Sequential Data Assimilation in Geotechnical Engineering and Its Application to Seepage Analysis

Kazuyuki Nakamura
Graduate School of Advanced Mathematical Sciences
Meiji University
Kanagawa, Japan.
Email: knaka@isc.meiji.ac.jp

Shinya Yamamoto
Shimizu Institute of Technology
Shimizu Corporation
Tokyo, Japan
Email: s-yama@shimz.co.jp

Makoto Honda
Shimizu Institute of Technology
Shimizu Corporation
Tokyo, Japan
Email: mac.honda@shimz.co.jp

Abstract—There are many uncertainties in the geotechnical engineering design, which lead to prediction errors and risks in the real world. Sequential data assimilation is one of the most applicable methods to lessen these errors and risks, for partial observations can be obtained online and required computation time in numerical analysis has similar time scale of real geotechnical phenomena. We summarize characteristics of sequential data assimilation from the viewpoint of its application to geotechnical engineering and apply them to a seepage analysis using realistic simulation system. The result shows that application of sequential data assimilation with the particle filter gives a good estimation in seepage analysis.

Keywords: Data assimilation, particle filter, seepage analysis, estimation.

I. INTRODUCTION

There are many uncertainties in the geotechnical engineering design because objectives are too large to obtain detailed measurements such as land subsidence, soil permeability and other soil mechanical parameters. There are also difficulties in that mechanics of sampled soil usually differ from that in which the environment samples originally exist because environments around samples such as stress fields and water pore pressures are different.

To overcome this difficulty in the geotechnical engineering design, numerical analyses are widely used. Many parameters such as compression index, initial pore pressure and initial stress parameters are combined into numerical simulation models. Those have large uncertainties which lead to gap between the result of numerical analysis and real phenomena. One of the solution is inversion analysis that is applied in geotechnical engineering. For example, the extended Kalman filtering (EKF) [1] is applied to estimate uncertain parameters of soil mechanics in that field [2], [3]. However, linear approximation of extended Kalman filter causes estimation error [1]. Also, physical variables should have spatial continuity, which makes this approximation inappropriate. Particularly in geotechnical engineering, there is also a difficulty in conserving spatial continuity of physical variables. For example, the advection of pore water should satisfy conservation laws. However, linear approximation destroys the laws. In addition, some governing equations lack of Markov property because time integration term is included in the governing differential equations.

In the geotechnical engineering, time scale of the system ranges from several hours to hundreds of days whereas a single simulation of the system needs several hours. Because batch-type DA such as 4DV AR [4] requires many time simulations to optimize physical variables after obtaining whole observations, it is difficult to apply this approach. On the other hand, some kind of sequential data assimilation techniques can reduce that repetition, and so it is effective for the geotechnical engineering design. In this paper, we discuss which algorithm is suitable from the viewpoint of applications of sequential data assimilation to geotechnical engineering. Based on this discussion, we apply the sequential importance resampling (SIR) [5], [6] to a seepage analysis. The result shows effectiveness of the SIR for seepage analysis.

II. SEQUENTIAL DATA ASSIMILATION IN GEOFANTICAL ENGINEERING

A. Sequential data assimilation and filtering algorithms

Data assimilation (DA) [7] is the concept that combines computational simulation models with observation data. Data assimilation is developed in the geophysics, especially in the atmosphere and marine sciences. It aims at providing initial conditions of a simulation model for forecast, spatio-temporal interpolation of unobservable physical variables for fact findings, and estimation of unknown parameters to make numerical model more precise.

Sequential data assimilation is online type of data assimilation, that is, physical variables in the numerical models are modified at each time step of observation (See Figure 1). Sequential data assimilation can be formulated as nonlinear filtering and smoothing of nonlinear non-Gaussian state space
where $x_t$ is the state vector which consists of all the simulation variables at time $t$, $y_t$ is the observation vector and $v_t$ and $w_t$ represent system and observation noises with covariance matrices $Q_t$ and $R_t$, respectively. The nonlinear function of the system model $f_t$ is modeled by discretized system equations and uncertainties in the system. The observation function $h_t$ is modeled by the relation between observations and physical variables that is known a priori.

To estimate state vector of nonlinear non-Gaussian state model (SSM)

\[
\begin{align*}
x_t &= f_t(x_{t-1}, v_t), \\ y_t &= h_t(x_t, w_t), \\ v_t &\sim N(0, Q_t), \\ w_t &\sim N(0, R_t)
\end{align*}
\]

(1) (2) (3) (4)

Table I shows pros and cons of filtering algorithm for geotechnical sequential data assimilation; continuity of physical variables, fitness for nonlinearity and ensemble efficiency in estimation. For comparison, 4DVAR which is not sequential data assimilation but batch-type data assimilation is also shown.

### B. Sequential data assimilation in geotechnical engineering

In the geotechnical engineering, there are many uncertainties in all scales. In the microscopic scale, parameters of soil mechanics have uncertainties. To study and lessen these uncertainties, we compare the result of laboratory experiments with physical model and/or parameters. For this purpose, we do not need online estimation because small samples do not have spatial dependency and environmental effects. For example, Markov Chain Monte Carlo sampling is one of the best method to obtain Bayes estimates of mechanical parameters of a material. Therefore, sequential data assimilation is not necessarily good for this scale of geotechnical problem.

