NORTHEASTERN UNIVERSITY
Department of Electrical and Computer Engineering

2019 PhD Qualifying Examination

in

Communications, Control, and Signal Processing

Mahdiar Sadeghi

Problem 4

03/11/2019

Problem 4: Nonlinear Dynamic State Estimation

All solutions in this report are based on a time-invariant nonlinear state space model of the form (1). The $f(.)$ and $h(.)$ are multivariate functions of the state space x_k, and additive noises $\{w_k\}$, $\{\nu_k\}$ are i.i.d processes independent of each other of the initial state x_0. Fig. 1 is an illustration of this model.

$$\begin{aligned} x_{k+1} &= f(x_k) + w_k, \\ y_k &= h(x_k) + v_k. \end{aligned} \qquad 0 \leq k \tag{1}$$

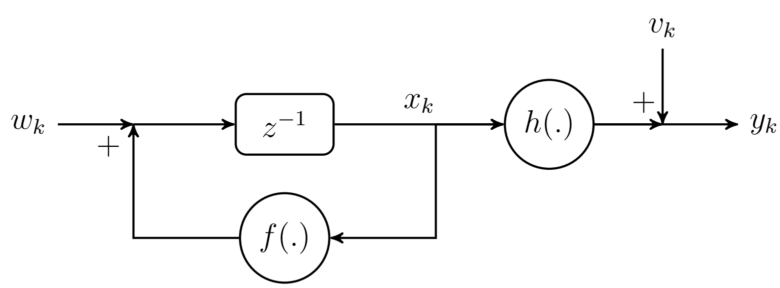

Figure 1: Block diagram of the nonlinear state space model (1).

Part A

Discussion of the *Extendted Kalman Filter* (EKF).

1) The linearized version of the nonlinear state-space model (1) subject to EKF is obtained by a first order Taylor expansion:

$$\begin{aligned} x_{k+1} &= f(\hat{x}_{k|k}) + F(\hat{x}_{k|k})(x_k - \hat{x}_{k|k}) + w_k, \\ y_k &= h(\hat{x}_{k|k-1}) + H(\hat{x}_{k|k-1})(x_k - \hat{x}_{k|k-1}) + v_k. \end{aligned} \tag{2}$$

   The $F(\hat{x}_{k|k}) \triangleq \frac{\partial f}{\partial x}\big|_{x=\hat{x}_{k|k}}$ and $H(\hat{x}_{k|k-1}) \triangleq \frac{\partial h}{\partial x}\big|_{x=\hat{x}_{k|k-1}}$ are Jacobian matrices. Both matrices are (i) time-invariant, and (ii) stochastic. EKF does not take into account that x_k is a random variable with inherent uncertainty, and this is true when the first two terms of the Taylor series are dominating the remaining terms.

2) The idea of EKF is to assume quasi-linear behavior for a nonlinear system. Assuming that the process noise w_k and the measurement noise v_k are zero mean with covariance W_k and V_k, and the initial state is x_0. The time-update (3) and measurement-update (4) are the corresponding EKF explicit recursions [1, 2] to the model (2).

$$\hat{x}_{k+1|k} = f(\hat{x}_{k|k}), \tag{3a}$$
$$P_{k+1|k} = F(\hat{x}_{k|k}) P_{k|k} F^T(\hat{x}_{k|k}) + W_k. \tag{3b}$$

$$P_{yy,k|k-1} = H(\hat{x}_{k|k-1})P_{k|k-1}H^T(\hat{x}_{k|k-1}) + V_k, \tag{4a}$$

$$P_{xy,k|k-1} = P_{k|k-1}H^T(\hat{x}_{k|k-1}), \tag{4b}$$

$$K_k = P_{xy,k|k-1}P^{-1}_{yy,k|k-1}, \tag{4c}$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k(y_k - h(\hat{x}_{k|k-1})), \tag{4d}$$

$$P_{k|k} = P_{k|k-1} - K_k P_{yy,k|k-1}K_k^T. \tag{4e}$$

Higher order EKF is defined based on second order Taylor expansion using Jacobian and Hessian matrices. EKF analysis here are based on (3) and (4).

3) Block diagram of EKF is represented in Fig. 2. The dashed blue rectangular is the time-update module. It takes the process noise variance W_k, the mean $\hat{x}_{k-1|k-1}$ and covariance $P_{k-1|k-1}$ of the previous estimation, then calculates predicts the mean $\hat{x}_{k|k-1}$ and covariance $P_{k|k-1}$ of the next state using Eq. (3). The dashed red rectangular is the measurement-update module. It takes the output of time-step along with the new measurement y_k, and the measurement noise variance V_k, then calculates the mean $\hat{x}_{k|k}$ and covariance $P_{k|k}$ of the next system state using Eq. (4).

