On the Observability of General Nonlinear Gaussian State Space Models Using Discrete Distributional Approximations

ARIANE HANEBECK
CLAUDIA CZADO

We consider arbitrary nonlinear stochastic discrete-time state space models (SSMs) with time-invariant parameters and nonadditive Gaussian disturbances. Given an observation trajectory, the goal is to obtain an estimate of the augmented state (the underlying state trajectory and the time-invariant parameters of the model). A numerical approach to checking this type of observability is given, and a quantitative assessment of the degree of observability is provided. In general, no global statements for all observation trajectories can be made on the observability of nonlinear SSMs. However, we can find regions of the state-observation space (consisting of all possible observation trajectories and corresponding state trajectories) in which an estimate of the augmented state can be obtained. This is achieved by approximating the continuous distribution of observation trajectories and state trajectories using an optimal discrete distribution. The associated locations of the point masses are called design values. For these design values, we can then check whether the augmented state can be recovered. We could also use random realizations of the observation trajectory. However, when using design values, a smaller number of considered observation trajectories is required to achieve a good coverage of the space. We illustrate our approach to checking observability for different specifications of discrete-time SSMs in univariate and multivariate settings.

Manuscript received October 9, 2023; revised July 8, 2024; released for publication January 10, 2025

Refereeing of this contribution was handled by Paolo Braca.

A. Hanebeck is with the Applied Mathematical Statistics, Technische Universität München, 85748 Garching, Germany (e-mail: Ariane.Hanebeck@tum.de).

C. Czado is with the Applied Mathematical Statistics and Munich Data Science Institute (MDSI), Technische Universität München, 85748 Garching, Germany (e-mail: czado@tum.de).

Comment: Preliminary considerations of this paper have been presented at the 25th International Conference on Information Fusion (FUSION 2022) in Linköping in [18]. The present paper expands the method from Gaussian models (which express Gaussian random variables as a function of desired parameters and a Gaussian disturbance term) to general nonlinear Gaussian state space models. The algorithm uses the distributional approximations presented in [32].

I. INTRODUCTION

In time series analysis, state space models (SSMs) are a common tool to model the dynamic behavior of observed multivariate data $\mathbf{Z}_t, t = 1, \ldots, T$, using multivariate latent states $\mathbf{X}_t, t = 0, \ldots, T$. As in [12], we consider arbitrary nonlinear stochastic discrete-time SSMs with unknown time-invariant parameters $\boldsymbol{\Omega}$ and nonadditive Gaussian disturbances $\boldsymbol{\epsilon}_t, \boldsymbol{\eta}_t, t = 1, \ldots, T$, of the form

$$\begin{aligned} &\text{Observation equation: } \mathbf{Z}_t = g(\boldsymbol{\Omega}, \mathbf{X}_t, \boldsymbol{\epsilon}_t) \\ &\text{State equation: } \mathbf{X}_t = h(\boldsymbol{\Omega}, \mathbf{X}_{t-1}, \boldsymbol{\eta}_t) \text{ for } t = 1, \ldots, T. \end{aligned} \tag{1}$$

More details on this model can be found in Section II.A.

For a given observation trajectory $\mathbf{z}_1, \ldots, \mathbf{z}_T$, the goal is to obtain an estimate of the augmented state $\boldsymbol{\Theta}$ consisting of initial state $\mathbf{x}_0$, state trajectory $\mathbf{x}_1, \ldots, \mathbf{x}_T$, and parameters $\boldsymbol{\Omega}$. Simulations show that the joint maximum a posteriori (MAP) estimator is biased in this situation. Instead, we choose the marginal MAP estimator to recover the augmented state. To avoid high-dimensional integration for the marginalization, a Bayesian Markov chain Monte Carlo (MCMC) setup is chosen.

In general, no global statements for all observation trajectories can be made on the observability of nonlinear SSMs [11], [13], [24]. This means that an estimate of the augmented state cannot necessarily be recovered for all observation trajectories (see [18, p. 2] for an intuitive example). We call a given SSM observable if the marginal MAP estimator exists, i.e., if the marginal MAP estimate can be recovered for at least one realization of the observation trajectory. Section II.B gives the exact theoretical definition of observability.

Our contribution is a novel approach to checking this type of observability in practice. As no global statements for all observation trajectories can be made, all possible realizations of the observation trajectory and corresponding state trajectory have to be analyzed. This is in general not possible, which is why we only consider a set of appropriately selected disturbance realizations called design disturbances. For obtaining design disturbances, we approximate the distribution of the Gaussian disturbances by a discrete distribution. For that, the approach proposed in [32] is used, which minimizes a distance measure similar to Cramér-von Mises between the continuous and the discrete distribution. The locations of the point masses of this discrete distribution are called the design disturbances.

Since observability properties depend on the parameters $\boldsymbol{\Omega}$, the initial state $\mathbf{x}_0$, and the number of time steps T, observability is always considered for fixed values of $\boldsymbol{\Omega}$, $\mathbf{x}_0$, and T. Each design disturbance leads to a corresponding design state trajectory and observation trajectory by plugging it into the corresponding model. The detailed construction of these design values is given in Section III.A.

The design disturbances are created to cover the space of disturbances homogeneously. Our simulations

show that they allow us to choose a smaller number of considered observation trajectories compared to using random values to obtain a satisfactory coverage of the space. Furthermore, they are determined offline before checking for observability and hence do not lead to additional computational effort.

For all chosen design observation trajectories, we execute two tasks. The first task is used to obtain information about the type of extrema of the posterior density (see Section III.B). In the second task, we then check whether the augmented state can be recovered by considering the marginal MAP estimate of the posterior density (see Section III.C). This tells us in which regions of the state-observation space the marginal MAP estimator is defined for the fixed time-invariant parameters $\mathbf{\Omega}$, initial state $\mathbf{x}_0$, and T.

There are many applications of SSMs in the literature. For example, [6] uses SSMs to describe flight control systems. In [36], the models are used to facilitate biological and biomedical modeling, while [17] use SSMs to analyze epidemiological data. Furthermore, [22] and [23] propose copula-based SSMs to capture the time dynamics in air pollution. Regarding observability, several definitions and approaches to checking it are developed. One approach is to use Kalman's observability matrix [3], [17], [20]. However, this approach is only used for linear SSMs. A definition for nonlinear observability based on the so-called indistinguishability is given in [20]. The models that are considered allow for deterministic input only in the state equation, which is described by a differential equation. The suggested approach to check this kind of observability is an extension of Kalman's observability matrix using Lie derivatives. An algebraic approach for nonlinear polynomial SSMs of this type is proposed in [15]. In [28], it is noted that the existing approaches checking for observability cannot identify for which observation trajectories the underlying parameters can be obtained. For models with uncertainties, which are restricted to bounded intervals, [28] solve this problem by using interval analysis. An overview of different observability definitions found in the literature is given in [37] and more recently [3]. In contrast to our approach, the literature imposes restrictions on the considered models, such as linearity or a specific form. Furthermore, only a deterministic input is allowed, in most cases only in the observation equation.

The remainder of the paper is organized as follows: In Section II, we introduce the type of model we consider and the theoretical definition of observability based on the existence of the marginal MAP estimator. Section III describes our proposal to check this definition of observability in practice using design observation trajectories. In Section IV, we illustrate how this approach can be used to decide upon observability of different examples of Gaussian SSMs. Section V concludes the paper and gives an outlook for possible adaptions and for future work to extend the approach to non-Gaussian SSMs. In Section VI, an overview over the most important definitions is given.

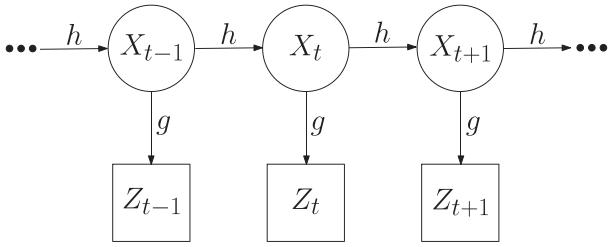

Figure 1. A general SSM for $d = p = 1$.

II. SSMS AND OBSERVABILITY NOTATIONS

A. Stochastic SSMs

We describe the model (1) in more detail now.

1) Definitions: The random variable $\mathbf{Z}_t \in \mathbb{R}^d$ denotes the observation at time t, while $\mathbf{X}_t \in \mathbb{R}^p$ represents the (latent) state at time t. This latent state is an underlying variable driving the behavior of the observations. The state equation describes the evolution of the unobserved latent state over time depending on the random disturbances $\boldsymbol{\eta}_t, t = 1, \ldots, T$, while the observation equation describes how the observation $\mathbf{Z}_t$ at time t is defined, given the latent state $\mathbf{X}_t$ and the random disturbance $\boldsymbol{\epsilon}_t$ at time t. We define the random observation trajectory from time point 1 to T by $\mathcal{Z}_T = (\mathbf{Z}_1^\top, \ldots, \mathbf{Z}_T^\top)^\top \in \mathbb{R}^{T \cdot d}$ with realization $z_T = (\mathbf{z}_1^\top, \ldots, \mathbf{z}_T^\top)^\top$. The random trajectory of the latent state is defined by $\mathcal{X}_T = (\mathbf{X}_1^\top, \ldots, \mathbf{X}_T^\top)^\top \in \mathbb{R}^{T \cdot p}$ with realization $\chi_T = (\mathbf{x}_1^\top, \ldots, \mathbf{x}_T^\top)^\top$. Figure 1 illustrates an SSM for $d = p = 1$.

We assume the initial value $\mathbf{x}_0$ to be fixed and unknown and denote by $\mathbf{x}_0^{true}$ the true underlying value. Furthermore, $\mathbf{\Omega}$ is a vector of unknown time-invariant parameters, where $\mathbf{\Omega}^{true}$ is the true underlying value of $\mathbf{\Omega}$.

For the disturbances $(\boldsymbol{\epsilon}_t)_{t=1,\ldots,T}$ and $(\boldsymbol{\eta}_t)_{t=1,\ldots,T}$, we assume that they are serially independent and independent of each other at all time points [12, Ch. 9] with $\boldsymbol{\epsilon}_t \sim \mathcal{N}_d(\mathbf{0}, R_t)$ and $\boldsymbol{\eta}_t \sim \mathcal{N}_p(\mathbf{0}, Q_t)$, where $R_t \in \mathbb{R}^{d \times d}$ and $Q_t \in \mathbb{R}^{p \times p}$ are known covariance matrices.

2) Underlying Parameters: Consider an SSM of the form (1) with fixed true underlying values $\mathbf{\Omega}^{true}$ and $\mathbf{x}_0^{true}$. Given an observation trajectory z_T from that model, the underlying unknown true parameters are $\mathbf{\Theta}^{true} = (\mathbf{\Omega}^{true}, \mathbf{x}_0^{true}, \chi_T)$. The values $\mathbf{\Omega}^{true}$ and $\mathbf{x}_0^{true}$ always stay the same, but the trajectory χ_T of the state is changing with every realization of the disturbances $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t, t = 1, \ldots, T$, and hence every observation trajectory z_T. Consequently, there is not one common true underlying value of χ_T that could be compared for different z_T's. Thus, we treat χ_T as a nuisance parameter vector, and the parameters of interest are $\mathbf{\Omega}$ and $\mathbf{x}_0$. Following [5] and [25], the vector $\mathbf{\Theta} = (\mathbf{\Omega}, \mathbf{x}_0, \chi_T)$ of unknown

parameters is called the *augmented state,* consisting of the time-invariant parameters, the initial state, and the state trajectory. Even though we primarily consider the parameters $\mathbf{\Omega}$ and $\mathbf{x}_0$ in our proposed approach, we are also interested in χ_T in applications to allow for predictive simulation.

For better readability, we write $\mathbf{\Theta} = (\mathbf{\Omega}, \mathbf{x}_0, \chi_T)$ instead of $\mathbf{\Theta} = (\mathbf{\Omega}^\top, \mathbf{x}_0^\top, \chi_T^\top)^\top$. This notation will be used throughout the paper, e.g, also for z_T, $\mathcal{Z}_T$, χ_T, and $\mathcal{X}_T$.

3) Posterior Density: The joint posterior density of (1) is defined by

$$\Pi(\mathbf{\Theta}|z_T) \propto \ell\left(\mathbf{\Theta}|z_T\right) \cdot p(\mathbf{\Theta}) = \left[\prod_{t=1}^{T} f_z\left(\mathbf{z}_t|\mathbf{x}_t, \mathbf{\Omega}\right)\right] \cdot \left[\prod_{t=1}^{T} p_x\left(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{\Omega}\right)\right] \cdot p_0(\mathbf{x}_0) \cdot p_{\mathbf{\Omega}}(\mathbf{\Omega}). \tag{2}$$

The corresponding log-posterior density is defined by $\pi(\mathbf{\Theta}|z_T) = \log(\Pi(\mathbf{\Theta}|z_T))$. Here, $\ell\left(\mathbf{\Theta}|z_T\right)$ is the likelihood of $\mathbf{\Theta}$ given z_T, $p(\mathbf{\Theta})$ is the prior of the parameter vector $\mathbf{\Theta}$, p_0 is the prior for $\mathbf{x}_0$, $\prod_{t=1}^{T} p_x\left(\cdot|\mathbf{x}_{t-1}, \mathbf{\Omega}\right)$ is the prior for χ_T given $\mathbf{x}_0$, and $p_{\mathbf{\Omega}}$ is the prior for $\mathbf{\Omega}$. The function $f_z\left(\mathbf{z}_t|\mathbf{x}_t, \mathbf{\Omega}\right)$ is the density of $\mathbf{Z}_T|\{\mathbf{X}_t = \mathbf{x}_t, \mathbf{\Omega}\}$ defined by the observation equation.

