Proceedings of the 2007 IEEE Symposium on Computational
Intelligence in Image and Signal Processing (CIISP 2007)

Evolution Strategies Based Particle Filters for Fault Detection

Katsuji Uosaki, *Member, IEEE*, and Toshiharu Hatanaka, *Member, IEEE*

Abstract— **Recent massive increase of the computational power has allowed to rebirth of Monte Carlo integration and its application of Bayesian filtering, or particle filters. Particle filters evaluate a posterior probability distribution of the state variable based on observations in Monte Carlo simulation using so-called importance sampling. However, the filter performance is deteriorated by degeneracy phenomena in the importance weights. Recognizing the similarities and the difference of the processes between the particle filters and Evolution Strategies, an Evolutionary Computaion approaches, a novel filter called the Evolution Strategies based particle filter (ESP) has been proposed to circumvent this difficulty and to improve the performance. Here, the ESP filter is applied to fault detection of nonlinear stochastic state space models. Its applicability is exemplified by numerical simulation studies.**

I. Introduction

The problem of fault detection in dynamic systems has attracted considerable attention in designing systems with safety and reliability. In the past two decades, a large number of methods have been proposed for solving the fault detection problem, see the survey papers [5], [10], [11], [24] and the books [6], [7], and references therein. These fault detection approaches fall into two major categories, i.e., model-based approaches which use the quantitative analytical model of the system to be monitored, and the knowledge-based approaches which do not need full analytical modeling and allow one to use qualitative models based on the available information and knowledge of the system to be monitored. In the case of information-rich systems the dynamic behavior of the system can be well described by a mathematical modelcan. Thus the model-based approaches are by nature the most powerful fault detection method. The decision of a fault is based on available observed input-output data and a mathematical model of the system for all model-based approaches, For the stochastic state space models, most of the fault detection schemes has relied on the system being linear and the system and observation noises being Gaussian, and decision making for fault detection is carried out based on the innovation from the Kalman filter [13], [1], [19].

Katsuji Uosaki is with Department of Management and Information Sciences, Fukui University of Technology, Gakuen, Fukui 910–8505, Japan (phone: +81–776–29-2521; fax: +81–776–29–7891; email: uosaki@fukui-ut.ac.jp and Toshiharu Hatanaka is with Department of Information and Physical Sciences, Graduate School of Information Science and Technology, Osaka University, Suita, Osaka 565–0871, Japan (phone: +81–6–6879–7834; fax: +81–6–6879–7836; email: hatanaka@ist.osaka-u.ac.jp).

Application of the idea used in the linear/Gaussian case mentioned above to nonlinear systems with non-Gaussian noises is generally difficult. Though extended Kalman filter (EKF) [12], [1], which uses the linear approximations of the nonlinear functions in system and observation equations around the estimate and applies the Kalman filter to obtain estimates for the state, can be used, it does not guarantee good result in many cases, i.e., divergence of the estimate of the state variable occurs due to the linearizations of nonlinear functions in case of severe nonlinearities in the models. Thus, the fault detection problem in general nonlinear/non-Gaussian stochastic systems are still open.

Recently, "particle filtering," a simulation-based method for Bayesian sequential analysis, attracts much attentions from the massive progress of computing ability [16], [8], [2]). This approach represents a probability density function by a weighted sum based on the discrete grid sequentially chosen by the importance sampling and the estimates are obtained based on corresponding importance weights. Though it can handle any functional nonlinearity and system and observation noises of any probability distribution, there is a common problem of the degeneracy phenomenon, where almost all importance weights tend to zero after some iteration and hence, a large computational effort is wasted to updating the particles with negligible weights. In order to resolve this difficulty, some modifications such as Resampling particle filter [17] has been proposed. Recognizing the similarities and differences of the operations in Resampling particle filter (SIR) and Evolution Strategies [18], which is an Evolutionary Computation approach [4], [9], [3], we also developed in [21] a novel particle filter called Evolution Strategies based Particle Filter (ESP). In this paper, ESP is employed to develop a new method for fault detection in general nonlinear/non-Gaussian stochastic systems. Numerical simulation studies have been conducted to exemplify the applicability of this approach.

II. Fault detection

A variety of fault detection methods has been developed for dynamic systems depending on the available knowledge about the system, fault type and noises [11], [10]. Consider here the following set of nonlinear state space models indexed by $m = 0, 1$.

$$x_{t+1} = f^{(m)}(x_t, u_t) + v_t \tag{1}$$

$$y_t = g^{(m)}(x_t) + w_t \qquad m = 0, 1 \tag{2}$$

where x_t, u_t, y_t are the state variable, input and observation, respectively, $f^{(m)}(\cdot)$ and $g^{(m)}(\cdot)$ are known possibly nonlinear functions, and v_t and w_t are independently identically distributed (i.i.d.) system noise and observation noise sequences, respectively. We assume v_t and w_t are mutually independent. The probability density functions (pdfs) of v_t and w_t are both assumed to be known as $p_v(v_t)$ and $p_w(w_t)$, respectively. The system works normally and its behavior is governed by the given normal mode model described as in (1) and (2) indexed by $m = 0$, and then the model may change to the fault mode model indexed by $m = 1$ at unknown time $t = \tau$. Then the fault detection problem to be considered here can be reduced to perform a hypothesis testing for the hypotheses:

$$\begin{aligned} &H_0 \text{ (Normal mode) : System model indexed by } m = 0 \\ &H_1 \text{ (Fault mode) : System model indexed by } m = 1 \end{aligned} \tag{3}$$

