

State Estimation And Fault Diagnosis Of Industrial Process

By Using of Particle Filters

SAREH BAHMANPOUR¹, MAHDI BASHOOKI¹, M. H. REFAN²

¹MAPNA Electrical and Control Company (MECO), Karaj, IRAN

²S. Rajae University, Lavizan, Tehran 16788, IRAN

Abstract: - It is no surprise that the general problem of fault detection has received considerable attention in a wide variety of industrial systems such as chemical and petrochemical industries, autonomous underwater vessels, robots, and even space vehicles. We have presented a probabilistic approach to state estimation and fault diagnosis in complex industrial processes. In particular, we adapted a Jump Markov Linear Gaussian (JMLG) model to describe a continuous stirred tank reactor. Expectation Maximization (EM) algorithm identifies the parameters of this process. After identification, real-time particle filtering algorithms were adapted to diagnosis of the state of operation of continuous stirred tank reactor. For this application, we compared two particle-filtering variants: standard particle filtering and Rao-Blackwellised particle filtering. The particle filtering estimates were then used to drive an automatic control system.

Keywords: - State estimation, Fault diagnosis, JMLG model, Particle filtering, continuous stirred tank reactor.

1 Introduction

In vast varieties of areas such as robot navigation and diagnosis of complex industrial systems, Real-time monitoring is very important [1, 2]. In this research, we concentrated on online monitoring of complex industrial processes. These processes have a number of discrete states, corresponding to different combinations of faults. Based on the discrete states, the dynamics can be very different. Even if there are very few discrete states, exact monitoring is computationally unfeasible as the state of the system depends on the history of the discrete states. However there is a need to monitor these systems in real time to determine what faults could have occurred.

Here, we proposed a real-time, automatic strategy in order to estimate the states of industrial processes using noisy measurements of continuous variables. This approach enables us to reduce the cognitive load experienced by human operators. It also serves to minimize the number of instruments and to open up room for sophisticated control strategies.

In particular, we adopt a Jump Markov Linear Gaussian (JMLG) model to describe a continuous stirred tank reactor with different linear regimes of operation. A discrete state variable controls the switching between the various linear regimes. The

parameters of each regime are identified off-line with the Expectation Maximization (EM) algorithm [3]. Once the stationary parameters have been identified, real-time Rao-Blackwellised Particle Filtering (RBPF) algorithms are used to execute on-line estimation of the continuous and discrete states of the system [4, 5]. These estimates are used to determine the control policy of a PID controller.

2 Monitoring Process

System monitoring and timely fault detection capabilities are critical requirements of many modern systems. For years, these features have only been of utmost importance in safety critical systems, such as civil and military aviation or nuclear power plants. However, recently other factors have been playing a major role in recognizing the need of these capabilities in other technical systems.

Processes in the chemical and petrochemical industries are becoming larger and more complex. Associated of this development, each hour of down time has a high cost, and the source of malfunction or fault is more difficult to locate. The purpose of monitoring of faults is to reduce the occurrence of sudden, disruptive, or dangerous outages, equipment

damage, and personal accidents, as well as assistance of the operation of the maintenance program.

Diagnosis can detect/determine of faults that could occur in the process itself, in its measuring instruments, or in its actuators. By state estimation we mean the identification of different operating conditions which the process can be in. In dynamic industrial processes, we consider a continuous stirred tank reactor, which has exclusively modeled with dynamic systems characterized by a continuous-time operation. If the dynamic system has non-linear behavior, it has to be modeled as a series of linear segments. Let us consider a multi-input multi-output continuous dynamic system, as defined below.

2.1 Continuous Stirred Tank Reactor

This reactor consists of a well -stirred tank containing the enzyme, which is normally immobilized. The substrate stream is continuously pumped into the reactor at the same time as the product stream is removed. If the reactor is behaving in an ideal manner, there is total back-mixing and the product stream is identical with the liquid phase within the reactor and invariant with respect to time. Some molecules of substrate may be removed rapidly from the reactor, whereas others may remain for substantial periods. A Continuous Stirred Tank Reactor (CSTR) is a complex nonlinear, multivariable system; as shown in Figure 1. It involves a second-order exothermic reaction ($2A \rightarrow B$), where 2 components of A react irreversibly at specific reaction rate of k, in order to form product of B [6].

