A VARIANCE-ADAPTIVE PARTICLE FILTER WITH APPLICATION TO TIME-VARYING PARAMETER ESTIMATION

Zhang, Bai Chen, Minze and Zhou, D. H.

Dept. of Automation, Tsinghua University, Beijing 100084, P.R. China
Email: zhangbai00@mails.tsinghua.edu.cn

Abstract: Particle filtering, as a new method to solve dynamic system filtering problems, has been applied with great success to many scientific and engineering fields. Particle filters have the ability to perform state estimation in nonlinear and non-Gaussian state space models. However, the standard particle filter algorithm is not applicable for time-varying parameter estimation problems, especially incompetent for abrupt parameters. In this paper, a variance-adaptive particle filter (VAPF) algorithm is proposed, and is applied to time-varying parameter estimation. A simulation example is also presented to demonstrate this method.

Keywords: particle filter; sequential Monte Carlo; parameter estimation; variance adaptive

1. INTRODUCTION

The first particle filter (PF) algorithm, or bootstrap filter, was proposed by N. J. Gordon in 1993 (Gordon, et al., 1993). Since then a number of alternative particle filter algorithms have been proposed, such as sampling importance sampling (SIS) particle filter, auxiliary sampling importance resampling (ASIR) particle filter, and regularized particle filter (RPF) (Arulampalam, *et al*., 2001). Particle filters follow Bayesian filtering formulae, which provide a rigorous general framework for dynamic state estimation. They use sequential Monte Carlo methods to approximate the optimal filtering by representing the probability density function (PDF) with a swarm of particles. Particle filters are particularly useful in dealing with nonlinear and non-Gaussian problems. Particle filter algorithms consist of two steps, prediction and update, which enable the particle filters to perform online state estimation recursively. Convergence results for particle filters have also been studied, which are reviewed in (Crisan and Doucet, 2002). Particle filters have been successfully applied to many scientific and engineering fields such as tracking problems (Arulampalam, *et al*., 2001), speech enhancement (Vermaak, et al., 2002), fault detection (Li and Kadirkamanathan, 2001) and fault prediction (Chen and Zhou, 2003).

In some application fields, there are demands to estimate the unknown or time-varying parameters online. The standard approach is to augment the state vector with the parameters, which is often used in Kalman filter or extended Kalman filter (EKF). However, it is not successful to apply this method directly in particle filters for parameter estimation, because the augmented state vector is lack of ergodicity. In order to solve this problem, several methods have been proposed (Andrieu, et al., 2004). But these methods are all designed for performing static parameter estimation, and the time-varying parameter estimation is still an open problem. Especially for the abrupt-change parameter estimation, there is no effective method at present.

In this paper, a variance-adaptive particle filter (VAPF) algorithm is proposed, which is able to estimate both slow-change and abrupt-change time-varying parameters. The basic idea is to adjust noise variance added on the parameters dynamically in order to achieve both the estimation accuracy and fast tracking ability for the abrupt-change parameters.

The paper is organized as follows. In section 2, the state-space model description and the problem statement are presented. In section 3, the variance-adaptive particle filter principle and its algorithm are introduced. Simulation results are presented in section 4. The last section is conclusions.

2. PROBLEM STATEMENT

In this paper, we will consider a class of nonlinear systems:

$$\mathbf{x}_{k+1} = f(\mathbf{x}_k, \boldsymbol{\theta}_k, \mathbf{u}_k) + \mathbf{w}_k \tag{1}$$

$$\mathbf{y}_k = h(\mathbf{x}_k) + \mathbf{v}_k \tag{2}$$

where

$\mathbf{x}_k \in \mathbf{R}^n$: state vector;

$\boldsymbol{\theta}_k \in \mathbf{R}^l$: time-varying unknown parameter vector;

$\mathbf{u}_k \in \mathbf{R}^p$: input vector;

$\mathbf{w}_k \in \mathbf{R}^n$: process noise vector independent of current state;

$f(\cdot,\cdot,\cdot)$: state transition function;

$\mathbf{y}_k \in \mathbf{R}^m$: output measurement vector;

$\mathbf{v}_k \in \mathbf{R}^m$: measurement noise vector independent of states and the system noise;

$h(\cdot,\cdot)$: measurement function.

Let D_k denote the available information of the measurement set at time *k*,

$$D_k = \{\mathbf{y}_i : i = 1,...,k\}$$

$\boldsymbol{\theta}_k$ is the time-varying unknown parameter to estimate. We augment the state vector with the parameter, denoting

$$\mathbf{z}_k \triangleq \begin{bmatrix} \mathbf{x}_k \\ \boldsymbol{\theta}_k \end{bmatrix}$$

Because $\{\mathbf{z}_k\}$ is not ergodic, the implementation of particle filters in such cases is bound to fail, even leads to divergence. In order to implement particle filter algorithm, a practical method is to add some artificial dynamic noise to the model of the unknown parameter $\boldsymbol{\theta}_k$:

$$\boldsymbol{\theta}_k = \boldsymbol{\theta}_{k-1} + \mathbf{w}_{k-1}^{\theta}$$

where $\mathbf{w}_{k-1}^{\theta}$ is the parameter noise.

However, the variance for the artificial dynamic noise is difficult to determine. If the variance is too small, it will be unable to track abrupt parameter change, which is illustrated in Fig. 1. If the variance is too large, the parameter estimation will be inaccurate, which is illustrated in Fig. 2. To tackle this problem, we propose a variance-adaptive particle filter, whose variance for the parameter noise is adjusted adaptively, ensuring that both the estimation accuracy and the parameter tracking ability are achieved.

3. VARIANCE-ADAPTIVE PARTICLE FILTER

Basic sampling importance resampling (SIR) algorithm was presented in (Gordon, et al., 1993), and other alternative algorithms were reviewed in (Arulampalam and et al., 2001). Here we will focus on the principle and algorithm of the variance-adaptive particle filter.

