Introduction to Bayesian Filtering

Andrea Arnold and Franz Hamilton

Department of Mathematics
Center for Quantitative Sciences in Biomedicine

North Carolina State University

July 30, 2016

Dynamic Model

NC STATE UNIVERSITY

Consider the system of differential equations

$$\frac{dx}{dt} = f(t, x, \theta), \qquad x(0) = x_0,$$

where

- $x = x(t) \in \mathbb{R}^d$ is the state vector
- $\theta \in \mathbb{R}^k$ is the unknown, or poorly known, parameter vector
- $f : \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^k \to \mathbb{R}^d$ is the known model function
- x_0 is the possibly unknown, or poorly known, initial value

Observation Model

The measured observations, assuming additive noise, are given by

$$b_j = g(x(t_j), \theta) + e_j \in \mathbb{R}^m, \qquad 0 < t_1 < t_2 < \ldots < t_T,$$

where

- $g : \mathbb{R}^d \times \mathbb{R}^k \to \mathbb{R}^m$, $m \leq d$, is the known observation function
- e_j is the observation error

The inverse problem: Estimate the parameter vector θ, the state vector $x(t)$ at given times, and possibly the initial value x_0, from the measurements b_j.

Discrete Time Markov Models

State evolution equation:

$$X_{j+1} = F(X_j, \theta) + V_{j+1}, \quad j = 0, 1, 2, \ldots$$

where

- F is a known propagation model
- V_{j+1} is an innovation process, independent of X_j
- θ is a parameter (vector)

Observation equation:

$$Y_j = G(X_j) + W_j, \quad j = 1, 2, \ldots$$

where

- G is a known operator
- W_j is the observation noise, independent of X_j

Denoting by D_j the accumulated observations up to time $t = t_j$,

$$D_j = \{y_1, y_2, \ldots, y_j\},$$

we aim to sequentially update the posterior distribution $\pi(x_j, \theta \mid D_j)$ using the following scheme:

$$\pi(x_j, \theta \mid D_j) \longrightarrow \pi(x_{j+1}, \theta \mid D_j) \longrightarrow \pi(x_{j+1}, \theta \mid D_{j+1})$$

Assume first that the parameter vector θ is fixed and known, so we aim to estimate only the states.

The sequential update of the posterior distribution $\pi(x_j \mid D_j)$ is then

$$\pi(x_j \mid D_j) \longrightarrow \pi(x_{j+1} \mid D_j) \longrightarrow \pi(x_{j+1} \mid D_{j+1})$$

NC STATE UNIVERSITY

Time evolution update: Chapman-Kolmogorov formula

$$\begin{aligned}\pi(x_{j+1} \mid D_j) &= \int \pi(x_{j+1} \mid x_j, D_j)\pi(x_j \mid D_j)dx_j \\ &= \int \pi(x_{j+1} \mid x_j)\pi(x_j \mid D_j)dx_j\end{aligned}$$

Observation update: Bayes' theorem

$$\begin{aligned}\pi(x_{j+1} \mid D_{j+1}) &= \pi(x_{j+1} \mid y_{j+1}, D_j) \\ &\propto \pi(y_{j+1} \mid x_{j+1}, D_j)\pi(x_{j+1} \mid D_j) \\ &= \pi(y_{j+1} \mid x_{j+1})\pi(x_{j+1} \mid D_j)\end{aligned}$$

In the special case when the state evolution equation and observation equation are both **linear** and **Gaussian**,

$$
\begin{aligned}
X_{j+1} &= \mathrm{F}X_j + V_{j+1}, \qquad V_{j+1} \sim \mathcal{N}(0, \mathrm{C}), \qquad j = 0, 1, 2, \ldots, \\
Y_{j+1} &= \mathrm{G}X_{j+1} + W_{j+1}, \quad W_{j+1} \sim \mathcal{N}(0, \mathrm{D}),
\end{aligned}
$$

the classical **Kalman filter** (KF) gives explicit updating formulas for both the prediction step and the analysis step.

We assume that the noise vectors V_{j+1} and W_{j+1} are mutually independent, and further that the distribution of the initial state X_0, $\pi(x_0) = \pi(x_0 \mid D_0)$, is known and Gaussian.

