Computational Data Assimilation: Project 2

October 25, 2022

1 Resources

The following resources may be useful while working on the project, beyond the class notes:

1. The syllabus contains pointers to three data assimilation textbooks, by Asch et al., Law et al., and Reich et al. The Asch et al. book is written at an introductory level and is particularly well accessible.

2. The book "Data Assimilation: A Mathematical Introduction" by K.J.H. Law et al., first part of which is available on arxiv at `https://arxiv.org/pdf/1506.07825.pdf`. See Chapters 4.2 for EnKF and 4.3 for particle filters, and Chapter 5 for Matlab programs.

3. The OTP Matlab package available from URL `https://github.com/ComputationalScienceLaboratory/ODE-Test-Problems`. It contains implementations of the Lorenz models, among others. Start with the readme file.

4. The DATools Matlab package available from `https://github.com/ComputationalScienceLaboratory/DATools`. It contains implementations of all the data assimilation algorithms we will discuss in class, including various particle filters. Start with the readme file.

5. The MATLODE Matlab package available from `https://github.com/ComputationalScienceLaboratory/MATLODE`. It contains implementations of many time stepping algorithms and direct and adjoint sensitivity analyses.

2 Models

2.1 Lorenz three-variables system

The Lorenz three-variables system Lorenz-63 [**?**] is described by the equations:

$$\begin{cases} x_1' = \sigma\,(x_2 - x_1), \\ x_2' = x_1\,(\rho - x_3) - x_2, \quad t_0 \le t \le t_{\mathrm{F}}, \quad \mathbf{x}(t_0) = \mathbf{x}_0, \\ x_3' = x_1\,x_2 - \beta\,x_3, \end{cases} \tag{1a}$$

with the parameter values

$$\sigma = 10, \quad \rho = 28, \quad \beta = 8/3. \tag{1b}$$

For these values of parameters the Lorenz-63 system has a chaotic dynamics. The chaotic solutions (started from appropriate initial values) fall onto the Lorenz attractor. This physical system is completely deterministic and yet inherently unpredictable.

We consider the evolution of nature (**??**) to be known and governed by the Lorenz equations (1a) with the system parameters (1b). The true state has $n = 3$ components. The initial conditions, however, are uncertain, and therefore the evolution of the entire system is uncertain.

At different times t_i we measure the first two components $x_1(t_i)$ and $x_2(t_i)$. The inverse problem is to estimate the entire state, and therefore to follow closely the model trajectory, from these measurements. The system in general ODE notation reads:

$$\mathbf{x}' = f(\mathbf{x})$$

where

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad f(\mathbf{x}) = \begin{bmatrix} \sigma\,(x_2 - x_1) \\ x_1\,(\rho - x_3) - x_2 \\ x_1\,x_2 - \beta\,x_3 \end{bmatrix}. \tag{2}$$

The Jacobian of the ODE function is the matrix of partial derivatives:

$$f_{\mathbf{x}}(\mathbf{x}) = \left(\frac{\partial f_i(\mathbf{x})}{\partial \mathbf{x}_j}\right)_{1 \le i,j \le n}.$$

For our function (2) the Jacobian matrix reads:

$$f_{\mathbf{x}}(\mathbf{x}) = \begin{bmatrix} -\sigma & \sigma & 0 \\ (\rho - x_3) & -1 & -x_1 \\ x_2 & x_1 & -\beta \end{bmatrix}. \tag{3}$$

2.2 The Lorenz-96 model

The Lorenz-96 model is described by the following system of ordinary differential equations [**?**]:

$$\frac{dx_i}{dt} = \begin{cases} (x_2 - x_{\mathrm{N_{state}}}) \cdot x_{\mathrm{N_{state}}} - x_1 + \varphi & \text{for } i = 1, \\ (x_{i+1} - x_{i-2}) \cdot x_{i-1} - x_i + \varphi & \text{for } 2 \le i \le \mathrm{N_{state}} - 1, \\ (x_1 - x_{\mathrm{N_{state}}-2}) \cdot x_{\mathrm{N_{state}}-1} - x_{\mathrm{N_{state}}} + \varphi & \text{for } i = \mathrm{N_{state}}, \end{cases} \tag{4}$$

which mimics fundamental properties of atmospheric dynamics. Typically $\mathrm{N_{state}} = 40$. This model exhibits extended chaos with an external forcing value $\varphi = 8.0$, when the solution is in the form of moving waves. For this reason, the model is adequate to perform basic studies of predictability. One time unit of the Lorenz-96 model (4) corresponds to 1.5 days of the atmosphere.

The observation operator for this problem is linear and selects only the odd-numbered states:

$$\mathbf{H} \in \mathbb{R}^{20\times 40}, \qquad \mathbf{H} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_{40} \end{bmatrix} = \begin{bmatrix} x_1 \\ x_3 \\ \vdots \\ x_{39} \end{bmatrix}.$$

3 Question

1. Consider the the Lorenz-96 model (4) with the setting used in Project 1. The data assimilation problem is posed over the interval $[0, T]$, starting with initial background states from $\mathcal{N}(\mathbf{x}_0^{\mathrm{b}}, \mathbf{B}_0)$. Observations are available at times $t_i = i\,\Delta t$, $i = 0, \ldots, N$, $t_N = T$. Use your synthetic observations, observation error covariance, and observation operator from Project 1.

2. Solve the data assimilation problem using the standard sequential importance with resampling (SIR) particle filter. Resample when the effective number of particles decreases below a fraction of the total number, $\mathrm{N}_{\mathrm{ens}}^{\mathrm{eff}} \le \alpha\,\mathrm{N}_{\mathrm{ens}}$, with $0 < \alpha < 1$ a tuning parameter under your control.

3. Solve the data assimilation problem using a sequential importance sampling (SIS) approach, with two different proposal transition densities (your choice) selected out of the following list:

   - EnKF;
   - Optimal proposal density formulated using Gaussian assumptions;
   - Model nudging;
   - Implicit sampling.

4. For each of the questions above please implement the specific filter under consideration, run it, and assess the quality of the results using the following metrics:

   - The time evolution of the error in the analysis ensemble mean, compared to the reference trajectory.
   - The rank histogram for one of the observations.

   How does the performance of particle filters compare against the performance of EnKF?

4 What to submit

1. A well articulated report detailing your work and the results you obtained, and
2. The well-organized, structured, and commented code you developed for each of the steps.