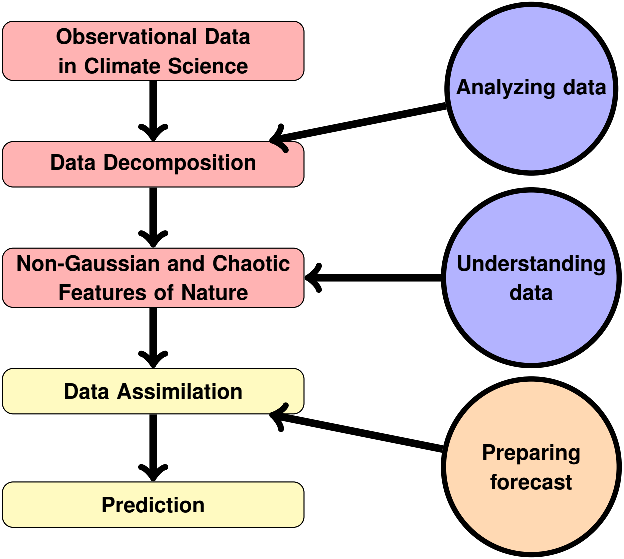
Introduction to Data Assimilation and Kalman Filter

Nan Chen

Department of Mathematics
University of Wisconsin-Madison

Math 717, Spring 2024

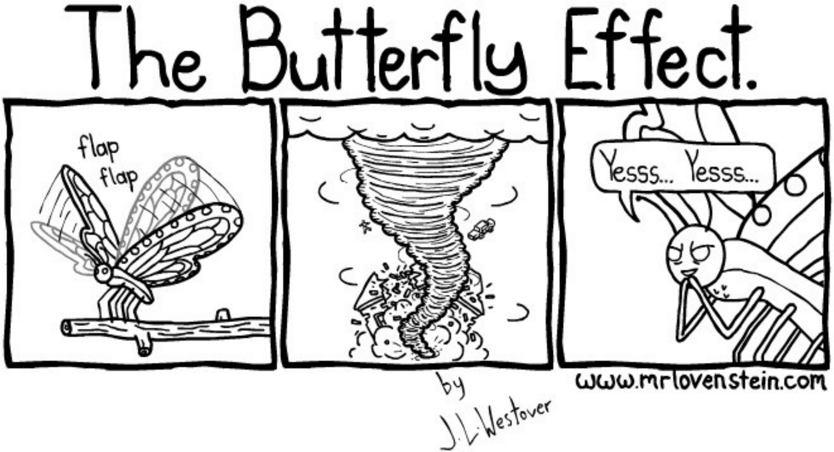

Observational Data
in Climate Science
Data Decomposition
Non-Gaussian and Chaotic
Features of Nature
Data Assimilation
Prediction
Analyzing data
Understanding
data
Preparing
forecast

Outline

- Main difficulties in predicting nature
- General data assimilation
- Filtering and Kalman filter
- Comments on nonlinear filters

Main difficulties in predicting nature

Nature is chaotic

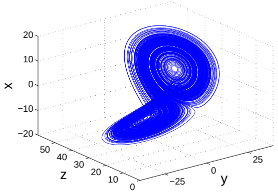


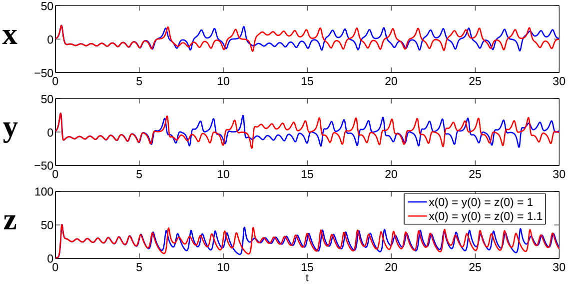
The flap of a butterfly's wings in Brazil set off a tornado in Texas.

In chaos theory, the butterfly effect is the sensitive dependence on initial conditions in which a small change in one state of a deterministic nonlinear system can result in large differences in a later state.

Example: The Lorenz 63 model

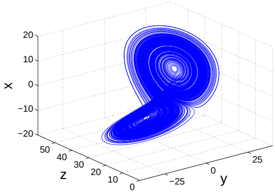
$$dx = \sigma(y - x)dt,$$
$$dy = \big(x(\rho - z) - y\big)dt,$$
$$dz = (xy - \beta z)dt,$$

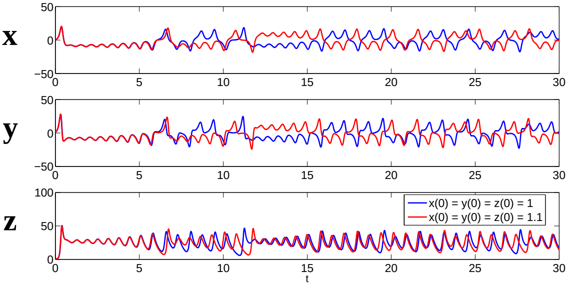




Example: The Lorenz 63 model

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



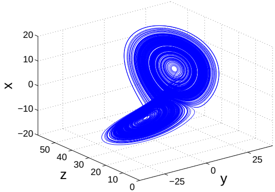


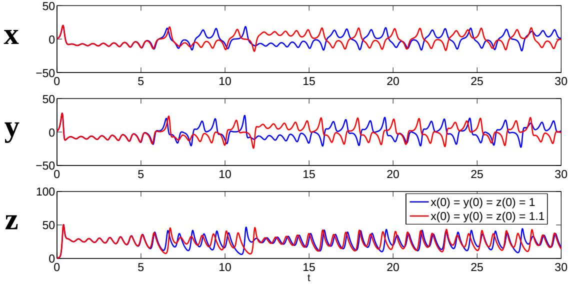
Many nature phenomena are chaotic systems. Predicting the chaotic systems is **extremely difficult!** We need **accurate initial conditions** for skillful predictions.

However, observations are noisy and models contain model error. Then, how to get accurate initial conditions?

