ESAIM: COCV 24 (2018) 1815–1847
https://doi.org/10.1051/cocv/2017077

ESAIM: Control, Optimisation and Calculus of Variations
www.esaim-cocv.org

A DISCRETE-TIME OPTIMAL FILTERING APPROACH FOR NON-LINEAR SYSTEMS AS A STABLE DISCRETIZATION OF THE MORTENSEN OBSERVER[☆],[☆☆]

P. Moireau[***]

Abstract. In this work, we seek exact formulations of the optimal estimator and filter for a non-linear framework, as the Kalman filter is for a linear framework. The solution is well established with the Mortensen filter in a continuous-time setting, but we seek here its counterpart in a discrete-time context. We demonstrate that it is possible to pursue at the discrete-time level an exact dynamic programming strategy and we find an optimal estimator combining a prediction step using the model and a correction step using the data. This optimal estimator reduces to the discrete-time Kalman estimator when the operators are in fact linear. Furthermore, the strategy that consists of discretizing the least square criterion and then finding the exact estimator at the discrete level allows to determine a new time-scheme for the Mortensen filter which is proven to be consistent and unconditionally stable, with also a consistent and stable discretization of the underlying Hamilton-Jacobi-Bellman equation.

Mathematics Subject Classification. 34K35, 93E24, 93B52, 49J21, 49L20

Received October 28, 2017. Accepted November 13, 2017.

1. Introduction

State and parameter model estimation based on observations is a very active subject with in particular new trends in computational engineering where data assimilation has become a popular topic with applications to environmental sciences or life sciences [10, 14]. In general, we make a distinction between two main mathematical contexts for estimation. On the one hand, it is common to consider a stochastic context [17, 36] where the state dynamics is defined by a stochastic differential equation and the uncertainties by white noises. The objectives are then to produce an estimate of the mean state and parameters using the data at hand in the sense of the maximum likelihood estimator, or a mean square estimator, or a least square estimator [7] – besides all three can be equivalent in certain cases. The estimator is typically associated with a conditional expectation whereas covariance operators characterize the uncertainty errors in this context. On the other hand, numerous contributions have been proposed in a purely deterministic context where the model is described as a dynamical system. The errors are then unknown quantities defined in a deterministic space with adequate norms characterizing

[☆] The author would like to deeply thank Dr. Dominique Chapelle for his very useful comments on this work.

[☆☆] The author would like to thank Dr. Vivien Mallet for letting him use his HJB implementation in the data assimilation library *Verdandi.*

Keywords and phrases: Deterministic observer, optimal filtering, time-discrete non-linear systems, Hamilton-Jacobi-Bellman.

Inria – LMS, Ecole Polytechnique, CNRS – Université Paris-Saclay, France.

[***] Corresponding author: philippe.moireau@inria.fr

Article published by EDP Sciences

their weight. The most classical deterministic estimation is certainly based on the minimization of a least square criterion aggregating the various sources of error (that we expect to minimize) in order to define an adequate trajectory [16, 27]. Numerous works [3, 24, 25] have demonstrated the connection between the stochastic context and the deterministic one. In particular, the deterministic context can be seen as an asymptotic regime for the probabilistic theory in the sense that noise amplitudes tend to 0 [25]. In the present work, we will concentrate on a deterministic description.

Once the context is set up, there are two categories of estimation strategies: variational and sequential. The variational strategy – which denomination is inspired from the popular 4D-Var algorithm in data assimilation [40] – consists in minimizing the least square criterion, often through an adjoint model integration linked to the dynamical model constraint under which the cost function is minimized. The estimated trajectory is therefore computed as a whole in numerous minimization iterations. The second strategy is called sequential since the estimation corrects the dynamics in time to take into account the possible discrepancy between the computed trajectory and the available data. The sequential estimator is also called observer in the deterministic context, and the feedback operator which corrects the dynamics from the discrepancy is called a gain or a filter – *ergo* the common name of filtering strategy instead of sequential estimation strategy. A very effective observer candidate can be derived from the least square criterion by considering at every time t the least square estimate minimizing the criterion aggregating the data up to the time t. Accordingly, the optimal observer is based on dynamic programming principles [5]. In a linear framework – namely, when all operators are assumed to be linear – it is well known that the optimal observer is given by the popular Kalman filter [29, 31]. The optimal Kalman gain it then computed from the solution of a Riccati equation. In a general non-linear framework, the deterministic non-linear optimal observer can also be formulated and was first introduced by Mortensen [39] – who called his filter the Maximum Likelihood filter – with several in depth-studies in [21, 25, 33]. Following [21], we will call the optimal observer in a non-linear framework the Mortensen observer. This time, the optimal gain is obtained through the solution of a Hamilton-Jacobi-Bellman (HJB) equation as we will recall in the first part of this work. This optimal deterministic observer is also called by various authors the Minimum Energy Estimator, see [33] which also introduces the corresponding HJB equation.

The last classification that we have to consider concerns the time evolution of the system. We have implicitly considered in the previous paragraphs that the evolution of the system was in continuous-time, but we can also consider the class of discrete-time systems and expect to verify with both classes all the equivalences mentioned previously. We find in [19] a seminal attempt to answer this question. However, to the author's best knowledge, the exact counterpart of the optimal observer for discrete-time non-linear systems is still missing and the present article intends to fill this gap by proposing a discrete-time Mortensen filter which strictly satisfies discrete-time dynamic programming principles at each time step. The filter proposed in this article will then reduce to the discrete-time Kalman-filter when a linear framework is considered. In particular, our estimator will be formulated with a prediction-correction scheme as it is classically the case for discrete-time Kalman like filters.

Finally, a discrete-time model can be considered in itself but also corresponds to the time-discretization of continuous-time model. This discretization should then satisfy the necessary properties of consistency and stability. The same question occurs with the observer. Hence, our discrete-time estimator can be seen as a new consistent and stable discretization of the Mortensen optimal observer. In fact, an efficient way to discretize the solution of a minimization problem is to discretize the initial problem and criterion and then solve the optimization problem at the discrete level [16]. As a result, the stability is directly deduced from the minimization, whereas the consistency is obtained through the consistency of the initial model and criterion discretization. Moreover, our approach can also be considered to formulate a stable discretization of HJB equations.

The article is divided as follows. In the next section we present the problem and summarize our main contribution, namely the discrete-time Mortensen observer formulation. Then, we recall some fundamentals on the Mortensen filter in a continuous-time setting in order to ease the understanding of our discrete-time Mortensen observer formulation, based on the strong similarity between discrete-time and continuous-time optimal filtering formulations. We then present a mathematical justification of our approach. Finally, we provide a complete numerical algorithm and illustrate the performance and properties of the observer on linear and non-linear systems of small dimensions.

2. Main contribution

2.1. Problem statement

2.1.1. Models settings

We consider a general class of dynamical model on a finite dimensional space $\mathcal{X} \sim \mathbb{R}^{\mathrm{N}_x}$. This model would be either defined in a continuous-time framework or a discrete-time framework. In the continuous-time framework we thus define

$$\begin{cases} \dot{x}_{|\zeta,\omega}(t) = F(x_{|\zeta,\omega}(t),t) + B(t)\omega(t), \quad t \in \mathbb{R}^+ \\ x_{|\zeta,\omega}(0) = x_\diamond + \zeta \end{cases} \tag{2.1}$$

where x is the state variable in $\mathcal{X}$, F is the mapping defining the dynamics, $\omega \in \mathcal{W} \sim \mathbb{R}^{N_\omega}$ represents an unknown contribution that can be seen as a model error, $x_\diamond$ is the known part of the initial condition but ζ_x is to be determined. In other words, the initial condition of this system is considered unknown around an *a priori* $x_\diamond$. For the general presentation of the equations, we choose a general notation for the model dynamics that includes potentially non-linear dynamics. However, for the sake of simplicity, we will assume in all proofs that $F(\cdot,t)$ is an affine mapping (hence $F \in \mathcal{C}^1(\mathcal{X})$). The linear part of $F(\cdot,t)$ is its differential with respect to x and is denoted by $\mathrm{d}_x F$, which is assumed to be bounded uniformly with respect of t. Note that we believe that the results presented in this work are valid for a large class of non-linear dynamics hence justifying the general notation $F(\cdot)$ for the dynamics. In addition, $B(t)$ is also considered to be bounded uniformly with respect of t. By the Cauchy-Lipschitz theorem we can thus consider unique global solutions of the system (2.1). In the discrete-time framework, the counterpart of (2.1) will be to consider a sequence in $\mathcal{X}$ defined recursively by

$$\begin{cases} x_{n+1|\zeta,(\omega_k)_{k=0}^{n}} = F_{n+1|n}(x_{n|\zeta,(\omega_k)_{k=0}^{n-1}}) + B_n\omega_n, \quad n \in \mathbb{N} \\ x_{0|\zeta,(\omega_k)_{k=0}^{n}} = x_\diamond + \zeta \end{cases} \tag{2.2}$$

with $F_{n+1|n}$ the transition mapping from step n to step $n+1$, $(\omega_k)_{k=0}^{n}$ a sequence of unknown contributions in $\mathcal{W}$ that can be seen as a discrete-time model noise definition and B_n the model noise operator at step n. Identically we will consider in our proofs that for all n, $F_{n+1|n}$ is an affine mapping (hence is $\mathcal{C}^1(\mathcal{X})$). Moreover, $\mathrm{d}F_{n+1|n}$ – which is the differential of $F_{n+1|n}$ – and B_n are assumed to be bounded uniformly with respect to n. Note that this discrete-time system can be considered in itself or as a consistent and stable discretization of (2.1). In this respect, we introduce a sequence of times $(t_n)_{n\in\mathbb{N}}$ and discretization parameter Δt considered here to be constant to simplify, *i.e.* $\Delta t = t_{n+1} - t_n$. Then the transition operator and the model noise operator would typically satisfy

$$\forall (x,t_n) \in \mathbb{R}^+, \quad \lim_{\Delta t \to 0} \frac{1}{\Delta t}(F_{n+1|n}(x) - \mathbb{1}) = F(x,t_n), \quad \lim_{\Delta t \to 0} \frac{1}{\Delta t} B_n = B(t_n).$$

As an example we could consider – when stable – an explicit discretization of the continuous-time model (2.1) and we would have then

$$F_{n+1|n}(x) = \mathbb{1} + \Delta t F(x,t_n), \quad B_n = \Delta t B(t_n).$$

We also note that, in the context of discretization, we can assume that $\mathrm{d}F_{n+1|n}$ is invertible (hence $F_{n+1|n}$ is a $C^1(\mathcal{X})$ diffeomorphism) without being too restrictive since it corresponds to a perturbation of the identity.

Note that the same type of assumption holds in the seminal work [19] about optimal estimation of non-linear time-discrete system. This assumption will be made in all the rest of the present article.

Considering this class of system, we assume to have at our disposal some measurements – or observations – on a particular instance of this model. In the continuous-time context, we thus introduce a target trajectory $\{\breve{x}(t), t \in \mathbb{R}^+\}$ and model the measurement generation in the general form

$$\forall t \in \mathbb{R}^+, \quad z(t) = H(\breve{x}(t), t) + \chi(t), \tag{2.3}$$

where H is called the observation operator from $\mathcal{X}$ to the observation space $\mathcal{Z} \sim \mathbb{R}^{\mathrm{N_{obs}}}$ and χ is an unknown function assimilated to a measurement error. Here we assume that $H(\cdot, t)$ is a $\mathcal{C}^1(\mathbb{R}^n)$-mapping and that $\mathrm{d}_x H$ is uniformly bounded for all t. Similarly in the discrete-time context, we consider the target sequence $(\breve{x}_n)_{n\in\mathbb{N}}$ and an observation operator H_n at each time step, so that some measurements z_n are given by

$$z_n = H_n(\breve{x}_n) + \chi_n,$$

with an additive noise sequence (χ_n). Mirroring the continuous setting, H_n is a $\mathcal{C}^1(\mathbb{R}^{\mathrm{N_{obs}}})$-mapping and $\mathrm{d}H_n$ is bounded uniformly with respect to n. Note that, in general, measurement procedures are in essence time-sampled and the observation discrete-time model should be more directly defined than the observation continuous-time model. Then the latter can be regenerated by interpolation from the sampled measurements with a measurement error ω incorporating some interpolation error [18]. Eventually, we can consider the two frameworks independently or assuming that for all $t \in [t_n, t_{n+1}]$, we have $\|H(x,t) - H_n(x)\| = O(\Delta t))$.

Finally, note that we must characterize the finite dimensional spaces $\mathcal{X}$, $\mathcal{W}$ and $\mathcal{Z}$ with adequate norms and, here, with finite dimensional spaces and for the sake of simplicity, we rely for $\|\cdot\|_{\mathcal{X}}$, $\|\cdot\|_{\mathcal{W}}$ and $\|\cdot\|_{\mathcal{Z}}$ on the standard euclidian norm.

2.1.2. Optimal estimation context

When seeking to reconstruct a complete trajectory from $\{z(t), t \in \mathbb{R}^+\}$ – *i.e.* without knowing specifically $\breve{\zeta}, \breve{\omega}$ which have generated $\{\breve{x}(t), t \in \mathbb{R}^+\}$ – one approach, often referred to as optimal, is to consider the least square estimator (LSE). In the literature [27, 43], the least square estimation problem is also called optimal estimation problem in the sense that the estimation relies on an optimal criterion minimization. We will thus adopt this denomination in the rest of the present article.

In the continuous-time framework we seek to minimize the criterion defined on a specific time window $[0, t]$

$$\min_{\zeta\in\mathcal{X}\omega\in L^2*([0,t],\mathcal{W})} \left\{ \mathscr{J}(\zeta, \omega, t) = \frac{1}{2}\|\zeta\|^2_{N_\diamond} + \frac{1}{2}\int_0^t \left(\|D(x_{|\zeta,\omega}(s), s)\|^2_M + \|\omega(s)\|^2_S \right) \mathrm{d}s \right\}, \tag{2.4}$$

where the discrepancy between the model and the measurement is computed through

$$\forall t \in \mathbb{R}^+, \forall x \in \mathcal{X}, \quad D(x, t) = z(t) - H(x, t),$$

and where we introduce weighted norms in each of these spaces with three operators, symmetric and invertible, $N_\diamond \in \mathbb{S}^+_{\mathrm{N}_x}(\mathbb{R})$, $S \in \mathbb{S}^+_{N_\omega}(\mathbb{R})$ and $M \in \mathbb{S}^+_{\mathrm{N_{obs}}}(\mathbb{R})$

$$\forall x \in \mathcal{X}, \|x\|^2_{N_\diamond} = \langle x, N_\diamond x\rangle_{\mathcal{X}} = x^\intercal N_\diamond x,$$
$$\forall \omega \in \mathcal{W}, \|\omega\|^2_S = \langle \omega, S\omega\rangle_{\mathcal{W}} = \omega^\intercal S\omega,$$
$$\forall z \in \mathcal{Z}, \|z\|^2_M = \langle z, Mz\rangle_{\mathcal{Z}} = z^\intercal Mz.$$

In the sequel, we will use a subscript to recall the operator defining the norm. The inverse of these three operators

$$P_\diamond = N_\diamond^{-1}, \quad W = M^{-1}, \text{ and } Q = S^{-1},$$

will be related to typical error covariances expected in these spaces in a probabilistic framework of estimation [27, 43]. For the sake of simplicity of our presentation, we will assume that $\|D(x,t)\|_M^2$ is convex with respect to the variable x. Therefore, for all t, our criterion $\mathscr{J}(\cdot,\cdot,t)$ is strictly convex with respect to (ζ,ω), hence we have the existence of unique minimizers denoted by

$$(\bar{\zeta}_{|t}, \bar{\omega}_{|t}) = \underset{\substack{\zeta\in\mathcal{X},\\ \omega\in L^2([0,t],\mathcal{W})}}{\operatorname{argmin}} \mathscr{J}(\cdot,\cdot,t),$$

with the associated trajectory written

$$\forall s \in [0,t], \quad \bar{x}_t(s) = x_{\bar{\zeta}_{|t},\bar{\omega}_{|t}}(s).$$

Note that we are here in a particular case of the more general framework defined in [23] where the existence of unique minimizers for non-linear dynamics is obtained thanks to Chapter 3 from [21] and [13].

Definition 2.1 (Continuous-time Mortensen Estimator)**.** When for all $t \in \mathbb{R}^+$, $\mathscr{J}(\cdot,\cdot,t)$ admits the unique minimizers $\bar{\zeta}_{|t} \in \mathcal{X}$ and $\bar{\omega}_{|t} \in L^2([0,t],\mathcal{W})$, the continuous-time Mortensen Estimator is defined as the optimal sequential estimator in the sense that

$$\forall t \in \mathbb{R}^+, \quad \hat{x}(t) = \bar{x}_{|t}(t). \tag{2.5}$$

Similarly, in the discrete-time framework, we introduce $D_k : x \mapsto z - H_k(x)$. The least square criterion minimization is typically given by

$$\min_{\substack{\zeta\in\mathcal{X},\\ (\omega_k)_{k<n}\in\mathcal{W}^n}} \left\{ \mathscr{J}_n^+(\zeta,(\omega_k)_{k<n}) = \frac{1}{2}\|\zeta\|_{N_\diamond}^2 + \frac{1}{2}\sum_{k=0}^{n}\|D_k(x_k)\|_{M_k}^2 + \frac{1}{2}\sum_{k=0}^{n-1}\|\omega_k\|_{S_k}^2 \right\}, \tag{2.6}$$

with, by extension,

$$\mathscr{J}_0^+(\zeta) = \frac{1}{2}\|\zeta\|_{N_\diamond}^2 + \|D_0(x_k)\|_{M_0}^2.$$

A variant is to consider

$$\min_{\substack{\zeta\in\mathcal{X},\\ (\omega_k)_{k\le n}\in\mathcal{W}^{n+1}}} \left\{ \mathscr{J}_{n+1}^-(\zeta,(\omega_k)_{k\le n}) = \frac{1}{2}\|\zeta\|_{N_\diamond}^2 + \frac{1}{2}\sum_{k=0}^{n}\|D_k(x_k)\|_{M_k}^2 + \frac{1}{2}\sum_{k=0}^{n}\|\omega_k\|_{S_k}^2 \right\}, \tag{2.7}$$

with

$$\mathscr{J}_0^-(\zeta) = \frac{1}{2}\|\zeta\|_{N_\diamond}^2,$$

and the comparative interest of the two criteria will appear in the next section. Here again, the continuous-time criterion and the discrete-time criterion are defined in contexts which can be completely independent or related as the discrete-time framework is defined after time-discretization of the continuous-time system. In the latter, the two criteria should be defined consistently as Δt goes to 0. In this respect, we could for instance consider

$$M_n = \Delta t\, M, \quad S_n = \Delta t\, S \quad \text{or equivalently} \quad W_n = \frac{1}{\Delta t} W, \quad Q_n = \frac{1}{\Delta t}\, Q,$$

so that $\lim_{\Delta t \to 0} \mathscr{J}_n^+(\cdot,\cdot) = \mathscr{J}(\cdot,\cdot,t)$, with $n\Delta t = t$.

