Adaptive Unscented Kalman Filter using Maximum Likelihood Estimation

Mahmoudi, Zeinab; Poulsen, Niels Kjølstad; Madsen, Henrik; Jørgensen, John Bagterp

Published in:
IFAC-PapersOnLine

Link to article, DOI:
10.1016/j.ifacol.2017.08.356

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 50-1 (2017) 3859–3864

Adaptive Unscented Kalman Filter using Maximum Likelihood Estimation ⋆

Zeinab Mahmoudi, Niels Kjølstad Poulsen, Henrik Madsen, John Bagterp Jørgensen

Department of Applied Mathematics and Computer Science, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark (e-mail: jbjo@dtu.dk).

Abstract: The purpose of this study is to develop an adaptive unscented Kalman filter (UKF) by tuning the measurement noise covariance. We use the maximum likelihood estimation (MLE) and the covariance matching (CM) method to estimate the noise covariance. The multi-step prediction errors generated by the UKF are used for covariance estimation by MLE and CM. Then we apply the two covariance estimation methods on an example application. In the example, we identify the covariance of the measurement noise for a continuous glucose monitoring (CGM) sensor. The sensor measures the subcutaneous glucose concentration for a type 1 diabetes patient. The root-mean square (RMS) error and the computation time are used to compare the performance of the two covariance estimation methods. The results indicate that as the prediction horizon expands, the RMS error for the MLE declines, while the error remains relatively large for the CM method. For larger prediction horizons, the MLE provides an estimate of the noise covariance that is less biased than the estimate by the CM method. The CM method is computationally less expensive though.

Keywords: Unscented Kalman filter, Maximum likelihood estimation, Covariance matching technique, Adaptive filtering, Covariance estimation, Continuous glucose monitors.

1. INTRODUCTION

Identifying the uncertainties that affect a system is fundamental for monitoring and designing an optimal and adaptive estimator. In order to have a filter that is close to optimal, we need to know the covariance of the process and measurement noise. Methods for identifying noise covariances often include maximum likelihood estimation (MLE) (Zagrobelny and Rawlings, 2015b; Jørgensen and Jørgensen, 2007), covariance matching (CM) techniques (Maybeck, 1982; Weige et al., 2015; Partovibakhsh and Liu, 2015), and correlation-based approaches such as the autocovariance least-squares (ALS) method (Åkesson et al., 2008; Odelson et al., 2006a,b; Zagrobelny and Rawlings, 2015a). These methods often deal with only linear or linearized systems. The CM technique is commonly used for the nonlinear systems, because it is computationally inexpensive and it is flexible to accommodate the nonlinear models. It is a suboptimal covariance estimation though. The literature on the use of optimization-based covariance estimation approaches for nonlinear systems is sparse.

The purpose of this study is to use an optimization-based estimator, i.e., the MLE method, for identification of the noise covariance in a nonlinear system. Furthermore, we compare the MLE approach with a suboptimal estimation method, i.e., the CM algorithm, in the context of an adaptive unscented Kalman filter (UKF). We employ the estimation of the measurement noise covariance as the basis for deriving filter adaptation.

The paper is structured as follows. First, we present the unscented Kalman filter (UKF) for prediction, filtering, and generating the covariance matrix of the prediction errors. Then, we develop the MLE problem and the CM technique to estimate the covariance of the measurement noise. The multi-step prediction errors and their covariances generated by the UKF are used in the MLE and CM methods. We then apply the MLE and CM algorithms on an example. The example is a nonlinear metabolic model of a patient with type 1 diabetes. In the example, we estimate the noise covariance of a continuous glucose monitoring (CGM) sensor, and derive the adaptive UKF to filter the sensor measurements.

2. MATERIALS AND METHODS

2.1 The unscented Kalman filter

In the UKF, a set of sigma points are deterministically chosen to represent the mean and covariance of the states. The sigma points therefore approximate the probability distribution of the states as it goes through the nonlinear transformation (Särkkä, 2007; Julier and Uhlmann, 2004). Approximating the probability distribution of the states by the sigma points in the UKF has shown to produce less estimation bias compared to the linearization in the extended Kalman filter (EKF) (Simon, 2006).

⋆ This work is funded by the Danish Diabetes Academy supported by the Novo Nordisk Foundation.

10.1016/j.ifacol.2017.08.356

Model The model of the state space in the stochastic differential equation (SDE) form and the measurement model are of the form

$$dx(t) = f\left(x(t), u(t), d(t)\right) dt + \sigma \cdot d\omega(t), \tag{1a}$$
$$y_k = g(x_k) + \xi_k, \tag{1b}$$
$$d\omega(t) \sim N_{iid}(0, Idt), \tag{1c}$$

in which x is the state, u is the input, d is the disturbance, and y is the measurement. We assume that ξ is a Gaussian zero-mean discrete-time measurement noise with covariance R. The stochastic noise ω is a standard Wiener process, and σ is the diffusion coefficient. I is an $n \times n$ identity matrix, where n is the number of state variables in the model.