Notation

- Current state: $X_{j|j}$
- Predicted state: $X_{j+1|j}$
- Posterior state: $X_{j+1|j+1}$

From the state evolution equation, we have

$$X_{j+1|j} = \mathsf{F} X_{j|j} + V_{j+1}.$$

Assuming $X_{j|j} \sim \mathcal{N}(\overline{x}_{j|j}, \Gamma_{j|j})$, it follows that

$$X_{j+1|j} \sim \mathcal{N}(\overline{x}_{j+1|j}, \Gamma_{j+1|j})$$

where

$$\overline{x}_{j+1|j} = \mathsf{F}\overline{x}_{j|j} \qquad \text{and} \qquad \Gamma_{j+1|j} = \mathsf{F}\Gamma_{j|j}\mathsf{F}^{\mathsf{T}} + \mathsf{C}.$$

Therefore the density prior to the data arriving is of the form

$$\pi(x_{j+1} \mid D_j) \propto \exp\left\{ -\frac{1}{2}(x_{j+1} - \overline{x}_{j+1|j})^{\mathsf{T}} \Gamma_{j+1|j}^{-1} (x_{j+1} - \overline{x}_{j+1|j}) \right\}.$$

KF Analysis Step

NC STATE UNIVERSITY

Since $W_{j+1} \sim \mathcal{N}(0, \mathrm{D})$, the observation equation defines a Gaussian likelihood of the form

$$\pi(y_{j+1} \mid x_{j+1}) \propto \exp\left\{ -\frac{1}{2}(y_{j+1} - \mathrm{G}x_{j+1})^{\mathsf{T}}\mathrm{D}^{-1}(y_{j+1} - \mathrm{G}x_{j+1}) \right\}.$$

Thus, from Bayes' theorem, it follows that

$$\begin{aligned}
\pi(x_{j+1} \mid D_{j+1}) &\propto \pi(y_{j+1} \mid x_{j+1})\pi(x_{j+1} \mid D_j) \\
&= \exp\left\{ -\frac{1}{2}(y_{j+1} - \mathrm{G}x_{j+1})^{\mathsf{T}}\mathrm{D}^{-1}(y_{j+1} - \mathrm{G}x_{j+1}) \right. \\
&\qquad \left. -\frac{1}{2}(x_{j+1} - \overline{x}_{j+1|j})^{\mathsf{T}}\Gamma_{j+1|j}^{-1}(x_{j+1} - \overline{x}_{j+1|j}) \right\}.
\end{aligned}$$

KF Analysis Step

The posterior state mean $\overline{x}_{j+1|j+1}$ and posterior error covariance $\Gamma_{j+1|j+1}$ can be obtained by maximizing $\pi(x_{j+1} \mid D_{j+1})$ with respect to x_{j+1}, which is equivalent to minimizing the cost function

$$\begin{aligned}\mathcal{J}(x) &= (y_{j+1} - \mathsf{G}x)^{\mathsf{T}}\mathsf{D}^{-1}(y_{j+1} - \mathsf{G}x) \\ &\quad + (x - \overline{x}_{j+1|j})^{\mathsf{T}}\Gamma_{j+1|j}^{-1}(x - \overline{x}_{j+1|j}),\end{aligned}$$

i.e.,

$$\overline{x}_{j+1|j+1} = \arg\min_{x} \mathcal{J}(x).$$

It follows that the posterior state mean and posterior error covariance are given by

$$\overline{x}_{j+1|j+1} = \overline{x}_{j+1|j} + \mathsf{K}_{j+1}(y_{j+1} - \mathsf{G}\overline{x}_{j+1|j})$$

and

$$\Gamma_{j+1|j+1} = (\mathsf{I} - \mathsf{K}_{j+1}\mathsf{G})\Gamma_{j+1|j},$$

respectively, where the matrix

$$\mathsf{K}_{j+1} = \Gamma_{j+1|j}\mathsf{G}^{\mathsf{T}}(\mathsf{G}\Gamma_{j+1|j}\mathsf{G}^{\mathsf{T}} + \mathsf{D})^{-1}$$

is called the **Kalman gain**.

Recap: KF Algorithm

Given an initial prior distribution

$$\pi(x_0) = \pi(x_0 \mid D_0) \sim \mathcal{N}(\overline{x}_{0|0}, \Gamma_{0|0})$$

set $j = 0$.

