RISK AVERSE DETERMINISTIC KALMAN FILTERS FOR UNCERTAIN DYNAMICAL SYSTEMS

KARL KUNISCH*,† AND JESPER SCHRÖDER†

ABSTRACT. Taking a deterministic viewpoint this work investigates extensions of the Kalman-Bucy filter for state reconstruction to systems containing parametric uncertainty in the state operator. The emphasis lies on risk averse designs reducing the probability of large reconstruction errors. 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 examples allowing for a comparison of risk neutral and risk averse estimators.

1. INTRODUCTION

This work proposes estimators for state reconstruction of linear perturbed dynamical systems of the form

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

with unknown, deterministic disturbances $v \in L^2(0,T;\mathbb{R}^m)$ and $\eta \in \mathbb{R}^n$. The system matrix A_σ depends on a parametric uncertainties indicated by $\sigma \in \Sigma$. The objective is an online approximation of the system's state x based on incomplete measurements given as

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

that are affected by the output disturbance $\mu \in L^2(0,T;\mathbb{R}^r)$.

Utilizing flawed measurements to estimate the state of systems suffering from perturbances is of high interest in the applied sciences at least since Wiener's seminal work [28]. Kalman's and Bucy's work in [16] is tailored to the linear setting, for known parameter σ, and modeling the disturbances as Gaussian white noise, see also [15] for the time-discrete analogue. The so-called Kalman-Bucy filter is easy to implement, numerically efficient, and widely applied.

Since the parametric uncertainty introduces nonlinearities and multiplicative noise, it is not directly applicable to the class of systems considered here. In this work a risk averse extension is proposed based on Mortensen's strategy of recasting the Kalman filter as the energy minimal state [21]. Specifically, we study probability measures with respect to the parameter of the energy. The measures of interest are the entropic risk measure and the essential supremum for which the minimizers are characterized based on first order optimality conditions. The proposed measures utilize the energies for each system corresponding to a parameter $\sigma \in \Sigma$ separately, followed by a variational criterion which determines a universal state-estimator for the ensemble. An important part of our work consists in establishing estimates on the difference between the proposed state estimator

and that estimator which would be optimal if we knew the exact parameter $\sigma \in \Sigma$. For both measures we analyze their temporal regularity. The entropic risk measure depends on a risk aversion parameter $\theta > 0$. We verify that for $\theta \to 0$ the resulting estimator converges to the risk neutral estimator and for $\theta \to \infty$ we obtain convergence to the minimizer of the essential supremum. In the numerical part of our work the proposed state estimators are compared to results emerging from the risk neutral approach associated with the expectation which we investigated in earlier work [17].

Let us further comment on the two measures which we investigate. The entropic risk measure was introduced in the context of financial mathematics to optimize portfolios with a flexible level of risk aversion [11]. More recently it was adapted for optimal control under uncertainty [12,13] leading to risk averse control strategies reducing the likelihood of large costs. This functional also arises in many other applied sciences including machine learning and neural networks [4], or statistical mechanics, [24], for example. It is well known that for infinitely large risk aversion the entropic risk measure corresponds to the essential supremum which in the framework of this work turns into a maximum. The associated estimator is hence given by the solution of a min-max problem which on a technical level is closely related to the minimum ball problem arising in operations research. In the 2-dimensional case this problem has a long history, see for instance [14] and the references provided there. The $\mathbb{R}^n$ case was investigated for instance in [5,7]. These papers also contain many references on the algorithmic developments for this class of problems. In this paper, to analyze the min-max formulation as well as the entropic risk formulation we follow a different analytical route by treating these problems by convex analysis techniques. In this way we can use results from [20] to obtain existence and uniqueness of solutions and from [22] to derive optimality conditions, in an elegant manner. The latter are of vital importance for the error analysis which we provide.

2. The Kalman filter

We begin with a brief presentation of two different derivations of the Kalman-Bucy filter for time-invariant systems. While our proposed estimators rely only on the deterministic formulation a recap of the stochastic formulation enriches the interpretation of our approach. We include these well known concepts for the purpose of a self-contained work.

2.1. **The stochastic formulation.** We illustrate the stochastic filter based on the original work [16] and refer to [23,30] for a more rigorous treatment in terms of SDEs. Consider the disturbed system with system state X_t and measured output Y_t modeled by

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

with system matrix $A \in \mathbb{R}^{n,n}$, noise input matrix $B \in \mathbb{R}^{n,m}$, and output matrix $C \in \mathbb{R}^{r,n}$. The dynamics and the measurement are disturbed by the multivariate Gaussian random variables $V_t \in \mathbb{R}^m$ and $W_t \in \mathbb{R}^r$, respectively. The so-called white noise terms have zero mean and covariances given by

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

where $R \in \mathbb{R}^{m,m}$ and $Q \in \mathbb{R}^{r,r}$ are given symmetric, positive definite matrices. The distribution of the state at time t is Gaussian and denoted by X_t, where the initial

distribution X_0 is characterized via its mean $x_0 \in \mathbb{R}^n$ and covariance $\Gamma \in \mathbb{R}^{n,n}$, symmetric, positive definite. The two noise terms and the initial distribution are assumed to be mutually independent and finally Y_t denotes the distribution of the output at time t.

According to [16] the goal is to optimally estimate the realized state $x(t)$ of the system given realized, measured outputs $y(s)$, $0 \leq s \leq t$. In their work Kalman and Bucy show that the optimal estimate $\widehat{x}(t)$ is given via

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

where Π is the unique solution to the differential Riccati equation

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

In the stochastic interpretation, given a measured output the state of the system at time t is estimated via the Gaussian distribution $\mathcal{N}(\widehat{x}(t), \Pi(t))$ where the mean and covariance are characterized via (2.2) and (2.3), respectively.

2.2. The deterministic formulation.
In this subsection we consider a deterministic model of a disturbed system brought forward by Mortensen in [21], see also [29]. It reads

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

where the system matrices A, B, and C are as in (2.1). The disturbances in dynamics, initial value, and output are modeled by $v \in L^2(0, T; \mathbb{R}^m)$, $\eta \in \mathbb{R}^n$, and $\mu \in L^2(0, T; \mathbb{R}^r)$ and are assumed to be deterministic and unknown. To increase readability for the remainder of this work we denote $\mathcal{L}^p_{t_1, t_2} = L^p(t_1, t_2; \mathbb{R}^d)$ and $\mathcal{H}^{t_2}_{t_1} = H^1(t_1, t_2; \mathbb{R}^n)$ for $0 \leq t_1 < t_2 \leq T$, $d \in \mathbb{N}$, and $0 < p \leq \infty$ and

The strategy for estimating $x(t)$, at $t \in (0, T]$ consists in identifying the energy minimal disturbances $\eta^*[t]$, $v^*[t]$, and $\mu^*[t]$ that fit the measured output $y(s)$, $0 \leq s \leq t$, i.e., in solving

$$\min_{\eta \in \mathbb{R}^n, v \in \mathcal{L}^2_{0,t}, \mu \in \mathcal{L}^2_{0,t}} \|\eta\|^2_{\Gamma^{-1}} + \int_0^t \|v(s)\|^2_{R^{-1}} + \|\mu(s)\|^2_{Q^{-1}} \, ds, \tag{2.5}$$
$$\text{subject to (2.4) on } (0, t).$$

This minimizer is used to solve (2.4) for the associated trajectory $x^*[t]$ and the estimator is defined as $\widehat{x}(t) = x^*[t](t)$. Here and in the following for a positive definite matrix M and vector x of fitting size we denote $\|x\|^2_M = x^\top M x$. The matrices Γ, R, and Q correspond to the noise covariances from the stochastic formulation. Using their inverses as weighting matrices is a strategy also used in generalized least squares [1]. In [29] Willems demonstrates that this estimator is given precisely by the Kalman filter equations known from the stochastic formulation. Indeed $\widehat{x}$ solves

$$\dot{\widehat{x}}(t) = A\widehat{x}(t) + \Pi(t)C^\top Q^{-1}(y(t) - C\widehat{x}(t)) \qquad t \in (0, T),$$
$$\widehat{x}(0) = x_0, \tag{2.6}$$

where Π is given as the unique solution of

$$\dot{\Pi}(t) = A\Pi(t) + \Pi(t)A^\top - \Pi(t)C^\top Q^{-1}C\Pi(t) + BRB^\top \qquad t \in (0, T),$$
$$\Pi(0) = \Gamma. \tag{2.7}$$

Our analysis follows the slightly different formulation of [21]. By adding a final state ξ and inserting the identities for $x(0)$ and y Mortensen recasts (2.5) into a problem of optimal control. It reads

$$\min_{x \in \mathcal{H}_0^t, v \in \mathcal{L}_{0,t}^2} J(x, v; t, \xi) = \|x(0) - x_0\|_{\Gamma^{-1}}^2 + \int_0^t \|v(s)\|_{R^{-1}}^2 + \|y - Cx(s)\|_{Q^{-1}}^2 \, ds, \quad (2.8)$$

$$\text{subject to} \qquad \dot{x}(s) \ = Ax(s) + Bv(s) \qquad s \in (0, t), \tag{2.9}$$
$$x(t) \ = \xi \tag{2.10}$$

with associated value function

$$\mathcal{V}(t, \xi) = \inf_{x \in \mathcal{H}_0^t, v \in \mathcal{L}_{0,t}^2} J(x, v; t, \xi) \quad \text{subject to } (2.9) - (2.10) \quad t \in (0, T],$$
$$\mathcal{V}(0, \xi) = \|\xi - x_0\|_{\Gamma^{-1}}^2. \tag{2.11}$$

It represents the minimal amount of energy required to ensure that at time t the system is in the state ξ. In [17, Lem. 2.5] it is shown that in the present case of linear dynamics the value function $\mathcal{V}$ is quadratic in space and given as

$$\mathcal{V}(t, \xi) = (\xi - \widehat{x}(t))^\top \Pi^{-1}(t)(\xi - \widehat{x}(t)) + \int_0^t \|y(s) - C\widehat{x}(s)\|_{Q^{-1}}^2 \, ds, \tag{2.12}$$

where $\widehat{x}$ and Π are given via the Kalman filter equations (2.6) and (2.7). Hence the estimator $\widehat{x}$ constructed via (2.5) can equivalently be characterized as

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

We point out that the control formulation at hand is merely a tool to reconstruct energy minimal disturbances that most likely caused the measured output.

Even though the following derivations and arguments are based on the deterministic perspective, we still refer to the solution Π of (2.7) as the (error) covariance associated with the Kalman filter allowing for an interpretation of our results. We denote the inverse of the covariance as $P := \Pi^{-1}$ and following [8, Sec. 5.4] we refer to it as the precision matrix. By multiplying (2.7) with $P(t)$ from the left and from the right we find that the precision matrix is characterized via the differential Riccati equation

$$\dot{P}(t) = -A^\top P(t) - P(t)A - P(t)BRB^\top P(t) + C^\top Q^{-1}C, \qquad t \in (0, T),$$
$$P(0) = \Gamma^{-1}. \tag{2.13}$$

3. KALMAN FILTERING FOR UNCERTAIN SYSTEMS

The Kalman filter described in the previous section assumes exact knowledge of both the system dynamics represented by A and the nature of the occurring noise represented by the covariance matrices Γ, R, and Q. In practical applications, however, this information may not be available in exact form and the matrices may depend on parameters, [19, 27].

3.1. **Modeling uncertain parameters.** Following the description in [17] the uncertainties of the system matrix A are modeled 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 $\mathcal{P}_A(\Sigma_A)$ denotes the power set of Σ_A and for $S \subset \Sigma_A$ 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

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

and a_σ denotes the system matrix that realizes with the parameter $\sigma \in \Sigma_A$. The uncertain weighting matrices γ_σ, r_σ, and q_σ are defined 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 as that for 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. Further 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 are defined as random variables on $(\Sigma, \mathcal{P}(\Sigma), P)$, where for $S \subset \Sigma$ we set $P(S) = \frac{1}{N}|S|$. Denote $\sigma = (\sigma_A, \sigma_\Gamma, \sigma_R, \sigma_Q) \in \Sigma$ and define

$$\sigma \in \Sigma \mapsto A_\sigma = a_{\sigma_A} \in \mathbb{R}^{n,n}, \quad \sigma \in \Sigma \mapsto \Gamma_\sigma = \gamma_{\sigma_\Gamma} \in \mathbb{R}^{n,n},$$

$$\sigma \in \Sigma \mapsto R_\sigma = r_{\sigma_R} \in \mathbb{R}^{m,m}, \quad \sigma \in \Sigma \mapsto Q_\sigma = q_{\sigma_Q} \in \mathbb{R}^{r,r}.$$

Note that by construction they are 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., $A_{\sigma_k} = A_k$. Also, for a given parameter σ the associated four-tuple consisting of system matrix and covariance matrices is denoted by $\mathcal{S}_\sigma = (A_\sigma, \Gamma_\sigma, R_\sigma, Q_\sigma) \in \mathbb{R}^{n,n} \times \mathbb{R}^{n,n} \times \mathbb{R}^{m,m} \times \mathbb{R}^{r,r}$. To quantify the distance between such tuples for $p \in \mathbb{N}$ we introduce the norm $\|\mathcal{S}_\sigma\|_p^p = \|A_\sigma\|_{\mathbb{R}^{n,n}}^p + \|\Gamma_\sigma\|_{\mathbb{R}^{n,n}}^p + \|R_\sigma\|_{\mathbb{R}^{m,m}}^p + \|Q_\sigma\|_{\mathbb{R}^{r,r}}^p$.