Assuming – as in the time-continuous case – that $\|D_n(x)\|^2_{M_n}$ are convex for all n, our discrete-time criteria are again strictly convex with respect to their variables, hence we have the existence of unique minimizers. Note that in a more general framework where the dynamics mapping is non-linear, the existence of unique minimizers can be much more intricate to obtain. A typical strategy is to use the same type of argument as in [13] or Chapter 3 from [23], which consists in rewriting the minimization problem as a Mayer problem. In this respect, we can cite [19, 20] that introduces a similar discrete-time least square criterion and establishes succinctly the existence of a unique minimizer.

As unique minimizers of (2.6) or (2.7) exist, we denote them by

$$(\bar{\zeta}^+_{|n}, (\bar{\omega}^+_{k|n})_{k<n}) = \underset{\substack{\zeta \in \mathcal{X} \\ (\omega_k)_{k<n} \in \mathcal{W}^n}}{\operatorname{argmin}} \mathscr{J}_n^+(\zeta, (\omega_k)_{k<n}),$$

and

$$(\bar{\zeta}^-_{|n}, (\bar{\omega}^-_{k|n})_{k\leq n}) = \underset{\substack{\zeta \in \mathcal{X} \\ (\omega_k)_{k\leq n} \in \mathcal{W}^{n+1}}}{\operatorname{argmin}} \mathscr{J}_{n+1}^-(\zeta, (\omega_k)_{k\leq n}),$$

respectively. The associated trajectories are then

$$\forall k \in [0,n], \quad \bar{x}^+_{k|n} = x_{k|\bar{\zeta}^+_{|n}, (\bar{\omega}^+_{j|n})_{j<n}},$$

and

$$\forall k \in [0,n+1], \quad \bar{x}^-_{k|n} = x_{k|\bar{\zeta}^-_{|n}, (\bar{\omega}^-_{j|n})_{j\leq n}},$$

respectively. Note finally that these sequences can be extended for all k by simply propagating them with the discrete-time dynamics (2.2) with a null model noise. In particular, we will rely on

$$\bar{x}^+_{n+1|n} = F_{n+1|n}(\bar{x}^+_{n|n}).$$

Then a counterpart of Definition 2.1 is given by the following definition. We call this new estimator *the discrete-time Mortensen estimator*.

Definition 2.2 (Discrete-time Mortensen Estimator)**.** When for all $n \in \mathbb{N}$, the criterion $\mathscr{J}_n^+$ – the criterion $\mathscr{J}_{n+1}^-$ respectively – admits unique minimizers $\bar{\zeta}^+_{|n} \in \mathcal{X}$ and $(\bar{\omega}^+_{k|n})_{k<n} \in \mathcal{W}^n$ – $\bar{\zeta}^-_{|n} \in \mathcal{X}$ and $(\bar{\omega}^-_{k|n})_{k\leq n} \in \mathcal{W}^{n+1}$ respectively – the discrete-time Mortensen Estimator is defined as the optimal sequential estimator in the sense that

$$\forall n \in \mathbb{N}, \quad \hat{x}^+_n = \bar{x}^+_{n|n} = x_{n|\bar{\zeta}^+_{|n}, (\bar{\omega}^+_{k|n})_{k<n}} \quad \text{and} \quad \hat{x}^-_{n+1} = \bar{x}^-_{n+1|n} = x_{n+1|\bar{\zeta}^-_{|n}, (\bar{\omega}^-_{k|n})_{k\leq n}}. \tag{2.8}$$

2.2. Optimal estimation in the linear context: the Kalman Filter

Before presenting the specific contribution of this paper, we want to present the full linear context when the optimal observers are well known in both the discrete-time and the continuous-time frameworks since the seminal work [29, 31] introducing the Kalman Filter in the discrete-time framework and its counterpart the Kalman-Bucy Filter in the continuous-time framework, see also Chapter 2 from [6], Chapter 9,16 from [27] or Part 2 from [43]. Hence in this section we assume that the dynamics mapping F is an affine mapping and the observation operator H is a linear operator. In the continuous-time setting, we thus introduce

$$\forall t \in \mathbb{R}^+, \quad F : x \to A(t) + R(t), \quad H : x \to H(t)x,$$

and assume that A and H are uniformly bounded with respect to time. Then the optimal estimator defined by (2.5) is given by the dynamics

$$\begin{cases} \dot{\hat{x}}(t) = A(t)\hat{x}(t) + R(t) + P(t)H(t)^\intercal M\big(z(t) - H(t)\hat{x}(t)\big), \quad t \in \mathbb{R}^+ \\ \hat{x}(0) = x_\diamond \end{cases} \tag{2.9}$$

where $G = P(t)H(t)^\intercal M$ is the (continuous-time) Kalman-Bucy Filter built from $P \in \mathcal{L}(\mathcal{X}, \mathcal{X})$ satisfying the following Riccati equation

$$\begin{cases} \dot{P}(t) = A(t)P(t) + P(t)A(t)^\intercal - P(t)H(t)^\intercal M H(t)P(t) + BQB^\intercal, \quad t \in \mathbb{R}^+ \\ P(0) = P_\diamond \end{cases} \tag{2.10}$$

In the discrete-time linear case, we introduce

$$\forall n \in \mathbb{N}, \quad F_{n+1|n} : x \to A_{n+1|n}x + R_n(t), \quad H_n : x \to H_n x,$$

and it is well-known that the discrete-time optimal estimator is given by

$$\begin{cases} \text{Initialization:} \\ \quad \hat{x}_0^- = x_\diamond \\ \text{Correction:} \\ \quad \hat{x}_n^+ = \hat{x}_n^- + G_n\big(z_n - H_n\hat{x}_n^-\big), \quad n \in \mathbb{N} \\ \text{Prediction:} \\ \quad \hat{x}_{n+1}^- = A_{n+1|n}\hat{x}_n^+ + R_n, \quad n \in \mathbb{N} \end{cases} \tag{2.11}$$

where the (discrete-time) Kalman filter is given by prediction-correction dynamics

$$G_n = P_n^+ H_n^\intercal M_n = P_n^- H_n^\intercal (W_n + H_n^\intercal P_n^- H_n)^{-1}$$

with the state covariances $P_n^\pm$ satisfying the discrete-time Riccati equation

$$\begin{cases} \text{Initialization:} \\ \quad P_0^- = P_\diamond \\ \text{Correction:} \\ \quad P_n^+ = P_n^- - P_n^- H_n^\intercal (H_n P_n H_n + M_n) H_n P_n, \quad n \in \mathbb{N} \\ \text{Prediction:} \\ \quad P_{n+1}^- = A_{n+1|n} P_n^+ A_{n+1|n}^\intercal + B_n Q_n B_n^\intercal, \quad n \in \mathbb{N} \end{cases} \tag{2.12}$$

Here we use a classical notation in Kalman filter theory [43] where the predicted state is denoted by a *minus* superscript whereas the correction step is identified by a *plus* superscript. It is optimal in the sense of (2.8) since, here, it can be easily proved – see for instance [27, 43] – that

$$\hat{x}_n^+ = \bar{x}_{n|n}^+ \text{ and } \hat{x}_{n+1}^- = \bar{x}_{n+1|n}^-.$$

Therefore, each filter can be shown to control the dynamics of an optimal state estimator in each context. But moreover, the discrete-time Kalman estimator can be seen as a consistent discretization of the continuous-time Kalman estimator as soon as the dynamics and observation operators are consistent in their discrete-time *versus* continuous-time version. Then, one possible time discretization of the continuous-time Kalman estimator (2.9) is directly furnished by the discrete-time Kalman (2.11). It corresponds to a splitting time-scheme.

2.3. Optimal estimation in the general non-linear context: the Mortensen Filter

When either the dynamics or the observation operator are non-linear, the extension of the Kalman estimator is more intricate. In a deterministic context and with continuous-time operators, the answer was given originally in [39] leading to the definition of the Mortensen filter. In essence, it was demonstrated in [21] that the dynamics of the optimal estimator defined in (2.5) can still rely on a correction at each time of the original dynamics with a feedback loop based on the discrepancy. To compute the feedback gain we introduce the following definition.

Definition 2.3 (Continuous-time Cost-To-Come)**.** The *continuous-time cost-to-come* is the function defined by

$$\forall (x,t) \in \mathscr{X} \times \mathbb{R}^+, \quad \mathscr{V}(x,t) = \min_{\substack{\omega \in L^2([0,T],\mathcal{W}), \\ \zeta \in \mathscr{X} \,|\, x(t) = x}} \mathscr{J}(\zeta, \omega, t). \tag{2.13}$$

Note that in our case, the *continuous-time cost-to-come* is well defined as in [21], which then proves the following result.

Theorem 2.4 (Continuous-time Mortensen Filter)**.** *The continuous-time* cost-to-come *is solution– in the viscosity sense – of a Hamilton-Jacobi-Bellman equation*

$$\begin{cases} \partial_t \mathscr{V}(x,t) - \bar{\mathscr{H}}\big(x, \nabla \mathscr{V}(x,t), t\big) = 0, \quad (x,t) \in \mathscr{X} \times \mathbb{R}^+ \\ \mathscr{V}(x,0) = \frac{1}{2} \| x - x_\diamond \|_{N_\diamond}^2 \end{cases} \tag{2.14}$$

where

$$\bar{\mathscr{H}}(x,p,t) = \frac{1}{2} \| D(x,t) \|_M^2 - \frac{1}{2} p^\intercal B Q B^\intercal p - p^\intercal F(x,t). \tag{2.15}$$

Then, when $\mathscr{V}(\cdot,t)$ is $\mathcal{C}^2(\mathcal{X})$ with its hessian invertible, the continuous-time Mortensen filter is solution of the dynamics

$$\begin{cases} \dot{\hat{x}}(t) = F(\hat{x},t) - (\nabla^2\mathscr{V}(\hat{x}(t),t))^{-1}\,\mathrm{d}_x D(\hat{x}(t),t)^\intercal M D(\hat{x}(t),t), \quad t \in \mathbb{R}^+ \\ \hat{x}(0) = x_\diamond \end{cases} \tag{2.16}$$

The full linear case – namely the case of affine dynamics and linear observation operator – can then be deduced from Theorem 2.4 as recalled by [21]. In addition to the assumptions on the operators, we suppose that observability and controllability conditions are satisfied [28, 30]. Then, we can prove the existence of a regular solution for the cost-to-come. Indeed, the operator P in (2.9) is invertible, see for instance [2, 8] and an analytical solution of (2.14) can be computed, namely

$$\forall (x,t) \in \mathcal{X} \times \mathbb{R}^+, \quad \mathscr{V}(x,t) = \frac{1}{2}\big(x - \hat{x}(t)\big)^\intercal P^{-1}(t)\big(x-\hat{x}(t)\big) + \frac{1}{2}\int_0^t \|z(s) - H(s)\hat{x}(s)\|_M^2 \,\mathrm{d}s. \tag{2.17}$$

from which we retrieve (2.9) from (2.16).

To the author's best knowledge, there exists no counterpart of the Mortensen Filter in the case of discrete-time operators in order to exactly define the optimal estimator in the sense of (2.8). Consequently, the only available results at the discrete-time level consist of standard discretizations of the continuous-time Mortensen estimator (2.16), requiring adequate discretizations of the related HJB (2.14) as proposed in [4] for example. Hence the optimality is no longer preserved at the discrete level and the discretization does not benefit from increased stability properties coming from the optimality as it is often noticed in similar optimal control theory problems [16].

2.4. Main result

The present article aims at filling the gap in deterministic optimal filtering for non-linear configurations by introducing the exact counterpart of (2.16) in the discrete-time context, hence finding the exact dynamics of the *discrete-time* Mortensen estimator. The estimator equations are presented in a general framework. However, for the sake simplicity, we limit our proofs to the case of an affine dynamics and a non-linear observation operator leading to convex least-square criteria. Note that this case goes already beyond the Kalman estimator formulation and approximate optimal estimators such as the extended Kalman estimator in a non-linear framework. Moreover, we believe that our result can be extended to a much more general framework by mimicking what has been done for the continuous-time case in [21]. Firstly, let us define the counterpart at the discrete-time level of the *cost-to-come*.

Definition 2.5 (Discrete-time Cost-To-Come)**.** The *discrete-time costs-to-come* - called the *prediction cost-to-come* and *the correction cost-to-come* – are defined for all $n \in \mathbb{N}$ by

$$\begin{cases} \mathscr{V}_n^+(x) = \min\limits_{\substack{(\omega_k)_{k<n} \in \mathcal{W}^n \\ \zeta \in \mathcal{X} \,|\, x_n = x}} \mathscr{J}_n^+(\zeta, (\omega_k)_{k<n}), \\ \mathscr{V}_{n+1}^-(x) = \min\limits_{\substack{(\omega_k)_{k\leq n} \in \mathcal{W}^{n+1} \\ \zeta \in \mathcal{X} \,|\, x_{n+1} = x}} \mathscr{J}_{n+1}^-(\zeta, (\omega_k)_{k\leq n}), \quad \mathscr{V}_0^-(x) = \frac{1}{2}\|x - x_\diamond\|_{N_\diamond}^2. \end{cases} \tag{2.18}$$

In our setting, the two costs-to-come are clearly well defined since our criteria have unique minimizers and the discrete dynamics is invertible. We also understand the interest of the introduction of two different criteria since we now consider a constraint at the end of the time window. Then our main result is given by the following theorem.

Theorem 2.6 (Discrete-time Mortensen Filter)**.** *Assuming that for all n, $\mathscr{V}_n^-$ and $\mathscr{V}_n^+$ are smooth enough – namely $\mathcal{C}^2(\mathcal{X})$ – they are solutions of the following system*

$$\begin{cases} \mathscr{V}_0^-(x) = \frac{1}{2}\|x - x_\diamond\|^2_{N_\diamond}, \\ \mathscr{V}_n^+(x) = \mathscr{V}_n^-(x) + \frac{1}{2}\|D_n(x)\|^2_{M_n}, \\ \mathscr{V}_{n+1}^-(x) = \mathscr{V}_n^+(y) + \frac{1}{2}\nabla\mathscr{V}_{n+1}^-(x)^\intercal B_n Q_n B_n^\intercal \nabla\mathscr{V}_{n+1}^-(x) \\ \qquad\qquad\qquad\qquad \text{with } x = F_{n+1|n}(y) + B_n Q_n B_n^\intercal \nabla\mathscr{V}_{n+1}^-(x). \end{cases} \tag{2.19}$$

Then the discrete-time Mortensen filter can be computed by the recursive procedure

$$\begin{cases} \textit{Initialization:} \\ \quad \hat{x}_0 = x_\diamond, \\ \quad \mathscr{V}_0^- = \frac{1}{2}\|x - \hat{x}_0^-\|^2_{N_\diamond}; \\ \textit{Correction:} \\ \quad \nabla\mathscr{V}_n^+(\hat{x}_n^+) = 0, \\ \textit{Prediction:} \\ \quad \hat{x}_{n+1}^- = F_{n+1|n}(\hat{x}_n^+), \quad n \in \mathbb{N} \end{cases} \tag{2.20}$$

Note that $\hat{x}_n^+$ is defined implicitly. Therefore, in practice and when the Hessian of $\mathscr{V}_n^+$ is invertible, we use a Newton-Raphson procedure to compute $\hat{x}_n^+$ as the limit of the recursive procedure

$$\begin{cases} \hat{x}_{0|n}^+ = \hat{x}_n^-, \quad n \in \mathbb{N} \\ \hat{x}_{k+1|n}^+ = \hat{x}_{k|n}^+ - (\nabla^2\mathscr{V}_n^+(\hat{x}_{k|n}^+))^{-1}\nabla\mathscr{V}_n^+(\hat{x}_{k|n}^+), \quad k \in \mathbb{N}. \end{cases}$$

We remark that we thus used the Hessian of the *cost-to-come* as in the continuous-time setting. Moreover, we will show that

$$\nabla\mathscr{V}_n^+(\hat{x}_{0|n}^+) = \mathrm{d}D_n(\hat{x}_n^-)^\intercal M_n D_n(\hat{x}_n^-), \tag{2.21}$$

therefore, after one iteration of the Newton-Raphson procedure, we have

$$\hat{x}_{1|n}^+ = \hat{x}_n^- - (\nabla^2\mathscr{V}_n^+(\hat{x}_n^-))^{-1}\,\mathrm{d}D_n(\hat{x}_n^-)^\intercal M_n D_n(\hat{x}_n^-),$$

a very similar expression to what we have in the Kalman context, or in the extended Kalman filtering context for non-linear systems [43].

