The Particle Filter for joint state & parameter estimation of nonlinear systems

E. Chatzi\textsuperscript{1} and A.W. Smyth\textsuperscript{2}

\textsuperscript{1}Institute of Structural Engineering, ETH Zürich
\textsuperscript{2}Department of Civil Engineering & Engineering Mechanics, Columbia University

Lecture 4 - 06 June, 2013
Energy dissipation is often concentrated at specific elements within a system (eg., joints, dampers). Nonlinear hysteretic behavior common.

Robust adaptive control - *realtime* tracking of degradation. On-Line Identification methods required - Reduced order models.
Problem definition

Formulation in State-Space

Consider the general dynamical system described by the following nonlinear continuous state-space (process) equation

$$\dot{x} = f(x(t), u(t), w(t))$$

and the nonlinear observation equation at time $t = k\Delta t$

$$y(t) = h(x(t), v(t))$$

or in discrete form:

$$x_{k+1} = F(x_k, u_k, w_k)$$

$$y_k = H(x_k, v_k)$$

where $w_k$ is the process noise vector with covariance matrix $Q_k$, $v_k$ is the observation noise vector with corresponding covariance matrix $R_k$, and function $F$ is obtained from $f$ via numerical integration.
Probabilistic Inference Problem

Adjustable Candidate Models of Nonlinear Structural System

Unknown Nonlinear Structural System

Measured Response Motions

\[ M_1 + \_ \]

Measured Restoring Force

Estimated Restoring Force

Output Error

Smoothness & Accuracy Criteria

... Optimal Model

Parameter Adjustment Procedure (UKF)

Adjustable Candidate Models of Nonlinear Structural System
Optimal Bayesian Solution

Predict

Assuming the prior $p(x_0)$ is known and that the required pdf $p(x_{k-1}|y_{1:k-1})$ at time $k-1$ is available, the prior probability $p(x_k|y_{1:k-1})$ can be obtained sequentially through prediction (Chapman-Kolmogorov equation):

$$p(x_k|y_{1:k-1}) = \int p(x_k|x_k-1)p(x_{k-1}|y_{1:k-1})dx_{k-1}$$

Update

Consequently, the prior (or prediction) is updated using the measurement $y_k$ at time $k$, as follows (Bayes Theorem):

$$p(x_k|y_{1:k}) = p(x_k|y_k, y_{1:k-1}) = \frac{p(y_k|x_k)p(x_k|y_{1:k-1})}{p(y_k|y_{1:k-1})}$$
Since the Bayesian solution is hard to compute analytically we have to resort to approximations or suboptimal Bayesian algorithms such as:

**Extended Kalman Filter (EKF)**

The most commonly used Bayesian Technique but its reliability is limited to almost linear systems.

**Unscented Kalman Filter (UKF)**

An improvement of the EKF especially for the case of higher order nonlinearities.

**Sequential Monte Carlo Methods (Particle Filters)**

An extension of the point mass filters using a large number of weighted particles, concentrated in regions of high probability.
The Particle Filter

Theoretical Concepts

Particle Filters approximate the posterior PDF $p(x_k|y_{1:k})$ by a set of support points $x^i_k, i = 1, ..., N$ with associated weights $w^i_k$. This means that the probability density function at time $k$ can be approximated as follows:

$$p(x_k|y_{1:k}) = \sum_{i=1}^{N} w^i_k \delta(x_k - x^i_k)$$

where $\delta(x)$ is the Dirac delta measure. Using the state space assumptions (1st order Markov / observational independence given state), the weights can be estimated recursively:

$$w^i_k \propto w^i_{k-1} \frac{p(y_k|x^i_k)p(x^i_k|x^i_{k-1})}{q(x^i_k|x^i_{k-1}, y_k)}$$

where $p(x^i_k|x^i_{k-1})$ is the transitional density, defined by the process equation and $p(y_k|x_k)$ is the likelihood function defined by the observation equation.
Theoretical Concepts
An important issue in the implementation of Particle Filters is the selection of the importance density. The idea behind importance sampling is that certain values of the input random variables in a simulation have more impact on the parameter being estimated than others. If these important” values are emphasized by sampling more frequently, then the estimator variance can be reduced. Hence, the basic methodology in importance sampling is to choose a distribution which “encourages” the important values.