For any $\sigma_k \in \Sigma$ the corresponding Kalman filter is denoted by $\widehat{x}_k$ and the associated covariance and precision matrices are denoted by Π_k and $P_k = \Pi_k^{-1}$, respectively. They are characterized via

$$
\begin{aligned}
\dot{\widehat{x}}_k(t) &= A_k\widehat{x}_k(t) + \Pi_k(t)C^\top Q_k^{-1}\left(y(t) - C\widehat{x}_k(t)\right) \qquad t \in (0, T),\\
\widehat{x}_k(0) &= x_0,
\end{aligned}
\tag{3.1}
$$

and

$$
\begin{aligned}
\dot{\Pi}_k(t) &= A_k\Pi_k(t) + \Pi_k(t)A_k^\top - \Pi_k(t)C^\top Q_k^{-1}C\Pi_k(t) + BR_kB^\top \qquad t \in (0, T),\\
\Pi_k(0) &= \Gamma_k.
\end{aligned}
\tag{3.2}
$$

We note that the solution Π_k is infinitely often continuously differentiable. Since the same holds for the mapping $M \mapsto M^{-1}$ defined on the set of symmetric positive definite matrices, we obtain that for all k we have

$$P_k \in C^\infty([0, T]; \mathbb{R}^{n,n}).\tag{3.3}$$

To make our arguments more concise, we denote for $k \in \{1, \ldots, N\}$ and $t \in [0, T]$

$$r_k(t) = \int_0^t \|y(s) - C\widehat{x}_k(s)\|_{Q_k^{-1}}^2 \, ds.$$

Analogous to (2.12) we obtain

$$\mathcal{V}_k(t, x) = \|x - \widehat{x}_k(t)\|_{P_k(t)}^2 + r_k(t). \tag{3.4}$$

Next the estimators which will be investigated are described.

3.2. **Minimizing the expected energy.** We begin by presenting a design that was proposed and studied in an earlier work, cf. [17]. For $t \in [0, T]$ consider the minimization problem

$$\min_{x \in \mathbb{R}^n} \ \mathbb{E}\left[\mathcal{V}_\sigma(t, x)\right] = \min_{x \in \mathbb{R}^n} \ \frac{1}{N} \sum_{k=1}^N \mathcal{V}_k(t, x). \tag{3.5}$$

This minimization problem defines the first estimator. In [17, Prop. 3.4] we showed existence and uniqueness of the solution of (3.5) and throughout this work we denote the estimator defined as the pointwise solution of (3.5) as $\widehat{x}_0(t)$. We further cite the following characterization.

Proposition 3.1. *The state estimator $\widehat{x}_0$ is well-defined and given in terms of the Kalman filter trajectories and precision matrices associated with the individual Kalman filters. For $t \in [0, T]$ it holds*

$$\widehat{x}_0(t) = \mathcal{P}^{-1}(t) \sum_{k=1}^N P_k(t)\, \widehat{x}_k(t), \tag{3.6}$$

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

This representation immediately implies a result on the regularity of the estimator.

Corollary 3.2. *The estimator $\widehat{x}_0$ is square integrable and admits a square integrable weak derivative, i.e., $\widehat{x}_0 \in \mathcal{H}_0^T$. Its weak derivative is given by*

$$\dot{\widehat{x}}_0(t) = -\mathcal{P}^{-1}(t)\, \dot{\mathcal{P}}(t)\, \widehat{x}_0(t) + \mathcal{P}^{-1}(t) \sum_{k=1}^N P_k(t)\dot{\widehat{x}}(t) + \dot{P}_k(t)\widehat{x}_k(t).$$

Proof. The regularity of P_k presented in (3.3) implies $\mathcal{P}^{-1} \in C^\infty([0, T]; \mathbb{R}^{n,n})$. For the solutions of (3.1) we have $\widehat{x}_k \in \mathcal{H}_0^T$. The claim for $\widehat{x}_0$ follows from the representation given in (3.6). The formula for the weak derivative is obtained by an application of the chain rule and (3.6). $\qquad \square$

We proceed by introducing more risk averse designs.

3.3. **Minimizing the worst case energy.** For the construction of the second estimator consider the problem of minimizing the worst case energy

$$\min_{x \in \mathbb{R}^n} \ \operatorname*{ess\,sup}_{\sigma \in \Sigma} \mathcal{V}_\sigma(t, x) = \min_{x \in \mathbb{R}^n} \ \max_{k=1,\ldots,N} \mathcal{V}_k(t, x). \tag{3.7}$$

We begin by ensuring existence of a unique solution at any time.

Lemma 3.3. *For every $t \in [0, T]$ the problem* (3.7) *admits a unique solution.*

Proof. The value functions $\mathcal{V}_k$ defined in (3.4) are strictly convex and coercive in x. Since the maximum in (3.7) is taken over finitely many indices k, the expression $\max_k \mathcal{V}_k(t, \cdot)$ inherits the strict convexity and coercivity properties. As a consequence, the existence of a unique minimizer then follows. □

This result enables the definition of the second estimator.

Definition 3.4. *The state estimator minimizing the worst case energy at time $t \in [0, T]$ is denoted as $\widehat{x}_\infty(t)$ and given by the unique solution of the minimization problem (3.7).*

In the following we characterize the minimizer via a representation in terms of the family members $\widehat{x}_k$. This is done for every time point individually. Hence, from here on we fix a time $t \in [0, T]$. To facilitate the discussion using tools from convex analysis we introduce the following notation. We define

$$f[t] \colon \mathbb{R}^n \to \mathbb{R}, \quad f[t](x) = \max_{k=1,\ldots,N} \mathcal{V}_k(t, x).$$

Now solving (3.7) for the fixed time t is equivalent to solving

$$\min_{x \in \mathbb{R}^n} f[t](x). \tag{3.8}$$

To formulate the first order optimality condition we introduce the set of active indices.

Definition 3.5. *For $x \in \mathbb{R}$ we define*

$$I[t](x) = \{k \in \{1, \ldots, N\} \colon \mathcal{V}_k(t, x) = f[t](x)\},$$

i.e., $I[t](x)$ is the set of all indices for which the maximum is attained.

With these definitions at hand we are able to characterize the minimizer $\widehat{x}_\infty(t)$.

Theorem 3.6. *Let $t \in [0, T]$ be arbitrary. There exists a subset $S(t) \subset I[t](\widehat{x}_\infty(t))$ with cardinality $|S(t)| = p \leq n + 1$ and coefficients $\alpha(t) \in [0, 1]^p$ such that it holds*

$$(i) \qquad \sum_{k \in S(t)} \alpha_k(t) = 1,$$

$$(ii) \qquad \{P_k(t)(\widehat{x}_\infty(t) - \widehat{x}_k(t))\}_{k \in S(t)} \text{ are affine independent,}$$

$$(iii) \qquad \widehat{x}_\infty(t) = \left(\sum_{k \in S(t)} \alpha_k(t) P_k(t) \right)^{-1} \sum_{k \in S} \alpha_k(t) P_k(t) \widehat{x}_k(t).$$

In other words the estimator can be characterized in terms of a subset of $n + 1$ or fewer family members in which the maximum is attained.

Proof. Fix an arbitrary $t \in [0, T]$. Then $\widehat{x}_\infty(t)$ is the unique solution of (3.8). As discussed in the proof of Lemma 3.3 $f[t]$ is strictly convex, hence $\widehat{x}_\infty(t)$ satisfies the optimality condition [9, Def. 5.1. and (5.5)] given by

$$0 \in \partial f[t](\widehat{x}_\infty(t)),$$

where $\partial f[t](\bar{x})$ denotes the subdifferential of $f[t]$ at $\bar{x}$. The subdifferential of a maximum is well known, cf. [25, Prop. 4.5.2 and Rem. 4.5.3]. Denoting the convex hull of a set X by $\mathrm{conv}(X)$ we indeed have

$$\partial f[t](\bar{x}) = \mathrm{conv}\{P_k(t)(\widehat{x}_\infty(t) - \widehat{x}_k(t)) \colon k \in I[t](\widehat{x}_\infty(t))\}$$
$$= \mathrm{conv}\{P_k(t)(\widehat{x}_\infty(t) - \widehat{x}_k(t)) \colon k \in S(t)\},$$

where the second equality is justified by Caratheodory's theorem with a set $S(t) \subset I[t](\widehat{x}_\infty(t))$ such that (ii) holds. The affine independence in particular implies $|S(t)| \le n+1$. The optimality condition now ensures existence of $\alpha(t) \in [0,1]^p$ such that (i) holds and

$$0 = \sum_{k \in S(t)} \alpha_k(t) P_k(t)(\widehat{x}_\infty(t) - \widehat{x}_k(t)).$$

Rearranging terms yields (iii) and the proof is complete. $\qquad\square$

Remark 3.7. *This representation does not yield any time regularity of $\widehat{x}_\infty$. Since there is no straight forward way of characterizing the time dependence of the coefficients α_k, at this point we can not even deduce measurability of the estimator. A result regarding this is obtained via a convergence argument in Corollary 4.2 below.*

The two presented estimators $\widehat{x}_0$ and $\widehat{x}_\infty$ are the two extremes on the spectrum of risk aversion. While the minimizer of the expected value $\widehat{x}_0$ does not take risk into account at all, the worst-case minimizer $\widehat{x}_\infty$ is by construction the most risk averse. In the following we introduce a concept filling the gap in between these two extremes.

3.4. **Minimizing the entropic risk of the energy.** In this subsection we introduce another estimator as the minimizer of the so-called *entropic risk measure*. To the best of the authors' knowledge it was introduced in the context of financial mathematics, cf. [11]. We commence with a brief presentation of the concept and its properties for the setting of this work. For a more general definition and detailed derivation we refer to [11]. This functional is also important in the context of machine learning, cf. for example [4].

Definition 3.8. *Let $X \colon \Sigma \to \mathbb{R}$ be a random variable on the probability space $(\Sigma, \mathcal{P}(\Sigma), P)$ defined in Subsection 3.1. The entropic risk measure of X with risk aversion $\theta \in (0, \infty)$ is defined as*

$$\rho_\theta(X) = \frac{1}{\theta} \ln\left(\mathbb{E}\left[e^{\theta X}\right]\right) = \frac{1}{\theta} \ln\left(\frac{1}{N} \sum_{k=1}^{N} e^{\theta X(\sigma_k)}\right).$$

We proceed by presenting some well known properties of the entropic risk measure. While we do not claim any novelty, for the sake of completeness we give a proof tailored to the setting of this work.

Lemma 3.9. *For all random variables X on the probability space $(\Sigma, \mathcal{P}(\Sigma), P)$ it holds*

$$\rho_\theta(X) \to \operatorname*{ess\,sup}_{\sigma \in \Sigma} X(\sigma) = \max_{k=1,\dots,N} X(\sigma_k) \qquad \text{for } \theta \to \infty.$$

Proof. Let X be such a fixed random variable and let $k^* \in \{1, \dots, N\}$ be such that $X(\sigma_{k^*}) = \max_{k=1,\dots,N} X(\sigma_k) = \operatorname{ess\,sup}_{\sigma \in \Sigma} X(\sigma)$. Since the exponential is positive, for all $\theta \in (0, \infty)$ we have

$$\frac{1}{N} \sum_{k=1}^{N} e^{\theta X(\sigma_k)} \ge \frac{1}{N} e^{\theta X(\sigma_{k^*})}.$$

Applying the logarithm on both sides and dividing by θ yields

$$\rho_\theta(X) \ge X(\sigma_{k^*}) + \frac{\ln(\frac{1}{N})}{\theta}.$$

We hence find that for all $\theta \in (0, \infty)$ it holds

$$X(\sigma_{k^*}) + \frac{\ln(\frac{1}{N})}{\theta} \le \rho_\theta(X) \le X(\sigma_{k^*}),$$

where the second estimate is justified by the monotonicity of the logarithm and the exponential. The assertion follows by considering the limit for $\theta \to \infty$. $\qquad\square$

We now turn to the construction of the risk averse estimator. For its definition we consider the following minimization problem. For a fixed $t \in [0, T]$ consider

$$\min_{x \in \mathbb{R}^n} \rho_\theta \left(\mathcal{V}_\sigma(t, x) \right) = \min_{x \in \mathbb{R}^n} \frac{1}{\theta} \ln \left(\frac{1}{N} \sum_{k=1}^{N} e^{\theta(\|x - \widehat{x}_k(t)\|_{P_k(t)}^2 + r_k(t))} \right). \tag{3.9}$$

We proceed to formally define the third estimator as the minimizer of the entropic risk applied to the energy.