1. **Prediction step**: Update the prior mean and covariance using the formulas

$$\overline{x}_{j+1|j} = \mathsf{F}\overline{x}_{j|j}$$

and

$$\Gamma_{j+1|j} = \mathsf{F}\Gamma_{j|j}\mathsf{F}^{\mathsf{T}} + \mathsf{C}.$$

Recap: KF Algorithm

2. **Observation update**: After observing y_{j+1}, update the posterior mean and error covariance via the formulas

$$\overline{x}_{j+1|j+1} = \overline{x}_{j+1|j} + \mathrm{K}_{j+1}(y_{j+1} - \mathrm{G}\overline{x}_{j+1|j})$$

   and

$$\Gamma_{j+1|j+1} = (\mathrm{I} - \mathrm{K}_{j+1}\mathrm{G})\Gamma_{j+1|j},$$

   where

$$\mathrm{K}_{j+1} = \Gamma_{j+1|j}\mathrm{G}^{\mathsf{T}}(\mathrm{G}\Gamma_{j+1|j}\mathrm{G}^{\mathsf{T}} + \mathrm{D})^{-1}.$$

3. If $j < T$, set $j = j + 1$ and repeat from Step 1; otherwise, stop.

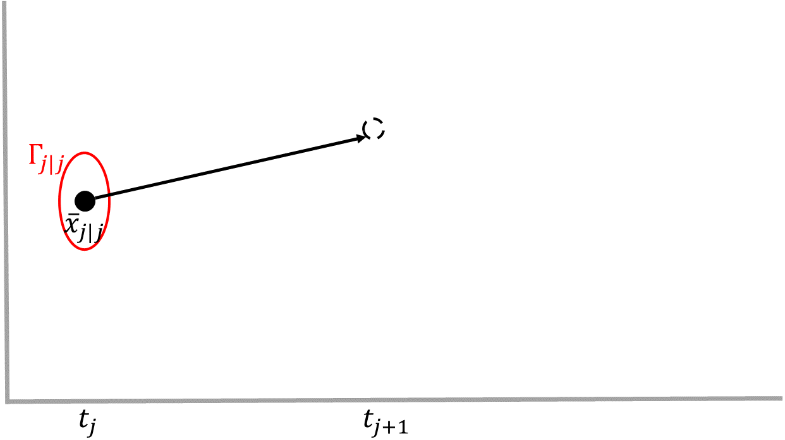
Schematic Representation of KF

Schematic Representation of KF

