DYNAMIC OUTPUT-FEEDBACK STABILIZATION OF UNCERTAIN LINEAR DYNAMICS VIA DIGITAL TWINS

PHILIPP A. GUTH[1], KARL KUNISCH[1,2], SÉRGIO S. RODRIGUES[3], AND JESPER SCHRÖDER[1]

Abstract. This work presents a digital twin framework for output-feedback stabilization and parameter identification in uncertain dynamical systems. A virtual model evolves in parallel with the physical process, assimilating measurement data in real time. By design, the digital twin reconstructs the system state and generates a stabilizing feedback, while model parameters are simultaneously inferred from data of the controlled dynamics using a Bayesian approach. Numerical results for the coupled physical-virtual dynamics demonstrate how digital twins can act jointly as observers, parameter estimators, and control agents, ensuring robust performance under uncertainty.

1. Introduction

Digital twins (DT) have become a useful tool in a multitude of applied sciences, including engineering, healthcare, supply chains, and environmental systems, for example. While important and practical experience has been obtained and documented, the mathematical analysis poses significant challenges to be overcome. In this research we take a step in this direction. To commence, let us specify the notion of DT that we shall follow.

> *A DT is a coupled system consisting of a set of virtual information constructs, as for instance an underdetermined dynamical system, that describes the structure and behavior of the physical system, from which it is updated by data as time progresses. The DT has the capability to influence the physical system to achieve predefined objectives. A bidirectional interaction between the virtual twin and the physical twin is central to the digital twin.*

This is a slightly modified description of a DT guided by the definition given by the National Academies of Science, Engineering, and Medicine [14], which itself is a slightly

---

MSC2020: 93C40, 93B52, 49N10, 93B51.

Keywords: model parameter uncertainty, feedback adaptive control, stabilization, continuous data assimilation, adaptive observer design, digital twin.

[1] Johann Radon Institute for Computational and Applied Mathematics, ÖAW, Altenbergerstrasse 69, 4040 Linz, Austria.

[2] Institute of Mathematics and Scientific Computing, Karl-Franzens University of Graz, Heinrichstrasse 36, 8010 Graz, Austria, and Johann Radon Institute for Computational and Applied Mathematics, ÖAW, Altenbergerstrasse 69, 4040 Linz, Austria.

[3] Departament of Mathematics, FCT, University NOVA of Lisbon, 2829-516 Caparica, Portugal

Emails: philipp.guth@ricam.oeaw.ac.at, karl.kunisch@uni-graz.at, ssi.rodrigues@fct.unl.pt, jesper.schroeder@ricam.oeaw.ac.at .

altered and extended form of an earlier definition by the American Institute of Aeronautics and Astronautics [1]. For an interesting introduction to mathematical aspects of DTs we also refer to [4].

For the successful analysis of DTs, many techniques of mathematical systems theory come into play. These include parameter estimation, inverse problems, uncertainty quantification, feedback and possibly optimal feedback control, state estimation, and reduced order modeling. Due to the coupling between the virtual and the physical twin, these methods must be employed in an interconnected manner, which implies a need for the development of new concepts. We attempt to go a step in this direction. There are many other important applied and numerical challenges which arise in the context of DTs, including large data sets, the curse of dimensionality, and multiscale aspects. These topics are not part of the present work.

In the development of digital twins, the nature of the underlying problem typically dictates whether rich data sets or highly-accurate physical models are available. Rarely, both are available simultaneously. While data-driven approaches leverage large datasets to infer system behavior, model-based strategies capitalize on detailed knowledge of the underlying physics and therefore remain reliable even when measurements are sparse, noisy, or only indirectly related to quantities of interest. In this work, we operate on the model-driven end of the spectrum, where the governing equations provide the structural backbone of the digital twin and the available measurements are used to calibrate the model within a holistic framework for prediction and control that allows for uncertainty quantification and ultimately leads to guaranteed stabilization of the physical system.

In this manuscript, the main purpose of the design of the DT is oriented towards the construction of controls stabilizing the physical system. In the context of uncertain dynamics, this requires the construction of the virtual twin to take into account the nontrivial task of online identification of appropriate estimates for unknown parameters of the dynamics. Since the data obtained from the physical twin only partially describe its state, state reconstruction has to be taken into consideration as well.

The literature on controlling dynamical systems under parametric uncertainty is as rich as it is diverse. The following overview does not claim completeness and is to be understood as a starting point for the interested reader. A broad overview from a system theoretic standpoint at the turn of the century is presented in [5]. This includes, in particular, *self-tuning regulators* [5, Ch. 3 and 4], based on the principle of estimating uncertain parameters and adjusting the controller as if these parameters were known. This approach, which is referred to as the *certainty equivalence principle*, does not take into account the uncertainties of the estimated parameters. For more recent accounts of adaptive control, we refer to [3, 10], for example. A conceptually different strategy is given by a stochastic formulation where controls are chosen according to a cost functional including a control goal and the parameter uncertainty. This leads to the concept of *dual control* [5, Ch. 7], where optimal controls strike a balance between achieving the desired behavior and properly identifying parameters. We refer to [12] for a survey on dual control and to [20] proposing an explicit construction of controls sequentially exploring parameters and controlling the system towards a desired behavior. In [7] the authors present an MPC-based strategy combining control, state estimation, and parameter estimation. Another approach is given by *gain scheduling* [5, Ch. 9], where controllers are adapted according to measurements of the operating environment, see also [19] for a more recent reference relying on this concept. Finally, we mention [18] presenting a fully Bayesian approach utilizing the Ensemble Kalman filter.

This work presents a novel design for simultaneous parameter identification and system stabilization. Given flawed information on the initial state and partial, noisy measurements of the evolving state of the physical twin, the virtual twin is updated iteratively using Bayesian inference. This identification procedure is carried out in conjunction with feedback control and state estimation which are performed continuously using Riccati gains based on the current parameter estimate of the virtual twin.

The paper is structured by the following sections. Section 2 provides a detailed description of our problem setting. The asymptotic behavior of the coupled physical-virtual twin system without parameter updates is investigated in Section 3. The Bayesian update strategy for the parameters of the virtual twin is described in Section 4. Section 5 explains the Kalman-filter estimation strategy for the coupled twin state. The final Section 6 is devoted to the investigation of the proposed strategies in numerical practice.

2. Design of the digital twin

Suppose a real-world phenomenon evolves, for time $t > 0$, according to

$$\tfrac{\mathrm{d}}{\mathrm{d}t} y(t) = A_\sigma y(t) + Bu(t), \quad \text{in } \mathbb{R}^n, \tag{2.1}$$

with unknown initial condition $y(0) = y_0$. Motivated by the fact that modeling errors are ubiquitous in applications, we consider an uncertain parameter $\sigma \in \mathbb{R}^p$ influencing the dynamics via A_σ. The linear operator A_σ defining the free dynamics is independent of time and depends continuously on σ. The control input is $u(t) \in \mathbb{R}^m$, and the linear control operator $B\colon \mathbb{R}^m \to \mathbb{R}^n$ is independent of time. Furthermore, we collect noisy output data according to

$$z(t_j) = Cy(t_j) + \eta_j, \quad \text{where} \quad \eta_j \sim \mathcal{N}(0, \Gamma_j), \quad t_j \in (t_0, \dots, t_s), \tag{2.2}$$

where the noise η_j, $j = 0, \dots, s$ is independently normally distributed with symmetric positive definite covariance matrix Γ_j. At every measurement point t_j, the vector z represents the output of ℓ sensor measurements, with linear output operator $C\colon \mathbb{R}^n \to \mathbb{R}^\ell$. Above, n, m, and ℓ are fixed positive integers.

Our goal is to find an input feedback operator $K_{\widehat{\sigma}}$ and take the input $u(t) = K_{\widehat{\sigma}} \widehat{y}(t)$, where $\widehat{y}(t)$ is an estimate for $y(t)$ and $\widehat{\sigma}$ is an estimate for σ, which is updated as time progresses, so that the solution of the real-world system (2.1)

$$\dot{y}(t) = A_\sigma y(t) + BK_{\widehat{\sigma}} \widehat{y}(t), \tag{2.3}$$

converges to zero as time increases, that is we want to stabilize the physical system (2.1). For brevity, we omit the explicit dependence on t whenever it is clear from the context.

The input $u = K_{\widehat{\sigma}}\widehat{y}$ is obtained based on a virtual model that sequentially assimilates data and mirrors the dynamics of the physical system. This model evolves simultaneously with the physical system and is governed by

$$\dot{\widehat{y}} = A_{\widehat{\sigma}}\widehat{y} + BK_{\widehat{\sigma}}\widehat{y} + L_{\widehat{\sigma}}C(\widehat{y} - y) - L_{\widehat{\sigma}}\eta, \qquad \widehat{y}(0) := \widehat{y}_0; \tag{2.4}$$

Designing the virtual twin in (2.4) entails addressing several interrelated challenges.

- Estimating $\widehat{\sigma}$ based on the output measurement data collected from the physical system (2.2). This involves solving a sequence of inverse problems, which are typically ill-posed and sensitive to noise, while ensuring that the estimated parameters produce a dynamically consistent model suitable for reliable controller synthesis and observer design.
- Constructing the state-feedback operator $K_{\widehat{\sigma}}$, a task commonly referred to as *controller synthesis* or *state-feedback design*. The goal is to shape the system's closed-loop behavior in accordance with performance criteria, which in this manuscript is stabilization. In the presence of uncertainty in the parameter estimate $\widehat{\sigma}$, this becomes a problem of *robust* or *risk-aware* control.
- Designing the observer gain $L_{\widehat{\sigma}}$ to ensure stable and accurate reconstruction of the system state from noisy and possibly partial measurements. This task is often referred to as *observer design* or *continuous data assimilation*. It involves balancing responsiveness to new data with robustness to measurement noise. It plays a crucial role in ensuring that the virtual state $\widehat{y}$ remains synchronized with the true physical state y.

The main contribution of this manuscript is a mathematical framework that jointly addresses the interacting challenges identified above, with the goal of stabilization of the physical system.

2.1. **The case without uncertain parameter and without measurement noise.** If we knew $\sigma \in \mathbb{R}^p$ and if we had exact observations, i.e., $z = Cy$, a classical strategy is to seek a feedback-input operator K_σ and an output-injection operator L_σ so that both operators $A_\sigma + BK_\sigma$ and $A_\sigma + L_\sigma C$ are exponentially stable. Once we find these operators, it follows that the coupled closed-loop system

$$\dot{y} = A_\sigma y + BK_\sigma\widehat{y}, \qquad y(0) := y_0; \tag{2.5a}$$

$$\dot{\widehat{y}} = A_\sigma\widehat{y} + BK_\sigma\widehat{y} + L_\sigma C(\widehat{y} - y), \qquad \widehat{y}(0) := \widehat{y}_0; \tag{2.5b}$$

provides us with the sought stabilizing input $K_\sigma\widehat{y}$. Here, $\widehat{y}_0$ can be taken as an initial guess that we might have for y_0.

As we can see, the estimate $\widehat{y}$ is given by a dynamic Luenberger observer, namely, by a dynamical system consisting of a copy of the physical system plus a correction term given by the injected forcing $L_\sigma C(\widehat{y} - y) = L_\sigma(C\widehat{y} - z)$.

2.2. **The case with an uncertain parameter and measurement noise.** If the parameter σ is unknown, the classical strategy as described above is not directly applicable, since the operators K_σ and L_σ in (2.5) rely on exact knowledge of σ. Furthermore, measurement noise in the output propagates through (2.5). Even if the parameter σ was known, this persistent noise generally prevents the state from converging to zero almost surely. To overcome these limitations, we propose an adaptive strategy that incorporates online parameter identification. Specifically, we compute an estimate $\widehat{\sigma}$ of the true parameter σ based on the system model (2.1) and the observation process (2.2). Using Bayes' law, our prior belief about the parameter is continuously updated as new measurement data become available. Given an estimate $\widehat{\sigma}$, we can mimic (2.5)

$$\dot{y} = A_\sigma y + BK_{\widehat{\sigma}}\widehat{y}, \qquad y(0) \coloneqq y_0; \tag{2.6a}$$

$$\dot{\widehat{y}} = A_{\widehat{\sigma}}\widehat{y} + BK_{\widehat{\sigma}}\widehat{y} + L_{\widehat{\sigma}}C(\widehat{y} - y) - L_{\widehat{\sigma}}\eta, \qquad \widehat{y}(0) \coloneqq \widehat{y}_0. \tag{2.6b}$$

Since the observer uses $A_{\widehat{\sigma}}$ instead of A_σ, it is not guaranteed that the input $K_{\widehat{\sigma}}\widehat{y}$ will be able to stabilize y. For this reason, it is necessary to update $\widehat{\sigma}$ online so that $A_{\widehat{\sigma}}$ progressively approaches A_σ. In doing so, the dynamics of (2.6) better approximate that of (2.5), up to the unavoidable influence of measurement noise. We shall compute a piecewise constant function $\widehat{\sigma}$ by updating $\widehat{\sigma}$ periodically over time. As feedback gain $K_{\widehat{\sigma}}$ and observer gain $L_{\widehat{\sigma}}$ we shall utilize the solutions to the appropriate algebraic Riccati equations.

2.3. **Our digital twin strategy.** In order to find the sought stabilizing control input in the presence of an uncertain parameter σ and incomplete noisy data $Cy+\eta$, we design an observer that operates in parallel with the physical system. It synchronizes with the physical system through output data collected according to (2.2), computes estimates $\widehat{y}$ of y using Kalman filtering, and $\widehat{\sigma}$ of σ using Bayesian estimation. Subsequently, the digital twin acts on the physical system via the input $u(t) = K_{\widehat{\sigma}(t)}\widehat{y}(t)$. In this sense, the observer acts as a *virtual twin* for the *physical twin* (2.1). Importantly, the approach illustrated in Figure 1 does not rely on numerical evaluations or direct knowledge of the physical system with the true parameter. Instead, the physical system is treated as an existing entity that receives the feedback input u and supplies the measurements z.

The reader may at this point already consult Algorithm 1 in Section 6.

2.4. **Mathematical formulation of the virtual twin.** Output measurements of the physical system are available only at discrete time instances, and the virtual twin itself is implemented as a virtual computational model. It is therefore natural to formulate the virtual twin as a discrete-time dynamical system. To make this precise, and consistent with (2.2), we assume that, between two consecutive parameter updates, with current estimate $\widehat{\sigma}$, there are s discrete observation times, see Algorithm 1. For $t \in [t_j, t_{j+1})$, $j = 0, \dots, s-1$ the controlled physical system is given as

$$\dot{y}(t) = A_\sigma y(t) + BK_{\widehat{\sigma}}\widehat{y}_j, \qquad y(t_j) \coloneqq y_j, \tag{2.7}$$

where the virtual twin input $\widehat{y}_j(t) = \widehat{y}_j$, for all $t \in [t_j, t_{j+1})$ is constant between the observation points. Let us assume an equidistant grid $t_j = j\Delta t$, $\forall j \geq 0$. The exact

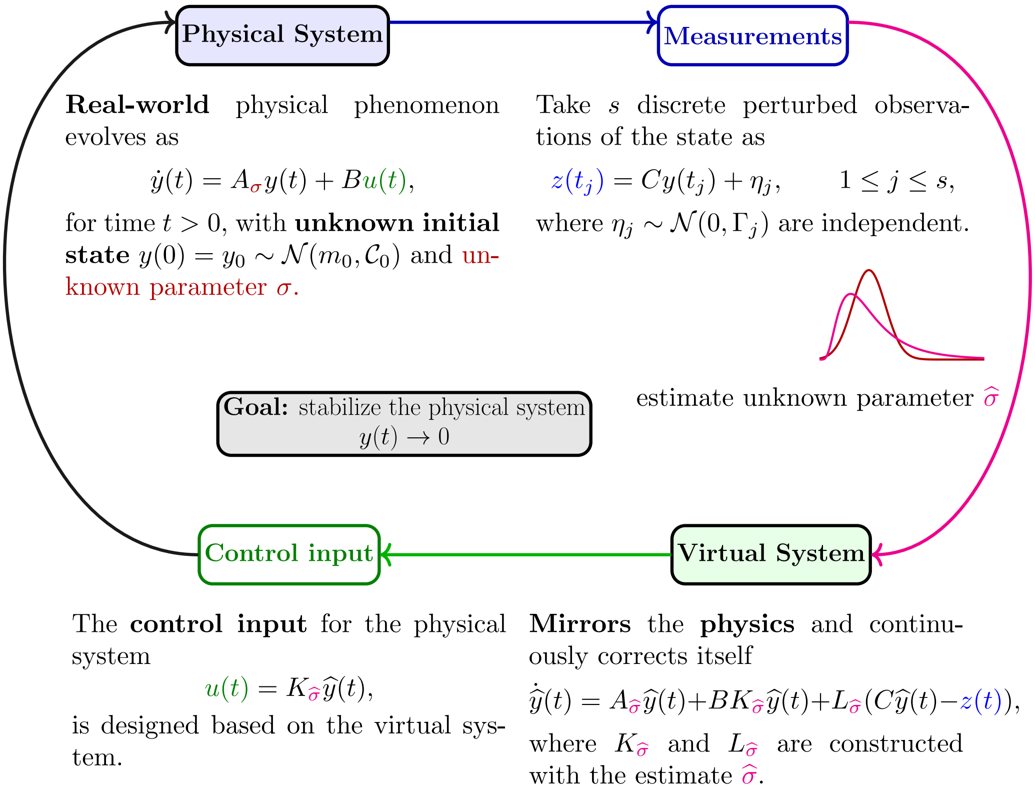


FIGURE 1. Illustration of the Digital Twin.

