Discrete-time formulations as time discretization strategies in data assimilation

Philippe Moireau

▶ **To cite this version:**

Philippe Moireau. Discrete-time formulations as time discretization strategies in data assimilation. Handbook of Numerical Analysis, Numerical Control: Part B, 2, Elsevier; Chapter - 9; Elsevier, pp.297-339, 2023, Handbook of Numerical Analysis, 10.1016/bs.hna.2022.11.005 . hal-03921465

HAL Id: hal-03921465
https://inria.hal.science/hal-03921465

Submitted on 3 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Discrete time formulations as time discretization strategies in data assimilation

Philippe Moireau[1,2]

[1] Inria, 1 rue Honoré d'Estienne d'Orves, 91128 Palaiseau, France

[2] LMS, Ecole Polytechnique, CNRS, Institut Polytechnique de Paris, 91128 Palaiseau, France

Handbook on Numerical Control and Beyond, Part B

Abstract

Data assimilation combines control theory and scientific computing to propose a set of methods for coupling dynamic models and data sequences for estimation and prediction in all engineering domains. Data assimilation naturally raises the question of how the developed control and optimization methods interact with the discretization of the underlying physical models, in particular their temporal discretization. We would like to present here some of the best known techniques developed for discrete-time models, which are essentially based on a mechanism involving model prediction on the one hand and data correction on the other. We show that they can be considered as specific discretizations of the data assimilation strategies proposed for continuous-time models in the sense of a *discretization-and-then-control* approach. This paradigm justifies the stability of these prediction-correction schemes, paving the way for convergence properties and justifying their popularity in practice.

Data assimilation emerged in the 1980s [75, 3] as the set of techniques for estimating the past, present, or future state of atmospheric models from partial observations and *a priori* knowledge. The range of applications then grew to include all environmental sciences and even new applications such as in biology and life sciences [39, 23], as well as in various engineering fields [83]. Data assimilation was also essentially the meeting point of control theory – more specifically, observation theory – and scientific computing, as the systems of interest were often represented by complex physical models with partial differential equations [60]. To strengthen the links between control and scientific computing, the question of numerical analysis developed for the underlying models should be generalized to include the study of discretization of data assimilation methods. Therefore, the question arose of the interaction of control theory and discretization, but here in the context of observation theory. Indeed, in data assimilation [85, 3, 58] methods were presented for continuous-time models, but also for discrete-time models, since they must ultimately be implemented in their discrete-time form. A fundamental question is then to use the proposed discrete-time forms as discretization of strategies of the continuous-time approaches.

As an illuminating example, let us briefly recall the historical – and formal – vision of the Kalman-Bucy estimator proposed in 1961 in [53] for systems described by the following dynamics

$$\begin{cases} \dot{z}(t) + A(t)z(t) = B(t)\nu(t), \quad t > 0, \\ z(0) = z_0, \end{cases} \tag{1}$$

where $\dot{z}$ stands for the time-derivative $\frac{\mathrm{d}}{\mathrm{d}t}z$, $A(t) \in C^0([0,T];\mathbb{M}_N(\mathbb{R}))$, $\nu \in \mathbb{R}^q$ and $B \in C^0([0,T];\mathbb{M}_{N,q}(\mathbb{R}))$. In [53], ν is presented formally as a "*white noise*" perturbation of 0 mean and covariance $Q \in \mathbb{M}_q(\mathbb{R})$ and ζ is an independent initial random variable of mean value $\hat{z}_0$ and covariance $\Pi_0 \in \mathbb{M}_N(\mathbb{R})$. Note that this historical presentation has led to a more justified mathematical reformulation of (1) in the context of Ordinary Differential Equations (ODE) perturbed by $L^2(0,T)$ errors in a deterministic context – see for instance [91] – or has led to a complete rewriting of (1) with stochastic processes z_t solution of the Stochastic Differential Equation (SDE)

$$\mathrm{d}z_t + A(t)z_t\mathrm{d}t = B(t)\mathrm{d}b_t^\nu, \tag{2}$$

where $\mathrm{d}b_t^\nu \in \mathbb{R}^q$ is a Wiener process of 0 mean and covariance $Q \in \mathbb{M}_q(\mathbb{R})$ and $B \in C^0(\mathbb{R}^+; \mathbb{M}_{N,q}(\mathbb{R}))$ – see for instance [58]. Moreover in the historical presentation [53], it is assumed that a physical system trajectory can be modeled by a realization $\check{z}$, solution of (1), albeit with unknown $\check{\zeta}$ and $t \mapsto \check{\nu}(t)$ and with observations y^δ of the form

$$t \mapsto y^\delta(t) = C(t)\check{z}(t) + \eta,$$

where again η is a "*white noise*" perturbation of 0 mean and covariance $W \in \mathbb{M}_q(\mathbb{R})$. An *estimation* $\hat{z}$ of $\check{z}$ is given by the so-called Kalman-Bucy estimator

$$\begin{cases} \dot{\hat{z}}(t) + A(t)\hat{z}(t) = \mathit{\Pi}(t)C(t)^\intercal W^{-1}(y^\delta(t) - C\hat{z}(t)), \quad t > 0, \\ \hat{z}(0) = \hat{z}_0 \end{cases} \tag{3}$$

where $\mathit{\Pi}(t)$, often called the *covariance operator*, is solution of the Riccati dynamics

$$\begin{cases} \dot{\mathit{\Pi}}(t) + A(t)\mathit{\Pi}(t) + \mathit{\Pi}(t)A(t)^\intercal + \mathit{\Pi}(t)C(t)^\intercal W^{-1}C(t)\mathit{\Pi}(t) - B(t)QB(t)^\intercal = 0, \quad t > 0, \\ \mathit{\Pi}(0) = \mathit{\Pi}_0 \end{cases} \tag{4}$$

By contrast, one year before in 1960, [54] proposed the so-called Kalman estimator for "time-discrete dynamics" of the form

$$z_{n+1} = \Phi_{n+1|n} z_n + B_{n+1}\nu_{n+1}, \quad n \in \mathbb{N}, \tag{5}$$

where $(z_n)_{n\geq 0}$ is a Markov chain, $\Phi_{n+1|n} \in \mathbb{M}_N(\mathbb{R})$ the transition operator, $B_{n+1} \in \mathbb{M}_{N,q}(\mathbb{R})$ a model error operator, and ν_n are i.i.d Gaussian variables $\mathcal{N}(0, Q_n)$ also independent of $z_0 \in \mathcal{N}(\hat{z}_0, \Pi_0)$. The available measurements sequence $(y_n^\delta)_{n\geq 0}$ is modeled a sample of random variables $(y_n)_{n\geq 0}$ linearly related to $(z_n)_{n\geq 0}$ by $y_n = C_n z_n + \eta_n$ where η_n are i.i.d Gaussian variables $\mathcal{N}(0, W_n)$. The estimator then reads

$$\begin{cases} \text{Initialization:} \\ \hat{z}_0^- = \hat{z}_0, \text{ and } \mathit{\Pi}_0^- = \mathit{\Pi}_0, \\ \text{Correction } (n \in \mathbb{N}): \\ G_n = \mathit{\Pi}_n^- C_n^\intercal (C_n \mathit{\Pi}_n^- C_n^\intercal + W_n)^{-1} \\ \hat{z}_n^+ = \hat{z}_n^- + G_n(y_n^\delta - C_n z_n^-) \\ \mathit{\Pi}_n^+ = \mathit{\Pi}_n^- - G_n C_n \mathit{\Pi}_n^- \\ \text{Prediction } (n \in \mathbb{N}): \\ \hat{z}_{n+1}^- = \Phi_{n+1|n}\hat{z}_n^+, \\ \mathit{\Pi}_{n+1}^- = \Phi_{n+1|n}\mathit{\Pi}_n^+\Phi_{n+1|n}^\intercal + B_{n+1}Q_{n+1}B_{n+1}^\intercal, \end{cases} \tag{6}$$

with a remarkable succession of model prediction and data correction. A notable result – that will here be reviewed for PDE formulation encountered in data assimilation – is that as the time discretization of (1) can lead to a convergent time-discrete system of the form (5), then (6) can be reinterpreted as a convergent time-discretization of (3)-(4). More generally, the ensemble of techniques developed for discrete-time models [85, 3] can in fact be viewed as a specific discretization of the methods developed for continuous-time models in the spirit of discretization-then-control approaches [96]. This strategy is shown to bring stability and convergence to the resulting time schemes, all of which take the form of a splitting strategy in which half a time step is devoted to forecasting – or predicting with – the model, the second half a time step is devoted to correcting – or updating – the models with the available observations.

Our goal, therefore, is to highlight the remarkable connection between continuous-time estimation and discrete-time estimation in general classes of dynamics, ranging from linear evolution equations in infinite-dimensional spaces to nonlinear formulations in finite dimension, from a deterministic point of view adapted to numerical analysis, albeit with a natural opening to problem formalization in a stochastic context. In the first section, we provide an overview of the Kalman estimator for parabolic PDE and its reduced-order variants, which were developed in the context of data assimilation to increase computational efficiency. In the second section, we present a set of estimators that are numerically better adapted to hyperbolic PDE than the Kalman estimators, but still represent a time discretization for prediction-correction. Finally, we will discuss classical filtering approaches and their discretization in the presence of nonlinear dynamics or nonlinear measurement procedures.

1 Least squares estimation and associated discretization for linear parabolic cases

1.1 The functional framework

In this section we restrict the presentation to dynamical systems modeled by linear partial differential equations of the parabolic type, more specifically first-order variational evolution equations [66]. The reason for this choice is twofold. First, this case is general enough to give a taste of the general linear-quadratic framework in data assimilation, encompassing in particular the finite-dimensional case while being compatible with a large class of infinite dimensional systems. Second, we will see that the regularization effect in parabolic problems allows to consider the discretization of a sequential strategy based on dynamic programming – leading to the famous Kalman filters – without being doomed by the classical curse of dimensionality of such methods for large dimensional systems, typically when refining the spatial discretization. On the contrary, conservative infinite dimensional systems of hyperbolic type will be covered in a second section.

Let $\mathcal{Z}$ be defined as a real-value state space, we introduce a subspace $\mathcal{V}$ with continuous injection in $\mathcal{Z}$ such that we can consider $\mathcal{V} \subset \mathcal{Z} \equiv \mathcal{Z}' \subset \mathcal{V}'$. We then consider an operator $A \in \mathcal{L}(\mathcal{V}, \mathcal{V}')$ assumed to be $\mathcal{V}$ - $\mathcal{Z}$ coercive, namely there exists $\rho > 0$ and $\lambda \in \mathbb{R}$ such that

$$\langle Av, v\rangle_{\mathcal{V}} + \lambda \|v\|^2_{\mathcal{Z}} \geq \rho \|v\|^2_{\mathcal{V}}. \tag{7}$$

where $\langle \cdot, \cdot \rangle$ denotes a duality bracket not to be confused with scalar products denoted $(\cdot, \cdot)$. Such an operator can be naturally extended to an operator $(A, \mathcal{D}(A))$ with value in $\mathcal{Z}$ and domain $\mathcal{D}(A) = \big\{v \in \mathcal{V} : \exists \beta \in \mathcal{Z} \text{ s.t. } \forall w \in \mathcal{V}, \langle Av, w\rangle_{\mathcal{V}} = (\beta, w)_{\mathcal{Z}}\big\}$. The unbounded operator $(A, \mathcal{D}(A))$ is maximal accretive and is the generator of an analytical semigroup $\Phi(t) = e^{-tA}$ see [14, Section 2.7].

We then consider the following dynamics

$$\begin{cases} \dot{z}(t) + Az(t) = g(t) + B\nu(t), \quad t \in (0, T), \\ z(0) = z_0. \end{cases} \tag{8}$$

parametrized by $z_0 \in \mathcal{Z}$, and $\nu \in L^2((0,T);\mathcal{U})$ where $\mathcal{U}$ is a Hilbert space and $B \in \mathcal{L}(\mathcal{U}, \mathcal{Z})$ and $g \in H^1((0,T);\mathcal{Z})$ is given. The case with time-varying operator $B(t)$ would be a simple extension. Classical results for evolution equations give the existence of a solution of (8) knowing z_0 and ν and denoted $z_{|z_0,\nu}$. First, using semigroup theory [78, 14], we have on the one hand if $z_0 \in \mathcal{Z}$ and $\nu \in L^2((0,T);\mathcal{U})$ then $z_{|z_0,\nu} \in C^0([0,T];\mathcal{Z})$ and on the other hand if $z_0 \in \mathcal{D}(A)$ and $\nu \in H^1((0,T);\mathcal{U})$ then $z_{|z_0,\nu} \in C^1([0,T];\mathcal{Z}) \cap C^0([0,T];\mathcal{D}(A))$. Moreover, using variational theory [66, 14], if $z_0 \in \mathcal{Z}$ and $\nu \in L^2((0,T);\mathcal{U})$ then $z_{|z_0,\nu} \in \mathcal{W}(0,T) = \{z \in L^2((0,T);\mathcal{V}), \dot{z} \in L^2((0,T);\mathcal{V}')\}$. We point out that existence results are more complex where B is unbounded and refer for example to [14, 57, 90] for such more general cases. Note finally that variational theory – and even mild evolution operator theory – can accommodate a time dependent operator $A(t)$ with additional refinements [87, 78].

We now assume that the dynamics (8) *accurately* models an observed physical system. The observed trajectory is modeled by a particular unknown realization $(\check{z}_0, \check{\nu})$ producing $\check{z} = z_{|\check{z}_0,\check{\nu}}$. The observations – also called measurements – are denoted by y^δ such that, in a deterministic approach, they belong to $L^2((0,T),\mathcal{Y})$ with $\mathcal{Y}$ a Hilbert space. Moreover, the measurement procedure is modeled via an observation operator $C \in \mathcal{L}(\mathcal{Z}, \mathcal{Y})$, such that $\check{y} : [0,T] \ni t \mapsto C\check{z}(t) \in \mathcal{Y}$ satisfies

$$\|y^\delta - \check{y}\|^2_{L^2((0,T),\mathcal{Y})} \leq \delta^2 T,$$

and δ quantifies the measurement error amplitude. The operator C is here assumed to be bounded to simplify the presentation but more general configurations where C is unbounded are also studied [14, 90]. The time-dependent case $C(t)$ is a simple extension.

Our objective is to estimate $\check{z}$ using y^δ despite the uncertainty $(\check{z}_0, \check{\nu})$. More precisely, we decompose $\check{z}_0 = \hat{z}_0 + \check{\zeta}$ where $\check{\zeta}$ is controlled in a certain space $\mathcal{V}_s \subset \mathcal{V}$ with $\|\check{\zeta}\|_{\mathcal{V}_s} \leq \alpha$. Identically, $\check{\nu}$ is also understood as a bounded perturbation of the right hand side g, with typically $\|\check{\nu}(t)\|^2_{L^2(0,T,\mathcal{U})} \leq \kappa^2 T$.

At this point we must alert the reader to the fact that our description is a purely deterministic vision of the data assimilation objectives. In this respect, the presentation is close to the well known review

[91], although we also cover infinite dimensional cases in the spirit of [65, 12]. One interest of such a deterministic vision is its compatibility with the classical paradigms of numerical analysis. We will later position this vision in terms of an alternative – and perhaps more common today – stochastic vision of data assimilation.

1.2 Data assimilation via optimal control

1.2.1 The variational method

In this deterministic vision of data assimilation, a first strategy is to estimate $\check{z}$ using an optimal control approach, namely by an estimation of $(\check{\zeta}, \check{\nu})$ from the minimizer of the functional $\mathscr{J}_T : \mathcal{V}_s \times L^2((0,T);\mathcal{U}) \to \mathbb{R}$ defined by

$$\mathscr{J}_T(\zeta,\nu) = \frac{1}{2} a_{\pi_0}(\zeta,\zeta) + \frac{1}{2}\int_0^T \left[\delta^{-2}\|y^\delta(s) - Cz_{|\zeta,\nu}(s)\|_{\mathcal{Y}}^2 + \kappa^{-2}\|\nu(s)\|_{\mathcal{U}}^2\right] \mathrm{d}s, \tag{9}$$

where $z_{|\zeta,\nu}$ is solution of (8) for an initial condition $z_0 = \hat{z}_0 + \zeta$ and a_{π_0} is a symmetric and coercive bilinear form on $\mathcal{V}_s$. For further computation, we introduce $\mathit{\Pi}_0 \in \mathcal{L}(\mathcal{Z})$ the well-defined inverse of the operator $\mathit{\Pi}_0^{-1} \in \mathcal{L}(\mathcal{V}_s, \mathcal{V}_s')$ such that

$$\forall (\zeta,\xi) \in \mathcal{V}_s^2, \quad a_{\pi_0}(\zeta,\xi) = \langle \mathit{\Pi}_0^{-1}\zeta, \xi\rangle_{\mathcal{V}_s',\mathcal{V}_s}.$$

In this generalized Tikhonov criterion, the uncertainty level of $\check{\zeta}$ is controlled by assuming $\langle \mathit{\Pi}_0^{-1}\check{\zeta}, \check{\zeta}\rangle_{\mathcal{V}_s',\mathcal{V}_s} \leq 1$. The quadratic functional $\mathscr{J}_T$ is strictly convex hence enforcing the existence of a unique minimizer $(\bar{\zeta}_T, \bar{\nu}_T)$. The minimization of $\mathscr{J}_T$ is performed under the dynamics (8) constraint which can be enforced by considering the saddle point $(\bar{z}_T, \bar{q}_T, \bar{\nu}_T)$ of the Lagrangian $\mathscr{L}_T : \mathcal{W}(0,T) \times L^2((0,T);\mathcal{V}) \times L^2((0,T);\mathcal{U}) \to \mathbb{R}$ defined by

$$\begin{aligned}\mathscr{L}_T(z,q,\nu) = \frac{1}{2}\langle \mathit{\Pi}_0^{-1}(z(0) - \hat{z}_0), (z(0) - \hat{z}_0)\rangle_{\mathcal{V}_s',\mathcal{V}_s} \\ + \frac{1}{2}\int_0^T \left[\delta^{-2}\|y^\delta(s) - Cz(s)\|_{\mathcal{Y}}^2 + \kappa^{-2}\|\nu(s)\|_{\mathcal{U}}^2\right] \mathrm{d}s \\ + \int_0^T \langle q(s), \dot{z}(s) + Az(s) - g(s) - B\nu(s)\rangle_{\mathcal{V},\mathcal{V}'} \,\mathrm{d}s. \end{aligned} \tag{10}$$

After justifying that the optimal solution $\bar{q}_T$ initially defined in $L^2((0,T);\mathcal{V})$ in fact belongs to $\mathcal{W}(0,T)$ where $\dot{\bar{q}}_T \in L^2((0,T);\mathcal{V}')$, the optimality conditions read

$$\begin{cases} \dot{\bar{z}}_T + A\bar{z}_T = g + BQB^*\bar{q}_T, & \text{in } (0,T), \\ \dot{\bar{q}}_T - A^*\bar{q}_T = -C^*R(y^\delta - C\bar{z}_T), & \text{in } (0,T), \\ \bar{z}_T(0) = \hat{z}_0 + \mathit{\Pi}_0\bar{q}_T(0), \\ \bar{q}_T(T) = 0, \end{cases} \tag{11}$$

where, here, $Q = \kappa^2 \mathrm{Id}_{\mathcal{U}}$ and $R = \delta^{-2}\mathrm{Id}_{\mathcal{Y}}$. In other words, the minimizer of $\mathscr{J}_T$ is characterized by

$$\bar{\zeta}_T = \mathit{\Pi}_0\bar{q}_T(0), \quad \bar{\nu}_T = QB^*\bar{q}_T. \tag{12}$$

The system (11) is called a two-ends problem in control theory [65] and in observation theory [12], as it differs from a Cauchy problem since the adjoint dynamics is defined backward in time from a terminal condition. Solving the estimation problem from (11) is often called a 4D-Var strategy in data assimilation [61] as it corresponds to a variational approach often applied to 3D complex systems modeled by PDEs.

