}/t_k}$, an adequate adaptive strategy is necessary. Since the equations (4.1)-(4.2) for the first two conditional moments of $\mathbf{y}$ are ordinary differential equations, conventional adaptive strategies for numerical integrators of such class of equations are useful. In what follows, the adaptive strategy described in [16] is adapted to the LL filter requirements.

Once the values for the relative and absolute tolerances $rtol_{\mathbf{y}}, rtol_{\mathbf{P}}$ and $atol_{\mathbf{y}}, atol_{\mathbf{P}}$ for the local errors of the first two conditional moments, for the maximum and minimum stepsizes $h_{\max}$ and $h_{\min}$, and for the floating point precision prs are set, an initial stepsize h_1 needs to be estimated. Specifically,

$$h_1 = \max\{h_{\min}, \min\{\delta(\mathbf{y}), \delta(vec(\mathbf{P})), t_1 - t_0\}\}$$

where

$$\delta(\mathbf{v}) = \min\{100\delta_1(\mathbf{v}), \delta_2(\mathbf{v})\}$$

with

$$\delta_1(\mathbf{v}) = \begin{cases} atol_{\mathbf{v}} & \text{if } d_0(\mathbf{v}) < 10 \cdot atol_{\mathbf{v}} \text{ or } d_1(\mathbf{v}) < 10 \cdot atol_{\mathbf{v}} \\ 0.01\frac{d_0(\mathbf{v})}{d_1(\mathbf{v})} & \text{otherwise} \end{cases}$$

and

$$\delta_2(\mathbf{v}) = \begin{cases} \max\{atol_{\mathbf{v}}, \delta_1 \cdot rtol_{\mathbf{v}}\}. & \text{if } max\{d_1(\mathbf{v}), d_2(\mathbf{v})\} \leq prs \\ (\dfrac{0.01}{\max\{d_1(\mathbf{v}), d_2(\mathbf{v})\}})^{\frac{1}{\beta+1}} & \text{otherwise} \end{cases}.$$

Here, $d_0(\mathbf{v}) = \left\|\mathbf{v}_{t_0/t_0}\right\|$, $d_1(\mathbf{v}) = \left\|\mathbf{F}(t_0, \mathbf{v}_{t_0/t_0})\right\|$ and $d_2(\mathbf{v}) = \left\|\dfrac{\partial \mathbf{F}(t_0, \mathbf{v}_{t_0/t_0})}{\partial t} + \dfrac{\partial \mathbf{F}(t_0, \mathbf{v}_{t_0/t_0})}{\partial \mathbf{v}} \mathbf{F}(t_0, \mathbf{v}_{t_0/t_0})\right\|$ are the norms of the filters and of their first two derivatives with respect to t at t_0, where $\mathbf{F}$ is the vector field of the equation for $\mathbf{v}$ (i.e., (4.1) for $\mathbf{y}$, and (4.2) for $\mathbf{P}$), and $\|\mathbf{v}\| = \sqrt{\dfrac{1}{\dim(\mathbf{v})} \sum_{i=1}^{\dim(\mathbf{v})} (\dfrac{\mathbf{v}^i}{\mathbf{sc}^i(\mathbf{v})})^2}$ with $\mathbf{sc}^i(\mathbf{v}) = atol_{\mathbf{v}} + rtol_{\mathbf{v}} \cdot \left|\mathbf{v}^i_{t_0/t_0}\right|$.

Starting with the filter estimates $\mathbf{y}_{t_k/t_k}$ and $\mathbf{P}_{t_k/t_k}$, the basic steps of the adaptive algorithm for determining $(\tau)_h$ and computing the predictions $\mathbf{y}_{t_{k+1}/t_k}$ and $\mathbf{P}_{t_{k+1}/t_k}$ between two consecutive observations t_k and t_{k+1} are the following:

1. Computation of $\mathbf{y}_{\tau_n/t_k}$ and $vec(\mathbf{P}_{\tau_n/t_k})$ at $\tau_n = \tau_{n-1} + 2h_n$ by the recursive evaluation of the expressions (5.1)-(5.2) at the two consecutive times $\tau_{n-1} + h_n$ and $(\tau_{n-1} + h_n) + h_n$. That is,

$$\mathbf{y}_{\tau_n/t_k} = \mathbf{y}_{\tau_{n-1}/t_k} + \mathbf{L}_2 e^{h_n \mathbf{M}(\tau_{n-1})} \mathbf{u}_{\tau_{n-1},t_k} + \mathbf{L}_2 e^{h_n \mathbf{M}(\tau_{n-1}+h_n)} \mathbf{u}_{\tau_{n-1}+h_n,t_k}$$

   and

$$vec(\mathbf{P}_{\tau_n/t_k}) = \mathbf{L}_1 e^{h_n \mathbf{M}(\tau_{n-1}+h_n)} \mathbf{u}_{\tau_{n-1}+h_n,t_k}.$$

2. Computation of an alternative estimate for the predictions at $\tau_n = \tau_{n-1} + 2h_n$ by means of the expressions

$$\widehat{\mathbf{y}}_{\tau_n/t_k} = \mathbf{y}_{\tau_{n-1}/t_k} + \mathbf{L}_2 e^{2h_n \mathbf{M}(\tau_{n-1})} \mathbf{u}_{\tau_{n-1},t_k}$$

   and

$$vec(\widehat{\mathbf{P}}_{\tau_n/t_k}) = \mathbf{L}_1 e^{2h_n \mathbf{M}(\tau_{n-1})} \mathbf{u}_{\tau_{n-1},t_k},$$

   which follow from the straightforward evaluation of (5.1)-(5.2) at $\tau_{n-1} + 2h_n$.
3. Evaluation of the error formulas

$$E_1 = \sqrt{\frac{1}{d} \sum_{i=1}^{d} (\frac{\mathbf{y}^i_{\tau_n/t_k} - \widehat{\mathbf{y}}^i_{\tau_n/t_k}}{\mathbf{sc}^i(\mathbf{y})})^2} \quad \text{and} \quad E_2 = \sqrt{\frac{1}{d^2} \sum_{i=1}^{d^2} (\frac{\mathbf{p}^i_{\tau_n/t_k} - \widehat{\mathbf{p}}^i_{\tau_n/t_k}}{\mathbf{sc}^i(\mathbf{p})})^2},$$

   where $\mathbf{p}_{\tau_n/t_k} = vec(\mathbf{P}_{\tau_n/t_k})$ and $\mathbf{sc}^i(\mathbf{v}) = atol_{\mathbf{v}} + rtol_{\mathbf{v}} \cdot \max\{\left|\mathbf{v}^i_{\tau_{n-1}/t_k}\right|, \left|\mathbf{v}^i_{\tau_n/t_k}\right|\}$.

4. Estimation of a new stepsize

$$h_{new} = \max\{h_{\min}, \min\{\delta_{new}(E_1), \delta_{new}(E_2)\}\}$$

where

$$\delta_{new}(E) = \begin{cases} h_n \cdot \min\{5, \max\{0.25, 0.8 \cdot (\frac{1}{E})^{\frac{1}{\beta+1}}\}\} & E \leq 1 \\ h_n \cdot \min\{1, \max\{0.1, 0.2 \cdot (\frac{1}{E})^{\frac{1}{\beta+1}}\}\} & E > 1 \end{cases}$$

5. Validation of $\mathbf{y}_{\tau_n/t_k}$ and $vec(\mathbf{P}_{\tau_n/t_k})$: if $max\{E_1, E_2\} \leq 1$ or $h_n = h_{\min}$, then accept $\mathbf{y}_{\tau_n/t_k}$ and $vec(\mathbf{P}_{\tau_n/t_k})$ as approximations to the first two conditional moments of $\mathbf{x}$ at $\tau_n = \tau_{n-1} + 2h_n$. Otherwise, return to step 1 with $h_n = h_{new}$.
6. Control of the final stepsize: if $\tau_n + 2h_n = t_{k+1}$, stop. If $\tau_n + 2h_n + h_{new} > t_{k+1}$, then redefine $h_{new} = t_{k+1} - (\tau_n + 2h_n)$.
7. Return to step 1 with $n = n + 1$ and $h_n = h_{new}$.

Clearly, in this adaptive strategy, the selected values for the relative and absolute tolerances will have a direct impact in the filtering performance expressed in terms of the filtering error and the computational time cost. Note that, under the assumed smoothness conditions for the first two conditional moments of the state equation, the adaptive algorithm provides an adequate estimation of the local errors of the approximate moments at each $\tau_n \in (\tau)_h$, and ensures that the relative and absolute errors of the approximate moments at τ_n are lower than the prearranged relative and absolute tolerance. This is done with a computational time cost that typically increases as the values of the tolerances decreases. Thus, for each filtering problem, adequate tolerance values should be carefully set in advance. In practical control engineering, these tolerances can be chosen by taking into account the level of accuracy required by the particular problem under consideration and the specific range of values of its state variables.

Remarks: It is worth to emphasize that the initial stepsize h_1 is computed just one time for computing the value of $\tau_1 \in [t_0, t_1]$. For other $\tau_n \in [t_k, t_{k+1}]$ with $n = n_{t_k} + 1$ and $k > 0$, the initial value for the corresponding h_n is set as $h_n = h_{new}$, where the value h_{new} was estimated when the previous stepsize h_{n-1} was accepted. Further note that, because the flow property of the exponential operator, only two exponential matrices need to be evaluated in steps 1 and 2, instead of three. These two exponential matrices can the efficiently computed through the well known Padé method for exponential matrices [37] or, alternatively, by means of the Krylov subspace method [37] in the case of high dimensional state equation. Even more, low order Padé and Krylov methods as suggested in [26] can be used as well for reducing the computation cost, but preserving the order-β of the LL filters. In step 4, the constant values in the formula for the new stepsize $\delta_{new}(E)$ were set according to the standard integration criteria oriented to reach an adequate balance of accuracy and computational cost with the adaptive strategy (see, e.g., [16]). These values might be adjusted for improving the filtering performance in some specific types of state equations.