Since however, sampling from the optimal importance density might not be straightforward, the transitional prior (essentially the process equation) is commonly used as the importance density function, leading to:

$$w_k^i = w_{k-1}^i p(y_k|x_k^i)$$

This means that at time step $k$ the samples $x_k^i$ are drawn from the transitional density or process equation and the selection of the importance weights is essentially dependent on the likelihood of the error between the estimate and the actual measurement as this is defined by the observation equation.
Re-sampling and Sample Impoverishment

Another common problem that is connected to the implementation of Particle Filters is that of **degeneracy**, meaning that after some time steps importance weights are unevenly distributed, thus considerable computational effort is spent on updating particles with “trivial” contribution to the approximation of $p(x_k|y_{1:k})$. Simply put, After only a few time steps, one weight approaches 1 whereas all the other weights approach zero. A measure of this degeneracy is the effective sample size:

$$N_{eff} = \frac{1}{\sum_{i=1}^{N} (w_i^k)^2}$$

**Re-sampling** is a technique aiming at the elimination of degeneracy. It discards those particles with negligible weights and enhances the ones with larger weights (usually duplicates large weight samples).
Re-sampling

Re-sampling takes place when $N_{\text{eff}}$ falls below some user defined threshold $N_T$. Re-sampling is performed by the generation of a new set $\{x^i_k, i = 1, .., N\}$ which occurs by replacement from the original set, so that $Pr(x^i_k = x^j_k) = w^j_k$. The weights are in this way reset to $w^i_k = 1/N$ and therefore become uniform.

The process of Re-sampling: the random variable $u_i$ uniformly distributed in [0,1], maps into index $j$, thus the corresponding particle $x^i_k$ is likely to be selected due to its considerable weight $w^j_k$. 
Re-sampling

Essentially, when the variance of the weights is too high, or in other words the entropy of the weights is below a value specified by the user, particles with small weights are killed and particles with large weights are replicated multiple times.

The underlying idea is to focus the computational efforts on the promising zones of the space. Finally one assigns equal weights $N = 1$ to each copy. The Re-sampling step is what makes SMC functional. Albeit it introduces additional Monte Carlo errors, due to basically sparser sampling, it can be shown that this process ensures that the algorithm does not “degenerate” over time.
Advantages

The particle Filter can deal with nonlinear systems with non Gaussian posterior probability unlike the Kalman Filter of the state, where it is often desirable to propagate the conditional PDF itself.

Like the Unscented kalman Filter (UKF) the computation can be performed in parallel, possibly resulting in a fast process (still slower than the UKF).
Important Note! - Disadvantages

The use of Re-sampling may lead to other problems. As the high weight particles are selected multiple times, diversity amongst particles is not maintained. This phenomenon known as sample impoverishment (or particle depletion), is most likely to occur in the case of small process noise.

In addition, even with a large number of particles, it may happen that there are no particles in the vicinity of the true state. Existing techniques for tackling the sample impoverishment problem include the use of evolutionary operators, the use of Support Vector Regression based re-weighting schemes, and finally the combination of PF techniques with other tools, such as the Kalman Filter, for the time update and importance density generation.
The Particle Filter

The Sigma Point Bayes Filter

The Sigma Point Bayes Filter or Gaussian Mixture Sigma-Point Particle Filter (GMSPPF), combines an importance sampling (IS) based measurement update step with an blue Unscented Point Kalman Filter for the time update and importance density generation. A Gaussian mixture model is then used for the representation of the posterior density. This is obtained from the weighted particle set that came from measurement update step using a weighted Expectation-Maximization algorithm (EM). The GMM is of the following form:

\[ p_G(x) = \sum_{g=1}^{G} \alpha^{(g)} N(x; m^{(g)}, P^{(g)}) \]

where \( G \) is the number of mixing components, \( \alpha^{(g)} \) are the mixing weights and \( N(x; m, P) \) is a normal distribution with mean vector \( m \) and positive definite covariance matrix \( P \).
The Particle Filter

The Sigma Point Bayes Filter

The EM step replaces the standard Re-sampling technique used in the Generic Particle Filter, thus mitigating the sample depletion problem. The Expectation Maximization algorithm recovers a maximum likelihood Gaussian Mixture Model (GMM) fit to the set of weighted samples, leading to both the smoothing of the posterior set (and the avoidance of the sample impoverishment problem) and the use of a reduced number of mixing components in the posterior, leading to a lower computational cost.
The **Generic Particle Filter** described earlier can be summarized as:

1. **Set of weighted particles** $\{\hat{x}_k^i, w_k^i\}$ at time $k-1$.
2. **Draw particles from Importance Density**, $p(x_k|x_{k-1})$:
   \[
   \hat{x}_k^i = F(\hat{x}_{k-1}^i) + \nu_k
   \]
3. **Evaluate importance weights using likelihood function**:
   \[
   w_k^i \propto p(y_k|x_k^i)
   \]
4. **Resample if below $N_{eff}$**.

