PROBABILISTIC MODELS – PART 6b: THE KALMAN FILTER

Gerhard Widmer

Institute of Computational Perception
Johannes Kepler University
Linz, Austria

gerhard.widmer@jku.at
www.cp.jku.at/people/widmer

January 12, 2026

Presentation partly based on and inspired by [Koller & Friedman, 2009] and [Russell & Norvig, 2021], including the use of some figures from their books and/or lecture slides.

Many thanks to Daphne Koller, Nir Friedman, Stuart Russell, and Peter Norvig for making these available (`pgm.stanford.edu`; `aima.cs.berkeley.edu`).

Goals of this Lecture

- Introduce a popular class of state-observation models for continuous systems
- Show how to track a dynamic process with the Kalman Filter
- Introduce some extended variants of the KF (briefly)
- Show how the Kalman Filter is a special case of Dynamic Bayes Nets (DBNs)
- Discuss some real-life applications.

Outline

Motivation

Consider the following scenarios:

- You watch a bird fly through the jungle; only see it through the leaves at some time points; try to guess where the bird is and where it will appear next
- You watch blips on a radar screen appearing once every 5 or 10 seconds; try to keep track of which blip is which plane, and how it is moving
- You are Johannes Kepler trying to reconstruct the motions of the planets from a collection of highly inaccurate measurements taken at irregular and imprecisely measured intervals.

Motivation

Consider the following scenarios:

- You watch a bird fly through the jungle; only see it through the leaves at some time points; try to guess where the bird is and where it will appear next
- You watch blips on a radar screen appearing once every 5 or 10 seconds; try to keep track of which blip is which plane, and how it is moving
- You are Johannes Kepler trying to reconstruct the motions of the planets from a collection of highly inaccurate measurements taken at irregular and imprecisely measured intervals.

Common problem:

- Try to estimate the state of a physical system (e.g., position and velocity) from noisy, incomplete observations over time.

Can be formulated as a state-observation model:

- *Transition Model* describes physics of the system (motion)
- *Observation Model* describes measurement process and its imprecision.

An Extremely Simple Example: Tracking a Bicycle

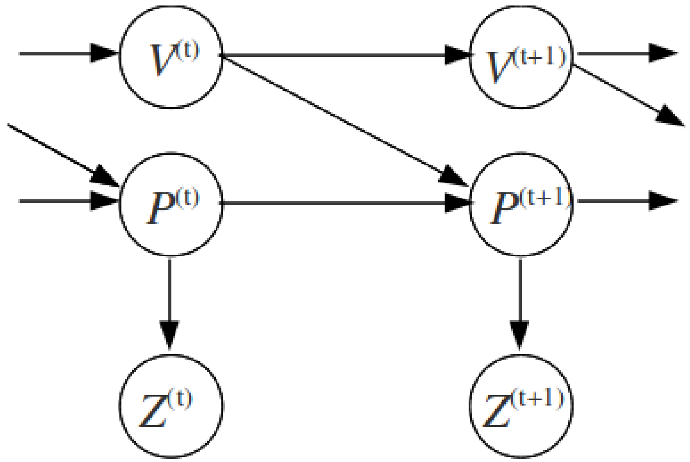
Simplified bike scenario:

- A bicyclist rides along a road
- For simplicity: the road is a straight line (consider only 1 single dimension)
- Position is measured via a GPS sensor once every second
- Model bike's position (in m) at second t by a continuous variable $P^{(t)}$
- Model the velocity of the bike (in m/sec) by a continuous variable $V^{(t)}$
- Model the GPS measurement of the position (in m) by variable $Z^{(t)}$

Note:

- True position and velocity are not observable!
- All we observe is (potentially noisy and unreliable) GPS readings

A State-Observation Model of the Bicycle Example



A dynamical system with

- Position $P^{(t)}$ (continuous state variable; hidden)
- Velocity $V^{(t)}$ (continuous state variable; hidden)
- noisy position measurements $Z^{(t)}$ (continuous observation variable; observed).

Continuous Systems and Variables

So far:

- Assumed that all variables are *discrete* with a finite domain $Val(X)$
- Represented conditional probability distributions $P(X \mid \mathrm{Pa}(X))$ as a *table*.

Problem:

- Many systems and processes involve *continuous* aspects and measurements that are best modelled by *real-valued* variables
- Example: Position and motion of an airplane or a robot (and our measurements of their position and motion).

Implications for State-Observation Models:

- Both internal system states and observations need to be modelled as continuous quantities
- Transition Model $P(\boldsymbol{S}' \mid \boldsymbol{S})$ and Observation Model $P(\boldsymbol{O} \mid \boldsymbol{S})$ become *continuous* conditional distributions
- HMMs cannot deal with this and are not applicable.

Continuous Systems and Variables[1]

Solution 1: Discretisation

- Split the domain of a variable into a fixed number of intervals; represent each continuous value by the label of its interval
- Advantage: All of our representation and reasoning machinery is directly applicable.

Problems

- Problem 1: Loss of information caused by this approximation
- Problem 2: Complexity
- Example: For reasonably accurate robot tracking, might need to discretise position variables into $1,000$ intervals
  $\Rightarrow 1,000,000$ different (x, y) positions
  $\Rightarrow$ Imagine the resulting CPDs ...

[1] Recapitulation from Chapter 3 (“Bayesian Networks – Representation and Semantics”)

Continuous Systems and Variables

Solution 2: Work with continuous variables and CPDs

- Permit continuous random variables X with $Val(X) \subseteq \mathbb{R}$
- Model CPDs as *continuous density functions* $p(X)$ and $p(X \mid \mathrm{pa}(X))$

Main Questions:

1. How to model the distribution of a continuous variable X?
2. How to define *conditional* distributions $p(X \mid \mathrm{pa}(X))$, when the parents of X are also continuous variables?