Wald's sequential probability ratio test (SPRT) [23] is common procedure for testing the above hypotheses. In the SPRT, we compute the logarithm of likelihood ratio function (LLR)

$$\lambda_t = \log \frac{p(y_{1:t}|H_1)}{p(y_{1:t}|H_0)} \tag{4}$$

and compare with two threshold values $B^* < 0 < A^*$ derived from the specified error probabilities for false alarm (α) and miss alarm (β), i.e.,

$$A^* = \log \frac{1-\beta}{\alpha}, \qquad B^* = \log \frac{\beta}{1-\alpha} \tag{5}$$

If the LLR exceeds the boundary A^* or falls below the boundary B^*, one terminates the observation with acceptance of the hypothesis H_1 (fault mode) or the hypothesis H_0 (normal mode), respectively. Otherwise, one continues the observation and defer the decision.

It is shown that for random variable λ_t, the following relation holds [7].

$$E_{H_0}(\lambda_t) < 0, \qquad E_{H_1}(\lambda_t) < 0 \tag{6}$$

where E_{H_m} denotes the expectation of random variables with the pdf under the hypothesis H_m. Since

$$p(y_{1:t}|H_m) = p(y_{1:t-1}|H_m)p(y_t|y_{1:t-1}, H_m) \quad m = 0, 1 \tag{7}$$

the LLR can be computed recursively by

$$\begin{aligned} \lambda_t &= \lambda_{t-1} + \log \frac{p(y_t|y_{1:t-1}, H_1)}{p(y_t|y_{1:t-1}, H_0)} = \lambda_t + \ell_t, \\ \ell_t &= \log \frac{p(y_t|y_{1:t-1}, H_1)}{p(y_t|y_{1:t-1}, H_0)} \end{aligned} \tag{8}$$

as the new observation comes in, and the test can be performed recursively.

Fault detection system based on the above mentioned Wald's SPRT formulation minimizes, on the average, the time to reach a decision for specified error probabilities if the system is either in the normal mode or the fault mode from the beginning of the test. However, the characteristics of the fault process differs from it; the system is initially operated in the normal mode and then transition occurs to the fault mode at time instant $\tau > 0$ during observations. When the system is in the normal mode, the LLR defined by (4) will show, on the average, a negative drift, and then the detection system suffers an extra time delay in compensating for a negative quantity accumulated in the period under the normal mode before the transition to the fault mode. Moreover, only the decision concerning to the fault mode is necessary in the usual fault detection, though the SPRT formulation considers the hypothesis corresponding to the normal mode. In the following, the idea of the backward SPRT (BSPRT) [20] is introduced to fit this situation.

A. Fault detection by Backward SPRT

Suppose the system fault occurs at time instant τ, the hypotheses representing normal and fault modes are given by

$$\begin{aligned} &H_0 \text{ (Normal mode) :} \\ &\quad \text{System model at time } k \text{ indexed by } m = 0 \\ &H_1 \text{ (Fault mode) :} \\ &\quad \text{System model at time } k \text{ indexed by } m = 1, \\ &\qquad k = \tau, \tau+1, \dots \end{aligned} \tag{9}$$

which can be restated as

$$\begin{aligned} &H_0 \text{ (Normal mode) :} \\ &\quad \text{System models at time } t-k+1 \text{ are indexed by } m = 0 \\ &H_1 \text{ (Fault mode) :} \\ &\quad \text{System models at time } t-k+1 \text{ are indexed by } m = 1, \\ &\qquad t > \tau,\ k = 1, \dots, t-\tau+1 \end{aligned} \tag{10}$$

Since, in this formulation, system model is indexed by $m = 1$ from the beginning $(k = 1)$ corresponding to fault mode, it agrees with the conventional SPRT formulation.

Define a backward LLR (BLLR), computed in reverse (*backward*) direction from the current observation to the past observations, by

$$\lambda_{t,k}^B = \log \frac{p(y_t, y_{t-1}, \ldots, y_{t-k+1}|H_1)}{p(y_t, y_{t-1}, \ldots, y_{t-k+1}|H_0)} \quad (11)$$

When the BLLR is applied to test the hypotheses (10), the test is called as the backward SPRT (BSPRT). We have interest only in detecting the fault mode, the decision rule is given as follows:

"*If* $\lambda_{t,k}^B > K$ *for some* $k = 1, \ldots, t$, *where* K *is a suitable constant, one terminates observation with acceptance of the hypothesis that the system is in the fault mode. Otherwise, one continue observations as the system is likely not in the fault mode.*"

Assuming $p(y_{1:t}) = p(y_{1:k})p(y_{k+1:t}|y_{1:k}) \approx p(y_{1:k}) \times p(y_{k+1:t})$ ($y_{1:k}$ and $y_{k+1:t}$ are independent), we can express the BLLR approximately with the conventional LLR as

$$\lambda_{t,k}^B = \lambda_t - \lambda_{t-k}, \quad k = 1, 2, \ldots, n \quad (12)$$

with $\lambda_0 = 0$. So the decision rule for acceptance of the hypothesis that the system is in the fault mode can be restated as

$$\lambda_{t,k}^B = \lambda_t - \lambda_{t-k} > K \text{ for some } k = 1, 2, \ldots, t \quad (13)$$

or,

$$\lambda_t - \min_{1 \le k \le t} \lambda_k > K \quad (14)$$

Introducing the statistics called the maximum BLLR,

$$\begin{aligned} &\Lambda_t = \max[0, \Lambda_{t-1} + \ell_t], \quad t = 1, 2, \ldots \\ &\Lambda_0 = 0 \end{aligned} \quad (15)$$

then the decision rule can be expressed as

"*If* $\Lambda_t > K$, *where* K *is a suitable constant, one terminates observation with acceptance of the hypothesis that the system is in the fault mode. Otherwise, one continue observations as the system is likely not in the fault mode.*"