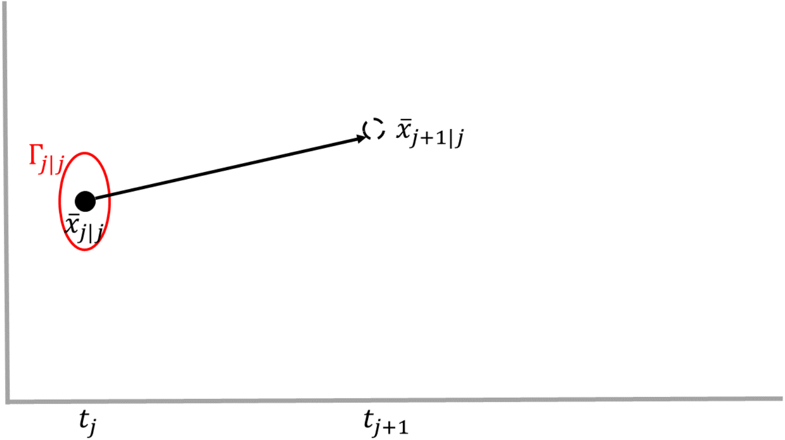
Schematic Representation of KF


$\Gamma_{j|j}$
$\bar{x}_{j|j}$
t_j

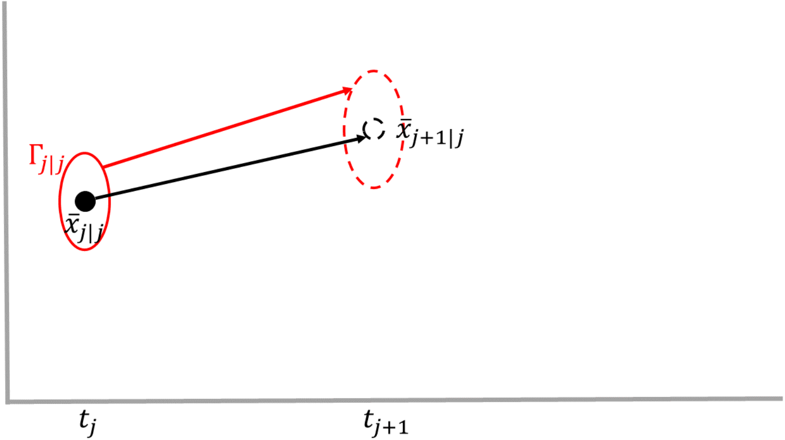
Schematic Representation of KF



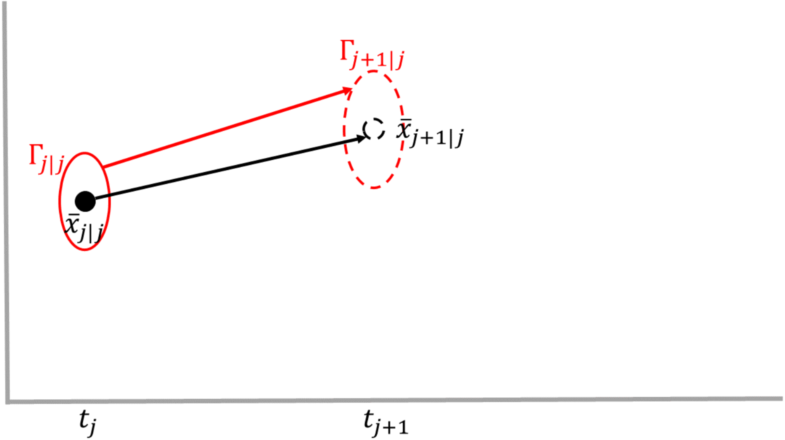
Schematic Representation of KF



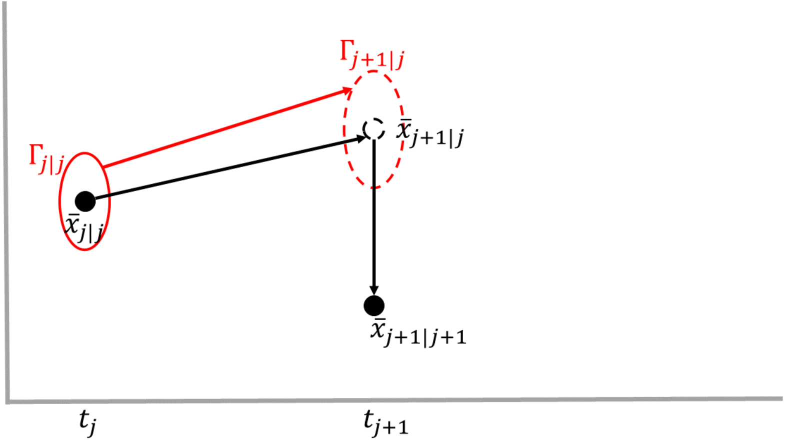
Schematic Representation of KF



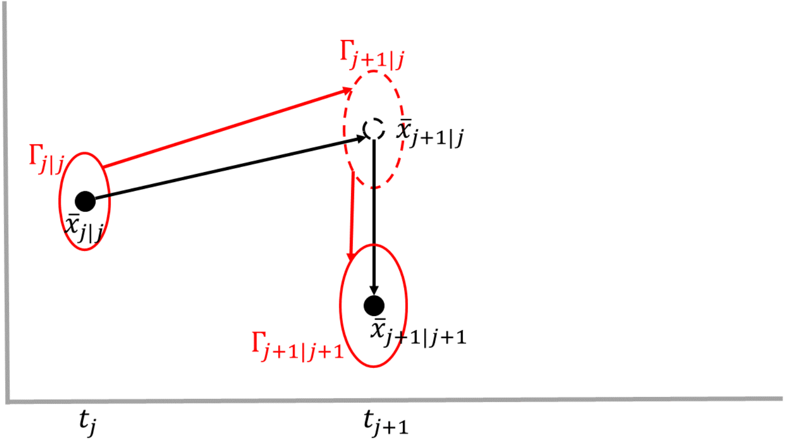
Schematic Representation of KF



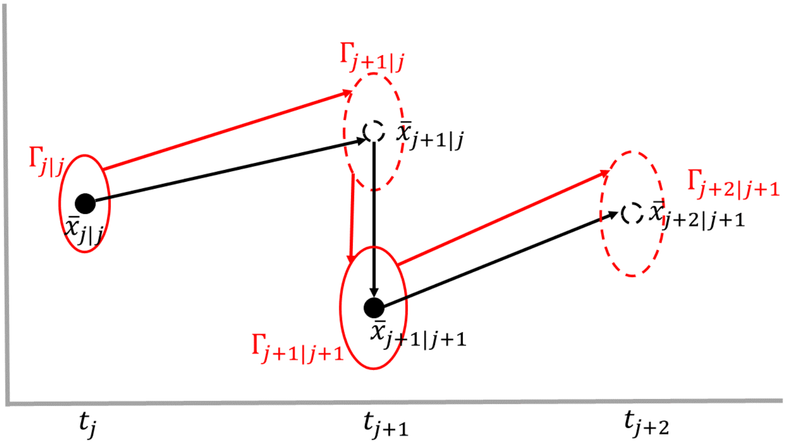
Schematic Representation of KF



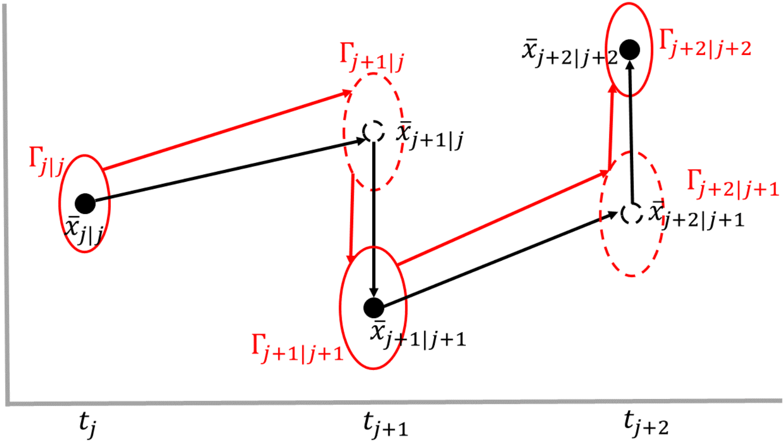
Schematic Representation of KF



Schematic Representation of KF



Schematic Representation of KF



Example: Mass-Spring System

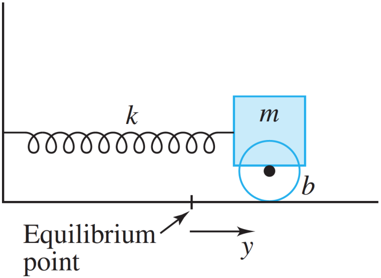


Figure 4.1 Damped mass–spring oscillator

[Figure from Nagle, Saff and Snider (2011)]

Example: Mass-Spring System

NC STATE UNIVERSITY

Consider the damped mass-spring oscillator