1.2.2 The sequential approach

As opposed to variational approaches, sequential approaches do not rely on an explicit optimization algorithm but rather on modifying the original dynamics (8) to account for the available measurements using a feedback control G:

$$\begin{cases} \dot{\hat{z}} + A\hat{z} = g + G(y^\delta - C\hat{z}), & \text{in } (0,T), \\ \hat{z}(0) = \hat{z}_0. \end{cases} \tag{13}$$

It should be proved that the resulting system $\hat{z}$ converges asymptotically in time - i.e., T is expected to tend toward $+\infty$ – to the actual trajectory and is stable with respect to measurement errors. If this is the case, it is called *an observer* according to the general definition of [55]. Following a dynamic programming principle approach [12], an observer can be designed from the optimal trajectory using the following definition.

Definition 1.1. *The optimal sequential estimator is defined by*

$$\forall t \in \mathbb{R}^+, \quad \hat{z}(t) = \bar{z}_t(t). \tag{14}$$

Indeed, as the criterion gives increasing importance to the available measurements, the above definition leads to the definition of an effective observer. Note again that the previously introduced time T will be considered as going to infinity for the observer to satisfy its objective.

To characterize this optimal observer $\hat{z}$, let us first introduce the Riccati dynamics

$$\begin{cases} \dot{\Pi} + A\Pi + \Pi A^* + \Pi C^* R C \Pi - BQB^* = 0, & \text{in } (0,T), \\ \Pi(0) = \Pi_0. \end{cases} \tag{15}$$

The conditions of existence of a solution are well known, see for instance [14, Section IV-1]. Assuming that Π_0 belongs to the set $\mathcal{S}^+(\mathcal{Z})$ of bounded symmetric and positive linear operators, then $\Pi \in C^0([0,T], \mathcal{S}^+(\mathcal{Z}))$, Moreover if Π_0 belongs to the space $\mathcal{D}(\mathcal{Z})$ of operators $\Upsilon \in \mathcal{S}^+(\mathcal{Z})$ such that there exists c_{st} so that

$$\forall (z_1, z_2) \in \mathcal{Z}^2, \quad (A^* z_1, \Upsilon z_2)_{\mathcal{Z}} + (\Upsilon z_1, A^* z_2)_{\mathcal{Z}} \leq c_{\text{st}} \|z\|_1 \|z\|_2,$$

then $\Pi \in C^1([0,T], \mathcal{S}^+(\mathcal{Z})) \cap C^0([0,T], \mathcal{D}(\mathcal{Z}))$. Finally in the variational setting, [41] proved additionally that for all time $t \in [0,T]$, $\Pi(t) \in \mathcal{L}(\mathcal{V}', \mathcal{V})$. From the Riccati solution, we define the Kalman gain $G = \Pi C^*$ leading to the Kalman estimator dynamics – generalizing (3) to parabolic PDEs –

$$\begin{cases} \dot{\hat{z}} + A\hat{z} = g + \Pi C^* R(y^\delta - C\hat{z}), & \text{in } (0,T), \\ \hat{z}(0) = \hat{z}_0. \end{cases} \tag{16}$$

The Kalman estimator dynamics admits one and only one solution with the same regularity as $\check{z}$ because $A + \Pi C^* R C$ is a bounded perturbation of A [14]. Then, we have the remarkable identity, called the *decoupling principle* [65, 12].

Theorem 1.1 (Decoupling principle)**.** *Considering $(\bar{z}_T, \bar{q}_T)$ solution of the two-ends problem (11) and $\hat{z}$ and Π solutions of the Cauchy problems (16) and (15), then we have*

$$\forall t \in [0,T], \quad \bar{z}_T(t) = \hat{z}(t) + \Pi(t)\bar{q}_T(t). \tag{17}$$

We also refer to [5] for a detailed proof adapted to the parabolic context when $\Pi_0 \in \mathcal{D}(\mathcal{Z})$. The idea is, as for bounded operators, to study the dynamics of $\eta = \bar{z}_T(t) - \hat{z}(t) - \Pi(t)\bar{q}_T(t)$ and to prove that η remains null over time. From (17), we directly infer that the Kalman filter is the optimal sequential estimator in the sense of Definition 1.1 since, at final time, $\bar{z}_T(T) = \hat{z}(T)$. This result provides a reinterpretation of the Kalman filter in a deterministic context [12, 31].

1.3 Discretization of the variational strategy

In the classic dilemma of control-then-discretization versus discretization-then-control [96], we emphasize, here in data assimilation, the interest in a discretization-then-control strategy, that can be shown to bring fundamental stability then convergence properties.

Let us start with the discretization of the model. Let $0 = t_0 < ... < t_k = k\tau < ... < t_n = T$ be a discretization of the time interval $[0,T]$ with fixed timestep τ to simplify the presentation. The time-scheme leads to a discrete-time dynamics of the general form

$$\begin{cases} z_{k+1}^{h,\tau} = \Phi_1^{h,\tau} z_k^{h,\tau} + g_{k+1}^{h,\tau} + B^{h,\tau} \nu_{k+1}^{\tau}, & 0 \leq k \leq n-1, \\ z_0^h = \hat{z}_0^h + \zeta^h, \end{cases} \tag{18}$$

where $\Phi_1^{h,\tau}$ is an approximation of the continuous semigroup $\Phi(\tau)$ and $\tau^{-1}B^{h,\tau}$ a consistent approximation of B, and $\tau^{-1}g_k^{h,\tau}$ a consistent approximation of $g(t)$.

To fix the idea, it can be useful to keep the following simple example of a Galerkin spatial discretization combined with an implicit time scheme. We introduce the finite element space $\mathcal{V}^h$, and the orthogonal projection from $\mathcal{Z}$ to $\mathcal{Z}^h$ is denoted by P^h. A discretized operator is defined by $(A^h u^h, v^h)_{\mathcal{Z}} = \langle Au^h, v^h\rangle_{\mathcal{V}}$ for all $(u^h, v^h) \in \mathcal{Z}^h$. Moreover, we introduce $g^h = P^h g$ and $B^h = P^h B$. Then, with a Backward-Euler time scheme, we have

$$\Phi_1^{h,\tau} = (\mathrm{Id}_{\mathcal{Z}^h} + \tau A^h)^{-1}, \text{ whereas } \begin{cases} B^{h,\tau} = \tau(\mathrm{Id}_{\mathcal{Z}^h} + \tau A^h)^{-1}B^h = \tau\Phi_1^{h,\tau}B^h, \\ g_{k+1}^{h,\tau} = \tau(\mathrm{Id}_{\mathcal{Z}^h} + \tau A^h)^{-1}g^h(t^{k+1}) = \tau\Phi_1^{h,\tau}g^h(t^{k+1}). \end{cases}$$

Note that we typically have the following approximation property for an analytic semigroup using Trotter-Kato results [44, Theorem 2.7]:

$$\forall z_0 \in \mathcal{Z}, \quad \|\Phi(t_k)z_0 - \Phi_k^{h,\tau}z_0\|_{\mathcal{Z}} \le c_{\mathrm{st}}\frac{h^2+\tau}{t_k}\|z_0\|_{\mathcal{Z}}, \quad 0 \le k \le n. \tag{19}$$

Note also that for time-dependent operators this can be generalized by introducing the transition operator from time step k to $k+1$, $\Phi_{k+1|k}^{h,\tau} = (\mathrm{Id}_{\mathcal{Z}^h} + \tau A^h(t_{k+1}))^{-1}$. Finally we define $C^h = CP^{h*}$ which implies that, here, we neglect the – yet fundamental – effect of the interplay between the spatial sampling of the measurement and the model discretization.

As in the continuous setting, the solution of (18) depends on ζ^h and $(\nu_k^\tau)_{1\le k\le n}$ and, to use a more compact notation, we denote by $\nu_{|n}^\tau = (\nu_k^\tau)_{1\le k\le n}$. Then, we discretize the criterion (9) using the simplest quadrature rule to approximate the integrals. We introduce $\Pi_0^h = P^h\Pi_0 P^{h*}$ and a discrete functional given, for all $n \ge 1$, by

$$\mathscr{J}_n^{h,\tau-}(\zeta^h, \nu_{|n}^\tau) = \frac{1}{2}\big(\zeta^h, (\Pi_0^h)^{-1}\zeta^h\big)_{\mathcal{Z}} + \frac{1}{2}\sum_{k=1}^{n}\kappa^{-2}\tau\|\nu_k^\tau\|_{\mathcal{U}}^2 + \frac{1}{2}\sum_{k=0}^{n-1}\tau\delta^{-2}\|y_k^\delta - C^h z_{k|\zeta^h,\nu_{|n}^\tau}^{h,\tau}\|_{\mathcal{Y}}^2. \tag{20}$$

This discrete functional is complemented by a second functional given by, for $n \ge 1$,

$$\mathscr{J}_n^{h,\tau+}(\zeta^h, \nu_{|n}^\tau) = \frac{1}{2}\big(\zeta^h, (\Pi_0^h)^{-1}\zeta^h\big)_{\mathcal{Z}} + \frac{1}{2}\sum_{k=1}^{n}\kappa^{-2}\tau\|\nu_k^\tau\|_{\mathcal{U}}^2 + \frac{1}{2}\sum_{k=0}^{n}\tau\delta^{-2}\|y_k^\delta - C^h z_{k|\zeta^h,\nu_{|n}^\tau}^{h,\tau}\|_{\mathcal{Y}}^2. \tag{21}$$

Intuitively, $\mathscr{J}_n^{h,\tau-}$ is the criterion that includes one last step of model prediction whereas $\mathscr{J}_n^{h,\tau+}$ includes the last available observations. These two functionals are consistent with respect to τ and we will see later the interplay between them.

Let us now identify the minimizer $(\bar\zeta_{|n}^{h,\tau-}, \bar\nu_{|n}^{h,\tau-})$ of $\mathscr{J}_n^{h,\tau-}$, the procedure being similar when considering $\mathscr{J}_n^{h,\tau+}$. We recall that we are in the case of the minimization of a strictly convex quadratic functional in finite dimension under a linear discrete-time dynamics constraint, hence we have one and only one minimizer. We then introduce the following Lagrangian

$$\begin{aligned}\mathscr{L}_n^{h,\tau-}(z_{|n}^{h,\tau}, q_{|n}^{h,\tau}, \nu_{|n}^\tau) = {}& \frac{1}{2}\big((z_{0|n}^{h,\tau} - \hat z_0^h), (\Pi_0^h)^{-1}(z_{0|n}^{h,\tau} - \hat z_0^h)\big)_{\mathcal{Z}} + \frac{1}{2}\sum_{k=1}^{n}\kappa^{-2}\tau\|\nu_k^\tau\|_{\mathcal{U}}^2 \\ & + \frac{1}{2}\sum_{k=0}^{n-1}\tau\delta^{-2}\|y_k^\delta - C^h z_{k|n}^{h,\tau}\|_{\mathcal{Y}}^2 \\ & + \sum_{k=0}^{n-1}\big(q_{k+1|n}^{h,\tau}, z_{k+1}^{h,\tau} - \Phi_1^{h,\tau}z_k^{h,\tau} - g_{k+1}^{h,\tau} - B^{h,\tau}\nu_{k+1}^\tau\big)_{\mathcal{Z}},\end{aligned} \tag{22}$$

and minimizing $\mathscr{J}_n^{h,\tau-}$ in the finite dimensional Hilbert space $\mathcal{Z}^h \times \mathcal{U}^n$ under the constraint of the discrete-time dynamics (18), is equivalent to finding the saddle point of $\mathscr{L}_n^{h,\tau-}$. The derivative of $\mathscr{L}_n^{h,\tau-}$ gives for $0 < k < n$,

$$\begin{aligned}\forall \xi \in \mathcal{Z}^h, \quad & \big\langle \mathrm{D}_{z_{k|n}^{h,\tau}}\mathscr{L}_n^{h,\tau-}(z_{|n}^{h,\tau}, q_{|n}^{h,\tau}, \nu_{|n}^\tau), \xi\big\rangle_{\mathcal{Z}^{h\prime},\mathcal{Z}^h} \\ & \qquad = -\tau\delta^{-2}\big(y_k^\delta - C^h z_{k|n}^{h,\tau}, C^h\xi\big)_{\mathcal{Y}} + \big(q_{k|n}^{h,\tau}, \xi\big)_{\mathcal{Z}} - \big(q_{k+1|n}^{h,\tau}, \Phi_1^{h,\tau}\xi\big)_{\mathcal{Z}},\end{aligned}$$

while for $k = n$,

$$\forall \xi \in \mathcal{Z}^h, \quad \big\langle \mathrm{D}_{z^{h,\tau}_{n|n}} \mathscr{L}^{h,\tau-}_n(z^{h,\tau}_{|n}, q^{h,\tau}_{|n}, \nu^{\tau}_{|n}), \xi \big\rangle_{\mathcal{Z}^{h\prime},\mathcal{Z}^h} = \big(q^{h,\tau}_{n|n}, \xi\big)_{\mathcal{Z}},$$

and for $k = 0$,

$$\begin{aligned}\forall \xi \in \mathcal{Z}^h, \quad & \big\langle \mathrm{D}_{z^{h,\tau}_{0|n}} \mathscr{L}^{h,\tau-}_n(z^{h,\tau}_{|n}, q^{h,\tau}_{|n}, \nu^{\tau}_{|n}), \xi \big\rangle_{\mathcal{Z}^{h\prime},\mathcal{Z}^h} \\ & = \big((z^{h,\tau}_{0|n} - \hat{z}^h_0), (\Pi^h_0)^{-1}\xi\big)_{\mathcal{Z}} - \tau\delta^{-2}\big(y^\delta_0 - C^h z^{h,\tau}_{0|n}, C^h\xi\big)_{\mathcal{Y}} - \big(q^{h,\tau}_{1|n}, \Phi^{h,\tau}_1\xi\big)_{\mathcal{Z}}, \\ & = \big((z^{h,\tau}_{0|n} - \hat{z}^h_0), (\Pi^h_0)^{-1}\xi\big)_{\mathcal{Z}} - \big(q^{h,\tau}_{0|n}, \xi\big)_{\mathcal{Z}},\end{aligned}$$

as soon as we define $q^{h,\tau}_{0|n} = \Phi^{h,\tau*}_1 q^{h,\tau}_{1|n} + \delta^{-2}\tau C^{h*}\big(y^\delta_0 - C^h z^{h,\tau}_{0|n}\big)$. Finally, we have for $k \geq 0$,

$$\forall \mu \in \mathcal{U}, \quad \big\langle \mathrm{D}_{\nu^{h,\tau}_{k+1|n}} \mathscr{L}^{h,\tau-}_n(z^{h,\tau}_{|n}, q^{h,\tau}_{|n}, \nu^{\tau}_{|n}), \mu \big\rangle_{\mathcal{U}',\mathcal{U}} = \kappa^{-2}\tau\big(\nu^{h,\tau}_{k+1|n}, \mu\big)_{\mathcal{U}} - \big(B^{h,\tau*} q^{h,\tau}_{k+1|n}, \mu\big)_{\mathcal{U}},$$

and also

$$\forall \lambda \in \mathcal{Z}^h, \quad \big\langle \mathrm{D}_{q^{h,\tau}_{k+1|n}} \mathscr{L}^{h,\tau-}_n(z^{h,\tau}_{|n}, q^{h,\tau}_{|n}, \nu^{\tau}_{|n}), \lambda \big\rangle_{\mathcal{Z}^{h\prime},\mathcal{Z}^h} = \big(z^{h,\tau}_{k+1} - \Phi^{h,\tau}_1 z^{h,\tau}_k - B^{h,\tau}\nu^{\tau}_{k+1}, \lambda\big)_{\mathcal{Z}}.$$

We end up by finding that the minimizer $(\bar{\zeta}^{h,\tau-}_{|n}, \bar{\nu}^{h,\tau-}_{|n})$ of $\mathscr{J}^{h,\tau-}_n$ satisfies

$$\bar{\zeta}^{h,\tau-}_{|n} = \Pi^h_0 \bar{q}^{h,\tau-}_{0|n}, \qquad \bar{\nu}^{h,\tau-}_{k|n} = Q^\tau B^{h,\tau*} \bar{q}^{h,\tau-}_{k|n}, \ 1 \leq k \leq n,$$

where $Q^\tau = \kappa^2\tau^{-1}\mathrm{Id}_{\mathcal{U}}$, $R^\tau = \delta^{-2}\tau\mathrm{Id}_{\mathcal{Y}}$ and

$$\begin{cases} \bar{z}^{h,\tau}_{k+1|n} = \Phi^{h,\tau}_1 \bar{z}^{h,\tau}_{k|n} + g^{h,\tau}_{k+1} + B^{h,\tau} Q^\tau B^{h,\tau*} \bar{q}^{h,\tau}_{k|n}, & 0 \leq k \leq n-1, \quad (23a) \\ \bar{z}^{h,\tau}_{0|n} = \hat{z}^h_0 + \Pi^h_0 \bar{q}^{h,\tau}_{0|n}, & (23b) \\ \bar{q}^{h,\tau}_{k|n} = \Phi^{h,\tau*}_1 \bar{q}^{h,\tau}_{k+1|n} + C^{h*} R^\tau\big(y^\delta_k - C^h \bar{z}^{h,\tau}_{k|n}\big), & 0 \leq k \leq n-1, \quad (23c) \\ \bar{q}^{h,\tau}_{n|n} = 0. & (23d) \end{cases}$$

The resulting system (23) benefits from very useful properties. The existence of a solution follows from the existence of a minimizer for the discrete criterion. Moreover, convergence properties to the continuous minimizer could be studied from Γ-convergence results [19], even if, to the author's knowledge the question was not specifically addressed with the proposed discretization.

