DETERMINISTIC KALMAN FILTERS FOR UNCERTAIN DYNAMICAL SYSTEMS

KARL KUNISCH*,† AND JESPER SCHRÖDER†

ABSTRACT. The Kalman(-Bucy) filter is the natural choice for the state reconstruction of disturbed, linear dynamical systems based on flawed and incomplete measurements. Taking a deterministic viewpoint this work investigates possible extensions of the concept to systems with uncertain dynamics and noise covariances. In a theoretical analysis error bounds in terms of the variance of the uncertainties are derived. The article concludes with a numerical implementation of two example systems allowing for a comparison of the estimators.

KEYWORDS: state estimation, filtering, parameter dependent systems

AMS subject classification: 34F05, 34H05, 49N10, 93B53, 93C15

1. INTRODUCTION

In this work we aim at reconstructing the state of a linear perturbed dynamical system of the form

$$\dot{x}(t) = A_\sigma x(t) + Bv(t) \qquad t \in (0, T),$$
$$x(0) = x_0 + \eta,$$

where the disturbances in the dynamics and the initial state are given by $v \in L^2(0, T; \mathbb{R}^m)$ and $\eta \in \mathbb{R}^n$, respectively. Additionally the system matrix A_σ depends on a parameter $\sigma \in \Sigma$ representing uncertainties of the system dynamics. We assume access to a measurement of the system given as

$$y(t) = Cx(t) + \mu(t) \qquad t \in (0, T),$$

affected by the disturbance $\mu \in L^2(0, T; \mathbb{R}^r)$. The disturbances are assumed to be unknown and of deterministic nature. They are associated with parameter dependent positive definite matrices $\Gamma_\sigma \in \mathbb{R}^{n,n}$, $R_\sigma \in \mathbb{R}^{m,m}$, and $Q_\sigma \in \mathbb{R}^{r,r}$ representing their magnitudes.

Reconstructing a system's state based on incomplete and disturbed measurements has a rich history going back at least to Wiener's seminal work [22], where disturbances are modeled as random processes. The pioneering works of Kalman and Bucy introduced the highly celebrated Kalman(-Bucy) filter for time discrete [9] and for time continuous [10] linear systems with Gaussian noise. Due to its effectiveness it is widely applicable in practice and it inspired numerous variations to address particularly challenging systems such as the extended Kalman filter for nonlinear dynamical systems and the ensemble Kalman filter [6] for systems of very large dimension.

The assumption of normally distributed noise, however, implies state independent disturbances. Hence exact knowledge of the system dynamics is required and modeling uncertainties and errors introduced by, e.g., material dependent parameters, moving sensors, or model reduction, are not addressed explicitly. Similarly, the Kalman filter relies on exact knowledge on the distribution of the errors, i.e., their respective covariances which, in

practice, are known only approximately. In [17], see also the monograph [18], the authors consider time continuous, quadratically stable dynamics subject to bounded uncertainties. They design an estimator based on solutions of associated Riccati equations and analyze the asymptotic behavior with respect to time. It is shown that the expected difference of estimator and disturbed state converges to a neighborhood of zero. For a discussion of time discrete systems affected by noise with uncertain covariances we refer to [14] and the references therein, see also [20] for a more recent analysis.

The remainder of the article is organized as follows. Section 2 gives an overview on the Kalman filter in the stochastic as well as in the deterministic setting. We compare the two formulations, point out connections, and set up the framework for our investigation. In Section 3 the uncertainties are formulated mathematically and three designs for Kalman filters for uncertain systems are proposed and characterized. In part we shall focus on the estimator which arises as the weighted average of the individual Kalman filters and minimizes the average energy. Section 4 contains the technical proof that the distances of the proposed estimators to the Kalman filter associated with the hidden parameter can be estimated from above in terms of the level of variance in the uncertain matrices. The article concludes with a numerical investigation of the designs at the hand of two examples in Section 5.

Throughout the article we use the following notation. We say that a multivariate random variable X is distributed according to $\mathcal{N}(m, \Sigma)$, where $m \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d,d}$ is symmetric and positive definite, if X is normally distributed with mean m and covariance matrix Σ. In short we write $X \sim \mathcal{N}(m, \Sigma)$. For a symmetric, positive definite matrix $P \in \mathbb{R}^{d,d}$ the associated norm of $x \in \mathbb{R}^d$ is defined via $\|x\|_M^2 = x^\top M x$. The cardinality of a finite set $\mathcal{M}$ is denoted by $|\mathcal{M}|$. Finally, we abbreviate some common L^p spaces, Sobolev spaces, and spaces of continuous functions equipped with their default norms as follows. We set $L^2(t_1, t_2; \mathbb{R}^d) = \mathcal{L}_{t_1}^{t_2}$, $H^1(t_1, t_2; \mathbb{R}^d) = \mathcal{H}_{t_1}^{t_2}$, $C([t_1; t_2]; \mathbb{R}^d) = \mathcal{C}_{t_1}^{t_2}$, and $C([0; t]; \mathbb{R}^{d,d}) = \mathcal{C}_t^{d,d}$. For the spaces of vector valued function the dimension d is suppressed in the notation and becomes clear from context.

2. The Kalman filter

We commence with a brief presentation of two different formulations of the Kalman filter. Even though our proposed Kalman filter for uncertain systems relies only on the deterministic formulation we include a recap of the stochastic formulation enriching the interpretation of our approach. For ease of presentation we restrict ourselves to time-invariant systems. While the concepts and results presented in this section are well known in the literature we include them for the purpose of a self-contained work.

2.1. **The stochastic formulation.** We illustrate the stochastic filter based on the original work [10]. Even though this formulation of the model lacks some mathematical rigor, the formal representation suffices to illustrate the underlying concept. For a more thorough treatment based on Itô integrals we refer to [16, 24].

Consider the disturbed system with system state X_t and measured output Y_t modeled by

$$\begin{aligned}
\tfrac{\mathrm{d}}{\mathrm{d}t} X_t &= A X_t + B V_t \quad t \in (0, T), \\
Y_t &= C X_t + W_t, \quad t \in (0, T),
\end{aligned} \tag{2.1}$$

where $A \in \mathbb{R}^{n,n}$ describes the linear dynamics of the system, $C \in \mathbb{R}^{r,n}$ describes the incomplete measurements taken from the state and $B \in \mathbb{R}^{n,m}$ encodes how the noise enters

the dynamics. The white noise affecting the dynamics and the measurement is modeled by the multivariate Gaussian random variables $V_t \in \mathbb{R}^m$ and $W_t \in \mathbb{R}^r$, respectively. They have zero mean and covariances given by

$$\mathrm{cov}[V_t, V_s] = R\,\delta(t-s),$$
$$\mathrm{cov}[W_t, W_s] = Q\,\delta(t-s),$$

with given symmetric, positive definite matrices $R \in \mathbb{R}^{m,m}$ and $Q \in \mathbb{R}^{r,r}$. The Gaussian distribution of the state at time t is described by X_t with initial Gaussian distribution X_0 with mean $x_0 \in \mathbb{R}^n$ and covariance $\Gamma \in \mathbb{R}^{n,n}$, symmetric, positive definite. Finally, V_t, W_t, and X_0 are assumed to be mutually independent and Y_t encodes the distribution of the output of the system at time t.

The objective, as stated in [10], is to optimally estimate the realized state $x(t)$, of the system based on realized, measured outputs $y(s)$, $0 \leq s \leq t$. Kalman and Bucy find that the optimal estimate $\widehat{x}(t)$ is characterized via

$$\begin{aligned}
\dot{\widehat{x}}(t) &= A\widehat{x}(t) + \Pi(t)C^\top Q^{-1}(y(t) - C\widehat{x}(t)) \qquad t \in (0,T), \\
\widehat{x}(0) &= x_0,
\end{aligned} \tag{2.2}$$

where Π is the unique solution to the differential Riccati equation

$$\begin{aligned}
\dot{\Pi}(t) &= A\Pi(t) + \Pi(t)A^\top - \Pi(t)C^\top Q^{-1}C\Pi(t) + BRB^\top \qquad t \in (0,T), \\
\Pi(0) &= \Gamma.
\end{aligned} \tag{2.3}$$

These formulas have since been confirmed by the stochastic community using rigorous mathematical concepts, see e.g. [16,24]. Interpreting (2.1) as stochastic differential equations with noise terms defined by Brownian motions and solutions defined via the Itô integral one characterizes the conditional expectation and covariance. More precisely, given a measured output the state of the system at time t is distributed according to $\mathcal{N}(\widehat{x}(t), \Pi(t))$ where the mean and covariance are characterized via (2.2) and (2.3), respectively. In particular, if one uses the mean $\widehat{x}$ as a state estimator the error has zero mean and a covariance characterized via the Riccati equation (2.3).

We conclude the subsection by introducing some well-known concepts connected to multivariate normal distributions as presented,

Definition 2.1. *Consider a random variable distributed according to $\mathcal{N}(m, \Sigma)$ with mean $m \in \mathbb{R}^d$ and covariance $\Sigma \in \mathbb{R}^{d,d}$ and assume Σ to be regular. We define the following concepts associated with it.*

(i) The precision matrix [4, Sec. 5.4] P is defined as the inverse of the covariance matrix, i.e., $P = \Sigma^{-1}$.

(ii) The generalized variance [7, Sec. 5] is defined as the determinant of the covariance matrix. Further the generalized precision p is defined as the reciprocal of the generalized variance i.e., $p = \frac{1}{\det(\Sigma)} = \det(P)$.

(iii) For any fixed vector $\xi \in \mathbb{R}^d$ its Mahalanobis distance [13] to the normal distribution is defined as its distance to the mean weighted by the precision, i.e.,

$$\|\xi - m\|_P = \sqrt{(\xi - m)^\top P(\xi - m)}.$$

The precision matrix and generalized precision extend the precision associated with univariate normal distributions to the multivariate case and the Mahalanobis distance introduces a distance from a point to a point to a given normal distribution. As described

above, the Kalman filter can be understood as a time dependent normal distribution. Hence these concepts enable us to quantify the precision of the Kalman filter as well as the distance of deterministic points to the Kalman filter.

2.2. **The deterministic formulation.** In this section we consider a deterministic model of a disturbed system brought forward by Mortensen in [15]. In his work he considers possibly nonlinear systems and derives a state estimator based on energy minimization. Unfortunately the obtained formulas for nonlinear systems rely on access to the value function arising from the optimal control problem minimizing the energy. While said value function is formally available as the solution of a Hamilton-Jacobi-Bellman equation its realization is afflicted by the curse of dimensionality. However, Mortensen notes that, when applied to linear dynamics, the formulas reduce to the Kalman filter derived in [10]. A detailed study of this concept for linear systems including a rigorous derivation of the filter equations is presented by Willems in [23].

The deterministic formulation of the dynamics and the output of the model reads

$$
\begin{aligned}
x(t) &= Ax(t) + Bv(t) & t &\in (0, T), \\
x(0) &= x_0 + \eta, \\
y(t) &= Cx(t) + \mu(t) & t &\in (0, T),
\end{aligned}
\tag{2.4}
$$

where the system matrices A, B, and C are the same as in the stochastic model (2.1). The disturbances in the dynamics, initial value, and output are represented by $v \in L^2(0, T; \mathbb{R}^m)$, $\eta \in \mathbb{R}^n$, and $\mu \in L^2(0, T; \mathbb{R}^r)$, respectively. They are assumed to be deterministic and unknown. For the sake of readability throughout the rest of this work for $0 \leq t_1 < t_2 \leq T$ and $d \in \mathbb{R}^d$ we denote $\mathcal{L}_{t_1}^{t_2} = L^2(t_1, t_2; \mathbb{R}^d)$. Note that the dimension of the image space is surpressed in the notation and is explicitly mentioned if not clear from context.

Remark 2.2. *On a formal level system* (2.4) *can be interpreted as a version of the stochastic model* (2.1) *in which the random terms have realized but are unknown. As Willems notes, due to a gap in regularity this connection can unfortunately not be established in a rigorous fashion. The assumed L^2-regularity of the errors v and μ is essential for the following analytical treatment. In contrast, realizations of the white noise in the stochastic model (interpreted as the derivative of a Brownian motion) must be expected to be of significantly lower regularity, cf. [11].*

Willems motivates the construction of the state estimator $\widehat{x}(t)$ at time $t \in (0, T]$ as follows. Given measured output data up until time t, i.e., $y(s)$, $0 \leq s \leq t$, one reconstructs the disturbances η, and $v(s)$, and $\mu(s)$, $0 \leq s \leq t$ as the ones that have minimal (weighted) energy among all possible disturbances that fit the given output in the sense of (2.4) restricted to $(0, t)$. Mathematically speaking, for a fixed $t \in (0, T]$ the state estimation at time t is defined via

$$
\min_{\eta \in \mathbb{R}^n, v \in \mathcal{L}_0^t, \mu \in \mathcal{L}_0^t} \frac{1}{2} \|\eta\|_{\Gamma^{-1}}^2 + \int_0^t \|v(s)\|_{R^{-1}}^2 + \|\mu(s)\|_{Q^{-1}}^2 \, \mathrm{d}s, \quad \text{subject to}
$$

$$
\begin{aligned}
x(s) &= Ax(s) + Bv(s) & s &\in (0, t), \\
x(0) &= x_0 + \eta, \\
y(s) &= Cx(s) + \mu(s) & s &\in (0, t),
\end{aligned}
\tag{2.5}
$$

where $\Gamma \in \mathbb{R}^{n,n}$, $R \in \mathbb{R}^{m,m}$, and $Q \in \mathbb{R}^{r,r}$ are symmetric positive definite matrices and in the following are referred to as weighting matrices. Denoting the resulting state trajectory

associated with the minimizing disturbances by x_t^* the estimator at time t is defined as $\widehat{x}(t) = x_t^*(t)$. We stress that for $0 \leq t_1 < t_2 \leq T$ in general it holds $\widehat{x}(t_1) = x_{t_1}^*(t_1) \neq x_{t_2}^*(t_1)$. Therefore, in order to obtain the state estimate $\widehat{x}(t)$ for all $t \in (0, T]$ the minimization must be carried out individually for each time point. Before discussing the derivation of the Kalman filter equations we comment on the weighting matrices Γ, R, and Q.

