# A NEW FORMULATION FOR NONLINEAR FORWARD-BACKWARD SMOOTHING

Anindya S. Paul and Eric A. Wan Computer Science and Electrical Engineering Department OGI School of Science and Engineering, Oregon Health & Science University (OHSU) 20000 NW Walker Road, Beaverton, OR 97006 {anindya, ericwan}@csee.ogi.edu

## ABSTRACT

A new formulation for nonlinear smoothing is derived using forward-backward sigma-point Kalman filtering (SPKF). The forward filter uses the standard SPKF. The backward filter requires the use of the inverse dynamics of the forward filter. While smoothers based on the extended Kalman filter (EKF) simply invert the linearized dynamics, with the SPKF the forward nonlinear dynamics are never analytically linearized. Thus the backward nonlinear dynamics are not well defined. In previous work, a sigma-point Kalman smoother (SPKS) was derived by learning a nonlinear model of the backward dynamics from empirical data. In this paper, we make use of the relationship between the SPKF and weighted statistical linear regression (WSLR). The resulting pseudo-linearized dynamics obtained by WSLR is more accurate in the statistical sense than using a first order truncated Taylor series expansion as with the EKF. A new backward information filter can then be derived, which is combined with the forward SPKF to form the smoothed estimates.

*Index Terms*—Kalman filter, sigma-point Kalman smoother, statistical linearized regression, forward-backward filter.

## **1. INTRODUCTION**

The Kalman filter provides an optimal recursive estimate for the state  $x_k$  of a linear state-space system driven by Gaussian noise given all noisy measurements  $Z_k = \{z_1, z_2, ..., z_k\}$  up to the current time *k*. In contrast, the Kalman smoother estimates the conditional expectation of the state given all past and future measurements  $Z_k = \{z_1, z_2, ..., z_N\}$ ,  $1 \le k \le N$ . Several common Kalman smoothing formulations are given in [2]-[5]. In this article we consider fixed interval smoothing where the final time *N* is fixed. For nonlinear state-space models, the extended Kalman filter (EKF) approach may be used whereby the nonlinear state-space model is linearized around the estimate of the current state using a first order Taylor series expansion. Linear Kalman smoothing equations are then applied.

The smoothing algorithm presented in this paper employs the sigma-point Kalman filter (SPKF). The SPKF, which includes the unscented Kalman filter (UKF) [6], central difference Kalman filter (CDKF) [7], and their square-root variants [8], has recently become a popular better alternative to the EKF. Like the EKF, the SPKF approximates the state distribution by a gaussian random variable (GRV). However, the probability distribution is represented by a set of carefully chosen deterministic sample points (known as sigma points). These sigma points are then propagated through the true nonlinear system, with the posterior

mean and covariance calculated using simple weighted averaging. This approach captures the posterior mean and covariance accurately to the 2nd-order (3rd-order is achieved for symmetric distribution) compared to EKF which linearizes the nonlinear systems and only achieves 1st-order accuracy.

Two SPKF-based smoother variants have appeared in the literature. In [1], the sigma-point Kalman smoother (SPKS) uses a forward-backward approach. A standard SPKF is run in the forward direction using the nonlinear model. A second SPKF is then run in the backward direction and the two estimates are combined. However, in order to run the backward SPKF, an inverse dynamic model of the forward filter is needed, which is derived by training a backward nonlinear predictor (e.g., a neural network model). As the backward model needs to be fit to the data, it is both application-specific and potentially time consuming to design. In [9], the unscented Rauch-Tung-Striebel smoother (URTSS) is proposed, which uses the joint distribution of the current and future state in order to obtain a smoothed estimate of the current state. While this avoids the need for inverse dynamics, the computational complexity (cube of the state dimension) increases significantly due to the doubling of the state dimension.

The proposed smoother presented in this paper follows the same forward-backward procedure described in [1]. However, instead of learning a new backward model from the data, our proposed SPKS makes use of the WSLR formulation of the filter. WSLR is a linearization technique that takes into account the uncertainty of the prior random variable (RV) when linearizing the nonlinear model [1]. In this way, WSLR is more accurate in the statistical sense than first-order linearization which does factor in the "probabilistic spread" at the point of linearization. By representing the forward dynamics in terms of WSLR, we are able to derive from first principles a backward information filter that does not require inverting the nonlinear dynamics. Estimates of the forward and backward filter are then optimally combined to generate smoothed estimates in the standard manner. To distinguish between the algorithms, we refer to the new filter as a forwardbackward statistical linearized sigma-point Kalman smoother (FBSL-SPKS). The original smoother presented in [1] is referred to as a forward-backward nonlinear sigma-point Kalman smoother (FBNL-SPKS).

