Bayesian Approach of the skewed Kalman filter applied to an elastically supported structure

K. Runtemund, G. Müller
Technische Universität München, Chair for Structural Mechanics
Arcisstraße 21, 80333 München, Germany
e-mail: **katrin.runtemund@bv.tum.de**

Abstract

In a Bayesian Approach the Kalman filter can be regarded as recursive Bayesian estimator and be described as Bayesian dynamic network. The linear dynamic system discretized in the time domain follows a first order hidden Markov process where uncertainties in the system model and the measurement model are assumed to be Gaussian and modeled as uncorrelated white noise processes. As the assumption of linearity and Gaussianity is often violated a Bayesian approach of an extended skewed Kalman filter is derived which allows to consider a nonlinear dynamic system excited by a process with skew-normal probability distributions.

1 Introduction

Slender structures such as pedestrian bridges, long-span frame structures or high-rise buildings under transient loads as wind, traffic, earthquake, etc. are excited to vibrations. Uncertainties in the model of the structure and the loads lead to variations in the prediction of the expected vibration and the associated radiation of sound in buildings.

Using the Bayesian Approach the uncertain parameters are modeled as random variables by incorporating prior knowledge and observational evidence: starting from a prior density function including all available information about the parameters, the posterior density function is estimated at each time step based on measurements accompanying the construction process. The parameters can thus be estimated evaluating the a posteriori density function. Assuming a linear Gaussian model where both, the measurement error and the system error are modeled as additive zero mean Gaussian noises, the Bayesian Approach leads to the Kalman filter, an unbiased minimum variance estimate.

Identification problems of unknown parameters are in general nonlinear and can be solved using the extended Kalman filter (EKF) which linearizes about the current mean and covariance. The method is applied to a single degree of freedom system in order to identify the stiffness and damping parameters using simulated measurement data.

As the EKF always approximates the posterior distribution as a Gaussian, the obtained estimate is biased and the filter may provide poor performance in some nonlinear problems where the true posterior is non-Gaussian (e.g. multimodal or heavily skewed) [1]. In order to apply the Kalman filter to a wider range of distributions than the normal one, Naveau introduced a skewed Kalman filter which is based on the skew-normal distributions [2]. The method is derived in a Bayesian framework and extended for a parameter identification problem.

2 The Kalman filter

The Kalman filter was developed in 1960 by Rudolf Kálmán. It is an optimal recursive algorithm to estimate the state $\mathbf{x}$ of a linear dynamic system discretized in the time domain using noisy measurement data $\mathbf{z}$ [3][4]. The state-space model is generated by two equations, the system and the observational

PREDICTION (TIME UPDATE)	CORRECTION (MEASUREMENT UPDATE)
1) Project the state ahead (*priori estimate*) $\bar{\mathbf{x}}_k = \mathbf{T}_k \hat{\mathbf{x}}_{k-1} + \mathbf{B}_k \mathbf{f}_{k-1}$ 2) Project the *priori* error covariance ahead $\boldsymbol{\Sigma}_{\overline{xx},k} = \mathbf{T}_k \Sigma_{\hat{x}\hat{x},k-1} \mathbf{T}_k^T + \mathbf{B}_k \boldsymbol{\Sigma}_{ff,k-1} \mathbf{B}_k^T + \mathbf{S}_k \boldsymbol{\Sigma}_{ww,k-1} \mathbf{S}_k^T$	1) Compute the Kalman gain matrix $\mathbf{K}_k = \boldsymbol{\Sigma}_{\overline{xx},k} \mathbf{A}_k^T \left(\boldsymbol{\Sigma}_{vv,k} + \mathbf{A}_k \boldsymbol{\Sigma}_{\overline{xx},k} \mathbf{A}_k^T\right)^{-1}$ 2) Update estimate with measurement (*posteriori estimate*) $\hat{\mathbf{x}}_k = \bar{\mathbf{x}}_k + \mathbf{K}_k (\mathbf{z}_k - \mathbf{A}_k \bar{\mathbf{x}}_k)$ 3) Update *a posteriori* error covariance $\boldsymbol{\Sigma}_{\hat{x}\hat{x},k} = \boldsymbol{\Sigma}_{\overline{xx},k} - \mathbf{K}_k \mathbf{A}_k^T \overline{\boldsymbol{\Sigma}}_{\overline{xx},k}$

Figure 1: Prediction-correction procedure of the Kalman filter

equations. The latter describe the linear relation between the n-dimensional vector of observations $\mathbf{z}_k$ at time k and the unobserved m-dimensional state vector $\mathbf{x}_k$

$$\mathbf{z}_k = \mathbf{A}_k \mathbf{x}_k + \mathbf{v}_k \tag{1}$$

where $\mathbf{A}_k$ is a $n \times m$ matrix of scalars and $\mathbf{v}_k$ is an added n-dimensional noise vector to consider random measurement errors e.g. due to sensor inaccuracy. The system equation

$$\mathbf{x}_k = \mathbf{T}_k \mathbf{x}_{k-1} + \mathbf{B}_k \mathbf{f}_{k-1} + \mathbf{S}_k \mathbf{w}_{k-1} \tag{2}$$

describes the temporal changes in the state of the linear dynamic system. The $m \times m$ transfer matrix $\mathbf{T}_k$ relates the state $k-1$ at the previous time step to the state at the current step k. Optionally a system input (control) vector (e.g. external loads) $\mathbf{f}_{k-1}$ of length f can be included which is related to the actual state $\mathbf{x}_k$ by the $m \times f$ input matrix $\mathbf{B}_k$.The model uncertainties or disturbances are represented by the added m-dimensional noise vector $\mathbf{w}_{k-1}$ which is related to the actual state by the $m \times m$ matrix $\mathbf{S}_k$.

The Kalman filter is based on a Gaussian noise model, i.e. the measurement error $\mathbf{v}_k$ as well as the state error $\mathbf{w}_k$ are modeled as independent, white noises with normal distribution

$$\begin{aligned} \mathbf{v}_k &\sim N(\mathbf{0}, \boldsymbol{\Sigma}_{vv,k}) \\ \mathbf{w}_k &\sim N(\mathbf{0}, \boldsymbol{\Sigma}_{ww,k}) \end{aligned} \tag{3}$$

with the $n \times n$ measurement noise covariance matrix $\boldsymbol{\Sigma}_{vv,k}$ and the $m \times m$ process noise covariance matrix $\boldsymbol{\Sigma}_{ww,k}$.

The algorithm is characterized by an iterative *prediction-correction* structure as shown in figure 1 [5]. The overline " ¯" indicates the prediction whereas the estimate is marked by the hat "^".
In the *prediction step* a time update of the current state and error covariance is taken in order to obtain a prior estimate of the process state $\bar{\mathbf{x}}_k$ and its associated error covariance for the next time step. The priori error and its covariance are given by

$$\begin{aligned} \boldsymbol{\varepsilon}_{\hat{x},k-1} &= \mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1} \\ \boldsymbol{\Sigma}_{\hat{x}\hat{x},k-1} &= \mathbb{E}\left[\boldsymbol{\varepsilon}_{\hat{x},k-1}^T \boldsymbol{\varepsilon}_{\hat{x},k-1}\right] \end{aligned} \tag{4}$$

The time-update of the current state is calculated from the undisturbed system equation

$$\bar{\mathbf{x}}_\mathrm{k} = \mathbf{T}_k \hat{\mathbf{x}}_{\mathrm{k}-1} + \mathbf{B}_k \mathbf{f}_{k-1} \tag{5}$$

where the prediction error

$$\boldsymbol{\varepsilon}_{\bar{\mathrm{x}},\mathrm{k}} = \mathbf{x}_k - \bar{\mathbf{x}}_\mathrm{k} = \mathbf{T}_k \underbrace{(\mathbf{x}_k - \hat{\mathbf{x}}_k)}_{\varepsilon_{\bar{x},k-1}} + \mathbf{B}_k \underbrace{\left(\mathbf{f}_{k-1} - \hat{\mathbf{f}}_{k-1}\right)}_{\varepsilon_{f,k-1}} + \mathbf{S}_k \underbrace{(\mathbf{w}_{k-1} - \hat{\mathbf{w}}_{k-1})}_{\varepsilon_{w,k-1}} \tag{6}$$

leads to the updated error covariance matrix

$$\boldsymbol{\Sigma}_{\overline{xx},k} = \mathbf{T}_k \boldsymbol{\Sigma}_{\hat{x}\hat{x},k-1} \mathbf{T}_k^T + \mathbf{B}_k \boldsymbol{\Sigma}_{ff,k-1} \mathbf{B}_k^T + \mathbf{S}_k \boldsymbol{\Sigma}_{ww,k-1} \mathbf{S}_k^T \tag{7}$$

The random noises $\boldsymbol{\varepsilon}_{f,k-1}$, $\boldsymbol{\varepsilon}_{w,k-1}$ and $\boldsymbol{\varepsilon}_{v,k-1} = \mathbf{z}_{k-1} - \hat{\mathbf{z}}_{k-1} = \mathbf{v}_{k-1}$ denote the input, state and measurement error which are described as stationary, uncorrelated and zero-mean white noise processes.

In the *correction step* the measurement equation is used to predict the likeliest measurement for the given prior state estimate. Once the actual measurement is obtained, the difference $\mathbf{d}_k$

$$\mathbf{d}_k = \mathbf{A}_k \bar{\mathbf{x}}_k - \mathbf{z}_k \tag{8}$$

between the predicted measurement and the actual measurement, also known as innovation or residual, is calculated. The Kalman gain matrix $\mathbf{K}_k$ is determined in order to correct the prior state estimate $\bar{\mathbf{x}}_k$ in the measurement update. It is the result of the minimization of the mean-square error of the posterior state estimate $\hat{\mathbf{x}}_k$

$$\boldsymbol{\varepsilon}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k \qquad \mathbb{E}\left[\boldsymbol{\varepsilon}_k^T \boldsymbol{\varepsilon}_k\right] \longrightarrow min. \tag{9}$$

which is equivalent to minimizing the trace of the posterior estimate covariance matrix [3]. It leads to

$$\mathbf{K}_k = \boldsymbol{\Sigma}_{\overline{xx},k} \mathbf{A}_k^T \left(\boldsymbol{\Sigma}_{vv,k} + \mathbf{A}_k \boldsymbol{\Sigma}_{\overline{xx},k} \mathbf{A}_k^T\right)^{-1} \tag{10}$$

which is used to update the prior estimate $\bar{\mathbf{x}}_k$ with measurement data and its associated posterior error covariance

$$\begin{aligned} \hat{\mathbf{x}}_k &= \bar{\mathbf{x}}_k + \mathbf{K}_k(\mathbf{z}_k - \mathbf{A}_k \bar{\mathbf{x}}_k) \\ \boldsymbol{\Sigma}_{\hat{x}\hat{x},k} &= \boldsymbol{\Sigma}_{\overline{xx},k} + \mathbf{K}_k \boldsymbol{A}^T \overline{\boldsymbol{\Sigma}}_{\overline{xx},k} \end{aligned} \tag{11}$$

2.1 Recursive Bayesian estimation

While in the standard Kalman filter the dynamic system is described by the Gaussian state space model, where $\mathbf{x}_k$ is the unobservable (hidden) state, $\mathbf{f}_{k-1}$ a given system input and $\mathbf{z}_k$ is the incoming measurement, the recursive Bayesian estimator is based on a probabilistic state space model

$$\begin{aligned} \mathbf{z}_k &\sim p(\mathbf{z}_k | \mathbf{x}_k) \\ \mathbf{x}_k &\sim p(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{f}_{k-1}) \end{aligned} \tag{12}$$

where $p(A|B)$ denotes the conditional probability density function (PDF) of the event A given that the event B occurred. Using the incoming measurements and a mathematical state space model, the recursive Bayesian estimator determines the estimates of the unknown probability density function recursively in time while the standard Kalman filter calculates recursively the true values of observations. In both cases the underlying dynamic system model can be graphically described by a dynamic Bayesian network as shown in figure. 2.

