Kalman Filter Variants in the Closed Skew Normal Setting

Javad Rezaie and Jo Eidsvik
Department of Mathematical Sciences, NTNU, Norway,
Email:rezaie@math.ntnu.no, joeid@math.ntnu.no

Abstract

The filtering problem, or dynamic data assimilation problem, is studied for linear and nonlinear systems with continuous state space and over discrete time steps. The paper presents filtering approaches based on the conjugate closed skew normal probability density. This distribution allows additional flexibility over the usual Gaussian approximations. With linear dynamic systems, the filtering distribution can now be computed in analytical form. With nonlinear dynamic systems, an ensemble-based version is proposed, fitting a closed skew normal distributions at each updating step. Numerical examples discuss various special cases of the methods.

Keywords:
Recursive Bayesian Estimation; Closed Skew Normal Distribution;
Ensemble Kalman Filter; Nonlinear System; Petroleum Reservoir.

1. Introduction

In this paper we consider the filtering problem under non-Gaussian and non-linear modeling assumptions. The challenge is to characterize the probability distribution of state variables over time, given the available information at that time instant. The underlying model comes from a physical system and is represented as a set of differential or difference equations, known as the process model. At each (discrete) time step sensors measure the state variable directly or indirectly, and we are interested in assimilating these data with the knowledge imposed by the process model and all previous measurements.

The history of the filtering problem goes back more than 50 years, when Kalman proposed his famous filtering solution for linear dynamical systems in optimal control literature (Kalman, 1960). The Kalman filter (KF) has shown extremely useful, but has strict assumptions about linearity and Gaussian noise. The Extended Kalman filter (EKF) is an extension handling nonlinearities (Jazwinsky, 1970), but if the nonlinearity of the system is high, this first order approximation diverges. Second order KF variants have also been proposed, but they may have similar challenges. Moreover, we need to calculate the Jacobian and Hessian of the system equations which may not be feasible. For instance, these derivative expressions are rarely available from implicitly formulated process models or black box models.

Particles filters (PF) were proposed to handle general distributions by Monte Carlo sampling. These approaches can approximate any distribution when the number of particles goes to infinity (Doucet et al., 2001). In practical systems there are limits in the available computation time and consequently the number of particles/samples which can be used. Albeit very popular in many applications, particle filters may suffer from sample degeneracy when the system dimension increases.

Ensemble Kalman filter (EnKF) was introduced as a sampling representation for very high dimensional systems, see e.g. Evensen (2001), Sakov and Oke (2008) and Evensen (2009). It incorporates the nonlinear process model, whereas a Gaussian approximation is used for the updating with respect to new measurements. This approach has been very useful for practical applications, but the filter solution may be biased or underestimate uncertainty. Gaussian mixture filters have been suggested to get benefits from both PF and EnKF, see e.g. Stordal et al. (2010) and Rezaie and Eidsvik (2012).

In this paper we introduce a new filter which captures skewness in the filtering solution. It is easy to implement and some of the mentioned filters are special cases of the suggested approach. The filter is based on the closed skew normal (CSN) distribution, which allows analytical solutions under certain modeling assumptions. This family of filters is named the closed skew normal Kalman filter (CSNKF).

The skew normal (SN) and CSN distribution are extensions of the normal or Gaussian distributions, see e.g. Azzalini and Dalla-Valle (1996) and Gupta et al. (2004). Skewness is incorporated by adding new parameters to the traditional Gaussian formulation (Genton, 2004). The CSN distribution has some useful properties similar to those of the Gaussian distribution, such as closedness under linear conditioning. Thus, we can extend the previous

Gaussian-based filters by introducing the CSN distribution into the filtering problem.

A skewed version of the KF for linear systems was proposed by Naveau et al. (2005). They defined the filter in an extended state space model. Our proposed algorithms work for linear and nonlinear systems in a unified setting, with structure similar to the KF and EnKF. Computational aspects are studied to handle the ensemble-based fitting, and challenges related to the skewness dimension over many filtering times are discussed for all the KF variants.

In Section 2 we outline the modeling assumptions and present the CSN distribution. In Section 3 we present the CSN filter under linear modeling assumptions. In Section 4 we similarly present the CSN filter under nonlinear modeling assumptions. In Section 5 we illustrate the methodologies using numeric examples.

2. Background

2.1. Notation

Throughout this paper, we use $\boldsymbol{x}_t \in \Re^{n\times 1}$ as a n_x dimensional distinction of interest at time $t = 1, \ldots, T$. We assume that the dynamics of these state variables are represented by a set of difference equations with additive noise $\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_{t-1}) + \boldsymbol{\eta}_t$, where $\boldsymbol{f}(\cdot) : \Re^{n_x\times 1} \longmapsto \Re^{n_x\times 1}$ is a general linear or nonlinear function and $\boldsymbol{\eta}_t \in \Re^{n_x\times 1}$ is independent additive noise with known distribution. If the system dynamics are linear, we use the notation $\boldsymbol{x}_t = \boldsymbol{F}\boldsymbol{x}_{t-1} + \boldsymbol{\eta}_t$, where $\boldsymbol{F} \in \Re^{n_x\times n_x}$.

The notation $\boldsymbol{x} \sim \pi(\cdot)$ is used to show that variable $\boldsymbol{x}$ is distributed according to the probability density functions (pdf) $\pi(\boldsymbol{x})$. The Markov assumption about the state means that $\pi(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{x}_{t-2}, \cdots \boldsymbol{x}_0) = \pi(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$, and we have joint pdf $\pi(\boldsymbol{x}_T, \boldsymbol{x}_{T-1}, \cdots \boldsymbol{x}_0) = \pi(\boldsymbol{x}_T|\boldsymbol{x}_{T-1}) \cdots \pi(\boldsymbol{x}_1|\boldsymbol{x}_0)\pi(\boldsymbol{x}_0)$, where $\pi(\boldsymbol{x}_0)$ is the initial pdf.

The observation equation is $\boldsymbol{d}_t = \boldsymbol{h}(\boldsymbol{x}_t) + \boldsymbol{\epsilon}_t$, where $\boldsymbol{d}_t \in \Re^{n_d\times 1}$ and $\boldsymbol{\epsilon}_t$ is the independent additive noise with known distribution. Thus, we assume that the data at different times are mutually independent given the state. The likelihood for the data is $\pi(\boldsymbol{d}_t|\boldsymbol{x}_t)$. The notation $\boldsymbol{D}_t = [\boldsymbol{d}_1, \boldsymbol{d}_2, \cdots \boldsymbol{d}_t]$ is used for the collection of data from time 1 to t. Here, we assume a linear or weakly nonlinear relationship between the observation and the state variable. Thus, we linearize the measurement equation using a first order Taylor series expansion to get $\boldsymbol{h}(\boldsymbol{x}_t) \approx \boldsymbol{h}_0 + \boldsymbol{H}\boldsymbol{x}_t$, where $\boldsymbol{H} \in \Re^{n_d\times n_x}$.

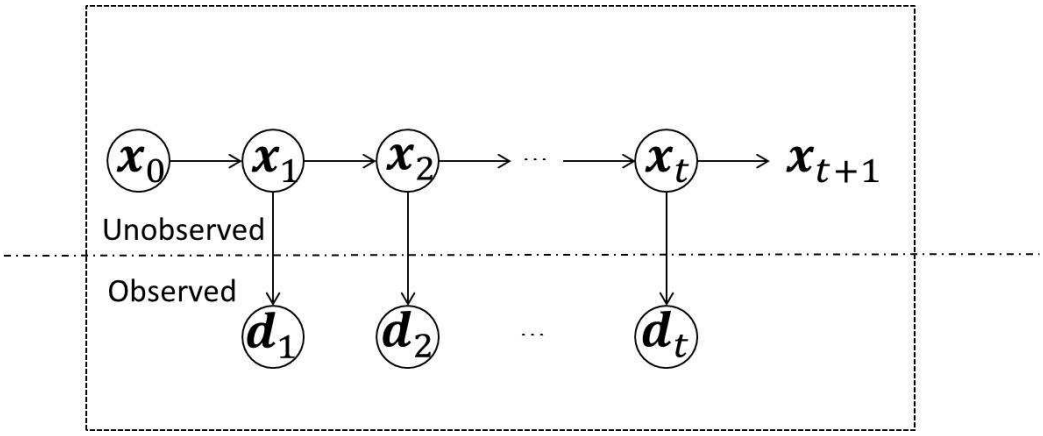

Figure 1: Graphical representation of state variables $\boldsymbol{x}_t$, at discrete time points $t = 0, 1, \ldots$ and observations $\boldsymbol{d}_t$, $t = 1, 2, \ldots$. The process model is assumed to follow a Markov structure. The data are assumed to be conditionally independent, given the state variable at the indicated time steps. The filtering problem characterizes the distribution of states over time, given all currently available data.

Figure 1 illustrates the modeling assumptions graphically. In the following we will specify the initial distribution of $\boldsymbol{x}_0$ in Figure 1 and the distributions for the additive noise terms $\boldsymbol{\eta}_t$ and $\boldsymbol{\epsilon}_t$. In the examples we assign specific relations $\boldsymbol{f}$ or $\boldsymbol{F}$, and $\boldsymbol{H}$. These operators could also depend on time, but we ignore this to simplify the mathematical notation.

2.2. The filtering problem

Our goal is to assess the distribution of the state vector given all the currently available observations, i.e. the pdf $\pi\left(\boldsymbol{x}_t|\boldsymbol{D}_t\right)$ at each time index t. A recursive formulation gives the exact solution to this filtering problem by the following two steps. First, derive the one-step predictive pdf:

$$\pi(\boldsymbol{x}_t|\boldsymbol{D}_{t-1}) = \int \pi\left(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}\right) \pi\left(\boldsymbol{x}_{t-1}|\boldsymbol{D}_{t-1}\right) d\boldsymbol{x}_{t-1}, \tag{1}$$

where we use the Markovian state model to propagate variables one time index forward. Next, Bayes rule is used to assimilate the current data:

$$\pi\left(\boldsymbol{x}_t|\boldsymbol{D}_t\right) = \frac{\pi\left(\boldsymbol{d}_t|\boldsymbol{x}_t\right) \pi\left(\boldsymbol{x}_t|\boldsymbol{D}_{t-1}\right)}{\pi\left(\boldsymbol{d}_t|\boldsymbol{D}_{t-1}\right)}. \tag{2}$$

Here, the one-step predictive pdf plays the role of a prior (like in prior to data $\boldsymbol{d}_t$), and the conditionally independent likelihood of current data is $\pi(\boldsymbol{d}_t|\boldsymbol{x}_t)$. Expression (1) and (2) are evaluated from time $t = 1$ to T, with initial distribution $\pi(\boldsymbol{x}_0|\boldsymbol{D}_0) = \pi(\boldsymbol{x}_0)$.