Prediction This section explains the multi-step prediction with the UKF. The prediction steps are $j = 1, 2, \ldots, N_p$ and N_p is the prediction horizon. The scaling parameter λ determines how far the sigma points are scattered away from the mean. λ and c are defined as

$$\lambda = \alpha^2(n+\kappa) - n, \qquad c = \alpha^2(n+\kappa). \tag{2a}$$

The associated weights, W, for the $2n+1$ sigma points are given by

$$W_m^{(0)} = \lambda/(n+\lambda), \tag{2b}$$
$$W_c^{(0)} = \lambda/\{(n+\lambda)(1-\alpha^2+\beta)\}, \tag{2c}$$
$$W_m^{(i)} = 1/\{2(n+\lambda)\}, \qquad i = 1, \ldots, 2n \tag{2d}$$
$$W_c^{(i)} = 1/\{2(n+\lambda)\}, \qquad i = 1, \ldots, 2n \tag{2e}$$
$$W_m = [W_m^{(0)} \ldots W_m^{(2n)}]^T. \tag{2f}$$

A deterministic approach, based on the Cholesky factorization of the covariance P, samples the probability distribution to generate the sigma points $\hat{X}$.

$$\begin{aligned}\hat{X}_{k+j-1} &= [\hat{x}_{k+j-1|k} \ \ldots \ \hat{x}_{k+j-1|k}] \\ &\quad + \sqrt{c}\,[0 \quad \sqrt{P_{k+j-1|k}} \quad -\sqrt{P_{k+j-1|k}}] \\ &= [m^{(i)} \ \ldots \ m^{(2n)}], \qquad i = 0, \ldots, 2n.\end{aligned} \tag{3a}$$

The nonlinear function f propagates each of the sigma points according to

$$\frac{d\hat{X}_{k+j-1}}{dt}(t) = f\left(\hat{X}_{k+j-1}(t), u(t), d(t)\right), \qquad t \in [t_{k+j-1} \ t_{k+j}] \tag{3b}$$
$$\hat{X}_{k+j} = \hat{X}_{k+j-1}(t_{k+j}). \tag{3c}$$

The parameters α, κ, and β are set to $\alpha = 0.01$, $\kappa = 0$, and $\beta = 2$. The weighted average of the transformed sigma points gives the predicted mean

$$\hat{x}_{k+j|k} = \hat{X}_{k+j} W_m. \tag{3d}$$

The covariance of the estimation error is computed by propagating dP_{k+j-1} according to

$$\begin{aligned}&\frac{dP_{k+j-1}(t)}{dt} \\ &= \sum_{i=0}^{2n} W_c^{(i)} \left(m^{(i)}(t) - m_x(t)\right)\left(f(m^{(i)}(t), u(t)) - m_f(t)\right)^T \\ &+ \sum_{i=0}^{2n} W_c^{(i)} \left(f(m^{(i)}(t), u(t)) - m_f(t)\right)\left((m^{(i)}(t) - m_x(t)\right)^T \\ &+ \sigma\sigma^T, \qquad t \in [t_{k+j-1} \ t_{k+j}]\end{aligned} \tag{3e}$$

where m_x and m_f are

$$m_x(t) = \sum_{i=0}^{2n} W_m^{(i)} \ m^{(i)}(t),$$
$$m_f(t) = \sum_{i=0}^{2n} W_m^{(i)} \ f(m^{(i)}(t), u(t)). \tag{3f}$$

The propagated error covariance is then

$$P_{k+j|k} = P_{k+j-1}(t_{k+j}). \tag{3g}$$

To increase accuracy, new sigma points are generated from the predicted state mean and covariance as indicated in

$$\begin{aligned}\tilde{X}_{k+j} &= [\hat{x}_{k+j|k} \ \ldots \ \hat{x}_{k+j|k}] \\ &\quad + \sqrt{c}\,[0 \quad \sqrt{P_{k+j|k}} \quad -\sqrt{P_{k+j|k}}] \\ &= [\tilde{m}^{(i)} \ \ldots \ \tilde{m}^{(2n)}], \qquad i = 0, \ldots, 2n.\end{aligned} \tag{3h}$$

The measurement model transforms each of the new sigma points

$$\hat{Y}_{k+j} = g(\tilde{X}_k) = [\mu^{(i)} \ \ldots \ \mu^{(2n)}], \qquad i = 0, \ldots, 2n. \tag{4a}$$