![Diagram](image)

**Discrete Monte Carlo Representation of** $p(x_{k-1}|y_{1:k-1})$

**Predict**

**Measure**

**Resample**

**Representation of** $p(x_k|y_{1:k})$
Numerical Application

The previously referenced 3dof system is used once again. However, in this case nonlinear response is assumed, resulting through the assumption of a nonlinear 1st dof corresponding to a “soft first storey”.

Hysteretic behavior may be simulated by various models. Here we choose the Bouc-Wen model for its simplicity and compact form. Then, the nonlinear dof $x_1$ is linked to a hysteretic term $r_1(t)$ (the Bouc-Wen hysteretic component) with:

$$\dot{r}_1(t) = \ddot{x}_1 - \beta |\ddot{x}_1| r_1^{n-1} r - \gamma (\dot{x}_1) |r_1|^n$$
Additionally, the restoring force will be equal to:

\[ R_1 = \alpha k_1 x_1 + (1 - \alpha) k_1 r_1 \] strongly nonlinear term

where \( \alpha = k_e l / k_p l \) is the post-yield to elastic stiffness ratio. Here it is assumed effectively 0.

The system’s state space equation can be then formulated as follows:

\[
\begin{bmatrix}
\dot{z}_1 \\
\dot{z}_2 \\
\dot{z}_3 \\
\dot{z}_4 \\
\dot{z}_5 \\
\dot{z}_6 \\
\dot{z}_7 \\
\dot{z}_8 \\
\dot{z}_9 \\
\dot{z}_{10} \\
\dot{z}_{11} \\
\dot{z}_{12} \\
\dot{z}_{13} \\
\dot{z}_{14} \\
\dot{z}_{15} \\
\dot{z}_{16}
\end{bmatrix} =
\begin{bmatrix}
\dot{x}_1 \\
\dot{x}_2 \\
\dot{x}_3 \\
\dot{r}_1 \\
\dot{x}_2 \\
\dot{x}_3 \\
\dot{r}_1 \\
\dot{k}_2 \\
\dot{r}_1 \\
\dot{c}_1 \\
\dot{c}_2 \\
\dot{c}_3 \\
\dot{\beta} \\
\dot{\gamma} \\
\dot{h}
\end{bmatrix} =
\begin{bmatrix}
z_5 \\
z_6 \\
z_7 \\
(z_5 - z_{14} | z_5 | z_4 | z_{16} - 1 | z_4 + z_{15} z_5 | z_4 | z_{16}) \\
(-z_8 \cdot z_4 - z_9 \cdot z_1 + z_9 \cdot z_2 + \text{(z11 + z12) \cdot z_5 + z_12 \cdot z_6})/m_1 \\
-w_1 + \frac{F_1(t)}{m_1} \\
\dot{x}_{m2} + w_2 \\
\dot{x}_{m3} + w_3
\end{bmatrix}
\]

Note how the system has been augmented to account for the constant parameters which are to be identified (joint state-parameter estimation).
The previous process equation can be brought into the discrete form by applying a simple integration rule (even forward Euler).

Additionally we can formulate the observation equation, considering that measurements of $x_1$, $\ddot{x}_2$ and $\ddot{x}_3$ are available:

$$y = \begin{bmatrix} x_{m1} \\ \dot{x}_{m2} \\ \ddot{x}_{m3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{k_2}{m_2} & \frac{-k_2+k_3}{m_2} & \frac{k_3}{m_2} & \frac{-c_2}{m_2} & \frac{-c_2+c_3}{m_2} & \frac{c_3}{m_2} \\ 0 & \frac{k_3}{m_3} & \frac{-k_3}{m_3} & \frac{c_3}{m_3} & \frac{-c_3}{m_3} & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix}$$

$$+ \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{F_2(t)}{m_2} \\ \frac{F_3(t)}{m_3} \end{bmatrix}$$

Finally, the **discrete system** can be compactly written in matrix form as:

$$z_{k+1} = F(z_k) + \ddot{x}_m + w_k + F^p_k$$

$$y_k = H(z_k) + v_k + F^m_k$$
Tested Methodologies

The **Unscented Kalman Filtered (UKF)** - demonstrated by Prof. Smyth, The **Generic Particle Filter (PF)** - described here -
and a mixture of the two,
The **Sigma Point Bayes Filter (GMSPPF)**
were used in order to identify the parameters describing the system and the unmeasured states.