Unfortunately, the solution provided by (1) and (2) is not applicable to most practical systems because it contains complicated multi-dimensional integrals. However, these integrals are possible by putting some limitations on the model. Under linear and Gaussian assumptions, the KF is the exact solution to the filtering problem. Another possibility is to approximate the complicated integrals, via Monte Carlo sampling or by Gaussian approximations. The EnKF is based on fitting a Gaussian approximation to the one-step predictive pdf. In the current paper we present methods that extend these popular algorithms. Our contribution allows skewed distributions of the type we will discuss next.

2.3. Closed skew normal distribution

The CSN distribution is an extension of the Gaussian pdf, see Genton (2004) and Flecher et al. (2009). It has many properties similar to the Gaussian (Domínguez-Molina et al., 2003). In particular, the CSN distribution is conjugate, i.e. when the prior pdf and likelihood are both CSN, and there is a linear relation in the likelihood, the posterior pdf is also CSN (Karimi et al., 2010).

Denote the n dimensional Gaussian pdf by $\pi(\boldsymbol{x}) = \phi_n(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ and the associated cumulative distribution function (cdf) by $\Phi_n(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$. A random vector $\boldsymbol{x} = (x_1, \ldots, x_n)^t$ is CSN distributed if its pdf is as follows:

$$\begin{aligned} \boldsymbol{x} &\sim CSN_{n,q}\left(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Gamma}, \boldsymbol{v}, \boldsymbol{\Delta}\right) \\ &= \left[\Phi_q\left(0; \boldsymbol{v}, \boldsymbol{\Delta} + \boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}^T\right)\right]^{-1} \Phi_q\left(\boldsymbol{\Gamma}\left(\boldsymbol{x} - \boldsymbol{\mu}\right); \boldsymbol{v}, \boldsymbol{\Delta}\right) \phi_n\left(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}\right) \end{aligned} \quad (3)$$

where $\boldsymbol{\mu} \in \Re^{n\times 1}$, $\boldsymbol{\Sigma} \in \Re^{n\times n}$, $\boldsymbol{\Gamma} \in \Re^{q\times n}$, $\boldsymbol{v} \in \Re^q$ and $\boldsymbol{\Delta} \in \Re^{q\times q}$. Here, $\boldsymbol{\Sigma}$ and $\boldsymbol{\Delta}$ are positive definite matrices. The integer q defines the skewness dimension of the CSN pdf. The $\boldsymbol{\mu}$ can be interpreted as a centralization parameter, and $\boldsymbol{\Sigma}$ is a scale matrix, but note that the mean is not $\boldsymbol{\mu}$ and the covariance matrix is not $\boldsymbol{\Sigma}$. If we set $\boldsymbol{\Gamma} = \boldsymbol{0}$, the CSN pdf reduces to the Gaussian with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. Figure 2 illustrates a bivariate CSN distribution.

A critical point in the evaluation of (3) is the two cdf terms. For large skewness dimension q, the Gaussian cdf is very hard to compute. This makes

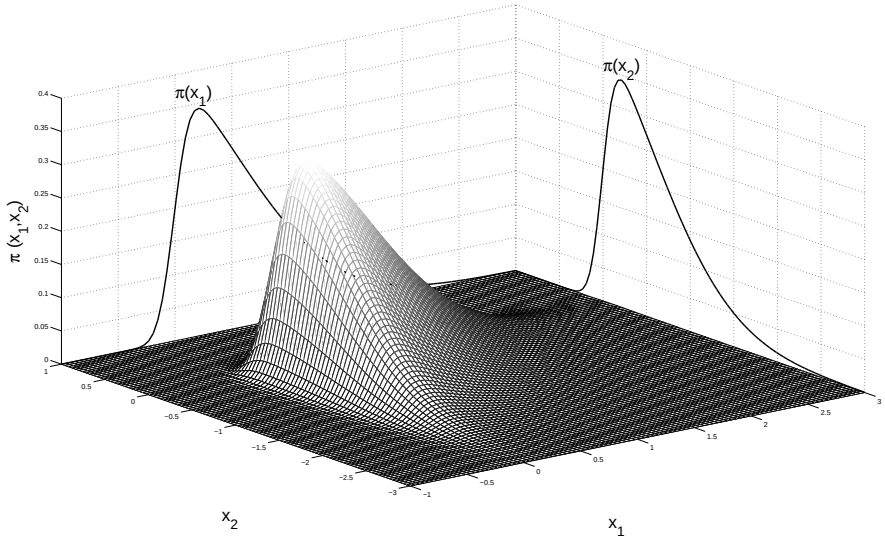

Figure 2: Bivariate CSN and its marginal pdfs.

parameter estimation, prediction and sampling challenging for the CSN distribution, and the issue will also be discussed for the filtering computations below.

We will use several important properties of the CSN distribution. The conjugacy was mentioned earlier. More over, if $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are independent CSN random variables, then $\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2$ is also CSN. If $\boldsymbol{x}$ is CSN, then a linear transformation of $\boldsymbol{x}$, $\boldsymbol{H}\boldsymbol{x}$, is also CSN. Further, we will use a conditional definition of the CSN distribution as follows: Assume that $\boldsymbol{x}$ and $\boldsymbol{t}$ are joint Gaussian distributed, i.e.

$$\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{t} \end{pmatrix} \sim \phi_{n+q} \left(\begin{bmatrix} \boldsymbol{\mu} \\ -\boldsymbol{v} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma}\boldsymbol{\Gamma}^T \\ \boldsymbol{\Gamma}\boldsymbol{\Sigma} & \boldsymbol{\Delta} + \boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}^T \end{bmatrix} \right). \tag{4}$$

The block diagonals of the covariance matrix are the marginal covariances of $\boldsymbol{x}$ and $\boldsymbol{t}$, and the off-diagonals are their cross-covariance. The conditional distribution $\boldsymbol{x}|\boldsymbol{t} > \boldsymbol{0}$ is the CSN pdf with parameterization defined above. See Appendix for further description of the CSN properties.

3. CSN filter for linear dynamical systems

We will here consider the linear process model from above; $\boldsymbol{x}_t = \boldsymbol{F}\boldsymbol{x}_{t-1} + \boldsymbol{\eta}_t$. We first present the main CSN filtering formulation, and then some special cases.

3.1. Prediction and update formula

The recursive solution to the filtering problem described in (1) and (2) is now possible to solve analytically under CSN assumptions. The prediction and update steps are as follows:

1) Prediction step:

Assume that $\pi\left(\boldsymbol{x}_{t-1}|\boldsymbol{D}_{t-1}\right) \sim CSN_{n_x,q_x}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Gamma}, \boldsymbol{v}, \boldsymbol{\Delta}\right)$ and the process noise $\pi\left(\boldsymbol{\eta}_t\right) \sim CSN_{n_x,q_{\boldsymbol{\eta}}}\left(\boldsymbol{\mu}_\eta, \boldsymbol{\Sigma}_\eta, \boldsymbol{\Gamma}_\eta, \boldsymbol{v}_\eta, \boldsymbol{\Delta}_\eta\right)$ are mutually independent.

The linear process model, in conjunction with the results in the Appendix, means the one-step predictive pdf is $\pi\left(\boldsymbol{x}_t|\boldsymbol{D}_{t-1}\right) \sim CSN_{n_x,q_x+q_{\boldsymbol{\eta}}}\left(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x, \boldsymbol{\Gamma}_x, \boldsymbol{v}_x, \boldsymbol{\Delta}_x\right)$, where:

$$
\begin{aligned}
\boldsymbol{\mu}_x &= \boldsymbol{F}\boldsymbol{\mu} + \boldsymbol{\mu}_\eta, \boldsymbol{\Sigma}_x = \boldsymbol{F}\boldsymbol{\Sigma}\boldsymbol{F}^T + \boldsymbol{\Sigma}_\eta \\
\boldsymbol{\Gamma}_x &= \begin{pmatrix} \boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{F}^T\boldsymbol{\Sigma}_x^{-1} \\ \boldsymbol{\Gamma}_\eta\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}_x^{-1} \end{pmatrix}, \boldsymbol{v}_x = \begin{pmatrix} \boldsymbol{v} \\ \boldsymbol{v}_\eta \end{pmatrix} \\
\boldsymbol{\Delta}_x &= \begin{pmatrix} \boldsymbol{\Delta} + \boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}^T - \boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{F}^T\boldsymbol{\Sigma}_x^{-1}\boldsymbol{F}\boldsymbol{\Sigma}\boldsymbol{\Gamma}^T & -\boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{F}^T\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Gamma}_\eta^T \\ -\boldsymbol{\Gamma}_\eta\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}_x^{-1}\boldsymbol{F}\boldsymbol{\Sigma}\boldsymbol{\Gamma}^T & \boldsymbol{\Delta}_\eta + \boldsymbol{\Gamma}_\eta\boldsymbol{\Sigma}_\eta\boldsymbol{\Gamma}_\eta^T - \boldsymbol{\Gamma}_\eta\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_\eta\boldsymbol{\Gamma}_\eta^T \end{pmatrix}
\end{aligned}
\tag{5}
$$

2) Update step:

Just like for the KF, the CSN pdf entails a conjugate update step of the filter. For the updating we integrate the data via the likelihood $\pi\left(\boldsymbol{d}_t|\boldsymbol{x}_t\right) \sim CSN_{n_d,q_\epsilon}(\boldsymbol{H}\boldsymbol{x}_t+\boldsymbol{\mu}_\epsilon, \boldsymbol{\Sigma}_\epsilon, \boldsymbol{\Gamma}_\epsilon, \boldsymbol{v}_\epsilon, \boldsymbol{\Delta}_\epsilon)$. The updated pdf is $\pi\left(\boldsymbol{x}_t|\boldsymbol{D}_t\right) \sim CSN_{n_x,q_x+q_{\boldsymbol{\eta}}+q_\epsilon}(\boldsymbol{\mu}_{x|d}, \boldsymbol{\Sigma}_{x|d}, \boldsymbol{\Gamma}$