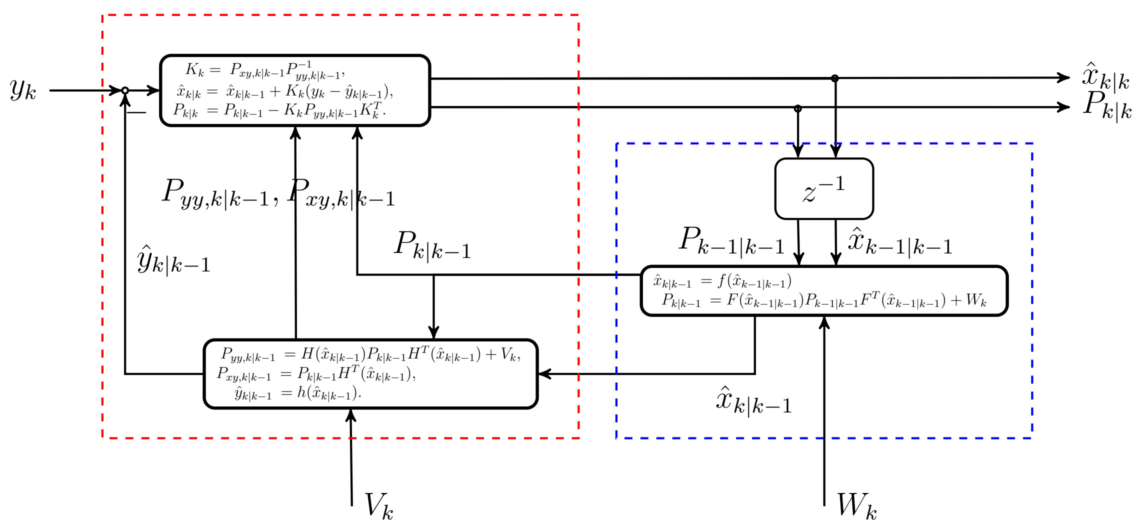

Figure 2: EKF block diagram. Time-update and measurement-update blocks are represented by blue and red dashed rectangular.

The measurement variable directly influences the estimated state mean $\hat{x}_{k|k}$ and maximum covariance $P_{k|k}$, therefore all variables are dependent of the measurement recursively, except the measurement noise v_k and process noise w_k.

The system will be independent of measurement in case of high measurement noise covariance V_k. In this case Kalman gain K_k goes to zero, and the estimated state will depend on predicted state $\hat{x}_{k|k}$ only. It is obvious that a measurement with high uncertainty is not valuable and KF reduces the weight of the measured (observed) value in consequence.

4) Kalman's original derivation did not apply Bayes' rule and does not require the exploitation of any specific error distribution information beyond the mean and covariance [3]. However, the filter yields the exact conditional probability estimate in the special case that all errors are Gaussian. KF can be derived as an application of Bayes' rule under the assumption that all estimates have independent, Gaussian distributed errors [4].

5) The innovation process can be rewritten as $y_k - \hat{y}_{k|k-1}$. It is the difference between the actual measurement y_k and its estimated prediction, based on the system model and previous measurement $\hat{y}_{k|k-1}$. The first two moments of the innovation signal is known. The mean is zero. The covariance matrix of innovation based on the linearized model is given by (4a). If the noise assumed to be Gaussian, then the innovation sequence is zero-mean, white (uncorrelated), with covariance equal to the measurement prediction covariance, and has a Gaussian distribution.

Part B

Discussion of the *Cubature Kalman Filter* (CKF).

1) *Bayesian filtering* aims to compute the posterior density $p(x_k|y_{1:k})$ of the state x_k at each time step k, given the history of the measurement up to the time step k and the prior density $p(x_{k-1}|y_{1:k-1})$ [2]. The recursions start from the initial distribution $p(x_0)$, then applying two step recursive algorithm: times-update (5) which is derived by *Chapman-Kolmogorov equation*, and measurement-step (6) which is derived by *Bayes' rule.*

$$p(x_k|y_{1:k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|y_{1:k-1})\,\mathrm{d}x_{k-1} \tag{5}$$

$$p(x_k|y_{1:k}) = \frac{1}{c_k}p(y_k|x_k)p(x_k|y_{1:k-1}) \tag{6a}$$

$$c_k = \int p(y_k|x_k)p(x_k|y_{1:k-1})\,\mathrm{d}x_k \tag{6b}$$

The integrals of Eq. (5) and (6b) need to be computed in the general Bayesian filtering. However, from practical perspective they are intractable [5]. Notable exceptions in which optimal solution is tractable are including: Linear-Gaussian dynamic system (KF [3]), discrete-valued state-space with a fixed number of states (Hidden-Markov model filter [6]), and *Benes type* of non-linearity [7].