solution of (2.7) evaluated on the interval boundaries gives the iteration map

$$y_{j+1} = e^{A_\sigma \Delta t} y_j + \int_0^{\Delta t} e^{A_\sigma(\Delta t - \tau)}\, \mathrm{d}\tau\, BK_{\widehat{\sigma}}\, \widehat{y}_j. \tag{2.8}$$

The dynamics (2.8) can be expressed as

$$y_{j+1} = \mathcal{A}_\sigma y_j + \mathcal{B}_\sigma K_{\widehat{\sigma}} \widehat{y}_j, \tag{2.9}$$

for $\mathcal{A}_\sigma = e^{A_\sigma \Delta t}$ and $\mathcal{B}_\sigma = \int_0^{\Delta t} e^{A_\sigma(\Delta t - \tau)}\, \mathrm{d}\tau\, B$ with initializations y_0 and $\widehat{y}_0$ as in (2.5). Based on $\mathcal{A}_\sigma$, $\mathcal{B}_\sigma$, and an estimate $\widehat{\sigma}$ of σ, we design the virtual twin as

$$\widehat{y}_{j+1} = \mathcal{A}_{\widehat{\sigma}} \widehat{y}_j + \mathcal{B}_{\widehat{\sigma}} K_{\widehat{\sigma}} \widehat{y}_j + L_{\widehat{\sigma}}(C\widehat{y}_j - z_j), \quad \widehat{y}_0 \text{ is known.}$$

In this way, we arrive at the coupled discrete-time system

$$y_{j+1} = \mathcal{A}_\sigma y_j + \mathcal{B}_\sigma K_{\widehat{\sigma}} \widehat{y}_j, \tag{2.10a}$$

$$z_j = Cy_j + \eta_j, \tag{2.10b}$$

$$\widehat{y}_{j+1} = \mathcal{A}_{\widehat{\sigma}} \widehat{y}_j + \mathcal{B}_{\widehat{\sigma}} K_{\widehat{\sigma}} \widehat{y}_j + L_{\widehat{\sigma}}(C\widehat{y}_j - z_j), \tag{2.10c}$$

with initial guess $\widehat{y}_0$ for the observer and unknown initial condition y_0 for the physical system.

Remark 2.1. Alternatively to the exact representation via the matrix exponential, we can use approximations of $\mathcal{A}_\sigma$ and $\mathcal{B}_\sigma$. Below we list two typical choices.

- For the choices

$$\mathcal{A}_\sigma := (I + \Delta t A_\sigma), \quad \text{and} \quad \mathcal{B} := \Delta t B,$$

the approximation (2.9) corresponds to a forward Euler discretization of the physical system (2.3).

- For the choices

$$\mathcal{A}_\sigma := \left(I - \frac{\Delta t}{2} A_\sigma\right)^{-1} \left(I + \frac{\Delta t}{2} A_\sigma\right), \quad \text{and,} \quad \mathcal{B}_\sigma := \Delta t \left(I - \frac{\Delta t}{2} A_\sigma\right)^{-1} B,$$

the approximation (2.9) corresponds to a semi-implicit Crank–Nicolson discretization of the physical system (2.7).

3. Infinite horizon asymptotics

Here, we provide an analysis for system (2.10) as $j \to \infty$. Throughout, σ is the fixed but unknown coefficient in (2.1). First, $\widehat{\sigma}$ remains constant, then in Subsection 3.1 the special case $\widehat{\sigma} = \sigma$ is considered, and subsequently a sequence of estimates $(\widehat{\sigma}_i)_{i \in \mathbb{N}}$ is considered in Subsection 3.2. Throughout we will assume that the estimates are sufficiently close to σ. All operators $\mathcal{A}_\varsigma, \mathcal{B}_\varsigma, K_\varsigma, L_\varsigma$ in (2.10) are assumed to depend continuously on ς in a neighborhood of σ.

For our analysis, it is convenient to introduce the tracking error $e_j = y_j - \hat{y}_j$, for $j = 0, 1, \ldots$. Then, from (2.10), we deduce that

$$\begin{aligned}
y_{j+1} &= \mathcal{A}_\sigma y_j + \mathcal{B}_\sigma K_{\widehat{\sigma}} \widehat{y}_j = \mathcal{A}_\sigma y_j + \mathcal{B}_\sigma K_{\widehat{\sigma}} y_j - \mathcal{B}_\sigma K_{\widehat{\sigma}} e_j \\
\widehat{y}_{j+1} &= (\mathcal{A}_{\widehat{\sigma}} + \mathcal{B}_{\widehat{\sigma}} K_{\widehat{\sigma}}) \widehat{y}_j + L_{\widehat{\sigma}} C (\widehat{y}_j - y_j) - L_{\widehat{\sigma}} \eta_j \\
e_{j+1} &= (\mathcal{A}_\sigma - \mathcal{A}_{\widehat{\sigma}}) y_j + (\mathcal{A}_{\widehat{\sigma}} + L_{\widehat{\sigma}} C) e_j + (\mathcal{B}_\sigma K_{\widehat{\sigma}} - \mathcal{B}_{\widehat{\sigma}} K_{\widehat{\sigma}}) y_j - (\mathcal{B}_\sigma K_{\widehat{\sigma}} - \mathcal{B}_{\widehat{\sigma}} K_{\widehat{\sigma}}) e_j + L_{\widehat{\sigma}} \eta_j,
\end{aligned}$$

where the conditions on the noise sequence $\{\eta_j\}$ in $\mathbb{R}^\ell$ are given below. Associated to this system, we define

$$F = F(\sigma, \widehat{\sigma}) = \begin{bmatrix} \mathcal{A}_\sigma + \mathcal{B}_\sigma K_{\widehat{\sigma}} & -\mathcal{B}_\sigma K_{\widehat{\sigma}} \\ \mathcal{A}_\sigma - \mathcal{A}_{\widehat{\sigma}} + (\mathcal{B}_\sigma - \mathcal{B}_{\widehat{\sigma}}) K_{\widehat{\sigma}} & \mathcal{A}_{\widehat{\sigma}} + L_{\widehat{\sigma}} C - (\mathcal{B}_\sigma - \mathcal{B}_{\widehat{\sigma}}) K_{\widehat{\sigma}} \end{bmatrix}. \tag{3.1}$$

Then, the iteration in the variables (y_j, e_j) can be expressed as

$$X_{j+1} = F(\sigma, \widehat{\sigma})\, X_j - \begin{bmatrix} 0 \\ L_{\widehat{\sigma}} \end{bmatrix} \eta_j, \text{ with } X_0 = \begin{bmatrix} y_0 \\ y_0 - \widehat{y}_0 \end{bmatrix}, \tag{3.2}$$

where

$$X_j = \begin{bmatrix} y_j \\ e_j \end{bmatrix} = \begin{bmatrix} y_j \\ y_j - \widehat{y}_j \end{bmatrix}.$$

Hereafter, we denote by $\varrho(M)$ the spectral radius of a matrix M.

Theorem 3.1. *Let σ be fixed. Assume existence of an open neighborhood $\mathcal{I}$ of σ and continuous, matrix valued functions $\varsigma \in \mathcal{I} \mapsto K_\varsigma$ and $\varsigma \in \mathcal{I} \mapsto L_\varsigma$, and that $\varrho(\mathcal{A}_\sigma + \mathcal{B}_\sigma K_\sigma) < 1$ and $\varrho(\mathcal{A}_\sigma + L_\sigma C) < 1$. Then, there exists $\mathcal{I}' \subset \mathcal{I}$ such that for all $\varsigma \in \mathcal{I}'$ it holds that $\varrho(F(\sigma, \varsigma)) < 1$.*

Remark 3.2. Before presenting the proof of Theorem 3.1, we point out that, under mild assumptions, the theorem can be applied with the gains obtained via standard discrete-time Riccati equations. To that end, let σ be fixed and assume that the pairs $(\mathcal{A}_\sigma, \mathcal{B}_\sigma)$ and $(\mathcal{A}_\sigma, C)$ are stabilizable and detectable, respectively. The continuity of $\mathcal{A}_\sigma$ and $\mathcal{B}_\sigma$ in σ implies existence of $\mathcal{I}$ containing σ, such that, for all $\varsigma \in \mathcal{I}$ the pairs $(\mathcal{A}_\varsigma, \mathcal{B}_\varsigma)$ and $(\mathcal{A}_\varsigma, C)$ are stabilizable and detectable. Further, assume a sufficiently small time step, such that $\mathcal{A}_\varsigma$ is invertible. Then, for any symmetric positive definite matrices $R \in \mathbb{R}^{m\times m}$, $Q \in \mathbb{R}^{n\times n}$, the discrete algebraic Riccati equations

$$\begin{aligned} X_\varsigma &= \mathcal{A}_\varsigma^\top X_\varsigma \mathcal{A}_\varsigma + C^\top C - (\mathcal{B}_\varsigma^\top X_\varsigma \mathcal{A}_\varsigma)^\top (R + \mathcal{B}_\varsigma^\top X_\varsigma \mathcal{B}_\varsigma)^{-1} (\mathcal{B}_\varsigma^\top X_\varsigma \mathcal{A}_\varsigma) \\ Y_\varsigma &= \mathcal{A}_\varsigma Y_\varsigma \mathcal{A}_\varsigma^\top + Q - (C Y_\varsigma \mathcal{A}_\varsigma^\top)^\top (\Gamma + C Y_\varsigma C^\top)^{-1} C Y_\varsigma \mathcal{A}_\varsigma^\top \end{aligned}$$

admit unique symmetric, positive semidefinite, stabilizing solutions X_ς and Y_ς, cf. [9, Cor. 13.5.3]. Additionally, X_ς and Y_ς depend continuously on ς, see [9, Thm. 14.2.1]. Hence, the associated gains

$$\begin{aligned} K_\varsigma &= -(R + \mathcal{B}_\varsigma^\top X_\varsigma \mathcal{B}_\varsigma)^{-1} \mathcal{B}_\varsigma^\top X_\varsigma \mathcal{A}_\varsigma, \\ L_\varsigma &= -\mathcal{A}_\varsigma Y_\varsigma C^\top (\Gamma + C Y_\varsigma C^\top)^{-1} \end{aligned}$$

satisfy the assumptions of Theorem 3.1.

Proof of Theorem 3.1. Let σ be fixed and $\mathcal{I}$ as in the assumption. For now, set $\mathcal{I}' = \mathcal{I}$. Define the continuous mappings

$$\varsigma \mapsto \mathcal{A}_1(\varsigma) := \mathcal{A}_\sigma + \mathcal{B}_\sigma K_\varsigma, \qquad \varsigma \mapsto \mathcal{A}_2(\varsigma) := \mathcal{A}_\varsigma + L_\varsigma C + (\mathcal{B}_\sigma - \mathcal{B}_\varsigma) K_\varsigma,$$

$$\varsigma \mapsto \mathcal{B}_1(\varsigma) := -\mathcal{B}_\sigma K_\varsigma, \qquad \varsigma \mapsto \mathcal{B}_2(\varsigma) := \mathcal{A}_\sigma - \mathcal{A}_\varsigma - (\mathcal{B}_\sigma - \mathcal{B}_\varsigma) K_\varsigma.$$

By assumption, there holds $\varrho(\mathcal{A}_\sigma + \mathcal{B}_\sigma K_\sigma) < 1$ and due to the continuity of K_σ a possible decrease of $\mathcal{I}'$ ensures that for $\varsigma \in \mathcal{I}'$ it follows that $\varrho(\mathcal{A}_1(\varsigma)) = \varrho(\mathcal{A}_\sigma + \mathcal{B}_\sigma K_\varsigma) < 1$. Analogous arguments yield that $\varrho(\mathcal{A}_2(\varsigma)) < 1$ for $\varsigma \in \mathcal{I}'$. We now proof the assertion under the technical assumption that there exists a submultiplicative norm $\|\cdot\|_*$ such that for all $\varsigma \in \mathcal{I}'$ there holds

$$\left\| (\lambda I - \mathcal{A}_2(\varsigma))^{-1} \mathcal{B}_2(\varsigma) (\lambda I - \mathcal{A}_1(\varsigma))^{-1} \mathcal{B}_1(\varsigma) \right\|_* < 1 \quad \text{for all } \lambda \in \Lambda := \{\lambda \in \mathbb{C} : |\lambda| \geq 1\}. \tag{3.3}$$

From now on, for ease of readability, we suppress the dependence on ς in the notation. Observe that

$$F(\sigma, \varsigma) = F = \begin{bmatrix} \mathcal{A}_1 & \mathcal{B}_1 \\ \mathcal{B}_2 & \mathcal{A}_2 \end{bmatrix}.$$

Since $\varrho(\mathcal{A}_i) < 1$, the spectrum of $\mathcal{A}_i$ is contained in the open unit disk. Hence, $\lambda I - \mathcal{A}_i$ is invertible for all $\lambda \in \Lambda$, for $i = 1, 2$. For $\lambda \in \Lambda$ we write

$$\begin{aligned} \lambda I - F &= \begin{bmatrix} \lambda I - \mathcal{A}_1 & -\mathcal{B}_1 \\ -\mathcal{B}_2 & \lambda I - \mathcal{A}_2 \end{bmatrix} \\ &= \begin{bmatrix} \lambda I - \mathcal{A}_1 & 0 \\ -\mathcal{B}_2 & \lambda I - \mathcal{A}_2 \end{bmatrix} \begin{bmatrix} I & -(\lambda I - \mathcal{A}_1)^{-1} \mathcal{B}_1 \\ 0 & I - (\lambda I - \mathcal{A}_2)^{-1} \mathcal{B}_2 (\lambda I - \mathcal{A}_1)^{-1} \mathcal{B}_1 \end{bmatrix}. \end{aligned}$$

By hypothesis, there holds

$$\left\|(\lambda I-\mathcal{A}_2)^{-1}\mathcal{B}_2(\lambda I-\mathcal{A}_1)^{-1}\mathcal{B}_1\right\|_* < 1,$$

so the lower-right block of the second factor is invertible via the Neumann series

$$\left(I-(\lambda I-\mathcal{A}_2)^{-1}\mathcal{B}_2(\lambda I-\mathcal{A}_1)^{-1}\mathcal{B}_1\right)^{-1} = \sum_{i=0}^{\infty}\left((\lambda I-\mathcal{A}_2)^{-1}\mathcal{B}_2(\lambda I-\mathcal{A}_1)^{-1}\mathcal{B}_1\right)^i.$$

Hence, the second factor is invertible. The first factor is invertible as it is block lower-triangular with invertible diagonal blocks. Therefore, $\lambda I - F$ is invertible for all $\lambda\in\Lambda$, so no eigenvalue of F lies on or outside the unit circle. We conclude that $\varrho(F)<1$.