Here, the parameters are as follows:

$$
\begin{aligned}
\boldsymbol{\mu}_{x|d} &= \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_x \boldsymbol{H}^T \left[\boldsymbol{H}\boldsymbol{\Sigma}_x \boldsymbol{H}^T + \boldsymbol{\Sigma}_\epsilon\right]^{-1} (\boldsymbol{d}_t - \boldsymbol{H}\boldsymbol{\mu}_x - \boldsymbol{\mu}_\epsilon) \\
\boldsymbol{\Sigma}_{x|d} &= \boldsymbol{\Sigma}_x - \boldsymbol{\Sigma}_x \boldsymbol{H}^T \left[\boldsymbol{H}\boldsymbol{\Sigma}_x \boldsymbol{H}^T + \boldsymbol{\Sigma}_\epsilon\right]^{-1} \boldsymbol{H}\boldsymbol{\Sigma}_x \\
\boldsymbol{\Gamma}_{x|d} &= \left[\begin{bmatrix}\boldsymbol{\Gamma}_x \boldsymbol{\Sigma}_x \\ 0\end{bmatrix} - \begin{bmatrix}\boldsymbol{\Gamma}_x \boldsymbol{\Sigma}_x \boldsymbol{H}^T \\ \boldsymbol{\Gamma}_\epsilon \boldsymbol{\Sigma}_\epsilon\end{bmatrix} \left[\boldsymbol{H}\boldsymbol{\Sigma}_x \boldsymbol{H}^T + \boldsymbol{\Sigma}_\epsilon\right]^{-1} \boldsymbol{H}\boldsymbol{\Sigma}_x\right] \boldsymbol{\Sigma}_{x|d}^{-1} \qquad (6) \\
\boldsymbol{v}_{x|d} &= \begin{bmatrix}-\boldsymbol{v}_x \\ -\boldsymbol{v}_\epsilon\end{bmatrix} + \begin{bmatrix}\boldsymbol{\Gamma}_x \boldsymbol{\Sigma}_x \boldsymbol{H}^T \\ \boldsymbol{\Gamma}_\epsilon \boldsymbol{\Sigma}_\epsilon\end{bmatrix} \left[\boldsymbol{H}\boldsymbol{\Sigma}_x \boldsymbol{H}^T + \boldsymbol{\Sigma}_\epsilon\right]^{-1} (\boldsymbol{d}_t - \boldsymbol{H}\boldsymbol{\mu}_x - \boldsymbol{\mu}_\epsilon) \\
\boldsymbol{\Delta}_{x|d} &= \begin{bmatrix}\boldsymbol{\Delta}_x + \boldsymbol{\Gamma}_x \boldsymbol{\Sigma}_x \boldsymbol{\Gamma}_x^T & 0 \\ 0 & \boldsymbol{\Delta}_\epsilon + \boldsymbol{\Gamma}_\epsilon \boldsymbol{\Sigma}_\epsilon \boldsymbol{\Gamma}_\epsilon^T\end{bmatrix} \\
&- \begin{bmatrix}\boldsymbol{\Gamma}_x \boldsymbol{\Sigma}_x \boldsymbol{H}^T \\ \boldsymbol{\Gamma}_\epsilon \boldsymbol{\Sigma}_\epsilon\end{bmatrix} \left[\boldsymbol{H}\boldsymbol{\Sigma}_x \boldsymbol{H}^T + \boldsymbol{\Sigma}_\epsilon\right]^{-1} \begin{bmatrix}\boldsymbol{\Gamma}_x \boldsymbol{\Sigma}_x \boldsymbol{H}^T \\ \boldsymbol{\Gamma}_\epsilon \boldsymbol{\Sigma}_\epsilon\end{bmatrix}^T - \boldsymbol{\Gamma}_{x|d} \boldsymbol{\Sigma}_{x|d} \boldsymbol{\Gamma}_{x|d}^T
\end{aligned}
$$

For details, see the Appendix.

We can divide the relations in (6) in two parts: i) Gaussian parts which consist of updating $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, and ii) skewed parts which consist of updating $\boldsymbol{\Gamma}$, $\boldsymbol{v}$ and $\boldsymbol{\Delta}$. Similar to the KF, we have linear updates for both $\boldsymbol{\mu}$ and $\boldsymbol{v}$. The other parameters do not depend on the data. By defining $(\boldsymbol{d}_t - \boldsymbol{H}\boldsymbol{\mu}_x - \boldsymbol{\mu}_\epsilon)$ as the innovation or measurement residual, we have two Kalman gains; one for the Gaussian and one for the skewed part:

$$
\begin{aligned}
K_{Gauss} &= \boldsymbol{\Sigma}_x \boldsymbol{H}^T \left[\boldsymbol{H}\boldsymbol{\Sigma}_x \boldsymbol{H}^T + \boldsymbol{\Sigma}_\epsilon\right]^{-1}, \\
K_{Skewed} &= \begin{bmatrix}\boldsymbol{\Gamma}_x \boldsymbol{\Sigma}_x \boldsymbol{H}^T \\ \boldsymbol{\Gamma}_\epsilon \boldsymbol{\Sigma}_\epsilon\end{bmatrix} \left[\boldsymbol{H}\boldsymbol{\Sigma}_x \boldsymbol{H}^T + \boldsymbol{\Sigma}_\epsilon\right]^{-1}. \qquad (7)
\end{aligned}
$$

The skewness dimension of the updated CSN is different from the predictive CSN. In the predictive pdf it is q_x, while it is $q_x + q_\eta + q_\epsilon$ for the updated pdf. This means that the skewness dimension explodes as the recursion proceeds over many time steps. As a result the matrix dimensions grow, parameter estimation gets more complicated, sampling is harder, and so on. Thus, for practical purposes we need to assume simplified conditions.

3.2. Special cases

Note that it is hard to grasp the influence of skewness in the prediction noise and the likelihood on the filtering solution. But some guidelines can be drawn from the update formulas in (6). For instance, if the prior is Gaussian

and the likelihood noise CSN, the posterior will be skewed, and the level of skewness clearly depends on the noise scale ($\boldsymbol{\Sigma}_\epsilon$) and the prior covariance ($\boldsymbol{\Sigma}_x$). Moreover, the anchor point of the skewness variable, $\boldsymbol{v}_{x|d}$, depends on the skewness and scale of the likelihood and prior covariance. However, in our experience it is hard to foresee effects of skewness over time without actually implementing the algorithm and checking the results. Instead we list some special cases of the filter below.

Kalman filter

If the initial pdf is Gaussian and all noise terms are Gaussian, the CSN filter formulation is identical to that of the KF. This occurs when $\boldsymbol{\Gamma} = \mathbf{0}$. Consequently we just have $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in the prediction and update equations, i.e. the Gaussian Kalman gain part.

Gaussian noise with skewed prior

The reason for the explosion in the skewness dimension is the process and observation noise terms, i.e. q_η and q_ϵ. If these dimensions are assumed to be zero, the proposed filter avoids explosion in the skewness. Setting $q_{\boldsymbol{\eta}} = 0$ and $q_\epsilon = 0$ is equivalent to assuming that the additive process and observation noise terms are Gaussian. Only the initial pdf $\pi(\boldsymbol{x}_0)$ is CSN. Over many time steps the initial distribution will lose influence, and the solution will go towards the Gaussian one.

Re-fitting the updated distribution

One idea for avoiding skewness explosion is to reset the parameters at each step in the recursion. This entails that we approximate the updated pdf with a new CSN distribution with structure similar to that of the previous one. This gives

$$CSN_{n_x,q_x+q_\eta+q_\epsilon}(\boldsymbol{\mu}_{x|d}, \boldsymbol{\Sigma}_{x|d}, \boldsymbol{\Gamma}_{x|d}, \boldsymbol{v}_{x|d}, \boldsymbol{\Delta}_{x|d}) \approx CSN_{n_x,q_x}(\boldsymbol{\mu}', \boldsymbol{\Sigma}', \boldsymbol{\Gamma}', \boldsymbol{v}', \boldsymbol{\Delta}'). \quad (8)$$

However, it is unclear how to fit the new CSN, and to understand its properties.

4. CSN filter for nonlinear dynamical systems

Nonlinearities in the propagation of state variables are always problematic because of the chaotic effects that may be induced. It is very hard to find a solution for all kinds of nonlinearity and it is common to find a solution for a family of nonlinearities. In general, if $\boldsymbol{x} \sim \pi(\boldsymbol{x})$ and $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$, it is very hard to find the exact distribution of $\boldsymbol{y} \sim \pi(\boldsymbol{y})$, for general function $\boldsymbol{f}(\cdot)$.

Fortunately, statisticians have a solution for handling nonlinear functions: We can approximate the distribution of $\boldsymbol{y}$ via Monte Carlo sampling. I.e. generate independent realizations from $\pi(\boldsymbol{x})$ and propagate them through the function $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$. The resulting samples represent the empirical distribution of $\boldsymbol{y}$. We may add noise to these samples to represent the predictive distribution in (1). A good approximation requires many samples. Generating realizations from $\pi(\boldsymbol{x})$ can usually be done efficiently, but evaluating $\boldsymbol{f}(\cdot)$ may be time-demanding for many practical applications. For filtering purposes, predicted samples must be updated using (2) when new data get available. For the PF algorithms, the number of samples must usually be very large in this setting to avoid degeneration, see e.g. Doucet et al. (2001). This means that the empirical approximation of $\pi(\boldsymbol{x}_t|\boldsymbol{D}_t)$, obtained by many recursive steps with moderate Monte Carlo sample size, may not be very reliable for practical applications.

Another approach is to consider an special parametric distribution for the samples $\boldsymbol{y}$. One would then attempt to fit the parameters in the selected parametric family using the Monte Carlo samples from the predictive distribution. This is done in the EnKF, which assumes a Gaussian distribution for the predictive distribution. The EnKF in this way estimates the mean and covariance of a Gaussian approximation to $\pi(\boldsymbol{x}_t|\boldsymbol{D}_{t-1})$ by using the empirical mean and covariance of the propagated samples.

We next study the nonlinear filtering problem in a CSN setting. This entails using a CSN approximation for the predictive distribution of the filter. The approximation will be an ensemble-based, in the spirit of the EnKF. The parameter estimation is harder than for the EnKF, and we apply computational ideas based on the conditional form of the CSN distribution.

4.1. Prediction and update formulas

Using Monte Carlo sampling and CSN approximations, the prediction and update steps are as follows:

1) Prediction step:

We assume we have B samples from the updated distribution at time $t-1$. Denote these by $\boldsymbol{x}_{t-1}^i, i = 1, 2, \cdots B$. Let $\boldsymbol{y}_t^i = \boldsymbol{f}\left(\boldsymbol{x}_{t-1}^i\right)$ represent the propagated state variables of each sample.

We next consider a CSN approximation for $\boldsymbol{y}_t^i$, i.e. we assume $\pi(\boldsymbol{y}_t) \sim CSN_{n_y,q_x}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Gamma}, \boldsymbol{v}, \boldsymbol{\Delta})$. The computational aspects related to this fitting step are described in Section 4.2. Since we assume additive CSN process

noise $\pi(\boldsymbol{\eta}_t) \sim CSN_{n_x,q_\eta}\left(\boldsymbol{\mu}_\eta, \boldsymbol{\Sigma}_{\boldsymbol{\eta}}, \boldsymbol{\Gamma}_{\boldsymbol{\eta}}, \boldsymbol{v}_{\boldsymbol{\eta}}, \boldsymbol{\Delta}_{\boldsymbol{\eta}}\right)$, the predictive distribution will be CSN as well, with skewness dimension $q_x + q_\eta$. See (5) and the Appendix.