$$mp''(t) + bp'(t) + kp(t) = 0$$

where

- $p(t)$ denotes the position of mass at time t
- $m > 0$ is the mass
- $b \geq 1$ is the damping coefficient
- $k > 0$ is the spring constant

Example: Mass-Spring System

This can be written as the first-order linear system

$$\begin{aligned} \frac{dp}{dt} &= v \\ \frac{dv}{dt} &= -\frac{k}{m}p - \frac{b}{m}v \end{aligned}$$

where $v(t) = p'(t)$ denotes the velocity of the mass at time t, and letting

$$x = \begin{bmatrix} p \\ v \end{bmatrix} \in \mathbb{R}^2$$

yields the matrix equation

$$\frac{dx}{dt} = \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{b}{m} \end{bmatrix} x = Ax.$$

NC STATE UNIVERSITY

State evolution equation:

$$X_{j+1} = (\mathrm{I} + \Delta t \mathrm{A})X_j + V_{j+1}$$

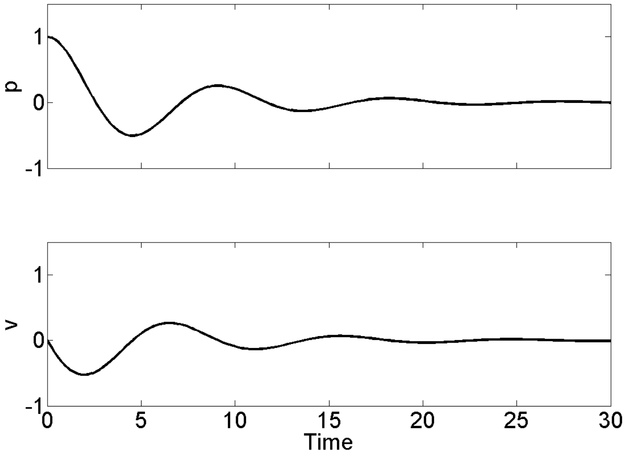
Assume noisy, direct observations of position p at discrete times t_j.