The weighted average of the measurement sigma points gives the predicted measurement

$$\hat{y}_{k+j|k} = \hat{Y}_{k+j} W_m, \tag{4b}$$

and the j-step prediction error is given by

$$e_{k+j|k} = y_{k+j} - \hat{y}_{k+j|k}. \tag{4c}$$

S_{k+j}^{yy} is the covariance of $\hat{Y}_{k+j}$ and is computed as

$$S_{k+j}^{yy} = \sum_{i=0}^{2n} W_c^{(i)} \left(\mu^{(i)} - \hat{y}_{k+j|k}\right)\left(\mu^{(i)} - \hat{y}_{k+j|k}\right)^T. \tag{4d}$$

S_{k+j} is the covariance of $e_{k+j|k}$ and is calculated by

$$S_{k+j} = S_{k+j}^{yy} + R_{k+j}. \tag{4e}$$

Filtering The equation set (5) describes filtering and measurement update with the UKF. S_{k+1}^{xy} is the cross-covariance of $\tilde{X}$ and $\hat{Y}_{k+j|k}$, and can be estimated as

$$S_{k+1}^{xy} = \sum_{i=0}^{2n} W_c^{(i)} \ \left(\tilde{m}^{(i)} - \hat{x}_{k+1|k}\right)\left(\mu^{(i)} - \hat{y}_{k+1|k}\right)^T, \tag{5a}$$

and K_{k+1} is the filter gain as follows

$$K_{k+1} = S_{k+1}^{xy} \ S_{k+1}^{-1}. \tag{5b}$$

The updated state mean is computed as

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_{k+1}(y_{k+1} - \hat{y}_{k+1|k}). \tag{5c}$$

The updated error covariance is given by

$$P_{k+1|k+1} = P_{k+1|k} - K_{k+1} S_{k+1} K_{k+1}^T. \tag{5d}$$

Multi-step prediction error and its covariance matrix Let $\{y_k\}_{k=1}^N$ denote the measurements and N_p denote the prediction horizon. Let the time indices be $k = 0, 1, \ldots N - N_p$, and the prediction index be $1 \le j \le N_p$. This implies that $1 \le k + j \le N$. Let ϵ_{k+N_p} denote the vector of the prediction errors in the N_p-sample prediction window as

$$\epsilon_{k+N_p} = \begin{bmatrix} e_{k+1|k} \\ e_{k+2|k} \\ \vdots \\ e_{k+N_p|k} \end{bmatrix} = \begin{bmatrix} y_{k+1} - \hat{y}_{k+1|k} \\ y_{k+2} - \hat{y}_{k+2|k} \\ \vdots \\ y_{k+N_p} - \hat{y}_{k+N_p|k} \end{bmatrix}. \tag{6}$$

Analogously to the linear systems, the cross-covariances of the multi-step prediction errors may be computed by (Jørgensen and Jørgensen, 2007; Kailath et al., 2000)

$$S_{k+i,k+j} = \langle e_{k+i|k}, e_{k+j|k} \rangle = \begin{cases} \hat{Y}_{k+i}\ W\ \hat{Y}_{k+j}^T & \text{if } i > j, \\ \hat{Y}_{k+i}\ W\ \hat{Y}_{k+j}^T + R_{k+j} & \text{if } i = j, \\ \hat{Y}_{k+j}\ W\ \hat{Y}_{k+i}^T & \text{if } i < j, \end{cases} \quad (7)$$

The covariance matrix of ϵ_{k+N_p} is calculated by

$$\mathcal{R}_{k+N_p} = \langle \epsilon_{k+N_p}, \epsilon_{k+N_p} \rangle = \begin{bmatrix} S_{k+1,k+1} & S_{k+1,k+2} & \cdots & S_{k+1,k+N_p} \\ S_{k+2,k+1} & S_{k+2,k+2} & \cdots & S_{k+2,k+N_p} \\ \vdots & \vdots & \vdots & \vdots \\ S_{k+N_p,k+1} & S_{k+N_p,k+2} & \cdots & S_{k+N_p,k+N_p} \end{bmatrix}. \quad (8)$$

The covariance $\mathcal{R}_{k+N_p}$ is an $n_y N_p \times n_y N_p$ matrix, in which n_y is the size of the measurement vector y. Under the assumption that the measurement noise ξ is zero-mean Gaussian, the prediction error ϵ_{k+N_p} is also Gaussian with the distribution $N\left([0, 0, \cdots 0]^T, \mathcal{R}_{k+N_p}\right)$. By having $\mathcal{R}$, ϵ, and e from the UKF, we estimate and tune the covariance of the measurement noise R by MLE and the CM method.

