Reduce Rank and Ensemble Kalman Filter: Analyse and its Application

E. APRILIANI, B.A. SANJOYO, D.K. ARIF, D. ADZKIYAH

Department of Mathematic, Sepuluh Nopember Institute of Technology

Abstract — In this paper we present the reduced rank ensemble Kalman filter (the RREnKF) algorithm and its application. Kalman filter is an estimation method of the state of stochastic dynamic system. The ensemble Kalman filter (EnKF) is modification of Kalman filter algorithm to estimate the non linear stochastic dynamic system. The Square root EnKF is an algorithm to keep the numeric stability, but need more computational time. Here we analyze the reduced rank EnKF (RR-EnKF) as an alternative to reduced the computational time, and keep the accuracy estimation. We applied the RR-EnKF algorithm in some cases.

Keywords — EnKF, reduced rank, computational time, accuracy

1. Introduction

Suppose we have a dynamic stochastic system

\[ x_{k+1} = f(x_k, u_k, k) + w_k \]  

and the measurement equation

\[ z_k = Hx_k + v_k \]

Where \( x_k, u_k, z_k \) is a state variable, input variable, and output or measurement data, respectively. \( w_k, v_k \) is a noise system and noise measurement.

To estimate the state variable of Eq. (1) based on measurement data Eq. (2), we use Kalman filter algorithm for Eq. (1) is linear[Lewis], but for nonlinear system. We use the modification of Kalman filter such as extended Kalman filter (EKF)[3], unscented Kalman filter (UKF) or ensemble Kalman filter (EnKF)[1].

There are some modifications of the EnKF, such as the square root ensemble Kalman filter [1] and the variance reduced ensemble Kalman filter [2]. All of these modifications are done to avoid the divergence of computation and to decrease the computational time.

We had been applied the EnKF and Reduced rank EnKF (RR-EnKF) in some cases. We analyze the accuracy and computational time.

2. The EnKF and RR-EnKF

To estimate the state variable of system Eq. (1) with measurement equation (2), we can use the ensemble Kalman filter algorithm as follow [1]:

a. Initial Estimation

Generate the N-ensembles of initial estimation

\[ x_0 = \begin{bmatrix} x_{0,1} & x_{0,2} & \cdots & x_{0,N-1} & x_{0,N} \end{bmatrix} \]

with \( x_{0,j} \sim N(\tilde{x}_0, P_0) \).

b. The prediction step

\[ \hat{x}_{k,j} = f(\hat{x}_{k-1,j}, u_k) + w_{k,j} \]  

where \( w_{k,j} \sim N(0, Q_k) \) is the ensemble of noise system

Mean of prediction step estimation

\[ \hat{x}_k = \frac{1}{N} \sum_{j=1}^{N} \hat{x}^{-}_{k,j} \]

Error covariance of prediction step estimation

\[ P_k = \frac{1}{N-1} \sum (\hat{x}_{k,j} - \hat{x}_k)(\hat{x}_{k,j} - \hat{x}_k)^T \]

c. The Correction step

Generate the ensemble of measurement data,\n
\[ z_{k,j} = z_k + v_{k,j} \]

Where \( v_{k,j} \sim N(0, R_k) \) is an ensemble of measurement noise

Kalman gain is defined as

\[ K_k = P_k H^T (HP_k H^T + R_k)^{-1} \]

Estimation of correction step is

\[ \hat{x}_{k,j} = \hat{x}_{k,j} + K_k (z_{k,j} - H\hat{x}_{k,j}) \]

(5)

Mean of correction step estimation:
\[ \hat{x}_k = \frac{1}{N} \sum_{j=1}^{N} \hat{x}_{k,j} \]

With error covariance
\[ P_k = [I - K_k H] P_k^- \]
d. Substitute Eq. (5) in prediction step Eq. (3)
c. Continuing until we get correction step estimation Eq. (3).

The reduced rank of matrix is the reducing rank of diagonal matrix D, where \( A = UDV^T \), \( U,V \) are singular vector of A and D is a diagonal matrix of singular value A. The reducing rank will be applied in matrix D, \( D^* = [D]_{n,q} \), is q first columns of D.