5.3. Adaptive LL filter algorithm. Starting with the initial filter values $\mathbf{y}_{t_0/t_0} = \mathbf{x}_{t_0/t_0}$ and $\mathbf{P}_{t_0/t_0} = \mathbf{Q}_{t_0/t_0}$, the adaptive LL filter algorithm performs the recursive computation of:

1. the predictions $\mathbf{y}_{\tau_n/t_k}$ and $\mathbf{P}_{\tau_n/t_k}$ for all $\tau_n \in \{(\tau)_h \cap (t_k, t_{k+1}]\}$ by means of the recursive formulas and the adaptive strategy of the last two subsections, and the prediction variance by

$$\mathbf{V}_{t_{k+1}/t_k} = \mathbf{P}_{t_{k+1}/t_k} - \mathbf{y}_{t_{k+1}/t_k} \mathbf{y}_{t_{k+1}/t_k}^{\intercal};$$

2. the filters

$$\begin{aligned} \mathbf{y}_{t_{k+1}/t_{k+1}} &= \mathbf{y}_{t_{k+1}/t_k} + \mathbf{K}_{t_{k+1}}(\mathbf{z}_{t_{k+1}} - \mathbf{C}\mathbf{y}_{t_{k+1}/t_k}), \\ \mathbf{V}_{t_{k+1}/t_{k+1}} &= \mathbf{V}_{t_{k+1}/t_k} - \mathbf{K}_{t_{k+1}}\mathbf{C}\mathbf{V}_{t_{k+1}/t_k}, \\ \mathbf{P}_{t_{k+1}/t_{k+1}} &= \mathbf{V}_{t_{k+1}/t_{k+1}} + \mathbf{y}_{t_{k+1}/t_{k+1}} \mathbf{y}_{t_{k+1}/t_{k+1}}^{\intercal}, \end{aligned}$$

with filter gain

$$\mathbf{K}_{t_{k+1}} = \mathbf{V}_{t_{k+1}/t_k}\mathbf{C}^{\intercal}(\mathbf{C}\mathbf{V}_{t_{k+1}/t_k}\mathbf{C}^{\intercal} + \Sigma_{t_{k+1}})^{-1};$$

for each k, with $k = 0, 1, \ldots, M - 2$.

6. Numerical Simulations. In this section, the performance of the approximate LMV filters introduced in this paper is illustrated, by means of simulations, with four examples of state space models. To do so, the prediction and filter values are computed in four different ways by means of: 1) the exact LMV filter formulas, when it is possible; 2) the conventional LL filter; when the exact filter formulas are available; 3) the order-1 LL filter with various uniform time discretizations; and 4) the adaptive order-1 LL filter. For each example, the error analysis for the estimated moments and the estimation of the weak convergence rate are carried out through the standard procedures (see, e.g., [33, 6]).

The state space models to be considered are the followings.

Example 1. State equation with multiplicative noise

$$dx = atxdt + \sigma\sqrt{t}xdw_1 \tag{6.1}$$

and observation equation

$$z_{t_k} = x(t_k) + e_{t_k}, \text{ for } k = 0, 1, .., M-1 \tag{6.2}$$

with $a = -0.1$, $\sigma = 0.1$, $t_0 = 0.5$, $\Sigma = 0.0001$, $x_{t_0/t_0} = 1$ and $Q_{t_0/t_0} = 1$. For this state equation, the predictions for the first two moments are

$$x_{t_{k+1}/t_k} = x_{t_k/t_k} e^{a(t_{k+1}^2 - t_k^2)/2} \quad \text{and} \quad Q_{t_{k+1}/t_k} = Q_{t_k/t_k} e^{(a+\sigma^2/2)(t_{k+1}^2 - t_k^2)},$$

where the filters x_{t_k/t_k} and Q_{t_k/t_k} are obtained from (2.5) and (2.6) for all $k = 0, 1, .., M-2$.

Example 2. State equation with two additive noise

$$dx = atxdt + \sigma_1 t^p e^{at^2/2} dw_1 + \sigma_2\sqrt{t}dw_2 \tag{6.3}$$

and observation equation

$$z_{t_k} = x(t_k) + e_{t_k}, \text{ for } k = 0, 1, .., M-1 \tag{6.4}$$

with $a = -0.25$, $p = 2$, $\sigma_1 = 5$, $\sigma_2 = 0.1$, $t_0 = 0.01$, $\Sigma = 0.0001$, $x_{t_0/t_0} = 10$ and $Q_{t_0/t_0} = 100$. For this state equation, the predictions for the first two moments are

$$x_{t_{k+1}/t_k} = x_{t_k/t_k} e^{a(t_{k+1}^2 - t_k^2)/2}$$

and

$$Q_{t_{k+1}/t_k} = (Q_{t_k/t_k} + \frac{\sigma_2^2}{2a}) e^{a(t_{k+1}^2 - t_k^2)} + \frac{\sigma_1^2}{2p+1}(t_{k+1}^{2p+1} - t_k^{2p+1}) e^{at_{k+1}^2} - \frac{\sigma_2^2}{2a},$$

where the filters x_{t_k/t_k} and Q_{t_k/t_k} are obtained from (2.5) and (2.6) for all $k = 0, 1, .., M-2$.

Example 3. Van der Pool oscillator with random input [14]

$$dx_1 = x_2 dt \tag{6.5}$$
$$dx_2 = (-(x_1^2 - 1)x_2 - x_1 + a)dt + \sigma dw \tag{6.6}$$

and observation equation

$$z_{t_k} = x_1(t_k) + e_{t_k}, \text{ for } k = 0, 1, .., M-1, \tag{6.7}$$

where $a = 0.5$ and $\sigma^2 = (0.75)^2$ are the intensity and the variance of the random input, respectively. In addition, $t_0 = 0$, $\Sigma = 0.001$, $\mathbf{x}_{t_0/t_0}^{\intercal} = [1\ 1]$ and $\mathbf{Q}_{t_0/t_0} = \mathbf{x}_{t_0/t_0}\mathbf{x}_{t_0/t_0}^{\intercal}$.

Example 4. Van der Pool oscillator with random frequency [14]

$$dx_1 = x_2 dt \tag{6.8}$$
$$dx_2 = (-(x_1^2 - 1)x_2 - \varpi x_1)dt + \sigma x_1 dw \tag{6.9}$$

Table 6.1

Confidence limits for the errors between the exact LMV filter $\mathbf{x}_{t_{k+1}/t_{k+1}}, \mathbf{U}_{t_{k+1}/t_{k+1}}$ *of (6.1)-(6.2) and the order-*1 *LL filter* $\mathbf{y}^h_{t_{k+1}/t_{k+1}}, \mathbf{V}^h_{t_{k+1}/t_{k+1}}$ *on* $(\tau)^u_h$ *with different value of* h*. Order* $\widehat{\beta}$ *of weak convergence estimated from the errors.*

$\mathbf{y}^h_{t_{k+1}/t_{k+1}}$	$h = 1/64$	$h = 1/128$	$h = 1/256$	$h = 1/512$	$\widehat{\beta}$
t_1/t_1	$1.36 \pm 0.03 \times 10^{-5}$	$6.73 \pm 0.13 \times 10^{-6}$	$3.35 \pm 0.06 \times 10^{-6}$	$1.67 \pm 0.03 \times 10^{-6}$	1.00
t_2/t_2	$5.35 \pm 0.11 \times 10^{-6}$	$2.66 \pm 0.06 \times 10^{-6}$	$1.33 \pm 0.03 \times 10^{-6}$	$6.64 \pm 0.14 \times 10^{-7}$	1.00
t_3/t_3	$3.65 \pm 0.06 \times 10^{-6}$	$1.82 \pm 0.03 \times 10^{-6}$	$9.09 \pm 0.16 \times 10^{-7}$	$4.54 \pm 0.08 \times 10^{-7}$	1.00
t_4/t_4	$3.32 \pm 0.10 \times 10^{-6}$	$1.66 \pm 0.05 \times 10^{-6}$	$8.28 \pm 0.25 \times 10^{-7}$	$4.14 \pm 0.12 \times 10^{-7}$	1.00
t_5/t_4	$3.54 \pm 0.09 \times 10^{-6}$	$1.77 \pm 0.04 \times 10^{-6}$	$8.82 \pm 0.22 \times 10^{-7}$	$4.41 \pm 0.11 \times 10^{-7}$	1.00
t_6/t_6	$3.98 \pm 0.09 \times 10^{-6}$	$1.98 \pm 0.05 \times 10^{-6}$	$9.91 \pm 0.23 \times 10^{-7}$	$4.95 \pm 0.12 \times 10^{-7}$	1.00
t_7/t_7	$3.42 \pm 0.11 \times 10^{-6}$	$1.71 \pm 0.05 \times 10^{-6}$	$8.52 \pm 0.26 \times 10^{-7}$	$4.26 \pm 0.13 \times 10^{-7}$	1.00
t_8/t_8	$2.00 \pm 0.05 \times 10^{-6}$	$9.96 \pm 0.26 \times 10^{-7}$	$4.98 \pm 0.13 \times 10^{-7}$	$2.49 \pm 0.06 \times 10^{-7}$	1.01
t_9/t_9	$8.34 \pm 0.33 \times 10^{-7}$	$4.17 \pm 0.16 \times 10^{-7}$	$2.09 \pm 0.08 \times 10^{-7}$	$1.05 \pm 0.04 \times 10^{-7}$	1.01
$\mathbf{V}^h_{t_{k+1}/t_{k+1}}$	$h = 1/64$	$h = 1/128$	$h = 1/256$	$h = 1/512$	$\widehat{\beta}$
t_1/t_1	$2.47 \pm 0.06 \times 10^{-5}$	$1.23 \pm 0.03 \times 10^{-5}$	$6.12 \pm 0.14 \times 10^{-6}$	$3.05 \pm 0.07 \times 10^{-6}$	1.01
t_2/t_2	$8.08 \pm 0.20 \times 10^{-6}$	$4.03 \pm 0.10 \times 10^{-6}$	$2.01 \pm 0.05 \times 10^{-6}$	$1.00 \pm 0.03 \times 10^{-6}$	1.00
t_3/t_3	$4.06 \pm 0.11 \times 10^{-6}$	$2.03 \pm 0.06 \times 10^{-6}$	$1.01 \pm 0.03 \times 10^{-6}$	$5.05 \pm 0.14 \times 10^{-7}$	1.00
t_4/t_4	$2.36 \pm 0.07 \times 10^{-6}$	$1.18 \pm 0.03 \times 10^{-6}$	$5.87 \pm 0.17 \times 10^{-7}$	$2.93 \pm 0.08 \times 10^{-7}$	1.00
t_5/t_4	$1.52 \pm 0.05 \times 10^{-6}$	$7.60 \pm 0.24 \times 10^{-7}$	$3.78 \pm 0.12 \times 10^{-7}$	$1.89 \pm 0.06 \times 10^{-7}$	1.00
t_6/t_6	$9.36 \pm 0.30 \times 10^{-7}$	$4.66 \pm 0.15 \times 10^{-7}$	$2.33 \pm 0.07 \times 10^{-7}$	$1.16 \pm 0.04 \times 10^{-7}$	1.00
t_7/t_7	$4.49 \pm 0.22 \times 10^{-7}$	$2.23 \pm 0.10 \times 10^{-7}$	$1.11 \pm 0.05 \times 10^{-7}$	$5.55 \pm 0.27 \times 10^{-8}$	1.00
t_8/t_8	$1.32 \pm 0.07 \times 10^{-7}$	$6.55 \pm 0.35 \times 10^{-8}$	$3.26 \pm 0.17 \times 10^{-8}$	$1.63 \pm 0.09 \times 10^{-8}$	1.01
t_9/t_9	$2.42 \pm 0.19 \times 10^{-8}$	$1.19 \pm 0.09 \times 10^{-8}$	$5.94 \pm 0.46 \times 10^{-9}$	$2.96 \pm 0.23 \times 10^{-9}$	1.01