Definition 3.10. *Let $y \in \mathcal{L}_{0,T}^2$ and $t \in [0, T]$. The state estimator minimizing the entropic risk of the energy with risk aversion parameter $\theta \in (0, \infty)$ is denoted by $\widehat{x}_\theta(t)$ and defined as the solution of* (3.9), *assuming it exists and is unique.*

The following lemma shows that the estimator is well defined and characterizes it via an implicit equation.

Lemma 3.11. *For every $t \in [0, T]$, $y \in \mathcal{L}_{0,T}^2$, and $\theta \in (0, \infty)$ there exists exactly one solution to the minimization problem* (3.9). *The solution, denoted by $\widehat{x}_\theta(t)$, satisfies*

$$\widehat{x}_\theta(t) = \left(\sum_{k=1}^{N} e^{\theta(\|\widehat{x}_\theta(t) - \widehat{x}_k(t)\|_{P_k(t)}^2 + r_k(t))} P_k(t) \right)^{-1} \sum_{k=1}^{N} e^{\theta(\|\widehat{x}_\theta(t) - \widehat{x}_k(t)\|_{P_k(t)}^2 + r_k(t))} P_k(t) \widehat{x}_k(t)$$

(3.10)

and is the only element of $\mathbb{R}^n$ satisfying this relation.

Proof. Let $t \in [0, T]$ be fixed. For the purpose of this proof we define $F \colon \mathbb{R}^n \to \mathbb{R}$ via

$$F(x) := \frac{1}{\theta} \ln \left(\frac{1}{N} \sum_{k=1}^{N} e^{\theta(\|x - \widehat{x}_k(t)\|_{P_k(t)}^2 + r_k(t))} \right).$$

In the following we show that F is strictly convex. To that end note that it is smooth and the chain rule yields its gradient as

$$\nabla F(x) = 2 \sum_{k=1}^{N} c_k[\theta, x](t) \ P_k(t) \widehat{x}_k(t). \tag{3.11}$$

To characterize the Hessian we denote the outer product of a vector $x \in \mathbb{R}^n$ with itself via $[x]_\otimes^2 = x \otimes x = xx^\top \in \mathbb{R}^{n,n}$ and define

$$0 \le c_k[\theta, x](t) = \frac{e^{\theta(\|x - \widehat{x}_k(t)\|_{P_k(t)}^2 + r_k(t))}}{\sum_{j=1}^{N} e^{\theta(\|x - \widehat{x}_j(t)\|_{P_j(t)}^2 + r_j(t))}} \le 1. \tag{3.12}$$

Note that for all x and t it holds $\sum_{k=1}^{N} c_k[\theta, x](t) = 1$. The chain rule then yields the Hessian as

$$\nabla^2 F(x) = 2 \sum_{k=1}^{N} c_k[\theta, x](t) P_k(t) + 4 \sum_{k=1}^{N} c_k[\theta, x](t) \left[P_k(t)(x - \widehat{x}_k(t)) \right]_\otimes^2$$

$$- 4\theta \left[\sum_{k=1}^{N} c_k[\theta, x](t) P_k(t) \widehat{x}_k(t) \right]_\otimes^2.$$

Now for any $z \in \mathbb{R}^n \backslash \{0\}$ we find

$$z^\top \nabla^2 F(x) z = 2 \sum_{k=1}^N c_k[\theta, x](t) \ z^\top P_k(t) z + 4\theta \sum_{k=1}^N c_k[\theta, x](t) \| z^\top P_k(t)(x - \widehat{x}_k(t)) \|^2$$

$$- 4\theta \| \sum_{k=1}^N c_k[\theta, x](t) z^\top P_k(t) \widehat{x}_k(t) \|^2.$$

Due to the positive definiteness of P_k the first summand is greater than zero. Since the c_k are non negative and sum up to one, the convexity of the squared norm ensures that the sum of the last two terms is greater or equal than zero. Hence, the Hessian $\nabla^2 F(x)$ is positive definite implying strict convexity of F. One can quickly verify that F also is coercive ensuring that it admits exactly one minimizer. Further, any element $\bar{x} \in \mathbb{R}^n$ minimizes F if and only if $\nabla F(\bar{x}) = 0$. Setting the gradient given in (3.11) to zero and performing standard calculus yields the claimed equation for $\widehat{x}_\infty(t)$. $\qquad \square$

Note that an alternative to the characterization via (3.10) is given by

$$\widehat{x}_\theta(t) = \left(\sum_{k=1}^N c_k[\theta, \widehat{x}_\theta(t)](t) P_k(t) \right)^{-1} \sum_{k=1}^N c_k[\theta, \widehat{x}_\theta(t)](t) P_k(t) \widehat{x}_k(t) \qquad (3.13)$$

with the coefficients given as in (3.12) satisfying $\sum_{k=1}^N c_k = 1$ and $c_k \in [0, 1]$.

We conclude the section by stating a result on the time regularity of $\widehat{x}_\theta$ depending on the regularity of the output y. The proof is presented in Appendix A.2.

Theorem 3.12. *Let $\theta \in (0, \infty)$.*

(i) Assuming $y \in \mathcal{L}_{0,T}^{2p}$ with an integer $1 \leq p < \infty$ it holds $\widehat{x}_\theta \in W^{1,p}(0, T; \mathbb{R}^n)$.

(ii) For $y \in C([0, T]; \mathbb{R}^r)$ it holds $\widehat{x}_\theta \in C^1([0, T]; \mathbb{R}^n)$.

In both settings the (weak) derivative is given via the formula

$$\dot{\widehat{x}}_\theta(t) = -M(t, \widehat{x}_\theta(t))^{-1} V(t, \widehat{x}_\theta(t)), \qquad (3.14)$$

where

$$M(x, t) = \sum_{k=1}^N e^{\theta \mathcal{V}_k(t,x)} \left(P_k(t) + 2\theta P_k(t)(x - \widehat{x}_k(t)) \otimes (x - \widehat{x}_k(t)) P_k(t) \right), \qquad (3.15)$$

and

$$V(t, x) = - \sum_{k=1}^N e^{\theta \mathcal{V}_k(t,x)} \left(P_k(t) \dot{\widehat{x}}_k(t) - \dot{P}_k(t)(x - \widehat{x}_k(t)) \right)$$

$$+ \theta \sum_{k=1}^N e^{\theta \mathcal{V}_k(t,x)} \left(\| y(t) - C \widehat{x}_k(t) \|_{Q_k^{-1}}^2 - 2\langle x - \widehat{x}_k(t), P_k(t) \dot{\widehat{x}}_k(t) \rangle \right) P_k(t)(x - \widehat{x}_k(t)). \qquad (3.16)$$

4. Continuity of the state estimator in the risk aversion parameter

In this section we analyze the behavior of the limit cases of the risk aversion parameter. We find that indeed, $\widehat{x}_\theta$ fills the gap between $\widehat{x}_0$ and $\widehat{x}_\infty$ in the sense that the estimators are continuous with respect to the risk aversion, in particular for the extreme cases.

Proposition 4.1. *Let $t \in [0, T]$ be fixed. It holds*

$$\widehat{x}_\theta(t) \to \widehat{x}_0(t), \quad \theta \to 0,$$
$$\widehat{x}_\theta(t) \to \widehat{x}_\infty(t), \quad \theta \to \infty,$$

where $\widehat{x}_0$ and $\widehat{x}_\infty$ are the estimators introduced in Subsection 3.2 and Definition 3.4, respectively. Moreover $\theta \mapsto \widehat{x}_\theta(t)$ is continuous in $(0, \infty)$.

Proof. To show the first limit we note that due to Lemma A.1 and Lemma A.3 in the appendix we have that $\|\widehat{x}_\theta(t) - \widehat{x}_k(t)\|^2_{P_k(t)}$ is bounded with respect to θ. Since additionally $r_k[y](t)$ is independent of θ, we find

$$\theta\left(\|\widehat{x}_\theta(t) - \widehat{x}_k(t)\|^2_{P_k(t)} + r_k[y](t)\right) \to 0 \quad \text{for } \theta \to 0.$$

Since all involved operations are continuous, passing to the limit in (3.10) reveals

$$\widehat{x}_\theta(t) \to \left(\sum_{k=1}^N P_k(t)\right)^{-1} \sum_{k=1}^N P_k(t)\, \widehat{x}_k(t) \quad \text{for } \theta \to 0$$

and the first claim follows with Proposition 3.1.

To show the second convergence we utilize the characterization of $\widehat{x}_\theta(t)$ given in (3.13) which reads as

$$\widehat{x}_\theta(t) = \left(\sum_{k=1}^N c_k[\theta, \widehat{x}_\theta(t)](t)P_k(t)\right)^{-1} \sum_{k=1}^N c_k[\theta, \widehat{x}_\theta(t)](t)P_k(t)\widehat{x}_k(t), \tag{4.1}$$

where the coefficients $0 \le c_k[\theta, x](t) \le 1$ are defined as in (3.12). Now let θ_l be an arbitrary sequence such that $\theta_l \to \infty$ for $l \to \infty$. Since the $c_k[\theta_l, \widehat{x}_{\theta_l}(t)](t)$ are from a compact interval, any subsequence admits yet another subsequence θ_g such that for any k we have that $c_k[\theta_g, \widehat{x}_{\theta_g}(t)](t)$ converges for $g \to \infty$. We denote the corresponding limits by $\beta_k(t)$. Since by construction for every $g \in \mathbb{N}$ and $t \in [0, T]$ the elements $\{c_k[\theta_g, \widehat{x}_{\theta_g}(t)](t)\}_{k=1}^N$ are non negative and sum up to one, the same holds for $\{\beta_k(t)\}_{k=1}^N$. With (4.1) it follows

$$\lim_{g\to\infty} \widehat{x}_{\theta_g}(t) = \left(\sum_{k=1}^N \beta_k(t)P_k(t)\right)^{-1} \sum_{k=1}^N \beta_k(t)P_k(t)\widehat{x}_k(t) =: \bar{x}.$$

Note that at this point $\bar{x}$ may depend on the choice of the subsequence θ_g. To remedy this, we now show that in fact $\bar{x} = \widehat{x}_\infty(t)$. To that end denote

$$F[\theta](x) = \frac{1}{\theta} \ln\left(\frac{1}{N}\sum_{k=1}^N e^{\theta(\|x-\widehat{x}_k(t)\|^2_{P_k(t)} + r_k(t))}\right) \quad \text{and} \quad F_\infty(x) = \max_k \mathcal{V}_k(t, x)$$

and consider

$$\left\|F[\theta_g](\widehat{x}_{\theta_g}(t)) - F_\infty(\bar{x})\right\| \le \left\|F[\theta_g](\widehat{x}_{\theta_g}(t)) - F[\theta_g](\bar{x})\right\| + \left\|F[\theta_g](\bar{x}) - F_\infty(\bar{x})\right\|. \tag{4.2}$$

Utilizing Lemma 3.9 we find that the second summand on the right hand side converges to zero. To show convergence of the first summand on the right hand side we first estimate the gradient $\|\nabla F[\theta](x)\|$ for some $x \in \mathbb{R}^n$ and $\theta \in (0, \infty)$. We have

$$\|\nabla F[\theta](x)\| = 2\left\|\sum_{k=1}^N c_k[\theta, x](t)P_k(t)\widehat{x}_k(t)\right\|,$$

where the coefficients $c_k[\theta, x]$ are defined as above and hence bounded by one. Due to the continuity of P_k and $\widehat{x}_k$ there exists a constant $L > 0$ independent of t, θ, x, and k such that
$$\|\nabla F[\theta](x)\| \leq L.$$
We now apply Taylor's theorem [31, Thm. 4.C] and obtain
$$\left\|F[\theta_g](\widehat{x}_{\theta_g}(t)) - F[\theta_g](\bar{x})\right\| \leq \int_0^1 \|\nabla F[\theta_g](\tau \widehat{x}_{\theta_g}(t) + (1-\tau)\bar{x})\| \, \mathrm{d}\tau \, \|\widehat{x}_{\theta_g}(t) - \bar{x}\|$$
$$\leq L\|\widehat{x}_{\theta_g}(t) - \bar{x}\|,$$
where the convergence of $\widehat{x}_{\theta_g}(t)$ implies convergence to zero of the right hand side. We have therefore shown that the right hand side of (4.2) converges to zero for $g \to \infty$, implying
$$\lim_{g \to \infty} F[\theta_g](\widehat{x}_{\theta_g}(t)) = F_\infty(\bar{x}).$$
As by construction $\widehat{x}_{\theta_g}$ minimizes $F[\theta_g]$ we have
$$F[\theta_g](\widehat{x}_{\theta_g}) \leq F[\theta_g](\widehat{x}_\infty(t)).$$
Applying again Lemma 3.9 we pass to the limit on both sides and obtain
$$F_\infty(\bar{x}) \leq F_\infty(\widehat{x}_\infty(t)).$$
Since $\widehat{x}_\infty(t)$ is the unique minimizer of F_∞, it follows $\bar{x} = \widehat{x}_\infty(t)$. We have now shown that any subsequence of $\widehat{x}_{\theta_l}(t)$ admits a subsequence $\widehat{x}_{\theta_g}(t)$ that converges to $\widehat{x}_\infty(t)$ and the second convergence claim is shown.

