How do we solve it and what does the solution look like?

KF/PFs offer solutions to dynamical systems, nonlinear in general, using prediction and update as data becomes available. Tracking in time or space offers an ideal framework for studying KF/PF.

Consider the discrete, linear system,

$$x_{k+1} = M_k x_k + w_k, \ k = 0, 1, 2, \ldots$$

where

- $x_k \in \mathbb{R}^n$ is the state vector at time $t_k$
- $M_k \in \mathbb{R}^{n \times n}$ is the state transition matrix (mapping from time $t_k$ to $t_{k+1}$) or model
- $\{w_k \in \mathbb{R}^n; k = 0, 1, 2, \ldots\}$ is a white, Gaussian sequence, with $w_k \sim N(0, Q_k)$, often referred to as model error
- $Q_k \in \mathbb{R}^{n \times n}$ is a symmetric positive definite covariance matrix (known as the model error covariance matrix).
The Observations

We also have discrete, linear observations that satisfy
\[ y_k = H_k x_k + v_k, \quad k = 1, 2, 3, \ldots, \]   \hspace{1cm} (2)

where
- \( y_k \in \mathbb{R}^p \) is the vector of actual measurements or observations at time \( t_k \)
- \( H_k \in \mathbb{R}^{n \times p} \) is the observation operator. Note that this is not in general a square matrix.
- \( \{v_k \in \mathbb{R}^p, k = 1, 2, \ldots\} \) is a white, Gaussian sequence, with \( v_k \sim N(0, R_k) \), often referred to as observation error.
- \( R_k \in \mathbb{R}^{p \times p} \) is a symmetric positive definite covariance matrix (known as the observation error covariance matrix).

We assume that the initial state, \( x_0 \) and the noise vectors at each step, \( \{w_k\}, \{v_k\} \), are assumed mutually independent.

Summary of the Kalman filter

**Prediction step**

Mean update:
\[ \hat{x}_{k+1|k} = M_k \hat{x}_{k|k} \]

Covariance update:
\[ P_{k+1|k} = M_k P_{k|k} M_k^T + Q_k. \]

**Observation update step**

Mean update:
\[ \hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k (y_k - H_k \hat{x}_{k|k-1}) \]

Kalman gain:
\[ K_k = P_{k|k-1} H_k^T (H_k P_{k|k-1} H_k^T + R_k)^{-1} \]

Covariance update:
\[ P_{k|k} = (I - K_k H_k) P_{k|k-1}. \]
Prediction step

We first derive the equation for one-step prediction of the mean using the state propagation model (1).

\[
\hat{x}_{k+1|k} = \mathbb{E}\left[ x_{k+1|1} | y_1, \ldots, y_k \right], \\
= \mathbb{E}\left[ M_k x_k + w_k \right], \\
= M_k \hat{x}_{k|k} 
\]

(5)

The one step prediction of the covariance is defined by,

\[
P_{k+1|k} = \mathbb{E}\left[ (x_{k+1|1} - \hat{x}_{k+1|k})(x_{k+1|1} - \hat{x}_{k+1|k})^T | y_1, \ldots, y_k \right]. 
\]

(6)

**Exercise:** Using the state propagation model, (1), and one-step prediction of the mean, (5), show that

\[
P_{k+1|k} = M_k P_{k|k} M_k^T + Q_k. 
\]

(7)
Bayesian Framework

- \( m \): model parameter vector (unknown parameters to be estimated)
- \( d \): data vector relating to \( m \) via an equation \( h(. \)
  \[ d = h(m) + \text{noise} \]

Classical parameter estimation framework: Unknown but deterministic \( m \)
Bayesian parameter estimation framework: Unknown and random variable \( m \)

Bayes' Formula

\[
p(m, d) = p(m | d)p(d) = p(d | m)p(m)
\]

Sequential updates

\[
p(m | d) \propto p(d | m)p(m)
\]

Consider \( d \) consisting of two independent data set

\[
p(d_1, d_2) = p(d_1)p(d_2)
\]

\[
p(m | d) = p(m | d_1, d_2)
\]

\[
= \frac{p(d_1 | d_2, m)p(m | d_2)}{p(d_1 | d_2)}
\]

\[
= \frac{p(d_1 | d_2, m)p(d_2 | m)p(m)}{p(d_1 | d_2)}
\]

\[
\propto p(d_1 | m)p(d_2 | m)p(m)
\]