In practice, as the two-ends problem (23) implies to introduce space-time discretization or – more commonly – an iterative procedure based on a gradient descent approach. Using the adjoint equation

$$\begin{cases} q^{h,\tau}_{k|n} = \Phi^{h,\tau*}_1 q^{h,\tau}_{k+1|n} + C^{h*} R^\tau\big(y^\delta_k - C^h z^{h,\tau}_{k|n}\big), & 0 \leq k \leq n-1, \\ q^{h,\tau}_{n|n} = 0, \end{cases}$$

the derivative of $\mathscr{J}^{h,\tau-}_n$ is given by

$$\forall \xi \in \mathcal{Z}^h, \quad \big\langle \mathrm{D}_{\zeta^h} \mathscr{J}^{h,\tau-}_n(\zeta^h, (\nu^{\tau}_{|n}), \xi \big\rangle_{\mathcal{Z}^{h\prime},\mathcal{Z}^h} = \big(z^{h,\tau}_{0|n}, (\Pi^h_0)^{-1}\xi\big)_{\mathcal{Z}} - \big(q^{h,\tau}_{0|n}, \xi\big)_{\mathcal{Z}},$$

and

$$\forall \mu \in \mathcal{U} \quad \big\langle \mathrm{D}_{\nu^{\tau}_{k+1}} \mathscr{J}^{h,\tau-}_n(\zeta^h, \nu^{\tau}_{|n}), \mu \big\rangle_{\mathcal{U}',\mathcal{U}} = \big(\nu^{h,\tau}_{k+1|n}, \mu\big)_{\mathcal{U}} - \big(B^{h,\tau*} q^{h,\tau}_{k+1|n}, \mu\big)_{\mathcal{U}},$$

giving the gradient from Riesz' representation theorem, for instance with respect to the norm induced by $(\Pi^h_0)^{-1}$. Then, various optimisation algorithms are used in practice to compute the limit system (23). We can think of a fixed step descent or a conjugate gradient descent, a Broyden–Fletcher–Goldfarb–Shanno algorithm, a Gauss-Newton or even a Newton-Raphson method [69, 25].

We want to finally underline that there is nowadays an effort to combine the regularization parameters choices in the initial least square formulation with the discretization parameters leading to very elegant strategies inspired from stabilization of mixed formulation theory. We refer to the recent work of [16] to understand how we can deal with the interplay of discretization and regularization in order to go beyond the current presentation.

1.4 The discrete time Kalman filter as a time discretization of the Kalman-Bucy filter

In the previous section, we presented the discretized version of the $4D$-Var approach. We can now mimic the result of optimal sequential estimation in continuous time to propose optimal sequential estimation in discrete time. We will see that this leads to the discrete-time Kalman filter equation, since the continuous-time Kalman filters were the dynamic programming version of the optimal control problem solved by the adjoint formulation in the $4D$-Var approach.

Indeed, the discrete-time Kalman filter reads

$$\begin{cases} \hat{z}_0^{h,\tau-} = \hat{z}_0^h, & (24a) \\ \hat{z}_n^{h,\tau+} = \hat{z}_n^{h,\tau-} + \Pi_n^{h,\tau+} C^{h*} R^\tau \left(y_n^\delta - C^h \hat{z}_n^{h,\tau-} \right), \quad n \in \mathbb{N}, & (24b) \\ \hat{z}_{n+1}^{h,\tau-} = \Phi_1^{h,\tau} \hat{z}_n^{h,\tau+} + g_{n+1}^{h,\tau}, \quad n \in \mathbb{N}, & (24c) \end{cases}$$

with a varying n following the observer definition, while

$$\begin{cases} \Pi_0^{h,\tau-} = \Pi_0^h, & (25a) \\ \Pi_n^{h,\tau+} = \left[(\Pi_n^{h,\tau-})^{-1} + C^{h*} R^\tau C^{h,\tau} \right]^{-1}, \quad n \in \mathbb{N}, & (25b) \\ \Pi_{n+1}^{h,\tau-} = \Phi_1^{h,\tau} \Pi_n^{h,\tau+} \Phi_1^{h,\tau*} + B^{h,\tau} Q^\tau B^{h,\tau*}, \quad n \in \mathbb{N}. & (25c) \end{cases}$$

Note that, here, $\Phi_1^{h,\tau}$ is invertible, hence $\Pi_{n+1}^{h,\tau-}$ is positive definite if $\Pi_n^{h,\tau+}$ is. This recursively ensures that $\Pi_n^{h,\tau-}$ is well defined and positive definite for all n, hence the discrete-time Kalman filter is well defined. The Kalman gain is given by $G_n^{h,\tau} = \Pi_n^{h,\tau+} C^{h*} R_n^\tau$ which can be proved [85] to be equivalently computed from

$$G_n^{h,\tau} = \Pi_n^{h,\tau-} C^{h*} \left(C^h \Pi_n^{h,\tau-} C^{h*} + W_n^\tau \right)^{-1}, \tag{26}$$

where $W_n^\tau = (R_n^\tau)^{-1}$, while

$$\Pi_n^{h,\tau+} = \Pi_n^{h,\tau-} - G_n^{h,\tau} \left[C^h \Pi_n^{h,\tau-} C^{h*} + W_n^{h,\tau} \right] G_n^{h,\tau*}, \tag{27}$$

identities that can also be used in more general configurations when $\Pi_n^{h,\tau-}$ or $\Pi_n^{h,\tau+}$ are not invertible [85].

From the definition of the discrete-time Kalman filter equations (24) and the corresponding discrete-time Riccati equation, one can prove a discrete-time counterpart of the fundamental identity (17).

Theorem 1.2. *Considering $(\bar{z}_{k|n}^{h,\tau}, \bar{q}_{k|n}^{h,\tau})_{1 \le k \le n}$ the solution of the two-ends problem (23), and $(\hat{z}_k^{h,\tau-})_{1 \le k \le n}$ and $(\Pi_k^{h,\tau-})_{1 \le k \le n}$ computed sequentially with (24) and (25), we have*

$$\bar{z}_{k|n}^{h,\tau} = \hat{z}_k^{h,\tau-} + \Pi_k^{h,\tau-} \bar{q}_{k|n}^{h,\tau}, \quad 0 \le k \le n. \tag{28}$$

This identity is easily obtained by induction, see [5] for details. Therefore, we directly obtain a characterization of the discrete-time Kalman filter in a deterministic setting as the discrete-time optimal filter. The advantage of relying on properties of dynamic programming at the discrete-time level is stability, since one easily has the existence of a minimizer for the quadratic functionals (20) and (21). Then, the convergence of the discrete-time Kalman filter through the continuous-time Kalman filter can be studied. Typically, if we assume that Π_0 is a bounded symmetric Hilbert-Schmidt operator, [18, Theorem 5.1] proves that Π remains a Hilbert-Schmidt operator over time, see also [89, 32, 17]. Moreover, [5] obtains the following convergence result in the space $\mathcal{I}_2(\mathcal{Z})$ of Hilbert-Schmidt operators, endowed with its natural norm

$$\|\Pi\|_2 = \sqrt{\mathrm{tr}(\Pi^* \Pi)} = \Big(\sum_{n \ge 0} (\Pi e_n, \Pi e_n)_{\mathcal{Z}} \Big)^{\frac{1}{2}},$$

with $(e_n)_{n \in \mathbb{N}}$ any orthonormal Hilbert basis of $\mathcal{Z}$.

Theorem 1.3. *Assuming $\Pi_0 \in \mathcal{S}^+(\mathcal{Z}) \cap \mathcal{I}_2(\mathcal{Z})$, then the solution of the Riccati dynamics (15) $\Pi \in C^0([0,T], I_2)$. Considering $(\Pi_k^{h,\tau-})_{1\leq k\leq n}$ the solution of the discrete-time scheme (25) associated with the time-grid $0 = t_0 < ... < t_k = k\tau < ... < t_n = T$, and P^h the projection from $\mathcal{V}$ to $\mathcal{V}_h$, we have*

$$\sup_{k\in[0,n]} \|\Pi_k^{h,\tau-} - P^h \Pi(t_k) P^{h*}\|_2 \xrightarrow[h,\tau\to 0]{} 0.$$

The proof is based on the original result of [46] – see also [82] – obtained for spatial discretization only, which [5] extends to a time-discretization based on the discrete-time Kalman filter. Convergence for the Riccati operator then easily implies convergence for the estimators.

1.5 Link with stochastic filtering

To conclude this overview of discrete-time Kalman filters, we recall that this presentation differs from the classical framework of stochastic filtering [20, 92] where (18) should be understood as the dynamics of a discrete-time Markov chain in a finite-dimensional space $\mathcal{Z}_h$ with $(\nu_k^\tau)_{k\geq 0}$ independent random variables $\mathcal{N}(0, Q^{h,\tau})$ and ζ an independent random variable $\mathcal{N}(0, \Pi_0^h)$. This linear Gaussian state-space model is partially observed through the process $y_k^{h,\tau} = C^h z_k + \eta_k^{h,\tau}$ where $(\eta_k^{h,\tau})_{k\geq 0}$ are independent random variables $\mathcal{N}(0, W^\tau)$ from which $(y_k^\delta)_{k\geq 0}$ is a random sample.

Then it is classically shown that the Kalman estimator is nothing but the conditional expectation knowing the measurements, and the Riccati solution is an estimation error covariance [20, 92, 13], viz.

$$\begin{cases} \text{Correction } (n \in \mathbb{N}): \\ \hat{z}_n^{h,\tau+} = \mathbb{E}(z_n^{h,\tau} | y_0^{h,\tau} = y_0^\delta, \ldots, y_n^{h,\tau} = y_n^\delta), \\ \Pi_n^{h,\tau+} = \mathrm{Cov}(z_n^{h,\tau} - \hat{z}_n^{h,\tau+}), \\ \text{Prediction } (n \in \mathbb{N}): \\ \hat{z}_{n+1}^{h,\tau-} = \mathbb{E}(z_{n+1}^{h,\tau} | y_0^{h,\tau} = y_0^\delta, \ldots, y_n^{h,\tau} = y_n^\delta), \\ \Pi_{n+1}^{h,\tau-} = \mathrm{Cov}(z_{n+1}^{h,\tau} - \hat{z}_{n+1}^{h,\tau-}). \end{cases}$$

Of note, the scaling in the operator $W^\tau = \tau^{-1} W$ and $Q^{h,\tau} = \kappa^{-1}\tau^{-1}\mathrm{Id}_{\mathcal{U}}$ are also interpretable in a stochastic context, since they tend to represent continuous-time white noises [12]. Moreover in this stochastic context, the convergence of the time discretization by a continuous-time version was also established as a connection between the original Kalman filter of [54] – the discrete-time Kalman filter – and the Kalman-Bucy filter [53] – the continuous-time Kalman filter. In finite dimensional spaces this has been studied, for example, in [84]. In infinite dimensional systems, the problem is more complicated since we need to introduce stochastic systems in infinite dimensional systems and associated input-output noises. Such systems have been the subject of intensive study, see for example [40, 12, 30]. Recent results [1, 2] have proved convergence results for dynamics in infinite dimensional Hilbert spaces with finite dimensional observation spaces, where the convergence rate depends on the regularity assumption and the semigroup analyticity.

Remarkably, the stochastic formulation in the linear Gaussian context does not differ from the deterministic formulation with optimal control. However, the stochastic framework will bring in the last section some additional flavor when considering nonlinear dynamics.

1.6 Alternative strategies for large dimensional discretized systems

The use of a Kalman filter for infinite dimensional systems leads to intractable computations after spatial discretization with the refinement of the spatial grid. Namely when discretized, the Riccati/covariance operator becomes a matrix with a dense pattern, hence limiting its use. To circumvent such a computational burden, several strategies have been considered in the literature: reducing the dimension of the initial model, reducing the dimension of the uncertainty space, or justified regularity properties for the covariance to adjust its discretization.

As for the first approach, this has led to the use of widely developed reduction methods that allow finite-dimensional spatial discretizations of reduced dimension [10]. For the linear problem presented in this section, a spectral discretization will strongly limit the dimension of the discretized system space

while preserving a good approximation rate. Of course, more specific reduction methods have been developed when dealing with nonlinear dynamics [71] in conjunction with data assimilation [70], especially when dealing with sequential estimation [24, 76]. The major advantage of model reduction is that it is directly compatible with a stable time-discretization of the data assimilation strategies. In particular, the Kalman filter is stable and converges when the model reduction converges.

An alternative to model reduction is covariance reduction where the covariance is projected into an uncertainty subspace. Low rank approximation of covariances is an important topic in particular often used in optimal control with differential Riccati equations, see for instance [11, 9] and references therein. We here present a strategy adapted to estimation and compatible with the prediction-correction strategy. In the absence of modeling error, namely $B = 0$, the criterion to be minimized becomes

$$\min_{\zeta_r \in \mathcal{V}_r} \Big\{ \mathscr{J}(\zeta_r) = \frac{\alpha^{-2}}{2} \|\zeta_r\|^2_{\mathcal{Z}} + \int_0^T \frac{\delta^{-2}}{2} \|y^\delta - C z_{|P_r^* \zeta_r}\|^2_{\mathcal{Y}} \, \mathrm{d}t \Big\},$$

where P_r is a projector into a space $\mathcal{V}_r$ of small finite dimension r and $z_{|P_r^*\zeta_r}$ is solution of (8) for an initial condition $z_0 = \hat{z}_0 + P_r^* \zeta_r$. Note that as the dimension remains finite in r, we can consider a classical regularization norm $\alpha^{-2}\|\cdot\|^2_{\mathcal{Z}}$ as all norms are equivalent in $\mathcal{V}_r$. The resulting Kalman observer can still be defined from the Riccati dynamics

$$\begin{cases} \dot{\Pi} + A\Pi + \Pi A^* + \Pi C^* R C \Pi = 0, \quad t > 0, \\ \Pi(0) = \alpha^{-2} P_r^* P_r. \end{cases} \tag{29}$$

But moreover, the Riccati solution can be deduced from a reduced covariance operator Λ, solution of a Riccati dynamics in the reduced space $\mathcal{V}_r$, viz

$$\begin{cases} \dot{\Lambda} + \Lambda P_r e^{-tA^*} C^* R C e^{-tA} P_r^* \Lambda = 0, \quad t > 0, \\ \Lambda(0) = \alpha^{-2} \mathrm{Id}_{\mathcal{V}_r}. \end{cases} \tag{30}$$

Theorem 1.4. *Given Λ a mild solution of (30) and $L : t \mapsto e^{-tA} P_r^* \in C^0([0,T], \mathcal{L}(\mathcal{V}_r, \mathcal{Z}))$, the mild solution of (29) is given by*

$$\forall t \geq 0, \quad \Pi(t) = L(t)\Lambda(t)L(t)^*.$$

Proof. For all $t \geq 0$, that $\Pi_\Lambda = L(t)\Lambda(t)L(t)^*$ satisfies

$$\begin{aligned} \Pi_\Lambda(t)z &= L(t)\Lambda(t)L(t)^* z = e^{-tA} P_r^* \Lambda(t) P_r e^{-tA^*} z \\ &= e^{-tA} P_r^* \Big[\Lambda(0) - \int_0^t \Lambda P_r e^{-sA^*} C^* R C e^{-sA} P_r^* \Lambda \, \mathrm{d}s \Big] P_r e^{-tA^*} z \\ &= e^{-tA} P_r^* \Lambda(0) P_r e^{-tA^*} z - \int_0^t e^{(s-t)A} \Pi_\Lambda(s) C^* R C \Pi_\Lambda(s) e^{(s-t)A^*} z \, \mathrm{d}s \end{aligned}$$

therefore by uniqueness of the mild solution of (29), we have that $\Pi_\Lambda(t)$ is solution of (29). □

Therefore, a numerical algorithm can be based only on the computation of $\Lambda(t)$ and $L(t)$ that are more tractable numerically. When $B \neq 0$, the reduction is more intricate as we may have an interplay between the model uncertainty space and the initial condition uncertainty space. However, if we minimize

$$\mathscr{J}(\zeta_r, \nu) = \frac{\alpha^{-2}}{2} \|\zeta_r\|^2_{\mathcal{Z}} + \int_0^T \Big(\frac{\delta^{-2}}{2} \|y^\delta - C z_{|P_r^*\zeta_r}\|^2_{\mathcal{Y}} + \frac{\kappa^{-2}}{2} \|\nu\|^2_{\mathcal{U}} \Big) \, \mathrm{d}t,$$

for a modified dynamics $\dot{z}(t) + Az(t) = g(t) + L(t)[L(t)^*L(t)]^{-1}L(t)^* B\nu(t)$, then, we can still decompose $\Pi(t) = L(t)\Lambda(t)L^*(t)$ with $\Lambda(t) \in \mathcal{L}(\mathcal{V}_r)$ following

$$\dot{\Lambda} + \Lambda L^*(t) C^* R C L(t) \Lambda - (L(t)^*L(t))^{-1} L(t)^* B Q B^* L(t) (L(t)^*L(t))^{-1} = 0. \tag{31}$$

The advantage of covariance reduction is then to maintain the original model accuracy and only reduce the cost of estimation. However, covariance reduction can lead to instabilities in the system, since errors in the complementary space to $\mathrm{Im}(L)$ are not stabilized by the Kalman filter. This could be the case

for modeling errors, but also for measurement errors entering the observer dynamics as a source term. Fortunately, for parabolic problems, the dynamics is exponentially stable in $\mathcal{V}_r^\perp$ at a controlled rate when $\mathcal{V}_r$ is formed from eigenvectors of A associated with the smallest eigenvalues.