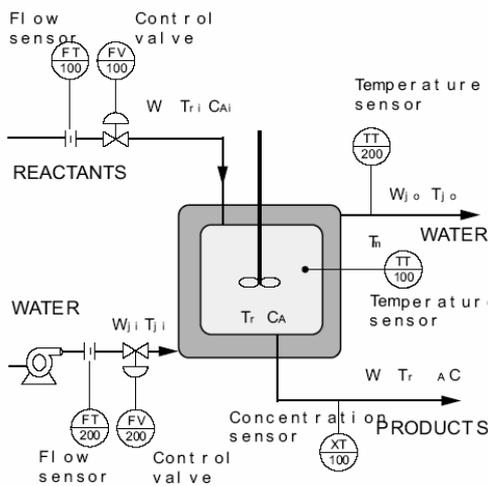


Figure 1. Continuous Stirred Tank Reactor

Where the reaction rate constant, k, follows the Arrhenius equation (1). According to this equation, the effect of the temperature, $T_r(t)$ on the specific reaction rate of k is usually exponential. This exponential temperature dependence represents one of the most severe nonlinearities in chemical engineering systems. The overall reaction rate, R, is defined as the rate of change of moles of any component per volume due to chemical reaction divided by that component's stoichiometric coefficient. Due to this reaction, we have $R = kC_A^2$. Then, the overall rate, R will vary with temperature, $T_r(t)$ and with the concentration of the reactant C_A raised to the 2nd power (second-order reaction). As we can see, this term R is highly nonlinear.

$$k(T_r(t)) = k_0 e^{\left[\frac{-a}{T_r(t)+460}\right]} \quad (1)$$

2.2 Modeling

The mathematical model for this CSTR involves a mass balance on A component, as expressed by equation (2), commonly known as a component continuity equation.

The first law of thermodynamics puts forward the principle of conservation of energy. The mathematical model must include an enthalpy balance on reaction mass, and an enthalpy balance on jacket (which water is flowing through in). In this case, the flow of internal energy into the system, minus the flow of internal energy out of the system, plus the heat added to the system by reaction is equal to the rate of change of internal energy inside the system. The balance on reacting mass is given by equation (3) and the balance on the jacket by equation (4).

$$V\rho \frac{dC_A}{dt} = W(C_{Ai} - C_A) - kC_A^2V\rho \quad (2)$$

$$V\rho C_p \frac{dT_r}{dt} = WC_p(T_{ri} - T_r) - UA(T_r - T_{jo}) + (-\Delta H)VkC_A^2 \quad (3)$$

$$M_j C_{pj} \frac{dT_{jo}}{dt} = UA(T_r - T_{jo}) - W_j C_{pj} (T_{jo} - T_{ji}) \quad (4)$$

Some assumptions were made to derive these equations. These equations represent a very simplified nonlinear CSTR model (the t functionality was omitted for clarity). Nevertheless, this simpler model captures the nonlinearity we are interested in. we will only measure the following 3 variables: the output concentration $C_A(t)$, the reactor temperature $T_r(t)$, and the output jacket temperature $T_{jo}(t)$.

Table 1 demonstrates a complete definition of the variables [9] and Figure 1 conveys their meaning graphically as well as some instruments for monitoring and control purposes. Table 2 gives a complete description of this instrumentation.