For the nonlinear stochastic state space model (1) and (2), we can apply the following extended Kalman filter [1] to estimate the state variables.

$$\begin{aligned} &\hat{x}_{t|t-1} = f(\hat{x}_{t-1|t-1}) \\ &\sigma_{t|t-1}^2 = \tilde{a}_{t-1}^2 \sigma_{t-1|t-1}^2 + \sigma_v^2 \\ &\hat{x}_{t|t} = \hat{x}_{t|t-1} + k_t(y_t - g(\hat{x}_{t|t-1})) \\ &\sigma_{t|t}^2 = (1 - k_t \tilde{c}_t)\sigma_{t|t-1}^2 \\ &k_t = \frac{\tilde{c}_t \sigma_{t|t-1}^2}{\tilde{c}_t^2 \sigma_{t|t-1}^2 + \sigma_w^2} \\ &\tilde{a}_t = \frac{df(x)}{dx}\Big|_{x=\hat{x}_{t|t}}, \quad \tilde{c}_t = \frac{dg(x)}{dx}\Big|_{x=\hat{x}_{t|t-1}} \end{aligned} \quad (16)$$

where σ_v^2 and σ_w^2 are variances of noises v_t and w_t, respectively. The predicted output based on the EKF state estimate is given by

$$\hat{y}_t = g(\hat{x}_{t|t-1}) \quad (17)$$

For linear Gaussian state space model, the innovations $\nu_t = y_t - \hat{y}_t$ is independent Gaussian random variables with mean zero and variance

$$\hat{\sigma}_\nu^2 = \tilde{c}_t^2 \sigma_{t|t-1}^2 + \sigma_w^2, \quad (18)$$

and the test of the BLLR is carried out by using

$$\begin{aligned} \Lambda_t = \max \Bigg[0, \Lambda_{t-1} + \log \frac{\sigma_\nu^{(0)}}{\sigma_\nu^{(1)}} \\ + \frac{1}{2} \left(\frac{(y_t - g^{(1)}(\hat{x}_{t|t-1}^{(0)}))^2}{\sigma_\nu^{(0)2}} - \frac{(y_t - g^{(1)}(\hat{x}_{t|t-1}^{(1)}))^2}{\sigma_\nu^{(1)2}} \right) \Bigg] \quad (19) \\ t = 1, 2, \ldots \\ \Lambda_0 = 0 \end{aligned}$$

For nonlinear state space models where the EKF state estimate is used, innovations ν_t are only approximations. It is known that the EKF estimate is often divergent due to linearization error for severe nonlinear system, and then the fault detection may fails. Hence, more stable approximation is required. An approach is to evaluate pdfs $p(y_t|y_{1:t-1}, H_m)$, $(m = 0, 1)$ approximately by using the particle filters for general nonlinear non-Gaussian models.

III. Particle Filters

Here, we briefly explain the particle filters, Sequential importance sampling particle filter (SIS), Sampling importance resampling particle filter (SIR) and theirs evolutionary computationally modification, Evolution Strategies based particle filter (ESP) that forms the basis for development of the new fault detection methods for nonlinear non-Gaussian state space models.

State estimation problem can be solved by calculating the posterior pdf of the state variable x_t of time instant t based on all the available data of observation sequence $y_{1:t}$.

The posterior pdf $p(x_t|y_{1:t})$ of x_t based on the observation sequence $y_{1:t}$ satisfies the following recursions:

$$p(x_t|y_{1:t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1})dx_{t-1} \quad \text{(Chapman-Kolmogorov equation)} \quad (20)$$

$$p(x_t|y_{1:t}) = \frac{p(y_t|x_t)p(x_t|y_{1:t-1})}{p(y_t|y_{1:t-1})} \quad \text{(Bayes' rule)} \quad (21)$$

with a prior pdf $p(x_0|y_0) \equiv p(x_0)$ of the initial state variable x_0. Here normalizing constant

$$p(y_t|y_{1:t-1}) = \int p(y_t|x_t)p(x_t|y_{1:t-1})dx_t$$

depends on the likelihood $p(y_t|x_t)$, which is determined by the observation equation (2).

In particle filters, the true posterior pdf is approximated by the following weighted empirical distribution of a set of $n \gg 1$ samples $\{x_{t|t}^{(i)}, \ (i = 1, \ldots, n)\}$ called as particles or discrete grids with associated importance weights $\{w_{t|t}^{(i)}, \ (i = 1, \ldots, n)\}, w_{t|t}^{(i)} > 0, \ \sum_{i=1}^n w_{t|t}^{(i)} = 1$,

$$p(x_t|y_{1:t}) \approx \sum_{i=1}^{n} w_{t|t}^{(i)} \delta(x_t - x_{t|t}^{(i)}) \quad (22)$$

where $\delta(\cdot)$ is Dirac's delta function ($\delta(x) = 1$ for $x = 0$ and $\delta(x) = 0$ otherwise).