The prior p_0 might depend on the true underlying value of $\mathbf{x}_0$ if information about it is available. The prior $\prod_{t=1}^{T} p_x\left(\cdot|\mathbf{x}_{t-1}, \mathbf{\Omega}\right)$ of χ_T might either just be the product of the densities $f_x\left(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{\Omega}\right)$ of $\mathbf{X}_t|\{\mathbf{X}_{t-1} = \mathbf{x}_{t-1}, \mathbf{\Omega}\}$ defined by the state equation, i.e., $p_x = f_x$, or more information might be available on χ_T that can be incorporated. For example, the support of f_x could be restricted to obtain p_x.

Example 1. *If the true underlying value Ω_i^{true} of a time-invariant parameter component $\Omega_i \in (-1, 1)$ of $\mathbf{\Omega}$ is known, we define the restricted prior*

$$p^{res}_{\Omega_i^{true}}(\Omega_i) = \begin{cases} U(0,1) & \text{for } \Omega_i^{true} > 0, \\ U(-1,0) & \text{for } \Omega_i^{true} < 0. \end{cases} \tag{3}$$

If Ω_i^{true} is not known, prior expert knowledge might be available on the sign of Ω_i.

4) Models Linear in State: One type of SSM, which is often studied, is *linear in the states* $\mathbf{X}_t$ and is given by

$$\begin{aligned} \mathbf{Z}_t &= A(\mathbf{\Omega}) \cdot \mathbf{X}_t + B(\mathbf{\Omega}) \cdot \boldsymbol{\epsilon}_t\,, \\ \mathbf{X}_t &= C(\mathbf{\Omega}) \cdot \mathbf{X}_{t-1} + D(\mathbf{\Omega}) \cdot \boldsymbol{\eta}_t \end{aligned} \tag{4}$$

for $t = 1, \ldots, T$, and $A(\mathbf{\Omega}) \in \mathbb{R}^{d\times p}$, $B(\mathbf{\Omega}) \in \mathbb{R}^{d\times d}$, $C(\mathbf{\Omega}) \in \mathbb{R}^{p\times p}$, and $D(\mathbf{\Omega}) \in \mathbb{R}^{p\times p}$ for unknown $\mathbf{\Omega}$. Note that the matrices $A(\mathbf{\Omega})$, etc, do not need to be linear in the parameters $\mathbf{\Omega}$ in order for the model to be linear in the states. The disturbance distributions are defined as before.

We introduce one example of this type now.

Example 2. *The model we consider with $d = p = 1$ is of the form*

$$\begin{aligned} &Z_t = aX_t + \sqrt{1-a^2} \cdot \epsilon_t\,,\ X_t = aX_{t-1} + \sqrt{1-a^2} \cdot \eta_t\,, \\ &t = 1, \ldots, T, \end{aligned} \tag{5}$$

with $\epsilon_t \sim \mathcal{N}(0,1)$ i.i.d. and $\eta_t \sim \mathcal{N}(0,1)$ i.i.d. independent. The parameter $\Omega = a \in (-1,1)$ is unknown, so the parameters of interest are a and x_0 with unknown nuisance parameters $x_1, \ldots, x_T$. Using $Z_t|\{X_t = x_t, a\} \sim \mathcal{N}(ax_t, 1-a^2)$, $X_t|\{X_{t-1} = x_{t-1}, a\} \sim \mathcal{N}(ax_{t-1}, 1-a^2)$, we determine the posterior for $\mathbf{\Theta} = (a, x_0, \chi_T)$ as

$$\Pi(\mathbf{\Theta}|z_T) \propto \left[\prod_{t=1}^{T} \frac{1}{\sqrt{1-a^2}} \cdot e^{-\frac{(z_t - ax_t)^2}{2(1-a^2)}}\right] \cdot \left[\prod_{t=1}^{T} \frac{1}{\sqrt{1-a^2}} \cdot e^{-\frac{(x_t - ax_{t-1})^2}{2(1-a^2)}}\right] \cdot p^{res}_{a^{true}}(a). \tag{6}$$

We use a noninformative prior for x_0. For known a^{true}, the prior of a is chosen by $p^{res}_{a^{true}}(a)$, as defined in (3). For small values of $|a|$, this restriction is necessary as otherwise the posterior becomes bimodal. For larger values of $|a|$, the restriction might not always be necessary. Furthermore, in general, the integral over all parameters $a \in (-1,1), x_t \in (-\infty, \infty), t = 0, \ldots, T$, of (6) with uniform prior on $(-1,1)$ for a might not exist.

The log-posterior is of the form

$$\pi(\mathbf{\Theta}|z_T) = -\frac{1}{2}\frac{1}{1-a^2}\left[\sum_{t=1}^{T}(z_t - ax_t)^2 + \sum_{t=1}^{T}(x_t - ax_{t-1})^2\right] - T\log(1-a^2) + \log(p^{res}_{a^{true}}(a)).$$

B. Theoretical Definition of Observability

For different values of $\mathbf{\Omega}^{true}$, $\mathbf{x}_0^{true}$, and T, the observability properties of a given SSM can vary [20]. For this reason, observability is always investigated for a fixed initial value $\mathbf{x}_0^{true}$ and fixed time-invariant parameters $\mathbf{\Omega}^{true}$. Hence, if certain values of the parameters $(\mathbf{\Omega}, \mathbf{x}_0)$ are of special interest, they can be studied in detail. Additionally, different numbers of time steps T can be considered.

To define our notion of observability, we first consider a fixed but arbitrary observation trajectory z_T with corresponding state trajectory χ_T of model (1) using $\mathbf{x}_0^{true}$ and $\mathbf{\Omega}^{true}$, i.e.,

$$\begin{aligned} \mathbf{Z}_t &= g(\mathbf{\Omega}^{true}, \mathbf{X}_t, \boldsymbol{\epsilon}_t) \text{ for } t = 1, \ldots, T, \\ \mathbf{X}_t &= h(\mathbf{\Omega}^{true}, \mathbf{X}_{t-1}, \boldsymbol{\eta}_t) \text{ for } t = 2, \ldots, T, \\ \mathbf{X}_1 &= h(\mathbf{\Omega}^{true}, \mathbf{x}_0^{true}, \boldsymbol{\eta}_1) \end{aligned}$$

and study the posterior $\Pi(\mathbf{\Theta}|z_T)$ defined in (2) using priors, which are specified specifically for each model. Then, two tasks are carried out. In the first task, we obtain information about the type of extrema of $\Pi(\mathbf{\Theta}|z_T)$

by checking if the joint MAP estimate can be obtained. In the second task, we consider the marginal MAP estimate. The existence of the marginal MAP estimate in the second task depends on the type of extrema of the joint posterior, which is investigated in the first task.

1) Existence of Joint MAP Estimate: We want to investigate if the joint MAP estimate

$$\hat{\boldsymbol{\Theta}}(z_T) = (\hat{\boldsymbol{\Omega}}(z_T), \hat{\mathbf{x}}_0(z_T), \widehat{\chi}_T(z_T)) = \underset{\boldsymbol{\Theta}}{\operatorname{argmax}}\, \Pi(\boldsymbol{\Theta}|z_T) \tag{7}$$

exists, from which we obtain information about the type of extrema of the posterior. For this, we maximize the log-posterior π, i.e., $\hat{\boldsymbol{\Theta}}(z_T) = (\hat{\boldsymbol{\Omega}}(z_T), \hat{\mathbf{x}}_0(z_T), \widehat{\chi}_T(z_T)) = \underset{\boldsymbol{\Theta}}{\operatorname{argmax}}\, \pi(\boldsymbol{\Theta}|z_T)$.

Given an observation trajectory z_T, the ideal case is to have a unique maximum of the corresponding log-posterior $\pi(\cdot|z_T)$. However, there are three further cases to be considered.

To define the three cases, we first have to define the notion of a ridge. For a function $f : \mathbb{R}^N \to \mathbb{R}$, a ridge is a curve consisting of local maxima in $N-1$ dimensions. A point $\mathbf{u}_0 \in \mathbb{R}^N$ is such a local maximum if $f(\mathbf{u}) < f(\mathbf{u}_0)$ for $\mathbf{u} \in \mathcal{S} \subset \mathbb{R}^N$ with $\dim(\mathcal{S}) = N-1$. We define the set of local maxima on a ridge by $\mathcal{S}_R$ and say a ridge is of constant height if $f(\mathbf{u}_1) = f(\mathbf{u}_2)$ for all $\mathbf{u}_1 \neq \mathbf{u}_2 \in \mathcal{S}_R$ and of variable height otherwise. For an example of a variable height ridge, see Figure 4, which shows that a ridge is in general a nonlinear feature of the posterior density. Examples where ridges are encountered are given in [29] and [30].

The three cases are given as follows:

1. No maximum: In the case of no finite maximum, a maximum can only be found on the boundary, and it is not possible to recover the true augmented state. This might happen if we have a ridge of variable height with no local maxima within the boundaries.

2. Finitely many maxima: In the case of two or more distinct maxima with the same log-posterior value, no unique joint MAP estimate $\hat{\boldsymbol{\Theta}}(z_T)$ of $\boldsymbol{\Theta}$ given z_T can be recovered. However, it might be possible to find a unique maximum $\operatorname{argmax}_{\boldsymbol{\Theta}} \pi(\boldsymbol{\Theta}|z_T)$ when constraining the parameter space using restricting priors on the parameters. In general, the value of the posterior in the distinct local maxima is not the same. Then, a constraint of the parameter space around the global maximum is usually carried out. This also leads to a posterior with a unique maximum.

3. Infinitely many maxima: In the case where there is a ridge of constant height of the log-posterior, the function does not have a distinct maximum $\operatorname{argmax}_{\boldsymbol{\Theta}} \pi(\boldsymbol{\Theta}|z_T)$. In that case, it is not possible to recover the true augmented state. However, it might be possible to find estimates for functions of parameters.

Example 3 (Finitely many maxima). *Consider a slightly changed version of the model in (5) with $d = p = 1$ given by*

$$Z_t = aX_t + \sqrt{1-a^2} \cdot \epsilon_t\,,\ X_t = a^2 X_{t-1} + \sqrt{1-a^4} \cdot \eta_t\,,\ t = 1, \ldots, T,$$

with $\epsilon_t \sim \mathcal{N}(0,1)$ i.i.d. and $\eta_t \sim \mathcal{N}(0,1)$ i.i.d. independent for unknown $a \in (-1,1)$. Then, given an observation trajectory z_T with underlying parameter values $(a^{true}, x_0^{true}, \chi_T)$, the parameter values $(-a^{true}, -x_0^{true}, -\chi_T)$ lead to the same posterior value. Hence, we can use the prior $p^{res}_{a^{true}}(a)$ to constrain the parameter space and to find the unique underlying parameter values $(a^{true}, x_0^{true}, \chi_T)$.

Example 4 (No maximum). *For $\epsilon_t \sim \mathcal{N}(0,1)$, $\eta_t \sim \mathcal{N}(0,1)$ i.i.d. independent, consider*

$$Z_t = aX_t + \epsilon_t\,,\ X_t = X_{t-1} + \eta_t\,,\ t = 1, \ldots, T. \tag{8}$$

We are interested in $a \in (-1,1)$, the initial state x_0, and the latent state trajectory χ_T, given an observation trajectory z_T. The log-posterior is given in (13). For a given z_T, the posterior has a ridge of variable height for small values of T. More details are given in Section IV.

A model of the form (4) is often called a linear SSM. However, it is only linear in the states $\mathbf{X}_t$. We call an SSM *linear* if it is linear in both the states $\mathbf{X}_t$ and the time-invariant parameters $\boldsymbol{\Omega}$.

For a linear model and a given observation trajectory z_T, the derivatives $\frac{\mathrm{d}}{\mathrm{d}\Theta_i}\pi(\boldsymbol{\Theta}|z_T)$, $i = 1, \ldots, |\boldsymbol{\Theta}|$, of the log-posterior form a system of linear equations for the parameters in $\boldsymbol{\Theta}$. Hence, it is only possible to have none, one unique, or infinitely many maxima. The second case of having finitely many maxima cannot occur [11], [20].

Consider, for example, the model given in (8), which is of the form (4) and hence linear in the states $X_t, t = 0, \ldots, T$. If the parameter a is known and not of interest, it is also a linear model because it is linear in all unknown parameters $X_0, \ldots, X_T$. However, if a is of interest and $\Omega = a$, the derivative of the log-posterior in (13) with respect to a is $\sum_{t=1}^{T}(z_t x_t - a x_t^2)$, which is obviously not linear in x_t. Hence, this model is nonlinear and only linear in the states X_t.