It remains to show the existence of norms as in (3.3). First, note that $\mathcal{A}_1(\sigma)=\mathcal{A}_\sigma+\mathcal{B}_\sigma K_\sigma$, $\mathcal{A}_2(\sigma)=\mathcal{A}_\sigma+L_\sigma C$, and $\mathcal{B}_2(\sigma)=0$. Now, for $\lambda\in\Lambda$ and $\varsigma\in\mathcal{I}'$, define

$$G(\lambda,\varsigma) := (\lambda I-\mathcal{A}_2(\varsigma))^{-1}\mathcal{B}_2(\varsigma)(\lambda I-\mathcal{A}_1(\varsigma))^{-1}\mathcal{B}_1(\varsigma).$$

Since $\varrho(\mathcal{A}_i(\sigma))<1$, there exists norms $\|\cdot\|_*$ and $\|\cdot\|_\dagger$ on $\mathbb{R}^n$, such that, for their respective induced matrix norms there holds $\|\mathcal{A}_1(\sigma)\|_\dagger<1$ and $\|\mathcal{A}_2(\sigma)\|_*<1$, [17, Sect. 6.9, (6.9.2) Thm.]. Also, possibly after further reducing $\mathcal{I}'$, we have that $\|\mathcal{A}_1(\varsigma)\|_\dagger, \|\mathcal{A}_2(\varsigma)\|_*<1$, for all $\varsigma\in\mathcal{I}'$. Since the matrix norms $\|\cdot\|_\dagger$ and $\|\cdot\|_*$ are equivalent, there exists $K>0$, such that $\|\cdot\|_*\le K\|\cdot\|_\dagger$. Consequently, for $\lambda\in\Lambda$, we estimate

$$\begin{aligned}\|G(\lambda,\varsigma)\|_* &\le \frac{K\|\mathcal{B}_2(\varsigma)\|_*}{|\lambda|-\|\mathcal{A}_2(\varsigma)\|_*}\left\|(\lambda I-\mathcal{A}_1(\varsigma))^{-1}\mathcal{B}_1(\varsigma)\right\|_\dagger \\ &\le \frac{K\|\mathcal{B}_2(\varsigma)\|_*\|\mathcal{B}_1(\varsigma)\|_\dagger}{(|\lambda|-\|\mathcal{A}_2(\varsigma)\|_*)(|\lambda|-\|\mathcal{A}_1(\varsigma)\|_\dagger)} \le \frac{K\|\mathcal{B}_2(\varsigma)\|_*\|\mathcal{B}_1(\varsigma)\|_\dagger}{(1-\|\mathcal{A}_2(\varsigma)\|_*)(1-\|\mathcal{A}_1(\varsigma)\|_\dagger)},\end{aligned}$$

and hence, due to the form of $\mathcal{B}_2$, after further reducing $\mathcal{I}'$, we have that $\|G(\lambda,\varsigma)\|_*<1$ and thus, (3.3) holds for all $\lambda\in\Lambda$. As an induced matrix norm $\|\cdot\|_*$ is submultiplicative, the proof is finished. □

Remark 3.3. If the operators $\mathcal{A}_\sigma$ and $\mathcal{B}_\sigma$ introduced below (2.9) are approximated by $\tilde{\mathcal{A}}_\sigma$ and $\tilde{\mathcal{B}}_\sigma$, as discussed in Remark 2.1, then Theorem 3.1 is again applicable with $\mathcal{A}_\sigma$ and $\mathcal{B}_\sigma$ replaced by $\tilde{\mathcal{A}}_\sigma$ and $\tilde{\mathcal{B}}_\sigma$, and analogously for $\mathcal{A}_{\widehat{\sigma}}$ and $\mathcal{B}_{\widehat{\sigma}}$. Indeed, for this purpose, one can again validate the steps in the proof of Theorem 3.1 provided that $\|\mathcal{A}_\sigma-\tilde{\mathcal{A}}_\sigma\|$, $\|\mathcal{B}_\sigma-\mathcal{B}_{\widehat{\sigma}}\|$, $\|\mathcal{A}_{\widehat{\sigma}}-\tilde{\mathcal{A}}_{\widehat{\sigma}}\|$, and $\|\mathcal{B}_{\widehat{\sigma}}-\tilde{\mathcal{B}}_{\widehat{\sigma}}\|$ are sufficiently small.

We turn to analyzing the convergence in distribution of system (2.10). First, the definition adapted to our setting is recalled.

Definition 3.4. Let X_k be a sequence of normally distributed random vectors with $X_k\sim\mathcal{N}(m_k,\Sigma_k)$. We say that X_k converges in distribution to the Gaussian variable $X\sim\mathcal{N}(m,\Sigma)$ if m_k and Σ_k converge to m and Σ, respectively.

Remark 3.5. For a more general definition of convergence in distribution see [6, Sec. 25]. By [6, Thm. 30.2, Ex. 30.1], for Gaussian random vectors, it is equivalent to our definition.

We henceforth assume that the initial condition y_0 of the state variable is distributed according to a Gaussian $\mathcal{N}(m_0, \mathcal{C}_0)$. For the measurement error we take $\eta_j \sim \mathcal{N}(0, \Gamma_j)$ mutually independent and $\Gamma_j \to \Gamma_\infty$. We shall also need the covariance matrix

$$\mathbf{\Pi}_0 = \begin{bmatrix} \mathcal{C}_0 & \mathcal{C}_0 \\ \mathcal{C}_0 & \mathcal{C}_0 \end{bmatrix}. \tag{3.4}$$

Theorem 3.6. *Given $\sigma \in \mathbb{R}^p$ and $\widehat{\sigma} \in \mathbb{R}^p$, consider system (2.10) with the assumptions in Theorem 3.1 holding for $\sigma, \widehat{\sigma}$ and initialize the virtual twin with $\widehat{y}_0 = m_0$.*

Then, the iterates X_j of (2.10) are normally distributed at each iteration level, and $X_j \sim \mathcal{N}(F^j \left[\begin{smallmatrix} m_0 \\ 0 \end{smallmatrix}\right], \mathbf{\Pi}_j)$, where

$$\mathbf{\Pi}_j = F(\sigma, \widehat{\sigma})^j \, \mathbf{\Pi}_0 \, F(\sigma, \widehat{\sigma})^{\top j} + \sum_{k=0}^{j-1} F(\sigma, \widehat{\sigma})^{j-1-k} \left[\begin{smallmatrix} 0 \\ L_{\widehat{\sigma}} \end{smallmatrix}\right] \Gamma_k \left[\begin{smallmatrix} 0 \\ L_{\widehat{\sigma}} \end{smallmatrix}\right]^\top F(\sigma, \widehat{\sigma})^{\top\, j-1-k},$$

for $j \geq 1$. Further, for $j \to \infty$, the pair (y_j, e_j) converges to

$$(y_\infty, e_\infty) \sim \mathcal{N}(0, \mathbf{\Pi}_\infty)$$

in distribution, where $\mathbf{\Pi}_\infty$ is the unique solution to the Lyapunov equation

$$\mathbf{\Pi}_\infty = F(\sigma, \widehat{\sigma}) \mathbf{\Pi}_\infty F(\sigma, \widehat{\sigma})^\top + \left[\begin{smallmatrix} 0 \\ L_{\widehat{\sigma}} \end{smallmatrix}\right] \Gamma_\infty \left[\begin{smallmatrix} 0 \\ L_{\widehat{\sigma}} \end{smallmatrix}\right]^\top. \tag{3.5}$$

Proof. From (3.2), for $j \geq 0$, we obtain the recursion formula

$$X_j = F(\sigma, \widehat{\sigma})^j \left[\begin{smallmatrix} y_0 \\ y_0 - m_0 \end{smallmatrix}\right] - \sum_{k=0}^{j-1} F(\sigma, \widehat{\sigma})^{j-1-k} \left[\begin{smallmatrix} 0 \\ L_{\widehat{\sigma}} \end{smallmatrix}\right] \eta_k.$$

The independence of y_0 and the η_j yields the characterization of X_j. Since $\Gamma_j \to \Gamma_\infty$ and $F(\sigma, \widehat{\sigma})$ is a contraction, by Theorem 3.1, a minor extension of the arguments in [16, Lemma 4.2, Corollary 2] allows us to conclude that $\mathbf{\Pi}_j$ converges to the solution of the Lyapunov equation (3.5). Uniqueness of $\mathbf{\Pi}_\infty$ follows from the contraction property of F. This ends the proof. □

We consider next the case of decaying measurement noise.

Corollary 3.7. *In addition to the assumptions of Theorem 3.6, let $\Gamma_j \to 0$, $j \to \infty$. Then, for any $\epsilon > 0$, there holds*

$$P(\|y_j\| + \|e_j\| > \epsilon) \to 0, \ \ j \to \infty.$$

Proof. Again we denote $\begin{bmatrix} y_j \\ e_j \end{bmatrix} = X_j$. From the previous theorem, we have that $X_j \sim \mathcal{N}(m_j, \mathbf{\Pi}_j)$, with $m_j \to 0$ for $j \to \infty$. Concerning $\mathbf{\Pi}_j$, we insert $\Gamma_\infty = 0$ into (3.5), and utilizing the uniqueness of the associated solution, we obtain that $\mathbf{\Pi}_j$ converges to $\mathbf{\Pi}_\infty = 0$. Utilizing Markov's inequality for every $\epsilon > 0$, we obtain that

$$P\left(\|X_j\| > \epsilon\right) \leq \epsilon^{-2} \, \mathbb{E}\left[\|X_j\|^2\right] = \epsilon^{-2} \, \left(\|m_j\|^2 + \mathrm{tr}(\mathbf{\Pi}_j)\right).$$

As $\mathbf{\Pi}_j$ converges to zero, so does its trace $\mathrm{tr}(\mathbf{\Pi}_j)$, and the assertion is shown. □

3.1. **The case $\widehat{\sigma} = \sigma$.** The following corollary addresses the case that the estimated parameter $\widehat{\sigma}$ coincides with the true parameter of the physical system. For convenience, we repeat the expression for F which in this case reads

$$F = F(\sigma,\sigma) = \begin{bmatrix} \mathcal{A}_\sigma + \mathcal{B}_\sigma K_\sigma & -\mathcal{B}_\sigma K_\sigma \\ 0 & \mathcal{A}_\sigma + L_\sigma C \end{bmatrix}.$$

Further, for $j = 1, \ldots,$ there holds

$$F^j = \begin{bmatrix} (\mathcal{A}_\sigma + \mathcal{B}_\sigma K_\sigma)^j & S_j \\ 0 & (\mathcal{A}_\sigma + L_\sigma C)^j \end{bmatrix}, \tag{3.6}$$

where $S_j = -\sum_{k=0}^{j-1} (\mathcal{A}_\sigma + \mathcal{B}_\sigma K_\sigma)^k \mathcal{B}_\sigma K_\sigma (\mathcal{A}_\sigma + L_\sigma C)^{j-1-k}$.

Corollary 3.8. *Consider the setting of the previous theorem, and, in addition, assume that $\widehat{\sigma} = \sigma$. Then, we have $y_j \sim \mathcal{N}(m_j, \Pi_j^y)$ and $e_j \sim \mathcal{N}(0, \Pi_j^e)$, where*

$$\begin{aligned} m_j &= (\mathcal{A}_\sigma + \mathcal{B}_\sigma K_\sigma)^j m_0 \\ \Pi_j^e &= (\mathcal{A}_\sigma + L_\sigma C)^j \mathcal{C}_0 (\mathcal{A}_\sigma + L_\sigma C)^{j^\top} \\ &\quad + \sum_{k=0}^{j-1} (\mathcal{A}_\sigma + L_\sigma C)^{j-1-k} L_\sigma \Gamma_k L_\sigma^\top (\mathcal{A}_\sigma + L_\sigma C)^{j-1-k^\top} \\ \Pi_j^y &= ((\mathcal{A}_\sigma + \mathcal{B}_\sigma K_\sigma)^j + S_j) \mathcal{C}_0 ((\mathcal{A}_\sigma + \mathcal{B}_\sigma K_\sigma)^j + S_j)^\top \\ &\quad + \sum_{k=0}^{j-1} S_{j-1-k} L_\sigma \Gamma_k {L_\sigma}^\top S_{j-1-k}^\top, \end{aligned}$$

for $j \geq 1$, with S_i as defined above. Further, we have convergence in distribution of e_j and y_j to $e_\infty \sim \mathcal{N}(0, \Pi_\infty^e)$ and $y_\infty \sim \mathcal{N}(0, \Pi_\infty^y)$, where Π_∞^e and Π_∞^y satisfy

$$\Pi_\infty^e = (\mathcal{A}_\sigma + L_\sigma C) \Pi_\infty^e (\mathcal{A}_\sigma + L_\sigma C)^\top + L_\sigma \Gamma_\infty L_\sigma^\top, \tag{3.7}$$

and

$$\begin{aligned} \Pi_\infty^y \quad &= (\mathcal{A}_\sigma + \mathcal{B}_\sigma K_\sigma) \Pi_\infty^y (\mathcal{A}_\sigma + \mathcal{B}_\sigma K_\sigma)^\top + \mathcal{B}_\sigma K_\sigma \Pi_\infty^e K_\sigma^\top \mathcal{B}_\sigma^\top \\ &\quad - (\mathcal{A}_\sigma + \mathcal{B}_\sigma K_\sigma) \Pi_\infty^{ye} (\mathcal{B}_\sigma K_\sigma)^\top - (\mathcal{B}_\sigma K_\sigma) \Pi_\infty^{ey} (\mathcal{A}_\sigma + \mathcal{B}_\sigma K_\sigma)^\top, \end{aligned} \tag{3.8}$$

with

$$\Pi_\infty^{ye} = \Pi_\infty^{ey\top} = (\mathcal{A}_\sigma + \mathcal{B}_\sigma K_\sigma) \Pi_\infty^{ye} (\mathcal{A}_\sigma + L_\sigma C)^\top - \mathcal{B}_\sigma K_\sigma \Pi_\infty^e (\mathcal{A}_\sigma + L_\sigma C)^\top.$$

Proof. For the sake of readability, within this proof, we drop the dependence on σ in the notation. Applying Theorem 3.6, for $j \geq 1$, we obtain that

$$X_j = \begin{bmatrix} y_j \\ e_j \end{bmatrix} \sim \mathcal{N}(F^j \left[\begin{smallmatrix} m_0 \\ 0 \end{smallmatrix}\right], \mathbf{\Pi}_j).$$

Utilizing (3.6), the announced expected values of y_j and e_j follow immediately. We proceed by characterizing $\mathbf{\Pi}_j$ via

$$\mathbf{\Pi}_j = F^j \, \mathbf{\Pi}_0 \, F^{\top j} + \sum_{k=0}^{j-1} F^{j-1-k} \left[\begin{smallmatrix} 0 \\ L \end{smallmatrix}\right] \Gamma_k \left[\begin{smallmatrix} 0 \\ L \end{smallmatrix}\right]^\top F^{\top\, j-1-k}.$$

After plugging in (3.4) and (3.6), basic calculus yields that

$$\mathbf{\Pi}_j = \begin{bmatrix} \Pi_j^{(1,1)} & \Pi_j^{(1,2)} \\ \Pi_j^{(2,1)} & \Pi_j^{(2,2)} \end{bmatrix},$$

with

$$
\begin{aligned}
\Pi_j^{(1,1)} &= \left((\mathcal{A}+\mathcal{B}K)^j + S_k\right)\mathcal{C}_0\left((\mathcal{A}+\mathcal{B}K)^j + S_k\right)^\top - \sum_{k=1}^{j-1} S_{j-1-k}L\Gamma_k L^\top S_{j-1-k}^\top,\\
\Pi_j^{(2,2)} &= (\mathcal{A}+LC)\mathcal{C}_0(\mathcal{A}+LC)^\top - \sum_{k=1}^{j-1}(\mathcal{A}+LC)^{j-1-k}L\Gamma_k L^\top {(\mathcal{A}+LC)^{j-1-k}}^\top,\\
\Pi_j^{(1,2)} &= {\Pi_j^{(2,1)}}^\top\\
&= (\mathcal{A}+LC)^j\mathcal{C}_0\left((\mathcal{A}+\mathcal{B}K)^j + S_k\right)^\top - \sum_{k=1}^{j-1}(\mathcal{A}+LC)^{j-1-k}L\Gamma_k L^\top S_{j-1-k}^\top,
\end{aligned}
$$

and the covariances of y_j and e_j are given by the diagonal blocks $\Pi_j^y = \Pi_j^{(1,1)}$ and $\Pi_j^e = \Pi_j^{(2,2)}$, respectively.

We turn to the verification of (3.7) and (3.8) and express (3.5) as

$$
\boldsymbol{\Pi}_\infty = \begin{bmatrix} \Pi_\infty^y & \Pi_\infty^{ye} \\ \Pi_\infty^{ey} & \Pi_\infty^e \end{bmatrix} =
$$

$$
\begin{bmatrix} \mathcal{A}+\mathcal{B}K & -\mathcal{B}K \\ 0 & \mathcal{A}+LC \end{bmatrix}\begin{bmatrix} \Pi_\infty^y & \Pi_\infty^{ye} \\ \Pi_\infty^{ey} & \Pi_\infty^e \end{bmatrix}\begin{bmatrix} (\mathcal{A}+\mathcal{B}K)^\top & 0 \\ (-\mathcal{B}K)^\top & (\mathcal{A}+LC)^\top \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & L\Gamma_\infty L^\top \end{bmatrix}.
$$

Solving this equation first for Π_∞^e, we obtain (3.7). Then, we solve for Π_∞^{ye}, and finally for Π_∞^y to obtain (3.8). □

3.2. **The case $\widehat{\sigma}$ is piecewise constant.** Eventually, we want to update the estimate of the uncertain parameter online. We point out that switching between stable systems may have destabilizing effects [2, Prop. 1], [11, Prob. A], [8, Sect. 4.3.3]. This does not happen if $\widehat{\sigma}$ remains in a sufficiently small neighborhood $\mathcal{I}$ of σ, because $\varrho(\mathcal{A}_\sigma + \mathcal{B}_\sigma K_\sigma) < 1$ implies $\|\mathcal{A}_\sigma + \mathcal{B}_\sigma K_\sigma\|_* < 1$ for some induced operator norm and, by continuity, $\|\mathcal{A}_{\widehat{\sigma}} + \mathcal{B}_{\widehat{\sigma}} K_{\widehat{\sigma}}\|_* < 1$ (by taking a smaller neighborhood $\widetilde{\mathcal{I}} \subset \mathcal{I}$, if necessary).