The recursive Bayesian filter is derived using two assumptions:

1. The state follows a first order hidden Markov process, i.e. the future state $\mathbf{x}_{k+1}$ is independent of the past states $\mathbf{x}_{1:k-1} = \mathbf{x}_1, \dots, \mathbf{x}_{k-1}$ given the present state $\mathbf{x}_k$ and the input $\mathbf{f}_{k-1}$

$$p(\mathbf{x}_{k+1} | \mathbf{x}_{1:k}, \mathbf{z}_{1:k}, \mathbf{f}_{1:k-1}) = p(\mathbf{x}_{k+1} | \mathbf{x}_k, \mathbf{f}_{k-1}) \tag{13}$$

2. The actual observation $\mathbf{z}_k$ is conditional independent of the previous ones if the present state $\boldsymbol{x}_\mathrm{k}$ is given

$$p(\mathbf{z}_k | \mathbf{x}_{1:k}, \mathbf{z}_{1:k-1}) = p(\mathbf{z}_k | \mathbf{x}_k) \tag{14}$$

The Bayesian estimator calculates the PDF of the posterior state at time k from the PDF of the prior estimate and the likelihood applying the Bayes' rule

$$\overbrace{p(\mathbf{x}_k | \mathbf{z}_{1:k}, \mathbf{f}_{1:k-1})}^{a\ posteriori} = \frac{\overbrace{p(\mathbf{z}_k | \mathbf{x}_k)}^{likelihood} \overbrace{p(\mathbf{x}_k | \mathbf{z}_{1:k-1}, \mathbf{f}_{1:k-1})}^{a\ priori}}{\underbrace{p(\mathbf{z}_k | \mathbf{z}_{1:k-1})}_{normalization}} \tag{15}$$

Similar to the Kalman filter the recursive Bayesian estimation is based on a *prediction step* prior to the observation and a *correction step* after obtaining the measurement data using the measurement likelihood. The underlying probabilistic state space model is given by

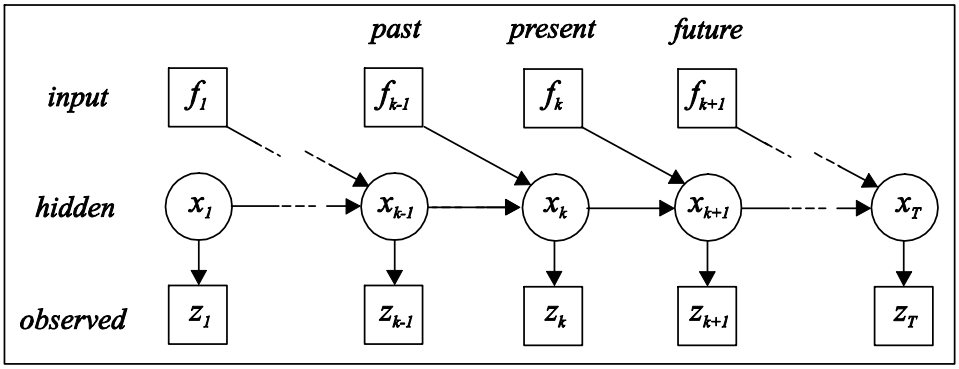

Figure 2: Dynamic Bayesian network

$$\begin{aligned} p(\mathbf{z}_k|\mathbf{x}_k) &\sim N\left(\mathbf{A}_k x_k, \boldsymbol{\Sigma}_{vv,k}\right) \\ p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{f}_{k-1}) &\sim N\left(\mathbf{T}_k \hat{\boldsymbol{x}}_{\mathrm{k}-1} + \mathbf{B}_k \mathbf{f}_{k-1}, \mathbf{S}_k \boldsymbol{\Sigma}_{ww,k-1} \mathbf{S}_k^T\right) \end{aligned} \tag{16}$$

applying the standard rules of linear transformation of random variables.
Starting at time $k-1$ the PDF of the actual state $\mathbf{x}_{k-1}$ is given by the posterior distribution of the previous time step

$$p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1}, \mathbf{f}_{1:k-2}) \sim N(\hat{\mathbf{x}}_{k-1}, \boldsymbol{\Sigma}_{\hat{x}\hat{x},k-1}) \tag{17}$$

which is assumed to be Gaussian with expectation $\hat{\mathbf{x}}_{k-1}$ and error covariance matrix $\boldsymbol{\Sigma}_{\hat{x}\hat{x},k-1}$.
In the *prediction step* first the joint density of $(\mathbf{x}_k,\ \mathbf{x}_{k-1})$ conditional on the input and the observation up to time $k-1$ is calculated using the PDF given in (16) and the posterior distribution from the previous time step

$$p(\mathbf{x}_k, \mathbf{x}_{k-1}|\mathbf{z}_{1:k-1}, \mathbf{f}_{1:k-1}) = p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{f}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1}, \mathbf{f}_{1:k-2}) \tag{18}$$

from which the prior distribution is obtained by marginalization, i.e. integration over $\mathbf{x}_{k-1}$

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}, \mathbf{f}_{1:k-1}) \sim N(\bar{\mathbf{x}}_k, \boldsymbol{\Sigma}_{\overline{xx},k}) \tag{19}$$

where $\bar{\mathbf{x}}_k, \boldsymbol{\Sigma}_{\overline{xx},k}$ denote the expectation and covariance of the predicted state as given in (5), (7).

In the *correction step* the likelihood $p(\mathbf{z}_k|\mathbf{x}_k)$ incorporates the new measurement into the posterior PDF. Once the actual measurement $\mathbf{z}_k$ is obtained, the residual $\mathbf{d}_k$

$$\mathbf{d}_k = \mathbf{z}_k - \hat{\mathbf{z}}_k = \mathbf{z}_k - \mathbf{A}_k \bar{\mathbf{x}}_k \tag{20}$$

between the predicted and the incoming measurement is calculated using the conditional expectation $\bar{\mathbf{x}}_k$ of the priori PDF. The latter as well as the transition matrix $\mathbf{A}_k$ are known, i.e. observing the error $\mathbf{d}_k$ and observing $\mathbf{z}_k$ are equally likely [6]. Using the observational equation (1) the likelihood is given by

$$p(\mathbf{z}_k|\mathbf{x}_k) \sim N\left(\mathbf{A}_k \bar{\mathbf{x}}_k, \boldsymbol{\Sigma}_{vv,k}\right) \tag{21}$$

In order to obtain the posterior distribution the normalization factor $p(\mathbf{z}_k|\mathbf{z}_{1:k-1})$ is computed by marginalization of

$$p(\mathbf{z}_k, \mathbf{x}_k|\mathbf{z}_{1:k-1}) = p(\mathbf{z}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{z}_{1:k-1}, \mathbf{f}_{1:k-1}) \tag{22}$$

over $\mathbf{x}_k$ and it follows

$$p(\mathbf{z}_k|\mathbf{z}_{1:k-1}) \sim N(\mathbf{A}_k \bar{\mathbf{x}}_k, \mathbf{A}_k \boldsymbol{\Sigma}_{\overline{xx},k} \mathbf{A}_k^T + \boldsymbol{\Sigma}_{vv,k}) \tag{23}$$

The posterior density is now computed applying the Bayes' rule leading to

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}, \mathbf{f}_{1:k-1}) \sim \mathrm{N}\left(\boldsymbol{\mu}_{x|z}, \boldsymbol{\Sigma}_{x|z}\right)$$

where

$$\begin{aligned} \boldsymbol{\mu}_{x|z} &= \bar{\mathbf{x}}_k + \mathbf{K}_k(\mathbf{z}_k - \mathbf{A}_k \bar{\mathbf{x}}_k) && \equiv \hat{\mathbf{x}}_k \\ \boldsymbol{\Sigma}_{x|z} &= \boldsymbol{\Sigma}_{vv,k} \boldsymbol{\Sigma}_{\overline{xx},k} \left(\mathbf{A}_k \boldsymbol{\Sigma}_{\overline{xx},k} \mathbf{A}_k^T + \boldsymbol{\Sigma}_{vv,k}\right)^{-1} && \equiv \boldsymbol{\Sigma}_{\hat{x}\hat{x},k} \\ \mathbf{K}_k &= \boldsymbol{\Sigma}_{\overline{xx},k} \mathbf{A}_k^T \left(\mathbf{A}_k \boldsymbol{\Sigma}_{\overline{xx},k} \mathbf{A}_k^T + \boldsymbol{\Sigma}_{vv,k}\right)^{-1} && \equiv \text{Kalman gain matrix} \end{aligned} \tag{24}$$

which is similar to the result (11) of the optimal state estimate given by the standard Kalman filter. If the assumption of Gaussianity holds the optimal estimate of $\mathbf{x}_k$ conditional on all the data $\mathbf{z}_{1:k-1}$ is given by

the expectation $\boldsymbol{\mu}_{x|z}$ of the posterior PDF of $p(\mathbf{x}_k|\mathbf{z}_{1:k}, \mathbf{f}_{1:k-1})$. For other types of statistics, e.g. skewed, multimodal PDFs or in the nonlinear case, this is not necessarily the case [7].

2.2 Parameter identification using the Extended Kalman filter

If the assumption of Gaussianity holds the Kalman filter is an optimal, unbiased minimum variance estimator of linear state-space models [1][2]. However, in practice identification problems of unknown parameters are in general nonlinear. Hence, the state has to be extended to include the model parameters to be identified [3][4]. The new extended state is defined as

$$\mathbf{x}_{ext,k} = \begin{bmatrix} \mathbf{x}_k \\ \mathbf{p}_k \end{bmatrix} \tag{25}$$

where $\mathbf{p}_k$ is the p-dimensional vector of the unknown parameters. The time update of the parameter vector is modeled as random walk process by adding a zero-mean Gaussian noise

$$\begin{aligned} \mathbf{p}_k &= \mathbf{p}_{k-1} + \mathbf{w}_{p,k} \\ \mathbf{w}_{p,k} &\sim N(\mathbf{0},\ \boldsymbol{\Sigma}_{\mathrm{ww,k}}) \end{aligned} \tag{26}$$

The extended state space system model is now given by

$$\begin{aligned} \mathbf{x}_{ext,k} &= \begin{bmatrix} \mathbf{T}_k(\mathbf{p}_{k-1} + \mathbf{w}_{p,k}) & \mathbf{0}_{m\times p} \\ \mathbf{0}_{p\times m} & \mathbf{I}_{p\times p} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{k-1} \\ \mathbf{p}_{k-1} \end{bmatrix} + \begin{bmatrix} \mathbf{B}_k(\mathbf{p}_{k-1} + \mathbf{w}_{p,k}) \\ \mathbf{0}_{p\times f} \end{bmatrix} \mathbf{f}_{k-1} + \begin{bmatrix} \mathbf{S}_k(\mathbf{p}_{k-1} + \mathbf{w}_{p,k}) & \mathbf{0}_{m\times p} \\ \mathbf{0}_{p\times f} & \mathbf{I}_{p\times p} \end{bmatrix} \begin{bmatrix} \mathbf{w}_k \\ \mathbf{w}_{p,k} \end{bmatrix} \\ \mathbf{z}_{ext,k} &= [\mathbf{A}_k \quad \mathbf{A}_{p,k}] \begin{bmatrix} \mathbf{x}_k \\ \mathbf{p}_k \end{bmatrix} + \mathbf{v}_k \end{aligned} \tag{27}$$

where $\mathbf{I}_{p\times p}$ denotes the $p \times p$ identity matrix and $\mathbf{0}_{a\times b}$ an $a \times b$ zero matrix. The time-invariant matrices $\mathbf{T}_k, \mathbf{B}_k$ and $\mathbf{S}_k$ are now given as functions in nonlinear dependency on the predicted state estimates $\mathbf{p}_k$. Assuming that the parameters are not directly observable, it follws $\mathbf{A}_{p,k} = \mathbf{0}_{n\times p}$.
Hence, the state space model can be rewritten as

$$\begin{aligned} \mathbf{z}_{ext,k} &= \mathbf{A}_{ext,k}\mathbf{x}_{ext,k} + \mathbf{v}_k \\ \mathbf{x}_{ext,k} &= f(\mathbf{x}_{k-1}, \mathbf{p}_{k-1}, \mathbf{f}_{k-1}, \mathbf{w}_{k-1}, \mathbf{w}_{p,k-1}) \end{aligned} \tag{28}$$

where $f(\cdot)$ is the nonlinear system equation of the extended state vector to be identified. In order to use the standard Kalman filter algorithm the system function is linearized by applying a first order Taylor expansion near the current state estimate

$$\mathbf{x}_{0,k-1} = [\mathbf{x}_{k-1} \ \ \mathbf{p}_{k-1} \ \ \mathbf{f}_{k-1} \ \ \mathbf{w}_{k-1} \ \ \mathbf{w}_{p,k-1}]^T \approx \left[\hat{\mathbf{x}}_{k-1} \ \ \hat{\mathbf{p}}_{k-1} \ \ \hat{\mathbf{f}}_{k-1} \ \ \hat{\mathbf{w}}_{k-1} \ \ \hat{\mathbf{w}}_{p,k-1}\right]^T = \hat{\mathbf{x}}_{0,k-1} \tag{29}$$