It remains to show the continuity for $\theta \in (0, \infty)$. This follows from the fact that for each $t \in [0, T]$ this mapping is uniformly bounded with respect to θ, see Lemma A.3. This allows to extract converging subsequences for families $\theta \to \widehat{x}_\theta(t)$ with $\theta \to \bar{\theta}$ for some $\bar{\theta} \in (0, \infty)$. Subsequently one can pass to the limit in (3.10) to assert that this limit is necessarily $\widehat{x}_{\bar{\theta}}(t)$. $\qquad\square$

The arguments made in the proof further yield a representation of $\widehat{x}_\infty(t)$ with coefficients that are measurable in time. We obtain the following result.

Corollary 4.2. *The worst case minimizer $\widehat{x}_\infty$ is measurable and essentially bounded, and hence $\widehat{x}_\infty \in \mathcal{L}_{0,T}^\infty$.*

Proof. Above it was shown that it holds
$$\widehat{x}_\infty(t) = \left(\sum_{k=1}^N \beta_k(t) P_k(t)\right)^{-1} \sum_{k=1}^N \beta_k(t) P_k(t)\widehat{x}_k(t),$$
where $\beta_k(t)$ are the pointwise limits of $c_k[\theta_g, \widehat{x}_{\theta_g}(t)](t)$ as defined in (3.12). Since the $c_k[\theta_g, \widehat{x}_\theta(t)]$ are measurable, so are the $\beta_k(t)$. As discussed in the proof of Proposition 4.1 we have $0 \leq \beta_k(t) \leq 1$. Moreover P_k and $\widehat{x}_k$ are continuous in $t \in [0, T]$. Consequently $\sum_{k=1}^N \beta_k(t) P_k(t)\widehat{x}_k(t)$ and $\sum_{k=1}^N \beta_k(t) P_k(t)$ are measurable. According to Lemma A.2 $\sum_{k=1}^N \beta_k(t) P_k(t)$ is invertible with an inverse bounded uniformly in t. The continuity of $M \mapsto M^{-1}$ on positive definite matrices implies measurability of $\sum_{k=1}^N \beta_k(t) P_k(t)$. These facts imply the measurability and uniform boundedness of $\hat{x}_\infty$. $\qquad\square$

Remark 4.3. *Using the affine independence established in Theorem 3.6 one could show a relation between the here discussed representation via β_k and the one via α_k as presented in Theorem 3.6. Since this does not offer any further insight towards the properties of $\widehat{x}_\infty$, we omit the required technical arguments.*

Due to the θ-uniform bound on $\widehat{x}_\theta(t)$ the pointwise convergence implies convergence in appropriate function spaces.

Proposition 4.4. *Let $y \in \mathcal{L}^2_{0,T}$ and $p \in \mathbb{N}$ such that $1 \le p < \infty$. It holds*

$$\|\widehat{x}_\theta - \widehat{x}_0\|_{W^{1,1}(0,T;\mathbb{R}^n)} \to 0 \qquad \text{for } \theta \to 0,$$
$$\|\widehat{x}_\theta - \widehat{x}_\infty\|_{\mathcal{L}^p_{0,T}} \to 0 \qquad \text{for } \theta \to \infty.$$

Proof. We first consider the limit for $\theta \to \infty$. The pointwise convergence established in Proposition 4.1 together with the θ-uniform bound from Lemma A.3 allow for an application of the dominated convergence theorem yielding the asserted convergence in $\mathcal{L}^p_{0,T}$ for any $1 \le p < \infty$.

The exact same argument yields convergence of $\widehat{x}_\theta$ to $\widehat{x}_0$ in $\mathcal{L}^1_{0,T}$ for $\theta \to 0$. It remains to show L^1 convergence of the derivatives. According to Theorem 3.12 for all $\theta \in (0, \infty)$ the weak derivative of $\widehat{x}_\theta$ exists, lies in $\mathcal{L}^1_{0,T}$ and for almost all $t \in [0, T]$ is given by the formula (3.14). Passing to the limit $\theta \to 0$ in said formula and comparing the outcome with the formula given in Corollary 3.2 we find that we have pointwise convergence almost everywhere, i.e., for almost all $t \in [0, T]$ it holds

$$\dot{\widehat{x}}_\theta(t) \to \dot{\widehat{x}}_0(t) \qquad \text{for } \theta \to 0.$$

It remains to establish an integrable upper bound that dominates $\dot{\widehat{x}}_\theta(t)$ almost everywhere. Denoting the uniform bound for $\widehat{x}_\theta(t)$ established in Lemma A.3 by $C = \hat{C}(\|x_0\|^2 + \|y\|_{\mathcal{L}^2_{0,T}})^2$ and utilizing (3.14) we find

$$\|\dot{\widehat{x}}_\theta(t)\| \le \|M(\widehat{x}_\theta(t), t)^{-1}\|_2 \|V(\widehat{x}_\theta(t), t)\|,$$

with M and V as in (3.15) and (3.16), respectively. We begin by estimating the first factor. In the following, for two symmetric matrices $M_1, M_2 \in \mathbb{R}^{n \times n}$ we write $M_1 \ge M_2$ whenever $M_1 - M_2$ is positive semi-definite. From the definition of M we deduce that

$$M(\widehat{x}_\theta(t), t) \ge \sum_{k=1}^N P_k(t) \ge P_1(t).$$

It follows

$$\|M(\widehat{x}_\theta(t), t)^{-1}\|_2 \le \|P_1(t)\|_2 \le \max_{s \in [0,T]} \|P_1(s)\|_2 = c_1 \tag{4.3}$$

with c_1 independent of t and θ. Turning to $V(\widehat{x}_\theta(t), t)$ the only critical term is

$$\theta \sum_{k=1}^N e^{\|\widehat{x}_\theta(t) - \widehat{x}_k(t)\|^2_{P_k(t)} + r_k(t)} \|y(t) - C\widehat{x}_k(t)\|^2 P_k(t)(\widehat{x}_\theta(t) - \widehat{x}_k(t)). \tag{4.4}$$

The remaining ones are either continuous in $t \in [0, T]$ or they can be uniformly estimated using Lemma A.3. Note in particular that θ appearing as factor does not pose any issue, as we are considering the limit for $\theta \to 0$. Turning to (4.4) we use the known bounds for P_k, $\widehat{x}_k$ and $\widehat{x}_\theta(t)$, so that since $y \in \mathcal{L}^2_{0,T}$ this term can be bounded by an L^1 function.

We have now shown that $\dot{\widehat{x}}_\theta$ converges pointwise almost everywhere and is pointwise almost everywhere bounded by an integrable function. Consequently the theorem of dominated convergence yields

$$\dot{\widehat{x}}_\theta \to \dot{\widehat{x}}_0 \qquad \text{for } \theta \to 0, \qquad\qquad \text{in } \mathcal{L}^1_{0,T}.$$

Together with the L^1 convergence of $\widehat{x}_\theta$ to $\widehat{x}_0$ this yields the asserted convergence of $\widehat{x}_\theta$ in $W^{1,1}$. $\square$

5. Error analysis

Throughout this section we denote the true, hidden parameter as $\bar{\sigma} \in \Sigma$. We establish a priori bounds for the weighted distance of each estimator to the Kalman filter $\widehat{x}_{\bar{\sigma}}$ associated with the true parameter characterized via

$$\begin{aligned}
\dot{\widehat{x}}_{\bar{\sigma}}(t) &= A_{\bar{\sigma}}\widehat{x}_{\bar{\sigma}}(t) + \Pi_{\bar{\sigma}}(t)C^\top Q_{\bar{\sigma}}^{-1}\left(y(t) - C\widehat{x}_{\bar{\sigma}}(t)\right) \qquad t \in (0,T), \\
\widehat{x}_{\bar{\sigma}}(0) &= x_0,
\end{aligned} \tag{5.1}$$

and

$$\begin{aligned}
\dot{\Pi}_{\bar{\sigma}}(t) &= A_{\bar{\sigma}}\Pi_{\bar{\sigma}}(t) + \Pi_{\bar{\sigma}}(t)A_{\bar{\sigma}}^\top - \Pi_{\bar{\sigma}}(t)C^\top Q_{\bar{\sigma}}^{-1}C\Pi_{\bar{\sigma}}(t) + BR_{\bar{\sigma}}B^\top \qquad t \in (0,T), \\
\Pi_{\bar{\sigma}}(0) &= \Gamma_{\bar{\sigma}}.
\end{aligned} \tag{5.2}$$

We begin by citing two results from an earlier work. The following auxiliary result is required for proving error bounds for the estimators. The proof is presented in [17, Lem. 4.6].

Lemma 5.1. *Let $\widehat{x}_k$, Π_k, $\widehat{x}_{\bar{\sigma}}$, and $\Pi_{\bar{\sigma}}$ be the solutions of* (3.1), (3.2), (5.1), *and* (5.2), *respectively. Then there exists a constant $c_k > 0$ independent of $t \in [0,T]$ such that*

$$\|\widehat{x}_k(t) - \widehat{x}_{\bar{\sigma}}(t)\| \le c_k\|\mathcal{S}_k - \mathcal{S}_{\bar{\sigma}}\|_1.$$

With this technical result at hand the following bound can be shown for the minimizer of the expected energy $\widehat{x}_0$ as defined in Subsection 3.2, for a proof see [17, Prop. 4.10].

Proposition 5.2. *There exists a constant $c_0 > 0$ independent of N such that for all $t \in [0,T]$ it holds*

$$\|\widehat{x}_0(t) - \widehat{x}_{\bar{\sigma}}(t)\|_{P_{\bar{\sigma}(t)}} \le c_0\,\mathbb{E}\left[\|\mathcal{S}_\sigma - \mathcal{S}_{\bar{\sigma}}\|_1\right].$$

We proceed to show similar bounds for the remaining two estimators.

Proposition 5.3. *There exists a constant $c_\infty > 0$ independent of N such that for all $t \in [0,T]$ and $\theta \in (0,\infty)$ it holds*

$$\|\widehat{x}_\infty(t) - \widehat{x}_{\bar{\sigma}}(t)\|_{P_{\bar{\sigma}(t)}} \le c_\infty \max_{j \in \{1,\ldots,N\}} \|\mathcal{S}_j - \mathcal{S}_{\bar{\sigma}}\|_1,$$

$$\|\widehat{x}_\theta(t) - \widehat{x}_{\bar{\sigma}}(t)\|_{P_{\bar{\sigma}(t)}} \le c_\infty \max_{j \in \{1,\ldots,N\}} \|\mathcal{S}_j - \mathcal{S}_{\bar{\sigma}}\|_1.$$

Proof. We begin by showing the first estimate. The proof is based on the representation of $\widehat{x}_\infty(t)$ given in Theorem 3.6. Let $\alpha_k(t)$ and $S(t)$ be as described there. For convenience we fix t and denote $\alpha_k = \alpha_k(t)$ and $S = S(t)$. With Lemma A.1, Lemma 5.1, and Lemma A.2

we obtain

$$\|\widehat{x}_\infty(t) - \widehat{x}_{\bar\sigma}(t)\|_{P_{\bar\sigma}(t)} = \left\|(\sum_{k\in S}\alpha_k P_k(t))^{-1}\sum_{k\in S}\alpha_k P_k(t)(\widehat{x}_k(t) - \widehat{x}_{\bar\sigma}(t))\right\|_{P_{\bar\sigma}(t)}$$

$$\leq \lambda_{\max}^{\frac{3}{2}}\lambda_{\min}^{-1}\sum_{k\in S}\alpha_k\|\widehat{x}_k(t) - \widehat{x}_{\bar\sigma}(t)\| \leq \lambda_{\max}^{\frac{3}{2}}\lambda_{\min}^{-1}\sum_{k\in S}\alpha_k c_k\|\mathcal{S}_k - \mathcal{S}_{\bar\sigma}\|_1.$$

The assertion follows by setting $c_\infty = \lambda_{\max}^{\frac{3}{2}}\lambda_{\min}^{-1}\max_{k\in S}c_k$ and estimating $\|\mathcal{S}_k - \mathcal{S}_{\bar\sigma}\|_1$ from above by the maximum over $k \in S$.

Using the characterization of $\widehat{x}_\theta$ given in (3.13) the proof for the second estimate can be carried out analogously. $\qquad\square$

We conclude the section by presenting another error bound for the minimizer of the entropic risk $\widehat{x}_\theta$ as defined in Definition 3.10.