One of the Northridge (1994) earthquake acceleration records was used as ground excitation. The “clean” displacement, velocity and acceleration response was obtained using 4th order Runge Kutta Integration. The assumed measurements are corrupted with white noise 10% RMS noise to signal ratio.
A process noise of 1% RMS noise-to-signal ratio was added only in the case of the state variables $z_5, z_6, z_7$ where data from acceleration measurements are utilized.
Simulation Results

State Estimation

![Graph of State Estimation (x₁)](image1)

- Observation
- UKF estimate
- PF estimate
- GMSPPF estimate

---

![Graph of State Estimation (x₂)](image2)

- Observation
- UKF estimate
- PF estimate
- GMSPPF estimate

---

![Graph of State Estimation (x₃)](image3)

- Observation
- UKF estimate
- PF estimate
- GMSPPF estimate

---

![Graph of State Estimation (r₁)](image4)

- Observation
- UKF estimate
- PF estimate
- GMSPPF estimate
Simulation Results

Parameter Estimation

Parameter Estimation ($c_3$)

Parameter Estimation ($\beta$)

Parameter Estimation ($\gamma$)

Parameter Estimation ($n$)

Identification Methods for SHM
Parameter Validation

**Validation Plot - Hysteretic Loops**

The finally estimated values are plugged into the model of the system (noise-free) in order to reproduce the nonlinear hysteretic response.
Comments

• The UKF method and the GMSPPF method provide us with very satisfactory results. The Generic PF method, yields inferior performance related to the sample depletion problem.

• The UKF method is considerably faster (runtimes: UKF 3.203 sec, PF 3200 sec, GMSPPF 770 sec)

• Parametric runs have shown that - the addition of process noise for the constant parameters can mildly improve the estimates
  - the Generic PF method is heavily dependent upon the choice of initial value interval
The Sample Depletion problem

Representation of the Initial and Final Sample Space of the stiffness-damping parameter pairs for the generic PF (square points) and the GMSPPF. The PF is represented by a single point in the final sample space (all points have collapsed to a single one!) and is evidently quite far from the actual value (circle point). This does not hold for the GMSPPF which manages to maintain diversity.
The Particle Filter with Mutation (MPF)

Improving the Standard PF

In order to tackle the sample impoverishment problem, especially for the case of state vectors with time invariant components, a modified PF approach incorporating a mutation operator has been introduced by Chatzi & Smyth (2012), termed the Particle Filter with Mutation (MPF).

It features a two component innovation:

**Replicate the previous estimate value**

Part of the formerly unfit particles is replaced by some uniform probability, $p_e$, by a prior optimal estimate of the state obtained as $F(x^{e}_{k-1})$. In this way the actual value of the estimate is incorporated in the particles and propagated through the non-linear system.

**Mutate**

The time invariant components are mutated. Mutation takes place by replacing the parameter components, by a mutation probability $p_m$, with a random number uniformly distributed in some interval that extends around the former value. The latter resembles the creep mutation operator of the traditional Genetic Algorithm.
The Particle Filter with Mutation (MPF)

**Improving the Standard PF**

The mutated particles are assigned a weight that is lower than the original uniform weight. The new weight is inversely proportional to the relative difference between the mutated vector and the original one ("parent" vector) according to the following relationship:

\[ w_k^i = \frac{1}{N} \left( \frac{1}{\| \Delta x_k^i \| + 1} \right) \]

whereas the weights of the non-mutated re-sampled particles remain equal to \( w_k^i = \frac{1}{N} \).

\( \Delta x_k^i \) is the difference between the mutated vector and the parent one and \( \| \cdot \| \) is the \( L_2 \) norm.

The full set of weights is then normalized before proceeding to the next time step by dividing by the total sum.
The Particle Filter with Mutation (MPF)

Improving the Standard PF - the MPF

Evaluate importance weights using likelihood function:

\[ w^i_k \propto p(y^i_k | x^i_k) \]

Resample if \(N_{eff} < N_T\)

Replace by the fit particles or by the prior estimate \(\hat{x}_k\) with probability \(p_e\)

\[ w^i_k = \frac{1}{N} \]

Mutate the time invariant component of the previously unfit particles, using a mutation probability \(p_m\).

Mutated weights:

\[ w^i_k = \frac{1}{N} \frac{1}{\left| \frac{\Delta x^i_k}{x^i_k} \right| + 1} \]
The MPF - Numerical Application

The same 3dof system as before is utilized for testing the performance of the suggested methodology.
The MPF - Numerical Application

Parameter Estimation ($k_j$)

Parameter Estimation ($c_j$)

Parameter Estimation ($\beta$)

Parameter Estimation ($\gamma$)

Parameter Estimation ($n$)