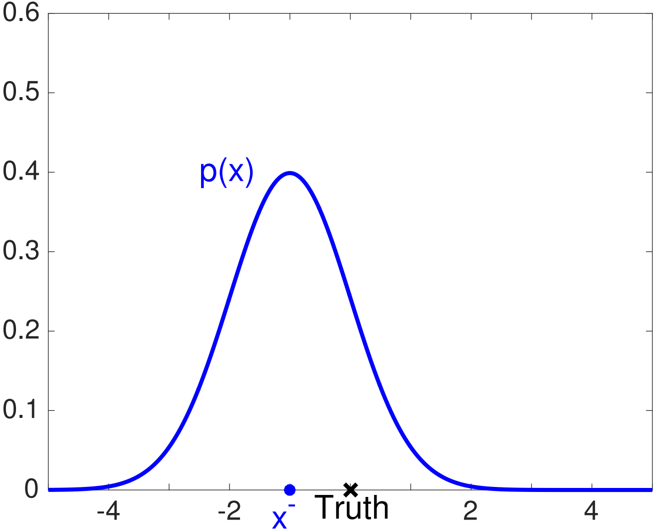
Introduction to Data Assimilation and Kalman Filtering

Tyrus Berry
Dept. of Mathematical Sciences
George Mason University

July 28, 2018

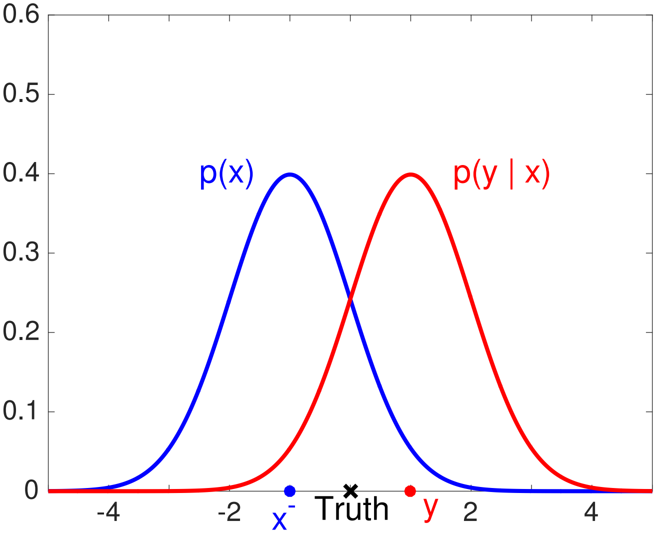
PRIOR

- We start out uncertain where *x* is



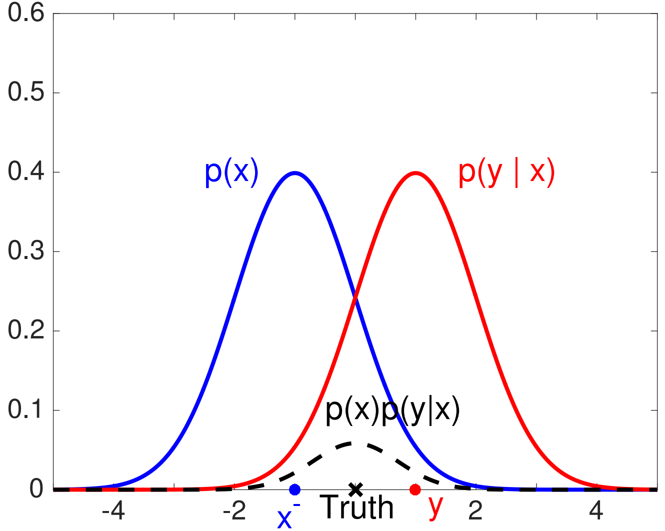
LIKELIHOOD

- Then we observe $y = x + noise$



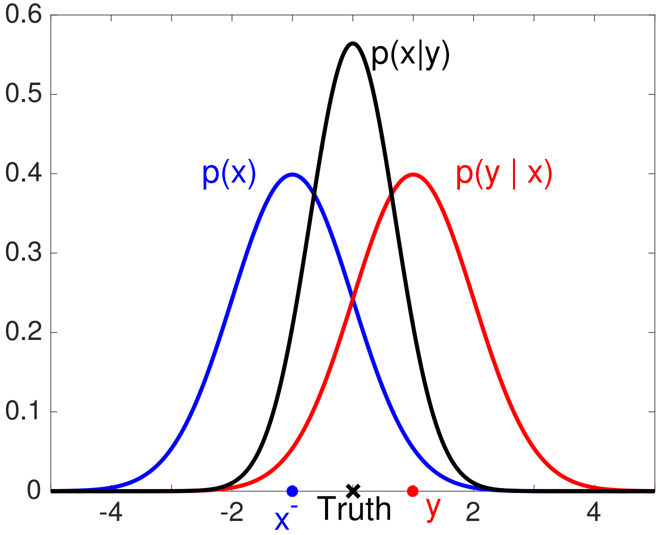
Bayes' Law

- Combine info by multiplying $p(x)p(y \mid x)$



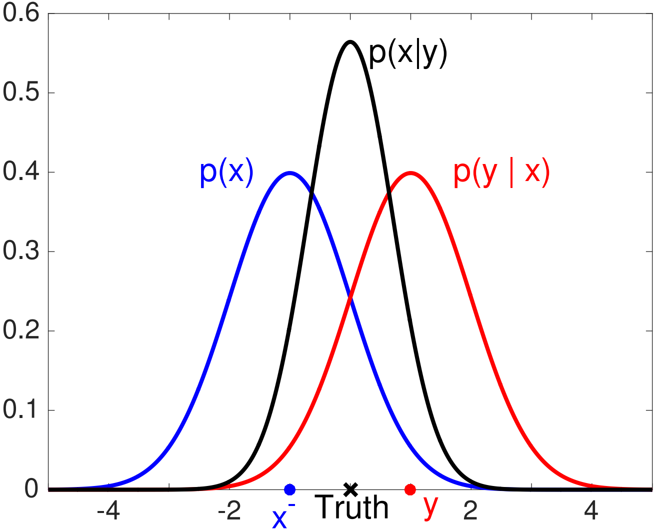
POSTERIOR

- Renormalize to get the 'posterior' distribution $p(x \mid y)$



Assimilating Information

- Notice: We have reduced our uncertainty!



ASSIMILATING INFORMATION

- We start out uncertain where x is
- Assume x has a Gaussian distribution, $x \sim \mathcal{N}(x^-, \sigma^2)$

$$p(x) = \frac{e^{-\frac{(x-x^-)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$$

Assimilating Information

- Now we observe $y = x + \nu$, where ν is noise