## 2. RELATIONSHIP BETWEEN SPKF AND WSLR

Consider a prior RV x that is propagated thorough a nonlinear function g(x) to obtain a posterior RV y. Sigma points  $\chi_i$ , i = 0, 1, ..., 2M are selected as the prior mean  $\overline{x}$  plus and minus

the columns of the square root of the prior covariance  $P_x$ 

$$\boldsymbol{\chi} = \begin{bmatrix} \overline{\boldsymbol{x}} & \overline{\boldsymbol{x}} + \gamma \sqrt{\boldsymbol{P}_x} & \overline{\boldsymbol{x}} - \gamma \sqrt{\boldsymbol{P}_x} \end{bmatrix},$$
(1)

where M is the RV dimension, and  $\gamma$  is a composite scaling parameter.

The weighted statistics of the sigma points capture the mean  $\overline{x}$  and prior covariance  $P_x$ .

$$\overline{\boldsymbol{x}} = \sum_{i=0}^{2M} w_i \boldsymbol{\chi}_i$$

$$\boldsymbol{P}_x = \sum_{i=0}^{2M} w_i (\boldsymbol{\chi}_i - \overline{\boldsymbol{x}}) (\boldsymbol{\chi}_i - \overline{\boldsymbol{x}})^T,$$
(2)

where  $w_i$  is the normalized scalar weight. Each prior sigma point is propagated through the nonlinearity to form the posterior sigma point set  $\gamma_i$ .

$$\boldsymbol{\gamma}_i = g(\boldsymbol{\chi}_i) \quad i = 0, \dots, 2M .$$
(3)

The posterior statistics can then be calculated using weighted averaging of the posterior sigma points,

$$\hat{\boldsymbol{y}} = \sum_{i=0}^{2M} w_i \boldsymbol{\gamma}_i$$

$$\hat{\boldsymbol{P}}_y = \sum_{i=0}^{2M} w_i (\boldsymbol{\gamma}_i - \hat{\boldsymbol{y}}) (\boldsymbol{\gamma}_i - \hat{\boldsymbol{y}})^T$$

$$\hat{\boldsymbol{P}}_{xy} = \sum_{i=0}^{2M} w_i (\boldsymbol{\chi}_i - \overline{\boldsymbol{x}}) (\boldsymbol{\gamma}_i - \hat{\boldsymbol{y}})^T.$$
(4)

An alternative view is to consider the estimates arising from the sigma point approach as a weighted statistical linearization of the nonlinear dynamics:

$$y = g(x) \cong Ax + b + \varepsilon$$
 (5)

where  $\boldsymbol{\varepsilon}$  is the linearization error. A and b are calculated to minimize the expected mean squared errors (MSE).

$$[\mathbf{A}, \mathbf{b}] = \arg\min(E\left[\mathbf{\varepsilon}^T \mathbf{W}\mathbf{\varepsilon}\right]) \tag{6}$$

The true expectation is replaced as a finite sample approximation,

$$E\left[\boldsymbol{\varepsilon}^{T}\boldsymbol{W}\boldsymbol{\varepsilon}\right] \approx \sum_{i=0}^{2M} w_{i}\boldsymbol{\varepsilon}_{i}^{T}\boldsymbol{\varepsilon}_{i} , \qquad (7)$$

where the point-wise linearization error  $\varepsilon_i = \gamma_i - A \chi_i - b$ . Taking partial derivatives with respect to *A* and *b*, we can show that,

where the necessary terms are calculated from the posterior sigma points. The linearization error  $\boldsymbol{\varepsilon}$  has zero mean and covariance  $\boldsymbol{P}_{\varepsilon} = \hat{\boldsymbol{P}}_{v} - \boldsymbol{A}\boldsymbol{P}_{x}\boldsymbol{A}^{T}$ .

