2011 American Control Conference
on O'Farrell Street, San Francisco, CA, USA
June 29 - July 01, 2011

A General Perspective on Gaussian Filtering and Smoothing: Explaining Current and Deriving New Algorithms

Marc Peter Deisenroth

Henrik Ohlsson

***Abstract*— We present a general probabilistic perspective on Gaussian filtering and smoothing. This allows us to show that common approaches to Gaussian filtering/smoothing can be distinguished solely by their methods of computing/approximating the means and covariances of joint probabilities. This implies that novel filters and smoothers can be derived straightforwardly by providing methods for computing these moments. Based on this insight, we derive the cubature Kalman smoother and propose a novel robust filtering and smoothing algorithm based on Gibbs sampling.**

I. Introduction

Filtering and smoothing in latent variable time series models, including hidden Markov models and dynamic systems, have played an important role in signal processing, control, and machine learning for decades [12], [15], [3].

In the context of dynamic systems, filtering is widely used in control and robotics for online Bayesian state estimation [18], while smoothing is commonly used in machine learning algorithms for parameter learning [3]. For computational efficiency reasons, many filters and smoothers approximate appearing probability distributions by Gaussians. This is why they are referred to as *Gaussian filters/smoothers*.

In this paper, we discuss Gaussian filtering and Rauch-Tung-Striebel (RTS) smoothing for nonlinear systems from a general probabilistic perspective, initially without focusing on particular implementations. We identify the high-level concepts for filtering and smoothing, while avoiding getting lost in computational details of particular algorithms (see, for example, the standard derivations of the Kalman filter [1], [18]). Based on these high-level concepts, we show that Gaussian filters/smoothers for (non)linear systems (including common algorithms such as the extended Kalman filter (EKF) [15], the cubature Kalman filter (CKF) [2], or the unscented Kalman filter (UKF) [11]) can be distinguished by their means to computing Gaussian approximations of one/two joint probability distributions. Our results also imply that novel filtering and smoothing algorithms can be derived straightforwardly, given a method to determining the moments of these joint distributions. Using this insight, we present and analyze the cubature Kalman smoother (CKS) and a filter and an RTS smoother based on Gibbs sampling.

MP Deisenroth is with the Department of Computer Science & Engineering, University of Washington, Seattle, USA, and the Department of Engineering, University of Cambridge, UK. H Ohlsson is with the Department of Electrical Engineering, Linköping University, Sweden.

MPD has been supported by ONR MURI grant N00014-09-1-1052 and by Intel Labs. HO has been partially supported by the Swedish foundation for strategic research in the center MOVIII and by the Swedish Research Council in the Linnaeus center CADICS.

The authors are grateful to Shakir Mohamed for valuable discussions.

II. Setup and Notation

We consider discrete-time stochastic dynamic systems of the form

$$\mathbf{x}_t = f(\mathbf{x}_{t-1}) + \mathbf{w}_t\,, \tag{1}$$

$$\mathbf{z}_t = g(\mathbf{x}_t) + \mathbf{v}_t\,, \tag{2}$$

where $\mathbf{x}_t \in \mathbb{R}^D$ is the state, $\mathbf{z}_t \in \mathbb{R}^E$ is the measurement at time step $t = 1, \ldots, T$, $\mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$ is i.i.d. Gaussian system noise, $\mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ is i.i.d. Gaussian measurement noise, f is the transition/system function and g is the measurement function. The noise covariance matrices $\mathbf{Q}$, $\mathbf{R}$, the system function f, and the measurement function g are assumed known. If not stated otherwise, we assume nonlinear functions f and g. The initial state $\mathbf{x}_0$ of the time series is distributed according to a Gaussian prior distribution $p(\mathbf{x}_0) = \mathcal{N}(\boldsymbol{\mu}_0^x, \boldsymbol{\Sigma}_0^x)$. The purpose of filtering and smoothing is to find approximations to the posterior distributions $p(\mathbf{x}_t|\mathbf{z}_{1:\tau})$, where a subscript $1:\tau$ abbreviates $1, \ldots, \tau$, with $\tau = t$ for filtering and $\tau = T$ for smoothing.

In this paper, we consider Gaussian approximations $\mathcal{N}(\mathbf{x}_t\,|\,\boldsymbol{\mu}_{t|\tau}^x, \boldsymbol{\Sigma}_{t|\tau}^x)$ of the latent state posteriors $p(\mathbf{x}_t|\mathbf{z}_{1:\tau})$. We use the shorthand notation $\mathbf{a}_{b|c}^d$ where $\mathbf{a} = \boldsymbol{\mu}$ denotes the mean $\boldsymbol{\mu}$ and $\mathbf{a} = \boldsymbol{\Sigma}$ denotes the covariance, b denotes the time step under consideration, c denotes the time step up to which we consider measurements, and $d \in \{x, z\}$ denotes either the latent space (x) or the observed space (z).

III. Gaussian Filtering

Given a prior $p(\mathbf{x}_0)$ on the initial state and a dynamic system (e.g., Eqs. (1)–(2)), the objective of *filtering* is to infer a posterior distribution $p(\mathbf{x}_t|\mathbf{z}_{1:t})$ of the hidden state $\mathbf{x}_t$, $t = 1, \ldots, T$, incorporating the evidence of the measurements $\mathbf{z}_{1:t}$. Specific for Gaussian filtering is that posterior distributions are approximated by Gaussians [18]. Approximations are required since generally a Gaussian distribution mapped through a nonlinear function does not stay Gaussian.

Assume a Gaussian filter distribution $p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1}) = \mathcal{N}(\boldsymbol{\mu}_{t-1|t-1}^x, \boldsymbol{\Sigma}_{t-1|t-1}^x)$ is given (if not, we employ the prior $p(\mathbf{x}_0) = p(\mathbf{x}_0|\emptyset) = \mathcal{N}(\boldsymbol{\mu}_{0|\emptyset}^x, \boldsymbol{\Sigma}_{0|\emptyset}^x)$) on the initial state. Using Bayes' theorem, the filter distribution at time t is

$$p(\mathbf{x}_t|\mathbf{z}_{1:t}) = \frac{p(\mathbf{x}_t, \mathbf{z}_t|\mathbf{z}_{1:t-1})}{p(\mathbf{z}_t|\mathbf{z}_{1:t-1})} \propto p(\mathbf{z}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{z}_{1:t-1})\,. \tag{3}$$

Proposition 1 (Filter Distribution): Gaussian filters approximate the filter distribution $p(\mathbf{x}_t|\mathbf{z}_{1:t})$ using a Gaussian

distribution $\mathcal{N}(\boldsymbol{\mu}^x_{t|t}, \boldsymbol{\Sigma}^x_{t|t})$. The moments of this approximation are in general computed through

$$\boldsymbol{\mu}^x_{t|t} = \hat{\boldsymbol{\mu}}^x_{t|t-1} + \hat{\boldsymbol{\Sigma}}^{xz}_{t|t-1}(\hat{\boldsymbol{\Sigma}}^z_{t|t-1})^{-1}(\mathbf{z}_t - \hat{\boldsymbol{\mu}}^z_{t|t-1}), \quad (4)$$

$$\boldsymbol{\Sigma}^x_{t|t} = \hat{\boldsymbol{\Sigma}}^x_{t|t-1} - \hat{\boldsymbol{\Sigma}}^{xz}_{t|t-1}(\hat{\boldsymbol{\Sigma}}^z_{t|t-1})^{-1}\hat{\boldsymbol{\Sigma}}^{zx}_{t|t-1}. \quad (5)$$

Since the true moments of the joint distribution $p(\mathbf{x}_t, \mathbf{z}_t|\mathbf{z}_{1:t-1})$ can in general not be computed analytically, approximations/estimates are used (hence the $\hat{}$-symbols).