3.1 Principle

To determine the parameter noise variance, we convert this problem to the following optimization problem:

$$\begin{aligned} \max_{\lambda_k} J &= \mathbf{E}\left\{\Pr\{(\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}_{k+1|k}^N)^T (\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}_{k+1|k}^N) < \varepsilon^2\}\right\} \\ &= \mathbf{E}\{\Pr\{(\boldsymbol{\theta}_k + \boldsymbol{\eta}_k - \boldsymbol{\theta}_k^N - \lambda_k \mathbf{w}_k^{\theta})^T \\ &\quad \cdot (\boldsymbol{\theta}_k + \boldsymbol{\eta}_k - \boldsymbol{\theta}_k^N - \lambda_k \mathbf{w}_k^{\theta}) < \varepsilon^2\}\} \end{aligned} \tag{3}$$

where $\boldsymbol{\theta}_{k+1}$ is the actual time-varying parameter, $\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \boldsymbol{\eta}_k$, $\boldsymbol{\eta}_k$ is the dynamic of $\boldsymbol{\theta}_k$ at time k, $\boldsymbol{\theta}_{k+1|k}^N = \boldsymbol{\theta}_k^N + \lambda_k \mathbf{w}_k^{\theta}$ is the prediction value calculated by our algorithm, $\boldsymbol{\theta}_k^N$ is the random samples ("particles") of the unknown parameter produced in the algorithm, ε is a small constant.

Since $\boldsymbol{\eta}_k$ is unknown at time k, we ignore it. The optimization problem becomes the following form:

$$\begin{aligned} \max_{\lambda_k} J = \mathbf{E}\{\Pr\{&(\boldsymbol{\theta}_k - \boldsymbol{\theta}_k^N - \lambda_k \mathbf{w}_k^{\theta})^T \\ &\cdot (\boldsymbol{\theta}_k - \boldsymbol{\theta}_k^N - \lambda_k \mathbf{w}_k^{\theta}) < \varepsilon^2\}\} \end{aligned} \tag{4}$$

For simplicity, we only consider one dimensional unknown parameter is in this paper.

Then, the equation (4) is equivalent to the following form:

$$\max_{\lambda_k} J = \mathbf{E}\left\{\Pr\left\{-\frac{\varepsilon}{\lambda_k} < w_k^{\theta} - \frac{\theta_k - \theta_k^N}{\lambda_k} < \frac{\varepsilon}{\lambda_k}\right\}\right\} \tag{5}$$

We obtain three useful results via deduction, and summarize them as follows (their proofs are presented in the Appendix):

Lemma 1 If stochastic matrices A and B are independent, then

$$\mathrm{E}\{\mathbf{ABA}^T\} = \mathrm{E}\{\mathbf{A} \cdot \mathrm{E}\{\mathbf{B}\} \cdot \mathbf{A}^T\} \tag{6}$$

■

Lemma 2 Consider the system (1), $\mathbf{z}_k \triangleq \begin{bmatrix} \mathbf{x}_k \\ \boldsymbol{\theta}_k \end{bmatrix}$,

$\mathbf{H}_{k+1} = \left.\frac{\partial h}{\partial x}\right|_{\mathbf{x}_{k+1} = \hat{\mathbf{x}}_{k+1|k}}$, $\mathbf{F}_k^i = \left.\frac{\partial f}{\partial x}\right|_{\mathbf{x}_k = \mathbf{x}_k^i}$, if $\mathbf{E}\{\mathbf{v}_{k+1}\} = 0$, $\mathbf{E}\{\mathbf{v}_{k+1}\mathbf{v}_{k+1}^T\} = \mathbf{R}$, $\mathbf{E}\{\mathbf{w}_k\} = 0$, $\mathbf{E}\{\mathbf{w}_k\mathbf{w}_k^T\} = \mathbf{Q}$, there are,

$$\begin{aligned} &\mathbf{E}\left\{(\mathbf{z}_k - \mathbf{z}_k^i)(\mathbf{z}_k - \mathbf{z}_k^i)^T\right\} \\ &\approx \mathbf{E}\left\{\tilde{\mathbf{A}}^i (\mathbf{y}_{k+1} - \mathbf{y}_{k+1}^i)(\mathbf{y}_{k+1} - \mathbf{y}_{k+1}^i)^T (\tilde{\mathbf{A}}^i)^T\right\} \\ &\quad - \mathbf{E}\left\{\tilde{\mathbf{A}}^i (2\mathbf{H}_{k+1}\mathbf{Q}\mathbf{H}_{k+1}^T + \mathbf{R})(\tilde{\mathbf{A}}^i)^T\right\} \end{aligned} \tag{7}$$

where $\tilde{\mathbf{A}}^i$ is the generalized inverse matrix of $\mathbf{H}_{k+1}\mathbf{F}_k^i$, $\mathbf{z}_k^i$ is the i th particle of $\mathbf{z}_k$, and $\mathbf{y}_{k+1}^i$ is output prediction of the i th particle at time k+1 .

■

Theorem 1 If w_k^θ has the Gaussian-density function $f_w(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}}$, then the approximate solution for this optimization problem is

$$\lambda_k=\frac{1}{\sigma}\sqrt{\frac{1}{N}\sum_{i=1}^{N}b_i^2} .$$

where $b_i=\theta_k-\theta_k^i, i=1,2,...N$.

■

3.2 Variance-adaptive particle filter algorithm

Step1: Initialization

Augment the state vector with the unknown parameter:

$$\mathbf{z}_k \triangleq \begin{bmatrix}\mathbf{x}_k\\ \theta_k\end{bmatrix}$$

Sampling N particles $\{\mathbf{z}_0^i, i=1,...,N\}$ from the supposed conditional PDF $p(\mathbf{z}_0 \mid D_0)$.

