Estimating States and Model Uncertainties Jointly by a Sparsity Promoting UKF

Ricarda-Samantha Götte, Julia Timmermann

Heinz Nixdorf Institute, Paderborn University, Paderborn, Germany
(e-mail: {rgoette,julia.timmermann}@hni.upb.de)

Abstract: State estimation when only a partial model of a considered system is available remains a major challenge in many engineering fields. This work proposes a joint, square-root unscented Kalman filter to estimate states and model uncertainties simultaneously by linear combinations of physics-motivated library functions. Using a sparsity promoting approach, a selection of those linear combinations is chosen and thus an interpretable model can be extracted. Results indicate a small estimation error compared to a traditional square-root unscented Kalman filter and exhibit the enhancement of physically meaningful models.

Keywords: joint estimation, unscented transform, Kalman filter, sparsity, data-driven, compressed sensing

1. INTRODUCTION

Estimating relevant quantities is an underlying key issue of many engineering problems, e.g. estimating a vehicle’s position or velocity for tracking control in driver assistance systems. Since its introduction by Kalman (1960), the Kalman filter has become one of the most prevalent estimators due to its optimality for linear systems. Naturally, numerous filter derivations for nonlinear dynamics were introduced thereafter, e.g. the extended Kalman filter (EKF), already used for the Apollo missions to the moon, and the unscented Kalman filter (UKF) by Julier and Uhlmann (1997). However, an accurate model is crucial for the estimation performance regardless of which model-based filter is deployed. Recalling complex plants or systems, the process of modeling does not become difficult or time-consuming only due to parameter identification but due to complicated partial dynamics, often containing highly nonlinear parts and being challenging to capture. Though many powerful filter techniques exist in control engineering, they strongly rely on the quality of the plant’s model and will provide faulty estimates if it is inaccurate. Therefore, a filter that is able to cope with a low-quality model and its uncertainty while also returning a small estimation error and a comprehensive model would be desirable.

Regarding parameters uncertainty, concepts like joint and dual estimation are sufficiently known, e.g. Wan et al. (1999), but only few studies on filter approaches dealing with complex model uncertainties and their detection have been carried out so far. Recently, only Khajenejad et al. (2021), Buisson-Fenet et al. (2021) and Kullberg et al. (2021) investigated situations when just a partial model of the system is known. Khajenejad et al. (2021) utilize a complex geometrical approach to estimate an upper and lower barrier for the approximated state to lie in between. The states and unknown partial dynamics are estimated very closely, but the approach does not enable an interpretable representation of the unknown dynamics. In Buisson-Fenet et al. (2021), a high gain observer is deployed to estimate states that are then utilized to infer a Gaussian process model of the plant. Subsequently, the model is used within the observer and vice versa, affecting and correlating with each other. Only Kullberg et al. (2021) strive for a similar strategy as we do: They propose linear combinations of radial basis functions (RBFs) with compact support to approximate the unknown partial dynamics. This is carried out by extending the state with the linear combinations’ parameters to infer dominant RBFs. Established within an EKF, the authors focus on an efficient computation by altering the EKF formulas rather than identifying and interpreting the unknown partial dynamics while simultaneously estimating the states.

In this work, we start from a problem formulation similar to Khajenejad et al. (2021), Buisson-Fenet et al. (2021) and Kullberg et al. (2021) and define a joint state, that additionally contains parameters of the linear combinations for the partial dynamics to be approximated. However, we use a square-root UKF (SQ-UKF) combined with a physics-motivated library for the linear combinations, simplifying the inference of physical coherences. Finally, arguing similar to Brunton et al. (2016) that most systems in nature and technology can be characterized by a small amount of dynamical terms, we identify a method from compressed sensing (CS), see e.g. Carmi et al. (2010), for our field of research and adjust it to estimate the full state and an interpretable representation of the system’s unknown partial dynamics. In particular, we deploy the idea of pseudo measurements within a SQ-UKF to enforce the linear combinations’ parameters being sparse. In contrast to the method in Brunton et al. (2016) and related approaches such as LASSO (Tibshirani (1996)), the model selection is carried out online while estimating the system’s states. Therefore, our main contributions are the following:

• Joint estimation of states and model uncertainties within a SQ-UKF, using physics-based and experts’ knowledge,
2. PROBLEM FORMULATION