Proof: Generally, filtering proceeds by alternating between predicting (*time update*) and correcting (*measurement update*) [1], [18]:

1) Time update (predictor)
   a) Compute the predictive distribution $p(\mathbf{x}_t|\mathbf{z}_{1:t-1})$.
2) Measurement update (corrector)
   a) Compute the joint distribution $p(\mathbf{x}_t, \mathbf{z}_t|\mathbf{z}_{1:t-1})$ of the next latent state and the next measurement.
   b) Measure $\mathbf{z}_t$.
   c) Compute the posterior $p(\mathbf{x}_t|\mathbf{z}_{1:t})$.

In the following, we detail these steps to prove Prop. 1.

1) Time Update (Predictor):

(a) Compute the predictive distribution $p(\mathbf{x}_t|\mathbf{z}_{1:t-1})$. The predictive distribution of state $\mathbf{x}$ at time t given the evidence of measurements up to time $t-1$ is

$$p(\mathbf{x}_t|\mathbf{z}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1})\,\mathrm{d}\mathbf{x}_{t-1}, \quad (6)$$

where $p(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t \mid f(\mathbf{x}_{t-1}), \mathbf{Q})$ is the transition probability. In Gaussian filters, the predictive distribution $p(\mathbf{x}_t|\mathbf{z}_{1:t-1})$ in Eq. (6) is approximated by a Gaussian distribution, whose exact mean and covariance are

$$\boldsymbol{\mu}^x_{t|t-1} \coloneqq \mathbb{E}_{\mathbf{x}_t}[\mathbf{x}_t|\mathbf{z}_{1:t-1}] = \mathbb{E}_{\mathbf{x}_{t-1},\mathbf{w}_t}[f(\mathbf{x}_{t-1}) + \mathbf{w}_t|\mathbf{z}_{1:t-1}]$$
$$= \int f(\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1})\,\mathrm{d}\mathbf{x}_{t-1}, \quad (7)$$

$$\boldsymbol{\Sigma}^x_{t|t-1} = \int f(\mathbf{x}_{t-1})f(\mathbf{x}_{t-1})^\top p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1})\,\mathrm{d}\mathbf{x}_{t-1} - \boldsymbol{\mu}^x_{t|t-1}(\boldsymbol{\mu}^x_{t|t-1})^\top + \mathbf{Q}, \quad (8)$$

respectively. In Eq. (7), we exploited that the noise term $\mathbf{w}_t$ in Eq. (1) has mean zero and is independent. A Gaussian approximation to the time update $p(\mathbf{x}_t|\mathbf{z}_{1:t-1})$ is then given by $\mathcal{N}(\mathbf{x}_t \mid \boldsymbol{\mu}^x_{t|t-1}, \boldsymbol{\Sigma}^x_{t|t-1})$.

2) Measurement Update (Corrector):

(a) Compute the joint distribution

$$p(\mathbf{x}_t, \mathbf{z}_t|\mathbf{z}_{1:t-1}) = p(\mathbf{z}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{z}_{1:t-1}). \quad (9)$$

In Gaussian filters, a Gaussian approximation to this joint is an intermediate step toward the desired Gaussian approximation of the posterior $p(\mathbf{x}_t|\mathbf{z}_{1:t})$. If the mean and the covariance of the joint in Eq. (9) can be computed or estimated, the desired filter distribution corresponds to the conditional $p(\mathbf{x}_t|\mathbf{z}_{1:t})$ and is given in closed form [3]. Our objective is to compute a Gaussian approximation

$$\mathcal{N}\left(\begin{bmatrix}\boldsymbol{\mu}^x_{t|t-1}\\ \boldsymbol{\mu}^z_{t|t-1}\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}^x_{t|t-1} & \boldsymbol{\Sigma}^{xz}_{t|t-1}\\ \boldsymbol{\Sigma}^{zx}_{t|t-1} & \boldsymbol{\Sigma}^z_{t|t-1}\end{bmatrix}\right) \quad (10)$$

to the joint $p(\mathbf{x}_t, \mathbf{z}_t|\mathbf{z}_{1:t-1})$ in Eq. (9). Since a Gaussian approximation $\mathcal{N}(\boldsymbol{\mu}^x_{t|t-1}, \boldsymbol{\Sigma}^x_{t|t-1})$ to the marginal $p(\mathbf{x}_t|\mathbf{z}_{1:t-1})$ is known from the time update, it remains to compute the marginal $p(\mathbf{z}_t|\mathbf{z}_{1:t-1})$ and the cross-covariance $\boldsymbol{\Sigma}^{xz}_{t|t-1} \coloneqq \mathrm{cov}_{\mathbf{x}_t,\mathbf{z}_t}[\mathbf{x}_t, \mathbf{z}_t|\mathbf{z}_{1:t-1}]$.

- The marginal $p(\mathbf{z}_t|\mathbf{z}_{1:t-1})$ of the joint in Eq. (10) is

$$p(\mathbf{z}_t|\mathbf{z}_{1:t-1}) = \int p(\mathbf{z}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{z}_{1:t-1})\,\mathrm{d}\mathbf{x}_t,$$

where the state $\mathbf{x}_t$ is integrated out according to the time update $p(\mathbf{x}_t|\mathbf{z}_{1:t-1})$. The measurement Eq. (2), yields $p(\mathbf{z}_t|\mathbf{x}_t) = \mathcal{N}(g(\mathbf{x}_t), \mathbf{R})$. Hence, the exact mean of the marginal is

$$\boldsymbol{\mu}^z_{t|t-1} \coloneqq \mathbb{E}_{\mathbf{z}_t}[\mathbf{z}_t|\mathbf{z}_{1:t-1}] = \mathbb{E}_{\mathbf{x}_t}[g(\mathbf{x}_t)|\mathbf{z}_{1:t-1}]$$
$$= \int g(\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{z}_{1:t-1})\,\mathrm{d}\mathbf{x}_t \quad (11)$$

since the noise term $\mathbf{v}_t$ in the measurement Eq. (2) is independent and has zero mean. Similarly, the exact covariance of the marginal $p(\mathbf{z}_t|\mathbf{z}_{1:t-1})$ is

$$\boldsymbol{\Sigma}^z_{t|t-1} = \int g(\mathbf{x}_t)g(\mathbf{x}_t)^\top p(\mathbf{x}_t|\mathbf{z}_{1:t-1})\,\mathrm{d}\mathbf{x}_t - \boldsymbol{\mu}^z_{t|t-1}(\boldsymbol{\mu}^z_{t|t-1})^\top + \mathbf{R}. \quad (12)$$

Hence, a Gaussian approximation to the marginal measurement distribution $p(\mathbf{z}_t|\mathbf{z}_{1:t-1})$ is

$$\mathcal{N}(\mathbf{z}_t \mid \boldsymbol{\mu}^z_{t|t-1}, \boldsymbol{\Sigma}^z_{t|t-1}), \quad (13)$$

with the mean and covariance given in Eqs. (11) and (12), respectively.