Step2: prediction

$\tilde{\mathbf{w}}_k=\left((\mathbf{w}_k^x)^T,(\mathbf{w}_k^\theta)^T\right)^T$, Sample N values $\{\tilde{\mathbf{w}}_k^i, i=1,...,N\}$ from the PDF of $\tilde{\mathbf{w}}_k$, where $\mathbf{w}_k^x$ is the system noise and $\mathbf{w}_k^\theta$ is the parameter noise.

Then calculate

$$\mathbf{z}_{k+1|k}^i = f(\mathbf{z}_k^i,\mathbf{u}_k)+\tilde{\mathbf{w}}_k^i .$$

Therefore

$$p(\mathbf{z}_{k+1} \mid D_k)=\frac{1}{N}\sum_{i=1}^{N}\delta(\mathbf{z}_{k+1|k}-\mathbf{z}_{k+1|k}^i) ,$$

where δ is the Dirac-delta function.

Step3: Update

On receipt of the measurement $\mathbf{y}_{k+1}$,

Let

$$\mathbf{A}^i=\mathbf{H}_{k+1}\mathbf{F}(\mathbf{z}_k^i) ,$$

and $\tilde{\mathbf{A}}^i$ is the generalized inverse matrix of $\mathbf{H}_{k+1}\mathbf{F}(\mathbf{z}_k^i)$.

Calculate

$$\begin{aligned}\mathbf{P}_{k+1}^i&=(\mathbf{z}_k-\mathbf{z}_k^i)(\mathbf{z}_k-\mathbf{z}_k^i)^T\\&=\sum_{i=1}^{N}\tilde{\mathbf{A}}^i(\mathbf{y}_{k+1}-\mathbf{y}_{k+1}^i)(\mathbf{y}_{k+1}-\mathbf{y}_{k+1}^i)^T(\tilde{\mathbf{A}}^i)^T\\&-\sum_{i=1}^{N}\tilde{\mathbf{A}}^i(2\mathbf{H}_{k+1}\mathbf{Q}\mathbf{H}_{k+1}^T+\mathbf{R})(\tilde{\mathbf{A}}^i)^T\end{aligned}$$

where Q is the process noise covariance matrix and R is the measurement noise covariance matrix.

Let

$$b_k^i=\theta_k^i-\theta_k ,$$

then

$$\frac{1}{N}\sum_{i=1}^{N}b_k^i=\frac{1}{N}\sum_{i=1}^{N}\mathbf{P}_{k+1}^i(\dim(\mathbf{z}_k),\dim(\mathbf{z}_k)) .$$

where $\dim(\mathbf{z}_k)$ means the dimension of $\mathbf{z}_k$

If $f_w(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}}$, the approximate solution for the equation is

$$\lambda_k=\frac{1}{\sigma}\sqrt{\frac{1}{N}\sum_{i=1}^{N}(b_k^i)^2} .$$

Sample new $(w_k^\theta)^i$ from w_k^θ , then let $\tilde{\theta}_{k+1|k}^i=\theta_{k|k}^i+\lambda_k(w_k^\theta)^i$, $\tilde{\mathbf{z}}_{k+1|k}^i=(\tilde{\mathbf{x}}_{k+1|k}^i,\tilde{\theta}_{k+1|k}^i)$.

Calculate the weight ζ^i by

$$\zeta^i=\frac{p_v\left(\mathbf{y}_{k+1}-h(\tilde{\mathbf{z}}_{k+1|k}^i)\right)\delta(\mathbf{z}_{k+1|k}-\tilde{\mathbf{z}}_{k+1|k}^i)}{\sum_{j=1}^{N}p_v\left(\mathbf{y}_{k+1}-h(\tilde{\mathbf{z}}_{k+1|k}^j)\right)\delta(\mathbf{z}_{k+1|k}-\tilde{\mathbf{z}}_{k+1|k}^j)} ,$$

we have

$$\tilde{p}(\mathbf{z}_{k+1} \mid D_{k+1})=\sum_{i=1}^{N}\zeta^i\delta(\mathbf{z}_{k+1}-\tilde{\mathbf{z}}_{k+1|k}^i) .$$

Step4: Resampling

Resample independently N times from the above discrete distribution. The resulting particles $\left\{\mathbf{z}_{k+1}^i\right\}_{i=1,...,N}$ satisfy $\Pr\left\{\mathbf{z}_{k+1}^i=\mathbf{z}_{k+1|k}^j\right\}=\zeta^j$, $j=1,...,N$. Then the updated PDF becomes

$$p(\mathbf{z}_{k+1} \mid D_{k+1})=\frac{1}{N}\sum_{i=1}^{N}\delta(\mathbf{z}_{k+1}-\mathbf{z}_{k+1}^i)$$

Step5: Iteration

Replacing *k* by *k*+1, go to step 2.

4. SIMULATION RESULTS

4.1 System Description

The mathematical model of a continuous stirred tank reactor (CSTR) is described as (Zhou and Frank, 1998):

$$\frac{dC_A}{dt}=\frac{q}{V}(C_{Af}-C_A)-k_0\exp(-\frac{E}{RT})C_A$$

$$\frac{dT}{dt}=\frac{q}{V}(T_f-T)+\frac{-\Delta H}{\rho C_p}k_0\exp(-\frac{E}{RT})C_A+\frac{UA}{V\rho C_p}(T_c-T)$$

Denotations are listed as:

C_A: reactant concentration

T: reaction temperature

T_c: refrigerant temperature

q: inflow velocity

C_{Af}: inflow concentration

T_f: inflow temperature

V: reactor volume

k_0: exponential coefficient

E: activation energy

ΔH: reaction heat

C_p: thermal capacity

The parameters of standard state are shown in Table 1.