2.2 Maximum likelihood estimation

By taking the negative log likelihood of the multivariate normal distribution of ϵ_{k+N_p}, we write the MLE optimization problem as

$$\min_{R \succeq 0} V(R) = \ln\ \det(\mathcal{R}_{k+N_p}) + \epsilon_{k+N_p}^T\ \mathcal{R}_{k+N_p}^{-1}\ \epsilon_{k+N_p}. \quad (9)$$

When N_p is large, computing $\ln\ \det(\mathcal{R}_{k+N_p})$ is challenging in terms of computational time and numerical accuracy. Alternatively, Cholesky factorization offers a faster approach. The Cholesky factorization decomposes the positive definite $\mathcal{R}_{k+N_p}$ into $\mathcal{R}_{k+N_p} = LL^T$ with L being a lower triangular matrix. Then

$$\ln\ \det(\mathcal{R}_{k+N_p}) = 2 \sum_{i=1}^{n_y N_p} ln(L_{ii}), \quad (10)$$

in which L_{ii} are the diagonal entries of L. Computing inverse of $\mathcal{R}_{k+N_p}$ for calculating $\epsilon_{k+N_p}^T\ \mathcal{R}_{k+N_p}^{-1}\ \epsilon_{k+N_p}$ is computationally heavy. To avoid matrix inversion, we compute $\epsilon_{k+N_p}^T\ \mathcal{R}_{k+N_p}^{-1}\ \epsilon_{k+N_p}$ by solving a system of linear equations $\mathcal{R}_{k+N_p} Z = \epsilon_{k+N_p}$ via back substitution and finding Z. Then

$$\epsilon_{k+N_p}^T \mathcal{R}_{k+N_p}^{-1} \epsilon_{k+N_p} = \epsilon_{k+N_p}^T Z. \quad (11)$$

2.3 Covariance matching technique

In the prediction window for the prediction steps $j = 1, 2, \ldots, N_p$, we compute S_{k+j} that is the theoretical covariance of $e_{k+j|k}$. The covariance matrix S_{k+j} is computed as

$$S_{k+j} = S_{k+j}^{yy} + R_{k+j}, \quad (12a)$$

in which the covariance matrix S_{k+j}^{yy} is calculated according to (4d). The sample covariance of $e_{k+j|k}$ is $\hat{S}_{k+j}$, which is given by

$$\hat{S}_{k+j} = \frac{1}{M-1} \sum_{q=k+j-M+1}^{k+j} e_{q|q-j}\ e_{q|q-j}^T, \quad (12b)$$

where M is the length of the data sequence used for estimating the sample covariance. We set M to 15. The estimated measurement noise covariance is

$$\hat{R} = \frac{1}{N_p - 1} \sum_{j=1}^{N_p} (\hat{S}_{k+j} - S_{k+j}^{yy}). \quad (12c)$$

In both estimation methods, the estimated R is used in the UKF for the next w samples. w is the size of the moving step of the prediction window and is set to 50% of N_p. The size of the prediction window is the same as the prediction horizon N_p.

The root-mean square (RMS) error of the estimated noise covariances evaluates the performance of the two covariance estimation methods. The RMS error is computed by

$$V_{rms} = \sqrt{\frac{1}{N} \sum_{k=1}^{N} \|R_{true,k} - \hat{R}_k\|_2^2}, \quad (13)$$

where N is length of the data sequence.

3. EXAMPLE: IDENTIFYING THE NOISE OF CONTINUOUS GLUCOSE MONITORING SENSOR

The CGM sensor measures interstitial glucose from the subcutaneous (SC) tissue. The sensor measurements are corrupted by random noise and artifacts originated from several sources including sensor electronics, miscalibration of the sensor, and biofouling.

3.1 The state space model

We used the Medtronic Virtual Patient (MVP) model in the SDE form for the state space representation of the patient's metabolism and also for simulating the CGM sensor (Kanderian et al., 2009). This model describes the pharmacokinetics (PK) of SC insulin and the insulin-glucose interaction. We also included the blood glucose -interstitial glucose dynamics in the model. The model also contains the two-compartments of carbohydrate (CHO) absorption (Wilinska et al., 2010). The model is described as

$$dI_{sc}(t) = \frac{1}{\tau_1}\left(\frac{ID(t)}{C_I} - I_{sc}(t)\right)dt + \sigma_{SC} \cdot d\omega_{SC}(t), \quad (14a)$$

$$dI_p(t) = \frac{1}{\tau_2}\left(I_{SC}(t) - I_p(t)\right)dt + \sigma_p \cdot d\omega_p(t), \quad (14b)$$

$$dI_{eff}(t) = \left(-P_2 \cdot I_{eff}(t) + P_2 \cdot S_I \cdot I_p(t)\right)dt + \sigma_{eff} \cdot d\omega_{eff}(t), \quad (14c)$$