- Assume ν has a Gaussian distribution, $\nu \sim \mathcal{N}(0, r^2)$

$$p(\nu) = \frac{e^{\frac{-\nu^2}{2r^2}}}{\sqrt{2\pi r^2}}$$

- Since $y = x + \nu$, we have $p(y \mid x) = \frac{e^{-\frac{(y-x)^2}{2r^2}}}{\sqrt{2\pi r^2}}$
- $p(y \mid x)$ is called a likelihood function for x

Assimilating Information

- Combine with Bayes' Law: $p(x \mid y) \propto p(x)p(y \mid x)$

$$p(x \mid y) \propto p(x)p(y \mid x) = \frac{e^{-\frac{(x-x^-)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} \frac{e^{-\frac{(y-x)^2}{2r^2}}}{\sqrt{2\pi r^2}} \propto e^{-\frac{(x-x^-)^2}{2\sigma^2} - \frac{(y-x)^2}{2r^2}}$$

- Complete the square: $-\left(\frac{1}{2\sigma^2} + \frac{1}{2r^2}\right) x^2 + \left(\frac{x^-}{\sigma^2} + \frac{y}{r^2}\right) x + c$
- Variance: $\sigma_+^2 = \left(\frac{1}{\sigma^2} + \frac{1}{r^2}\right)^{-1}$
- Mean: $x^+ = \sigma_+^2 \left(\frac{x^-}{\sigma^2} + \frac{y}{r^2}\right)$
- After assimilation: $p(x \mid y) = \frac{e^{-\frac{(x-x^+)^2}{2\sigma_+^2}}}{\sqrt{2\pi\sigma_+^2}}$