Remark 2.3. *While the weighting matrices can, in principal, be chosen freely they correspond to the covariance matrices Γ, R, and Q appearing in the stochastic formulation. The intuitive connection is very straight forward if the covariance matrices are diagonal. If, e.g., Q is diagonal the r components of the stochastic measurement error W_t in (2.1) are uncorrelated and their respective variances are found on the diagonal of Q. Weighting the (distance of the) corresponding deterministic measurement error μ (to its zero mean) in (2.5) with Q^{-1} has the following effect. Any component with a small variance (relative to other components of μ and the components of η and v) is weighted with the reciprocal, i.e., is strongly penalized in the minimization and the minimizer is promoted to display a small value in said component. This exactly fits the small variance of the stochastic formulation. It is well-known that this concept transfers to the case of non-diagonal covariance matrices Γ, R, and Q allowing for the components of each error to be correlated and it is applied, e.g., in generalized least squares [1]. To stress this principle, for the remainder of this work we refer to Γ, R, and Q as covariance rather than weighting matrices.*

In [23] Willems treats the minimization (2.5) directly. While his work considers the slightly less general setting with fixed $Q = \mathrm{Id}$ and $R = \mathrm{Id}$ a straight forward extension of his arguments leads to analogous results for the setting presented here. One finds that the state estimator $\widehat{x}$ is characterized via the Kalman filter equations (2.2)-(2.3), where now x_0 denotes the undisturbed initial condition rather than the mean of the initial distribution.

For our purposes we follow Mortensen's approach and reformulate (2.5) into a problem of optimal control. More precisely, we insert the identities for $x(0)$ and y into the cost and additionally fix a final state $\xi \in \mathbb{R}^n$ for the state trajectory to obtain

$$\min_{x \in \mathcal{H}_0^t, v \in \mathcal{L}_0^t} J(x, v; t, \xi) = \frac{1}{2}\|x(0) - x_0\|_{\Gamma^{-1}}^2 + \frac{1}{2}\int_0^t \|v(s)\|_{R^{-1}}^2 + \|y - Cx(s)\|_{Q^{-1}}^2 \, \mathrm{d}s, \quad (2.6)$$

$$\text{subject to} \qquad \dot{x}(s) = Ax(s) + Bv(s) \qquad s \in (0, t), \tag{2.7}$$

$$x(t) = \xi. \tag{2.8}$$

Remark 2.4. *We point out that, even though the construction of the estimator is based on the optimal control problem (2.6)-(2.8) and v takes on the role of a control, we do not assume any possibility to influence the system. The control formulation at hand is merely a tool to reconstruct energy minimal disturbances that most likely caused the measured output.*

The value function associated with the control problem is defined as

$$\mathcal{V}(t, \xi) = \inf_{x \in \mathcal{H}_0^t, v \in \mathcal{L}_0^t} J(x, v; t, \xi) \quad \text{subject to } (2.7) - (2.8) \quad t \in (0, T],$$

$$\mathcal{V}(0, \xi) = \frac{1}{2}\|\xi - x_0\|_{\Gamma^{-1}}^2, \tag{2.9}$$

and represents the minimal amount of energy required to ensure that at time t the system is in the state ξ. It turns out that for all $t \in [0, T]$ the state estimator $\widehat{x}$ defined above can equivalently be defined as

$$\widehat{x}(t) = \arg\min_{\xi \in \mathbb{R}^n} \mathcal{V}(t, \xi).$$

We now show that the value function can be given explicitly in terms of the solution of the Kalman filter equations which are

$$
\begin{aligned}
\dot{\widehat{x}}(t) &= A\widehat{x}(t) + \Pi(t)C^\top Q^{-1}\left(y(t) - C\widehat{x}(t)\right) \qquad t \in (0, T), \\
\widehat{x}(0) &= x_0,
\end{aligned}
\tag{2.10}
$$

and

$$
\begin{aligned}
\dot{\Pi}(t) &= A\Pi(t) + \Pi(t)A^\top - \Pi(t)C^\top Q^{-1}C\Pi(t) + BRB^\top \qquad t \in (0, T), \\
\Pi(0) &= \Gamma.
\end{aligned}
\tag{2.11}
$$

Even though the following derivations and arguments are based on the deterministic perspective, we still refer to the solution Π of (2.11) and its inverse $\Pi^{-1} = P$ as the (error) covariance and the precision matrix associated with the Kalman filter allowing for an interpretation of our results. Multiplying (2.11) by $P(t)$ from the left and from the right we find that the precision matrix is characterized via the differential Riccati equation

$$
\begin{aligned}
\dot{P}(t) &= -A^\top P(t) - P(t)A - P(t)BRB^\top P(t) + C^\top Q^{-1}C, \qquad t \in (0, T), \\
P(0) &= \Gamma^{-1}.
\end{aligned}
\tag{2.12}
$$

Lemma 2.5. *Let $\widehat{x}$ and P be given as the solutions of (2.10) and (2.12), respectively. Then for all $t \in [0, T]$ and $\xi \in \mathbb{R}^n$ it holds*

$$
\mathcal{V}(t, \xi) = \frac{1}{2}(\xi - \widehat{x}(t))^\top P(t)(\xi - \widehat{x}(t)) + \frac{1}{2}\int_0^t \|y(s) - C\widehat{x}(s)\|_{Q^{-1}}^2 \, \mathrm{d}s.
\tag{2.13}
$$

In particular the value function is quadratic in ξ implying that its Hessian is independent of ξ and its Hessian $\nabla_{\xi\xi}^2 \mathcal{V}(t, \xi)$ is given by the solution P of (2.12).

Proof. The result is shown in [2, Sec. 2.3] for continuous outputs y and weighting matrices $\Gamma = \mathrm{Id}$, $R = \mathrm{Id}$, and $Q = \mathrm{Id}$. A straight forward extension yields the result for the weighted setting. A standard sensitivity analysis, e.g., via an application of the inverse mapping theorem to the first order optimality system ensures that $\mathcal{V}(t, \xi)$ depends continuously on $y \in \mathcal{L}_0^t$. A density argument completes the proof. $\square$

This representation of the value functions allows for an interesting interpretation of the energy.

Corollary 2.6. *Let $\widehat{x}$ and P be given as the solution of (2.10) and (2.12), respectively. Then for all $t \in [0, T]$ and $\xi \in \mathbb{R}^n$ it holds*

$$
\mathcal{V}(t, \xi) - \mathcal{V}(t, \widehat{x}(t)) = \frac{1}{2}\|\xi - \widehat{x}(t)\|_{P(t)}^2.
$$

In view of the discussion in Subsection 2.1 we find that the difference in energies of a state ξ at time t and the Kalman trajectory $\widehat{x}(t)$ coincides with their squared Mahalanobis distance.

3. Kalman filtering for uncertain systems

The Kalman filter described in the previous section assumes exact knowledge of both the system dynamics represented by the matrix A and the nature of the occurring noise represented by the weighting/covariance matrices Γ, R, and Q. In practical applications, however, this information may not be available in exact form as the dynamics of the system

may depend on various uncertain parameters [20]. Further, as described, e.g. in [14], the covariance matrices associated with the distribution of the noise and the initial state may be subject to uncertainties.

3.1. **Modeling uncertain parameters.** We model the uncertainties of the system matrix A as follows. Let $s_A \in \mathbb{N}$ and consider a matrix valued function

$$a \colon \mathbb{R}^{s_A} \to \mathbb{R}^{n,n}.$$

Further let $N_A \in \mathbb{N}$ and $\Sigma_A \subset \mathbb{R}^{s_A}$ be of cardinality $|\Sigma_A| = N_A$. Consider the discrete probability space $(\Sigma_A, \mathcal{P}(\Sigma_A), P_A)$, where $\mathcal{P}_{\mathcal{A}}(\Sigma_A)$ denotes the power set of Σ_A and for $S \subset \Sigma_A$ we define the probability function $P_A(S) = \frac{1}{N_A}|S|$. Then the linear dynamics of the system are characterized by the matrix valued random variable

$$\sigma \mapsto a_\sigma = a(\sigma), \quad \Sigma_A \to \mathbb{R}^{n,n},$$

and a_σ denotes the system matrix that realizes with the parameter $\sigma \in \Sigma_A$. We analogously define the uncertain weighting matrices γ_σ, r_σ, and q_σ as random variables on Σ_Γ, Σ_R, and Σ_Q with respective cardinalities and dimensions N_Γ, s_Γ, N_R, s_R, and N_Q, s_Q. The associated probabilities are defined and denoted analogously to A. We denote $N = N_A\, N_\Gamma\, N_R\, N_Q$ and $s = s_A\, s_\Gamma\, s_R\, s_Q$. Then

$$\Sigma = \Sigma_A \times \Sigma_\Gamma \times \Sigma_R \times \Sigma_Q \subset \mathbb{R}^s$$

has cardinality N and we define the product space $(\Sigma, \mathcal{F}, P)$, via

$$\mathcal{F} = \mathcal{P}(\Sigma_A) \otimes \mathcal{P}(\Sigma_\Gamma) \otimes \mathcal{P}(\Sigma_R) \otimes \mathcal{P}(\Sigma_Q) = \mathcal{P}(\Sigma),$$

$$P(S_A \otimes S_\Gamma \otimes S_R \otimes S_Q) = P_A(S_A)\, P_\Gamma(S_\Gamma)\, P_R(S_R)\, P_Q(S_Q) = \frac{1}{|S_A \otimes S_\Gamma \otimes S_R \otimes S_Q|}.$$

Finally the uncertain system and weighting matrices we work with are defined as random variables on Σ. Denote $\sigma = (\sigma_A, \sigma_\Gamma, \sigma_R, \sigma_Q) \in \Sigma$ and define

$$\Sigma \to \mathbb{R}^{n,n}, \quad \sigma \mapsto A_\sigma = a_{\sigma_A},$$
$$\Sigma \to \mathbb{R}^{n,n}, \quad \sigma \mapsto \Gamma_\sigma = \gamma_{\sigma_\Gamma},$$
$$\Sigma \to \mathbb{R}^{m,m}, \quad \sigma \mapsto R_\sigma = r_{\sigma_R},$$
$$\Sigma \to \mathbb{R}^{r,r}, \quad \sigma \mapsto Q_\sigma = q_{\sigma_Q}.$$

Note that they are, by construction, mutually independent. For convenience we denote the N elements of Σ by σ_k, $k = 1, \ldots, N$ and dependencies on σ_k may be indicated by indexing k, e.g., $A_{\sigma_k} = A_k$.

In the following we present three different estimators for the reconstruction of the state of the uncertain system affected by noise with uncertain covariances.

3.2. **The Kalman filter for expected system and covariance matrices.** We begin with the most straight forward and computationally affordable approach. Namely, we form the expectation of the four uncertain matrices and realize the associated Kalman filter. More precisely, define

$$A_{\mathbb{E}} = \mathbb{E}[A_\sigma] = \frac{1}{N}\sum_{k=1}^{N} A_k, \quad \Gamma_{\mathbb{E}} = \mathbb{E}[\Gamma_\sigma] = \frac{1}{N}\sum_{k=1}^{N} \Gamma_k,$$

$$R_{\mathbb{E}} = \mathbb{E}[R_\sigma] = \frac{1}{N}\sum_{k=1}^{N} R_k, \quad Q_{\mathbb{E}} = \mathbb{E}[Q_\sigma] = \frac{1}{N}\sum_{k=1}^{N} Q_k.$$

The associated state estimator is constructed as the solution to

$$\begin{aligned}
\dot{\widehat{x}}_{\mathbb{E}}(t) &= A_{\mathbb{E}}\widehat{x}_{\mathbb{E}}(t) + \Pi_{\mathbb{E}}(t)C^{\top}Q_{\mathbb{E}}^{-1}\left(y(t) - C\widehat{x}_{\mathbb{E}}(t)\right) \qquad t \in (0, T), \\
\widehat{x}(0) &= x_0,
\end{aligned} \tag{3.1}$$

where $\Pi_{\mathbb{E}}$ is given via the Riccati equation

$$\begin{aligned}
\dot{\Pi}_{\mathbb{E}}(t) &= A_{\mathbb{E}}\Pi_{\mathbb{E}}(t) + \Pi_{\mathbb{E}}(t)A_{\mathbb{E}}^{\top} - \Pi_{\mathbb{E}}(t)C^{\top}Q_{\mathbb{E}}^{-1}C\Pi_{\mathbb{E}}(t) + BR_{\mathbb{E}}B^{\top} \qquad t \in (0, T), \\
\Pi(0) &= \Gamma_{\mathbb{E}}.
\end{aligned} \tag{3.2}$$

3.3. The expected Kalman filter.
For the second approach we realize the Kalman filter associated with each possible parameter $\sigma \in \Sigma$ allowing for the construction of an estimator $\widehat{x}_{\sigma}$ depending on the uncertainty. Formulating the uncertain trajectory as a random variable its associated expectation yields the state estimator.