2) Update step:

By conjugacy, assuming a linear relation for the center parameter in the CSN likelihood, the updated distribution is also CSN. We have $\pi(\boldsymbol{x}_t|\boldsymbol{D}_t) \sim CSN_{n,q_x+q_{\boldsymbol{\eta}}+q_\epsilon}(\boldsymbol{\mu}_{x|d}, \boldsymbol{\Sigma}_{x|d}, \boldsymbol{\Gamma}_{x|d}, \boldsymbol{v}_{x|d}, \boldsymbol{\Delta}_{x|d})$, where the parameters are analytically available by (6).

4.2. Computational methods for fitting the CSN predictive distribution

In this part, we introduce a method for fitting the distribution of a function of CSN random variables. A good starting point for the exposition is that of a linear function, i.e. $\boldsymbol{y} = \boldsymbol{F}\boldsymbol{x}$ and $\boldsymbol{x} \sim CSN_{n_x,q_x}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x, \boldsymbol{\Gamma}_x, \boldsymbol{v}_x, \boldsymbol{\Delta}_x)$. From the Appendix we know that $\boldsymbol{y}$ is also CSN. According to the conditional formulation of the CSN distribution in Section 2.3 we have:

$$\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{t}_1 \end{pmatrix} \sim \phi_{n_x+q_x}\left(\begin{bmatrix} \boldsymbol{\mu}_x \\ -\boldsymbol{v}_x \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_x & \boldsymbol{\Sigma}_x\boldsymbol{\Gamma}_x^T \\ \boldsymbol{\Gamma}_x\boldsymbol{\Sigma}_x & \boldsymbol{\Delta}_x + \boldsymbol{\Gamma}_x\boldsymbol{\Sigma}_x\boldsymbol{\Gamma}_x^T \end{bmatrix}\right). \tag{9}$$

For the variable $\boldsymbol{y} = \boldsymbol{F}\boldsymbol{x}$ we similarly have

$$\begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{t}_2 \end{pmatrix} \sim \phi_{p+q}\left(\begin{bmatrix} \boldsymbol{F}\boldsymbol{\mu}_x \\ -\boldsymbol{v}_x \end{bmatrix}, \begin{bmatrix} \boldsymbol{F}\boldsymbol{\Sigma}_x\boldsymbol{F}^T & \boldsymbol{F}\boldsymbol{\Sigma}_x\boldsymbol{\Gamma}_x^T \\ \boldsymbol{\Gamma}_x\boldsymbol{\Sigma}_x\boldsymbol{F} & \boldsymbol{\Delta}_x + \boldsymbol{\Gamma}_x\boldsymbol{\Sigma}_x\boldsymbol{\Gamma}_x^T \end{bmatrix}\right). \tag{10}$$

Considering the linear relationship between $\boldsymbol{x}$ and $\boldsymbol{y}$ in (9) and (10) we set

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{t}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{F} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{t}_1 \end{bmatrix}. \tag{11}$$

Now, for generating samples from $\pi(\boldsymbol{y})$ we have two options. The first one is deriving the distribution of $\boldsymbol{y}$ analytically, which is CSN, and then generating samples from it. The second method is to generate samples from $\pi(\boldsymbol{x})$ and then propagating them according to $\boldsymbol{y} = F\boldsymbol{x}$. In this latter case we have samples from $\pi(\boldsymbol{y})$ without using the actual pdf of $\boldsymbol{y}$. Generating variables from a CSN distribution is usually done by rejection sampling. Independent samples are drawn from the unconditional joint Gaussian distribution of $[\boldsymbol{x}^T, \boldsymbol{t}_1^T]^T$, and then we choose the samples which satisfy $\boldsymbol{x}|\boldsymbol{t}_1 > 0$. In our setting we generate $[\boldsymbol{x}^T, \boldsymbol{t}_1^T]^T$, and then transform them by diag$[\boldsymbol{F}, \boldsymbol{I}]$ in (11). This gives samples of $[\boldsymbol{y}^T, \boldsymbol{t}_2^T]^T$. This approach sets $\boldsymbol{t}_1 = \boldsymbol{t}_2$, and it does not

matter if we condition on $\boldsymbol{t}_1$ or $\boldsymbol{t}_2$. We are postponing the conditioning on this skewness variable being positive.

Returning to nonlinear transformations $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$. We generate $[\boldsymbol{x}^T, \boldsymbol{t}_1^T]^T$ and then propagate them through $[\boldsymbol{f}^T(\boldsymbol{x}), \boldsymbol{t}_1^T]^T$. From these predictive samples we fit a joint Gaussian for $(\boldsymbol{y}^T, \boldsymbol{t}_2^t)^t$, i.e. we estimate the mean $(\boldsymbol{\mu}_1^T, \boldsymbol{\mu}_2^T)^T$ and covariance $\begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_2 \end{bmatrix}$. Once we condition on $\boldsymbol{t}_2 > \boldsymbol{0}$, we have the approximate CSN distribution.

The benefit of using this fitting approach is that we avoid complicated maximization. For the joint Gaussian distribution it is easy to fit the mean and the covariance, while the explicit CSN form is hard to fit. For instance, using maximum likelihood to estimate the five parameters directly is not very stable. Note, however, that we have limitations related to the skewness dimension. It increases at each time step when we have additive CSN noise terms. Moreover, if the forward model is extremely nonlinear, the skewness variables may get quite extreme, and it is not easy to generate CSN samples by rejection sampling (see Appendix). The approach based on postponing the samples may then not be accurate enough.

Algorithm 1 and 2 provide summaries of the approach in the form of pseudocode. The code in Algorithm 1 is written as a prediction and updating step, inside the recursion over time. Note how the code relies on Monte Carlo samples for capturing the predictive distribution. A CSN distribution is fitted from the predicted ensembles. The steps required for this are written out in Algorithm 2. The updating step relies on conjugate forms in the CSN family.

Algorithm 1 CSN Nonlinear

Draw B samples $\boldsymbol{x}_0^1, \boldsymbol{x}_0^2, \cdots, \boldsymbol{x}_0^B \sim CSN_{n_x,q_x}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \boldsymbol{\Gamma}_0, \boldsymbol{v}_0, \boldsymbol{\Delta}_0)$ from the initial distribution.
for $t = 1$ to T **do**
    Prediction step:
    Propagate to achieve predicted samples $\boldsymbol{y}_t^i = \boldsymbol{f}\left(\boldsymbol{x}_{t-1}^i\right), i = 1, 2, \cdots B$.
    Fit a $CSN_{n_x,q_x}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Gamma}, \boldsymbol{v}, \boldsymbol{\Delta})$ distribution to the predicted samples (See Algorithm 2).
    Find the predictive distribution analytically by (1): $\pi(\boldsymbol{x}_t|\boldsymbol{Y}_{t-1}) \sim CSN_{n_x,q_x+q_\eta}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x, \boldsymbol{\Gamma}_x, \boldsymbol{v}_x, \boldsymbol{\Delta}_x)$.
    Update step: (when new observations $\boldsymbol{d}_t$ arrive)
    Find the posterior distribution $\pi(\boldsymbol{x}_t|\boldsymbol{D}_t) \sim CSN_{n,q_x+q_{\boldsymbol{\eta}}+q_\epsilon}(\boldsymbol{\mu}_{x|d}, \boldsymbol{\Sigma}_{x|d}, \boldsymbol{\Gamma}_{x|d}, \boldsymbol{v}_{x|d}, \boldsymbol{\Delta}_{x|d})$ according to (6).
    Generate B samples from the posterior $\boldsymbol{x}_t^1, \boldsymbol{x}_t^2, \cdots, \boldsymbol{x}_t^B \sim \pi(\boldsymbol{x}_t|\boldsymbol{D}_t)$.
end for

Algorithm 2 Predictive distribution estimation

1: Construct the prior Gaussian of the augmented state using (9).
2: Generate samples from this joint Gaussian distribution, $[(\boldsymbol{x}_{t-1}^i)^T, (\boldsymbol{t}_1^i)^T]^T, i = 1, 2, \cdots B$.
3: Propagate these samples through $[\boldsymbol{f}^T(\boldsymbol{x}_{t-1}^i), (\boldsymbol{t}_1^i)^T]^T, i = 1, 2, \cdots B$.
4: Empirically estimate the mean and covariance of these propagated samples to construct the predictive Gaussian distribution of the augmented state.
5: Use the conditional definition of the CSN distribution to get predictive CSN distribution, $CSN_{n_x,q_x}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Gamma}, \boldsymbol{v}, \boldsymbol{\Delta})$.

4.3. Special cases

For the nonlinear case it is even harder than in the linear case to provide guidelines for the filtering performance. But we have the similar special cases as in the linear situation in 3.2.

Ensemble Kalman filter

When we assume no skewness in the predictive distributions, and Gaussian noise terms, we get the EnKF. This special case occurs by enforcing $\boldsymbol{\Gamma}_x = 0$ in the CSN formulation.

Gaussian prediction and skewed noise

We may fit a Gaussian to the predictive samples, but allow skewness in the noise terms. This means an extension over the EnKF to accomodate more general kinds of additive noise, but not treating the dynamical system differently.

Gaussian noise with skewed prior

Like we discussed above, $q_\eta > 0$ and $q_\epsilon > 0$ mean an explosion in the skewness dimension. If we assume $q_\eta = 0$ and $q_\epsilon = 0$, the proposed filter avoids explosion in the skewness. Then we always fit a CSN of skewness dimension identical to that of the initial pdf $\pi(\boldsymbol{x}_0)$.

Re-fitting the updated distribution

One may try to reset the parameters at each step in the recursion. By enforcing $q < q_x + q_\eta$ the CSN approximation avoids skewness dimension explosion.

5. Numeric examples

This part consists of a couple of numeric examples in order to evaluate the CSNKF. We compare the results of the KF and EnKF with the CSNKF.