and observation equation

$$z_{t_k} = x_1(t_k) + e_{t_k}, \text{ for } k = 0, 1, .., M-1, \tag{6.10}$$

where $\varpi = 1$ and $\sigma^2 = 1$ are the frequency mean value and variance, respectively. In addition, $t_0 = 0$, $\Sigma = 0.001$, $\mathbf{x}^{\intercal}_{t_0/t_0} = [1\ 1]$ and $\mathbf{Q}_{t_0/t_0} = \mathbf{x}_{t_0/t_0}\mathbf{x}^{\intercal}_{t_0/t_0}$.

For each example, 2000 realizations of the state equation solution were computed by means of the Euler [33] or the Local Linearization scheme [26] for the equations with multiplicative or additive noise, respectively. For each example, the realizations were computed over the thin time partition $\{t_0 + n\delta : \delta = 10^{-4}, n = 0, .., 9 \times 10^4\}$ for guarantee a precise simulation of the stochastic solutions on the time interval $[t_0, t_0 + 9]$. A subsample of each realization at the time instants $\{t\}_{M=10} = \{t_k = t_0 + k : k = 0, .., M-1\}$ was taken to evaluate the corresponding observation equation. In this way, 2000 time series $\{z^i_{t_k}\}_{k=0,..,M-1}$, with $i = 1, ..2000$, of 10 values each one were finally available for every state space example.

For each time series of the first two examples, the values of the exact LMV filter, the conventional LL filter on $\{t\}_M$, the order-1 LL filter on uniform time discretization $(\tau)^u_h = \{\tau_n = t_0 + nh : n = 0, .., (M-1)/h\} \supset \{t\}_M$ with $h = 1/64, 1/128, 1/256, 1/512$, and the adaptive order-1 LL filter were computed.

For each time series $\{z^i_{t_k}\}_{k=0,..,M-1}$, four type of errors were evaluated: the errors $\left|\mathbf{x}^i_{t_{k+1}/t_{k+1}} - \mathbf{y}^i_{t_{k+1}/t_{k+1}}\right|$ and $\left|\mathbf{U}^i_{t_{k+1}/t_{k+1}} - \mathbf{V}^i_{t_{k+1}/t_{k+1}}\right|$ between each approximate filter and the exact one, and the errors $\left|\mathbf{x}^i_{t_{k+1}/t_k} - \mathbf{y}^i_{t_{k+1}/t_k}\right|$ and $\left|\mathbf{U}^i_{t_{k+1}/t_k} - \mathbf{V}^i_{t_{k+1}/t_k}\right|$ between the predictions, for all $k = 0, .., M-2$.

The 2000 errors of each type were arranged into $L = 20$ batches with $K = 100$ values each one, which are denoted by $\widehat{e}_{l,j}$, $l = 1, .., L$; $j = 1, ..., K$. Then, the sample mean of the l-*th* batch and of all batches

Table 6.2

Confidence limits for the errors between the exact LMV predictions $\mathbf{x}_{t_{k+1}/t_k}, \mathbf{U}_{t_{k+1}/t_k}$ *of (6.1)-(6.2) and their approximations* $\mathbf{y}^h_{t_{k+1}/t_k}, \mathbf{V}^h_{t_{k+1}/t_k}$ *obtained by the order-1 LL filter on* $(\tau)^u_h$ *with different value of h. Order* $\widehat{\beta}$ *of weak convergence estimated from the errors.*

$\mathbf{y}^h_{t_{k+1}/t_k}$	$h=1/64$	$h=1/128$	$h=1/256$	$h=1/512$	$\widehat{\beta}$
t_1/t_0	$7.35 \pm 0.00 \times 10^{-7}$	$1.84 \pm 0.00 \times 10^{-7}$	$4.60 \pm 0.00 \times 10^{-8}$	$1.15 \pm 0.00 \times 10^{-8}$	2.00
t_2/t_1	$1.11 \pm 0.02 \times 10^{-5}$	$5.52 \pm 0.10 \times 10^{-6}$	$2.74 \pm 0.05 \times 10^{-6}$	$1.37 \pm 0.03 \times 10^{-6}$	1.01
t_3/t_2	$4.22 \pm 0.09 \times 10^{-6}$	$2.02 \pm 0.04 \times 10^{-6}$	$9.95 \pm 0.21 \times 10^{-7}$	$4.94 \pm 0.10 \times 10^{-7}$	1.03
t_4/t_3	$2.75 \pm 0.05 \times 10^{-6}$	$1.27 \pm 0.02 \times 10^{-6}$	$6.20 \pm 0.11 \times 10^{-7}$	$3.07 \pm 0.05 \times 10^{-7}$	1.05
t_5/t_4	$2.20 \pm 0.06 \times 10^{-6}$	$1.03 \pm 0.03 \times 10^{-6}$	$5.07 \pm 0.15 \times 10^{-7}$	$2.52 \pm 0.08 \times 10^{-7}$	1.04
t_6/t_5	$2.06 \pm 0.05 \times 10^{-6}$	$9.88 \pm 0.26 \times 10^{-7}$	$4.88 \pm 0.12 \times 10^{-7}$	$2.43 \pm 0.06 \times 10^{-7}$	1.03
t_7/t_6	$2.02 \pm 0.05 \times 10^{-6}$	$9.93 \pm 0.24 \times 10^{-7}$	$4.94 \pm 0.12 \times 10^{-7}$	$2.46 \pm 0.06 \times 10^{-7}$	1.01
t_8/t_7	$1.57 \pm 0.05 \times 10^{-6}$	$7.74 \pm 0.24 \times 10^{-7}$	$3.84 \pm 0.12 \times 10^{-7}$	$1.92 \pm 0.06 \times 10^{-7}$	1.01
t_9/t_8	$8.18 \pm 0.22 \times 10^{-7}$	$4.06 \pm 0.11 \times 10^{-7}$	$2.03 \pm 0.05 \times 10^{-7}$	$1.01 \pm 0.03 \times 10^{-7}$	1.00
$\mathbf{V}^h_{t_{k+1}/t_k}$	$h=1/64$	$h=1/128$	$h=1/256$	$h=1/512$	$\widehat{\beta}$
t_1/t_0	$1.22 \pm 0.00 \times 10^{-4}$	$6.14 \pm 0.00 \times 10^{-5}$	$3.08 \pm 0.00 \times 10^{-5}$	$1.54 \pm 0.00 \times 10^{-5}$	1.01
t_2/t_1	$7.61 \pm 0.02 \times 10^{-5}$	$3.85 \pm 0.00 \times 10^{-5}$	$1.94 \pm 0.00 \times 10^{-5}$	$9.71 \pm 0.02 \times 10^{-6}$	1.00
t_3/t_2	$4.05 \pm 0.04 \times 10^{-5}$	$2.06 \pm 0.02 \times 10^{-5}$	$1.04 \pm 0.00 \times 10^{-5}$	$5.22 \pm 0.05 \times 10^{-6}$	1.00
t_4/t_3	$1.77 \pm 0.04 \times 10^{-5}$	$9.06 \pm 0.17 \times 10^{-6}$	$4.59 \pm 0.09 \times 10^{-6}$	$2.31 \pm 0.04 \times 10^{-6}$	1.00
t_5/t_4	$6.10 \pm 0.14 \times 10^{-6}$	$3.16 \pm 0.07 \times 10^{-6}$	$1.61 \pm 0.04 \times 10^{-6}$	$8.09 \pm 0.18 \times 10^{-7}$	1.00
t_6/t_5	$1.68 \pm 0.05 \times 10^{-6}$	$8.81 \pm 0.23 \times 10^{-7}$	$4.51 \pm 0.12 \times 10^{-7}$	$2.28 \pm 0.06 \times 10^{-7}$	1.00
t_7/t_6	$3.79 \pm 0.11 \times 10^{-7}$	$1.99 \pm 0.06 \times 10^{-7}$	$1.02 \pm 0.03 \times 10^{-7}$	$5.17 \pm 0.15 \times 10^{-8}$	1.00
t_8/t_7	$8.01 \pm 0.34 \times 10^{-8}$	$4.14 \pm 0.18 \times 10^{-8}$	$2.10 \pm 0.09 \times 10^{-8}$	$1.06 \pm 0.04 \times 10^{-8}$	1.00
t_9/t_8	$1.58 \pm 0.07 \times 10^{-8}$	$7.83 \pm 0.35 \times 10^{-9}$	$3.90 \pm 0.17 \times 10^{-9}$	$1.95 \pm 0.09 \times 10^{-9}$	1.00