This leads to the following linear approximated model of the extended state vector

$$\mathbf{x}_{ext,k} \approx f(\hat{\mathbf{x}}_{0,k-1}) + \mathbf{T}_{ext,k} \underbrace{\begin{bmatrix} \mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1} \\ \mathbf{p}_{k-1} - \hat{\mathbf{p}}_{k-1} \end{bmatrix}}_{\varepsilon_{ext,\hat{x},k-1}} + \mathbf{B}_{ext,k} \underbrace{(\mathbf{f}_{k-1} - \hat{\mathbf{f}}_{k-1})}_{\varepsilon_{ext,f,k-1}} + \mathbf{S}_{ext,k} \underbrace{\begin{bmatrix} \mathbf{w}_{k-1} - \hat{\mathbf{w}}_{k-1} \\ \mathbf{w}_{k-1} - \hat{\mathbf{w}}_{p,k-1} \end{bmatrix}}_{\varepsilon_{ext,w,k-1}} \tag{30}$$

where $\mathbf{T}_{ext,k}$, $\mathbf{B}_{ext,k}$ and $\mathbf{S}_{ext,k}$ indicate the linearized transfer matrices. The derivatives with respect to the parameter vector $\mathbf{p}_{k-1}$ and the associated noise vector $\mathbf{w}_{p,k-1}$ are expressed using the $m \times m$ Jacobian matrix $\mathcal{J}_{f,k}(\cdot)$ of the nonlinear function $f(\cdot)$. Hence, the extended matrices are time-variant and have to be calculated at each time step. It follows

$$\begin{aligned} \mathbf{T}_{ext,k} &= \begin{bmatrix} \mathbf{T}_k(\mathbf{p}_{k-1}) & \mathcal{J}_{f,k}(p_{k-1}) \\ \mathbf{0}_{p\times m} & \mathbf{I}_{p\times p} \end{bmatrix} && \text{with} \quad \mathcal{J}_{f,k}(p_{k-1}) = \left.\frac{\partial f(x_{0,k-1})}{\partial p_{k-1}}\right|_{\hat{x}_{0i,k-1}} \\ \mathbf{S}_{ext,k} &= \begin{bmatrix} \mathbf{S}_k(\hat{\mathbf{p}}_{k-1}) & \mathcal{J}_{f,k}(w_{p,k-1}) \\ \mathbf{0}_{p\times m} & \mathbf{I}_{p\times p} \end{bmatrix} && \text{with} \quad \mathcal{J}_{f,k}(w_{p,k-1}) = \left.\frac{\partial f(x_{0,k-1})}{\partial w_{p,k-1}}\right|_{\hat{x}_{0i,k-1}} \\ \mathbf{B}_{ext,k} &= \begin{bmatrix} \mathbf{B}_k(\hat{\mathbf{p}}_{k-1}) \\ \mathbf{0}_{p\times f} \end{bmatrix} \end{aligned} \tag{31}$$

where $f(\hat{\mathbf{x}}_{0,k-1})$ denotes the prior extended state estimate defined by

$$f(\hat{\mathbf{x}}_{0,k-1}) = \mathbf{T}_k(\hat{\mathbf{p}}_{k-1})\hat{\mathbf{x}}_{k-1} + \mathbf{B}_k(\hat{\mathbf{p}}_{k-1})\hat{\mathbf{f}}_k = \bar{\mathbf{x}}_{ext,k} \tag{32}$$

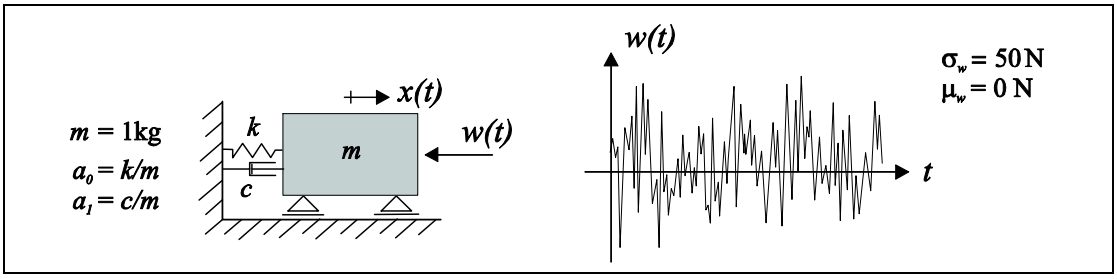

Figure 3: Damped single degree of freedom system excited by a white noise process

The approximated prediction error of the extended state estimate is now given by

$$\hat{\boldsymbol{\varepsilon}}_{ext,\bar{x},k} = \mathbf{x}_{ext,k} - \bar{\mathbf{x}}_{ext,k} \approx \mathbf{T}_{ext,k}\boldsymbol{\varepsilon}_{ext,\hat{x},k-1} + \mathbf{B}_{ext,k}\boldsymbol{\varepsilon}_{ext,f,k-1} + \mathbf{S}_{ext,k}\boldsymbol{\varepsilon}_{ext,w,k-1} \tag{33}$$

Using the extended vectors and matrices the minimization of the mean-square error of the extended posterior state estimate $\boldsymbol{\varepsilon}_{ext,k} = \mathbf{x}_{ext,k} - \hat{\mathbf{x}}_{ext,k}$ leads to the extended Kalman gain matrix. The above described standard Kalman filter algorithm and its Bayesian approach can be applied to the linearized model.

In contrast to the linear problem the estimate given by the extended Kalman filter is biased as in general the assumption

$$\mathbb{E}[f(x_k)] = f(\mathbb{E}[x_k]) \tag{34}$$

is violated. Moreover the filter approximates the posterior distribution of the estimate always as Gaussian which leads to poor results if the true posterior distribution is for instance heavily tailed or multimodal [1].

2.3 Numerical Example

In order to illustrate the introduced method the extended Kalman filter is used to identify the stiffness and damping parameters of a single degree of freedom system excited by an ambient load as shown in figure 3 using noisy measurement data of the displacement x_k.

The equation of motion is given by

$$m\ddot{x} + c\dot{x} + kx = w_f(t) \tag{35}$$

where m, c, k denote the mass, damping and stiffness parameter of the system, respectively. Assuming that only the system response is measureable, the unknown system input $f(t)$ is regarded as system error $\mathrm{w}_f(t)$ which is described as zero-mean white noise process. Using the state vector $\mathbf{x}(t) = [x(t), \dot{x}(t)]^T$ the eq. (35) is transformed to the state-space form

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & 1 \\ -k/m & -c/m \end{bmatrix}\mathbf{x}(t) + \begin{bmatrix} 0 \\ 1/m \end{bmatrix} w_f(t) \tag{36}$$

Assuming that the external load is constant within a time interval $\Delta t = t_k - t_{k-1}$ eq. (36) is discretized to a difference equation and one obtained the linear system equation

$$\mathbf{x}_\mathrm{k} = \mathbf{T}_k\mathbf{x}_{k-1} + \mathbf{S}_k w_{f,k-1} \quad \text{where} \quad \mathbf{x_k} = \mathbf{x}(k\Delta t) \tag{37}$$

The transfer matrix $\mathbf{T}_k$ gives the relation between the system response at time k for a given displacement and velocity at time $k-1$. Hence, it can be derived from the homogenous solution of the equation of motion (35) where the state $\mathbf{x}_{k-1}$ gives the initial conditions. It follows

$$\mathbf{T}_k = e^{-\delta\Delta t}\begin{bmatrix} \frac{\delta}{\omega_d}\sin(\omega_d\Delta t) + \cos(\omega_d\Delta t) & \frac{1}{\omega_d}\sin(\omega_d\Delta t) \\ -\frac{\omega_0^2}{\omega_d}\sin(\omega_d\Delta t) & \cos(\omega_d\Delta t) - \frac{\delta}{\omega_d}\sin(\omega_d\Delta t) \end{bmatrix} \tag{38}$$

where $\omega_0^2 = k/m$, $\delta = c/2m$, $\omega_d = \sqrt{\omega_0^2 - \delta^2}$.

In a similar way the input matrix $\mathbf{S}_k$ is obtained from the steady state solution for transient loading. Assuming that the force is constant within the time interval Δt, the system response is calculated from the Duhamel integral and it follows

$$\mathbf{S}_k = \int_{t_{k-1}}^{t_k} \mathbf{T}_k(t_k - \tau)\,\mathbf{G}d\tau \;\text{ with }\; \mathbf{G}^T = [0 \quad 1/m]^T \tag{39}$$

error		initial values	assumption
$\hat{\varepsilon}_z$	$\mathrm{N}(0, \sigma_z)$	$\sigma_z = 5$ mm	5 % of the static displacement due to a load of 100 N (97th percentile)
$\hat{\varepsilon}_{\hat{x}}$	$\mathrm{N}(0, \Sigma_{\hat{x}\hat{x}})$	$\sigma_x = 0.5$ m $\sigma_{\dot{x}} = 1$ m/s $\sigma_{a_0} = 1500\ \mathrm{s}^{-2}$ $\sigma_{a_1} = 3.0\ \mathrm{s}^{-2}$	Arbitrarily chosen
$\hat{\varepsilon}_{w_f}$	$\mathrm{N}(0, \sigma_u)$	$\sigma_{w_f} = 50$ N	Estimated deviation of the white noise process
$\hat{\varepsilon}_{w_p}$	$\mathrm{N}(0, \Sigma_{w_p})$	$\Sigma_{w_p} = 0_{2\times 2}$	

Table 1: Stochastic error model

	true values	initial values	initial error	identified parameters	identification error
$a_0[\mathrm{s}^{-2}]$	1000	1500	50 %	1008	0,8 %
$a_1\ [\mathrm{s}^{-1}]$	1	1.5	50 %	1,06	6 %

Table 2: Initial values and identification result after $\mathrm{t} = 150$ s

As described in section 2.2 the state vector of the extended Kalman filter has to be extended by the parameters to be estimated. While considering the mass parameter m to be known, the system response is mainly depending on the parameters

$$a_0 = \frac{k}{m} \quad \text{and} \quad a_1 = \frac{c}{m} \tag{40}$$

and the extended state vector is given as

$$\mathbf{x}_{ext,k}^T = [x_k \quad \dot{x}_k \quad a_{0,k} \quad a_{1,k}]^{\mathrm{T}} \tag{41}$$

The transfer matrices $\mathbf{T}_k$, and $\mathbf{S}_k$ of the system model are now nonlinear functions of the unknown parameters $a_{0,k}$ and $a_{1,k}$. The nonlinear system equations are approximated by a Taylor series around the prior estimates $\hat{a}_{0,k-1}$ and $\hat{a}_{1,k-1}$ which leads to a linear state space model. Consequently the extended matrices $\mathbf{T}_{ext,k}$ and $\mathbf{S}_{ext,k}$ are time variant and have to be calculated in each time step.

The parameter identification is based on the measured displacements of the system. Hence, the observational equation (28) reduces to

$$\mathbf{z}_{ext,k} = \mathbf{x}_{k-1} + \mathbf{v}_k \tag{42}$$

As the unknown parameters a_0, a_1 cannot be measured themselves, the identification will just succeed as long as there exists a dependency between the observable quantities and the unknown parameters. In this example the dependency is given by the Jacobian matrices included in the extended transfer matrices.

In order to start the filter algorithm the initial parameters and the initial stochastic model based on the error covariances have to be defined. The chosen values are shown in table 1 and 2. The measurements were simulated based on the analytical solution of the given example excited by a simulated white noise process as input force. Hereby a random measurement error of 5 % of the static displacement due to a load of 100 N (97th percentile of the load). The errors in the estimated parameters $\hat{\boldsymbol{\varepsilon}}_{w_p}$ were set to zero, as there's no need to disturb the identification result.

The simulated measurement data, the estimation of the displacement and velocity and the results of the parameter identification are shown in figures 4 and 5, respectively. The initial model parameters where chosen taking into account an initial error of 50% of the true model parameters. Table 2 shows the initial values and the identified parameters after $\mathrm{t} = 150$ s. As just a small damping ratio of $\mathrm{D} = 0.5\%$ is considered the response is mainly depending on the stiffness parameter a_0. Hence, the difference between the observed and the estimated displacement, e.g. the prediction error of the state, is dominated by the error of the estimated stiffness. Consequently the identification of the parameter a_0 requires significant less iterations as can be seen from figure 5.