For state estimation we wish to obtain the state $\mathbf{x}_k \in \mathbb{R}^{n_x}$ at time step $k$ of a real system, whose dynamics are modeled by the following discrete system with control input $u_k \in \mathbb{R}$:

$$
\mathbf{x}_{k+1} = f(\mathbf{x}_k, u_k) + \mathbf{w}_k,
$$

$$
\mathbf{y}_k = h(\mathbf{x}_k, u_k) + \mathbf{v}_k. \tag{1}
$$

The real system is assumed to be observable and can be observed by measurements $\mathbf{y}_k \in \mathbb{R}^{n_y}$. Here, $f$ and $h$ denote the dynamical and observational models, respectively, with additive process and measurement noise $\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$ and $\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ assumed to be Gaussian distributed. However, if the underlying model (1) lacks of relevant quantities, the estimator based upon that model will not be able to capture the non-measurable states closely or will diverge.

Nonetheless, we could assume the unknown partial dynamics $g(\mathbf{x}, u)$ as a mapping that can be approximated by linear combinations of suitable functions $\Psi_i$, stored in a library $\Psi$, and evaluated by parameters $\Theta \in \mathbb{R}^{n_{\Theta}}$. These functions contain prior knowledge, resulting from observational experience or experts’ hypotheses, so that a close approximation by $g(\mathbf{x}, u) \approx \Theta^T \cdot \Psi(\mathbf{x}, u)$ can be expected. Presuming that the real system’s order and its states can be defined, $\Psi$’s minimal number of elements is $n_x + 2$, as it contains at least constants, the states and the control input. However, $\Theta$’s values are initially unknown and need to be calculated. Regarding parameter identification the joint estimation scheme is sufficiently known for estimating states and parameters simultaneously during the filter iterations, see e.g. Wan et al. (1999) and van der Merwe and Wan (2001). Therefore, we extend the model (1) to estimate unknown partial dynamics by defining the parameters $\Theta$ within a joint state vector $\tilde{x}_k = (x_k^T, \Theta_k^T) \in \mathbb{R}^{n}$ with $\tilde{n} = n_x + n_{\Theta}$:

$$
\tilde{x}_{k+1} = \left( f(\mathbf{x}_k, u_k, \Theta_k) + \mathbf{w}_k^x \right)
$$

$$
\Theta_k + \mathbf{w}_k^\Theta,
$$

$$
\mathbf{y}_k = h(\mathbf{x}_k, u_k) + \mathbf{v}_k. \tag{2}
$$

Remark that $\Theta$’s evolution is modeled by stationary dynamics and additive Gaussian noise $\mathbf{w}_k^\Theta \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_\Theta)$, whereas it holds $\mathbf{w}_k^x \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_x)$. Hence, the process covariance matrix of the joint state is a block matrix $\tilde{\mathbf{Q}} = \text{blkdiag}(\mathbf{Q}_x, \mathbf{Q}_\Theta)$ with zeros elsewhere. Note within the scope of this paper the unknown partial dynamics are considered as one dimensional, namely $g(\mathbf{x}, u)$, and $h$ is linear. To receive a small estimation error $e_k = \hat{x}_k - \tilde{x}_k$, any Kalman filter relies on the minimum-mean-squared-error to estimate $\hat{x}_k$ as stated e.g. in Gibbs (2011):

$$
\arg\min_{\hat{x}} \frac{1}{2} \mathbb{E}[(\hat{x} - \tilde{x})^T(\hat{x} - \tilde{x})]. \tag{3}
$$

Thus, the lift towards a higher dimensional state $\tilde{x}$, including the unknown partial dynamics identifying parameters $\Theta$, affects the optimization problem (3) by the cost of increased model complexity and computational burden.

3. JOINT ESTIMATION BY SPARSITY

After the model’s reformulation to estimate states and unknown partial dynamics simultaneously, this section copes with the model’s application within an SQ-UKF and presents a sparse solution as remedy towards the computational burden and the degrees of freedom due to the high dimensional state.

3.1 Square-root unscented Kalman filter

The unscented Kalman filter (UKF), established by Julier (2002); Julier and Uhlmann (1997), is a powerful nonlinear filter design method. In contrast to the extended Kalman filter (EKF), which linearizes nonlinear dynamics along a single sample point by the first order Taylor series, the UKF seeks to estimate the moments of the nonlinear probability distribution by multiple sample points. Therefore, it is independent of the analytical differentiation calculations the EKF requires. This enables a higher estimation accuracy when systems with strong nonlinearities are considered. However, the calculation of the covariance matrix in every filter iteration for $2n + 1$ samples, called sigma points, is computational expensive. But, as presented in van der Merwe (2004); van der Merwe and Wan (2001), the UKF’s efficiency and its numerical stability can be improved by using the square-root covariance matrix $\mathbf{S}_k$ instead of $\mathbf{P}_k = \mathbf{S}_k \mathbf{S}_k^T$, leading to the square-root UKF (SQ-UKF) algorithm.