Example: $X \longrightarrow Y$, with both X and Y continuous

- CPD $P(Y|X)$ needs a conditional distribution $p(Y|x)$ over Y, for *every possible value* $x \in Val(X) \subseteq \mathbb{R}$
  $\Rightarrow$ need to define an *infinite* number of conditional distributions $p(Y|x)$!

General Approach:

1. Use a parameterisable family of density functions to model distributions
2. Design a general rule that uniquely **computes** the parameters of conditional distribution $p(Y|x)$ as a **function of the specific parent value** x.

Linear Gaussian Models

Basic Idea

- Model all (prior and conditional) distributions over single continuous variables X as Gaussians $\mathcal{N}(\mu;\sigma^2)$:

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

- Model *conditional* distributions $p(Y|X)$ as Gaussians whose parameters μ, σ^2 depend on the value of X
- Need a set of parameters $\{\mu_i, \sigma_i^2\}$ for every one of the infinitely many values $x_i \in Val(X)$
- Solution: Assume that the **mean** μ of Y in $p(Y|X)$ is a **linear function of** X, and that the **variance** σ^2 of Y **is independent of** X (i.e., fixed)[2]

Example:

$$p(Y \mid x) = \mathcal{N}(\underbrace{3.5x - 0.9}_{\mu}; \underbrace{1.0}_{\sigma^2})$$

[2]Note that this is the same model that underlies *Least Squares Linear Regression* (see slide set 05b)

Linear Gaussian Models

Definition

Let Y be a continuous variable with continuous parents $X_1, \ldots, X_k$.
We say that Y has a **LINEAR GAUSSIAN MODEL** if there are parameters $\beta_0, \beta_1, \ldots, \beta_k$ and σ^2 such that

$$p(Y \mid x_1, \ldots, x_k) = \mathcal{N}(\underbrace{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}_{\mu};\ \sigma^2)$$

In vector notation:

$$p(Y \mid \boldsymbol{x}) = \mathcal{N}(\beta_0 + \boldsymbol{\beta}^T \boldsymbol{x};\ \sigma^2)$$

Alternative Interpretation:
Y is a linear function of its parents $X_1, \ldots, X_k$, with the addition of Gaussian noise (randomly distributed “error”) ϵ:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon$$

where ϵ is a normally distributed random variable with mean 0 and variance σ^2.

The Bicycle Case as a Simple Linear Gaussian System

Simplified bike scenario:

- Bike's position (in m) at second t modelled by a continuous variable $P^{(t)}$
- Velocity of the bike (in m/sec) modelled by a continuous variable $V^{(t)}$

Consider only the transition model:

- Under constant motion, we would expect that $P^{(t+1)} = P^{(t)} + V^{(t)}$
  (If the bicycle is at meter 510.4 at time t, and its velocity is $5m/sec$, then we expect its position $P^{(t+1)}$ the next second to be at meter 515.4) ...
- *but* there is inevitably some stochasticity (random imprecision) in motion.

$\Rightarrow$ More realistic to say that the bicycle's position $P^{(t+1)}$ is described by a normal distribution whose mean is 515.4 and whose variance is 0.8 meters.

The Kalman Filter

Definition

A **KALMAN FILTER** is a state-observation model for systems described by one or several continuous state variables, and one or several continuous observation variables, where

- all state and observation distributions are modelled as (multivariate) Gaussians
- All random variables (both transition model and observation model) have **linear Gaussian models.**

This implies

- linear state transitions
- linear dependency of observations on states, and
- Gaussian noise.

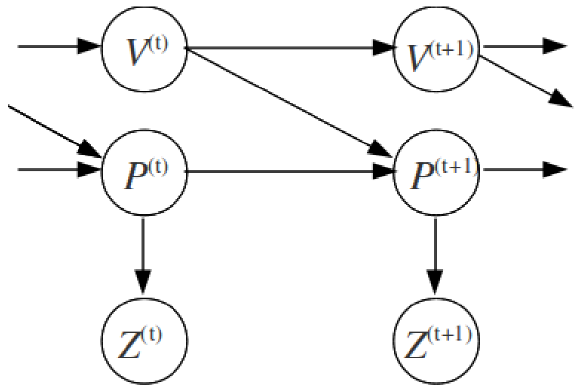
Kalman filters are mainly used for real-time tracking tasks.

The Bicycle Case as a Kalman Filter

TRANSITION MODEL:

$$\begin{aligned} p^{(t+1)} &= p^{(t)} + v^{(t)} + \epsilon_p \\ v^{(t+1)} &= v^{(t)} + \epsilon_v \end{aligned}$$

with $\epsilon_p \sim \mathcal{N}(0; \sigma_p^2); \quad \epsilon_v \sim \mathcal{N}(0; \sigma_v^2)$



In Matrix Notation:

$$\begin{pmatrix} p^{(t+1)} \\ v^{(t+1)} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \times \begin{pmatrix} p^{(t)} \\ v^{(t)} \end{pmatrix} + \begin{pmatrix} \epsilon_p \\ \epsilon_v \end{pmatrix} \qquad \begin{pmatrix} \epsilon_p \\ \epsilon_v \end{pmatrix} \sim \mathcal{N}(\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \sigma_p^2 & 0 \\ 0 & \sigma_v^2 \end{pmatrix})$$

General Formulation:

$$\boxed{\boldsymbol{x}^{(t+1)} = \boldsymbol{A}\boldsymbol{x}^{(t)} + \boldsymbol{\epsilon_x} \qquad \boldsymbol{\epsilon_x} \sim \mathcal{N}(\boldsymbol{0}; \boldsymbol{\Sigma_x})}$$