Generalizing

\[
p(m | d) \propto \prod_{i=1}^{N} p(d_i | m)p(m)
\]

Thus, in principle with no measurement equation, you can update sequentially or just at once
Inversion, Filtering and Smoothing

\( p(x_t \mid y_t) \): Inversion, Only observations at time \( t \)

\( p(x_t \mid y_{1:t}) \): Filter, Observations from time 1 : \( t \)

\( p(x_t \mid y_{1:T}) \): Smoother, Observations from time 1 : \( T \)

\[ x_k = f_{k-1}(x_{k-1}, v_k) \]
\[ y_k = h_k(x_k, w_k) \]

A Single Kalman Iteration

1. Predict the mean \( \hat{x}_{k|k-1} \) using previous history.
   \[ p(x_k \mid x_{k-1}) \]
   \[ \hat{x}_{k|k-1} = E\{x_k \mid x_{k-1}\} = \int x_k p(x_k \mid x_{k-1}) dx_k \]

2. Predict the covariance \( P_{k|k-1} \) using previous history.

3. Correct/update the mean using new data \( y_k \)
   \[ p(x_k \mid Y_k) \]
   \[ \hat{x}_{k|k} = E\{x_k \mid Y_k\} = \int x_k p(x_k \mid Y_k) dx_k \]

4. Correct/update the covariance \( P_{k|k} \) using \( y_k \)

\[ \cdots \Rightarrow p(x_{k-1} \mid Y_{k-1}) \Rightarrow p(x_k \mid Y_k) \Rightarrow p(x_k \mid Y_k) \Rightarrow \cdots \]

PREDICTOR-CORRECTOR  DENSITY PROPAGATOR
Product of Gaussians=Gaussian:

For the general linear inverse problem we would have

Prior: \[ p(m) \propto \exp\left\{ -\frac{1}{2}(m - m_0)^T C_m^{-1}(m - m_0) \right\} \]

Likelihood: \[ p(d|m) \propto \exp\left\{ -\frac{1}{2}(d - Gm)^T C_d^{-1}(d - Gm) \right\} \]

Posterior PDF: \[ \propto \exp\left\{ -\frac{1}{2}(d - Gm)^T C_d^{-1}(d - Gm) + (m - m_0)^T C_m^{-1}(m - m_0) \right\} \]

\[ \propto \exp\left\{ -\frac{1}{2}(m - \hat{m})^T S^{-1}(m - \hat{m}) \right\} \]

\[ S^{-1} = G^T C_d^{-1} G + C_m^{-1} \]

\[ \hat{m} = \left( G^T C_d^{-1} G + C_m^{-1} \right)^{-1} \left( G^T C_d^{-1} d + C_m^{-1} m_0 \right) \]

\[ = m_0 + \left( G^T C_d^{-1} G + C_m^{-1} \right)^{-1} G^T C_d^{-1} (d - G m_0) \]
Basic estimation theory

Observation: \( T_0 = T + e_0 \)  \( E\{e_0\} = 0 \)  \( E\{e_0^2\} = s_0^2 \)
First guess: \( T_m = T + e_m \)  \( E\{e_m\} = 0 \)  \( E\{e_m^2\} = s_m^2 \)
\( E\{e_0e_m\} = 0 \)

Assume a linear best estimate: \( T_n = a T_0 + b T_m \)
with \( T_n = T + e_n \).

Find \( a \) and \( b \) such that
1) \( E\{e_n\} = 0 \)  2) \( E\{e_n^2\} \) minimal

1) Gives: \( E\{e_n\} = E\{T_n - T\} = E\{a T_0 + b T_m - T\} = E\{a e_0 + b e_m + (a + b - 1) T\} = (a + b - 1) T = 0 \)
Hence \( b = 1 - a \).