In Alg. 1 the SQ-UKF’s procedure is outlined. To ensure a symmetric distribution during the unscented transformation, the UKF’s parameters $\lambda = \alpha^2(n + \kappa) - \tilde{n}$ and $\eta = \sqrt{n + \lambda}$ define the weights $W_i^{(c)}$ and $W_i^{(m)}$:

$$
W_i^{(m)} = \frac{\lambda}{\lambda + \bar{n}}, \quad W_i^{(c)} = \frac{\lambda}{\lambda + \bar{n}} + 1 - \alpha^2 + \beta,$$

$$
W_i^{(m)} = W_i^{(c)} = \frac{1}{2(n + \kappa)} \quad \text{for } i = 1, \ldots, 2\tilde{n}. \tag{4}
$$

The parameters $\alpha$, $\beta$ and $\kappa$ need to be chosen beforehand, since they reflect how the nonlinear distribution is covered. In Nielsen et al. (2021) it has been shown that the choice of these is problem-dependent and has a significant influence on the filter performance. However, for the scope of this paper, we choose $\alpha = 10^{-3}, \beta = 2$ and $\kappa = 0$. 

- a sparse strategy to promote interpretable models by methods of CS,
- demonstration and discussion on relevant application systems.

The paper proceeds as follows: In Sec. 2 the challenge of estimating states while only a partial model of the considered system being available is discussed. Hence, the strategy of a sparse, joint SQ-UKF, that estimates states and model uncertainties simultaneously, is outlined and clarified by notions of CS in Sec. 3. The proposed filter is then demonstrated for experiments in Sec. 4. A short discussion and outlook in Sec. 5 concludes this work.

Notation: $\cdot$ denotes the joint state vector or relating variables, such as the covariance matrix. A subscript $ij(k-1)$ represents the variable at time step $k$ with observations up to time step $k-1$. Bold variables refer to vectors or matrices.
Algorithm 1 SQ-UKF

1: **Initialize:** \( \alpha, \beta, \kappa, \tilde{\alpha}_0 = \mathbb{E}[\tilde{x}_0] \),
\( S_0 = \text{chol}(\mathbb{E}(\tilde{x}_0 - \tilde{\alpha}_0)^T(\tilde{x}_0 - \tilde{\alpha}_0)) \)

2: for \( k \in \{1, \ldots, \infty\} \) 
   Calculation of sigma points:
3: \( X_{k-1} = [\tilde{x}_{k-1} + \eta S_{k-1} \tilde{x}_{k-1} - \eta S_{k-1}] \)
4: Prediction:
5: \( \tilde{x}_k = \sum_{i=0}^{2n} W_i^{(m)} X_{i,k|k-1} \)
6: \( S_k = q r \left( \begin{bmatrix} W_i^{(c)} (X_{i,k|k-1} - \tilde{x}_k) \end{bmatrix} \sqrt{Q} \right) \)
7: \( S_y_k = \text{cholupdate}(S_k, X_{0,k} - \tilde{x}_k, W_0^{(c)}) \)
8: \( Y_{k|k-1} = h(X_{k|k-1}, u_{k-1}) \)
9: Correction:
10: \( S_{yk} = \text{cholupdate}(S_{yk}, Y_{0,k} - \tilde{y}_k, W_0^{(c)}) \)
11: \( P_{xy} = \sum_{i=0}^{2n} W_i^{(c)} [X_{i,k|k-1} - \tilde{x}_k] [Y_{i,k|k-1} - \tilde{y}_k]^T \)
12: \( K_k = P_{xy} P_{yy}^{-1} \)
13: \( \tilde{x}_k = \tilde{x}_k + K_k (y_k - \tilde{y}_k) \)
14: \( U = K_k S_{yk} \)
15: \( S_k = \text{cholupdate}(S_k, U, -1) \)

If the SQ-UKF is utilized in its original form (Algo. 1) without any modifications for the additional degrees of freedom resulting from \( \theta \), it has been observed that the filter tends to become locally unstable or divergent, which may lead to faulty state estimates. Depending on the number of design degrees \( n_\theta \), the calculation of \( \theta \) comes close to an underdetermined optimization problem. Hence, a strategy to choose only a selection of \( \theta \) while retaining the design degrees is required. Therefore, notions from the field of compressed sensing are adopted.