As for larger scale, subsidence analysis is one of the most important problem in geotechnical engineering. For this problem sequential data assimilation can give useful information. The reason is that many physical variables are unobservable and have large uncertainties because they are under the ground. In addition, as time scale of the subsidence ranges from several
hours to hundreds of days, we can use online estimation results of sequential data assimilation for decision making such as change of construction method. From that point, online property of sequential data assimilation is important. For the subsidence problem, we can employ elastic deformation model. In this case, inversion analysis is easy because we only need current strain variables of each grids. It means that the simulation model has simple Markov property. In addition, it is linear or nearly linear equations, we can apply simple regression or linear Kalman filter. In the real subsidence problem, however, the ground has elasto-plastic property. Realistic governing equations of elasto-plastic ground, integration part exists. That makes the problem hard in that the phenomena is highly nonlinear and non-Markov. Because of this property, the SIS with small or no system noise is only applicable, which preserves physical validity in sequential data assimilation.

Seepage analysis is another important problem in geotechnical engineering. Internal water flow causes piping in embankment erosion which leads to embankment failure. To prevent this disaster, online forecast of seepage analysis can be used. It is the reason why the sequential data assimilation is required. In seepage analysis, governing equation has weak nonlinearity (we will show this property later). It also has spatial extent and locality around “singular point” which is originated from boundary conditions. As a result, estimated results can have several likely modes. These modes have important information for risks and should be evaluated through posterior probability. To obtain estimates of these analyses, we need some online Bayes type estimation which can deal with multi-modality. Because the nonlinearity is weak in seepage analysis, we can use the EnKF and the SIR. The EnKF can give effective estimation in small samples, whereas evaluation of multi-modality is difficult. The SIR requires more samples than the EnKF, but the multi-modality and risks can be analyzed. The SIS needs much more samples than the SIR and the nonlinearity and continuity can be managed by the system noise level. As a result, the SIS is unsuitable from the viewpoint of estimation accuracy. Therefore, the SIR is the best choice for the seepage analysis.

The discussion of this subsection is summarized as follows; we focused on the important applications of sequential data assimilation for geotechnical engineering, subsidence analysis and seepage analysis. In view of pros and cons of estimation algorithm, the SIS is suitable for subsidence analysis and the SIR or the EnKF is applicable for seepage analysis. Especially, the SIR is suitable for seepage analysis.

III. APPLICATION TO SEEPAGE ANALYSIS

A. Twin experiment

Based on the discussions in the previous sections, the SIR algorithm would be suitable for the seepage analysis. Therefore, we applied sequential data assimilation to seepage analysis of a fill dam.

To test the validity of the application, we employed a twin experiment (TE). TE is the procedure that evaluates validity of introduced DA algorithms. The procedure of TE is as follows (See Figure 2). TE consists of two phases, the simulation phase (TE-1,2,3) and the assimilation and evaluation phase (TE-4,5). At first, “true” conditions such as initial conditions and parameters are set (TE-1). In next, we run the simulation model and obtain true results (TE-2). From the true results and the observation model, an observation data set is generated (TE-3). Simulation phase ends at this point. In next, we assimilate a biased (erroneous) simulation model and observations (TE-4) and check whether the states and/or the parameters in the simulation model are appropriately corrected by comparing with the true states and parameters (TE-5). If the states are modified appropriately, the introduced method is valid.

B. Settings of numerical experiment

We tested twin experiment of two dimensional seepage analysis of a fill dam case. In this experiment, time dependent flow vectors and time invariant parameters (permeability coefficient) are estimated.

Governing equation of the seepage analysis is as follows:

$$\frac{\partial}{\partial t} \left( K \frac{\partial h}{\partial x} \right) = (C + \beta S_s) \frac{\partial h}{\partial t}$$

(5)

where $h$ is hydraulic head (water energy), $S_s$ is coefficient of specific storage, $\beta$ is characteristic function of specific storage whether soil is saturated ($\beta = 1$) or unsaturated ($\beta = 0$), $K$ is permeability coefficient, and $\xi$ is spatial coordinates. Among these variables, hydraulic $h$ is a time dependent parameter and the others are time invariant parameters. This equation is discretized by Euler scheme in time

$$\frac{\partial h}{\partial t} \approx \frac{h_t - h_{t-1}}{\Delta t}$$

(6)

which leads to the following update equation form:

$$h_t = \left\{ 1 + \frac{\Delta t}{C + \beta S_s} \frac{\partial}{\partial x} \left( K \frac{\partial}{\partial x} \right) \right\} h_{t-1}.$$  

(7)

For space discretization, finite element method (FEM) is applied. In this twin experiment, two dimensional model is used and the analysis grid is shown in Figure 3. The number of nodes is 1313. Initial conditions of water level and true permeability coefficient parameters are also shown in Figure 3. In the TE-1 step, we gave these conditions and run the simulation. Observations are sum of outflows at three regions of outside core zone, shown in Figure 3. In the simulation, water level ranges from 40(m) to 85(m).