Example: The Lorenz 63 model

$$dx = \sigma(y - x)dt,$$
$$dy = \big(x(\rho - z) - y\big)dt,$$
$$dz = (xy - \beta z)dt,$$



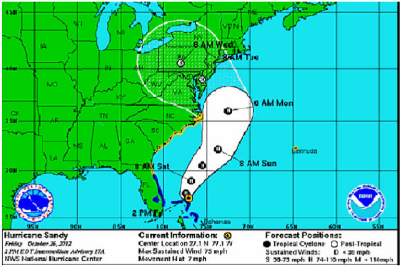


Many nature phenomena are chaotic systems. Predicting the chaotic systems is **extremely difficult!** We need **accurate initial conditions** for skillful predictions.

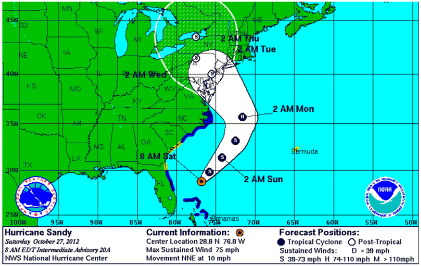
However, observations are noisy and models contain model error. Then, how to get accurate initial conditions? **Data Assimilation!**

Example: a). Hurricane Sandy (Oct 30, 2012)

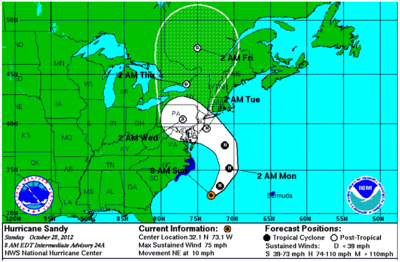
Oct 26



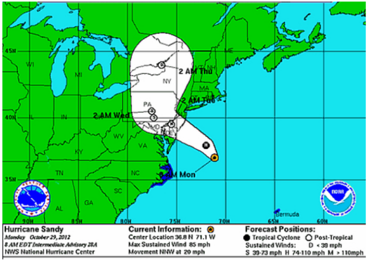
Oct 27



Oct 28



Oct 29



NYU Classes are Cancelled and Offices are Closed Tomorrow Inbox x

Jules Martin- NYU Vice President for Global Security and Crisis Management <jules.martin@nyu.edu>
to

Oct 28, 2012, 12:02 PM

NEW YORK UNIVERSITY

TO: THE NYU COMMUNITY

FROM: Jules Martin, Vice President for Global Security and Crisis Management

RE: Cancellation of Classes and Normal University Operations on Monday, Oct 29.

===

The safety and well-being of the NYU community is of foremost concern to us. Due to the hazards presented by Hurricane Sandy and the announced shut-down this evening at 7:00 pm of the New York's transit systems, all classes, activities, and events will be cancelled tomorrow, Monday, Oct 29. University offices will be closed.

This communication does not apply to NYU Langone Medical Center, including the School of Medicine, will separately be making determinations and communicating directly with the Med Ctr. community.


Crackers
Oct 28

Oct 28

you thought


Oct 28


truth

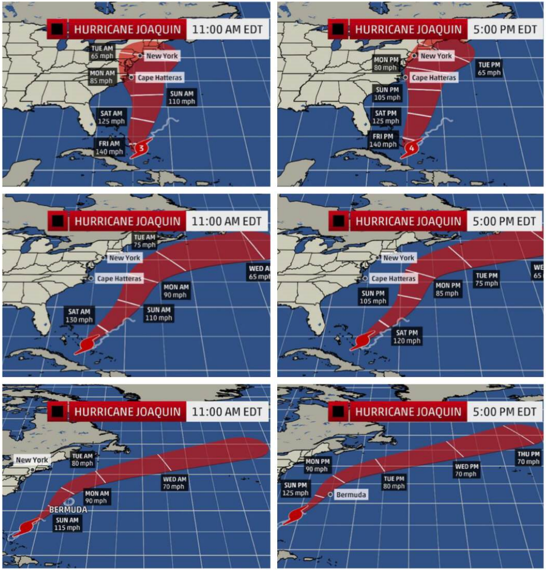

you thought

b). Hurricane Joaquin (Oct 3, 2015)

Oct 1

Oct 2

Oct 3

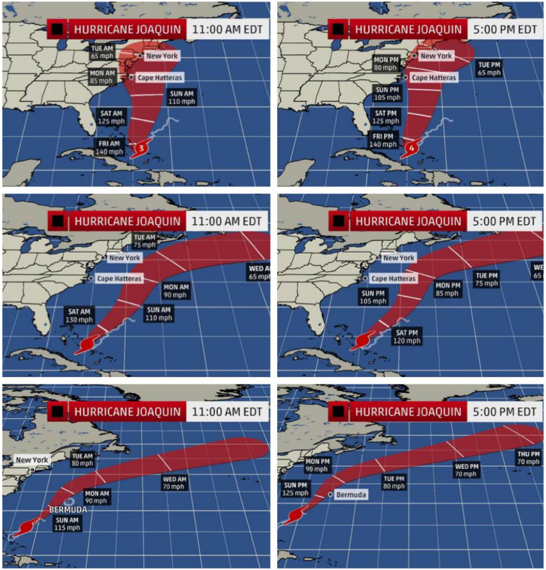


b). Hurricane Joaquin (Oct 3, 2015)

Oct 1

Oct 2

Oct 3



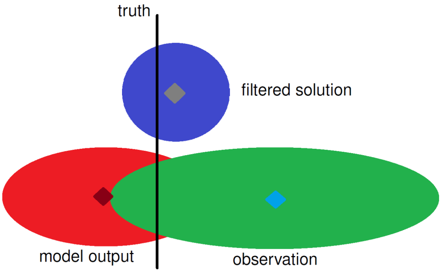
Note: The initial condition (on Oct 1) of the Joaquin was almost the same as that of the Sandy! But the two events afterwards were completely different!

General data Assimilation

What is data assimilation?

Data assimilation (DA) seeks to optimally combine a numerical model with observations to improve results, where

- the model is typically chaotic and has uncertainties/model errors, while
- the observations contain noise and are often available only for a subset of the state variables.



Why models are not accurate? Because we do not perfectly know nature.

A famous aphorism by George Box: **"All models are wrong, but some are useful"**.

Data assimilation is the process of obtaining the best statistical estimate of a natural system from partial observations of the true signal from nature.

- Model + Data = Data Assimilation.