Here, the particles are generated and associated weights are chosen using the principle of "importance sampling"[8]:

If the samples $x_{t|t}^{(i)}$ in (22) were drawn from an importance density $q(x_t|y_{1:t})$, then the associated normalized weights are defined by

$$w_{t|t}^{(i)} \propto \frac{p(x_{t|t}^{(i)}|y_{1:t})}{q(x_{t|t}^{(i)}|y_{1:t})}. \quad (23)$$

When the importance density $q(x_t|y_{1:t-1})$ is chosen to factorize such that

$$q(x_t|y_{1:t}) = q(x_t|x_{t-1}, y_{1:t})q(x_{t-1}|y_{1:t-1}). \quad (24)$$

Then we can obtain samples $x_{t|t}^{(i)}$ by augmenting each of the existing samples $x_{t-1|t-1}^{(i)}$ sampled from the importance density $q(x_{t-1}|y_{1:t-1})$ with the new state sampled from $q(x_t|x_{t-1}, y_{1:t})$.

Noting that

$$\begin{aligned} p(x_t|y_{1:t}) &= \frac{p(y_t|x_t, y_{1:t-1})p(x_t|y_{1:t-1})}{p(y_t|y_{1:t-1})} \\ &\propto p(y_t|x_t)p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1}) \end{aligned} \quad (25)$$

we have

$$\begin{aligned} w_{t|t}^{(i)} &\propto \frac{p(y_t|x_{t|t}^{(i)})p(x_{t|t}^{(i)}|x_{t-1|t-1}^{(i)})p(x_{t-1|t-1}^{(i)}|y_{1:t-1})}{q(x_{t|t}^{(i)}|x_{t-1|t-1}^{(i)}, y_{1:t})q(x_{t-1|t-1}^{(i)}|y_{1:t-1})} \\ &= w_{t-1|t-1}^{(i)} \frac{p(y_t|x_{t|t}^{(i)})p(x_{t|t}^{(i)}|x_{t-1|t-1}^{(i)})}{q(x_{t|t}^{(i)}|x_{t-1|t-1}^{(i)}, y_{1:t})}. \end{aligned} \quad (26)$$

The particle filter with these steps is called "Sequential importance sampling particle filter" (SIS).

It is known that the SIS filter suffers from the degeneracy phenomenon, where all but one of the normalized importance weights are very close to zero after a few iterations. By this degeneracy, a large computational effort is wasted to updating trajectories whose contribution to the final estimate is almost zero. In order to prevent this phenomenon, resampling process is usually introduced. Its idea is to eliminate trajectories whose normalized importance weights are small and to concentrate upon the trajectories with larger weights. It involves generating new grid points ${x_{t|t}^*}^{(i)}$ $(i = 1, \ldots, n)$ by resampling from the grid approximation (22) randomly with probability

$$\Pr({x_{t|t}^*}^{(i)} = x_{t|t}^{(j)}) = w_{t|t}^{(j)} \quad (27)$$

and the weights are reset to ${w_{t|t}^*}^{(i)} = 1/n$. If the effective sample size N_{eff} [14] defined by

$$\hat{N}_{eff} = \frac{1}{\sum_{i=1}^n (w_{t|t}^{(i)})^2} \quad (28)$$

with the associated normalized weight $w_{t|t}^{(i)}$, is smaller than a predetermined threshold $N_{thres} \in [1, n]$, it is decided that the severe degeneracy is occurred and resampling step should be introduced. Particle filter with this resampling process is called "Sampling importance resampling particle filter" (SIR).

A. Evolution Strategies Based Particle Filters

A novel particle filter called Evolution strategies based particle filter (ESP) is proposed in [21] by recognizing that the importance sampling and resampling processes in SIR filter are corresponding to mutation and selection processes in Evolution Strategies (ES) [18], an Evolutionary Computation approach [4], [9], [3]. Resampling process in SIR filter selects offspring with probability (26), and this corresponds to selection process in ES with fitness function $w_{t|t}^{(i)}$. On the other hand, the importance sampling process in SIR filter samples $x_{t|t}^{(i)}$ according to the importance density $q(x_t|x_{t-1|t-1}^{(i)}, y_{1:t})$, and this corresponds to mutation process in ES from the viewpoint of generating offspring $x_{t|t}^{(i)}$ from the extrapolated parents $f(x_{t-1|t-1}^{(i)})$ with perturbation by v_t. The main difference is resampling in SIR is carried out probabilistically and the weights are reset as $1/n$, while the selection in ES is deterministic and the fitness function is never reset. Hence, by replacing the resampling process in SIR by the deterministic selection process in ES, we can derive a new particle filter as follows.

Based on the particles $x^{(i)}_{t-1|t-1}$ $(i = 1,\ldots,n)$ sampled from the importance density $q(x_{t-1}|y_{1:t-1})$, we generate r samples $x^{(i,j)}_{t|t}$, $(j = 1,\ldots,r)$ from the importance density function $q(x_t|x^{(i)}_{t-1|t-1}, y_{1:t})$. Corresponding weights $w^{(i,j)}_{t|t}$ are evaluated by

$$w^{(i,j)}_{t|t} = w^{(i)}_{t-1|t-1} \frac{p(y_t|x^{(i,j)}_{t|t})p(x^{(i,j)}_{t|t}|x^{(i)}_{t-1|t-1})}{q(x^{(i,j)}_{t|t}|x^{(i)}_{t-1|t-1}, y_{1:t})}$$
$$i = 1,\ldots,n,\ j = 1,\ldots,r$$