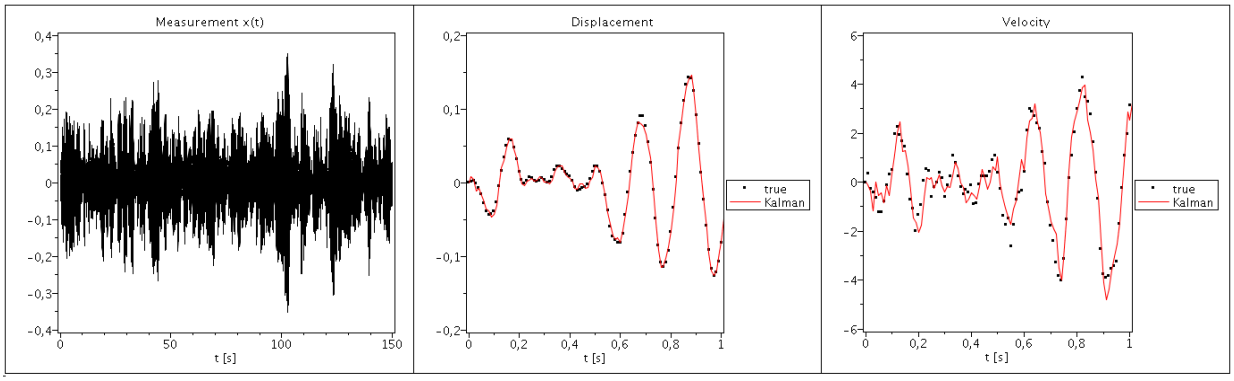

Figure 4: Simulated measurement data of the displacement; estimated/true value of the displacement and velocity

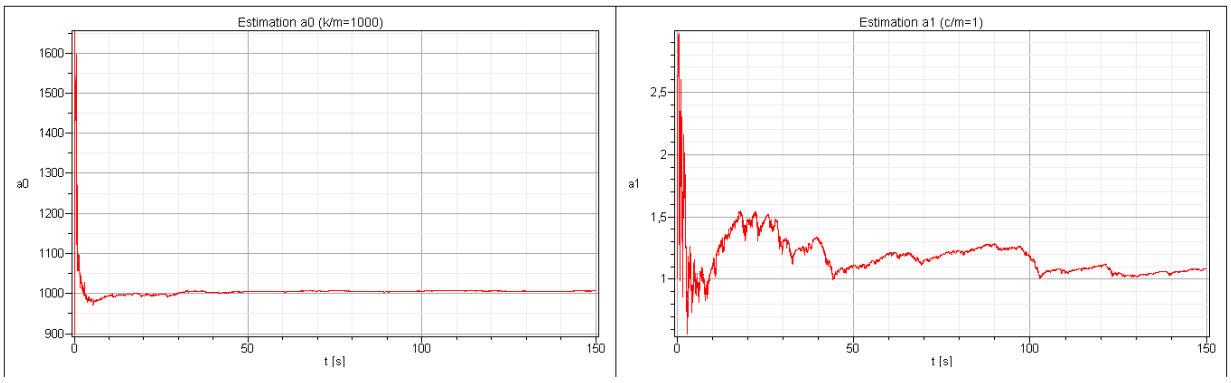

Figure 5: Identified values of the stiffness and damping parameters a_0 and a_1 after $\mathrm{t} = 150$ s

As the prediction error, on which the time update of the parameters depends, decreases quickly the damping parameter a_1 converge slowly to the true value. Hence, the identification of the damping parameter requires more iterations.

3 Skewed Kalman filter

3.1 Closed skew-normal distribution

The skewed Kalman filter is a modification of the standard Kalman algorithm in order to introduce skewness to the state space model. It is based on the so called closed skew-normal distribution (CSN) which allows to model skewness while preserving the advantageous properties of the Gaussian distribution as the closure under conditioning, linear transformations and marginalization [2][8][9]. This allows to derive a filter algorithm where the posterior density takes the prior form, so that the recursion of the Bayesian estimation reduces to an algebraic recursion operation on covariance, mean and - in addition to the standard Kalman filter - on skewing parameters.

The closed (multivariate) skew-normal density function of a n-dimensional vector $\mathbf{X}$ was introduced by González-Farías, Domínguez-Molina and Gupta [9] and is defined as follows:

Definition: For $m \geq 1$, $n \geq 1$, $\boldsymbol{\mu} \in \mathbb{R}^n$, $\boldsymbol{\nu} \in \mathbb{R}^m$, an arbitrary matrix $\mathbf{D} \in \mathbb{R}^{m \times n}$ and positive definite covariance matrices $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ and $\boldsymbol{\Delta} \in \mathbb{R}^{m \times m}$ the CSN is given by

$$\mathbf{X} \sim CSN_{n,m}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{D}, \boldsymbol{\nu}, \boldsymbol{\Delta})$$

$$p_{n,m}(\mathbf{x}) = C^{-1}\phi_n(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi_m(\mathbf{D}(\mathbf{x} - \boldsymbol{\mu});\ \boldsymbol{\nu}, \boldsymbol{\Delta})$$

with

$$C = \Phi_m(\mathbf{0};\ \boldsymbol{\nu}, \boldsymbol{\Delta} + \mathbf{D}\boldsymbol{\Sigma}\mathbf{D}^T) \quad \text{and} \quad \mathbf{X} \in \mathbb{R}^n \tag{43}$$

where $\phi_n(*\,; \boldsymbol{\eta}, \boldsymbol{\Omega})$ and $\Phi_n(*\,; \boldsymbol{\eta}, \boldsymbol{\Omega})$ are the probability density function and the cumulative distribution function (CDF) of a n-dimensional normal distribution with mean vector $\boldsymbol{\eta} \in \mathbb{R}^n$ and covariance matrix

$\boldsymbol{\Omega} \in \mathbb{R}^{n\times n}$ [7]. The matrix $\mathbf{D}$ regulates the skewness of the distribution and allows to vary continuously from the normal PDF ($\mathbf{D} = \mathbf{0}$) to a half normal distribution whereas the constraints $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are location and scale parameters. The remaining parameters secure the closure properties of the CSN: the parameter $\boldsymbol{\nu}$ ensure the closure under conditioning, the parameter $\boldsymbol{\Delta}$ ensure the closure under marginalization and the parameter C allows the closure under summation of independent CSN random variables [8]. In the following some important properties of the skewed Kalman filter are summarized which allows to implement a recursive filter procedure comparable to the standard Kalman filter. The proofs of the following properties are given in the appendix.

Property A - Closure under linear transformation
Let $\mathbf{X} \sim CSN_{n,m}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{D}, \boldsymbol{\nu}, \boldsymbol{\Delta})$ and let $\mathbf{A}$ be a $r \times n$ matrix of rank n and $r < n$, then

$$\mathbf{AX} \sim CSN_{r,m}(\boldsymbol{\mu}_\mathrm{A}, \boldsymbol{\Sigma}_\mathrm{A}, \mathbf{D}_\mathrm{A}, \boldsymbol{\nu}, \boldsymbol{\Delta}_\mathrm{A})$$

where

$$\boldsymbol{\mu}_\mathrm{A} = \mathbf{A}\boldsymbol{\mu} \qquad \boldsymbol{\Sigma}_\mathrm{A} = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\mathrm{T} \qquad \mathbf{D}_\mathrm{A} = \mathbf{D}\boldsymbol{\Sigma}\mathbf{A}^\mathrm{T}\boldsymbol{\Sigma}_\mathrm{A}^{-1} \qquad \boldsymbol{\Delta}_\mathrm{A} = \boldsymbol{\Delta} + (\mathbf{D} - \mathbf{D}_\mathrm{A}\mathbf{A})\boldsymbol{\Sigma}\mathbf{D}^\mathrm{T} \tag{44}$$

As long as the conditions $\Sigma > 0$ and $rank(\mathbf{A}) = n$ holds $\boldsymbol{\Sigma}_\mathrm{A}$ is a non-singular matrix [8]. When $\mathbf{A}$ is a $n \times n$ non-singular matrix, then $\mathbf{D}_\mathrm{A}$, $\boldsymbol{\Delta}_\mathrm{A}$ reduces to $\mathbf{D}_\mathrm{A} = \mathbf{D}\mathbf{A}^{-1}$ and $\boldsymbol{\Delta}_\mathrm{A} = \boldsymbol{\Delta}$.

Property B - Closure under marginalization
Let $\mathbf{X} \sim CSN_{n,m}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{D}, \boldsymbol{\nu}, \boldsymbol{\Delta})$ which is partitioned as $\mathbf{X} = [\mathbf{x}_1^T \;\; \mathbf{x}_2^T]^T$ where $\mathbf{x}_1 \in \mathbb{R}^k$ and $\mathbf{x}_2 \in \mathbb{R}^{n-k}$ then

$$\mathbf{x}_1 \sim CSN_{n,m}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \mathbf{D}^*, \boldsymbol{\nu}, \boldsymbol{\Delta}^*)$$

with

$$\mathbf{D}^* = \mathbf{D}_1 + \mathbf{D}_2\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11} \text{ and } \boldsymbol{\Delta}^* = \boldsymbol{\Delta} + \mathbf{D}_2(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})\mathbf{D}_2^T$$

$$\boldsymbol{\mu} = \begin{bmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_1\end{bmatrix}\begin{matrix}k\\ n-k\end{matrix} \qquad \boldsymbol{\Sigma} = \begin{array}{cc} k & n-k \\ \end{array}\begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}\begin{matrix}k\\ n-k\end{matrix} \qquad \mathbf{D} = \begin{array}{cc} k & n-k \\ \end{array}[\mathbf{D}_1 \;\; \mathbf{D}_2]\;\; m \tag{45}$$

Property C - Closer under conditioning
Let $\mathbf{X} \sim CSN_{n,m}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{D}, \boldsymbol{\nu}, \boldsymbol{\Delta})$ which is partitioned as $\mathbf{X} = [\mathbf{x}_1^T \;\; \mathbf{x}_2^T]^T$ where $\mathbf{x}_1 \in \mathbb{R}^k$ and $\mathbf{x}_2 \in \mathbb{R}^{n-k}$, then the density of $\mathbf{x}_2$ conditional on $\mathbf{x}_1 = \mathbf{x}_{10}$ is derived using the result of *property B* and the Bayes' rule

$$p(\mathbf{x}_2|\mathbf{x}_1) = \frac{p(\mathbf{x}_1, \mathbf{x}_2)}{p(\mathbf{x}_1)}$$

$$\Rightarrow \quad \mathbf{x}_2|\mathbf{x}_1 \sim CSN_{n,m}(\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}_{10} - \boldsymbol{\mu}_1),\ \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\Sigma_{12}, \mathbf{D}_2, \boldsymbol{\nu} - \mathbf{D}^*(\mathbf{x}_{10} - \boldsymbol{\mu}_1), \boldsymbol{\Delta}) \tag{46}$$

Property D – Closure under summation
Let $\mathbf{X}_n \sim CSN_{n,m}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x, \mathbf{D}, \boldsymbol{\nu}, \boldsymbol{\Delta})$ and $\mathbf{Y}_n \sim N(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_\mathrm{y})$, then the sum $\mathbf{Z}_n = \mathbf{X}_n + \mathbf{Y}_n$ follows a closed skew-normal distribution

$$\mathbf{Z}_n \sim CSN_{n,m}(\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_\mathrm{z}, \mathbf{D}_z, \boldsymbol{\nu}_z, \boldsymbol{\Delta}_\mathrm{z})$$

where $\boldsymbol{\mu}_z = \boldsymbol{\mu}_x + \boldsymbol{\mu}_y \quad \boldsymbol{\Sigma}_\mathrm{z} = \boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_y \quad \mathbf{D}_z = \mathbf{D}\boldsymbol{\Sigma}_x\boldsymbol{\Sigma}_z^{-1} \quad \boldsymbol{\nu}_z = \boldsymbol{\nu} \quad \boldsymbol{\Delta}_z = \boldsymbol{\Delta} + (\mathbf{D} - \mathbf{D}_z)\boldsymbol{\Sigma}_x\mathbf{D}^T$ (47)

As it is shown in [9] the closer under summation of p independent random variable with CSN of dimension (n, m) leads to another CSN of dimension (n, pm). Hence, in the recursive procedure of the Kalman filter this would lead to a rapid increase of the dimensions of the skewness parameters $\mathbf{D}$, $\boldsymbol{\nu}$ and $\boldsymbol{\Delta}$ at each time step. Due to *property D* this dimensional problem can be avoided by choosing a Gaussian error model where the system error as well as the measurement error are modeled as white noise [2].