Data assimilation is the process of obtaining the best statistical estimate of a natural system from partial observations of the true signal from nature.

- Model + Data = Data Assimilation.
- Model with bias and uncertainty + Data with noise = Data Assimilation.

Data assimilation is the process of obtaining the best statistical estimate of a natural system from partial observations of the true signal from nature.

- Model + Data = Data Assimilation.
- Model with bias and uncertainty + Data with noise = Data Assimilation.

e.g.,

$$
\begin{aligned}
& \text{Model} && \text{(30\% error + 50\% uncertainty)} \\
+\ & \text{Data} && \text{(30\% error + 50\% uncertainty)} \\
\xrightarrow{\text{data assimilation}}\ & \text{Estimation} && \text{(10\% error + 20\% uncertainty)}
\end{aligned}
$$

Data assimilation can be achieved by the Bayesian formula:

$$\mathbf{p(u|v) \sim p(u)p(v|u)}$$

- $p(u|v)$: posterior, this is what we want.
- $p(u)$: prior, from model.
- $p(v|u)$: likelihood, from observation.

Consider a time interval $[0, T]$. Consider time discretizations of the state variable and observation with gap Δt such that $(I+1)\Delta t = T$,

$$u = (u_0, u_1, \dots, u_I) \qquad v = (v_0, v_1, \dots, v_I)$$

Two widely used data assimilation methods

$$\begin{aligned} \text{Filtering :} &\qquad p(u_i|v_0, \dots, v_i) \\ \text{Smoothing :} &\qquad p(u_i|v_0, \dots, v_I) \end{aligned}$$

or in general filtering/smoothing can be represented as

$$p(u_1, \dots, u_I|v_0, \dots, v_I)$$

With this in hand, we can easily take the marginal distribution to reach $p(u_i|v_0, \dots, v_I)$
If $i < I$, then it is smoothing. If $i = I$ (e.g., the end point), then it is filtering.

However, $p(u_1, \dots, u_I|v_0, \dots, v_I)$ is high dimensional and computationally challenging. Nevertheless, we can solve it sequentially. Similar to the filtering.

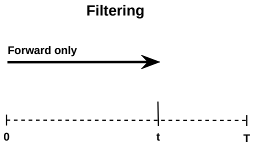
Filtering

using the information only in the past

cheaper but less accurate

online

the precursor of real-time prediction



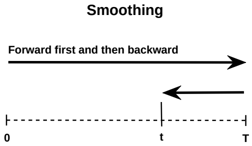
Smoothing

using the entire available information

more accurate but more expensive

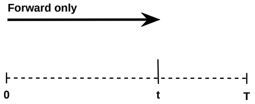
offline

applying for post-processing of data

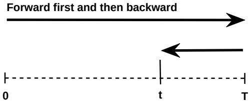


Filtering	**Smoothing**
using the information only in the past	using the entire available information
cheaper but less accurate	more accurate but more expensive
online	offline
the precursor of real-time prediction	applying for post-processing of data

Filtering

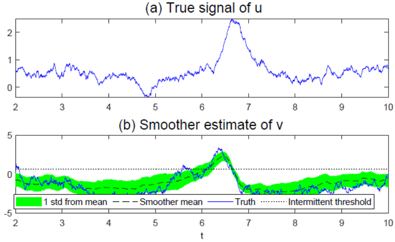


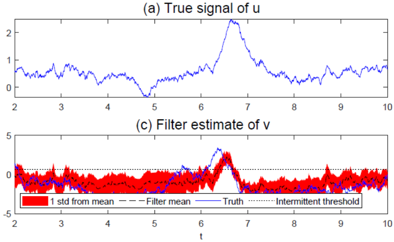
Smoothing



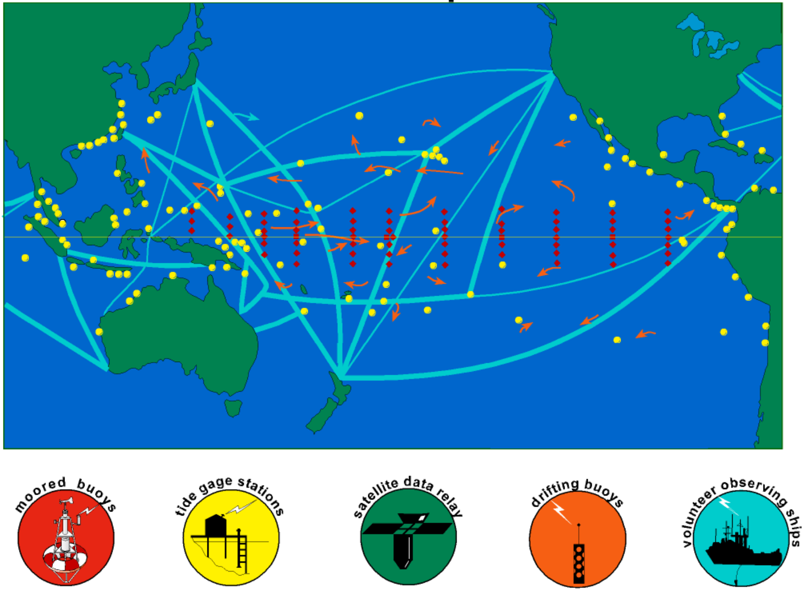
Example: Recovering the states of v given the observed time series of u in the following *nonlinear* model,

$$\mathrm{d}u = ((-d_u + cv)u + F_u)\,\mathrm{d}t + \sigma_u\,\mathrm{d}W_u, \qquad \mathrm{d}v = (-d_v v - cu^2)\,\mathrm{d}t + \sigma_v\,\mathrm{d}W_v.$$



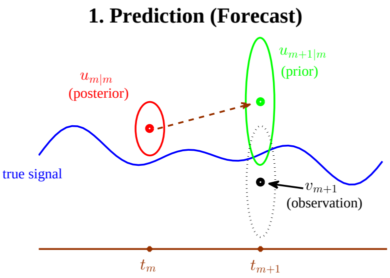


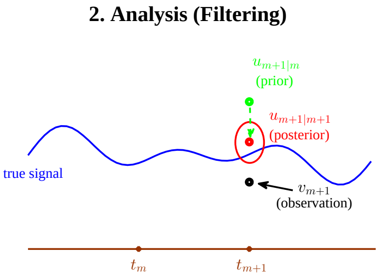
TOGA In Situ Ocean Observing System
Global Tropics