Table1. Variables

Var	Value	Units	Definitions
C_A	3.5955	lb / ft^3	Concentration of reactant A in reactor and exit stream
C_{Ai}	10.8	lb / ft^3	Concentration of reactant A in feed
a	2.560	$^{\circ}R$	Constant in Arrhenius expression For reaction rate
k	0.0278	$ft^3 / lb - min$	Reaction rate constant
k_0	1.43	$ft^3 / lb - min$	Constant in Arrhenius expression
$-\Delta H$	867	Btu / lb_A	Heat of reaction
C_p	0.9	$Btu / lb^{\circ}F$	Specific heat of reacting mixture
C_{pj}	1.0	$Btu / lb^{\circ}F$	Specific heat of water
A	500	ft^2	Effective jacket transfer area
ρ	80	lb / ft^3	Density of reacting mixture
U	1.2	$Btu / min ft^2 ^{\circ}F$	Heat transfer coefficient
T_{ri}	150	$^{\circ}F$	Input reactants temperature
T_r	190.0611	$^{\circ}F$	Reactor temperature
T_{rm}	190.0611	$^{\circ}F$	Measured reactor temperature
T_{jo}	120.0222	$^{\circ}F$	Outlet Jacket temperature
T_{ji}	80	$^{\circ}F$	Inlet Jacket temperature
V	250	ft^3	Reactor Volume
W	1000	lb / min	Feed mass flow rate
W_j	1050	lb / min	Water cooling rate at jacket
M_j	4000	lb	Mass of jacket water

The nonlinear model described by equations (2-4) was linearized to build the JMLG model. Then, four discrete modes were tested.

Table2. CSTR Instrumentation

Tag-name	Functional name	Description
FT-100	Flow sensor/transmitter	Input reactants flow
FT-200	Flow sensor/transmitter	Input water flow
FV-100	Control valve	Reactants flow valve
FV-200	Control valve	Water flow valve
TT-100	Temperature sensor/transmitter	Reactor temperature
TT-200	Temperature sensor/transmitter	Output water temperature
XT-100	Analyzer sensor/transmitter	Output products concentration

We consider a fouled surface (dirty surface) in the jacket as a possible faulty point (of course, there are many possible faulty points in this system). A fouled surface can be caused by normal operating conditions over an extended time, or by stochastic problems such as cooling water with a high concentration of minerals or salts. Surface fouling reduces the global heat transfer coefficient $U(t)$ in the mathematical model. We defined four of the possible discrete modes for this nonlinear multivariable system as below in Table 3.

Table3. CSTR Operating conditions

Z	Model Name	Description	Variation
1	Normal	Clean heat transfer area	none
2	Fouling-1	Dirty heat transfer area	5% fouling
3	Fouling-2	Dirty heat transfer area	10%Fouling
4	Fouling-3	Dirty heat transfer area	15%Fouling

We obtain the sampled state-space representation using the continuous state-space representation. The continuous state-space is generated by the system of linear differential equations.

For the “normal” discrete mode, $z_i = 1$:

$$x_{t+1} = A(z_{t+1})x_t + B(z_{t+1})w_{t+1} + F(z_{t+1})u_{t+1} \tag{5}$$

$$y_t = C(z_t)x_t + D(z_t)v_t + G(z_t)u_t \tag{6}$$

Where:

$$x_t = \begin{bmatrix} C_A(t) \\ T_r(t) \\ T_{jo}(t) \end{bmatrix} \quad y_t = \begin{bmatrix} C_A(t) \\ T_r(t) \\ T_{jo}(t) \end{bmatrix} \quad u_t = \begin{bmatrix} u_1(t) \\ u_2(t) \\ u_3(t) \end{bmatrix}$$

w_t and v_t are the process and measurement noises; both follow $N(0,1)$. The matrices are:

$$A(z_t = 1) = \begin{bmatrix} 0.9752 & -0.000214 & -3.55(10)^{-7} \\ 0.2376 & 0.9943 & 0.00325 \\ 0.00176 & 0.01465 & 0.9596 \end{bmatrix}$$

$$F(z_t = 1) = \begin{bmatrix} 0.1299 & 0 & 0 \\ -0.1635 & 0 & 0 \\ 2.0560 & 0 & 0 \end{bmatrix}$$

$$B(z_t = 1) = n_{process} I^{3 \times 3}, \quad C(z_t = 1) = I^{3 \times 3}, \quad D(z_t = 1) = n_{measurement} I^{3 \times 3}, \quad G(z_t = 1) = 0^{3 \times 3}.$$

The values of $n_{process}$ and $n_{measurement}$ are fixed for each test. Corresponding results were obtained for each faulty discrete mode ($z_t = 2, 3, \text{ and } 4$). Where

the initial states are $x_0 \sim N(\mu_0, \Sigma_0)$ and $z_0 \sim P(z_0)$. That is important to notice that for each realization of z_t , we have a single linear-Gaussian model. If we know z_t , we could solve x_t exactly, using the Kalman filter algorithm.