Fix $\Delta t > 0$, we shall consider the case where $\widehat{\sigma}$ is updated at instants of time multiples of Δt:

$$
\widehat{\sigma}(t) = \widehat{\sigma}_i, \qquad i\Delta t \le t < (i+1)\Delta t, \quad i \in \mathbb{N}.
$$

Now, let

$$
F_j = F(\sigma, \widehat{\sigma}_j) = \begin{bmatrix} \mathcal{A}_\sigma + \mathcal{B}_{\widehat{\sigma}_j} K_{\widehat{\sigma}_j} & -\mathcal{B}_{\widehat{\sigma}_j} K_{\widehat{\sigma}_j} \\ 0 & \mathcal{A}_{\widehat{\sigma}_j} + L_{\widehat{\sigma}_j} C \end{bmatrix}, \qquad j \in \mathbb{N}.
$$

We have the following generalization of Theorem 3.6.

Theorem 3.9. *Given $\sigma \in \mathbb{R}^p$ and $\widehat{\sigma}_i \in \mathbb{R}^p$, $i \in \mathbb{N}$, consider system* (2.10) *with the assumptions in Theorem 3.1 holding in a neighborhood of σ. Initialize the virtual twin with $\widehat{y}_0 = m_0$. Then, the iterates X_j of* (2.10) *are normally distributed at each iteration* *level, and* $X_j \sim \mathcal{N}(\Phi_{j|0}\left[\begin{smallmatrix} m_0 \\ 0 \end{smallmatrix}\right], \mathbf{\Pi}_j)$*, where*

$$\Phi_{0|i} = \mathbf{1}; \quad \Phi_{m+1|i} \coloneqq F_{i+m}\Phi_{m|i}, \qquad i \in \mathbb{N}, \quad m \in \mathbb{N};$$

$$\mathbf{\Pi}_j = \Phi_{j|0}\,\mathbf{\Pi}_0\,\Phi_{j|0}^\top + \sum_{k=0}^{j-1} \Phi_{j-1-k|k}\left[\begin{smallmatrix} 0 \\ L_{\widehat{\sigma}_k} \end{smallmatrix}\right]\Gamma_k\left[\begin{smallmatrix} 0 \\ L_{\widehat{\sigma}_k} \end{smallmatrix}\right]^\top \Phi_{j-1-k|k}^\top, \qquad j \in \mathbb{N}_+.$$

Further, there exists a matrix norm $\|\cdot\|$ *and* $\lambda < 1$ *such that, for every* $\widehat{\sigma}_j$ *in a small neighborhood of* σ*,*

$$\begin{aligned} &\left\|\Phi_{j|0}\left[\begin{smallmatrix} m_0 \\ 0 \end{smallmatrix}\right]\right\| \le \lambda^j \left\|\left[\begin{smallmatrix} m_0 \\ 0 \end{smallmatrix}\right]\right\|; \\ &\|\mathbf{\Pi}_j\| \le \lambda^{2j}\,\|\mathbf{\Pi}_0\| + \frac{1}{1-\lambda^2}\sup_{k\in\mathbb{N}}\left\|\left[\begin{smallmatrix} 0 \\ L_{\widehat{\sigma}_k} \end{smallmatrix}\right]\Gamma_k\left[\begin{smallmatrix} 0 \\ L_{\widehat{\sigma}_k} \end{smallmatrix}\right]^\top\right\|. \end{aligned} \tag{3.9}$$

Proof. From the analogue to (3.2),

$$X_0 = \begin{bmatrix} y_0 \\ y_0 - \widehat{y}_0 \end{bmatrix}, \qquad X_{j+1} = F(\sigma, \widehat{\sigma}_j)\,X_j - \begin{bmatrix} 0 \\ L_{\widehat{\sigma}_j} \end{bmatrix}\eta_j, \quad j \in \mathbb{N}, \tag{3.10}$$

we find, for $j \ge 1$,

$$\begin{aligned} X_{j+1} &= F(\sigma, \widehat{\sigma}_j)\left(F(\sigma, \widehat{\sigma}_{j-1})\,X_{j-1} - \begin{bmatrix} 0 \\ L_{\widehat{\sigma}_{j-1}} \end{bmatrix}\eta_{j-1}\right) - \begin{bmatrix} 0 \\ L_{\widehat{\sigma}_j} \end{bmatrix}\eta_j \\ &= F(\sigma, \widehat{\sigma}_j)F(\sigma, \widehat{\sigma}_{j-1})\,X_{j-1} - F(\sigma, \widehat{\sigma}_j)\begin{bmatrix} 0 \\ L_{\widehat{\sigma}_{j-1}} \end{bmatrix}\eta_{j-1} - \begin{bmatrix} 0 \\ L_{\widehat{\sigma}_j} \end{bmatrix}\eta_j \\ &= \Phi_{2|j-1}\,X_{j-1} - \Phi_{1|j}\begin{bmatrix} 0 \\ L_{\widehat{\sigma}_{j-1}} \end{bmatrix}\eta_{j-1} - \begin{bmatrix} 0 \\ L_{\widehat{\sigma}_j} \end{bmatrix}\eta_j. \end{aligned}$$

Repeating the argument, for $j \ge 2$,

$$\begin{aligned} X_{j+1} &= \Phi_{3|j-2}\,X_{j-2} - \Phi_{2|j-1}\begin{bmatrix} 0 \\ L_{\widehat{\sigma}_{j-2}} \end{bmatrix}\eta_{j-2} - \Phi_{1|j}\begin{bmatrix} 0 \\ L_{\widehat{\sigma}_{j-1}} \end{bmatrix}\eta_{j-1} - \begin{bmatrix} 0 \\ L_{\widehat{\sigma}_j} \end{bmatrix}\eta_j \\ &= \Phi_{j+1|0}\,X_0 - \sum_{k=0}^{j}\Phi_{k|j-k+1}\begin{bmatrix} 0 \\ L_{\widehat{\sigma}_{j-k}} \end{bmatrix}\eta_{j-k}. \end{aligned}$$

Following again the arguments in [16, Lem. 4.2, Rem. 3, Eq. (4.18)], now in the context of time-varying systems, we find $X_j = \mathcal{N}(M_j, \mathbf{\Pi}_j)$, where

$$M_j = \Phi_{j|0}\left[\begin{smallmatrix} m_0 \\ 0 \end{smallmatrix}\right], \qquad R_j = \left[\begin{smallmatrix} 0 \\ L_{\widehat{\sigma}_j} \end{smallmatrix}\right]\Gamma_j\left[\begin{smallmatrix} 0 \\ L_{\widehat{\sigma}_j} \end{smallmatrix}\right]^\top,$$

and

$$\begin{aligned} \mathbf{\Pi}_j &= \Phi_{1|j-1}\,\mathbf{\Pi}_{j-1}\,\Phi_{1|j-1}^\top + R_{j-1} = \Phi_{2|j-2}\,\mathbf{\Pi}_{j-2}\,\Phi_{2|j-2}^\top + \Phi_{1|j-1}\,R_{j-2}\,\Phi_{1|j-1}^\top + R_{j-1} \\ &= \Phi_{3|j-3}\,\mathbf{\Pi}_{j-3}\,\Phi_{3|j-3}^\top + \Phi_{2|j-2}\,R_{j-3}\,\Phi_{2|j-3}^\top + \Phi_{1|j-1}\,R_{j-2}\,\Phi_{1|j-1}^\top + Rj - 1, \\ &= \Phi_{j|0}\,\mathbf{\Pi}_0\,\Phi_{j|0}^\top + \sum_{k=0}^{j-1}\Phi_{j-k-1|k+1}\,R_k\,\Phi_{j-k-1|k+1}^\top. \end{aligned}$$

By Theorem 3.1 we have that $\varrho(F(\sigma,\sigma)) < 1$, which implies that there exists an induced matrix $\|\cdot\|$ such that $\|F(\sigma,\sigma))\| < 1$. By continuity, $\|F(\sigma,\varsigma))\| \le \lambda < 1$ for every ς in a small neighborhood of σ, where $\lambda < 1$. Therefore, for $(\widehat{\sigma}_j)_{j\in\mathbb{N}}$ contained in that neighborhood of σ, for the mean we obtain

$$\|M_j\| \leq \left\|\Phi_{j|0}\right\| \left\|\left[\begin{smallmatrix} m_0 \\ 0 \end{smallmatrix}\right]\right\| \leq \lambda^j \left\|\left[\begin{smallmatrix} m_0 \\ 0 \end{smallmatrix}\right]\right\|,$$

and for the covariance,

$$\begin{aligned}\|\mathbf{\Pi}_j\| &\leq \left\|\Phi_{j|0}\right\|^2 \|\mathbf{\Pi}_0\| + \sum_{k=0}^{j-1} \left\|\Phi_{j-k-1|k+1}\right\|^2 \|R_k\| \leq \lambda^{2j} \|\mathbf{\Pi}_0\| + \sum_{k=0}^{j-1} \lambda^{2(j-k-1)} \|R_k\| \\ &\leq \lambda^{2j} \|\mathbf{\Pi}_0\| + \frac{1}{1-\lambda^2} \sup_{k\in\mathbb{N}} \|R_k\|,\end{aligned}$$

which finishes the proof. □

In practice, the parameter will not be updated at every time step, but only at multiples $t = \kappa s$ of $s \in \mathbb{N}_+$, $\kappa \in \mathbb{N}$, cf., the bullet points in Figure 2 below.

4. Parameter estimation

For Theorem 3.6 above, it is essential to have a good estimate $\widehat{\sigma}$ of the unknown parameter $\sigma \in \mathbb{R}^p$. We adopt a Bayesian perspective: starting with a prior distribution of the unknown parameter, we update this prior belief sequentially based on new data from the controlled physical-virtual-twin system. We denote the estimate after the k-th update by $\widehat{\sigma}_k$, resulting in a sequence of estimates $(\widehat{\sigma}_k)_{k\geq 0}$. Each $\widehat{\sigma}_k$ is computed as the posterior mean from a Bayesian inverse problem using time-discrete, noisy, partial observations of the physical twin. This results in two nested time grids, as illustrated in Figure 2: a sequence of parameter estimation intervals indexed by $k \geq 1$, each of which contains a sequence of observations z_j, $j = 0, \ldots, s$. For each Bayesian update, a distribution of the unknown initial state $y_0^{(k)}$ conditioned on σ in the corresponding interval is required. It turns out that, this distribution is Gaussian and can be obtained by Kalman filtering as outlined in Section 5 below. This section is structured as follows: we commence with details on a single Bayesian parameter update for our specific model in Subsection 4.1 and Subsection 4.2. In Subsection 4.3, we present an illustrative explanation for the parameter estimation for two consecutive estimation intervals before we explain how the method is applied sequentially for an arbitrary number of estimation intervals in Subsection 4.4.

4.1. **Bayesian inverse problem.** In this and the subsequent subsection, we focus on a single parameter estimation, hence, k and $\widehat{\sigma}^{(k)}$ are fixed. To improve clarity, we omit the superscript notation indicating the k-th estimation interval. Let us assume a joint prior measure of the unknown initial condition y_0 and the unknown parameter σ

$$\mu_{\text{prior}}(\mathrm{d}\sigma, \mathrm{d}y_0) \;=\; \mu_{\text{prior}}(\mathrm{d}y_0 \mid \sigma)\, \mu_{\text{prior}}(\mathrm{d}\sigma),$$

where $\mu_{\text{prior}}(\mathrm{d}\sigma)$ is a probability measure on $\mathbb{R}^p$ (with Lebesgue density $\rho_{\text{prior}}(\sigma)$), and $\mu_{\text{prior}}(\mathrm{d}y_0 \mid \sigma)$ is a probability measure on $\mathbb{R}^n$. The joint posterior measure on (σ, y_0)

given data $\boldsymbol{z} = \begin{bmatrix} z_0^\top & \dots & z_s^\top \end{bmatrix}^\top$ is determined in the sense of Radon–Nikodym derivatives by

$$\frac{\mathrm{d}\mu_{\mathrm{post}}(\cdot \mid \boldsymbol{z})}{\mathrm{d}\mu_{\mathrm{prior}}}(\sigma, y_0) \ \propto \ \rho_{\mathrm{like}}(\boldsymbol{z} \mid \sigma, y_0),$$

provided that ρ_{like} is nonnegative and belongs to $L^1(\mu_{\mathrm{prior}})$. Equivalently, we can write

$$\mu_{\mathrm{post}}(\mathrm{d}\sigma, \mathrm{d}y_0 \mid \boldsymbol{z}) = \frac{1}{Z(\boldsymbol{z})} \rho_{\mathrm{like}}(\boldsymbol{z} \mid \sigma, y_0)\, \mu_{\mathrm{prior}}(\mathrm{d}y_0 \mid \sigma)\, \mu_{\mathrm{prior}}(\mathrm{d}\sigma),$$
$$Z(\boldsymbol{z}) = \int_{\mathbb{R}^p} \int_{\mathbb{R}^n} \rho_{\mathrm{like}}(\boldsymbol{z} \mid \sigma, y_0)\, \mu_{\mathrm{prior}}(\mathrm{d}y_0 \mid \sigma)\, \mu_{\mathrm{prior}}(\mathrm{d}\sigma).$$

Here, $0 < Z(\boldsymbol{z}) < \infty$ is the normalizing constant (sometimes called model evidence). Computing the joint posterior is challenging as the forward map $(\sigma, y_0) \mapsto \boldsymbol{z}$ is nonlinear and only implicitly defined through the solution of the coupled dynamical system, which makes the posterior non-Gaussian.

For the stabilization problem, the initial condition acts as a nuisance parameter: it is not of direct interest for the output-feedback design but enters the likelihood through the model dynamics and therefore must be accounted for in the inference. For this reason, we pursue inference on σ based on its marginal posterior obtained by integrating out the latent variable y_0

$$\mu_{\mathrm{post}}(\mathrm{d}\sigma \mid \boldsymbol{z}) \ = \ \frac{1}{Z(\boldsymbol{z})} \left(\int_{\mathbb{R}^n} \rho_{\mathrm{like}}(\boldsymbol{z} \mid \sigma, y_0)\, \mu_{\mathrm{prior}}(\mathrm{d}y_0 \mid \sigma) \right) \mu_{\mathrm{prior}}(\mathrm{d}\sigma).$$

The inner integral defines the marginal likelihood (see (4.7)) of the current estimation interval

$$\rho_{\mathrm{like}}(\boldsymbol{z} \mid \sigma) \ := \ \int_{\mathbb{R}^n} \rho_{\mathrm{like}}(\boldsymbol{z} \mid \sigma, y_0)\, \mu_{\mathrm{prior}}(\mathrm{d}y_0 \mid \sigma),$$

so that the marginal posterior of σ has Lebesgue density

$$\rho_{\mathrm{post}}(\sigma \mid \boldsymbol{z}) \ = \ \frac{1}{Z(\boldsymbol{z})} \rho_{\mathrm{like}}(\boldsymbol{z} \mid \sigma)\, \rho_{\mathrm{prior}}(\sigma), \quad Z(\boldsymbol{z}) = \int_{\mathbb{R}^p} \rho_{\mathrm{like}}(\boldsymbol{z} \mid \sigma)\, \rho_{\mathrm{prior}}(\sigma)\, \mathrm{d}\sigma. \tag{4.1}$$

Hence, for the inference of the posterior, the marginal likelihood ρ_{like} is required, which in turn requires the conditional prior on the initial condition $y_0 \mid \sigma$. While the prior on $y_0 \mid \sigma$ is a Gaussian measure on $\mathbb{R}^n$

$$y_0 \mid \sigma \sim \mathcal{N}\big(m_0(\sigma), \mathcal{C}_0(\sigma)\big), \tag{4.2}$$

the likelihood ρ_{like} depends crucially on the forward model, as explained in the subsequent subsection.

Remark 4.1. At the initialization step, see Figure 2, we assume that $y_0^{(1)} \sim \mathcal{N}(m_0^{(1)}, \mathcal{C}_0^{(1)})$. Thus, for the first parameter estimation, it can be assumed that $y_0^{(1)}$ is independent of σ. In this case, it holds that

$$\mu_{\mathrm{prior}}(\mathrm{d}y_0 \mid \sigma)\, \mu_{\mathrm{prior}}(\mathrm{d}\sigma) \ = \ \mu_{\mathrm{prior}}(\mathrm{d}\sigma)\, \mu_{\mathrm{prior}}(\mathrm{d}y_0).$$

For the k-th parameter estimation ($k \geq 2$), the distribution of the initial condition $y_0^{(k)} \mid \sigma$ is a Gaussian determined by the Kalman filter (see Section 5 below), and does depend on the unknown σ, see (5.1).