2) Existence of Marginal MAP Estimate: This approach of considering the marginal MAP estimate is also used in [22], [23] to estimate the unknown parameters. The differences between joint and marginal MAP estimates are discussed in [14, Ch. 13] and [27].

Consider the marginal posterior densities

$$\begin{aligned}\Pi(\Omega_j|z_T) &= \int \Pi(\boldsymbol{\Omega}, \mathbf{x}_0, \chi_T|z_T)\,\mathrm{d}\boldsymbol{\Omega}_{-j}\,\mathrm{d}\mathbf{x}_0\,\mathrm{d}\chi_T,\\ &\qquad j = 1,\dots,|\boldsymbol{\Omega}|,\\ \Pi(x_{0,i}|z_T) &= \int \Pi(\boldsymbol{\Omega}, \mathbf{x}_0, \chi_T|z_T)\,\mathrm{d}\boldsymbol{\Omega}\,\mathrm{d}\mathbf{x}_{0,-i}\chi_T,\\ &\qquad i = 1,\dots,p,\\ \Pi(x_{t,i}|z_T) &= \int \Pi(\boldsymbol{\Omega}, \mathbf{x}_0, \chi_T|z_T)\,\mathrm{d}\boldsymbol{\Omega}\,\mathrm{d}\mathbf{x}_0\,\mathrm{d}\mathbf{x}_{t,-i}\,\mathrm{d}\chi_{T,-t},\\ &\qquad i = 1,\dots,p, t = 1,\dots,T,\end{aligned} \tag{9}$$

where $\mathbf{x}_t = (x_{t,1},\dots,x_{t,p})$, $\mathbf{x}_{t,-i} = (x_{t,1},\dots,x_{t,i-1},x_{t,i+1},\dots,x_{t,p}) \in \mathbb{R}^{p-1}$ for $t = 0,\dots,T$, $\boldsymbol{\Omega}_{-j} = \boldsymbol{\Omega} \setminus \{\Omega_j\} \in \mathbb{R}^{|\boldsymbol{\Omega}|-1}$, $\chi_{T,-t} = (\mathbf{x}_1,\dots,\mathbf{x}_{t-1},\mathbf{x}_{t+1},\dots,\mathbf{x}_T) \in \mathbb{R}^{p(T-1)}$.

We maximize the marginal posterior densities defined in (9) over Ω_j, $j = 1,\dots,|\boldsymbol{\Omega}|$, and $x_{t,i}, t = 0,\dots,T, i = 1,\dots,p$, respectively, and obtain $\hat{\Omega}_j(z_T)_{\mathrm{mar}} = \operatorname{argmax}_{\Omega_j}\Pi(\Omega_j|z_T)$ and $\hat{x}_{t,i}(z_T)_{\mathrm{mar}} = \operatorname{argmax}_{x_{t,i}}\Pi(x_{t,i}|z_T)$. The marginal MAP estimate is then denoted by $\hat{\boldsymbol{\Theta}}(z_T)_{\mathrm{mar}} = (\hat{\boldsymbol{\Omega}}(z_T)_{\mathrm{mar}}, \hat{\mathbf{x}}_0(z_T)_{\mathrm{mar}}, \widehat{\chi_T}(z_T)_{\mathrm{mar}})$. If one of the three cases of (1) no maximum, (2) finitely many maxima, or (3) infinitely many maxima occurs and the joint MAP estimate cannot be recovered, then it might also not be possible to recover the marginal MAP estimate. How to obtain the marginal MAP estimate in practice is discussed in Section III.C.

3) The Definition of Observability: Until now, we have considered a fixed but arbitrary observation trajectory z_T with underlying χ_T. Now, we define the notion of observability for a given SSM of the form (1).

Definition 1. *Define the* state-observation space *as the space of all possible realizations of the observation trajectory $\mathcal{Z}_T$ and corresponding state trajectory $\mathcal{X}_T$, i.e., for $\mathbf{x}_0^{true} \in \mathbb{R}$, $\boldsymbol{\Omega}^{true} \in \mathbb{R}^{|\boldsymbol{\Omega}|}$, define*

$$\begin{aligned}&SO(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T) =\\ &\{\textit{Realizations } (\chi_T, z_T) \textit{ of } (\mathcal{X}_T, \mathcal{Z}_T):\\ &\quad \mathbf{Z}_t = g(\boldsymbol{\Omega}^{true}, \mathbf{X}_t, \boldsymbol{\epsilon}_t), t = 1,\dots,T,\\ &\quad \mathbf{X}_t = h(\boldsymbol{\Omega}^{true}, \mathbf{X}_{t-1}, \boldsymbol{\eta}_t) \textit{ for } t = 2,\dots,T,\\ &\quad \mathbf{X}_1 = h(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, \boldsymbol{\eta}_t),\\ &\boldsymbol{\epsilon}_t \sim \mathcal{N}_d(0, R_t) \textit{ i.i.d., and } \boldsymbol{\eta}_t \sim \mathcal{N}_p(0, Q_t)\\ &\quad \textit{i.i.d. independent}\}.\end{aligned} \tag{10}$$

Remark 1. *The joint MAP estimator and the marginal MAP estimator of a nonlinear SSM are not necessarily defined for all elements of $SO(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T)$.*

The phenomenon of Remark 1 can also be found in the literature on SSMs, where the state equation is described by a differential equation, for which the concepts can be transferred to our type of model. In [24], it is stated that *the observability for nonlinear systems is, in general, not only a local property but also depends on the input of the system*. The authors of [13] talk about *bad inputs*, and the same arguments can be found in [11]. These "bad" inputs lead to observation trajectories for which the augmented state cannot be recovered. In the setup of this paper, they correspond to "bad" realizations of the disturbances, inducing realizations in $SO(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T)$, for which the augmented state cannot be recovered. This leads to the following definitions:

Definition 2 (Existence of the joint MAP estimator). *If the log-posterior $\pi(\cdot|z_T)$ does not have a distinct maximum for a given z_T, no unique joint MAP estimate $\hat{\boldsymbol{\Theta}}(z_T)$ of $\boldsymbol{\Theta}$ can be recovered. Let $\mathcal{E}(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T) \subset SO(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T) \subset \mathbb{R}^{T(p+d)}$ be the set of realizations z_T and corresponding χ_T such that $\hat{\boldsymbol{\Theta}}(z_T) = \operatorname{argmax}_{\boldsymbol{\Theta}}\pi(\boldsymbol{\Theta}|z_T)$ can be uniquely recovered. If $\mathcal{E}(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T) = \emptyset$, the estimator $\hat{\boldsymbol{\Theta}}(\mathcal{Z}_T)$ does not exist. If $\mathcal{E}(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T) \neq \emptyset$, the estimator*

$$\hat{\boldsymbol{\Theta}}(\mathcal{Z}_T) = \begin{cases}\hat{\boldsymbol{\Theta}}(z_T), & (\mathcal{X}_T, \mathcal{Z}_T) = (\chi_T, z_T)\\ & \quad \in \mathcal{E}(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T)\\ \textit{undefined}, & \textit{otherwise}\end{cases}$$

exists but is not necessarily defined for every realization $(\chi_T, z_T) \in SO(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T)$.

If the joint MAP estimator exists, it is often biased, and a Bayesian analysis for all parameters jointly is not the best approach [14, Ch. 13.4]. However, the existence of $\hat{\boldsymbol{\Theta}}(\mathcal{Z}_T)$ and the type of extrema of the posterior provide information for the marginal MAP approach.

Definition 3 (Observability based on marginal MAP estimator). *Let $\mathcal{E}^{\mathrm{mar}}(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T) \subset SO(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T)$ be the set of realizations (χ_T, z_T) such that the marginal MAP estimate $\hat{\boldsymbol{\Theta}}(z_T)_{\mathrm{mar}}$ can be uniquely recovered. If $\mathcal{E}^{\mathrm{mar}}(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T) = \emptyset$, $\hat{\boldsymbol{\Theta}}(\mathcal{Z}_T)_{\mathrm{mar}}$ does not exist, and we call the model unobservable. If $\mathcal{E}^{\mathrm{mar}}(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T) \neq \emptyset$, we call the model observable. Then,*

$$\hat{\boldsymbol{\Theta}}(\mathcal{Z}_T)_{\mathrm{mar}} = \begin{cases}\hat{\boldsymbol{\Theta}}(z_T)_{\mathrm{mar}}, & (\mathcal{X}_T, \mathcal{Z}_T) = (\chi_T, z_T)\\ & \quad \in \mathcal{E}^{\mathrm{mar}}(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T)\\ \textit{undefined}, & \textit{otherwise}\end{cases}$$

exists but is not necessarily defined for every $(\chi_T, z_T) \in SO(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T)$. If we constrain our parameter space to obtain a unique marginal MAP estimate (recall case 2. Finitely many maxima), we call our model locally observable.

Remark 2. *In Definition 3, we define a model to be locally observable if the parameter space has to be constrained due to finitely many maxima of the posterior density. This should not be confused with a different type of locality that is described in Remark 1, stating that the (marginal) MAP estimator is not necessarily defined for all realizations $(\chi_T, z_T) \in SO(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T)$.*

In the ideal case, the estimator $\hat{\boldsymbol{\Theta}}(\mathcal{Z}_T)_{\mathrm{mar}}$ is consistent for the parameters of interest, i.e.,

$(\hat{\boldsymbol{\Omega}}(\mathcal{Z}_T)_{\text{mar}}, \hat{\mathbf{x}}_0(\mathcal{Z}_T)_{\text{mar}}) \xrightarrow{P} (\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true})$ for $T \to \infty$. If not, a bias would be detected by the algorithm and could be taken into account.

For an observable model, we do not only want to answer the question of observability but also want to know how well it is observable. As measures, we consider the cardinality of $\mathcal{E}(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T)$ and $\mathcal{E}^{\text{mar}}(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T)$, $\mathbb{E}(\hat{\boldsymbol{\Theta}}(\mathcal{Z}_T))$, $\text{Var}(\hat{\boldsymbol{\Theta}}(\mathcal{Z}_T))$, and $\text{MSE}(\hat{\boldsymbol{\Theta}}(\mathcal{Z}_T))$ as well as $\mathbb{E}(\hat{\boldsymbol{\Theta}}(\mathcal{Z}_T)_{\text{mar}})$, $\text{Var}(\hat{\boldsymbol{\Theta}}(\mathcal{Z}_T)_{\text{mar}})$, and $\text{MSE}(\hat{\boldsymbol{\Theta}}(\mathcal{Z}_T)_{\text{mar}})$. We introduce the local variance, which is also used to quantify the degree of observability: Given an observation trajectory z_T, we want to approximate the posterior $\Pi(\cdot|z_T)$ by a Gaussian density in $\boldsymbol{\Theta}$ at the joint MAP estimate $\hat{\tilde{\boldsymbol{\Theta}}}(z_T)$ if it exists, using a second-order Taylor polynomial. For given z_T, the local variances of the parameter estimates $\hat{\Theta}_j(z_T)$ are then defined by the variances of the fitted Gaussian density. Hence, they are given by $\text{LVar}(\hat{\Theta}_j(z_T)) := -\mathbf{H}_\pi^{-1}(\hat{\tilde{\boldsymbol{\Theta}}}(z_T))_{jj}$, $j = 1, \ldots, |\boldsymbol{\Theta}|$, where $\mathbf{H}_\pi(\cdot) \in \mathbb{R}^{|\boldsymbol{\Theta}|\times|\boldsymbol{\Theta}|}$ is the Hessian matrix of $\pi(\cdot|z_T)$. In [18], the local variance is derived in detail for the one-dimensional parameter case.

Hence, we now have two types of variances. The first one is the variance $\text{Var}(\hat{\boldsymbol{\Theta}}(\mathcal{Z}_T))$ of the estimator $\hat{\boldsymbol{\Theta}}(\mathcal{Z}_T)$. In contrast to $\text{Var}(\hat{\boldsymbol{\Theta}}(\mathcal{Z}_T))$, the local variance considers how peaked the maximum of $\pi(\cdot|z_T)$ in the joint MAP estimate $\hat{\tilde{\boldsymbol{\Theta}}}(z_T)$ is, given one specific observation trajectory z_T. The maximum of $\pi(\cdot|z_T)$ is getting more peaked if the local variances are decreasing (see [18, Fig. 2]).