In this analysis, we estimate the flow vector of each grid and the permeability coefficient of the core zone, curtain grouting zone and foundation ground. To estimate time invariant parameters online by the SIR, we need to modify model or algorithm. We employed self organizing state space model [10] for this purpose. Time invariant parameters are transformed into time dependent state variables which follow random walk. For the estimation of permeability coefficient, we formulate a new model

$$K_{t,j} = K_{t-1,j} + v_{t,j}$$

(8)

where $j(=1,2,3)$ represents the region index, $K_{t,j}$ is “time dependent” permeability coefficient of region $j$ at time $t$ and
“true” condition
(“true” initial states
“true” parameters)

set
conditions
(TE-1)

Simulation

obtain
results
(TE-2)

“true” results
(“true” states
“true” parameters)

“biased” condition
(“biased” initial states
errornious parameters)

set
conditions
(TE-4)

Data assimilation
(state space model)

obtain
results
(TE-4)

assimilated results
(assimilated states
modified parameters)

obtain
observations
(TE-3)

Figure 2. Twin experiment procedure

$v_{t,j}$ is a system noise artificially added. If $v_{t,j}$ is equal to zero, estimates of $K_{t,j}$ is equivalent to the estimates of $K_j$ that is a time invariant permeability coefficient of region $j$. In this analysis log-normal distribution with $\sigma = 0.5$ is used for $v_{t,j}$. By combining hydraulic heads $h_t$ of all grid points and permeability coefficients $K_{t,j}$ we obtain state vector $x_t$.

On the other hand, observation model is given by

$$y_t = H_t x_t + w_t$$

where $H_t$ is $0-1$ matrix that is determined by the corresponding grid point of observation flow.

To estimate $K_j$, we only need to give filtering distributions or estimates of $K_{t,j}$ at last time step. Strictly speaking, the model is different from original one. In addition it is non-stationary model. Therefore, applicability and validity of this model should be also checked by convergence of estimates. We will focus on this point later.

C. Results

In the TE-1 step, the water level setting is determined as shown in Figure 4. It is put into a simulation model. Based on this setting, a numerical analysis is conducted (TE-2). From that result we obtain three time series of observation flow (TE-3). It corresponds to leakage of dam water. As the level of water reservoir increases, the leakage also increases.

In TE-4 and TE-5 step, observation flows are assimilated by the SIR. The number of ensemble members are set to 500. The assimilated result is shown in Figure 5 and 6. Solid and dashed lines in Figure 5 show estimated and true states of flow at core zone, directly beneath core zone and downstream area. At the core zone and beneath core zone, assimilated results fit data as soon as data assimilation starts whereas the flow at downstream does not fit soon. The reason can be considered that the flow of downstream area has less sensitivity from

<table>
<thead>
<tr>
<th>Flow Observation 1</th>
<th>Flow Observation 2</th>
<th>Flow Observation 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>40m (Low water level)</td>
<td>85m (High water level)</td>
<td>15, 10, 15</td>
</tr>
</tbody>
</table>

Figure 3. Analysis grid and settings
permeability coefficients, especially from that of core zone.

Figure 6 shows the estimation results of permeability coefficient parameters of foundation ground, core zone and curtain grouting zone. Solid lines are estimates of parameters and dashed lines are true values of them. The estimates of all regions are finally converged to true one. The result means estimation of permeability coefficients by the SIR and self-organizing model works well.

Figure 7, 8, 9 show the number of ensemble members of specified interval at all three regions. Shown time steps are 2, 3, 30 and 70 from left to right. These histograms correspond with an approximation of estimated filter PDF of permeability. The true value is also shown by vertical line at each step. The results show that filter PDF can be obtained correctly. From the results of core zone and curtain grouting zone at step 2 and foundation ground at step 30, introduced framework works well for the case of multi-modal estimates. Those modes represent possible scenarios at corresponding steps and so we can evaluate future probability of water leaks or other risks based on these estimates.

IV. CONCLUSION

We summarized sequential data assimilation for geotechnical simulations and discussed which of the algorithms should be used from the viewpoint of continuity of physical variables, fitness for nonlinearity and ensemble efficiency in estimation. The discussion result confirmed that the SIS is suitable for subsidence analysis. It is also confirmed that the SIR and the EnKF are suitable for seepage analysis and especially the SIR is the best choice. We applied the SIR to the seepage analysis in which water flow and unknown parameters are corrected. Validity of this approach is checked by a twin experiment whose result shows that the SIR works well for identification of correct parameters and water flows in seepage analysis. The result also shows that bimodality of the estimates can be appropriately detected by the SIR. That property enables us to give more accurate evaluation of risks in the geotechnical engineering.

REFERENCES

Figure 7. Posterior distribution of permeability coefficient at foundation ground

Figure 8. Posterior distribution of permeability coefficient at core zone

Figure 9. Posterior distribution of permeability coefficient at curtain grouting zone