The aim of the analysis is to compute the marginal posterior distribution of the discrete states $P(z_{0:t} | y_{1:t})$. This distribution can be derived from the posterior distribution $P(dx_{0:t}, z_{0:t} | y_{1:t})$ by standard marginalization. The posterior density satisfies the following recursion:

$$p(x_{0:t}, z_{0:t} | y_{1:t}) = p(x_{0:t-1}, z_{0:t-1} | y_{1:t-1}) \frac{p(y_t | x_t, z_t) p(x_t, z_t | x_{t-1}, z_{t-1})}{p(y_t | y_{1:t-1})} \quad (7)$$

This recursion involves intractable integrals. One, therefore, has to resort to some form of numerical approximation scheme.

3 Problem Solutions

Most existing model-based fault diagnosis methods use a technique called analytical redundancy [7]. Real measurements of a process variable are compared to analytically calculated values. The resulting differences, named residuals, are indicative of faults in the process. Many of these methods rely on simplifications and heuristics. Here, we propose a principled probabilistic approach to this problem.

3.1 Particle Filtering

In the PF setting, we used a weighted set of samples (particles) $\{(x_{0:t}^{(i)}, z_{0:t}^{(i)})\}_{i=1}^N$ to approximate the posterior with the following point-mass distribution

$$\hat{P}_N(dx_{0:t}, z_{0:t} | y_{1:t}) = \sum_{i=1}^N w_t^{(i)} \delta_{x_{0:t}^{(i)}, z_{0:t}^{(i)}}(dx_{0:t}, z_{0:t}) \quad (8)$$

Where $\delta_{x_{0:t}^{(i)}, z_{0:t}^{(i)}}(dx_{0:t}, z_{0:t})$ denotes the Dirac-Delta function.

Given N Particles $\{(x_{0:t-1}^{(i)}, z_{0:t-1}^{(i)})\}_{i=1}^N$ at time t-1, approximately distributed according to $P(dx_{0:t-1}, z_{0:t-1} | y_{1:t-1})$, PF enables us to compute N particles $\{(x_{0:t}^{(i)}, z_{0:t}^{(i)})\}_{i=1}^N$ approximately distributed according to $P(dx_{0:t}, z_{0:t} | y_{1:t})$, at time t. Since we cannot take samples from the posterior directly, the PF update is accomplished by introducing an appropriate importance proposal distribution $Q(dx_{0:t}, z_{0:t})$ from

which we can obtain samples. The basic algorithm, Figure 2, consists of two steps: sequential importance sampling and selection (see [5] for a detailed derivation). This algorithm uses the transition priors as proposal distributions; for the selection step, we used a state-of-the-art minimum variance resampling algorithm [8].

3.2 Rao-Blackwellised Particle Filtering

By considering the factorization $p(x_{0:t}, z_{0:t} | y_{1:t}) = p(x_{0:t} | y_{1:t}, z_{0:t}) p(z_{0:t} | y_{1:t})$, it is possible to design more efficient PF algorithms.

The density $p(x_{0:t} | y_{1:t}, z_{0:t})$ is Gaussian and can be computed analytically if we know the marginal posterior density, $p(z_{0:t} | y_{1:t})$. This density satisfies the alternative recursion:

$$p(z_{0:t} | y_{1:t}) = p(z_{0:t-1} | y_{1:t-1}) \frac{p(y_t | y_{1:t-1}, z_{0:t}) p(z_t | z_{t-1})}{p(y_t | y_{1:t-1})} \quad (9)$$

Sequential Importance Sampling

For $i=1, \dots, N$, sample from the transition priors

$$\hat{z}_t^{(i)} \sim p(z_t | z_{t-1}^{(i)})$$

$$\hat{x}_t^{(i)} \sim p(dx_t | z_t^{(i)})$$

And set $(\hat{x}_{0:t}^{(i)}, \hat{z}_{0:t}^{(i)}) = (\hat{x}_t^{(i)}, \hat{z}_t^{(i)}, x_{0:t-1}^{(i)}, z_{0:t-1}^{(i)})$.