- Due to the independence of $\mathbf{v}_t$, the exact cross-covariance terms of the joint in Eq. (10) are

$$\boldsymbol{\Sigma}^{xz}_{t|t-1} = \mathrm{cov}_{\mathbf{x}_t,\mathbf{z}_t}[\mathbf{x}_t, \mathbf{z}_t|\mathbf{z}_{1:t-1}]$$
$$= \iint \mathbf{x}_t\mathbf{z}_t^\top p(\mathbf{x}_t, \mathbf{z}_t|\mathbf{z}_{1:t-1})\,\mathrm{d}\mathbf{z}_t\,\mathrm{d}\mathbf{x}_t - \boldsymbol{\mu}^x_{t|t-1}(\boldsymbol{\mu}^z_{t|t-1})^\top.$$

Plugging in the measurement Eq. (2), we obtain

$$\boldsymbol{\Sigma}^{xz}_{t|t-1} = \int \mathbf{x}_t g(\mathbf{x}_t)^\top p(\mathbf{x}_t|\mathbf{z}_{1:t-1})\,\mathrm{d}\mathbf{x}_t - \boldsymbol{\mu}^x_{t|t-1}(\boldsymbol{\mu}^z_{t|t-1})^\top. \quad (14)$$

(b) Measure $\mathbf{z}_t$.

(c) Compute a Gaussian approximation of the posterior $p(\mathbf{x}_t|\mathbf{z}_{1:t})$. This boils down to computing a conditional from the Gaussian approximation to the joint distribution $p(\mathbf{x}_t, \mathbf{z}_t|\mathbf{z}_{1:t-1})$ in Eq. (10). The expressions from Eqs. (7), (8), (11), (12), and (14), yield a Gaussian approximation $\mathcal{N}(\mathbf{x}_t \mid \boldsymbol{\mu}^x_{t|t}, \boldsymbol{\Sigma}^x_{t|t})$ to the filter distribution $p(\mathbf{x}_t|\mathbf{z}_{1:t})$, where

$$\boldsymbol{\mu}^x_{t|t} = \boldsymbol{\mu}^x_{t|t-1} + \boldsymbol{\Sigma}^{xz}_{t|t-1}\big(\boldsymbol{\Sigma}^z_{t|t-1}\big)^{-1}(\mathbf{z}_t - \boldsymbol{\mu}^z_{t|t-1}), \quad (15)$$

$$\boldsymbol{\Sigma}^x_{t|t} = \boldsymbol{\Sigma}^x_{t|t-1} - \boldsymbol{\Sigma}^{xz}_{t|t-1}\big(\boldsymbol{\Sigma}^z_{t|t-1}\big)^{-1}\boldsymbol{\Sigma}^{zx}_{t|t-1}. \quad (16)$$

Generally, the required integrals in Eqs. (7), (8), (11), (12), and (14) cannot be computed analytically. Hence, approximations of the moments are typically used in Eqs. (15) and (16). This concludes the proof of Prop. 1. ■

A. Sufficient Conditions for Gaussian Filtering

In any Bayes filter [18], the sufficient components to computing the Gaussian filter distribution in Eqs. (15) and (16) are the mean and the covariance of the joint distribution $p(\mathbf{x}_t, \mathbf{z}_t|\mathbf{z}_{1:t-1})$. Generally, the required integrals in Eqs. (7), (8), (11), (12), and (14) cannot be computed analytically. One exception are *linear* functions f and g, where the analytic solutions to the integrals are embodied in the Kalman filter [12]. In many *nonlinear* dynamic systems, filtering algorithms approximate probability distributions (see e.g., the UKF [11] and the CKF [2]) or the functions f and g (see e.g., the EKF [15] or the GP-Bayes filters [6], [13]). Using the means and (cross-)covariances computed by these algorithms and plugging them into Eqs. (15)–(16), recovers the corresponding filter update equations for the EKF, the UKF, the CKF, and the GP-Bayes filters.

IV. Gaussian RTS Smoothing

In this section, we present a general probabilistic perspective on Gaussian RTS smoothers and derive sufficient conditions for Gaussian smoothing.

The smoothed state distribution is the posterior distribution of the hidden state given *all* measurements

$$p(\mathbf{x}_t|\mathbf{z}_{1:T})\,,\ t = T, \dots, 0\,. \tag{17}$$

Proposition 2 (Smoothing Distribution): For Gaussian smoothers, the mean and the covariance of a Gaussian approximation to the distribution $p(\mathbf{x}_t|\mathbf{z}_{1:T})$ are generally computed as

$$\boldsymbol{\mu}^x_{t-1|T} = \hat{\boldsymbol{\mu}}^x_{t-1|t-1} + \hat{\mathbf{J}}_{t-1}(\hat{\boldsymbol{\mu}}^x_{t|T} - \hat{\boldsymbol{\mu}}^x_{t|t-1})\,, \tag{18}$$

$$\boldsymbol{\Sigma}^x_{t-1|T} = \hat{\boldsymbol{\Sigma}}^x_{t-1|t-1} + \hat{\mathbf{J}}_{t-1}(\hat{\boldsymbol{\Sigma}}^x_{t|T} - \hat{\boldsymbol{\Sigma}}^x_{t|t-1})\hat{\mathbf{J}}^\top_{t-1}\,, \tag{19}$$

$$\begin{aligned}\mathbf{J}_{t-1} &= \operatorname{cov}[\mathbf{x}_{t-1}, \mathbf{x}_t|\mathbf{z}_{1:t-1}]\operatorname{cov}[\mathbf{x}_t|\mathbf{z}_{1:t-1}]^{-1} \\ &= \boldsymbol{\Sigma}^x_{t-1,t|t-1}(\boldsymbol{\Sigma}^x_{t|t-1})^{-1}\,. \end{aligned}\tag{20}$$

Proof: The smoothed state distribution at the terminal time step T is equivalent to the filter distribution $p(\mathbf{x}_T|\mathbf{z}_{1:T})$ [1], [3]. The distributions $p(\mathbf{x}_{t-1}|\mathbf{z}_{1:T})$, $t = T, \dots, 1$, of the smoothed states can be computed recursively according to

$$p(\mathbf{x}_{t-1}|\mathbf{z}_{1:T}) = \int p(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{z}_{1:t-1})p(\mathbf{x}_t|\mathbf{z}_{1:T})\,\mathrm{d}\mathbf{x}_t \tag{21}$$

by integrating out the smoothed hidden state at time step t. In Eq. (21), we exploited that $\mathbf{x}_{t-1}$ is conditionally independent of the future measurements $\mathbf{z}_{t:T}$ given $\mathbf{x}_t$.