5.1. Synthetic linear model

Consider the following two dimensional linear system:

$$\begin{aligned}
\pi\left(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}\right) &\sim CSN_{2,1}\left(\begin{bmatrix}0.7 & -0.1\\ -0.1 & 0.5\end{bmatrix}\boldsymbol{x}_{t-1}, \operatorname{diag}\begin{bmatrix}0.005\\ 0.02\end{bmatrix}, \begin{bmatrix}0.05\\ -0.01\end{bmatrix}^T, 0, 1\right),\\
\pi\left(\boldsymbol{d}_t|\boldsymbol{x}_t\right) &\sim CSN_{2,1}\left(\boldsymbol{x}_t, \operatorname{diag}\begin{bmatrix}0.05\\ 0.2\end{bmatrix}, \begin{bmatrix}0.05\\ -0.01\end{bmatrix}^T, 0, 1\right),\\
\pi(\boldsymbol{x}_0) &\sim CSN_{2,1}\left(\begin{bmatrix}-5\\ 5\end{bmatrix}, \operatorname{diag}\begin{bmatrix}0.05\\ 0.2\end{bmatrix}, \begin{bmatrix}0.05\\ -0.01\end{bmatrix}^T, 0, 1\right). \qquad (12)
\end{aligned}$$

The dynamic system is stable because its modes are inside the unit circle. In addition, the system is observable because the state vector is measured directly. We choose the mean and variance of the Gaussian prior, process and observation distributions in the KF such that they match the related first and second moment of the CSN pdfs.

We compare the filtering results of three filters: i) KF ii) CSNKF based on analytical formulation for the predictive distribution and iii) The ensemble-based CSNKF using $B = 100$. We run 1000 replicates in the comparison. We compare results in absolute error and total mean square error (MSE). The absolute error is the absolute value of the difference between the estimated values and true value from the model at each time index. The total MSE at each time is the mean of cumulative MSE of the estimated values from beginning until current time index.

Figure 3 shows that the analytical CSN filter is very similar to the ensemble-based CSNKF. The negligible difference is caused by Monte Carlo error. The absolute error and total MSE for the CSNKF is lower than for the KF. The errors decrease for the KF as well, but not so much. The KF solution appears biased since the centered Gaussian pdfs cannot capture the true skewness.

5.2. Re entering body to the atmosphere

This example is from Julier (1998). Assume a body re-enters the atmosphere at a very high altitude and high velocity (assume the gravity is

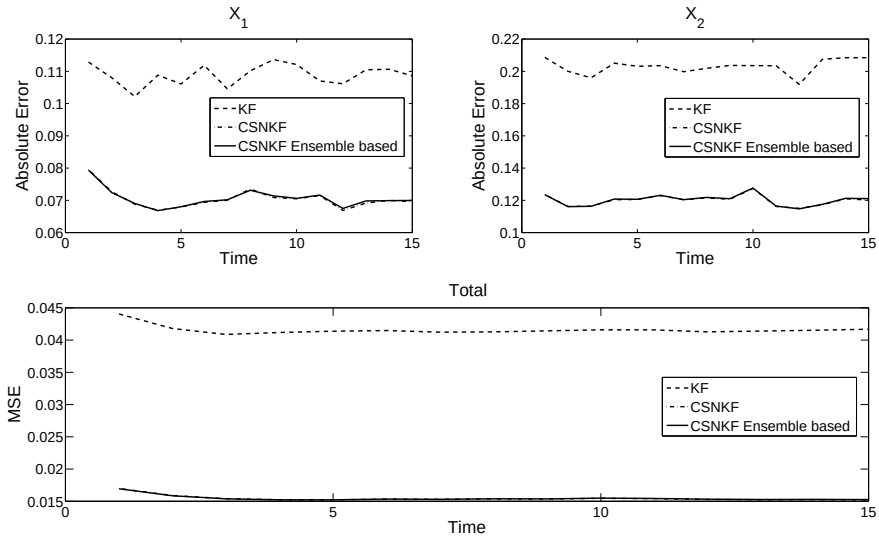

Figure 3: Linear Case Simulation: the upper plots show the absolute errors and lower plot shows the total MSE for the KF (dash line), the CSNKF based on analytically achieved predictive distribution (dash-dot line) and the ensemble based CSNKF based on postponing the conditioning idea.

negligible). The time dynamics of this nonlinear system are:

$$\begin{bmatrix} \dfrac{dx_{1,t}}{dt} \\ \dfrac{dx_{2,t}}{dt} \end{bmatrix} = \begin{bmatrix} -x_{2,t} \\ -\beta x_{2,t}^2 \exp\left(-\gamma x_{1,t}\right) \end{bmatrix} \tag{13}$$

where $\beta = 2 \cdot 10^{-3}$ is the constant ballistic coefficient and $\gamma = 5 \cdot 10^{-5}$ is a constant related to the air density with altitude. We are interested in estimating the body altitude $x_{1,t}$ and velocity $x_{2,t}$. A fourth order Runge-Kutta method is used for the dynamics equation giving $\boldsymbol{x}_t = \boldsymbol{f}\left(\boldsymbol{x}_{t-1}\right) + \boldsymbol{\eta}_t$, where η_t is additive process noise with known pdf. The position of the body is measured by radar every second, with additive noise, i.e. $d_t = x_{1,t} + \epsilon_t$.

We compare the EnKF and the proposed ensemble-based CSNKF for this nonlinear system. We consider the following initial pdf, process and

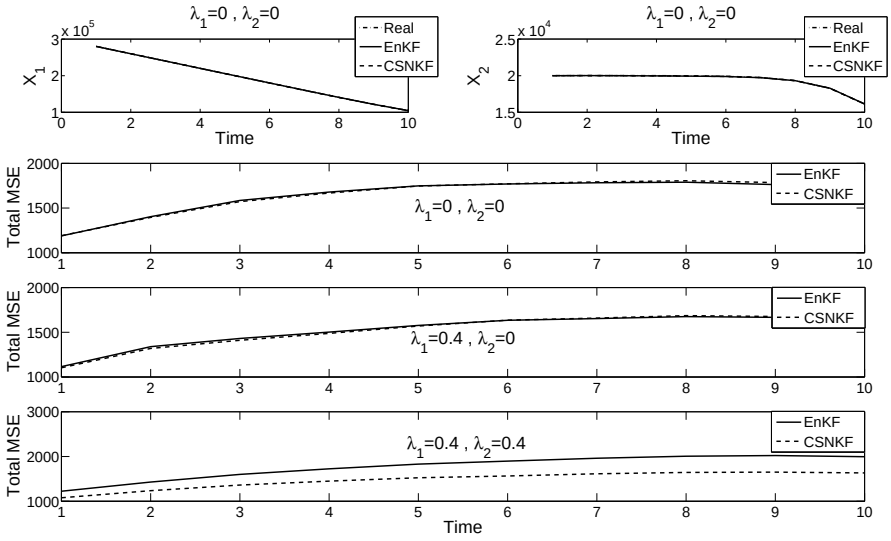

Figure 4: The re-enter body's motion simulation for different λ_1 and λ_2: upper plots show the estimated values from the EnKF (solid line), the CSNKF (dash line) and the exact value (dash-dot line) for the position (left) and velocity (right) when $\lambda_1 = \lambda_2 = 0$; the lower plots are the total MSE for the EnKF (solid line), the CSNKF (dash line) for different values of λ_1 and λ_2.

observation distributions:

$$\begin{aligned}
\pi(\boldsymbol{x}_0) &\sim CSN_{2,2}\left(\begin{bmatrix}30\\2\end{bmatrix}\cdot 10^4, \boldsymbol{\Sigma}_x, \lambda_1\boldsymbol{\Sigma}_x^{-\frac{1}{2}}, \begin{bmatrix}0\\0\end{bmatrix}, (1-\lambda_1^2)\boldsymbol{I}\right)\\
\pi(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) &\sim CSN_{2,2}\left(\boldsymbol{f}(\boldsymbol{x}_{t-1}), \boldsymbol{\Sigma}_x, \lambda_2\boldsymbol{\Sigma}_x^{-\frac{1}{2}}, \begin{bmatrix}0\\0\end{bmatrix}, (1-\lambda_2^2)\boldsymbol{I}\right)\\
\pi(d_t|\boldsymbol{x}_t) &\sim \phi_1\left([1,0]\boldsymbol{x}_t, 4\cdot 10^3\right)
\end{aligned} \tag{14}$$

Where, $\boldsymbol{\Sigma}_x = \text{diag}[10^3, 4\cdot 10^2]$, and $0 \leq \lambda_1, \lambda_2 < 1$ controls the skewness of the distributions (Flecher et al., 2009). For $\lambda_1 = 0$ or $\lambda_2 = 0$, there is no skewness and the CSN distribution is exactly the Gaussian, but note that $q = 2$ and the ensemble-based CSNKF still has the capability of adding skewness in the nonlinear prediction. The observation noise is Gaussian.

Figure 4, shows the simulation results for this case with different prior and process noise skewness's with $B = 100$ ensemble members. The upper plots present the estimated values and real value when $\lambda_1 = \lambda_2 = 0$, while the lower

plots show the filter's performance in the total MSE sense. For $\lambda_1 = \lambda_2 = 0$, the performance of both filters are almost equal (the second plot from top), as it should be. The CSNKF recognizes Gaussianity. The prior skewness increased by putting $\lambda_1 = 0.4$. The CSNKF is better than the EnKF in the beginning, but after some steps the results are the same (the third plot from top). This plot indicates that some skewness is imposed by the dynamics in the initial phase when there is much uncertainty in the predicted distribution. But since we measure the altitude directly with Gaussian symmetric noise, the skewness seems to die out over time in the filter and the CSNKF converges to EnKF as time goes.

Consider the case where $\lambda_1 = \lambda_2 = 0.4$. The lowest plot in Figure 4 shows the simulation results for this case with more skewness. This plot shows that the differences between two filters can be quite significant. When we have additive CSN noise terms in the process model, the observations induce less symmetry than they did before. Note that, in order to reduce the random effects all simulations are done for 500 replicates (Monte Carlo runs).

5.3. Saturation estimation in petroleum reservoir

We consider a petroleum reservoir model with nonlinear dynamics $\boldsymbol{f}(\boldsymbol{x}_t)$ consisting of the partial differential equations governing fluid flow in porous media. In our setting the dynamic model takes the form of a black-box routine. There are several commercial and non-commercial reservoir simulators (i.e. ECLIPSE by GeoQuest). We use the MATLAB Reservoir Simulation Toolbox (MRST). For more details see www.sintef.no/Projectweb/MRST/. The EnKF is a typical tool for the state estimation in very high dimensional systems (Sætrom and Omre, 2011; Rezaie and Eidsvik, 2012). The state estimation in petroleum reservoirs is also known as history matching. We are going to compare the results of the EnKF and the proposed ensemble based CSNKF.