### 3.2 Compressed sensing

In the domain signal processing methods of compressed sensing cope with the reconstruction of signals \( x = \Omega s \) with \( \Omega \in \mathbb{R}^{n \times n} \) a suitable basis matrix and \( s \in \mathbb{R}^n \) the sparse representation of the signal \( x \in \mathbb{R}^n \). Based on early works of Donoho (2006) and Candes et al. (2006), the linear underdetermined problem with the measurement \( y = H x \) and its measurement matrix \( H \)

\[
\text{argmin}_{s} \| H \Omega s - y \|_2 + \lambda \| s \|_1, \tag{5}\]

which seeks for a sparse solution \( s \) and is non-convex due to the \( \ell_1 \)-norm leading to a high combinatorial effort, can be transformed into

\[
\text{argmin}_{s} \| H \Omega s - y \|_2 + \tilde{\lambda} \| s \|_1. \tag{6}\]

Here, \( 0 < \tilde{\lambda} \ll 1 \) denotes the sparsity balancing factor. The minimization problem (6) is now convex and provides a solution if sufficiently many measurements \( y \) are given and the rows of \( H \) are independent of \( \Omega \)'s columns, for details see Brunton and Kutz (2019); Carni et al. (2010). Further, the problem can be reformulated into a constrained optimization problem

\[
\text{argmin}_{\tilde{s}} \| H \Omega \tilde{s} - y \|_2, \tag{7}\]

\[
s.t. \| \tilde{s} \|_1 \leq \epsilon, \]

with \( 0 < \epsilon \ll 1 \).

### 3.3 Promoting sparsity by pseudo measurements

Considering the degrees of freedom introduced by the parameters \( \theta \), an accurate estimation by the filter problem (3) cannot be guaranteed any more as discussed in the previous sections. However, using the same argument as in Brunton et al. (2016), most phenomena in nature or technology can be represented rather by a small amount of dominating terms. Therefore, only a selection of \( \Psi \) is needed to capture the system's behavior. To enforce this assumption, the filter problem (3) is reformulated by adding a term that encourages a sparse vector \( \tilde{\theta} \). Adopting the notions (5) and (6) from Sec. 3.2 to efficiently propagate a sparse \( \theta \) within the iterative filter Algo. 1, the minimization problem can be formulated by

\[
\text{argmin}_{\tilde{s}} \frac{1}{2} \mathbb{E}[ \tilde{x} - \tilde{x} ]^T(\tilde{x} - \tilde{x}) + \tilde{\lambda} \| \tilde{x}_{n+1|n} \|_1 \tag{8}\]

where \( \tilde{\lambda} \in \mathbb{R} \) as the sparsity promoting term. However, how to incorporate this approach within the iterative structure of the SQ-UKF is not straight-forward. Recalling the previous section and adjusting (7) towards the problem (8) yields a sparse mean-squared-error for joint estimation

\[
\text{argmin}_{\tilde{s}} \frac{1}{2} \mathbb{E}[ \tilde{x} - \tilde{x} ]^T(\tilde{x} - \tilde{x}) \tag{9}\]

\[
s.t. \| \tilde{s}_{n+1|n} \|_1 \leq \epsilon. \]

By interpreting the soft constraint as a fictitious, additional measurement \( 0 = \| \tilde{x}_{n+1|n} \|_1 - \epsilon \) it can be utilized during estimation schemes, see e.g. Julier and LaViola (2007), Carni et al. (2010), Hage et al. (2020). Different approaches, e.g. as shown in Julier and LaViola (2007), exist how to incorporate the pseudo measurement within a filter iteration. For this paper, we define a pseudo measurement function \( \tilde{h}(\tilde{x}) \) by

\[
\tilde{h}(\tilde{x}) = \| \tilde{x}_{n+1|n} \|_1 - \epsilon, \tag{10}\]

where \( \epsilon \sim \mathcal{N}(0, R_{pm}) \) now represents the fictitious, additive measurement noise and controls the tightness of the constraint. Since the SQ-UKF is based on the unscented transformation, the nonlinear observational model (10) can be easily applied without any further adaption in contrast to its use within an EKF, e.g. Julier and LaViola (2007).