4.2. **The forward mapping.** The representation (2.10a) together with (2.10c) enables us to define the coupled state

$$x_j := \begin{bmatrix} y_j \\ \widehat{y}_j \end{bmatrix}, \quad 0 \leq j \leq s,$$

such that (2.10) can be reformulated as the linear time-discrete system

$$x_{j+1} = A_{\mathrm{cpl}}(\sigma, \widehat{\sigma})\, x_j + B_{\mathrm{cpl}}(\widehat{\sigma})\, \eta_j, \tag{4.3}$$

$$z_j = C_{\mathrm{cpl}}\, x_j + \eta_j, \tag{4.4}$$

where $x_0 = \begin{bmatrix} y_0^\top & \widehat{y}_0^\top \end{bmatrix}^\top$ and

$$A_{\mathrm{cpl}}(\sigma, \widehat{\sigma}) = \begin{bmatrix} \mathcal{A}_\sigma & \mathcal{B}_\sigma K_{\widehat{\sigma}} \\ -L_{\widehat{\sigma}} C & \mathcal{A}_{\widehat{\sigma}} + \mathcal{B}_{\widehat{\sigma}} K_{\widehat{\sigma}} + L_{\widehat{\sigma}} C \end{bmatrix}, \qquad B_{\mathrm{cpl}}(\widehat{\sigma}) = \begin{bmatrix} 0 \\ -L_{\widehat{\sigma}} \end{bmatrix}, \qquad C_{\mathrm{cpl}} = \begin{bmatrix} C & 0 \end{bmatrix}.$$

Iterating (4.3) yields that

$$x_j = A_{\mathrm{cpl}}^j(\sigma, \widehat{\sigma})\, x_0 + \sum_{r=0}^{j-1} A_{\mathrm{cpl}}^{j-1-r}(\sigma, \widehat{\sigma})\, B_{\mathrm{cpl}}(\widehat{\sigma})\, \eta_r.$$

Substituting into (4.4) and stacking the observations $\boldsymbol{z} = \begin{bmatrix} z_0^\top, \dots, z_s^\top \end{bmatrix}^\top$ and noises $\boldsymbol{\eta} = \begin{bmatrix} \eta_0^\top, \dots, \eta_s^\top \end{bmatrix}^\top$ yields that

$$\boldsymbol{z} = G_x(\sigma, \widehat{\sigma})\, x_0 + G_\eta(\sigma, \widehat{\sigma})\, \boldsymbol{\eta}, \tag{4.5}$$

where

$$G_x(\sigma, \widehat{\sigma}) = \begin{bmatrix} C_{\mathrm{cpl}} \\ C_{\mathrm{cpl}} A_{\mathrm{cpl}}(\sigma, \widehat{\sigma}) \\ C_{\mathrm{cpl}} A_{\mathrm{cpl}}^2(\sigma, \widehat{\sigma}) \\ \vdots \\ C_{\mathrm{cpl}} A_{\mathrm{cpl}}^s(\sigma, \widehat{\sigma}) \end{bmatrix} \in \mathbb{R}^{\ell(s+1) \times 2n},$$

$$G_\eta(\sigma, \widehat{\sigma}) = \begin{bmatrix} \mathrm{Id}_\ell & 0 & 0 & \cdots & 0 \\ C_{\mathrm{cpl}} B_{\mathrm{cpl}}(\widehat{\sigma}) & \mathrm{Id}_\ell & 0 & \cdots & 0 \\ C_{\mathrm{cpl}} A_{\mathrm{cpl}}(\sigma, \widehat{\sigma}) B_{\mathrm{cpl}} & C_{\mathrm{cpl}} B_{\mathrm{cpl}}(\widehat{\sigma}) & \mathrm{Id}_\ell & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ C_{\mathrm{cpl}} A_{\mathrm{cpl}}^{s-1}(\sigma, \widehat{\sigma}) B_{\mathrm{cpl}}(\widehat{\sigma}) & C_{\mathrm{cpl}} A_{\mathrm{cpl}}^{s-2}(\sigma, \widehat{\sigma}) B_{\mathrm{cpl}}(\widehat{\sigma}) & \cdots & C_{\mathrm{cpl}} B_{\mathrm{cpl}}(\widehat{\sigma}) & \mathrm{Id}_\ell \end{bmatrix},$$

where $G_\eta(\sigma, \widehat{\sigma}) \in \mathbb{R}^{\ell(s+1) \times \ell(s+1)}$ and Id_ℓ denotes the identity matrix in $\mathbb{R}^\ell$.

Recall that in this section, we consider a single parameter estimation, for which $\widehat{\sigma}$ is fixed. Furthermore, when using the data $\boldsymbol{z} = \begin{bmatrix} z_0{}^\top, \dots, z_s{}^\top \end{bmatrix}^\top$, the virtual twin trajectory at the time points $t_0, \dots, t_j$ is determined. In particular, $\widehat{y}_0$ is known. So, according to (4.2), given σ, the coupled initial state is Gaussian

$$x_0 = \begin{bmatrix} y_0 \\ \widehat{y}_0 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m_0(\sigma) \\ \widehat{y}_0 \end{bmatrix}, \begin{bmatrix} \mathcal{C}_0(\sigma) & 0 \\ 0 & 0 \end{bmatrix} \right).$$

By linearity, we conclude that

$$\begin{aligned} G_x(\sigma,\widehat{\sigma})x_0 &\sim \mathcal{N}\left(G_x(\sigma,\widehat{\sigma}) \begin{bmatrix} m_0(\sigma) \\ \widehat{y}_0 \end{bmatrix}, G_x(\sigma,\widehat{\sigma}) \begin{bmatrix} \mathcal{C}_0(\sigma) & 0 \\ 0 & 0 \end{bmatrix} G_x(\sigma,\widehat{\sigma})^\top \right), \\ G_\eta(\sigma,\widehat{\sigma})\boldsymbol{\eta} &\sim \mathcal{N}\left(0, G_\eta(\sigma,\widehat{\sigma})\boldsymbol{\Gamma}_{s+1}G_\eta(\sigma,\widehat{\sigma})^\top\right), \end{aligned}$$

where $\boldsymbol{\Gamma}_{s+1}$ is the block diagonal matrix containing the noise covariances $\Gamma_0, \dots, \Gamma_s$. From (4.5), we conclude that

$$\boldsymbol{z} \mid \sigma \sim \mathcal{N}\left(\bar{\boldsymbol{z}}_{\mathrm{cpl}}(\sigma,\widehat{\sigma}), \boldsymbol{\Gamma}_\eta(\sigma,\widehat{\sigma})\right),$$

where

$$\begin{aligned} \bar{\boldsymbol{z}}_{\mathrm{cpl}}(\sigma,\widehat{\sigma}) &:= G_x(\sigma,\widehat{\sigma}) \begin{bmatrix} m_0(\sigma) \\ \widehat{y}_0 \end{bmatrix}, \\ \boldsymbol{\Gamma}_\eta(\sigma,\widehat{\sigma}) &:= G_x(\sigma,\widehat{\sigma}) \begin{bmatrix} \mathcal{C}_0(\sigma) & 0 \\ 0 & 0 \end{bmatrix} G_x(\sigma,\widehat{\sigma})^\top + G_\eta(\sigma,\widehat{\sigma})\boldsymbol{\Gamma}_{s+1}G_\eta(\sigma,\widehat{\sigma})^\top. \end{aligned}$$

Then, the marginal likelihood of the data $\boldsymbol{z}$ given σ is the Gaussian density

$$\rho_{\mathrm{like}}(\boldsymbol{z} \mid \sigma) = \frac{\exp\Big(-\frac{1}{2}\big(\boldsymbol{z} - \bar{\boldsymbol{z}}_{\mathrm{cpl}}(\sigma,\widehat{\sigma})\big)^\top \boldsymbol{\Gamma}_\eta(\sigma,\widehat{\sigma})^{-1}\big(\boldsymbol{z} - \bar{\boldsymbol{z}}_{\mathrm{cpl}}(\sigma,\widehat{\sigma})\big)\Big)}{\sqrt{(2\pi)^{-s\ell}\,\det\big(\boldsymbol{\Gamma}_\eta(\sigma,\widehat{\sigma})\big)}}. \tag{4.6}$$

In the k-th parameter estimation step, this marginal likelihood ρ_{like} is the likelihood $\rho^{(k)}_{\mathrm{like}}(\boldsymbol{z}^{(k)} \mid \boldsymbol{z}^{(1)}, \dots, \boldsymbol{z}^{(k-1)}, \sigma)$ in (4.8), which is used to sample from the posterior distribution of the unknown parameter in the k-th parameter estimation problem. This corresponds to step 8 in Algorithm 1.

4.3. **Parameter estimation: two step illustration.** To illustrate the parameter estimation procedure introduced in this section, we present the method for two subsequent parameter estimation steps choosing $s+1 = 6$ observations in each estimation interval.

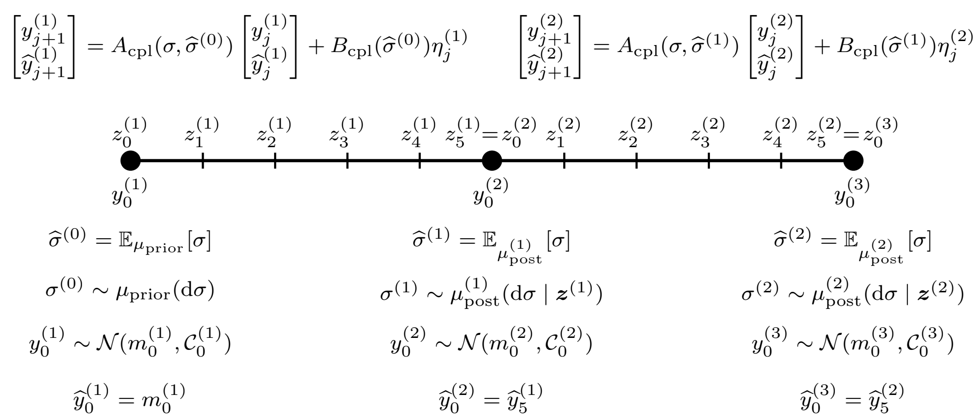


FIGURE 2. Illustration for $k = 1, 2$.

Our strategy is as follows:

Initialization. We are at the most left point in Figure 2. We are given $\mu_{\text{prior}}(\mathrm{d}\sigma)$ and $y_0^{(1)} \sim \mathcal{N}(m_0^{(1)}, \mathcal{C}_0^{(1)})$ by assumption. We initialize $\widehat{y}_0^{(1)} = m_0^{(1)}$ and $\widehat{\sigma}^{(0)} = \mathbb{E}_{\mu_{\text{prior}}}[\sigma]$ and run the coupled system characterized by $(A_{\text{cpl}}(\sigma, \widehat{\sigma}^{(0)}), B_{\text{cpl}}(\widehat{\sigma}^{(0)}))$ for $s = 5$ steps. While the system is running, we collect the performed measurements $\boldsymbol{z}^{(1)} = \left[z_0^{(1)\top}, \dots, z_5^{(1)\top}\right]^\top$.
1st estimation. We are at the center in Figure 2. We are given $y_0^{(1)} \sim \mathcal{N}(m_0^{(1)}, \mathcal{C}_0^{(1)})$ and $\mu_{\text{prior}}(\mathrm{d}\sigma)$. Using this, we can sample from the posterior $\sigma^{(1)} \sim \mu_{\text{post}}^{(1)}(\mathrm{d}\sigma \mid \boldsymbol{z}^{(1)})$ (marginalized over $y_0^{(1)}$) as detailed in Subsection 4.1 (and in Subsection 4.4 later on). We set

$$\widehat{\sigma}^{(1)} = \mathbb{E}_{\mu_{\text{post}}^{(1)}}[\sigma].$$

At this point, $\widehat{y}_j^{(1)}$ are known for $j \in \{0, \dots, 5\}$. They will be used to determine the distribution of the initial condition $y_0^{(2)}$ of the subsequent interval. We update $(A_{\text{cpl}}(\sigma, \widehat{\sigma}^{(1)}), B_{\text{cpl}}(\widehat{\sigma}^{(1)}))$ and run the coupled system $s = 5$ additional steps. While the system is running, we collect measurements $\boldsymbol{z}^{(2)} = \left[z_0^{(2)\top}, \dots, z_5^{(2)\top}\right]^\top$.
2nd estimation. We are at the most right point in Figure 2. As before, updating the parameter

$$\widehat{\sigma}^{(2)} = \mathbb{E}_{\mu_{\text{post}}^{(2)}}[\sigma],$$

requires samples from the posterior $\mu_{\text{post}}^{(2)}(\mathrm{d}\sigma \mid \boldsymbol{z}^{(2)})$. For this purpose, we use the posterior from the previous iteration $\mu_{\text{post}}^{(1)}(\mathrm{d}\sigma \mid \boldsymbol{z}^{(1)})$ as prior and marginalize over the initial condition $y_0^{(2)}$. To this end, we use Kalman filtering to obtain the conditional distribution $\mu(\mathrm{d}y_0^{(2)} \mid \boldsymbol{z}^{(1)}, \sigma)$ of $y_0^{(2)}$ given σ, and the data that is available at this point $\boldsymbol{z}^{(1)}$, i.e., the Kalman filter propagates $m_0^{(k)}$ and $\mathcal{C}_0^{(k)}$. Below, in Section 5, we provide more details on the Kalman filter. At this point, the virtual twin trajectory is determined completely.
For more than two estimation intervals, the exact same strategy is repeated: we update the matrices $(A_{\text{cpl}}(\sigma, \widehat{\sigma}^{(2)}), B_{\text{cpl}}(\widehat{\sigma}^{(2)}))$ and run the coupled system $s = 5$ additional steps. While the system is running, we collect measurements $\boldsymbol{z}^{(3)} = \left[z_0^{(3)\top}, \dots, z_5^{(3)\top}\right]^\top$, which we will use in the next parameter estimation, and repeat the steps described above. The repeated application of this method for an arbitrary number of estimation intervals is described in the subsequent subsection.

4.4. **Parameter estimation: general case.** Suppose we are at time t_j, $j = sk$ for $k \geq 2$ (for $k = 2$, this is the most right point in Figure 2). At this point, we wish to estimate $\widehat{\sigma}^{(k)}$, and we are given a distribution $\mu_{\text{post}}^{(k-1)}(\mathrm{d}\sigma \mid \boldsymbol{z}^{(k-1)})$ from the previous estimation interval, all measurements $\boldsymbol{z}^{(i)}$, $i \leq k$, and all values of the virtual twin up to this point. Let us denote the virtual twin evolution in the k-th interval by $\widehat{\boldsymbol{y}}^{(k)} = (\widehat{y}_0, \dots, \widehat{y}_{s-1})$ and let us point out that the virtual twin is a deterministic function of the observations, see (2.10c). Thus, given $\boldsymbol{z}^{(k-1)}$ and the initial condition $\widehat{y}_0$, the virtual twin $\widehat{\boldsymbol{y}}^{(k-1)}$ is fully determined. Conditioning on the virtual twin $\widehat{\boldsymbol{y}}^{(1)}, \dots, \widehat{\boldsymbol{y}}^{(k-1)}$ in addition to $\boldsymbol{z}^{(1)}, \dots, \boldsymbol{z}^{(k-1)}$ provides no additional information compared to conditioning on $\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(k-1)}$ only, that is

$$\rho_{\text{like}}^{(k)}(\boldsymbol{z}^{(k)} \mid \boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(k-1)}, \sigma, \widehat{\boldsymbol{y}}^{(1)}, \ldots, \widehat{\boldsymbol{y}}^{(k-1)}) = \rho_{\text{like}}^{(k)}(\boldsymbol{z}^{(k)} \mid \boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(k-1)}, \sigma).$$

Therefore, conditioning on $\widehat{\boldsymbol{y}}^{(1)}, \ldots, \widehat{\boldsymbol{y}}^{(k-1)}$ will be omitted in the following.