Table 1 normal CSTR parameters

Variable	Value
q	100 L/min
C_{Af}	1 mol/L
T_f	400 K
V	100 L
ρ	1 kg/L
Cp	0.239 J/(g•K)
E/R	5360 K
UA	11950J/(min•K)
-ΔH	17835.82 J/mol
Tc	419 K
k_0	exp(13.4) min^{-1}
dt	0.2 min

Based on prior knowledge on CSTR, system fault is most likely to occur in the inflow velocity *q*. Therefore it is necessary to estimate inflow velocity *q* online.

Control variable is $u = T_c$. The state vector is augmented with parameter q:

$$x = [x_1 \quad x_2 \quad q]^T = [C_A \quad T \quad q]^T$$

Output vector is

$$y = [y_1 \quad y_2]^T = [C_A \quad T]^T$$

Discretize the continuous state model with Euler method, and consider the effects of system noise and measurement noise. The CSTR model is described as:

$$x(k+1) = x(k) + dt \times g(x(k), u(k)) + w(k)$$

where

$$g(x(k),u(k)) = \begin{bmatrix} \frac{x_3(k)}{V}(C_{Af} - x_1(k)) - k_0 \exp(-\frac{E}{Rx_2(k)})x_1(k) \\ \frac{x_3(k)}{V}(T_f - x_2(k)) + \frac{-\Delta H}{\rho C_p} k_0 \exp(-\frac{E}{Rx_2(k)})x_1(k) \\ + \frac{UA}{V\rho C_p}(u(k) - x_2(k)) \\ 0 \end{bmatrix}$$

$$x(k) = (C_A, T, q)_k^T, w(k) = (w_1, w_2, w_k^{\theta})_k^T$$

$$y(k) = \left[x_1(k) \quad x_2(k)\right]^T + v(k)$$

Variance matrices for system noise *w*(*k*) and observation noise *v*(*k*) is Q and R, respectively:

$$Q = \begin{bmatrix} 0.005^2 & 0 & 0 \\ 0 & 0.5^2 & 0 \\ 0 & 0 & \lambda^2 \end{bmatrix}, R = \begin{bmatrix} 0.005^2 & 0 \\ 0 & 0.5^2 \end{bmatrix}$$

The CSTR initial states are:

$x_1(0) = 0.2 mol/L$, $x_2(0) = 400K$, $x_3(0) = 100 L/\min$

In order to initialize the variance-adaptive particle filter algorithm, we suppose the state vector initial distribution to be Gaussian with mean and variance

$$x(0) = \begin{bmatrix} 0.15 \\ 420 \\ 100 \end{bmatrix}, \quad Q = \begin{bmatrix} 0.005^2 & 0 & 0 \\ 0 & 0.5^2 & 0 \\ 0 & 0 & 0.6^2 \end{bmatrix}.$$

The control objective of the system is to track the reactant concentration with setpoint $x_1(k)$=0.2. Control method for this CSTR model is numerical PID control based on state feedback. Control algorithm is:

$$u(k) = u(k-1) + A_0\varepsilon(k) - A_1\varepsilon(k-1) + A_2\varepsilon(k-2)$$

$$\varepsilon(k) = x_1^d(k) - \hat{x}_1(k \mid k), \quad A_0 = K_p(1 + \frac{dt}{T_I} + \frac{T_d}{dt})$$

$$A_1 = K_p(1 + 2\frac{T_d}{dt}), \quad A_2 = K_p\frac{T_d}{dt}$$

where

$K_p = 100 K \bullet L/mol$, $T_i = 0.4 \min$,

$T_d = 0.1 \min$, $u(0) = 419K$.

The inflow velocity *q* is time-varying as

$$q(k) = \begin{cases} 100, & k<50 \\ 100+0.3\times(k-50), & 50 \le k<130 \\ 125, & 130 \le k<150 \\ 125-12.5\times(k-149), & 150 \le k<152 \\ 100, & k>152 \end{cases}$$

First, we use standard SIR particle filter algorithm to estimate the state and parameter *q*. In this case, the parameter noise variance is set to 0.6^2. The result is shown in Fig. 1. It is clear that the SIR particle filter is able to track the slow-change of the parameter *q*, but is unable to track the abrupt change. After about 50 steps from the abrupt parameter change, the estimation of *q* is close to the true value.

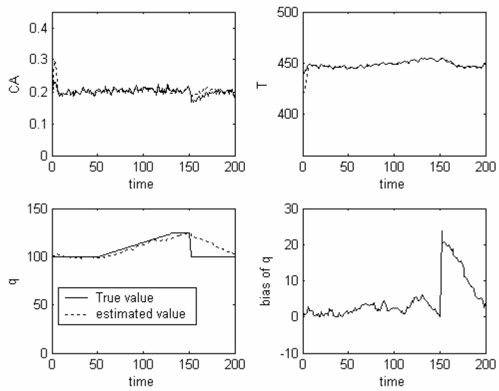

Fig. 1 State and parameter estimation using SIR algorithm

In order to achieve better tracking ability, we increase the parameter noise variance, which is set to 10^2.

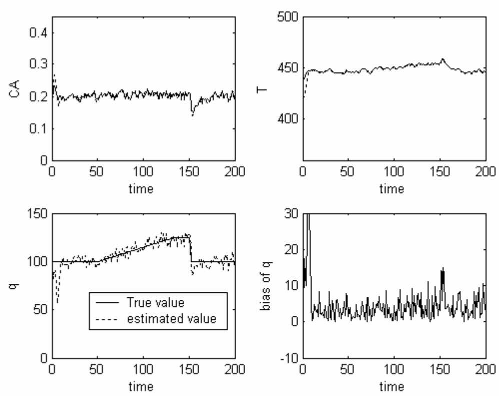

Fig. 2 State and parameter estimation using SIR algorithm with large parameter noise variance

The result is shown in Fig. 2. Although tracking ability is enhanced, the estimation accuracy is seriously damaged. The parameter estimation bias fluctuates notably, causing poor estimation accuracy.