Here we applied the rank reducing in correction step.

The algorithm of RR-EnKF has same algorithm with EnKF for initial step and prediction step. In the correction step, the RR-EnKF algorithm is below:

Generated the N ensemble measurement data such as eq. (3)
Define \( S_k = H\hat{x}_{k,j} \), \( E_k = v_{k,j} \) and
\[ C_k = S_k S_k^T + E_k E_k^T \]

The ensemble correction estimation
\[ \bar{x}_{k,j} = \hat{x}_{k,j} + \bar{x}_{k,j} S_k^{-1} (z_{k,j} - H\hat{x}_{k,j}) \]

Before we reduced the rank, we decomposed the matrix \( C_k \) by using the singular value decomposition:

\[ [U,D,V] = svd(C_k) \]
Reduced the rank is obtained with using q first column of square root inverse of matrix D,
\[ D^* = [D^{-1/2}]_{1m,1q} \]
Define matrix \( M_k^* = D^* U^T S_k \), decompose matrix \( M_k^*, [U_1,D_1,V_1] = svd(M_k^*) \)
The error of ensemble estimations are
\[ \tilde{x}_{k,j} = \hat{x}_{k,j} V_1 (I - D_1^T D_1)^{1/2} \]
The ensemble of estimations are
\[ \hat{x}_{k,j} = \tilde{x}_{k,j} + \bar{x}_{k,j} \]

Repeat the prediction step and correction step such that we get the time step which desired.

3. The Application of RR-EnKF

In our research, we applied the EnKF and RR-EnKF in some cases, such as the groundwater pollution estimation, the heat transfer in one dimensional rod, in the stirred tank reactor, in the diffusion problem. The diffusion problem can be written as mathematical model below:

\[ \frac{\partial C}{\partial t} + u \frac{\partial C}{\partial x} + v \frac{\partial C}{\partial y} = \mu \frac{\partial^2 C}{\partial x^2} + \vartheta \frac{\partial^2 C}{\partial y^2} + S \]

Where, \( C \) is a concentration of air pollution, \( u,v \) are the velocities in x and y directions, \( \mu, \vartheta \) are the dispersion coefficient in x and y directions and S is a source of pollution.

The stirred tank reactor has mathematical model as follows

\[ \frac{dC_A}{dt} = \frac{F}{A} (C_{Ain} - C_A) - 2k_0 e^{\frac{E}{RT}} C_A^2 \]
\[ \frac{dT}{dt} = \frac{F}{V} (T_{in} - T) + 2 \frac{\Delta H}{\rho C_p} R k_0 e^{\frac{E}{RT}} C_A^2 - \frac{l}{V} \]
\[ \frac{dT_j}{dt} = \frac{F_w}{V_w} (T_{jin} - T_j) + \frac{UA}{V_w \rho_w C_{pw}} (T - T_j) \]

Where, \( C_A, T, T_j \) is reactant concentration, temperature of tank and temperature of cooling jacket, respectively.

The one dimensional heat conduction distribution is

\[ \frac{\partial U}{\partial t} = c \frac{\partial^2 U}{\partial x^2} \]

Where \( u(x,t) \) is temperature in time t and position x.

The groundwater pollution distribution is
\[ D_x \frac{\partial^2 C}{\partial x^2} + D_y \frac{\partial^2 C}{\partial y^2} - v_x \frac{\partial C}{\partial x} - v_y \frac{\partial C}{\partial y} = \frac{\partial C}{\partial t} \]

Where \( C \) is concentration of groundwater pollution, \( D_x, D_y, v_x, v_y \) are coefficient of diffusion and groundwater flow velocities in the \( x, y \) direction, respectively.

All of those problems have partial differential form. Before we applied the EnKF or RR-EnKF to estimate those state variables, we discretize those equations respect to time \( t \) and position \( x, y \) such that we have Eq. (1). Beside that, we must define the measurement equation which is connected the measurement data and the state variables such as Eq. (2).