By the law of total probability we have that

$$\begin{aligned} &\rho_{\text{like}}^{(k)}(\boldsymbol{z}^{(k)} \mid \boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(k-1)}, \sigma) \\ &= \int_{\mathbb{R}^n} \rho(\boldsymbol{z}^{(k)} \mid y_0^{(k)}, \boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(k-1)}, \sigma)\, \rho_{\text{Kal}}(y_0^{(k)} \mid \boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(k-1)}, \sigma)\, \mathrm{d}y_0^{(k)}. \end{aligned}$$

The second term, $\rho_{\text{Kal}}(y_0^{(k)} \mid \boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(k-1)}, \sigma)$, is obtained by the Kalman filter as presented in Section 5 below. According to the measurement model, once the state y_j and the unknown parameter σ are known, the observation z_j does not depend on past states, the virtual twin trajectory, and past observations. Hence, the first term in the likelihood simplifies to

$$\rho^{(k)}(\boldsymbol{z}^{(k)} \mid y_0^{(k)}, \boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(k-1)}, \sigma) = \rho(\boldsymbol{z}^{(k)} \mid y_0^{(k)}, \sigma),$$

which yields that

$$\rho_{\text{like}}^{(k)}(\boldsymbol{z}^{(k)} \mid \boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(k-1)}, \sigma) = \int_{\mathbb{R}^n} \rho(\boldsymbol{z}^{(k)} \mid y_0^{(k)}, \sigma)\, \rho_{\text{Kal}}(y_0^{(k)} \mid \boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(k-1)}, \sigma)\, \mathrm{d}y_0^{(k)}. \tag{4.7}$$

Using the likelihood density of future observations conditioned on past data and the unknown parameter

$$\rho_{\text{like}}^{(k)}(\boldsymbol{z}^{(k)} \mid \boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(k-1)}, \sigma),$$

we can construct the posterior density of σ, and given all observations $\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(k)}$ using Bayes' theorem

$$\rho_{\text{post}}^{(k)}(\sigma \mid \boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(k)}) = \frac{\rho_{\text{like}}^{(k)}(\boldsymbol{z}^{(k)} \mid \boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(k-1)}, \sigma)\, \rho_{\text{post}}^{(k-1)}(\sigma \mid \boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(k-1)})}{\rho(\boldsymbol{z}^{(k)} \mid \boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(k-1)})}, \tag{4.8}$$

where the posterior from the previous step $\rho_{\text{post}}^{(k-1)}(\sigma \mid \boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(k-1)})$ serves as prior for the k-th update, and $\rho_{\text{post}}^{(0)} = \rho_{\text{prior}}$.

5. Estimation of the initial state distribution using Kalman-filtered Gaussians

5.1. **Kalman filtering on multiple intervals.** With reference to Figure 2, we refer to the intervals between the big bullets as estimation intervals. Each of these is divided into s time steps at which data are collected. For the parameter estimation at time t_j with $j = (k-1)s$ for $k \geq 2$, a distribution of the unknown initial condition of the physical system $y_0^{(k)}$ given all available data $\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(k-1)}$ and the unknown parameter σ is required, see (4.7). For the estimation of the first $\widehat{\sigma}^{(k)}$, i.e., with $k = 1$, we assume that $y_0 = y_0^{(1)} \sim \mathcal{N}(m_0^{(1)}, \mathcal{C}_0^{(1)})$. For the subsequent estimations of $\widehat{\sigma}^{(k)}$, we utilize the distribution

$$\mu(\mathrm{d}y_{(k-1)s} \mid \boldsymbol{z}^{(1)},\dots,\boldsymbol{z}^{(k-1)},\sigma) = \mathcal{N}(m_0^{(k)}(\sigma),\mathcal{C}_0^{(k)}(\sigma)), \tag{5.1}$$

with mean $m_0^{(k)}(\sigma) = m_{y|\widehat{y};s}^{(k-1)}(\sigma)$ and covariance $\mathcal{C}_0^{(k)}(\sigma) = C_{y|\widehat{y};s}^{(k-1)}(\sigma)$ to be defined in Subsection 5.2. In this subsection and in Subsection 5.2 below we explain how this distribution can be computed using the Kalman filter on every parameter estimation interval. We denote the Lebesgue density of $\mu(\mathrm{d}y_{(k-1)s} \mid \boldsymbol{z}^{(1)},\dots,\boldsymbol{z}^{(k-1)},\sigma)$ by $\rho_{\mathrm{Kal}}(y_0^{(k)} \mid \boldsymbol{z}^{(1)},\dots,\boldsymbol{z}^{(k-1)},\sigma)$, see (4.7).

By the law of total probability, conditioning on the previous interval's initial state $y_0^{(k-1)}$ gives

$$\begin{aligned}&\rho_{\mathrm{Kal}}\big(y_0^{(k)} \mid \boldsymbol{z}^{(1)},\dots,\boldsymbol{z}^{(k-1)},\sigma\big) \\ &\quad= \int_{\mathbb{R}^n} \rho\big(y_0^{(k)} \mid y_0^{(k-1)},\boldsymbol{z}^{(1)},\dots,\boldsymbol{z}^{(k-1)},\sigma\big)\rho\big(y_0^{(k-1)} \mid \boldsymbol{z}^{(1)},\dots,\boldsymbol{z}^{(k-1)},\sigma\big)\,\mathrm{d}y_0^{(k-1)}.\end{aligned} \tag{5.2}$$

In the multi-interval setting with $y_0^{(k)} = y_{(k-1)s}$, we have the Markov property

$$y_0^{(k)} \;\perp\; \boldsymbol{z}^{(1)},\dots,\boldsymbol{z}^{(k-2)} \mid y_0^{(k-1)},\boldsymbol{z}^{(k-1)},\sigma,$$

i.e., the initial state of the k-th interval depends on all past measurements only through the final state and measurements of the previous interval. Hence, for the first factor on the right-hand side of (5.2), it holds that

$$\rho\big(y_0^{(k)} \mid y_0^{(k-1)},\boldsymbol{z}^{(1)},\dots,\boldsymbol{z}^{(k-1)},\sigma\big) = \rho\big(y_0^{(k)} \mid y_0^{(k-1)},\boldsymbol{z}^{(k-1)},\sigma\big).$$

Regarding the second factor on the right-hand side of (5.2), given σ and $\boldsymbol{z}^{(k-2)}$, the initial condition $y_0^{(k-1)}$ is clearly independent of the future measurements $\boldsymbol{z}^{(k-1)}$, so

$$\rho\big(y_0^{(k-1)} \mid \boldsymbol{z}^{(1)},\dots,\boldsymbol{z}^{(k-1)},\sigma\big) = \rho\big(y_0^{(k-1)} \mid \boldsymbol{z}^{(1)},\dots,\boldsymbol{z}^{(k-2)},\sigma\big).$$

This yields the recursion

$$\begin{aligned}&\rho_{\mathrm{Kal}}\big(y_0^{(k)} \mid \boldsymbol{z}^{(1)},\dots,\boldsymbol{z}^{(k-1)},\sigma\big) \\ &\quad= \int_{\mathbb{R}^n} \rho\big(y_0^{(k)} \mid y_0^{(k-1)},\boldsymbol{z}^{(k-1)},\sigma\big)\rho\big(y_0^{(k-1)} \mid \boldsymbol{z}^{(1)},\dots,\boldsymbol{z}^{(k-2)},\sigma\big)\,\mathrm{d}y_0^{(k-1)}.\end{aligned}$$

Solving this recursion until the first estimation interval with $y_0^{(1)} = y_0$ gives

$$\begin{aligned}&\rho_{\mathrm{Kal}}\big(y_0^{(k)} \mid \boldsymbol{z}^{(1)},\dots,\boldsymbol{z}^{(k-1)},\sigma\big) \\ &\quad= \int_{\mathbb{R}^{n\times(k-1)}} \left(\prod_{i=1}^{k-1} \rho(y_0^{(i+1)} \mid y_0^{i},\boldsymbol{z}^{(i)},\sigma)\right) \rho(y_0^{(1)})\,\mathrm{d}y_0^{(1)}\cdots\mathrm{d}y_0^{(k-1)}.\end{aligned}$$

For each k, this conditional density can be computed exactly by means of the Kalman filter. In particular, the Kalman filter admits a natural extension across multiple estimation intervals: the density over all k estimation intervals defined by the above recursion is obtained by successive applications of the Kalman filter, using the posterior from the previous estimation interval as the initial condition for the next one. Accordingly, in the following section we present the Kalman filter formulation for a single estimation interval.

5.2. **Kalman filtering on a single estimation interval.** The aim of this section is to compute the distribution (5.1) for a single estimation interval. We will use the Kalman filter to obtain the distribution of the coupled virtual-physical twin state and then condition on the realization of the virtual twin. This is equivalent to interpreting the virtual twin as an additional noise-free observation and applying the Kalman filter to this augmented observation model.

Given σ, used in $\mathcal{A}_\sigma$, the coupled system after the k-th parameter update is linear and Gaussian. Precisely, it is

$$\begin{aligned}
\begin{bmatrix} y_{j+1}^{(k)} \\ \widehat{y}_{j+1}^{(k)} \end{bmatrix} &= A_{\mathrm{cpl}}(\sigma, \widehat{\sigma}^{(k)}) \begin{bmatrix} y_{j}^{(k)} \\ \widehat{y}_{j}^{(k)} \end{bmatrix} + B_{\mathrm{cpl}}(\widehat{\sigma}^{(k)})\eta_j, && j = 0, \dots, s-1, \\
\begin{bmatrix} y_{0}^{(k)} \\ \widehat{y}_{0}^{(k)} \end{bmatrix} &\sim \mathcal{N}\left(\begin{bmatrix} m_0^{(k)}(\sigma) \\ \widehat{y}_0^{(k)} \end{bmatrix}, \begin{bmatrix} \mathcal{C}_0^{(k)}(\sigma) & 0 \\ 0 & 0 \end{bmatrix} \right), \\
z_j^{(k)} &= C_{\mathrm{cpl}} \begin{bmatrix} y_{j}^{(k)} \\ \widehat{y}_{j}^{(k)} \end{bmatrix} + \eta_j, && j = 0, \dots, s,
\end{aligned}$$

where A_{cpl}, B_{cpl}, and C_{cpl} are as in Subsection 4.2. Recall that $\widehat{y}_0^{(k)}$, the initial condition of the virtual twin, is known when we estimate the unknown parameter σ. This is modeled by the covariance identical to zero above.

For fixed $\widehat{\sigma}$, the conditional distribution of the pair $x_j^{(k)} = \begin{bmatrix} y_j^{(k)} & \widehat{y}_j^{(k)} \end{bmatrix}$, given the data $z_j^{(k)}$ and the parameter σ, is again Gaussian. It is determined by the Kalman filter in terms of the mean and covariance. The Kalman filter reads [15, Chapter 5]:
Initialization:

$$m_{0|0}^{(k)}(\sigma) = \begin{bmatrix} m_0^{(k)}(\sigma) \\ \widehat{y}_0^{(k)} \end{bmatrix}, \quad \text{and} \quad \mathcal{C}_{0|0}^{(k)}(\sigma) = \begin{bmatrix} \mathcal{C}_0^{(k)}(\sigma) & 0 \\ 0 & 0 \end{bmatrix}.$$

Then, for $j = 1, \dots, s$:
Prediction:

$$\begin{aligned}
&m_{j|j-1}^{(k)}(\sigma) = A_{\mathrm{cpl}}(\sigma, \widehat{\sigma}^{(k)}) m_{j-1|j-1}^{(k)}(\sigma), \\
&\mathcal{C}_{j|j-1}^{(k)}(\sigma) = A_{\mathrm{cpl}}(\sigma, \widehat{\sigma}^{(k)}) \mathcal{C}_{j-1|j-1}^{(k)}(\sigma) A_{\mathrm{cpl}}(\sigma, \widehat{\sigma}^{(k)})^\top + B_{\mathrm{cpl}}(\widehat{\sigma}^{(k)}) \Gamma_j B_{\mathrm{cpl}}(\widehat{\sigma}^{(k)})^\top.
\end{aligned}$$

Innovation and Kalman Gain:

$$\begin{aligned}
\text{innovation: } \nu_j(\sigma) &= z_j^{(k)} - C_{\mathrm{cpl}}\, m_{j|j-1}^{(k)}(\sigma), \\
\text{innovation covariance: } S_j(\sigma) &= C_{\mathrm{cpl}}\, \mathcal{C}_{j|j-1}^{(k)}(\sigma)\, C_{\mathrm{cpl}}^\top + \Gamma_j, \\
\text{Kalman gain: } K_j(\sigma) &= \mathcal{C}_{j|j-1}^{(k)}(\sigma)\, C_{\mathrm{cpl}}^\top\, S_j(\sigma)^{-1}.
\end{aligned}$$

Coupled update:

$$\begin{aligned}
\mathfrak{m}_{j|j}(\sigma) &= m_{j|j-1}^{(k)}(\sigma) + K_j(\sigma)\, \nu_j(\sigma), \\
\mathfrak{C}_{j|j}(\sigma) &= \mathcal{C}_{j|j-1}^{(k)}(\sigma) - K_j(\sigma)\, S_j(\sigma)\, K_j(\sigma)^\top.
\end{aligned}$$

Conditioning on the virtual twin:

$$m^{(k)}_{y|\widehat{y};j} = [\mathfrak{m}_{j|j}(\sigma)]_{1:d} - [\mathfrak{C}_{j|j}(\sigma)]_{d+1:2d,1:d}[\mathfrak{C}_{j|j}(\sigma)]^{-}_{d+1:2d,d+1:2d}\left(\widehat{y}^{(k)}_j - [\mathfrak{m}_{j|j}(\sigma)]_{d+1:2d}\right),$$
$$C^{(k)}_{y|\widehat{y};j} = [\mathfrak{C}_{j|j}(\sigma)]_{1:d,1:d} - [\mathfrak{C}_{j|j}(\sigma)]_{1:d,d+1:2d}[\mathfrak{C}_{j|j}(\sigma)]^{-}_{d+1:2d,d+1:2d}[\mathfrak{C}_{j|j}(\sigma)]_{d+1:2d,1:d},$$

where $[v]_{n:l}$ denotes a vector containing components $v_n, v_{n+1}, \dots, v_l$ of a vector v, and $[A]_{n:l,p:q}$ denotes the submatrix containing the rows $n, n+1, \dots, l$ and columns $p, p+1, \dots, q$ of a matrix A. Here, $[\mathfrak{C}_{j|j}(\sigma)]^{-}_{d+1:2d,d+1:2d}$ denotes the generalized inverse of $[\mathfrak{C}_{j|j}(\sigma)]_{d+1:2d,d+1:2d}$, see [13, Prop. 3.13].
Conditioned update:

$$m^{(k)}_{j|j}(\sigma) = \begin{bmatrix} m^{(k)}_{y|\widehat{y};j} \\ \widehat{y}^{(k)}_j \end{bmatrix} \quad \text{and} \quad C^{(k)}_{j|j}(\sigma) = \begin{bmatrix} \mathcal{C}^{(k)}_{y|\widehat{y};j} & 0 \\ 0 & 0 \end{bmatrix}.$$

In the coupled update step we compute, for fixed $\widehat{\sigma}$ and $1 \le j \le s$, the conditional distribution of the coupled state given the data $(z^{(k)}_0, \dots, z^{(k)}_j)$ and given the parameter σ. Since the virtual twin component $\widehat{y}^{(k)}_j$ is known at the time of the observation $z^{(k)}_j$, we then compute the conditional distribution of the physical state $y^{(k)}_j$ given the virtual twin $\widehat{y}^{(k)}_j$.

In the next estimation interval, we initialize with $m^{(k+1)}_{0|0}(\sigma) = m^{(k)}_{s|s}(\sigma)$ and $\mathcal{C}^{(k+1)}_{0|0}(\sigma) = C^{(k)}_{s|s}(\sigma)$. This procedure provides us with the distribution of the initial condition (5.1) on each estimation interval. It enters the parameter estimation via its density denoted by ρ_{Kal} in (4.7).

6. Numerical experiments

Our digital twin strategy is summarized in Algorithm 1. In Algorithm 2, a detailed description of the computation of the posterior distribution in Step 13 of Algorithm 1 is provided. With reference to Remark 3.2, unless otherwise specified, we set $R = 1$ and $Q = 10^{-12} \cdot \mathrm{Id}_2$. Further details on the implementation of the examples listed above are provided with the examples. These are a harmonic oscillator, a spring-damper-system, and a finite element discretization of a parabolic partial differential equation. The oscillator example and the spring-damper-system are illustrated in Figure 3.

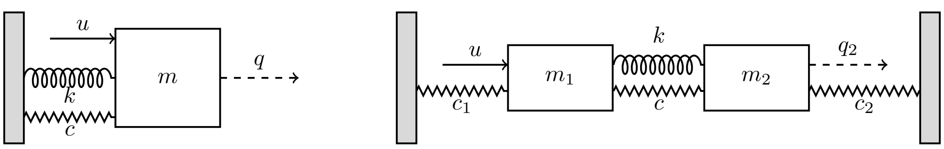


Figure 3. Comparison of the single mass oscillator system (left) and a two-mass (right) spring–damper system.

Algorithm 1 Digital twin stabilization

Require: Prior distribution of the parameters, prior distribution of the initial condition $\mathcal{N}(m_0^{(1)}, \mathcal{C}^{(1)})$, control operator B, output operator C, initial guess for the digital twin, e.g., $\widehat{y}_0^{(1)} = m_0^{(1)}$.

Ensure: Feedback control input $u = K_{\widehat{\sigma}} \widehat{y}$.