$$dG_B(t) = \left(-(GEZI + I_{eff}(t)) \cdot G_B(t) + EGP + R_a(t)\right)dt + \sigma_G \cdot d\omega_G(t), \quad (14d)$$

$$dG_I(t) = -\frac{1}{\tau_3}\left(G_B(t) - G_I(t)\right)dt + \sigma_{G_I} \cdot d\omega_{G_I}(t), \quad (14e)$$

$$dD_1(t) = \left(q(t) - \frac{1}{\tau_m} D_1(t)\right)dt + \sigma_{D1} \cdot d\omega_{D_1}(t), \quad (14f)$$

$$dD_2(t) = \frac{1}{\tau_m}\left(D_1(t) - D_2(t)\right)dt + \sigma_{D2} \cdot d\omega_{D_2}(t), \quad (14g)$$

$$R_a(t) = \frac{1}{V_G \tau_m} D_2(t). \quad (14h)$$

ID is the SC insulin input (μU/min), I_{sc}, I_p, and I_{eff} are the SC insulin concentration (mU/L), the plasma insulin concentration (mU/L), and the effect of insulin (min^{-1}), respectively. G_B is the blood glucose concentration and G_I is the interstitial glucose concentration both in mg/dL. $q(t)$ is the CHO ingestion rate (g/min), D_1 and D_2 are the glucose masses (mg) in the accessible and inaccessible compartments, and R_a is the glucose appearance rate (mg/dL/min). τ_1 (49 min) is the time constant to the insulin movement from administration site to the SC tissue, τ_2 (47 min) is the time constant related to the insulin movement from the SC tissue to plasma, τ_3 (10 min) is the time constant related to the glucose movement from plasma to SC tissue, C_I (2010 mL/min) is the insulin clearance, P_2 (1.06 10^{-2} min^{-1}) is the delayed insulin action on the blood glucose, S_I (8.11 10^{-4} mL/μU/min) is the insulin sensitivity, $GEZI$ (2.20 10^{-3} min^{-1}) is the glucose effectiveness at zero insulin, EGP (1.33 mg/dL/min) is the endogenous glucose production rate at zero insulin, τ_m (47 min) is the peak time of meal absorption, and V_g (253 dL) is the volume of distribution for glucose.

ID contains basal and bolus insulin. The patient eats breakfast at 8:00 hrs, lunch at 13:15 hrs, dinner at 18:00 hrs, and snack at 22:00 hrs. The CHO content of the meals are 72 g for breakfast, 131 g for lunch, 51 g for dinner, and 70 g for snack, respectively. The insulin bolus to cover meal are 3.5 U for breakfast, 7 U for lunch, 2.5 U for dinner, and 3.5 U for snack, respectively.

3.2 The measurement model

The CGM sensor samples interstitial glucose. Therefore, the measurement model comprises G_I measurements which are affected by the measurement noise ϕ as indicated by

$$y_k = G_{I,k} + \phi_k. \tag{15a}$$

The measurement noise ϕ has covariance R_ϕ, and Facchinetti et al. identified it as the sum of two autoregressive processes given by (Facchinetti et al., 2014)

$$\phi_k = c_k + \hat{\vartheta}_k, \tag{15b}$$
$$c_k = 1.23c_{k-1} - 0.3995c_{k-1} + \delta_{c,k}, \tag{15c}$$
$$\hat{\vartheta}_k = 1.013\hat{\vartheta}_{k-1} - 0.2135\hat{\vartheta}_{k-2} + \delta_{\vartheta,k}, \tag{15d}$$
$$\delta_{c,k} \sim N(0, 11.3), \qquad \delta_{c,k} \sim N(0, 14.45).$$

The model (1) corresponds to the model (14) with the state variables $x = [I_{sc}\ \ I_p\ \ I_{eff}\ \ G_B\ \ G_I\ \ D_1\ \ D_2]^T$, the input $u = ID$, the disturbance $d = q$, the noise $\xi = \phi$, and the measurement y being the CGM data. In this example, the measurement model g is linear and R_ϕ is a scaler. For simulating the measurements y, we first simulated the model in (14) by using Euler Maruyama method (Higham, 2001). Then we added noise ϕ to the simulated G_I. We simulated one day of one-minute CGM data that consists of 1440 measurements. The aim is to estimate the unknown R_ϕ and adopt the UKF accordingly.

4. RESULTS AND DISCUSSION

We considered six different values for the prediction horizon N_p (see Table 1). For each prediction horizon we performed 50 experiments, each experiment consists of simulating the model in (14) to generate one day (1440 min) CGM data. For simulation, we set σ to 0.5% of x_{ss}, in which x_{ss} is the steady state of the model. The experiments had different realizations of the measurement noise ϕ and process noise $d\omega$. Then we applied the MLE and the CM method on each experiment to estimate the covariance R_ϕ. We initialized the UKF from the steady state of the model, and $P_0 = \sigma\sigma^T$. Table 1 and Table 2 summarize the results which are averaged over 50 experiments for each prediction horizon.