To form the sigma-point Kalman filter, we consider the nonlinear state-space model:

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{f}_k(\mathbf{x}_k, \mathbf{v}_k) \\ \mathbf{z}_k &= \mathbf{h}_k(\mathbf{x}_k, \mathbf{n}_k) \end{aligned} \tag{9}$$

where  $\mathbf{x}_k \in \mathbb{R}^M$  is the state,  $\mathbf{z}_k \in \mathbb{R}^P$  is the measurement at time k,  $\mathbf{v}_k$  and  $\mathbf{n}_k$  are the gaussian-distributed process and observation noises,  $\mathbf{f}(.)$  is the dynamic model function and  $\mathbf{h}(.)$  is the observation model function. The process and observation noise has zero mean and covariances  $\mathbf{Q}_k$  and  $\mathbf{R}_k$  respectively. The SPKF is then derived by recursively applying the sigma point selection scheme to these dynamic equations (see [6], [10] for details).

Alternatively, we can form the statistically linearized statespace:

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{A}_{f,k} \mathbf{x}_k + \mathbf{b}_{f,k} + \mathbf{G}_f \mathbf{v}_k + \mathbf{G}_f \mathbf{\varepsilon}_{f,k} \\ \mathbf{z}_k &= \mathbf{A}_{h,k} \mathbf{x}_k + \mathbf{b}_{h,k} + \mathbf{n}_k + \mathbf{\varepsilon}_{h,k} \end{aligned} \tag{10}$$

where  $A_{f,k}$ ,  $A_{h,k}$ ,  $b_{f,k}$ ,  $b_{h,k}$  are the statistical linearization parameters

and  $\boldsymbol{\varepsilon}_{f,k}$ ,  $\boldsymbol{\varepsilon}_{h,k}$  are the linearization error with mean zero and covariances  $\boldsymbol{P}_{\varepsilon_{f,k}}$ , and  $\boldsymbol{P}_{\varepsilon_{h,k}}$ . Deriving the Kalman filter using this linearized state-space also leads to the SPKF (see again [1] for details). However, the advantage of the statistically linearized form is that it allows for the formulation of the backward information filter necessary for smoothing, as detailed in the next section.

# 3. FORWARD-BACKWARD STATISTICAL LINEARIZED SIGMA-POINT KALMAN SMOOTHER (FBSL-SPKS)

### **3.1 Forward Filter**

A standard SPKF is used as the forward filter with the nonlinear state-space model shown in (9). The forward filter estimates the state  $x_k$  at time index k given all the past and present measurements. The pseudo code for the SPKF with WSLR is shown below:

Initialization:

$$\hat{\boldsymbol{x}}_{0} = E[\boldsymbol{x}_{0}], \boldsymbol{P}_{\boldsymbol{x}_{0}} = E\left[\left(\boldsymbol{x}_{0} - \hat{\boldsymbol{x}}_{0}\right)\left(\boldsymbol{x}_{0} - \hat{\boldsymbol{x}}_{0}\right)^{T}\right]$$

$$\hat{\boldsymbol{x}}_{0}^{a} = E\left[\boldsymbol{x}_{0}^{a}\right] = \left[\hat{\boldsymbol{x}}_{0}^{T} \ \hat{\boldsymbol{v}}_{0}^{T} \ \hat{\boldsymbol{n}}_{0}^{T}\right]^{T}$$

$$P_{0}^{a} = E\left[\left(\boldsymbol{x}_{0}^{a} - \hat{\boldsymbol{x}}_{0}^{a}\right)\left(\boldsymbol{x}_{0}^{a} - \hat{\boldsymbol{x}}_{0}^{a}\right)^{T}\right] = \begin{bmatrix}\boldsymbol{P}_{\boldsymbol{x}_{0}} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{R}_{0} \end{bmatrix}.$$

$$(11)$$

For  $k = 1 \dots N$ *1.* Calc

Calculate sigma-points:  

$$\boldsymbol{\chi}_{k}^{a} = \begin{bmatrix} \hat{\boldsymbol{x}}_{k}^{a} & \hat{\boldsymbol{x}}_{k}^{a} + \sqrt{(L+\lambda)\boldsymbol{P}_{k}^{a}} & \hat{\boldsymbol{x}}_{k}^{a} - \sqrt{(L+\lambda)\boldsymbol{P}_{k}^{a}} \end{bmatrix}$$
(12)

where  $\boldsymbol{\chi}^{a} = \left[ \left( \boldsymbol{\chi}^{x} \right)^{T} \left( \boldsymbol{\chi}^{v} \right)^{T} \left( \boldsymbol{\chi}^{n} \right)^{T} \right]^{T}$  is the prior sigma point set. 2. *Time-update equations:* 