To compute the smoothed state distribution in Eq. (21), we need to multiply a distribution in $\mathbf{x}_t$ with a distribution in $\mathbf{x}_{t-1}$ and integrate over $\mathbf{x}_t$. To do so, we follow the steps:

(a) Compute the conditional $p(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{z}_{1:t-1})$.
(b) Formulate $p(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{z}_{1:T})$ as an unnormalized distribution in $\mathbf{x}_t$.
(c) Multiply the new distribution with $p(\mathbf{x}_t|\mathbf{z}_{1:T})$.
(d) Solve the integral in Eq. (21).

We now examine these steps in detail. Assume a known (Gaussian) smoothed state distribution $p(\mathbf{x}_t|\mathbf{z}_{1:T})$.

(a) Compute a Gaussian approximation to the conditional $p(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{z}_{1:t-1})$. We compute the conditional in two steps: First, we compute a Gaussian approximation to the joint distribution $p(\mathbf{x}_t, \mathbf{x}_{t-1}|\mathbf{z}_{1:t-1})$. Second, we apply the rules of computing conditionals to this joint Gaussian. Let us start with a Gaussian approximation

$$\mathcal{N}\left(\begin{bmatrix}\boldsymbol{\mu}^x_{t-1|t-1}\\ \boldsymbol{\mu}^x_{t|t-1}\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}^x_{t-1|t-1} & \boldsymbol{\Sigma}^x_{t-1,t|t-1}\\ (\boldsymbol{\Sigma}^x_{t-1,t|t-1})^\top & \boldsymbol{\Sigma}^x_{t|t-1}\end{bmatrix}\right) \tag{22}$$

to the joint $p(\mathbf{x}_{t-1}, \mathbf{x}_t|\mathbf{z}_{1:t-1})$ and have a closer look at its components: A Gaussian approximation of the filter distribution $p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1})$ at time step $t-1$ is known and is the first marginal distribution in Eq. (22). The second marginal $\mathcal{N}(\boldsymbol{\mu}^x_{t|t-1}, \boldsymbol{\Sigma}^x_{t|t-1})$ is the time update and also known from filtering. To fully determine the joint in Eq. (22), we require the cross-covariance matrix

$$\begin{aligned}\boldsymbol{\Sigma}^x_{t-1,t|t-1} = &\iint \mathbf{x}_{t-1} f(\mathbf{x}_{t-1})^\top p(\mathbf{x}_{t-1}|\mathbf{x}_{t-1})\,\mathrm{d}\mathbf{x}_{t-1} \\ &- \boldsymbol{\mu}^x_{t-1|t-1}(\boldsymbol{\mu}^x_{t|t-1})^\top\,,\end{aligned}\tag{23}$$

where we used the means $\boldsymbol{\mu}^x_{t-1|t-1}$ and $\boldsymbol{\mu}^x_{t|t-1}$ of the measurement update and the time update, respectively. The zero-mean independent noise in the system Eq. (1) does not influence the cross-covariance matrix. The cross-covariance matrix in Eq. (23) can be pre-computed during filtering since it does not depend on future measurements. This concludes the first step (computation of the joint Gaussian) of the computation of the desired conditional. In the second step, we apply the rules of Gaussian conditioning to obtain the desired conditional distribution $p(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{z}_{1:t-1})$. For a shorthand notation, we define

$$\mathbf{J}_{t-1} \coloneqq \boldsymbol{\Sigma}^x_{t-1,t|t-1}(\boldsymbol{\Sigma}^x_{t|t-1})^{-1}\,, \tag{24}$$

and obtain a Gaussian approximation $\mathcal{N}(\mathbf{x}_{t-1}\,|\,\mathbf{m}, \mathbf{S})$ of the conditional distribution $p(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{z}_{1:t-1})$ with

$$\mathbf{m} = \boldsymbol{\mu}^x_{t-1|t-1} + \mathbf{J}_{t-1}(\mathbf{x}_t - \boldsymbol{\mu}^x_{t|t-1})\,, \tag{25}$$

$$\mathbf{S} = \boldsymbol{\Sigma}^x_{t-1|t-1} - \mathbf{J}_{t-1}(\boldsymbol{\Sigma}^x_{t-1,t|t-1})^\top\,. \tag{26}$$

(b) Formulate $\mathcal{N}(\mathbf{x}_{t-1}\,|\,\mathbf{m}, \mathbf{S})$ as an unnormalized distribution in $\mathbf{x}_t$. The square-root of the exponent of $\mathcal{N}(\mathbf{x}_{t-1}\,|\,\mathbf{m}, \mathbf{S})$ contains

$$\mathbf{x}_{t-1} - \mathbf{m} = \mathbf{r}(\mathbf{x}_{t-1}) - \mathbf{J}_{t-1}\mathbf{x}_t$$

with $\mathbf{r}(\mathbf{x}_{t-1}) = \mathbf{x}_{t-1} - \boldsymbol{\mu}^x_{t-1|t-1} + \mathbf{J}_{t-1}\boldsymbol{\mu}^x_{t|t-1}$, which is a linear function of both $\mathbf{x}_{t-1}$ and $\mathbf{x}_t$. We now reformulate the conditional Gaussian $\mathcal{N}(\mathbf{x}_{t-1}\,|\,\mathbf{m}, \mathbf{S})$ as a Gaussian in $\mathbf{J}_{t-1}\mathbf{x}_t$ with mean $\mathbf{r}(\mathbf{x}_{t-1})$ and the unchanged covariance matrix $\mathbf{S}$. We obtain the conditional

$$\begin{aligned}\mathcal{N}(\mathbf{x}_{t-1}\,|\,\mathbf{m}, \mathbf{S}) &= c_1\mathcal{N}(\mathbf{x}_t\,|\,\mathbf{a}, \mathbf{A})\,, \\ \text{with}\quad c_1 &= \sqrt{|2\pi(\mathbf{J}^\top_{t-1}\mathbf{S}^{-1}\mathbf{J}_{t-1})^{-1}|/|2\pi\mathbf{S}|}\,,\end{aligned}\tag{27}$$

and $\mathbf{a} = \mathbf{J}^{-1}_{t-1}\mathbf{r}(\mathbf{x}_{t-1})\,, \mathbf{A} = (\mathbf{J}^\top_{t-1}\mathbf{S}^{-1}\mathbf{J}_{t-1})^{-1}$. Note that $\mathcal{N}(\mathbf{x}_{t-1}\,|\,\mathbf{m}, \mathbf{S})$ is an unnormalized Gaussian in $\mathbf{x}_t$, see Eq. (27). The matrix $\mathbf{J}_{t-1}$ defined in Eq. (24) is quadratic, but not necessarily invertible, in which case we

take the pseudo-inverse. However, we will see that this inversion will be unnecessary to obtain the final result.

(c) Multiply the new distribution with $p(\mathbf{x}_t|\mathbf{z}_{1:T})$. To determine $p(\mathbf{x}_{t-1}|\mathbf{z}_{1:T})$, we multiply the Gaussian in Eq. (27) with the smoothed Gaussian state distribution $\mathcal{N}(\mathbf{x}_t \,|\, \boldsymbol{\mu}^x_{t|T}, \boldsymbol{\Sigma}^x_{t|T})$, which yields the Gaussian approximation

$$c_1\mathcal{N}(\mathbf{x}_t \,|\, \mathbf{a}, \mathbf{A})\mathcal{N}(\mathbf{x}_t \,|\, \boldsymbol{\mu}^x_{t|T}, \boldsymbol{\Sigma}^x_{t|T}) = c_1 c_2(\mathbf{a})\mathcal{N}(\mathbf{x}_t \,|\, \mathbf{b}, \mathbf{B}) \tag{28}$$

of $p(\mathbf{x}_{t-1}, \mathbf{x}_t|\mathbf{z}_{1:T})$, for some $\mathbf{b}$, $\mathbf{B}$, where $c_2(\mathbf{a})$ is the inverse normalization constant of $\mathcal{N}(\mathbf{x}_t \,|\, \mathbf{b}, \mathbf{B})$.