2) For the case of Gaussian state-spaces, the predictive PDF $p(x_k|y_{1:k-1})$ and the filter likelihood PDF $p(x_k|y_{1:k})$ are both assumed to be Gaussian. In the time-update step, the $p(x_k|y_{1:k-1})$ is $\mathcal{N}(\hat{x}_{k|k-1}, P_{k|k-1})$. Where the mean and covariance are defined by

the statistical expectation operator $\mathbb{E}[.]$ as following.

$$
\begin{aligned}
\hat{x}_{k|k-1} &= \mathbb{E}[x_k|y_{1:k-1}] = \mathbb{E}[f(x_{k-1}) + w_k|y_{1:k-1}] && \text{(7a)}\\
&= \mathbb{E}[f(x_{k-1})|y_{1:k-1}] && \text{(7b)}\\
&= \int f(x_{k-1})p(x_{k-1}|y_{1:k-1})\,\mathrm{d}x_{k-1} && \text{(7c)}\\
&= \int f(x_{k-1}) \times \mathcal{N}(x_{k-1}|\hat{x}_{k-1|k-1}, P_{k-1|k-1})\,\mathrm{d}x_{k-1} && \text{(7d)}
\end{aligned}
$$

It is possible to write (7a) from (7b) by assuming that the process noise $w_k \sim \mathcal{N}(0, W_k)$ is zero-mean and uncorrelated with the past measurements.

$$
\begin{aligned}
P_{k|k-1} &= \mathbb{E}[(x_k - \hat{x}_{k|k-1})(x_k - \hat{x}_{k|k-1})^T|y_{1:k-1}] && \text{(8a)}\\
&= \mathbb{E}[(f(x_{k-1}) + w_{k-1} - \hat{x}_{k|k-1})(f(x_{k-1}) + w_{k-1} - \hat{x}_{k|k-1})^T|y_{1:k-1}] && \text{(8b)}\\
&= \mathbb{E}[f(x_{k-1})f^T(x_{k-1}) + f(x_{k-1})w_{k-1}^T - f(x_{k-1})\hat{x}_{k|k-1}^T \\
&\qquad + w_{k-1}f^T(x_{k-1}) + w_{k-1}w_{k-1}^T - w_{k-1}\hat{x}_{k|k-1}^T && \text{(8c)}\\
&\qquad\qquad - \hat{x}_{k|k-1}f^T(x_{k-1}) - \hat{x}_{k|k-1}w_{k-1}^T + \hat{x}_{k|k-1}\hat{x}_{k|k-1}^T|y_{1:k-1}] \\
&= \int f(x_{k-1})f^T(x_{k-1}) \times \mathcal{N}(x_{k-1}|\hat{x}_{k-1|k-1}, P_{k-1|k-1})\,\mathrm{d}x_{k-1} \\
&\qquad - \hat{x}_{k-1|k-1}\hat{x}_{k-1|k-1}^T + W_{k-1} && \text{(8d)}
\end{aligned}
$$

In the measurement-update step, the $p(y_k|y_{1:k-1})$ is $\mathcal{N}(\hat{y}_{k|k-1}, P_{yy,k|k-1})$.

$$
\begin{aligned}
\hat{y}_{k|k-1} &= \mathbb{E}[y_k|y_{1:k-1}] = \mathbb{E}[h(x_k) + v_k|y_{1:k-1}] && \text{(9a)}\\
&= \mathbb{E}[h(x_k)|y_{1:k-1}] && \text{(9b)}\\
&= \int h(x_k) \times \mathcal{N}(x_k|\hat{x}_{k|k-1}, P_{k|k-1})\,\mathrm{d}x_k && \text{(9c)}
\end{aligned}
$$