$$\chi_{i,k+1|k}^{x} = f_{k} \left( \chi_{i,k}^{x}, \chi_{i,k}^{v} \right) \quad i = 0, 1..., 2L$$

$$\hat{\chi}_{k+1}^{-} = \sum_{i=0}^{2L} w_{i}^{m} \chi_{i,k+1|k}^{x} \qquad (13)$$

$$P_{\chi_{k+1}}^{-} = \sum_{i=0}^{2L} \sum_{j=0}^{2L} w_{ij}^{c} \left( \chi_{i,k+1|k}^{x} - \hat{\chi}_{k+1}^{-} \right) \left( \chi_{j,k+1|k}^{x} - \hat{\chi}_{k+1}^{-} \right)^{T}.$$
3. Weighted Statistical linearization of  $f(.)$ :

$$\boldsymbol{P}_{x_{k}x_{k+1}^{-}} = \sum_{i=0}^{2L} \sum_{j=0}^{2L} w_{ij}^{c} (\boldsymbol{\chi}_{j,k}^{x} - \hat{\boldsymbol{x}}_{k}) (\boldsymbol{\chi}_{i,k+1|k}^{x} - \hat{\boldsymbol{x}}_{k+1}^{-})^{T}$$

$$\boldsymbol{A}_{f,k} = \boldsymbol{P}_{x_{k}x_{k+1}}^{T} \boldsymbol{P}_{x_{k}}^{-1}$$

$$\boldsymbol{b}_{f,k} = \hat{\boldsymbol{x}}_{k+1}^{-} - \boldsymbol{A}_{f,k} \hat{\boldsymbol{x}}_{k}$$

$$\boldsymbol{P}_{\varepsilon_{f,k}} = \boldsymbol{P}_{x_{k+1}}^{-} - \boldsymbol{A}_{f,k} \boldsymbol{P}_{x_{k}} \boldsymbol{A}_{f,k}^{T} .$$

$$(14)$$

4. Measurement update equations:  

$$\begin{aligned} \boldsymbol{\gamma}_{i,k+1|k} &= \boldsymbol{h}_{k} \left( \boldsymbol{\chi}_{i,k+1|k}^{x}, \boldsymbol{\chi}_{i,k}^{n} \right) & i = 0, 1..., 2L \\ \hat{\boldsymbol{z}}_{k+1}^{-} &= \sum_{i=0}^{2L} w_{i}^{m} \boldsymbol{\gamma}_{i,k+1|k} \\ \boldsymbol{P}_{\tilde{\boldsymbol{z}}_{k+1}\tilde{\boldsymbol{z}}_{k+1}} &= \sum_{i=0}^{2L} \sum_{j=0}^{2L} w_{ij}^{c} \left( \boldsymbol{\gamma}_{i,k+1|k} - \hat{\boldsymbol{z}}_{k+1}^{-} \right) \left( \boldsymbol{\gamma}_{j,k+1|k} - \hat{\boldsymbol{z}}_{k+1}^{-} \right)^{T} \\ \boldsymbol{P}_{\boldsymbol{x}_{k+1}\tilde{\boldsymbol{z}}_{k+1}} &= \sum_{i=0}^{2L} \sum_{j=0}^{2L} w_{ij}^{c} \left( \boldsymbol{\chi}_{i,k+1|k}^{x} - \hat{\boldsymbol{x}}_{k+1}^{-} \right) \left( \boldsymbol{\gamma}_{j,k+1|k} - \hat{\boldsymbol{z}}_{k+1}^{-} \right)^{T} \end{aligned}$$
(15)

$$K_{k+1} = P_{x_{k+1}z_{k+1}} P_{\bar{z}_{k+1}\bar{z}_{k+1}}^{-1}$$

$$\hat{x}_{k+1} = \hat{x}_{k+1}^{-} + K_{k+1} \Big( z_{k+1} - \hat{z}_{k+1}^{-} \Big)$$

$$P_{x_{k+1}} = P_{x_{k+1}}^{-} - K_{k+1} P_{\bar{z}_{k+1}\bar{z}_{k+1}} K_{k+1}^{T}$$
(16)

Weighted Statistical linearization of h(.): 5.

$$A_{h,k} = P_{x_{k+1}\bar{z}_{k+1}}^{T} \left( P_{x_{k+1}}^{-} \right)^{-}$$
  

$$b_{h,k} = \hat{z}_{k+1}^{-} - A_{h,k} \hat{x}_{k+1}^{-}$$
  

$$P_{\varepsilon_{h,k}} = P_{\bar{z}_{k+1}\bar{z}_{k+1}}^{-} - A_{h,k} P_{x_{k+1}}^{-} A_{h,k}^{T} .$$
(17)

*Parameters*:  $\lambda$ ,  $w_i^{\text{m}}$ , and  $w_{ii}^{\text{c}}$  are scalar weights as defined in [1]. L is the dimension of the augmented state. Here we assumed that the length of the observation sequence is N.