Proposition 5.4. *Let c_∞ be the constant from Proposition 5.3. For all $\theta \in (0,\infty)$ and $t \in [0,T]$ it holds*

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

where $J(t,\theta) = \max_{k,j\in\{1,\dots,N\}}\left[\mathcal{V}_k(t,\widehat{x}_\theta(t)) - \mathcal{V}_j(t,\widehat{x}_\theta(t))\right]$.

Proof. Note first that with Lemma A.2 we obtain

$$\|(\sum_{k=1}^N P_k(t))^{-1}\|_2 = \frac{1}{N}\|(\sum_{k=1}^N\frac{1}{N}P_k(t))^{-1}\|_2 \leq (N\lambda_{\min})^{-1}.$$

Now utilizing (3.10) and Lemma A.1 we find

$$\|\widehat{x}_\theta(t) - \widehat{x}_{\bar\sigma}(t)\|_{P_{\bar\sigma}(t)} \leq \lambda_{\max}^{\frac{3}{2}}\|(\sum_{k=1}^N e^{\theta\mathcal{V}_k(t,\widehat{x}_\theta(t))}P_k(t))^{-1}\|_2\sum_{k=1}^N e^{\theta\mathcal{V}_k(t,\widehat{x}_\theta(t))}\|\widehat{x}_k(t) - \widehat{x}_{\bar\sigma}(t)\|$$

$$\leq \lambda_{\min}^{-1}\lambda_{\max}^{\frac{3}{2}}e^{\theta(\mathcal{V}_{k_{\max}}(t,\widehat{x}_\theta(t)) - \mathcal{V}_{k_{\min}}(t,\widehat{x}_\theta(t)))}\frac{1}{N}\sum_{k=1}^N\|\widehat{x}_k(t) - \widehat{x}_{\bar\sigma}(t)\|,$$

where $k_{\min}$ is chosen such that $\min_k \mathcal{V}_k(t,\widehat{x}_\theta(t)) = \mathcal{V}_{k_{\min}}(t,\widehat{x}_\theta(t))$ and $k_{\max}$ is chosen analogously. Applying Lemma 5.1 yields the assertion. $\qquad\square$

6. Numerical experiments

This section presents numerical results illustrating and comparing the proposed estimators using two examples.

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

$$\begin{aligned}\dot{x}(t) &= A_\sigma x(t) + Bv(t) & t\in(0,T), & \qquad x(0) = x_0 + \eta, \\ y(t) &= Cx(t) + \mu(t) & t\in(0,T),\end{aligned} \tag{6.1}$$

where $\sigma \in \Sigma_A \subset \mathbb{R}^{s_A}$ with $|\Sigma_A| = N_A$, as in Subsection 3.1. The errors are associated with positive definite covariance matrices Γ, R, and Q independent of σ.

All ODEs are solved using the MATLAB® solver `ode15s` with a relative tolerance of 10^{-8} and an equidistant time grid $0 = t_0 < t_1 < \cdots < t_{1000} = T$. Kalman filter equations such as (2.6)-(2.7) are transformed into $n + n^2$ dimensional ODEs.

To generate outputs $y \in \mathcal{L}^2_{0,T}$ we utilize the MATLAB® function `normrnd` to 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 $k = 0, \ldots, 1000$. The time continuous errors $v \in \mathcal{L}^2_{0,T}$ and $\mu \in \mathcal{L}^2_{0,T}$ are obtained via linear interpolation. After choosing a $\bar{\sigma} \in \Sigma_A$ to be the true, hidden parameter y is obtained by solving (6.1) with the constructed errors and $\bar{\sigma}$. Note that this construction of v and μ only works in discrete time and the extension via interpolation is of heuristic nature and does not converge to any function for $\Delta t \to 0$. In spirit this limit leads to the stochastic formulation via random processes and SDEs.

To construct the estimators one needs to solve for the individual Kalman filters $\hat{x}_k$ and P_k, $k = 1, \ldots, N_A$. As they are independent of each other this is done in parallel using `parfor` from the MATLAB® Parallel Computing Toolbox.

With the $\hat{x}_k$ and P_k at hand, the risk neutral estimator $\hat{x}_0$ is realized directly via (3.6). The realization of the risk averse estimators $\hat{x}_\theta$, $\theta \in (0, \infty)$, however, requires a more elaborate approach. We opt for a gradient descent scheme to approximate $\hat{x}_\theta(t_k)$, $k = 0, \ldots, 1000$ as the minimizer of $\rho_\theta(\mathcal{V}_\sigma(t_k, x))$, cf. (3.9). For each iteration the stepsize is constructed using the Barzilai-Borwein method [2,3] combined with Armijo backtracking. Denoting the risk aversion parameters of interest by $\theta_1, \ldots, \theta_l$, sorted increasingly, we initialize the gradient descent for θ_1 with $\hat{x}_0$ and in an iterative manner the construction of $\hat{x}_{\theta_g}$ is initialized by $\hat{x}_{\theta_{g-1}}$. For the evaluation of $\rho_\theta(\mathcal{V}_\sigma)$ and its gradient we employ the well known log-sum-exp trick to avoid overflow. The realizations were implemented in MATLAB® R2024b and the code is available in [26].

6.2. **Harmonic oscillator with uncertain damping.** As a first example we consider a harmonic oscillator which in first order form is modeled as

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

The two components of the state vector $x = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^\top$ represent the position and velocity of the system. Further $m > 0$, $c \geq 0$, and $k \geq 0$ represent the mass, damping coefficient, and spring constant, respectively. The function $v \in \mathcal{L}^2_{0,T}$ 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. Finally, y denotes the measurement of the position with a deterministic but unknown output error $\mu \in \mathcal{L}^2_{0,T}$.

For our experiments we fix the mass and the spring constant to one, i.e., $m = k = 1$, the time horizon is set to $T = 5$. As an undisturbed initial state we use $x_0 = \begin{bmatrix} 1 & 0 \end{bmatrix}^\top$ and for the covariances we set $\Gamma = 0.1$ Id, $R = 0.05$, and $Q = 0.05$. In the following we present results for two sets of possible damping parameters each containing $N_A = 100$ elements.

The first parameter set is given via uniform samples from the interval $[0.1, 3]$, cf. Figure 1a, the second parameter set is sampled via a log-normal distribution with mean -0.25 and variance 0.5, both in the logarithmic scale, illustrated in Figure 1b. For both sets the parameter of maximal value is designated as the true underlying parameter $\bar{\sigma}$. The resulting disturbed state x, and associated state estimators are presented in Figure 1c and Figure 1d, respectively. We observe that the Kalman filter $\hat{x}_{\bar{\sigma}}$ associated with the hidden parameter yields a satisfying reconstruction of the hidden state x validating the Kalman

based approach. However, since $\bar{\sigma}$ is not known, $\widehat{x}_{\bar{\sigma}}$ can not be constructed and we turn to the uncertainty based estimators.

Comparing the risk neutral estimator $\widehat{x}_0$ with the risk averse estimators $\widehat{x}_\theta$, for $\theta = 0.1, 0.5, 1, 20, 750, 1000$ we find that if the underlying true parameter is an outlier then the risk averse approach yields noticeably more accurate reconstructions than the risk neutral one, cf. Figure 1d. In the case of uniformly distributed parameters we observe relatively large similarities between the risk neutral and risk averse approaches. In both phase plots, see Figure 1c and Figure 1d, it appears that convergence for $\theta \to \infty$ has almost been reached. Indeed, the difference of the trajectories $\widehat{x}_{750}$ and $\widehat{x}_{1000}$ is close to zero.

A deeper analysis of the numerical results is offered in Table 1 where we present the time integral over the entropic risk evaluated along a selection of state estimates. More specifically for $\theta = 0, 0.5, 20, 1000$ and $\tau = 0, 0.5, 20, 1000, \infty$ we present the values

$$\int_0^T \rho_\tau(\mathcal{V}_\sigma(t, \widehat{x}_\theta(t))) \, \mathrm{d}t,$$

where for $\tau \in (0, \infty)$ we denote by ρ_τ the entropic risk as defined in Definition 3.8, and ρ_0 and ρ_∞ denote the expectation and essential supremum, respectively. By construction it holds that $\widehat{x}_\theta(t)$ minimizes $\rho_\theta(\mathcal{V}_\sigma(t, \cdot))$. Consequently, for both parameter sets the minima in rows one to four (printed in bold face) lie on the diagonal. Additionally, the largest risk aversion parameter $\theta = 1000$ leads to the smallest value of the essential supremum, reiterating the proximity of $\widehat{x}_{1000}$ to the limit $\widehat{x}_\infty$. The table further illustrates the interplay between risk aversion and the underlying parameter set. For the case of uniformly distributed parameters the level of risk aversion is of limited influence. In row 1 we find that the risk neutral estimator ($\theta = 0$) and the strongly risk averse estimator ($\theta = 1000$) perform similarly well with respect to the expected value. In fact, the minimizer $\widehat{x}_0$ performs only about

$$\frac{6.3765 - 6.0716}{6.3765} \approx 4.8\%$$

better than $\widehat{x}_{1000}$. Similarly, row 5 shows that the performance of $\widehat{x}_0$ with respect to the essential supremum is almost as good as the one of the risk averse estimator $\widehat{x}_{1000}$. Specifically, the latter offers an improvement over the former of about

$$\frac{10.589 - 9.4986}{10.589} \approx 10.3\%.$$

The results are drastically different for the second parameter set. In row 6 we observe that $\widehat{x}_0$ performs about

$$\frac{9.2433 - 6.7095}{9.2433} \approx 27.4\%$$

better than $\widehat{x}_{1000}$ when considering the expected value. On the other hand, in terms of the essential supremum the risk averse estimator offers an improvement of about

$$\frac{38.15 - 11.945}{38.15} \approx 68.7\%.$$

Summarizing the results we observe that the parameter set may play a crucial role towards the effect of the risk averse estimators. In our study we find that the relative differences for the two parameter sets differ by a factor of 6. Further note that, by construction moving from risk neutral to strongly risk averse estimation improves the performance with respect to the essential supremum (see rows five and ten) while the performance with respect to
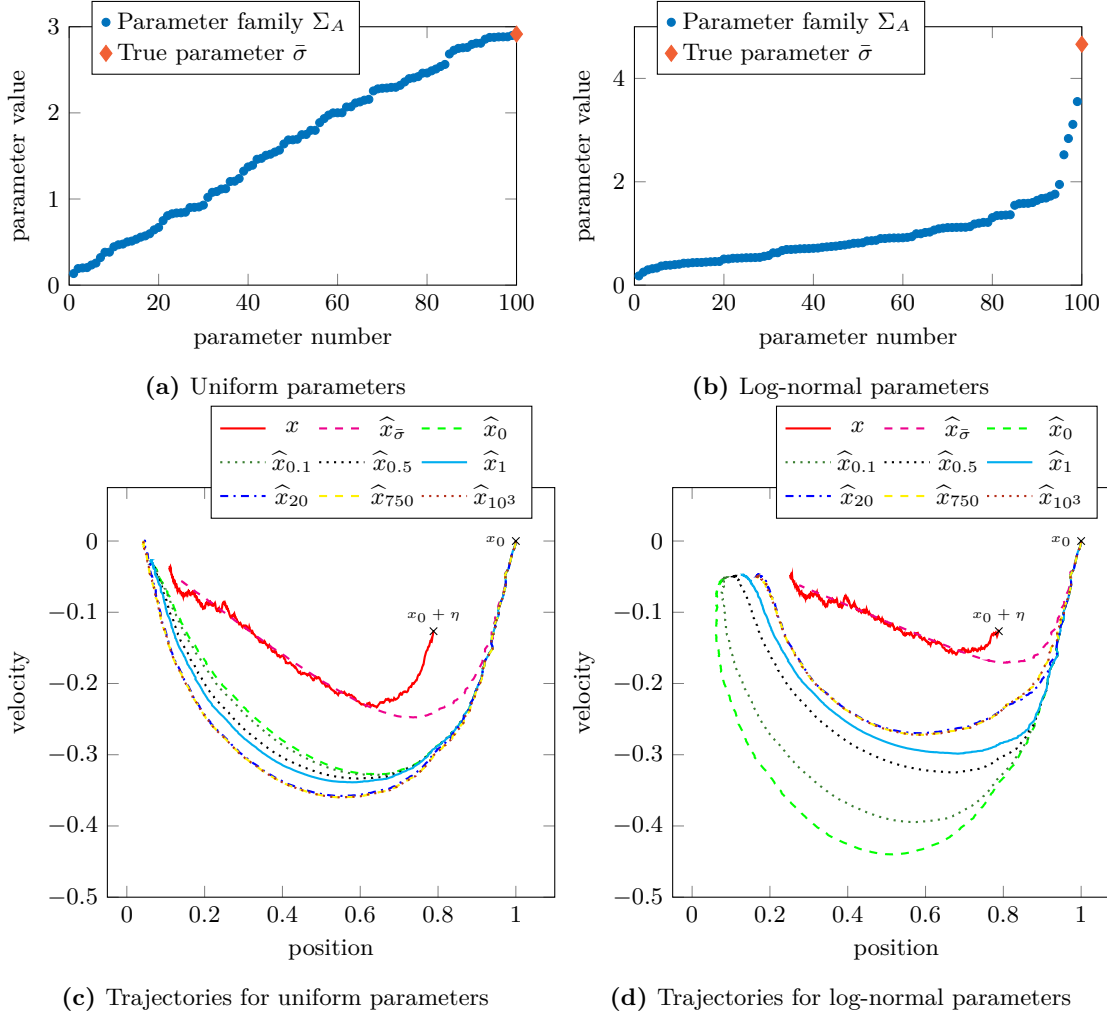
(a) Uniform parameters

(b) Log-normal parameters

(c) Trajectories for uniform parameters

(d) Trajectories for log-normal parameters

Figure 1 – Harmonic oscillator with uncertain damping parameter

the expectation might suffer, as shown in rows one and six. For this example, however, the improvement with respect to ess sup is noticeably larger than the decrease in terms of $\mathbb{E}$.