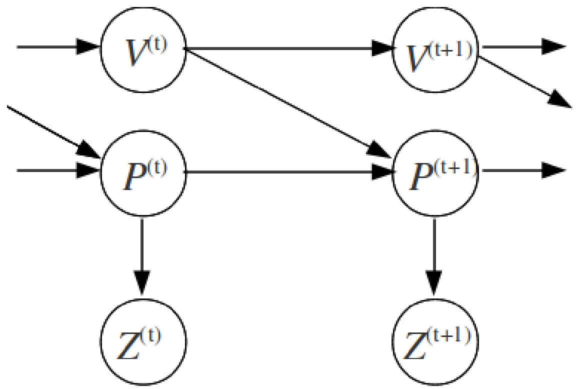
$$\boxed{p(\boldsymbol{x}^{(t+1)} \mid \boldsymbol{x}^{(t)}) = \mathcal{N}(\boldsymbol{A}\boldsymbol{x}^{(t)}; \boldsymbol{\Sigma_x})}$$

The Bicycle Case as a Kalman Filter

OBSERVATION MODEL:

$$z^{(t)} = p^{(t)} + \epsilon_z$$

with $\epsilon_z \sim \mathcal{N}(0; \sigma_z^2)$



In Matrix Notation:

$$z^{(t)} = \begin{pmatrix} 1 & 0 \end{pmatrix} \times \begin{pmatrix} p^{(t)} \\ v^{(t)} \end{pmatrix} + \epsilon_z \qquad \epsilon_z \sim \mathcal{N}(0; \sigma_z^2)$$

General Formulation:

$$\boxed{\boldsymbol{z}^{(t)} = \boldsymbol{B}\boldsymbol{x}^{(t)} + \epsilon_z \qquad \epsilon_z \sim \mathcal{N}(\mathbf{0}; \boldsymbol{\Sigma}_z)}$$

$$\boxed{p(\boldsymbol{z}^{(t)} \mid \boldsymbol{x}^{(t)}) = \mathcal{N}(\boldsymbol{B}\boldsymbol{x}^{(t)}; \boldsymbol{\Sigma}_z)}$$

The Kalman Filter: Formal Specification

Definition

A **KALMAN FILTER** is a state-observation model defined by

- continuous hidden state variables $\boldsymbol{X}$ following a joint normal distribution
- continuous observation variables $\boldsymbol{Z}$ following a joint normal distribution
- an initial state distribution $p(\boldsymbol{X}^{(0)}) = \mathcal{N}(\boldsymbol{\mu}^{(0)}, \boldsymbol{\Sigma}^{(0)})$
- a matrix $\boldsymbol{A}$ and a covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{x}}$ that specify the **linear Gaussian state transition model**

$$p(\boldsymbol{X}' \mid \boldsymbol{x}) = \mathcal{N}(\boldsymbol{A}\boldsymbol{x}; \boldsymbol{\Sigma}_{\boldsymbol{x}})$$

- a matrix $\boldsymbol{B}$ and a covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{z}}$ that specify the **linear Gaussian observation model**

$$p(\boldsymbol{Z} \mid \boldsymbol{x}) = \mathcal{N}(\boldsymbol{B}\boldsymbol{x}; \boldsymbol{\Sigma}_{\boldsymbol{z}})$$

Tracking with the Kalman Filter

The Filtering/Tracking Task

Given:

- Observation sequence $\boldsymbol{O} = \langle \boldsymbol{z}^{(1)} \ldots \boldsymbol{z}^{(t)} \rangle$

Compute:

- $p(\boldsymbol{X}^{(t)} \mid \boldsymbol{z}^{(1:t)}) \stackrel{?}{=} \mathcal{N}(\boldsymbol{\mu}^{(t)}; \boldsymbol{\Sigma}^{(t)})$

Main Problem:

- How to compute the forward message $\mathrm{f}^{(t)}$ (as a multivariate Gaussian) with linear Gaussian models?
- ... and is this going to be Gaussian at all, with well-defined $\boldsymbol{\mu}^{(t)}$, $\boldsymbol{\Sigma}^{(t)}$?

In typical applications, tracking must work in **real time**, **on-line**, for each $t = 1, \ldots$

Tracking with the Kalman Filter

Remember: Tracking in state-observation models involves 2 steps

STEP 1: State Propagation:

Given $P(\boldsymbol{S}^{(t)} \mid \boldsymbol{o}^{(1:t)})$, compute

$$P(\boldsymbol{S}^{(t+1)} \mid \boldsymbol{o}^{(1:t)}) = \sum_{\boldsymbol{s}^{(t)}} P(\boldsymbol{S}^{(t+1)}|\boldsymbol{s}^{(t)})P(\boldsymbol{s}^{(t)}|\boldsymbol{o}^{(1:t)})$$

For the continuous Kalman Filter case, this translates into:

$$p(\boldsymbol{x}^{(t+1)} \mid \boldsymbol{z}^{(1:t)}) = \int_{\boldsymbol{x}^{(t)}} p(\boldsymbol{x}^{(t+1)}|\boldsymbol{x}^{(t)})p(\boldsymbol{x}^{(t)}|\boldsymbol{z}^{(1:t)})d\boldsymbol{x}^{(t)}$$

... because we have an infinite number of possible predecessor states $\boldsymbol{x}^{(t)}$

Tracking with the Kalman Filter

STEP 2: Conditioning on new observation:

$$P(\boldsymbol{S}^{(t+1)} \mid \boldsymbol{o}^{(1:t+1)}) \quad = \quad \frac{1}{Z} P(\boldsymbol{o}^{(t+1)} | \boldsymbol{S}^{(t+1)}) P(\boldsymbol{S}^{(t+1)} | \boldsymbol{o}^{(1:t)})$$