$$
\begin{aligned}
P_{yy,k|k-1} &= \mathbb{E}[y_k - \hat{y}_{k|k-1})(y_k - \hat{y}_{k|k-1})^T|y_{1:k-1}] && \text{(10a)}\\
&= \mathbb{E}[(h(x_k) + v_k - \hat{y}_{k|k-1})(h(x_k) + v_k - \hat{y}_{k|k-1})^T|y_{1:k-1}] && \text{(10b)}\\
&= \int h(x_k)h^T(x_k) \times \mathcal{N}(x_k|\hat{x}_{k|k-1}, P_{k|k-1})\,\mathrm{d}x_k - \hat{y}_{k|k-1}\hat{y}_{k|k-1}^T + V_k && \text{(10c)}
\end{aligned}
$$

$$
\begin{aligned}
P_{xy,k|k-1} &= \mathbb{E}[(x_k - \hat{x}_{k|k-1})(y_k - \hat{y}_{k|k-1})^T|y_{1:k-1}] && \text{(11a)}\\
&= \mathbb{E}[(x_k - \hat{x}_{k|k-1})(h(x_k) + v_k - \hat{y}_{k|k-1})^T|y_{1:k-1}] && \text{(11b)}\\
&= \mathbb{E}[x_k h^T(x_k) + x_k v_k^T - x_k\hat{y}_{k|k-1}^T - \hat{x}_{k|k-1}h^T(x_k) \\
&\qquad - \hat{x}_{k|k-1}v_k^T + \hat{x}_{k|k-1}\hat{y}_{k|k-1}^T|y_{1:k-1}] && \text{(11c)}\\
&= \int x_k h^T(x_k) \times \mathcal{N}(x_k|\hat{x}_{k|k-1}, P_{k|k-1})\,\mathrm{d}x_k - \hat{x}_{k|k-1}\hat{y}_{k|k-1}^T && \text{(11d)}
\end{aligned}
$$

Therefore, the conditional Gaussian density of the joint state and the measurement is:

$$p\left(\left[\begin{array}{c} x_k \\ y_k \end{array}\right]\middle| y_{1:k-1}\right) = \mathcal{N}\left(\left(\begin{array}{c} \hat{x}_{k|k-1} \\ \hat{y}_{k|k-1} \end{array}\right), \left(\begin{array}{cc} P_{k|k-1} & P_{xy,k|k-1} \\ P^T_{xy,k|k-1} & P_{yy,k|k-1} \end{array}\right)\right) \tag{12}$$

So, the Bayesian filter computes the posterior density $p(x_k|y_k)$ from Eq. (12) as:

$$p(x_k|y_k) = \mathcal{N}(x_k|\hat{x}_{k|k}, P_{k|k}) \tag{13}$$

Upon receiving the new measurement y_k, the KF equations will be used to find the posterior density mean $\hat{x}_{k|k}$ and convariance $P_{k|k}$:

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k(y_k - \hat{y}_{k|k-1}), \tag{14a}$$

$$P_{k|k} = P_{k|k-1} - K_k P_{yy,k|k-1} K_k^T, \tag{14b}$$

$$K_k = P_{xy,k|k-1} P^{-1}_{yy,k|k-1}. \tag{14c}$$

The main challenge and the heart of the Bayesian filter is to compute Gaussian weighted integrals of (7d), (8d), (9c), (10c), and (11d). They are all in a combination of non-linear function and a Gaussian PDF. It is not easy to derive this integral analytically for general non-linear state space model, for a linear state space model:

$$x_{k+1} = F_k x_k + w_k, \tag{15a}$$

$$y_k = H_k x_k + v_k. \tag{15b}$$

The time-update step derivations can be rewritten as:

$$\begin{aligned} \hat{x}_{k|k-1} &= \mathbb{E}[F_{k-1}x_{k-1} + w_{k-1}|y_{1:k-1}] = F_{k-1}\mathbb{E}[x_{k-1}|y_{1:k-1}] \\ &= F_{k-1}\hat{x}_{k-1|k-1} \end{aligned} \tag{16}$$