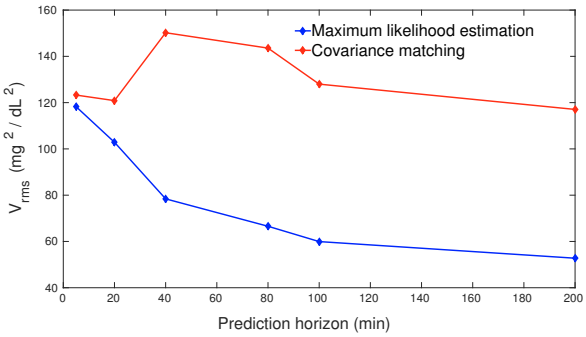


Fig. 1. Root-mean square error of covariance estimation averaged over 50 experiments.

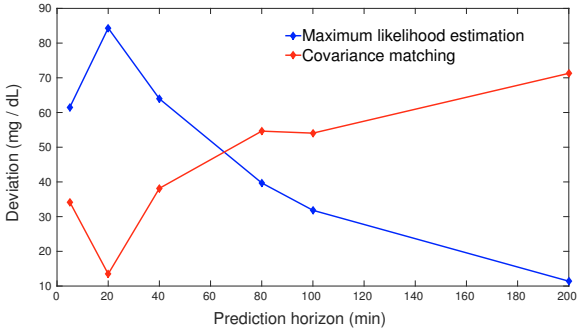


Fig. 2. Deviation of the estimated covariance from the true covariance averaged over 50 experiments.

Fig. 1 compares the performance of the two covariance estimation methods in terms of the root-mean square error. Fig. 2 illustrates the absolute deviation of the estimated covariance from the true covariance, for the two estimation methods. Fig. 3 depicts the histogram of the estimated covariance over 50 experiments for the prediction horizon = 200 min. Fig. 4 shows the histogram of the CPU time for covariance estimation for 50 experiments and the prediction horizon = 200 min. Fig. 5 indicates the result of applying MLE and CM methods on an example experiment. Fig. 1 and Fig. 2 indicate that as the prediction horizon N_p expands, the bias of the covariance estimate declines. This originates from the consistency property of the MLE. Because MLE is a consistent estimator, increasing the number of measurements by expanding the prediction horizon improves the estimation precision and reduces bias. This is well demonstrated in Fig. 1 and Fig. 2. The figures also indicate that the decrease in V_{rms} and bias is approximately exponential for the MLE. Fig. 3 also shows that for sufficiently large N_p, the MLE has considerably less bias than CM method.

Table 1. Estimating measurement noise covariance by maximum likelihood estimation*.

Prediction horizon (min)	$R_{\phi,true}$ (mg^2/dL^2)	$\hat{R}_\phi$	$(\lvert R_{\phi,true} - \hat{R}_\phi \rvert / R_{\phi,true})$ %	V_{rms}	CPU time (s)
5	103.8	42.3	59.3	118.2	378.0
20	114.0	29.6	74.0	102.9	284.3
40	105.9	41.9	60.4	78.4	277.1
80	104.9	65.3	37.7	66.6	289.9
100	102.6	70.8	31.0	59.9	301.2
200	106.9	98.8	7.5	52.8	378.5

* The values are the mean for one-day data which is the average of the 50 one-day experiments.

Table 2. Estimating measurement noise covariance by covariance matching*.

Prediction horizon (min)	$R_{\phi,true}$ (mg^2/dL^2)	$\hat{R}_\phi$	$(\lvert R_{\phi,true} - \hat{R}_\phi \rvert / R_{\phi,true})$ %	V_{rms}	CPU time (s)
5	103.8	69.9	32.6	123.3	220.5
20	114.0	105.5	7.4	120.9	179.2
40	105.9	143.3	35.4	150.2	172.7
80	104.9	159.6	52.1	143.6	161.0
100	102.6	156.7	52.6	128.0	157.2
200	106.9	178.2	66.7	117.1	142.6

* The values are the mean for one-day data which is the average of the 50 one-day experiments.