can be computed by

$$\widehat{e}_l = \frac{1}{K}\sum_{j=1}^{K} \widehat{e}_{l,j}, \text{ and } \widehat{e} = \frac{1}{L}\sum_{l=1}^{L} \widehat{e}_l,$$

respectively. The confidence interval for each type of error is computed as

$$[\widehat{e} - \Delta, \widehat{e} + \Delta],$$

where

$$\Delta = t_{1-\alpha/2,L-1}\sqrt{\frac{\widehat{\sigma}_e^2}{L}}, \ \widehat{\sigma}_e^2 = \frac{1}{L-1}\sum_{i=1}^{L} \left|\widehat{e}_i - \widehat{e}\right|^2,$$

and $t_{1-\alpha/2,L-1}$ denotes the $1-\alpha/2$ percentile of the Student's t distribution with $L-1$ degrees for the significance level $0<\alpha<1$. The 90% confidence interval (i.e., the values Δ for $\alpha = 0.1$) was chosen.

6.1. Results for Example 1. Tables 6.1-6.3 show the estimated errors for the state space model (6.1)-(6.2). Specifically, Table 6.1 shows the confidence limits for the errors between the exact LMV filter $\mathbf{x}_{t_{k+1}/t_{k+1}}, \mathbf{U}_{t_{k+1}/t_{k+1}}$ and the order-1 LL filter $\mathbf{y}_{t_{k+1}/t_{k+1}}, \mathbf{V}_{t_{k+1}/t_{k+1}}$ on the time discretization $(\tau)^u_h$, with $h = 1/64, 1/128, 1/256, 1/512$. Table 6.2 shows the confidence limits for the errors between the exact LMV predictions $\mathbf{x}_{t_{k+1}/t_k}, \mathbf{U}_{t_{k+1}/t_k}$ and their approximations $\mathbf{y}_{t_{k+1}/t_k}, \mathbf{V}_{t_{k+1}/t_k}$ obtained by the order-1 LL filter on $(\tau)^u_h$. Table 6.3 shows the confidence limits for the errors between the moments of the exact LMV filter and their respective approximations obtained by the conventional LL filter and the adaptive LL filter. The average of accepted and fail steps of the adaptive LL filter at each $t_k \in \{t\}_M$ is given in Figure 6.1. The absolute and relative tolerances for the first and second moments were set as

Table 6.3

Confidence limits for the errors between the exact LMV filter and predictions of (6.1)-(6.2) with their corresponding approximations obtained by the conventional LL filter and the adaptive LL filter, which are denoted with superscripts 0 *and* A*, respectively.*

k	$\mathbf{y}^0_{t_{k+1}/t_k}$	$\mathbf{y}^A_{t_{k+1}/t_k}$	$\mathbf{V}^0_{t_{k+1}/t_k}$	$\mathbf{V}^A_{t_{k+1}/t_k}$
0	$2.79 \pm 0.00 \times 10^{-3}$	$5.09 \pm 0.00 \times 10^{-10}$	$1.75 \pm 0.00 \times 10^{-3}$	$3.23 \pm 0.00 \times 10^{-6}$
1	$5.62 \pm 0.13 \times 10^{-3}$	$2.86 \pm 0.05 \times 10^{-7}$	$5.42 \pm 0.15 \times 10^{-3}$	$2.09 \pm 0.00 \times 10^{-6}$
2	$6.09 \pm 0.05 \times 10^{-3}$	$1.06 \pm 0.02 \times 10^{-8}$	$4.04 \pm 0.07 \times 10^{-3}$	$1.16 \pm 0.01 \times 10^{-6}$
3	$5.74 \pm 0.06 \times 10^{-3}$	$6.75 \pm 0.12 \times 10^{-8}$	$3.16 \pm 0.07 \times 10^{-3}$	$5.29 \pm 0.10 \times 10^{-7}$
4	$4.54 \pm 0.05 \times 10^{-3}$	$5.73 \pm 0.17 \times 10^{-8}$	$1.70 \pm 0.04 \times 10^{-3}$	$1.92 \pm 0.04 \times 10^{-7}$
5	$3.17 \pm 0.04 \times 10^{-3}$	$5.72 \pm 0.15 \times 10^{-8}$	$7.21 \pm 0.20 \times 10^{-4}$	$5.62 \pm 0.15 \times 10^{-8}$
6	$2.01 \pm 0.03 \times 10^{-3}$	$6.07 \pm 0.14 \times 10^{-8}$	$2.44 \pm 0.07 \times 10^{-4}$	$1.34 \pm 0.04 \times 10^{-8}$
7	$1.24 \pm 0.02 \times 10^{-3}$	$5.02 \pm 0.15 \times 10^{-8}$	$7.37 \pm 0.27 \times 10^{-5}$	$2.85 \pm 0.12 \times 10^{-9}$
8	$7.32 \pm 0.14 \times 10^{-4}$	$2.82 \pm 0.07 \times 10^{-8}$	$1.81 \pm 0.09 \times 10^{-5}$	$5.39 \pm 0.23 \times 10^{-10}$
k	$\mathbf{y}^0_{t_{k+1}/t_{k+1}}$	$\mathbf{y}^A_{t_{k+1}/t_{k+1}}$	$\mathbf{V}^0_{t_{k+1}/t_{k+1}}$	$\mathbf{V}^A_{t_{k+1}/t_{k+1}}$
0	$3.94 \pm 0.08 \times 10^{-3}$	$3.50 \pm 0.07 \times 10^{-7}$	$7.22 \pm 0.17 \times 10^{-3}$	$6.38 \pm 0.14 \times 10^{-7}$
1	$6.25 \pm 0.13 \times 10^{-4}$	$1.43 \pm 0.03 \times 10^{-7}$	$9.61 \pm 0.25 \times 10^{-4}$	$2.16 \pm 0.05 \times 10^{-7}$
2	$3.58 \pm 0.07 \times 10^{-4}$	$1.01 \pm 0.02 \times 10^{-7}$	$4.12 \pm 0.12 \times 10^{-4}$	$1.12 \pm 0.03 \times 10^{-7}$
3	$3.09 \pm 0.09 \times 10^{-4}$	$9.44 \pm 0.29 \times 10^{-8}$	$2.33 \pm 0.07 \times 10^{-4}$	$6.69 \pm 0.19 \times 10^{-8}$
4	$3.50 \pm 0.09 \times 10^{-4}$	$1.04 \pm 0.03 \times 10^{-7}$	$1.64 \pm 0.06 \times 10^{-4}$	$4.45 \pm 0.14 \times 10^{-8}$
5	$4.49 \pm 0.12 \times 10^{-4}$	$1.22 \pm 0.03 \times 10^{-7}$	$1.16 \pm 0.04 \times 10^{-4}$	$2.86 \pm 0.09 \times 10^{-8}$
6	$5.93 \pm 0.10 \times 10^{-4}$	$1.12 \pm 0.03 \times 10^{-7}$	$7.96 \pm 0.28 \times 10^{-5}$	$1.45 \pm 0.07 \times 10^{-8}$
7	$6.61 \pm 0.12 \times 10^{-4}$	$6.93 \pm 0.18 \times 10^{-8}$	$3.92 \pm 0.17 \times 10^{-5}$	$4.50 \pm 0.23 \times 10^{-9}$
8	$5.91 \pm 0.09 \times 10^{-4}$	$2.89 \pm 0.11 \times 10^{-8}$	$1.48 \pm 0.06 \times 10^{-5}$	$8.17 \pm 0.63 \times 10^{-10}$

$rtol_{\mathbf{y}} = rtol_{\mathbf{P}} = 5 \times 10^{-9}$ and $atol_{\mathbf{y}} = 5 \times 10^{-9}$, $atol_{\mathbf{P}} = 5 \times 10^{-12}$. Note as the accuracy of the LL filter on uniform discretizations $(\tau)^u_h$ improve as h decreases, and the large difference among the accuracy of the conventional and the adaptive LL filter.

For each approximate conditional moment, the estimated order $\widehat{\beta}$ of weak convergence were obtained as the slope of the straight line fitted to the set of four points $\{\log_2(h_j),\, \log_2(\widehat{e}(h_j))\}_{j=1,..,4}$ taken from their corresponding errors tables 6.1 and 6.2. The values $\widehat{\beta}$ are shown in these tables as well. The estimates $\widehat{\beta} \approx 1$ corroborate the theoretical value for β given in Theorem 4.5. The estimate $\widehat{\beta} = 2.00$ corresponding to $\mathbf{y}^h_{t_1/t_0}$ in Table 6.2 agrees with the expected estimate of β for the equation (6.1) on $[t_0,t_1]$. In this particular situation, the exact prediction $\mathbf{x}_{t_1/t_0}$ given by (2.3) reduces to an ordinary differential equation and the LL prediction formula (4.1) reduces to the classical order-2 LL integrator for such class of equations (see, e.g., [24]). In the others subintervals $[t_k,t_{k+1}]$ with $k \neq 0$, the prediction $\mathbf{y}^h_{t_{k+1}/t_k}$ depends nonlinearly of $\mathbf{y}$ through the initial value $\mathbf{y}^h_{t_{k+1}/t_{k+1}}$.