III. PROPOSAL TO CHECK OBSERVABILITY IN PRACTICE

From Remark 1 we know that the joint and marginal MAP estimators might not be defined everywhere. In particular, the existence of $\hat{\tilde{\boldsymbol{\Theta}}}(z_T)$ depends on the realization (χ_T, z_T) of $(\mathcal{X}_T, \mathcal{Z}_T)$, i.e., $\mathcal{E}(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T) \neq SO(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T)$ (see Definition 2). The same holds for the marginal MAP estimator $\hat{\tilde{\boldsymbol{\Theta}}}(\mathcal{Z}_T)_{\text{mar}}$, i.e., $\mathcal{E}^{\text{mar}}(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T) \neq SO(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T)$ (see Definition 3). We want to know for which elements in $SO(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T)$ the estimators are defined if they exist. In general, the estimators cannot be determined analytically to see where they are defined. Hence, in theory, we have to try all realizations $(\chi_T, z_T) \in SO(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T)$ to check if a joint or marginal MAP estimate of $\pi(\cdot|z_T)$ can be recovered. That is not possible in practice as there are infinitely many elements in $SO(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T)$. Instead, for fixed values of $\boldsymbol{\Omega}^{true}$, $\mathbf{x}_0^{true}$, and T, we construct K design observation trajectories $\tilde{z}_T^{(k)} = \tilde{z}_T^{(k)}(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true})$ with corresponding design state trajectory $\tilde{\chi}_T^{(k)} = \tilde{\chi}_T^{(k)}(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true})$, $k = 1, \ldots, K$, approximately representing all possible elements in $SO(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T)$.

For this, the disturbance distribution is approximated by a discrete distribution and the corresponding locations of the point masses are used as designed disturbance realizations, from which we construct the design values $\tilde{z}_T^{(k)}$ and $\tilde{\chi}_T^{(k)}$. Then, given one design observation trajectory $\tilde{z}_T^{(k)}$, the goal is to check whether we can estimate the augmented state $(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, \tilde{\chi}_T^{(k)})$. We discuss the construction of these design observation trajectories in detail now.

A. Construction of the Design Observation Trajectories

The construction of the K design observations $\tilde{\mathbf{z}}_t^{(k)}$ and the corresponding design states $\tilde{\mathbf{x}}_t^{(k)}$, $t = 1, \ldots, T$, $k = 1, \ldots, K$, approximately representing $SO(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T)$, given $\boldsymbol{\Omega}^{true}$, $\mathbf{x}_0^{true}$, T, and K, will now be discussed in detail. We only consider $R_t = I_d$ and $Q_t = I_p$. For different covariance matrices, transformations or other approaches to approximate non-standard multivariate normal distributions could be used.

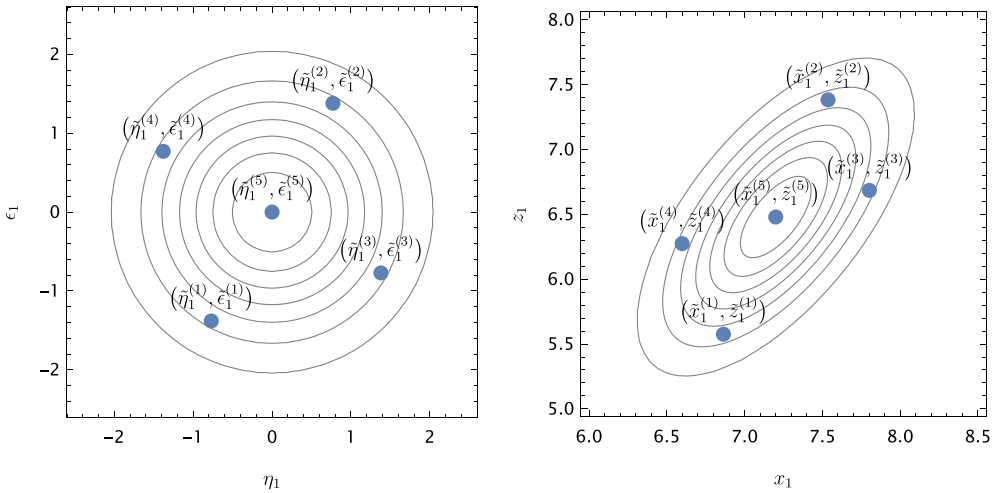

Figure 2. Left: Illustration of the design disturbance vectors $(\tilde{\eta}_1^{(k)}, \tilde{\epsilon}_1^{(k)})$ for $k = 1, \ldots, 5$, together with the contour plot of the standard normal density. Right: Illustration of the corresponding $(\tilde{x}_1^{(k)}, \tilde{z}_1^{(k)})$, $k = 1, \ldots, 5$, together with the contour plot of the theoretical density for $a^{true} = 0.9$, $x_0^{true} = 8$.

1) Case: $d = p = 1$: The parts of Model (1) with $d = p = 1$ on which we have an influence are the disturbances ϵ_t and $\eta_t, t = 1, \ldots, T$, because they are not fixed to a specific value. Hence, we need to find suitable values for them. In order to obtain K design disturbance vectors, we use the fact that $(\eta_1, \ldots, \eta_T, \epsilon_1, \ldots, \epsilon_T) \sim \mathcal{N}_{2T}(\mathbf{0}, I_{2T})$. Then, we approximate the distribution of $\mathcal{N}_{2T}(\mathbf{0}, I_{2T})$ by a discrete distribution with K point mass vectors in $\mathbb{R}^{2T}$ and obtain $(\tilde{\eta}_1^{(k)}, \ldots, \tilde{\eta}_T^{(k)}, \tilde{\epsilon}_1^{(k)}, \ldots, \tilde{\epsilon}_T^{(k)}), k = 1, \ldots, K$. To do this, the algorithm introduced in [32] is used.

The idea of this algorithm is to approximate the continuous distribution of $\mathcal{N}_q(\mathbf{0}, I_q)$ by a discrete distribution that takes on K point-symmetric point masses with equal probability. The locations of the point masses of this discrete distribution are called discrete approximations or deterministic samples. For an even number of values $K = 2N$, the discrete distribution takes on the values $\{\mathbf{s}_1, \ldots, \mathbf{s}_N, -\mathbf{s}_1, \ldots, -\mathbf{s}_N\}$. For an odd number $K = 2N + 1$, we add $\mathbf{0}$. Hence, the vectors we have to find are given by $\mathcal{S}_N^q = \{\mathbf{s}_1, \ldots, \mathbf{s}_N\}$. To find the optimal values of the vectors in $\mathcal{S}_N^q$, we need a distance measure between a continuous and a discrete distribution. For this, the so-called localized cumulative distribution function of [19] is utilized. Then, a distance measure based on the Cramér-von Mises distance for the localized cumulative distribution function is defined. Minimizing the distance between $\mathcal{N}_q(\mathbf{0}, I_q)$ and the desired discrete distribution over $\mathcal{S}_N^q$ leads to an optimal discrete approximation.

We obtain a matrix with dimensions $2T \times K$ consisting of the design disturbance values. The kth column corresponds to one design disturbance vector $(\tilde{\eta}_1^{(k)}, \ldots, \tilde{\eta}_T^{(k)}, \tilde{\epsilon}_1^{(k)}, \ldots, \tilde{\epsilon}_T^{(k)})$ for $k = 1, \ldots, K$, where the first half belongs to the state disturbance $\eta_t, t = 1, \ldots, T$, and the second half belongs to the observation disturbance $\epsilon_t, t = 1, \ldots, T$. Given these design disturbance vectors, the design observations and corresponding design states are defined by

$$\begin{aligned}
\tilde{z}_t^{(k)} &= g(\boldsymbol{\Omega}^{true}, \tilde{x}_t^{(k)}, \tilde{\epsilon}_t^{(k)}), t = 1, \ldots, T, \\
\tilde{x}_t^{(k)} &= h(\boldsymbol{\Omega}^{true}, \tilde{x}_{t-1}^{(k)}, \tilde{\eta}_t^{(k)}), t = 2, \ldots, T, \\
\tilde{x}_1^{(k)} &= h(\boldsymbol{\Omega}^{true}, x_0^{true}, \tilde{\eta}_1^{(k)})
\end{aligned}$$

for fixed values x_0^{true} and $\boldsymbol{\Omega}^{true}$ of the initial state x_0 and the time-invariant parameters $\boldsymbol{\Omega}$.

Example 5 *.For $T = 1$, we approximate the distribution of $\mathcal{N}_2(\mathbf{0}, I_2)$ by $K = 5$ values. This leads to the approximations of the disturbances (η_1, ϵ_1) shown on the left side of Fig. 2. Then we obtain $\tilde{z}_1^{(k)} = g(\boldsymbol{\Omega}^{true}, \tilde{x}_1^{(k)}, \tilde{\epsilon}_1^{(k)})$ and $\tilde{x}_1^{(k)} = h(\boldsymbol{\Omega}^{true}, x_0^{true}, \tilde{\eta}_1^{(k)})$ for $k = 1, \ldots, 5$. For the model given in 2, this means $\tilde{z}_1^{(k)} = a^{true}\tilde{x}_1^{(k)} + \sqrt{1 - (a^{true})^2} \cdot \tilde{\epsilon}_1^{(k)}$ and $\tilde{x}_1^{(k)} = a^{true}x_0^{true} + \sqrt{1 - (a^{true})^2} \cdot \tilde{\eta}_1^{(k)}$. On the right side of Fig. 2, the result is plotted together with the contour plot of the theoretical density of (X_1, Z_1) for $a^{true} = 0.9, x_0^{true} = 8$.*

2) Case: General d and p: For general d and p, the construction of the design observation trajectories can be done analogously. For $R_t = I_d$ and $Q_t = I_p$, the joint distribution of $(\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_T, \boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_T)$ is $\mathcal{N}_{(d+p)T}(\mathbf{0}, I_{(d+p)T})$, which we approximate by a discrete distribution with K values to obtain design observation disturbances that we can insert into the model. We obtain a matrix with dimensions $(d + p)T \times K$ consisting of the design disturbance values. This leads to design observation trajectories $\tilde{z}_T^{(k)} = (\tilde{\mathbf{z}}_1^{(k)}, \ldots, \tilde{\mathbf{z}}_T^{(k)})$, based on the design state trajectories $\tilde{x}_T^{(k)} = (\tilde{\mathbf{x}}_1^{(k)}, \ldots, \tilde{\mathbf{x}}_T^{(k)}), k = 1, \ldots, K$.

B. Finding the Joint MAP Estimate Using Deterministic Numerical Optimization

As described in Section II.B , we want to investigate whether the joint MAP estimator exists to obtain information about the type of extrema of the posterior. Furthermore, we are interested in the properties of the joint MAP estimator, such as mean, variance, and local variance. Using the design observation trajectories $\tilde{z}_T^{(k)}, k = 1, \ldots, K$, we maximize the K posterior densities $\Pi(\cdot|\tilde{z}_T^{(k)}), k = 1, \ldots, K$, defined in (2), numerically over $\boldsymbol{\Theta}$.

The numerical maximization of $\Pi(\cdot|\tilde{z}_T^{(k)})$ is done using a primal barrier method (see [26, Sec. 19.6]) for $k = 1, \ldots, K$. When obtaining the optimization result, we have to perform checks to make sure that we have found a maximum. Due to the phenomenon described in Remark 1, it might be possible that for some design vectors $\tilde{z}_T^{(k)}$ with corresponding $\tilde{x}_T^{(k)}$, no maximum can be found. This leads to the estimates $\hat{\boldsymbol{\Theta}}(\tilde{z}_T^{(k)}), k \in \mathcal{K}^{\max} \subset \{1, \ldots, K\}$, with $|\mathcal{K}^{\max}| \leq K$.

The optimization procedure searches for minima; hence, we use the function $-2\pi(\cdot|\tilde{z}_T^{(k)})$ for the optimization. Then, given a design observation trajectory $\tilde{z}_T^{(k)}$, we check the following properties of the result $\hat{\boldsymbol{\Theta}}(\tilde{z}_T^{(k)})$ to ensure that the optimizer has converged to a minimum. The thresholds are chosen based on results of numerical studies we executed. For the kth optimization, we choose $(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, \tilde{x}_T^{(k)})$ as initial value.

Numerical Checks. Hessian: *Is the Hessian $H_{-2\pi}(\hat{\boldsymbol{\Theta}}(\tilde{z}_T^{(k)}))$ of $-2\pi(\cdot|\tilde{z}_T^{(k)})$ in $\hat{\boldsymbol{\Theta}}(\tilde{z}_T^{(k)})$ positive definite? Then, we have a minimum of $-2\pi(\cdot|\tilde{z}_T^{(k)})$ and hence a maximum of $\Pi(\cdot|\tilde{z}_T^{(k)})$.*

Gradients: *Are the gradients of $-2\pi(\hat{\boldsymbol{\Theta}}(\tilde{z}_T^{(k)}))$ close to zero, i.e., smaller than a threshold? The threshold we use is 10^{-5}.*

Eigenvalues: *Is the result a ridge? The ratio of the smallest eigenvalue to the largest eigenvalue of $H_{-2\pi}(\hat{\boldsymbol{\Theta}}(\tilde{z}_T^{(k)}))$ should not be too small. We want a ratio of $> r_0 = 0.00001$. The ratio we consider is the inverse of the condition number of the Hessian, which is often considered in*

the literature [33], [34] and can be used in sensitivity analysis the way we also use it [16, Sec. 8.3.3]. As the threshold of r_0 is chosen based on results of numerical studies, one more step is applied to ensure that a ridge is present. In the case of a ratio $< r_0$, another starting point for the optimization is chosen. That starting point is close to the found optimum but shifted in the direction of the ridge. If this new optimization leads to a different value than before, we conclude that a ridge is present.