#### 3.2 Backward Filter

An information filter is used to estimate the states from the backward direction with the statistically linearized state-space shown in (10). As the statistically linearized state-space is different than the standard linear state-space used by the Kalman filter [3], [11], we must derive the time update and measurement update equations from first principles.

 $S_{N}^{-}=0, y_{N}^{-}=0$ (18)where the information matrix is denoted as the inverse of the state error covariance *i.e.*,  $S_k^- = (P_k^-)^{-1}$  and  $\hat{y}_k^- = S_k^- \hat{x}_k^-$  is defined as information state.

Time update for information matrix: 2 From (10),

$$\hat{\mathbf{x}}_{k}^{-} = A_{f,k}^{-1} \left( \hat{\mathbf{x}}_{k+1} - \boldsymbol{b}_{f,k} \right)$$

$$P_{\mathbf{x}_{k}}^{-} = P_{k}^{-} = E \left[ \left( \mathbf{x}_{k} - \hat{\mathbf{x}}_{k}^{-} \right) \left( \mathbf{x}_{k} - \hat{\mathbf{x}}_{k}^{-} \right)^{T} \right].$$
Assuming  $\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1} = \mathbf{e}_{k+1}$ 

$$P_{k}^{-} = E \begin{bmatrix} \left( A_{f,k}^{-1} \left( \mathbf{e}_{k+1} \right) - A_{f,k}^{-1} \left( G_{f} \mathbf{v}_{k} + G_{f} \mathbf{\varepsilon}_{f,k} \right) \right) \\ \left( A_{f,k}^{-1} \left( \mathbf{e}_{k+1} \right) - A_{f,k}^{-1} \left( G_{f} \mathbf{v}_{k} + G_{f} \mathbf{\varepsilon}_{f,k} \right) \right)^{T} \end{bmatrix}$$

$$= A_{f,k}^{-1} P_{k+1} A_{f,k}^{-T} + A_{f,k}^{-1} G_{f} (P_{\epsilon_{f,k}} + Q_{k}) G_{f}^{T} A_{f,k}^{-T}$$
(19)

where  $P_{\varepsilon_{f,k}} = E \left[ \boldsymbol{\varepsilon}_{f,k} \boldsymbol{\varepsilon}_{f,k}^T \right]$  is the covariance of the linearization error,  $P_{k+1} = E \left[ e_{k+1} e_{k+1}^T \right]$  is the error covariance of the state prediction and  $Q_k = E \left[ \mathbf{v}_k \mathbf{v}_k^T \right]$  is the process noise covariance. Apply

$$\mathbf{S}_{k} = \mathbf{A}_{f,k}^{t} \, \mathbf{S}_{k+1} \mathbf{A}_{f,k} - \mathbf{A}_{f,k}^{t} \, \mathbf{S}_{k+1} \mathbf{G}_{f}.$$

 $\left[ \left( \boldsymbol{P}_{\varepsilon_{f},k} + \boldsymbol{Q}_{k} \right)^{-1} + \boldsymbol{G}_{f}^{T} \boldsymbol{S}_{k+1} \boldsymbol{G}_{f} \right]^{-1} \boldsymbol{G}_{f}^{T} \boldsymbol{S}_{k+1} \boldsymbol{A}_{f,k} .$ 

Define  $K_{b,k}$  as the backward gain matrix,

K

$$\boldsymbol{K}_{b,k} = \boldsymbol{S}_{k+1} \boldsymbol{G}_{f} \left[ (\boldsymbol{P}_{\varepsilon_{f},k} + \boldsymbol{Q}_{k})^{-1} + \boldsymbol{G}_{f}^{T} \boldsymbol{S}_{k+1} \boldsymbol{G}_{f} \right]^{\top}.$$
 (21)  
Putting (21) into (20) we obtain,

$$\boldsymbol{S}_{k}^{-} = \boldsymbol{A}_{f,k}^{T} \left[ \boldsymbol{I} - \boldsymbol{K}_{b,k} \boldsymbol{G}_{f}^{T} \right] \boldsymbol{S}_{k+1} \boldsymbol{A}_{f,k} .$$
(22)