(d) Solve the integral in Eq. (21). Since we integrate over $\mathbf{x}_t$ in Eq. (21), we are solely interested in the parts that make Eq. (28) unnormalized, i.e., the constants c_1 and $c_2(\mathbf{a})$, which are independent of $\mathbf{x}_t$. The constant $c_2(\mathbf{a})$ in Eq. (28) can be rewritten as $c_2(\mathbf{x}_{t-1})$ by reversing the step that inverted the matrix $\mathbf{J}_{t-1}$, see Eq. (27). Then, $c_2(\mathbf{x}_{t-1})$ is given by

$$c_2(\mathbf{x}_{t-1}) = c_1^{-1}\mathcal{N}(\mathbf{x}_{t-1} \,|\, \boldsymbol{\mu}^x_{t-1|T}, \boldsymbol{\Sigma}^x_{t-1|T})\,, \tag{29}$$

$$\boldsymbol{\mu}^x_{t-1|T} = \boldsymbol{\mu}^x_{t-1|t-1} + \mathbf{J}_{t-1}(\boldsymbol{\mu}^x_{t|T} - \boldsymbol{\mu}^x_{t|t-1})\,, \tag{30}$$

$$\boldsymbol{\Sigma}^x_{t-1|T} = \boldsymbol{\Sigma}^x_{t-1|t-1} + \mathbf{J}_{t-1}(\boldsymbol{\Sigma}^x_{t|T} - \boldsymbol{\Sigma}^x_{t|t-1})\mathbf{J}^\top_{t-1}\,. \tag{31}$$

Since $c_1 c_1^{-1} = 1$ (plug Eq. (29) into Eq. (28)), the desired smoothed state distribution is

$$p(\mathbf{x}_{t-1}|\mathbf{z}_{1:T}) = \mathcal{N}(\mathbf{x}_{t-1} \,|\, \boldsymbol{\mu}^x_{t-1|T}, \boldsymbol{\Sigma}^x_{t-1|T})\,, \tag{32}$$

where the mean and the covariance are given in Eq. (30) and Eq. (31), respectively.

This result concludes the proof of Prop. 2. ■

A. Sufficient Conditions for Smoothing

After filtering, to determine a Gaussian approximation to the distribution $p(\mathbf{x}_{t-1}|\mathbf{z}_{1:T})$ of the smoothed state at time $t-1$, only a few additional ingredients are required: the matrix $\mathbf{J}_{t-1}$ in Eq. (24) and Gaussian approximations to the smoothed state distribution $p(\mathbf{x}_t|\mathbf{z}_{1:T})$ at time t and the predictive distribution $p(\mathbf{x}_t|\mathbf{z}_{1:t-1})$. Everything but the matrix $\mathbf{J}_{t-1}$ can be precomputed either during filtering or in a previous step of the smoothing recursion. Note that $\mathbf{J}_{t-1}$ can also be precomputed during filtering.

Hence, for Gaussian RTS smoothing it is sufficient to determine Gaussian approximations to both the joint distribution $p(\mathbf{x}_t, \mathbf{z}_t|\mathbf{z}_{1:t-1})$ of the state and the measurement for the filter step and the joint distribution $p(\mathbf{x}_{t-1}, \mathbf{x}_t|\mathbf{z}_{1:t-1})$ of two consecutive states.

V. Implications and Theoretical Results

Using the results from Secs. III and IV, we conclude that for filtering and RTS smoothing it is sufficient to compute or estimate the means and the covariances of the joint distribution $p(\mathbf{x}_{t-1}, \mathbf{x}_t|\mathbf{z}_{1:t-1})$ between two consecutive states (smoothing) and the joint distribution $p(\mathbf{x}_t, \mathbf{z}_t|\mathbf{z}_{1:t-1})$ between a state and the subsequent measurement (filtering and smoothing). This result has two implications:

1) Gaussian filters/smoothers can be distinguished by their approximations to these joint distributions.
2) If there exists an algorithm to compute or to estimate the means and the covariances of the joint distributions $p(\mathbf{x}, h(\mathbf{x}))$, where $h \in \{f, g\}$, the algorithm can be used for filtering and RTS smoothing.

In the following, we first consider common filtering and smoothing algorithms and describe how they compute Gaussian approximations to the joint distributions $p(\mathbf{x}_{t-1}, \mathbf{x}_t|\mathbf{z}_{1:t-1})$ and $p(\mathbf{x}_t, \mathbf{z}_t|\mathbf{z}_{1:t-1})$, respectively, which emphasizes the first implication (Sec. V-A). After that, for the second implication of our results, we take an algorithm for estimating means and covariances of joint distributions and turn this algorithm into a filter/smoother (Sec. V-B).

A. Current Algorithms for Computing the Joint Distributions

Fig. 1(a) gives an overview of how the Kalman filter, the EKF, the UKF, and the CKF represent the means and the (cross-)covariances of the joint distributions $p(\mathbf{x}_t, \mathbf{z}_t|\mathbf{z}_{1:t-1})$ and $p(\mathbf{x}_{t-1}, \mathbf{x}_t|\mathbf{z}_{1:t-1})$. In Fig. 1(a), we use the shorthand notation $\mathbf{a}^2 \coloneqq \mathbf{a}\mathbf{a}^\top$. For example, we defined $(f(\mathbf{X}^{(i)}_{t-1|t-1}) - \boldsymbol{\mu}^x_{t|t-1})^2 \coloneqq (f(\mathbf{X}^{(i)}_{t-1|t-1}) - \boldsymbol{\mu}^x_{t|t-1})(f(\mathbf{X}^{(i)}_{t-1|t-1}) - \boldsymbol{\mu}^x_{t|t-1})^\top$.

In the Kalman filter, the transition function f and the measurement function are linear and represented by the matrices $\mathbf{F}$ and $\mathbf{G}$, respectively. The EKF linearizes f and g resulting in the matrices $\tilde{\mathbf{F}}$ and $\tilde{\mathbf{G}}$, respectively. The UKF computes $2D+1$ sigma points $\mathbf{X}$ and uses their mappings through f and g to compute the desired moments, where w_m and w_c are the weights used for computing the mean and the covariance, respectively (see [18], pp. 65). The CKF computations are nearly equivalent to the UKF's computations with slight modifications: First, the CKF only requires $2D$ cubature points $\mathbf{X}$. Thus, the sums run from 1 to $2D$. Second, the weights $w_c = 1/D = w_m$ are all equal [2].