For $i=1, \dots, N$, evaluate and normalize the importance weights

$$w_t^{(i)} \propto p(y_t | \hat{x}_t^{(i)}, \hat{z}_t^{(i)})$$

Selection (Resampling step)

Multiply/Discard particles $\{\hat{x}_{0:t}^{(i)}, \hat{z}_{0:t}^{(i)}\}_{i=1}^N$ with respect to high/low importance weights $w_t^{(i)}$ to obtain N particles

$$\{\hat{x}_{0:t}^{(i)}, \hat{z}_{0:t}^{(i)}\}_{i=1}^N.$$

Figure2. Particle Filtering Algorithm

If equation (7) does not admit a closed-form expression, then equation (8) does not admit one either and sampling-based methods are still required. (Also note that the term $p(y_t | y_{1:t-1}, z_{0:t})$ in equation (9) does not simplify to $p(y_t | z_t)$ because there is a dependency on past values through $x_{0:t}$. Now assuming that we can use a weighted set of samples $\{z_{0:t}^{(i)}, w_t^{(i)}\}_{i=1}^N$ to represent the marginal posterior distribution

$$\hat{P}_N(z_{0:t} | y_{1:t}) = \sum_{i=1}^N w_t^{(i)} \delta_{z_{0:t}^{(i)}}(z_{0:t}) \quad (10)$$

The marginal density of $x_{0:t}$ is a Gaussian mixture

$$\hat{P}_N(x_{0:t} | y_{1:t}) = \int p(x_{0:t} | z_{0:t}, y_{1:t}) dP(z_{0:t} | y_{1:t}) = \sum_{i=1}^N w_t^{(i)} p(x_{0:t} | y_{1:t}, z_{0:t}^{(i)}) \tag{11}$$

That can be computed efficiently with a stochastic bank of Kalman filters. That is, we use PF to estimate the distribution of z_t and exact computations (Kalman filter) to estimate the mean and variance of z_t . In particular, we sample $z_t^{(i)}$ and then propagate the mean $\mu_t^{(i)}$ and covariance $\Sigma_t^{(i)}$ of x_t with a Kalman filter:

$$\mu_{t|t-1}^{(i)} = A(z_t^{(i)})\mu_{t-1}^{(i)} + F(z_t^{(i)})u_t \tag{12}$$

$$\Sigma_{t|t-1}^{(i)} = A(z_t^{(i)})\Sigma_{t-1}^{(i)}A(z_t^{(i)})^T + B(z_t^{(i)})B(z_t^{(i)})^T \tag{13}$$

$$S_t^{(i)} = C(z_t^{(i)})\Sigma_{t|t-1}^{(i)}C(z_t^{(i)})^T + D(z_t^{(i)})D(z_t^{(i)})^T \tag{14}$$

$$y_{t|t-1}^{(i)} = C(z_t^{(i)})\mu_{t|t-1}^{(i)} + G(z_t^{(i)})u_t \tag{15}$$

$$\mu_t^{(i)} = \mu_{t|t-1}^{(i)} + \Sigma_{t|t-1}^{(i)}C(z_t^{(i)})^T S_t^{-1(i)}(y_t - y_{t|t-1}^{(i)}) \tag{16}$$

$$\Sigma_t^{(i)} = \Sigma_{t|t-1}^{(i)} - \Sigma_{t|t-1}^{(i)}C(z_t^{(i)})^T S_t^{-1(i)}C(z_t^{(i)})\Sigma_{t|t-1}^{(i)} \tag{17}$$