To maintain such an adequate decomposition after time-discretization, the strategy remains the same. First discretize the dynamics and the criterion and then re-apply optimal control results at the discrete-time level. For instance, the discretized criterion $\mathscr{J}_n^{h,\tau+}$ becomes

$$\mathscr{J}_n^{h,\tau+}(\zeta_r,(\nu_k^\tau)_{1\leq k\leq n})=\frac{\alpha^{-2}}{2}\|\zeta_r\|_{\mathcal{Z}}^2+\frac{1}{2}\sum_{k=1}^n\kappa^{-2}\tau\|\nu_k^\tau\|_{\mathcal{U}}^2+\frac{1}{2}\sum_{k=0}^n\tau\delta^{-2}\|y_k^\delta-C^h z_{k|\zeta_r,(\nu_k^\tau)_{0\leq k\leq n}}^{h,\tau}\|_{\mathcal{Y}}^2,$$

subject to the discrete dynamics

$$\begin{cases}z_{k+1}^{h,\tau}=\Phi_1^{h,\tau}z_k^{h,\tau}+g_{k+1}^{h,\tau}+L_n^{h,\tau}[L_n^{h,\tau*}L_n^{h,\tau}]^{-1}L_n^{h,\tau*}B^{h,\tau}\nu_{k+1}^\tau, & 0\leq k\leq n-1,\\ z_0^h=\hat{z}_0^h+P_hP_r^*\zeta_r,\end{cases}$$

with $L_0^{h,\tau}=P_hP_r^*$ and $L_{n+1}^{h,\tau}=(\Phi_1^{h,\tau})^{n+1}P_hP_r^*=\Phi_1^{h,\tau}L_n^{h,\tau}$. We end up with the same optimal estimator dynamics (24) where $\Pi_n^{h,\tau+}$ is computed from

$$\begin{cases}U_n^{h,\tau+}=U_n^{h,\tau-}+L_n^{h,\tau*}C^{h*}R^\tau C^hL_n^{h,\tau*}, & n\in\mathbb{N},\\ U_{n+1}^{h,\tau-}=(U_n^{h,\tau+}+[L_n^{h,\tau*}L_n^{h,\tau}]^{-1}L_n^{h,\tau*}B^{h,\tau}Q^\tau B^{h,\tau*}L_n^{h,\tau*}[L_n^{h,\tau*}L_n^{h,\tau}]^{-1})^{-1} & n\in\mathbb{N},\end{cases}$$

since, following [80], one can easily prove recursively the following time-discrete counterpart of Theorem 1.4, with an additional modeling error as in (31).

Theorem 1.5. *Assuming that we have the initial decomposition $\Pi_0^{h,\tau+}=L_0^{h,\tau}(U_0^{h,\tau+})^{-1}L_0^{h,\tau*}$ and a n-dependent model error operator $B_n^{h,\tau}=L_n^{h,\tau}[L_n^{h,\tau*}L_n^{h,\tau}]^{-1}L_n^{h,\tau*}B^{h,\tau}$, then the recursive time-discrete Riccati solution (25) is given for all $n\in\mathbb{N}$ by*

$$\Pi_n^{h,\tau+}=L_n^{h,\tau}(U_n^{h,\tau+})^{-1}L_n^{h,\tau*}\text{ and }\Pi_n^{h,\tau-}=L_n^{h,\tau}(U_n^{h,\tau-})^{-1}L_n^{h,\tau*}.$$

To conclude this section on computational issues and related overcoming strategies, we can mention recent attempts to avoid model or covariance reduction. If the covariance is a Hilbert-Schmidt operator, we know that it is associated with a kernel. After analyzing the regularity of the kernel, the idea is to propose a spatial discretization of the kernel using the $\mathcal{H}$-matrix algebra, as it is now known for the discretization of integral equations. Such a strategy has been known since [64] and was mathematically analyzed in [5] for parabolic problems. Note also that there is a recent literature on the discretization of the Riccati solution using the "tensor" decomposition, which has the potential to be adapted to the Riccati dynamics (15) and its commonly used time discretization (25), see for example the recent attempts [67, 8].

2 Luenberger observer strategies and discretization for linear hyperbolic cases

In this section, we turn to the study of data assimilation strategies adapted to hyperbolic cases in the sense that in the dynamics (8) the operator $(A,\mathcal{D}(A))$ is now only the generator of a group. We typically have Schrödinger-like equations or wave-like equations in mind. In both cases, we continue to define these systems in their first-order form with the same unique solution given by semigroup theory. Note, however, that in this hyperbolic context we can no longer use the variational theory since A does not satisfy a coercivity estimate of the form (7). Here we will see that the 4D-Var approach remains, but with a different rationale. Moreover, there are still Kalman filters for such models, but when the covariance operator is discretized, the memory required to store the corresponding covariance matrix can be prohibitive since, hyperbolic systems inherently tend to exhibit sharp structures that propagate in the domain and correlate distant points without dissipation in time, so the entire discretization must be preserved for approximating the covariance operator. Preferably, other types of observers, called Luenberger observers [68], have been developed [4, 81, 21], with new challenges in terms of formalism and discretization.

2.1 Optimal control strategy formalism for hyperbolic problems

Returning to our deterministic vision of data assimilation, we could continue to follow a 4D-Var strategy justified for hyperbolic equations by minimizing a criterion

$$\mathscr{J}_T(\zeta,\nu)=\frac{\alpha^{-2}}{2}\|\zeta\|^2_{\mathcal{Z}}+\frac{1}{2}\int_0^T\left[\delta^{-2}\|y^\delta(s)-Cz_{|\zeta,\nu}(s)\|^2_{\mathcal{Y}}+\kappa^{-2}\|\nu(s)\|^2_{\mathcal{U}}\right]\,\mathrm{d}s, \tag{32}$$

Note that here we choose a classical Tikhonov regularization in $\mathcal{Z}$ as opposed to the generalized regularization in (9). This is due to the fact that inverse problems associated with such hyperbolic problems are often mildly ill-posed whereas parabolic systems are severely ill-posed. Therefore, we may avoid overregularization of the estimated initial condition. The minimization leads to the same optimal system given by (11). However, it is worth noting that the rationale differs since we can no longer benefit from the fact that A is a variational operator. To avoid using the Lagrangian definition, the Duhamel formula allows to directly introduce the dependency with respect to (ζ,ν) without relying on a formulation under constraint. We have for all $\xi\in\mathcal{Z}$, indeed,

$$\langle \mathrm{D}_\zeta\mathscr{J}_T(\zeta,\nu),\xi\rangle_{\mathcal{Z}',\mathcal{Z}}=\int_0^T\delta^{-2}(\xi,e^{-sA^*}C^*(Cz_{|\zeta,\nu}-y^\delta))_{\mathcal{Z}}\,\mathrm{d}s+\alpha^{-2}(\xi,\zeta)_{\mathcal{Z}},$$

whereas for all $\mu\in L^2((0,T);\mathcal{U})$,

$$\langle \mathrm{D}_\nu\mathscr{J}_T(\zeta,\nu),\mu\rangle_{L^2((0,T);\mathcal{U}'),L^2((0,T);\mathcal{U})}=\int_0^T\delta^{-2}(\xi,e^{-sA^*}C^*(Cz_{|\zeta,\nu}-y^\delta))_{\mathcal{Z}}\,\mathrm{d}s+\alpha^{-2}(\xi,\zeta)_{\mathcal{Z}}.$$

Therefore with $R=\delta^{-2}\mathrm{Id}_{\mathcal{Y}}$, introducing the adjoint dynamics for any model solution $z_{|\zeta,\nu}\in L^2((0,T);\mathcal{Z})$ and $y^\delta\in L^2((0,T);\mathcal{Y})$,

$$\begin{cases}\dot{q}_T(t)-A^*q_T(t)=-C^*R(y^\delta(t)-Cz_{|\zeta,\nu}(t)), & t\in(0,T),\\ q_T(T)=0,\end{cases} \tag{33}$$

whose mild solution in $C([0,T];\mathcal{Z})$ is given by the Duhamel formula – adapted to this backward formulation

$$q_T(t)=\int_t^T e^{(t-s)A^*}C^*R(y^\delta(s)-Cz_{|\zeta,\nu}(s))\,\mathrm{d}s,\quad t\in[0,T],$$

we find

$$\forall\xi\in\mathcal{Z},\quad \langle \mathrm{D}_\zeta\mathscr{J}_T(\zeta,\nu),\xi\rangle_{\mathcal{Z}',\mathcal{Z}}=\alpha^{-2}(\zeta,\xi)_{\mathcal{Z}}-(q_T(0),\xi)_{\mathcal{Z}},$$

and

$$\forall\mu\in L^2((0,T);\mathcal{U}),\quad \langle \mathrm{D}_\nu\mathscr{J}_T(\zeta,\nu),\mu\rangle_{L^2((0,T);\mathcal{U}'),L^2((0,T);\mathcal{U})}=\int_0^T-(q_T(t),B\mu)_{\mathcal{U}}+\kappa^{-2}(\nu,\mu)_{\mathcal{U}}\,\mathrm{d}t.$$

This leads to the same optimality system (11) with, here, $\mathit{\Pi}_0=\alpha^2\mathrm{Id}_{\mathcal{Z}}$. Note that minimizing (32) is the most common variational approach in data assimilation but not the only one, as one can use duality principle to define alternative optimal control strategies, see for instance [28] for hyperbolic problems.

Since on the one hand, the minimization leads to the same optimality system and, on the other hand, the Riccati dynamics (15) with $\mathit{\Pi}_0=\alpha^2\mathrm{Id}$ still admits a mild solution, then a Kalman filter can still be defined for such a system leading to a sequential alternative to the 4D-Var approach. However, this view is naive, since the resulting Kalman filter does not benefit from the regularization properties observed in the parabolic case, allowing effective computation. The reduced order formulation is not adapted. Indeed, when we reduce the covariance operator to a subspace, we implicitly assume that any error in the orthogonal of this space is stabilized by the dynamics itself. This property cannot be satisfied for a conservative system. Therefore, the covariance operator must be computed for the entire space. Similarly, the H-matrix formulations rely on the regularity of the covariance operator to compress its numerical storage.

2.2 Luenberger observers as an optimal filtering alternative

If we hold on to sequential approaches, we need to find an alternative to the Kalman filter. This is where observer approaches come in, based not on optimal control considerations, but on stabilization theory. The principle is to design a gain G that is adapted to the dynamics and can be computed, so that the observer with dynamics of the form

$$\begin{cases} \dot{\hat{z}} + A\hat{z} = g + G(y^\delta - C\hat{z}), & \text{in } (0,T), \\ \hat{z}(0) = \hat{z}_0. \end{cases}$$

tracks the target system through time. We refer to [55] for an exhaustive definition of observer systems. In essence, G should be designed so that, in the absence of measurement error – i.e. $\eta \equiv 0$ – and model noise error – i.e. $\nu \equiv 0$ – the error system $\tilde{z} = \check{z} - \hat{z}$ is asymptotically – ideally exponentially – stable to 0. This error system is solution of

$$\begin{cases} \dot{\tilde{z}} + (A + GC)\tilde{z} = 0, & \text{in } (0,T), \\ \tilde{z}(0) = \check{z}_0 - \hat{z}_0. \end{cases}$$

For a conservative system, namely with A skew-adjoint, it is well known that the simplest choice, namely $G = \gamma C^*$ with $\gamma \in \mathbb{R}$, can be very efficient when the system is observable as studied in [49]. For instance, if for all modes $(\varphi, \lambda) \in \mathcal{Z} \times \mathbb{R}$ solution of $A\varphi = \lambda\varphi$, we have $C\varphi = 0 \Rightarrow \varphi = 0$, then this Hautus test implies the error asymptotic stability [90]. Moreover, if we have the observability condition

$$\exists (T_0, c_{\text{st}}) \text{ such that } \forall T \geq T_0, \forall z_0 \in \mathcal{Z}, \quad \int_0^T \|Ce^{-tA} z_0\|_{\mathcal{Y}}^2 \, \mathrm{d}t \geq c_{\text{st}} \|z_0\|_{\mathcal{Z}}^2,$$

then, the error $\tilde{z}$ is exponentially stable to 0 [49]. The resulting gain $G = \gamma C^*$ should be compared with the Kalman gain $G = \Pi C^*$ considering a relation between efficiency and complexity. In data assimilation, such a strategy is called a nudging approach [4], while in observation theory for partial differential equations it is more commonly called a Luenberger approach [81, 21].

2.3 Time discretization: from observability conditions to numerical analysis

Let us now consider the time discretization of the presented data assimilation methods for conservative systems. As far as the 4D-Var algorithm is concerned, the strategy does not change, since only the mathematical proofs have been adapted. Therefore, the discretization strategy remains, namely, first discretize the system, then the criterion, and then compute a discrete-time adjoint equation, which is the Lagrange multiplier of the constraints associated with the discrete-time dynamics. The discrete-time adjoint equation allows the exact computation of the functional gradient that must be integrated in any gradient descent strategy.

As for the sequential estimation strategies, we also propose to first discretize the model and then define a Luenberger filter fitted to this discretization. For this purpose, the filter must be dissipative for the discrete estimation error. Moreover, the observability conditions should be satisfied, which leads to additional exponential stability.

For conservative systems defined by A being skew-adjoint, numerous time schemes can be chosen. To present these ideas on an illustrative example, let us consider a mid-point discretization which has the advantage of respecting the conservative nature of the underlying system:

$$\frac{z_{k+1}^{h,\tau} - z_k^{h,\tau}}{\tau} + A^h \frac{z_{k+1}^{h,\tau} + z_k^{h,\tau}}{2} = g_{k+\frac{1}{2}}^h + B^h \nu_{k+1}, \quad 0 \leq k \leq n, \tag{34}$$

with $g_{k+\frac{1}{2}}^h = g^h(\frac{1}{2}(t^{k+1} + t^k))$. This system can be rewritten in the form (18) with a transition operator from time-step n to time step $n+1$ and a model noise operator given by

$$\Phi_1^{h,\tau} = \Big(\mathrm{Id}_{\mathcal{Z}^h} + \frac{\tau}{2} A^h\Big)^{-1} \Big(\mathrm{Id}_{\mathcal{Z}^h} - \frac{\tau}{2} A^h\Big), \text{ and } B^{h,\tau} = \tau \Big(\mathrm{Id}_{\mathcal{Z}^h} + \frac{\tau}{2} A^h\Big)^{-1} B^h.$$

Considering such transition dynamics, the optimality system associated with the 4D-Var approach reproduces exactly (23). As for the continuous-time system, this optimality system is reached through a gradient descent approach computed from the adjoint dynamics

$$\begin{cases} q^{h,\tau}_{k|n} = \Phi_1^{h,\tau *} q^{h,\tau}_{k+1|n} + C^{h*} R^\tau \big(y^\delta_k - C^h z^{h,\tau}_k\big), \quad 0 \le k \le n-1, \\ q^{h,\tau}_{n|n} = 0, \end{cases}$$

for any forward dynamics $(z^{h,\tau}_k)_{0\le k\le n}$. The previous system might appear as less practical than the original dynamics (34). In fact, this impression is wrong because, by defining $w^{h,\tau}_k = (\mathrm{Id}_{\mathcal{Z}^h} + \frac{\tau}{2}A^h)^{*-1} q^{h,\tau}_{k|n}$, we get the following dynamics for $(w^{h,\tau}_k)_{0\le k\le n}$:

$$(\mathrm{Id}_{\mathcal{Z}^h} - \frac{\tau}{2}A^h) w^{h,\tau}_k = (\mathrm{Id}_{\mathcal{Z}^h} + \frac{\tau}{2}A^h) w^{h,\tau}_{k+1} + C^{h*} R^\tau \big(y^\delta_k - C^h z^{h,\tau}_k\big),$$

leading to

$$\frac{w^{h,\tau}_{k+1} - w^{h,\tau}_k}{\tau} + A^h \frac{w^{h,\tau}_{k+1} + w^{h,\tau}_k}{2} = -C^{h*} R^\tau \big(y^\delta_k - C^h z^{h,\tau}_k\big). \tag{35}$$

Therefore, $(w^{h,\tau}_k)_{0\le k\le n}$ is a discretization of the continuous-time adjoint variable, just as $(z^{h,\tau}_k)_{0\le k\le n}$ is a discretization of the continuous-time state, with the same time scheme. This property is very general and can be obtained regardless of the chosen time discretization. It allows the adjoint equation to be computed with the same implementation as for the direct model, an essential requirement when turning to more complex physical, modeling, and scientific computing requirements.