Although none of these algorithms explicitly computes the joint distributions $p(\mathbf{x}_t, \mathbf{z}_t|\mathbf{z}_{1:t-1})$ and $p(\mathbf{x}_{t-1}, \mathbf{x}_t|\mathbf{z}_{1:t-1})$, they all implicitly do so. Using the means and covariances in Fig. 1(a) in the filtering and smoothing Eqs. (4), (5), (18), and (19), the results from the original papers [12], [16], [15], [11], [17], [2] are recovered. To the best of our knowledge, Fig. 1(a) is the first presentation of the CKS.

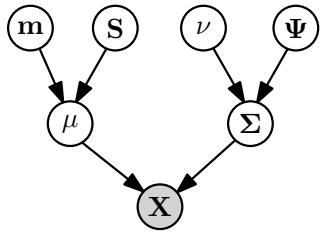
B. Gibbs-Filter and Gibbs-RTS Smoother

We now derive a Gaussian filter and RTS smoother based on Gibbs sampling [9]. In the context of filtering and RTS smoothing, we use Gibbs sampling for inferring the mean and the covariance of the distributions $p(\mathbf{x}_{t-1}, \mathbf{x}_t|\mathbf{z}_{1:t-1})$ and $p(\mathbf{x}_t, \mathbf{z}_t|\mathbf{z}_{1:t-1})$, respectively, which is sufficient for Gaussian filtering and RTS smoothing, see Sec. V.

At each time step, we use Gibbs sampling to infer the moments of the joint distributions $p(\mathbf{x}_{t-1}, \mathbf{x}_t|\mathbf{z}_{1:t-1})$ and $p(\mathbf{x}_t, \mathbf{z}_t|\mathbf{z}_{1:t-1})$. Fig. 1(b) shows the graphical model for inferring the mean $\boldsymbol{\mu}$ and the covariance $\boldsymbol{\Sigma}$ from the joint data set $\mathbf{X}$ using Gibbs sampling. The parameters of the conjugate hyper-priors on the mean $\boldsymbol{\mu}$ and the covariance $\boldsymbol{\Sigma}$ are denoted by $\mathbf{m}, \mathbf{S}$ and $\boldsymbol{\Psi}, \nu$, respectively.

	Kalman filter/smoother	EKF/EKS	UKF/URTSS and CKF/CKS*
$\hat{\mu}^x_{t\|t-1}$	$\mathbf{F}\mu^x_{t-1\|t-1}$	$\tilde{\mathbf{F}}\hat{\mu}^x_{t-1\|t-1}$	$\sum_{i=0}^{2D} w_m^{(i)} f(\mathbf{X}^{(i)}_{t-1\|t-1})$
$\hat{\mu}^z_{t\|t-1}$	$\mathbf{G}\mu^x_{t\|t-1}$	$\tilde{\mathbf{G}}\hat{\mu}^x_{t\|t-1}$	$\sum_{i=0}^{2D} w_m^{(i)} g(\mathbf{X}^{(i)}_{t\|t-1})$
$\hat{\mathbf{\Sigma}}^x_{t\|t-1}$	$\mathbf{F}\mathbf{\Sigma}^x_{t-1\|t-1}\mathbf{F}^\top + \mathbf{Q}$	$\tilde{\mathbf{F}}\hat{\mathbf{\Sigma}}^x_{t-1\|t-1}\tilde{\mathbf{F}}^\top + \mathbf{Q}$	$\sum_{i=0}^{2D} w_c^{(i)} (f(\mathbf{X}^{(i)}_{t-1\|t-1}) - \mu^x_{t\|t-1})^2 + \mathbf{Q}$
$\hat{\mathbf{\Sigma}}^z_{t\|t-1}$	$\mathbf{G}\mathbf{\Sigma}^x_{t\|t-1}\mathbf{G}^\top + \mathbf{R}$	$\tilde{\mathbf{G}}\hat{\mathbf{\Sigma}}^x_{t\|t-1}\tilde{\mathbf{G}}^\top + \mathbf{R}$	$\sum_{i=0}^{2D} w_c^{(i)} (g(\mathbf{X}^{(i)}_{t\|t-1}) - \mu^z_{t\|t-1})^2 + \mathbf{R}$
$\hat{\mathbf{\Sigma}}^{xz}_{t\|t-1}$	$\mathbf{\Sigma}^x_{t\|t-1}\mathbf{G}^\top$	$\hat{\mathbf{\Sigma}}^x_{t\|t-1}\tilde{\mathbf{G}}^\top$	$\sum_{i=0}^{2D} w_c^{(i)} (\mathbf{X}^{(i)}_{t\|t-1} - \mu^x_{t\|t-1})(g(\mathbf{X}^{(i)}_{t\|t-1}) - \mu^z_{t\|t-1})^\top$
$\hat{\mathbf{\Sigma}}^x_{t-1,t\|t}$	$\mathbf{\Sigma}^x_{t-1\|t-1}\mathbf{F}^\top$	$\hat{\mathbf{\Sigma}}^x_{t-1\|t-1}\tilde{\mathbf{F}}^\top$	$\sum_{i=0}^{2D} w_c^{(i)} (\mathbf{X}^{(i)}_{t-1\|t-1} - \mu^x_{t-1\|t-1})(f(\mathbf{X}^{(i)}_{t\|t-1}) - \mu^x_{t\|t-1})^\top$

(a) Approximating the means and the covariances of $p(\mathbf{x}_t, \mathbf{z}_t|\mathbf{z}_{1:t-1})$ and $p(\mathbf{x}_{t-1}, \mathbf{x}_t|\mathbf{z}_{1:t-1})$.



(b) Gibbs-filter/RTSS graphical model.

Fig. 1. (a) Summary of how standard filters/smoothers compute joint probabilities. (b) Graphical model for the Gibbs-filter/RTSS.

To infer the moments of the joint $p(\mathbf{x}_{t-1}, \mathbf{x}_t|\mathbf{z}_{1:t-1})$, we first generate i.i.d. samples from the filter distribution $p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1})$ and map them through the transition function f. The samples and their mappings serve as samples $\mathbf{X}$ from the joint distribution $p(\mathbf{x}_{t-1}, \mathbf{x}_t|\mathbf{z}_{1:t-1})$. With a conjugate Gaussian prior $\mathcal{N}(\boldsymbol{\mu}\,|\,\mathbf{m}, \mathbf{S})$ on the joint mean, and a conjugate inverse Wishart prior distribution $\mathcal{IW}(\mathbf{\Sigma}|\mathbf{\Psi}, \nu)$ on the joint covariance matrix, we infer the hyper-posterior distributions on $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$. By sampling from these posterior distributions, we obtain unbiased estimates of the desired mean and the covariance of the joint $p(\mathbf{x}_{t-1}, \mathbf{x}_t|\mathbf{z}_{1:t-1})$ as the sample average (after a burn in).

To infer the mean and the covariance of the joint $p(\mathbf{x}_t, \mathbf{z}_t|\mathbf{z}_{1:t-1})$, we proceed similarly: We generate i.i.d. samples from the distribution $p(\mathbf{x}_t|\mathbf{z}_{1:t-1})$, which are subsequently mapped through the measurement function. The combined data set of i.i.d. samples and their mappings define the joint data set $\mathbf{X}$. Again, we choose a conjugate Gaussian prior on the mean vector and a conjugate inverse Wishart prior on the covariance matrix of the joint $p(\mathbf{x}_t, \mathbf{z}_t|\mathbf{z}_{1:t-1})$. Using Gibbs sampling, we sample means and covariances from the hyper-posteriors and obtain unbiased estimates for the mean and the covariance of the joint $p(\mathbf{x}_t, \mathbf{z}_t|\mathbf{z}_{1:t-1})$.