**-**--1

Note the presence of the correction term  $P_{\varepsilon_{c,k}}$  in the right hand side of (20) and (21). This term does not appear in the standard information filter formulation [3], [11]. The more severe the nonlinearity is over the uncertainty region of the state, the higher will be the linearization error covariance matrix  $P_{\varepsilon_{\epsilon,k}}$ . A firstorder Taylor series expansion is less accurate because it does not consider this error term.

3. Time update for information state:  

$$\hat{x}_{k}^{-} = A_{f,k}^{-1} (\hat{x}_{k+1} - b_{f,k})$$

$$\hat{y}_{k}^{-} = S_{k}^{-} \hat{x}_{k}^{-}$$

$$= S_{k}^{-} A_{f,k}^{-1} (S_{k+1}^{-1} \hat{y}_{k+1} - b_{f,k}).$$
(23)

Now putting the value of  $S_k^-$  from (22) into (23), we obtain:

$$\hat{\boldsymbol{y}}_{k}^{-} = \boldsymbol{A}_{f,k}^{T} \left( \boldsymbol{I} - \boldsymbol{K}_{b,k} \boldsymbol{G}_{f}^{T} \right) \left( \hat{\boldsymbol{y}}_{k+1} - \boldsymbol{S}_{k+1} \boldsymbol{b}_{f,k} \right).$$
(24)

Note the correction term  $S_{k+1}b_{f,k}$  is subtracted out from the previous information state. Again, this term is not present in the standard information filter.

## 4. Measurement update: By applying (19)-(24):

$$S_{k} = S_{k}^{-} + A_{h,k}^{T} (\boldsymbol{P}_{\varepsilon_{h},k} + \boldsymbol{R}_{k})^{-1} A_{h,k}$$
  

$$\hat{\boldsymbol{y}}_{k} = \hat{\boldsymbol{y}}_{k}^{-} + A_{h,k}^{T} (\boldsymbol{P}_{\varepsilon_{h},k} + \boldsymbol{R}_{k})^{-1} (\boldsymbol{z}_{k} - \boldsymbol{b}_{h,k})$$
(25)

where  $z_k$  is the measurement at time k. Details of this last derivation are omitted due to space limitations.

## 3.3 Smoothing

The SPKF is run in the forward direction on the interval [0, N] to compute the forward estimates  $\hat{\boldsymbol{x}}_k^{\mathrm{f}}$ . The information filter is then run backwards to compute the backward estimates  $\hat{x}_{k}^{\mathrm{b-}}$ . Note that the forward estimates are posterior estimates (after the measurement), whereas the backward estimates are prior estimates (after the time update). The smoother optimally combines the forward and backward estimates and their respective error covariances to determine an overall estimate  $\hat{x}_{i}$  and corresponding error covariance. Specifically,

$$\boldsymbol{P}_{k} = \left[ \left( \boldsymbol{P}_{k}^{f} \right)^{-1} + \boldsymbol{S}_{k}^{-} \right]^{-1}$$

$$\hat{\boldsymbol{x}}_{k} = \left( \boldsymbol{I} + \boldsymbol{P}_{k}^{f} \boldsymbol{S}_{k}^{-} \right)^{-1} \hat{\boldsymbol{x}}_{k}^{f} + \boldsymbol{P}_{k} \hat{\boldsymbol{y}}_{k}^{-} .$$
(26)

These estimates now incorporate all measurement data  $z_{\theta:N}$ .

# 4. EXPERIMENTAL RESULTS

We evaluate the proposed FBSL smoother to estimate an underlying clean Mackey-Glass-30 chaotic time series corrupted by additive Gaussian white noise (SNR = 3 db). The clean and noisy time series are shown in Fig. 1. The clean time series was modeled as nonlinear autoregression f using a 6-5-1 (input-hiddenoutput nodes) feed forward neural network as in [1]. The corresponding state-space equations are shown in (27).