$$\begin{aligned} P_{k|k-1} &= \mathbb{E}[(x_k - \hat{x}_{k|k-1})(x_k - \hat{x}_{k|k-1})^T|y_{1:k-1}] \\ &= \mathbb{E}[(F_{k-1}x_{k-1} + w_{k-1} - F_{k-1}\hat{x}_{k-1|k-1})(F_{k-1}x_{k-1} + w_{k-1} - F_{k-1}\hat{x}_{k-1|k-1})^T|y_{1:k-1}] \\ &= \mathbb{E}[(F_{k-1}x_{k-1})(F_{k-1}x_{k-1})^T + (F_{k-1}x_{k-1})w_{k-1}^T - (F_{k-1}x_{k-1})(F_{k-1}\hat{x}_{k-1})^T \\ &\quad + w_{k-1}(F_{k-1}x_{k-1})^T + w_{k-1}w_{k-1}^T - w_{k-1}(F_{k-1}\hat{x}_{k-1|k-1})^T \\ &\quad + (F_{k-1}\hat{x}_{k-1|k-1})(F_{k-1}x_{k-1})^T + (F_{k-1}\hat{x}_{k-1|k-1})w_{k-1}^T \\ &\quad - (F_{k-1}\hat{x}_{k-1|k-1})(F_{k-1}\hat{x}_{k-1|k-1})^T|y_{1:k-1}] \\ &= \mathbb{E}[(F_{k-1}x_{k-1} - F_{k-1}\hat{x}_{k-1|k-1})(F_{k-1}x_{k-1} - F_{k-1}\hat{x}_{k-1|k-1})^T + w_{k-1}w_{k-1}^T|y_{1:k-1}] \\ &= F_{k-1}\mathbb{E}[(x_{k-1} - \hat{x}_{k-1|k-1})(x_{k-1} - \hat{x}_{k-1|k-1})^T|y_{1:k-1}]F_{k-1}^T + \mathbb{E}[w_{k-1}w_{k-1}^T|y_{1:k-1}] \\ &= F_{k-1}P_{k-1|k-1}F_{k-1}^T + W_{k-1} \end{aligned} \tag{17}$$

$$\begin{aligned} \hat{y}_{k|k-1} &= \mathbb{E}[H_k x_k + v_k|y_{1:k-1}] = H_k\mathbb{E}[x_k|y_{1:k-1}] \\ &= H_k\hat{x}_{k|k-1} \end{aligned} \tag{18}$$

And similar to derivation of (17), the measurement-update step can be rewritten as:

$$\begin{aligned}\hat{P}_{yy,k|k-1} &= \mathbb{E}[(y_k - \hat{y}_{k|k-1})(y_k - \hat{y}_{k|k-1})^T | y_{1:k-1}] \\ &= \mathbb{E}[(H_k x_k + v_k - H_k \hat{x}_{k|k-1})(H_k x_k + v_k - H_k \hat{x}_{k|k-1})^T | y_{1:k-1}] \\ &= H_k P_{k|k-1} H_k^T + V_k\end{aligned} \tag{19}$$

$$\begin{aligned}\hat{P}_{xy,k|k-1} &= \mathbb{E}[(x_k - \hat{x}_{k|k-1})(y_k - \hat{y}_{k|k-1})^T | y_{1:k-1}] \\ &= \mathbb{E}[(x_k - \hat{x}_{k|k-1})(H_k x_k + v_k - H_k \hat{x}_{k|k-1})^T | y_{1:k-1}] \\ &= \mathbb{E}[(x_k - \hat{x}_{k|k-1})(H_k x_k - H_k \hat{x}_{k|k-1})^T | y_{1:k-1}] \\ &= \mathbb{E}[(x_k - \hat{x}_{k|k-1})(x_k - \hat{x}_{k|k-1})^T | y_{1:k-1}] H_k^T \\ &= P_{k|k-1} H_k^T\end{aligned} \tag{20}$$

Now that the predicted density (12) is derived for linear state space. The KF equations are proven to be the same as the conditional distribution of Gaussian variables x_k and y_k for linear Gaussian systems [2].

Lemma 0.1 *If the random variables x and y have the joint Gaussian probability distribution:*

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} A & C \\ C^T & B \end{pmatrix}\right) \tag{21}$$

The marginal distribution of x and y are given as:

$$x \sim \mathcal{N}(a, A), \tag{22a}$$

$$y \sim \mathcal{N}(b, B), \tag{22b}$$

$$x|y \sim \mathcal{N}(a + CB^{-1}(y - b), A - CB^{-1}C^T), \tag{22c}$$

$$y|x \sim \mathcal{N}(b + C^T A^{-1}(x - a), B - C^T A^{-1} C). \tag{22d}$$

The marginal distribution (22) is proven in Ch. 2 of [8]. Now, the idea is to write probability density function of variable x from Eq. (21) and consider y as a constant variable to find mean and covariance of $p(x|y)$.