One way to quantify the observability of an SSM is the number $|\mathcal{K}^{\max}|$ of design observation trajectories $\tilde{z}_T^{(k)}$ with corresponding $\tilde{x}_T^{(k)}$, for which a unique estimate $\hat{\boldsymbol{\Theta}}(\tilde{z}_T^{(k)})$ can be recovered. We denote the set of values k for which a maximum can be recovered by $\mathcal{K}^{\max} = \mathcal{K}^{\max}(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T) = \{k \in \{1, \dots, K\} : \operatorname{argmax}_{\boldsymbol{\Theta}} \Pi(\boldsymbol{\Theta} | \tilde{z}_T^{(k)}(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}))$ fulfills numerical checks listed above$\} \subset \{1, \dots, K\}$ with $|\mathcal{K}^{\max}| \leq K$. For further analyses, only the results $\hat{\boldsymbol{\Theta}}(\tilde{z}_T^{(k)}), k \in \mathcal{K}^{\max}$, are considered.

We say that the degree of observability for the fixed values $\boldsymbol{\Omega}^{true}$, $\mathbf{x}_0^{true}$, and T increases with $|\mathcal{K}^{\max}|$. Given the estimates $\hat{\boldsymbol{\Theta}}(\tilde{z}_T^{(k)}), k \in \mathcal{K}^{\max}$, that were recovered, the mean, variance, and MSE of the corresponding estimator $\hat{\boldsymbol{\Theta}}(\mathcal{Z}_T)$ can be estimated. The local variance can be used as another indicator of the degree of observability. A lower local variance indicates a higher degree of observability as the joint MAP estimator has higher precision.

C. Finding the Marginal MAP Estimate Using MCMC Sampling

In general, the integrals in the marginal posterior densities in (9) cannot be determined analytically. One solution to this is to use MCMC methods to approximately solve the integrals. We use a variant of Hamiltonian Monte Carlo (HMC), more specifically the No-U-Turn-Sampler [21] implemented in `Stan` [8], a platform for statistical modeling. Using `Stan`, we obtain MCMC samples from the posterior. Then, we can determine marginal MAP estimates from univariate kernel density estimates.

The three cases of (1) no maximum, (2) finitely many maxima, and (3) infinitely many maxima of the posterior can lead to convergence problems in the MCMC sampler. According to [1], *ridges in the posterior density [...] wreak havoc with both sampling and inference*. Additionally, multimodal posteriors lead to problems, which can however be avoided by constraining the parameter space if the posterior is not highly multimodal. Furthermore, there are examples where there are *no posterior modes and numerical stability issues can arise as sampled parameters approach constraint boundaries* [1]. That is why we first obtain information about the type of extrema by considering the joint MAP estimate to understand when such situations occur.

Sampling with `Stan` from the log-posterior $\pi(\cdot|\tilde{z}_T^{(k)})$ for one fixed $\tilde{z}_T^{(k)}$ outputs the samples

$$\boldsymbol{\Theta}(\tilde{z}_T^{(k)})^{(r)} = \left(\boldsymbol{\Omega}(\tilde{z}_T^{(k)})^{(r)}, \mathbf{x}_0(\tilde{z}_T^{(k)})^{(r)}, \mathbf{x}_1(\tilde{z}_T^{(k)})^{(r)}, \dots, \mathbf{x}_T(\tilde{z}_T^{(k)})^{(r)}\right), \quad r = 1, \dots, R,$$

of $\boldsymbol{\Theta}$, where R is the number of MCMC samples we obtain after discarding the burn-in phase.

Define $\mathcal{K}^{\texttt{Stan}} = \mathcal{K}^{\texttt{Stan}}(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T) = \{k \in \{1, \dots, K\} :$ `Stan` converges when sampling from $\pi(\cdot|\tilde{z}_T^{(k)}(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}))\}$ with $\left|\mathcal{K}^{\texttt{Stan}}\right| \leq K$. To check convergence, no divergent transitions after warmup are allowed. Furthermore, the convergence diagnostics R-hat [14] as well as bulk and tail effective sample size [35] are considered. Additionally, the maximum treedepth and the Bayesian Fraction of Missing Information (BFMI) [7] are checked. For more information on convergence diagnostics and the exact thresholds used in `Stan` to check convergence, see [2]. Only the results $\hat{\boldsymbol{\Theta}}(\tilde{z}_T^{(k)})_{\text{mar}}, k \in \mathcal{K}^{\texttt{Stan}}$, are considered for further analyses. Hence, the diagnostics of all considered `Stan` runs are within the desired thresholds.

If $\left|\mathcal{K}^{\texttt{Stan}}\right|$ is close to K, we say that the model has a high degree of observability. If no marginal MAP estimate is recovered, i.e., $\left|\mathcal{K}^{\texttt{Stan}}\right| = 0$, we conclude that the model is not observable for the chosen values of $\boldsymbol{\Omega}^{true}$, $\mathbf{x}_0^{true}$, and T.

With the output of `Stan`, it is possible to approximate the marginal MAP estimate of the parameters as it is done in [22] and [23]. We obtain the marginal MAP estimate of Θ_j by

$$\hat{\Theta}_j(\tilde{z}_T^{(k)})_{\text{mar}} = \operatorname{argmax} \operatorname{kde}((\Theta_j(\tilde{z}_T^{(k)})^{(r)})_{r=1,\dots,R}), \quad j = 1, \dots, |\boldsymbol{\Theta}|, \tag{11}$$

where kde is the kernel density estimate [31]. The values $\hat{\boldsymbol{\Theta}}(\tilde{z}_T^{(k)})_{\text{mar}}, k \in \mathcal{K}^{\texttt{Stan}}$, can be used to estimate the mean and variance of the marginal MAP estimator $\hat{\boldsymbol{\Theta}}(\mathcal{Z}_T)_{\text{mar}}$.

For $j = 1, \dots, |\boldsymbol{\Theta}|$, we can determine 90% credible intervals $(\hat{\Theta}_j(\tilde{z}_T^{(k)})_{q5}, \hat{\Theta}_j(\tilde{z}_T^{(k)})_{q95})$ for Θ_j by

$$\begin{aligned}\hat{\Theta}_j(\tilde{z}_T^{(k)})_{q5} &= \text{empirical 5\%-quantile of } (\Theta_j(\tilde{z}_T^{(k)})^{(r)})_{r=1,\dots,R},\\ \hat{\Theta}_j(\tilde{z}_T^{(k)})_{q95} &= \text{empirical 95\%-quantile of } (\Theta_j(\tilde{z}_T^{(k)})^{(r)})_{r=1,\dots,R}.\end{aligned} \tag{12}$$

For every parameter Θ_j, we determine the percentage of iterations $k \in \mathcal{K}^{\texttt{Stan}}$, for which the true value Θ_j^{true} is in the interval $(\hat{\Theta}_j(\tilde{z}_T^{(k)})_{q5}, \hat{\Theta}_j(\tilde{z}_T^{(k)})_{q95})$, $j = 1, \dots, |\boldsymbol{\Theta}|$, i.e.,

$$\widehat{\text{Cover}}_{90}(\Theta_j^{true}) = \frac{|\{k \in \mathcal{K}^{\texttt{Stan}} : \hat{\Theta}_j(\tilde{z}_T^{(k)})_{q5} < \Theta_j^{true} < \hat{\Theta}_j(\tilde{z}_T^{(k)})_{q95}\}|}{|\mathcal{K}^{\texttt{Stan}}|}.$$

To get an idea how the marginal MAP estimator performs for the latent states $x_{t,i}, t = 1, \dots, T, i = 1, \dots, p$,

we estimate the bias by

$$\widehat{\text{Bias}}(x_{t,i}) = \frac{\sum_{k\in\mathcal{K}^{\texttt{Stan}}}\left(\hat{x}_{t,i}(\tilde{z}_T^{(k)})_{\text{mar}} - \tilde{x}_{t,i}^{(k)}\right)}{|\mathcal{K}^{\texttt{Stan}}|},$$
$$t = 1, \ldots, T, i = 1, \ldots, p,$$

where $\tilde{\mathbf{x}}_t^{(k)} = (\tilde{x}_{t,1}^{(k)}, \ldots, \tilde{x}_{t,p}^{(k)})$, and the MSE by

$$\widehat{\text{MSE}}(x_{t,i}) = \frac{\sum_{k\in\mathcal{K}^{\texttt{Stan}}}\left(\hat{x}_{t,i}(\tilde{z}_T^{(k)})_{\text{mar}} - \tilde{x}_{t,i}^{(k)}\right)^2}{|\mathcal{K}^{\texttt{Stan}}|},$$
$$t = 1, \ldots, T, i = 1, \ldots, p.$$

IV. ILLUSTRATIONS

We consider the different scenarios given in Table I.

In the following, we give an overview of the SSMs we investigate together with the respective definition of $\boldsymbol{\Theta}$ and the log-posterior $\pi(\boldsymbol{\Theta}|z_T)$. The prior $p_{\boldsymbol{\Omega}}(\boldsymbol{\Omega})$ is abbreviated by $p(\boldsymbol{\Omega})$. The disturbances are always assumed to be standard normal i.i.d. and independent and the prior p_0 of $\mathbf{x}_0$ is set to the uniformative prior. For sampling using `Stan`, we always consider 4 chains with 10 000 iterations and a burn-in of 4000, respectively, leading to $R = 24000$. Note here that running `Stan` on $\pi(\cdot|\tilde{z}_T^{(k)})$ can be parallelized for each $k = 1, \ldots, K$. We allow for random initial values in order to investigate the behavior when the true underlying values are not known.

We do not show the detailed results for all the considered models. However, they are available from the authors upon request.

1) Model `(Same,d=p=1)` (Same dynamics for observations and states for d=p=1):

$$Z_t = aX_t + \sqrt{1-a^2}\cdot\epsilon_t\,,\ X_t = aX_{t-1} + \sqrt{1-a^2}\cdot\eta_t\,,$$
$$t = 1, \ldots, T.$$

Log-posterior of $\boldsymbol{\Theta} = (a, x_0, \chi_T)$:

$$\pi(\boldsymbol{\Theta}|z_T) = -\frac{1}{2}\frac{1}{1-a^2}\left[\sum_{t=1}^{T}(z_t - ax_t)^2 + \sum_{t=1}^{T}(x_t - ax_{t-1})^2\right] - T\log(1-a^2) + \log(p^{res}_{a^{true}}(a)).$$

2) Model `(Sep,d=p=1)` (Separate dynamics for observations and states for d=p=1):

$$Z_t = bX_t + \sqrt{1-b^2}\cdot\epsilon_t\,,\ X_t = aX_{t-1} + \sqrt{1-a^2}\cdot\eta_t\,,$$
$$t = 1, \ldots, T.$$

Log-posterior of $\boldsymbol{\Theta} = (a, b, x_0, \chi_T)$:

$$\pi(\boldsymbol{\Theta}|z_T) = -\left[\sum_{t=1}^{T}\frac{(z_t - bx_t)^2}{2(1-b^2)} + \frac{(x_t - ax_{t-1})^2}{2(1-a^2)}\right] - \frac{T}{2}\left[\log(1-a^2) + \log(1-b^2)\right] + \log(p(a,b)).$$

3) Model `(Sep,d=2,p=1)` (Separate dynamics for bivariate observations driven by single states):

$$Z_{t,1} = aX_t + \sqrt{1-a^2}\cdot\epsilon_{t,1}\,,\ Z_{t,2} = bX_t + \sqrt{1-b^2}\cdot\epsilon_{t,2}\,,$$
$$X_t = cX_{t-1} + \sqrt{1-c^2}\cdot\eta_t.$$

Log-posterior of $\boldsymbol{\Theta} = (a, b, c, x_0, \chi_T)$:

$$\pi(\boldsymbol{\Theta}|z_T) = -\frac{1}{2}\frac{1}{1-a^2}\sum_{t=1}^{T}(z_{t,1} - ax_t)^2 - \frac{T}{2}\log(1-a^2) - \frac{1}{2}\frac{1}{1-b^2}\sum_{t=1}^{T}(z_{t,2} - bx_t)^2 - \frac{T}{2}\log(1-b^2) - \frac{1}{2}\frac{1}{1-c^2}\sum_{t=1}^{T}(x_t - cx_{t-1})^2 - \frac{T}{2}\log(1-c^2) + \log(p(a,b,c)).$$

4) Model `(Sep,d=1,p=2)` (Separate dynamics for univariate observations and bivariate states):

$$Z_t = aX_{t,1} + bX_{t,2} + \sqrt{1-a^2-b^2}\cdot\epsilon_t$$
$$X_{t,1} = cX_{t-1,1} + \sqrt{1-c^2}\cdot\eta_{t,1}\,,\ X_{t,2} = dX_{t-1,2} + \sqrt{1-d^2}\cdot\eta_{t,2}\,,\ t = 1, \ldots, T$$