6.3. **Connected amplidynes with uncertain inductances.** For our second example we consider an electrical circuit amplifying a given input. The so-called amplidyne amplifies a given input and can be described via a linear dynamical system, see [18, Ch. 1.12].KwaSiv72. For our experiments we consider two such amplidynes that are connected such that the output of the first one acts as an input of the second one. The total system then consists

		$\theta = 0$	$\theta = 0.5$	$\theta = 20$	$\theta = 1000$
	Uniform parameters				
1	$\int \mathbb{E}[\mathcal{V}_\sigma(t, \widehat{x}_\theta(t))]\,\mathrm{d}t$	**6.0716**	6.1175	6.3561	6.3765
2	$\int \rho_{0.5}(\mathcal{V}_\sigma(t, \widehat{x}_\theta(t)))\,\mathrm{d}t$	6.7182	**6.6248**	6.8269	6.8508
3	$\int \rho_{20}(\mathcal{V}_\sigma(t, \widehat{x}_\theta(t)))\,\mathrm{d}t$	10.171	9.6972	**9.1291**	9.148
4	$\int \rho_{1000}(\mathcal{V}_\sigma(t, \widehat{x}_\theta(t)))\,\mathrm{d}t$	10.579	10.111	9.5187	**9.4896**
5	$\int \operatorname{ess\,sup} \mathcal{V}_\sigma(t, \widehat{x}_\theta(t))\,\mathrm{d}t$	10.589	10.12	9.5282	**9.4986**
	Log-normal parameters				
6	$\int \mathbb{E}[\mathcal{V}_\sigma(t, \widehat{x}_\theta(t))]\,\mathrm{d}t$	**6.7095**	7.7027	9.2433	9.2433
7	$\int \rho_{0.5}(\mathcal{V}_\sigma(t, \widehat{x}_\theta(t)))\,\mathrm{d}t$	20.581	**8.521**	9.7391	9.738
8	$\int \rho_{20}(\mathcal{V}_\sigma(t, \widehat{x}_\theta(t)))\,\mathrm{d}t$	37.676	18.529	**11.58**	11.588
9	$\int \rho_{1000}(\mathcal{V}_\sigma(t, \widehat{x}_\theta(t)))\,\mathrm{d}t$	38.14	19.007	11.958	**11.935**
10	$\int \operatorname{ess\,sup} \mathcal{V}_\sigma(t, \widehat{x}_\theta(t))\,\mathrm{d}t$	38.15	19.016	11.968	**11.945**

Table 1 – Various risk measures integrated along selected state estimates

of four components and is modeled as

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

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

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

where $x(t) \in \mathbb{R}^4$ encodes the current of the individual components. The resistances and inductances of the individual components are given by $\rho_i > 0$ and $L_i > 0$, respectively. Further for $i = 1, 2, 3, 4$ we have $e_i = k_i x_i$, i.e., $k_i > 0$ describe the proportion of the currents x_i and the voltages e_i. By e_0 we denote the known, time-dependent input entering the first component. It is subject to the disturbance v. Finally, the measured output is the output of the second amplidyne $e_4 = k_4 x_4$. The known forcing term given by e_0 can be incorporated in our formulation of the Kalman filter in a straightforward fashion. Our numerical experiments are conducted with $\rho_1 = \rho_3 = 5$, $\rho_2 = \rho_4 = 10$, $k_1 = k_3 = 20$, $k_2 = k_4 = 50$, $L_1 = L_3 = 0.5$, $T = 10$, and $e_0(t) \equiv 1$ and undisturbed initial state $x_0 = \begin{bmatrix} 0.5 & 1 & 10 & 20 \end{bmatrix}^\top$. The uncertainty of the system lies in the inductances $\sigma = \begin{bmatrix} L_2 & L_4 \end{bmatrix}^\top \in \mathbb{R}^2$. As noise covariances we use

$$\Gamma = 0.25 \operatorname{diag}(|x_{0,1}|, |x_{0,2}|, |x_{0,3}|, |x_{0,4}|) = \operatorname{diag}(0.125, 0.25, 2.5, 5),$$

$$R = (0.1|e_0(0)|)^2 = 0.01,$$

$$Q = (0.1 * 400|e_0(0)|)^2 = 1600,$$

where the different magnitudes are due to the scales of the different components.

Again we construct two parameter sets to compare their effects on the risk averse state estimators. Both contain $N_A = 100$ elements sampled according to given bivariate distributions. The first set of parameters is sampled from a uniform distribution on the rectangle $[10, 40]^2$ and illustrated in Figure 2a. The second parameter set is sampled from

a Gaussian mixture characterized via

$$0.95\,\mathcal{N}\left(\begin{bmatrix}15\\35\end{bmatrix},\begin{bmatrix}2&0\\0&2\end{bmatrix}\right)+0.05\,\mathcal{N}\left(\begin{bmatrix}35\\15\end{bmatrix},\begin{bmatrix}1&0\\0&1\end{bmatrix}\right),$$

i.e., every drawn sample has a chance of 0.95 to be drawn from a normal distribution with mean $\begin{bmatrix}15 & 35\end{bmatrix}^\top$ and covariance 2Id and a 0.05 chance to be drawn from a normal distribution with mean $\begin{bmatrix}35 & 15\end{bmatrix}^\top$ and covariance Id. The resulting parameter set is shown in Figure 2b. For both sets the parameter $\sigma = (L_2, L_4)$ that maximizes $L_2 - L_4$ is set to be the true underlying parameter.

To illustrate the qualitative behavior of the trajectories we plot the second component of the resulting state x and the state estimators in Figure 2c and Figure 2d. Again we observe satisfying results obtained from the Kalman filter constructed based on the hidden parameter. Further, for both parameter sets the risk averse estimators seem to outperform the risk neutral option in terms of euclidean distance. We note that for this example convergence for $\theta \to \infty$ is approached for far lower θ when compared to the oscillator, showing that the effective range of the risk aversion parameter is highly problem dependent. To further investigate the performances of the risk neutral estimator $\widehat{x}_0$ and the risk averse estimator $\widehat{x}_4$ we present the energies according to each parameter $\sigma \in \Sigma_A$ or equivalently for each σ_k, $k = 1, \ldots, N_A$. More precisely, in Figure 2e and Figure 2f we plot the values

$$\mathcal{V}_{\sigma_k}(t, \widehat{x}_\theta(t)) = \mathcal{V}_k(t, \widehat{x}_\theta(t))$$

along $t \in [0, T]$ and for $\theta = 0$ and $\theta = 4$. Additionally, we present the resulting risk measures

$$\mathbb{E}\left[\mathcal{V}_{\sigma_k}(t, \widehat{x}_\theta(t))\right] \quad \text{and} \quad \max_k \mathcal{V}_{\sigma_k}(t, \widehat{x}_\theta(t)),$$

again for $\theta = 0$ and $\theta = 4$. Comparing the dashed lines in Figure 2e we find that for the uniform parameter set $\widehat{x}_4$ significantly outperforms $\widehat{x}_0$ with respect to the maximum over the family members, showing that for this example the risk averse state estimation works exactly as intended. Interestingly, $\widehat{x}_4$ still performs relatively well with respect to the expectation when compared to the minimizer $\widehat{x}_0$, cf. solid orange and dotted red line. Turning to Figure 2f we observe that the risk aversion takes an even stronger effect. It appears that the energies corresponding to the four outliers in the parameter set, cf. Figure 2b, evaluated along $\widehat{x}_0$ clearly stand out resulting in a poor performance with respect to the maximum. On the other hand, in this example the relative difference of the expected energies along $\widehat{x}_4$ and $\widehat{x}_0$ is larger when compared to the uniform parameter set.

7. Conclusion

This work presents a novel approach for risk averse state estimation under uncertainty for linear, time-invariant, finite-dimensional systems. Investigating analogous concepts for time-dependent, nonlinear systems as well as for systems governed by PDEs could be the focus of future research.

Acknowledgement

(a) Uniform parameters

(b) Parameters via Gaussian mixture

(c) Second component along time

(d) Second component along time

(e) Energies via uniform parameters
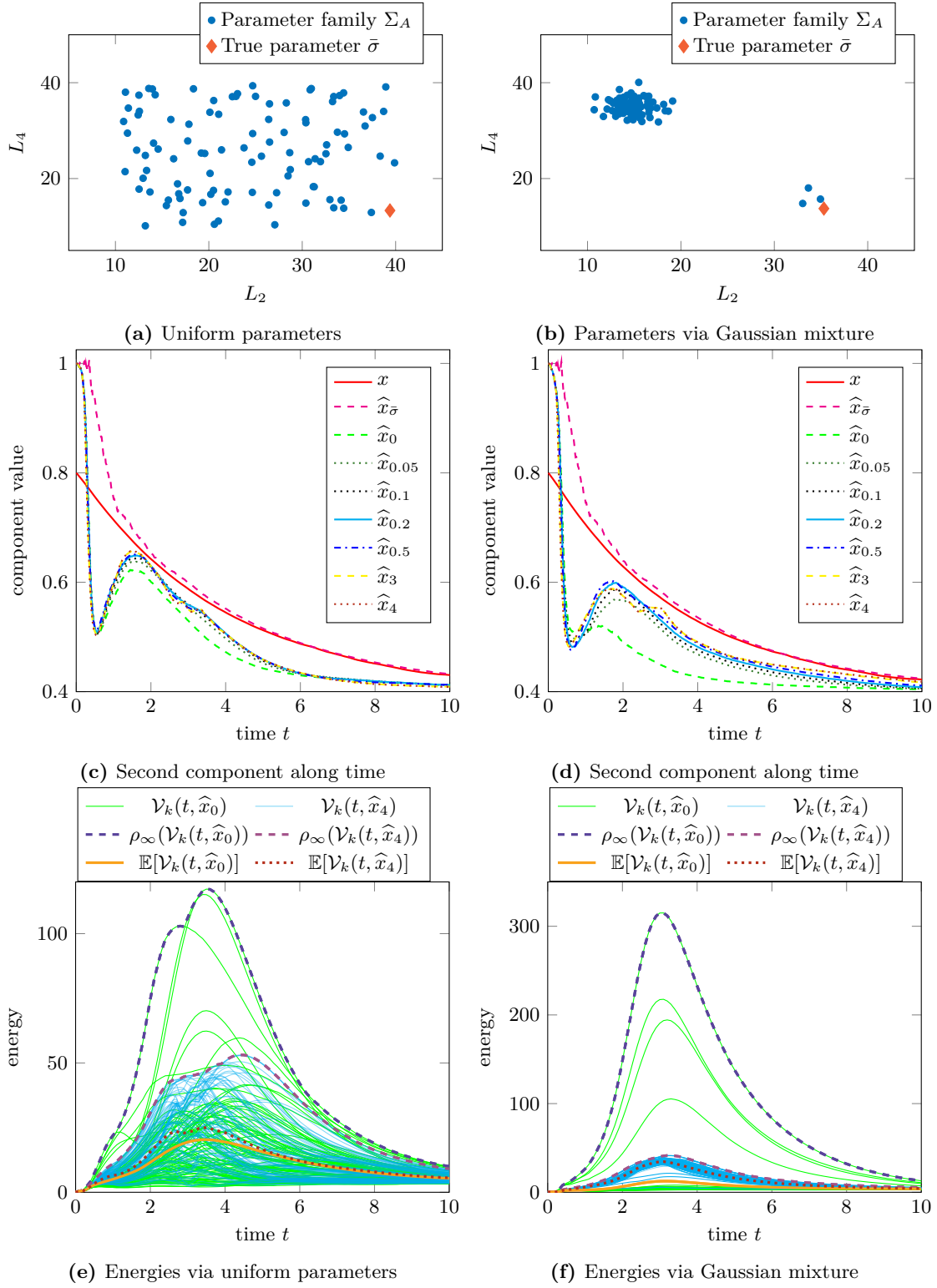
(f) Energies via Gaussian mixture

Figure 2 – Amplidyne with uncertain inductance

References

[1] A. C. Aitken, *On least squares and linear combination of observations*, Proc. R. Soc. Edinb., 55 (1936), p. 42–48.