Basic estimation theory

2) \( E\{e_n^2\} \) minimal gives:
\[
E\{e_n^2\} = E\{(T_n - T)^2\} = E\{(a T_0 + b T_m - T)^2\} = E\{(a e_0 + b e_m)^2\} = a^2 E\{e_0^2\} + b^2 E\{e_m^2\} = a^2 s_0^2 + (1 - a)^2 s_m^2
\]

This has to be minimal, so the derivative wrt \( a \) has to be zero:
\[
2 a s_0^2 - 2(1 - a) s_m^2 = 0, \quad \text{so} \quad (s_0^2 + s_m^2) a - s_m^2 = 0, \quad \text{hence:}
\]
\[
a = \frac{s_m^2}{s_0^2 + s_m^2} \quad \text{and} \quad b = 1 - a = \frac{s_0^2}{s_0^2 + s_m^2}
\]
\[
s_n^2 = E\{e_n^2\} = \frac{s_m^4 s_0^2 + s_0^4 s_m^2}{(s_0^2 + s_m^2)^2} = \frac{s_0^2 s_m^2}{s_0^2 + s_m^2}
\]
Basic estimation theory

Solution: \[ T_n = \frac{s_m^2}{s_0^2 + s_m^2} T_0 + \frac{s_0^2}{s_0^2 + s_m^2} T_m \]

and \[ \frac{1}{s_n^2} = \frac{1}{s_0^2} + \frac{1}{s_m^2} \]

Note: \( s_n \) smaller than \( s_0 \) and \( s_m \) !!

Best Linear Unbiased Estimate \( \text{BLUE} \)

Just least squares!!!
Data assimilation fuses information from (1) prior, (2) model, (3) observations to obtain consistent description of a physical system

Source of information #1: the prior encapsulates our current knowledge about the state of the system

- Background (prior) pdf: $P^b(x)$
- Current best estimate: background state $x^b$.
- Typical assumption:
  $$\varepsilon^b = x^b - S(x^{true}) \in N(0, B).$$
- With nonlinear models the normality assumption is difficult to justify, but is nevertheless used because of its convenience.
Correct models of background (prior) errors are very important for data assimilation

- Background error representation determines the spread of information, and impacts the assimilation results
- Needs: high rank, capture dynamic dependencies, efficient computations
- Traditionally estimated empirically (NMC, Hollingsworth-Lonnberg)

1. Tensor products of 1d correlations, decreasing with distance (Singh et al, 2010)
2. Multilateral AR model (Constantinescu et al 2007)
3. Hybrid methods in the context of 4D-Var (Cheng et al, 2009)

Source of information #2: the model encapsulates our knowledge about the physical laws that govern the evolution of the system

- The model evolves an initial state \( x_0 \in \mathbb{R}^n \) to future times
  \[ x_i = M_{t_{i-1} \to t_i} \left( x_0 \right). \]
- The model is imperfect
  \[ S \left( x_i^{\text{true}} \right) = M_{t_{i-1} \to t_i} \cdot S \left( x_{i-1}^{\text{true}} \right) - \eta_i, \]
  where \( \eta_i \) is the model error in step \( i \).

How large are the models of interest? Typically \( O(10^8) \) variables, and \( O(10) \) different physical processes
Source of information #3: **the observations** are sparse and noisy snapshots of reality

- Measurements $\mathbf{y}_i \in \mathbb{R}^m (m \ll n)$ taken at times $t_1, \ldots, t_N$

  $$\mathbf{y}_i = \mathcal{H}^t (\mathbf{x}_i^{\text{true}}) - \varepsilon_i^{\text{instrument}} = \mathcal{H} (S(\mathbf{x}_i^{\text{true}})) - \varepsilon_i^{\text{obs}}, \quad i = 1, \ldots, N.$$ 

- Observation operators
  - $\mathcal{H}^t$: physical space $\rightarrow$ observation space, while
  - $\mathcal{H}$: the model space $\rightarrow$ observation space.