6.2. Results for Example 2. Tables 6.4-6.6 show the estimated errors for the state space model (6.3)-(6.4). In particular, Table 6.4 shows the confidence limits for the errors between the exact LMV filter $\mathbf{x}_{t_{k+1}/t_{k+1}}, \mathbf{U}_{t_{k+1}/t_{k+1}}$ and the order-1 LL filter $\mathbf{y}_{t_{k+1}/t_{k+1}}, \mathbf{V}_{t_{k+1}/t_{k+1}}$ on the time discretization $(\tau)^u_h$, with $h = 1/64, 1/128, 1/256, 1/512$. Table 6.5 shows the confidence limits for the errors between the exact LMV predictions $\mathbf{x}_{t_{k+1}/t_k}$, $\mathbf{U}_{t_{k+1}/t_k}$ and their approximations $\mathbf{y}_{t_{k+1}/t_k}$, $\mathbf{V}_{t_{k+1}/t_k}$ obtained by the order-1 LL filter on $(\tau)^u_h$. Table 6.6 shows the confidence limits for the errors between the moments of the exact LMV filter and their respective approximations obtained by the conventional LL filter and the adaptive LL filter. The average of accepted and fail steps of the adaptive LL filter at each $t_k \in \{t\}_M$ is given in Figure 6.1. The absolute and relative tolerances for the first and second moments for this filter were set as $rtol_{\mathbf{y}} = rtol_{\mathbf{P}} = 5 \times 10^{-8}$ and $atol_{\mathbf{y}} = 5 \times 10^{-8}$, $atol_{\mathbf{P}} = 5 \times 10^{-11}$. Note as the accuracy of the LL filter on uniform discretizations $(\tau)^u_h$ improve as h decreases, and the large difference among the accuracy of the conventional and the adaptive LL filter.

For each approximate conditional moment, the estimated order $\widehat{\beta}$ of weak convergence were obtained as the slope of the straight line fitted to the set of four points $\{\log_2(h_j),\quad \log_2(\widehat{e}(h_j))\}_{j=1,..,4}$ taken

TABLE 6.4

Confidence limits for the errors between the exact LMV filter $\mathbf{x}_{t_{k+1}/t_{k+1}}, \mathbf{U}_{t_{k+1}/t_{k+1}}$ of (6.3)-(6.4) and the order-1 LL filter $\mathbf{y}^h_{t_{k+1}/t_{k+1}}, \mathbf{V}^h_{t_{k+1}/t_{k+1}}$ on $(\tau)^u_h$ with different value of h. Order $\widehat{\beta}$ of weak convergence estimated from the errors.

$\mathbf{y}^h_{t_{k+1}/t_{k+1}}$	$h = 1/64$	$h = 1/128$	$h = 1/256$	$h = 1/512$	$\widehat{\beta}$
t_1/t_1	$2.00 \pm 0.04 \times 10^{-8}$	$1.17 \pm 0.02 \times 10^{-8}$	$6.23 \pm 0.11 \times 10^{-9}$	$3.22 \pm 0.06 \times 10^{-9}$	0.95
t_2/t_2	$1.31 \pm 0.03 \times 10^{-8}$	$6.44 \pm 0.14 \times 10^{-8}$	$3.20 \pm 0.07 \times 10^{-9}$	$1.59 \pm 0.04 \times 10^{-9}$	1.02
t_3/t_3	$1.12 \pm 0.03 \times 10^{-8}$	$5.52 \pm 0.14 \times 10^{-8}$	$2.74 \pm 0.06 \times 10^{-9}$	$1.36 \pm 0.03 \times 10^{-9}$	1.02
t_4/t_4	$1.56 \pm 0.03 \times 10^{-8}$	$7.74 \pm 0.12 \times 10^{-8}$	$3.85 \pm 0.06 \times 10^{-9}$	$1.92 \pm 0.03 \times 10^{-9}$	1.01
t_5/t_4	$2.95 \pm 0.07 \times 10^{-8}$	$1.47 \pm 0.03 \times 10^{-8}$	$7.38 \pm 0.17 \times 10^{-9}$	$3.69 \pm 0.08 \times 10^{-9}$	1.01
t_6/t_6	$7.85 \pm 0.19 \times 10^{-8}$	$3.98 \pm 0.09 \times 10^{-8}$	$2.01 \pm 0.05 \times 10^{-8}$	$1.01 \pm 0.02 \times 10^{-8}$	0.99
t_7/t_7	$2.65 \pm 0.06 \times 10^{-7}$	$1.37 \pm 0.03 \times 10^{-7}$	$6.94 \pm 0.15 \times 10^{-8}$	$3.50 \pm 0.07 \times 10^{-8}$	0.99
t_8/t_8	$5.46 \pm 0.16 \times 107$	$2.79 \pm 0.08 \times 10^{-7}$	$1.41 \pm 0.04 \times 10^{-7}$	$7.09 \pm 0.21 \times 10^{-8}$	0.99
t_9/t_9	$4.76 \pm 0.13 \times 10^{-7}$	$2.37 \pm 0.06 \times 10^{-7}$	$1.18 \pm 0.03 \times 10^{-7}$	$5.91 \pm 0.16 \times 10^{-8}$	1.01
$\mathbf{V}^h_{t_{k+1}/t_{k+1}}$	$h = 1/64$	$h = 1/128$	$h = 1/256$	$h = 1/512$	$\widehat{\beta}$
t_1/t_1	$3.48 \pm 0.09 \times 10^{-7}$	$2.03 \pm 0.05 \times 10^{-7}$	$1.09 \pm 0.03 \times 10^{-7}$	$5.60 \pm 0.14 \times 10^{-8}$	0.88
t_2/t_2	$2.66 \pm 0.11 \times 10^{-7}$	$1.31 \pm 0.05 \times 10^{-7}$	$6.51 \pm 0.26 \times 10^{-8}$	$3.24 \pm 0.13 \times 10^{-8}$	1.01
t_3/t_3	$2.97 \pm 0.12 \times 10^{-7}$	$1.46 \pm 0.06 \times 10^{-7}$	$7.24 \pm 0.30 \times 10^{-8}$	$3.61 \pm 0.15 \times 10^{-8}$	1.01
t_4/t_4	$3.46 \pm 0.11 \times 10^{-7}$	$1.71 \pm 0.05 \times 10^{-7}$	$8.53 \pm 0.27 \times 10^{-8}$	$4.26 \pm 0.13 \times 10^{-8}$	1.01
t_5/t_4	$3.44 \pm 0.16 \times 10^{-7}$	$1.73 \pm 0.08 \times 10^{-7}$	$8.65 \pm 0.41 \times 10^{-8}$	$4.33 \pm 0.21 \times 10^{-8}$	1.01
t_6/t_6	$3.58 \pm 0.15 \times 10^{-7}$	$1.83 \pm 0.07 \times 10^{-7}$	$9.21 \pm 0.38 \times 10^{-8}$	$4.63 \pm 0.19 \times 10^{-8}$	0.98
t_7/t_7	$3.57 \pm 0.14 \times 10^{-7}$	$1.85 \pm 0.07 \times 10^{-7}$	$9.42 \pm 0.38 \times 10^{-8}$	$4.75 \pm 0.19 \times 10^{-8}$	0.97
t_8/t_8	$2.35 \pm 0.13 \times 10^{-7}$	$1.21 \pm 0.07 \times 10^{-7}$	$6.11 \pm 0.34 \times 10^{-8}$	$3.08 \pm 0.17 \times 10^{-8}$	0.98
t_9/t_9	$1.67 \pm 0.09 \times 10^{-7}$	$8.31 \pm 0.04 \times 10^{-8}$	$4.15 \pm 0.22 \times 10^{-8}$	$2.07 \pm 0.11 \times 10^{-8}$	1.00

from their corresponding errors tables 6.4 and 6.5. The values $\widehat{\beta}$ are included in these tables too. The estimates $\widehat{\beta} \approx 1$ corroborate the theoretical value for β given in Theorem 4.5. The estimate $\widehat{\beta} \approx 2.00$ corresponding to $\mathbf{y}^h_{t_{k+1}/t_k}$ in Table 6.5 agrees with the expected estimate of β for the equation (6.3) on $[t_k, t_{k+1}]$, for all k. Similarly to the previous example, the exact prediction $\mathbf{x}_{t_{k+1}/t_k}$ given by (2.3) reduces to an ordinary differential equation and the LL prediction formula (4.1) reduces as well to the classical order-2 LL integrator for all k. Contrary to the first example, in this one, the prediction $\mathbf{y}^h_{t_{k+1}/t_k}$ with $k \neq 0$ does not depend of $\mathbf{y}$ through the initial value $\mathbf{y}^h_{t_{k+1}/t_{k+1}}$ and so the estimate $\widehat{\beta} \approx 2.00$ is preserved.