Observation equation:

$$Y_j = \begin{bmatrix} 1 & 0 \end{bmatrix} X_j + W_{j+1}$$

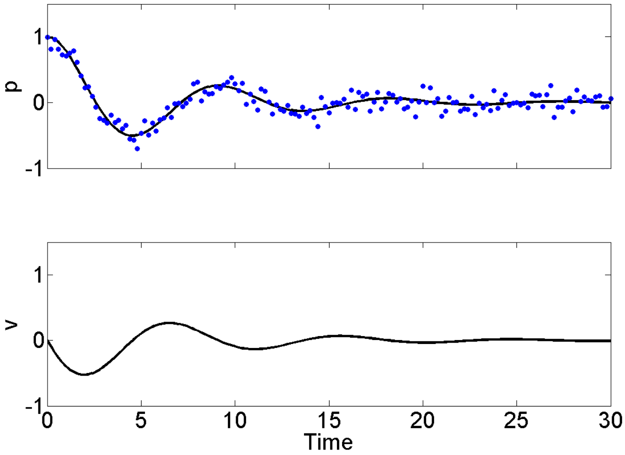
Goal: Use KF to estimate $x(t)$ at times t_j from observations y_j.

Example: Mass-Spring System



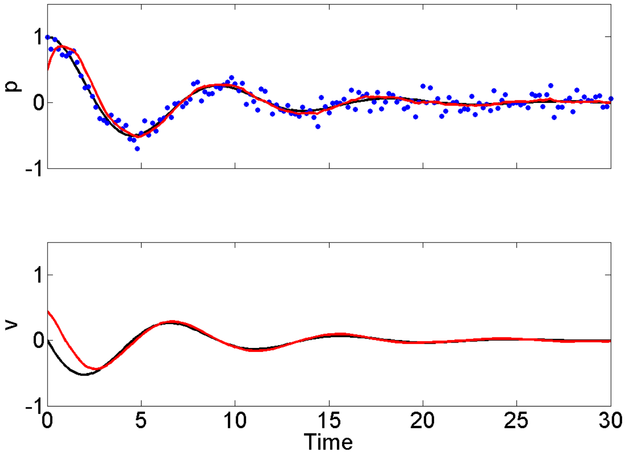
True State

Example: Mass-Spring System



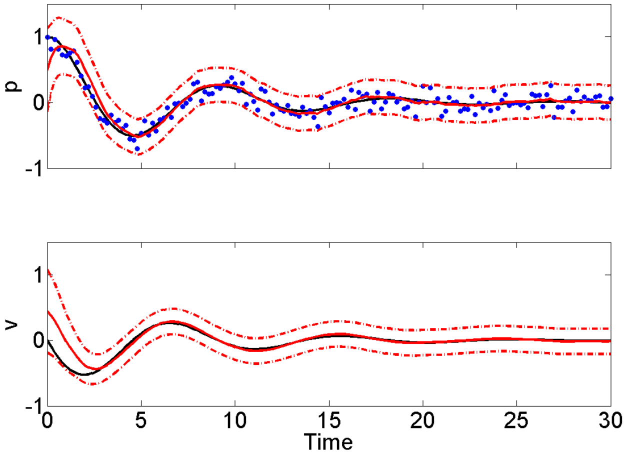
True State **Noisy Obs**

Example: Mass-Spring System



True State **Noisy Obs** **Filter**

Example: Mass-Spring System



True State **Noisy Obs** **Filter (w/ UQ)**

Kalman equations designed for **linear systems**

Problem: How to propogate mean and covariance of nonlinear function?