From Theorem 2.6, we deduce two important propositions. The first one is a characterization of the (prediction) estimator based on the *prediction cost-to-come.*

Corollary 2.7. *Under the assumptions of Theorem 2.6, we have*

$$\nabla\mathscr{V}_{n+1}^-(\hat{x}_{n+1}^-) = 0. \tag{2.22}$$

Finally we link Theorem 2.6 to the Kalman estimator introduced in [29] when the dynamics is affine and the observation operator is linear under the Kalman observability and controllability conditions [1].

Proposition 2.8. *For all $n \in \mathbb{N}$, we consider an affine mapping transition $F_{n+1|n} : x \mapsto A_{n+1|n}x + R_n$, a linear observation operator $H_n \in \mathcal{L}(\mathcal{X}, \mathcal{Z})$ and a model noise operator $B_n \in \mathcal{L}(\mathcal{W}, \mathcal{X})$. We assume that these operators* *satisfy the Kalman observability and controllability conditions. Then, we consider the covariance operators $P_n^\pm$ given by the discrete Riccati equations* (2.12) *and the estimators $\hat{x}_n^\pm$ given by the Kalman Filter* (2.11)*. The discrete* costs-to-come *exist, are $\mathcal{C}^2(\mathcal{X})$, and satisfy*

$$\begin{cases} \mathscr{V}_n^-(x) = \dfrac{1}{2}(x-\hat{x}_n^-)^\intercal (P_n^-)^{-1}(x-\hat{x}_n^-) + \mathscr{V}_{n-1}^0, \\ \mathscr{V}_n^+(x) = \dfrac{1}{2}(x-\hat{x}_n^+)^\intercal (P_n^+)^{-1}(x-\hat{x}_n^+) + \mathscr{V}_n^0, \end{cases} \tag{2.23}$$

with $\mathscr{V}_{-1}^0 = 0$,

$$\mathscr{V}_n^0 = \frac{1}{2}\sum_{k=0}^n \Big(\|z_k - H_k \hat{x}_k^-\|_{M_k}^2 - \|\hat{x}_k^+ - \hat{x}_k^-\|_{(P_k^+)^{-1}}^2 \Big), \tag{2.24}$$

$$= \frac{1}{2}\sum_{k=0}^n \Big(\|z_k - H_k \hat{x}_k^+\|_{M_k}^2 + \|\hat{x}_k^+ - \hat{x}_k^-\|_{(P_k^-)^{-1}}^2 \Big). \tag{2.25}$$

Moreover, $\hat{x}_n^\pm$ can be computed by the discrete-time Mortensen procedure (2.20)*.*

In order to prove all the results announced in this section, we organize the next sections as follows. Theorem 2.6 will be proved in Section 4. Our proof follows a strategy similar to what has been done for the continuous-time Mortensen filter, hence we have chosen to recall succinctly the justification of Theorem 2.4 as obtained in [21]. This is the objective of the next section. In section 5, we will discuss the consistency of our discrete-time results with respect to the continuous-time setting when Δt goes to 0. Finally, we will present in Section 6 a complete numerical procedure to compute the discrete-time Mortensen estimator for a non-linear dynamical system in small dimension.

3. Background on the Mortensen filter

The most common approach for solving the continuous-time least square problem (2.4) – called the variational strategy in data assimilation [37] – consists in minimizing the criterion with the help of the adjoint equation associated with the dynamics constraint under which we minimize the criterion. In fact, defining the adjoint variable from a given trajectory $\{x_t(s), s \in [0,t]\}$ by

$$\begin{cases} \dot{p}_t(s) + \mathrm{d}_x F(x_t(s),s)^\intercal p_t = \mathrm{d}_x D(x_t(s),s)^\intercal M D(x_t(s),s), \quad s \in [0,t] \\ p_t(t) = 0 \end{cases} \tag{3.1}$$

the adjoint equation is linear and generated by an operator which we assumed to be bounded uniformly in the time variable. Hence, we have a unique global solution and the trajectory of System (2.1) minimizing the criterion $\mathscr{J}$ introduced in (2.4) is given by

$$\begin{cases} \dot{\bar{x}}_t(s) = F(\bar{x}_t, s) + BQB^\intercal \bar{p}_t(s), \quad \forall s \in [0,t] \\ \bar{x}_t(0) = x_\diamond + P_\diamond \bar{p}_t(0) \end{cases} \tag{3.2}$$

where $\bar{p}_t$ is the adjoint variable associated with $\bar{x}_t$ [6]. The estimator dynamics (3.2) can be summarized by introducing

$$\mathscr{L} : \left| \begin{array}{l} \mathcal{X} \times \mathcal{W} \times \mathbb{R}^+ \to \mathbb{R} \\ \qquad (x,\omega,t) \mapsto \dfrac{1}{2}\|D(x,t)\|_M^2 + \dfrac{1}{2}\|\omega\|_S^2 \end{array} \right. \tag{3.3}$$

hence the system Hamiltonian functional

$$\mathscr{H}: \left| \begin{array}{l} \mathcal{X} \times \mathcal{X} \times \mathcal{W} \times \mathbb{R}^+ \to \mathbb{R} \\ \qquad (x,p,\omega,t) \mapsto \mathscr{L}(x,\omega,t) - p^\intercal (F(x,t) + B\omega) \end{array} \right. \tag{3.4}$$

giving for all $s \in [0,t]$ that

$$\dot{\bar{x}}_t = -\nabla_p \mathscr{H}(\bar{x}_t, \bar{p}_t, \bar{\omega}_t, s), \quad \dot{\bar{p}}_t = \nabla_x \mathscr{H}(\bar{x}_t, \bar{p}_t, \bar{\omega}_t, s), \quad \nabla_\omega \mathscr{H}(\bar{x}_t, \bar{p}_t, \bar{\omega}_t, s) = 0 \tag{3.5}$$

where the gradient ∇ is considered as the transpose of the underlying differential.

Following the dynamic programming method – see for instance Chapter 4 from [23] or Chapter 3 from [9], as reference works – this minimization problem can be characterized through the solution of a Hamilton-Jacobi-Bellman (HJB) equation. Indeed, assuming that the minimization problem (2.4) has a unique solution, then [21] recalls – see also [25] for a more detailed justification – that the *cost-to-come* defined in (2.13) is a solution – in the viscosity sense – of the Hamilton-Jacobi-Bellman (HJB) equation,

$$\partial_t \mathscr{V}(x,t) - \min_{\omega \in \mathcal{W}} \mathscr{H}(x, \nabla \mathscr{V}(x,t), \omega, t) = 0,$$

which can be simplified by directly finding the minimum in ω, namely using (2.15)

$$\min_{\omega \in \mathcal{W}} \mathscr{H}(x,p,\omega,t) = \bar{\mathscr{H}}(x,p,t) = \frac{1}{2}\|D(x,t)\|_M^2 - \frac{1}{2} p^\intercal BQB^\intercal p - p^\intercal F(x,t),$$

hence, the HJB equation reduces to (2.14).

Then, as recalled in [21], the adjoint variable can be characterized by the *cost-to-come* as

$$\forall s \in [0,t], \quad \nabla \mathscr{V}(\bar{x}_t(s), s) = \bar{p}_t(s). \tag{3.6}$$

With a view to a similar proof in the discrete-time context, we should give a few hints on how (3.6) can be derived when $\mathscr{V}$ is smooth enough. In fact, regarding the initial condition in (2.14), we have

$$\nabla \mathscr{V}(\bar{x}_t(0), 0) = N_\diamond(\bar{x}_t(0) - x_\diamond) = N_\diamond P_\diamond \bar{p}_t(0) = \bar{p}_t(0).$$

Regarding then the dynamics, we consider the HJB equation (2.14) where, since $\mathscr{V}$ is assumed to be $\mathcal{C}^2$, we have for all $(x,t) \in \mathcal{X} \times \mathbb{R}^+$,

$$\nabla \partial_t \mathscr{V}(x,t) - \nabla_x \bar{\mathscr{H}}(x, \nabla \mathscr{V}(x,t), t) - \nabla^2 \mathscr{V}(x,t) \cdot \nabla_p \bar{\mathscr{H}}(x, \nabla \mathscr{V}(x,t), t) = 0,$$

meaning that for $s \in [0,t]$,

$$\begin{aligned} &\nabla \partial_t \mathscr{V}(\bar{x}_t(s), s) + \Big(\mathrm{d}_x F(\bar{x}_t(s), s)^\intercal \nabla \mathscr{V}(\bar{x}_t(s), s) - \mathrm{d}_x D(\bar{x}_t(s), s)^\intercal M D(\bar{x}_t(s), s) \Big) + \nabla^2 \mathscr{V}(\bar{x}_t(s), s) \cdot (F(\bar{x}_t(s), s) \\ &\quad + BQB^\intercal \nabla \mathscr{V}(\bar{x}_t(s), s)) = 0. \end{aligned}$$

Using the chain rule, we get

$$\frac{\mathrm{d}}{\mathrm{d}s} \Big(\nabla \mathscr{V}(\bar{x}_t(s), s) \Big) = \nabla \partial_t \mathscr{V}(\bar{x}_t(s), s) + \nabla^2 \mathscr{V}(\bar{x}_t(s), s) \cdot \dot{\bar{x}}_t(s),$$

which, re-injected in the previous relation, gives

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big(\nabla\mathscr{V}(\bar{x}_t(s),s)\Big) + \mathrm{d}_x F(\bar{x}_t(s),s)^\intercal \nabla\mathscr{V}(\bar{x}_t(s),s) = \mathrm{d}_x D(\bar{x}_t(s),s)^\intercal M D(\bar{x}_t(s),s) + \nabla^2\mathscr{V}(\bar{x}_t(s),s) \\ \cdot BQB^\intercal\big(\bar{p}_t(s) - \nabla\mathscr{V}(\bar{x}_t(s),s)\big), \quad s\in[0,t].$$

Therefore introducing for all s, $m(s) = \nabla\mathscr{V}(\bar{x}_t(s),s) - \bar{p}_t(s)$ that satisfies the dynamics

$$\begin{cases} \dot{m}(s) + (\,\mathrm{d}_x F(\bar{x}_t(s),s)^\intercal - \nabla^2\mathscr{V}(\bar{x}_t(s),s)\cdot BQB^\intercal)m(s) = 0, \quad s\in[0,t]. \\ m(0) = 0 \end{cases}$$

We get that m is null hence we have

$$\forall t\in\mathbb{R}^+,\ \forall s\in[0,t], \quad \nabla\mathscr{V}(\bar{x}_t(s),s) = \bar{p}_t(s). \tag{3.7}$$

Given the (LSE) trajectory $\{\bar{x}_t(s), s\in[0,t]\}$, associated with the minimizers $(\bar{\zeta}_t,\bar{\omega}_t)$, the *optimal* estimator is defined in (2.5) such that

$$\hat{x}(t) = \bar{x}_t(t). \tag{3.8}$$

Injecting (3.8) into the identity (3.7) we characterize $\hat{x}$ by

$$\forall t\in\mathbb{R}^+, \quad \nabla\mathscr{V}(\hat{x}(t),t) = \nabla\mathscr{V}(\bar{x}_t(t),t) = \bar{p}_t(t) = 0, \tag{3.9}$$

Once again, if the cost-to-come is $\mathcal{C}^2$ around the estimated trajectory, we can then characterize the dynamics of the optimal observer by now simply considering (3.9) as a function of t and using the chain rule:

$$\begin{aligned} 0 = \frac{\mathrm{d}}{\mathrm{d}t}\Big(\nabla\mathscr{V}(\hat{x}(t),t)\Big) &= \nabla\partial_t\mathscr{V}(\hat{x}(t),t) + \nabla^2\mathscr{V}(\hat{x}(t),t)\cdot\dot{\hat{x}}(t) \\ &= \nabla\bar{\mathscr{H}}(\hat{x}(t),\nabla\mathscr{V}(\hat{x}(t),t),t) \\ &\qquad \nabla^2\mathscr{V}(\hat{x}(t),t)\cdot\nabla_p\bar{\mathscr{H}}(\hat{x}(t),\nabla\mathscr{V}(\hat{x}(t),t),t) + \nabla^2\mathscr{V}(\hat{x}(t),t)\cdot\dot{\hat{x}}(t), \end{aligned}$$

and from the partial derivatives of $\bar{\mathscr{H}}$, we finally obtain the filter dynamics

$$\begin{cases} \dot{\hat{x}}(t) = F(\hat{x},t) - (\nabla^2\mathscr{V}(\hat{x}(t),t))^{-1}\,\mathrm{d}_x D(\hat{x}(t),t)^\intercal M D(\hat{x}(t),t), \quad t\in\mathbb{R}^+ \\ \hat{x}(0) = x_\diamond \end{cases}$$

and the optimal gain in (2.16) is thus given by $G = (\nabla^2\mathscr{V}(\hat{x}(t),t))^{-1}\,\mathrm{d}_x D(\hat{x}(t),t)^\intercal M$.

4. Proof of Theorem 2.6 and Corollary 2.7

Inspired by the previous section, we now proceed to the proof of Theorem 2.6 under our assumptions of affine dynamics and non-linear observation operator leading to convex least-square criteria.

4.1. Discrete-time Bellman equations

In this first step, we aim at proving the dynamics (2.19) for the costs-to-come. We recall that in Theorem 2.6, we assume that $\mathscr{V}_n^+$ and $\mathscr{V}_n^-$ are $\mathcal{C}^2(\mathcal{X})$ for all $n\in\mathbb{N}$. Let us now introduce the following *discrete Lagrangians*

$\mathscr{L}^+$ and $\mathscr{L}^-$ defined by

$$\mathscr{L}^\pm : \left| \begin{array}{l} \mathcal{X}\times\mathcal{W}\times\mathbb{N}^* \to \mathbb{R} \\ \qquad (x,\omega,n)\mapsto \mathscr{L}^\pm_{n+\frac12\mp\frac12}(x,\omega) = \dfrac12\|D_n(x)\|^2_{M_n} + \dfrac12\|\omega\|^2_{S_{n-\frac12\mp\frac12}}, \end{array}\right. \tag{4.1}$$

with the convention that the model noise norm $S_{-1} = 0$. The discrete dynamics is invertible, hence we directly infer that the Bellman equations satisfied by these *costs-to-come* functions are

$$\mathscr{V}^+_{n+1}(x) = \min_{\omega\in\mathcal{W}}\Big\{\mathscr{V}^+_n(y) + \mathscr{L}^+_{n+1}(x,\omega)\Big\} \quad \text{for } x = F_{n+1|n}(y) + B_n\omega, \tag{4.2}$$

and

$$\mathscr{V}^-_{n+1}(x) = \min_{\omega\in\mathcal{W}}\Big\{\mathscr{V}^-_n(y) + \mathscr{L}^-_{n+1}(y,\omega)\Big\} \quad \text{for } x = F_{n+1|n}(y) + B_n\omega, \tag{4.3}$$

and we denote by $\bar\omega^+_n(x)\in\mathcal{W}$ and $\bar\omega^-_n(x)\in\mathcal{W}$ the respective minimizers of (4.2) and (4.3), which are unique by uniqueness of the minimizers of $\mathscr{J}^+_n$ and $\mathscr{J}^-_{n+1}$.