Alg. 1 outlines the steps for computing the joint distribution $p(\mathbf{x}_t, \mathbf{z}_t|\mathbf{z}_{1:t-1})$. Since the chosen hyper-priors for the mean and the covariance are conjugate priors, all updates of the posterior hyper-parameters can be computed analytically [10]. The moments of $p(\mathbf{x}_{t-1}, \mathbf{x}_t|\mathbf{z}_{1:t-1})$, which are required for smoothing, are computed similarly by exchanging the pass-in distributions and the mapping function.

Algorithm 1 Inferring the mean $\boldsymbol{\mu}_{x,z}$ and the covariance $\mathbf{\Sigma}_{x,z}$ of $p(\mathbf{x}_t, \mathbf{z}_t|\mathbf{z}_{1:t-1})$ using Gibbs sampling

pass in marginal distribution $p(\mathbf{x}_t|\mathbf{z}_{1:t-1})$, burn-in period B, number L of Gibbs iterations, size N of data set
init. hyper-priors on joint mean and covariance $\mathcal{N}(\boldsymbol{\mu}_{x,z}\,|\,\mathbf{m}, \mathbf{S})$ and $\mathcal{IW}(\mathbf{\Sigma}_{x,z}|\mathbf{\Psi}, \nu)$
$\mathbf{X} \coloneqq [\mathbf{x}_t^{(i)}, g(\mathbf{x}_t^{(i)}) + \mathbf{v}_t^{(i)}]_{i=1}^N$ ▷ generate joint data set
sample $\boldsymbol{\mu}_1 \sim \mathcal{N}(\mathbf{m}, \mathbf{S})$, sample $\mathbf{\Sigma}_1 \sim \mathcal{IW}(\mathbf{\Psi}, \nu)$
for $j = 1$ **to** L **do** ▷ for L Gibbs iterations do
  update $\mathbf{m}|\mathbf{X}, \boldsymbol{\mu}_j, \mathbf{\Sigma}_j$ ▷ posterior mean of $\boldsymbol{\mu}_j$
  update $\mathbf{S}|\mathbf{X}, \boldsymbol{\mu}_j, \mathbf{\Sigma}_j$ ▷ posterior cov. of $\boldsymbol{\mu}_j$
  sample $\boldsymbol{\mu}_{j+1} \sim \mathcal{N}(\boldsymbol{\mu}_{j+1}\,|\,\mathbf{m}, \mathbf{S})$
  update $\mathbf{\Psi}|\mathbf{X}, \boldsymbol{\mu}_{j+1}, \mathbf{\Sigma}_j$ ▷ posterior scale of $\mathbf{\Sigma}_j$
  update $\nu|\mathbf{X}, \boldsymbol{\mu}_{j+1}, \mathbf{\Sigma}_j$ ▷ posterior DoF of $\mathbf{\Sigma}_j$
  sample $\mathbf{\Sigma}_{j+1} \sim \mathcal{IW}(\mathbf{\Sigma}_{j+1}|\mathbf{\Psi}, \nu)$
$\boldsymbol{\mu}_{x,z} \coloneqq \mathbb{E}[\boldsymbol{\mu}_{B+1:L}]$ ▷ unbiased estimate of joint mean
$\mathbf{\Sigma}_{x,z} \coloneqq \mathbb{E}[\mathbf{\Sigma}_{B+1:L}]$ ▷ unbiased estimate joint cov.
return $\boldsymbol{\mu}_{x,z}, \mathbf{\Sigma}_{x,z}$ ▷ return inferred statistics

VI. Numerical Evaluation

As a proof of concept, we show that the Gibbs-RTSS proposed in Sec. V-B performs well in linear and nonlinear systems. As performance measures, we consider the expected root mean square error (RMSE) and the expected negative log-likelihood (NLL) per data point in the trajectory. While the RMSE solely penalizes the distance of the true state and the mean of the filtering/smoothing distribution, the NLL measures the coherence of the filtering/smoothing distributions.In our experiments, we chose a time horizon $T = 50$.

A. Proof of Concept: Linear System

First, we tested the performance of the Gibbs-filter/RTSS in the linear system $x_t = x_{t-1} + w_t\,, z_t = -2\,x_t + v_t\,$, where $w_t \sim \mathcal{N}(0, 1), v_t \sim \mathcal{N}(0, 10), p(x_0) = \mathcal{N}(0, 5)$. In a linear system, the (E)KF is optimal and unbiased [1]. The Gibbs-filter/RTSS perform as well as the EKF/EKS as shown in Fig. 2, which shows the expected performances (with the corresponding standard errors) of the filters/smoothers over 100 independent runs, where $x_0 \sim p(x_0)$. The Gibbs-sampler parameters were set to $(N, L, B) = (1000, 200, 100)$, Alg. 1.

	EKF	Gibbs-filter*	EKS	Gibbs-RTSS*
RMSE	1.11 ± 0.014	1.12 ± 0.014	0.88 ± 0.011	0.89 ± 0.011
NLL	1.52 ± 0.012	1.52 ± 0.012	1.30 ± 0.013	1.30 ± 0.012

Fig. 2. Expected performances (*linear* system) with standard error.

B. Nonlinear System: Non-stationary Growth Model

As a nonlinear example, we consider the dynamic system

$$x_t = \tfrac{x_{t-1}}{2} + \tfrac{25 x_{t-1}}{1 + x_{t-1}^2} + 8\cos(1.2\,(t-1)) + w_t\,, \tag{33}$$

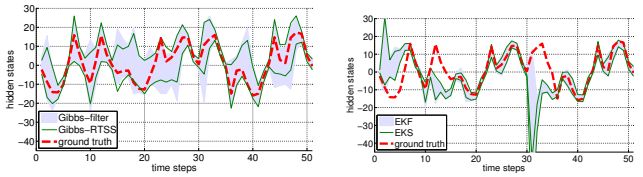
$$z_t = \tfrac{x_t^2}{20} + v_t\,, \tag{34}$$

with exactly the same setup as in [8]: $w_t \sim \mathcal{N}(0, 1)$, $v_t \sim \mathcal{N}(0, 10)$, and $p(x_0) = \mathcal{N}(x_0|0, 5)$. This system is challenging for Gaussian filters due to its quadratic measurement equation and its highly nonlinear system equation.

We run the Gibbs-RTSS, the EKS, the CKS, and the URTSS [17] for comparison. We chose the Gibbs parameters $(N, L, B) = (1000, 200, 100)$. For 100 independent runs starting from $x_0 \sim p(x_0)$, we report the expected RMSE and NLL performance measures in Fig. 3.