Solution: Non-trivial (in fact we end up **approximating** this forward propagation)

- Linearization (Extended Kalman filter)
- Statistical sampling (e.g. Ensemble Kalman filter, Unscented Kalman filter)

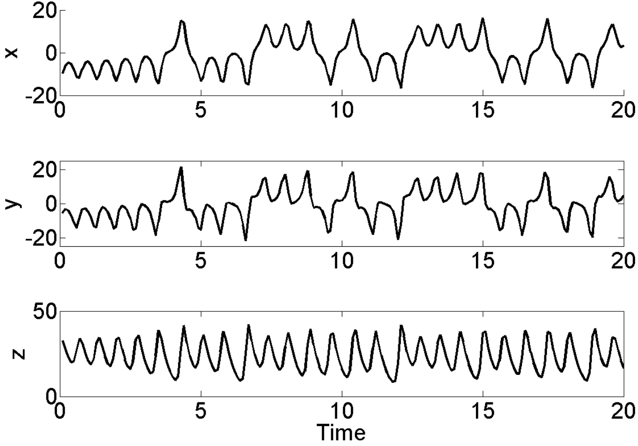
Example: Lorenz-63 System

Consider the classical Lorenz-63 system modeling atmospheric convection

$$\begin{aligned} \frac{dx}{dt} &= \sigma(y - x) \\ \frac{dy}{dt} &= x(\rho - z) - y \\ \frac{dz}{dt} &= xy - \beta z \end{aligned}$$

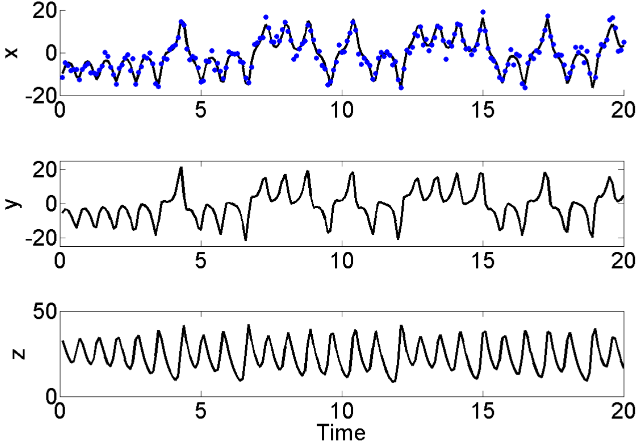
with nominal parameters

- $\sigma = 10$
- $\rho = 28$
- $\beta = \frac{8}{3}$



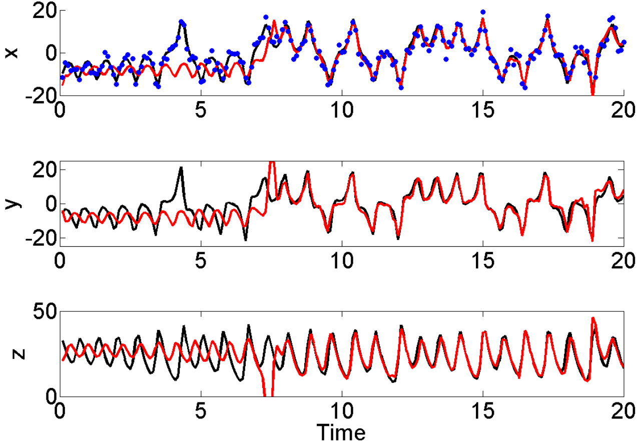
True State

Example: Lorenz-63 System



True State **Noisy Obs**

Example: Lorenz-63 System



True State **Noisy Obs** **Filter Estimate**

Estimation of States and Parameters

NC STATE UNIVERSITY

If parameters θ unknown, assume

$$\frac{d\theta}{dt} = \begin{bmatrix} \frac{d\theta_1}{dt} \\ \vdots \\ \frac{d\theta_m}{dt} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

and form an augmented state vector

$$z = \begin{bmatrix} x_1 \\ \vdots \\ x_n \\ \theta_1 \\ \vdots \\ \theta_m \end{bmatrix}$$