Filtering and Kalman filter

Filtering: A Two-Step Procedure

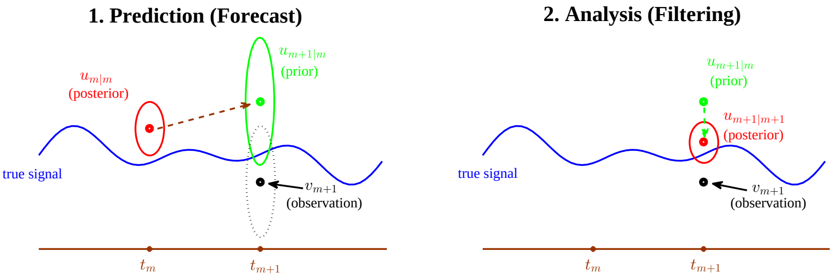




Two steps at each time instant $t_m = m\Delta t$.

Step 1. A statistical prediction of a probability distribution $u_{m+1|m}$ starting from the initial value $u_{m|m}$ using the given dynamical model.

Filtering: A Two-Step Procedure

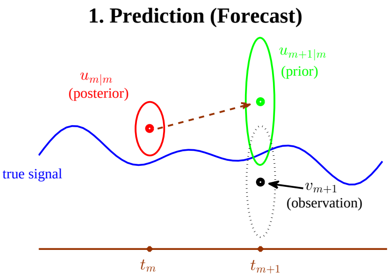


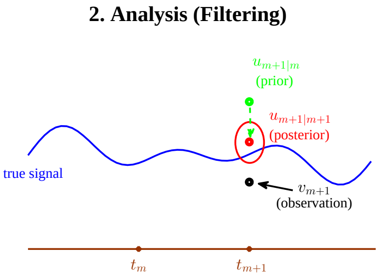
Two steps at each time instant $t_m = m\Delta t$.

Step 1. A statistical prediction of a probability distribution $u_{m+1|m}$ starting from the initial value $u_{m|m}$ using the given dynamical model.

Step 2. $u_{m+1|m}$ is corrected on the basis of the statistical input of noisy observation v_{m+1}, which results in $u_{m+1|m+1}$.

Filtering: A Two-Step Procedure





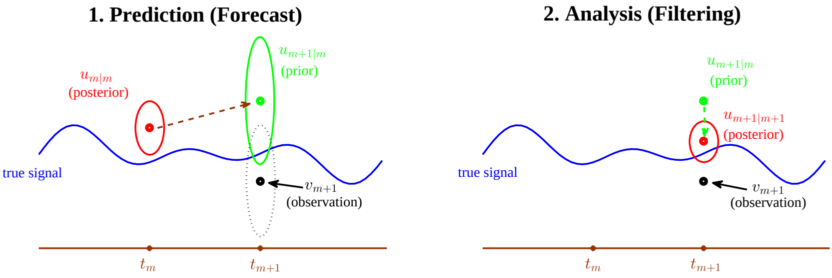
Two steps at each time instant $t_m = m\Delta t$.

Step 1. A statistical prediction of a probability distribution $u_{m+1|m}$ starting from the initial value $u_{m|m}$ using the given dynamical model.

Step 2. $u_{m+1|m}$ is corrected on the basis of the statistical input of noisy observation v_{m+1}, which results in $u_{m+1|m+1}$.

For linear system with Gaussian noise, the above procedure is known as the *Kalman filter*.

Filtering: 1-D Complex Case

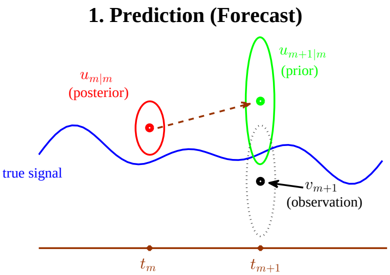


Let $u_m \in \mathbb{C}$ be a complex random variable whose dynamics is given by the following,

$$u_{m+1} = Fu_m + \mathcal{F}_{m+1} + \sigma_{m+1},$$

- σ_{m+1} is a complex Gaussian noise with $\sigma_{m+1} = (\sigma_{1,m+1} + i\sigma_{2,m+1})/\sqrt{2}$ and it has zero mean and variance $r = \langle \sigma_{m+1}\sigma^*_{m+1} \rangle = \frac{1}{2}\sum_{j=1}^{2}\langle \sigma^2_{j,m+1} \rangle$.
- F is a complex number known as the forward operator.
- $\mathcal{F}$ is an external forcing which can vary in time.

Filtering: 1-D Complex Case

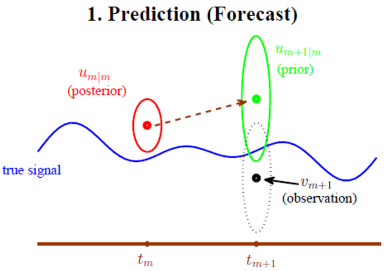


The goal of the Kalman filter is to estimate the unknown true state u_{m+1}, given noisy observations

$$v_{m+1} = g u_{m+1} + \sigma^o_{m+1},$$

- g is a linear observation operator.
- $\sigma^o_m \in \mathbb{C}$ is an unbiased Gaussian noise with variance $r^o = \langle \sigma^o_m (\sigma^o_m)^* \rangle$.

Filtering: 1-D Complex Case



Model:	$u_{m+1} = Fu_m + \mathcal{F}_{m+1} + \sigma_{m+1},$
Obs:	$v_{m+1} = gu_{m+1} + \sigma^o_{m+1}.$

Assume the filter model is perfectly specified.

An estimate of the true state prior to knowledge of the observation at time t_{m+1}, which is known as the prior state or forecast state, is given by

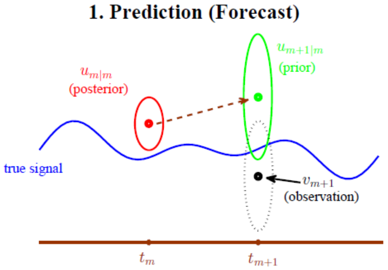
$$u_{m+1|m} = Fu_{m|m} + \mathcal{F}_{m+1} + \sigma_{m+1}.$$

This prior estimate can be represented with a probability density

$$p(u_{m+1}) \sim \mathcal{N}(\bar{u}_{m+1|m}, r_{m+1|m}),$$

which accounts only for the earlier observations up to time t_m.