Let us now characterize $\bar\omega^+_n$ and $\bar\omega^-_n$. On the one hand, we introduce for all $n\in\mathbb{N}^*$

$$\mathcal{F}^+_n : \left| \begin{array}{l} \mathcal{X}\times\mathcal{W} \to \mathbb{R} \\ \quad (x,\omega)\mapsto \mathscr{V}^+_n(F^{-1}_{n+1|n}(x - B_n\omega)) + \mathscr{L}^+_{n+1}(x,\omega), \end{array}\right. \tag{4.4}$$

Since $\mathcal{F}^+_n\in\mathcal{C}^1(\mathcal{X}\times\mathcal{W})$ and $\bar\omega^+_n(x)$ is the minimum for a given x, we have $\mathrm{d}_\omega\mathcal{F}^+_n(x,\bar\omega^+_n) = 0$, hence giving

$$\bar\omega^+_n = Q_n B_n^\intercal(\,\mathrm{d}F_{n+1|n}(\bar y^+_n))^{-\intercal}\nabla\mathscr{V}^+_n(\bar y^+_n), \tag{4.5}$$

with $\bar y^+_n(x) = F^{-1}_{n+1|n}(x - B_n\bar\omega^+_n(x))$. On the other hand, we introduce for all $n\in\mathbb{N}^*$

$$\mathcal{F}^-_n : \left| \begin{array}{l} \mathcal{X}\times\mathcal{W} \to \mathbb{R} \\ \quad (x,\omega)\mapsto \mathscr{V}^-_n(F^{-1}_{n+1|n}(x - B\omega)) + \mathscr{L}^-_{n+1}(F^{-1}_{n+1|n}(x - B_n\omega)),\omega). \end{array}\right. \tag{4.6}$$

Since $\mathcal{F}^-_n\in C^1(\mathcal{X}\times\mathcal{W})$, we have $\mathrm{d}_\omega\mathcal{F}^-_n(x,\bar\omega^-_n) = 0$, hence giving this time

$$\bar\omega^-_n = Q_n B_n^\intercal(\,\mathrm{d}F_{n+1|n}(y^-_n))^{-\intercal}\Big(\nabla\mathscr{V}^-_n(y^-_n) + \,\mathrm{d}_x D_n^\intercal M_n D_n\Big), \tag{4.7}$$

with $\bar y^-_n(x) = F^{-1}_{n+1|n}(x - B_n\bar\omega^-_n(x))$. Then, the expression (4.7) can be simplified by computing the gradient of $\mathscr{V}^-_{n+1}(x) = \mathcal{F}^-_n(x,\bar\omega^-_n(x))$. We have

$$\begin{aligned} \nabla\mathscr{V}^-_{n+1}(x) &= \,\mathrm{d}_x\mathcal{F}^-_n(x,\bar\omega^-_n)^\intercal + \nabla\bar\omega^-_n\,.\,\mathrm{d}_\omega\mathcal{F}^-_n(x,\bar\omega^-_n)^\intercal \\ &= \,\mathrm{d}_x\mathcal{F}^-_n(x,\bar\omega^-_n)^\intercal \\ &= (\,\mathrm{d}F_{n+1|n}(y^-_n(x)))^{-\intercal}\Big(\nabla\mathscr{V}^-_n(y^-_n(x)) + \,\mathrm{d}_x D_n^\intercal M_n D_n\Big). \end{aligned}$$

As a consequence, (4.7) simplifies into

$$\bar\omega^-_n = Q_n B_n^\intercal\nabla\mathscr{V}^-_{n+1}(x). \tag{4.8}$$

Introducing now

$$\bar{\mathscr{L}}^{\pm} : \left| \begin{array}{l} \mathcal{X} \times \mathcal{X} \times \mathbb{N}^* \to \mathbb{R} \\ (x,p,n) \mapsto \bar{\mathscr{L}}^{\pm}_{n+\frac{1}{2}\mp\frac{1}{2}}(x,p) = \frac{1}{2}\|D_n(x)\|^2_{M_n} + \frac{1}{2}p^{\intercal} B_{n-\frac{1}{2}\mp\frac{1}{2}} Q_{n-\frac{1}{2}\mp\frac{1}{2}} B^{\intercal}_{n-\frac{1}{2}\mp\frac{1}{2}} p \end{array} \right.$$

with the convention $Q_{-1} = 0$, we infer from (4.5) and (4.8) the following two Bellman equations

$$\begin{cases} \mathscr{V}^+_{n+1}(x) = \mathscr{V}^+_n(y) + \bar{\mathscr{L}}^+_{n+1}(x, (\,\mathrm{d}F_{n+1|n}(y))^{-\intercal}\nabla\mathscr{V}^+_n(y)) \\ \qquad\qquad\qquad \text{with } y \,|\, x = F_{n+1|n}(y) + B_n Q_n B_n^{\intercal}(\,\mathrm{d}F_{n+1|n}(y))^{-\intercal}\nabla\mathscr{V}^+_n(y) \\ \mathscr{V}^-_{n+1}(x) = \mathscr{V}^-_n(y) + \bar{\mathscr{L}}^-_{n+1}(y, \nabla\mathscr{V}^-_{n+1}(x)) \\ \qquad\qquad\qquad \text{with } y \,|\, x = F_{n+1|n}(y) + B_n Q_n B_n^{\intercal}\nabla\mathscr{V}^-_{n+1}(x) \end{cases} \tag{4.9}$$

In addition, crossed-recursive relations between $\mathscr{V}^+_n$ and $\mathscr{V}^-_n$ are also available. From

$$\begin{aligned} \mathscr{V}^+_n(x) &= \min_{\substack{(\omega_k)_{k<n}\in\mathcal{W}^n \\ \zeta\in\mathcal{X},\, x_n = x}} \mathscr{J}^+_n(\zeta, (\omega_k)_{k<n}) \\ &= \min_{\substack{(\omega_k)_{k<n}\in\mathcal{W}^n \\ \zeta\in\mathcal{X},\, x_n = x}} \Big(\mathscr{J}^-_n(\zeta, (\omega_k)_{k<n}) + \frac{1}{2}\|D_n(x)\|^2_{M_n} \Big), \end{aligned}$$

we deduce that

$$\mathscr{V}^+_n(x) = \mathscr{V}^-_n(x) + \frac{1}{2}\|D_n(x)\|^2_{M_n}.$$

Moreover, from

$$\forall x \in \mathcal{X}, \quad \mathscr{V}^-_{n+1}(x) = \mathscr{V}^+_n(y) + \frac{1}{2}\|\bar{\omega}^-_n\|^2_{S_n} \quad \text{with} \quad x = F_{n+1|n}(y) + B_n\bar{\omega}^-_n \tag{4.10}$$

and the characterization (4.8), we have

$$\mathscr{V}^-_{n+1}(x) = \mathscr{V}^+_n(y) + \frac{1}{2}\nabla\mathscr{V}^-_{n+1}(x)^{\intercal} B_n Q_n B_n^{\intercal}\nabla\mathscr{V}^-_{n+1}(x) \quad \text{with} \quad x = F_{n+1|n}(y) + B_n Q_n B_n^{\intercal}\nabla\mathscr{V}^-_{n+1}(x).$$

Therefore, we have proved (2.19).

4.2. Estimator dynamics

We can now proceed to the justification of the estimator dynamics (2.20). To this end, we first define the adjoint system of the discrete-time dynamics and express – as we did in the continuous-time framework – the discrete-time estimator dynamics from the discrete-time adjoint characterization.

Let us first introduce the discrete-time adjoint system that, for any sequence $(x_k)_{0\le k\le n}$, is the unique solution of the discrete-time dynamics

$$\begin{cases} -p^+_{k|n} + \mathrm{d}F^{\intercal}_{k+1|k} p^+_{k+1|n} = \mathrm{d}D_k(x_k)^{\intercal} M_k D_k(x_k), \quad k \in [0|n] \\ p^+_{n+1|n} = 0. \end{cases} \tag{4.11}$$

Then, we prove the following result, the counterpart of the corresponding result in the continuous-time framework.

Proposition 4.1. *For all $n \in \mathbb{N}$, the trajectory associated with the unique minimizers of $\mathscr{J}_n^+(\cdot,\cdot)$ satisfies the following dynamics*

$$\begin{cases} \bar{x}^+_{k+1|n} = F_{k+1|k}(\bar{x}^+_{k|n}) + B_k Q_k B_k^\intercal \bar{p}^+_{k+1|n}, \quad k \in [0,n] \\ \bar{x}^+_{0|n} = x_\diamond + P_\diamond \bar{p}^+_{0|n} \end{cases} \tag{4.12}$$

where $(\bar{p}^+_{k|n})_{0\le k\le n}$ is the adjoint associated with $(\bar{x}^+_{k|n})_{0\le k\le n}$.

Proof. The proof is classical, see for instance Chapter 9 from [6]. $\mathscr{J}_n^+$ is differentiable since the system dynamics and the observation operator are. Using the adjoint variable dynamics (4.11), we easily compute the Fr?(c)chet derivative of the criterion with respect to ζ. We have for all $\zeta \in \mathcal{X}, (\omega_\ell)_{\ell<n} \in \mathcal{W}^n$,

$$\forall \delta\zeta \in \mathcal{X}, \quad \mathrm{d}_\zeta \mathscr{J}_n^+(\zeta,(\omega_\ell)_{\ell<n})(\delta\zeta) = \zeta^\intercal N_\diamond \delta\zeta - p^{+\intercal}_{0|n}\delta\zeta.$$

Moreover, by differentiating the system dynamics with respect to ω_k for $0 \le k < n$, we obtain

$$\begin{cases} \mathrm{d}_{\omega_k} x_{j+1} = \mathrm{d}F_{j+1|j}\, \mathrm{d}_{\omega_k} x_j + \delta_{k,j} B_j, \quad j \in [k,n] \\ \mathrm{d}_{\omega_k} x_0 = 0 \end{cases}$$

where $\delta_{k,j}$ is the Kronecker symbol. Using again the adjoint dynamics (4.11), we then show that for all $0 \le k < n$, $\zeta \in \mathcal{X}, (\omega_\ell)_{\ell<n} \in \mathcal{W}^n$

$$\forall \delta\zeta \in \mathcal{X}, \quad \mathrm{d}_{\omega_k} \mathscr{J}_n^+(\zeta,(\omega_\ell)_{\ell<n})(\delta\omega_k) = \omega_k^\intercal S_k(\delta\omega_k) - p^{+\intercal}_{k+1|n} B_k(\delta\omega_k).$$

The criterion minimizers $\bar{\zeta}^+_{|n} \in \mathcal{X}$, $(\bar{\omega}^+_{\ell|n})_{\ell<n} \in \mathcal{W}^n$ are then given by

$$\Big(\forall \delta\zeta \in \mathcal{X},\ \mathrm{d}_\zeta \mathscr{J}_n^+(\bar{\zeta}^+_{|n},(\bar{\omega}^+_{\ell|n})_{\ell<n})(\delta\zeta) = 0\Big) \Rightarrow \bar{\zeta}^+_{|n} = P_\diamond \bar{p}^+_{0|n},$$

and for all $0 \le k < n$

$$\Big(\forall \delta\omega_k \in \mathcal{W},\ \mathrm{d}_{\omega_k} \mathscr{J}_n^+(\bar{\zeta}^+_{|n},(\bar{\omega}^+_{\ell|n})_{\ell<n})(\delta\omega_k) = 0\Big) \Rightarrow \bar{\omega}^+_{k|n} = Q_k B_k^\intercal \bar{p}^+_{k+1|n},$$

which concludes the proof. □

A similar proposition can be obtained for the second functional $\mathscr{J}^-_{n+1}$. In particular, we can directly establish here that $\bar{\omega}^-_{n|n+1} = 0$, hence

$$\min_{\zeta,(\omega_k)_{k<n}} \mathscr{J}_n^+(\zeta,(\omega_k)_{k<n}) = \min_{\zeta,(\omega_k)_{k\le n}} \mathscr{J}^-_{n+1}(\zeta,(\omega_k)_{k\le n}).$$

Therefore, we infer that a shift in the index notation suffices to define the sequences $(p^-_{k|n+1})_{0\le k\le n+1}$ with the relation $p^-_{k|n+1} = p^+_{k|n}$, and consequently the optima $(\bar{x}^-_{k|n+1})$ and $(\bar{p}^-_{k|n+1})$ can be defined with a similar shift.

From the Bellman equations, it is now possible to prove the next proposition which is very similar to what we have in the continuous-time context.

Proposition 4.2. *Assuming that the Bellman equation*

$$\begin{cases} \mathscr{V}^-_{n+1}(x) = \mathscr{V}^-_n(y) + \bar{\mathscr{L}}^-_{n+1}(y, \nabla \mathscr{V}^-_{n+1}(x)) \text{ with } y \,|\, x = F_{n+1|n}(y) + B_n Q_n B_n^\intercal \nabla \mathscr{V}^-_{n+1}(x) \\ \mathscr{V}^-_0(x) = \frac{1}{2}\|x - x_\diamond\|^2_{N_\diamond} \end{cases} \tag{4.13}$$

has a $\mathcal{C}^2(\mathcal{X})$ *solution for all* $n \in \mathbb{N}$, *then*

$$\nabla \mathscr{V}^-_k(\bar{x}^+_{k|n}) = \bar{p}^+_{k|n}, \quad k \in [0, n+1]. \tag{4.14}$$

Proof. Let us introduce the sequence $(\check{x}_{k|n})_{k \leq n}$ defined by

$$\check{x}_{k+1|n} = A_{k+1|k}(\check{x}_{k|n}) + B_k Q_k B_k^\intercal \nabla \mathscr{V}^-_{k+1}(\check{x}_{k+1|n}), \quad \check{x}_{0|n} = \bar{x}^+_{0|n},$$

To enforce that the previous sequence is well defined we rely on the following assumptions. First we have assumed that $\mathscr{V}^-_{k+1}$ is $\mathcal{C}^2$. Then we have considered that the discrete-time system discretized the continuous-time system. Hence for Δt small enough with our choice that $B_k Q_k B_k^\intercal = O(\Delta t^2)$ we can define $\check{x}_{k+1|n}$ uniquely from $\check{x}_{k|n}$. Our objective is now to jointly prove by induction the relation (4.14) and

$$\check{x}_{k|n} = \bar{x}^+_{k|n}, \quad k \in [0, n+1].$$

For $k = 0$, we have by definition $\check{x}_{0|n} = \bar{x}^+_{0|n}$ and

$$\nabla \mathscr{V}^-_0(\bar{x}^+_{0|n}) = N_\diamond(\bar{x}^+_{0|n} - x_\diamond) = N_\diamond P_\diamond \bar{p}^+_{0|n} = \bar{p}^+_{0|n}.$$

Then we consider a given $k \in [1, n]$. By differentiating the equation $(4.13)_1$ with respect to y, we obtain

$$\mathrm{d}_y x^\intercal \nabla \mathscr{V}^-_{k+1}(x) - \nabla \mathscr{V}^-_k(y) + \nabla_y \bar{\mathscr{L}}^-_{k+1}(y, \nabla \mathscr{V}^-_{k+1}(x)) + \nabla^2 \mathscr{V}_{k+1}(x) \cdot\, \mathrm{d}_y x^\intercal \nabla_p \bar{\mathscr{L}}(y, \nabla \mathscr{V}^-_{k+1}(x)) = 0,$$

giving

$$\begin{aligned} &\mathrm{d}F_{k+1|k}(y)^\intercal \nabla \mathscr{V}^-_{k+1}(x) + \nabla^2 \mathscr{V}^-_{k+1}(x) \cdot\, \mathrm{d}_y x^\intercal B_k Q_k B_k^\intercal \nabla \mathscr{V}^-_{k+1}(x) \\ &\qquad = \nabla \mathscr{V}^-_k(y) + \nabla^2 \mathscr{V}_{k+1}(x) \cdot\, \mathrm{d}_y x^\intercal B_k Q_k B_k^\intercal \nabla \mathscr{V}^-_{k+1}(x) + \mathrm{d}D_k(y)^\intercal M_k D_k(y), \end{aligned}$$

which simplifies into

$$\mathrm{d}F_{k+1|k}(y)^\intercal \nabla \mathscr{V}^-_{k+1}(x) = \nabla \mathscr{V}^-_k(y) + \mathrm{d}D_k(y)^\intercal M_k D_k(y).$$

We thus obtain

$$-\nabla \mathscr{V}^-_k(\check{x}_{k|n}) + \mathrm{d}F_{k+1|k}(\check{x}_{k|n})^\intercal \nabla \mathscr{V}^-_{k+1}(\check{x}_{k+1|n}) = -\,\mathrm{d}D_k(\check{x}_{k|n})^\intercal M_k D_k(\check{x}_{k|n}).$$

which implies that

$$\nabla \mathscr{V}^-_{k+1}(\check{x}_{k+1|n}) = \bar{p}^+_{k+1|n} \text{ and } \check{x}_{k+1|n} = \bar{x}^+_{k+1|n}.$$

□

We can now derive the equation of the non-linear discrete-time estimator – namely the discrete-time equivalent of system (2.16). Let us consider the two sequences $(\hat{x}_n^-)_{n\in\mathbb{N}}$ and $(\hat{x}_n^+)_{n\in\mathbb{N}}$ defined in (2.8). First, we recall that by definition

$$\hat{x}_{n+1}^- = \bar{x}_{n+1|n}^- = F_{n+1|n}(\bar{x}_{n|n}^-).$$

But $\bar{x}_{n|n}^- = \hat{x}_n^+$ since

$$\min_{\substack{\zeta\in\mathcal{X}\\(\omega_k)_{k\leq n}\in\mathcal{W}^n}} \mathscr{J}_{n+1}^-(\zeta,(\omega_k)_{k\leq n}) = \min_{\substack{\zeta\in\mathcal{X}\\(\omega_k)_{k<n}\in\mathcal{W}^n}} \mathscr{J}_n^+(\zeta,(\omega_k)_{k<n}) + \min_{\omega_n\in\mathcal{W}} \frac{1}{2}\|\omega_n\|_S^2$$

with unique minimizers $\bar{\zeta}_{|n}^- = \bar{\zeta}_{|n}^+$, $(\bar{\omega}_{k|n}^-)_{k<n} = (\bar{\omega}_{k|n}^+)_{k<n}$ and $\bar{\omega}_{n|n}^- = 0$. Therefore we have

$$\hat{x}_{n+1}^- = F_{n+1|n}(\hat{x}_n^+).$$

as announced in (2.20).

Then by Proposition 4.2, we obtain

$$\nabla\mathscr{V}_{n+1}^-(F_{n+1|n}(\hat{x}_n^+)) = \bar{p}_{n+1|n}^+ = 0,$$

which proves Corollary 2.7. Moreover, by differentiating $(2.19)_3$, with $x = \hat{x}_{n+1}^-$ and $y = \hat{x}_n^+$, we then get

$$\nabla\mathscr{V}_n^+(\hat{x}_n^+) = \,\mathrm{d}F_{n+1|n}(\hat{x}_n^+)^\intercal\nabla\mathscr{V}_{n+1}^-(\hat{x}_{n+1}^-) = 0. \tag{4.15}$$

which justifies the characterization of $\hat{x}_n^+$ in (2.20) and concludes the proof of Theorem 2.6.

We now end this section by finally justifying (2.21). Indeed, by differentiating $(2.19)_1$ around $x = \hat{x}_{n+1}^-$ and $y = \hat{x}_n^+$ we obtain

$$\nabla\mathscr{V}_{n+1}^+(\hat{x}_{n+1}^-) = \,\mathrm{d}D_{n+1}(\hat{x}_{n+1}^-)^\intercal M_{n+1}D_{n+1}(\hat{x}_{n+1}^-). \tag{4.16}$$

4.3. The full linear case

We now proceed to the justification of Proposition 2.8. From the discrete-time model dynamics (2.20), we should be able to retrieve the classic Kalman filter in a linear context. We thus consider for all n an affine dynamics and a linear observation operator, namely

$$F_{n+1|n}(x) = A_{n+1|n}x + R_n, \quad \text{and} \quad H_n(x) = H_n x.$$

Moreover we continue to assume that they are bounded and that $A_{n+1|n}$ is invertible. In order to be compatible with the assumptions of our Theorem 2.6, we consider the case where the discrete system is observable and controllable, namely when the Kalman conditions of observability and controllability are satisfied [27, 43]. In this case, we have unique minimizers for the criteria $\mathscr{J}_n^-$ and $\mathscr{J}_n^+$. The optimal observer exists and is defined by the Kalman Filter dynamics (2.11) and the covariance operators, P_n^- and P_n^+, solutions of (2.12), are invertible, see Section 9.5.3 from [27] – or in a more general context of potentially infinite dimensional systems see [2] or Part II–Chapter 1 from [8].