Table I
The Considered Scenarios

Dim. d of observations	Dim. p of state	Linearity	Considered models
1	1	Nonlinear (linear only in state)	`(Same,d=p=1)`, `(Sep,d=p=1)`, `(Sep,d=p=1,ranWalk)`
2	1	Nonlinear (linear only in state)	`(Sep,d=2,p=1)`
1	2	Nonlinear (linear only in state)	`(Sep,d=1,p=2)`
1	1	Nonlinear	`(Sep,d=p=1, Multiplicative)`

Log-posterior of $\boldsymbol{\Theta} = (a, b, c, d, \mathbf{x}_0, \chi_T)$:

$$\begin{aligned}\pi(\boldsymbol{\Theta}|z_T) = &-\frac{1}{2}\frac{1}{1-a^2-b^2}\sum_{t=1}^{T}(z_t - ax_{t,1} - bx_{t,2})^2 \\ &- \frac{T}{2}\log(1-a^2-b^2) \\ &-\frac{1}{2}\frac{1}{1-c^2}\sum_{t=1}^{T}(x_{t,1} - cx_{t-1,1})^2 \\ &- \frac{T}{2}\log(1-c^2) \\ &-\frac{1}{2}\frac{1}{1-d^2}\sum_{t=1}^{T}(x_{t,2} - dx_{t-1,2})^2 \\ &- \frac{T}{2}\log(1-d^2) + \log(p(a,b,c,d)).\end{aligned}$$

5) Model `(Sep,d=p=1,ranWalk)` (Univariate observations with random walk state dynamics):

$$Z_t = aX_t + \epsilon_t\,,\ X_t = X_{t-1} + \eta_t\,,\ t = 1, \ldots, T$$

Log-posterior of $\boldsymbol{\Theta} = (a, \mathbf{x}_0, \chi_T)$:

$$\begin{aligned}\pi(\boldsymbol{\Theta}|z_T) = &-\frac{1}{2}\sum_{t=1}^{T}(z_t - ax_t)^2 - \frac{1}{2}\sum_{t=1}^{T}(x_t - x_{t-1})^2 \\ &+ \log(p^{res}_{a^{true}}(a)).\end{aligned} \tag{13}$$

6) Model `(Sep,d=p=1,Multiplicative)` (Additive state and multiplicative observation equation):

$$Z_t = \sqrt{a}\cdot X_t \cdot \epsilon_t\,,\ X_t = a + X_{t-1} + \eta_t\,,\ t = 1, \ldots, T.$$

Log-posterior of $\boldsymbol{\Theta} = (a, \mathbf{x}_0, \chi_T)$:

$$\begin{aligned}\pi(\boldsymbol{\Theta}|z_T) = &-\frac{1}{2}\sum_{t=1}^{T}\left(\frac{z_t^2}{a\cdot x_t^2} + \log(a\cdot x_t^2)\right) \\ &-\frac{1}{2}\sum_{t=1}^{T}(x_t - a - x_{t-1})^2 + \log(p(a)).\end{aligned}$$

A. Model `(Same,d=p=1)`

To decide on the number of design points K, we consider $\boldsymbol{V}_{K_i} = (\frac{|\mathcal{K}^{\max}|}{K_i}, \hat{\mathbb{E}}(\hat{a}(\mathcal{Z}_T)), \hat{\mathbb{E}}(\hat{x}_0(\mathcal{Z}_T)), \widehat{\mathrm{Var}}(\hat{a}(\mathcal{Z}_T)), \widehat{\mathrm{Var}}(\hat{x}_0(\mathcal{Z}_T)), \overline{\mathrm{LVar}}(\hat{a}(\tilde{z}_T^{(k)})), \overline{\mathrm{LVar}}(\hat{x}_0(\tilde{z}_T^{(k)}))) \in \mathbb{R}^7$ of the joint MAP estimator $(\hat{a}(\mathcal{Z}_T), \hat{x}_0(\mathcal{Z}_T))$ when using $K_i = i \cdot 100$ design observation trajectories. $\hat{\mathbb{E}}$ and $\widehat{\mathrm{Var}}$ denote the empirical mean and variance (see Table II). $\overline{\mathrm{LVar}}(\hat{\Theta}_i(\tilde{z}_T^{(k)}))$ is the mean of the local variances over k.

We want to find the value of i for which $\boldsymbol{V}_{K_i}$ has converged. For that, define the relative difference $RD(x, y) = \frac{|x-y|}{\max(|x|,|y|)}$ between two values $x, y \in \mathbb{R}$. For vectors, RD is defined component-wise. Then, we determine the smallest value i, for which $RD(\boldsymbol{V}_{K_i}, \boldsymbol{V}_{K_j}) < (c_1, c_1, c_1, c_2, c_2, c_2, c_2)$ for all $j > i$ and choose the number of design values as $K = i \cdot 100$. The values $(c_1, c_2) \in \mathbb{R}^2$ are chosen thresholds. We choose two different values because $|\mathcal{K}^{\max}|/K_i$ and the empirical means are converging faster than the empirical and the local variance.

Using the design disturbance values allows us to choose a smaller value of K compared to using random values of the disturbances. Numerical studies show that when using random values of the disturbances, a higher value of i has to be chosen to ensure convergence of $\boldsymbol{V}_{K_i}$. For example, for $(a^{true}, x_0^{true}) = (0.9, 8)$, $T = 4$, we choose $K = 300$ using $(c_1, c_2) = (0.01, 0.01)$ and the design disturbances. When using random values of the disturbances for the same setup, we have to choose

Table II
Top and Middle: The Empirical Mean and Variance of the Joint MAP Estimator (Left) and the Marginal MAP Estimator (Right)

	Joint MAP estimation	Marginal MAP estimation
T	$\left(\hat{\mathbb{E}}(\hat{a}(\mathcal{Z}_T)), \hat{\mathbb{E}}(\hat{x}_0(\mathcal{Z}_T))\right)$	$\left(\hat{\mathbb{E}}(\hat{a}(\mathcal{Z}_T)_{\mathrm{mar}}), \hat{\mathbb{E}}(\hat{x}_0(\mathcal{Z}_T)_{\mathrm{mar}})\right)$
	$(\bar{\mu}, \bar{\nu}) = \frac{\sum_{k\in\mathcal{K}^{\max}}(\hat{a}(\tilde{z}_T^{(k)}), \hat{x}_0(\tilde{z}_T^{(k)}))}{\lvert\mathcal{K}^{\max}\rvert}$	$(\bar{\mu}_{\mathrm{mar}}, \bar{\nu}_{\mathrm{mar}}) = \frac{\sum_{k\in\mathcal{K}^{\mathtt{Stan}}}(\hat{a}(\tilde{z}_T^{(k)})_{\mathrm{mar}}, \hat{x}_0(\tilde{z}_T^{(k)})_{\mathrm{mar}})}{\lvert\mathcal{K}^{\mathtt{Stan}}\rvert}$
20	(0.941, 7.15)	(0.896, 8.07)
50	(0.949, 6.99)	(0.898, 8.05)
T	$\left(\widehat{\mathrm{Var}}(\hat{a}(\mathcal{Z}_T)), \widehat{\mathrm{Var}}(\hat{x}_0(\mathcal{Z}_T))\right)$	$\left(\widehat{\mathrm{Var}}(\hat{a}(\mathcal{Z}_T)_{\mathrm{mar}}), \widehat{\mathrm{Var}}(\hat{x}_0(\mathcal{Z}_T)_{\mathrm{mar}})\right)$
	$\frac{\sum_{k\in\mathcal{K}^{\max}}\left((\hat{a}(\tilde{z}_T^{(k)})-\bar{\mu})^2, (\hat{x}_0(\tilde{z}_T^{(k)})-\bar{\nu})^2\right)}{\lvert\mathcal{K}^{\max}\rvert-1}$	$\frac{\sum_{k\in\mathcal{K}^{\mathtt{Stan}}}\left((\hat{a}(\tilde{z}_T^{(k)})_{\mathrm{mar}}-\bar{\mu}_{\mathrm{mar}})^2, (\hat{x}_0(\tilde{z}_T^{(k)})_{\mathrm{mar}}-\bar{\nu}_{\mathrm{mar}})^2\right)}{\lvert\mathcal{K}^{\mathtt{Stan}}\rvert-1}$
20	(0.0001960, 0.468)	(0.000605, 0.728)
50	(0.0000756, 0.374)	(0.000355, 0.602)
T	$\left(\widehat{\mathrm{MSE}}(\hat{a}(\mathcal{Z}_T)), \widehat{\mathrm{MSE}}(\hat{x}_0(\mathcal{Z}_T))\right)$	$\left(\widehat{\mathrm{MSE}}(\hat{a}(\mathcal{Z}_T)_{\mathrm{mar}}), \widehat{\mathrm{MSE}}(\hat{x}_0(\mathcal{Z}_T)_{\mathrm{mar}})\right)$
20	(0.00184, 1.18)	(0.000621, 0.733)
50	(0.00248, 1.40)	(0.000360, 0.605)

The Empirical Mean is Closest to the True Values $(a^{true}, x_0^{true}) = (0.9, 8)$ When Using the Marginal MAP Estimate. Bottom: The Estimated MSE for the Two Estimators.

$(c_1, c_2) = (0.01, 0.03)$, and it only converges for values between $K = 2500$ and $K = 4500$ or higher depending on the realizations.

In the following, we investigate observability properties for different choices of (a^{true}, x_0^{true}).

$(a^{true}, x_0^{true}) = (0.9, 8)$. For $T = 20$ and $T = 50$, we determine the values of K as described above. For $T = 20$, we choose $K = 1300$ with $(c_1, c_2) = (0.01, 0.03)$, and for $T = 50$, we choose $K = 3100$ with $(c_1, c_2) = (0.01, 0.04)$. K increases with T because more values are necessary to accurately cover a higher-dimensional space.

Maximizing the log-posterior densities $\pi(\boldsymbol{\Theta}|\tilde{z}_T^{(k)})$, $k = 1, \ldots, K$, leads to the left side of Fig. 3, showing the estimates $(\hat{a}(\tilde{z}_T^{(k)}), \hat{x}_0(\tilde{z}_T^{(k)}))$, $k \in \mathcal{K}^{\max}$, of (a, x_0). The joint MAP estimator exists with $|\mathcal{K}^{\max}|/K = 1$ but is biased. The empirical expected value and variance of the corresponding estimator $(\hat{a}(Z_T), \hat{x}_0(Z_T))$ are reported in the left part of Table II. The bias is increasing in T, while the empirical variance is decreasing. As expected, the mean of the K local variances is decreasing for increasing T (not shown here).

Using `Stan` to obtain the marginal MAP estimates as discussed in Section III.C, the right side of Fig. 3 shows that the issue of the bias gets solved. However, there is a trade-off with the variance. This can also be seen on the right side of Table II.

To study the joint behavior of bias and variance, we report estimates for the MSE of a and x_0 for the joint and marginal MAP estimation, respectively (bottom of Table II).

The estimated MSE of the marginal MAP estimator is always smaller. This is due to the fact that the bias is a lot smaller for the marginal MAP estimator compared to the joint MAP estimator, and the variance is not increasing enough to obtain a higher value of the MSE values. The values of the estimated coverage of a, x_0, and $x_1, \ldots, x_T$, as well as the estimated average bias and MSE of x_t, $t = 1, \ldots, T$, are satisfactory (Table III).

$(a^{true}, x_0^{true}) = (0.7, 8)$. We study the behavior of the model for a smaller value of a. We need a larger value of T in order to ensure a good degree of observability. Hence, we consider $T = 100$ and choose $K = 3600$ with $(c_1, c_4) = (0.01, 0.06)$. The joint MAP approach leads again to an obvious bias with $|\mathcal{K}^{\max}| = K$. Compared to $(a^{true}, x_0^{true}) = (0.9, 8)$, the local variance shows that the maximum of the posterior here is on average less peaked.

We use `Stan` to obtain the marginal MAP estimates, leading to a value of $|\mathcal{K}^{\texttt{Stan}}| = 2910$ smaller than $|\mathcal{K}^{\max}|$. As before, using the marginal MAP estimator solves the issue of the bias, but we have a trade-off with the variance. However, the estimated MSE is smaller for the marginal estimator. The average MSE of x_t, $t = 1, \ldots, T$, is slightly higher compared to the case $(a^{true}, x_0^{true}) = (0.9, 8)$. The estimated coverage of a, x_0, and $x_1, \ldots, x_T$ is satisfactory.