6.3. Results for Examples 3 and 4. Since explicit formulas of the LMV filter for the state space models (6.5)-(6.7) and (6.8)-(6.10) are not available, the error analysis of the previous examples should be adjusted. In this situation, by taking into account the results of the previous examples, the moments estimated by the adaptive LL filter with small tolerance can be used as a precise estimation for the moments of the exact LMV filter. By doing this, the confidence interval for the errors can similarly be computed as before for estimate the order $\widehat{\beta}$ of weak convergence of the order-1 LL filter. Table 6.7 shows the estimated order $\widehat{\beta}$ of weak convergence obtained, as explained above, as the slope of the straight line fitted to the set of four points $\{\log_2(h_j), \quad \log_2(\widehat{e}(h_j))\}_{j=1,..,4}$, where $\widehat{e}(h_j)$ denotes the error between the order-1 LL filter on $(\tau)^u_{h_j}$, with $h_j = 1/2^{5+j}$, and the adaptive LL filter with small tolerance. The tolerances for the adaptive filter were set as $rtol_{\mathbf{y}} = rtol_{\mathbf{P}} = 5 \times 10^{-8}$ and $atol_{\mathbf{y}} = 5 \times 10^{-8}$, $atol_{\mathbf{P}} = 5 \times 10^{-11}$ in the model (6.5)-(6.7), and as $rtol_{\mathbf{y}} = rtol_{\mathbf{P}} = 10^{-7}$ and $atol_{\mathbf{y}} = 10^{-7}$, $atol_{\mathbf{P}} = 10^{-10}$ in the model (6.8)-(6.10). For each model, the average of accepted and fail steps of the adaptive LL filter at each $t_k \in \{t\}_M$ is given in Figure 6.1. Notice that, for both examples, the estimates $\widehat{\beta} \approx 1$ corroborate the theoretical value for β stated in Theorem 4.5.

6.4. Supplementary simulations. As mentioned above, the approximate LMV filters play a central role in the effective implementation of the innovation method for the parameter estimation of diffusion

TABLE 6.5

Confidence limits for the errors between the exact LMV predictions $\mathbf{x}_{t_{k+1}/t_k}, \mathbf{U}_{t_{k+1}/t_k}$ of (6.3)-(6.4) and their approximations $\mathbf{y}^h_{t_{k+1}/t_k}, \mathbf{V}^h_{t_{k+1}/t_k}$ obtained by the order-1 LL filter on $(\tau)^u_h$ with different value of h. Order $\widehat{\beta}$ of weak convergence estimated from the errors.

$\mathbf{y}^h_{t_{k+1}/t_k}$	$h=1/64$	$h=1/128$	$h=1/256$	$h=1/512$	$\widehat{\beta}$
t_1/t_0	$2.28 \pm 0.00 \times 10^{-5}$	$5.70 \pm 0.00 \times 10^{-6}$	$1.43 \pm 0.00 \times 10^{-6}$	$3.57 \pm 0.00 \times 10^{-7}$	2.00
t_2/t_1	$4.63 \pm 0.03 \times 10^{-5}$	$1.16 \pm 0.00 \times 10^{-5}$	$2.89 \pm 0.02 \times 10^{-6}$	$7.22 \pm 0.05 \times 10^{-7}$	2.00
t_3/t_2	$5.44 \pm 0.11 \times 10^{-5}$	$1.36 \pm 0.03 \times 10^{-5}$	$3.39 \pm 0.07 \times 10^{-6}$	$8.47 \pm 0.17 \times 10^{-7}$	2.00
t_4/t_3	$6.91 \pm 0.13 \times 10^{-5}$	$1.72 \pm 0.03 \times 10^{-5}$	$4.30 \pm 0.08 \times 10^{-6}$	$1.07 \pm 0.02 \times 10^{-6}$	2.00
t_5/t_4	$5.85 \pm 0.12 \times 10^{-5}$	$1.46 \pm 0.03 \times 10^{-5}$	$3.64 \pm 0.08 \times 10^{-6}$	$9.09 \pm 0.19 \times 10^{-7}$	2.00
t_6/t_5	$3.10 \pm 0.08 \times 10^{-5}$	$7.73 \pm 0.21 \times 10^{-6}$	$1.93 \pm 0.05 \times 10^{-6}$	$4.81 \pm 0.13 \times 10^{-7}$	2.00
t_7/t_6	$1.13 \pm 0.03 \times 10^{-5}$	$2.82 \pm 0.06 \times 10^{-6}$	$7.01 \pm 0.16 \times 10^{-7}$	$1.74 \pm 0.04 \times 10^{-7}$	2.01
t_8/t_7	$3.07 \pm 0.07 \times 10^{-6}$	$7.56 \pm 0.18 \times 10^{-7}$	$1.84 \pm 0.04 \times 10^{-7}$	$4.40 \pm 0.11 \times 10^{-8}$	2.04
t_9/t_8	$7.63 \pm 0.24 \times 10^{-7}$	$1.75 \pm 0.05 \times 10^{-7}$	$3.73 \pm 0.11 \times 10^{-8}$	$6.97 \pm 0.16 \times 10^{-9}$	2.25
$\mathbf{V}^h_{t_{k+1}/t_k}$	$h=1/64$	$h=1/128$	$h=1/256$	$h=1/512$	$\widehat{\beta}$
t_1/t_0	$2.43 \pm 0.00 \times 10^{-3}$	$1.28 \pm 0.00 \times 10^{-3}$	$6.56 \pm 0.00 \times 10^{-4}$	$3.32 \pm 0.00 \times 10^{-4}$	0.88
t_2/t_1	$7.22 \pm 0.00 \times 10^{-2}$	$3.54 \pm 0.00 \times 10^{-2}$	$1.75 \pm 0.00 \times 10^{-2}$	$8.73 \pm 0.00 \times 10^{-3}$	1.01
t_3/t_2	$1.69 \pm 0.00 \times 10^{-1}$	$8.29 \pm 0.00 \times 10^{-2}$	$4.11 \pm 0.00 \times 10^{-2}$	$2.04 \pm 0.00 \times 10^{-2}$	1.01
t_4/t_3	$1.16 \pm 0.00 \times 10^{-1}$	$5.73 \pm 0.00 \times 10^{-2}$	$2.84 \pm 0.00 \times 10^{-2}$	$1.42 \pm 0.00 \times 10^{-2}$	1.01
t_5/t_4	$3.38 \pm 0.00 \times 10^{-2}$	$1.68 \pm 0.00 \times 10^{-2}$	$8.36 \pm 0.00 \times 10^{-3}$	$4.17 \pm 0.00 \times 10^{-3}$	1.00
t_6/t_5	$4.81 \pm 0.00 \times 10^{-3}$	$2.41 \pm 0.00 \times 10^{-3}$	$1.21 \pm 0.00 \times 10^{-3}$	$6.05 \pm 0.00 \times 10^{-4}$	0.99
t_7/t_6	$3.77 \pm 0.00 \times 10^{-4}$	$1.91 \pm 0.00 \times 10^{-4}$	$9.62 \pm 0.00 \times 10^{-5}$	$4.83 \pm 0.00 \times 10^{-5}$	0.97
t_8/t_7	$3.27 \pm 0.00 \times 10^{-5}$	$1.65 \pm 0.00 \times 10^{-5}$	$8.28 \pm 0.00 \times 10^{-6}$	$4.15 \pm 0.00 \times 10^{-6}$	0.98
t_9/t_8	$1.70 \pm 0.00 \times 10^{-5}$	$8.44 \pm 0.00 \times 10^{-6}$	$4.21 \pm 0.00 \times 10^{-6}$	$2.10 \pm 0.00 \times 10^{-6}$	1.00

TABLE 6.6

Confidence limits for the errors between the exact LMV filter and predictions of (6.3)-(6.4) with their corresponding approximations obtained by the conventional LL filter and the adaptive LL filter, which are denoted with superscripts 0 and A, respectively.

k	$\mathbf{y}^0_{t_{k+1}/t_k}$	$\mathbf{y}^A_{t_{k+1}/t_k}$	$\mathbf{V}^0_{t_{k+1}/t_k}$	$\mathbf{V}^A_{t_{k+1}/t_k}$
0	$7.69 \pm 0.00 \times 10^{-2}$	$2.17 \pm 0.00 \times 10^{-6}$	2.63 ± 0.00	$3.72 \pm 0.00 \times 10^{-4}$
1	$2.09 \pm 0.01 \times 10^{-1}$	$2.14 \pm 0.04 \times 10^{-7}$	8.01 ± 0.03	$1.85 \pm 0.00 \times 10^{-3}$
2	$2.81 \pm 0.06 \times 10^{-1}$	$8.41 \pm 0.38 \times 10^{-8}$	$4.93 \pm 0.13 \times 10^{2}$	$3.24 \pm 0.02 \times 10^{-3}$
3	$4.02 \pm 0.07 \times 10^{-1}$	$1.26 \pm 0.07 \times 10^{-7}$	$3.22 \pm 0.17 \times 10^{2}$	$2.33 \pm 0.02 \times 10^{-3}$
4	$3.82 \pm 0.08 \times 10^{-1}$	$1.55 \pm 0.08 \times 10^{-7}$	$6.18 \pm 0.10 \times 10^{1}$	$7.45 \pm 0.06 \times 10^{-4}$
5	$2.27 \pm 0.06 \times 10^{-1}$	$1.06 \pm 0.06 \times 10^{-7}$	$6.23 \pm 0.25 \times 10^{-1}$	$1.15 \pm 0.01 \times 10^{-4}$
6	$9.36 \pm 0.21 \times 10^{-2}$	$4.68 \pm 0.22 \times 10^{-8}$	$8.34 \pm 0.25 \times 10^{-2}$	$9.74 \pm 0.11 \times 10^{-6}$
7	$2.89 \pm 0.07 \times 10^{-2}$	$1.23 \pm 0.08 \times 10^{-8}$	$8.19 \pm 0.28 \times 10^{-3}$	$7.43 \pm 0.08 \times 10^{-7}$
8	$8.63 \pm 0.28 \times 10^{-3}$	$1.10 \pm 0.06 \times 10^{-9}$	$2.73 \pm 0.02 \times 10^{-3}$	$2.86 \pm 0.00 \times 10^{-7}$
k	$\mathbf{y}^0_{t_{k+1}/t_{k+1}}$	$\mathbf{y}^A_{t_{k+1}/t_{k+1}}$	$\mathbf{V}^0_{t_{k+1}/t_{k+1}}$	$\mathbf{V}^A_{t_{k+1}/t_{k+1}}$
0	$1.75 \pm 0.03 \times 10^{-3}$	$3.29 \pm 0.06 \times 10^{-9}$	$3.08 \pm 0.08 \times 10^{-2}$	$5.73 \pm 0.14 \times 10^{-8}$
1	$9.47 \pm 0.24 \times 10^{-7}$	$3.37 \pm 0.07 \times 10^{-10}$	$1.62 \pm 0.07 \times 10^{-5}$	$6.84 \pm 0.26 \times 10^{-9}$
2	$2.17 \pm 0.06 \times 10^{-6}$	$2.16 \pm 0.05 \times 10^{-10}$	$5.58 \pm 0.23 \times 10^{-5}$	$5.75 \pm 0.25 \times 10^{-9}$
3	$2.73 \pm 0.04 \times 10^{-6}$	$3.15 \pm 0.05 \times 10^{-10}$	$5.70 \pm 0.17 \times 10^{-5}$	$7.10 \pm 0.26 \times 10^{-9}$
4	$3.02 \pm 0.05 \times 10^{-6}$	$6.58 \pm 0.14 \times 10^{-10}$	$2.79 \pm 0.11 \times 10^{-5}$	$8.01 \pm 0.42 \times 10^{-9}$
5	$1.35 \pm 0.03 \times 10^{-5}$	$1.94 \pm 0.06 \times 10^{-9}$	$7.37 \pm 0.28 \times 10^{-5}$	$9.32 \pm 0.44 \times 10^{-9}$
6	$1.16 \pm 0.14 \times 10^{-3}$	$7.07 \pm 0.20 \times 10^{-9}$	$1.89 \pm 0.09 \times 10^{-4}$	$1.04 \pm 0.06 \times 10^{-8}$
7	$1.01 \pm 0.03 \times 10^{-3}$	$1.26 \pm 0.05 \times 10^{-8}$	$3.18 \pm 0.16 \times 10^{-5}$	$5.56 \pm 0.36 \times 10^{-9}$
8	$6.91 \pm 0.18 \times 10^{-5}$	$8.05 \pm 0.21 \times 10^{-9}$	$2.01 \pm 0.12 \times 10^{-5}$	$2.83 \pm 0.16 \times 10^{-9}$