We do some simulation to observe the accuracy and the computational time of EnKF and RR-EnKF. We also do some number of ensemble such as 100, 200 and 500.

Table 1. shows the feasibilities of algorithm to the problems above.

<table>
<thead>
<tr>
<th>Problem</th>
<th>The EnKF</th>
<th>The RR-EnKF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diffusion problem</td>
<td>Feasible</td>
<td>Feasible</td>
</tr>
<tr>
<td>Stirred tank reactor</td>
<td>Feasible</td>
<td>Not feasible</td>
</tr>
<tr>
<td>Heat conduction</td>
<td>Feasible</td>
<td>Feasible</td>
</tr>
<tr>
<td>Groundwater pollution</td>
<td>Feasible</td>
<td>Feasible</td>
</tr>
</tbody>
</table>

From table 1 the RR-EnKF can’t be applied to estimate the state variables of the stirred tank reactor, because the stirred tank reactor has small dimension (three dimension), but the RR-EnKF can be applied for others problem because those problem are the large scale (dimension) problems.

The EnKF and the RR-EnKF had been applied in the diffusion problem. The error of estimation of EnKF is 0.01571 and RR-EnKF is 0.01487. For the stirred tank reactor only the EnKF can be apply with error of estimation 0.023. For the heat conduction distribution and the groundwater pollution, the estimation of state variables by the RR-EnKF is doing simulation.

Figure 1 shows the correlation between the numbers of ensemble with the estimation error of the heat conduction distribution problem. Figure 2 shows the correlation between the numbers of ensemble with the computational time.

Figure 1. Relationship between size of ensemble and error estimates

Figure 2. Relationship between size of ensemble and computational time

Figure 3. The concentration of groundwater pollution estimation by EnKF
Figure 4. The estimation of the stirred tank reactor by EnKF.

Figure 3 and Figure 4 are the estimation of the groundwater pollution and the stirred tank reactor by EnKF. In this research, the real system is taken from Matlab simulation. And the EnKF can be applied in those problems.

Table 2 The error estimation of stirred tank reactor

<table>
<thead>
<tr>
<th>Number of Ensemble</th>
<th>variable</th>
<th>( H=[1 \quad 0 \quad 0 \quad 1 \quad 0] )</th>
<th>( H=[1 \quad 0 \quad 0 \quad 0 \quad 1] )</th>
<th>( H=[0 \quad 1 \quad 0 \quad 0 \quad 0] )</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>C_A</td>
<td>0.0662</td>
<td>0.0556</td>
<td>0.3005</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>0.0001</td>
<td>0.0006</td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td>T_j</td>
<td>0.0023</td>
<td>0.0011</td>
<td>0.0019</td>
</tr>
<tr>
<td>100</td>
<td>C_A</td>
<td>0.0414</td>
<td>0.00435</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>0.0001</td>
<td>0.0008</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>T_j</td>
<td>0.0008</td>
<td>0.0021</td>
<td>0.001</td>
</tr>
</tbody>
</table>

From our research, we know that the size (number) of ensemble is not influence the estimation error. But the size of ensemble is influence the computational time.

4. Concluding Remark

From our research we conclude that
a. The EnKF can be applied to estimate the state variables of the diffusion, heat conduction, stirred tank reactor, and groundwater pollution problems.
b. The RR-EnKF can be applied to estimate the state variables of the diffusion, heat conduction, and groundwater pollution problems, because these problems have large dimension.
c. The number of ensemble isn’t influence the accuracy of estimation, but it is influence the computational time.

d. For small dimension system, the RR-EnKF can’t be applied to estimate these state variables of this system.

5. Acknowledgment

This research is funded by research grant, “Hibah Bersaing 2009”. I would like say thanks to LPPM ITS, DIKTI and also to our students in Mathematic Department ITS which did this topic as their final project and their thesis.

Bibliography

[1] G. Evensen, Sampling Strategies and square root analysis schemes for the Ensemble Kalman Filter (EnKF), Hydro Research Centre, 2004