Where

$$\mu_{t|t-1} \stackrel{\Delta}{=} E(x_t | y_{1:t-1}), \mu_t \stackrel{\Delta}{=} E(x_t | y_{1:t}), y_{t|t-1} \stackrel{\Delta}{=} E(y_t | y_{1:t-1})$$

$$\Sigma_{t|t-1} \stackrel{\Delta}{=} \text{cov}(x_t | y_{1:t-1}), \Sigma_t \stackrel{\Delta}{=} \text{cov}(x_t | y_{1:t}) \quad \text{and} \quad S_t \stackrel{\Delta}{=} \text{cov}(y_t | y_{1:t-1}).$$

This is a basis of the RBPF algorithm that was adopted in [4].

4 Results

We tested the 2 inference algorithms for N=100 and T=50. These simulations were designed using the transition matrix and prior probabilities shown below [9]:

$$P(z_t | z_{t-1}) = \begin{bmatrix} 0.9983 & 0.001 & 0.0005 & 0.00025 \\ 0.001 & 0.9975 & 0.001 & 0.0005 \\ 0.0005 & 0.001 & 0.9975 & 0.001 \\ 0.00025 & 0.0005 & 0.001 & 0.9983 \end{bmatrix}$$

$$P(z_0) = \begin{bmatrix} 0.9983 \\ 0.001 \\ 0.0005 \\ 0.00025 \end{bmatrix}$$

Figure 3 plots the tracking error for each algorithm. As it can be seen, the RBPF algorithm can track the state better than PF algorithm. If we have several states, for example 3 or 4, Figure 4 will be resulted. In this plot, it is clear that RBPF algorithm is again better than the other. Figure 5 and 6 show the states, outputs and discrete modes uses in simulation.

Figure 7 and 8 plots the Probability density of PF and RBPF algorithms.

5 Conclusions

Results show that the RBPF algorithm gives a very low diagnosis error per number of particles. It works significantly better than standard Pf. RBPF also gives a very low diagnosis error per unit of computing time, despite its greater computational expense per particle compared with standard PF. Faulty conditions usually have a very low probabilities. Standard numerical approximations have trouble with this situation because of a very small number of particles are assigned to a faulty discrete mode, despite the observations. However, RBPF samples the possible discrete modes from their true posterior distribution, capturing evidence of faulty conditions and allowing them to be identified. RBPF also gives lower variance than standard PF per number of particles. This advantage, based on the Rao-Blackwell formula, grows as the number of particle is increased.

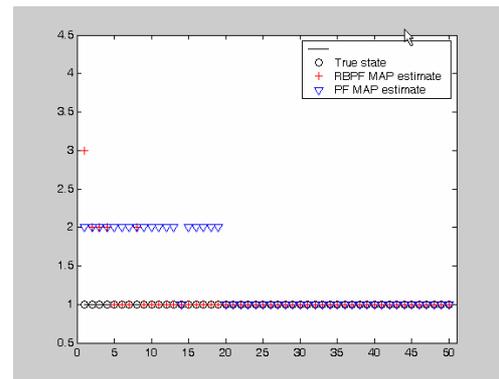


Figure3. CSTR (1 discrete Mode)

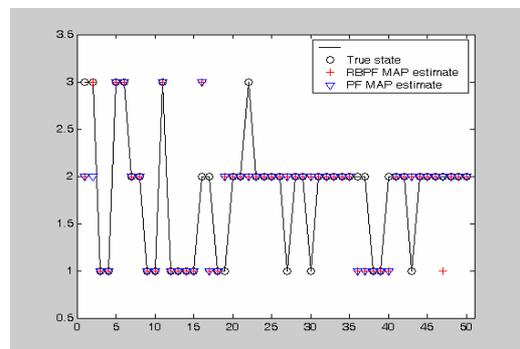


Figure4. CSTR (several discrete modes)