Using the (16), (17), (18), (19), and (20) as the joint Gaussian PDF of (12), gives $a = \hat{x}_{k|k-1}$, $b = \hat{y}_{k|k-1}$, $A = P_{k|k-1}$, $B = P_{yy,k|k-1}$, and $C = P_{xy,k|k-1}$. The marginal distribution formula (22) gives the KF equations:

$$x_k \sim \mathcal{N}(\hat{x}_{k|k}, P_{k|k}). \tag{23a}$$

$$\begin{aligned}\hat{x}_{k|k} &= \hat{x}_{k|k-1} + P_{xy,k|k-1} P_{yy,k|k-1}^{-1} [y_k - \hat{y}_{k|k-1}], \\ &= \hat{x}_{k|k-1} + K_k [y_k - \hat{y}_{k|k-1}].\end{aligned} \tag{23b}$$

$$\begin{aligned}P_{k|k} &= P_{k|k-1} - P_{xy,k|k-1} P_{yy,k|k-1}^{-1} P_{xy,k|k-1}^T, \\ &= P_{k|k-1} - K_k P_{yy,k|k-1} K_k^T.\end{aligned} \tag{23c}$$

Where (23) is rewritten in the form of standard measurement-update KF (4) by defining Kalman gain as $K_k P_{xy,k|k-1} P_{yy,k|k-1}^{-1}$. The $\hat{x}_{k|k-1}$ and $P_{k|k-1}$ are the predicted variables from the time-update step (16) and (17).

3) Fig. 3 is the block diagram of CKF. Similar to the EKF block diagram, the filter is taking the previous state estimate mean $\hat{x}_{k-1|k-1}$ and covariance $P_{k-1|k-1}$ to calculate the new estimate state mean $x_{k|k}$ and covariance $P_{k|k}$.

   In the time-update state (illustrated with dashed blue rectangular), the sigma-points $\xi^{(i)}$ are derived based on spherical-radial cubature rules [5]. Then the sigma points will propagate in the system $f(.)$, and used for calculation of mean $x_{k|\hat{k}-1}$ and covariance $P_{k|k-1}$ of the predicted state $x_{k|k-1}$.

   In the measurement-update state (illustrated with dashed red rectangular), the propagated sigma-points $\xi^{(i)}$ are taken to derive the new sampling points using the square root matrix $S_{k|k-1}$. The propagated sigma-points in the system $h(.)$ will be used for calculation of the mean $\hat{y}_{k|k-1}$, covariance $P_{yy,k|k-1}$, and cross covariance $P_{xy,k|k-1}$. On the top left of the Fig. 3, the KF block takes the new measurement, with the first two moments of predicted measurement and it cross covariance with the predicted state to calculate the best state estimate of the mean $\hat{x}_{k|k}$ and covariance $P_{k|k}$ using represented analytic formulas of standard KF.