In Algo. 2 the procedure of the proposed sparsity promoting, joint SQ-UKF (J-SQ-UKF) is depicted. Within every filter iteration \( k \) additional updates for the estimated state \( \tilde{x}_k \) and its square-root covariance \( S_k \) might be enclosed depending on the sparsity of \( \theta_k \). If the number \( n_{\theta, act} \) of non-sparse elements, that are greater than the sparsity barrier \( \tilde{\lambda} \), is exceeded (line 18), a filter update is calculated again but now using the pseudo measurement (10) as fictitious observational model instead of the actual observational model \( h \) (line 19). This step is followed by additional checks and updates. The sparsity promoting...
loop terminates either when the maximal iteration of pseudo measurements \( j = N \) is reached or the number of allowed non-sparse elements is less or equal \( n_{\theta, \text{act}} \). In contrast to Sec. 3.2 online estimation is possible because of the evolving covariance that contains information about past measurements. Since with every filter iteration \( k \) the dominant parameters of \( \theta \) could completely change due to the sparsity promoting loop and the degrees of freedom at the beginning of every iteration \( k \), the stability of the filter algorithm may be easily affected as well as the model’s interpretability might get lost. Hence, to avoid local instabilities or divergent behavior, a soft switching method for \( \theta \) is enforced afterwards (line 23) by balancing the state estimate before and after the pseudo measurement correction. This guarantees a more continuous estimation of the dominant parameters, resulting in a more continuous state estimation itself. Different methods of soft switching could be applied, e.g., batch updates, but herein a simple weighted approach is employed by a factor \( 0 < \gamma < 1 \). It is recommended to enforce a higher emphasis on the newly corrected estimate after the pseudo measurement loop.

Algorithm 2 J-SQ-UKF

15: 
16: \( S_k = \text{cholupdate}(S_k, U, -1) \)
17: 
18: \textbf{Initialize:} \( \hat{x}, h, N, n_{\theta, \text{act}}, \gamma, j = 1 \)
19: \( S_{pm,0} = S_k, \hat{x}_{pm,0} = \hat{x}_k \)
20: \textbf{while} \( \#(\theta_j \theta_j > \lambda) > n_{\theta, \text{act}} \) \textbf{and} \( j < N \)
21: \( \hat{x}_{pm,j}, S_{pm,j} \leftarrow \text{SQ-UKF with} \)
22: \( (\hat{x}_{pm,j-1}, S_{pm,j-1}, f, h, \tilde{Q}, R_{pm}) \)
23: \textbf{end}
24: 
25: \( S_{k,\text{final}} = S_{pm,j} \)
26: 
27: \( \hat{x}_{k,\text{final}} = \hat{x}_{pm,j}, \)
28: \( \hat{x}_{k,\text{final}} = (1 - \gamma)[\hat{x}_{pm,j}]_{n_x+1:n} + \gamma[\hat{x}_k]_{n_x+1:n} \)

3.4 Curse of dimension

As generally known, any Kalman filter’s performance depends strongly on the choice of its covariance matrices. While the measurement covariance \( R \in \mathbb{R}^{m \times m} \) often is easier to determine by taking a closer look at the plant’s measurements, the initial square-root covariance \( S_0 \) and the process covariance \( \tilde{Q} \in \mathbb{R}^{\tilde{n} \times \tilde{n}} \) remain a major challenge in filter design, see e.g., Chen et al. (2021); Nielsen et al. (2021). When promoting sparsity within the SQ-UKF even an additional covariance \( R_{pm} \) comes along which determines the pseudo measurements’ performance. With increasing numbers of states \( n_x \) and possibly uncertain partial dynamics, expressed by \( n_{\theta} \) numerous parameters \( \theta \), not only the curse of dimension but also the difficulty to initialize the performance parameters become obvious. Common approaches in filter design include manual tuning, grid or random search, which are unsatisfying in terms of effort and benefit while possibly resulting in local minima. Lately, Bayesian optimization becomes the state-of-the-art method for hyperparameter optimization and could be applied herein when offline estimation is pursued. For the scope of this paper, we focus on online estimation and therefore set the parameters as \( P_0 = \tilde{Q} = \text{blkdiag}(10^{-6}I_{n_x}, 10^{-4}I_{n_{\theta}}), R = 10^{-4}I_{m}, \) and \( R_{pm} = 1 \).