$k = 0$;
Compute an estimate $\widehat{\sigma}^{(k)}$ for $\sigma \sim \mu_{\text{prior}}$, e.g., $\widehat{\sigma}^{(k)} = \mathbb{E}_{\mu_{\text{prior}}}[\sigma]$;
Assemble $A_{\widehat{\sigma}^{(k)}}$ and compute $\mathcal{A}_{\widehat{\sigma}^{(k)}}$ and $\mathcal{B}_{\widehat{\sigma}^{(k)}}$ as in Remark 2.1;
Construct $K_{\widehat{\sigma}^{(k)}}$ and $L_{\widehat{\sigma}^{(k)}}$ as in Remark 3.2;
while the coupled system (2.10) is running **do**
  $k \leftarrow k + 1$;
  for $j = 0 : s$ **do**
    Collect data $z_j^{(k)} = z(t_j)$ from the physical model according to (2.2);
  end for
  if $k > 1$ **then**
    Run Kalman filter (see (5.1)) to obtain $y_0^{(k)} \sim \mathcal{N}(m_0^{(k)}, \mathcal{C}_0^{(k)})$;
  end if
  Update $\widehat{\sigma}$ based on the posterior $\mu_{\text{post}}^{(k)}$ (see (4.8)), e.g., $\widehat{\sigma}^{(k)} = \mathbb{E}_{\mu_{\text{post}}^{(k)}}[\sigma]$;
  Construct $K_{\widehat{\sigma}}^{(k)}$ and $L_{\widehat{\sigma}}^{(k)}$, assemble $A_{\widehat{\sigma}}^{(k)}$ and compute $\mathcal{A}_{\widehat{\sigma}^{(k)}}$ and $\mathcal{B}_{\widehat{\sigma}^{(k)}}$;
end while

6.1. **Oscillator.** We consider a mass attached to a spring, as depicted in Figure 3 (left).

Let $x(t)$ be the position of the mass m and let x_0 be an equilibrium position at rest. Now, consider the relative position $q(t) = x(t) - x_0$. For the dynamics, we have the oscillator model (for small q)

$$m\ddot{q} = -c\dot{q} - kq + u. \tag{6.1}$$

The input control forcing u acts on the mass m and we measure the (relative) position q. Hereafter we take $k = m = 1$ and consider the damping constant $c = \sigma$ as the uncertain parameter. Denoting $Q = (q, \dot{q})$, we write the system as

$$\dot{Q} = A_\sigma Q + Bu, \tag{6.2a}$$

$$\text{with} \quad A_\sigma = \begin{bmatrix} 0 & 1 \\ -1 & -\sigma \end{bmatrix}, \qquad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \qquad C = \begin{bmatrix} 1 & 0 \end{bmatrix}.$$

The Kalman controllability $\mathcal{K}_{\mathcal{C}}$ and observability $\mathcal{K}_{\mathcal{O}}$ matrices are given by

$$\mathcal{K}_{\mathcal{C}} = \begin{bmatrix} 0 & 1 \\ 1 & -\sigma \end{bmatrix}, \qquad \mathcal{K}_{\mathcal{O}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

which are both full-rank. In particular, (A_σ, B) is stabilizable and (A_σ, C) is detectable for all $\sigma \in \mathbb{R}$.

Remark 6.1. Usually, the damping parameter is positive $\sigma > 0$. The damping free case $\sigma = 0$ is relevant as well. We also test our strategy with the academic case $\sigma < 0$, which is of interest because in this case the free dynamics is exponentially unstable and the input will have to be able to counteract this instability.

Algorithm 2 Sequential Monte Carlo with Resample–Move Step

Require: Prior particles $\Sigma_0 = \{\sigma_i\}_{i=1}^N$, observations $\boldsymbol{z}^{(k)}$ on intervals $\{\mathcal{I}_{\text{obs}}^k\}$, number of MCMC steps n_{MCMC}, proposal scale ε.

Ensure: Particle approximation Σ_k of the posterior at each interval.

Initialize sequential likelihoods $\rho_{\text{like}}^0(\boldsymbol{z}^{(0)} \mid \sigma_i) = 1$ and particle set $\Sigma_1 = \Sigma_0$;
for $k = 1, 2, \dots$ **do**
  for $i = 1, \dots, N$ **do**
    Compute likelihood increment $\rho_{\text{like}}^{\text{new}}(\boldsymbol{z}^{(k)} \mid \sigma_i)$ on $\mathcal{I}_{\text{obs}}^k$ as in (4.6);
    Update sequential likelihood:

$$\rho_{\text{like}}^{(k)}(\boldsymbol{z}^{(k)} \mid \sigma_i) = \rho_{\text{like}}^{(k-1)}(\boldsymbol{z}^{(k-1)} \mid \sigma_i) \cdot \rho_{\text{like}}^{\text{new}}(\boldsymbol{z}^{(k)} \mid \sigma_i);$$

    Compute normalized weights:

$$w_i \propto \rho_{\text{like}}^{(k)}(\boldsymbol{z}^{(k)} \mid \sigma_i), \quad \sum_{i=1}^N w_i = 1;$$

  end for
  Resample N particles according to $w_1, \dots, w_N$:

$$\Sigma_k = \{\sigma_{\iota_i} \in \Sigma_{k-1} \,:\, i = 1, \dots, N\}, \quad \text{with } \Pr(\iota_i = \ell) = w_\ell \text{ for } \ell = 1, \dots, N;$$

  for $s = 1, \dots, n_{\text{MCMC}}$ **do**
    for $i = 1, \dots, N$ **do**
      Propose candidate $\sigma_i' = \sigma_i + \varepsilon\,\xi$, $\xi \sim \mathcal{N}(0, I)$;
      Evaluate likelihood increment for candidate $\rho_{\text{like}}^{\text{new}}(\boldsymbol{z}^{(k)} \mid \sigma_i')$ according to (4.6):
      Approximate candidate sequential likelihood:

$$\rho_{\text{like}}'(\boldsymbol{z}^{(k)} \mid \sigma_i') = \rho_{\text{like}}^{(k-1)}(\boldsymbol{z}^{(k-1)} \mid \sigma_i) \cdot \rho_{\text{like}}^{\text{new}}(\boldsymbol{z}^{(k)} \mid \sigma_i');$$

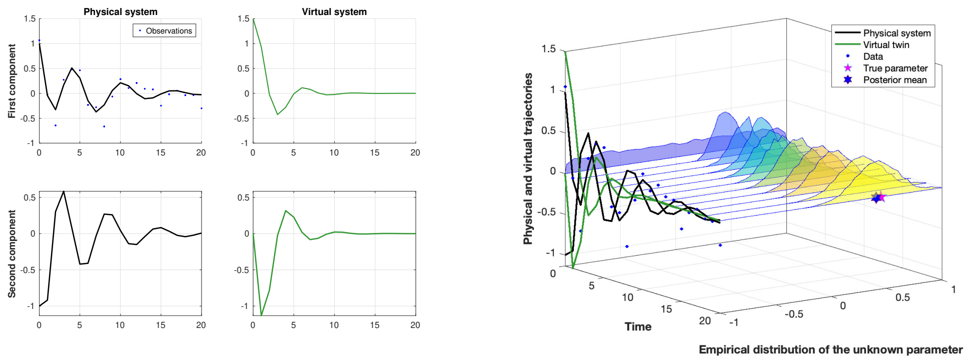
      Accept candidate with probability

$$\alpha = \min\left(1, \frac{\rho_{\text{like}}'(\boldsymbol{z}^{(k)} \mid \sigma_i')\rho_{\text{prior}}(\sigma_i')}{\rho_{\text{like}}^{(k)}(\boldsymbol{z}^{(k)} \mid \sigma_i)\rho_{\text{prior}}(\sigma_i)}\right);$$

      if accepted **then**
        $\sigma_i = \sigma_i'$;
        $\rho_{\text{like}}^{(k)}(\boldsymbol{z} \mid \sigma_i) = \rho_{\text{like}}'(\boldsymbol{z} \mid \sigma_i')$;
      end if
    end for
  end for
end for

The numerical results displayed in Figure 4 and Figure 5 are obtained using $N = 5000$ particles initialized from a uniform prior distribution on the interval $[-1, 1]$. The true parameter is $\sigma = \frac{\pi}{7} \approx 0.4488$ in the stable case and $\sigma = -\frac{\pi}{7} \approx -0.4488$ in the unstable case. In the stable case, the estimate $\widehat{\sigma}$ is updated at times $2, 4, 6, \dots, 20$ with observations at times $1, 2, 3, \dots, 20$, while in the unstable case, $\widehat{\sigma}$ is updated more frequently at times $1, 2, 3, \dots, 20$ with more observations at times $0.25, 0.5, 0.75, \dots, 20$. In both cases, we initialize with $y_0 = \begin{bmatrix} 1 & -1 \end{bmatrix}^\top$ and $\widehat{y}_0 = \begin{bmatrix} 1.5 & 0 \end{bmatrix}^\top$. Further, we assume that the distribution of the initial condition is $\mathcal{N}\Big(\begin{bmatrix} 1.5 & -1.1 \end{bmatrix}^\top, \mathrm{Id}_2\Big)$ and that the distribution of the noise is $\mathcal{N}(0, 0.015\|y_0\|)$. The matrices $\mathcal{A}_\sigma, \mathcal{B}_\sigma$ are computed using the matrix exponential as $\mathcal{A}_\sigma = e^{A_\sigma \Delta t}$ and $\mathcal{B}_\sigma = \int_0^{\Delta t} e^{A_\sigma(\Delta t - \tau)}\,\mathrm{d}\tau\, B$, whereas the matrices $\mathcal{A}_{\widehat{\sigma}}, \mathcal{B}_{\widehat{\sigma}}$ are computed using a Crank–Nicolson discretization as in Remark 2.1. We use $n_{\mathrm{MCMC}} = 50$ steps with proposal scale $\epsilon = 0.015$.

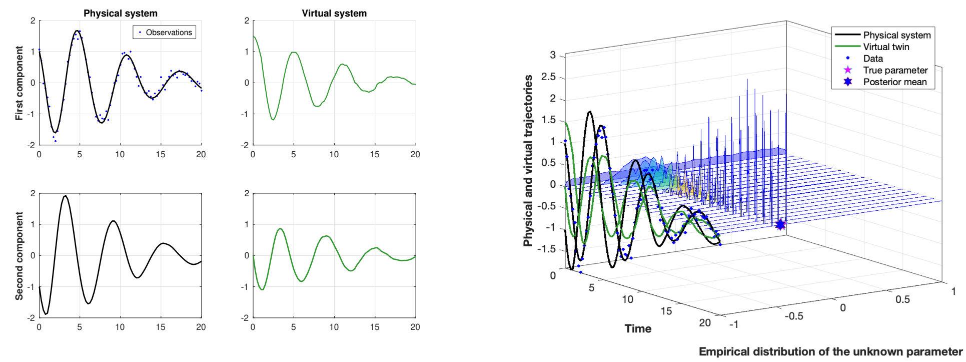
Comparing the stable and unstable cases, we make the following observations. For unstable systems, differences caused by small parameter variations amplify over time. As a consequence, the outputs $m_0^{(k+1)}$ of the Kalman filters (see Subsection 5.2) vary a lot for different parameters. Hence, their difference to the data, which occurs in the likelihood (4.6), varies significantly across different realizations of the unknown parameter. This leads to notably centered posterior density functions in Figure 5. On the other hand, for stable systems, the differences caused by small parameter variations vanish over time, leading to less centered posterior densities, see Figure 4).



(A) Trajectories of the coupled physical-virtual system.

(B) Online estimation of the empirical densities of the uncertain parameter.

FIGURE 4. The true parameter yields a stable system.



(A) Trajectories of the coupled physical-virtual system.

(B) Online estimation of the empirical densities of the uncertain parameter.

FIGURE 5. The true parameter yields an unstable system.

In further numerical experiments, which are not displayed in this manuscript, we discovered that the update frequency of $\widehat{\sigma}_k$ is crucial for the parameter estimation quality and stabilization. For instance, decreasing the update frequency, estimating at times $4, 8, 12, 16, 20$ results in failure of stabilization in the unstable case. Compared to the influence of the update frequency, the noise level and the number of observation have a smaller effect on the results.

6.2. **Spring-damper-system.** We next consider a finite-dimensional system which is both stabilizable and detectable, but neither controllable nor observable, see Figure 3 (right).

Let $(x_1(t), x_2(t)) \in [0, L] \times [0, L]$ be the position of the pair of masses (m_1, m_2), let $(x_{1,0}, x_{2,0})$ be an equilibrium position at rest, and define the relative position $q(t) = (q_1(t), q_2(t)) = (x_1(t) - x_{1,0}, x_2(t) - x_{2,0})$. For the dynamics, we have the model (for small q)

$$m_1\ddot{q}_1 = -c_1\dot{q}_1 - c(\dot{q}_1 - \dot{q}_2) - k(q_1 - q_2) + u, \tag{6.3}$$

$$m_2\ddot{q}_2 = -c_2\dot{q}_2 - c(\dot{q}_2 - \dot{q}_1) - k(q_2 - q_1). \tag{6.4}$$

The input control forcing u acts on the mass m_1 and we measure the (relative) position q_2 of mass m_2. Hereafter we consider the case $m_1 = m_2 = c_1 = c = c_2 = 1$ with the uncertain parameter $\sigma = k > 0$. Denoting $Q = (q_1, q_2, \dot{q}_1, \dot{q}_2)$, we write the system as

$$\dot{Q} = A_\sigma Q + Bu, \tag{6.5a}$$

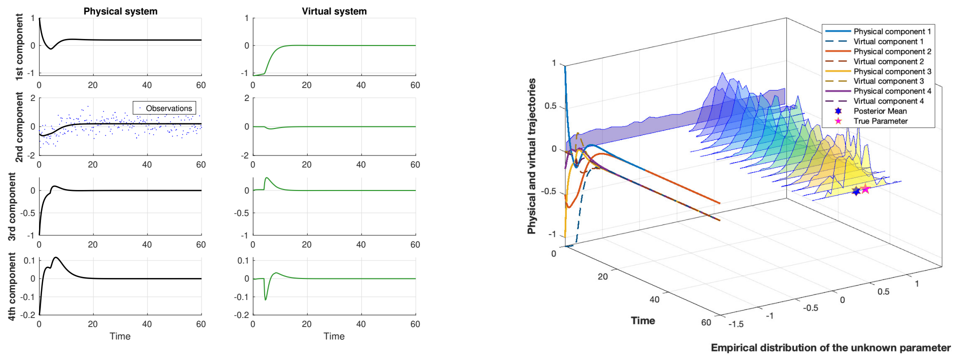
$$\text{with} \quad A_\sigma = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -\sigma & \sigma & -2 & 1 \\ \sigma & -\sigma & 1 & -2 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix}. \tag{6.5b}$$

Though the physical model above is considered with $\sigma > 0$, here we allow σ to vary in all of $\mathbb{R}$. This is motivated by numerical tests which include the challenging case where the dynamics is exponentially unstable, if $\sigma < 0$.

It can be shown that the system (6.5) is stabilizable for all $\sigma \in \mathbb{R}$ and detectable for all $\sigma \neq 0$. Further, in case $\sigma = 1$ the system is neither controllable nor observable.

The numerical results displayed in Figure 6 and Figure 7 are obtained using $N = 5000$ particles initialized from a uniform prior distribution on the interval $[-1.5, 1.5]$. The true parameter is $\sigma = \frac{\pi}{7} \approx 0.4488$ in the stable case and $\sigma = -0.1$ in the unstable case. In the stable case, the estimate $\widehat{\sigma}$ is updated at times $4, 8, 12, \ldots, 60$ with observations at times $0.25, 0.5, 0.75, \ldots, 60$, while in the unstable case, $\widehat{\sigma}$ is updated more frequently at times $0.5, 1, 1.5, \ldots, 60$ with more observations at times $1/16, 2/16, 3/16, \ldots, 60$. In both cases, we initialize with $y_0 = \begin{bmatrix} 1 & -0.5 & -1 & -0.2 \end{bmatrix}^\top$ and $\widehat{y}_0 = \begin{bmatrix} -1.1 & 0 & 0 & 0 \end{bmatrix}^\top$. Furthermore, we assume that the distribution of the initial condition is the Gaussian $\mathcal{N}\Big(\begin{bmatrix} 1.5 & -1.1 & -0.8 & -0.18 \end{bmatrix}^\top, \mathrm{Id}_4\Big)$ and that the distribution of the noise is $\mathcal{N}(0, 0.15\|y_0\|)$. The matrices $\mathcal{A}_\sigma, \mathcal{B}_\sigma$ are computed using the matrix exponential as $\mathcal{A}_\sigma = e^{A_\sigma \Delta t}$ and $\mathcal{B}_\sigma = \int_0^{\Delta t} e^{A_\sigma(\Delta t - \tau)}\,\mathrm{d}\tau\, B$, whereas the matrices $\mathcal{A}_{\widehat{\sigma}}, \mathcal{B}_{\widehat{\sigma}}$ are computed using a Crank–Nicolson discretization as in Remark 2.1. We use $n_{\mathrm{MCMC}} = 10$ steps with proposal scale $\epsilon = 0.1$.