(20)

$$\mathbf{x}_{k} = \begin{bmatrix} x_{k+1} \\ x_{k} \\ \vdots \\ x_{k-M} \end{bmatrix} = \begin{bmatrix} f(x_{k}, x_{k-1}, \cdots, x_{k-M+1}) \\ 1 & 0 \cdots & 0 \\ 0 & 1 \cdots & 0 \\ \vdots \\ 0 \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} x_{k} \\ x_{k-1} \\ \vdots \\ x_{k-M+1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} v_{k}$$
(27)  
$$z_{k} = \begin{bmatrix} 1 & 0 \cdots & 0 \end{bmatrix} \mathbf{x}_{k} + n.$$

Fig. 2 shows a segment of the FBSL-SPKS estimates compared to the standard SPKF. Fig. 3 displays the squared error between the true and estimated time series. The superior performance of the proposed SPKS over the SPKF is clearly visible. Table 1 summarizes performance of the different filters in terms of mean of the MSE and standard deviation of the MSE for a Monte-Carlo run of 100 randomly initialized experiments. Large errors in the EKF and EKS estimates are due to the inaccurate linearizationbased approach. As seen in Table 1, both sigma-point Kalman smoothers perform comparably. Whereas the FBNL-SPKS requires a separate learned model of the nonlinear backward dynamics, the new FBSL-SPKS formulates the backward information filter directly using the WSLR technique during the forward pass.



Fig. 1: Clean and noisy time series data



Fig. 2: Estimated time series using SPKF and FBSL-SPKS



Fig. 3: Squared error between true and estimated time series

Table 1: Mean and standard deviations over 100 Monte-Carlo runs

Estimator	Mean(MSE)	Std(MSE)
EKF	1.20	0.25
EKS	0.725	0.18
SPKF	0.236	0.05
FBNL-SPKS	0.106	0.02
FBSL-SPKS	0.098	0.02

# **5. CONCLUSION**

A new smoother formulation for nonlinear state-space models has been derived. The FBSL-SPKS uses a forward-backward approach. The forward filter is an SPKF, which operates on the original nonlinear state-space. The backward filter is an information filter, which uses a linear state-space derived using the WSLR technique. The estimates of the two filters are then optimally combined to generate the smoothed estimate. The performance of the FBSL-SPKS has been demonstrated and compared with other approaches. It was also shown that the FBSL-SPKS performs comparably to the original FBNL-SPKS, with the advantage of not requiring an extra step to learn the backward nonlinear dynamics. Using similar WSLR techniques, we can also derive a new Rauch-Tung-Striebel (RTS) type of smoother. This RTS smoother differs from the form in [9], and will be detailed in a future publication.

## 6. REFERENCES

[1] R. van der Merwe, "Sigma-Point Kalman filters for probabilistic inference in dynamic state-space models," *PhD thesis*, OGI School of Science and Engineering at OHSU, Beaverton, OR, April 2004.

[2] H. E. Rauch, F. Tung and C.T. Striebel, "Maximum likelihood estimates of linear dynamic systems," *AIAA Journal*, vol. 3(8), pp. 1445-1450, 1965.

[3] F. L. Lewis, *Optimal estimation with an introduction to stochastic control theory*. John Wiley & Sons, 1986.

[4] D. Fraser and J. Potter, "The optimum linear smoother as a combination of two optimum linear filters," *IEEE Transactions on Automatic Control*, vol. 14(4), pp. 387-390, 1969.

[5] C. T. Leondes, J.B. Peller and E.B. Stear, "Nonlinear smoothing theory," *IEEE Transactions on System, Science and Cybernetics*, vol. 6(1), January 1970.

[6] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," In *Proceedings of the IEEE*, vol. 92, pp. 401-422, March 2004.

[7] M. Norgaard, N. Poulsen, and O. Ravn, "New developments in state estimation for nonlinear systems," *Automatica*, vol. 36, pp. 1627-1638, November 2000.

[8] R. van der Merwe and E. Wan, "The square-root unscented kalman filter for state and parameter estimation," In Proceedings of the *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Salt Lake City, UT, May 2001, vol. 6, pp. 3461-3464,.

[9] S. Sarkka, "Unscented Rauch-Tung-Striebel smoother," Accepted for publications in *IEEE Transactions on Automatic Control.* 

[10] R. van der Merwe, "Sigma-Point Kalman filters for probabilistic inference in dynamic state space models," in Workshop on Advances in Machine Learning, June 2003.

[11] D. Simon, *Optimal state estimation kalman*  $H\infty$  and nonlinear approaches. Wiley-Interscience, 2006.