- The **observation error**

  $$\varepsilon_i^{\text{obs}} = \varepsilon_i^{\text{instrument}} + \mathcal{H} (S(\mathbf{x}_i^{\text{true}})) - \mathcal{H}^t (\mathbf{x}_i^{\text{true}})$$

- Typical assumptions:

  $$\varepsilon_i^{\text{obs}} \in \mathcal{N} (\mathbf{0}, \mathbf{R}_i), \quad \varepsilon_i^{\text{obs}}, \varepsilon_j^{\text{obs}} \quad \text{independent for } t_i \neq t_j.$$ 

How many observations? ECMWF: O(10^7)

---

Some conventional and remote data sources used at ECMWF for numerical weather prediction

- SYNOP/METAR/SHIP: pres., wind, RH
- Aircraft: wind, temperature
- 13 Sounders: NOAA AMSU-A/B, HIRS, AIRS, ...
- Geostationary, 4 IR and 5 winds

---

To allow model-data comparison, **observation operators** map the model state space to observation space.

![Model state space](image1)

Result of DA: **the analysis**, which encapsulates our enhanced knowledge about the state of the system.

- **The analysis** (posterior) probability density $P^a(x)$:
  
  Bayes: \[ P^a(x) = P(x|y) = \frac{P(y|x) \cdot P^b(x)}{P(y)} \]

- Best posterior state estimate: the **analysis** $x^a$.
- Analysis estimation errors $e^a = x^a - S(x^{true})$ characterized by bias $\beta^a = \mathbb{E}[e^a]$, covariance $A = \text{cov}(e^a - \beta^a) \in \mathbb{R}^{n \times n}$.

- **Kalman filter**: analytical solution for $P^a(x)$ in Gaussian, linear case
- **Methods of practical interest**:
  - Suboptimal and Ensemble Kalman filters ($\sim$ min. var.)
  - Variational methods (MAP)
Result of DA: **the analysis**, which encapsulates our enhanced knowledge about the state of the system

- The analysis (posterior) probability density $\mathcal{P}^a(\mathbf{x})$:

\[
\text{Bayes: } \quad \mathcal{P}^a(\mathbf{x}) = \mathcal{P}(\mathbf{x}|\mathbf{y}) = \frac{\mathcal{P}(\mathbf{y}|\mathbf{x}) \cdot \mathcal{P}^b(\mathbf{x})}{\mathcal{P}(\mathbf{y})}. 
\]

- Best posterior state estimate: the **analysis** $\mathbf{x}^a$.
- Analysis estimation errors $\mathbf{e}^a = \mathbf{x}^a - S(\mathbf{x}^{\text{true}})$ characterized by **bias** $\beta^a = \mathbb{E}^a[\mathbf{e}^a]$, **covariance** $\mathbf{A} = \text{cov}(\mathbf{e}^a - \beta^a) \in \mathbb{R}^{n \times n}$.
- **Kalman filter**: analytical solution for $\mathcal{P}^a(\mathbf{x})$ in Gaussian, linear case
The ensemble Kalman filter (EnKF) is based on EKF, and uses a MC approach to propagate covariances.

Sequential approach to DA:
Incorporates data in succession

\[
x_{k+1}^{b(i)} = M_{t_k \rightarrow t_{k+1}} (x_k^{a(i)}) + \eta_{k+1}^{(i)}
\]

\[
P_k^b = \frac{1}{K} \sum_{i=1}^{K} (x_k^{b(i)} - \bar{x}_k^b) (x_k^{b(i)} - \bar{x}_k^b)^T
\]

The ensemble Kalman filter (EnKF) is based on EKF, and uses a MC approach to propagate covariances.

Convert each model state to an expected observation
\[ y = h(x) \]

\[
P_k^b H_k^T \approx \frac{1}{K} \sum_{i=1}^{K} (x_k^{b(i)} - \bar{x}_k^b) (H_k (x_k^{b(i)}) - H_k (x_k^b))^T
\]

\[
H_k P_k^b H_k^T \approx \frac{1}{K} \sum_{i=1}^{K} (H_k (x_k^{b(i)}) - H_k (x_k^b)) (H_k (x_k^{b(i)}) - H_k (x_k^b))^T
\]
The ensemble Kalman filter (EnKF) is based on EKF, and uses a MC approach to propagate covariances.

\[ x_k^a = x_k^b + K_k \left( y_k - H_k \left( x_k^b \right) \right) \]
The ensemble Kalman filter (EnKF) is based on EKF, and uses a MC approach to propagate covariances.