Table 6.7

Estimate order of convergence $\widehat{\beta}$ for the moments of the order-1 LL filter applied to the state space models (6.5)-(6.7) and (6.8)-(6.10) corresponding to the Van der Pool oscillator with additive (Add) and multiplicative (Mul) noise, respectively.

$k\backslash Add$	$\mathbf{y}_{t_{k+1}/t_k}$	$\mathbf{V}_{t_{k+1}/t_k}$	$\mathbf{y}_{t_{k+1}/t_{k+1}}$	$\mathbf{V}_{t_{k+1}/t_{k+1}}$	$k\backslash Mul$	$\mathbf{y}_{t_{k+1}/t_k}$	$\mathbf{V}_{t_{k+1}/t_k}$	$\mathbf{y}_{t_{k+1}/t_{k+1}}$	$\mathbf{V}_{t_{k+1}/t_{k+1}}$
0	1.10	1.04	1.11	1.04	0	1.08	1.01	1.03	1.01
1	1.04	1.05	1.04	1.05	1	1.03	1.03	1.04	1.03
2	1.03	1.03	1.03	1.03	2	1.03	1.04	1.05	1.04
3	1.02	1.02	1.03	1.02	3	1.03	1.03	1.04	1.02
4	1.01	1.01	1.01	0.97	4	1.02	1.02	1.02	1.02
5	1.01	1.03	1.01	1.01	5	1.02	0.94	1.02	0.83
6	1.02	1.01	1.01	0.98	6	1.01	0.98	1.01	0.97
7	1.02	1.04	1.02	1.06	7	0.97	1.00	1.02	0.99
8	1.03	1.02	1.02	1.02	8	1.02	1.01	1.03	0.99

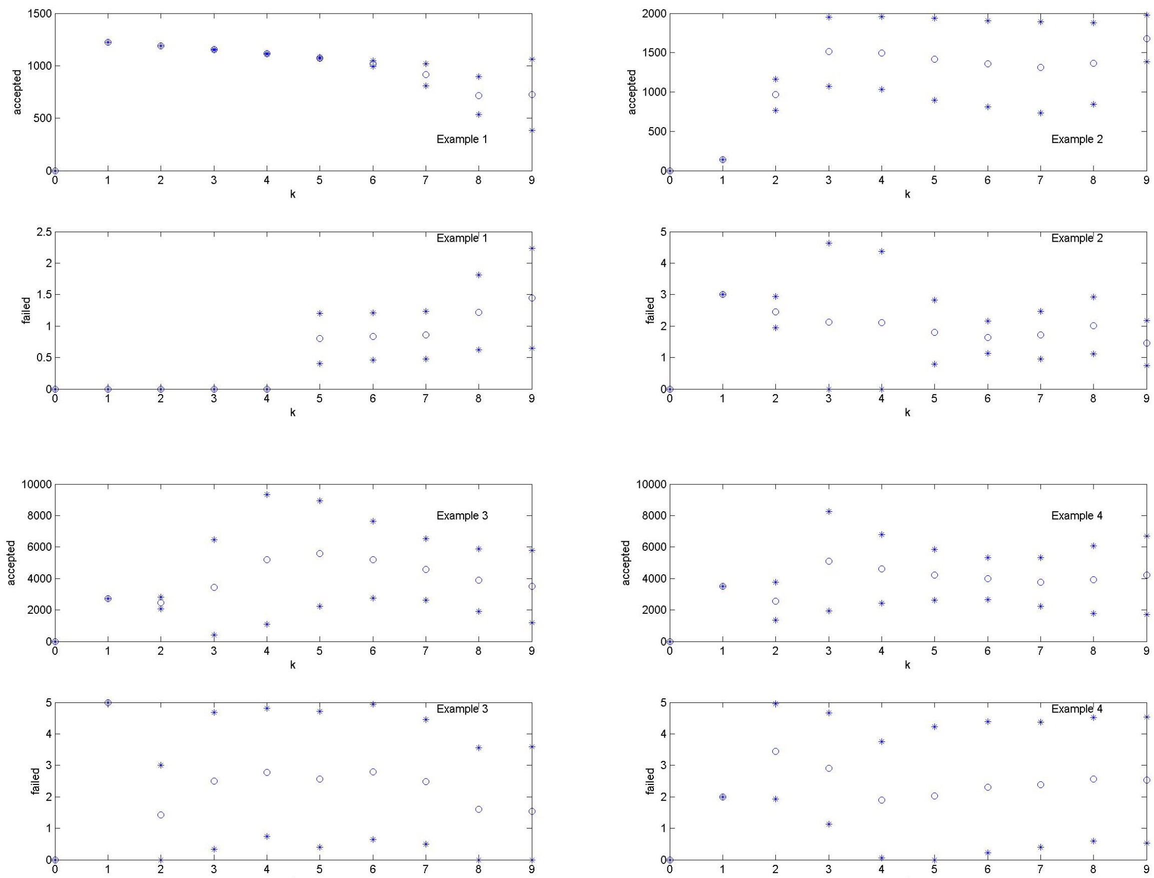

Fig. 6.1. *Average (o) and 90% confidence limits (*) of accepted and failed steps of the adaptive LL filter at each $t_k \in \{t\}_M$ in the four examples.*

processes given a set of partial and noisy observations. Recently, in [22], the performance of the innovation method based on different approximations to the LMV filter has been evaluated by means of simulations. In that paper, the parameters of the four state space models considered in this section were estimated. The results show that the estimators based on the order-β LMV filters are significantly more unbiased and efficient than the estimators based on conventional approximations to the LMV filter, which clearly illustrate the relevance of the approximate filters introduced here. The reader interested in this type of identification problem is encouraged to consider these simulations.

7. Conclusions. Approximate Linear Minimum Variance filters for continuous-discrete state space models were introduced and their order of convergence is stated. As particular instance, the order-β Local Linearization filters were studied in detail. For them, practical algorithms were also provided and their performance in simulation illustrated with various examples. Simulations show that: 1) with thin time discretizations between observations, the order-1 LL filter provides accurate approximations to the exact LMV filter; 2) the convergence of the order-1 LL filter to the exact LMV filter when the maximum stepsize of the time discretization between observations decreases; 3) with respect to the conventional LL filter, the order-1 LL filter significantly improves the approximation to the exact LMV filter; 4) with an adequate tolerance, the adaptive LL filter provides an automatic, accurate and computationally efficient approximation to the LMV filtering problem; and 5) the effectiveness of the order-1 LL filter for the accurate identification of nonlinear stochastic systems from a reduced number of partial and noisy observations distant in time. Finally, it is worth noting that the approximate filters introduced here have already been used in [22] for the implementation of computational efficient parameter estimators of diffusion processes from partial and noisy observations, which would have a positive impact in a variety of applications. Further, they could be easily extended to deal with network-induced phenomena (i.e., missing measurements and communication delays as considered in [49, 17, 18]), which is currently a hot research topic.

Acknowledgments. This work was concluded on July 2012 within the framework of the Associateship Scheme of the Abdus Salam International Centre for Theoretical Physics (ICTP), Trieste, Italy. The author thanks to the ICTP for the partial support to this work. The author is also grateful to the referees for their valuable and helpful comments that contributed to improve the paper.