To that end for every element $\sigma_k \in \Sigma$ solve for the associated Kalman filter $\widehat{x}_k = \widehat{x}_{\sigma_k}$ via

$$\begin{aligned}
\dot{\widehat{x}}_k(t) &= A_k\widehat{x}_k(t) + \Pi_k(t)C^{\top}Q_k^{-1}\left(y(t) - C\widehat{x}_k(t)\right) \qquad t \in (0, T), \\
\widehat{x}_k(0) &= x_0,
\end{aligned} \tag{3.3}$$

where Π_k is given via the Riccati equation

$$\begin{aligned}
\dot{\Pi}_k(t) &= A_k\Pi_k(t) + \Pi_k(t)A_k^{\top} - \Pi_k(t)C^{\top}Q_k^{-1}C\Pi_k(t) + BR_kB^{\top} \qquad t \in (0, T), \\
\Pi_k(0) &= \Gamma_k.
\end{aligned} \tag{3.4}$$

For later reference we denote the associated precision matrix by $P_k(t) := \Pi_k^{-1}(t)$. It is given as the unique solution of

$$\begin{aligned}
\dot{P}_k(t) &= -A_k^{\top}P_k(t) - P_k(t)A_k - P_k(t)BR_kB^{\top}P_k(t) + C^{\top}Q_k^{-1}C, \qquad t \in (0, T), \\
P_k(0) &= \Gamma_k^{-1}.
\end{aligned} \tag{3.5}$$

Now $\sigma_k \mapsto \widehat{x}_{\sigma_k}$, $\Sigma \to \{\widehat{x}_{\sigma_k} : \sigma_k \in \Sigma\} \subset \mathcal{H}_0^T$ is considered a random variable where the image space is equipped with its power set as a σ-algebra. The state estimator then is defined as the expectation, i.e.,

$$\widehat{x}_{\varnothing} := \mathbb{E}[\widehat{x}_{\sigma}] = \frac{1}{N}\sum_{k=1}^{N}\widehat{x}_k. \tag{3.6}$$

We point out that, by construction, for any $t \in [0, T]$ the estimator $\widehat{x}_{\varnothing}(t)$ minimizes the expected squared Euclidean distance, i.e., it is the unique solution of

$$\min_{\xi \in \mathbb{R}^n}\mathbb{E}\left[\|\xi - \widehat{x}_{\sigma}(t)\|^2\right] = \min_{\xi \in \mathbb{R}^n}\frac{1}{N}\sum_{k=1}^{N}\|\xi - \widehat{x}_k(t)\|^2.$$

Finally, note that the realization of this option is computationally more demanding than the one in the previous subsection. Indeed, it requires solving N Riccati equations and ODEs.

3.4. Minimizing the expected energy.
The third and final estimator is defined as the minimizer of the expected energy. Throughout the rest of this work we use the notation established in the previous subsection, i.e., $\widehat{x}_k$, Π_k, and P_k are the Kalman filter trajectory, the error covariance, and the precision matrix associated with the matrices A_k, Γ_k, R_k,

and Q_k. Additionally, throughout the remainder of this work we denote the associated value function according to (2.9) by $\mathcal{V}_k = \mathcal{V}_{\sigma_k}$. For $t \in [0, T]$ we obtain

$$\widehat{x}_k(t) = \underset{\xi \in \mathbb{R}^n}{\arg \min}\, \mathcal{V}_k(t, \xi),$$

and

$$\mathcal{V}_k(t, \xi) = \frac{1}{2}(\xi - \widehat{x}_k(t))^\top P_k(t)(\xi - \widehat{x}_k(t)) + \frac{1}{2} \int_0^t \|y(s) - C\widehat{x}_k(s)\|_{Q_k^{-1}}^2 \, \mathrm{d}s, \qquad (3.7)$$

with $\widehat{x}_k$ and P_k given via (3.3) and (3.5), respectively. Note that analogously to $\widehat{x}_k = \widehat{x}_{\sigma_k}$ the value function $\mathcal{V}_k = \mathcal{V}_{\sigma_k}$ can be considered as a random variable. The reader is further reminded that intuitively speaking $\mathcal{V}_k(t, \xi)$ represents the minimum amount of energy required to ensure that the system associated with σ_k at time t is in the state ξ.

For the uncertain system we now define the expected energy as follows.

Definition 3.1. *The expected energy $E \colon [0, T] \times \mathbb{R}^n \to \mathbb{R}$ is defined via*

$$
\begin{aligned}
E(t, \xi) &= \mathbb{E}\left[\mathcal{V}_\sigma(t, \xi)\right] = \frac{1}{N} \sum_{k=1}^{N} \mathcal{V}_k(t, \xi) \\
&= \frac{1}{2N} \sum_{k=1}^{N} \left((\xi - \widehat{x}_k(t))^\top P_k(t)(\xi - \widehat{x}_k(t)) + \int_0^t \|y(s) - C\widehat{x}_k(s)\|_{Q_k^{-1}}^2 \, \mathrm{d}s\right).
\end{aligned}
\qquad (3.8)
$$

Intuitively, $E(t, \xi)$ represents the amount of energy that is required in expectation to ensure that at time t the uncertain system is in the state ξ. This motivates the formal definition of the third state estimator discussed in this work.

Definition 3.2. *For $t \in [0, T]$ we define the state estimator as the state minimizing the expected energy via*

$$\widehat{x}_{\mathrm{E}}(t) = \underset{\xi \in \mathbb{R}^n}{\arg \min}\, E(t, \xi).$$

In the following we show that for each $t \in [0, T]$ the function $E(t, \cdot)$ admits a unique minimizer ensuring that $\widehat{x}_{\mathrm{E}}$ is in fact well-defined. We further derive a characterization in terms of the individual Kalman trajectories $\widehat{x}_k = \widehat{x}_{\sigma_k}$ allowing for a numerical realization with cost comparable to the realization of $\widehat{x}_\varnothing$.

The representation via the individual value functions shows that E is quadratic in ξ allowing for straight forward differentiation of E with respect to ξ.

Lemma 3.3. *For any fixed $t \in [0, T]$ the mapping $\xi \mapsto E(t, \xi)$ is of class C^∞. Its gradient and Hessian are given as*

$$\nabla E(t, \xi) = \frac{1}{N} \sum_{k=1}^{N} P_k(t)(\xi - \widehat{x}_k(t)), \qquad \text{and} \qquad \nabla^2 E(t, \xi) = \frac{1}{N} \sum_{k=1}^{N} P_k(t),$$

respectively.

Proof. According to Theorem 3.1 the average energy is a sum of functions quadratic in ξ and differentiability follows. The formulas are obtained by differentiation with respect to ξ. $\qquad \square$

Existence, uniqueness, and the characterization of a minimizer follow directly.

Proposition 3.4. *The state estimator defined in Theorem 3.2 is well-defined and given in terms of the Kalman filter trajectories and covariance matrices associated with the individual Kalman filters. For $t \in [0, T]$ it holds*

$$\widehat{x}_{\mathrm{E}}(t) = \mathcal{P}^{-1}(t) \sum_{k=1}^{N} P_k(t) \, \widehat{x}_k(t), \tag{3.9}$$

where $\mathcal{P}(t) = \sum_{k=1}^{N} P_k(t)$.

Proof. We fix $t \in [0, T]$ and begin by showing uniqueness. Due to the regularity of E established in the previous lemma any minimizer ξ^* must satisfy $\nabla E(t, \xi^*) = 0$. Rearranging the terms in the formula for the gradient yields

$$\sum_{k=1}^{N} P_k(t) \, \xi^* = \sum_{k=1}^{N} P_k(t) \, \widehat{x}_k(t).$$

For all $k = 1, \ldots, N$ we have that P_k is given as the solution to a differential Riccati equation with positive definite initial value, hence $P_k(t)$ is positive definite for all $t \in [0, T]$, cf. [5, Prop. 1.1]. The positive definiteness then carries over to the to the sum denoted by $\mathcal{P}$. Hence $\mathcal{P}$ is invertible and $\widehat{x}_{\mathrm{E}}(t)$ is the only state satisfying the necessary condition to be a minimizer. Together with the formula for the Hessian these considerations additionally ensure positive definiteness of the Hessian and it is shown that $\widehat{x}(t)$ is in fact the unique minimizer. $\qquad \square$

Remark 3.5. *According to this formula $\widehat{x}_{\mathrm{E}}$ can be interpreted as a generalized weighted mean of the individual Kalman filter trajectories. Indeed, it is given as the sum of each Kalman trajectory multiplied from the left by its associated precision matrix. The sum is then multiplied by the inverse of the sum of those precision matrices. We note that the expression for $\widehat{x}_{\mathrm{E}}$ as a weighted average of the family $\widehat{x}_k$ resembles the statistical method of inverse-variance weighting were a family of random variables is aggregated with the goal to minimize the variance of the weighted average.*

Interestingly $\widehat{x}_{\mathrm{E}}$ not only minimizes the expected energy, but also the expected squared Mahalanobis distance.

Corollary 3.6. *Let $t \in [0, T]$ be arbitrary. Then $\widehat{x}_{\mathrm{E}}(t)$ is the unique solution of the minimization problem*

$$\min_{\xi \in \mathbb{R}^n} \mathbb{E} \left[\|\xi - \widehat{x}_\sigma(t)\|_{P_\sigma(t)}^2 \right] = \min_{\xi \in \mathbb{R}^n} \frac{1}{N} \sum_{k=1}^{N} \|\xi - \widehat{x}_k(t)\|_{P_k(t)}^2.$$

Proof. The assertion is a direct consequence of Theorem 2.6 and Theorem 3.1. $\qquad \square$

Before we turn to the error analysis we mention a fourth estimator.

Remark 3.7. *Another possible design for the reconstruction of the state of the uncertain system is obtained by solving the Kalman filter equation for the mean (3.1) where for the gain one replaces the solution $\Pi_{\mathbb{E}}$ of the associated Riccati equation (3.2) by the expected gain, i.e., $\frac{1}{N} \sum_{k=1}^{N} \Pi_k$. It can be understood as an intermediate option combining the designs of Subsection 3.2 and Subsection 3.3.*

4. Error analysis

In this section we discuss the error that is introduced by the uncertainty in system and noise covariances. Our analysis is based on the following premise. If one had access to the true, hidden parameter $\bar\sigma \in \Sigma$ the obvious choice for the state estimator would be the associated Kalman filter $\widehat{x}_{\bar\sigma}$ with error covariance $\Pi_{\bar\sigma}$ given via

$$
\begin{aligned}
\dot{\widehat{x}}_{\bar\sigma}(t) &= A_{\bar\sigma}\widehat{x}_{\bar\sigma}(t) + \Pi_{\bar\sigma}(t)C^\top Q_{\bar\sigma}^{-1}\left(y(t) - C\widehat{x}_{\bar\sigma}(t)\right) \qquad t \in (0,T),\\
\widehat{x}_{\bar\sigma}(0) &= x_0,
\end{aligned}
\tag{4.1}
$$

and

$$
\begin{aligned}
\dot{\Pi}_{\bar\sigma}(t) &= A_{\bar\sigma}\Pi_{\bar\sigma}(t) + \Pi_{\bar\sigma}(t)A_{\bar\sigma}^\top - \Pi_{\bar\sigma}(t)C^\top Q_{\bar\sigma}^{-1}C\Pi_{\bar\sigma}(t) + BR_{\bar\sigma}B^\top \qquad t \in (0,T),\\
\Pi_{\bar\sigma}(0) &= \Gamma_{\bar\sigma}.
\end{aligned}
\tag{4.2}
$$

Hence the quality of the three estimators presented in Section 3 is measured in terms of their respective distances to $\widehat{x}_{\bar\sigma}$. More precisely, for a given estimator $\widehat{x}$ and time $t \in [0,T]$ we analyze the distance

$$
\|\widehat{x}(t) - \widehat{x}_{\bar\sigma}(t)\|_{\Pi_{\bar\sigma}^{-1}(t)} = \|\widehat{x}(t) - \widehat{x}_{\bar\sigma}(t)\|_{P_{\bar\sigma}(t)}.
\tag{4.3}
$$

The justification for the particular choice of weighted norm lies in the stochastic interpretation. In this case the estimator of the state at time t associated with the true parameter $\bar\sigma$ is given by a normal distribution with mean $\widehat{x}_{\bar\sigma}(t)$ and covariance $\Pi_{\bar\sigma}(t)$. The weighting then is motivated by the reasoning illustrated in Theorem 2.3. We further note that in the spirit of Theorem 2.6 it holds that

$$
\frac{1}{2}\|\widehat{x}(t) - \widehat{x}_{\bar\sigma}(t)\|_{P_{\bar\sigma}(t)}^2 = \mathcal{V}_{\bar\sigma}(t, \widehat{x}(t)) - \mathcal{V}_{\bar\sigma}(t, \widehat{x}_{\bar\sigma}(t)).
$$

Hence by estimating the weighted difference we simultaneously estimate the surplus of energy when compared to the minimal energy.

To facilitate a concise presentation of the results we introduce the following notation. For a given parameter σ the associated four-tuple consisting of system matrix and covariance matrices is denoted by $\mathcal{S}_\sigma = (A_\sigma, \Gamma_\sigma, R_\sigma, Q_\sigma) \in \mathbb{R}^{n,n} \times \mathbb{R}^{n,n} \times \mathbb{R}^{m,m} \times \mathbb{R}^{r,r}$. For $p \in \mathbb{N}$ we introduce the norm $\|\mathcal{S}_\sigma\|_p^p = \|A_\sigma\|_{\mathbb{R}^{n,n}}^p + \|\Gamma_\sigma\|_{\mathbb{R}^{n,n}}^p + \|R_\sigma\|_{\mathbb{R}^{m,m}}^p + \|Q_\sigma\|_{\mathbb{R}^{r,r}}^p$ and note that sums of systems are understood componentwise.