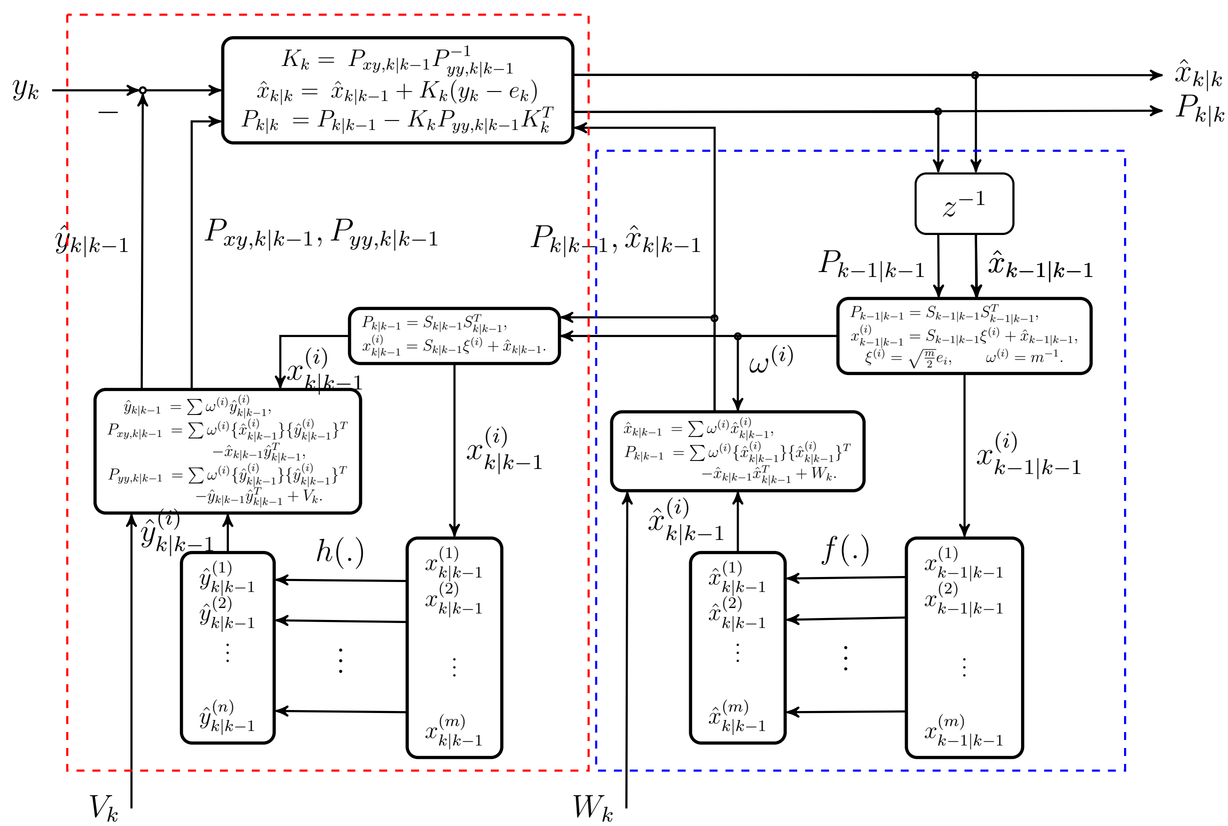

Figure 3: CKF block diagram. Time-update and measurement-update blocks are represented by blue and red dashed rectangular.

EKF and CKF filters aim to apply the extension form of the KF to nonlinear systems. However, EKF utilizes the first two moments of the state in its update rule. It is simple and offers a number of important practical benefits, however it does not guarantee the convergence and derivation of its Jacobian matrices might not be trivial. On the other hand, CKF uses sampling points to find the most accurate solution of Bayesian filter. The number of sampling points are not significantly increase the computational cost in comparison with EKF. Fig. 3 and Fig. 2 will be used in the next part for comparison between EKF and CKF.

4) The main difference between CKF and UKF is the way that they generate deterministic weighted sigma-points sigma-points $\xi^{(i)}$ [5, 4]. UKF considers a symmetric prior PDF within which the Gaussian is a special case, while CKF is an approximation of Bayesian filter under the Gaussian domain. Both algorithms can be used with discontinuous transformations.

   - Assuming that the n-dimensional prior expected state x_k having mean of $\hat{x}_{k|k}$ and covariance of $P_{k|k}$. UKF and CKF both use a deterministic weighted sigma-points $\xi^{(i)}$. The weighted sigma points of (24) are derived analytically from spherical-radial cubature rules based on the third-degree spherical-radial rule $m = 2n$. A third order characteristic of the cubature integration rule is exact to determines the mean for third order polynomials. But, it is exact to determine the covariance for the first order polynomials.

$$\xi^{(i)} = \sqrt{\frac{m}{2}} e_i, \tag{24a}$$

$$\omega^{(i)} = \frac{1}{m}, \qquad i = 1 : m. \tag{24b}$$

     On the other hand, the unscented transform of UKF algorithm should only satisfy the two moment-matching conditions (25). The PDF is assumed to be symmetric, so the odd moments are zero (similar to Gaussian). The weighted sigma-points in this approach are not unique, and it offers enough flexibility to allow information beyond mean and covariance to be incorporated into the set of sigma points. In this line of thinking, weights of unscented transform can be negative as long as the moment matching equations (25) hold.

     Also, it is important to mention that UKF can be seen as a generalization of the CKF , because CKF is just a UKF with specific parameters [2]. The weighted sigma points of (26) is one example of unscented transform $m = 2n+1$ [4]. Where $\bar{\omega}$ is a tuning parameter and $i \in \{1, \ldots, n\}$.