When we turn to the discretization of the observer, it is natural to use a mid-point discretization also for the additional feedback, viz.

$$\frac{z^{h,\tau}_{n+1} - z^{h,\tau}_n}{\tau} + A^h \frac{z^{h,\tau}_{n+1} + z^{h,\tau}_n}{2} = g^h_{n+\frac12} + \gamma C^{h*} \Big(y^\delta_n - C^h \frac{z^{h,\tau}_{n+1} + z^{h,\tau}_n}{2}\Big), \quad n \in \mathbb{N},$$

so that the error $\tilde z^{h,\tau}_n = \check z^{h,\tau}_n - \hat z^{h,\tau}_n$ satisfies the energy identity,

$$\begin{aligned} \frac12 \|\tilde z^{h,\tau}_{n+1}\|^2_{\mathcal{Z}} - \frac12 \|\tilde z_n\|^2_{\mathcal{Z}} = -\gamma \Big\| C^h \frac{\tilde z^{h,\tau}_{n+1} + \tilde z^{h,\tau}_n}{2} \Big\|^2_{\mathcal{Y}} \\ + \Big(B\nu_{n+1}, \frac{\tilde z^{h,\tau}_{n+1} + \tilde z^{h,\tau}_n}{2}\Big)_{\mathcal{Z}} + \gamma \Big(C^{h*}\eta^{h,\tau}_n, \frac{\tilde z^{h,\tau}_{n+1} + \tilde z^{h,\tau}_n}{2}\Big)_{\mathcal{Z}}. \end{aligned}$$

where $\eta^{h,\tau}_n$ gather measurement errors and discretization errors. First assuming $\eta^{h,\tau}_n = 0$ and $\nu_n = 0$, the energy of the error decreases. Moreover, it converges exponentially to 0 if the following discrete observation inequality is satisfied:

$$\exists (n_0, c_{\mathrm{st}}) \text{ such that } \forall n \ge n_0, \quad \sum_{k=0}^{n} \Big\| C^h \frac{\tilde z^{h,\tau}_{k+1} + \tilde z^{h,\tau}_k}{2} \Big\|^2_{\mathcal{Y}} \ge c_{\mathrm{st}} \|\tilde z^{h,\tau}_0\|^2_{\mathcal{Z}}.$$

Unfortunately, such an observability inequality is often not satisfied for popular discretization schemes, even if the observability inequality was satisfied for the continuous time system. This phenomenon is due to unwanted spurious high frequencies coming from the discretization not being stabilized, and has been the subject of an extensive literature, whether for spatial discretization [47, 6, 88] or temporal discretization [36, 95], see also the detailed review [96].

When observability is not satisfied, many strategies rely on adjusting the time- scheme to stabilize the spurious high frequencies responsible for the lack of observability. A typical example of such a strategy was proposed in [22, 27] for observers of wave systems, drawing inspiration from what was originally proposed in [37] for control problems. In [22], the resulting time scheme is a prediction-correction splitting timing scheme of the form

$$\begin{cases} \dfrac{\hat z^{h,\tau-}_{n+1} - \hat z^{h,\tau+}_n}{\tau} + A^h \dfrac{\hat z^{h,\tau-}_{n+1} + \hat z^{h,\tau+}_n}{2} = g^h_{n+\frac12} + \gamma C^{h*} \Big(y^\delta_n - C^h \dfrac{\hat z^{h,\tau-}_{n+1} + \hat z^{h,\tau+}_n}{2}\Big), & n \in \mathbb{N}, \\ \dfrac{\hat z^{h,\tau+}_{n+1} - \hat z^{h,\tau-}_{n+1}}{\tau} + V^{h,\tau}_\epsilon \hat z^{h,\tau+}_{n+1} = 0, & n \in \mathbb{N}, \end{cases} \tag{36}$$

where $V_\epsilon^{h,\tau}$ is a vanishing viscosity operator that is positive definite in $\mathcal{Z}_h$ and commutes with the projector to the first modes of A from eigenvalues below $\frac{1}{\epsilon}$. The parameter ϵ is small and consistent with respect to the discretization parameters h and τ and controls the threshold for cutting at high frequencies. A typical choice of vanishing viscosity operator from [37] is $V_\epsilon^{h,\tau} = -\epsilon(A^h)^2$. By choosing $\epsilon = \max(h,\tau)$, [22] obtains the full time and space analysis in the absence of measurement errors, resulting in an estimate of the form

$$\|\hat{z}_n^{h,\tau+} - \check{z}(n\tau)\|_{\mathcal{Z}} \leq c_{\mathrm{st}}(\hat{z}_0)\max(\epsilon, \epsilon^2 h^{-1}\tau), \quad n \in \mathbb{N}. \tag{37}$$

Basically, this estimate is based on the fact that [37] proves that the scheme (36) is a uniform exponentially stable approximation of the damped system

$$\dot{\tilde{z}} + A\tilde{z} - \epsilon A^2\tilde{z} + C^*C\tilde{z} = 0,$$

due to a satisfied discrete observability inequality of the form

$$\exists (n_0, c_{\mathrm{st}}), \forall n \geq n_0, \quad \sum_{k=0}^{n} \left\| C^h \frac{\tilde{z}_{k+1}^{h,\tau-} + \tilde{z}_k^{h,\tau+}}{2} \right\|_{\mathcal{Y}}^2 + \epsilon \left\| A^h \tilde{z}_{k+1}^{h,\tau+} \right\|_{\mathcal{Z}}^2 \geq c_{\mathrm{st}} \|\tilde{z}_0^{h,\tau}\|_{\mathcal{Z}}^2.$$

The estimates (37) must be compared with the estimates from classical numerical analysis that could have been obtained for the target system if we had known the initial conditions and the model error. We would then have obtained

$$\|\check{z}_n^{h,\tau+} - \check{z}(n\tau)\|_{\mathcal{Z}} \leq c_{\mathrm{st}}(T)(h + \tau^2),$$

with a constant $c_{\mathrm{st}}(T)$ that deteriorates the estimation as T grows. Since the observer benefits from the available measurement and under an observability condition, we indeed have a better numerical estimate of the trajectory than with the direct system. Note that the observer estimate is naturally perturbed by the measurement noise, leading to a tradeoff through the Grönwall inequality between exponential stability and additional errors coming from the measurement procedure.

In the case of the wave-like equation, it was found in [51] that the simpler choice

$$A = \begin{pmatrix} 0 & -\mathrm{Id} \\ A_0 & 0 \end{pmatrix} \Rightarrow V_\epsilon^{h,\tau} = \begin{pmatrix} \mathrm{Id}^h & 0 \\ 0 & A_0^h \end{pmatrix}, \tag{38}$$

yields similar results and has the advantage of being easy to implement, in particular when considering elasticity. Alternatively to (36), the observation can be considered in the correction term as in [27, 51], resulting in

$$\begin{cases} \dfrac{\hat{z}_{n+1}^{h,\tau-} - \hat{z}_n^{h,\tau+}}{\tau} + A^h \dfrac{\hat{z}_{n+1}^{h,\tau-} + \hat{z}_n^{h,\tau+}}{2} = g_{n+\frac{1}{2}}^h, & n \in \mathbb{N}, \\ \dfrac{\hat{z}_{n+1}^{h,\tau+} - \hat{z}_{n+1}^{h,\tau-}}{\tau} + V_\epsilon^{h,\tau}\hat{z}_{n+1}^{h,\tau+} = \gamma C^{h*}\Big(y_n^\delta - C^h\hat{z}_{n+1}^{h,\tau+}\Big), & n \in \mathbb{N}, \end{cases} \tag{39}$$

or the vanishing viscosity operator can be used in the prediction step

$$\begin{cases} \dfrac{\hat{z}_{n+1}^{h,\tau-} - \hat{z}_n^{h,\tau+}}{\tau} + A^h \dfrac{\hat{z}_{n+1}^{h,\tau-} + \hat{z}_n^{h,\tau+}}{2} + V_\epsilon^{h,\tau}\hat{z}_{n+1}^{h,\tau-} = g_{n+\frac{1}{2}}^h & n \in \mathbb{N}, \\ \dfrac{\hat{z}_{n+1}^{h,\tau+} - \hat{z}_{n+1}^{h,\tau-}}{\tau} = \gamma C^{h*}(y_n^\delta - C^h\hat{z}_{n+1}^{h,\tau+}), & n \in \mathbb{N}. \end{cases} \tag{40}$$

With this last choice, the correction can be rewritten into the following form

$$\hat{z}_{n+1}^{h,\tau+} = \hat{z}_{n+1}^{h,\tau-} + \Pi_\infty^{h,\tau} C^{h*}(y_n^\delta - C^h\hat{z}_{n+1}^{h,\tau-}),$$

where $\Pi_\infty^{h,\tau} = (\gamma^{-1}\mathrm{Id}_h + \tau C^{h*}C^h)^{-1}$ and we obtain a very similar structure to the Kalman filter, but instead of having to compute the inverse of a full covariance matrix, in this case it is replaced by inverting a local operator, i.e. a sparse matrix defined over the degrees of freedom of the discrete state. For a wave-like equation where $V_\epsilon^{h,\tau}$ is given by (38), we ultimately need to solve a Kalman-like algorithm where

the covariance is a computable steady-state operator and the model incorporates a vanishing structure damping, hence becoming a variational system [14, II -2-3].

Let us conclude this section by mentioning a result that explores the implications of data sampling [27]. Given a discretization of time-step τ, how should we typically assimilate data sampled on a grid $0 < T_0 < ... < T_r = r\Delta T < ... < T_m = T$ with $\Delta T/\tau \in \mathbb{N}$. The question is whether we should interpolate the data so that it is available at each time step of the discretization, or whether we should use the data only when they are available. In the first case, we benefit from exponential stability at each time step, but at the cost of additional measurement error due to interpolation. Mathematically, to account for sampling of the data, the time scheme (39) is changed into

$$\begin{cases} \dfrac{\hat{z}_{n+1}^{h,\tau-} - \hat{z}_n^{h,\tau+}}{\tau} + A\dfrac{\hat{z}_{n+1}^{h,\tau-} + \hat{z}_n^{h,\tau+}}{2} = g_{n+\frac{1}{2}}^h, & n \in \mathbb{N}, \\ \dfrac{\hat{z}_{n+1}^{h,\tau+} - \hat{z}_{n+1}^{h,\tau-}}{\tau} = \rho^{n+1}\gamma C^* \left(d^{n+1} - C\hat{z}_{n+1}^{h,\tau+}\right) + V_\epsilon^{h,\tau}\hat{z}_{n+1}^{h,\tau+}, & n \in \mathbb{N}, \\ \hat{z}_0^{h,\tau+} = \hat{z}_0^h, \end{cases}$$

where for interpolated data $\rho^n \equiv 1$ and between two successive data indexed by r and $r+1$ of corresponding indexes j_r and j_{r+1} in the simulation grid indexed by n,

$$d_n = \frac{n - j_r}{j_{r+1} - j_r} y_{r+1}^\delta + \left(1 - \frac{n - j_r}{j_{r+1} - j_r}\right) y_r^\delta \qquad j_r \leq n \leq j_{r+1}.$$

Alternatively when using only the available data

$$\rho^n = \begin{cases} 1, \\ 0, \end{cases} \qquad d^n = \begin{cases} y_r^\delta & \text{if } \exists r \in \mathbb{N} \ : \ n = j_r, \\ 0 & \text{otherwise.} \end{cases}$$

Using the uniform observability conditions associated with schemes with multiple time steps [95],[27] quantifies the dilemma as a function of the sampling time step and the measurement error with the following conclusion: in the case of a reasonable time-sampling of the data and with potentially high noise, we should interpolate the data, while in the case of poor data availability, using data only when they are available is more robust.

2.4 Coupling Luenberger observers with optimal filtering

2.4.1 The continuous-time setting

For conservative systems, the use of an optimal filtering strategy is counterproductive because the uncertainties pollute the entire state space and force the computation of a full covariance with the associated curse of dimensionality. Therefore, we should rely on Luenberger approaches with their appropriate time discretization. However, by using a Luenberger observer for the state, we lose the advantage of generality offered by optimal control methods. In particular, how can we deal with a coupled system that is hyperbolic-parabolic, or how can we perform joint state and parameter sequential estimation – also known as adaptive estimation [34] – for a conservative system? And once the strategy is defined, how can we discretize it appropriately?

Let us start by the second problem of jointly estimating the state and identifying parameters. With a strategy only relying on optimal control, the strategy is straightforward as we only need to complement the initial dynamics (8) with the parameter – gathered in a variable p belonging to a Hilbert space $\mathcal{P}$ – dynamics which, by definition, reads $\dot{p} = 0$. Gathering the state and parameter in a vector $\boldsymbol{z} = (z, p)$, forgetting the model noise to simplify the notation without loss of generality and considering a parameter dependency in the source term to keep a linear problem, the dynamics now reads

$$\underbrace{\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} z \\ p \end{pmatrix}}_{\dot{\boldsymbol{z}}} + \underbrace{\begin{pmatrix} A & B \\ 0 & 0 \end{pmatrix}\begin{pmatrix} z \\ p \end{pmatrix}}_{\boldsymbol{A}\boldsymbol{z}} = \underbrace{\begin{pmatrix} g \\ 0 \end{pmatrix}}_{\boldsymbol{g}}, \quad \text{with} \quad \underbrace{\begin{pmatrix} z(0) \\ p(0) \end{pmatrix}}_{\boldsymbol{z}(0)} = \underbrace{\begin{pmatrix} \hat{z}_0 \\ \hat{p}_0 \end{pmatrix}}_{\hat{\boldsymbol{z}}_0} + \underbrace{\begin{pmatrix} \zeta \\ \theta \end{pmatrix}}_{\boldsymbol{\zeta}}.$$

With this new dynamics, the optimal control strategy is unchanged and consists in minimizing the functional $\mathscr{J}_T$ under the constraint of the joint dynamics, namely

$$\min_{\zeta,\theta}\left\{ \mathscr{J}_T(\zeta,\theta) = \frac{1}{2}(\Pi_0^{-1}\zeta,\zeta)_{\mathcal{Z}} + \frac{1}{2}(\Lambda_0^{-1}\theta,\theta)_{\mathcal{P}} + \int_0^T \frac{\delta^{-2}}{2}\|y^\delta - \underbrace{\begin{pmatrix} C & 0 \end{pmatrix}}_{\boldsymbol{C}} \boldsymbol{z}_{|(\zeta,\theta)}\|^2 \,\mathrm{d}t \right\}.$$

The minimization procedure remains unchanged with a Pontryagin principle leading to a two-ends problem or a dynamic programming approach leading to a coupled Kalman filter.

When considering a conservative system, a way to avoid the use of a Kalman filter on the state equation consists in modifying the dynamics in the minimization in order to justify a reduced-order minimization on the parameter space. This reads

$$\min_{\theta\in\mathcal{P}}\left\{ \mathscr{J}_T(\theta) = \frac{1}{2}\langle\Lambda_0^{-1}\theta,\theta\rangle_{\mathcal{P}} + \int_0^T \frac{\delta^{-2}}{2} e^{-\varrho(T-t)}\|y^\delta - \boldsymbol{C}\breve{\boldsymbol{z}}_{|\theta}\|^2 \,\mathrm{d}t \right\}, \tag{41}$$

to be minimized under the constraint that $\breve{\boldsymbol{z}}_{|\theta}$ is a solution, for a given θ, of the dynamics

$$\begin{cases} \dot{\breve{\boldsymbol{z}}} + \boldsymbol{A}\breve{\boldsymbol{z}} = \boldsymbol{g} + \gamma\boldsymbol{C}^*(y^\delta - \boldsymbol{C}\breve{\boldsymbol{z}}), \quad \text{in } (0,T), \\ \breve{\boldsymbol{z}}(0) = \hat{\boldsymbol{z}} + \begin{pmatrix} 0 \\ \theta \end{pmatrix}. \end{cases}$$

We recognize here the dynamics of the Luenberger observer converging asymptotically to the solution without initial state error for a known θ. In principle, the scaling $e^{\varrho(T-t)}$ must be understood as a way to increase the weight of the present measurement from the past measurement, since the Luenberger observer filters the initial state error only asymptotically. As formally shown in [73], the associated optimal observer is given by the following dynamics, originally proposed by [94] in the context of adaptive observers,

$$\begin{cases} \dot{\hat{z}} + A\hat{z} + B\hat{p} = g + \gamma C^*(y^\delta - C\hat{z}) + L\dot{\hat{p}}, \quad t > 0, \\ z(0) = \hat{z}_0 \\ \dot{\hat{p}} = \Lambda L^* C^* R(y^\delta - C\hat{z}), \quad t > 0, \\ \hat{p}(0) = \hat{p}_0 \\ \dot{\Lambda} + \Lambda L^* C^* R C L \Lambda - \varrho\Lambda = 0, \quad t > 0, \\ \Lambda(0) = \Lambda_0 \\ \dot{L} + (A + \gamma C^* C)L + B = 0, \quad t > 0, \\ L(0) = 0 \end{cases} \tag{42}$$

Namely, we have the following theorem.