Since observability and controllability conditions imply that the covariance operators, P_n^- and P_n^+, solutions of (2.12) are invertible, we have from $(2.11)_2$ that

$$H_n^\intercal M_n(z_n - H_n\hat{x}_n^-) = (P_n^+)^{-1}(\hat{x}_n^+ - \hat{x}_n^-),$$

thus we have

$$\frac{1}{2}(x-\hat{x}_n^+)^\intercal (P_n^+)^{-1}(x-\hat{x}_n^+) = \frac{1}{2}(x-\hat{x}_n^-)^\intercal (P_n^+)^{-1}(x-\hat{x}_n^-) + \frac{1}{2}(\hat{x}_n^+-\hat{x}_n^-)^\intercal (P_n^+)^{-1}(\hat{x}_n^+-\hat{x}_n^-) \\ -(x-\hat{x}_n^-)^\intercal H_n^\intercal M_n(z_n - H_n\hat{x}_n^-).$$

Then we verify, as we did in the continuous setting, that

$$-\frac{1}{2}\|z_n - H_n x\|^2_{M_n} + \frac{1}{2}\|H_n(x-\hat{x}_n^-)\|^2_{M_n} - (x-\hat{x}_n^-)^\intercal H_n^\intercal M_n(z_n-H_n\hat{x}_n^-) + \frac{1}{2}\|z_n - H_n\hat{x}_n^-\|^2_{M_n} = 0.$$

Therefore, using from (2.12)$_2$ that

$$(P_n^+)^{-1} = (P_n^-)^{-1} + H_n^\intercal M_n H_n,$$

we get

$$\begin{aligned}&\frac{1}{2}(x-\hat{x}_n^+)^\intercal (P_n^+)^{-1}(x-\hat{x}_n^+) - \frac{1}{2}\|z_n - H_n x\|^2_{M_n} \\ &= \frac{1}{2}(x-\hat{x}_n^-)^\intercal (P_n^-)^{-1}(x-\hat{x}_n^-) \underbrace{-\frac{1}{2}\|z_n - H_n\hat{x}_n^-\|_{M_n} + \frac{1}{2}(\hat{x}_n^+-\hat{x}_n^-)^\intercal (P_n^+)^{-1}(\hat{x}_n^+-\hat{x}_n^-)}_{=\mathscr{V}^0_{n-1}-\mathscr{V}^0_n \text{ from (2.24)}},\end{aligned} \tag{4.17}$$

which means that

$$\mathscr{V}_n^+(x) - \frac{1}{2}\|z_n - H_n x\|^2_{M_n} = \mathscr{V}_n^-(x),$$

and proves (2.19)$_2$. Moreover, applying the identity (4.17) at every step k with $x = \hat{x}_k^+$, we reach the identity

$$\forall k > 0, \quad \|z_k - H_k\hat{x}_k^+\|^2_{M_k} + \|\hat{x}_k^+ - \hat{x}_k^-\|^2_{(P_k^+)^{-1}} = \|z_k - H_k\hat{x}_k^-\|^2_{M_k} - \|\hat{x}_k^+ - \hat{x}_k^-\|^2_{(P_k^-)^{-1}},$$

which proves (2.25) from (2.24). Now, we can show that the candidates *cost-to-come* (2.23) follow the recursive relations (2.19). First we write

$$y - \hat{x}_n^+ = F_{n+1|n}^{-1}\big(\mathbb{1} - B_n Q_n B_n^\intercal (P_{n+1}^-)^{-1}\big)(x - \hat{x}_{n+1}^-).$$

Then, using the classical identity

$$F_{n+1|n}^{-\intercal}(P_n^+)^{-1}F_{n+1|n}^{-1} = \big(P_{n+1}^- - B_n Q_n B_n^\intercal\big)^{-1},$$

we get

$$(y-\hat{x}_n^+)^\intercal (P_n^+)^{-1}(y-\hat{x}_n^+) = (x-\hat{x}_{n+1}^-)^\intercal (P_{n+1}^-)^{-1}\big(P_{n+1}^- - B_n Q_n B_n^\intercal\big)(P_{n+1}^-)^{-1}(x-\hat{x}_{n+1}^-),$$

which demonstrates that

$$(x-\hat{x}_{n+1}^-)^\intercal (P_{n+1}^-)^{-1}(x-\hat{x}_{n+1}^-) = (y-\hat{x}_n^+)^\intercal (P_n^+)^{-1}(y-\hat{x}_n^+) \\ +(x-\hat{x}_{n+1}^-)^\intercal (P_{n+1}^-)^{-1}B_n Q_n B_n^\intercal (P_{n+1}^-)^{-1}(x-\hat{x}_{n+1}^-). \tag{4.18}$$

Therefore with

$$\nabla \mathscr{V}_{n+1}^{-}(x) = (P_{n+1}^{-})^{-1}(x - \hat{x}_{n+1}^{-}),$$

we prove (2.19)$_3$. We have verified the two recursive relations satisfied by the *costs-to-come*. The initial conditions are easy to verify, hence the costs-to-come solution of Proposition 2.8 are proved by induction.

Finally, we infer that the Hessian of the *costs-to-come* are constant. In particular,

$$\nabla \mathscr{V}_n^{+}(\hat{x}_n^{-}) - \underbrace{\nabla \mathscr{V}_n^{+}(\hat{x}_n^{+})}_{=0} = \nabla^2 \mathscr{V}_n^{+}(\hat{x}_n^{+}) \cdot (\hat{x}_n^{-} - \hat{x}_n^{+}),$$

which, combined with (4.16), gives

$$\begin{aligned}\hat{x}_n^{+} &= \hat{x}_n^{-} + (\nabla^2 \mathscr{V}_n^{+}(\hat{x}_n^{+}))^{-1} H_n^{\intercal} M_n (z_n - H_n \hat{x}_n^{-}) \\ &= \hat{x}_n^{-} + P_n^{+} H_n^{\intercal} M_n (z_n - H_n \hat{x}_n^{-}),\end{aligned}$$

as expected in the classical Kalman Filter formulation. This last result concludes our proof of Proposition 2.8.

Moreover, the linear case has two benefits. It shows that our set of assumptions can be simultaneously satisfied. In addition, it can be used to validate our numerical implementation in Section 6.

Remark 4.3 (Consistency with the continuous-time solution)**.** It is worth noticing that in the linear case, the discrete-time *costs-to-come* (2.23) have a similar solution when compared with the continuous-time cost-to-come analytical solution (2.17) up to a consistency term $\frac{1}{2}\sum_{k=0}^{n} \|\hat{x}_k^{+} - \hat{x}_k^{-}\|^2_{(P_k^{+})^{-1}}$. We observe that this term does not modify the fundamental positivity property of the *costs-to-come*. Indeed, we directly have from (2.25) that $\forall n > 0, \mathscr{V}_n^0 \geq 0$. Such consistency can in fact be generalized in the non-linear configuration as demonstrated in the next section.

5. The discrete-time Mortensen filter as a time-discretization of the continuous-time one

As presented in the introduction, the discrete-time model (2.2) can correspond to a consistent discretization with a fixed time-step Δt of the continuous-time system given by (2.1). Then, we want to establish that the discrete-time HJB formulation and the discrete-time Mortensen estimator converge to their continuous-time counterparts. Similar types of results have been sought in the literature [20] and required developments that are out of reach for this paper. Here, we limit our presentation to demonstrate that the discrete-time HJB formulation and the discrete-time Mortensen estimator are consistent with respect to their continuous-time counterparts. Without much loss of generality, the consistency of (2.2) can be obtained by assuming a time discretization of (2.1) using an Euler time-scheme

$$\forall n \in \mathbb{N}, \quad \frac{x_{n+1} - x_n}{\Delta t} = F(x_{n+1}, t_n) + B\omega,$$

leading to the definition of the discrete operator

$$F_{n+1|n}(\cdot) = (\mathbb{1} - \Delta t F(\cdot, t_n))^{-1}, \quad B_n = \Delta t B,$$

and for the observations we can keep $D_n = D$. To simplify the expression D and B can be chosen as time-independent. Beside the model, the discrete criterion considered should be consistent with its limit when Δt

tends to 0. Hence, we assume that, for all $n \in \mathbb{N}$, the discrete norms read

$$M_n = \Delta t\, M, \quad S_n = \Delta t\, S \quad \text{and} \quad W_n = \frac{1}{\Delta t} W, \quad Q_n = \frac{1}{\Delta t} Q,$$

so that $\lim_{\Delta t \to 0} \mathscr{J}_n^-(\cdot,\cdot) = \lim_{\Delta t \to 0} \mathscr{J}_n^+(\cdot,\cdot) = \mathscr{J}(\cdot,\cdot,t)$, with $n\Delta t = t$.

5.1. Consistency of the discretization of the estimator and HJB equation

We can now infer the consistency of the (discrete-time) Bellman equations (4.13) with respect to the continuous-time HJB formulation (2.14). Let us first focus on the *cost-to-come* prediction equation $(4.13)_1$. The consistency error is then defined by

$$\epsilon_n^- = \frac{1}{\Delta t}\Big(-\mathscr{V}(x,t_{n+1}) + \mathscr{V}(y,t_n) + \frac{\Delta t}{2}\|D(y,t_{n+1})\|_M^2 + \frac{\Delta t}{2}\nabla_x \mathscr{V}(x,t_{n+1})^\intercal BQB^\intercal \nabla_x \mathscr{V}(x,t_{n+1})\Big),$$
$$\text{with } \; x - \Delta t F(x,t_n) = y + \Delta t BQB^\intercal \nabla_x \mathscr{V}(x,t_n). \tag{5.1}$$

By a Taylor expansion, we get

$$\mathscr{V}(x,t_{n+1}) = \mathscr{V}(x,t_n) + \Delta t \partial_t \mathscr{V}(x,t_n) + O(\Delta t^2),$$

and

$$\mathscr{V}(y,t_n) = \mathscr{V}(x,t_n) - \Delta t \nabla_x \mathscr{V}(x,t_n)^\intercal F(x,t_n) - \Delta t \nabla_x \mathscr{V}(x,t_n)^\intercal BQB^\intercal \nabla_x \mathscr{V}(x,t_n) + O(\Delta t^2),$$

but also

$$\frac{\Delta t}{2}\|D(y,t_{n+1})\|_M^2 = \frac{\Delta t}{2}\|D(x,t_n)\|_M^2 + O(\Delta t^2),$$

and

$$\frac{\Delta t}{2}\nabla_x \mathscr{V}(x,t_{n+1})^\intercal BQB^\intercal \nabla_x \mathscr{V}(x,t_{n+1}) = \frac{\Delta t}{2}\nabla_x \mathscr{V}(x,t_n)^\intercal BQB^\intercal \nabla_x \mathscr{V}(x,t_n) + O(\Delta t^2).$$

Hence, we find that

$$\epsilon_n^- = -\partial_t \mathscr{V}(x,t_n) - \nabla_x \mathscr{V}(x,t_n)^\intercal F(x,t_n) + \frac{1}{2}\|D(x,t_n)\|_M^2 - \frac{1}{2}\nabla_x \mathscr{V}(x,t_n)^\intercal BQB^\intercal \nabla_x \mathscr{V}(x,t_n) + O(\Delta t) = O(\Delta t),$$

which shows that the Bellman equation satisfied by the discrete-time cost-to-come is first-order consistent with the continuous-time HJB equation. Therefore, the Bellman equation (4.13) can be seen as a consistent time scheme of (2.14). However, we mention that the time scheme for the *cost-to-come* appears to be first-order in Δt independently of the order of approximation of the model, observation and model noise operator.

Focusing now on the state estimator dynamics equation, we note from (2.20) that

$$\nabla \mathscr{V}_n^+(\hat{x}_n^+) - \nabla \mathscr{V}_n^+(\hat{x}_n^-) = -\,\mathrm{d}D_n(\hat{x}_n^-)^\intercal M_n D_n(\hat{x}_n^-), \tag{5.2}$$

and by a simple Taylor expansion

$$\nabla \mathscr{V}_n^+(\hat{x}_n^+) - \nabla \mathscr{V}_n^+(\hat{x}_n^-) = \nabla^2 \mathscr{V}_n^+(\hat{x}_n^+) \cdot (\hat{x}_n^+ - \hat{x}_n^-) + O(\|\hat{x}_n^+ - \hat{x}_n^-\|^2). \tag{5.3}$$

From these two relations – and recalling that $M_n = \Delta t\, M$ – we directly infer that

$$\hat{x}_n^+ - \hat{x}_n^- = O(\Delta t).$$

Then, by using the model prediction,

$$\hat{x}_n^+ = F_{n|n-1}(\hat{x}_{n-1}^+, t_{n-1}) - \big(\nabla^2 \mathscr{V}_n^+(\hat{x}_n^+)\big)^{-1} \, \mathrm{d}D_n(\hat{x}_n^-)^\intercal M_n D_n(\hat{x}_n^-) + O(\Delta t^2),$$

this shows that

$$\frac{\hat{x}_n^+ - \hat{x}_{n-1}^+}{\Delta t} = \frac{1}{\Delta t}(F_{n|n-1}(\cdot, t_{n-1}) - \mathbb{1})(\hat{x}_n^+) - \big(\nabla^2 \mathscr{V}_n^+(\hat{x}_n^+)\big)^{-1} \, \mathrm{d}_x D(\hat{x}_n^-)^\intercal M D(\hat{x}_n^-) + O(\Delta t),$$

and leads directly for $\hat{x}_n^+$ – and consequently for $\hat{x}_n^-$ – to a first order consistent time scheme of the continuous-time estimator $\hat{x}(t)$ – of dynamics (2.16).

5.2. Liapunov property consistency

More than the consistency, the most difficult part when trying to discretize the continuous-time Mortensen estimator is to define stable discretizations, especially for the HJB equation. Without completing such a proof which may require a full dedicated paper, we would like to show how our proposed discretization offers new perspectives in order to establish the stability of the time discretization jointly with the convergence of the estimator to a target trajectory. In this respect, we recall that that the observer purpose is at least to be able to *converge* in time to the following ideal target trajectory

$$\dot{\check{x}} = F(\check{x}, t), \quad D(\check{x}, t) = 0,$$

which, for example, can be generated synthetically. When observability and controllability conditions are satisfied [33], it means that we should expect that the error $\tilde{x}(t) = \check{x}(t) - \hat{x}(t)$ tends to 0. When observability and controllability conditions are not satisfied we expect at least a weaker property of the type [34] where there exists a contraction mapping β

$$\|\tilde{x}(t)\| \leq \beta(\|\tilde{x}(s)\|, t - s), \quad \forall t \geq s.$$

One tremendous advantage of optimal filtering is that the convergence property can be obtained in a very general class of system by considering a Liapunov functional of the form

$$\forall x \in \mathcal{X}, \quad \tilde{\mathscr{V}}(x, t) = \mathscr{V}(x + \hat{x}(t), t) - \mathscr{V}(\hat{x}(t), t), \tag{5.4}$$

which satisfies

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{dt}} \tilde{\mathscr{V}}(\tilde{x}(t), t) &= \partial_t \mathscr{V}(\check{x}(t), t) + \nabla \mathscr{V}(\check{x}(t), t)^\intercal \dot{\check{x}}(t) - \partial_t \mathscr{V}(\hat{x}(t), t) - \nabla \mathscr{V}(\hat{x}(t), t)^\intercal \dot{\hat{x}}(t) \\
&= -\frac{1}{2} \nabla \mathscr{V}(\check{x}(t), t)^\intercal B Q B^\intercal \nabla \mathscr{V}(\check{x}(t), t) + \underbrace{\frac{1}{2}\|D(\check{x}, t)\|_M^2}_{=0} \\
&\quad + \underbrace{\frac{1}{2} \nabla \mathscr{V}(\hat{x}(t), t)^\intercal B Q B^\intercal \nabla \mathscr{V}(\hat{x}(t), t)}_{=0} - \frac{1}{2}\|D(\hat{x}(t), t)\|_M^2 \\
&= -\frac{1}{2} \nabla \mathscr{V}(\check{x}(t), t)^\intercal B Q B^\intercal \nabla \mathscr{V}(\check{x}(t), t) - \frac{1}{2}\|D(\hat{x}(t), t)\|_M^2 \leq 0.
\end{aligned} \tag{5.5}$$

This type of property is typically required to prove that the error converges in time to 0 after considering adequate conditions of existence on the criterion minimum see [32, 33, 44] for such complete error analysis.

Without entering into too much details on the difficult question of observer convergence, our purpose is to show that the same type of Liapunov property exists with our discrete-time estimator which offers new perspectives of convergence at the discrete-time level. Formally, we follow the same approach as in the continuous-time formulation with the definition of a Liapunov function – here however, we will consider a discrete-time Liapunov function. Let us consider an ideal discrete-time target system

$$\forall n, \quad \breve{x}_{n+1} = F_{n+1|n}(\breve{x}_n), \quad D_n(\breve{x}_n) = 0.$$

We define the discrete-time Liapunov function

$$\forall x, \quad \tilde{\mathscr{V}}_n(x) = \mathscr{V}_n^-(x + \hat{x}_n^-) - \mathscr{V}_n^-(\hat{x}_n^-).$$

This function is going to be evaluated on the error

$$\forall n \quad \tilde{x}_n = \breve{x}_n - \hat{x}_n^-,$$

as in the continuous formulation to satisfy here the hypothesis of discrete time Liapunov stability theorems for non autonomous systems [32, 44].

