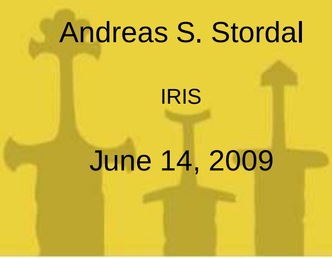
Gaussian Mixture filters for high-dimensional dynamic systems



Andreas S. Stordal

IRIS

June 14, 2009

Nonlinear filtering

- High-dimensional nonlinear dynamic system with Gaussian observation noise

$$\begin{aligned} X_t &= M_t(X_{t-1}) + \eta_t, \quad & \eta_t &\sim p_t(\eta) \\ Y_t &= H_t X_t + \epsilon_t, & \epsilon_t &\sim \Phi(0, R_t) \\ X_0 &\sim P_0 \end{aligned}$$

- Estimate the posterior distribution sequentially in time

$$p(x_t | y_{1:t})$$

- The EnKF does not produce the correct posterior in general

Sequential Importance Resampling filter

- Sample $\{x_t^i\}_{i=1}^N \sim p(x_t|x_{t-1}^i)$
- Assign weights to the sample $w_t^i \propto w_{t-1}^i p(y_t|x_t^i)$
- If the estimated effective ensemble size is too small we resample from the set of particles with probabilities equal to the weights

$$\hat{N}_{eff} = \frac{1}{\sum_{i=1}^N (w_t^i)^2}$$

- If both N and d (system dimension) tend to infinity, and $\frac{\log N}{d} \to 0$, we have $\max_i(w_t^i) \xrightarrow{p} 1$ (Bengtsson et al. 08)

Improved SIR

- Assume that η_t is Gaussian, $\eta_t \sim N(0, Q_t)$
- In this system it is possible to find the optimal importance function (the one that minimizes the variance of the weights conditioned on y_t and x_{t-1})
- Sample $\{x_t^i\}_{i=1}^N \sim p(x_t|x_{t-1}^i, y_t)$
- Assign weights to the sample $w_t^i \propto w_{t-1}^i p(y_t|x_{t-1}^i)$
- $p(x_t|x_{t-1}, y_t)$ is Gaussian with mean and covariance

$$\mu_t = M_t(x_{t-1}) + K_t(y_t - H_t M_t(x_{t-1}))$$

$$P_t = (I - K_t H_t) Q_t$$

$$K_t = Q_t H_t^T (H_t Q_t H_t^T + R)^{-1}$$

$$p(y_t|x_{t-1}) = \Phi(M_t(x_{t-1}), H_t Q_t H_t^T + R)$$

- The weights still suffer from degeneracy
- If model error is unknown, adding Gaussian may cause unphysical behavior
- Without model error, no Kalman update
- If we work with densities instead of particles, we may approximate the model error without adding noise

Gaussian Mixture filter

- At each time t we approximate the prior density with a Gaussian mixture

$$\hat{p}(x_t|y_{1:t-1}) = \sum_{i=1}^{N} w_{t-1}^{i} \Phi\left(x_t - M_t(x_{t-1}^{i}), P_{t|t-1}\right)$$

- When y_t arrives, we may update this according to Bayes theorem

$$\hat{p}\left(x_t|y_{1:t}\right) \sum_{i=1}^{N} w_{t-1}^{i} \Phi\left(x_t - M_t(x_{t-1}^{i}), P_{t|t-1}\right) \Phi\left(y_t - H_t x_t, R_t\right),$$

- which may be written as

$$
\begin{aligned}
&\sum_{i=1}^{N} w_{t-1}^{i} \Phi\left(x_t - x_t^i, P_t\right) \Phi\left(y_t - H_t M_t(x_{t-1}^i), \Sigma_t\right),\\
&x_t^i = M_t(x_{t-1}^i) + G_t(y_t - H_t M_t(x_{t-1}^i)),\\
&G_t = P_{t|t-1} H_t^T \Sigma_t^{-1},\\
&\Sigma_t = H_t P_{t|t-1} H_t^T + R_t\\
&P_t = (I - G_t H_t) P_{t|t-1}
\end{aligned}
$$

- $P_0 = h^2 S_0$
- h determines the size of the Kalman update.
- Increasing h leads to more uniform weights.
- We want to keep h as small as possible to capture more of the non-Gaussian aspects of the posterior distribution.