Theorem 2.1. *The mild solution of* (42) *is an optimal observer in the sense that*

$$\forall t > 0, \quad \hat{p}(t) = \hat{p}_0 + \bar{\theta}_t,$$

where $\bar{\theta}_t = \mathrm{argmin}_{\theta\in\mathcal{P}}\, \mathscr{J}_t(\theta)$*, the criterion defined in* (41)*.*

Proof. The proof is essentially based on the decomposition of the Riccati solution $\boldsymbol{\Pi}$ which decouples – again in the sense of [65, 12] – the two-ends problem associated with (41), namely

$$\begin{aligned} \dot{\boldsymbol{\Pi}} - \varrho\boldsymbol{\Pi} + (\boldsymbol{A} + \gamma\mathbf{C}\mathbf{C}^*)\boldsymbol{\Pi} + \boldsymbol{\Pi}(\boldsymbol{A}^* + \gamma\mathbf{C}\mathbf{C}^*) + \boldsymbol{\Pi}\mathbf{C}^* R\mathbf{C}\boldsymbol{\Pi} \\ = (\mathbf{A} + \gamma\mathbf{C}\mathbf{C}^* - \frac{\varrho}{2}\mathrm{Id})\boldsymbol{\Pi} + \boldsymbol{\Pi}(\mathbf{A}^* + \gamma\mathbf{C}\mathbf{C}^* - \frac{\varrho}{2}\mathrm{Id}) + \boldsymbol{\Pi}\mathbf{C}^* R\mathbf{C}\boldsymbol{\Pi} = 0, \end{aligned}$$

initialized from

$$\boldsymbol{\Pi}(0) = \begin{pmatrix} 0 & 0 \\ 0 & \Lambda_0 \end{pmatrix} = \begin{pmatrix} 0 \\ \mathrm{Id} \end{pmatrix} \Lambda_0 \begin{pmatrix} 0 & \mathrm{Id} \end{pmatrix}.$$

Using similar argument than in Theorem 1.4 but here with the reduced space corresponding to the parameter space, we find that

$$\begin{aligned}\boldsymbol{\Pi}(t) &= e^{-t(\mathbf{A}+\gamma\mathbf{C}\mathbf{C}^*-\frac{\varrho}{2}\mathbf{Id})}\boldsymbol{\Lambda}_0 e^{-t(\mathbf{A}^*+\gamma\mathbf{C}\mathbf{C}^*-\frac{\varrho}{2}\mathbf{Id})} \\ &\quad - \int_0^t e^{(s-t)(\mathbf{A}+\gamma\mathbf{C}\mathbf{C}^*-\frac{\varrho}{2}\mathbf{Id})}\boldsymbol{\Pi}\mathbf{C}^* R\mathbf{C}\boldsymbol{\Pi} e^{(s-t)(\mathbf{A}^*+\gamma\mathbf{C}\mathbf{C}^*-\frac{\varrho}{2}\mathbf{Id})}\,\mathrm{d}s \\ &= \begin{pmatrix} L\Lambda L^* & L\Lambda \\ \Lambda L^* & \Lambda \end{pmatrix} = \begin{pmatrix} L \\ \mathrm{Id} \end{pmatrix} \Lambda \begin{pmatrix} L^* & \mathrm{Id} \end{pmatrix}.\end{aligned}$$

From the covariance decomposition, simple computations finally give that $\hat{p}(t) = \hat{p}_0 + \operatorname{argmin}_{\theta\in\mathcal{P}} \mathscr{J}_t(\theta)$. □

The proof readily extends to a case where $B(t)$ is time-dependent, by considering Riccati dynamics defined from the mild evolution operator defined from $\boldsymbol{A}(t)$ [78, Chapter 5] in place of $e^{-t(\mathbf{A}^*+\gamma\mathbf{C}\mathbf{C}^*-\frac{\varrho}{2}\mathbf{Id})}$. Furthermore, the case of time-dependent sources is fundamental from an observability perspective as observability for joint state and parameter systems are usually covered by assumptions of the persistence of the excitation [34, 35] only satisfied with a time-dependent $B(t)$.

Note finally that in [29] a similar strategy has been formally proposed for a weakly coupled parabolic and hyperbolic system of the form

$$\underbrace{\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} z \\ p \end{pmatrix}}_{\dot{\boldsymbol{z}}} + \underbrace{\begin{pmatrix} A & B \\ D & 0 \end{pmatrix}\begin{pmatrix} z \\ p \end{pmatrix}}_{\boldsymbol{A}\boldsymbol{z}} = \underbrace{\begin{pmatrix} g \\ 0 \end{pmatrix}}_{\boldsymbol{g}},$$

with application to a coupled electromechanics system arising in cardiac modeling. The case of a fully general coupled parabolic-hyperbolic system remains to be studied.

2.4.2 The discrete-time setting

Here again, the discretization follows the same repeated principle, namely discretize the dynamics and criterion and then apply optimal control results. The discretization of the parameter dynamics is obviously $p_{n+1}^{h,\tau} = p_n^{h,\tau} \in \mathcal{P}^h$ and the discretization of the criterion (41) is

$$\mathscr{J}_n^{h,\tau}(\theta^{h,\tau}) = \frac{1}{2}((\Lambda_0^h)^{-1}\theta^{h,\tau},\theta^{h,\tau})_{\mathcal{P}^h} + \frac{1}{2}\sum_{k=0}^{n-1}\tau\delta^{-2}e^{-\varrho(n-k)\tau}\|y_k^\delta - \boldsymbol{C}^h\breve{\boldsymbol{z}}_{k|\theta^{h,\tau}}^{h,\tau}\|_{\mathcal{Y}}^2.$$

Then, the optimal parameter $\hat{p}_n^{h,\tau} = \operatorname{argmin}_{\theta^{h,\tau}\in\mathcal{P}^h} \mathscr{J}_n^{h,\tau}(\theta^{h,\tau})$ is solution of a splitting scheme. For all $n \in \mathbb{N}$, the correction-step is given by

$$\begin{cases} \hat{p}_n^{h,\tau+} = \hat{p}_n^{h,\tau-} + \Lambda_{n+1}^{h,\tau}L_{n+1}^{h,\tau*}C^{h*}R^\tau(y^\delta - C^h\hat{z}_n^{h,\tau-}), & n\in\mathbb{N}, \\ \hat{z}_n^{h,\tau+} = \hat{z}_n^{h,\tau-} + L_{n+1}^{h,\tau}\Lambda_{n+1}^{h,\tau}L_{n+1}^{h,\tau*}C^{h*}R^\tau(y^\delta - C^h\hat{z}_n^{h,\tau-}), & n\in\mathbb{N}, \end{cases}$$

where

$$\begin{cases} \Lambda_{n+1}^{h,\tau} = [e^{-\varrho\tau}(\Lambda_n^{h,\tau})^{-1} + L_{n+1}^{h,\tau*}C^{h*}R^\tau C^h L_{n+1}^{h,\tau}]^{-1}, & n\in\mathbb{N}, \\ L_{n+1}^{h,\tau} = \Phi_1^{h,\tau}L_n^{h,\tau} - \gamma\tau C^{h*}C^h L_{n+1}^{h,\tau} + B^{h,\tau}, & n\in\mathbb{N}. \end{cases}$$

The correction step is then followed by a prediction-step

$$\begin{cases} \hat{z}_{n+1}^{h,\tau-} = \Phi_1^{h,\tau}\hat{z}_n^{h,\tau+} + B^{h,\tau}\hat{p}_n^{h,\tau+} + \gamma\tau C^{h*}(y_{n+1}^\delta - C^h\hat{z}_{n+1}^{h,\tau-}), & n\in\mathbb{N}, \\ \hat{p}_{n+1}^{h,\tau-} = \hat{p}_n^{h,\tau+}, & n\in\mathbb{N}. \end{cases}$$

Theorem 2.2. *Denoting $\bar{\theta}_n^{h,\tau} = \operatorname{argmin}_{\theta^{h,\tau}\in\mathcal{P}^h} \mathscr{J}_n^{h,\tau}(\theta^{h,\tau})$, we have the fundamental identity*

$$\forall n\in\mathbb{N},\quad \hat{p}_n^{h,\tau-} = \hat{p}_0 + \bar{\theta}_n^{h,\tau}.$$

Again Theorem 2.2, counterpart of (2.3), brings stability to this prediction-correction time-scheme. The proof of Theorem 2.2 is analogous to that of Theorem 1.5. Indeed such joint state and parameter estimation proposed in [73], is an adaptation of the general reduced approach proposed by [79] and called Singular Evolutive Extended Kalman Filter.

3 Least squares estimation and associated discretization for non-linear models

We now turn to data assimilation involving nonlinear systems, and still in a deterministic context. Since the non-linear infinite dimensional setting remains to be covered in its full generality in the literature , we restrict the presentation of this last part to finite dimensional systems. To establish the notation, we consider

$$\begin{cases} \dot{z}(t) = f(z(t),t) + B(t)\nu(t), \quad t > 0, \\ z(0) = \hat{z}_0 + \zeta \end{cases} \tag{43}$$

where at time t the state variable $z(t) \in \mathcal{Z} \simeq \mathbb{R}^{\mathrm{N}}$, $f : \mathcal{Z} \to \mathcal{Z}$ and is a nonlinear mapping, for instance C^1 of bounded derivatives. We keep an additive linear perturbation as a model error with $B \in C^0([0,T], \mathbb{M}_{\mathrm{N},\mathrm{N}_\nu}(\mathbb{R}))$, but one can also think of more general settings with $B(z,t)$ or even non-additive model error models entering f, see for instance [85, Chapter 13]. We also consider a measurement procedure $\forall t \geq 0, \quad y^\delta(t) = h(z(t),t) + \eta(t)$, where this time $h : \mathcal{Z} \times \mathbb{R}^+ \to \mathcal{Y} \simeq \mathbb{R}^{\mathrm{m}}$ is a nonlinear mapping, for instance C^1 of bounded derivative.

Again, in data assimilation, an estimation can be produced from a least-square minimization

$$(\bar{\zeta}_T, \bar{\nu}_T) = \underset{\substack{\zeta \in \mathcal{Z} \\ \nu \in L^2(0,T;\mathcal{U})}}{\operatorname{argmin}} \left\{ \mathscr{J}_T(\zeta,\nu) = \mathscr{V}_0(\zeta) + \int_0^t \frac{\delta^{-2}}{2} \|y^\delta(t) - h(z_{|\zeta,\nu}(t),t)\|^2 + \frac{\kappa^{-2}}{2} \|\nu(s)\|^2 \, \mathrm{d}s \right\}, \tag{44}$$

with $\mathscr{V}_0$ a convex penalty function. When uniqueness is not guaranteed the following construction will not be unique [42, 25]. The associated two-ends problem is *formally* given by

$$\begin{cases} \dot{\bar{z}}_T(t) = f(\bar{z}(t),t) + B(t)QB(t)^* \bar{q}_T(t), & t \in [0,T], \\ \dot{\bar{q}}_T(t) + \mathrm{D}f(\bar{z}_T(t),t)^* \bar{q}_T(t) = -\mathrm{D}h(\bar{z}_T(t),t))^*(y^\delta - h(\bar{z}_T(t),t)), & t \in [0,T], \\ \bar{z}_T(0) = \hat{z}_0 + (\mathrm{D}\mathscr{V}_0)^{-1}(\bar{q}_T(0)), \\ \bar{q}_T(0) = 0. \end{cases} \tag{45}$$

As for the linear case, minimizing (44) is often performed using a gradient descent approach – with the risk of finding only a local minimum [25]. In the family of gradient descent methods, we would like to highlight the interest in using the Gauss-Newton approach based on a well-posed approximated linearization of the non-linear PDE model [48, 59]. There, each iteration of the Gauss-Newton approach is reduced to a linear quadratic problem solved by the methods presented earlier. As a consequence, this Gauss-Newton strategy reveals to be a good combination of mathematical reasoning and numerical efficiency for data assimilation purposes.

3.1 The Mortensen observer

In the linear context, the Kalman observer was presented as the optimal observer in the sense that

$$\forall t \geq 0, \quad \hat{z}(t) = \bar{z}_t(t) := z_{|\bar{\zeta}_t, \bar{\nu}_t}(t).$$

In [74], this definition is conserved but with nonlinear dynamics and observation mapping. From dynamics programming results [43], it was then shown [42, 72] that for all time t, $\hat{z}(t)$ can be proved to be the following minimizer:

$$\forall t \geq 0, \quad \hat{z}(t) = \underset{z \in \mathcal{Z}}{\operatorname{argmin}} \, \mathscr{V}(z,t), \tag{46}$$

defined from the so-called *cost-to-come*

$$\mathscr{V}(z,t) = \inf_{(\zeta,\nu) \in \mathcal{A}_{x,t}} \left[\mathscr{V}_0(\zeta) + \int_0^t \frac{\delta^{-2}}{2} \|y^\delta(s) - h(z_{|\zeta,\nu}(s),s)\|^2 + \frac{\kappa^{-2}}{2} \|\nu(s)\|^2 \, \mathrm{d}s \right], \tag{47}$$

where the infimum is taken over the pre-image set

$$\mathcal{A}(z,t) = \left\{ (\zeta,\nu) \in \mathcal{Z} \times L^2((0,t);\mathcal{U}) : z_{|\zeta,\nu} \text{ follows (43) with } z_{|\zeta,\nu}(0) = \zeta, z_{|\zeta,\nu}(t) = z \right\}.$$

From [7], the cost-to-come is the solution in the viscosity sense of the following Hamilton-Jacobi-Bellman dynamics

$$\begin{cases} \partial_t \mathscr{V}(z,t) - \mathscr{H}(z, \nabla \mathscr{V}(z,t), t) = 0, & (z,t) \in \mathcal{Z} \times \mathbb{R}^+, \\ \mathscr{V}(z,0) = \mathscr{V}_0(z - \hat{z}_0), & z \in \mathcal{Z}, \end{cases} \tag{48}$$

where the Hamiltonian is given by

$$\mathscr{H}(z,q,t) = \frac{\delta^{-2}}{2} \|y^\delta(t) - h(z(t),t)\|^2 - \frac{\kappa^2}{2} (B(t)q, B(t)^* q)_{\mathcal{U}} - (q, f(z,t))_{\mathcal{Z}}.$$

Therefore, by computing the cost-to-come, we can compute the optimal observer for nonlinear systems from (46), and formally, if $\mathscr{V}$ is regular enough and has an invertible Hessian at any time, we retrieve an observer formulation [42], namely

$$\begin{cases} \dot{\hat{z}}(t) = f(\hat{z}(t),t) + (\nabla^2 \mathscr{V}(\hat{z}(t),t))^{-1} \big(y^\delta(t) - h(\hat{z}(t),t) \big), \quad t \geq 0, \\ \hat{z}(0) = \hat{z}_0 \end{cases} \tag{49}$$

This dynamics came from the generalization of the *decoupling principle* identity (17), given when $\mathscr{V}$ is regular enough, by $\bar{q}_T(t) = \nabla \mathscr{V}(\bar{z}_T(t), t)$, for all $0 \leq t \leq T$.

As the cost-to-come is solution in the viscosity sense of (48), one can also see it as a viscosity limit. This then allows an elegant parallel with stochastic filtering derived in [50, 7, 42] in the case where $f : z \mapsto f(z)$ and $h : z \mapsto h(z)$ do not depend on time and $B = \mathrm{Id}$, $\kappa = \delta = 1$. In this case, one can introduce a shifted functional $\mathscr{S}(z,t) = \mathscr{V}(z,t) - \int_0^t \frac{1}{2} \|y^\delta\|^2 \, \mathrm{d}s - h(z) \ell^\delta(t)$ where $\ell^\delta = \int_0^t y^\delta \, \mathrm{d}t$ is a primitive of y^δ. A simple computation gives that $\mathscr{S}$ is also solution of a Hamilton-Jacobi-Bellman equation

$$\begin{cases} \partial_t \mathscr{S}(z,t) - \mathscr{H}^s(z, \nabla \mathscr{S}(z,t), t) = 0, & (z,t) \in \mathcal{Z} \times \mathbb{R}^+, \\ \mathscr{S}(z,0) = \mathscr{V}_0(z - \hat{z}_0), & z \in \mathcal{Z}, \end{cases} \tag{50}$$

with the modified Hamiltonian $\mathscr{H}^s(z,q,t) = -\frac{1}{2}\|q\|_{\mathcal{Z}}^2 - (q, f(z,t) - \ell^\delta \nabla h(z))_{\mathcal{Z}} + \mathscr{P}(z,t)$ and $\mathscr{P}(z,t) = \frac{1}{2}\|h(z)\|_{\mathcal{Y}}^2 + (y^\delta, \mathrm{D}h(z) f(z)) - \frac{1}{2}\|y^\delta\|_{\mathcal{Y}}^2 \|\nabla h\|_{\mathcal{Y}}^2$. Moreover, $\mathscr{S}$ can be seen as the vanishing viscosity limit solution of

$$\begin{cases} \partial_t \mathscr{S}_\varepsilon(z,t) - \mathscr{H}_\varepsilon^s(z, \nabla \mathscr{S}_\varepsilon(z,t), t) - \frac{\epsilon}{2} \Delta \mathscr{S}_\epsilon = 0, & (z,t) \in \mathcal{Z} \times \mathbb{R}^+, \\ \mathscr{S}_\varepsilon(z,0) = \mathscr{V}_0(z - \hat{z}_0), & z \in \mathcal{Z}, \end{cases} \tag{51}$$

where $\mathscr{H}_\varepsilon^s(z,q,t) = \mathscr{H}(z,q,t) - \mathscr{P}^\varepsilon(z,t)$ and $\mathscr{P}^\varepsilon(z,t) = \mathscr{P}(z,t) + \varepsilon \nabla \cdot f(z) - \frac{\varepsilon}{2}(y^\delta(t), \Delta h(z))_{\mathcal{Z}}$.

The Hopf-Cole Transform of $\mathscr{S}^\varepsilon$, $\mu_\varepsilon(z,t) = \exp(-\varepsilon^{-1} \mathscr{S}^\varepsilon(z,t))$ is the solution of the robust form of the Zakai equation [93] associated with the stochastic filtering problem – see a detailed exposition in [77] – of the following stochastic process

$$\begin{cases} \mathrm{d}z_t = f(z(t))\mathrm{d}t + \sqrt{\varepsilon}\mathrm{d}b_t^\nu, \quad z_0 = \hat{z}_0 + \zeta, \\ \mathrm{d}\ell_t = h(z(t))\mathrm{d}t + \sqrt{\varepsilon}\mathrm{d}b_t^\eta, \quad \ell_t(0) = 0, \end{cases} \tag{52}$$

where now b_t^ν and b_t^η are independent Wiener processes, independent of the random variable ζ of unnormalized density proportional to $\mu_\epsilon(z,0)$. In this framework, ℓ^δ is particular sample trajectory of the stochastic process ℓ_t. Moreover, [7] justifies a large deviation principle showing that the Mortensen observer is the deterministic limit of the more general stochastic filtering problem associated with (52).

Note finally that, for linear operators $f(z,t) = -A(t)z + g$ and $h(z,t) = C(t)z$, the Mortensen estimator reduces to the Kalman estimator since we easily prove that the unique solution of (48) is given by

$$\mathscr{V}(z,t) = \frac{1}{2}\Big(z - \hat{z}(t), \Pi^{-1}(t)(z - \hat{z}(t)) \Big) + \int_0^t \frac{1}{2\delta^2} \|y^\delta(s) - C(s)\hat{z}(s)\|^2 \, \mathrm{d}s,$$

where Π is the solution of the Riccati dynamics (15) and $\hat{z}$ is the Kalman estimator. Moreover in this case, the large deviation principle being independent of ε, we retrieve the strict equivalence between the deterministic estimator and the stochastic estimator.