First, since $\nabla \mathscr{V}_{n+1}^-(\hat{x}_{n+1}^-) = 0$ implies $\mathscr{V}_{n+1}^-(\hat{x}_{n+1}^-) = \mathscr{V}_n^+(\hat{x}_n^+)$ due to $(2.19)_3$, we see, as in the continuous formulation, that the Liapunov function estimator part satisfies

$$\begin{aligned}
-\mathscr{V}_{n+1}^-(\hat{x}_{n+1}^-) + \mathscr{V}_n^-(\hat{x}_n^-) &= -\mathscr{V}_n^+(\hat{x}_n^+) + \mathscr{V}_n^-(\hat{x}_n^-) \\
&= -\mathscr{V}_n^-(\hat{x}_n^+) - \frac{1}{2}\|D_n(\hat{x}_n^-)\| + \mathscr{V}_n^-(\hat{x}_n^-) \leq -\frac{1}{2}\|D_n(\hat{x}_n^-)\|_{M_n},
\end{aligned}$$

$\hat{x}_n^-$ being the minimizer of $\mathscr{V}_{n+1}^-$. Then, on the target system part, we have by definition

$$\mathscr{V}_{n+1}^-(\breve{x}_{n+1}) \leq \mathscr{V}_n^+(\breve{x}_n) + \mathscr{L}_{n+1}^-(\breve{x}_n, 0) \leq \mathscr{V}_n^+(\breve{x}_n).$$

In our setting, the *cost-to-come* $\mathscr{V}_{n+1}^-$ functional is convex. Indeed, it is obvious for $n = 0$. Then, for $n \geq 0$, we consider $(x_1, x_2) \in \mathcal{X}^2$, and $\lambda \in [0, 1]$ and consider

$$\mathscr{V}_{n+1}^-(\lambda x_1 + (1-\lambda)x_2) = \min_{\substack{(\omega_k)_{k \leq n} \\ x_{n+1|\zeta,(\omega_k)_{k\leq n}} = \lambda x_1 + (1-\lambda)x_2}} \mathscr{J}_{n+1}^-(\zeta, (\omega_k)_{k\leq n})$$

Then, since we consider affine mapping for the dynamics and $\mathscr{J}_{n+1}^-$ is convex, we can prove that

$$\begin{aligned}
\mathscr{V}_{n+1}^-(\lambda x_1 + (1-\lambda)x_2) &\leq \min_{\substack{(\omega_{1,k})_{k\leq n}, (\omega_{2,k})_{k\leq n} \\ x_1 = x_{n+1|\zeta_1,(\omega_{1,k})_{k\leq n}} \\ x_2 = x_{n+1|\zeta_2,(\omega_{2,k})_{k\leq n}}}} \mathscr{J}_{n+1}^-(\lambda\zeta_1 + (1-\lambda)\zeta_2, (\lambda\omega_{1,k} + (1-\lambda)\omega_{2,k})_{k\leq n}) \\
&\leq \lambda \mathscr{V}_{n+1}^-(x_1) + (1-\lambda)\mathscr{V}_{n+1}^-(x_2).
\end{aligned}$$

Hence we have,

$$\mathscr{V}_{n+1}^-(\breve{x}_{n+1} + B_n Q_n B_n^\intercal \nabla \mathscr{V}_{n+1}^-(\breve{x}_{n+1})) \geq \mathscr{V}_{n+1}^-(\breve{x}_{n+1}) + \nabla \mathscr{V}_{n+1}^-(\breve{x}_{n+1})^\intercal B_n Q_n B_n^\intercal \nabla \mathscr{V}_{n+1}^-(\breve{x}_{n+1}),$$

and get

$$\begin{aligned}\mathscr{V}_{n+1}^{-}(\breve{x}_{n+1}) &\leq \mathscr{V}_{n}^{+}(\breve{x}_{n}) + \mathscr{L}_{n+1}^{-}(\breve{x}_{n}, Q_n B_n^{\intercal} \nabla \mathscr{V}_{n+1}^{-}(\breve{x}_{n+1})) - \nabla \mathscr{V}_{n+1}^{-}(\breve{x}_{n+1})^{\intercal} B_n Q_n B_n^{\intercal} \nabla \mathscr{V}_{n+1}^{-}(\breve{x}_{n+1})\\ &\leq \mathscr{V}_{n}^{-}(\breve{x}_{n}) - \frac{1}{2} \nabla \mathscr{V}_{n+1}^{-}(\breve{x}_{n+1})^{\intercal} B_n Q_n B_n^{\intercal} \nabla \mathscr{V}_{n+1}^{-}(\breve{x}_{n+1}).\end{aligned}$$

We obtain in this case the exact analogous of the continuous identity (5.5)

$$\tilde{\mathscr{V}}_{n+1}(\tilde{x}_{n+1}) - \tilde{\mathscr{V}}_{n}(\tilde{x}_{n}) \leq -\frac{1}{2} \nabla \mathscr{V}_{n+1}^{-}(\breve{x}_{n+1})^{\intercal} B_n Q_n B_n^{\intercal} \nabla \mathscr{V}_{n+1}^{-}(\breve{x}_{n+1}) - \frac{1}{2} \| D_n(\hat{x}_n^{-}) \|_{M_n},$$

with, here, an inequality instead of the equality in (5.5). We point out that this inequality is sharp – at least in the linear framework – with respect to the additive consistent terms that we have found in the *costs-to-come* expressions (2.24) and (2.25).

6. Numerical implementation

To demonstrate the discrete optimality of our proposed discrete-time Mortensen estimator, we now propose a full discretization strategy and numerical algorithm of practical use. This will allow us to illustrate and verify numerically the various properties of the discrete-time Mortensen estimator seen in the previous sections.

6.1. Resolution of HJB prediction by a Newton-Raphson algorithm

We assume that the state lies in a bounding box $[a_1, b_1] \times \cdots \times [a_d, b_d] \subset \mathbb{R}^d$. For the spatial discretization we consider a regular grid and a Lagrange interpolation rule, typically cubic. We denote by g the number of points in the grid. A function $\phi : \mathbb{R}^d \to \mathbb{R}^r$ of the state $x \in \mathbb{R}^d$ is discretized on the grid and therefore represented by a vector $\vec{\Phi} \in \mathbb{R}^{g \times r}$. For any index k ($\mathbb{1} \leq k \leq g \times r$), there exists an equivalent couple $\{\ell, i\}$ such that $[\vec{\Phi}]_k$ corresponds to the i-th component of ϕ at the ℓ-th point of the grid. Therefore, for the sake of clarity in the sequel we will use the couple $\{\ell, i\}$ notation to refer to the global index k. Furthermore, $[\vec{\Phi}]_{\{\ell,\cdot\}}$ will be considered as a vector of $\mathbb{R}^r$ whereas $[\vec{\Phi}]_{\{\cdot,i\}}$ will refer to the corresponding vector of $\mathbb{R}^g$.

Let us give two specific examples of the use of the above notation. The function that associates with each point of the grid the corresponding state is defined by $\vec{X} \in \mathbb{R}^{g \times d}$ and we will define by $[\vec{X}]_{(\ell,i)}$ the i-th coordinate of the ℓ-th point of the grid. Then $[\vec{X}]_{\{\ell,\cdot\}}$ is the state of the ℓ-th point of the grid. Besides, a scalar function f is represented by $\vec{F} \in \mathbb{R}^g$ and we can assimilate the initial function f with its interpolation reconstructed from $\vec{F}$.

Considering the derivatives of this scalar function f, we choose an adequate finite differentiation rule allowing to define each Gâteaux derivatives $\overrightarrow{\nabla^{(i)} F}$, $(1 \leq i \leq d)$ from $\vec{F}$. This generates d derivative operators $\boldsymbol{\nabla}^{(i)} \in \mathrm{M}_g(\mathbb{R})$. Here also we neglect the interpolation errors to directly associate $\boldsymbol{\nabla}^{(i)} \vec{F} = \overrightarrow{\nabla^{(i)} f}$. The total gradient is then defined from the d partial derivatives to obtain an operator $\boldsymbol{\nabla}$ such that $[\boldsymbol{\nabla} \vec{F}]_{\{\ell,\cdot\}} = \nabla f([\vec{X}]_{\{\ell,\cdot\}})$.

We apply the rules presented above for the discretization of the *costs-to-come* functions $\mathscr{V}_n^-$ and $\mathscr{V}_n^+$ to define two vectors $\vec{\mathscr{V}}_n^- \in \mathbb{R}^g$ and $\vec{\mathscr{V}}_n^+ \in \mathbb{R}^g$ as the degrees of freedom of the *costs-to-come* functions and the corresponding derivatives by computing $\boldsymbol{\nabla}^{(i)} \vec{\mathscr{V}}_n^-$ or $\boldsymbol{\nabla}^{(i)} \vec{\mathscr{V}}_n^+$.

Meanwhile, let us associate with $\vec{X}$ the vector $\vec{Y}$ such that

$$\forall 1 \leq \ell \leq g, \quad [\vec{X}]_{\{\ell,\cdot\}} = F_{n+1|n}([\vec{Y}]_{\{\ell,\cdot\}}) + B_n Q_n B_n^{\intercal} \nabla \mathscr{V}_{n+1}^{-}([\vec{X}]_{\{\ell,\cdot\}}),$$

to discretize the field $y(x)$ in (2.19). As a consequence the *cost-to-come* prediction is computed with the formulation

$$\forall 1 \leq \ell \leq g, \quad [\vec{\mathscr{V}}^-_{n+1}]_{(\ell)} = \mathscr{V}_n^+(\vec{Y}_{\{\ell,\cdot\}}) + \frac{1}{2} \sum_{1\leq i,j\leq d} [\boldsymbol{\nabla}^{(i)}\vec{\mathscr{V}}^-_{n+1}]_{(\ell)}[B_nQ_nB_n^\intercal]_{(i,j)}[\boldsymbol{\nabla}^{(j)}\vec{\mathscr{V}}^-_{n+1}]_{(\ell)},$$

In other words, for the prediction, we compute $\vec{\mathscr{V}}^-_{n+1}$ as the component $\vec{\mathscr{V}}$ of the solution of

$$\forall n \in \mathbb{N}, \quad \text{Find } (\vec{\mathscr{V}}, \vec{Y}) \text{ such that } \left| \begin{array}{l} \vec{\mathcal{F}}_{\text{hjb}}(\vec{\mathscr{V}}, \vec{Y}) = 0 \\ \vec{\mathcal{F}}_{\text{dyn}}(\vec{\mathscr{V}}, \vec{Y}) = 0 \end{array} \right., \tag{6.1}$$

where for $\ell \in [1, g]$

$$\begin{cases} [\vec{\mathcal{F}}_{\text{hjb}}(\vec{\mathscr{V}}, \vec{Y})]_{(\ell)} & = -[\vec{\mathscr{V}}]_{(\ell)} + \mathscr{V}_n^+([\vec{Y}]_{\{\ell,\cdot\}}) + \frac{1}{2}\sum_{1\leq i,j\leq d}[\boldsymbol{\nabla}^{(i)}\vec{\mathscr{V}}]_{(\ell)}[B_nQ_nB_n^\intercal]_{(i,j)}[\boldsymbol{\nabla}^{(j)}\vec{\mathscr{V}}]_{(\ell)} \\ [\vec{\mathcal{F}}_{\text{dyn}}(\vec{\mathscr{V}}, \vec{Y})]_{\{\ell,\cdot\}} & = -[\vec{X}]_{\{\ell,\cdot\}} + F_{n+1|n}([\vec{Y}]_{\{\ell,\cdot\}}) + B_nQ_nB_n^\intercal[\boldsymbol{\nabla}\vec{\mathscr{V}}]_{\{\ell,\cdot\}} \end{cases}$$

We solve (6.1) with a Newton algorithm, ergo we compute the sequence $(\vec{\mathscr{V}}^m, \vec{Y}^m)_{m\in\mathbb{N}}$ starting from

$$\left| \begin{array}{l} \vec{\mathscr{V}}_0 = \vec{\mathscr{V}}_n^+ \\ \vec{Y}_0 = \vec{X} \end{array} \right.$$

such that

$$\begin{bmatrix} \mathbf{d}_{\vec{\mathscr{V}}}\vec{\mathcal{F}}_{\text{hjb}}(\vec{\mathscr{V}}^m, \vec{Y}^m) & \mathbf{d}_{\vec{Y}}\vec{\mathcal{F}}_{\text{hjb}}(\vec{\mathscr{V}}^m, \vec{Y}^m) \\ \mathbf{d}_{\vec{\mathscr{V}}}\vec{\mathcal{F}}_{\text{dyn}}(\vec{\mathscr{V}}^m, \vec{Y}^m) & \mathbf{d}_{\vec{Y}}\vec{\mathcal{F}}_{\text{dyn}}(\vec{\mathscr{V}}^m, \vec{Y}^m) \end{bmatrix} \begin{bmatrix} \vec{\mathscr{V}}^{m+1} - \vec{\mathscr{V}}^m \\ \vec{Y}^{m+1} - \vec{Y}^m \end{bmatrix} = - \begin{bmatrix} \vec{\mathcal{F}}_{\text{hjb}}(\vec{\mathscr{V}}^m, \vec{Y}^m) \\ \vec{\mathcal{F}}_{\text{dyn}}(\vec{\mathscr{V}}^m, \vec{Y}^m) \end{bmatrix}.$$

Here we compute for all $(k, \ell) \in [1, g]^2$,

$$[\,\mathbf{d}_{\vec{\mathscr{V}}}\vec{\mathcal{F}}_{\text{hjb}}(\vec{\mathscr{V}}^m, \vec{Y}^m)]_{(k,\ell)} = -\delta_{k,\ell} + \sum_{1\leq i,j\leq d}[\boldsymbol{\nabla}^{(i)}\vec{\mathscr{V}}^m]_{(k)}[B_nQ_nB_n^\intercal]_{(i,j)}[\boldsymbol{\nabla}^{(j)}]_{(k,\ell)},$$

and for all $(k, \ell, i) \in [1, g]^2 \times [1, d]$,

$$[\,\mathbf{d}_{\vec{\mathscr{V}}}\vec{\mathcal{F}}_{\text{dyn}}(\vec{\mathscr{V}}^m, \vec{Y}^m)]_{(\{\ell,i\},k)} = \sum_{1\leq j\leq d}[B_nQ_nB_n^\intercal]_{(i,j)}[\boldsymbol{\nabla}^{(j)}]_{(k,\ell)}.$$

Moreover we get for all $\ell \in [1, g]$

$$[\,\mathbf{d}_{\vec{Y}}\vec{\mathcal{F}}_{\text{hjb}}(\vec{\mathscr{V}}^m, \vec{Y}^m)]_{(\ell,\{\ell,\cdot\})} = \nabla^{(i)}\mathscr{V}_n^+([\vec{Y}^m]_{\{\ell,\cdot\}}),$$

whereas, for all $(\ell, k) \in [1, g]^2$ with $k \neq \ell$,

$$[\,\mathbf{d}_{\vec{Y}}\vec{\mathcal{F}}_{\text{hjb}}(\vec{\mathscr{V}}^m, \vec{Y}^m)]_{(k,\{\ell,\cdot\})} = 0.$$

In the same manner we have, for all $\ell \in [1, g]$,

$$[\,\mathbf{d}_{\vec{Y}} \vec{\mathcal{F}}_{\mathrm{dyn}}(\vec{\mathscr{V}}^m, \vec{Y}^m)]_{(\{\ell,\cdot\},\{\ell,\cdot\})} = \,\mathrm{d}F_{n+1|n}([\vec{Y}^m]_{\{\ell,\cdot\}}),$$

whereas, for all $(\ell, k) \in [1, g]^2$ with $k \neq \ell$,

$$[\,\mathbf{d}_{\vec{Y}} \vec{\mathcal{F}}_{\mathrm{dyn}}(\vec{\mathscr{V}}^m, \vec{Y}^m)]_{(\{\ell,\cdot\},\{k,\cdot\})} = 0.$$

These computations introduce block diagonal terms allowing to solve the complete Newton through the use of Schur complements. In fact for any grid point $[\vec{X}]_k$, we need to solve *locally* an inverse dynamics to reconstruct the corresponding $[\vec{Y}]_k$. For these Schur complement computations, it is convenient to introduce the vector residual

$$\vec{R}_{\mathrm{sch}}(\vec{\mathscr{V}}^m, \vec{Y}^m) = \big(\,\mathbf{d}_{\vec{Y}} \vec{\mathcal{F}}_{\mathrm{dyn}}(\vec{\mathscr{V}}^m, \vec{Y}^m)\big)^{-1} \cdot \vec{\mathcal{F}}_{\mathrm{dyn}}(\vec{\mathscr{V}}^m, \vec{Y}^m),$$

and the operator

$$\mathbf{T}_{\mathrm{sch}}(\vec{\mathscr{V}}^m, \vec{Y}^m) = \big(\,\mathbf{d}_{\vec{Y}} \vec{\mathcal{F}}_{\mathrm{dyn}}(\vec{\mathscr{V}}^m, \vec{Y}^m)\big)^{-1} \cdot \,\mathbf{d}_{\vec{\mathscr{V}}} \vec{\mathcal{F}}_{\mathrm{dyn}}(\vec{\mathscr{V}}^m, \vec{Y}^m).$$