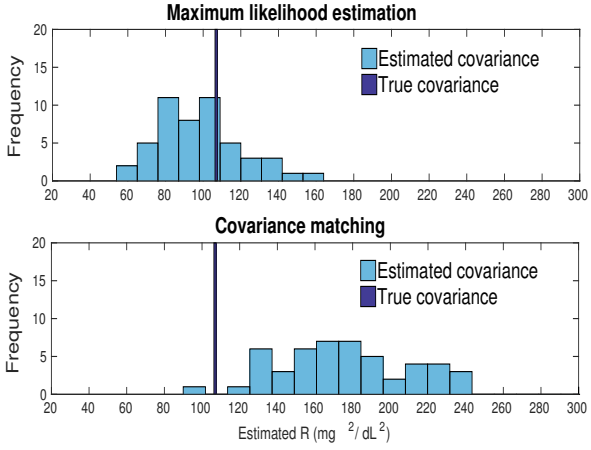


Fig. 3. Histogram of the estimated covariance compared to the true covariance based on 50 experiments and the prediction horizon = 200 min.

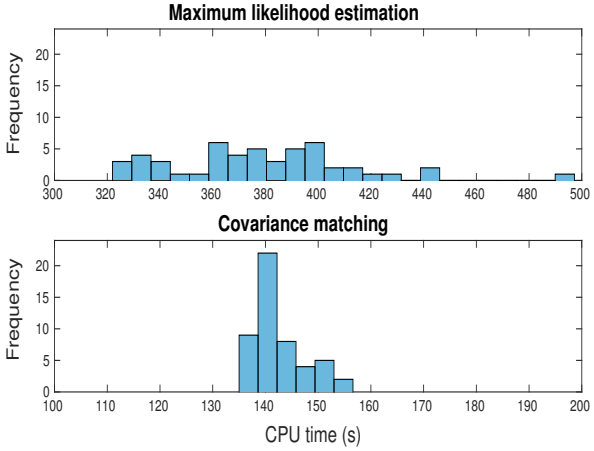
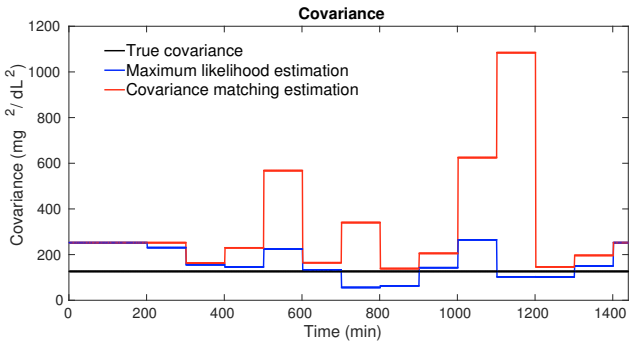
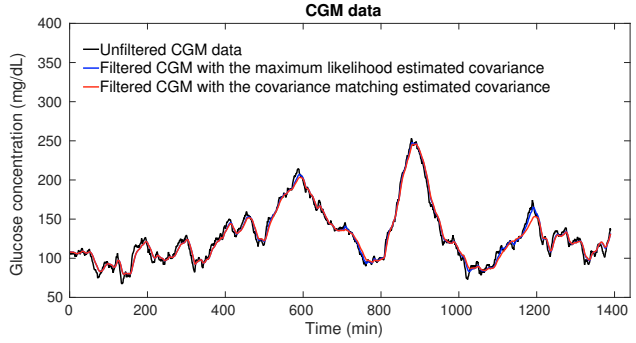


Fig. 4. Histogram of CPU time for covariance estimation based on 50 experiments and the prediction horizon = 200 min.

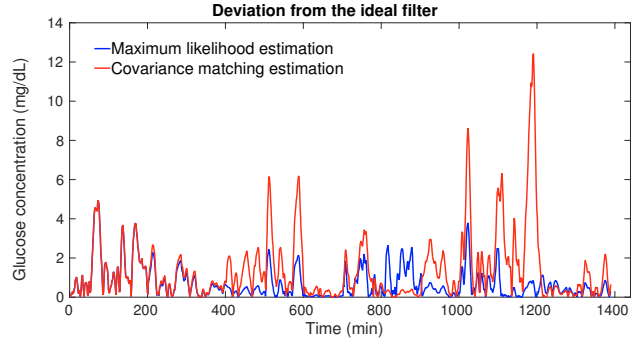
Fig. 4 implies that the CPU time for the MLE is around 1.5 times greater than that for the CM estimation. However, this CPU time seems reasonable for the example application in Section 3. If the filter was optimal, the filter innovations (the one-step ahead prediction error) were white and could serve as the orthogonal basis for the LDL decomposition of $\mathcal{R}_{k+N_p}$ (Jørgensen and Jørgensen, 2007; Kailath et al., 2000). In this case, there is no need for the Cholesky decomposition of $\mathcal{R}_{k+N_p}$. In addition we could spare us the multi-step predictions, because the filter routine is sufficient to generate the innovations sequence. This would result in a computationally more efficient MLE. However, the innovations sequence is not white for two reasons. First, we did not assume ξ in (1b) to be a white noise. Second, we do not process the data with the optimal filter. The optimal filter is unknown, because the true noise covariance is not known in the prediction window and is to be estimated. Furthermore, the assumption of optimality for the nonlinear filters, i.e. the UKF and EKF, is not valid in general. This is due to the fact that the UKF approximates the state probability distribution and the EKF linearizes the state-space model, both procedures make the filter deviate from optimality. As Fig. 5(c) illustrates, the CGM data filtered with the maximum likelihood estimated covariance is closer to the ideally (known covariance) filtered CGM, compared to the CGM data filtered with the CM estimated covariance. The improvement is modest though. This is because the process noise covariance moderates the effect of variation of the measurement noise covariance on filtering. When the process noise is relatively small (small σ in (14)), the filtered measurements are close to the one step model-predicted measurements, due to the small filter gain, without being