3.2 Bayesian approach to the skewed Kalman filter

A linear state space model as given in (1), (2) with n-dimensional vector of observations $\mathbf{z}_k$ and m-dimensional state vector $\mathbf{x}_k$ is assumed where both the observation as well as the system equation are subjected to zero-mean Gaussian noise. Hence the probabilistic state space model is given by

$$\begin{aligned} p(\mathbf{z}_k|\mathbf{x}_k) &\sim N\big(\mathbf{A}_k\bar{\mathbf{x}}_k, \boldsymbol{\Sigma}_{vv,k}\big) \\ p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{f}_{k-1}) &\sim N\big(\mathbf{T}_k\mathbf{x}_{k-1} + \mathbf{B}_k\mathbf{f}_{k-1}, \mathbf{S}_k\boldsymbol{\Sigma}_{ww,k-1}\mathbf{S}_k^T\big) \end{aligned} \tag{48}$$

PREDICTION (TIME UPDATE)	CORRECTION (MEASUREMENT UPDATE)
Project the state ahead (*a priori estimate*) $\mathbf{x}_k\|\mathbf{z}_{1:k-1},\mathbf{u}_{1:k-1}\sim CSN_{m,q}\left(\bar{\boldsymbol{\mu}}_{x,k},\bar{\boldsymbol{\Sigma}}_{xx,k},\bar{\mathbf{D}}_{x,k},\bar{\boldsymbol{\nu}}_{x,k},\bar{\boldsymbol{\Delta}}_{x,k}\right)$ $\bar{\boldsymbol{\mu}}_{x,k}=\mathbf{T}_k\mu_{x,k-1}+\mathbf{B}_k f_{k-1}$ $\bar{\boldsymbol{\Sigma}}_{xx,k}=\mathbf{T}_k\boldsymbol{\Sigma}_{xx,k-1}\mathbf{T}_k^T+\mathbf{B}_k\boldsymbol{\Sigma}_{ff,k-1}\mathbf{B}_k^T+\mathbf{S}_k\boldsymbol{\Sigma}_{ww,k-1}\mathbf{S}_k^T$ $\bar{\mathbf{D}}_{x,k}=\mathbf{D}_{x,k-1}\boldsymbol{\Sigma}_{xx,k-1}\mathbf{T}_k^T\bar{\boldsymbol{\Sigma}}_{xx,k}^{-1}$ $\bar{\boldsymbol{\nu}}_{x,k}=\boldsymbol{\nu}_{x,k-1}$ $\bar{\boldsymbol{\Delta}}_{x,k}=\boldsymbol{\Delta}_{x,k-1}+\left(\mathbf{D}_{x,k-1}-\bar{\mathbf{D}}_{x,k}\mathbf{T}_k\right)\boldsymbol{\Sigma}_{xx,k-1}\mathbf{D}_{x,k-1}^T$	1) Compute the Kalman gain matrix $\mathbf{K}_k=\bar{\boldsymbol{\Sigma}}_{xx,k}\mathbf{A}_k^T\left(\boldsymbol{\Sigma}_{vv,k}+\mathbf{A}_k\bar{\boldsymbol{\Sigma}}_{xx,k}\mathbf{A}_k^T\right)^{-1}$ 2) Update estimate with measurement (*a posteriori estimate*) $\mathbf{x}_k\|\mathbf{z}_{1:k},\mathbf{f}_{1:k-1}\sim CSN_{m,q}\left(\boldsymbol{\mu}_{x,k},\boldsymbol{\Sigma}_{xx,k},\mathbf{D}_{x,k},\boldsymbol{\nu}_{x,k},\boldsymbol{\Delta}_{x,k}\right)$ $\boldsymbol{\mu}_{x,k}=\bar{\boldsymbol{\mu}}_{x,k}+\mathbf{K}_k\left(\mathbf{z}_k-\mathbf{A}_k\bar{\boldsymbol{\mu}}_{x,k}\right)$ $\boldsymbol{\Sigma}_{xx,k}=\bar{\boldsymbol{\Sigma}}_{xx,k}-\mathbf{K}_k\mathbf{A}_k^T\bar{\boldsymbol{\Sigma}}_{xx,k}$ $\mathbf{D}_{x,k}=\mathbf{D}_{x,k-1}\boldsymbol{\Sigma}_{xx,k-1}\mathbf{T}_k^T\bar{\boldsymbol{\Sigma}}_{xx,k}^{-1}$ $\boldsymbol{\nu}_{x,k}=\boldsymbol{\nu}_{x,k-1}-\mathbf{D}_{x,k}\mathbf{K}_k\left(z_k-\mathbf{A}_k\bar{\mu}_{x,k}\right)$ $\boldsymbol{\Delta}_{x,k}=\boldsymbol{\Delta}_{x,k-1}+\left(\mathbf{D}_{x,k-1}-\mathbf{D}_{x,k}\mathbf{T}_k\right)\boldsymbol{\Sigma}_{xx,k-1}\mathbf{D}_{x,k-1}^T$

Figure 6: Prediction-correction procedure of the skewed Kalman filter

In contrast to the standard Kalman filter the initial state vector $\mathbf{x}_k|\mathbf{z}_{1:k},\mathbf{f}_{1:k-1}$ is assumed to follow a CSN. The skewness is implemented by assuming that the initial state vector $\mathbf{x}_0$ follows a skew-normal distribution.

Using $\mathbf{x}_k|\mathbf{z}_{1:k-1},\mathbf{f}_{1:k-1}\sim CSN_{m,q}\left(\boldsymbol{\mu}_{x,k-1},\boldsymbol{\Sigma}_{xx,k-1},\mathbf{D}_{x,k-1},\boldsymbol{\nu}_{x,k-1},\boldsymbol{\Delta}_{x,k-1}\right)$ the prior distribution is a direct consequence of *property A* and *D* and it follows

$$\mathbf{x}_k|\mathbf{z}_{1:k-1},\mathbf{f}_{1:k-1}\sim CSN_{n,q}\left(\bar{\boldsymbol{\mu}}_{x,k},\bar{\boldsymbol{\Sigma}}_{xx,k},\bar{\mathbf{D}}_{x,k},\bar{\boldsymbol{\nu}}_{x,k},\bar{\boldsymbol{\Delta}}_{x,k}\right)$$

where

$$\begin{aligned}
&\bar{\boldsymbol{\mu}}_{x,k}=\mathbf{T}_k\boldsymbol{\mu}_{x,k-1}+\mathbf{B}_k\mathbf{f}_{k-1} && \bar{\mathbf{D}}_{x,k}=\mathbf{D}_{x,k-1}\boldsymbol{\Sigma}_{xx,k-1}\mathbf{T}_k^T\bar{\boldsymbol{\Sigma}}_{xx,k}^{-1}\\
&\bar{\boldsymbol{\Sigma}}_{xx,k}=\mathbf{T}_k\boldsymbol{\Sigma}_{xx,k-1}\mathbf{T}_k^T+\mathbf{B}_k\boldsymbol{\Sigma}_{ff,k-1}\mathbf{B}_k^T+\mathbf{S}_k\boldsymbol{\Sigma}_{ww,k-1}\mathbf{S}_k^T && \bar{\boldsymbol{\nu}}_{x,k}=\boldsymbol{\nu}_{x,k-1}\\
&\bar{\boldsymbol{\Delta}}_{x,k}=\boldsymbol{\Delta}_{x,k-1}+\left(\mathbf{D}_{x,k-1}-\bar{\mathbf{D}}_{x,k}\mathbf{T}_k\right)\boldsymbol{\Sigma}_{xx,k-1}\mathbf{D}_{x,k-1}^T
\end{aligned}\tag{49}$$

The measurement distribution can be derived from (48), (49)

$$p(\mathbf{z}_k,\mathbf{x}_k|\mathbf{z}_{1:k},\mathbf{f}_{1:k-1})=p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k},\mathbf{f}_{1:k-1})$$

$$\begin{matrix}\mathbf{x}_k\\ \mathbf{z}_k\end{matrix}\Big|\,\mathbf{z}_{1:k-1},\mathbf{f}_{1:k-1}\sim \mathrm{CSN}_{m+n,q}\left(\begin{bmatrix}\bar{\boldsymbol{\mu}}_{x,k}\\ \mathbf{A}_k\bar{\boldsymbol{\mu}}_{x,k}\end{bmatrix},\begin{bmatrix}\bar{\boldsymbol{\Sigma}}_{xx,k} & \bar{\boldsymbol{\Sigma}}_{xx,k}\mathbf{A}_k^T\\ \mathbf{A}_k\bar{\boldsymbol{\Sigma}}_{xx,k} & \boldsymbol{\Sigma}_{vv,k}+\mathbf{A}_k\bar{\boldsymbol{\Sigma}}_{xx,k}\mathbf{A}_k^T\end{bmatrix},\begin{bmatrix}\bar{\mathbf{D}}_{x,k}^T\\ \mathbf{0}_{n\times m}\end{bmatrix}^T,\begin{bmatrix}\bar{\boldsymbol{\nu}}_{x,k}\\ \mathbf{0}_{m\times 1}\end{bmatrix},\bar{\boldsymbol{\Delta}}_{x,k}\right)\tag{50}$$

by marginalization over $\mathbf{x}_k$ which leads to

$$\mathbf{z}_k\big|\mathbf{z}_{1:k-1},\mathbf{f}_{1:k-1}\sim CSN_{n,q}\left(\boldsymbol{\mu}_{z,k},\boldsymbol{\Sigma}_{z,k},\mathbf{D}_{z,k},\boldsymbol{\nu}_{z,k},\boldsymbol{\Delta}_{z,k}\right)$$

with

$$\begin{aligned}
&\boldsymbol{\mu}_{z,k}=\mathbf{A}_k\bar{\boldsymbol{\mu}}_{x,k} && \mathbf{D}_{z,k}=\bar{\mathbf{D}}_{x,k}\bar{\boldsymbol{\Sigma}}_{xx,k}\mathbf{A}_k^T\Sigma_{zz,k}^{-1}\\
&\boldsymbol{\Sigma}_{zz,k}=\mathbf{A}_k\bar{\Sigma}_{xx,k}\mathbf{A}_k^T+\boldsymbol{\Sigma}_{vv,k-1} && \boldsymbol{\nu}_{z,k}=\boldsymbol{\nu}_{x,k-1}\\
&\boldsymbol{\Delta}_{z,k}=\bar{\boldsymbol{\Delta}}_{x,\mathrm{k}}+\left(\bar{\mathbf{D}}_{x,k}-\mathbf{D}_{z,k}\mathbf{A}_k\right)\bar{\boldsymbol{\Sigma}}_{xx,k}\bar{\mathbf{D}}_{x,k}^T
\end{aligned}\tag{51}$$

Applying the Bayes' rule using the results (48), (49) and (51), the posterior distribution is a direct consequence of *property C* and finally defined as

$$\mathbf{x}_k|\mathbf{z}_{1:k},\mathbf{f}_{1:k-1}\sim CSN_{m,q}\left(\boldsymbol{\mu}_{x,k},\boldsymbol{\Sigma}_{x,k},\mathbf{D}_{x,k},\boldsymbol{\nu}_{x,k},\boldsymbol{\Delta}_{\mathrm{x,k}}\right)$$

with

$$\begin{aligned}
&\boldsymbol{\mu}_{x,k}=\mathbf{T}_k\boldsymbol{\mu}_{x,k-1}+\mathbf{B}_k\mathbf{f}_{k-1}+\mathbf{K}_k\left(\mathbf{z}_k-\mathbf{A}_k\bar{\boldsymbol{\mu}}_{x,k}\right)\\
&\boldsymbol{\Sigma}_{xx,k}=\bar{\boldsymbol{\Sigma}}_{xx,k}-\mathbf{K}_k\mathbf{A}_k^T\bar{\boldsymbol{\Sigma}}_{xx,k}\\
&\mathbf{D}_{x,k}=\mathbf{D}_{x,k-1}\boldsymbol{\Sigma}_{xx,k-1}\mathbf{T}_k^T\bar{\boldsymbol{\Sigma}}_{xx,k}^{-1}\\
&\boldsymbol{\nu}_{x,k}=\boldsymbol{\nu}_{x,k-1}-\mathbf{D}_{x,k}\mathbf{K}_k\left(\mathbf{z}_k-\mathbf{A}_k\bar{\boldsymbol{\mu}}_{x,k}\right)
\end{aligned}$$

$$\boldsymbol{\Delta}_{x,k} = \boldsymbol{\Delta}_{x,k-1} + \left(\mathbf{D}_{x,k-1} - \mathbf{D}_{x,k}\mathbf{T}_k\right)\boldsymbol{\Sigma}_{xx,k-1}\mathbf{D}_{x,k-1}^T \tag{52}$$

Using the Kalman gain matrix $\mathbf{K}_k = \overline{\boldsymbol{\Sigma}}_{xx,k}\mathbf{A}_k^T\left(\boldsymbol{\Sigma}_{vv,k} + \mathbf{A}_k\overline{\boldsymbol{\Sigma}}_{xx,k}\mathbf{A}_k^T\right)^{-1}$ the strong resemblance to the standard Kalman filter algorithm is apparent. The only differences are the additional skewness parameters $\mathbf{D}_{x,k}$, $\boldsymbol{\nu}_{x,k}$ and $\boldsymbol{\Delta}_{x,k}$ to be updated at each time step. The closure properties of the skew-normal distribution allow to reduce the filter algorithm to the recursive calculation of the parameters. The algorithm of the skewed Kalman filter is summarized in figure 6.

The update of the skewness parameters $\mathbf{D}_{x,k}$ at time k depends on $\mathbf{D}_{x,k-1}$ from the previous time step, i.e. if $\mathbf{D}_{x,k-1} = \mathbf{0}$ at any time then $\mathbf{D}_{x,k} = \mathbf{0}$ for all times k. Hence, the initial skewness might get lost during the filter process and one obtains the standard Kalman filter. Furthermore assuming a Gaussian distributed measurement error, the state converge against a normal distribution after some time.

In the following an extension of the method is introduced which allows to implement a time variate skewness (if needed) directly in the observational equation.

3.3 Bayesian approach of an extension of the skewed Kalman filter

In order to handle skewed measurement data, Naveau developed an extension of the skewed Kalman algorithm which allows to introduce skewness directly into the observational model.