For the continuous Kalman Filter case, this translates into:

$$\begin{aligned} p(\boldsymbol{x}^{(t+1)} \mid \boldsymbol{z}^{(1:t+1)}) \quad &= \quad \frac{1}{Z} p(\boldsymbol{z}^{(t+1)} | \boldsymbol{x}^{(t+1)}) p(\boldsymbol{x}^{(t+1)} | \boldsymbol{z}^{(1:t)}) \\ &\overset{?}{=} \quad \mathcal{N}(\boldsymbol{\mu}^{(t+1)}; \boldsymbol{\Sigma}^{(t+1)}) \end{aligned}$$

Need to show

- that the resulting (conditional) distributions always remain Gaussian!
- how to compute the new $\boldsymbol{\mu}^{(t+1)}$ and $\boldsymbol{\Sigma}^{(t+1)}$ that define $p(\boldsymbol{x}^{(t+1)} \mid \boldsymbol{z}^{(1:t+1)})$, from $\boldsymbol{\mu}^{(t)}$ and $\boldsymbol{\Sigma}^{(t)}$

An Important Property of the Gaussian Family of Distributions

Theorem

The Gaussian Family of Distributions remains **closed under the update operations** used in Dynamic Bayesian Networks:[a]

1. If the distribution $p(\boldsymbol{x}^{(t)} \mid \boldsymbol{z}^{(1:t)})$ is Gaussian and the transition model $p(\boldsymbol{x}^{(t+1)} \mid \boldsymbol{x}^{(t)})$ is linear Gaussian, the one-step predicted distribution

$$p(\boldsymbol{x}^{(t+1)} \mid \boldsymbol{z}^{(1:t)}) = \int_{\boldsymbol{x}^{(t)}} p(\boldsymbol{x}^{(t+1)}|\boldsymbol{x}^{(t)})p(\boldsymbol{x}^{(t)}|\boldsymbol{z}^{(1:t)})d\boldsymbol{x}^{(t)}$$

   is also a Gaussian distribution.

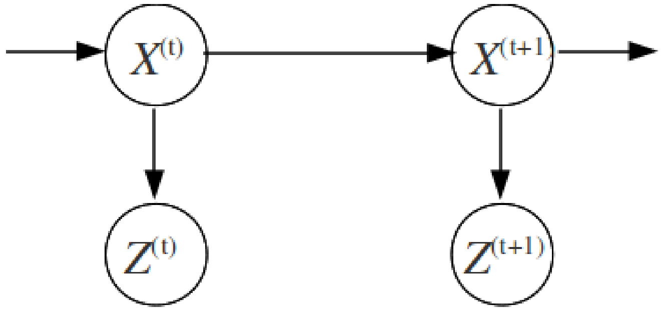
2. If the predicted distribution $p(\boldsymbol{x}^{(t+1)} \mid \boldsymbol{z}^{(1:t)})$ is Gaussian and the observation model $p(\boldsymbol{Z} \mid \boldsymbol{X})$ is linear Gaussian, then the distribution obtained by conditioning on a new observation $\boldsymbol{z}^{(t+1)}$

$$p(\boldsymbol{x}^{(t+1)} \mid \boldsymbol{z}^{(1:t+1)}) = \frac{1}{Z}p(\boldsymbol{z}^{(t+1)}|\boldsymbol{x}^{(t+1)})p(\boldsymbol{x}^{(t+1)}|\boldsymbol{z}^{(1:t)})$$

   is also a Gaussian distribution.

[a] Proof omitted due to complexity.

A Trivially Simple One-dimensional Example



Transition Model:

$$x^{(t+1)} = x^{(t)} + \epsilon_x \quad \text{with} \quad \epsilon_x \sim \mathcal{N}(0; \sigma_x^2)$$

Observation Model:

$$z^{(t)} = x^{(t)} + \epsilon_z \quad \text{with} \quad \epsilon_z \sim \mathcal{N}(0; \sigma_z^2)$$

Possible interpretation of the story:

X is a “consumer confidence index” that undergoes a random Gaussian-distributed change each month, and that is measured by a random consumer survey Z that also introduces Gaussian sampling noise. *(Example due to Russell & Norvig, 2003)*

A Simple One-dimensional Example

Let's simulate the FORWARD step from $t = 0$ to 1:

STEP 1: State Propagation:

$$\boxed{p(x^{(1)} \mid \{\})} = \int_{-\infty}^{\infty} p(x^{(1)}|x^{(0)})p(x^{(0)})dx^{(0)}$$

$$= \frac{1}{Z} \times \int_{-\infty}^{\infty} e^{-\frac{(x^{(1)}-x^{(0)})^2}{2\sigma_x^2}} e^{-\frac{(x^{(0)}-\mu_0)^2}{2\sigma_0^2}} dx^{(0)}$$

= *... some rather complex calculations ...*

$$= \boxed{\frac{1}{Z} \times e^{-\frac{(x^{(1)}-\mu_0)^2}{2(\sigma_0^2+\sigma_x^2)}}}$$

In Words:

- The one-step predicted state distribution is a Gaussian with the same mean μ_0 and with a variance that is the sum of the original variance σ_0 and the transition variance σ_x
- State transition model adds variance (uncertainty).