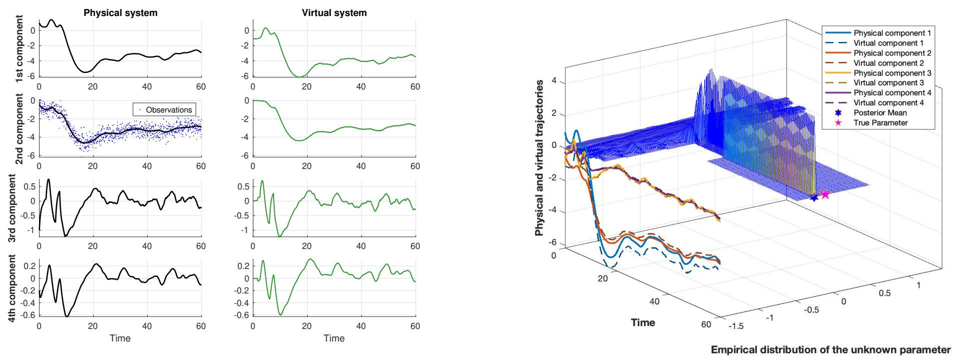
We observe that the system is stabilized in both the stable as well as the unstable case. In the unstable case, the posterior densities are more centered as compared to the stable case. Furthermore, we observe that the numerical stabilization of the unstable system fails if we update the estimate less frequently or have fewer observations.



(A) Trajectories of the coupled digital-physical system.

(B) Online estimation of the empirical densities of the uncertain parameter.

FIGURE 6. The true parameter yields a stable system.



(A) Trajectories of the coupled digital-physical system.

(B) Online estimation of the empirical densities of the uncertain parameter.

FIGURE 7. The true parameter yields an unstable system.

6.3. **Diffusion-reaction equation.** We consider a finite element semi-discretization of a parameterized diffusion-reaction equation.

Let $D = (0,1)$ and, for every σ let y_σ solve

$$\tfrac{\partial}{\partial t} y_\sigma(t,\xi) - \nabla\cdot(a_\sigma(\xi)\nabla y_\sigma(t,\xi)) + c\, y_\sigma(t,\xi) = \sum_{i=1}^{N_{\rm a}} u_i(t)\mathbf{1}_{O_i}, \quad (t,\xi)\in(0,T]\times D, \tag{6.6a}$$

$$\frac{\partial y_\sigma(t,\xi)}{\partial {\rm n}} = 0, \qquad (t,\xi)\in[0,T]\times\partial D, \tag{6.6b}$$

$$y_\sigma(t,\xi) = y_0, \qquad (t,\xi)\in\{t=0\}\times D, \tag{6.6c}$$

where $\mathbf{1}_{O_i}$ denotes the support of the i-th actuator on the open set $O_i \subset D$, for $1 \leq i \leq N_\mathrm{a}$, and $c \in \mathbb{R}$ is a constant reaction coefficient. The uncertain, spatially varying diffusion coefficient is parameterized log-linearly by $\sigma = [\sigma_1, \dots, \sigma_p] \in \Sigma_p \subset \mathbb{R}^p$ via

$$a_\sigma(\xi) = \exp\Big(\sum_{j=1}^{p} \sigma_j \psi_j(\xi)\Big),$$

with basis functions

$$\psi_{2j-1}(\xi) = (2j-1)^{-\vartheta} \cos(j\pi\xi), \qquad \psi_{2j}(\xi) = (2j)^{-\vartheta} \sin(j\pi\xi),$$

for $1 \leq j \leq \lceil p/2 \rceil$, so that $\vartheta > 0$ controls the decay of the parametric influence of higher modes, guaranteeing $0 < a_{\min} \leq a_\sigma(\xi) \leq a_{\max} < \infty$ for σ in a bounded set, such as $\Sigma_p := [-10, 10]$.

We discretize (6.6) in space using continuous piecewise-linear finite elements on a uniform mesh of D with d nodes and mesh width $h = 1/d$, with nodal basis $\{\phi_\ell\}_{\ell=1}^d$. Let $M \in \mathbb{R}^{d\times d}$ denote the finite element mass matrix, $M_{\ell,\ell'} = \int_D \phi_\ell \phi_{\ell'} \,\mathrm{d}\xi$, and let $K_{a_\sigma} \in \mathbb{R}^{d\times d}$ denote the stiffness matrix associated with the weighted diffusion form, $(K_{a_\sigma})_{\ell,\ell'} = \int_D a_\sigma \nabla\phi_\ell \cdot \nabla\phi_{\ell'} \,\mathrm{d}\xi$. The homogeneous Neumann condition enters (6.6) naturally, without an explicit boundary term. Writing $Y_\sigma(t) \in \mathbb{R}^d$ for the vector of nodal values of the finite element approximation of $y_\sigma(t,\cdot)$, the semi-discrete system reads $M\dot{Y}_\sigma = -\big(K_{a_\sigma} + cM\big)Y_\sigma + \mathcal{B}u$, i.e.,

$$\dot{Y}_\sigma = A_\sigma Y_\sigma + Bu, \quad A_\sigma := -M^{-1}\big(K_{a_\sigma} + cM\big),$$

with $B \in \mathbb{R}^{d\times N_\mathrm{a}}$ the discretized actuator operator, whose i-th column approximates a smoothed indicator of O_i. In our experiments, $N_\mathrm{a} = 3$ actuators of width 0.1, uniformly distributed over D are utilized.

The output operator $C \in \mathbb{R}^{d\times d}$ is taken as the M-orthogonal projection onto the span of the leading N_eig eigenfunctions $\{e_j\}_{j=1}^{N_\mathrm{eig}}$ of the Neumann Laplacian on D (here, $e_j(\xi) = \cos((j-1)\pi\xi)$), i.e.,

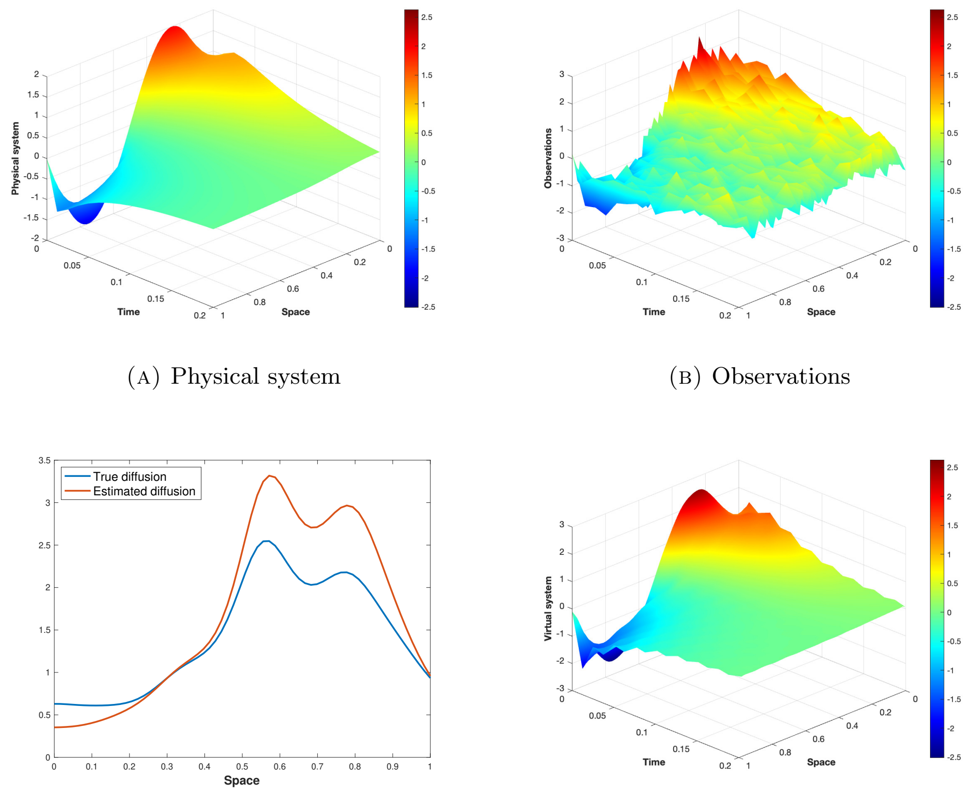
$$C = ME\big(E^\top ME\big)^{-1}E^\top, \qquad E = [e_1, \dots, e_{N_\mathrm{eig}}] \in \mathbb{R}^{d\times N_\mathrm{eig}},$$

so that Cy retains the projection of y onto the N_eig dominant spectral modes and discards the remainder. We choose $d = 64$ and $N_\mathrm{eig} = 16$. Furthermore, we take $p = 20$, decay exponent $\vartheta = 2$, constant reaction $c = -1$ (so that the uncontrolled system is unstable), true parameter

$$\sigma_\mathrm{true} = (-\tfrac{\pi}{7},\ \sqrt{2},\ -3,\ -1,\ 6,\ \xi_1, \dots, \xi_{15}), \qquad \xi_i \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}(-10, 10),$$

initial condition $y_0(\xi) = 2\sin(2\pi\xi)$, and prior $\sigma \sim \mathcal{N}(\sigma_\mathrm{true}, 4\cdot \mathrm{Id}_p)$. The distribution of the initial condition is $\mathcal{N}(1.25 \cdot y_0, \mathrm{Id}_\mathrm{d})$. The observation noise level is $\eta \sim \mathcal{N}(0,\ 0.0075 \cdot \|y_0\|)$. The matrices $\mathcal{A}_\sigma, \mathcal{B}_\sigma$ are computed using the matrix exponential as $\mathcal{A}_\sigma = e^{A_\sigma \Delta t}$ and $\mathcal{B}_\sigma = \int_0^{\Delta t} e^{A_\sigma(\Delta t - \tau)}\,\mathrm{d}\tau\, B$, whereas the matrices $\mathcal{A}_{\widehat{\sigma}}, \mathcal{B}_{\widehat{\sigma}}$ are computed using a Crank–Nicolson discretization as in Remark 2.1. We use $n_\mathrm{MCMC} = 10$ steps with proposal scale $\epsilon = 0.5$. The parameter estimate is updated at times $0.05, 0.1, 0.15, 0.2$ with observations at times $1/100, 2/100, 3/100, \dots, 0.2$.

We observe that the physical system is stabilized (see Figure 8a), while the virtual system tracks the physical one (compare Figure 8a with Figure 8d) and the diffusion coefficient is estimated simultaneously (see Figure 8c).



(A) Physical system

(B) Observations

(C) Diffusion coefficient at $t = 0.2$

(D) Virtual system

Figure 8. Results for the semi-discretized diffusion-reaction equation.

7. Outlook

We provided a digital twin framework involving a virtual and a physical twin and a bidirectional coupling between them. As a prototypical design objective we chose the stabilization of the physical twin. Mathematically, this involves simultaneous Bayesian inference, (optimal) feedback control, and state estimation. A significant difficulty arises due to the necessity of parameter updating, which consistently affects and changes the optimal feedback structure of the physical twin and the system operator of the virtual twin.

We identify several open challenges for possible further research. A convergence analysis of the parameter update procedure detailed in Section 4 is a challenging future goal. Similarly, for the estimation procedure, a finite set of data is utilized for each fixed control strategy. An asymptotic analysis with respect to the cardinality of the data set could be of interest, in the presence of updated controls. An efficient and constructive strategy for sensor placement, in relation to dynamical systems properties, for example, as well as the parameter update frequency, deserves deep investigation.

Acknowledgements

Part of this research was stimulated by the conducive atmosphere at the Institute for Mathematical and Statistical Innovation (IMSI), during workshops within the special semester on Digital Twins in the fall of 2025.
This work was initiated when S.R. was with RICAM, Austrian Academy of Sciences, part of his work is funded by the FCT – Portuguese Foundation for Science and Technology, I.P., under the scope of the projects UID/00297/2025 (doi: 10.54499/UID/00297/2025) and UID/PRR/00297/2025 (doi: 10.54499/UID/PRR/00297/2025) (Center for Mathematics and Applications – NOVA Math).

References

[1] AIAA Digital Engineering Integration Committee. Digital twin: Definition & value. Position paper, American Institute of Aeronautics and Astronautics (AIAA) and Aerospace Industries Association (AIA), 2020. URL: `https://www.aia-aerospace.org/publications/digital-twin-definition-value-an-aiaa-and-aia-position-paper/`.

[2] M. Akar, A. Paul, M. G. Safonov, and U. Mitra. Conditions on the stability of a class of second-order switched systems. *IEEE Trans. Autom. Control*, 51(2), 2006. `doi:10.1109/TAC.2005.863512`.

[3] A. M. Annaswamy and A. L. Fradkov. A historical perspective of adaptive control and learning. *Annu. Rev. Control*, 52:18–41, 2021. `doi:10.1016/j.arcontrol.2021.10.014`.

[4] H. Antil. Mathematical opportunities in digital twins (MATH-DT). Preprint arXiv:2402.10326v2 [math.OC], 2024. `doi:10.48550/arXiv.2402.10326`.

[5] K. J. Åström and B. Wittenmark. *Adaptive control.* Dover Publications, 2nd edition, 2013. Reprint of the Addison-Wesley 1995 edition. URL: `https://store.doverpublications.com/products/9780486319148`.

[6] P. Billingsley. *Probability and Measure.* Wiley Series in Probability and Statistics. Wiley, New York, 1995. URL: `https://books.google.pt/books?id=z39jQgAACAAJ`.

[7] A. Dey, A. Dhar, and S. Bhasin. Adaptive output feedback model predictive control. *IEEE Control Syst. Lett.*, 7:1129–1134, 2023. `doi:10.1109/LCSYS.2022.3231837`.

[8] P. A. Guth, K. Kunisch, and S. S. Rodrigues. Stabilization of uncertain linear dynamics: an offline-online strategy. *Math. Control and Related Fields*, 15(2):640–669, 2025. `doi:10.3934/mcrf.2024032`.

[9] P. Lancaster and L. Rodman. *Algebraic Riccati equations.* Oxford science publications. Clarendon Press, Oxford, 1995. URL: `https://global.oup.com/academic/product/algebraic-riccati-equations-9780198537953?cc=pt&lang=en&#`.

[10] D. Levanony and P. E. Caines. *Stochastic Lagrangian adaptation.* SpringerBriefs in Mathematics. Springer, Cham, 2024. `doi:10.1007/978-3-031-73758-9`.

[11] D. Liberzon and A. S. Morse. Basic problems in stability and design of switched systems. *IEEE Control Syst. Mag.*, 19(5):59–70, 1999. doi:10.1109/37.793443.

[12] A. Mesbah. Stochastic model predictive control with active uncertainty learning: A survey on dual control. *Annu. Rev. Control*, 45:107–117, 2018. doi:10.1016/j.arcontrol.2017.11.001.

[13] E. L. Morris. *Multivariate Statistics: A Vector Space Approach*, volume 53 of *IMS Lecture Notes–Monograph Series*. Institute of Mathematical Statistics, 2007. doi:10.1214/lnms/1196285102.

[14] National Academies of Sciences, Engineering, and Medicine. Foundational research gaps and future directions for digital twins. Technical report, The National Academies Press, Washington, DC, 2024. doi:10.17226/26894.

[15] D. Simon. *Optimal State Estimation: Kalman, H_∞, and Nonlinear Approaches*. Wiley-Interscience, Hoboken, NJ, 2006. URL: https://books.google.pt/books?id=UiMVoP_7TZkC.

[16] T. Söderström. *Discrete-Time Stochastic Systems: Estimation and Control*. Springer-Verlag, Berlin, Heidelberg, 2nd edition, 2002. doi:10.1007/978-1-4471-0101-7.

[17] J. Stoer and R. Bulirsch. *Introduction to Numerical Analysis*. Texts in Applied Mathematics, 12. Springer, New York, second edition, 1993. doi:10.1007/978-0-387-21738-3.

[18] J. R. Stroud, M. Katzfuss, and C. K. Wikle. A Bayesian adaptive ensemble Kalman filter for sequential state and parameter estimation. *Mon. Weather Rev.*, 146(1):373–386, 2018. doi:10.1175/MWR-D-16-0427.1.

[19] J. Venkatasubramanian, J. Köhler, J. Berberich, and F. Allgöwer. Robust dual control based on gain scheduling. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 2270–2277, 2020. doi:10.1109/CDC42340.2020.9304336.

[20] J. Venkatasubramanian, J. Köhler, J. Berberich, and F. Allgöwer. Sequential learning and control: Targeted exploration for robust performance. *IEEE Trans. Autom. Control*, 70(1):307–322, 2025. doi:10.1109/TAC.2024.3430088.