4. RESULTS

The proposed filter algorithm is now demonstrated for two application examples. Compared to a traditional SQ-UKF, the advantages of the sparse, joint SQ-UKF filter become very clear. Within each application example both filters start with a corrupted initial state \( \hat{x}_0 \neq \bar{x}_0 \). For the J-SQ-UKF the parameters \( \theta_0 \) are initialized equally with small values. Within all experiments discussed in this paper it holds \( n_{\theta, \text{act}} = 3 \) and \( \lambda = 0.1 \).

4.1 Duffing oscillator

The Duffing oscillator is a common example for nonlinear dynamical systems. Its dynamics are described in the following with parameters \( p = (-1, 3, 0.1)^T \):

\[
\frac{d^2}{dt^2} x_2 = -p_3 x_2 - p_1 x_1 - p_2 x_1^3 + u \quad (11)
\]

\[
y = x_1.
\]

Here, we assume the nonlinear term \( g(x, u) = p_2 x_1^3 \) as unknown and formulate the joint model, lacking of the mentioned term, analogously to the description in Sec. 2. Formulating (11) as a discrete model and using the explicit Euler scheme and the library \( \Psi_1 = (1, x_1, x_2, x_3, \sin(x_2), x_1^3, x_1, x_2, \cos(x_1), u)^T \) to evaluate the model, we can deploy it within the J-SQ-UKF.

As a result, we see through the red dashed line in Fig. 1 that the J-SQ-UKF is able to cope with the faulty model and to deliver a correct estimate. This is due to the selection of the true dynamics, which is illustrated in the top plot in Fig. 2 by \( \theta \)’s magnitude over time. Here, the line, that represents \( \theta_6 \) and therefore indicates that \( \Psi_6(x, u) = x_1^3 \) is mainly present, is of the highest magnitude. Further, it is able to deal with incorrect initial states and indicates a fast transient behavior towards the
modeling effort of (12) lies mainly in modeling the friction $M$ term. The model that completely lacks of the term $2022$; Schön et al. (2022). Here, the joint model (2) is alternatives to interpret the system’s behavior, namely in detecting either the correct partial dynamics or finding Therefore, the proposed method appears to be confident with $\tilde{\Psi}$ in approximating the system’s behavior, and still delivers an accurate state estimate, though a small degradation can be observed. When utilizing a library without the true term $g$, e.g. $\Psi_2 = (1, x_1, x_2, x_2^2, \sin(x_2), x_1 x_2, \cos(x_1), u)^T$ or $\Psi_3 = (1, x_1, x_2, x_2^2, \sin(x_2), x_1^2, x_2 x_1, \cos(x_1), u)^T$, the J-SQ-UKF is yet able to estimate the true dynamics. As illustrated in Fig. 1 and 2, the filter clearly extracts other physical terms that are close to the true missing term $g$ and still delivers an accurate state estimate, though a small degradation compared to the J-SQ-UKF with $\Psi_1$ can be observed. Therefore, the proposed method appears to be confident in detecting either the correct partial dynamics or finding alternatives to interpret the system’s behavior, namely $\Psi_2(x, u) = x_1$ or $\Psi_6(x, u) = x_2^2$. These interpretations can then be utilized for further model usage or refinement.

4.2 Golf robot

The golf robot, shown in Fig. 3, is a test bench at our laboratory to investigate the capability of machine learning methods on real world problems. After extensive modeling effort the robot’s dynamics are described by

$$
\dot{x} = \begin{pmatrix}
    x_2 \\
    J^{-1} (-m g \dot{\theta} \sin(x_1) - M_F + 4u)
\end{pmatrix},
$$

$$
M_F = dx_2 + 2r\mu \arctan(10^3 x_2) \pi^{-1} \left| m x_2^2 a + m \tilde{g} \cos(x_1) \right|,
$$

with $\tilde{g}$ as gravity constant and the same parameters $p = (m, a, d, J, r, \mu)^T$ utilized as in Götte and Timmermann (2022); Schön et al. (2022). Here, the joint model (2) is the model that completely lacks of the term $M_F$. Since the modeling effort of (12) lies mainly in modeling the friction terms, this assumption is reasonable and close to real world situations. Further, the library is defined as $\Psi(x, u) = (1, x_1, x_2, x_2^2, x_1^3, \sin(x_2), \cos(x_1), u)^T$ and the model is again discretized and evaluated with the explicit Euler scheme.