A Simple One-dimensional Example

STEP 2: Conditioning on the Observation:

$$
\begin{aligned}
\boxed{p(x^{(1)}|z^{(1)})} &= \frac{1}{Z} \times p(z^{(1)}|x^{(1)})p(x^{(1)}) \\
&= \frac{1}{Z} \times e^{-\frac{(z^{(1)}-x^{(1)})^2}{2\sigma_z^2}} \; e^{-\frac{(x^{(1)}-\mu_0)^2}{2(\sigma_0^2+\sigma_x^2)}} \\
&= \text{\textit{... some even more complex calculations ...}} \\
&= \boxed{\frac{1}{Z} \times e^{-\frac{(x^{(1)}-\mu_1)^2}{2\sigma_1^2}}}
\end{aligned}
$$

where

$$
\boxed{\mu_1 = \frac{(\sigma_0^2+\sigma_x^2)z^{(1)} + \sigma_z^2\mu_0}{\sigma_0^2+\sigma_x^2+\sigma_z^2} \qquad \sigma_1^2 = \frac{(\sigma_0^2+\sigma_x^2)\sigma_z^2}{\sigma_0^2+\sigma_x^2+\sigma_z^2}}
$$

In Words:

- The complete forward step produces a new Gaussian distribution $p(x^{(1)})$ whose parameters μ_1 and σ_1 are derived from the parameters of the previous distribution $p(x^{(0)})$ and the transition and observation models
- This combination *makes a lot of intuitive sense* (see next slide)

A Simple One-dimensional Example

$$\mu_1 = \frac{(\sigma_0^2 + \sigma_x^2) z^{(1)} + \sigma_z^2 \mu_0}{\sigma_0^2 + \sigma_x^2 + \sigma_z^2} \qquad \sigma_1^2 = \frac{(\sigma_0^2 + \sigma_x^2)\sigma_z^2}{\sigma_0^2 + \sigma_x^2 + \sigma_z^2}$$

Intuitive Interpretation of the Forward Update Step:

- New mean μ_1 is a *weighted average* of new observation $z^{(1)}$ and old mean (= prediction of state transition model) μ_0:
- If observations are unreliable (large σ_z^2), pay more attention to state forward prediction μ_0
- If the old mean is unreliable (σ_0^2 is large) or the process is highly unpredictable (σ_x^2 is large), pay more attention to the observation $z^{(1)}$
- Note: Update for the variance σ_1^2 is *independent of the observation* $z^{(1)}$

  Actually: will quickly converge to a fixed value that depends only on σ_x^2 and σ_z^2 (transition and observation noise).

A Simple One-dimensional Example[3]

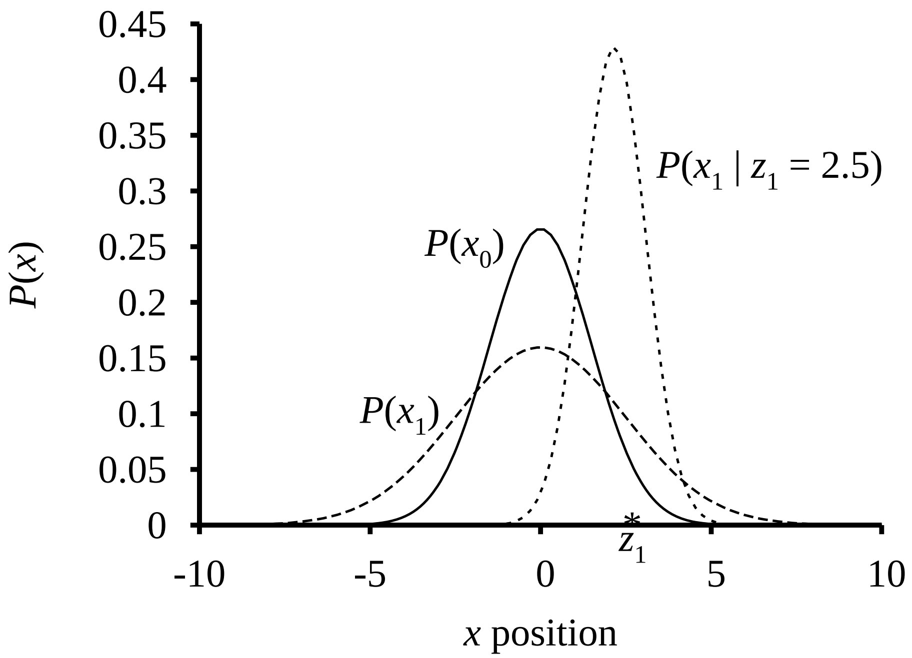
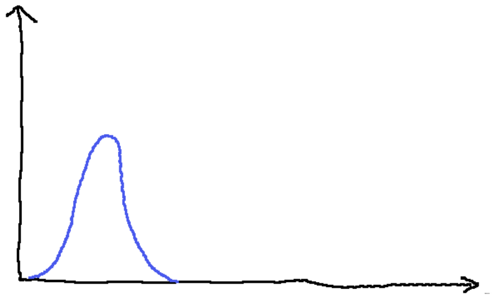


Figure : The two steps in the Kalman filter update from $t = 0$ to 1 for a random walk with prior given by $\mu_0 = 0.0$ and $\sigma_0 = 1.0$, transition noise $\sigma_x = 2.0$, observation noise $\sigma_z = 1.0$, and first observation $z_1 = 2.5$ (marked on the x axis). Notice how the prediction $p(x_1)$ is flattened out, relative to $p(x_0)$, by the transition noise. Notice also that the mean of the posterior $P(x_1|z_1)$ is slightly to the left of the observation z_1 because the mean is a weighted average of the prediction and the observation.