Indeed, when computing on the one hand

$$\vec{\mathscr{V}}^{m+1} = \vec{\mathscr{V}}^m - \Big(\,\mathbf{d}_{\vec{\mathscr{V}}} \vec{\mathcal{F}}_{\mathrm{hjb}}(\vec{\mathscr{V}}^m, \vec{Y}^m) - \,\mathbf{d}_{\vec{Y}} \vec{\mathcal{F}}_{\mathrm{hjb}}(\vec{\mathscr{V}}^m, \vec{Y}^m)\big(\,\mathbf{d}_{\vec{Y}} \vec{\mathcal{F}}_{\mathrm{dyn}}(\vec{\mathscr{V}}^m, \vec{Y}^m)\big)^{-1} \,\mathbf{d}_{\vec{\mathscr{V}}} \vec{\mathcal{F}}_{\mathrm{dyn}}(\vec{\mathscr{V}}^m, \vec{Y}^m)\Big)^{-1} \tag{6.2}$$

$$\cdot\Big(\vec{\mathcal{F}}_{\mathrm{hjb}}(\vec{\mathscr{V}}^m, \vec{Y}^m) - \,\mathbf{d}_{\vec{Y}} \vec{\mathcal{F}}_{\mathrm{hjb}}(\vec{\mathscr{V}}^m, \vec{Y}^m)\big(\,\mathbf{d}_{\vec{Y}} \vec{\mathcal{F}}_{\mathrm{dyn}}(\vec{\mathscr{V}}^m, \vec{Y}^m)\big)^{-1} \cdot \vec{\mathcal{F}}_{\mathrm{dyn}}(\vec{\mathscr{V}}^m, \vec{Y}^m)\Big), \tag{6.3}$$

this reads

$$\vec{\mathscr{V}}^{m+1} = \vec{\mathscr{V}}^m - \Big(\,\mathbf{d}_{\vec{\mathscr{V}}} \vec{\mathcal{F}}_{\mathrm{hjb}}(\vec{\mathscr{V}}^m, \vec{Y}^m) - \,\mathbf{d}_{\vec{Y}} \vec{\mathcal{F}}_{\mathrm{hjb}}(\vec{\mathscr{V}}^m, \vec{Y}^m)\mathbf{T}_{\mathrm{sch}}(\vec{\mathscr{V}}^m, \vec{Y}^m)\Big)^{-1} \tag{6.4}$$

$$\cdot\Big(\vec{\mathcal{F}}_{\mathrm{hjb}}(\vec{\mathscr{V}}^m, \vec{Y}^m) - \,\mathbf{d}_{\vec{Y}} \vec{\mathcal{F}}_{\mathrm{hjb}}(\vec{\mathscr{V}}^m, \vec{Y}^m) \cdot \vec{R}_{\mathrm{sch}}(\vec{\mathscr{V}}^m, \vec{Y}^m)\Big). \tag{6.5}$$

On the other hand, we have

$$\vec{Y}^{m+1} = \vec{Y}^m - \big(\,\mathbf{d}_{\vec{Y}} \vec{\mathcal{F}}_{\mathrm{dyn}}(\vec{\mathscr{V}}^m, \vec{Y}^m)\big)^{-1} \cdot \Big(\vec{\mathcal{F}}_{\mathrm{dyn}}(\vec{\mathscr{V}}^m, \vec{Y}^m) + \,\mathbf{d}_{\vec{\mathscr{V}}} \vec{\mathcal{F}}_{\mathrm{dyn}}(\vec{\mathscr{V}}^m, \vec{Y}^m) \cdot (\vec{\mathscr{V}}^{m+1} - \vec{\mathscr{V}}^m)\Big),$$

which can be rewritten in the form

$$\vec{Y}^{m+1} = \vec{Y}^m - \vec{R}_{\mathrm{sch}}(\vec{\mathscr{V}}^m, \vec{Y}^m) - \mathbf{T}_{\mathrm{sch}}(\vec{\mathscr{V}}^m, \vec{Y}^m) \cdot (\vec{\mathscr{V}}^{m+1} - \vec{\mathscr{V}}^m).$$

As a result, we need to compute for all $((k, i), \ell) \in ([1, g] \times [1, d]) \times [1, g]$,

$$[\mathbf{T}_{\mathrm{sch}}(\vec{\mathscr{V}}^m, \vec{Y}^m)]_{((k,i),\ell)} = \sum_{1 \leq j \leq d} [(\,\mathrm{d}F_{n+1|n}([\vec{Y}^m]_{\{k,\cdot\}}))^{-1} B_n Q_n B_n^{\intercal}]_{(i,j)} [\boldsymbol{\nabla}^{(j)}]_{(k,\ell)},$$

so that for all $(k, \ell) \in [1, g]^2$,

$$\Big[\,\mathbf{d}_{\vec{Y}} \vec{\mathcal{F}}_{\mathrm{hjb}}(\vec{\mathscr{V}}^m, \vec{Y}^m)\mathbf{T}_{\mathrm{sch}}(\vec{\mathscr{V}}^m, \vec{Y}^m)\Big]_{(k,\ell)} = \sum_{1 \leq i,j \leq d} \nabla^{(i)} \mathscr{V}^m([\vec{Y}^m]_{\{k,\cdot\}})[(\,\mathrm{d}F_{n+1|n}([\vec{Y}^m]_{\{k,\cdot\}}))^{-1} B_n Q_n B_n^{\intercal}]_{(i,j)} [\boldsymbol{\nabla}^{(j)}]_{(k,\ell)}.$$

Meanwhile, we get, for all $(\ell, i) \in [1, g] \times [1, d]$, the residuals

$$[\vec{R}_{\mathrm{sch}}(\vec{\mathscr{V}}^m, \vec{Y}^m)]_{(\ell,i)} = \sum_{1 \le j \le d} [(\,\mathrm{d}F_{n+1|n}([\vec{Y}^m]_{\{\ell,\cdot\}}))^{-1}]_{(i,j)} [\vec{\mathcal{F}}_{\mathrm{dyn}}(\vec{\mathscr{V}}^m, \vec{Y}^m)]_{(\ell,j)},$$

and for all $\ell \in [1, g]$

$$\Big[\mathbf{d}_{\vec{Y}} \vec{\mathcal{F}}_{\mathrm{hjb}}(\vec{\mathscr{V}}^m, \vec{Y}^m) \cdot \vec{R}_{\mathrm{sch}}(\vec{\mathscr{V}}^m, \vec{Y}^m)\Big]_{(\ell)} = \sum_{1 \le i,j \le d} \nabla^{(i)} \mathscr{V}^m([\vec{Y}^m]_{\{\ell,\cdot\}})[(\,\mathrm{d}F_{n+1|n}([\vec{Y}^m]_{\{\ell,\cdot\}}))^{-1}]_{(i,j)} [\vec{\mathcal{F}}_{\mathrm{dyn}}(\vec{\mathscr{V}}, \vec{Y})]_{(\ell,j)}.$$

All these vectors and matrices are finally combined in the Newton loop.

Concerning the boundary conditions, there is no easy way to handle the unbounded domain on which the HJB equation is computed. We therefore follow a classical strategy by assuming that outside the grid the *costs-to-come* are extrapolated linearly from their values and gradients on the bounding box boundary. In other words, we bound the domain with simplified Robin boundary conditions.

6.1.1. Numerical results

A scalar quadratic example. We start our numerical investigations by focusing our attention on the computation of the *cost-to-come*. For that matter, we will compare our algorithm with a more standard discretization of the Hamilton-Jacobi-Bellman solution in one dimension of space. We consider, here, the model associated with

$$F : (x, \omega) \mapsto a_0 + a_1 x(t) + a_2 x(t)^2 + b\omega(t),$$

where

$$a_0 = 1; \quad a_1 = -1; \quad a_2 = 1; \quad b = 1.$$

With this model, we generate a scalar observation z for $\omega = 0$ starting from $x_\diamond = 0.3$ such that in this case

$$\begin{cases} \dot{z}(t) = a_0 + a_1 z(t) + a_2 z(t)^2 \\ z(0) = x_\diamond \end{cases}$$

Setting all normalization coefficients to 1, this leads to the following HJB equation

$$\begin{cases} \partial_t \mathscr{V}(x,t) + (a_0 + a_1 x + a_2 x^2)\partial_x \mathscr{V}(x,t) + \dfrac{1}{2}(\partial_x \mathscr{V}(x,t))^2 - \dfrac{1}{2}(z(t) - x)^2 = 0, \quad (x,t) \in \mathcal{X} \times \mathbb{R}^+ \\ \mathscr{V}(x,0) = \frac{1}{2}(x - x_\diamond)^2 \end{cases}$$

The space-time domain computation is $(x,t) \in [-1,1] \times [0,1]$.

In the data assimilation library *Verdandi*[1] [14], a discretization of the same HJB equation introduced in [4][2] is performed using standard spatial interpolation rules in space and a Godunov time discretization. Boundary conditions consist of a linear extrapolation on each boundary. The results are presented in Figure 1. The first two plots correspond to the prediction $\mathscr{V}^-$ and correction $\mathscr{V}^+$ functions of our method, whereas the last one is given by the Godunov time scheme.

In a second step, we study the convergence of our time scheme by computing error convergences for several time discretizations. In fact from $\mathscr{V}^-_{\Delta t, \Delta x}$ with $(\Delta t, \Delta x) = (10^{-2}, 10^{-4})$, we compute the error for solutions with

[1] http://verdandi.sourceforge.net.

[2] In fact in [4] the HJB equation definition leads to a solution which is half the value of our HJB solution.

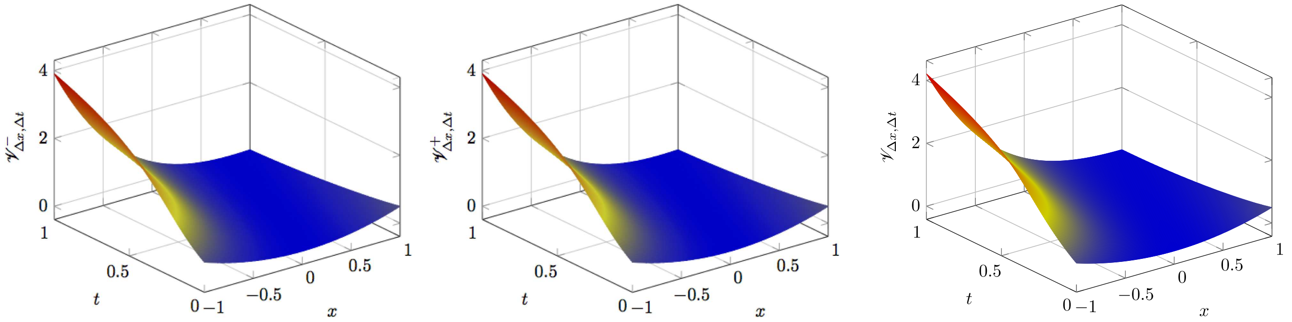

FIGURE 1. Comparison for $(\Delta x, \Delta t) = (10^{-2}, 10^{-4})$ of the proposed splitting time scheme (prediction (*left*) and correction (*center*)) with respect to a Godunov time-scheme (*right*). (Color online.)

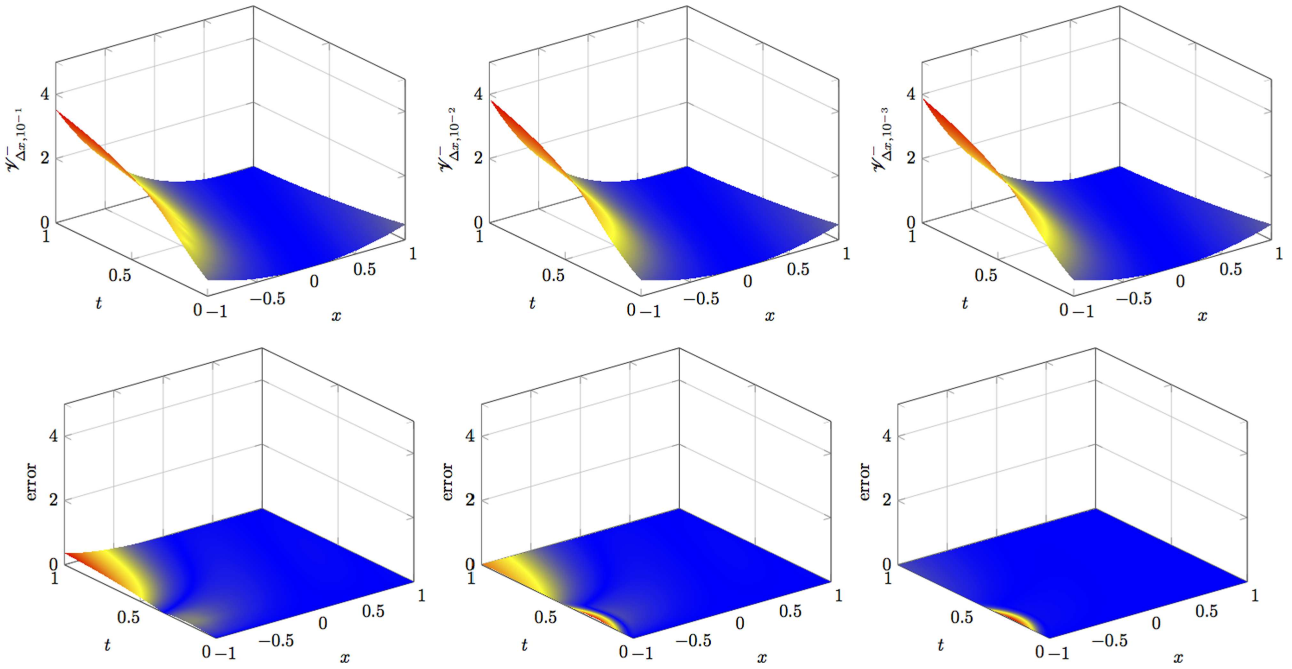

FIGURE 2. Time convergence of the HJB solution for $\Delta x = 10^{-2}$ and $\Delta t = 10^{-1}, 10^{-2}$, and 10^{-3} with respect to the solution computed with $(\Delta t, \Delta x) = (10^{-2}, 10^{-4})$. (Color online.)

$\Delta x = 10^{-2}$ and $\Delta t = 10^{-1}$, 10^{-2}, and 10^{-3}. The results are presented in Figure 2 where we plot the solution and the corresponding error. Note that the spatial discretization $\Delta x = 10^{-2}$ was chosen to balance spatial interpolation accuracy and computational complexity. We justify our spatial discretization choice in Figure 3 with two different spatial steps: $\Delta x = 10^{-2}$ and 10^{-3}, $\Delta t = 10^{-3}$. We point out that our time-scheme does not have a CFL condition and can then be computed for any time and space discretization which is not the case for more classical approaches as a Godunov time discretization. Typically in our configuration the CFL condition for the Godunov time scheme imposes at least $\Delta t < 10^{-2}$.

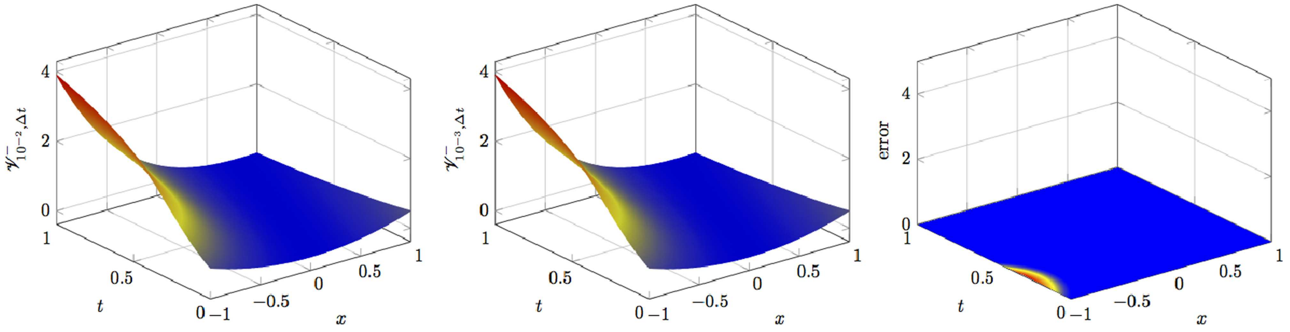

FIGURE 3. Two spatial discretizations $\Delta x = 10^{-2}$ and $\Delta x = 10^{-3}$ for $\Delta t = 10^{-3}$ and the corresponding error. (Color online.)

The pendulum. In a linear configuration we have shown that the optimal filter reduces to the classical Kalman filter and we expect to verify this property numerically. We consider a simple pendulum problem

$$\ddot{y} + \mu y = f + b\omega,$$

that we rewrite in a first order form $\dot{x} = Fx + R$ with $v = \dot{y}$,

$$x = \begin{pmatrix} y \\ v \end{pmatrix}, \quad F = \begin{pmatrix} 0 & 1 \\ \mu & 0 \end{pmatrix}, \quad R = \begin{pmatrix} 0 \\ f \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ b \end{pmatrix}.$$

We choose a mid-point rule stable time-discretization

$$\begin{cases} \frac{y_{n+1}-y_n}{\Delta t} = \frac{v_{n+1}+v_n}{2}, \\ \frac{v_{n+1}-v_n}{\Delta t} + \mu\frac{y_{n+1}+y_n}{2} = f, \end{cases}$$

such that $F_{n+1|n} = A_1^{-1}A_0$ and $R_n = A_1^{-1}R$ with

$$A_1 = \begin{pmatrix} \frac{1}{\Delta t} & -\frac{1}{2} \\ -\frac{\mu}{2} & \frac{1}{\Delta t} \end{pmatrix}, \quad A_0 = \begin{pmatrix} \frac{1}{\Delta t} & \frac{1}{2} \\ -\frac{\mu}{2} & \frac{1}{\Delta t} \end{pmatrix}.$$

In Figure 4, we present a direct simulation generated with $\mu = 0.2$ and an initial condition of $(y(0), v(0)) = (1, 0)$. The model noise coefficient is $b = 0.5$ but the reference solution will be generated without model noise *i.e.* $\omega = 0$. The time discretization step is $\Delta t = 0.1$. From this solution, we generate observations of the displacement only, and we add an observation noise of covariance $1e^{-3}$. We then consider a second solution starting from $(y(0), v(0)) = (0.5, 0)$ and an initial covariance $P_\diamond = \mathbb{1}$. Starting from the *a priori*, the estimator retrieves the target trajectory and the convergence of the observer is illustrated in Figure 4 for a spatial grid of 20 by 20 points on $[-1, 1] \times [-1, 1]$ - namely $\Delta x_1 = \Delta x_2 = 0.1$.