4.1. Technical preparations. Consider $\widehat{x}_*$ and Π_* given as the unique solutions of

$$
\begin{aligned}
\dot{\widehat{x}}_*(t) &= A_*\widehat{x}_*(t) + \Pi_*(t)C^\top Q_*^{-1}\left(y(t) - C\widehat{x}_*(t)\right) \qquad t \in (0,T),\\
\widehat{x}_*(0) &= x_0,
\end{aligned}
\tag{4.4}
$$

and

$$
\begin{aligned}
\dot{\Pi}_*(t) &= A_*\Pi_*(t) + \Pi_*(t)A_*^\top - \Pi_*(t)C^\top Q_*^{-1}C\Pi_*(t) + BR_*B^\top \qquad t \in (0,T),\\
\Pi_*(0) &= \Gamma_*.
\end{aligned}
\tag{4.5}
$$

These systems will be used with $(\widehat{x}_*, \Pi_*) = (\widehat{x}_\mathbb{E}, \Pi_\mathbb{E})$ and $(\widehat{x}_*, \Pi_*) = (\widehat{x}_k, \Pi_k)$ to obtain error estimates for the three state estimators presented in Section 3. In order to derive an estimate for $\|\Pi_* - \Pi_{\bar\sigma}\|_{\mathcal{C}_T^{n,n}}$ we first link the covariance matrices to the optimal control problems dual to the state estimation problem. For this purpose we define $\Sigma_*(t) = \Pi_*(T-t)$

for $t \in [0, T]$. By differentiating $t \mapsto \Pi_*(T-t)$ and utilizing the equation for $\dot{\Pi}_*$ we find that Σ_* is the unique solution to

$$\dot{\Sigma}_*(t) = -A_*\Sigma_*(t) - \Sigma_*A_*^\top + \Sigma_*(t)C^\top Q_*^{-1}C\Sigma_*(t) - BR_*B^\top \quad t \in (0, T),$$

$$\Sigma_*(T) = \Gamma_*.$$

This is the Riccati equation associated to the optimal control problem

$$\min_{x \in \mathcal{H}_t^T, u \in \mathcal{L}_t^T} \mathcal{J}_*(x, u; t, x_0) = \frac{1}{2}\int_t^T \|u(s)\|_{Q_*}^2 + \|B^\top x(s)\|_{R_*}^2 \, \mathrm{d}s + \frac{1}{2}\|x(T)\|_{\Gamma_*}^2, \quad (4.6)$$

$$\text{subject to} \quad \dot{x}(s) = A_*^\top x(s) + C^\top u(s) \quad s \in (t, T),$$

$$x(t) = x_0, \quad (4.7)$$

for $t \in [0, T]$. In the following lemma we recall that the associated value function

$$\nu_*(t, x_0) = \inf_{x, u \text{ s.t.}(4.7)} \mathcal{J}_*(x, u; t, x_0),$$

can be represented via the solution of the associated Riccati equation, cf. [21, Thm. 37].

Lemma 4.1. *Let Σ_* be as introduced above. Then for all $t \in [0, T)$ and $x \in \mathbb{R}^n$ it holds*

$$\nu_*(t, x) = \frac{1}{2}\left(x, \Sigma_*(t)x\right)$$

We also define $\Sigma_{\bar{\sigma}}$ and $J_{\bar{\sigma}}$ associated with the state equation $\dot{x}(s) = A_{\bar{\sigma}}^\top x(s) + C^\top u(s)$ and obtain an analogous representation of the related value function $\nu_{\bar{\sigma}}$ in terms of $\Sigma_{\bar{\sigma}} = \Pi_{\bar{\sigma}}(T - \cdot)$. Utilizing these characterizations we can represent the difference of covariance matrices via the difference of value functions.

Lemma 4.2. *Let $\Pi_{\bar{\sigma}}$ and Π_* be the solutions of (4.2) and (4.5), respectively. Then for all $t \in (0, T]$ it holds*

$$\frac{1}{2}\|\Pi_*(t) - \Pi_{\bar{\sigma}}(t)\| = \sup_{\|x\|=1} |\nu_*(T - t, x) - \nu_{\bar{\sigma}}(T - t, x)|,$$

where ν_* and $\nu_{\bar{\sigma}}$ are the value functions defined above.

Proof. Since both Σ_* and $\Sigma_{\bar{\sigma}}$ are symmetric, the same holds for their difference. For all $t \in [0, T]$ it follows

$$\frac{1}{2}\|\Pi_*(t) - \Pi_{\bar{\sigma}}(t)\| = \frac{1}{2}\|\Sigma_*(T - t) - \Sigma_{\bar{\sigma}}(T - t)\| = \frac{1}{2}\sup_{\|x\|=1} |(x, (\Sigma_*(T - t) - \Sigma_{\bar{\sigma}}(T - t))x)|$$

$$= \sup_{\|x\|=1} \left|\frac{1}{2}(x, \Sigma_*(T - t)x) - \frac{1}{2}(x, \Sigma_{\bar{\sigma}}(T - t)x)\right| = \sup_{\|x\|=1} |\nu_*(T - t, x) - \nu_{\bar{\sigma}}(T - t, x)|,$$

and the proof is complete. $\qquad\square$

To obtain the desired estimate for the difference of covariances we now estimate the difference of value functions. First we summarize some estimates for the solution of the optimal control problem associated with ν_*.

Lemma 4.3. *Let $(\bar{x}_*, \bar{u}_*)$ be the unique solution to (4.6)-(4.7) with associated adjoint state p^*. There exists $c_* > 0$ independent of t such that*

$$\|\bar{x}_*\|_{\mathcal{H}_t^T}^2 + \|\bar{x}_*\|_{\mathcal{C}_t^T}^2 + \|\bar{u}_*\|_{\mathcal{L}_t^T}^2 + \|p_*\|_{\mathcal{H}_t^T}^2 + \|p_*\|_{\mathcal{C}_t^T}^2 \leq c_*\|x_0\|^2.$$

Proof. Since this results can be proven using standard techniques from optimal control, we only sketch the proof. First note that

$$\|\bar{u}_*\|^2_{\mathcal{L}^T_t} \le 2\mathcal{J}_*(\bar{x}_*, \bar{u}_*; t, x_0) \le 2\mathcal{J}_*(x^0, 0; t, x_0), \tag{4.8}$$

where x^0 is the unique solution on (t, T) to

$$\dot{x}^0 = A_*^\top x^0, \qquad \dot{x}^0(t) = x_0.$$

Using standard arguments including an application of Gronwall's Lemma one can show the existence of $c > 0$ independent of t such that $\|x^0\|^2_{\mathcal{H}^T_t} \le c\|x_0\|^2$. Together with (4.8) and the continuous embedding $\mathcal{H}^T_t \hookrightarrow \mathcal{C}^T_t$ with constant c' we get

$$\|\bar{u}_*\|^2_{\mathcal{L}^T_t} \le \int_t^T \|Bx^0\|^2_{R^*} \, \mathrm{d}s + \|x^0(T)\|^2 \le c^2\|B\|^2 \|R^{*\frac{1}{2}}\|^2\|x_0\|^2 + c^2{c'}^2\|x_0\|^2.$$

The estimate transfers to $\bar{x}_*$ and p_* using the state and adjoint equations and arguments involving Gronwall's Lemma. $\qquad\square$

We now employ the strategy presented in [8, Lem. 3.2] to show the desired estimate for the value functions.

Lemma 4.4. *Let $t \in [0, T)$ and $x_0 \in \mathbb{R}^n$ with $\|x_0\| \le 1$. Then there exists a constant $c_\mathcal{J} > 0$ independent of t such that*

$$\mathcal{J}_{\bar{\sigma}}(\bar{x}_* - \bar{x}_{\bar{\sigma}}, \bar{u}_* - \bar{u}_{\bar{\sigma}}; t, x_0) \le c_\mathcal{J}\|\mathcal{S}_* - \mathcal{S}_{\bar{\sigma}}\|^2_2$$

where $(\bar{x}_*, \bar{u}_*)$ and $(\bar{x}_{\bar{\sigma}}, \bar{u}_{\bar{\sigma}})$ are the optimal pairs associated with $\mathcal{J}_*$ and $\mathcal{J}_{\bar{\sigma}}$, respectively.

Proof. We have $\nu_*(t, x_0) = \mathcal{J}_*(\bar{x}_*, \bar{u}_*; t, x_0)$, where $(\bar{x}_*, \bar{u}_*)$ is the unique minimizer of (4.6)-(4.7) which is characterized via the optimality system

$$\begin{aligned} \dot{\bar{x}}_*(s) &= A_*^\top \bar{x}_*(s) + C^\top \bar{u}_*(s), & \bar{x}_*(t) &= x_0, \\ \dot{p}_*(s) &= -A_* p_*(s) - BR_* B^\top \bar{x}_*(s), & p_*(T) &= \Gamma_* \bar{x}_*(T), \\ \bar{u}_*(s) &= -Q_*^{-1}Cp_*(s), \end{aligned}$$

where the time dependent equations hold for $s \in (t, T)$. Analogously we have $\nu_{\bar{\sigma}}(t, x_0) = \mathcal{J}_{\bar{\sigma}}(\bar{x}_{\bar{\sigma}}, \bar{u}_{\bar{\sigma}}; t, x_0)$, where $(\bar{x}_{\bar{\sigma}}, \bar{u}_{\bar{\sigma}})$ is the corresponding unique minimizer characterized via the optimality system

$$\begin{aligned} \dot{\bar{x}}_{\bar{\sigma}}(s) &= A_{\bar{\sigma}}^\top \bar{x}_{\bar{\sigma}}(s) + C^\top \bar{u}_{\bar{\sigma}}(s), & \bar{x}_{\bar{\sigma}}(t) &= x_0, \\ \dot{p}_{\bar{\sigma}}(s) &= -A_{\bar{\sigma}} p_{\bar{\sigma}}(s) - BR_{\bar{\sigma}} B^\top \bar{x}_{\bar{\sigma}}(s), & p_{\bar{\sigma}}(T) &= \Gamma_{\bar{\sigma}} \bar{x}_{\bar{\sigma}}(T), \\ \bar{u}_{\bar{\sigma}}(s) &= -Q_{\bar{\sigma}}^{-1}Cp_{\bar{\sigma}}(s). \end{aligned}$$

We define

$$\begin{aligned} x_\delta &:= \bar{x}_* - \bar{x}_{\bar{\sigma}}, & u_\delta &:= \bar{u}_* - \bar{u}_{\bar{\sigma}}, & p_\delta &:= \bar{p}_* - \bar{p}_{\bar{\sigma}}, \\ \Gamma_\delta &:= \Gamma_* - \Gamma_{\bar{\sigma}}, & R_\delta &:= R_* - R_{\bar{\sigma}}, & Q_{\delta^{-1}} &:= Q_*^{-1} - Q_{\bar{\sigma}}^{-1}, & A_\delta &= A_* - A_{\bar{\sigma}}, \end{aligned}$$

and obtain

$$\begin{aligned} \dot{x}_\delta &= A_{\bar{\sigma}}^\top x_\delta + A_\delta^\top \bar{x}_* + C^\top u_\delta, & x_\delta(t) &= 0, \\ \dot{p}_\delta &= -A_{\bar{\sigma}} p_\delta - A_\delta p_* - BR_{\bar{\sigma}} B^\top x_\delta - BR_\delta B^\top \bar{x}_*, & p_\delta(T) &= \Gamma_{\bar{\sigma}} x_\delta(T) + \Gamma_\delta \bar{x}_*(T), \\ u_\delta &= -Q_{\bar{\sigma}}^{-1}Cp_\delta - Q_{\delta^{-1}}Cp_*. \end{aligned}$$

Now a combination of standard calculus, Gronwall's Lemma, the norm bound on x_0, and Theorem 4.3 yield existence of a constant $c_\delta > 0$ depending on T, c_*, $\|A_{\bar\sigma}\|$, $\|B\|$, $\|C\|$, $\|R_{\bar\sigma}\|$, and $\|\Gamma_{\bar\sigma}\|$ but independent of t such that

$$\|x_\delta\|^2_{\mathcal{C}^T_t} + \|x_\delta\|^2_{\mathcal{H}^T_t} \leq c_\delta \left(\|A_\delta\|^2 + \|u_\delta\|^2_{\mathcal{L}^T_t} \right),$$
$$\|p_\delta\|^2_{\mathcal{C}^T_t} + \|p_\delta\|^2_{\mathcal{H}^T_t} \leq c_\delta \left(\|u_\delta\|^2_{\mathcal{L}^T_t} + \|A_\delta\|^2 + \|R_\delta\|^2 + \|\Gamma_\delta\|^2 \right). \tag{4.9}$$

Further, by testing the equation for $\dot{p}_\delta$ with x_δ we find

$$-\int_t^T (\dot{p}_\delta, x_\delta)\, \mathrm{d}s = \int_t^T (A_{\bar\sigma} p_\delta, x_\delta) + (A_\delta p_*, x_\delta) + (BR_{\bar\sigma}B^\top x_\delta, x_\delta) + (BR_\delta B^\top \bar{x}_*, x_\delta)\, \mathrm{d}s. \tag{4.10}$$

On the other hand, with $x_\delta(0) = 0$ and the identity for $p_\delta(T)$ we find testing the equation for $\dot{x}_\delta$ with p_δ

$$-\int_t^T (\dot{p}_\delta, x_\delta)\, \mathrm{d}s = \int_t^T (\dot{x}_\delta, p_\delta)\, \mathrm{d}s - (p_\delta(T), x_\delta(T))$$
$$= \int_t^T (A_{\bar\sigma}^\top x_\delta, p_\delta) + (A_\delta^\top \bar{x}_*, p_\delta) + (C^\top u_\delta, p_\delta)\, \mathrm{d}s - (\Gamma_{\bar\sigma} x_\delta(T), x_\delta(T)) - (\Gamma_\delta \bar{x}_*, x_\delta(T)). \tag{4.11}$$

By subtracting (4.10) from (4.11), rearranging terms and utilizing the identity for u_δ we arrive at

$$\delta\mathcal{J}_{\bar\sigma} := \int_t^T \|u_\delta\|^2_{Q_{\bar\sigma}} + \|B^\top x_\delta\|^2_{R_{\bar\sigma}}\, \mathrm{d}s + \|x_\delta(T)\|^2_{\Gamma_{\bar\sigma}}$$
$$= -\int_t^T (A_\delta p_*, x_\delta) + (BR_\delta B^\top \bar{x}_*, x_\delta) - (A_\delta^\top \bar{x}_*, p_\delta) + (Q_{\bar\sigma} u_\delta, Q_{\delta^{-1}} p_*)\, \mathrm{d}s$$
$$- (\Gamma_\delta \bar{x}_*(T), x_\delta(T)).$$