Filtering: 1-D Complex Case



Model:	$u_{m+1} = Fu_m + \mathcal{F}_{m+1} + \sigma_{m+1},$
Obs:	$v_{m+1} = gu_{m+1} + \sigma^o_{m+1}.$

The prior mean and prior variance

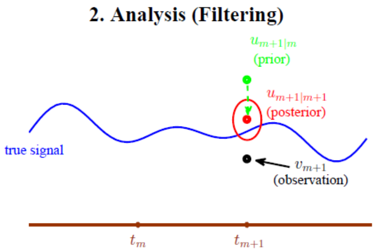
$$\bar{u}_{m+1|m} \equiv \langle u_{m+1|m} \rangle,$$
$$r_{m+1|m} \equiv \langle (u_{m+1} - \bar{u}_{m+1|m})(u_{m+1} - \bar{u}_{m+1|m})^* \rangle,$$

are solved via

$$\bar{u}_{m+1|m} = F\bar{u}_{m|m} + \mathcal{F}_{m+1},$$
$$r_{m+1|m} = Fr_{m|m}F^* + r,$$

with $r_{m|m} = \langle (u_m - \bar{u}_{m|m})(u_m - \bar{u}_{m|m})^* \rangle$. Note that in order to solve the prior distribution $p(u_{m+1|m})$, the posterior information in the previous step $\bar{u}_{m|m}, r_{m|m}$ has been used.

Filtering: 1-D Complex Case



Model:	$u_{m+1} = Fu_m + \mathcal{F}_{m+1} + \sigma_{m+1},$
Obs:	$v_{m+1} = gu_{m+1} + \sigma^o_{m+1}.$

The posterior state (or the filtered state) combines the prior information $p(u_{m+1|m})$ with the observation v_{m+1} at t_{m+1}.

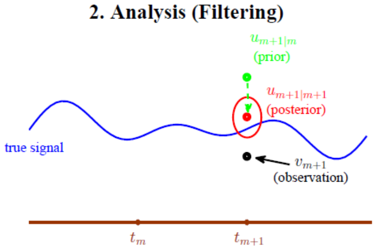
This estimate is given in the probabilistic sense by the Bayesian update through maximizing the following conditional density,

$$p(u_{m+1}|v_{m+1}) \sim p(u_{m+1})p(v_{m+1}|u_{m+1}) = e^{-\frac{1}{2}J(u_{m+1})},$$

which is equivalent to minimizing

$$J(u) = \frac{(u - \bar{u}_{m+1|m})^*(u - \bar{u}_{m+1|m})}{r_{m+1|m}} + \frac{(v_{m+1} - gu)^*(v_{m+1} - gu)}{r^o}.$$

Filtering: 1-D Complex Case



Model:	$u_{m+1} = Fu_m + \mathcal{F}_{m+1} + \sigma_{m+1},$
Obs:	$v_{m+1} = gu_{m+1} + \sigma^o_{m+1}.$

$$J(u) = \frac{(u - \bar{u}_{m+1|m})^*(u - \bar{u}_{m+1|m})}{r_{m+1|m}} + \frac{(v_{m+1} - gu)^*(v_{m+1} - gu)}{r^o}.$$

The value of u at which $J(u)$ attains its minimum is the estimate for the mean

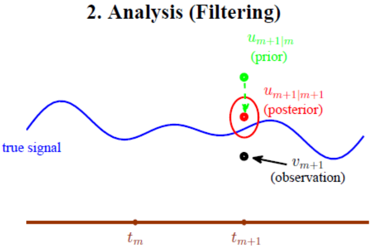
$$\bar{u}_{m+1|m+1} = (1 - K_{m+1}g)\bar{u}_{m+1|m} + K_{m+1}v_{m+1},$$

where

$$K_{m+1} = \frac{gr_{m+1|m}}{r^o + g^2 r_{m+1|m}}$$

is the Kalman gain. Note that $0 \leq K_{m+1}g \leq 1$.

Filtering: 1-D Complex Case



Model:	$u_{m+1} = F u_m + \mathcal{F}_{m+1} + \sigma_{m+1},$
Obs:	$v_{m+1} = g u_{m+1} + \sigma^o_{m+1}.$

$$J(u) = \frac{(u - \bar{u}_{m+1|m})^*(u - \bar{u}_{m+1|m})}{r_{m+1|m}} + \frac{(v_{m+1} - gu)^*(v_{m+1} - gu)}{r^o}.$$

The value of u at which $J(u)$ attains its minimum is the estimate for the mean

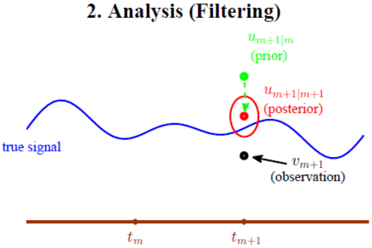
$$\bar{u}_{m+1|m+1} = (1 - K_{m+1} g)\bar{u}_{m+1|m} + K_{m+1} v_{m+1},$$

where

$$K_{m+1} = \frac{g r_{m+1|m}}{r^o + g^2 r_{m+1|m}}$$

is the Kalman gain. Note that $0 \le K_{m+1} g \le 1$.
The filter fully weighs to the model or prior forecast when $K_{m+1} g = 0$ and fully weighs to the observation when $K_{m+1} g = 1$. Such weights depend on the ratio of the uncertainty (reflected by the noise) in the observations and the model.