Assimilating Information: Example

- Prior: $x^- = -1, \sigma^2 = 1$

$$p(x) = \frac{e^{-(x+1)^2/2}}{\sqrt{2\pi}}$$

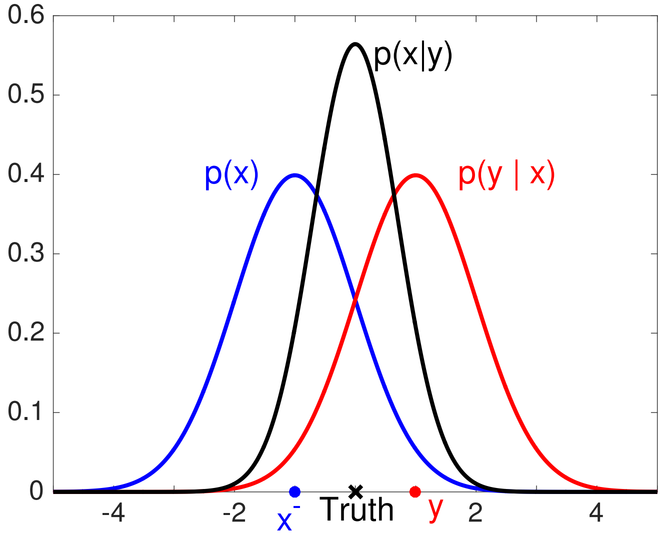
- Likelihood: $y = 1, s^2 = 1$

$$p(y \mid x) = \frac{e^{-(x-1)^2/2}}{\sqrt{2\pi}}$$

- Posterior: $\sigma_+^2 = \left(\frac{1}{\sigma^2} + \frac{1}{r^2}\right)^{-1} = \frac{1}{2}$, $x^+ = \sigma_+^2 \left(\frac{x^-}{\sigma^2} + \frac{y}{r^2}\right) = 0$

$$p(x \mid y) = \frac{e^{-x^2}}{\sqrt{\pi}}$$

Assimilating Information: Example



Assimilating Information: The 'update'

- Variance: $\sigma_+^2 = \left(\frac{1}{\sigma^2} + \frac{1}{r^2}\right)^{-1} = \sigma^2(r^2 + \sigma^2)^{-1} r^2$
- Mean:

$$x^+ = \sigma_+^2 \left(\frac{x^-}{\sigma^2} + \frac{y}{r^2}\right) = \sigma_+^2 \left(\frac{x^-}{\sigma^2} + \frac{x^-}{r^2} - \frac{x^-}{r^2} + \frac{y}{r^2}\right)$$
$$= \sigma_+^2 \left(\left(\frac{1}{\sigma^2} + \frac{1}{r^2}\right) x^- + \frac{y - x^-}{r^2}\right)$$
$$= x^- + \frac{\sigma_+^2}{r^2}(y - x^-)$$

- Define the **Kalman gain**: $K = \sigma_+^2 r^{-2} = \sigma^2(r^2 + \sigma^2)^{-1}$
- Variance update: $\sigma_+^2 = K(r^2 + \sigma^2) - K\sigma^2 = \sigma^2 - K\sigma^2$
- Mean update: $x^+ = x^- + K(y - x^-)$

Assimilating Information: The 'update'

- **Kalman gain**: $K = \sigma^2(r^2 + \sigma^2)^{-1}$
- Variance update: $\sigma_+^2 = (1 - K)\sigma^2$
- Mean update: $x^+ = x^- + K(y - x^-)$

Multivariable Update

- Prior: Mean vector x^-, covariance matrix P^-

$$p(x) \propto \exp(-(x - x^-)^\top (P^-)^{-1} (x - x^-))$$

- Observation Noise: Mean vector 0, covariance matrix R

$$p(y \mid x) \propto \exp(-(x - y)^\top R^{-1} (x - y))$$

- **Kalman gain**: $K = P^-(R + P^-)^{-1}$
- Variance update: $P^+ = (I - K)P^-$
- Mean update: $x^+ = x^- + K(y - x^-)$

Linear Observations

- Instead of observing x directly, we observe $y = Hx + \nu$

$$p(y \,|\, x) \propto \exp(-(Hx - y)^\top R^{-1}(Hx - y))$$

- Distribution of $Hx \sim \mathcal{N}(Hx^-, HP^-H^\top)$

- **Kalman gain**: $K = P^-H^\top(R + HP^-H^\top)^{-1}$
- Variance update: $P^+ = (I - KH)P^-$
- Mean update: $x^+ = x^- + K(y - Hx^-)$

SUMMARY SO FAR...