As illustrated in Fig. 4 the J-SQ-UKF provides a close estimation of the robot’s states and even shows a faster transient behavior than the traditional filter that delivers faulty angular velocities. The corresponding dominant parameters $\theta$ are depicted in Fig. 5 and reveal that more than one term of the library is relevant to approximate $g$, in fact these are $\Psi_1, \Psi_2, \Psi_3, \Psi_4$. Further, it is observed that these terms seem to vary in time which is reasonable since friction forces are dependent on velocity and are often highly nonlinear. It is also noteworthy that the dominant terms are close to the ones which have been identified so far by extensive modeling effort in model (12) or correlate to those by factor, e.g. $x_1 \approx \cos(x_1)$ when $x_1$ is small. However, extracting a more accurate representation for those dynamics requires a follow-up step like a modal transformation when there is more than one identified dominant term $\Psi_i, \Psi_j$.

4.2 Golf robot

The golf robot, shown in Fig. 3, is a test bench at our laboratory to investigate the capability of machine learning methods on real world problems. After extensive modeling effort the robot’s dynamics are described by

$$
\dot{x} = \begin{pmatrix}
    x_2 \\
    J^{-1} (-m g \dot{\theta} \sin(x_1) - M_F + 4u)
\end{pmatrix},
$$

$$
M_F = dx_2 + 2r\mu \arctan(10^3 x_2) \pi^{-1} \left| m x_2^2 a + m \tilde{g} \cos(x_1) \right|,
$$

with $\tilde{g}$ as gravity constant and the same parameters $p = (m, a, d, J, r, \mu)^T$ utilized as in Götte and Timmermann (2022); Schön et al. (2022). Here, the joint model (2) is the model that completely lacks of the term $M_F$. Since the modeling effort of (12) lies mainly in modeling the friction terms, this assumption is reasonable and close to real world situations. Further, the library is defined as $\Psi(x, u) = (1, x_1, x_2, x_2^2, x_1^3, \sin(x_2), \cos(x_1), u)^T$ and the model is again discretized and evaluated with the explicit Euler scheme.

As illustrated in Fig. 4 the J-SQ-UKF provides a close estimation of the robot’s states and even shows a faster transient behavior than the traditional filter that delivers faulty angular velocities. The corresponding dominant parameters $\theta$ are depicted in Fig. 5 and reveal that more than one term of the library is relevant to approximate $g$, in fact these are $\Psi_1, \Psi_2, \Psi_3, \Psi_4$. Further, it is observed that these terms seem to vary in time which is reasonable since friction forces are dependent on velocity and are often highly nonlinear. It is also noteworthy that the dominant terms are close to the ones which have been identified so far by extensive modeling effort in model (12) or correlate to those by factor, e.g. $x_1 \approx \cos(x_1)$ when $x_1$ is small. However, extracting a more accurate representation for those dynamics requires a follow-up step like a modal transformation when there is more than one identified dominant term $\Psi_i, \Psi_j$.

5. DISCUSSION AND OUTLOOK

The proposed method has shown for several application examples its power to jointly estimate states and model uncertainties. We could observe two cases: Either the library $\Psi$ contains the unknown partial dynamics $g$ and the algorithm extracts the correct term or close approximations of it, or it provides a suggestion for $g$ based on its library terms. Further, it clearly outperforms the traditional filter when only an incomplete model of the system is available. However, the interpretability has not been automated so far and needs the user to interpret and extract the dominant $\theta_i$ and its corresponding terms by sight.
Additionally, parameters like the number of allowed sparse elements \( n_{\theta,\text{act}} \), the sparsity barrier \( \lambda \) and the terms of \( \Psi \) currently have to be set by the user. Correlations between those parameters are to be expected. Moreover, it has not yet been investigated if there exists a dependency of \( n_{\theta} \) and \( n_{x} \), limiting possibly the number of \( \Psi 's \) elements. Nonetheless, the sparse, joint SQ-UKF allows a deeper insight into the system’s dynamics, while simultaneously providing sufficient estimates. This insight could be then concretized offline by a modal transformation method, e.g. proper orthogonal decomposition, to automate the model interpretations and utilize an updated model within the filter. However, experiments have shown to be real-time capable and encourage the usage within online estimation. Therefore, future research focuses on the relation of \( n_{\theta} \) and \( n_{x} \) and on a higher dimensional \( g \).

ACKNOWLEDGMENT

This work was developed in the junior research group DART (Datengebundene Methoden in der Regelungstechnik), University Paderborn, and funded by the Federal Ministry of Education and Research of Germany (BMBF - Bundesministerium für Bildung und Forschung) under the funding code 01IS20052. The responsibility for the content of this publication lies with the authors.