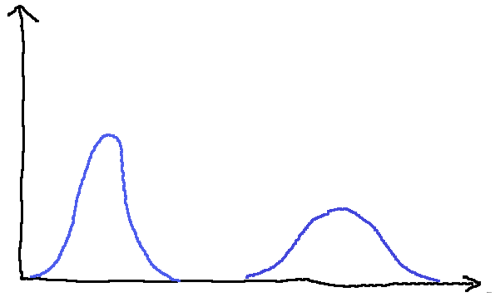
[3] Figure and caption from (Russell & Norvig, 2003, ch.15)

Tracking with the Kalman Filter: The General Loop

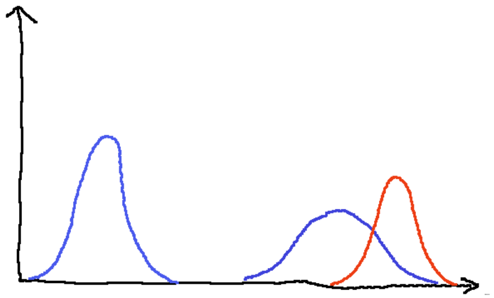
1. State hypothesis $p(\boldsymbol{x}^{(t)}|\boldsymbol{z}^{(1:t)})$ at time t (= forward message $\mathrm{f}^{(1:t)}$)



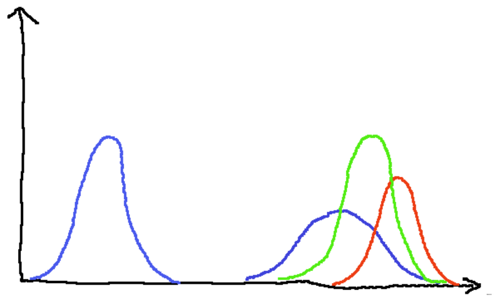
2. Next state $p(\boldsymbol{x}^{(t+1)}|\boldsymbol{z}^{(1:t)})$, predicted by transition model (*Prior Belief State*); implies a distribution $p(\hat{\boldsymbol{z}}^{(t+1)}|\boldsymbol{x}^{(t+1)})$ over **expected next observation** $\hat{\boldsymbol{z}}$.



3. **Actual next observation** $\boldsymbol{z}$ (possibly different from expected $\hat{\boldsymbol{z}}$) implies a different distribution $p(\boldsymbol{x}^{(t+1)}|\boldsymbol{z}^{(t+1)})$ (by observation model)



4. Update: Combine prediction and observation into *Posterior Belief State* $p(\boldsymbol{x}^{(t+1)}|\boldsymbol{z}^{(1:t+1)})$



The General Filtering Algorithm

FORWARD ALGORITHM for KALMAN FILTERING

Given:

- Transition model $[\boldsymbol{A}; \boldsymbol{\Sigma}_{\boldsymbol{x}}]$, observation model $[\boldsymbol{B}; \boldsymbol{\Sigma}_{\boldsymbol{z}}]$
- Current state distribution $p(\boldsymbol{x}^{(t)}|\boldsymbol{z}^{(1:t)}) = \mathcal{N}(\boldsymbol{\mu}^{(t)}; \boldsymbol{\Sigma}^{(t)})$

The Main Recursive Step:

1. Predict next state distribution $p(\boldsymbol{x}^{(t+1)}|\boldsymbol{z}^{(1:t)})$:

$$\boldsymbol{\mu}^* = \boldsymbol{A}\boldsymbol{\mu}^{(t)}; \quad \boldsymbol{\Sigma}^* = \boldsymbol{A}\boldsymbol{\Sigma}^{(t)}\boldsymbol{A}^T + \boldsymbol{\Sigma}_{\boldsymbol{x}}$$

2. Predict most likely next observation: $\hat{\boldsymbol{z}} = \boldsymbol{B}\boldsymbol{\mu}^*$
3. Observe actual observation $\boldsymbol{z}^{(t+1)}$
   $\Rightarrow (\boldsymbol{z}^{(t+1)} - \hat{\boldsymbol{z}}) = (\boldsymbol{z}^{(t+1)} - \boldsymbol{B}\boldsymbol{\mu}^*) =$ ***prediction error***
4. Update state distribution, based on prediction and error:

$$\begin{aligned} \boldsymbol{\mu}^{(t+1)} &= \boldsymbol{\mu}^* + \boldsymbol{K}^{(t+1)}(\boldsymbol{z}^{(t+1)} - \hat{\boldsymbol{z}}) \\ \boldsymbol{\Sigma}^{(t+1)} &= (\boldsymbol{I} - \boldsymbol{K}^{(t+1)}\boldsymbol{B})\boldsymbol{\Sigma}^* \end{aligned}$$

The Kalman Gain Matrix

... where $\boldsymbol{K}^{(t+1)}$ is called the **Kalman Gain Matrix** and is defined as (don't be afraid):

$$\boldsymbol{K}^{(t+1)} = \boldsymbol{\Sigma}^* \boldsymbol{B}^T (\boldsymbol{B}\boldsymbol{\Sigma}^* \boldsymbol{B}^T + \boldsymbol{\Sigma}_{\boldsymbol{z}})^{-1}$$

The Kalman Gain Matrix tells us how seriously to take the next observation $\boldsymbol{z}^{(t+1)}$, in relation to the predicted new state $\boldsymbol{\mu}^*$, when calculating our new state hypothesis $p(\boldsymbol{x}^{(t+1)}|\boldsymbol{z}^{(1:t+1)})$

(analogous to (much simpler) "consumer confidence index" example above).

No need to learn this by heart ...