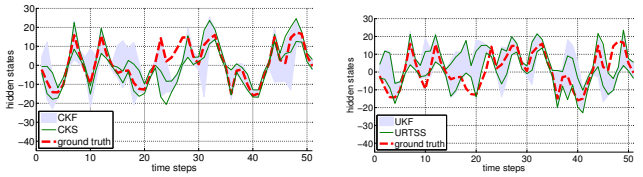
filters	Gibbs-filter*	EKF	CKF	UKF
RMSE	$\mathbf{5.04 \pm 0.088}$	11.1 ± 0.29	6.18 ± 0.17	8.57 ± 0.16
NLL	$\mathbf{2.87 \pm 0.12}$	26.1 ± 1.18	9.96 ± 0.75	13.6 ± 0.68
smoothers	Gibbs-RTSS*	EKS	CKS*	URTSS
RMSE	$\mathbf{4.01 \pm 0.085}$	10.6 ± 0.28	5.66 ± 0.20	8.02 ± 0.16
NLL	$\mathbf{2.78 \pm 0.15}$	$\mathbf{90.6 \pm 10.3}$	$\mathbf{28.9 \pm 3.31}$	$\mathbf{16.3 \pm 0.16}$

Fig. 3. Expected performances (*nonlinear* system) with standard error.



(a) Gibbs-filter (Gibbs-RTSS). RMSE: 5.56 (4.18), **NLL: 2.65 (2.45)**.

(b) EKF (EKS). RMSE: 11.3 (14.7), **NLL: 16.5 (20.8)**.



(c) CKF (CKS). RMSE: 5.66 (5.96), **NLL: 7.32 (20.7)**.

(d) UKF (URTSS). RMSE: 7.87 (7.18), **NLL: 8.66 (9.93)**.

Fig. 4. Example trajectories of filtering/smoothing in the nonlinear growth model using (a) Gibbs-RTSS, (b) EKS, (c) CKS, (d) URTSS. The filter distributions are represented by the shaded areas (95% confidence area), the smoothing distributions are shown by solid green lines (95% confidence area). The actual realization of the latent state is the dashed red graph.

Both high expected NLL-values and the fact that smoothing makes them even higher hint at the incoherencies of the EKF/EKS, the CKF/CKS, and the UKF/URTSS. The Gibbs-RTSS was the only considered smoother that consistently improved the results of the filtering step. Therefore, we conclude that the Gibbs-filter/RTSS is coherent.

Fig. 4 shows example realizations of filtering and smoothing using the Gibbs-filter/RTSS, the EKF/EKS, the CKF/CKS, and the UKF/URTSS, respectively. The Gibbs-filter/RTSS appropriately inferred the variances of the latent state while the other filters/smoothers did not (neither of them is moment-preserving), which can lead to incoherent filtering/smoothing distributions [5], see also Fig. 3.

VII. Discussion

Our Gibbs-filter/RTSS differs from [4], where Gibbs sampling is used to infer the noise in a linear system. Instead, we infer the means and covariances of the full joint distributions $p(\mathbf{x}_{t-1}, \mathbf{x}_t | \mathbf{z}_{1:t-1})$ and $p(\mathbf{x}_t, \mathbf{z}_t | \mathbf{z}_{1:t-1})$ in nonlinear systems from data. Neither the Gibbs-filter nor the Gibbs-RTSS require to know the noise matrices $\mathbf{R}, \mathbf{Q}$, but they can be inferred as a part of the joint distributions if access to the dynamic system is given. Unlike the Gaussian particle filter [14], the proposed Gibbs-filter is not a particle filter. Therefore, it does not suffer from degeneracy due to importance sampling.

Although the Gibbs-filter is computationally more involved than the EKF/UKF/CKF, it can be used as a baseline method to evaluate the accuracy and coherence of more efficient algorithms: When using sufficiently many samples the Gibbs-filter can be considered a close approximation to a moment-preserving filter in nonlinear stochastic systems.

The Gibbs-filter/RTSS only need to be able to evaluation the system and measurement functions. No further requirements such as differentiability are needed.

The Gibbs-RTSS code is publicly available at `mloss.org` and an extended version of this paper can be obtained from [7].

VIII. Conclusion

Using a general probabilistic perspective on Gaussian filtering and smoothing, we first showed that it is sufficient to determine Gaussian approximations to two joint probability distributions to perform Gaussian filtering and smoothing. Computational approaches to Gaussian filtering and Rauch-Tung-Striebel smoothing can be distinguished by their respective methods used to determining two joint distributions.

Second, our results allow for a straightforward derivation and implementation of novel Gaussian filtering and smoothing algorithms, e.g., the cubature Kalman smoother. Additionally, we presented a filtering smoothing algorithm based on Gibbs sampling as an example. Our experimental results show that the proposed Gibbs-filter/Gibbs-RTSS compares well with state-of-the-art Gaussian filters and RTS smoothers in terms of robustness and accuracy.

References

[1] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Dover Publications, 2005.

[2] I. Arasaratnam and S. Haykin, "Cubature Kalman Filters," *IEEE Trans. on Automatic Control*, vol. 54, pp. 1254–1269, 2009.

[3] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer-Verlag, 2006.

[4] C. K. Carter and R. Kohn, "On Gibbs Sampling for State Space Models," *Biometrika*, vol. 81, pp. 541–553, 1994.

[5] M. P. Deisenroth, *Efficient Reinforcement Learning using Gaussian Processes*. KIT Scientific Publishing, 2010, iSBN 978-3-86644-569-7.

[6] M. P. Deisenroth, M. F. Huber, and U. D. Hanebeck, "Analytic Moment-based Gaussian Process Filtering," in *Proc. of the 26th International Conference on Machine Learning*, pp. 225–232.

[7] M. P. Deisenroth and H. Ohlsson, *A Probabilistic Perspective on Gaussian Filtering and Smoothing*. http://arxiv.org/abs/1006.2165

[8] A. Doucet, S. J. Godsill, and C. Andrieu, "On Sequential Monte Carlo Sampling Methods for Bayesian Filtering," *Statistics and Computing*, vol. 10, pp. 197–208, 2000.

[9] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721–741, 1984.

[10] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, Eds., *Markov Chain Monte Carlo in Practice*. Chapman & Hall, 1996.

[11] S. J. Julier and J. K. Uhlmann, "Unscented Filtering and Nonlinear Estimation," *Proc. of the IEEE*, vol. 92, pp. 401–422, 2004.

[12] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Trans. of the ASME—Journal of Basic Engineering*, vol. 82, pp. 35–45, 1960.

[13] J. Ko and D. Fox, "GP-BayesFilters: Bayesian Filtering using Gaussian Process Prediction and Observation Models," *Autonomous Robots*, vol. 27, pp. 75–90, 2009.

[14] J. H. Kotecha and P. M. Djuric, "Gaussian Particle Filtering," *IEEE Trans. on Signal Processing*, vol. 51, pp. 2592–2601, 2003.

[15] P. S. Maybeck, *Stochastic Models, Estimation, and Control*. Academic Press, Inc., 1979, vol. 141.

[16] H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum Likelihood Estimates of Linear Dynamical Systems," *AIAA Journal*, pp. 1445–1450, 1965.

[17] S. Särkkä, "Unscented Rauch-Tung-Striebel Smoother," *IEEE Trans. on Automatic Control*, vol. 53, pp. 845–849, 2008.

[18] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. The MIT Press, 2005.