REFERENCES

[1] L. Arnold, Stochastic Differential Equations: Theory and Applications, Wiley-Interscience Publications, New York, 1974.
[2] R.H. Battin, Statistical optimizing navigation procedure for space flights, ARS J. 32 (1962) 1681-1696.
[3] D. Brigo, B. Hanzon and F. Le Gland, Approximate nonlinear filtering by projection on exponential manifolds of densities. Bernoulli, 5 (1999) 495-534.
[4] C.P. Calderon, N.C. Harris, C.H. Kiang and D.D. Cox, Analyzing single-molecule manipulation experiments, Journal of Molecular Recognition, 22 (2009) 356 - 362.
[5] F. Carbonell and J.C. Jimenez, Weak local linear discretizations for stochastic differential equations with jumps, J. Appl. Prob., 45 (2008) 201-210.
[6] F. Carbonell, J.C. Jimenez and R.J. Biscay, Weak local linear discretizations for stochastic differential equations: convergence and numerical schemes, J. Comput. Appl. Math., 197 (2006) 578-596.
[7] F. Carravetta, A. Germani and M. Raimondi, Polynomial filtering of discrete-time stochastic linear systems with multiplicative state noise, IEEE Trans. Autom. Control, 42 (1997) 1106-1126.
[8] C. Chiarella, H. Hung and T.D. To, The volatility structure of the fixed income market under the HJM framework: A nonlinear filtering approach. Comput. Stat. Data Anal., 53 (2009) 2075-2088.
[9] P. Date and K. Ponomareva, Linear and non-linear filtering in mathematical finance: A review, IMA Journal Management Mathematics 22 (2011) 195-211.
[10] W.L. de Koning, Optimal estimation of linear discrete-time systems with stochastic parameters, Automatica, 20 (1984) 113-115.
[11] A. de Santis, A. Germani and M. Raimondi, Optimal quadratic filtering of linear discrete time non-Gaussian systems, IEEE Trans. Autom. Control, 40 (1995).
[12] P. del Moral, J. Jacod, and P. Protter, The Monte-Carlo method for filtering with discrete-time observations. Prob. Theory Relat. Fields, 120 (2001) 346-368.
[13] G.B. Durham and A.R. Gallant, Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes, J. Buss. Econ. Stats., 20 (2002) 297-316.
[14] M. Gitterman, The noisy oscillator, World Scientific, 2005.
[15] N.R. Hansen, Geometric ergodicity of discrete-time approximations to multivariate diffusions, Bernoulli, 9 (2003) 725-743.
[16] E. Hairer, S.P. Norsett and G. Wanner, Solving Ordinary Differential Equations I, 2nd ed., Springer-Verlag: Berlin, 1993.
[17] J. Hu, Z. Wang, H. Gao and L.K Stergioulas, Extended Kalman filtering with stochastic nonlinearities and multiple missing measurements, Automatica, 48 (2012) 2007-2015.

[18] J. Hu, Z. Wang, H. Gao and L.K Stergioulas, Probability-guaranteed H-infinity finite-horizon filtering for a class of nonlinear time-varying systems with sensor saturations, Systems and Control Letters, 61 (2012), 477-484.
[19] A. S. Hurn, J. I. Jeisman and K. A. Lindsay, Seeing the wood for the trees: a critical evaluation of methods to estimate the parameters of stochastic differential equations. J. Financial Econometrics, 5 (2007) 390-455.
[20] A.H. Jazwinski, Stochastic Processes and Filtering Theory. Academic Press, 1970.
[21] J.C. Jimenez, Simplified formulas for the mean and variance of linear stochastic differential equations, Institute of Statistical Mathematics Research Memo. No.1153, Japan, Febrary 21, 2012. Submitted. http://arxiv.org/pdf/1207.5067.pdf.
[22] J.C. Jimenez, Approximate continuous-discrete filters for the estimation of diffusion processes from partial and noisy observations. Submitted. http://arxiv.org/pdf/1212.3721v1.pdf.
[23] J.C. Jimenez and R. Biscay, Approximation of continuous time stochastic processes by the Local Linearization method revisited. Stochast. Anal. & Appl., 20 (2002) 105-121.
[24] J.C. Jimenez, R. Biscay, C. Mora and L.M. Rodriguez, Dynamic properties of the Local Linearization method for initial-value problems. Appl. Math. Comput., 126 (2002) 63-81.
[25] J.C. Jimenez, R. Biscay and T. Ozaki, Inference methods for discretely observed continuous-time stochastic volatility models: A commented overview, Asia-Pacific Financial Markets, 12 (2006) 109-141.
[26] J.C. Jimenez and H. de la Cruz, Convergence rate of strong Local Linearization schemes for stochastic differential equations with additive noise, BIT, 52 (2012) 357-382.
[27] J.C. Jimenez and T. Ozaki, Linear estimation of continuous-discrete linear state space models with multiplicative noise, Systems & Control Letters, 47 (2002) 91-101.
[28] J.C. Jimenez and T. Ozaki, Local Linearization filters for nonlinear continuous-discrete state space models with multiplicative noise. Int. J. Control, 76 (2003) 1159-1170.
[29] J.C. Jimenez and T. Ozaki, An approximate innovation method for the estimation of diffusion processes from discrete data, J. Time Series Analysis, 27 (2006) 77-97.
[30] R.E. Kalman and R.S. Bucy, A new results in linear filtering and prediction problems, J. Basic Eng., 83 (1961) 95-108.
[31] R.E. Kalman, New methods in Wiener filtering theory, in: Proc. Symp. Eng. Appl. Random Function Theory and Probability, Bogdanoff J.L. and Kozin F. (eds.) Wiley, New York, 1963.
[32] S.C.L. Kamerlin, S. Vicatos, A. Dryga and A. Warshel, Coarse-grained (multiscale) simulations in studies of biophysical and chemical systems, Annual Review of Physical Chemistry 62 (2011) 41-64.
[33] P.E. Kloeden and E. Platen, Numerical Solution of Stochastic Differential Equations, Springer-Verlag, Berlin, Second Edition, 1995.
[34] G.N. Mil'shtein and L.B. Ryashko, Estimation in controlled stochastic systems with multiplicative noise, Autom Remote Control, 6 (1984) 759-765.
[35] G.N. Milstein and M.V. Tretyakov, Numerical integration of stochastic differential equations with nonglobally Lipschitz coefficients, SIAM J. Numer. Anal., 43 (2005) 1139-1154.
[36] R.R. Mohler, Bilinear Control Processes. Academic, 1973.
[37] C. Moler and C. Van Loan, Nineteen dubious ways to compute the exponential of a matrix, SIAM Review, 45 (2003) 3-49.
[38] R.R. Mohler and W.J. Kolodziej, An overview of stochastic bilinear control processes, IEEE Trans. Syst. Man, Cyber., SMC-10 (1980) 913-918.
[39] J. Nicolau, A new technique for simulating the likelihood of stochastic differential equations, Econom. J., 5 (2002) 91-103.
[40] J.N. Nielsen and H. Madsen, Applying the EKF to stochastic differential equations with level effects, Automatica, 37 (2001) 107-112.
[41] J.N. Nielsen, H. Madsen and P.C. Young, Parameter estimation in stochastic differential equations: an overview. Annual Review of Control, 24 (2000) 83-94.
[42] J.N. Nielsen, M. Vestergaard and H. Madsen, Estimation in continuous-time stochastic volatility models using nonlinear filters, Int. J. Theor. Appl. Finance, 3 (2000) 279-308.
[43] T. Ozaki, A local linearization approach to nonlinear filtering, Int. J. Control, 57 (1993) 75-96.
[44] T. Ozaki, The local linearization filter with application to nonlinear system identification. In Bozdogan, H. (ed.) Proceedings of the first US/Japan Conference on the Frontiers of Statistical Modeling: An Informational Approach, 1994, 217-240. Kluwer Academic Publishers.
[45] T. Ozaki, J.C. Jimenez and V. Haggan-Ozaki, Role of the likelihood function in the estimation of chaos models, J. Time Series Analysis, 21 (2000) 363-387.
[46] P.V. Pakshin, State estimation and control synthesis for discrete linear systems with additive and multiplicative noise, Autom. Remote Control, 43 (1978) 526-534.
[47] Y.A. Phillis and V.S. Kouikoglou, Minimax estimation and control of multiplicative systems, Control and Dynamics Systems, Vol. 31, 1989.
[48] J.J. Riera, J. Watanabe, K. Iwata, N. Miura, E. Aubert, T. Ozaki and R. Kawashima, A state-space model of the hemodynamic approach: nonlinear filtering of BOLD signals. Neuroimage, 21 (2004) 547-567.
[49] J.J. Riera, X. Wan, J.C. Jimenez and R. Kawashima, Nonlinear local electro-vascular coupling. Part II: : from data to neural masses, Human Brain Mapping, 28 (2007) 335-354.
[50] G.O. Roberts and O. Stramer, On inference for partially observed nonlinear diffusion models using the Metropolis-Hasting algorithm, Biometrika, 88 (2001) 603-621.
[51] S.F. Schmidt, Application of state-space methods to navigation problems, Advan. Control Syst. 3 (1966) 293-340.
[52] I. Shoji and T. Ozaki, Comparative study of estimation methods for continuous time stochastic process, J. Time Series Anal., 18 (1997) 485-506.
[53] I. Shoji and T. Ozaki, Estimation for nonlinear stochastic differential equations by a local linearization method, Stoch.

Anal. Appl., 16 (1998) 733-752.

[54] I. Shoji, A comparative study of maximum likelihood estimators for nonlinear dynamical systems, Int. J. Control, 71 (1998) 391-404.

[55] H. Singer, Parameter estimation of nonlinear stochastic differential equations: Simulated maximum likelihood versus extended Kalman filter and Ito-Taylor expansion, J. Comput. Graph. Stats., 11 (2002) 972-995.

[56] H.W. Sorenson, Kalman filtering techniques, Advan. Control Syst. 3 (1966).

[57] O. Stramer and R.L. Tweedie, Langevin-type models I: diffusion with given stationary distributions and their discretizations, Meth. Comput. Appl. Prob., 1 (1999) 283-306.

[58] O. Stramer, The local linearization scheme for nonlinear diffusion models with discontinuous coefficients, Stat. Prob. Letters, 42 (1999) 249-256.