B. Model `(Sep,d=p=1)`

For this model, we face the problem of having multiple modes with equal height. Consider

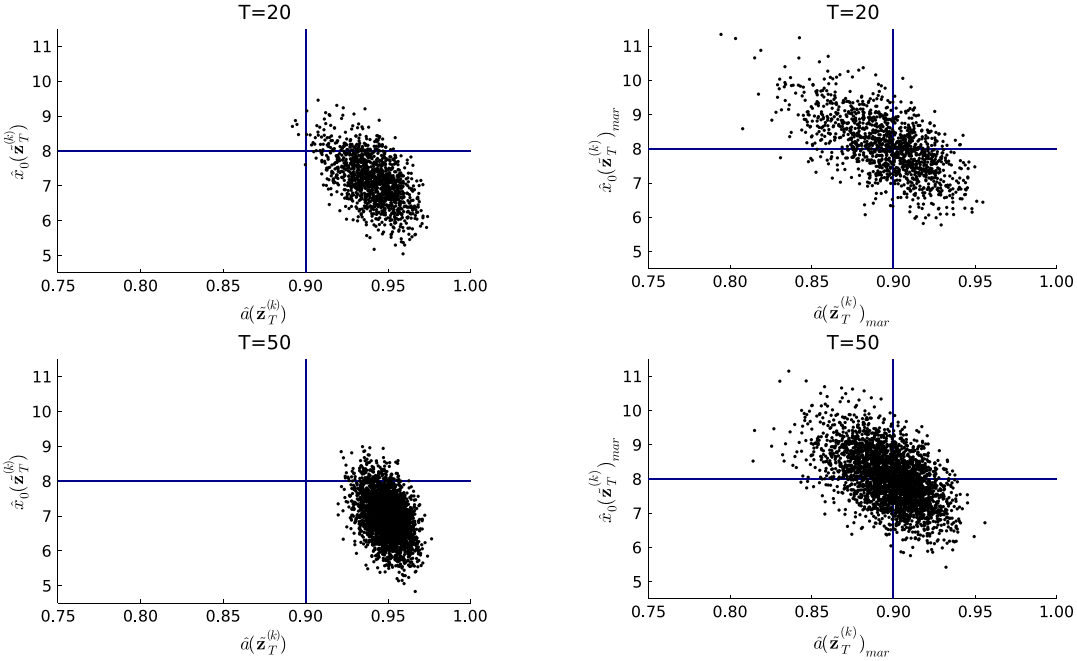

Figure 3. **$(a^{true}, x_0^{true}) = (0.9, 8)$:** Pairwise scatter plots of the joint MAP estimates $(\hat{a}(\tilde{z}_T^{(k)}), \hat{x}_0(\tilde{z}_T^{(k)}))$, $k \in \mathcal{K}^{\max}$, with $|\mathcal{K}^{\max}| = K$ (left) and marginal MAP estimates $(\hat{a}(\tilde{z}_T^{(k)})_{\text{mar}}, \hat{x}_0(\tilde{z}_T^{(k)})_{\text{mar}})$, $k \in \mathcal{K}^{\texttt{Stan}}$, with $|\mathcal{K}^{\texttt{Stan}}| = K$ (right) for different values of T. The true values are indicated by the lines.

Table III
The Estimated Coverage of the Parameters $\boldsymbol{\Theta} = (a, x_0, x_1, \ldots, x_T)$ as well as the Average Estimated Bias and MSE of $x_t, t = 1, \ldots, T$, for the Marginal MAP Estimates

T	$\widehat{\text{Cover}}_{90}(a^{true}, x_0^{true})$	$\frac{1}{T}\sum_{t=1}^{T} \widehat{\text{Cover}}_{90}(x_t^{true})$	$\frac{1}{T}\sum_{t=1}^{T} \lvert\widehat{\text{Bias}}(x_t)\rvert$	$\frac{1}{T}\sum_{t=1}^{T} \widehat{\text{MSE}}(x_t)$
20	0.870, 0.895	0.910	0.00892	0.116
50	0.893, 0.905	0.909	0.00335	0.105

$\boldsymbol{\Theta}_1 = (a, b, x_0, \chi_T)$ and $\boldsymbol{\Theta}_2 = (a, -b, -x_0, -\chi_T)$. As $\pi(\boldsymbol{\Theta}_1|z_T) = \pi(\boldsymbol{\Theta}_2|z_T)$, the posterior has two distinct modes, so we have to constrain our parameter space. One solution is to constrain both a and b while it is also possible to only pose restrictions on b. We consider both options.

Kreuzer *et al.* [22] approach the problem slightly different. The correlation between two consecutive observations Z_{t-1} and Z_t is $\text{Cor}(Z_{t-1}, Z_t) = b^2 \cdot a$. From this formula, it is obvious that there is an identifiability problem in b. As a solution, the model is reduced to a problem with one time-invariant parameter by setting $b = a^c$ for some $c \geq 1$ and the remaining parameter a is restricted to (0,1). Then, $\text{Cor}(Z_{t-1}, Z_t) = a^{2c+1}$, so that the identifiability problem is solved. With $c = 1$, this leads to Model `(Same,d=p=1)`.

$(a^{true}, b^{true}, x_0^{true}) = (0.9, 0.9, 8)$: Restricting both a and b to (0,1). We consider $T = 75$ and $K = 4000$ and choose the prior of a and b by $p(a, b) = p^{res}_{a^{true}}(a) \cdot p^{res}_{b^{true}}(b)$. We obtain $|\mathcal{K}^{\max}| = 1956$, hence the value of $|\mathcal{K}^{\max}|/K$ is significantly smaller than 1. There is an obvious bias of the joint MAP estimator again.

Using the marginal MAP estimates, we obtain $3680 = \left|\mathcal{K}^{\texttt{Stan}}\right| > |\mathcal{K}^{\max}|$. Using the marginal approach solves the issue of the bias again but leads to a trade-off with the variance. However, the MSE is smaller for the marginal approach again. The estimated coverage of a, b, x_0, and $x_1, \ldots, x_T$ as well as the average bias and MSE of $x_t, t = 1, \ldots, T$, are satisfactory.

$(a^{true}, b^{true}, x_0^{true}) = (0.9, 0.9, 8)$: Restricting b to (0,1) only. The prior of a and b is now chosen by $p(a, b) = 1_{(-1,1)}(a) \cdot p^{res}_{b^{true}}(b)$, and we consider again $T = 75$ and $K = 4000$. Compared to the case where we restrict a and b, the value of $\left|\mathcal{K}^{\texttt{Stan}}\right|$ decreases from 3680 to 2109, i.e., the degree of observability is decreasing. However, the values of the marginal MAP approach when only restricting b are very similar to restricting both a and b.

C. Model `(Sep,d=2,p=1)`

As noted by [23], the parameters $\boldsymbol{\Theta}_1 = (a, b, c, x_0, \chi_T)$ and $\boldsymbol{\Theta}_2 = (-a, -b, c, -x_0, -\chi_T)$ give the same posterior value and thus two distinct maxima. Hence, we have to restrict our parameter space. [23] propose to restrict only a to (0,1) and impose no further restrictions. We show now that this leads to an observable model. However, $\left|\mathcal{K}^{\texttt{Stan}}\right|$ increases significantly when also restricting the variables b and c.

$(a^{true}, b^{true}, c^{true}, x_0^{true}) = (0.9, 0.9, 0.9, 8)$: Restricting a, b, and c to (0,1). The prior of a, b, and c is chosen by $p(a, b, c) = p^{res}_{a^{true}}(a) \cdot p^{res}_{b^{true}}(b) \cdot p^{res}_{c^{true}}(c)$. We investigate $T = 50$ and $K = 4000$. We are now in the case of $d = 2$ and $p = 1$. The joint MAP approach leads to a biased estimator again, which is solved by the marginal MAP approach, for which we obtain $3995 = \left|\mathcal{K}^{\texttt{Stan}}\right| > |\mathcal{K}^{\max}| = 1833$. The values of the estimated coverage of a, b, x_0, and $x_1, \ldots, x_T$, as well as the average bias and MSE of $x_t, t = 1, \ldots, T$, are satisfactory.

$(a^{true}, b^{true}, c^{true}, x_0^{true}) = (0.9, 0.9, 0.9, 8)$: Restricting only a. The prior we use is $p(a, b, c) = p^{res}_{a^{true}}(a) \cdot 1_{(-1,1)}(b) \cdot 1_{(-1,1)}(c)$. Then, all the values of the marginal approach are practically identical to the case before. The difference is that $\left|\mathcal{K}^{\texttt{Stan}}\right| = 2972$, which is lower than the value before. Hence, the degree of observability increases when setting more restrictions.

D. Model `(Sep,d=1,p=2)`

As a prior, we use $p(a, b, c, d) = p^{res}_{a^{true}}(a) \cdot p^{res}_{b^{true}}(b) \cdot p^{res}_{c^{true}}(c) \cdot p^{res}_{d^{true}}(d)$. For this model, we have $d = 1$ and $p = 2$, which means that the dimension of the latent variable is higher than the dimension of the observation. This leads to unobservability of the model, i.e., $\left|\mathcal{K}^{\texttt{Stan}}\right| = 0$. As an additional step to setting no initial values for the MCMC sampling, we also set the initial value to the true parameters to make sure that no marginal MAP estimates can be recovered.

E. Model `(Sep,d=p=1,ranWalk)`

As a prior for a, we use

$$p_{a^{true}}(a) = \begin{cases} 1, a \in (0, \infty) & a^{true} > 0 \\ 1, a \in (-\infty, 0) & a^{true} < 0. \end{cases} \tag{14}$$

$(a^{true}, x_0^{true}) = (0.9, 8) : T = 4$. For $T = 4$, this model is not observable. No maximum of $\pi(\boldsymbol{\Theta}|\tilde{z}_T^{(k)})$, $k = 1, \ldots, K$, that fulfills our checks is found on the bounded support when using $K = 2000$ design observation trajectories, i.e., $|\mathcal{K}^{\max}| = 0$. The reason is that the posterior is a ridge of variable height, as shown in Fig. 4 for design observation $\tilde{z}_4^{(1)}$. The figure shows the profile posterior $\pi(a, x_0|\tilde{z}_4^{(1)}) = \max_{x_1, \ldots, x_4} \pi(a, x_0, x_1, \ldots, x_4|\tilde{z}_4^{(1)})$. Fur-

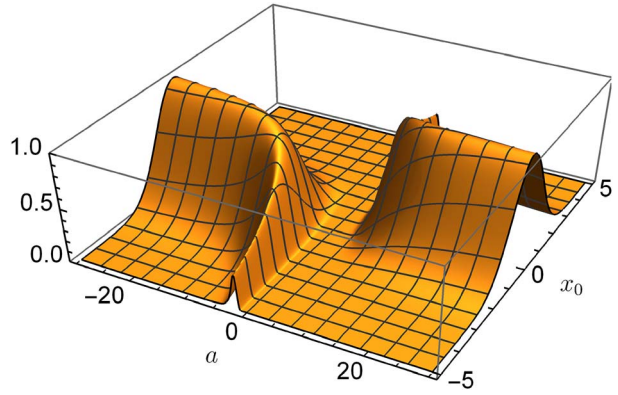

Figure 4. $(a^{true}, x_0^{true}) = (0.9, 8)$: Profile posterior $\pi(a, x_0|\tilde{z}_4^{(1)})$.

thermore, `Stan` does not converge for any $k = 1, \ldots, K$, when sampling from $\pi(\boldsymbol{\Theta}|\tilde{z}_T^{(k)})$, i.e., $\left|\mathcal{K}^{\texttt{Stan}}\right| = 0$. This makes sense with the information we have on the type of the extrema of the posterior.

$(a^{true}, x_0^{true}) = (0.9, 8) : T = 100$. For $T = 100$, the optimization procedure still does not find any joint maximum of the log-posteriors $\pi(\boldsymbol{\Theta}|\tilde{z}_T^{(k)})$, $k = 1, \ldots, 4000$. However, the marginal approach converges for 1575 out of 4000 design observation trajectories. Hence, for $T = 100$, the model is observable, but the degree of observability is low.

F. Model `(Sep,d=p=1,Multiplicative)`

As the prior of a, we choose (14) again and $(a^{true}, x_0^{true}) = (0.9, 8)$. In contrast to the previous examples, we set the initial values for sampling in `Stan` to the true underlying values. We investigate $T = 50$ and $K = 4000$.

For the joint MAP approach, we obtain $|\mathcal{K}^{\max}| = 3995$. Using the marginal MAP approach, we obtain $\left|\mathcal{K}^{\texttt{Stan}}\right| = 2598$ and that the bias as well as the variance and the MSE are higher compared to the other models we considered. The estimated MSE values are better for the marginal MAP approach. While the values of the estimated bias and MSE of $x_t, t = 1, \ldots, T$, are high compared to the other examples, the coverage is satisfactory.

V. CONCLUSIONS AND OUTLOOK

For a statistical SSM, we propose a definition of observability based on the existence or non-existence of a chosen parameter estimator. In particular, we consider the marginal MAP estimator in this paper. We provide an algorithm to check this definition of observability in practice. The algorithm allows us to check the observability properties of very general SSMs with Gaussian disturbances. Furthermore, we do not only answer the question of observability, but we also provide a quantitative observability measure given by the values $|\mathcal{K}^{\max}|$ and $\left|\mathcal{K}^{\texttt{Stan}}\right|$, the local variance, as well as the properties of the joint and marginal MAP estimator.

In general, observability is not a global property for all observation trajectories and has to be checked for every realization of the observation trajectory with corresponding state trajectory. Given this insight, the key idea of the proposed algorithm is to use deterministic approximations of the distribution of observation trajectories and state trajectories. Our simulations show that one advantage over using random observations is that the necessary number of samples is reduced in order to guarantee convergence of the properties of the joint MAP estimator (see Model `(Same, d=p=1)`). To the best of the authors' knowledge, this is the first time that deterministic approximations are used in order to obtain observability properties of stochastic SSMs.