The variance-adaptive particle filters adjust the parameter noise variance adaptively. Therefore, it is able to achieve both fast tracking ability and high estimation accuracy. Fig. 3 shows the estimation result using the variance-adaptive particle filter in CSTR.

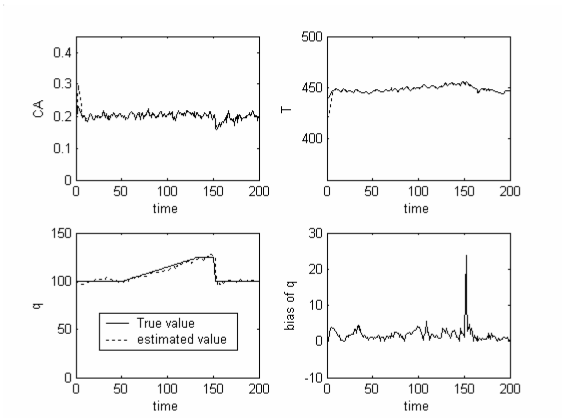

Fig. 3 State and parameter estimation using VAPF

As Fig. 3 shows, the VAPF is able to track both slow and abrupt parameter change. At the same time, it provides the estimation with high accuracy.

5 CONCLUSIONS

In this paper, we have proposed a variance-adaptive particle filter algorithm to estimate the unknown time-varying parameters. The VAPF is able to track both slow-change and abrupt-change parameters. It is an improvement on the standard particle filter.

The simulation results demonstrate that the variance adaptive particle filter has better performance than that of the standard particle filter in both tracking ability and estimation accuracy. Therefore this method will have many potential applications. In this paper, we only consider the case that the parameter is one dimensional. How to solve the optimization problem efficiently in multidimensional case still requires further research.

ACKNOWLEDGMENT

This work was mainly supported by NSFC (Grant No. 60025307, 60234010), partially supported by the national 863 program, the NSFH (Grant No. F2004000180) and the national 973 program (Grant No. 2002CB312200) of China.

REFERENCES

Andrieu C., A. Doucet, S. S. Singh, and V. B. Tadic (2004). Particle methods for change detection, system identification, and control. *IEEE Proceedings*, **92**(3), 423-438.

Arulampalam, M.S. and *et al*. (2001). A tutorial on particle filters for on-line nonlinear non-Gaussian Bayesian tracking. *IEEE Trans. On signal processing*, **50**(2), 174-188.

Chen, M. Z. and D. H. Zhou (2003). Particle filtering based fault prediction of nonlinear systems. *Proceedings of IFAC Symposium on Safeprocess*.

Crisan, D. and A. Doucet (2002). A survey of convergence results on particle filtering methods for practitioners. *IEEE Trans. On signal processing*, **50**(3), 736-746.

Gordon, N. J., D.J. Salmond and A.F.M. Smith (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F,* **140**, 107-113.

Li, P. and V. Kadirkamanathan (2001). Particle filtering based likelihood ratio approach to fault diagnosis in nonlinear stochastic systems. *IEEE transactions on systems, man, and cybernetics-part* c: *application and reviews*, **31**(3), 337-343.

Vermaak, J., C. Andrieu, and et al (2002). Particle methods for Bayesian modeling and enhancement of speech signals. *IEEE Transactions on Speech and audio processing*, **10**(3), 173-185.

Zhou, D.H. and P.M. Frank (1998). Fault diagnostics and fault tolerant control. *IEEE Transactions on Aerospace and electronic systems*, **34**(2), 420-427.