Intuitive Summary

The FORWARD **Operator for Kalman Filters**

- takes a Gaussian forward message $\mathrm{f}^{(0:t)} = \mathcal{N}(\boldsymbol{\mu}^{(t)}, \boldsymbol{\Sigma}^{(t)})$
- and produces new Gaussian forward message $\mathrm{f}^{(0:t+1)} = \mathcal{N}(\boldsymbol{\mu}^{(t+1)}, \boldsymbol{\Sigma}^{(t+1)})$
- New mean $\boldsymbol{\mu}^{(t+1)}$ is predicted mean ($\boldsymbol{\mu}^* = \boldsymbol{A}\boldsymbol{\mu}^{(t)}$), corrected by prediction error (= difference between actual and predicted observation $\boldsymbol{z}^{(t+1)} - \boldsymbol{B}\boldsymbol{\mu}^*$)
- Strength of correction is modulated by the **Kalman Gain Matrix** $\boldsymbol{K}^{(t+1)}$, which takes into account the uncertainty $\boldsymbol{\Sigma}^*$ of our forward-predicted state hypothesis (which includes the uncertainty $\boldsymbol{\Sigma}_{\boldsymbol{x}}$ of the transition model), and the uncertainty $\boldsymbol{\Sigma}_{\boldsymbol{z}}$ introduced by the observation model.

A 2-D Example[4]

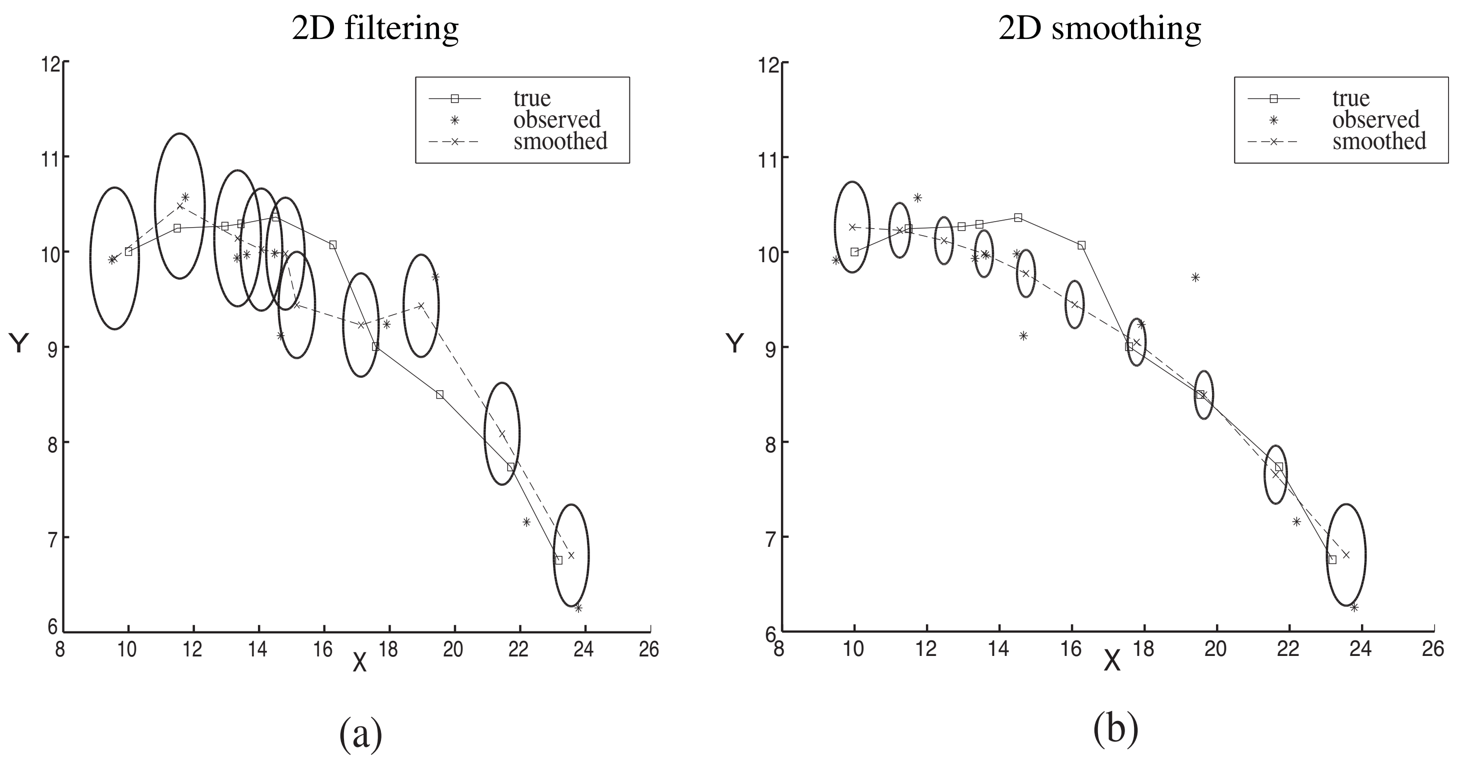
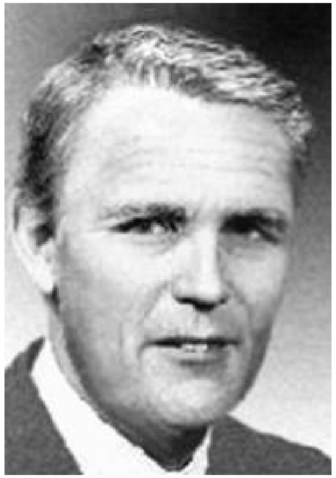


Figure : (a) Results of Kalman filtering for an object moving on the $X - Y$ plane ($\Rightarrow 2$ observation variables X, Y), showing the true trajectory (left to right), a series of noisy observations, and the trajectory estimated by Kalman filtering. Ovals indicate the variance in the position estimate. (b) Result of Kalman *smoothing* (not treated in class), for the same observation sequence.

[4] Figure from (Russell & Norvig, 2003, ch.15)

Credits



R.E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Transaction of the ASME-Journal of Basic Engineering* 35-45, 1960.