From the set of nr particles and weights $\{x^{(i,j)}_{t|t}, w^{(i,j)}_{t|t}, (i = 1,\ldots,n,\ j = 1,\ldots,r)\}$, we choose n sets with the larger weights, and set as $x^{(i)}_{t|t}, w^{(i)}_{t|t} (i = 1,\ldots,n)$. This process corresponds to (n, nr)-selection in ES. Hence, we call this particle filter using (n, nr)-selection in ES as Evolution strategies based particle filter comma (ESP(,)). When we add the particles $x^{(i,0)}_{t|t} = f(x^{(i)}_{t-1|t-1})$, $(i = 1,\ldots,n)$ in addition to nr samples $x^{(i,j)}_{t|t}$, $(i = 1,\ldots,n,\ j = 1,\ldots,r)$ from the importance density function $q(x_t|x^{(i)}_{t-1|t-1}, y_{1:t})$ as above and evaluate the weights $w^{(i,j)}_{t|t}, (i = 1,\ldots,n, j = 0,\ldots,r)$ by (29), and then choose n sets of $(x^{(i)}_{t|t}, w^{(i,j)}_{t|t})$ with larger weights from the ordered set of $n(r+1)$ particles $\{x^{(i,j)}_{t|t}, w^{(i,j)}_{t|t}, (i = 1,\ldots,n,\ j = 0,\ldots,r)\}$, we can obtain another ESP filter. Since this ESP filter uses the selection corresponding to $(n + nr)$-selection in ES, we can call this filter as Evolution strategies based particle filter plus (ESP(+)). The algorithms are summarized in Fig.1.

As shown in the 2-dimensional plots of squared errors at $t = 1000$ and processing time [s] until $t = 1000$ (Fig.2), the ESP filters behave more stable than SIR both in squared estimation errors and processing time by their deterministic selection process [22], we will develop fault detection methods using the ESP.

IV. Fault Detection by Evolution Strategies Based Particle Filters

In the fault detection by using the BLLR (14), the statistics (15) is used and it is necessary to compute

$$\ell_t = \log \frac{p(y_t|y_{1:t-1}, H_1)}{p(y_t|y_{1:t-1}, H_0)} \quad t = 1, 2, \ldots \tag{29}$$

where $p(y_t|y_{1:t-1}, H_m)$ is the one step output prediction density of y_t under the hypothesis H_m, $(m = 0, 1)$.

Using the grid approximation (22)

$$p(x_t|y_{1:t}, H_m) \approx \sum_{i=1}^{n} w^{(i,m)}_{t|t} \delta(x_t - x^{(i,m)}_{t|t}), \quad (m = 0, 1) \tag{30}$$

Procedure *ESP*

```
For   t = 0
      i = 1,...,n, sample x^(i)_{0|0} ~ q(x_0|y_0);
      i = 1,...,n, evaluate the weight
        w^(i)_{0|0} = p(y_0|x^(i)_{0|0}) p(x^(i)_{0|0}) / q(x^(i)_{0|0}|y_0).
For   t ≥ 1
      i = 1,...,n
        set x^(i,0)_{t|t} = f(x^(i)_{t-1|t-1})     [underlined]
      j = 1,...,r
        sample x̃^(i,j)_{t|t} ~ q(x_t|x^(i)_{t-1|t-1}, y_{1:t});
      i = 1,...,n and j = 0 [underlined],1,...,r,
        evaluate the weight
        w^(i,j)_{t|t} = w^(i)_{t-1|t-1} p(y_t|x̃^(i,j)_{t|t}) p(x̃^(i,j)_{t|t}|x^(i)_{t-1|t-1}) / q(x̃^(i)_{t|t}|x̃^(i)_{t|t}, y_{1:t})).
      Sort the set of pairs {x̃^(i,j)_{t|t}, w^(i,j)_{t|t}
      (i = 1,...,n, j = 0 [underlined],1,...,r)}
      by the size of w^(i,j)_{t|t} in descending
      order.
      Take the first n x^(i)_{t|t} from the
      ordered set {x̃^(i)_{t|t}, w̃^(i)_{t|t}}.
      i = 1,...,n, normalize the weight
        w^(i)_{t|t} = w^(i)_{t|t} / Σ_{i=1}^n w^(i)_{t|t}.
      Let p(x_t|y_{1:t}) ≈ Σ_{i=1}^n w^(i)_{t|t} δ(x_t − x^(i)_{t|t})
```

Fig. 1. Algorithm for ESP filters. ESP(+): with the underlined part; ESP(,): without the underlined part

where the second superscript m is corresponding to the models, the pdf $p(x_t|y_{1:t-1}, H_m)$ can be approximated as

$$\begin{aligned} & p(x_t|y_{1:t-1}, H_m) \\ &= \int p(x_t|x_{t-1}, H_m)p(x_{t-1}|y_{1:t-1}, H_m)dx_t \\ &= \int p_v(x_t - f^{(m)}(x_{t-1}))p(x_{t-1}|y_{1:t-1}, H_m)dx_t \\ &\approx \sum_{i=1}^{n} w^{(i,m)}_{t-1|t-1} p_v(x_t - f^{(m)}(x^{(i,m)}_{t-1|t-1})),\ (m = 0, 1) \end{aligned} \tag{31}$$