3.2 Discretization of the Mortensen filter

Following the principle of discretization, which relies on first discretizing and then controlling, [56, 72] explored the possibility of extending the discrete-time/continuous-time Kalman filter connection to nonlinear configurations with the discrete-time Mortensen filter, which reduces to the discrete-time Kalman filter when all operators are linear.

Let us then consider a discretization of (43) of the form

$$\begin{cases} z_{k+1}^\tau = \varphi_{k+1|k}^\tau(z_k^\tau) + B_{k+1}^\tau \nu_{k+1}^\tau, \quad 0 \leq k \leq n-1, \\ z_0^\tau = \hat{z}_0 + \zeta, \end{cases} \tag{53}$$

where $\varphi_{k+1|k}^\tau$ is a nonlinear transition map from step k to step $k+1$. The observations are also discretized with $y_n^\delta = h_n^\tau(z_n^\tau) + \eta_n^\tau$. Again, we consider the two least-squares functionals

$$\mathscr{J}_n^{\tau+}(\zeta, (\nu_k)_{k\leq n}) = \mathscr{V}_0(\zeta) + \frac{1}{2}\sum_{k=1}^{n} \kappa^{-2}\tau \|\nu_k^\tau\|_{\mathcal{U}}^2 + \frac{1}{2}\sum_{k=0}^{n} \tau\delta^{-2}\|y_k^\delta - h_k(z_{k|n}^\tau)\|_{\mathcal{Y}}^2,$$

and

$$\mathscr{J}_{n+1}^{\tau-}(\zeta, (\nu_k)_{k\leq n+1}) = \mathscr{V}_0(\zeta) + \frac{1}{2}\sum_{k=1}^{n+1} \kappa^{-2}\tau \|\nu_k^\tau\|_{\mathcal{U}}^2 + \frac{1}{2}\sum_{k=0}^{n} \tau\delta^{-2}\|y_k^\delta - h_k(z_{k|n}^\tau)\|_{\mathcal{Y}}^2,$$

from which we are going to define two costs-to-come. First, we introduce the pre-image set

$$\mathcal{A}_{n|m}^\tau(z) = \left\{ (\zeta, (\nu_k^\tau)_{m<k\leq n}) : (z_{k|n}^\tau)_{m\leq k\leq n} \text{ follows (53) with } z_{m|n}^\tau = \zeta, z_{n|n}^\tau = z \right\},$$

to then define

$$\begin{cases} \mathscr{V}_n^{\tau+}(z) = \min\limits_{(\zeta,(\nu_k)_{k\leq n})\in\mathcal{A}_{n|0}^\tau(z)} \mathscr{J}_n^{\tau+}(\zeta, (\nu_k^\tau)_{k\leq n}), & n \in \mathbb{N}, \\ \mathscr{V}_{n+1}^{\tau-}(z) = \min\limits_{(\zeta,(\nu_k)_{k\leq n+1})\in\mathcal{A}_{n+1|0}^\tau(z)} \mathscr{J}_{n+1}^{\tau-}(\zeta, (\nu_k^\tau)_{k\leq n+1}), & n \in \mathbb{N}, \\ \mathscr{V}_0^{\tau-}(z) = \mathscr{V}_0(z - \hat{z}_0). \end{cases} \tag{54}$$

These two costs-to-come are interconnected and it was shown in [56, 72] that

$$\forall n \geq 0, \quad \mathscr{V}_n^{\tau+}(z) = \mathscr{V}_n^{\tau-}(z) + \frac{\tau}{2\delta^2}\|y^\delta - h_n^\tau(z)\|_{\mathcal{Y}}^2,$$

whereas, from [56], $\mathscr{V}_{n+1}^{\tau-}(z)$ follows the Bellman equation

$$\forall n \geq 0, \quad \mathscr{V}_{n+1}^{\tau-}(z) = \min_{(\xi,\nu)\in\mathcal{A}_{n+1|n}^\tau(z)} \left\{ \mathscr{V}_n^{\tau+}(\xi) + \frac{\tau}{2\kappa^2}\|\nu\|^2 \right\}.$$

From the discrete-time evolution of these two costs-to-come, the prediction-step and correction-step of the discrete-time optimal filters are defined by

$$\hat{z}_n^{\tau+} = \operatorname{argmin} \mathscr{V}_n^{\tau+}(z) \text{ and } \hat{z}_n^{\tau-} = \operatorname{argmin} \mathscr{V}_n^{\tau-}(z), \tag{55}$$

and they can be computed from $\mathscr{V}_n^{\tau+}$ and $\mathscr{V}_n^{\tau-}$ with a prediction-correction approach [72], which reads when $\mathscr{V}_n^{\tau+} \in C^1(\mathcal{Z})$,

$$\begin{cases} \text{Initialization:} \\ \quad \hat{z}_0^{\tau-} = \hat{z}_0, \\ \text{Correction:} \\ \quad \nabla\mathscr{V}_n^{+}(\hat{z}_n^{\tau+}) = 0, \quad n \in \mathbb{N}, \\ \text{Prediction:} \\ \quad \hat{z}_{n+1}^{\tau-} = \varphi_{n+1|n}^\tau(\hat{z}_n^{\tau+}), \quad n \in \mathbb{N}, \end{cases} \tag{56}$$

which in fact generalizes to nonlinear dynamics the equation of the Kalman estimator in the deterministic setting.

Theorem 3.1. *Assuming that there exists a unique minimizer* $(\bar{\zeta}_n^{\tau-}, (\bar{\nu}_{k|n}^{\tau-})_{k\leq n}) = \operatorname{argmin} \mathscr{J}_n^{\tau-}$*, with associated trajectory* $(\bar{z}_{k|n}^{\tau-})_{0\leq k\leq n}$*, then the solution of* (56) *is an optimal observer in the sense that*

$$\hat{z}_n^{\tau-} = \bar{z}_{n|n}^{\tau-}.$$

The estimator (56) again has nice properties, since it is based on the optimality principle at the discrete level. Moreover, it has been proved in [72] that it is a consistent approximation of the continuous-time Mortensen dynamics. And as for stability, the cost-to-come can be used as a Lyapunov functional to obtain the error stabilization, which justifies some stability property at the discrete level.

In the literature, the Mortensen estimator, and hence its discretization, has not received as much attention as the Zakai equation in the stochastic context. This is understandable, since the stochastic context is certainly more general. After time discretization, i.e., for discrete-time systems in the nonlinear context where (53) is reformulated in a stochastic context with a Hidden Markov Model [92, 20, 58], the stochastic filtering approach is based on particle filters corresponding to Monte Carlo methods for solving nonlinear filtering problems [33, 26]. These particle methods are in turn based on a prediction-correction algorithm that allows the distributions $\pi_n^-(z_n|y_0, ..., y_{n-1})$ and $\pi_n^+(z_n|y_0, ..., y_n)$ to be sampled. To the author's knowledge, it is still not established that, at the discrete-time level, a large deviation principle is associated with the discrete-time cost-to-come $\mathscr{V}_n^+$ and $\mathscr{V}_{n+1}^-$, as this has been done for continuous-time systems in [50, 7]. However, we believe that such a result may follow a similar path of proof and ultimately link particle filters to the discrete-time Mortensen filter.

3.3 Approximated optimal approaches and discretization

The Mortensen estimator sheds light on the classical approximated optimal approaches used in practice, which can be reinterpreted as ways to avoid calculating the cost-to-come. This is especially true for the continuous-time Extended Kalman Filter [7] and its corresponding discrete-time Extended Kalman Filter [86]. At the continuous-time level, the EKF observer reads

$$\begin{cases} \dot{\hat{z}} = f(\hat{z}(t), t) + \mathit{\Pi} \mathrm{D}h(\hat{z}(t), t)^* R(y^\delta(t) - h(\hat{z}(t), t)), & \text{in } (0, T), \\ \hat{z}(0) = \hat{z}_0 \end{cases} \tag{57}$$

where $\mathit{\Pi}$ is given by

$$\begin{cases} \dot{\mathit{\Pi}} - \mathrm{D}f(\hat{z}, t)\mathit{\Pi} - \mathit{\Pi}\mathrm{D}f(\hat{z}, t)^* + \mathit{\Pi}\mathrm{D}h(\hat{z}, t)^* R\, \mathrm{D}h(\hat{z}, t)\mathit{\Pi} - B(t)QB(t)^* = 0, & \text{in } (0, T), \\ \mathit{\Pi}(0) = \mathit{\Pi}_0. \end{cases}$$

Therefore, we understand from (49) that $\mathit{\Pi}$ plays the role of an approximation of the inverse of the Hessian – if the Hessian exists and is invertible – of $\mathscr{V}(\hat{z}, t)$. More precisely, the Riccati solution gives a local approximation of the cost-to-come as a quadratic functional. Such an approximation was mathematically justified for a bilinear optimal control problem [15]. The observer problem should deserve a similar study.

At the discrete-time level, the discretization-then-control approach leads to the same kind of approximation. The discrete-time Extended Kalman Filter follows the now classical prediction-correction splitting scheme with

$$\begin{cases} \hat{z}_0^{\tau-} = \hat{z}_0, & & \text{(58a)} \\ \hat{z}_n^{\tau+} = \hat{z}_n^{\tau-} + \mathit{\Pi}_n^{\tau+} \mathrm{D}h_n^{\tau*}(\hat{z}_n^{\tau-}) R^\tau \big(y_n^\delta - h_n^\tau(\hat{z}_n^{\tau-})\big), & n \in \mathbb{N}, & \text{(58b)} \\ \hat{z}_{n+1}^{\tau-} = \varphi_{n+1|n}^\tau(\hat{z}_n^{\tau+}), & n \in \mathbb{N}, & \text{(58c)} \end{cases}$$

while

$$\begin{cases} \mathit{\Pi}_0^{\tau-} = \mathit{\Pi}_0, & & \text{(59a)} \\ \mathit{\Pi}_n^{\tau+} = \big[(\mathit{\Pi}_n^{\tau-})^{-1} + \mathrm{D}h_n^{\tau*}(\hat{z}_n^{\tau-}) R^\tau \mathrm{D}h_n^\tau(\hat{z}_n^{\tau-})\big]^{-1}, & n \in \mathbb{N}, & \text{(59b)} \\ \mathit{\Pi}_{n+1}^{\tau-} = \mathrm{D}\varphi_{n+1|n}^\tau(\hat{z}_n^{\tau+}) \mathit{\Pi}_n^{\tau+} \mathrm{D}\varphi_{n+1|n}^\tau(\hat{z}_n^{\tau+})^* + B_{n+1}^\tau Q^\tau B_{n+1}^{\tau*}, & n \in \mathbb{N}. & \text{(59c)} \end{cases}$$

Again $\mathit{\Pi}_n^{\tau-}$ can be seen as an approximation of the Hessian of $\mathscr{V}_n^{\tau-}$ while $\mathit{\Pi}_n^{\tau+}$ can be seen as an approximation of the Hessian of $\mathscr{V}_n^{\tau+}$. This is particularly striking when the correction step (56) is solved by a Newton algorithm with, until convergence, the successive iterations

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

Seeing that

$$\nabla \mathscr{V}_n^{\tau+}(\hat{z}_{n,0}^{\tau+}) = \mathrm{D}h_n^{\tau}(\hat{z}_n^{\tau-})^* R_n^{\tau}\big(y_n^{\delta} - h_n^{\tau}(\hat{z}_n^{\tau-})\big),$$

we have for the discrete-time Mortensen filter, after one iteration of the Newton-Raphson procedure,

$$\hat{z}_{n,1}^{\tau+} = \hat{z}_n^{\tau-} - (\nabla^2 \mathscr{V}_n^{+}(\hat{z}_n^{\tau-}))^{-1} \mathrm{D}h_n^{\tau}(\hat{z}_n^{\tau-})^* R_n^{\tau}\big(y_n^{\delta} - h_n^{\tau}(\hat{z}_n^{\tau-})\big), \tag{60}$$

a structure very similar to the EKF correction. Note that high-order approaches for non-linear Kalman filtering have also been proposed [85, Section 13.3] and it would be interesting to also compare them with successive iterations of such a Newton-Raphson procedure.

By comparison with EKF, the Unscented Kalman Filter (UKF) [52] has the same objective but uses a finite difference stencil to avoid computing the tangent $\mathrm{D}\varphi_{n+1|n}^{\tau}(\hat{z}_n^{\tau+})$ and $\mathrm{D}h^{\tau}(\hat{z}_n^{\tau-})$. The idea behind UKF is to assume a decomposition of the covariance over the so-called sigma points $(e^{(i)})_{1\leq i\leq N_s} \in \mathcal{Z}^{N_s}$, with associated weights $(\omega_i)_{1\leq i\leq N_s} \in \mathbb{R}_s^N$, satisfying

$$\sum_{1\leq i\leq N_s} \omega_i e^{(i)} = 0, \qquad \sum_{1\leq i\leq N_s} \omega_i e^{(i)} \otimes e^{(i)} = \mathrm{Id}_{\mathcal{Z}}.$$

Therefore, by defining $v_n^{(i)+} = \sqrt{\mathit{\Pi}_n^{\tau+}} e^{(i)}$, we have $\mathit{\Pi}_n^{\tau+} = \sum_{1\leq i\leq N_s} \omega_i v^{(i)} \otimes v^{(i)}$. In the prediction step approximation (59c), we then recognize

$$\begin{aligned} &\mathrm{D}\varphi_{n+1|n}^{\tau}(\hat{z}_n^{\tau+}) \mathit{\Pi}_n^{\tau+} \mathrm{D}\varphi_{n+1|n}^{\tau}(\hat{z}_n^{\tau+})^* \\ &\qquad \simeq \sum_{1\leq i\leq N_s} \omega_i [\varphi_{n+1|n}^{\tau}(\hat{z}_n^{\tau+} + v_n^{(i)+}] - \varphi_{n+1|n}^{\tau}(\hat{z}_n^{\tau+})) \otimes [\varphi_{n+1|n}^{\tau}(\hat{z}_n^{\tau+} + v_n^{(i)+}) - \varphi_{n+1|n}^{\tau}(\hat{z}_n^{\tau+})]. \end{aligned}$$

Identically by using Woodbury formula, we develop the correction step (59b) in the form of (27), namely

$$\mathit{\Pi}_n^{\tau+} = \mathit{\Pi}_n^{\tau-} - G_n^{\tau}\big[\mathrm{D}h_n^{\tau}(\hat{z}_n^{\tau-}) \mathit{\Pi}_n^{\tau-} \mathrm{D}h_n^{\tau}(\hat{z}_n^{\tau-})^* + W_n^{\tau}\big] G_n^{\tau *},$$

with $G_n^{\tau} = \mathit{\Pi}_n^{\tau-} \mathrm{D}h_n^{\tau}(\hat{z}_n^{\tau-})^* \left(\mathrm{D}h_n^{\tau}(\hat{z}_n^{\tau-}) \mathit{\Pi}_n^{\tau-} \mathrm{D}h_n^{\tau}(\hat{z}_n^{\tau-})^* + W_n^{\tau}\right)^{-1}$. Therefore with $v_n^{(i)+} = \sqrt{\mathit{\Pi}_n^{\tau-}} e^{(i)}$, we replace

$$\mathrm{D}h_n^{\tau}(\hat{z}_n^{\tau-}) \mathit{\Pi}_n^{\tau-} \mathrm{D}h_n^{\tau}(\hat{z}_n^{\tau-})^* \simeq \sum_{1\leq i\leq N_s} \omega_i [h_n^{\tau}(\hat{z}_n^{\tau-} + v_n^{(i)+}) - h_n^{\tau}(\hat{z}_n^{\tau-})] \otimes [h_n^{\tau}(\hat{z}_n^{\tau-} + v_n^{(i)+}) - h_n^{\tau}(\hat{z}_n^{\tau-})],$$

while

$$\mathit{\Pi}_n^{\tau-} \mathrm{D}h_n^{\tau}(\hat{z}_n^{-})^* \simeq \sum_{1\leq i\leq N_s} \omega_i v_n^{(i)+} \otimes [h_n^{\tau}(\hat{z}_n^{-} + v_n^{(i)+}) - h_n^{\tau}(\hat{z}_n^{-})].$$

In short, the UKF estimator is a discrete-time filter that can be understood as a tangent-free approximation of the discrete-time Mortensen estimator, which is itself a discretization of the continuous-time Mortensen estimator by a splitting algorithm. Note also that the UKF was originally proposed with a stochastic perspective [52], of criticized mathematical foundations [63]. We believe that the vision of UKF as an approximation to the deterministic discrete-time Mortensen estimator overcomes these criticisms.

An important alternative to EKF and UKF is the Ensemble Kalman Filter (EnKF) [38], which also has a discrete-time version with a prediction-correction form. There are intensive efforts to understand EnKF as a mean-field approximation of the particle filter at the continuous-time level [45] and at the discrete-time level [62], justifying the stochastic flavor of the Ensemble Kalman Filter. Unifying the mean-field vision from the continuous-time setting up to the discrete-time setting will finalize the overall understanding and positioning of all these filters together.

4 Conclusion

We here reviewed some of the most well-known data assimilation approaches in their continuous-time formulation and their discrete-time counterparts. We have seen that the discrete-time methods, applied to a time discretization of a continuous-time model, are a very powerful strategy to discretize in time the continuous-time data assimilation formulations. This paradigm is summarized in Figure 1 and is valid for a large class of problems from infinite-dimensional linear problems to nonlinear finite dimensional systems. For sequential approaches, we end up in each case with the same very general structure of the time scheme with a prediction-correction form. In essence, this justifies the well-known data assimilation recipe: First predict – or forecast – using the discretized model and then correct – or update or analyze – by incorporating the new available observations.