As shown in [9] the sum of i skew-normal distributed variables of dimension (n, m_i) follows a CSN distribution of dimension $\left(n, \sum_{i=1}^{k} m_i\right)$. Hence, modeling the skewness by adding a skew-normal measurement noise vector would lead to an increase of the size of the matrices $\mathbf{D}_{x,k}$ and $\boldsymbol{\Delta}_{\mathrm{x,k}}$ at each time step if the above described recursive procedure is used. In order to avoid the dimensional problem, in [1] the linear state space model is modified by splitting up the observational model in a linear part and a skewed part

$$\mathbf{z}_k = \mathbf{G}_k\mathbf{x}_k + \mathbf{v}_k \equiv \underbrace{\mathbf{A}_k\mathbf{u}_k}_{linear} + \underbrace{\mathbf{S}_k\mathbf{s}_k}_{skewed} + \underbrace{\mathbf{v}_k}_{noise} \tag{53}$$

where $\mathbf{G}_k = [\mathbf{A}_k, \mathbf{B}_k]$ and $\mathbf{x}_k = \left[\mathbf{u}_k^T, \mathbf{s}_k^T\right]^T$ with the vectors $\mathbf{s}_k \in \mathbb{R}^s$, $\mathbf{u}_k \in \mathbb{R}^u$ and the matrices of scalars $\mathbf{A}_k \in \mathbb{R}^{n\times s}$, $\mathbf{S}_k \in \mathbb{R}^{n\times u}$. Here the linear vector $\mathbf{u}_k$ as well as the zero-mean measurement noise $\mathbf{v}_k$ is assumed to be Gaussian distributed and independent of the skewness vector $\mathbf{s}_k$ with a CSN distribution. Hence, the dimension n of the observation vector $\mathbf{z}_k$ stays unchanged.
In [1] the linear part is used to describe a slowly changing trend and the skewed part corresponds to a time variate skewed process. In the investigated identification problem, the method is used to determine the system response due to an unknown ambient load function with CSN distribution. Hence the linear vector corresponds to the homogeneous part of the system response while the skewed vector describes the steady state response due to a skewed process describing the load. In order to generate a skewed process the following lemma to be found in [9] is used:

Property E: If $\mathbf{X} \in \mathbb{R}^n$ and $\mathbf{Y} \in \mathbb{R}^m$ are two random variables with joint normal distribution

$$\begin{bmatrix}\mathbf{X}\\ \mathbf{Y}\end{bmatrix} \sim N_{n+m}\left(\begin{bmatrix}\boldsymbol{\mu}\\ \boldsymbol{\nu}\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma} & -\boldsymbol{\Sigma}\mathbf{D}^T\\ -\mathbf{D}\boldsymbol{\Sigma} & \boldsymbol{\Delta} + \mathbf{D}\boldsymbol{\Sigma}\mathbf{D}^T\end{bmatrix}\right)$$

then the conditional distribution of $\mathbf{X}$ given $\mathbf{Y} \leq \mathbf{D}\boldsymbol{\mu}$ is skew normally distributed

$$\mathbf{X}|\ \mathbf{Y} \leq \mathbf{D}\boldsymbol{\mu} \sim CSN_{n,m}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{D}, \mathbf{v}, \boldsymbol{\Delta}) \tag{54}$$

Defining

$$\mathbf{u}_k = \mathbf{T}_k\mathbf{u}_{k-1} + \mathbf{S}_{u,k}\mathbf{w}_{u,k} \quad \text{and} \quad \mathbf{w}_{u,k} \sim N\left(\boldsymbol{\mu}_{w_u,k}, \boldsymbol{\Sigma}_{w_u,k}\right) \tag{55}$$

$$\mathbf{y}_k = -\mathbf{L}_k\mathbf{y}_{k-1} + \mathbf{w}_{y,k} \quad \text{and} \quad \mathbf{w}_{y,k} \sim N\left(\boldsymbol{\mu}_{w_y,k}, \boldsymbol{\Sigma}_{w_y,k}\right) \tag{56}$$

where $\mathbf{L}_k \in \mathbb{R}^{s\times s}$, $\mathbf{T}_k \in \mathbb{R}^{u\times u}$ are matrices of scalars, $\mathbf{w}_{u,k}$ and $\mathbf{w}_{y,k}$ are independent Gaussian distributed noise vectors. The joint distribution $(\mathbf{u}_k, \mathbf{y}_k)$ is given by

$$\begin{bmatrix}\mathbf{u}_k\\ \mathbf{y}_k\end{bmatrix} \sim N_{u+s}\left(\begin{bmatrix}\boldsymbol{\mu}_{u,k}\\ \boldsymbol{\mu}_{y,k}\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{uu,k} & \boldsymbol{\Sigma}_{uy,k}\\ \boldsymbol{\Sigma}_{yu,k} & \boldsymbol{\Sigma}_{yy,k}\end{bmatrix}\right) \tag{57}$$

The skewed process $\mathbf{s}_k$ is generated from the joint distribution $(\mathbf{y}_k, \mathbf{y}_{k-1})$ given by

$$\begin{bmatrix}\mathbf{y}_k\\ \mathbf{y}_{k-1}\end{bmatrix} \sim N_{s+s}\left(\begin{bmatrix}-\mathbf{L}_k\boldsymbol{\mu}_{y,k-1}+\boldsymbol{\mu}_{w_{y,k}}\\ \boldsymbol{\mu}_{y,k-1}\end{bmatrix}, \begin{bmatrix}\mathbf{L}_k\boldsymbol{\Sigma}_{yy,k-1}\mathbf{L}_k^T+\boldsymbol{\Sigma}_{w_k} & -\boldsymbol{\Sigma}_{yy,k-1}\mathbf{L}_k^T\\ -\mathbf{L}_k\boldsymbol{\Sigma}_{yy,k-1} & \boldsymbol{\Sigma}_{yy,k-1}\end{bmatrix}\right)$$

Due to *property E* the conditional distribution of $\mathbf{s}_k \sim \mathbf{y}_k|\, \mathbf{y}_{k-1} \leq \mathbf{D}_{y,k}\,\boldsymbol{\mu}_{y,k}$ follows a $\mathrm{CSN}_{n,m}$ distribution and one obtains

$$\begin{aligned}\mathbf{s}_k &= -\mathbf{L}_k\,\mathbf{y}_k^{cond}+\mathbf{w}_{y,k}\\ \mathbf{s}_k &\sim CSN_{s,s}(\boldsymbol{\mu}_{s,k}, \boldsymbol{\Sigma}_{ss,k}, \mathbf{D}_{s,k}, \boldsymbol{v}_{s,k}, \boldsymbol{\Delta}_{s,k})\end{aligned} \tag{58}$$

where

$$\begin{aligned}&\mathbf{y}_k^{cond} = \mathbf{y}_k|\, \mathbf{y}_{k-1} \leq \mathbf{D}_{y,k}\,\boldsymbol{\mu}_{y,k} && \mathbf{D}_{s,k} = \boldsymbol{\Sigma}_{yy,k-1}\mathbf{L}_k^T\boldsymbol{\Sigma}_{ss,k}^{-1}\\ &\boldsymbol{\mu}_{s,k} = -\mathbf{L}_k\boldsymbol{\mu}_{y,k-1}+\boldsymbol{\mu}_{w_{y,k}} && \mathbf{v}_{s,k} = \boldsymbol{\mu}_{y,k-1}-\mathbf{D}_{s,k}\boldsymbol{\mu}_{s,k}\\ &\boldsymbol{\Sigma}_{ss,k} = \mathbf{L}_k\boldsymbol{\Sigma}_{yy,k-1}\mathbf{L}_k^T+\boldsymbol{\Sigma}_{w_{y,k}} && \boldsymbol{\Delta}_{s,k} = \boldsymbol{\Sigma}_{yy,k-1}-\mathbf{D}_{s,k}\boldsymbol{\Sigma}_{ss,k}\mathbf{D}_{s,k}^T\end{aligned} \tag{58}$$

Using the inverse of *property B* the joint distribution $\mathbf{x}_k = (\mathbf{u}_k, \mathbf{s}_k)$ follows a $CSN_{n+m\times n+m}(\boldsymbol{\mu}_{x,k}, \boldsymbol{\Sigma}_{xx,k}, \mathbf{D}_{x,k}, \boldsymbol{\nu}_{x,k}, \boldsymbol{\Delta}_{x,k})$ given by

$$\mathbf{x}_k = \begin{bmatrix}\mathbf{u}_k\\ \mathbf{s}_k\end{bmatrix} \sim CSN_{u+s,u+s}\left(\begin{bmatrix}\boldsymbol{\mu}_{u,k}\\ \boldsymbol{\mu}_{s,k}\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{uu,k} & \mathbf{0}_{u\times s}\\ \mathbf{0}_{s\times u} & \boldsymbol{\Sigma}_{ss,k}\end{bmatrix}, \begin{bmatrix}\mathbf{0}_{u\times u} & \mathbf{0}_{u\times s}\\ \mathbf{0}_{s\times u} & \mathbf{D}_{s,k}\end{bmatrix}, \begin{bmatrix}\mathbf{0}_{u\times 1}\\ \boldsymbol{v}_{s\times 1}\end{bmatrix}, \begin{bmatrix}\mathbf{I}_{u\times u} & \mathbf{0}_{u\times s}\\ \mathbf{0}_{s\times u} & \boldsymbol{\Delta}_{s,k}\end{bmatrix}\right) \tag{60}$$

Hence, using *property D* and the result (51) the observation $\mathbf{z}_k|\mathbf{z}_{k-1}$ follows

$$\mathbf{z}_k|\mathbf{z}_{k-1} \sim CSN_{n,n}(\boldsymbol{\mu}_{z,k}, \boldsymbol{\Sigma}_{zz,k}, \mathbf{D}_{z,k}, \boldsymbol{\nu}_{z,k}, \boldsymbol{\Delta}_{z,k})$$

with

$$\begin{aligned}&\boldsymbol{\mu}_{z,k} = \mathbf{G}_k\boldsymbol{\mu}_{x,k} && \boldsymbol{\nu}_{z,k} = \boldsymbol{\nu}_{x,k}\\ &\boldsymbol{\Sigma}_{zz,k} = \mathbf{G}_k\boldsymbol{\Sigma}_{xx,k}\mathbf{G}_k^T+\boldsymbol{\Sigma}_{vv,k} && \boldsymbol{\Delta}_{z,k} = \boldsymbol{\Delta}_{x,k}+(\mathbf{D}_{x,k}-\mathbf{D}_{z,k}\mathbf{G}_k)\boldsymbol{\Sigma}_{xx,k}\mathbf{D}_{x,k}^T\\ &\mathbf{D}_{z,k} = \mathbf{D}_{x,k}\boldsymbol{\Sigma}_{xx,k}\mathbf{G}_k^T\boldsymbol{\Sigma}_{zz,k}^{-1}\end{aligned} \tag{61}$$

In contrast to the linear skewed Kalman filter described in section 3.2, where the skewness parameter $\mathbf{D}_k$ was sequently derived from the previous time step, the additive skewed process $\mathbf{s}_k$ of the extended algorithm allows to implement a different skewness at each time step by introducing a temporal structural of the matrix $\mathbf{L}_k$ [2].

3.4 Recursive algorithm of the extended skewed Kalman filter

Due to eq. (60) the extended skewed Kalman filter is initialized by the posterior CSN given by

$$\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1} \sim CSN_{m+s\times m+s}(\boldsymbol{\mu}_{x,k-1}, \boldsymbol{\Sigma}_{xx,k-1}, \mathbf{D}_{x,k-1}, \boldsymbol{\nu}_{x,k-1}, \boldsymbol{\Delta}_{x,k-1}) \tag{62}$$

In contrast to the linear skewed Kalman filter the prior time update of the state $\mathbf{x}_k = (\mathbf{u}_k, \mathbf{s}_k)^T$ cannot be calculated directly using eq. (49) as $\mathbf{s}_k$ has to be generated at each time step from the joint normal distribution $(\mathbf{y}_k, \mathbf{y}_{k-1})^T$ using eq. (58). Hence, the state space model defined by the observational equation (53), the linear system (55), (56) and the equation (58) of the skewed process $\mathbf{s}_k$ has a nonlinear structure in contrast to the model described in section 3.2.