We study a reservoir with $15 \times 15 \times 4$ grid cells. For each grid cell we set the porosity and permeability values from the SPE10 data set, see Christie and Blunt (2001). The porosity and permeability are static variables, and do not change during production. Upper plots in Figure 5 show the fixed porosity and permeability values on the grid. Petroleum production is done by one injection well and one production well (see lower left plot of Figure 5). At time $t = 0$ water is pumped at the injection well for replacing and moving oil to the production well. Based on fluid dynamics, the flow is faster where the permeability and porosity is high. The lower plots of Figure 5 show the

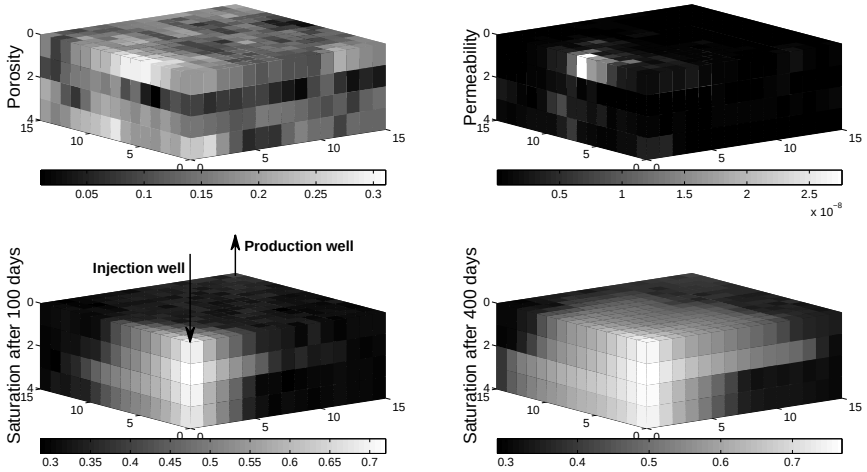

Figure 5: The reservoir description: upper plots are the reference porosity and permeability from SPE10 data set; lower left plot shows the position of the injection and production wells; the lower left plot is the saturation after 100 days of running the simulator and the lower right plot is the saturation after 400 days.

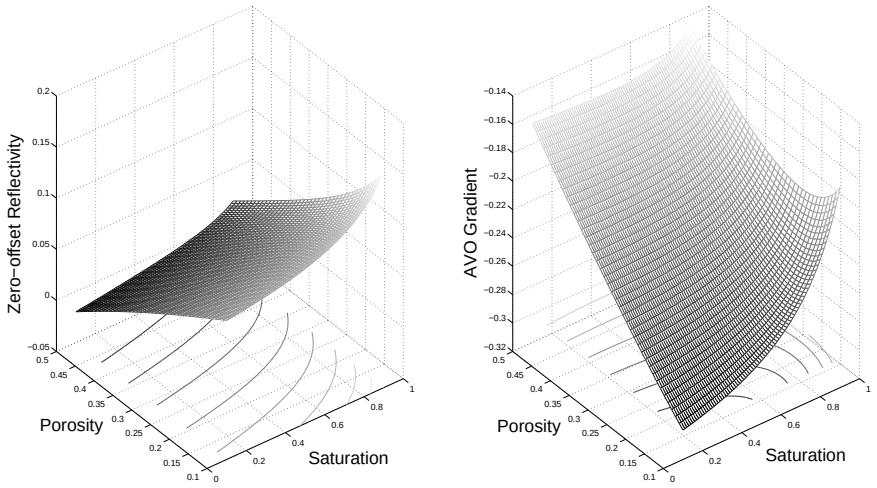

Figure 6: AVO seismic: zero-offset reflectivity (left) and AVO gradient (right) versus saturation and porosity.

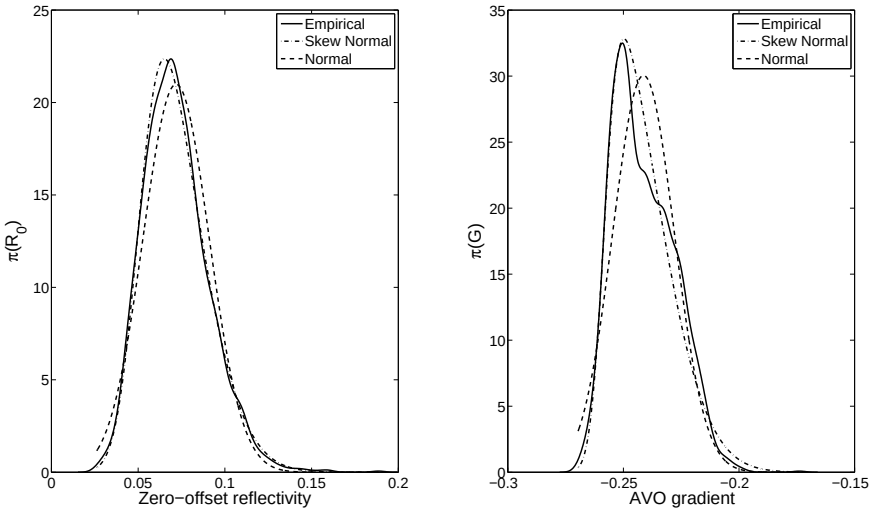

Figure 7: AVO seismic data distribution fitting based on SPE10 data set: the empirical distribution of the AVO seismic measurements (solid line); the Gaussian fitting results (dash line) and the skew normal fitting results (dash-dot line), the skew normal distribution has closer distribution to the empirical distribution than the Gaussian.

saturations of all grid cells after 100 days and 400 days of production. The oil in grid cells near the injection well have been replaced by the injected water, pushing oil towards the production well.

In petroleum exploration one goal is to predict the oil production, another goal is to locate the remaining oil in the reservoir, over time. Thus, the saturation at all grid cells is considered as the state vector $\boldsymbol{x}_t$. For achieving these filtering and prediction purposes it is useful to acquire seismic data ($\boldsymbol{d}_t$) over time. The data we consider here consists of seismic amplitude versus offset (AVO) observations: zero-offset reflectivity (R_0) and AVO gradient (G) at the top reservoir. These data are informative of the elastic properties of the reservoir, and this again depends on the saturation, as well as other reservoir variables. Figure 6 presents the expected AVO response, i.e. $E(\boldsymbol{d}_{t,i}|x_{t,i})$ for any grid cell i, for a range of saturation and porosity levels. These seismic response levels are obtained using a Reuss fluid mixing model and Gassmann's formula for fluid substitution, see e.g. Mavko et al. (2003) and Eidsvik et al. (2008). We assume the elastic properties of the cap rock above the reservoir unit are fixed. We assume that a baseline seismic survey is performed at the

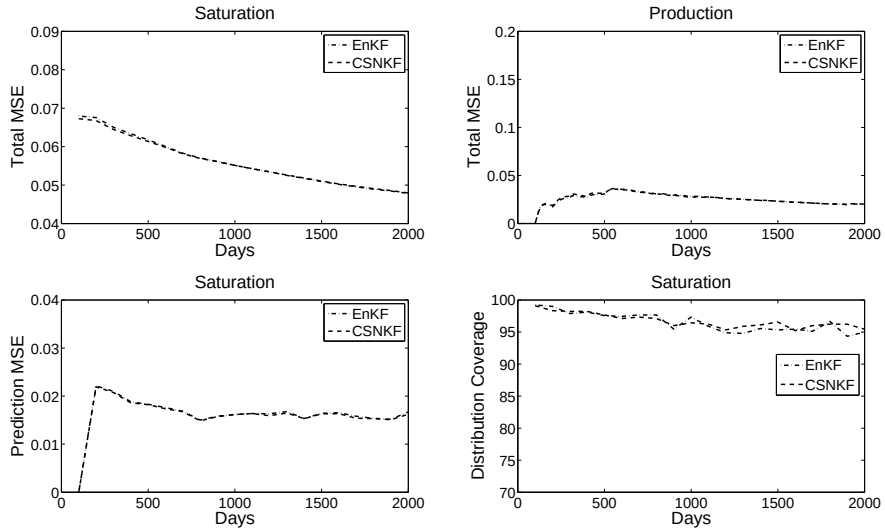

Figure 8: Reservoir parameter estimation for $\lambda = 0$ case: the performance of the CSNKF (dash line) is similar to the EnKF (dash-dot line).

initial time $t = 0$. Subsequently, monitoring surveys are performed every 100 days for $T = 20$ time steps, i.e. day $100, 200 \cdots 2000$ of production.

We consider two cases: i) additive Gaussian process and observation noises ii) additive Gaussian process noise and additive CSN observation noises. Figure 7 indicates that a skewed model is a more realistic fit to the seismic AVO data R_0 (left) and G (right). The EnKF and the CSNKF algorithms are compared, using $B = 100$ samples. We study the filtering and one-step predictive distribution of the saturation. We also look at the production in the reservoir, and compare the results with the true realizations. For all output variables of interest we compare the total MSE of estimated values and the 95 % distribution coverage.

Consider Gaussian observation noise terms. Figure 8 shows the results for EnKF and CSNKF. We note only negligible differences due to Monte Carlo error. We next include skewness in the seismic AVO measurement noise terms, setting $\lambda = 0.4$ which is a value inspired from the SPE10 data, see Figure 7. In Figure 9 the EnKF performance is worse than the CSNKF. We see that CSNKF has smaller filtering total MSE and one-step prediction MSE (left) for the saturation variables. In terms of predicting the production, the CSNKF also has smaller MSE. However, for all these attributes, the CSNKF

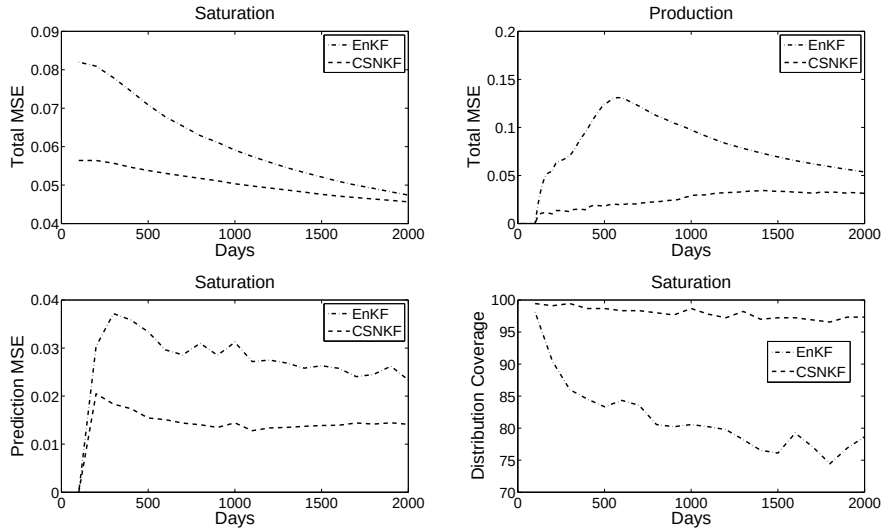

Figure 9: Reservoir parameter estimation for $\lambda = 0.4$ case: the performance of the CSNKF (dash line) is superior to the EnKF (dash-dot line).

again appears to get closer to the EnKF over time. The 95 % distribution coverage (lower right plot of Figure 9) of both filters are high at the beginning steps but it goes down fast for the EnKF, which undercovers a lot here.