Utilizing Theorem 4.3, (4.9), and Young's inequality it follows that for any $\epsilon > 0$ it holds

$$\delta\mathcal{J}_{\bar\sigma} \leq c_* \int_t^T \|A_\delta\|\|x_\delta\| + \|B\|^2 \|R_\delta\|\|x_\delta\| + \|A_\delta\|\|p_\delta\| + \|Q_{\bar\sigma}\|\|Q_{\delta^{-1}}\|\|u_\delta\|\, \mathrm{d}s$$
$$+ c_* \|\Gamma_\delta\|\|x_\delta\|_{\mathcal{C}^T_t}$$
$$\leq c_* \int_t^T \frac{1}{\epsilon}\|A_\delta\|^2 + \epsilon\|x_\delta\|^2 + \frac{\|B\|^4}{2\epsilon}\|R_\delta\|^2 + \frac{\epsilon}{2}\|p_\delta\|^2 + \frac{\|Q_{\bar\sigma}\|^2}{2\epsilon}\|Q_{\delta^{-1}}\|^2 + \frac{c_*\epsilon}{2}\|u_\delta\|^2\, \mathrm{d}s$$
$$+ \frac{c_*}{2\epsilon}\|\Gamma_\delta\|^2 + \frac{\epsilon}{2}\|x_\delta\|^2_{\mathcal{C}^T_t}$$
$$\leq \frac{c_* T}{\epsilon}\|A_\delta\|^2 + \frac{c_* T\|B\|^4}{2\epsilon}\|R_\delta\|^2 + \frac{c_* T\|Q_{\bar\sigma}\|^2}{2\epsilon}\|Q_{\delta^{-1}}\|^2 + \frac{c_*}{2\epsilon}\|\Gamma_\delta\|^2 + \frac{c_*\epsilon}{2}\|u_\delta\|^2_{\mathcal{L}^T_t}$$
$$+ (\epsilon c_* c_\delta + \frac{\epsilon c_\delta}{2})(\|A_\delta\|^2 + \|u_\delta\|^2_{\mathcal{L}^T_t}) + \frac{\epsilon c_* c_\delta}{2}\left(\|u_\delta\|^2_{\mathcal{L}^T_t} + \|A_\delta\|^2 + \|R_\delta\|^2 + \|\Gamma_\delta\|^2 \right).$$

Since

$$\|u_\delta\|^2_{\mathcal{L}^T_t} \leq \|Q_{\bar\sigma}^{-\frac{1}{2}}\|^2 \int_t^T \|u_\delta\|^2_{Q_{\bar\sigma}}\, \mathrm{d}s,$$

for a sufficiently small $\epsilon > 0$ all terms on the right hand side depending on u_δ can be absorbed in the left hand side, and with appropriate constants $c_1, c_2 > 0$ it follows that

$$c_1 \delta\mathcal{J}_{\bar\sigma} \leq c_2 \left(\|A_\delta\| + \|\Gamma_\delta\| + \|R_\delta\| + \|Q_{\delta^{-1}}\| \right).$$

Finally, note that the mapping $\mathcal{F}(A) = A^{-1}$ is continuously differentiable on the open set of positive definite matrices with derivative $D\mathcal{F}(A)B = -A^{-1}BA^{-1}$. Hence an application of Taylor's theorem yields

$$\|Q_{\delta^{-1}}\| = \|Q_*^{-1} - Q_{\bar{\sigma}}^{-1}\| \le \|Q_{\bar{\sigma}}^{-1}\|^2 \|Q_* - Q_{\bar{\sigma}}\|$$

and the assertion follows.   □

With this estimate at hand we are in a position to estimate the difference of value functions.

Lemma 4.5. *Let $\Pi_{\bar{\sigma}}$ and Π_* be the solutions of* (4.2) *and* (4.5), *respectively. Then there exists a constant $c_\Pi > 0$ such that the following inequality holds*

$$\|\Pi_* - \Pi_{\bar{\sigma}}\|_{\mathcal{C}_t^{n,n}} \le c_\Pi \|\mathcal{S}_* - \mathcal{S}_{\bar{\sigma}}\|_1.$$

Proof. Due to Theorem 4.2 it suffices to estimate the difference of the value functions. For this purpose let $(\bar{x}_*, \bar{u}_*)$ and $(\bar{x}_{\bar{\sigma}}, \bar{u}_{\bar{\sigma}})$ be as in Theorem 4.4, let $x_0 \in \mathbb{R}^n$ with $\|x_0\| = 1$, and let $t \in [0, T)$ be fixed. It follows

$$|\nu_*(t, x_0) - \nu_{\bar{\sigma}}(t, x_0)| = |\mathcal{J}_*(\bar{x}_*, \bar{u}_*; t, x_0) - \mathcal{J}_{\bar{\sigma}}(\bar{x}_{\bar{\sigma}}, \bar{u}_{\bar{\sigma}}; t, x_0)|$$

$$\le \frac{1}{2} \int_t^T |\|\bar{u}_*\|_{Q_*}^2 - \|\bar{u}_{\bar{\sigma}}\|_{Q_{\bar{\sigma}}}^2| + |\|B^\top \bar{x}_*\|_{R_*}^2 - \|B^\top \bar{x}_{\bar{\sigma}}\|_{R_{\bar{\sigma}}}^2| \, \mathrm{d}s + \frac{1}{2}|\|\bar{x}_*(T)\|_{\Gamma_*}^2 - \|\bar{x}_{\bar{\sigma}}(T)\|_{\Gamma_{\bar{\sigma}}}^2|.$$
$$(4.12)$$

We estimate the terms on the right hand side individually. For the term involving the controls consider

$$\int_t^T |\|\bar{u}_*\|_{Q_*}^2 - \|\bar{u}_{\bar{\sigma}}\|_{Q_{\bar{\sigma}}}^2| \, \mathrm{d}s \le \int_t^T |(\bar{u}_*, Q_* \bar{u}_*) - (\bar{u}_*, Q_{\bar{\sigma}} \bar{u}_*)| + |(\bar{u}_*, Q_{\bar{\sigma}} \bar{u}_*) - (\bar{u}_{\bar{\sigma}}, Q_{\bar{\sigma}} \bar{u}_{\bar{\sigma}})| \, \mathrm{d}s$$

$$\le \|Q_* - Q_{\bar{\sigma}}\| \int_t^T \|\bar{u}_*\|^2 \, \mathrm{d}s + \|Q_{\bar{\sigma}}^{\frac{1}{2}}\| \int_t^T \|\bar{u}_* + \bar{u}_{\bar{\sigma}}\| \, \|\bar{u}_* - \bar{u}_{\bar{\sigma}}\|_{Q_{\bar{\sigma}}} \, \mathrm{d}s.$$

With Theorem 4.3 and an analogous result for the solution of the optimal control problem formulated in terms of $\mathcal{J}_{\bar{\sigma}}$ with some constant $c_{\bar{\sigma}} > 0$ we obtain

$$\int_t^T |\|\bar{u}_*\|_{Q_*}^2 - \|\bar{u}_{\bar{\sigma}}\|_{Q_{\bar{\sigma}}}^2| \, \mathrm{d}s \le c_* \|Q_* - Q_{\bar{\sigma}}\| + (\sqrt{c_*} + \sqrt{c_{\bar{\sigma}}})\sqrt{T} \, \|Q_{\bar{\sigma}}^{\frac{1}{2}}\| \sqrt{\int_t^T \|\bar{u}_* - \bar{u}_{\bar{\sigma}}\|_{Q_{\bar{\sigma}}}^2 \, \mathrm{d}s}.$$

Similarly we find

$$\int_t^T |\|B^\top \bar{x}_*\|_{R_*}^2 - \|B^\top \bar{x}_{\bar{\sigma}}\|_{R_{\bar{\sigma}}}^2| \, \mathrm{d}s$$

$$\le c_* \|B\|^2 \|R_* - R_{\bar{\sigma}}\| + (\sqrt{c_*} + \sqrt{c_{\bar{\sigma}}})\sqrt{T} \, \|B\| \, \|R_{\bar{\sigma}}^{\frac{1}{2}}\| \sqrt{\int_t^T \|B^\top(\bar{x}_* - \bar{x}_{\bar{\sigma}})\|_{R_{\bar{\sigma}}}^2 \, \mathrm{d}s}$$

and

$$|\|\bar{x}_*(T)\|_{\Gamma_*}^2 - \|\bar{x}_{\bar{\sigma}}(T)\|_{\Gamma_{\bar{\sigma}}}^2| \le c_* \|\Gamma_* - \Gamma_{\bar{\sigma}}\| + \|\Gamma_{\bar{\sigma}}^{\frac{1}{2}}\|(\sqrt{c_*} + \sqrt{c_{\bar{\sigma}}})\|\bar{x}_*(T) - \bar{x}_{\bar{\sigma}}(T)\|_{\Gamma_{\bar{\sigma}}}.$$

Hence with (4.12) and $c_1 = c_* \max(\|B\|^2, 1)$ and an appropriate $c_2 > 0$ it follows

$$|\nu_*(t, x_0) - \nu_{\bar{\sigma}}(t, x_0)| \le c_1 (\|\Gamma_* - \Gamma_{\bar{\sigma}}\| + \|R_* - R_{\bar{\sigma}}\| + \|Q_* - Q_{\bar{\sigma}}\|)$$

$$+ c_2 \sqrt{\mathcal{J}_{\bar{\sigma}}(\bar{x}_* - \bar{x}_{\bar{\sigma}}, \bar{u}_* - \bar{u}_{\bar{\sigma}}; t, x_0)}.$$

Finally with Theorem 4.2 and Theorem 4.4 we obtain

$$\|\Pi_*(T-t) - \Pi_{\bar{\sigma}}(T-t)\| \leq c_1 \left(\|\Gamma_* - \Gamma_{\bar{\sigma}}\| + \|R_* - R_{\bar{\sigma}}\| + \|Q_* - Q_{\bar{\sigma}}\|\right)$$
$$+ c_2\sqrt{c_{\mathcal{J}}} \left(\|A_* - A_{\bar{\sigma}}\| + \|\Gamma_* - \Gamma_{\bar{\sigma}}\| + \|R_* - R_{\bar{\sigma}}\| + \|Q_* - Q_{\bar{\sigma}}\|\right),$$

implying the estimate for $t \in (0, T]$. Due to the initial condition of the Riccati equations the estimate holds in $t = 0$ as well and the assertion is shown. $\qquad\square$

Combining the estimate for the difference of error covariances with another application of Gronwall's Lemma we estimate the difference of the state reconstruction trajectories.

Lemma 4.6. *Let $\widehat{x}_*$, Π_*, $\widehat{x}_{\bar{\sigma}}$, and $\Pi_{\bar{\sigma}}$ be the solutions of* (4.4), (4.5), (4.1), *and* (4.2) *respectively. Then there exists a constant $c_{x_*} > 0$ independent of $t \in [0, T]$ such that*

$$\|\widehat{x}_*(t) - \widehat{x}_{\bar{\sigma}}(t)\| \leq c_{x_*}\|\mathcal{S}_* - \mathcal{S}_{\bar{\sigma}}\|_1.$$

Proof. For $s \in [0, T]$ denote $e(s) = \widehat{x}_*(s) - \widehat{x}_{\bar{\sigma}}(s)$. Subtracting (4.1) from (4.4) and rearranging terms shows that e is characterized as the unique solution to

$$\dot{e}(s) = \left(A_{\bar{\sigma}} - \Pi_{\bar{\sigma}}(s)C^\top Q_{\bar{\sigma}}^{-1}C\right)e(s) + \left(\Pi_*(s) - \Pi_{\bar{\sigma}}(s)\right)C^\top Q_{\bar{\sigma}}^{-1}\left(y(s) - C\widehat{x}_*(s)\right)$$
$$+ (A_* - A_{\bar{\sigma}})\widehat{x}_*(s) + \Pi_*(s)C^\top\left(Q_*^{-1} - Q_{\bar{\sigma}}^{-1}\right)(y(s) - C\widehat{x}_*(t)), \qquad s \in (0, T),$$
$$e(0) = 0.$$

Testing the first equation with $e(s)$ and integrating over $(0, t)$ for some fixed $t \in (0, T]$ yields

$$\frac{1}{2}\|e(t)\|^2 \leq \int_0^t \left(\|A_{\bar{\sigma}}\| + \|C\|^2\|Q_{\bar{\sigma}}^{-1}\|\|\Pi_{\bar{\sigma}}(s)\|\right)\|e(s)\|^2 + \|A_* - A_{\bar{\sigma}}\|\|e(s)\|$$
$$+ \|\Pi_*(s) - \Pi_{\bar{\sigma}}(s)\|\|C\|\|Q_{\bar{\sigma}}^{-1}\|\|y(s) - C\widehat{x}_*(s)\|\|e(s)\|$$
$$+ \|\Pi_*(s)\|\|C\|\|Q_*^{-1} - Q_{\bar{\sigma}}^{-1}\|\|y(s) - C\widehat{x}_*(s)\|\|e(s)\| \, \mathrm{d}s.$$

With Young's inequality we find

$$\|e(t)\|^2 \leq \int_0^t \left(2\|A_{\bar{\sigma}}\| + 2\|C\|^2\|Q_{\bar{\sigma}}^{-1}\|\|\Pi_{\bar{\sigma}}(s)\| + 3\right)\|e(s)\|^2 \, \mathrm{d}s + T\|A_* - A_{\bar{\sigma}}\|^2$$
$$+ \|C\|^2\|Q_{\bar{\sigma}}^{-1}\|^2\|y - C\widehat{x}_*\|^2_{\mathcal{L}_0^T}\|\Pi_* - \Pi_{\bar{\sigma}}\|^2_{\mathcal{C}_T^{n,n}}$$
$$+ \|C\|^2\|\Pi_*\|^2_{\mathcal{C}_T^{n,n}}\|y - C\widehat{x}_*\|^2_{\mathcal{L}_0^T}\|Q_*^{-1} - Q_{\bar{\sigma}}^{-1}\|^2,$$

and finally Gronwall's inequality yields

$$\|e(t)\| \leq c\left(\|A_* - A_{\bar{\sigma}}\| + \|Q_*^{-1} - Q_{\bar{\sigma}}^{-1}\| + \|\Pi_* - \Pi_{\bar{\sigma}}\|_{\mathcal{C}_T^{n,n}}\right),$$

for some appropriate constant $c > 0$. Again we estimate $\|Q_*^{-1} - Q_{\bar{\sigma}}^{-1}\|$ in terms of $\|Q_* - Q_{\bar{\sigma}}\|$ as we did in the proof of Theorem 4.4. The assertion now follows with Theorem 4.5 and a constant $c_{x_*} > 0$ depending on T, $\|C\|$, $\|A_{\bar{\sigma}}\|$, $\|Q_{\bar{\sigma}}^{-1}\|$, $\|y - C\widehat{x}_*\|_{\mathcal{L}_0^T}$, $\|\Pi_*\|_{\mathcal{C}_T^{n,n}}$, and $\|\Pi_{\bar{\sigma}}\|_{\mathcal{C}_T^{n,n}}$. $\qquad\square$

This concludes the technical preparations and we turn to the error estimation for the three state estimators discussed in Section 3.