allowing for simultaneous state and parameter estimation

Example: Lorenz-63 State Parameter Estimation

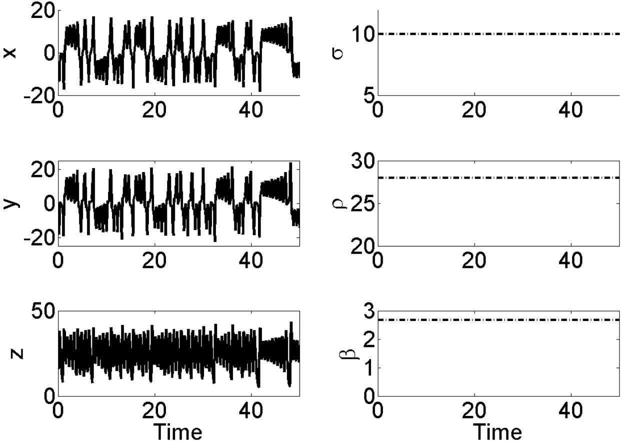
NC STATE UNIVERSITY

Idea: Extend the system equations

$$\begin{aligned} \frac{dx}{dt} &= \sigma(y - x) \\ \frac{dy}{dt} &= x(\rho - z) - y \\ \frac{dz}{dt} &= xy - \beta z \end{aligned}$$

Idea: Extend the system equations

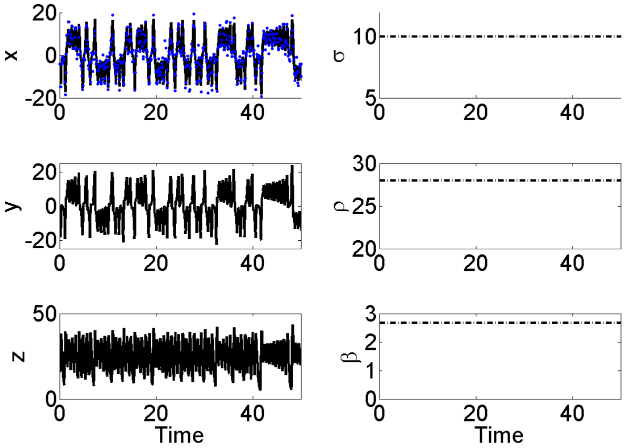
$$
\begin{aligned}
\frac{dx}{dt} &= \sigma(y - x) \\
\frac{dy}{dt} &= x(\rho - z) - y \\
\frac{dz}{dt} &= xy - \beta z \\
\frac{d\sigma}{dt} &= 0 \\
\frac{d\rho}{dt} &= 0 \\
\frac{d\beta}{dt} &= 0
\end{aligned}
$$



True State

Example: Lorenz-63 State Parameter Estimation

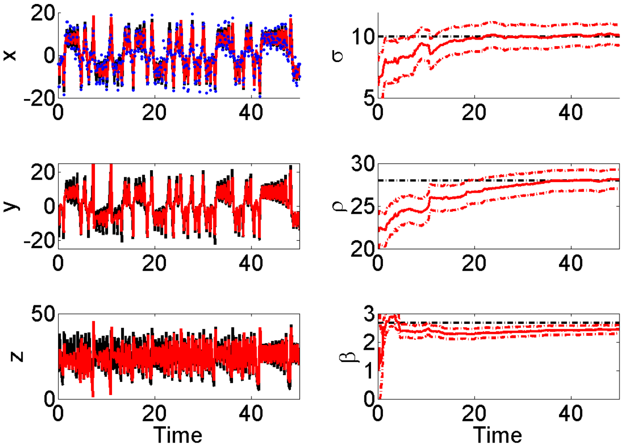
NC STATE UNIVERSITY



True State **Noisy Obs**

Example: Lorenz-63 State Parameter Estimation

NC STATE UNIVERSITY



True State **Noisy Obs** **Filter Estimate**

References

- J. Kaipio and E. Somersalo (2005) Statistical and Computational Inverse Problems. Springer, New York.
- D. Simon (2006) Optimal State Estimation: Kalman, H_∞, and Nonlinear Approaches. John Wiley and Sons, New York.
- A. J. Majda and J. Harlim (2012) Filtering Complex Turbulent Systems. Cambridge University Press.
- H. Voss, J. Timmer and J. Kurths (2004) Nonlinear Dynamical System Identification from Uncertain and Indirect Measurements. International Journal of Bifurcation and Chaos, 14, 1905-1933.
- A. Arnold, D. Calvetti and E. Somersalo (2014) Parameter estimation for stiff deterministic dynamical systems via ensemble Kalman filter. Inverse Problems, 30 (10), 105008.

Questions?