- We start with Gaussian information about $x \sim \mathcal{N}(x^-, P^-)$
- We make a noisy observation $y = Hx + \nu$, $\nu \sim \mathcal{N}(0, R)$
- Assimilate y to form the posterior $p(x \mid y) \sim \mathcal{N}(x^+, P^+)$

- **Kalman gain**: $K = P^- H^\top (R + HP^- H^\top)^{-1}$
- Variance update: $P^+ = (I - KH)P^-$
- Mean update: $x^+ = x^- + K(y - Hx^-)$

WHY DOES IT WORK...

- Linear combinations of Gaussians are Gaussian
- Bayesian combinations of Gaussians are Gaussian
- Now we can continue to assimilate more observations
- Each observation improves our estimate!

- Next step: What if x evolves before we make the next observation?

Kalman Filter

- Assume linear dynamics/obs and additive Gaussian noise

$$
\begin{aligned}
x_k &= Fx_{k-1} + \omega_k & \omega_k \sim \mathcal{N}(0, Q) \\
y_k &= Hx_k + \nu_k & \nu_k \sim \mathcal{N}(0, R)
\end{aligned}
$$

- Just like before, except now x is changing
- $x_k^- = Fx_{k-1}^+$
- $P_k^- = FP_{k-1}^+F^\top + Q$
- Notice that Q increases the uncertainty of x

Kalman Filtering: An Intuitive Idea

Filter tracks two things

1. Estimate of state x over time
2. Uncertainty of state estimate, covariance matrix P

This is accomplished using a predictor-corrector methodology at each observation time k

1. **Predict** an estimate of state (x_k^-) and covariance (P_k^-)
2. **Observe** data y_k
3. **Correct** state estimate (x_k^+) and covariance (P_k^+)

Kalman Filtering: An Intuitive Idea

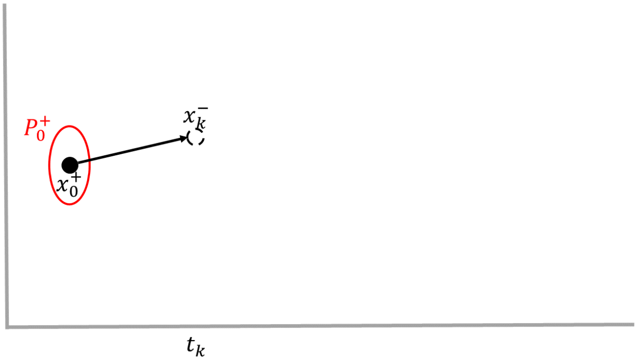
Initialize!

Kalman Filtering: An Intuitive Idea

Initialize!

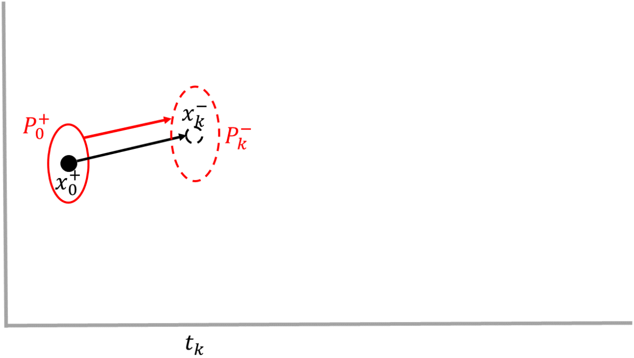
Kalman Filtering: An Intuitive Idea

Predict!



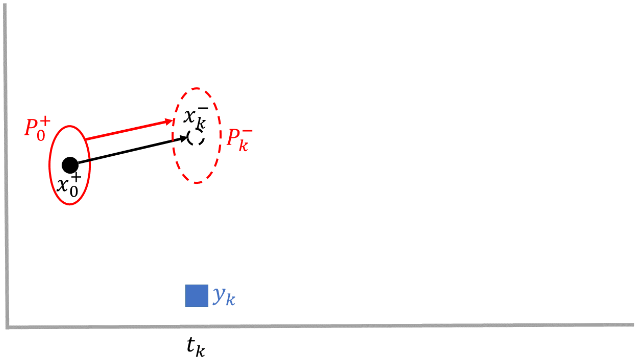
Kalman Filtering: An Intuitive Idea

Predict!