To obtain marginal MAP estimates, the programming language `Stan` is used to sample from the posterior density, which works well for the considered examples. Alternatives could be the iterated batch importance sampling algorithm [9], the particle Markov chain Monte Carlo algorithm [4], or the SMC^2 algorithm [10].

In numerical studies, we check the observability properties of interesting SSMs with random disturbances using the proposed definition of observability and our algorithm. The numerical studies show that the approach works and that using the marginal MAP estimator leads to satisfactory results for the considered examples. Note that the definition of observability can also be adapted to other types of estimators.

In this work, the initial state $\mathbf{x}_0$ is treated as an unknown fixed value. This information might, for example, be given in a scenario where a mobile robot always starts from the same position. However, our approach can be adapted to consider observability for unknown random values of $\mathbf{x}_0$. The next step is to extend this approach to copula SSMs [22], [23], where observability is still an open question. For that, the design values have to be modified.

VI. GLOSSARY OF DEFINITIONS

Notation	Explanation
$\mathcal{Z}_T = (\mathbf{Z}_1^\top, \ldots, \mathbf{Z}_T^\top)^\top \in \mathbb{R}^{T \cdot d}$	Random observation trajectory $\mathcal{Z}_T$ for $t = 1, \ldots, T$, consisting of the observations $\mathbf{Z}_t$ at time point t
$z_T = (\mathbf{z}_1^\top, \ldots, \mathbf{z}_T^\top)^\top$	Realization of $\mathcal{Z}_T$
$\mathcal{X}_T = (\mathbf{X}_1^\top, \ldots, \mathbf{X}_T^\top)^\top \in \mathbb{R}^{T \cdot p}$	Random trajectory $\mathcal{X}_T$ of the latent state, consisting of the latent states $\mathbf{X}_t$ at time point t
$x_T = (\mathbf{x}_1^\top, \ldots, \mathbf{x}_T^\top)^\top$	Realization of $\mathcal{X}_T$
$\mathbf{X}_0$	Initial state
$\mathbf{x}_0^{true}$	True underlying value of $\mathbf{X}_0$
$\boldsymbol{\Omega}$	Vector of unknown time-invariant parameters
$\boldsymbol{\Omega}^{true}$	True underlying value of $\boldsymbol{\Omega}$

Notation	Explanation
$(\boldsymbol{\epsilon}_t)_{t=1,\dots,T}$ and $(\boldsymbol{\eta}_t)_{t=1,\dots,T}$	Disturbances, serially independent and independent of each other at all time points with $\boldsymbol{\epsilon}_t \sim \mathcal{N}_d(\mathbf{0}, R_t)$ and $\boldsymbol{\eta}_t \sim \mathcal{N}_p(\mathbf{0}, Q_t)$, where $R_t \in \mathbb{R}^{d\times d}$ and $Q_t \in \mathbb{R}^{p\times p}$ are known covariance matrices
$\boldsymbol{\Theta}^{true} = (\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, \chi_T)$	Underlying unknown true parameters
$\Pi(\boldsymbol{\Theta}\|z_T)$	Joint posterior density
$\pi(\boldsymbol{\Theta}\|z_T)$	Log-posterior density
$\hat{\boldsymbol{\Theta}}(z_T) = \underset{\boldsymbol{\Theta}}{\operatorname{argmax}}\, \Pi(\boldsymbol{\Theta}\|z_T)$	Joint MAP estimate
$\hat{\boldsymbol{\Theta}}(z_T)_{\text{mar}}$	Marginal MAP estimate
$SO(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T)$	State-observation space
$\mathcal{E}(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T) \subset \mathbb{R}^{T(p+d)}$	Set of realizations z_T and corresponding χ_T such that $\hat{\boldsymbol{\Theta}}(z_T)$ can be uniquely recovered
$\hat{\boldsymbol{\Theta}}(Z_T)$	Joint MAP estimator
$\mathcal{E}^{\text{mar}}(\boldsymbol{\Omega}^{true}, \mathbf{x}_0^{true}, T) \subset \mathbb{R}^{T(p+d)}$	Set of realizations (χ_T, z_T) such that $\hat{\boldsymbol{\Theta}}(z_T)_{\text{mar}}$ can be uniquely recovered
$\hat{\boldsymbol{\Theta}}(Z_T)_{\text{mar}}$	Marginal MAP estimator
$\text{LVar}\left(\hat{\Theta}_j(z_T)\right)$	Local variance of $\hat{\Theta}_j(z_T)$
$(\tilde{\eta}_1^{(k)}, \dots, \tilde{\eta}_T^{(k)}, \tilde{\epsilon}_1^{(k)}, \dots, \tilde{\epsilon}_T^{(k)})$	Design disturbance vectors for $d = p = 1$
$\tilde{z}_T^{(k)} = (\tilde{\mathbf{z}}_1^{(k)}, \dots, \tilde{\mathbf{z}}_T^{(k)})$	Design observation trajectories
$\tilde{\chi}_T^{(k)} = (\tilde{\mathbf{x}}_1^{(k)}, \dots, \tilde{\mathbf{x}}_T^{(k)})$	Design state trajectory
$\mathcal{K}^{\max}$	Set of values k for which a maximum is recovered and $\|\mathcal{K}^{\max}\| \leq K$
$\boldsymbol{\Theta}(\tilde{z}_T^{(k)})^{(r)}$	MCMC samples
$\mathcal{K}^{\texttt{Stan}}$	Set of values k for which `Stan` converges and $\|\mathcal{K}^{\texttt{Stan}}\| \leq K$

REFERENCES

[1] https://mc-stan.org/docs/stan-users-guide/problematic-posteriors.html.

[2] https://mc-stan.org/misc/warnings.html. Accessed: March 10, 2024.

[3] L. A. Aguirre, L. L. Portes, and C. Letellier "Structural, dynamical and symbolic observability: From dynamical systems to networks," *PLOS ONE*, vol. 13, 2018, Art. no. e0206180.

[4] C. Andrieu, A. Doucet, and R. Holenstein "Particle Markov chain Monte Carlo methods," *J. Roy. Stat. Soc. Ser. B: Stat. Methodol.*, vol. 72, pp. 269–342, 2010.

[5] A. Balakrishnan and V. Peterka Identification in Automatic Control Systems, *Automatica*, vol. 5, pp. 817–829, 1969.

[6] A. V. Balakrishnan "Identification and adaptive control: An application to flight control systems," *J. Optim. Theory Appl.*, vol. 9, pp. 187–213, 1972.

[7] M. Betancourt "A. Conceptual Introduction to Hamiltonian Monte Carlo"2018, arXiv:1701.02434.

[8] B. Carpenter et al. "Stan: A probabilistic programming language," *J. Stat. Softw.*, vol. 76, pp. 1–32, 2017.

[9] N. Chopin "A sequential particle filter method for static models," *Biometrika*, vol. 89, pp. 539–552, 2002.

[10] N. Chopin, P. E. Jacob, and O. Papaspiliopoulos "SMC2: An efficient algorithm for sequential analysis of state space models," J. Roy. Statist. Soc., Ser. B, vol. 75, no. 3, pp. 397–426, 2013.

[11] O. Datcu, R. Tauleigne, A. Vlad, and J.-P. Barbot "Observability-singularity manifolds in the context of chaos based cryptography," in *Proc. 3rd Int. Conf. Syst. Control*, 2013, pp. 105–110.

[12] J. Durbin and S. J. Koopman *Time Series Analysis by State Space Methods*, 2nd revised ed. London, U.K.: Oxford University Press, 2012.

[13] J. Gauthier, H. Hammouri, and S. Othman "Simple observer for nonlinear systems applications to bioreactors," *IEEE Trans. Autom. Control*, vol. 37, pp. 875–880, 1992.

[14] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin Bayesian Data Analysis, 3rd ed. London, U.K: Chapman & Hall (Ltd.), 2013.

[15] D. Gerbet and K. Röbenack "'An algebraic approach to identifiability," *Algorithms*, vol. 14, pp. 255, 2021.

[16] P. E. Gill, W. Murray, and M. H. Wright *Practical Optimization.* London, U.K.: Emerald Group Publishing Limited, 1982.

[17] F. Hamelin, A. Iggidr, A. Rapaport, and G. Sallet *Observability, Identifiability Epidemiol. a Surv.*, arXiv:2011.12202 [math], (2021), https://arxiv.org/abs/2011.12202.

[18] A. Hanebeck and C. Czado "On the observability of Gaussian models using discrete density approximations," *Proc. 25th Int. Conf. Inf. Fusion*, 2022, pp. 1–8.

[19] U. Hanebeck and V. Klumpp "Localized cumulative distributions and a multivariate generalization of the Cramér-Von Mises distance," in *IEEE Int. Conf. Multisensor Fusion Integration Intell. Syst.*, 2008, pp. 33–39.

[20] R. Hermann and A. Krener "Nonlinear controllability and observability," *IEEE Trans. Autom. Control*, vol. 22, pp. 728–740, Oct. 1977.

[21] M. D. Hoffman and A. Gelman "The No-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo," *J. Mach. Learn. Res.*, vol. 15, pp. 1593–1623, 2014.

[22] A. Kreuzer, L. Dalla Valle, and C. Czado "A Bayesian non-linear state space copula model for air pollution in Beijing," *J. Roy. Stat. Society: Ser. C. (Applied Statistics)*, vol. 71, no. 3, pp. 613–638, 2022.

[23] A. Kreuzer, L. Dalla Valle, and C. Czado "Bayesian multivariate nonlinear state space copula models," *Comput. Statist. Data Anal.*, vol. 188, 2023, Art. no. 107820.

[24] K. Langueh, O. Datcu, J.-P. Barbot, G. Zheng, and K. Busawon "Observabilty Singularities and Observer Design: Dual Immersion Approach," *IFAC-PapersOnLine*, vol. 49, pp. 511–516, 2016.

[25] R. Mohler and W. Kolodziej "An overview of bilinear system theory and applications," *IEEE Trans. Syst., Man, Cybern.*, vol. 10, pp. 683–688, Oct. 1980.

[26] J. Nocedal and S. J. Wright Numerical optimization, in *Springer Series in Operations Research* T. V. Mikosch, S. Resnick, B. Zwart, and T. Dieker, Eds., 2nd ed., Berlin, Germany: Springer-Verlag, 2006.

[27] A. O'Hagan "On Posterior Joint and Marginal Modes," *Biometrika*, vol.

63, pp. 329–333, 1976.
[28] T. Paradowski, S. Lerch, M. Damaszek, R. Dehnert, and B. Tibken
"Observability of uncertain nonlinear systems using interval analysis,"
Algorithms, vol. 13, pp. 66, 2020.
[29] L. Polansky, P. de Valpine, J. O. Lloyd-Smith, and W. M. Getz
"Likelihood ridges and multimodality in population growth rate models,"
Ecology, vol. 90, pp. 2313–2320, 2009.
[30] A. E. Raftery and L. Bao
"Estimating and projecting trends in HIV/AIDS generalized epidemics using incremental mixture importance sampling,"
Biometrics, vol. 66, pp. 1162–1173, 2010.
[31] B. Silverman
"Density estimation for statistics and data analysis,"
Biometrical J., vol. 30, pp. 876–877, 1986.
[32] J. Steinbring, M. Pander, and U. D. Hanebeck
"The smart sampling Kalman filter with symmetric samples,"
J. Adv. Inf. Fusion, vol. 11, pp. 71–90, 2016.
[33] J. M. Tabeart, S. L. Dance, A. S. Lawless, N. K. Nichols, and J. A. Waller
"New bounds on the condition number of the Hessian of the preconditioned variational data assimilation problem,"
Numer. Linear Algebra with Appl., vol. 29 2022.
[34] W. N. van Wieringen
"Lecture notes on ridge regression," 2021, arXiv:1509.09169.
[35] A. Vehtari, A. Gelman, D. Simpson, B. Carpenter, and P.-C Bürkner
"Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of MCMC (with discussion),"
Bayesian Anal., vol. 16, pp. 667–718, 2021.
[36] A. F. Villaverde
"Observability and structural identifiability of nonlinear biological systems,
Complexity," vol. 2019, 2019, Art. no. e8497093.
[37] E. Walter
Identifiability of State Space Models with Applications to Transformation Systems, no. 46 in
Lecture Notes in Biomathematics, vol. 46. Berlin, Germany: Springer-Verl.

Ariane Hanebeck received the M.Sc. degree in mathematics from Karlsruhe Institute of Technology, Karlsruhe, Germany, in 2020. She is currently pursuing the Ph.D. degree with the Chair of Applied Mathematical Statistics, Technical University of Munich, Germany. Her research interests include copula-based state space models and MCMC sampling for vine copula models. Applications are in the fields of mortality analysis and flight safety.

Claudia Czado received the Ph.D. degree in operations research and industrial engineering from Cornell University, Ithaca, NY, USA, in 1989. Since 1998, she has held an Associate Professorship of Applied Mathematical Statistics at the Technical University of Munich, Germany. Her research interests include statistical dependence modeling using vine copulas (vine-copula.org) and their application to engineering, finance, and health.