The temporal structure of the state $\mathbf{x}_k|\mathbf{z}_{1:k}$ is defined by the variable $(\mathbf{u}_k, \mathbf{y}_k, \mathbf{y}_{k-1}|\, \mathbf{z}_{1:k})$ conditionally on $\mathbf{y}_{k-1} \leq \mathbf{D}_{y,k}\,\boldsymbol{\mu}_{y,k}$. Using the Bayes' rule the posterior distribution of $p(\mathbf{u}_k, \mathbf{y}_k, \mathbf{y}_{k-1}|\mathbf{z}_{1:k})$ is obtained by

$$p(\mathbf{u}_k, \mathbf{y}_k, \mathbf{y}_{k-1}|\mathbf{z}_{1:k}) = \frac{p(\mathbf{z}_k|\, \mathbf{u}_k, \mathbf{y}_k, \mathbf{y}_{k-1})p(\mathbf{u}_k, \mathbf{y}_k, \mathbf{y}_{k-1}|\mathbf{z}_{1:k-1})}{p(\mathbf{z}_k|\mathbf{z}_{1:k-1})} \tag{63}$$

The prior estimate of the multivariate normal distributed variable $(\mathbf{u}_k, \mathbf{y}_k, \mathbf{y}_{k-1}|\, \mathbf{z}_{1:k-1})$ conditional on the observations $\mathbf{z}_{1:k-1}$ up to time $t = k-1$ is given by

$$\left(\begin{matrix}\mathbf{u}_k\\ \mathbf{y}_k\\ \mathbf{y}_{k-1}\end{matrix}\middle|\, \mathbf{z}_{1:k-1}\right) = \left(\begin{matrix}\mathbf{T}_k\mathbf{u}_{k-1}+\mathbf{S}_{u,k}\mathbf{w}_{u,k}\\ -\mathbf{L}_k\mathbf{y}_{k-1}+\mathbf{w}_{y,k}\\ \mathbf{y}_{k-1}\end{matrix}\middle|\, \mathbf{z}_{1:k-1}\right)$$

$$p(\mathbf{u}_k,\mathbf{y}_k,\mathbf{y}_{k-1}|\,\mathbf{z}_{1:k-1}) \sim N\left(\begin{bmatrix}\mathbf{T}_k\boldsymbol{\mu}_{u,k-1}+\mathbf{S}_{u,k}\boldsymbol{\mu}_{w_{u,k}}\\ -\mathbf{L}_k\boldsymbol{\mu}_{y,k-1}+\boldsymbol{\mu}_{w_{y,k}}\\ \boldsymbol{\mu}_{y,k-1}\end{bmatrix},\begin{bmatrix}\overline{\boldsymbol{\Sigma}}_{uu,k} & -\mathbf{T}_k\boldsymbol{\Sigma}_{uy,k-1}\mathbf{L}_k^T & \mathbf{T}_k\boldsymbol{\Sigma}_{uy,k-1}\\ -\mathbf{L}_k\boldsymbol{\Sigma}_{uy,k-1}\mathbf{T}_k^T & \overline{\boldsymbol{\Sigma}}_{yy,k} & -\mathbf{L}_k\overline{\boldsymbol{\Sigma}}_{yy,k}\\ \boldsymbol{\Sigma}_{uy,k-1}\mathbf{T}_k^T & -\overline{\boldsymbol{\Sigma}}_{yy,k}\mathbf{L}_k^T & \boldsymbol{\Sigma}_{yy,k-1}\end{bmatrix}\right)$$

where

$$\overline{\boldsymbol{\Sigma}}_{uu,k}=\mathbf{T}_k\boldsymbol{\Sigma}_{uu,k-1}\mathbf{T}_k^T+\mathbf{S}_{u,k}\boldsymbol{\Sigma}_{w_{u,k}}\mathbf{S}_k^T \quad\text{and}\quad \overline{\boldsymbol{\Sigma}}_{yy,k}=\mathbf{L}_k\boldsymbol{\Sigma}_{yy,k-1}\mathbf{L}_k^T+\boldsymbol{\Sigma}_{w_{y,k}}. \tag{64}$$

Once the new measurement data $\mathbf{z}_k$ is available the likelihood $p(\mathbf{z}_k|\mathbf{u}_k,\mathbf{y}_k,\mathbf{y}_{k-1})$ is derived from the observational equation (53). Assuming that $(\mathbf{u}_k,\mathbf{y}_k,\mathbf{y}_{k-1})$ are given, the only random variable in (53) is the zero-mean measurement noise $\mathbf{v}_k$, which leads to

$$p(\mathbf{z}_k|\mathbf{u}_k,\mathbf{y}_k,\mathbf{y}_{k-1},\mathbf{z}_{1:k-1}) \sim N\big(\mathbf{A}_k\overline{\boldsymbol{\mu}}_{u,k-1}+\mathbf{S}_k\mathrm{E}[\mathbf{s}_k|\mathbf{z}_{1:k-1}],\boldsymbol{\Sigma}_{\mathrm{vv},k}\big) \tag{65}$$

where $\overline{\boldsymbol{\mu}}_{u,k-1}=\mathbf{T}_k\boldsymbol{\mu}_{u,k-1}+\mathbf{S}_{u,k}\boldsymbol{\mu}_{w_{u,k}}$. In contrast to the linear case the conditional expectation $\mathrm{E}[\mathbf{s}_k|\mathbf{z}_{1:k-1}]$ of the variable $\mathbf{s}_k$ has to be included. Equation (64) and (65) leads to the joint distribution of $p(\mathbf{u}_k,\mathbf{y}_k,\mathbf{y}_{k-1},\mathbf{z}_k|\,\mathbf{z}_{1:k-1})$ using

$$p(\mathbf{u}_k,\mathbf{y}_k,\mathbf{y}_{k-1},\mathbf{z}_k|\,\mathbf{z}_{1:k-1})=p(\mathbf{z}_k|\mathbf{u}_k,\mathbf{y}_k,\mathbf{y}_{k-1},\mathbf{z}_{1:k-1})p\big(\mathbf{u}_k,\mathbf{y}_k,\mathbf{y}_{k-1}|\,\mathbf{z}_{1:k-1}\big) \tag{66}$$

and it follows

$$p(\mathbf{u}_k,\mathbf{y}_k,\mathbf{y}_{k-1},\mathbf{z}_k|\,\mathbf{z}_{1:k-1}) \sim N\left(\begin{bmatrix}\mathbf{T}_k\boldsymbol{\mu}_{u,k-1}+\mathbf{S}_{u,k}\boldsymbol{\mu}_{w_{u,k}}\\ -\mathbf{L}_k\boldsymbol{\mu}_{y,k-1}+\boldsymbol{\mu}_{w_{y,k}}\\ \boldsymbol{\mu}_{y,k-1}\\ \boldsymbol{\mu}_z\end{bmatrix},\begin{bmatrix}\overline{\boldsymbol{\Sigma}}_{uu,k} & -\mathbf{T}_k\boldsymbol{\Sigma}_{uy,k-1}\mathbf{L}_k^T & \mathbf{T}_k\boldsymbol{\Sigma}_{uy,k-1} & \overline{\boldsymbol{\Sigma}}_{uu,k}\mathbf{A}_k^T\\ -\mathbf{L}_k\boldsymbol{\Sigma}_{uy,k-1}\mathbf{T}_k^T & \overline{\boldsymbol{\Sigma}}_{yy,k} & -\mathbf{L}_k\overline{\boldsymbol{\Sigma}}_{yy,k} & \mathbf{C}_k\mathbf{S}_k^T\\ \boldsymbol{\Sigma}_{uy,k-1}\mathbf{T}_k^T & -\overline{\boldsymbol{\Sigma}}_{yy,k}\mathbf{L}_k^T & \boldsymbol{\Sigma}_{yy,k-1} & \mathbf{C}_{k-1}\mathbf{S}_k^T\\ \mathbf{A}_k\overline{\boldsymbol{\Sigma}}_{uu,k} & \mathbf{S}_k\mathbf{C}_k & \mathbf{S}_k\mathbf{C}_{k-1} & \boldsymbol{\Sigma}_{zz,k}\end{bmatrix}\right)$$

where

$$\begin{aligned}&\boldsymbol{\Sigma}_{zz,k}=\mathbf{A}_k\overline{\boldsymbol{\Sigma}}_{uu,k}\mathbf{A}_k^T+\mathbf{S}_k\mathrm{var}[\mathbf{s}_k|\mathbf{z}_{1:k-1}]\mathbf{S}_k^T+\boldsymbol{\Sigma}_{\mathrm{vv},k} && \mathbf{C}_{k-1}=\mathrm{Cov}[\mathbf{y}_{k-1},\mathbf{s}_k|\mathbf{z}_{1:k-1}]\\ &\boldsymbol{\mu}_z=\mathbf{A}_k\overline{\boldsymbol{\mu}}_{u,k-1}+\mathbf{S}_k\mathrm{E}[\mathbf{s}_k|\mathbf{z}_{1:k-1}], && \mathbf{C}_k=\mathrm{Cov}[\mathbf{y}_k,\mathbf{s}_k|\mathbf{z}_{1:k-1}]=-\mathbf{L}_k\mathbf{C}_{k-1}+\boldsymbol{\Sigma}_{w_{y,k}}\end{aligned} \tag{67}$$

Both the condition expectation $\mathrm{E}[\mathbf{s}_k|\mathbf{z}_{1:k-1}]$ as well as the conditional variance $\mathrm{Var}[\mathbf{s}_k|\mathbf{z}_{1:k-1}]$ of the skewed process $\mathbf{s}_k$ can be derived from the moment generating function given by [8]

$$M(\boldsymbol{\theta})=\frac{\Phi_S\big(\mathbf{D}_{s,k}\boldsymbol{\Sigma}_{ss,k}\boldsymbol{\theta};\,\boldsymbol{\nu}_{\mathrm{s,k}},\boldsymbol{\Delta}_{s,k}+\mathbf{D}_{s,k}\boldsymbol{\Sigma}_{ss,k}\mathbf{D}_{s,k}^T\big)}{\Phi_S\big(\mathbf{0};\,\boldsymbol{\nu}_{\mathrm{s,k}},\boldsymbol{\Delta}_{s,k}+\mathbf{D}_{s,k}\boldsymbol{\Sigma}_{ss,k}\mathbf{D}_{s,k}^T\big)}e^{\boldsymbol{\theta}^T\boldsymbol{\mu}_{s,k}+\boldsymbol{\theta}^T\boldsymbol{\Sigma}_{ss,k}\boldsymbol{\theta}/2} \tag{68}$$

After marginalization over $\mathbf{y}_{k-1}$ the updated distribution of the variable $(\mathbf{u}_k,\mathbf{y}_k)$ conditional on $\mathbf{z}_k$ is defined as

$$p(\mathbf{u}_k,\mathbf{y}_k|\,\mathbf{z}_{1:k}) \sim N\left(\begin{bmatrix}\boldsymbol{\mu}_{u,k}\\ \boldsymbol{\mu}_{y,k}\end{bmatrix},\begin{bmatrix}\boldsymbol{\Sigma}_{uu,k} & \boldsymbol{\Sigma}_{uy,k}\\ \boldsymbol{\Sigma}_{yu,k} & \boldsymbol{\Sigma}_{yy,k}\end{bmatrix}\right)$$

where

$$\begin{aligned}&\mathbf{e}_k=\mathbf{y}_k-\mathbf{A}_k\overline{\boldsymbol{\mu}}_{u,k-1}-\mathbf{S}_k\mathrm{E}[\mathbf{s}_k|\mathbf{z}_{1:k-1}]\\ &\boldsymbol{\Sigma}_{uy,k}=-\big(\mathbf{T}_k\boldsymbol{\Sigma}_{uy,k-1}\mathbf{L}_k^T+\overline{\boldsymbol{\Sigma}}_{uu,k}\mathbf{A}_k^T\boldsymbol{\Sigma}_{zz,k}^{-1}\mathbf{S}_k\mathbf{C}_k\big)\end{aligned} \tag{69}$$

Hence, the updated parameters of the linear part $\mathbf{u}_k|\,\mathbf{z}_{1:k}$ are given by

$$\begin{cases}\boldsymbol{\mu}_{u,k}=\mathbf{T}_k\boldsymbol{\mu}_{u,k-1}+\mathbf{S}_{u,k}\boldsymbol{\mu}_{w_{u,k}}+\overline{\boldsymbol{\Sigma}}_{uu,k}\mathbf{A}_k^T\boldsymbol{\Sigma}_{zz,k}^{-1}\mathbf{e}_k\\ \boldsymbol{\Sigma}_{uu,k}=\overline{\boldsymbol{\Sigma}}_{uu,k}-\overline{\boldsymbol{\Sigma}}_{uu,k}\mathbf{A}_k^T\boldsymbol{\Sigma}_{zz,k}^{-1}\mathbf{A}_k\overline{\boldsymbol{\Sigma}}_{uu,k}\end{cases} \tag{70}$$

and for the update of the skewed part $\mathbf{y}_k|\,\mathbf{z}_{1:k}$ one obtains

$$\begin{cases}\boldsymbol{\mu}_{y,k}=-\mathbf{L}_k\boldsymbol{\mu}_{y,k-1}+\boldsymbol{\mu}_{w_{y,k}}+\mathbf{C}_k\mathbf{S}_k^T\boldsymbol{\Sigma}_{zz,k}^{-1}\mathbf{e}_k\\ \boldsymbol{\Sigma}_{yy,k}=\overline{\boldsymbol{\Sigma}}_{yy,k}-\mathbf{C}_k\mathbf{S}_k^T\boldsymbol{\Sigma}_{zz,k}^{-1}\mathbf{S}_k\mathbf{C}_k\end{cases} \tag{71}$$