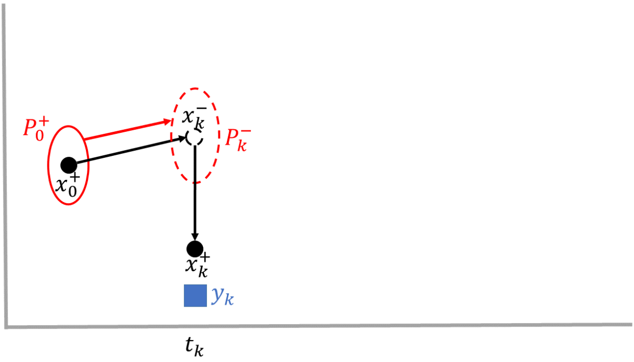
KALMAN FILTERING: AN INTUITIVE IDEA

Observe!



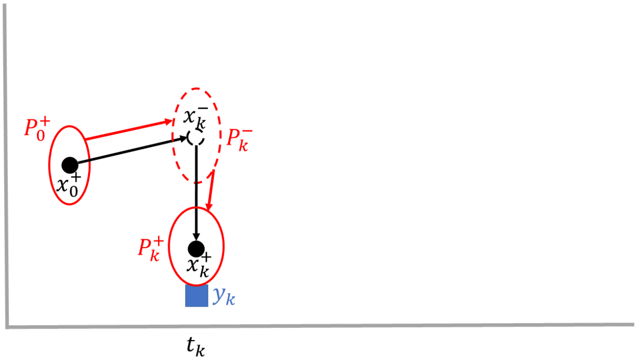
Kalman Filtering: An Intuitive Idea

Correct!



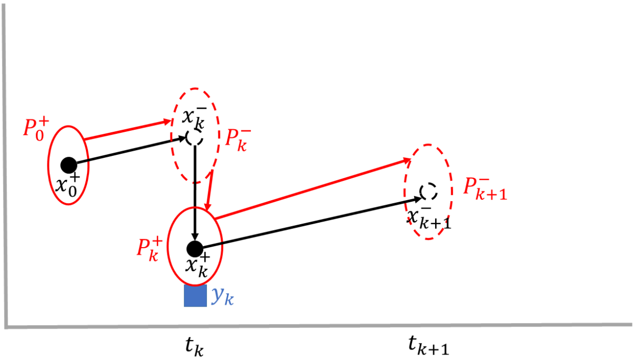
KALMAN FILTERING: AN INTUITIVE IDEA

Correct!



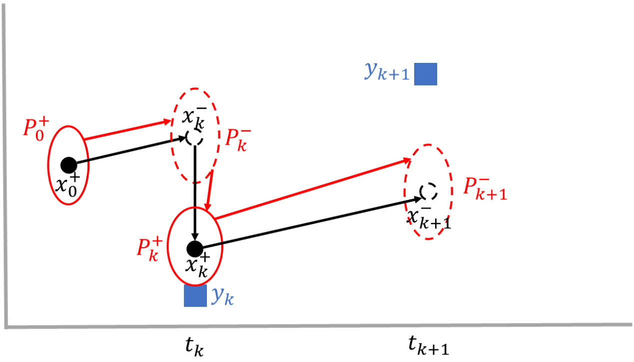
Kalman Filtering: An Intuitive Idea

Predict!



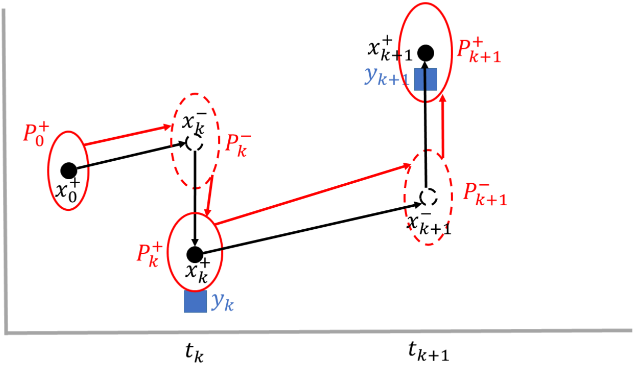
Kalman Filtering: An Intuitive Idea

Observe!