[2] B. Azmi and K. Kunisch, *Analysis of the Barzilai-Borwein step-sizes for problems in Hilbert spaces*, J. Optim. Theory Appl., 185 (2020), pp. 819–844.

[3] J. Barzilai and J. M. Borwein, *Two-point step size gradient methods*, IMA J. Numer. Anal., 8 (1988), pp. 141–148.

[4] G. C. Calafiore, S. Gaubert, and C. Possieri, *Log-sum-exp neural networks and posynomial models for convex and log-log-convex data*, IEEE Trans. Neural Netw. Learn. Syst., 31 (2020), pp. 827–838.

[5] M. E. Cawood and P. M. Dearing, *The weighted Euclidean one-center problem in $\mathbb{R}^n$*, Comput. Optim. Appl., 89 (2024), pp. 553–574.

[6] 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.

[7] P. M. Dearing and M. E. Cawood, *The minimum covering Euclidean ball of a set of euclidean balls in $\mathbb{R}^n$*, Ann. Oper. Res., 322 (2023), pp. 631–659.

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

[9] I. Ekeland and R. Temam, *Convex analysis and variational problems*, North-Holland Publ. a. o., Amsterdam a. o., first ed., 1976.

[10] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*, Wiley, New York a. o., second ed., 1999.

[11] H. Föllmer and A. Schied, *Stochastic finance: an introduction in discrete time*, 2016.

[12] P. A. Guth, V. Kaarnioja, F. Y. Kuo, C. Schillings, and I. H. Sloan, *Parabolic PDE-constrained optimal control under uncertainty with entropic risk measure using quasi-Monte Carlo integration*, Numer. Math., 156 (2024), pp. 565—-608.

[13] P. A. Guth and K. Kunisch, *Approximation of risk-averse optimal feedback control*, prepr., (2025).

[14] D. W. Hearn and J. Vijay, *Efficient algorithms for the (weighted) minimum circle problem*, Oper. Res., 30 (1982), pp. 777–795.

[15] 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.

[16] 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.

[17] K. Kunisch and J. Schröder, *Deterministic Kalman filters for uncertain dynamical systems*, prepr., (2025).

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

[19] R. Mehra, *On the identification of variances and adaptive Kalman filtering*, IEEE Trans. Autom. Control, 15 (1970), pp. 175–184.

[20] B. Mordukhovich, N. M. Nam, and C. Villalobos, *The smallest enclosing ball problem and the smallest intersecting ball problem: existence and uniqueness of solutions*, Optim. Lett., 7 (2013), pp. 839—853.

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

[22] N. M. Nam, T. A. Nguyen, and J. Salinas, *Applications of convex analysis to the smallest intersecting ball problem*, J. Convex Anal., 19 (2012), pp. 497–518.

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

[24] R. K. Pathria and P. D. Beale, *Statistical Mechanics*, Elsevier, New York a. o., third ed., 2011.

[25] W. Schirotzek, *Nonsmooth analysis*, Springer-Verlag, Berlin, Heidelberg, first ed., 2007.

[26] J. Schröder, *Code for the paper "Risk averse deterministic Kalman filters for uncertain dynamical systems"*, 2025. doi: 10.5281/zenodo.17581815.

[27] 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.

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

[29] J. C. Willems, *Deterministic least squares filtering*, J. Econom., 118 (2004), pp. 341–373.

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

[31] E. Zeidler, *Applied Functional Analysis: Main Principles and Their Applications*, Springer, New York, first ed., 1995.

Appendix A. Technical proofs

A.1. **Technical auxiliary results.** This section contains auxiliary results required for the analysis of this work. The first lemma is concerned with the eigenvalues of the precision matrices and their impact on the weighted norms.

Lemma A.1. *Let $k \in \{1, \ldots, N\}$, $t \in [0, T]$, and $\lambda_k(t) \in \mathbb{R}$ be an eigenvalue of $P_k(t)$, where P_k are the precision matrices as introduced in Subsection 3.1. There exist real numbers $\lambda_{\min}, \lambda_{\max} > 0$ independent of t and k such that*

$$\lambda_{\min} \leq \lambda_k(t) \leq \lambda_{\max}.$$

In particular for every $x \in \mathbb{R}^n$ it holds

$$\lambda_{\min} \|x\|^2 \leq \|x\|^2_{P_k(t)} \leq \lambda_{\max} \|x\|^2.$$

Proof. The boundedness of the eigenvalues follows from their continuous dependence on the matrix entries and the regularity of P_k presented in (3.3). The positivity of the lower bound is a consequence of the positive definiteness of P_k. The estimates for the weighted norm are a consequence of the symmetry of P_k. $\square$

Next we estimate the spectral norm of the inverse of a convex combination of the precision matrices.

Lemma A.2. *Let P_k be the precision matrices and let $S \subset \{1, \ldots, N\}$ and $(\gamma_k)_{k \in S}$ be such that $\sum_{k \in S} \gamma_k = 1$. Then for every $t \in [0, T]$ the matrix $\sum_{k \in S} \gamma_k P_k(t)$ is non singular and*

$$\|(\sum_{k \in S} \gamma_k P_k(t))^{-1}\|_2 \leq \lambda_{\min}^{-1}.$$

Proof. For every $x \in \mathbb{R}^n$ we have

$$\left\langle x, \sum_{k \in S} \gamma_k P_k(t) x \right\rangle \geq \sum_{k \in S} \gamma_k \lambda_{\min} \|x\|^2 = \lambda_{\min} \|x\|^2.$$

The assertion follows with the Lax-Milgram Lemma [6, Thm. 1.1.3 & Rem. 1.1.3]. $\square$

A.2. **Proof of Theorem 3.12.** This subsection contains the proof of the regularity of the estimator minimizing the entropic risk of the energy. We begin by showing the result for continuous outputs y. The result for p-integrable outputs will follow using an approximation by continuous elements.

Throughout this subsection we use the following notation. For a given $y \in \mathcal{L}^2_{0,T}$ and $\theta \in (0, \infty)$ we denote by $\widehat{x}_k[y]$ the unique solution associated with (3.1), we set

$$r_k[y](t) = \int_0^t \|y(s) - C\widehat{x}_k[y](s)\|^2_{Q_k^{-1}} \, \mathrm{d}s. \tag{A.1}$$

By $\widehat{x}_\theta[y]$ we denote the associated minimizer of the entropic risk as defined in Definition 3.10. We begin with a a uniform bound of the estimator.

Lemma A.3. *Let $t \in [0, T]$, $\theta \in (0, \infty)$, and $y \in \mathcal{L}^2_{0,T}$. There exists a constant $\widehat{C} > 0$ independent of t, θ, and y such that*

$$\|\widehat{x}_\theta[y](t)\|^2 \leq \widehat{C} \left(\|x_0\|^2 + \|y\|^2_{\mathcal{L}^2_{0,T}} \right).$$

In particular $\widehat{x}_\theta[y] \in \mathcal{L}^\infty_{0,T}$.

Proof. Using the characterization of $\widehat{x}_\theta[y](t)$ given in (3.13) the proof can be conducted via the arguments used in the proof of Corollary 4.2. The particular bound in terms of x_0 and y follows from the estimate of $\widehat{x}_k$ obtained via a Gronwall estimate. $\square$

Next we show a regularity result for $\widehat{x}_\theta[y]$ under the assumption of continuous data. To that end we define the mapping

$$G[\theta, y] \colon [0, T] \times \mathbb{R}^n \to \mathbb{R}^n$$

$$G[\theta, y](t, x) := \sum_{k=1}^N e^{\theta \|x - \widehat{x}_k[y](t)\|^2_{P_k(t)} + \theta r_k[y](t)} P_k(t)(x - \widehat{x}_k[y](t)). \tag{A.2}$$

The following lemma investigates its regularity depending on the smoothness of the output. Again for $x \in \mathbb{R}^n$ we denote $x^2_\otimes = x \otimes x \in \mathbb{R}^{n,n}$.

Lemma A.4. *Let $\theta \in (0, \infty)$ and $y \in \mathcal{L}^2_{0,T}$.*

(i) *For every $t \in [0, T]$ the mapping $G[\theta, y](t, \cdot)$ is of class C^∞. Its first derivative is given by*

$$D_x G[\theta, y](t, x)$$
$$= \sum_{k=1}^N e^{\theta \|x - \widehat{x}_k[y](t)\|^2_{P_k(t)} + \theta r_k[y](t)} \left(P_k(t) + 2\theta \left[P_k(t)(x - \widehat{x}_k[y](t)) \right]^2_\otimes \right). \tag{A.3}$$

(ii) *For every $t \in [0, T]$ and $x \in \mathbb{R}^n$ the matrix $D_x G[\theta, y](x, t)$ is invertible. Further there exists a constant c independent of θ, y, t, and x such that*

$$\|D_x G[\theta, y](t, x)^{-1}\|_2 \leq c.$$

(iii) *Assuming continuity of the output, i.e., $y \in C([0, T]; \mathbb{R}^r)$ it holds that $G[\theta, y]$ is continuously differentiable in $[0, T] \times \mathbb{R}^n$. Its partial time derivative is given as*

$$\partial_t G[\theta, y](t, x) = \sum_{k=1}^N e^{\theta \|x - \widehat{x}_k[y](t)\|^2_{P_k(t)} + \theta r_k[y](t)} \left(-P_k(t) \dot{\widehat{x}}_k[y](t) + \dot{P}_k(t)(x - \widehat{x}_k[y](t)) \right.$$
$$- 2\theta \langle x - \widehat{x}_k[y](t), P_k(t) \dot{\widehat{x}}_k[y](t) \rangle P_k(t)(x - \widehat{x}_k[y](t)) \tag{A.4}$$
$$\left. + \theta \|y(t) - C\widehat{x}_k[y](t)\|^2_{Q_k^{-1}} P_k(t)(x - \widehat{x}_k[y](t)) \right).$$

Proof. The regularity with respect to x and the formula for the associated partial derivative are a direct consequence of the chain rule and the regularity of the exponential and the squared weighted norm proving (i).

To show (ii) let $t \in [0, T]$ and $x, z \in \mathbb{R}^n$ be arbitrary. Utilizing (A.3), the positive definiteness of $P_k(t)$, and the fact that $e^a \geq 1$ for $a \geq 0$ we find

$$z^\top D_x G[\theta, y](t, x) z \geq \sum_{k=1}^{N} z^\top P_k(t) z + 2\theta \left| z^\top P_k(t)(x - \widehat{x}_k[y](t)) \right|^2 \geq \min_{t \in [0,T]} z^\top P_1(t) z =: c',$$

where the continuity of P_1 ensures that the minimum on the right hand side is attained. The Lax-Milgram Lemma [6, Thm. 1.1.3 & Rem. 1.1.3] yields the invertibility and the bound with $c := \frac{1}{c'}$.

Turning to the regularity in time we note that Π_k given as a solution of (3.2) is continuously differentiable. Further, for $y \in C([0, T]; \mathbb{R}^r)$ we find that the right hand side of (3.1) is continuous. For the associated weak solution it follows $\widehat{x}_k[y] \in C^1([0, T]; \mathbb{R}^n)$. Additionally, the integrant in (A.1) is continuous, ensuring $r_k[y] \in C^1([0, T]; \mathbb{R})$. It follows that $t \to G[\theta, y](t, x)$ is a composition of continuously differentiable functions. Hence for $y \in C([0, T]; \mathbb{R}^r)$ the formula for the time derivative follows with the chain rule. Since both partial derivatives exist and are continuous in $[0, T] \times \mathbb{R}^n$, we obtain $G[\theta, y] \in C^1([0, T] \times \mathbb{R}^n; \mathbb{R}^n)$. $\qquad\square$

With these results at hand we can show the first regularity result

Proof of Theorem 3.12(ii)

Proof. The assertion is a direct consequence of the implicit function theorem. Lemma 3.11 shows that $\widehat{x}_\theta[y]$ is implicitly defined by (A.2), i.e., for a given $t \in [0, T]$ we have that $x = \widehat{x}_\theta[y](t)$ is the unique solution of

$$0 = G[\theta, y](t, x).$$

With the results from Lemma A.4 we can apply the implicit function theorem [31, Thm. 4.E] and obtain $\widehat{x}_\theta[y] \in C^1([0, T]; \mathbb{R}^n)$ and the following formula for the derivative

$$\dot{\widehat{x}}_\theta[y](t) = -D_x G[\theta, y](t, \widehat{x}_\theta[y](t))^{-1} \partial_t G[\theta, y](t, \widehat{x}_\theta[y](t)). \tag{A.5}$$

In light of the identities (A.3) and (A.4) this is the formula for the derivative as announced in Theorem 3.12. $\qquad\square$

This result will be carried over to less regular measurements y via a density argument. To that end we first ensure that the terms appearing in the energy $\mathcal{V}_k$ depend continuously on the output in an appropriate sense.