Filtering: 1-D Complex Case



Model: $u_{m+1} = Fu_m + \mathcal{F}_{m+1} + \sigma_{m+1},$

Obs: $v_{m+1} = gu_{m+1} + \sigma^o_{m+1}.$

Posterior mean:

$$\bar{u}_{m+1|m+1} = (1-K_{m+1}g)\bar{u}_{m+1|m} + K_{m+1}v_{m+1}.$$

Finally, the posterior variance is calculated via the following

$$\begin{aligned} u_{m+1} - \bar{u}_{m+1|m+1} &= u_{m+1} - \bar{u}_{m+1|m} - K_{m+1}(v_{m+1} - gu_{m+1} - g(\bar{u}_{m+1} - u_{m+1})) \\ e_{m+1|m+1} &= (1 - K_{m+1}g)e_{m+1|m} - K_{m+1}\sigma^o_{m+1}. \end{aligned}$$

These result in the expression of the posterior variance

$$r_{m+1|m+1} = (1 - K_{m+1}g)r_{m+1|m}.$$

Note that the above Kalman filter is designed for linear system with Gaussian noise. In practice, different generalizations of the Kalman filter such as the ensemble Kalman filter, particle filter and blended filtering techniques are applied to nonlinear and non-Gaussian systems.

Example.

Consider a complex scalar forced OU process,

$$\frac{du}{dt} = (-\gamma + i\omega_0)u + f_0 + f_1 e^{i\omega_1 t} + \sigma \dot{W},$$

- γ and ω_0 are the damping and oscillation frequency.
- f_0 and $f_1 e^{i\omega_1 t}$ are a constant and time-periodic large-scale forcing.
- $\sigma \dot{W}$ is stochastic noise.

Example.

Consider a complex scalar forced OU process,

$$\frac{du}{dt} = (-\gamma + i\omega_0)u + f_0 + f_1 e^{i\omega_1 t} + \sigma \dot{W},$$

- γ and ω_0 are the damping and oscillation frequency.
- f_0 and $f_1 e^{i\omega_1 t}$ are a constant and time-periodic large-scale forcing.
- $\sigma \dot{W}$ is stochastic noise.

Despite the simplicity of this model, it can be used to mimic some climate physics, e.g.,

- The deterministic forcing $f_0 + f_1 e^{i\omega_1 t}$ can be regarded as the annual cycle.
- ω_0 can be treated as internal oscillation which may occur in the intreaseasonal time scale.
- The damping term γ measures the system memory and the stochastic term represents the input to the system from small or unresolved scales.
- The model can also be regarded as one Fourier mode of complex spatially-extended systems.

The simple test model:

$$\frac{du}{dt} = (-\gamma + i\omega_0)u + f_0 + f_1 e^{i\omega_1 t} + \sigma \dot{W}.$$

How to put this model into the framework of Kalman filter?

$$\begin{aligned}
\text{Model:} \qquad & u_{m+1} = F u_m + \mathcal{F}_{m+1} + \sigma_{m+1}, \\
\text{Obs:} \qquad & v_{m+1} = g u_{m+1} + \sigma^o_{m+1}.
\end{aligned}$$

$$\frac{du}{dt} = (-\gamma + i\omega_0)u + f_0 + f_1 e^{i\omega_1 t} + \sigma \dot{W} \qquad \Longrightarrow \qquad u_{m+1} = F u_m + \mathcal{F}_{m+1} + \sigma_{m+1}$$

The exact solution can be written down explicitly,

$$u(t) = u(t_0)e^{(-\gamma+i\omega_0)(t-t_0)} + \frac{f_0}{\gamma - i\omega_0}\left(1 - e^{(-\gamma+i\omega_0)(t-t_0)}\right)$$
$$+ \frac{f_1 e^{i\omega_1 t}}{\gamma + i(-\omega_0 + \omega_1)}\left(1 - e^{-(\gamma+i\omega_1-i\omega_0)(t-t_0)}\right) + \sigma \int_{t_0}^{t} e^{(-\gamma+i\omega_0)(t-s)} dW(s),$$

which provides the analytical forms of the time evolution of the forecast mean $\overline{u}(t)$ and the forecast variance $r(t)$,

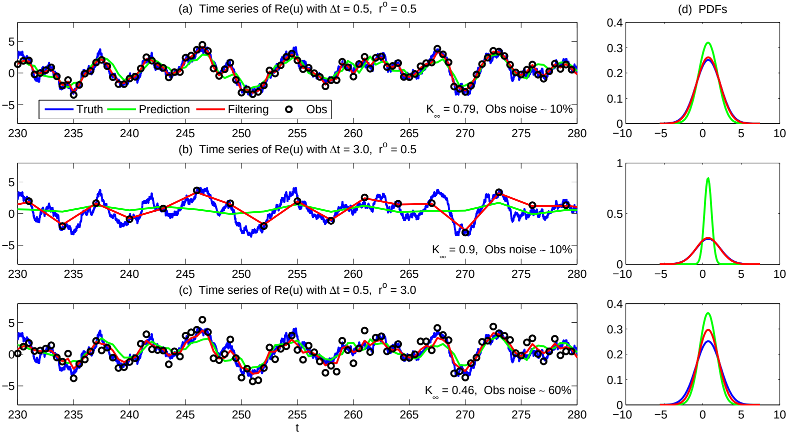
$$\overline{u}(t) = \overline{u}(t_0)e^{(-\gamma+i\omega_0)(t-t_0)} + \frac{f_0}{\gamma - i\omega_0}\left(1 - e^{(-\gamma+i\omega_0)(t-t_0)}\right)$$
$$+ \frac{f_1 e^{i\omega_1 t}}{\gamma + i(-\omega_0 + \omega_1)}\left(1 - e^{-(\gamma+i\omega_1-i\omega_0)(t-t_0)}\right),$$
$$r(t) = r(t_0)e^{-2\gamma(t-t_0)} + \frac{\sigma^2}{2\gamma}\left(1 - e^{-2\gamma(t-t_0)}\right).$$

The test model

$$\frac{du}{dt} = (-\gamma + i\omega_0)u + f_0 + f_1 e^{i\omega_1 t} + \sigma \dot{W}$$

with parameters

$$\gamma = 0.4, \quad \omega_0 = 1, \quad f_0 = 2, \quad f_1 = 0, \quad \omega_1 = 0, \quad \sigma = 1, \quad g = 1.$$



The N-dimensional Kalman filter

Our goal is to obtain the posterior distribution $p(\vec{u}|\vec{v}_{m+1})$ through the Bayesian formula assuming a knowledge of the prior distribution $p(\vec{u})$ of a true signal $\vec{u}_{m+1} \in \mathbb{R}^N$ and the observed value $\vec{v}_{m+1} \in \mathbb{R}^M$.