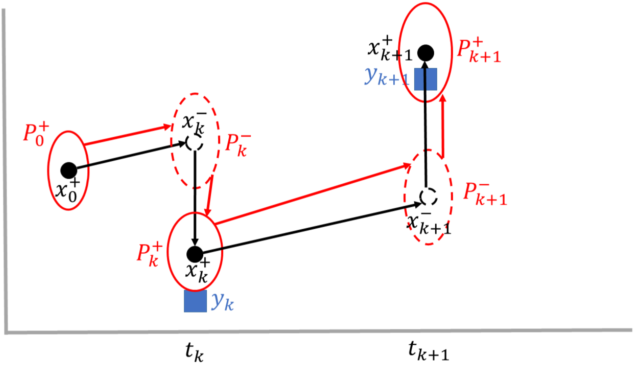


Kalman Filtering: An Intuitive Idea

Correct!



KALMAN FILTERING: AN INTUITIVE IDEA



And so on, and so on, and so on...

Kalman Filter Summary

Forecast Step

$$
\begin{aligned}
x_k^- &= F x_{k-1}^+ \\
P_k^- &= F P_{k-1}^+ F^T + Q \\
P_k^y &= H P_k^- H^T + R
\end{aligned}
$$

Assimilation Step

$$
\begin{aligned}
K_k &= P_k^- H^T (P_k^y)^{-1} \\
P_k^+ &= (I - K_k H) P_k^- \\
x_k^+ &= x_k^- + K_k (y_k - H x_k^-)
\end{aligned}
$$

NON-AUTONOMOUS KALMAN FILTER

- Assume linear dynamics/obs and additive Gaussian noise

$$
\begin{aligned}
x_k &= F_{k-1}x_{k-1} + \omega_k & \omega_k \sim \mathcal{N}(0, Q) \\
y_k &= H_k x_k + \nu_k & \nu_k \sim \mathcal{N}(0, R)
\end{aligned}
$$

- Just like before, except now F, H are changing

Non-Autonomous Kalman Filter

Forecast Step
$$\begin{cases} x_k^- = F_{k-1} x_{k-1}^+ \\ P_k^- = F_{k-1} P_{k-1}^+ F_{k-1}^T + Q \\ P_k^y = H_k P_k^- H_k^T + R \end{cases}$$

Assimilation Step
$$\begin{cases} K_k = P_k^- H_k^T (P_k^y)^{-1} \\ P_k^+ = (I - K_k H_k) P_k^- \\ x_k^+ = x_k^- + K_k (y_k - H_k x_k^-) \end{cases}$$

Nonlinear Kalman Filtering

- Consider a discrete time dynamical system:

$$\begin{aligned} x_k &= f(x_{k-1}, \omega_k) \\ y_k &= h(x_k, \nu_k) \end{aligned}$$

- We will convert this to a linear non-autonomous system
- Two methods:
- Extended Kalman Filter (EKF)
- Ensemble Kalman Filter (EnKF)

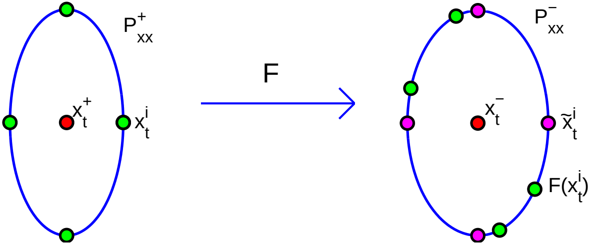
EXTENDED KALMAN FILTER (LINEARIZE DYNAMICS)

- Consider a system of the form:

$$\begin{aligned} x_{k+1} &= f(x_k) + \omega_{k+1} & \omega_{k+1} \sim \mathcal{N}(0, Q) \\ y_{k+1} &= h(x_{k+1}) + \nu_{k+1} & \nu_{k+1} \sim \mathcal{N}(0, R) \end{aligned}$$

- **Extended Kalman Filter (EKF)**:
  - Linearize $F_{k-1} = Df(x_{k-1}^+)$ and $H_k = Dh(x_k^-)$
- Problem: State estimate x_{k-1}^+ may not be well localized
- Solution: Ensemble Kalman Filter (EnKF)