REFERENCES

Brunton, S.L. and Kutz, J.N. (2019). *Data-driven Science and Engineering*. Cambridge University Press.

Brunton, S.L., Proctor, J.L., and Kutz, J.N. (2016). Sparse identification of nonlinear dynamics with control (sindyc). In *IFAC-PapersOnline*, volume 49, 710–715.

Buisson-Fenet, M., Morgenthaler, V., Trimpe, S., and Di Meglio, F. (2021). Joint state and dynamics estimation with high-gain observers and gaussian process models. In *IEEE Control Systems Letters*, volume 5, 1627–1632.

Candès, E.J., Romberg, J., and Tao, T. (2006). Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. In *IEEE Transactions on Information Theory*, volume 52, 489–509.

Carni, A., Gurfil, P., and Kanevsky, D. (2010). Methods for sparse signal recovery using kalman filtering with embedded pseudo-measurement norms and quasi-norms. In *IEEE Transactions on Signal Processing*, volume 58, 2405–2409.

Chen, Z., Heckman, C., Julier, S., and Ahmed, N. (2021). Time dependence in kalman filter tuning. Submitted and accepted by FUSION 2021.

Donoho, D.L. (2006). Compressed sensing. In *IEEE Transactions on Information Theory*, volume 52, 1289–1306.

Gibbs, B.P. (2011). Least-squares estimation, kalman filtering, and modeling. Wiley.

Götte, R.-S. and Timmermann, J. (2022). Composed physics- and data-driven system identification for non-autonomous systems in control engineering. Accepted for: 2022 3rd International Conference on Artificial Intelligence, Robotics and Control (AIRC 2022).

Hage, D.A., Heredia Conde, M., and Llofled, O. (2020). Sparse signal recovery via kalman-filter-based \( \ell_1 \) minimization. In *Signal Processing*, volume 171, Article 107487.

Julier, S.J. (2002). The scaled unscented transformation. In *2002 IEEE American Control Conference (ACC)*, 4555–4559.

Julier, S.J. and Uhlmann, J.K. (1997). New extension of the kalman filter to nonlinear systems. In *Proceedings Signal Processing, Sensor Fusion, and Target Recognition VI*, volume 3068.

Julier, S.J. and LaViola, J.J. (2007). On kalman filtering with nonlinear equality constraints. In *IEEE Transactions on Signal Processing*, volume 55, 2774–2784.

Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. In *ASME Journal of Basic Engineering*, volume 82, 35–45.

Khajenijnad, M., Jin, Z., and Yong, S.Z. (2021). Interval observers for simultaneous state and model estimation of partially known nonlinear systems. In *2021 American Control Conference (ACC)*, 2848–2854.

Kullberg, A., Skog, I., and Hendeby, G. (2021). Online joint state inference and learning of partially unknown state-space models. In *IEEE Transactions on Signal Processing*, volume 69, 4149–4161.

Nielson, K., Svahn, C., Rodriguez-Deniz, H., and Hendeby, G. (2021). Ukf parameter tuning for local variation smoothing. In *2021 IEEE International Conference, 1–8*.

Schön, O., Götte, R.-S., and Timmermann, J. (2022). Multi-objective physics-guided recurrent neural networks for identifying non-autonomous dynamical systems. Accepted for 14th IFAC Workshop on Adaptive and Learning Control Systems (ALCOS 2022).

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. In *Journal of the Royal Statistical Society*, volume 58, 267–288.

van der Merwe, R. (2004). Sigma-Point Kalman Filters or Probabilistic Inference in Dynamic State-Space Models. Oregon Health & Science University.

van der Merwe, R. and Wan, E.A. (2001). The square-root unscented kalman filter for state and parameter-estimation. In *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 3461–3464.

Wan, E.A., van der Merwe, R., and Nelson, A.T. (1999). Dual estimation and the unscented transformation. In *Advances in Neural Information Processing, 12*.

Fig. 5. Course of \( \theta \) corresponding to library \( \Psi \): Dominant \( \theta_i \) indicate that \( \Psi_1(x,u) = 1, \Psi_2(x,u) = x_1 \) and \( \Psi_4(x,u) = x_2^2 \) are the best terms to approximate \( g \). The barrier due to \( \lambda = 0.1 \) is illustrated by the transparent gray rectangle.