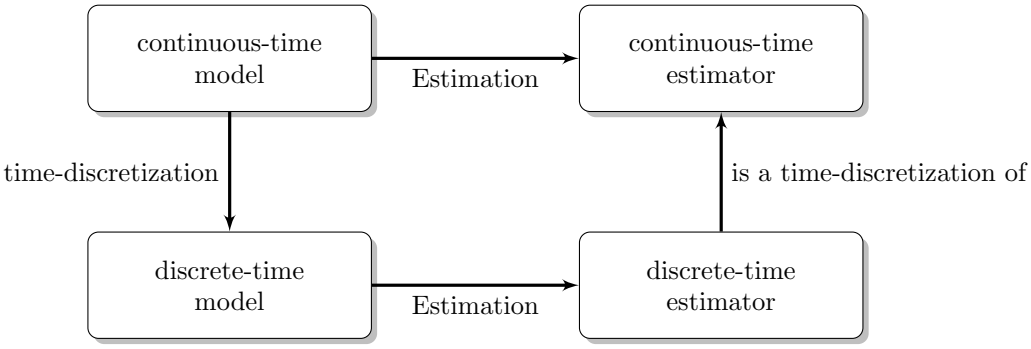

Figure 1: The discretization-then-control strategy for time-discretizing data assimilation problem

References

[1] A. Aalto. Convergence of discrete-time Kalman filter estimate to continuous time estimate. *International Journal of Control*, 89(4):668–679, 2014.

[2] A. Aalto. Convergence of discrete-time Kalman filter estimate to continuous-time estimate for systems with unbounded observation. *Mathematics of Control, Signals, and Systems*, 30(2):9, 2018.

[3] M. Asch, M. Bocquet, and M. Nodet. *Data assimilation: methods, algorithms, and applications*. SIAM, 2016.

[4] D. Auroux and J. Blum. A nudging-based data assimilation method: the Back and Forth Nudging (BFN) algorithm. *Nonlinear Processes In Geophysics*, 15(2):305–319, 2008.

[5] M. Aussal and P. Moireau. Kernel representation of Kalman observer and associated H-matrix based discretization. hal-03658937.

[6] H. T. Banks, K. Ito, and C. Wang. Exponentially stable approximations of weakly damped wave equations. volume 100 of *Estimation and control of distributed parameter systems (Vorau, 1990)*, pages 1 – 33. Birkhäuser Basel, 1991.

[7] J. S. Baras, A. Bensoussan, and M. R. James. Dynamic observers as asymptotic limits of recursive filters: special cases. *SIAM Journal of Applied Mathematics*, 48(5):1147–1158, 1988.

[8] K. Batselier, Z. Chen, and N. Wong. A tensor network Kalman filter with an application in recursive MIMO Volterra system identification. *Automatica*, 84:17–25, 2017.

[9] P. Benner and H. Mena. Numerical solution of the infinite-dimensional LQR problem and the associated Riccati differential equations. *Journal of Numerical Mathematics*, 26(1):1–20, 2018.

[10] P. Benner, M. Ohlberger, A. Cohen, and K. Willcox. *Model reduction and approximation: theory and algorithms*. SIAM, 2017.

[11] P. Benner and J. Saak. Numerical solution of large and sparse continuous time algebraic matrix Riccati and Lyapunov equations: a state of the art survey. *GAMM-Mitteilungen*, 36(1):32–52, 2013.

[12] A. Bensoussan. *Filtrage optimal des systèmes linéaires*. Dunod, 1971.

[13] A. Bensoussan. *Estimation and Control of Dynamical Systems*. Interdisciplinary Applied Mathematics. Springer, Cham, 2018.

[14] A. Bensoussan, M. C. Delfour, G. Da Prato, and S. K. Mitter. *Representation and Control of Infinite Dimensional Systems*. Birkhauser Verlag, Boston, 2nd edition, 2007.

[15] T. Breiten, K. Kunisch, and L. Pfeiffer. Taylor expansions of the value function associated with a bilinear optimal control problem. *Annales de l'Institut Henri Poincaré C, Analyse non linéaire*, 2019.

[16] E. Burman and L. Oksanen. Data assimilation for the heat equation using stabilized finite element methods. *Numerische Mathematik*, 139(3):505–528, February 2018.

[17] J. A. Burns and C. N. Rautenberg. Solutions and approximations to the Riccati integral equation with values in a space of compact operators. *SIAM Journal on Control and Optimization*, 53(5):2846–2877, 2015.

[18] J. A. Burns and C. N. Rautenberg. The infinite-dimensional optimal filtering problem with mobile and stationary sensor networks. *Numerical Functional Analysis and Optimization*, 36(2):181–224, 2015.

[19] G. Buttazzo and G. Dal Maso. Γ-convergence and optimal control problems. *Journal of optimization theory and applications*, 38(3):385–407, 1982.

[20] O. Cappé, E. Moulines, and T. Rydén. *Inference in Hidden Markov Models*. Springer Series in Statistics. Springer, 2005.

[21] D. Chapelle, N. Cîndea, M. de Buhan, and P. Moireau. Exponential convergence of an observer based on partial field measurements for the wave equation. *Mathematical Problems in Engineering*, 2012:12, October 2012.

[22] D. Chapelle, N. Cîndea, and P. Moireau. Improving Convergence in Numerical Analysis Using Observers — the Wave-like Equation Case. *Mathematical Models and Methods in Applied Sciences*, 22(12):1250040, 2012.

[23] 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. *Medical & Biological Eng & Computing*, 51(11):1221–1233, 2012.

[24] 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: Mathematical Modelling and Numerical Analysis*, 2013.

[25] G. Chavent. *Nonlinear Least Squares for Inverse Problems*. Springer, 2010.

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

[27] N. Cîndea, A. Imperiale, and P. Moireau. Data assimilation of time under-sampled measurements using observers, the wave-like equation example. *ESAIM: Control, Optimisation and Calculus of Variation*, 21(3):635–669, 2015.

[28] N. Cîndea and A. Münch. Inverse problems for linear hyperbolic equations using mixed formulations. *Inverse Problems*, 31(7):075001, 2015.

[29] C. Corrado, J.-F. Gerbeau, and P. Moireau. Identification of weakly coupled multiphysics problems. Application to the inverse problem of electrocardiography. *Journal Of Computational Physics*, 283(C):271–298, 2015.

[30] R. F. Curtain. A survey of infinite-dimensional filtering. *SIAM Review*, 17:395–411, 1975.

[31] R. F. Curtain. Infinite-dimensional filtering. *SIAM Journal on Control and Optimization*, 13:89–104, 1975.

[32] R. F. Curtain, K. Mikkola, and A. Sasane. The Hilbert-Schmidt property of feedback operators. *Journal of Mathematical Analysis and Applications*, 329(2):1145 – 1160, 2007.

[33] P. Del Moral. Nonlinear filtering: Interacting particle resolution. *Comptes Rendus de l'Académie des Sciences-Series I-Mathematics*, 325(6):653–658, 1997.

[34] M. A. Demetriou and I. G. Rosen. Adaptive identification of second-order distributed parameter systems. *Inverse Problems*, 10(2):261–294, 1994.

[35] M. A. Demetriou and I. G. Rosen. On the persistence of excitation in the adaptive estimation of distributed parameter systems. *IEEE Transactions on Automatic Control*, 39(5):1117–1123, 1994.

[36] S. Ervedoza, C. Zheng, and E. Zuazua. On the observability of time-discrete conservative linear systems. *Journal Of Functional Analysis*, 254(12):3037 – 3078, 2008.

[37] S. Ervedoza and E. Zuazua. Uniformly exponentially stable approximations for a class of damped systems. *Journal de Mathématiques Pures et Appliquées*, 91(1):20–48, 2009.

[38] G. Evensen. *Data Assimilation – The Ensemble Kalman Filter*. Springer Verlag, 2007.

[39] G. Evensen, J. Amezcua, M. Bocquet, A. Carrassi, A. Farchi, A. Fowler, P. L. Houtekamer, Jones. C. K., R. J. de Moraes, M. Pulido, C. Sampson, and F. C. Vossepoel. An international initiative of predicting the SARS-CoV-2 pandemic using ensemble data assimilation. *Foundations of Data Science*, 3(3):413–477, 2021.

[40] P. L. Falb. Infinite-dimensional filtering: the Kalman-Bucy filter in Hilbert space. *Information and Control*, 11:102–137, 1967.

[41] F. Flandoli. On the semigroup approach to stochastic evolution equations. *Stochastic Analysis and Applications*, 10(2):181–203, 1992.

[42] W.H. Fleming. Deterministic nonlinear filtering. *The Annali della Scuola Normale Superiore di Pisa, Classe di Scienze*, 25(3-4):435–454, 1997.

[43] W.H. Fleming and R.W. Rischel. *Deterministic and Stochastic Optimal Control*. Springer-Verlag, 1975.

[44] H. Fujita, N. Saito, and T. Suzuki. *Operator Theory and Numerical Methods*. Studies in Mathematics and Its Applications. North Holland, 2001.

[45] A. Garbuno-Inigo, F. Hoffmann, W. Li, and A. M. Stuart. Interacting Langevin diffusions: Gradient structure and Ensemble Kalman sampler. *SIAM Journal on Applied Dynamical Systems*, 19(1):412–441, 2020.

[46] A. Germani, L. Jetto, and M. Piccioni. Galerkin approximation for optimal linear filtering of infinite-dimensional linear systems. *SIAM journal on control and optimization*, 26(6):1287–1305, 1988.

[47] R. Glowinski, W. Kinton, and M. F. Wheeler. A mixed finite element formulation for the boundary controllability of the wave equation. *International Journal for Numerical Methods in Engineering*, 27(3):623–635, 1989.

[48] S. Gratton, A. S. Lawless, and N. K. Nichols. Approximate Gauss–Newton methods for nonlinear least squares problems. *SIAM Journal of Control and Optimization*, 18(1), 2007.

[49] A. Haraux. Une remarque sur la stabilisation de certains systèmes du deuxième ordre en temps. *Portugal. Math.*, 46(3):245–258, 1989.

[50] O. Hijab. Asymptotic nonlinear filtering and large deviations. *Advances in Filtering and Optimal Stochastic Control*, pages 170–176, 1982.

[51] A. Imperiale, D. Chapelle, and P. Moireau. Sequential data assimilation for mechanical systems with complex image data: application to tagged-MRI in cardiac mechanics. *Advanced Modeling and Simulation in Engineering Sciences*, 8, 2021.

[52] S. J. Julier, J. K. Uhlmann, and H. F. Durrant-Whyte. A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Transactions on Automatic Control*, 45(3):477–482, 2000.

[53] R. E. Kalman and R. Bucy. New results in linear filtering and prediction theory. *Journal of Basic Engineering*, 83:95–108, 1961.

[54] R.E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 1960.

[55] A. J. Krener. A Lyapunov theory of nonlinear observers. In G. G. Yin and Q. Zhang, editors, *Stochastic analysis, control, optimization and applications*, pages 409–420. Springer, 1998.

[56] A. J. Krener. Minimum energy estimation and moving horizon estimation. *54th IEEE Conference on Decision and Control (CDC)*, pages 4952–4957, 2015.

[57] I. Lasiecka and R. Triggiani. *Differential and algebraic Riccati equations with application to boundary/point control problems: Continuous theory and approximation theory.* Lecture Notes in Control and Information Sciences. Springer Berlin Heidelberg, 1991.

[58] K. Law, A. Stuart, and K. Zygalakis. *Data assimilation: A mathematical introduction*, volume 62 of *Texts in Applied Mathematics*. Springer, Cham, 2015.

[59] A. S. Lawless, S. Gratton, and N. K. Nichols. An investigation of incremental 4D-Var using non-tangent linear models. *Quarterly Journal of the Royal Meteorological Society*, 131(606):459–476, 2005.

[60] F.-X. Le Dimet. Optimal control for data assimilation in meteorology. In *Control theory of distributed parameter systems and applications (Shanghai, 1990)*, volume 159 of *Lecture Notes in Control and Inform. Sci.*, pages 51–60. Springer, Berlin, 1991.

[61] F.-X. Le Dimet and O. Talagrand. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A*, 38(2):97–110, 2010.

[62] Fr. Le Gland, V. Monbet, and V.-D. Tran. *Large sample asymptotics for the ensemble Kalman filter*, chapter 22, pages 598–631. Number RR-7014. Oxford University Press, 2011.

[63] T. Lefebvre, H. Bruyninckx, and J. de Schuller. Comment on "A new method for the nonlinear transformation of means and covariances in filters and estimators" [and authors' reply]. *IEEE Transactions on Automatic Control*, 47(8):1406–1409, 2002.

[64] J. Y. Li, S. Ambikasaran, E. F Darve, and P. K. Kitanidis. A Kalman filter powered by H2-matrices for quasi-continuous data assimilation problems. *Water Resources Research*, 50(5):3734–3749, May 2014.

[65] J.-L. Lions. *Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles.* Avant propos de P. Lelong. Dunod, Paris, 1968.

[66] J.-L. Lions and E. Magenes. *Non-Homogeneous Boundary Value Problems and Applications*, volume 1. Springer-Verlag, 1972.

[67] D. Lombardi. State estimation in nonlinear parametric time dependent systems using tensor train. *International Journal for Numerical Methods in Engineering*, 2022.

[68] D. G. Luenberger. An introduction to observers. *IEEE Transactions on Automatic Control*, 16:596–602, 1971.

[69] D. G. Luenberger and Y. Ye. *Linear and nonlinear programming*. Springer, 2011.

[70] Y. Maday and O. Mula. A generalized empirical interpolation method: application of reduced basis techniques to data assimilation. In F. Brezzi, P. Colli Franzone, U. Gianazza, and G. Gilardi, editors, *Springer INdAM Series 4*, pages 221–235, 2013.

[71] Y. Maday and A. T. Patera. Reduced basis methods. In Grivet-Talocia S. Quarteroni A. Rozza G. Schilders W. Silveira L. Benner, P., editor, *Model Order Reduction*, pages 139–179, 2020.

[72] P. Moireau. A discrete-time optimal filtering approach for non-linear systems as a stable discretization of the Mortensen observer. *ESAIM: Control, Optimisation and Calculus of Variations*, 24(4):1815–1847, 2018.

[73] P. Moireau, D. Chapelle, and P. Le Tallec. Joint state and parameter estimation for distributed mechanical systems. *Comp. Meth. In App. Mech. And Eng.*, 197(6-8):659 – 677, 2008.

[74] R. E. Mortensen. Maximum-likelihood recursive nonlinear filtering. *J. Optim. Theory Appl.*, 2(6):386–394, 1968.

[75] I. M. Navon. Data assimilation for numerical weather prediction: A review. In Seon K Park and Liang Xu, editors, *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[76] S. Pagani, A. Manzoni, and A. Quarteroni. Efficient State/Parameter Estimation in Nonlinear Unsteady PDEs by a Reduced Basis Ensemble Kalman Filter. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):890–921, 2017-01.

[77] E. Pardoux. Stochastic partial differential equations and filtering of diffusion processes. *Stochastics*, 3(1-4):127–167, 1980.

[78] A. Pazy. *Semigroups of Linear Operators and Applications to Partial Differential Equations*, volume 44 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1983.

[79] D. T. Pham, J. Verron, and L. Gourdeau. Filtres de Kalman singuliers évolutifs pour l'assimilation de données en océanographie. *Comptes Rendus de l'Académie des Sciences-Series IIA-Earth and Planetary Science*, 326(4):255 – 260, 1998.

[80] D. T. Pham, J. Verron, and M. C. Roubaud. A singular evolutive extended Kalman filter for data assimilation in oceanography. *Journal of Marine systems*, 16(3-4):323–340, 1998.

[81] K. Ramdani, M. Tucsnak, and G. Weiss. Recovering the initial state of an infinite-dimensional system using observers. *Automatica*, 46(10):1616–1625, 2010.

[82] I. G. Rosen. Convergence of Galerkin approximations for operator Riccati equations – A nonlinear evolution equation approach. *Journal of Mathematical Analysis and Applications*, 155(1):226–248, 11 1991.

[83] P.-B. Rubio, L. Chamoin, and F. Louf. Real-time data assimilation and control on mechanical systems under uncertainties. *Advanced Modeling and Simulation in Engineering Sciences*, 8(1):4, 2021.

[84] M. Salgado, R. Middleton, and G. C. Goodwin. Connection between continuous and discrete Riccati equations with applications to Kalman filtering. *IEE Proceedings D (Control Theory and Applications)*, 135(1):28, 1988.

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

[86] Y. Song and J. W. Grizzle. The extended Kalman filter as a local asymptotic observer for nonlinear discrete-time systems. *Journal of Mathematical Systems, Estimation and Control*, 5(1):59 – 78, 1995.

[87] H. Tanabe. *Equations of evolution*, volume 6 of *Monographs and Studies in Mathematics*. Pitman (Advanced Publishing Program), Boston, Mass.-London, 1979.

[88] L. T. Tebou and E. Zuazua. Uniform boundary stabilization of the finite difference space discretization of the 1d wave equation. *Advances In Computational Mathematics*, 26(1-3):337–365, 2007.

[89] R. Temam. Sur l'équation de Riccati associée à des opérateurs non bornés, en dimension infinie. *Journal Of Functional Analysis*, 7:85–115, 1971.

[90] M. Tucsnak and G. Weiss. *Observation and control for operator semigroups*. Birkhäuser Advanced Texts: Basler Lehrbücher. Birkhäuser Verlag, Basel, 2009.

[91] J C Willems. Deterministic least squares filtering. *Journal of Econometrics*, 118(1-2):341–373, 2004.

[92] J. Xiong. *An introduction to stochastic filtering theory*, volume 18. OUP Oxford, 2008.

[93] M. Zakai. On the optimal filtering of diffusion processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 11(3):230–243, 1969.

[94] Q. Zhang. Adaptive observer for multiple-input-multiple-output (MIMO) linear time-varying systems. *IEEE Transactions on Automatic Control*, 47(3):525–529, 2002.

[95] X. Zhang, C. Zheng, and E. Zuazua. Exact controllability of the time discrete wave equation: A multiplier approach. volume 15 of *Applied and Numerical Partial Differential Equations*, pages 229–245. Springer Netherlands, 2009.

[96] E. Zuazua. Propagation, Observation, and Control of Waves Approximated by Finite Difference Methods. *SIAM Review*, 47(2):197–243, 2005.