Ensemble Kalman Filter (EnKF)



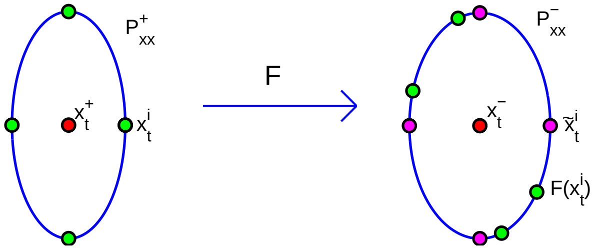
Generate an ensemble with the current statistics (use matrix square root):

$$x_k^i = \text{"sigma points" on semimajor axes}$$

$$x_k^- = \frac{1}{2n}\sum f(x_k^i)$$

$$P_k^- = \frac{1}{2n-1}\sum (f(x_k^i) - x_k^-)(f(x_k^i) - x_k^-)^T + Q$$

Ensemble Kalman Filter (EnKF)



Calculate $y_k^i = H(F(x_k^i))$. Set $y_k^- = \frac{1}{2n}\sum_i y_k^i$ (note: $t = k$)

$$
\begin{aligned}
P_k^y &= \frac{1}{2n-1}\sum_i (y_k^i - y_k^-)(y_k^i - y_k^-)^T + R \\
P_k^{xy} &= \frac{1}{2n-1}\sum_i (f(x_k^i) - x_k^-)(y_k^i - y_k^-)^T \\
K_k &= P_k^{xy}(P_k^y)^{-1} \text{ and } P_k^+ = P_k^- - KP_k^y K^T \\
x_{k+1}^+ &= x_k^- + K_k(y_k - y_k^-)
\end{aligned}
$$

Parameter Estimation

- When the model has parameters p,

$$x_{k+1} = f(x_k, p) + \omega_{k+1}$$

- Can *augment* the state $\tilde{x}_k = [x_k, p_k]$
- Introduce trivial dynamics for p

$$x_{k+1} = f(x_k, p_k) + \omega_{k+1}$$
$$p_{k+1} = p_k + \omega^p_{k+1}$$

- Need to tune the covariance of ω^p_{k+1}

EXAMPLE OF PARAMETER ESTIMATION

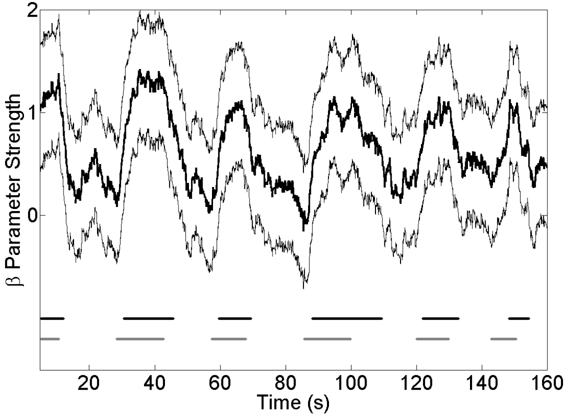
Consider the Hodgkin-Huxley neuron model, expanded to a network of n equations

$$
\begin{aligned}
\dot{V}_i &= -g_{Na} m^3 h(V_i - E_{Na}) - g_K n^4 (V_i - E_K) - g_L (V_i - E_L) \\
&\quad + I + \sum_{j \neq i}^{n} \Gamma_{\mathrm{HH}}(V_j) V_j \\
\dot{m}_i &= a_m(V_i)(1 - m_i) - b_m(V_i) m_i \\
\dot{h}_i &= a_h(V_i)(1 - h_i) - b_h(V_i) h_i \\
\dot{n}_i &= a_n(V_i)(1 - n_i) - b_n(V_i) n_i \\
\Gamma_{\mathrm{HH}}(V_j) &= \beta_{ij} / (1 + e^{-10(V_j + 40)})
\end{aligned}
$$

Only observe the voltages V_i, recover the hidden variables and the connection parameters β