4.2. **Worst case error estimation.** We begin by estimating the difference of the estimator $\widehat{x}_{\mathbb{E}}$ introduced in Subsection 3.2.

Proposition 4.7. *There exists a constant $c_{\mathbb{E}} > 0$ independent of N such that for all $t \in [0, T]$ it holds*

$$\|\widehat{x}_{\mathbb{E}}(t) - \widehat{x}_{\bar{\sigma}}(t)\|_{P_{\bar{\sigma}(t)}} \leq c_{\mathbb{E}} \; \mathbb{E}\left[\|\mathcal{S}_{\sigma} - \mathcal{S}_{\bar{\sigma}}\|_1\right].$$

Proof. We apply the results of Subsection 4.1 with

$$A_* = A_{\mathbb{E}} = \frac{1}{N} \sum_{k=1}^{N} A_k, \qquad \Gamma_* = \Gamma_{\mathbb{E}} = \frac{1}{N} \sum_{k=1}^{N} \Gamma_k,$$

$$R_* = R_{\mathbb{E}} = \frac{1}{N} \sum_{k=1}^{N} R_k, \qquad Q_* = Q_{\mathbb{E}} = \frac{1}{N} \sum_{k=1}^{N} Q_k,$$

and Theorem 4.6 yields existence of $c_{x_*} > 0$ independent of N such that for any $t \in [0, T]$ it holds

$$\|\widehat{x}_{\mathbb{E}}(t) - \widehat{x}_{\bar{\sigma}}(t)\|_{P_{\bar{\sigma}(t)}} \leq \|P_{\bar{\sigma}}^{\frac{1}{2}}(t)\| \; \|\widehat{x}_{\mathbb{E}}(t) - \widehat{x}_{\bar{\sigma}}(t)\|$$

$$\leq c_{x_*} \|P_{\bar{\sigma}}^{\frac{1}{2}}(t)\| \left(\|A_{\mathbb{E}} - A_{\bar{\sigma}}\| + \|\Gamma_{\mathbb{E}} - \Gamma_{\bar{\sigma}}\| + \|R_{\mathbb{E}} - R_{\bar{\sigma}}\| + \|Q_{\mathbb{E}} - Q_{\bar{\sigma}}\|\right).$$

We note that $P_{\bar{\sigma}}$ is continuous in $[0, T]$ and hence admits a maximum. Inserting the definition for $A_{\mathbb{E}}$, $\Gamma_{\mathbb{E}}$, $R_{\mathbb{E}}$, and $Q_{\mathbb{E}}$ and applying the Cauchy Schwarz inequality yields the assertion. $\qquad\square$

Similar estimates can be established for the reconstruction of the state via $\widehat{x}_{\varnothing}$ and $\widehat{x}_{\mathbb{E}}$ as presented in Subsection 3.3 and Subsection 3.4, respectively. For this purpose, let $k \in \{1, \ldots, N\}$. Applying the results of Subsection 4.1 with

$$A_* = A_k, \quad \Gamma_* = \Gamma_k, \quad R_* = R_k, \quad Q_* = Q_k,$$

we find that Theorem 4.6 yields the existence of $c_{x_k} > 0$ such that for any $t \in [0, T]$ it holds

$$\|\widehat{x}_k(t) - \widehat{x}_{\bar{\sigma}}(t)\| \leq c_{x_k} \left(\|A_k - A_{\bar{\sigma}}\| + \|\Gamma_k - \Gamma_{\bar{\sigma}}\| + \|R_k - R_{\bar{\sigma}}\| + \|Q_k - Q_{\bar{\sigma}}\|\right). \qquad (4.13)$$

The estimate for $\widehat{x}_{\varnothing}$ readily follows.

Proposition 4.8. *There exists a constant $c_{\varnothing} > 0$ independent of N such that for all $t \in [0, T]$ we obtain that*

$$\|\widehat{x}_{\varnothing}(t) - \widehat{x}_{\bar{\sigma}}(t)\|_{P_{\bar{\sigma}(t)}} \leq c_{\varnothing} \; \mathbb{E}\left[\|\mathcal{S}_{\sigma} - \mathcal{S}_{\bar{\sigma}}\|_1\right].$$

Proof. Looking at the definition of $\widehat{x}_{\varnothing}$ in (3.6) and utilizing (4.13) we find

$$\|\widehat{x}_{\varnothing}(t) - \widehat{x}_{\bar{\sigma}}(t)\|_{P_{\bar{\sigma}(t)}} \leq \|P_{\bar{\sigma}}^{\frac{1}{2}}(t)\| \; \|\frac{1}{N} \sum_{k=1}^{N} \widehat{x}_k(t) - \widehat{x}_{\bar{\sigma}}(t)\|$$

$$\leq \|P_{\bar{\sigma}}^{\frac{1}{2}}(t)\| \; \frac{1}{N} \sum_{k=1}^{N} c_{x_k} \left(\|A_{\sigma} - A_{\bar{\sigma}}\| + \|\Gamma_{\sigma} - \Gamma_{\bar{\sigma}}\| + \|R_{\sigma} - R_{\bar{\sigma}}\| + \|Q_{\sigma} - Q_{\bar{\sigma}}\|\right).$$

The assertion follows with $c_{\varnothing} = \left(\max_{t \in [0,T]} \|P_{\bar{\sigma}}^{\frac{1}{2}}(t)\|\right) \left(\max_{k \in \{1,\ldots,N\}} c_{x_k}\right)$. $\qquad\square$

Before we can establish an analogous estimate for the minimizer of the expected energy $\widehat{x}_{\mathrm{E}}$ discussed in Subsection 3.4 we take a closer look at the precision matrices $P_k(t)$. In the following, let $\mathcal{P}(\cdot) = \sum_{k=1}^{N} P_k(t)$ be as in Theorem 3.4.

Lemma 4.9. *There exists a constant $c_{\mathcal{P}} > 0$ independent of N such that for all $t \in [0,T]$ it holds*

$$\|\mathcal{P}^{-1}(t)\| \leq \frac{c_{\mathcal{P}}}{N}.$$

Proof. For all non-trivial $x \in \mathbb{R}^n$ and $t \in [0,T]$ we find

$$(x, \mathcal{P}(t)x) = \sum_{k=1}^{N} (x, P_k(t)x) \geq \sum_{k=1}^{N} \min_{s \in [0,T]} (x, P_k(s)x),$$

where the existence of the minima is ensured by the continuity of the matrices. We denote the (possibly non-unique) time at which the minimum of the k-th summand is assumed by s_k^*. In addition we denote by $\lambda_k > 0$ the smallest eigenvalue of $P_k(s_k^*)$, where the positivity follows from the positive definiteness of the precision matrices. Denoting $\bar{\lambda} = \max\limits_{k=1,\dots,N} \lambda_k > 0$ it follows

$$(x, \mathcal{P}(t)x) \geq \sum_{k=1}^{N} \lambda_k \|x\|^2 \geq N\bar{\lambda}\|x\|^2.$$

The proof is completed with an application of the Lax-Milgram Lemma [3, Thm. 1.1.3 & Rem. 1.1.3]. $\square$

We conclude this section with the announced estimate for $\widehat{x}_{\mathrm{E}}$.

Proposition 4.10. *There exists a constant $c_{\mathrm{E}} > 0$ independent of N such that for all $t \in [0,T]$ it holds*

$$\|\widehat{x}_{\mathrm{E}}(t) - \widehat{x}_{\bar{\sigma}}(t)\|_{P_{\bar{\sigma}(t)}} \leq c_{\mathrm{E}} \, \mathbb{E}\left[\|\mathcal{S}_\sigma - \mathcal{S}_{\bar{\sigma}}\|_1\right].$$

Proof. With Theorem 3.4 for every $t \in [0,T]$ we obtain

$$\|\widehat{x}_{\mathrm{E}}(t) - \widehat{x}_{\bar{\sigma}}(t)\|_{P_{\bar{\sigma}(t)}} \leq \|P_{\bar{\sigma}}^{\frac{1}{2}}(t)\| \left\|\mathcal{P}^{-1}(t)\right\| \sum_{k=1}^{N} \|P_k(t)\| \, \|\widehat{x}_k(t) - \widehat{x}_{\bar{\sigma}}(t)\|.$$

The continuity of $P_k(\cdot)$ implies existence of $M > 0$ such that for all $t \in [0,T]$ and $k = 1, \dots, N$ it holds $\|P_k(t)\| \leq M$. Utilizing (4.13) and Theorem 4.9 yields

$$\|\widehat{x}_{\mathrm{E}}(t) - \widehat{x}_{\bar{\sigma}}(t)\|_{P_{\bar{\sigma}(t)}}$$
$$\leq M\frac{c_{\mathcal{P}}}{N}\|P_{\bar{\sigma}}^{\frac{1}{2}}(t)\| \sum_{k=1}^{N} \left(\|A_k - A_{\bar{\sigma}}\| + \|\Gamma_k - \Gamma_{\bar{\sigma}}\| + \|R_k - R_{\bar{\sigma}}\| + \|Q_k - Q_{\bar{\sigma}}\|\right)$$

and the claim follows with $c_{\mathrm{E}} = Mc_{\mathcal{P}} \max\limits_{s \in [0,T]} \|P_{\bar{\sigma}}^{\frac{1}{2}}(s)\|$. $\square$

The worst case error estimates presented in Theorem 4.7, Theorem 4.8, and Theorem 4.10 immediately imply bounds for the expected errors.

Corollary 4.11. *Let $\widehat{x}_{\mathbb{E}}$, $\widehat{x}_{\varnothing}$, and $\widehat{x}_{\mathrm{E}}$ be the estimators introduced in Section 3 and let $c_{\mathbb{E}}$, $c_{\varnothing}$, and c_{E} be the constants from Theorem 4.7, Theorem 4.8, and Theorem 4.10, respectively. For all $t \in [0,T]$ it holds*

$$\mathbb{E}\left[\|\widehat{x}_{\mathbb{E}}(t) - \widehat{x}_{\sigma}(t)\|_{P_{\sigma(t)}}\right] \leq c_{\mathbb{E}} \, \mathbb{E}_{\bar{\sigma}}\left[\mathbb{E}_{\sigma}\left[\|\mathcal{S}_{\sigma} - \mathcal{S}_{\bar{\sigma}}\|_1\right]\right],$$

$$\mathbb{E}\left[\|\widehat{x}_{\varnothing}(t) - \widehat{x}_{\sigma}(t)\|_{P_{\sigma(t)}}\right] \leq c_{\varnothing} \, \mathbb{E}_{\bar{\sigma}}\left[\mathbb{E}_{\sigma}\left[\|\mathcal{S}_{\sigma} - \mathcal{S}_{\bar{\sigma}}\|_1\right]\right],$$

$$\mathbb{E}\left[\|\widehat{x}_{\mathrm{E}}(t) - \widehat{x}_{\sigma}(t)\|_{P_{\sigma(t)}}\right] \leq c_{\mathrm{E}} \, \mathbb{E}_{\bar{\sigma}}\left[\mathbb{E}_{\sigma}\left[\|\mathcal{S}_{\sigma} - \mathcal{S}_{\bar{\sigma}}\|_1\right]\right].$$

Proof. The proof follows by forming the expectation on both sides of the corresponding worst case error estimates. □

We found that all three state estimators discussed in this work satisfy the same qualitative worst case error bound. They express the fact that these errors can be quantified by the level of variance of the parameter-dependent matrices $A_{\bar{\sigma}}, \Gamma_{\bar{\sigma}}, R_{\bar{\sigma}}, Q_{\bar{\sigma}}$.

5. Numerical experiments

In this section we numerically realize and compare the three estimators for two examples of uncertain linear dynamical systems.

5.1. **General setup.** We consider disturbed uncertain linear systems of the form

$$\begin{aligned}
\dot{x}(t) &= A_{\sigma}x(t) + Bv(t) &\quad t &\in (0,T),\\
x(0) &= x_0 + \eta,\\
y(t) &= Cx(t) + \mu(t) &\quad t &\in (0,T),
\end{aligned} \tag{5.1}$$

where $\sigma \in \Sigma_A \subset \mathbb{R}^{s_A}$ with $|\Sigma_A| = N_A$, as in Subsection 3.1. We associate positive definite weighting/covariance matrices Γ, R, and Q with the three unknown errors η, v, and μ. Here we do not investigate uncertainties in the noise covariance matrices implying $N_{\Gamma} = N_R = N_Q = 1$ leading to $N = N_A$.

All ODEs are solved using an equidistant grid $0 = t_0 < t_1 < \cdots < t_{1000} = T$ and the MATLAB® solver `ode15s` with a relative tolerance of 10^{-8}. In particular, Kalman filter equations of the form (2.10)-(2.11) are treated as $n + n^2$ dimensional ODEs. For systems of higher dimension n this can certainly be improved upon by taking advantage of the decoupling of the two equations and utilizing solvers tailored to Riccati equations.