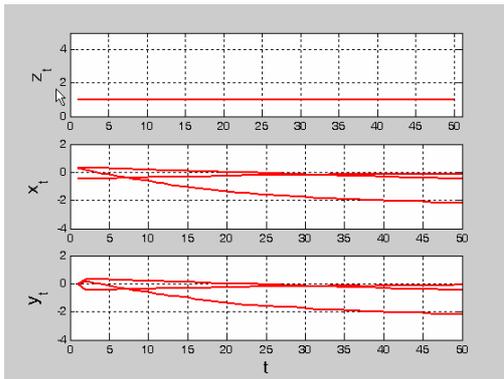


Figure 5. 1 Discrete mode, states, output

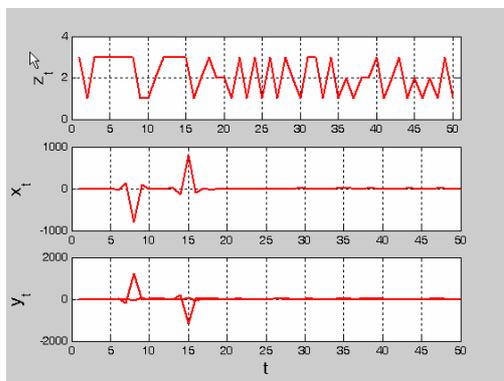


Figure 6. 3 discrete modes, states, outputs

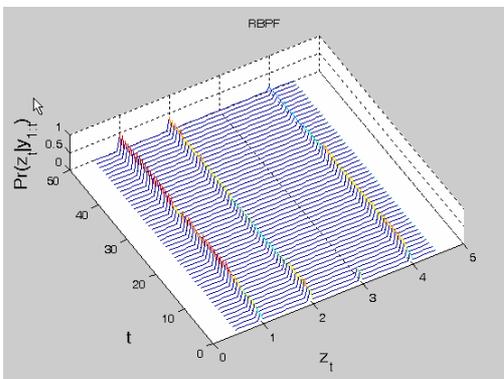


Figure 7. Probability distribution (RBPF algorithm), 1 discrete mode

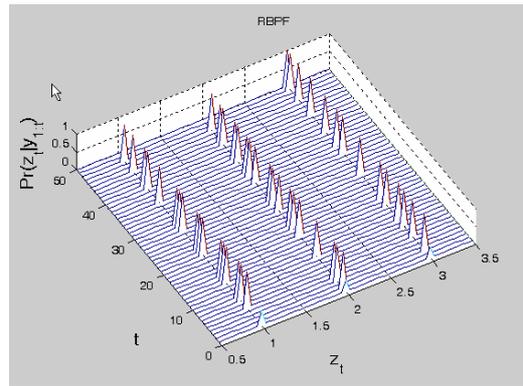


Figure 8. Probability Distribution (RBPF algorithm), 3 discrete modes

References:

- [1] J Chen and J Howell. A self-validating control system based approach to plant fault detection and diagnosis. *Computers and chemical Engineering*, 25:337-358, 2001.
- [2] S Thruan, J Langford, and V Verma. Risk sensitive particle filters. In S Becker T. G Dietterich and Z Ghahramani, editors, *Advances in Neural Information Processing Systems 14*, Cambridge, MA, 2002. MIT Press.
- [3] Z Ghahramani and G E Hinton. Parameter estimation for linear dynamical system. Technical Report CRG-TR-96-2, Department of Computer Science University of Toronto, Toronto, 1996.
- [4] N de Freitas. Rao-Blackwellised particle filtering for fault diagnosis. In *IEEE Aerospace Conference*, 2001.
- [5] A Doucet, N Freitas, and N J Gordon, editors. *Sequential Monte Carlo Methodes in Practice*. Springer-Verlag, 2001.
- [6] W L Luyben. *Process Modeling, Simulation and Control for chemical Engineers*. Mc GrawHill Publishing Company, second edition, 1989.
- [7] J Gertler. *Fault detection and diagnosis in engineering systems*. Marcel dekker, Inc., 1998.
- [8] G Kitagawa. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5: 1-25, 1996.
- [9] R M Menendez. *Real-time monitoring and diagnosis in dynamic systems using particle filtering methods*. PHD Thesis in Intelligent systems, Monterry Institute, Mexico, 2003.