Lemma A.5. *For all $k \in \{1, \dots, N\}$, $t \in (0, T]$, and $s \in [0, t]$ the mappings*

$$y \in \mathcal{L}_{0,t}^2 \mapsto \widehat{x}_k[y] \in \mathcal{H}_0^t, \qquad\qquad y \in \mathcal{L}_{0,t}^2 \mapsto \widehat{x}_k[y](s) \in \mathbb{R}^n$$

and

$$y \in \mathcal{L}_{0,t}^2 \mapsto r_k[y] \in C([0, t]; \mathbb{R}), \qquad\qquad y \in \mathcal{L}_{0,t}^2 \mapsto r_k[y](s) \in \mathbb{R},$$

are continuous.

Proof. Let k and $t \in [0, T]$ be fixed and let $(y_j) \subset \mathcal{L}_{0,T}^2$ be a sequence converging to some $\bar{y}$ in $\mathcal{L}_{0,T}^2$. Using a standard Gronwall argument one can show existence of a constant $c > 0$ such that

$$\|\widehat{x}_k[y_j] - \widehat{x}_k[\bar{y}]\|_{\mathcal{H}_0^T} \leq c \|y_j - \bar{y}\|_{\mathcal{L}_{0,t}^2},$$

implying continuity of the first mapping. The claim for the second mapping follows from the continuous embedding $\mathcal{H}_0^t \hookrightarrow C([0,t]; \mathbb{R}^n)$ and the continuity of the point evaluation.

Now let $y_1, y_2 \in \mathcal{L}_{0,t}^2$. We find

$$\sup_{s \in [0,t]} |r_k[y_j](s) - r_k[\bar{y}](s)|$$

$$\leq \sup_{s \in [0,t]} \int_0^s |\|y_j - C\widehat{x}_k[y_j]\|_{Q_k^{-1}}^2 - \|\bar{y} - C\widehat{x}_k[\bar{y}]\|_{Q_k^{-1}}^2| \, \mathrm{d}\tau$$

$$\leq \sup_{s \in [0,t]} \int_0^s |\langle y_j - \bar{y} + C(\widehat{x}_k[y_j] - \widehat{x}_k[\bar{y}]), y_j + \bar{y} + C(\widehat{x}_k[y_j] + \widehat{x}_k[\bar{y}]) \rangle| \, \mathrm{d}\tau$$

$$\leq c \left(\|y_j - \bar{y}\|_{\mathcal{L}_{0,t}^2} + \|\widehat{x}_k[y_j] - \widehat{x}_k[\bar{y}]\|_{\mathcal{L}_{0,t}^2} \right) \left(\|y_j + \bar{y}\|_{\mathcal{L}_{0,t}^2} + \|\widehat{x}_k[y_j] + \widehat{x}_k[\bar{y}]\|_{\mathcal{L}_{0,t}^2} \right),$$

where the τ dependence of the integrand is suppressed in the notation and c is an appropriate constant. Due to the convergence of y_j to $\bar{y}$ the right hand side converges to zero ensuring continuity of the third mapping. The assertion for the fourth mapping follows directly. □

With this result at hand we are able to show that the estimator depends continuously on the output in an L^2 sense.

Proposition A.6. *Let $\theta \in (0, \infty)$ and $(y_j) \subset \mathcal{L}_{0,T}^2$ be a sequence converging to some $\bar{y}$ in $\mathcal{L}_{0,T}^2$. Then it holds that*

$$\forall t \in [0,T] \qquad \|\widehat{x}_\theta[y_j](t) - \widehat{x}_\theta[\bar{y}](t)\| \to 0 \text{ as } j \to \infty,$$

$$\|\widehat{x}_\theta[y_j] - \widehat{x}_\theta[\bar{y}]\|_{\mathcal{L}_{0,T}^1} \to 0 \text{ as } j \to \infty,$$

Proof. Assume for now that $\widehat{x}_\theta[y_j]$ converges to $\widehat{x}_\theta[\bar{y}]$ pointwise for all $t \in [0,T]$. We establish an integrable upper bound for $\|\widehat{x}_\theta[y_j](t)\|$ to apply the dominated convergence. The convergence of the sequence (y_j) implies the existence of a constant C such that $\|y_j\|_{\mathcal{L}_{0,T}^2} \leq C$ for all $j \in \mathbb{N}$. Utilizing Lemma A.3 we obtain that for all t it holds

$$\|\widehat{x}_\theta[y_j](t)\| \leq \sqrt{\widehat{C}\left(\|x_0\|^2 + \|y_j\|_{\mathcal{L}_{0,T}^2}^2 \right)} \leq \sqrt{\widehat{C}\left(\|x_0\|^2 + C^2 \right)}. \tag{A.6}$$

Since the right hand side is constant, the integrable upper bound is found and the dominated convergence theorem yields the desired convergence in $\mathcal{L}_{0,T}^p$.

It remains to show the pointwise convergence. Let $t \in [0,T]$ be fixed. By (A.6) the family $\{\widehat{x}_\theta[y_j](t)\}_{j \in \mathbb{N}}$ is bounded in $\mathbb{R}^n$. Therefore each arbitrary subsequence admits another subsequence (also denoted by $\widehat{x}_\theta[y_j](t)$) that converges to some $\bar{x} \in \mathbb{R}^n$. Next we recall from Lemma 3.11 that for all j

$$\widehat{x}_\theta[y_j](t) = \left(\sum_{k=1}^N e^{\theta \|\widehat{x}_\theta[y_j](t) - \widehat{x}_k[y_j](t)\|_{P_k(t)}^2 + \theta r_k[y_j](t)} P_k(t) \right)^{-1}$$

$$\sum_{k=1}^N e^{\theta \|\widehat{x}_\theta[y_j](t) - \widehat{x}_k[y_j](t)\|_{P_k(t)}^2 + \theta r_k[y_j](t)} P_k(t) \widehat{x}_k[y_j](t)$$

holds. Considering the convergent subsequence and utilizing the regularity results established in Lemma A.5 and the continuity of the squared norm, the exponential, and the

matrix inverse we find that the right hand side converges and we obtain

$$\bar{x} = \left(\sum_{k=1}^{N} e^{\theta \|\bar{x} - \widehat{x}_k[\bar{y}](t)\|_{P_k(t)}^2 + \theta r_k[\bar{y}](t)} P_k(t) \right)^{-1} \sum_{k=1}^{N} e^{\theta \|\bar{x} - \widehat{x}_k[\bar{y}](t)\|_{P_k(t)}^2 + \theta r_k[\bar{y}](t)} P_k(t) \widehat{x}_k[\bar{y}](t).$$

By Lemma 3.11 it follows that $\bar{x} = \widehat{x}_\theta[\bar{y}](t)$ and we have shown that any subsequence of $\widehat{x}_\theta[y_j](t)$ admits a subsequence converging to $\widehat{x}_\theta[\bar{y}](t)$. By the subsequence principle the pointwise convergence follows and the proof is concluded. $\qquad \square$

Next we show convergence of the time derivatives of an approximating sequence.

Lemma A.7. *Let $\theta \in (0, \infty)$ and $(y_j) \subset C([0,T]; \mathbb{R}^r)$ be a sequence converging to some $\bar{y}$ in $\mathcal{L}_{0,T}^2$. Define $\bar{x}$ as*

$$\bar{x}(t) = D_x G[\theta, \bar{y}](t, \widehat{x}_\theta[\bar{y}](t))^{-1} \partial_t G[\theta, \bar{y}](t, \widehat{x}_\theta[\bar{y}](t)). \tag{A.7}$$

Then $\bar{x} \in \mathcal{L}_{0,T}^1$ and (y_j) admits a subsequence (denoted by (y_j)) such that

$$\int_0^T \|\dot{\widehat{x}}_\theta[y_j](t) - \bar{x}(t)\| \, \mathrm{d}t \to 0 \qquad \text{for } j \to \infty. \tag{A.8}$$

Proof. The assertion that $\bar{x} \in \mathcal{L}_{0,T}^1$ follows from Lemma A.4 (ii) and (iii). Turning to (A.8) we first note that by assumption we have convergence of y_j in $\mathcal{L}_{0,T}^2$ and by Lemma A.5 we have convergence of $\dot{\widehat{x}}_k[y_j]$ in $\mathcal{L}_{0,T}^2$. Hence [10, Cor. 2.32] allows the extraction of a subsequence such that y_j and $\dot{\widehat{x}}_k[y_j]$ converge pointwise almost everywhere. Further by Lemma A.5 and Lemma A.7 we have pointwise convergence of $\widehat{x}_k[y_j]$, $r_k[y_j]$, and $\widehat{x}_\theta[y_j]$. With (A.3), (A.4), and (A.5) it follows that for almost all $t \in [0,T]$ it holds

$$\dot{\widehat{x}}_\theta[y_j](t) = -D_x G[\theta, y_j](t, \widehat{x}_\theta[y_j](t))^{-1} \partial_t G[\theta, y_j](t, \widehat{x}_\theta[y_j](t)) \to \bar{x}(t) \qquad \text{for } j \to \infty.$$

An integrable function dominating $\dot{\widehat{x}}_\theta[y_j]$ can be derived using the boundedness of $\|y_j\|_{\mathcal{L}_{0,T}^2}$, Lemma A.3, Lemma A.5, and the continuity of P_k and $\dot{P}_k$. The assertion follows with dominated convergence. $\qquad \square$

We now identify the limit $\bar{x}$ as the weak derivative of $\widehat{x}_\theta[\bar{y}]$ to show the asserted regularity of the estimator.

Proof of Theorem 3.12(i)

Proof. We begin by showing the result for $p = 1$. Hence let $\bar{y} \in \mathcal{L}_{0,T}^2$. From Lemma A.3 we know that $\widehat{x}_\theta[\bar{y}] \in \mathcal{L}_{0,T}^\infty$. To obtain the assertion it remains to show that its weak derivative exists and is integrable. We denote $\bar{x} \in \mathcal{L}_{0,T}^1$ as in (A.7). Then the density of $C([0,T]; \mathbb{R}^r)$ in $\mathcal{L}_{0,T}^2$ together with Proposition A.6 and Lemma A.7 ensures existence of a sequence $(y_j) \subset C([0,T]; \mathbb{R}^r)$ such that

$$\widehat{x}_\theta[y_j] \to \widehat{x}_\theta[\bar{y}] \qquad \text{for } j \to \infty \qquad \text{in } \mathcal{L}_{0,T}^1, \tag{A.9}$$

$$\dot{\widehat{x}}_\theta[y_j] \to \bar{x} \qquad \text{for } j \to \infty \qquad \text{in } \mathcal{L}_{0,T}^1. \tag{A.10}$$

Now for any test function $\varphi \in C_0^\infty([0,T]; \mathbb{R}^n)$ we obtain

$$\int_0^T \langle \widehat{x}_\theta[\bar{y}](t), \dot{\varphi}(t) \rangle \, \mathrm{d}t = \lim_{j \to \infty} \int_0^T \langle \widehat{x}_\theta[y_j](t), \dot{\varphi}(t) \rangle \, \mathrm{d}t$$

$$= -\lim_{j \to \infty} \int_0^T \langle \dot{\widehat{x}}_\theta[y_j](t), \varphi(t) \rangle \, \mathrm{d}t = -\int_0^T \langle \bar{x}(t), \varphi(t) \rangle \, \mathrm{d}t,$$

where the first equality is justified by (A.9), the second equality follows from Theorem 3.12(iii) and partial integration, and the third equality holds due to (A.10). By definition of the weak derivative we have shown that $\bar{x} \in \mathcal{L}^1_{0,T}$ is the weak derivative of $\widehat{x}_\theta[\bar{y}]$, in short $\dot{\widehat{x}}_\theta[\bar{y}] = \bar{x}$. It follows that $\widehat{x}_\theta[\bar{y}] \in W^{1,1}(0,T;\mathbb{R}^n)$. Further, by the definition of $\bar{x}$ the formula for the weak derivative given in (3.14) is proven to hold for square integrable outputs.

It remains to show the result for $2 \leq p < \infty$. For $y \in \mathcal{L}^{2p}_{0,T} \hookrightarrow \mathcal{L}^2_{0,T}$ the first part of this proof ensures weak differentiability in $\mathcal{L}^1_{0,T}$, and (3.14) for its derivative. From this formula it can subsequently be confirmed that $\dot{\widehat{x}}_\theta[y] \in \mathcal{L}^p_{0,T}$ and the assertion is shown. $\qquad\square$

[*] Institute of Mathematics and Scientific Computing, University of Graz, A-8010 Graz, Austria

[†] 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