(a)



(b)



(c)

Fig. 5. Covariance estimation and filtering for an example experiment and the prediction horizon = 200 min. a) Estimated covariance. b) The filtered CGM data. c) The absolute deviation of the filtered CGM data from the ideally filtered CGM. The ideal filter is the UKF with the actual measurement noise covariance.

profoundly affected by the variations of the measurement noise covariance.

5. CONCLUSIONS

We presented an adaptive UKF by tuning the measurement noise covariance. A method based on the MLE estimates the noise covariance. The inputs of the ML objective function are the multi-step prediction errors and their covariance matrix generated by the UKF. We also compared the method with the CM algorithm which is a suboptimal estimation technique. The results generally show that the MLE method outperforms the CM method. However, the computational cost associated to the MLE method is somewhat larger than the computational cost of the CM method.

REFERENCES

Facchinetti, A., Del Favero, S., Sparacino, G., Castle, J., Ward, W., and Cobelli, C. (2014). Modeling the glucose sensor error. *IEEE Transactions on Biomedical Engineering*, 61(3), 620–629.

Higham, D.J. (2001). An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review*, 43(3), 525–546.

Jørgensen, J.B. and Jørgensen, S.B. (2007). Comparison of prediction-error-modelling criteria. In *Proceedings of the American Control Conference*, 5300–5306.

Julier, S. and Uhlmann, J. (2004). Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3), 401–422.

Kailath, T., Sayed, A.H., and Hassibi, B. (2000). *Linear estimation*. Prentice Hall,.

Kanderian, S., Weinzimer, S., Voskanyan, G., and Steil, G. (2009). Identification of intraday metabolic profiles during closed-loop glucose control in individuals with type 1 diabetes. *Journal of Diabetes Science and Technology*, 3(5), 1047–1057.

Maybeck, P.S. (1982). *Stochastic models, estimation, and control*. Academic Press, Chapter 10.

Odelson, B.J., Lutz, A., and Rawlings, J.B. (2006a). The autocovariance least-squares method for estimating covariances: Application to model-based control of chemical reactors. *IEEE Transactions on Control Systems Technology*, 14(3), 532–540.

Odelson, B.J., Rajamani, M.R., and Rawlings, J.B. (2006b). A new autocovariance least-squares method for estimating noise covariances. *Automatica*, 42(2), 303–308.

Partovibakhsh, M. and Liu, G. (2015). An adaptive unscented Kalman filtering approach for online estimation of model parameters and state-of-charge of lithium-ion batteries for autonomous mobile robots. *IEEE Transactions on Control Systems Technology*, 23(1), 357–363.

Åkesson, B.M., Jørgensen, J.B., Poulsen, N.K., and Jørgensen, S.B. (2008). A generalized autocovariance least-squares method for kalman filter tuning. *Journal of Process Control*, 18(7), 769–779.

Särkkä, S. (2007). On unscented Kalman filtering for state estimation of continuous-time nonlinear systems. *IEEE Transactions on Automatic Control*, 52(9), 1631–1641.

Simon, D. (2006). *Optimal state estimation : Kalman, H∞, and nonlinear approaches*. John Wiley & Sons.

Weige, Z., Wei, S., and Zeyu, M. (2015). Adaptive unscented Kalman filter based state of energy and power capability estimation approach for lithium-ion battery. *Journal of Power Sources*, 289, 50–62.

Wilinska, M.E., Chassin, L.J., Acerini, C.L., Allen, J.M., Dunger, D.B., and Hovorka, R. (2010). Simulation environment to evaluate closed-loop insulin delivery systems in type 1 diabetes. *Journal of Diabetes Science and Technology*, 4(1), 132–144.

Zagrobelny, M.A. and Rawlings, J.B. (2015a). Practical improvements to autocovariance least-squares. *AIChE Journal*, 61(6), 1840–1855.

Zagrobelny, M.A. and Rawlings, J.B. (2015b). Identifying the uncertainty structure using maximum likelihood estimation. *Proceedings of the American Control Conference*, 2015-, 422–427.