Bernoulli model

- Estimate the posterior distributin $p(y_{10}|d_{1:10})$ using EnKF, PF and GMF to the Bernoulli equation

$$\frac{dy}{dt} - y = -y^3, \tag{1}$$

- Analytical solution

$$y(n) = g(y_0) = y_0 \times \left(y_0^2 + \left(1 - y_0^2\right) e^{-2n}\right)^{-\frac{1}{2}}, \tag{2}$$

$$y_0 \sim N(-0.1, 0.2^2) \tag{3}$$

- The measurements are given by

$$d_k = y(k\delta) + \epsilon, \tag{4}$$

  where ϵ is a zero mean Gaussian noise with standard deviation 0.8 , $k = 1, \ldots, 10$ and δ is the frequency of observations chosen as 0.3
- 1000 ensemble members

Posterior density estimates

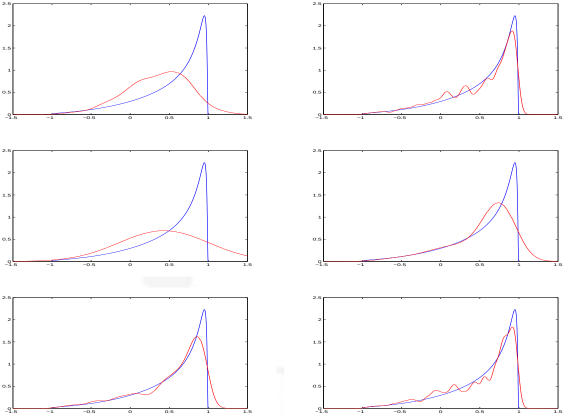


Figure: True posterior (blue) versus EnKF (top left), PF (top right) and GMF for h=1, 0.2, 0.1 and 0.05

Adaptive Gaussian Mixture Filter

- To avoid collapse of the weights we introduce a linear combination of the GM filter weights and the uniform weights from the EnKF.
- New weights

$$w_i(\alpha) = \alpha w_i + (1 - \alpha)N^{-1}$$

- If only a few of the weights are zero, α should be close to one.
- If only a few weights are nonzero, α should be close to zero.

- α is chosen to minimize:

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\alpha w_i - w_j\right)^2 = \sum_{\substack{j=1\\ j\neq i}}^{N}\left(\alpha w_i - w_j\right)^2 + \sum_{i=1}^{N}\left(\alpha w_i - w_i\right)^2$$

- with solution

$$\alpha_{\text{min}} = \frac{1}{N}\frac{1}{\sum_{i=1}^{N} w_i^2} = \frac{\hat{N}_{\mathit{eff}}}{N}$$

Lorenz40 model

- 40 coupled differential equations with cyclic boundary conditions

$$\dot{x}_i = (x_{i+1} - x_{i-1})\, x_{i-1} - y_i + 8, \quad i = 1, \ldots, 40; \qquad (5)$$

$$x_0 = x_{40}, \quad x_{-1} = x_{39}, \quad x_{41} = x_1 \qquad (6)$$

- The complete dynamical system is given by

$$X_{t_i} = M\left(X_{t_{i-1}}\right) + \eta_{t_i},$$
$$Y_{t_i} = X_{t_i} + \epsilon_{t_i}, \quad t_i = 0.05i, \quad i = 1, 2, \ldots 10000$$

where M is the solution of (5), $\eta_{t_i} \sim N\left(0, 0.01\mathbf{I}_{40}\right)$ and $\epsilon_{t_i} \sim N\left(0, \mathbf{I}_{40}\right)$.

- Run GMF for different values of h and α
- For each value of h we also run with adaptive α
- Compare the results with EnKF
- Repeat each run ten times to avoid too much Monte Carlo effect
- Convergence criteria:RMSE=$\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\bar{x}_t - x_{true}\right)^2} < 1$

RMSE table GM

α h	0.4	0.5	0.8
0.2	4.2788	0.2995	0.3134
0.5	4.7648	0.6224	0.3183
0.8	4.9382	4.6938	1.3753

Table: RMSE GM

RMSE table EnKF and adaptive GM

h	0.4	0.6	0.8	EnKF
RMSE	3.6669	0.2895	0.3126	0.6439

Table: RMSE Adaptive GM, EnKF

- Fill out table with *h* from 0.4 to 1 and α from 0 to 1
- Run a particle filter with huge amount of particles in order to find the true posterior distribution