APPENDIX

A. Proof of Lemma 2:

$$
\begin{aligned}
&\left(\mathbf{y}_{k+1}-\mathbf{y}_{k+1}^{i}\right)\left(\mathbf{y}_{k+1}-\mathbf{y}_{k+1}^{i}\right)^{T}\\
&=\left(h(\mathbf{z}_{k+1})-h(\mathbf{z}_{k+1}^{i})+\mathbf{v}_{k+1}\right)\left(h(\mathbf{z}_{k+1})-h(\mathbf{z}_{k+1}^{i})+\mathbf{v}_{k+1}\right)^{T}\\
&=\left(h(\mathbf{z}_{k+1})-h(\mathbf{z}_{k+1}^{i})\right)\left(h(\mathbf{z}_{k+1})-h(\mathbf{z}_{k+1}^{i})\right)^{T}\\
&\quad+\mathbf{v}_{k+1}\left(h(\mathbf{z}_{k+1})-h(\mathbf{z}_{k+1}^{i})\right)^{T}+\left(h(\mathbf{z}_{k+1})-h(\mathbf{z}_{k+1}^{i})\right)\mathbf{v}_{k+1}^{T}\\
&\quad+\mathbf{v}_{k+1}\mathbf{v}_{k+1}^{T}\\
&\approx\left(\mathbf{H}_{k+1}\mathbf{F}_{k}^{i}(\mathbf{z}_{k}-\mathbf{z}_{k}^{i})+\mathbf{H}_{k+1}(\mathbf{w}_{k}-\mathbf{w}_{k}^{i})\right)\\
&\quad\cdot\left(\mathbf{H}_{k+1}\mathbf{F}_{k}^{i}(\mathbf{z}_{k}-\mathbf{z}_{k}^{i})+\mathbf{H}_{k+1}(\mathbf{w}_{k}-\mathbf{w}_{k}^{i})\right)^{T}\\
&\quad+\mathbf{v}_{k+1}\left(\mathbf{H}_{k+1}\mathbf{F}_{k}^{i}(\mathbf{z}_{k}-\mathbf{z}_{k}^{i})+\mathbf{H}_{k+1}(\mathbf{w}_{k}-\mathbf{w}_{k}^{i})\right)^{T}\\
&\quad+\left(\mathbf{H}_{k+1}\mathbf{F}_{k}^{i}(\mathbf{x}_{k}-\mathbf{x}_{k}^{i})+\mathbf{H}_{k+1}(\mathbf{w}_{k}-\mathbf{w}_{k}^{i})\right)\mathbf{v}_{k+1}^{T}+\mathbf{v}_{k+1}\mathbf{v}_{k+1}^{T}\\
&=\left(\mathbf{H}_{k+1}\mathbf{F}_{k}^{i}(\mathbf{z}_{k}-\mathbf{z}_{k}^{i})\right)\left(\mathbf{H}_{k+1}\mathbf{F}_{k}^{i}(\mathbf{z}_{k}-\mathbf{z}_{k}^{i})\right)^{T}\\
&\quad+\mathbf{H}_{k+1}(\mathbf{w}_{k}-\mathbf{w}_{k}^{i})\left(\mathbf{H}_{k+1}\mathbf{F}_{k}^{i}(\mathbf{z}_{k}-\mathbf{z}_{k}^{i})\right)^{T}\\
&\quad+\left(\mathbf{H}_{k+1}\mathbf{F}_{k}^{i}(\mathbf{z}_{k}-\mathbf{z}_{k}^{i})\right)\left(\mathbf{H}_{k+1}(\mathbf{w}_{k}-\mathbf{w}_{k}^{i})\right)^{T}\\
&\quad+\mathbf{H}_{k+1}(\mathbf{w}_{k}-\mathbf{w}_{k}^{i})\mathbf{H}_{k+1}(\mathbf{w}_{k}-\mathbf{w}_{k}^{i})^{T} \qquad\qquad (\mathrm{A.1})\\
&\quad+\mathbf{v}_{k+1}\left(\mathbf{H}_{k+1}\mathbf{F}_{k}^{i}(\mathbf{z}_{k}-\mathbf{z}_{k}^{i})+\mathbf{H}_{k+1}(\mathbf{w}_{k}-\mathbf{w}_{k}^{i})\right)^{T}\\
&\quad+\left(\mathbf{H}_{k+1}\mathbf{F}_{k}^{i}(\mathbf{z}_{k}-\mathbf{z}_{k}^{i})+\mathbf{H}_{k+1}(\mathbf{w}_{k}-\mathbf{w}_{k}^{i})\right)\mathbf{v}_{k+1}^{T}\\
&\quad+\mathbf{v}_{k+1}\mathbf{v}_{k+1}^{T}
\end{aligned}
$$

Left multiplying equation (A.1) by $\tilde{\mathbf{A}}^{i}$ and right multiplying it by $(\tilde{\mathbf{A}}^{i})^{T}$ leads to

$$
\tilde{\mathbf{A}}^{i}\left(\mathbf{y}_{k+1}-\mathbf{y}_{k+1}^{i}\right)\left(\mathbf{y}_{k+1}-\mathbf{y}_{k+1}^{i}\right)^{T}\left(\tilde{\mathbf{A}}^{i}\right)^{T}
$$

$$
\begin{aligned}
&= (\mathbf{z}_k - \mathbf{z}_k^i)(\mathbf{z}_k - \mathbf{z}_k^i)^T + \tilde{\mathbf{A}}^i \mathbf{H}_{k+1}(\mathbf{w}_k - \mathbf{w}_k^i)(\mathbf{z}_k - \mathbf{z}_k^i)^T \\
&\quad + (\mathbf{z}_k - \mathbf{z}_k^i)\left(\mathbf{H}_{k+1}(\mathbf{w}_k - \mathbf{w}_k^i)\right)^T (\tilde{\mathbf{A}}^i)^T \\
&\quad + \tilde{\mathbf{A}}^i \mathbf{H}_{k+1}(\mathbf{w}_k - \mathbf{w}_k^i)\left(\mathbf{H}_{k+1}(\mathbf{w}_k - \mathbf{w}_k^i)\right)^T (\tilde{\mathbf{A}}^i)^T \\
&\quad + \tilde{\mathbf{A}}^i \mathbf{v}_{k+1}\left(\mathbf{H}_{k+1}\mathbf{F}_k^i(\mathbf{z}_k - \mathbf{z}_k^i) + \mathbf{H}_{k+1}(\mathbf{w}_k - \mathbf{w}_k^i)\right)^T (\tilde{\mathbf{A}}^i)^T \\
&\quad + \tilde{\mathbf{A}}^i \left(\mathbf{H}_{k+1}\mathbf{F}_k^i(\mathbf{z}_k - \mathbf{z}_k^i) + \mathbf{H}_{k+1}(\mathbf{w}_k - \mathbf{w}_k^i)\right)\mathbf{v}_{k+1}^T (\tilde{\mathbf{A}}^i)^T \\
&\quad + \tilde{\mathbf{A}}^i \mathbf{v}_{k+1}\mathbf{v}_{k+1}^T (\tilde{\mathbf{A}}^i)^T
\end{aligned}
$$