On the other hand, we can approximate the pdf $p(y_t|y_{1:t-1}, H_m)$ in (29) by

$$p(y_t|y_{1:t-1}, H_m) \approx \frac{1}{n}\sum_{i=1}^{n} p_w(y_t - g^{(m)}(x^{(i,m)}_{t|t-1})) \tag{32}$$

where $x^{(i,m)}_{t|t-1}$, $(i = 1,\ldots,n)$ are samples from the pdf

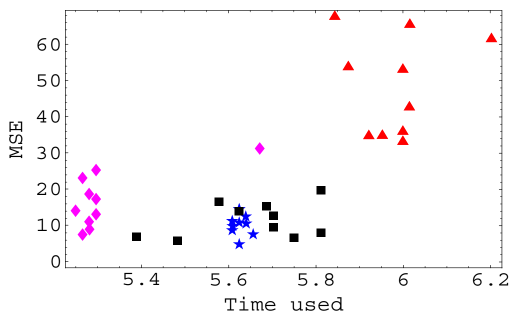


Fig. 2. Squared estimation errors and processing time (triangle: SIS, box: SIR, star: ESP(,), diamond: ESP(+))

$p(x_{t-1}|y_{1:t-1}, H_m)$ given by (31) since

$$\begin{aligned} &p(y_t|y_{1:t-1}, H_m) \\ &= \int p(y_t|x_t, H_m)p(x_t|y_{1:t-1}, H_m)dx_t \\ &= \int p_w(y_t - g^{(m)}(x_t))p(x_t|y_{1:t-1}, H_m)dx_t \end{aligned} \tag{33}$$

Thus, we can conduct the fault detection by evaluating the BLLR Λ_t with the pdf estimates obtained by two ESP filters under the system model H_m, $(m = 0, 1)$ and compare Λ_t with suitable threshold K. The fault detection procedure is summarized in Fig. 3.

V. Numerical Examples

To exemplify the applicability of the proposed ESP filters, we carried out a numerical simulation. We consider the following nonlinear state space model with known parameters.

$$\begin{aligned} x_t &= \frac{x_{t-1}}{2} + \frac{a^{(m)}x_{t-1}}{1 + x_{t-1}^2} + 8\cos(1.2t) + v_t \\ &= f^{(m)}(x_{t-1}) + v_t \quad m = 0, 1 \\ y_t &= \frac{x_t^2}{20} + w_t = g^{(m)}(x_t) + w_t \end{aligned} \tag{34}$$

with $a^{(0)} = 25$ for normal mode and $a^{(1)} = 12.5$ for fault mode, and v_t and w_t are i.i.d. zero-mean Gaussian random variates with variance 10 and 1, respectively. We assume that the fault occurs at $t = \tau = 101$. A sample behavior of the true state and corresponding observation processes is shown in Fig.4. The Gaussian distribution with mean $f(x_{t-1|t-1}^{(i)})$ and variance 10 is chosen as the importance density $q(x_t|x_{t-1|t-1}^{(i)}, y_{1:t})$.

Sample behaviors of state estimates by ESP(,) with $n = 10, r = 2$ based on the model H_m, $(m = 0, 1)$, and BLLR Λ_t and λ_t are given in Fig. 5 with corresponding results by EKF as well for comparison. Since the test statistics BLLR Λ_t takes positive value and is growing up rapidly after the change point τ both in ESP and EKF, we can detect the model change when the BLLR exceeds the suitable threshold K. It is found that the conventioanl LLR λ_t accumulates negative values before τ, and it takes long time to recover this and make delay in detection. This fact indicates the superiority of BLLR Λ_t to the conventional LLR λ_t. On the other hand, it should be noted that, as shown in Fig. 5, the state estimate by EKF shows poor behavior and hence the behavior of test statistics sometimes provides poor detection result. Eventually, the rate of false alarm[1] and miss alarm[2] are higher by the detection procedure using EKF than by the procedure using ESP as shown in Table I that summarizes 10 simulation results of fault detection with the

Procedure *Fault detection by ESP*

```
For t = 0
  set Λ_0 = 0;
  m = 0,1
    approximate the pdf p(x_0|y_0, H_m)
    as ESP by
    p(x_0|y_0, H_m) ≈ Σ_{i=1}^n w_{0|0}^{(i,m)} δ(x_t − x_{0|0}^{(i,m)}).
For t ≥ 1
  m = 0,1
  approximate the pdf p(x_t|y_{1:t−1}, H_m) by
    p(x_t|y_{1:t−1}, H_m)
      ≈ Σ_{i=1}^n w_{t−1|t−1}^{(i,m)} p_v(x_t − f^{(m)}(x_{t−1|t−1}^{(i,m)})).
  i = 1,...,n
    sample x_{t|t−1}^{(i,m)} ~ p(x_t|y_{1:t−1}, H_m).
    evaluate
    p(y_t|y_{1:t−1}, H_m) = Σ_{i=1}^n p_w(y_t − g^{(m)}(x_{t|t−1}^{(i,m)}))/n.
  evaluate
```