In order to evaluate the accuracy of our discretization, we plot in Figure 5 the numerical values computed for $\|\nabla\mathscr{V}^-(\hat{x}_n^-)\|$ with two HJB discretization grids: 10 by 10 points and 20 by 20 points. We found that the corresponding error with respect to 0 of the equality (2.22) is of the order of magnitude of our spatial discretization as presented in the first plot of Figure 5. Finally, we know that in this particular case we should retrieve exactly the values of the discrete covariance computed by the discrete-time Kalman filter. This is the case as shown in the second plot of Figure 5 that displays the difference $\|\nabla^2\mathscr{V}_n^+(\hat{x}_n^+) - P_n^+\|_2$ for the two grid discretizations.

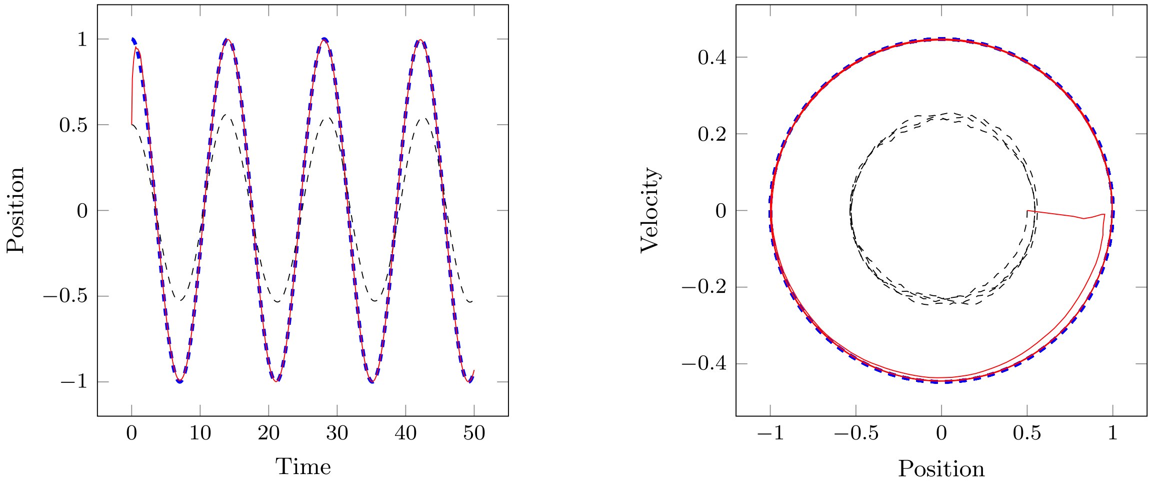

FIGURE 4. Trajectory (*left*) and phase portrait (*right*) of the target trajectory (blue, dashed), the estimator (red), and the uncorrected trajectory (black dashed). (Color online.)

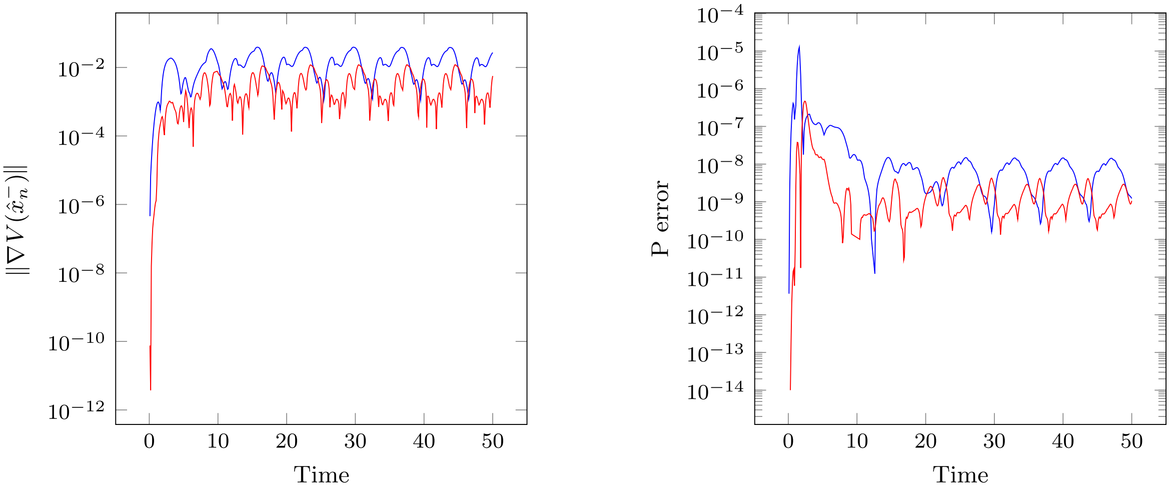

FIGURE 5. Optimality verification from the time plot of $\|\nabla\mathscr{V}^-(\hat{x}_n^-)\|$ (*left*) and Covariance computation verification with discrete-time Riccati solution (*right*). 10×10 discretization (blue) and 20×20 discretization (red). (Color online.)

The Van Der Pol oscillator. Our last illustration deals with the classical non-linear Van Der Pol oscillator defined by

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = -\mu(1-x_1^2)x_2 + x_1 + b\omega, \end{cases}$$

discretized using a mid-point scheme and solved with a Newton algorithm. Note that this example is compatible with the algorithm formulation given in Theorem 2.6, but goes beyond the proof that we have presented in Section 4.

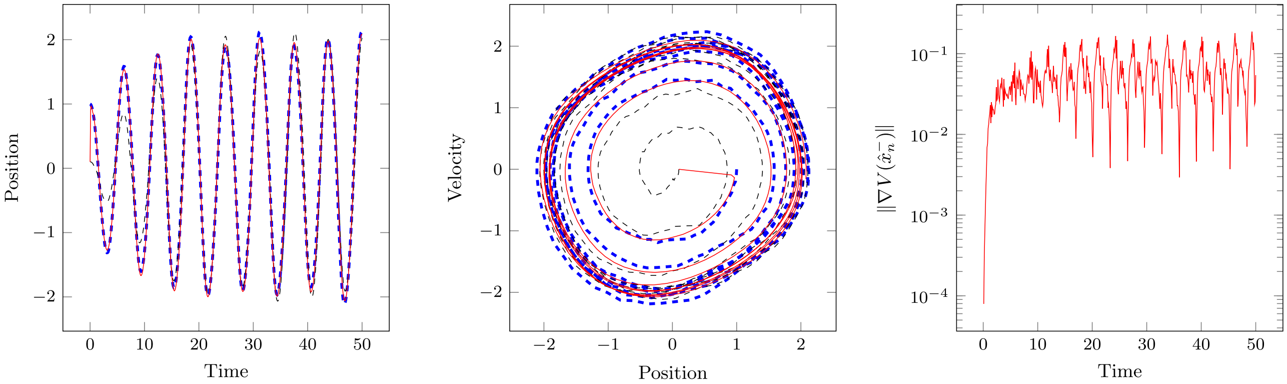

FIGURE 6. Trajectory (*left*) and phase portrait (*middle*) of the target trajectory (blue, dashed), the estimator (red), and the uncorrected trajectory (black dashed). Optimality verification from the time plot of $\|\nabla \mathscr{V}^-(\hat{x}_n^-)\|$ (*right*). (Color online.)

We choose in this example $\mu = 0.2$, $b = 0.1$ and $x_\diamond = \begin{pmatrix} 0.1 \\ 0 \end{pmatrix}$. The target trajectory is generated with $\zeta = \begin{pmatrix} 0.9 \\ 0 \end{pmatrix}$ and ω is defined as a 10 % white noise. The observations are generated from the first variable and perturbed with a 10 % white noise also.

In Figure 6, we present the target trajectory and the estimator computed with an HJB grid of 30 by 30 points in $[-3, 3]^2$. We point out that each Newton iteration of the estimator requires to compute the *costs-to-come* by Newton iterations which themselves necessitate solving the model at each point of the grid also by Newton iterations. We verify the accuracy of our algorithm by computing again $\|\nabla \mathscr{V}^-(\hat{x}_n^-)\|$.

7. Conclusion

In this work, we have presented and fully justified an exact optimal deterministic observer in a discrete-time non-linear framework. We completed the proof for an affine dynamics and a non-linear observation operator. However our estimator formulation is compatible with more general non-linear formulations and our numerical examples illustrate that the optimality of our estimator can be preserved in such cases. This observer is based on a prediction-correction evolution and the correction step is solved using a Newton algorithm. This observer reduces to the classical Kalman observer for linear systems. In non-linear configurations, it allows to understand in a deterministic framework the level of approximation made by approximate optimal filters such as the Extended Kalman Filter (EKF) [43] or the Unscented Kalman Filter (UKF) [26]. Moreover, our discrete-time formulation can be considered as a time discretization of the original Mortensen observer – and associated HJB equation – defined in continuous time. Our time discretization reveals to be consistent and unconditionally stable under some assumptions associated with the well-posedness of the minimization problem. This new discrete-time counterpart of the Mortensen filter allows to fill the gap between continuous-time and discrete-time for deterministic filters. A unified version of the two filters can now be envisioned using the time scales formalism as it is for control problems [41]. For practical use, we have presented a complete algorithm based on a simple spatial interpolation rule. The curse of dimensionality remains when using this observer in practice. However, it can be considered on a reduced model, even in order to validate the use of an EKF or UKF on a more complex model. Moreover, as in [35] where an HJB based feedback control is computed for PDEs based on reduced basis – Proper Orthogonal Decomposition in this case – this type of model reduction can also be applied for observers [15]. Besides, reduced-order observers with adequate discrete-time formulation can now be formulated similarly to the already existing RoEKF or RoUKF filters [14, 38, 42]. Finally, some improvements of the spatial discretization can be directly considered to reasonably increase the HJB dimension in the numerical simulations, for example sparse grids interpolation methods [11, 12] should allow to consider a state dimension of up to 10. We can also mention max-plus based strategies for handling high-dimensional problems [22].

References

[1] B.D.O. Anderson and J.B. Moore, Detectability and stabilizability of time-varying discrete-time linear systems. *SIAM J. Control Optim.* **19** (1981) 20–32.

[2] J.S. Baras and A. Bensoussan, On Observer Problems for Systems Governed by Partial Differential Equations. Technical Report. Maryland Univ., College Park (1987).

[3] J.S. Baras, A. Bensoussan and M.R. James, Dynamic observers as asymptotic limits of recursive filters: special cases. *SIAM J. Appl. Math.* **48** (1988) 1147–1158.

[4] J.S. Baras and A. Kurzhanski, Nonlinear Filtering: The Set-Membership (Bounding) and the H8 Techniques. Technical Report TR 1995-40, ISR (1995).

[5] R.E. Bellman, Dynamic Programming. Princeton University Press (1957).

[6] A. Bensoussan, Filtrage Optimal des Systèmes Linéaires. Dunod (1971).

[7] A. Bensoussan, Stochastic Control of Partially Observable Systems. Cambridge University Press, Cambridge (1992).

[8] A. Bensoussan, G. Da Prato, M.C. Delfour and S.K. Mitter, Representation and Control of Infinite-Dimensional Systems. Vol. II of Systems & Control: Foundations & Applications. Birkhäuser Boston Inc., Boston, MA (1993).

[9] D.P. Bertsekas, Dynamics Programming and Optimal Control. 3rd edn. Athena Scientific, Vol. 1 (2005).

[10] J. Blum, F.-X. Le Dimet and I.M. Navon, Data assimilation for geophysical fluids. *Comput. Methods Atmos. Ocean* **14** (2009) 385–441

[11] O. Bokanowski, J. Garcke, M. Griebel and I. Klompmaker, An adaptive sparse grid semi-Lagrangian scheme for first order Hamilton-Jacobi Bellman equations. *J. Sci. Comput.* **55** (2013) 575–605.

[12] H.-J. Bungartz and M. Griebel, Sparse grids. *Acta Numer.* **13** (2004) 147–269.

[13] L. Cesari, Existence theorems for weak and usual optimal solutions in Lagrange problems with unilateral constraints. II. Existence theorems for weak solutions. *Trans. Am. Math. Soc.* **124** (1966) 413–430.

[14] D. Chapelle, M. Fragu, V. Mallet and P. Moireau, Fundamental principles of data assimilation underlying the verdandi library: applications to biophysical model personalization within euheart. *Med. Biol. Eng. Comput.* **51** (2012) 1221–1233.

[15] D. Chapelle, A. Gariah, P. Moireau and J. Sainte-Marie, A Galerkin strategy with Proper Orthogonal Decomposition for parameter-dependent problems – analysis, assessments and applications to parameter estimation. *ESAIM: M2AN* **47** (2013) 1821–1843.

[16] G. Chavent, Nonlinear Least Squares for Inverse Problems. Springer (2010).

[17] Z. Chen, Bayesian filtering: From Kalman filters to particle filters, and beyond. *Statistics* **182** (2003) 1–69.

[18] N. Cîndea, A. Imperiale and P. Moireau, Data assimilation of time under-sampled measurements using observers, the wave-like equation example. *ESAIM: COCV* **21** (2015) 635–669.

[19] H. Cox, On the estimation of state variables and parameters for noisy dynamic systems. *IEEE Trans. Autom. Control* (1964).

[20] A.L. Dontchev, Discrete approximations in optimal control, in Nonsmooth Analysis and Geometric Methods in Deterministic Optimal Control. Springer, New York, NY (1996) 59–80.

[21] W.H. Fleming, Deterministic nonlinear filtering. *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* **25** (1997) 435–454.

[22] W.H. Fleming and W.M. McEneaney, A max-plus-based algorithm for a Hamilton–Jacobi–Bellman equation of nonlinear filtering. *SIAM J. Control Optim.* **38** (2000) 683–710.

[23] W.H. Fleming and R.W. Rischel, Deterministic and Stochastic Optimal Control. Springer-Verlag (1975).

[24] O. Hijab, Asymptotic nonlinear filtering and large deviations. *Adv. Filter. Optim. Stoch. Control* (1982) 170–176.

[25] M.R. James and J.S. Baras, Nonlinear filtering and large deviations: a PDE-control theoretic approach. *Stochastics* **23** (1988) 391–412.

[26] S.J. Julier and J.K. Uhlmann, New extension of the Kalman filter to nonlinear systems. *Proc. SPIE* **3068** (1997) 182–193.

[27] T. Kailath, A.H. Sayed and B. Hassibi, Linear Estimation. Prentice Hall, New Jersey, Vol. 1 (2000).

[28] R.E. Kalman, Contributions to the theory of optimal control. *Bol. Soc. Mat. Mexicana* **5** (1960) 102–119.

[29] R.E. Kalman, A new approach to linear filtering and prediction problems. *J. Basic Eng.* **82** (1960) 35–45.

[30] R.E. Kalman, Mathematical description of linear dynamical systems. *J. SIAM Control Ser. A* **1** (1963) 152–192.

[31] R.E. Kalman and R. Bucy, New results in linear filtering and prediction theory. *Trans. ASME J. Basic Eng.* **83** (1961) 95–108.

[32] A.J. Krener, A Lyapunov theory of nonlinear observers, in G.G Yin and Q. Zhang eds. Stochastic Analysis, Control, Optimization and Applications. Springer (1998) 409–420.

[33] A.J. Krener, The convergence of the minimum energy estimator, in New Trends in Nonlinear Dynamics and Control, and their Applications. Springer, Berlin (2003).

[34] A.J. Krener and A. Duarte, A hybrid computational approach to nonlinear estimation, in Proceedings of the 35th IEEE Decision and Control, 1996 (1996) 1815–1819.

[35] K. Kunisch, S. Volkwein and L. Xie, HJB-POD-based feedback design for the optimal control of evolution problems. *SIAM J. Appl. Dyn. Syst.* **3** (2004) 701–722.

[36] H.J. Kushner, Dynamical equations for optimal nonlinear filtering. *J. Differ. Equ.* **3** (1967) 179–190.

[37] F.-X. Le Dimet and O. Talagrand, Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A* **38** (2010) 97–110.

[38] P. Moireau and D. Chapelle, Reduced-order Unscented Kalman Filtering with application to parameter identification in large-dimensional systems. *ESAIM: COCV* **17** (2011) 380–405.

[39] R.E. Mortensen, Maximum-likelihood recursive nonlinear filtering. *J. Optim. Theory Appl.* **2** (1968) 386–394.

[40] I.M. Navon, Data assimilation for numerical weather prediction: a review, in S.K. Park and L. Xu eds. Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications. Springer, Berlin, Heidelberg (2009).

[41] Y. Peng, X. Xiang and Y. Jiang, Nonlinear dynamic systems and optimal control problems on time scales. *ESAIM: COCV* **17** (2010) 654–681.

[42] D.T. Pham, J. Verron and M.C. Roubaud, A singular evolutive extended Kalman filter for data assimilation in oceanography. *J. Mar. Syst.* **16** (1998) 323–340.

[43] D. Simon, Optimal State Estimation: Kalman, H^∞, and Nonlinear Approaches. Wiley-Interscience (2006).

[44] M. Vidyasagar, Nonlinear Systems Analysis. Prentice-Hall Internaltional Editions, Englewood Cliffs, NJ (1993).