so,

$$
\begin{aligned}
&\mathbf{E}\left\{\tilde{\mathbf{A}}^i(\mathbf{y}_{k+1} - \mathbf{y}_{k+1}^i)(\mathbf{y}_{k+1} - \mathbf{y}_{k+1}^N)^T (\tilde{\mathbf{A}}^i)^T\right\} \\
&= \mathbf{E}\left\{(\mathbf{z}_k - \mathbf{z}_k^i)(\mathbf{z}_k - \mathbf{z}_k^i)^T\right\} \\
&+ \mathbf{E}\left\{\tilde{\mathbf{A}}^i \mathbf{H}_{k+1}(\mathbf{w}_k - \mathbf{w}_k^i)(\mathbf{z}_k - \mathbf{z}_k^i)^T\right\} \\
&+ \mathbf{E}\left\{(\mathbf{z}_k - \mathbf{z}_k^i)\left(\mathbf{H}_{k+1}(\mathbf{w}_k - \mathbf{w}_k^i)\right)^T (\tilde{\mathbf{A}}^i)^T\right\} \\
&+ \mathbf{E}\left\{\tilde{\mathbf{A}}^i\left(\mathbf{H}_{k+1}(\mathbf{w}_k - \mathbf{w}_k^i)\right)\left(\mathbf{H}_{k+1}(\mathbf{w}_k - \mathbf{w}_k^i)\right)^T (\tilde{\mathbf{A}}^i)^T\right\} \\
&+ \mathbf{E}\left\{\tilde{\mathbf{A}}^i \mathbf{v}_{k+1}\left(\mathbf{H}_{k+1}\mathbf{F}_k^i(\mathbf{z}_k - \mathbf{z}_k^i) + \mathbf{H}_{k+1}(\mathbf{w}_k - \mathbf{w}_k^i)\right)^T (\tilde{\mathbf{A}}^i)^T\right\} \\
&+ \mathbf{E}\left\{\tilde{\mathbf{A}}^i\left(\mathbf{H}_{k+1}\mathbf{F}_k^i(\mathbf{z}_k - \mathbf{z}_k^i) + \mathbf{H}_{k+1}(\mathbf{w}_k - \mathbf{w}_k^i)\right)\mathbf{v}_{k+1}^T (\tilde{\mathbf{A}}^i)^T\right\} \\
&+ \mathbf{E}\left\{\tilde{\mathbf{A}}^i \mathbf{v}_{k+1}\mathbf{v}_{k+1}^T (\tilde{\mathbf{A}}^i)^T\right\}
\end{aligned}
$$

because $\mathbf{w}_k$ and $\mathbf{w}_k^i$ are independent identical distribution, $\mathbf{E}\{\mathbf{v}_{k+1}\} = 0$, $\mathbf{E}\{\mathbf{v}_{k+1}\mathbf{v}_{k+1}^T\} = \mathbf{R}$, $\mathbf{E}\{\mathbf{w}_k\} = 0$, $\mathbf{E}\{\mathbf{w}_k\mathbf{w}_k^T\} = \mathbf{Q}$, according to Lemma 1, there are:

$$
\mathbf{E}\left\{\tilde{\mathbf{A}}^i \mathbf{H}_{k+1}(\mathbf{w}_k - \mathbf{w}_k^i)(\mathbf{x}_k - \mathbf{x}_k^i)^T\right\} = 0,
$$

$$
\mathbf{E}\left\{(\mathbf{z}_k - \mathbf{z}_k^i)\left(\mathbf{H}_{k+1}(\mathbf{w}_k - \mathbf{w}_k^i)\right)^T (\tilde{\mathbf{A}}^i)^T\right\} = 0,
$$

$$
\begin{aligned}
&\mathbf{E}\left\{\tilde{\mathbf{A}}^i \mathbf{v}_{k+1}\left(\mathbf{H}_{k+1}\mathbf{F}_k^i(\mathbf{z}_k - \mathbf{z}_k^i) + \mathbf{H}_{k+1}(\mathbf{w}_k - \mathbf{w}_k^i)\right)^T (\tilde{\mathbf{A}}^i)^T\right\} \\
&= \mathbf{E}\left\{\tilde{\mathbf{A}}^i \mathbf{v}_{k+1}\left(\mathbf{H}_{k+1}\mathbf{F}_k^i(\mathbf{z}_k - \mathbf{z}_k^i)\right)^T (\tilde{\mathbf{A}}^i)^T\right\} \\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
&\mathbf{E}\left\{\tilde{\mathbf{A}}^i\left(\mathbf{H}_{k+1}\mathbf{F}_k^i(\mathbf{z}_k - \mathbf{z}_k^i) + \mathbf{H}_{k+1}(\mathbf{w}_k - \mathbf{w}_k^i)\right)\mathbf{v}_{k+1}^T (\tilde{\mathbf{A}}^i)^T\right\} \\
&= \mathbf{E}\left\{\tilde{\mathbf{A}}^i\left(\mathbf{H}_{k+1}\mathbf{F}_k^i(\mathbf{z}_k - \mathbf{z}_k^i)\right)\mathbf{v}_{k+1}^T (\tilde{\mathbf{A}}^i)^T\right\} \\
&= 0
\end{aligned}
$$

$$
\mathbf{E}\left\{\tilde{\mathbf{A}}^i \mathbf{v}_{k+1}\mathbf{v}_{k+1}^T (\tilde{\mathbf{A}}^i)^T\right\} = \mathbf{E}\left\{\tilde{\mathbf{A}}^i \mathbf{R} (\tilde{\mathbf{A}}^i)^T\right\}
$$

$$
\begin{aligned}
&\mathbf{E}\left\{\tilde{\mathbf{A}}^i\left(\mathbf{H}_{k+1}(\mathbf{w}_k - \mathbf{w}_k^i)\right)\left(\mathbf{H}_{k+1}(\mathbf{w}_k - \mathbf{w}_k^i)\right)^T (\tilde{\mathbf{A}}^i)^T\right\} \\
&= \mathbf{E}\left\{\tilde{\mathbf{A}}^i \mathbf{H}_{k+1} 2\mathbf{Q}\left(\tilde{\mathbf{A}}^i \mathbf{H}_{k+1}\right)^T\right\}
\end{aligned}
$$