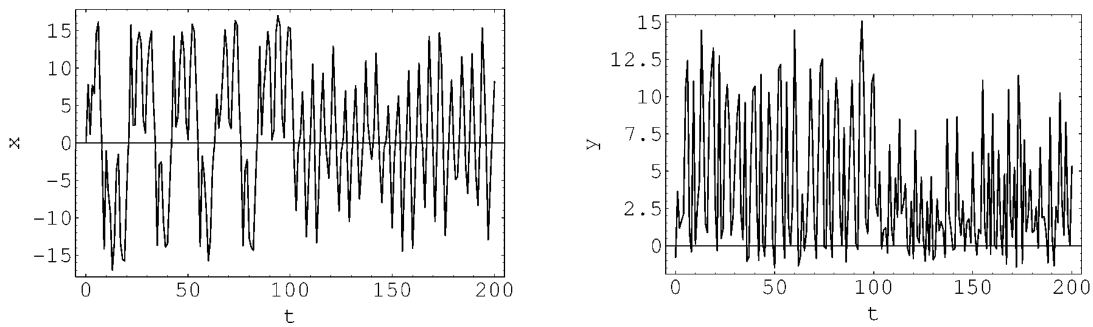
$$\Lambda_t = \max\left[0, \Lambda_{t-1} + \log\frac{\sum_{i=1}^{n} p_w(y_t - g^{(1)}(x_{t|t-1}^{(i,1)}))}{\sum_{i=1}^{n} p_w(y_t - g^{(0)}(x_{t|t-1}^{(i,0)}))}\right].$$

```
  if Λ_t > k then stop observations with
  acceptance of the hypothesis that
  the system is in the fault mode.
  m = 0,1
    approximate the pdf p(x_t|y_{1:t}, H_m) as
    ESP by
    p(x_t|y_{1:t}, H_m) ≈ Σ_{i=1}^n w_{t|t}^{(i,m)} δ(x_t − x_{t|t}^{(i,m)}).
```

Fig. 3. Fault detection procedure by ESP filter

[1]The decision that the system model has changed is made even when the system model does not change. In this example, the test statistics exceeds the threshold between $t = 0$ and $t = 100 < \tau$.

[2]The decision that the system model does not change is made even when model changes. In this example, the test statistics never exceeds the threshold beween $t = 101 = \tau$ and $t = 200$.



(a) State

(b) Observation

Fig. 4. Sample behavior of state and observation processes

TABLE I

FAULT DETECTION RESULT

	False alarm rate		Miss alarm rate	
	$K = 10$	$K = 25$	$K = 10$	$K = 25$
Fault detection by ESP	5/20	1/20	0/20	0/20
Fault detection by EKF	12/20	5/20	0/20	1/20

threshold[3] K =?10 and $K = 25$. These results illustrate the applicability of the proposed approach for fault detection of nonlinear stochastic state space models. By introducing the other choice of evolution processes such as crossover and suitable choice of evolution parameters it is expected the improvement of the performance, and their better choice will be pursued.

VI. Conclusions

Fault detection in dynamic systems have attracted considerable attention in designing systems with safety and reliability. Though a large number of methods have been proposed for solving the fault detection problem, it is hardly apply to nonlinear stochastic state space models. A novel filter called the Evolution Strategies based particle filter (ESP) proposed by recognizing the similarities and the difference of the processes between the particle filters and Evolution Strategies is applied here to fault detection of nonlinear stochastic state space models. Numerical simulation studies have been conducted to exemplify the applicability of this approach.

References

[1] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Englewood Cliffs: Prentice-Hall, 1979.

[2] S. Arulampalam, S. Maskell, N. Gordon and T. Clapp, "A tutorial on particle filters for on-line non-linear/non-Gaussian Bayesian tracking," *IEEE Trans. on Signal Processing*, vol. SP-50, no.2, pp. 174–188, 2002.

[3] T. Bäck, *Evolutionary Computation*, London: Oxford University Press, 1996.

[4] T. Bäck and H.-P. Schwefel, "An overview of evolutionary algorithms for parameter optimization," *Evolutionary Computation*, vol. 1, no. 1, pp. 1–23, 1993.

[5] M. Basseville, "Detecting changes in signals and systems — A survey," *Automatica*, vol. 24, no. 3, pp. 309–326, 1988.

[6] M. Basseville and A. Benveniste (Eds.), *Detection of Abruptly Changes in Signals and Dynamic Systems*, Berlin: Springer, 1986.

[7] M. Basseville and I. Nikiforov, *Detection of Abrupt Changes — Theory and Application*, Englewood Cliffs: Prentice-Hall, 1993.

[8] A. Doucet, N. de Freitas and N. Gordon (Eds.), *Sequential Monte Carlo Methods in Practice*, Berlin: Springer-Verlag, 2001.

[9] D. B. Fogel, *Evolutionary Computation*, New York: IEEE Press, 1995.

[10] P. M. Frank, "Fault diagnosis in dynamic systems using analytical knowledge based redundancy — A survey and new results," *Automatica*, vol. 26, pp. 459–474, 1990.

[11] R. Isermann, "Process fault detection based on modelling and estimation methods — A survey," *Automatica*, vol. 20, pp. 387–404, 1984.

[12] A. H. Jazwinski, *Stochastic Process and Filtering Theory*, New York: Academic Press, 1970.

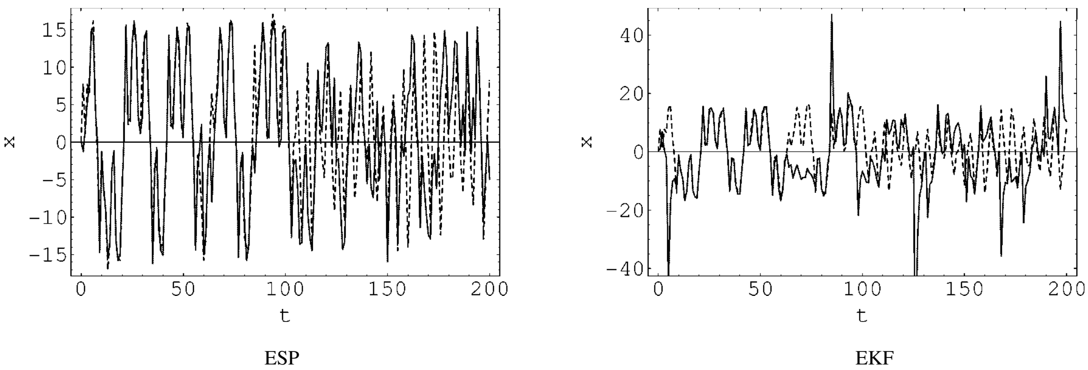
[13] R. Kalman and R. Bucy, "New results in linear filtering and prediction theory," *J. of Basic Engineering, Trans. ASME Series D*, vol. 83, pp. 95–108, 1961.