$$\hat{x}_{k|k} = \sum_{j=1}^{m} \omega_j \xi^{(j)}, \tag{25a}$$

$$P_{k|k} = \sum_{j=1}^{m} \omega_i (\xi^{(j)} - \hat{x}_{k|k})(\xi^{(j)} - \hat{x}_{k|k})^T \xi^{(j)}. \tag{25b}$$

$$\xi^{(0)} = \hat{x}_{k|k}, \qquad \omega^{(0)} = \bar{\omega}, \tag{26a}$$

$$\xi^{(i)} = \hat{x}_{k|k} + \sqrt{\frac{n}{1-\omega^{(0)}} P_{x|x}}, \qquad \omega^{(i)} = \frac{1-\omega^{(0)}}{2n}, \tag{26b}$$

$$\xi^{(i+n)} = \hat{x}_{k|k} - \sqrt{\frac{n}{1-\omega^{(0)}} P_{x|x}}, \qquad \omega^{(i+n)} = \frac{1-\omega^{(0)}}{2n}. \tag{26c}$$

- For the given examples, the third order spherical-radial cubature rules (24) requires $m = 2n$ sigma-points, and the unscented transform requires (24) $m = 2n+1$ sigma points. Although they look similar, but they result in a different set of points.

Part C

Comparison of the CKF with the EKF.

1) Fig. 2 and Fig. 3 are the block diagrams for EKF and CKF algorithms. The main difference between the two filters is the way that, they calculate first two moments of predicted state $\hat{x}_{k|k-1}$ and measurement $\hat{y}_{k|k-1}$ the first two moments of the previous state. The same KF equations (14) are used in both filters to find the best k^{th} state estimate of the mean and covariance from y_k measurement.

   EKF utilizes the first two moments of the state in its update rule, and propagate the single mean point through the system model $f(.)$ to find the mean. It is simple and successful to compromise between computational complexity and representation flexibility, when the mean and covariance are linearly transformable quantities with linear approximation of the system (Jacobian). The estimated mean and covariance from EKF algorithm are accurate for first order polynomials, so it is reliable when the error propagation can be well approximated by linear functions. Like the original KF [3], EKF does not require the exploitation of any specific error distribution information beyond the mean and covariance.

   CKF is a more accurate for the mean $\hat{x}_{k|k}$(3rd order polynomials). It assumes the noise to be in Gaussian form, then exploits the properties of highly efficient numerical integration methods known as cubature rules, to calculate the Bayesian filter integrals: (7d), (8d), (9c), (10c), and (11d). It produces weighted sigma-points $\xi^{(i)}$ using spherical-radial cubature rules based on the third-degree spherical-radial rule. The mean is exact for thirst order polynomials (better than EKF), while the covariance is exact for the first order polynomial.

2) CKF estimated mean is exact for the third order polynomials, and it does not require the model to be continuous or diffierentiable. It is better to be used for highly nonlinear functions. However EKF is a better choice when the system is almost linear. EKF is extendable to higher orders with Hessian matrices (third term of the Taylor series), but calculation of Jaboian and Hessian can be a very difficult and error-prone process [4, 9].

3) Although UKF and CKF are choosing sigma-points in a fundamentally different way, their computation cost is almost. The most expensive operations of these filters is in the calculation of the matrix square root and the outer product required to compute the covariance of the projected sigma-points. Matrix square root should be calculated using numerically efficient and stable methods such as the *Cholesky decomposition*, [8, 10].

   In this line of thinking, the computational cost of the UKF and CKF operations are in the same order $O(n^3)$, which is the same evaluations as evaluation the $n \times n$ matrix multiplications needed to calculate the EKF predicted covriance. More complicated versions of CKF e. g. Gauss-Hermite quadrature are computationally expensive as the number of points scales geometrically with the number of dimension [4].

References

[1] Fredrik Gustafsson. Particle filter theory and practice with positioning applications. *IEEE Aerospace and Electronic Systems Magazine*, 25(7):53–82, 2010.

[2] Simo Särkkä. *Bayesian filtering and smoothing*, volume 3. Cambridge University Press, 2013.

[3] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960.

[4] Simon J Julier and Jeffrey K Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401–422, 2004.

[5] Ienkaran Arasaratnam and Simon Haykin. Cubature kalman filters. *IEEE Transactions on automatic control*, 54(6):1254–1269, 2009.

[6] Jun S Liu. *Monte Carlo strategies in scientific computing*. Springer Science & Business Media, 2008.

[7] VE Beneš. Exact finite-dimensional filters for certain diffusions with nonlinear drift. *Stochastics: An International Journal of Probability and Stochastic Processes*, 5(1-2):65–92, 1981.

[8] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

[9] Peter Dulimov. *Estimation of ground parameters for the control of a wheeled vehicle*. PhD thesis, University of Sydney, 1998.

[10] Jeffrey K Uhlmann. Simultaneous map building and localization for real time applications. Technical report, Technical report, University of Oxford, 1994. Transfer thesis, 1994.