so,

$$
\begin{aligned}
&\mathbf{E}\left\{\tilde{\mathbf{A}}^i(\mathbf{y}_{k+1} - \mathbf{y}_{k+1}^i)(\mathbf{y}_{k+1} - \mathbf{y}_{k+1}^i)^T (\tilde{\mathbf{A}}^i)^T\right\} \\
&= \mathbf{E}\left\{(\mathbf{z}_k - \mathbf{z}_k^i)(\mathbf{z}_k - \mathbf{z}_k^i)^T\right\} \\
&\quad + \mathbf{E}\left\{\tilde{\mathbf{A}}^i \mathbf{H}_{k+1} 2\mathbf{Q}\left(\tilde{\mathbf{A}}^i \mathbf{H}_{k+1}\right)^T\right\} + \mathbf{E}\left\{\tilde{\mathbf{A}}^i \mathbf{R} (\tilde{\mathbf{A}}^i)^T\right\}
\end{aligned}
$$

then,

$$
\begin{aligned}
&\mathbf{E}\left\{(\mathbf{z}_k - \mathbf{z}_k^i)(\mathbf{z}_k - \mathbf{z}_k^i)^T\right\} \\
&= \mathbf{E}\left\{\tilde{\mathbf{A}}^i(\mathbf{y}_{k+1} - \mathbf{y}_{k+1}^i)(\mathbf{y}_{k+1} - \mathbf{y}_{k+1}^i)^T (\tilde{\mathbf{A}}^i)^T\right\} \\
&\quad - \mathbf{E}\left\{\tilde{\mathbf{A}}^i\left(2\mathbf{H}_{k+1}\mathbf{Q}\mathbf{H}_{k+1}^T + \mathbf{R}\right)(\tilde{\mathbf{A}}^i)^T\right\}
\end{aligned}
$$

■

B. Proof of Theorem 1:

let $b_N = \theta_k - \theta_k^N$,

$$
\begin{aligned}
J &= \mathbf{E}\left\{\Pr\left\{-\frac{\varepsilon}{\lambda} < w_k^\theta - \frac{b_N}{\lambda} < \frac{\varepsilon}{\lambda}\right\}\right\} \\
&= \mathbf{E}\left\{\Pr\left\{-\frac{\varepsilon}{\lambda} + \frac{b_N}{\lambda} < w_k^\theta < \frac{\varepsilon}{\lambda} + \frac{b_N}{\lambda}\right\}\right\} \\
&= \mathbf{E}\left\{F_w\left(\frac{b_N + \varepsilon}{\lambda}\right) - F_w\left(\frac{b_N - \varepsilon}{\lambda}\right)\right\} \\
&\approx \mathbf{E}\left\{\frac{2\varepsilon}{\lambda} f_w\left(\frac{b_N}{\lambda}\right)\right\} \\
&= \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\frac{2\varepsilon}{\lambda} f_w\left(\frac{b_i}{\lambda}\right)\right]
\end{aligned}
$$

where F_w is the probability distribution function of parameter noise w_k^θ.

If $f_w(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$, there are,

$$
\begin{aligned}
J &\approx \frac{1}{N}\sum_{i=1}^{N}\left[\frac{2\varepsilon}{\lambda} f_w\left(\frac{b_i}{\lambda}\right)\right] \\
&= \frac{1}{N}\sum_{i=1}^{N}\left[\frac{2\varepsilon}{\lambda}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{b_i^2}{2\sigma^2\lambda^2}\right)\right]
\end{aligned}
$$

According to Taylor expansion formula,

let $b_0^2 = \frac{1}{N}\sum_{i=1}^{N} b_i^2$,

$$
\begin{aligned}
&\exp\left(-\frac{b_i^2}{2\sigma^2\lambda^2}\right) - \exp\left(-\frac{b_0^2}{2\sigma^2\lambda^2}\right) \\
&= -\exp\left(-\frac{b_0^2}{2\sigma^2\lambda^2}\right)\left(\frac{b_i^2}{2\sigma^2\lambda^2} - \frac{b_0^2}{2\sigma^2\lambda^2}\right) \\
&\quad + O\left(\frac{b_i^2}{2\sigma^2\lambda^2} - \frac{b_0^2}{2\sigma^2\lambda^2}\right)
\end{aligned}
$$

so,

$$
\exp\left(-\frac{b_i^2}{2\sigma^2\lambda^2}\right) \approx \exp\left(-\frac{b_0^2}{2\sigma^2\lambda^2}\right)\left(1 - \frac{b_i^2}{2\sigma^2\lambda^2} + \frac{b_0^2}{2\sigma^2\lambda^2}\right)
$$

$$
\begin{aligned}
J &\approx \lim_{N\to\infty}\frac{1}{N}\frac{2\varepsilon}{\lambda}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{b_0^2}{2\sigma^2\lambda^2}\right) \\
&\quad \cdot \sum_{i=1}^{N}\left(1 - \frac{b_i^2}{2\sigma^2\lambda^2} + \frac{b_0^2}{2\sigma^2\lambda^2}\right) \\
&= \frac{2\varepsilon}{\lambda}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{b_0^2}{2\sigma^2\lambda^2}\right)
\end{aligned}
$$

let $\frac{\partial J}{\partial \lambda} = \frac{b_0^2 - \lambda^2\sigma^2}{\lambda^4\sigma^3}\exp\left(-\frac{b_0^2}{2\sigma^2\lambda^2}\right) = 0$

then, $\lambda^* = \sqrt{\frac{b_0^2}{\sigma^2}}$, and

$$
\left.\frac{\partial^2 J}{\partial \lambda^2}\right|_{\lambda=\lambda^*} = \frac{1}{\lambda^3\sigma}\left(2 - \frac{5b_0^2}{\lambda^2\sigma^2} + \frac{b_0^4}{\lambda^4\sigma^4}\right)\exp\left(-\frac{b_0^2}{2\sigma^2\lambda^2}\right) < 0
$$

so,

when $\lambda = \frac{1}{\sigma}\sqrt{\frac{1}{N}\sum_{i=1}^{N} b_i^2}$, J is maximal.

■