By marginalization of the distribution (67) over $\mathbf{u}_k$ and conditioning with respect to $\mathbf{z}_k$, one obtains after some algebraic simplification

$$p(\mathbf{y}_k,\mathbf{y}_{k-1}|\,\mathbf{z}_{1:k}) \sim N\left(\begin{bmatrix}\boldsymbol{\mu}_{y,k}\\ \boldsymbol{\mu}_{y,k-1}^*\end{bmatrix},\begin{bmatrix}\boldsymbol{\Sigma}_{yy,k} & -\mathbf{L}_k^*\boldsymbol{\Sigma}_{yy,k-1}^*\\ -\boldsymbol{\Sigma}_{yy,k-1}^*\mathbf{L}_k^{*T} & \boldsymbol{\Sigma}_{yy,k-1}^*\end{bmatrix}\right)$$

where

$$\begin{aligned}&\boldsymbol{\mu}_{y,k-1}^*=\boldsymbol{\mu}_{y,k-1}+\mathbf{C}_{k-1}\mathbf{S}_k^T\boldsymbol{\Sigma}_{zz,k}^{-1}\mathbf{e}_k && \mathbf{L}_k^*=\mathbf{L}_k+\boldsymbol{\Sigma}_{w_{y,k}}\mathbf{S}_k^T\boldsymbol{\Sigma}_{zz,k}^{-1}\mathbf{S}_k\mathbf{C}_{k-1}\big(\boldsymbol{\Sigma}_{yy,k-1}^*\big)^{-1}\\ &\boldsymbol{\Sigma}_{yy,k-1}^*=\boldsymbol{\Sigma}_{yy,k-1}-\mathbf{C}_{k-1}\mathbf{S}_k^T\boldsymbol{\Sigma}_{zz,k}^{-1}\mathbf{S}_k\mathbf{C}_{k-1}.\end{aligned} \tag{72}$$

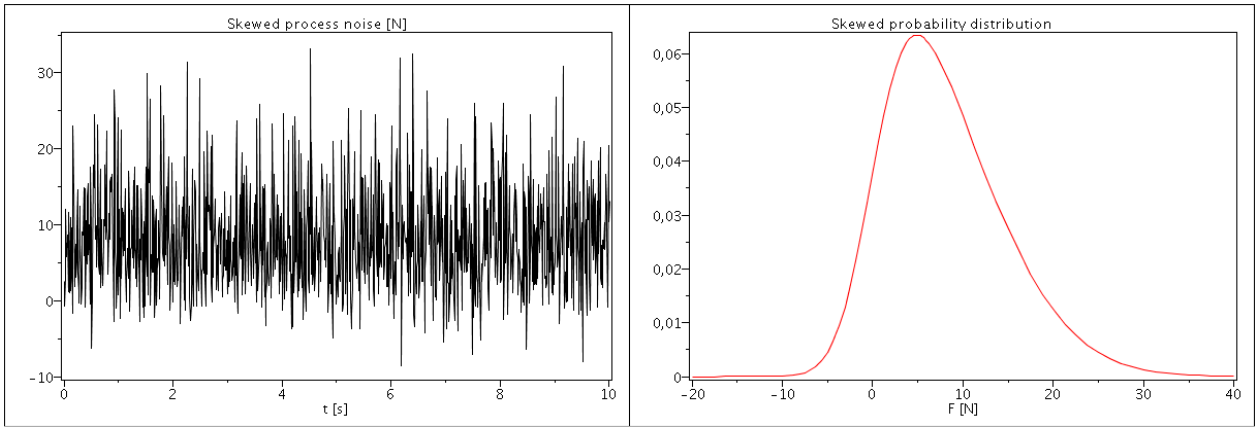

Figure 7: Simulated process noise $\mathbf{s}_k$ and corresponding probability distribution

Using *property E* and eq. (58) the skewed process $\mathbf{s}_k$ follows a CSN distribution and one obtains

$$\mathbf{s}_k|\mathbf{z}_{1:k} \sim CSN_{s,s}\left(\boldsymbol{\mu}_{s,k}, \boldsymbol{\Sigma}_{ss,k}, \mathbf{D}_{s,k}, \boldsymbol{\nu}_{s,k}, \boldsymbol{\Delta}_{s,k}\right)$$

where

$$\begin{aligned} \boldsymbol{\Sigma}_{ss,k} &= \boldsymbol{\Sigma}_{yy,k} & \mathbf{v}_{s,k} &= \boldsymbol{\mu}^*_{y,k-1} - \mathbf{D}_{s,k}\boldsymbol{\mu}_{y,k} \\ \boldsymbol{\mu}_{s,k} &= \boldsymbol{\mu}_{y,k} & \boldsymbol{\Delta}_{s,k} &= \boldsymbol{\Sigma}^*_{yy,k-1} - \mathbf{D}_{s,k}\boldsymbol{\Sigma}_{yy,k}\mathbf{D}^T_{s,k} \\ \mathbf{D}_{s,k} &= \boldsymbol{\Sigma}^*_{yy,k-1}\mathbf{L}^{*T}_k\boldsymbol{\Sigma}^{-1}_{yy,k} \end{aligned} \tag{73}$$

3.5 Numerical example

The SDOF system of the example given in 2.3 is now investigated using the extended skewed Kalman filter. The linear part $\mathbf{u}_k$ of the filter describes the system response while the process $\mathbf{s}_k$ is used to describe an ambient load modeled as skewed noise process. In order to apply the introduced algorithm to the parameter identification problem the system response is nonlinearly depending on the unknown stiffness and damping parameter. Hence, the linear part $\mathbf{u}_k$ is obtain by approximating the nonlinear system equation by a first order Taylor series as described in section 2.3. Consequently, the vector $\mathbf{u}_k$ is extended by the parameters to be identified and the transfer matrices $\mathbf{T}_k$ and $\mathbf{S}_k$ are replaced by the extended, linearised matrices $\mathbf{T}_{ext,k}$ and $\mathbf{S}_{ext,k}$.

The figure 7 shows the simulated process noise $\mathbf{s}_k$ and the underlying probability density function with mean $\mu_s = 8$ N and deviation $\sigma_s = 7$ N. The estimation of the temporal evolution of the state $\mathbf{x}_k$ by the standard Kalman filter (red line) and the extended skewed Kalman filter (green line) as well as the simulated (true) system response are shown in figure 8. The figure 9 illustrates the results of the parameter identification by the standard Kalman filter and the non-linear skewed filter. In the first case the unknown load was assumed to be a zero-mean white noise process. As in the numerical example 2.3 the unknown load is regarded as system error w_f with deviation $\sigma_{w_f} = 12$ N. A measurement error of $\sigma_v = 1$ mm was assumed and the error of the state $\varepsilon_{\hat{x},k}$ was chosen as in the numerical example in section 2. The initial model parameters where estimated taking into account an error of 50% of the true model parameters.

The table 3 shows the initial values and the identified parameters after $t = 190$ s. While the identification of the stiffness parameter a_0 leads to similar results in both methods the error in the estimation of the damping parameter a_1 is much smaller in the non-linear skewed Kalman filter.

	true values	**initial values**	**initial error**	**identified parameters**		**identification error**	
				Standard Kalman	Skewed Kalman	Standard Kalman	Skewed Kalman
$a_0[\mathrm{s}^{-2}]$	1000	1500	50 %	998	999	0,2 %	0,1%
$a_1\ [\mathrm{s}^{-1}]$	1	1.5	50 %	1,17	0.99	17 %	1%

Table 3: Initial values and identification result after $t = 190$ s

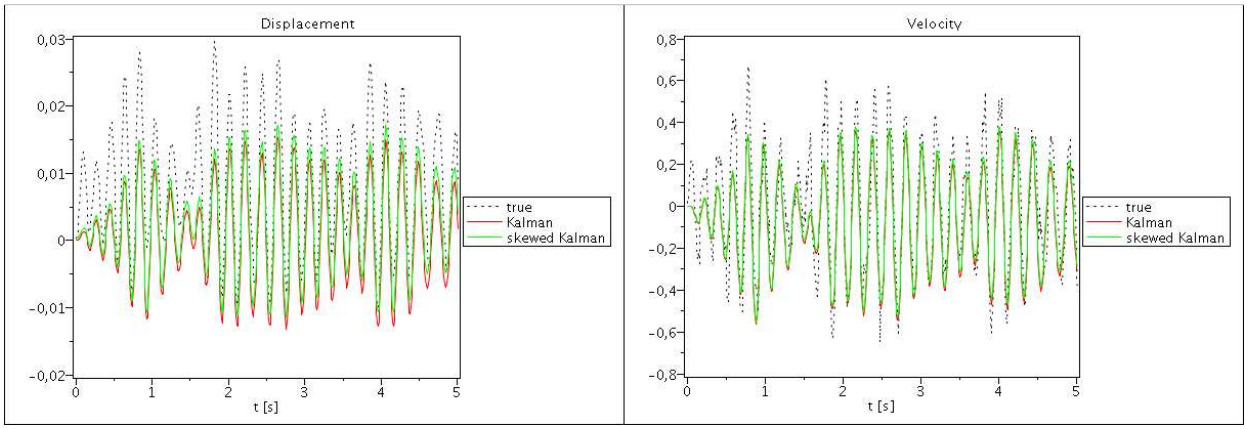

Figure 8: Simulated (true) and estimated temporal evolution of the state $\mathbf{x}_k$ using the standard Kalman filter (red line) and the extended skewed Kalman filter (green line)

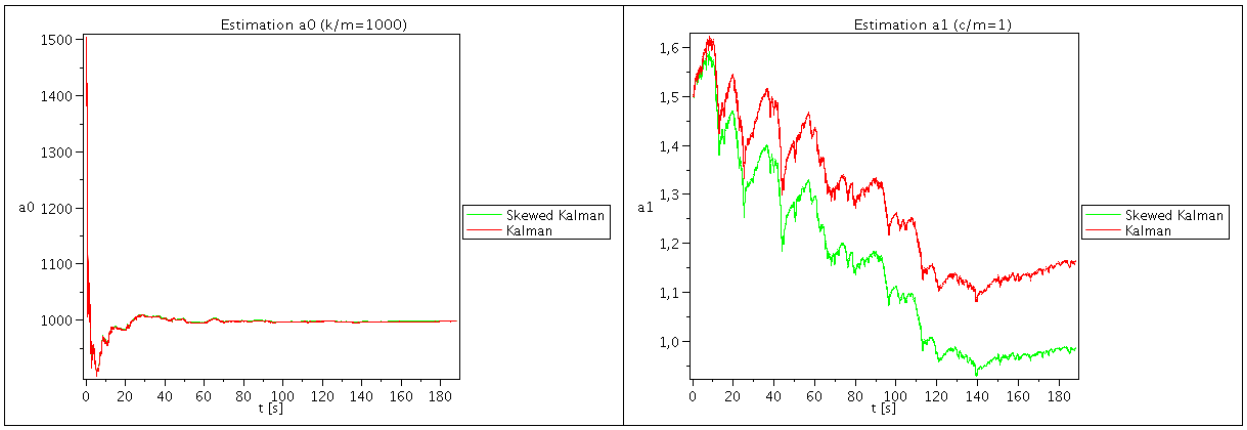

Figure 9: Identified values of the stiffness and damping parameters a_0 and a_1 after $\mathrm{t} = 150$ s

References

[1] Z. Chen, Bayesian *Filtering: From Kalman Filters to Particle Filters, and Beyond*, Adaptive Systems Laboratory, McMaters University, 2003

[2] P. Naveau, M. G. Genton, and X. Shen, *A skewed Kalman Filter*. Journal of Multivariate Analysis, Vol. 94, No. 2, pp. 382–400, 2005

[3] R. Isermann, *Identifikation dynamischer Systeme*, Band 2, Springer-Verlag Berlin, Heidelberg, 1988

[4] A. Eichhorn, *Ein Beitrag zur Identifikation von dynamischen Strukturmodellen mit Methoden der adaptiven Kalman-Filterung*, Deutsche Geodätische Kommission, Reihe C, Nr. 585, 2005

[5] G. Bishop, G, Welch, An introduction to the Kalman filter. SIGGRAPH 2001, Course Notes, 2001

[6] R. J. Meinhold, N.D. Singpurwalla, *Understanding the Kalman filter*, The American Statistician, Vol. 37, No. 2, pp. 123-127 (1983)

[7] R. M. du Plessis, *Poor man's explanation of Kalman Filters or How I stopped worrying and learned to love matrix inversion*, Taygeta Scientific Incorporated (May 1997), June 1967

[8] N. G. Genton, *Skew-elliptical distributions and their applications: a journey beyond normality*, Chapman & Hall/CRC, London, pp. 25, 2004

[9] G. Gonzáles-Farías, A. Domínguez-Molina, A. K. Gupta, *Additive properties of skew normal random vectors*, Journal of Statistical Planning and Inference Vol. 126, pp. 521-534, 2004