Project 5C

Our goal is to develop algorithms to solve the following type of "data assimilation" problem. Suppose we have a differential equation model of the form

$$\frac{d\vec{\mathrm{x}}}{dt} = \vec{\mathrm{f}}(\vec{\mathrm{x}}, \vec{\mathrm{c}})$$

where $\vec{\mathrm{x}}$ is a vector representing the "state" of a system and $\vec{\mathrm{c}}$ is a vector of parameters. Suppose there is a solution of the model that we will call the "truth", but we don't know the initial conditions nor the parameters for this true solution. What we do know are some measurements of the system at various times; mathematically, suppose we have an observing system of the form

$$\vec{\mathrm{y}}(t) = \vec{\mathrm{h}}(\vec{\mathrm{x}}(t)) + \vec{\varepsilon}(t), \qquad \vec{\varepsilon} \sim N(0, R),$$

where $\vec{\varepsilon}$ represents normally distributed measurement errors (with covariance matrix R) and $\vec{\mathrm{h}}$ is a function that expresses the measured quantities in terms of the system state, and we are given the vector of observations $\vec{\mathrm{y}}$ at times t_n. The problem is to estimate the true system state $\vec{\mathrm{x}}(t_n)$ and the parameters $\vec{\mathrm{c}}$ from the sequence of measurements $\vec{\mathrm{y}}(t_1), \vec{\mathrm{y}}(t_2), \ldots \vec{\mathrm{y}}(t_n)$, and to do this in "real time", so that when $\vec{\mathrm{y}}(t_{n+1})$ becomes available, we estimate $\vec{\mathrm{x}}(t_{n+1})$ using our previous estimate of $\vec{\mathrm{x}}(t_n)$ rather than solving the problem from scratch.

A specific model we will work with is a system of differential equations formulated by meteorologist Edward Lorenz in 1995:

$$\frac{dx_k}{dt} = (x_{k+1} - x_{k-2})x_{k-1} - x_k + c_k$$

where k runs from 1 to N and indices are taken modulo N; that is, $x_{-1} = x_{N-1}$, $x_0 = x_N$, and $x_{N+1} = x_1$. Typically N is on the order of 100 and the parameters c_k are on the order of 10 (usually c_k is taken to be independent of k). An observing system will be

$$y_k(t) = x_k(t) + \varepsilon_k(t)$$

but the measurements will be given only for some values of k; part of the problem will be to infer values for coordinates of the state vector $\vec{\mathrm{x}}$ that are never measured directly. Goals will include seeing how well you can estimate $\vec{\mathrm{x}}$ when you have plenty of high-quality measurements, how much measurement noise your methods can withstand, how few observations you can get by with, how large an N you can handle, etc.

We'll start with a simpler model, namely

$$x_{n+1} = ax_n + \delta_n, \qquad \delta_n \sim N(0, b)$$

with observations of the form

$$y_n = x_n + \varepsilon_n, \qquad \varepsilon_n \sim N(0, r).$$

Your first task will be to learn and apply the Kalman Filter in this scenario to optimally estimate x_n from $y_1, y_2, \ldots, y_n$. Some data will be given to you to start with, but as we go along you will start to generate your own test data sets for your models and algorithms.