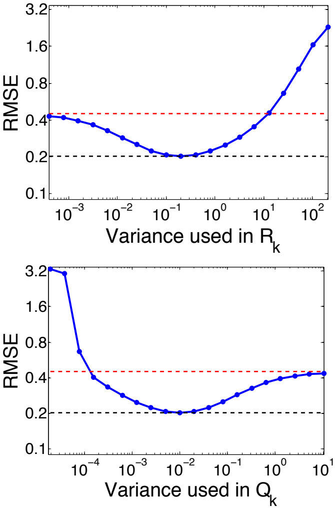
EXAMPLE OF PARAMETER ESTIMATION

Can even turn connections on and off (grey dashes)
Variance estimate ⇒ statistical test (black dashes)



NONLINEAR KALMAN-TYPE FILTER: INFLUENCE OF Q AND R

- Simple example with full observation and diagonal noise covariances
- Red indicates RMSE of unfiltered observations
- Black is RMSE of 'optimal' filter (true covariances known)



Further Topics

- **Inflation:** Improving stability by artificially increasing P^-, Q

  Jeff Anderson, An adaptive covariance inflation error correction algorithm for ensemble filters. Tellus A.

- **Adaptive Filtering:** Automatically learning/tuning Q, R

  R. Mehra, On the identification of variances and adaptive Kalman filtering. IEEE Trans. Auto. Cont.

  T. Berry, T. Sauer, Adaptive ensemble Kalman filtering of nonlinear systems. Tellus A.

- **Localization:** High-dimensional problems (eg. LETKF)

  B. Hunt, E. Kostelich, I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. Physica D.

  E. Ott, et al. 2004: A local ensemble Kalman filter for atmospheric data assimilation. Tellus A.

- **Efficiency:** Avoiding large matrix inverses (EAKF)

  Jeff Anderson, An Ensemble Adjustment Kalman Filter for Data Assimilation. Monthly Weather Review.

What is the filtering problem?

- Consider a discrete time dynamical system:

$$\begin{aligned} x_k &= f_k(x_{k-1}, \omega_k) \\ y_k &= h_k(x_k, \nu_k) \end{aligned}$$

- Given the observations $y_1, ..., y_k$ we define three problems:

- **Filtering:** Estimate the current state $p(x_k \,|\, y_1, ..., y_k)$
- **Forecasting:** Estimate a future state $p(x_{k+\ell} \,|\, y_1, ..., y_k)$
- **Smoothing:** Estimate a past state $p(x_{k-\ell} \,|\, y_1, ..., y_k)$

Two Step Filtering to Find $p(x_k \mid y_1, ..., y_k)$

- Assume we have $p(x_{k-1} \mid y_1, ..., y_{k-1})$
- **Forecast Step:** Find $p(x_k \mid y_1, ..., y_{k-1})$
- **Assimilation Step:** Perform a Bayesian update,

$$p(x_k \mid y_1, ..., y_k) \propto p(x_k \mid y_1, ..., y_{k-1}) p(y_k \mid x_k, y_1, ..., y_{k-1})$$

Posterior $\propto$ **Prior** $\times$ **Likelihood**

Alternative: Variational Filtering

- Given observations $y_1, ..., y_k$ write an error function:

$$J(x_1, ..., x_k) = \sum_{i=1}^{T-1} ||x_{i+1} - f(x_i)||_Q^2 + \sum_{i=1}^{T} ||y_i - h(x_i)||_R^2$$

- When noise is Gaussian $-J$ is the log-likelihood function
- When dynamics/obs are linear, Kalman filter provably minimizes J (maximal likelihood)
- Variational filtering explicitly minimizes J, often with Newton-type methods

Papers with Franz Hamilton and Tim Sauer

http://math.gmu.edu/˜berry/

- Ensemble Kalman filtering without a model. Phys. Rev. X (2016).
- Adaptive ensemble Kalman filtering of nonlinear systems. Tellus A (2013).
- Real-time tracking of neuronal network structure using data assimilation. Phys. Rev. E (2013).

Related/Background Material

- R. Mehra, 1970: On the identification of variances and adaptive Kalman filtering.
- P. R. Bélanger, 1974: Estimation of noise covariance matrices for a linear time-varying stochastic process.
- J. Anderson, 2007: An adaptive covariance inflation error correction algorithm for ensemble filters.
- H. Li, E. Kalnay, T. Miyoshi, 2009: Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter.
- B. Hunt, E. Kostelich, I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter.
- E. Ott, et al. 2004: A local ensemble Kalman filter for atmospheric data assimilation.