6. Closing remarks

The closed skew normal distribution was introduced as an extension of the Gaussian distributions for linear and nonlinear filtering. The suggested approach includes skewed versions of the Kalman filter and the ensemble Kalman filter. The common Gaussian Kalman filter variants are special cases of the proposed filters. We implemented the proposed methods on a linear model, a bivariate nonlinear falling body example and a high dimensional nonlinear petroleum reservoir model. In these examples we tried to study the various effects of using a closed skew normal distribution in the filtering problem.

Our presentation relies on additive noise terms, and that these are Gaussian or closed skew normal distributed. We believe this could be extended, but in general it is not straightforward to assign the skewness dimensions of the predictive distribution. In the current formulation we aimed to tune

the skewness with a flexible formulation used before, which ties the skewness dimension to the square root of the covariance matrix. The augmented Gaussian formulation of state and skewness variable then allows fast fitting of the predictive closed skew normal distribution at each step.

Recall that the computational methods for fitting the ensemble-based version of the filter are based on rejection sampling. In high dimensional situations, with much nonlinearity, we experienced some challenges in the acceptance rate here. Other sampling methods may be more useful.

7. Acknowledgments

We thank the sponsors of the Uncertainty in Reservoir Evaluation (URE) project at NTNU.

Appendix A. CSN definition and properties

Assume $\boldsymbol{x} \in \Re^{n_x \times 1}$ and $\boldsymbol{y} \in \Re^{n_y \times 1}$ are joint Gaussian random vectors:

$$\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} \sim \phi_{n_x+n_y} \left(\begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_x & \boldsymbol{\Sigma}_{x,y} \\ \boldsymbol{\Sigma}_{y,x} & \boldsymbol{\Sigma}_y \end{bmatrix} \right). \tag{A.1}$$

We then have:

$$\begin{aligned} \pi(\boldsymbol{z}) &= \pi(\boldsymbol{x}|\boldsymbol{y}>0) = \frac{\pi(\boldsymbol{y}>0|\boldsymbol{x})\,\pi(\boldsymbol{x})}{\pi(\boldsymbol{y}>0)} \qquad (A.2) \\ &= \left[1-\Phi_{n_y}\left(0;\boldsymbol{\mu}_y,\boldsymbol{\Sigma}_y\right)\right]^{-1}\left[1-\Phi_{n_y}\left(0;\boldsymbol{\mu}_{y|x},\boldsymbol{\Sigma}_{y|x}\right)\right]\phi_{n_x}(\boldsymbol{x};\boldsymbol{\mu}_x,\boldsymbol{\Sigma}_x), \end{aligned}$$

where $\boldsymbol{\mu}_{y|x} = \boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{y,x}\boldsymbol{\Sigma}_x^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_x)$, and $\boldsymbol{\Sigma}_{y|x} = \boldsymbol{\Sigma}_y - \boldsymbol{\Sigma}_{y,x}\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{x,y}$. Simple re-writing now gives the standard CSN parameterization $\boldsymbol{z} \sim CSN_{n_z,q_z}(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\Gamma},\boldsymbol{v},\boldsymbol{\Delta})$. We have $n_z = n_x$, $q_z = n_y$, $\boldsymbol{\mu} = \boldsymbol{\mu}_x$, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_x$, $\boldsymbol{\Gamma} = \boldsymbol{\Sigma}_{y,x}\boldsymbol{\Sigma}_x^{-1}$, $\boldsymbol{v} = -\boldsymbol{\mu}_y$ and $\boldsymbol{\Delta} = \boldsymbol{\Sigma}_y - \boldsymbol{\Sigma}_{y,x}\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{x,y}$.

The following properties of CSN random variables are derived in González-Farías et al. (2004).

Property 1: If $\boldsymbol{x}_1 \sim CSN_{b_x,q_{x_1}}(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_1,\boldsymbol{\Gamma}_1,\boldsymbol{v}_1,\boldsymbol{\Delta}_1)$ and $\boldsymbol{x}_2 \sim CSN_{n_x,q_{x_2}}(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_2,\boldsymbol{\Gamma}_2,\boldsymbol{v}_2,\boldsymbol{\Delta}_2)$

are independent, then $\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2 \sim CSN_{n_x, q_{x_1}+q_{x_2}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Gamma}, \boldsymbol{v}, \boldsymbol{\Delta})$, where:

$$\begin{aligned} \boldsymbol{\mu} &= \boldsymbol{\mu}_1 + \boldsymbol{\mu}_2, \quad \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2, \quad \boldsymbol{\Gamma} = \begin{pmatrix} \boldsymbol{\Gamma}_1 \boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}^{-1} \\ \boldsymbol{\Gamma}_2 \boldsymbol{\Sigma}_2 \boldsymbol{\Sigma}^{-1} \end{pmatrix}, \quad \boldsymbol{v} = \begin{pmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \end{pmatrix} \\ \boldsymbol{\Delta} &= \begin{pmatrix} \boldsymbol{\Delta}_1 + \boldsymbol{\Gamma}_1 \boldsymbol{\Sigma}_1 \boldsymbol{\Gamma}_1^T - \boldsymbol{\Gamma}_1 \boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_1 \boldsymbol{\Gamma}_1^T & -\boldsymbol{\Gamma}_1 \boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_2 \boldsymbol{\Gamma}_2^T \\ -\boldsymbol{\Gamma}_2 \boldsymbol{\Sigma}_2 \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_1 \boldsymbol{\Gamma}_1^T & \boldsymbol{\Delta}_2 + \boldsymbol{\Gamma}_2 \boldsymbol{\Sigma}_2 \boldsymbol{\Gamma}_2^T - \boldsymbol{\Gamma}_2 \boldsymbol{\Sigma}_2 \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_2 \boldsymbol{\Gamma}_2^T \end{pmatrix}. \end{aligned} \tag{A.3}$$

Property 2: If $\boldsymbol{x} \sim CSN_{n_x,q_x}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x, \boldsymbol{\Gamma}_x, \boldsymbol{v}_x, \boldsymbol{\Delta}_x)$ and $\boldsymbol{F}$ is $n_y \times n_x$ matrix $(n_y \leq n_x)$ then $\boldsymbol{y} = \boldsymbol{F}\boldsymbol{x} \sim CSN_{n_y,q_x}(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y, \boldsymbol{\Gamma}_y, \boldsymbol{v}_y, \boldsymbol{\Delta}_y)$, where:

$$\begin{aligned} \boldsymbol{\mu}_y &= \boldsymbol{F}\boldsymbol{\mu}_x, \quad \boldsymbol{\Sigma}_y = \boldsymbol{F}\boldsymbol{\Sigma}_x\boldsymbol{F}^T, \quad \boldsymbol{\Gamma}_y = \boldsymbol{\Gamma}_x\boldsymbol{\Sigma}_x\boldsymbol{F}^T\boldsymbol{\Sigma}_y^{-1}, \quad \boldsymbol{v}_y = \boldsymbol{v}_x \\ \boldsymbol{\Delta}_y &= \boldsymbol{\Delta}_x + \boldsymbol{\Gamma}_x\boldsymbol{\Sigma}_x\boldsymbol{\Gamma}_x^T - \boldsymbol{\Gamma}_x\boldsymbol{\Sigma}_x\boldsymbol{F}^T\boldsymbol{\Sigma}_y^{-1}\boldsymbol{F}\boldsymbol{\Sigma}_x\boldsymbol{\Gamma}_x^T. \end{aligned} \tag{A.4}$$

Property 3: Sampling from a CSN distribution: Let $\boldsymbol{E}_1 \sim \phi_p(\boldsymbol{E}_1; 0, \boldsymbol{\Sigma}_x)$ and $\boldsymbol{E}_2 \sim \phi_q(\boldsymbol{E}_2; 0, \boldsymbol{\Delta}_x)$ be independent random vectors, then $\boldsymbol{x} = \boldsymbol{\mu}_x + \boldsymbol{E}_1 \sim CSN_{p,q}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x, \boldsymbol{\Gamma}_x, \boldsymbol{v}_x, \boldsymbol{\Delta}_x)$ provided that $\boldsymbol{y} = -\boldsymbol{v}_x + \boldsymbol{\Gamma}_x\boldsymbol{E}_1 + \boldsymbol{E}_2 > 0$.
Property 4: If $\boldsymbol{x} \sim CSN_{n,q_x}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x, \boldsymbol{\Gamma}_x, \boldsymbol{v}_x, \boldsymbol{\Delta}_x)$ and $\pi(\boldsymbol{d}|\boldsymbol{x}) \sim CSN_{m,q_e}(\boldsymbol{H}\boldsymbol{x}, \boldsymbol{\Sigma}_e, \boldsymbol{\Gamma}_e, \boldsymbol{v}_e, \boldsymbol{\Delta}_e)$, then $\pi(\boldsymbol{x}|\boldsymbol{d}) \sim CSN_{n,q_x+q_e}(\boldsymbol{\mu}_{x|d}, \boldsymbol{\Sigma}_{x|d}, \boldsymbol{\Gamma}_{x|d}, \boldsymbol{v}_{x|d}, \boldsymbol{\Delta}_{x|d})$, where the parameters are defined in (6)

For completeness we sketch a proof of this last property. Define $\boldsymbol{x} = [\boldsymbol{t}|\boldsymbol{u} > 0]$ and $\boldsymbol{e} = [\boldsymbol{s}|\boldsymbol{v} > 0]$. Let $\boldsymbol{t}$, $\boldsymbol{s}$ and $\boldsymbol{v}$ be mutually independent and $\boldsymbol{u}$ and $\boldsymbol{v}$ mutually independent. Then we have

$$\begin{bmatrix} \boldsymbol{t} \\ \boldsymbol{r} = \boldsymbol{H}\boldsymbol{t} + \boldsymbol{s} \\ \boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix} \sim \phi_{n+m+q_x+q_e}\left(\begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{H}\boldsymbol{\mu}_x \\ -\boldsymbol{v}_x \\ -\boldsymbol{v}_e \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_x & \boldsymbol{\Sigma}_x\boldsymbol{H}^T & \boldsymbol{\Sigma}_x\boldsymbol{\Gamma}_x^T & 0 \\ \boldsymbol{H}\boldsymbol{\Sigma}_x & \boldsymbol{H}\boldsymbol{\Sigma}_x\boldsymbol{H}^T + \boldsymbol{\Sigma}_e & \boldsymbol{H}\boldsymbol{\Sigma}_x\boldsymbol{\Gamma}_x^T & \boldsymbol{\Sigma}_e\boldsymbol{\Gamma}_e^T \\ \boldsymbol{\Gamma}_x\boldsymbol{\Sigma}_x & \boldsymbol{\Gamma}_x\boldsymbol{\Sigma}_x\boldsymbol{H}^T & \boldsymbol{\Delta}_x + \boldsymbol{\Gamma}_x\boldsymbol{\Sigma}_x\boldsymbol{\Gamma}_x^T & 0 \\ 0 & \boldsymbol{\Gamma}_e\boldsymbol{\Sigma}_e & 0 & \boldsymbol{\Delta}_e + \boldsymbol{\Gamma}_e\boldsymbol{\Sigma}_e\boldsymbol{\Gamma}_e^T \end{bmatrix} \right). \tag{A.5}$$

Besides, from the multivariate normal distribution;

$$\begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix} \sim \phi_{n_1+n_2}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_2 \end{bmatrix} \right), \tag{A.6}$$

then $[\boldsymbol{x}_1|\boldsymbol{x}_2] \sim \phi_{n_1}\left(\boldsymbol{x}_1; \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_2^{-1}\left(\boldsymbol{x}_2 - \boldsymbol{\mu}_2\right), \boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_{21}\right)$.