The model is

$$\vec{u}_{m+1|m} = F\vec{u}_{m|m} + \vec{\sigma}_{m+1}, \qquad R = \langle \vec{\sigma}_m \otimes \vec{\sigma}_m^T \rangle.$$

The observational process is

$$\vec{v}_{m+1} = G\vec{u}_{m+1} + \vec{\sigma}^o_{m+1}, \qquad R^o = \langle \vec{\sigma}^o_m \otimes (\vec{\sigma}^o_m)^T \rangle.$$

The N-dimensional Kalman filter

Our goal is to obtain the posterior distribution $p(\vec{u}|\vec{v}_{m+1})$ through the Bayesian formula assuming a knowledge of the prior distribution $p(\vec{u})$ of a true signal $\vec{u}_{m+1} \in \mathbb{R}^N$ and the observed value $\vec{v}_{m+1} \in \mathbb{R}^M$.

The model is

$$\vec{u}_{m+1|m} = F\vec{u}_{m|m} + \vec{\sigma}_{m+1}, \qquad R = \langle \vec{\sigma}_m \otimes \vec{\sigma}_m^T \rangle.$$

The observational process is

$$\vec{v}_{m+1} = G\vec{u}_{m+1} + \vec{\sigma}^o_{m+1}, \qquad R^o = \langle \vec{\sigma}^o_m \otimes (\vec{\sigma}^o_m)^T \rangle.$$

The posterior mean and posterior covariance are given by

$$\begin{aligned}
\bar{\vec{u}}_{m+1|m+1} &= \bar{\vec{u}}_{m+1|m+1} + K_{m+1}(\vec{v}_{m+1} - G\bar{\vec{u}}_{m+1|m}), \\
R_{m+1|m+1} &= (I - K_{m+1}G)R_{m+1|m}
\end{aligned}$$

where the Kalman gain is

$$K_{m+1} = R_{m+1|m}G^T(GR_{m+1|m}G^T + R^o)^{-1}.$$

Time Evolution of the Posterior Statistics

In general, the Kalman filter system is given by

$$\mathbf{u_I}(t_{i+1}) = \mathbf{G}\mathbf{u_{II}}(t_{i+1}) + \mathbf{B_1}\boldsymbol{\varepsilon}_\mathbf{1},$$
$$\mathbf{u_{II}}(t_{i+1}) = \mathbf{a_0} + \mathbf{a_1}\mathbf{u_{II}}(t_i) + \mathbf{b_2}\boldsymbol{\varepsilon}_\mathbf{2},$$

Define $\mathbf{A_0} = \mathbf{G}\mathbf{a_0}, \mathbf{A_1} = \mathbf{G}\mathbf{a_1}$ and $\mathbf{B_2} = \mathbf{G}\mathbf{b_2}$. Then the above system becomes

$$\mathbf{u_I}(t_{i+1}) = \mathbf{A_0} + \mathbf{A_1}\mathbf{u_{II}}(t_i) + \mathbf{B_1}\boldsymbol{\varepsilon}_\mathbf{1} + \mathbf{B_2}\boldsymbol{\varepsilon}_\mathbf{2},$$
$$\mathbf{u_{II}}(t_{i+1}) = \mathbf{a_0} + \mathbf{a_1}\mathbf{u_{II}}(t_i) + \mathbf{b_2}\boldsymbol{\varepsilon}_\mathbf{2},$$

The time evolutions of the posterior mean and posterior covariance are given by

$$\begin{aligned}\boldsymbol{\mu}(t_{i+1}) =& \mathbf{a_0} + \mathbf{a_1}\boldsymbol{\mu}(t_i) + (\mathbf{b_2}\mathbf{B}_\mathbf{2}^* + \mathbf{a_1}\mathbf{R}(t_i)\mathbf{A}_\mathbf{1}^*)\times \\ & (\mathbf{B_1}\mathbf{B}_\mathbf{1}^* + \mathbf{B_2}\mathbf{B}_\mathbf{2}^* + \mathbf{A_1}\mathbf{R}(t_i)\mathbf{A}_\mathbf{1}^*)^{-1}(\mathbf{u_I}(t_{i+1}) - \mathbf{A_0} - \mathbf{A_1}\boldsymbol{\mu}(t_i)), \\ \mathbf{R}(t_{i+1}) =& \mathbf{a_1}\mathbf{R}(t_i)\mathbf{a}_\mathbf{1}^* + \mathbf{b_2}\mathbf{b}_\mathbf{2}^* - (\mathbf{b_2}\mathbf{B}_\mathbf{2}^* + \mathbf{a_1}\mathbf{R}(t_i)\mathbf{A}_\mathbf{1}^*)\times \\ & (\mathbf{B_1}\mathbf{B}_\mathbf{1}^* + \mathbf{B_2}\mathbf{B}_\mathbf{2}^* + \mathbf{A_1}\mathbf{R}(t_i)\mathbf{A}_\mathbf{1}^*)^{-1}(\mathbf{b_2}\mathbf{B}_\mathbf{2}^* + \mathbf{a_1}\mathbf{R}(t_i)\mathbf{A}_\mathbf{1}^*)^*.\end{aligned}$$

The equation of **R** is called the Riccati equation.

What's new in N-dimensional Kalman filter?

Case 1. Model:

$$u_{m+1}^1 = F_{11} u_m^1 + F_{12} u_m^2 + \sigma_m^1$$
$$u_{m+1}^2 = F_{21} u_m^1 + F_{22} u_m^2 + \sigma_m^2$$

Observation:

$$v_{m+1}^1 = g_1 u_{m+1}^1 + \sigma_m^{o,1}$$
$$v_{m+1}^2 = g_2 u_{m+1}^2 + \sigma_m^{o,2}$$

Case 2. Model:

$$u_{m+1}^1 = F_{11} u_m^1 + F_{12} u_m^2 + \sigma_m^1$$
$$u_{m+1}^2 = F_{21} u_m^1 + F_{22} u_m^2 + \sigma_m^2$$

Observation:

$$v_{m+1}^1 = g_1 u_{m+1}^1 + \sigma_m^{o,1}$$

Case 3. Model:

$$u_{m+1}^1 = F_{11} u_m^1 + F_{12} u_m^2 + \sigma_m^1$$
$$u_{m+1}^2 = F_{21} u_m^1 + F_{22} u_m^2 + \sigma_m^2$$

Observation:

$$v_{m+1}^1 = g_1 u_{m+1}^1 + g_2 u_{m+1}^2 + \sigma_m^{o,1}$$

Can we filter u_{m+1}^1? What about filtering u_{m+1}^2?