The measured output $y \in \mathcal{L}_0^T$ is artificially generated as follows. Using the MATLAB® function `normrnd` we construct realizations of $\eta \sim \mathcal{N}(0,\Gamma)$, $v(t_k) \sim \mathcal{N}(0,R)$, and $\mu(t_k) \sim \mathcal{N}(0,Q)$ for $k = 0, \ldots, 1000$ from which $v \in \mathcal{L}_0^T$, $\mu \in \mathcal{L}_0^T$ are obtained via linear interpolation. Finally, we designate some $\bar{\sigma} \in \Sigma_A$ to be the true, unknown parameter and obtain y by solving (5.1) based on the constructed errors and the parameter $\bar{\sigma}$.

Remark 5.1. *We point out that this realization of the unknown errors based on normal distributions can only be done rigorously in discrete time. The extension to the continuous time setting via interpolation is of heuristic nature as a construction in the spirit of $v(t) \sim \mathcal{N}(0,R)$, $t \in (0,T)$ leads to the stochastic formulation and the regularity gap discussed in Theorem 2.2.*

The construction of the estimators requires the realization of the N_A individual Kalman filters $\widehat{x}_k$. These are independent of each other, and therefore they are computed in parallel using `parfor` from the MATLAB® Parallel Computing Toolbox.

The results of this section were realized in MATLAB® R2024b and the computations were executed on a Lenovo ThinkPad E14 Intel(R) Core(TM) Ultra 7 155H 1.40 GHz with 32 GB memory. The code is available in [19].

5.2. **Harmonic oscillator.** As a first example we consider a disturbed harmonic oscillator modeled as

$$m\ddot{z}(t) + \sigma\dot{z}(t) + kz(t) = v(t), \qquad t \in (0, T)$$
$$z(0) = x_{0,1} + \eta_1, \ \dot{z}(0) = x_{0,2} + \eta_2,$$

where $z(t)$ represents the position at time t and $m > 0$, $\sigma \geq 0$, and $k \geq 0$ represent the mass, damping coefficient, and spring constant, respectively. The function $v \in \mathcal{L}_0^T$ represents the unknown disturbance in the dynamics. The modeled initial position and velocity are given by $x_{0,1}$ and $x_{0,2}$ and are subject to the initial errors η_1 and η_2. Further, we assume access to a disturbed measurement of the position modeled as

$$y(t) = z(t) + \mu(t) \quad t \in (0, T),$$

with a deterministic but unknown output error $\mu \in \mathcal{L}_0^T$. In first order form the system reads

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{\sigma}{m} \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} v(t) \quad t \in (0, T),$$
$$x(0) = x_0 + \eta,$$
$$y(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} x(t) + \mu(t) \qquad t \in (0, T),$$

with modeled initial state $x_0 = \begin{bmatrix} x_{0,1} & x_{0,2} \end{bmatrix}^\top$.

For our experiments we fix the mass and spring constant to $m = k = 1$ and set the time horizon to $T = 10$. As an undisturbed initial state we set $x_0 = \begin{bmatrix} 1 & 0 \end{bmatrix}^\top$ and the covariances of the three errors are given as $\Gamma = 0.1 \,\mathrm{Id}$, $R = 0.05$, and $Q = 0.05$. The uncertainty of the dynamics lies in the damping parameter σ and we set $N = N_A = 101$ and $\Sigma_A = \{0.1 + \frac{k}{N_A-1} 2.9 \colon k = 0, \ldots, N_A - 1\}$. We generate two outputs based on the parameters $\bar{\sigma} = 3$ and $\bar{\sigma} = 0.1$ as described in the previous subsection. The results are presented in Figure 1 and we observe the following.

We first consider the setting of $\bar{\sigma} = 3$, i.e., the underlying true damping parameter is the largest appearing in the family Σ_A. In Figure 1a we present the phase plots of the three estimators $\widehat{x}_{\mathbb{E}}$, $\widehat{x}_\varnothing$, and $\widehat{x}_{\mathrm{E}}$ resulting from the associated output y and compare them with the Kalman filter $\widehat{x}_{\bar{\sigma}}$ constructed based on the hidden parameter $\bar{\sigma} = 3$ and the associated state trajectory x. The trajectories starting in their respective initial states marked as crosses stabilize in a region around the origin. As expected, the Kalman filter $\widehat{x}_{\bar{\sigma}}$ successfully approximates the state trajectory x. The plot further illustrates that among the estimators constructed without knowledge of the hidden parameter the energy minimizer $\widehat{x}_{\mathrm{E}}$ yields the best approximation of $\widehat{x}_{\bar{\sigma}}$ with respect to the Euclidean distance. In addition Figure 1c shows that the energy minimizer outperforms the two remaining estimators with respect to the Mahalanobis distance. We proceed with the results associated with $\bar{\sigma} = 0.1$. Here the true damping parameter is the smallest appearing in Σ_A. The associated phase plot presented in Figure 1b shows that all three estimators $\widehat{x}_{\mathbb{E}}$, $\widehat{x}_\varnothing$, and $\widehat{x}_{\mathrm{E}}$ noticeably deviate from the Kalman filter $\widehat{x}_{\bar{\sigma}}$. The associated Mahalanobis differences are displayed in Figure 1d and we find that, in contrast to the case of $\bar{\sigma} = 3$, here the energy minimizer $\widehat{x}_{\mathrm{E}}$ does not outperform $\widehat{x}_{\mathbb{E}}$ and $\widehat{x}_\varnothing$. Note, however, that the accuracies of the three are noticeably more similar than in the previous setting.

In an attempt to interpret these results the reader is reminded of the formulas for $\widehat{x}_\varnothing$ and $\widehat{x}_\mathrm{E}$. According to (3.6) the estimator $\widehat{x}_\varnothing$ is given as an (unweighted) mean of the family of Kalman filters x_k. On the other hand the energy minimizer $\widehat{x}_\mathrm{E}$ is characterized as a weighted mean where the precision matrices P_k act as weights. In fact, if all P_k were diagonal, each component of $\widehat{x}_\mathrm{E}$ would be a mean of the associated components of $\widehat{x}_k$ weighted by their respective precisions. In such a setting one expects $\widehat{x}_\mathrm{E}$ to strongly resemble the family members $\widehat{x}_k$ associated with high precisions, while family members with lower precisions are neglected.

While the precision matrices can not be expected to be diagonal, we do find that for this example they are diagonally dominant. To illustrate this, for a given parameter $\sigma \in \Sigma_A$ we consider the quantity

$$0 \le d_\sigma(t) = \min_{j=1,\dots,n} \frac{|p_{jj}(t)|}{\sum_{i=1}^n |p_{ji}(t)|} \le 1,$$

where $pij(t)$ are the entries of $P_\sigma(t)$. Note that since the precision matrices are positive definite, the denominator is non zero. Now $P_\sigma(t)$ is diagonally dominant if $d_\sigma(t) \ge 0.5$ and diagonal if $d_\sigma(t) = 1$. The announced diagonal dominance of the precision matrices is illustrated in Figure 1f. In Figure 1e we display the generalized precisions (cf., Theorem 2.1)

$$p_\sigma(t) = \det(P_\sigma(t))$$

to quantify the precisions associated with the family members and find that it depends monotonously increasing on the damping parameter. This fits the intuition that larger damping parameters lead to stronger dissipation of errors. Note also that the precisions increase over time reflecting the fact that the effect of the initial error η decreases over time.
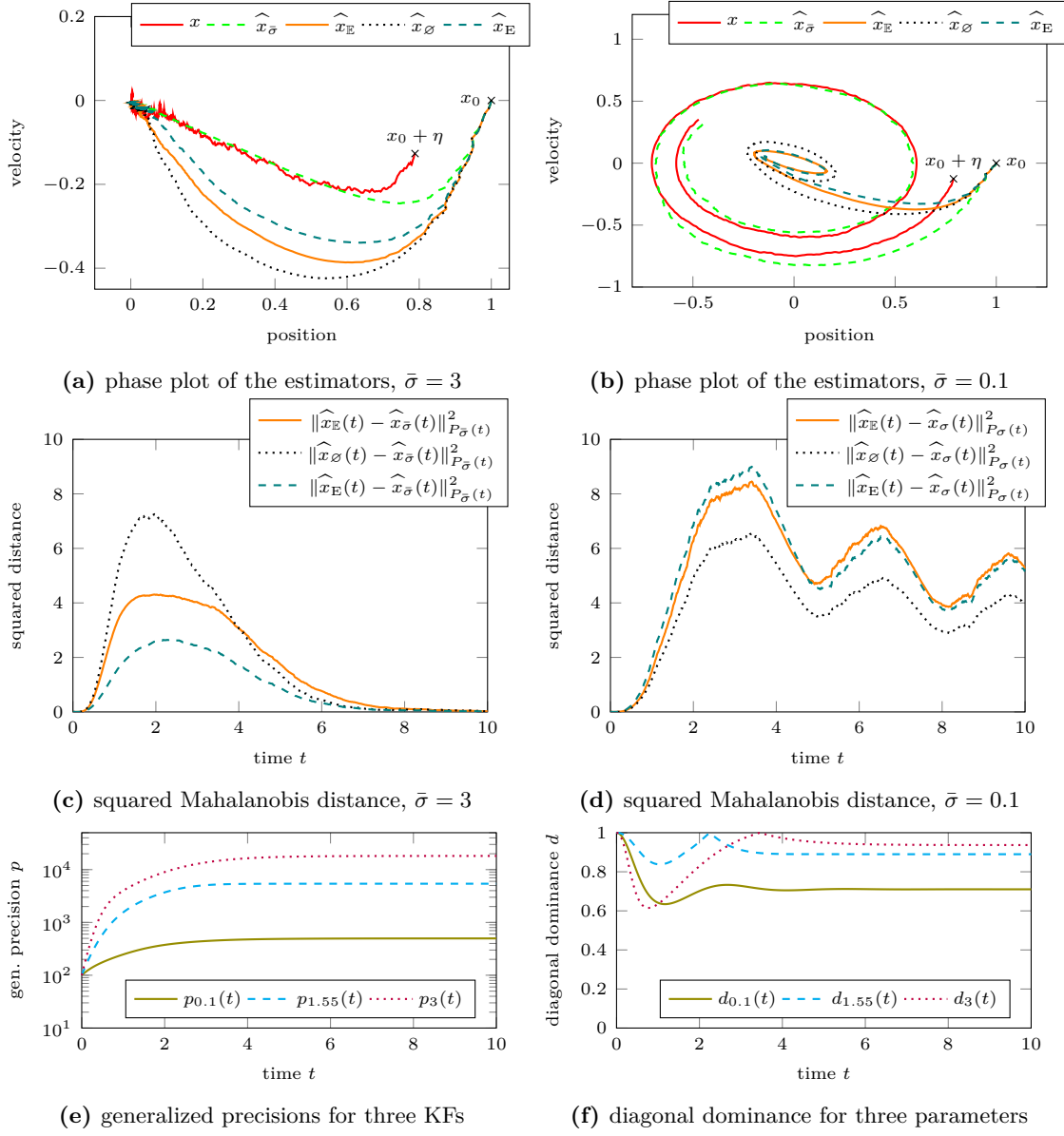
These considerations offer an explanation for the observations made while comparing the effects of the hidden parameters $\bar{\sigma} = 3$ and $\bar{\sigma} = 0.1$. Since the precision matrices are diagonally dominant, we expect the estimators to behave similarly to the setting of diagonal precision matrices. Hence the weighted mean $\widehat{x}_\mathrm{E}$ favors family members with higher precisions, i.e., trajectories associated with larger damping parameters. Finally, we note that all three designed estimators display noticeable differences from the Kalman filter based on the hidden parameter. This, however, is not surprising as the set of damping parameters leads to a family of system matrices A_σ representing a wide range of dynamics, making the state reconstruction a rather challenging task.

5.3. **Connected amplidynes.** Our second example is given by an electrical circuit that amplifies a given input. An amplidyne returns an amplified copy of the input signal and can be modeled as a linear dynamical system, see [12, Ch. 1.12]. We consider a system given by two connected amplidynes, i.e., the output of the first one serves as an input of the second one. Hence the resulting circuit consists of four components and the associated dynamics are modeled as

$$\dot{x}(t) = \begin{bmatrix} -\frac{\rho_1}{L_1} & 0 & 0 & 0 \\ \frac{k_1}{L_2} & -\frac{\rho_2}{L_2} & 0 & 0 \\ 0 & \frac{k_2}{L_3} & -\frac{\rho_3}{L_3} & 0 \\ 0 & 0 & \frac{k_3}{L_4} & -\frac{\rho_4}{L_4} \end{bmatrix} x(t) + \begin{bmatrix} \frac{e_0(t)}{L_1} \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{1}{L_1} \\ 0 \\ 0 \\ 0 \end{bmatrix} v(t) \quad t \in (0,T),$$

$$x(0) = x_0 + \eta,$$

$$y(t) = \begin{bmatrix} 0 & 0 & 0 & k_4 \end{bmatrix} x(t) + \mu(t) \quad\quad\quad\quad\quad\quad t \in (0,T),$$

(a) phase plot of the estimators, $\bar{\sigma} = 3$

(b) phase plot of the estimators, $\bar{\sigma} = 0.1$

(c) squared Mahalanobis distance, $\bar{\sigma} = 3$

(d) squared Mahalanobis distance, $\bar{\sigma} = 0.1$

(e) generalized precisions for three KFs

(f) diagonal dominance for three parameters

Figure 1 – Harmonic oscillator with uncertain damping

where $x(t) \in \mathbb{R}^4$ describes the current of the four components at time t. The positive parameters ρ_i and L_i represent the resistance and inductance of the i-th component, respectively. Further the constants $k_i > 0$ describe the proportion of the currents x_i and the voltages e_i in the components, i.e., for $i = 1, 2, 3, 4$ it holds $e_i = k_i x_i$. The known, time-dependent input entering the first component is denoted by e_0 and is subject to the disturbance v. Finally, as an output of the system we measure the output of the second amplidyne $e_4 = k_4 x_4$. We note that strictly speaking our formulation of the Kalman filter does not allow for a known forcing term like the input e_0 of the first amplidyne. This, however, can be incorporated without any major challenges.