Thus, we have

$$\begin{bmatrix} \boldsymbol{t}|\boldsymbol{r} \\ \boldsymbol{u}|\boldsymbol{r} \\ \boldsymbol{v}|\boldsymbol{r} \end{bmatrix} \sim \phi_{n+q_x+q_e}\left(\begin{bmatrix} \boldsymbol{\mu}_{t|r} \\ \boldsymbol{\mu}_{u|r} \\ \boldsymbol{\mu}_{v|r} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{t|r} & \boldsymbol{\Sigma}_{tu|r} & \boldsymbol{\Sigma}_{tv|r} \\ \boldsymbol{\Sigma}_{ut|r} & \boldsymbol{\Sigma}_{u|r} & \boldsymbol{\Sigma}_{uv|r} \\ \boldsymbol{\Sigma}_{vt|r} & \boldsymbol{\Sigma}_{vu|r} & \boldsymbol{\Sigma}_{v|r} \end{bmatrix}\right), \tag{A.7}$$

$$\begin{aligned}
\boldsymbol{\mu}_{t|r} &= \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_x\boldsymbol{H}^T\left[\boldsymbol{H}\boldsymbol{\Sigma}_x\boldsymbol{H}^T + \boldsymbol{\Sigma}_e\right]^{-1}\left(\boldsymbol{r} - \boldsymbol{H}\boldsymbol{\mu}_x\right) \\
\boldsymbol{\mu}_{u|r} &= -\boldsymbol{v}_x + \boldsymbol{\Gamma}_x\boldsymbol{\Sigma}_x\boldsymbol{H}^T\left[\boldsymbol{H}\boldsymbol{\Sigma}_x\boldsymbol{H}^T + \boldsymbol{\Sigma}_e\right]^{-1}\left(\boldsymbol{r} - \boldsymbol{H}\boldsymbol{\mu}_x\right) \\
\boldsymbol{\mu}_{v|r} &= -\boldsymbol{v}_e + \boldsymbol{\Gamma}_e\boldsymbol{\Sigma}_e\left[\boldsymbol{H}\boldsymbol{\Sigma}_x\boldsymbol{H}^T + \boldsymbol{\Sigma}_e\right]^{-1}\left(\boldsymbol{r} - \boldsymbol{H}\boldsymbol{\mu}_x\right) \\
\boldsymbol{\Sigma}_{t|r} &= \boldsymbol{\Sigma}_x - \boldsymbol{\Sigma}_x\boldsymbol{H}^T\left[\boldsymbol{H}\boldsymbol{\Sigma}_x\boldsymbol{H}^T + \boldsymbol{\Sigma}_e\right]^{-1}\boldsymbol{H}\boldsymbol{\Sigma}_x \\
\boldsymbol{\Sigma}_{u|r} &= \left[\boldsymbol{\Delta}_x + \boldsymbol{\Gamma}_x\boldsymbol{\Sigma}_x\boldsymbol{\Gamma}_x^T\right] - \boldsymbol{\Gamma}_x\boldsymbol{\Sigma}_x\boldsymbol{H}^T\left[\boldsymbol{H}\boldsymbol{\Sigma}_x\boldsymbol{H}^T + \boldsymbol{\Sigma}_e\right]^{-1}\boldsymbol{H}\boldsymbol{\Sigma}_x\boldsymbol{\Gamma}_x^T \\
\boldsymbol{\Sigma}_{v|r} &= \left[\boldsymbol{\Delta}_e + \boldsymbol{\Gamma}_e\boldsymbol{\Sigma}_e\boldsymbol{\Gamma}_e^T\right] - \boldsymbol{\Gamma}_e\boldsymbol{\Sigma}_e\left[\boldsymbol{H}\boldsymbol{\Sigma}_x\boldsymbol{H}^T + \boldsymbol{\Sigma}_e\right]^{-1}\boldsymbol{\Sigma}_e\boldsymbol{\Gamma}_e^T \\
\boldsymbol{\Sigma}_{ut|r} &= \boldsymbol{\Gamma}_x\boldsymbol{\Sigma}_x - \boldsymbol{\Gamma}_x\boldsymbol{\Sigma}_x\left[\boldsymbol{H}\boldsymbol{\Sigma}_x\boldsymbol{H}^T + \boldsymbol{\Sigma}_e\right]^{-1}\boldsymbol{H}\boldsymbol{\Sigma}_x \\
\boldsymbol{\Sigma}_{vt|r} &= 0 - \boldsymbol{\Gamma}_e\boldsymbol{\Sigma}_e\left[\boldsymbol{H}\boldsymbol{\Sigma}_x\boldsymbol{H}^T + \boldsymbol{\Sigma}_e\right]^{-1}\boldsymbol{H}\boldsymbol{\Sigma}_x \\
\boldsymbol{\Sigma}_{vt|r} &= 0 - \boldsymbol{\Gamma}_x\boldsymbol{\Sigma}_x\boldsymbol{H}^T\left[\boldsymbol{H}\boldsymbol{\Sigma}_x\boldsymbol{H}^T + \boldsymbol{\Sigma}_e\right]^{-1}\boldsymbol{\Sigma}_e\boldsymbol{\Gamma}_e^T
\end{aligned} \tag{A.8}$$

Let $\boldsymbol{w} = \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix}$, then $\begin{bmatrix} \boldsymbol{t}|\boldsymbol{r} \\ \boldsymbol{w}|\boldsymbol{r} \end{bmatrix} \sim \phi_{n+q_x+q_e}\left(\begin{bmatrix} \boldsymbol{\mu}_{t|r} \\ \boldsymbol{\mu}_{w|r} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{t|r} & \boldsymbol{\Sigma}_{tw|r} \\ \boldsymbol{\Sigma}_{wt|r} & \boldsymbol{\Sigma}_{w|r} \end{bmatrix}\right)$.

From the conditional definition of the CSN model, the posterior distribution is $\pi(\boldsymbol{x}|\boldsymbol{d}) = \pi(\boldsymbol{t}|\boldsymbol{r}, \boldsymbol{w} > 0) = CSN_{n,q_x+q_e}(\boldsymbol{\mu}_{x|d}, \boldsymbol{\Sigma}_{x|d}, \boldsymbol{\Gamma}_{x|d}, \boldsymbol{v}_{x|d}, \boldsymbol{\Delta}_{x|d})$.

References

Azzalini, A., Dalla-Valle, A., 1996. The multivariate skew-normal distribution. Biometrika 83, 715–726.

Christie, M.A., Blunt, M.J., 2001. Tenth spe comparative solution project: A comparison of upscaling techniques. SPE Reservoir Engineering and Evaluation 4, 308–317.

Domínguez-Molina, J.A., González-Farías, G., Gupta, A.K., 2003. The multivariate closed skew normal distribution. Technical Report 3, –.

Doucet, A., de Freitas, N., Gordon, N., 2001. Sequential Monte-Carlo Methods in Practice. Springer.

Eidsvik, J., Bhattacharjya, D., Mukerji, T., 2008. Value of information of seismic amplitude and CSEM resistivity. Geophysics 73, 59–69.

Evensen, G., 2001. The ensemble kalman filter: Theoretical formulation and practical implementation. Ocean Dynamics 53, 343–367.

Evensen, G., 2009. Data assimilation, The Ensemble Kalman Filter. Springer.

Flecher, C., Naveau, P., Allard, D., 2009. Estimating the closed skew-normal distribution parameters using weighted moments. Statistics and Probability Letters 79, 19771984.

Genton, M.G., 2004. Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality - ch: The closed skew-normal distribution, G. Gonzalez-Farias and J. A. Domnguez-Molina and A. K. Gupta. Chapman and Hall / CRC.

González-Farías, G., Domínguez-Molina, J.A., Gupta, A.K., 2004. Additive properties of skew normal random vectors. Statistical Planning and Inference 126, 521534.

Gupta, A.K., González-Farías, G., Domínguez-Molina, J.A., 2004. A multivariate skew normal distribution. Multivariate Analysis 89, 181–190.

Jazwinsky, A., 1970. Stochastic Processes and Filtering Theory. Academic Press, New York.

Julier, S.J., 1998. Skewed approach to filtering, in: Aerospace-Defense Sensing and Controls, International Society for Optics and Photonics. pp. 271–282.

Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. ASME J. Basic Eng. 82, 35–45.

Karimi, O., Omre, H., Mohammadzadeh, M., 2010. Bayesian closed-skew gaussian inversion of seismic avo data into elastic material properties. Geophysics 75, 1–11.

Mavko, G., Mukerji, T., Dvorkin, J., 2003. Rock Physics Handbook - Tools for Seismic Analysis in Porous Media. Cambridge University Press.

Naveau, P., Genton, M.G., Shen, X., 2005. A skewed kalman filter. Journal of Multivariate Analysis 94, 382 – 400.

Rezaie, J., Eidsvik, J., 2012. Shrinked $(1-\alpha)$ ensemble kalman filter and α gaussian mixture filter. Computational Geosciences 16, 837–852.

Sætrom, J., Omre, H., 2011. Ensemble kalman filtering with shrinkage regression techniques. Computational Geosciences 15, 271–292.

Sakov, P., Oke, P.R., 2008. Implications of the form of the ensemble transformation in the ensemble square root filters. Monthly Weather Review 136, 1042–1053.

Stordal, A.S., Karlsen, H.A., Nævdal, G., Skaug, H.J., Valls, B., 2010. Bridging the ensemble kalman filter and particle filters: the adaptive gaussian mixture filter. Computational Geosciences 15, 293–305.