What's new in N-dimensional Kalman filter?

Case 4. Model:

$$
\begin{aligned}
u^1_{m+1} &= F_{11} u^1_m \qquad\qquad + \sigma^1_m \\
u^2_{m+1} &= F_{21} u^1_m + F_{22} u^2_m + \sigma^2_m
\end{aligned}
$$

Observation:

$$v^1_{m+1} = g_1 u^1_{m+1} + \sigma^{o,1}_m$$

Case 5. Model:

$$
\begin{aligned}
u^1_{m+1} &= F_{11} u^1_m + F_{12} u^2_m + \sigma^1_m \\
u^2_{m+1} &= \qquad\qquad F_{22} u^2_m + \sigma^2_m
\end{aligned}
$$

Observation:

$$v^1_{m+1} = g_1 u^1_{m+1} + \sigma^{o,1}_m$$

Case 6. Model:

$$
\begin{aligned}
u^1_{m+1} &= F_{11} u^1_m \qquad\qquad + \sigma^1_m \\
u^2_{m+1} &= \qquad\qquad F_{22} u^2_m + \sigma^2_m
\end{aligned}
$$

Observation:

$$v^1_{m+1} = g_1 u^1_{m+1} + \sigma^{o,1}_m$$

Can we filter u^1_{m+1}? What about filtering u^2_{m+1}?

In addition to state estimation, what are the other direct applications of data assimilation?

In addition to state estimation, what are the other direct applications of data assimilation?

- parameter estimation

In addition to state estimation, what are the other direct applications of data assimilation?

- parameter estimation
- dynamical interpolation

In addition to state estimation, what are the other direct applications of data assimilation?

- parameter estimation
- dynamical interpolation
- improving the training data for machine learning forecast

In addition to state estimation, what are the other direct applications of data assimilation?

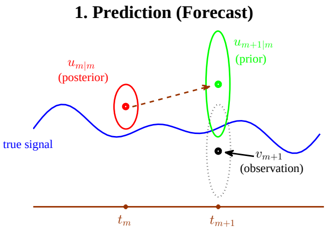
- parameter estimation
- dynamical interpolation
- improving the training data for machine learning forecast
- ...

Comments on nonlinear filters

Recall the classical Kalman filter (in 1D case):

$$\text{Model:} \qquad u_{m+1} = Fu_m + \mathcal{F}_{m+1} + \sigma_{m+1},$$
$$\text{Obs:} \qquad v_{m+1} = gu_{m+1} + \sigma^o_{m+1}.$$

- σ_{m+1} has zero mean and variance $r = \langle \sigma_{m+1}\sigma^*_{m+1}\rangle = \frac{1}{2}\sum_{j=1}^2 \langle \sigma^2_{j,m+1}\rangle$.
- σ^o_m has zero mean and variance $r^o = \langle \sigma^o_m(\sigma^o_m)^*\rangle$.



The posterior mean and posterior variance are given by

$$\bar{u}_{m+1|m+1} = (1 - K_{m+1}g)\bar{u}_{m+1|m} + K_{m+1}v_{m+1},$$
$$r_{m+1|m+1} = (1 - K_{m+1}g)r_{m+1|m} \qquad \to \text{a constant as } m \to \infty.$$

where K_{m+1} is the Kalman gain

$$K_{m+1} = \frac{g r_{m+1|m}}{r^o + g^2 r_{m+1|m}} \qquad \to \text{a constant as } m \to \infty$$

The classical Kalman filter:

$$\begin{aligned} \text{Model:} \qquad u_{m+1} &= Fu_m + \mathcal{F}_{m+1} + \sigma_{m+1}, \\ \text{Obs:} \qquad v_{m+1} &= gu_{m+1} + \sigma^o_{m+1}. \end{aligned}$$

- The Kalman filter is optimal when the underlying dynamics is linear and noise is Gaussian.
- Since the dynamics is linear and Gaussian, the time evolution of the mean and variance can be written down explicitly.
- Dynamical systems in real applications are usually of dimension $O(10^6)$–$O(10^9)$.

The classical Kalman filter:

$$\begin{aligned} \text{Model:} \qquad u_{m+1} &= F u_m + \mathcal{F}_{m+1} + \sigma_{m+1}, \\ \text{Obs:} \qquad v_{m+1} &= g u_{m+1} + \sigma^o_{m+1}. \end{aligned}$$

- The Kalman filter is optimal when the underlying dynamics is linear and noise is Gaussian.
- Since the dynamics is linear and Gaussian, the time evolution of the mean and variance can be written down explicitly.
- Dynamical systems in real applications are usually of dimension $O(10^6)$–$O(10^9)$.

Then what's the problems for the Kalman filter in real applications?

Practical issues in the Kalman filter with high-dimensional nonlinear dynamics (and/or non-Gaussian observations):

1. There is in general no analytic solutions in the forecast stage (nonlinear model).

$$\vec{u}_{m+1} = \vec{F}(\vec{u}_m) + \vec{\sigma}_{m+1}.$$

2. The posterior states have no analytic solution either (non-Gaussian statistics).

$$p(\vec{u}_{m+1}|\vec{v}_{m+1}) \sim p(\vec{u}_{m+1})p(\vec{v}_{m+1}|\vec{u}_{m+1}).$$

3. It is impossible to use numerical methods to solve either the prior or the posterior solutions (high dimensionality).

What if we still apply the Kalman filter to nonlinear problems with non-Gaussian statistics?

What if we still apply the Kalman filter to nonlinear problems with non-Gaussian statistics?

Then we need to linearize the nonlinear models.

- Dynamical properties are completely changed.
- Statistics beyond mean and covariance are ignored.

Some alternative methods are presented in the next lecture.