For our implementation we fix $\rho_1 = \rho_3 = 5$, $\rho_2 = \rho_4 = 10$, $k_1 = k_3 = 20$, $k_2 = k_4 = 50$, $L_1 = 0.5$, $T = 10$, and $e_0(t) \equiv 1$. The undisturbed initial state is set to $x_0 = \begin{bmatrix} 0.5 & 1 & 10 & 20 \end{bmatrix}^\top$ and the uncertainty of the system lies in the inductances $\sigma = \begin{bmatrix} L_2 & L_3 & L_4 \end{bmatrix}^\top \in \mathbb{R}^3$. We define $\Sigma_1 = \{10, 12.5, 15, 17.5, 20\}$, $\Sigma_2 = \{0.5, 0.75, 1, 1.25, 1.5\}$, and $\Sigma_3 = \{10, 17.5, 25, 32.5, 40\}$ and construct $\Sigma_A = \{\sigma \in \mathbb{R}^3 \colon \sigma_k \in \Sigma_k, k = 1, \dots, 3\}$. We obtain $N = N_A = 125$ and $s_A = 3$. The output y is generated as described in Subsection 5.1 according to the covariance matrices

$$\Gamma = 0.25 \, \mathrm{diag}(|x_{0,1}|, |x_{0,2}|, |x_{0,3}|, |x_{0,4}|) = \mathrm{diag}(0.125, 0.25, 2.5, 5),$$
$$R = (0.1|e_0(0)|)^2 = 0.01,$$
$$Q = (0.1 * 400|e_0(0)|)^2 = 1600,$$

and the true parameter $\bar{\sigma} = \begin{bmatrix} 10 & 0.5 & 10 \end{bmatrix}^\top$. We note that the covariance matrices are of different magnitudes to fit the scales of the different components. We illustrate the relationship between the input e_0 and the output $k_4 x_4 = e_4$ and the impact of the errors in Figure 2a and Figure 2b. We observe that the output of the two connected amplidynes associated with $\bar{\sigma}$ stabilizes at a copy of e_0 increased by a factor of 400.

Our focus lies on the investigation of the structural differences to the first example. In Figure 2c we present the second component of the disturbed trajectory and the various estimators. Additionally we plot in magenta color the minima and maxima of the family of Kalman filters denoted by

$$\widehat{x}_{-,2}(t) = \min_{k=1,\dots,N} \widehat{x}_{k,2}(t) \qquad \text{and} \qquad \widehat{x}_{+,2}(t) = \max_{k=1,\dots,N} \widehat{x}_{k,2}(t),$$

illustrating their convex hull. As expected the unweighted average $\widehat{x}_\varnothing$ lies in this convex hull. This is different from the energy minimizer $\widehat{x}_{\mathrm{en}}$ which for some times t has values outside this hull. This indicates that the off diagonal entries of the precision matrices P_k play an important role. This corresponds to the fact that the precision matrices are by no means diagonal dominant. In Figure 2f we present the quantities

$$d_k = d_{\sigma_k}(t) = \min_{j=1,\dots,n} \frac{|p_{jj}^{\sigma_k}(t)|}{\sum_{i=1}^n |p_{ji}^{\sigma_k}(t)|} \quad k = 1, \dots, 5,$$
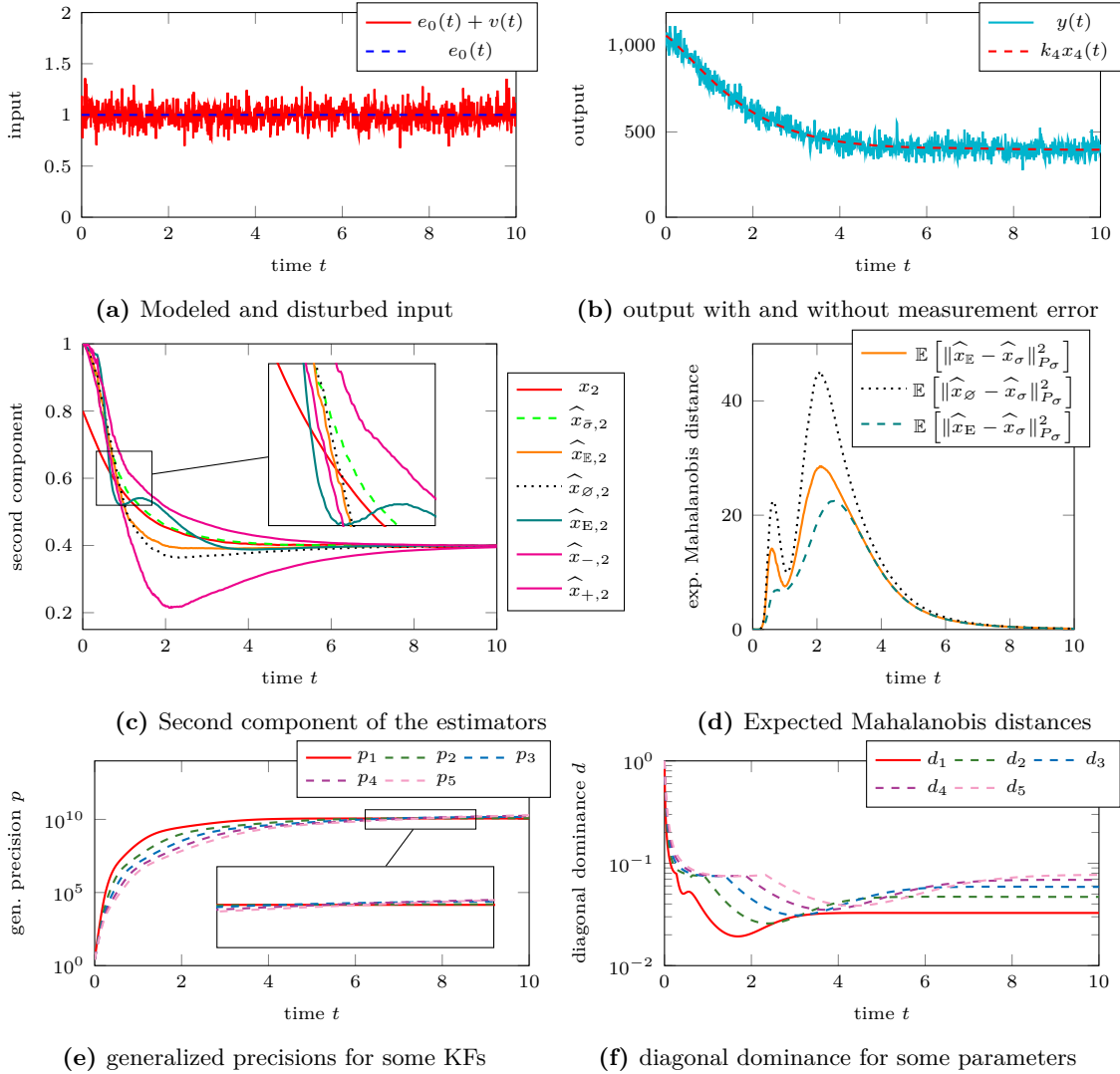
where $p_{ji}^{\sigma_k}(t)$ are the entries of $P_{\sigma_k}(t)$ and $\sigma_k = \begin{bmatrix} 10 & 0.5 & 10 \end{bmatrix}^\top + (k-1)\begin{bmatrix} 2.5 & 0.25 & 7.5 \end{bmatrix}^\top$. In particular we have $\sigma_1 = \bar{\sigma}$. Similarly, in Figure 2e we plot the associated generalized precisions

$$p_k(t) = \det(P_{\sigma_k}(t)), \quad k = 1, \dots, 5,$$

and in contrast to the first example we find no monotone dependence of the precision on the parameters. We conclude that in this example the interpretation of the results is less straight forward and the system displays a structure noticeably different from the behavior of the scalar case. The energy minimizer $\widehat{x}_{\mathrm{E}}$ is particularly interesting as its dependence on the precision matrices becomes apparent. Finally, we point out that, as predicted in Theorem 3.6, $\widehat{x}_{\mathrm{E}}$ offers the best approximation in terms of expected squared Mahalanobis distance, cf. Figure 2d.

6. Conclusion

In this work Kalman filter based concepts for the state estimation of uncertain linear systems are introduced, investigated, and numerically implemented. A particular focus

(a) Modeled and disturbed input

(b) output with and without measurement error

(c) Second component of the estimators

(d) Expected Mahalanobis distances

(e) generalized precisions for some KFs

(f) diagonal dominance for some parameters

Figure 2 – Connected amplidynes with uncertain inductances

lies on the minimizer of the expected energy which is characterized as the mean of the individual Kalman filters weighted with their precision matrices. It is theoretically proven and numerically illustrated that it minimizes the expected squared Mahalanobis distance to the Kalman filter arising from the realized, unknown parameter. We find that in particular settings the described estimator is slightly outperformed by less elaborate approaches which is not surprising since it is defined as the minimizer of an averaged quantity. This issue might be addressed by considering more risk averse concepts optimizing metrics other than the average of the parameter family. Another interesting strain of research is the case of continuous parameter distributions such as normally distributed uncertain parameters. One could also investigate the impact of uncertainties on nonlinear systems leading to (higher order) extended Kalman filters or the so called Mortensen observer.

Acknowledgement

References

[1] A. C. Aitken, *On least squares and linear combination of observations*, Proceedings of the Royal Society of Edinburgh, 55 (1936), p. 42–48.

[2] T. Breiten and K. Kunisch, *Neural network based nonlinear observers*, System & Control Letters, 148 (2021).

[3] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, Studies in Mathematics and Its Applications, North-Holland, Amsterdam New York Oxford, first ed., 1978.

[4] M. H. DeGroot, *Optimal statistical decisions*, McGraw-Hill, New York a. o., first ed., 1970.

[5] L. Dieci and T. Eirola, *Positive definiteness in the numerical solution of Riccati differential equations*, Numerische Mathematik, 67 (1994), p. 303–313.

[6] G. Evensen, *Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics*, J. Geophys. Res., 99 (1994), pp. 10143–10162.

[7] V. L. Girko, *Random matrices*, in Handbook of Algebra, M. Hazewinkel, ed., vol. 1 of Handbook of Algebra, North-Holland, 1996, pp. 27–78.

[8] P. A. Guth, K. Kunisch, and S. S. Rodrigues, *Tracking optimal feedback control under uncertain parameters*, Physica D: Nonlinear Phenomena, 467 (2024), p. 134245.

[9] R. E. Kalman, *A new approach to linear filtering and prediction problems*, Transactions of the ASME–Journal of Basic Engineering, 82 (1960), pp. 35–45.

[10] R. E. Kalman and R. S. Bucy, *New results in linear filtering and prediction theory*, Transactions of the ASME–Journal of Basic Engineering, 83 (1961), pp. 95–108.

[11] H. Kempka, C. Schneider, and J. Vybiral, *Path regularity of the Brownian motion and the Brownian sheet*, Constructive Approximation, 59 (2024), pp. 485–539.

[12] H. Kwakernaak and R. Sivan, *Linear optimal control systems*, Wiley-Interscience, a division of John Wiley & Sons, Inc., New York, Chichester, Brisbane, Toronto, Singapore, first ed., 1972.

[13] P. C. Mahalanobis, *Reprint of: Mahalanobis, P. C. (1936) "On the generalised distance in statistics"*, The Indian Journal of Statistics, 80 (2019), pp. S1–S7.

[14] R. Mehra, *On the identification of variances and adaptive Kalman filtering*, IEEE Transactions on Automatic Control, 15 (1970), pp. 175–184.

[15] R. E. Mortensen, *Maximum-likelihood recursive nonlinear filtering*, Journal of Optimization Theory and Applications, 2 (1968), pp. 386–394.

[16] B. Øksendal, *Stochastic differential equations*, Universitext, Springer-Verlag, Berlin, fifth ed., 1998.

[17] I. R. Petersen and D. C. McFarlane, *Optimal guaranteed cost control and filtering for uncertain linear systems*, IEEE Transactions on Automatic Control, 39 (1994), pp. 1971–1977.

[18] I. R. Petersen and A. V. Savkin, *Robust Kalman Filtering for Signals and Systems with Large Uncertainties*, Birkhäuser, Boston, first ed., 1999.

[19] J. Schröder, *Code for the paper "Deterministic Kalman filters for uncertain dynamical systems"*, 2025. doi: 10.5281/zenodo.15519854.

[20] L. Shi, K. H. Johansson, and R. M. Murray, *Kalman filtering with uncertain process and measurement noise covariances with application to state estimation in sensor networks*, in 2007 IEEE International Conference on Control Applications, 2007, pp. 1031–1036.

[21] E. D. Sontag, *Mathematical control theory: Deterministic finite dimensional systems*, Texts in Applied Mathematics 6, Springer, New York, second ed., 1998.

[22] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, The MIT Press, Cambridge, first ed., 1949.

[23] J. C. Willems, *Deterministic least squares filtering*, Journal of Econometrics, 118 (2004), pp. 341–373.

[24] J. Xiong, *An introduction to stochastic filtering theory*, Oxford University Press, Oxford, first ed., 2008.

⋆ Institute of Mathematics and Scientific Computing, University of Graz, A-8010 Graz, Austria

KARL KUNISCH[⋆,†] AND JESPER SCHRÖDER[†]

[†] Johann Radon Institute, Austrian Academy of Sciences, A-4040 Linz, Austria

Email address: karl.kunisch@uni-graz.at

Email address: jesper.schroeder@ricam.oeaw.ac.at