[14] A. Kong, J. S. Liu and W.H. Wong, "Sequential imputations and Bayesian missing data problems," *J. of American Statistical Association*, vol. 89, pp. 278–288, 1994.

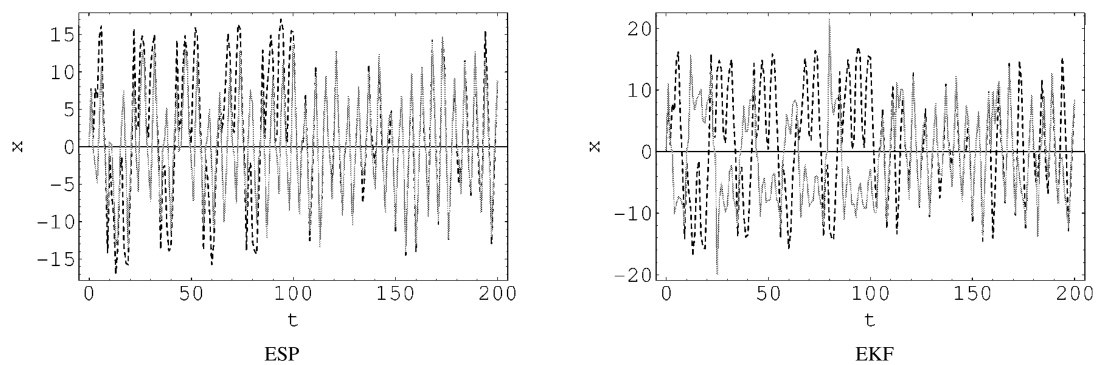
[15] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, Boston: MIT Press, 1992.

[16] J. Liu and R. Chen, "Sequential Monte Carlo Methods for Dynamic

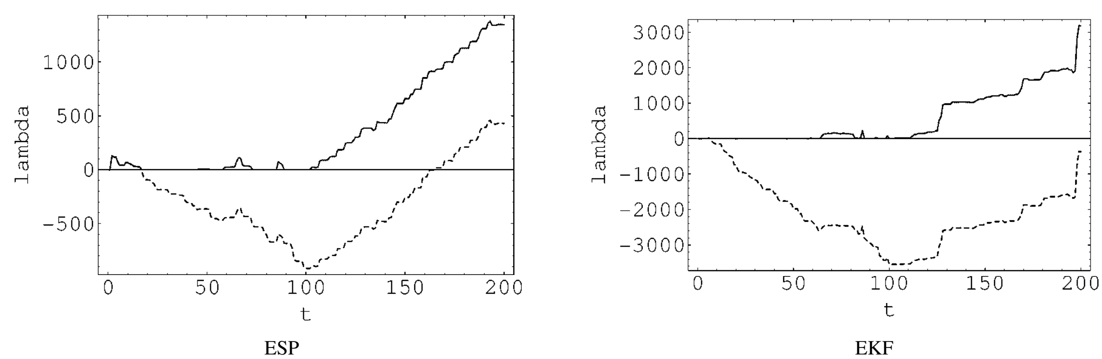
[3]This threshold is determined by trial and error.



(a) State estimate with model H_1 (solid line : estimate; dotted line : true state)



(b) State estimate with model H_2 (solid line : estimate; dotted line : true state)



(c) BLLR Λ_t and conventional LLR λ_t (solid line : Λ_t; dotted line : λ_t)

Fig. 5. Sample behaviors of state estimates and test statistics by ESP and EKF

Systems," *J. of American Statistical Association*, vol. 93, pp. 1032–1044, 1998.

[17] D. B. Rubin, "Using the SIR algorithm to simulate posterior distribution," in *Bayesian Statistics*, Edited by J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, vol. 3, pp. 395–402, London: Oxford University Press, 1988.

[18] H.-P. Schwefel, *Evolution and Optimum Seeking*, New York: J. Wiley, 1995.

[19] H. W. Sorenson, *Kalman Filtering: Theory and Application*, New York: IEEE Press, 1985.

[20] K. Uosaki, "Failure detection using backwatrd SPRT," *Proc. 7th IFAC Symp. System Identification and Parameter Estimation*, pp. 1619–1624, 1985.

[21] K. Uosaki, Y. Kimura and T. Hatanaka, "Nonlinear state estimation by evolution strategies based particle filters," *Proc. Congress on Evolutionary Computation 2003*, pp. 2102–2109, 2003.

[22] K. Uosaki and T. Hatanaka, "Nonlinear state estimation by evolution strategies based particle filters," *Proc. 16th IFAC World Congress on Automatic Control*, 2005.

[23] A. Wald, *Sequential Analysis*, New York: J. Wiley, 1947.

[24] A. S. Willsky, "A survey of design methods for failure detection in dynamic systems," *Automatica*, vol. 12, pp. 601–611, 1976.