R.L. Stratonovich. Application of the Markov Processes Theory to Optimal Filtering. *Radio Engineering and Electronic Physics* 5(11):1-19, 1960. Translated from Russian.

The Kalman Filter: Applications and Limitations

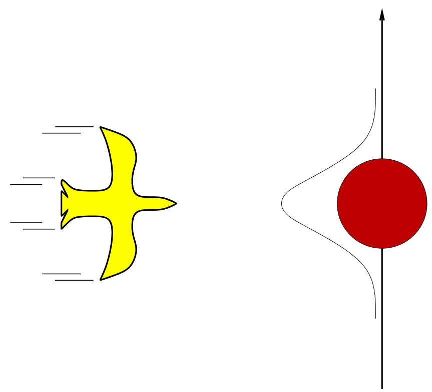
Large Number of Practical Applications

- *Civilian and military aviation:* Radar tracking of aircraft and missiles; acoustic tracking of submarines, ...
- *Aerospace:* Used, e.g., in guidance and navigation systems of the NASA Space Shuttle and the control and navigation systems of the International Space Station
- *Physics:* Reconstructing trajectories of sub-atomic particles in bubble chambers
- *Meteorology:* Reconstructing ocean currents from satellite surface measurements, ...
- *Industry:* Process monitoring in chemical plants, nuclear reactors, national economies, ...

Limitations

- The Gaussian distribution, linear dependency, and fixed variance assumptions are rather severe restrictions
- May be satisfied (or nearly satisfied) in some processes and applications, and not in others.

The Problem with Nonlinear Systems[5]

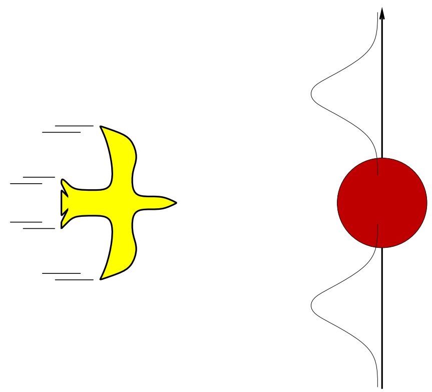


The Problem:

- A bird flies toward a tree (viewed from the top)
- If bird is equally likely to take left or right turn, Kalman filter will predict the location of the bird using a single Gaussian centered directly on the tree!

[5] Figure from (Russell & Norvig, 2003, ch.15)

The Problem with Nonlinear Systems



More realistic model:

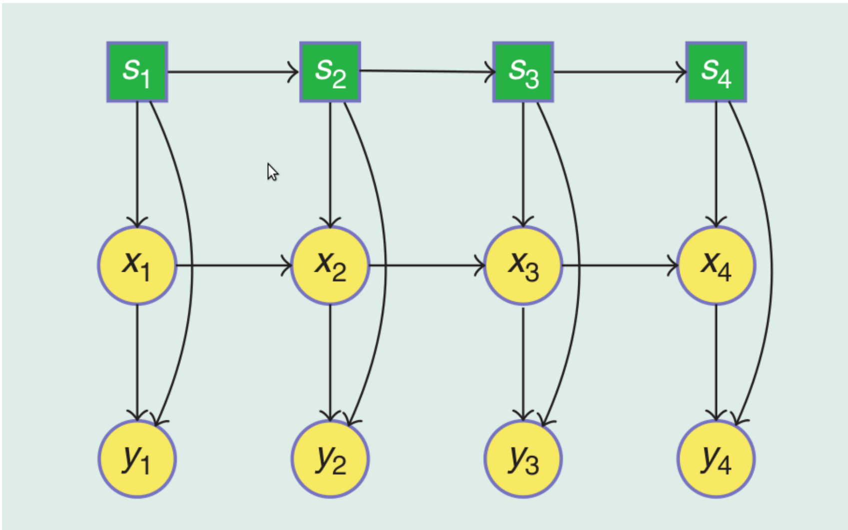
- should allow to model and predict that bird will fly either left or right.
- Example of a *highly non-linear* system: bird's decision (and next position) will differ sharply depending on its exact position relative to the tree.

Solution: Switching Kalman Filter

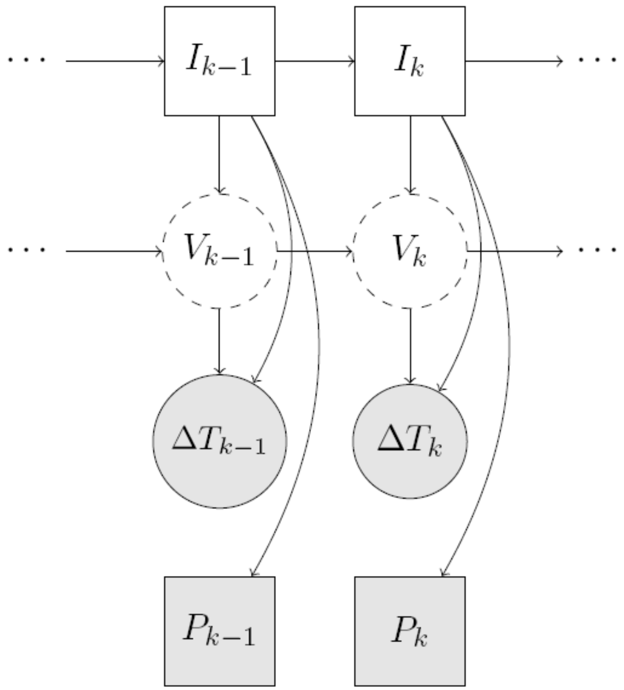
The Switching Kalman Filter (Briefly)

Basic Idea:

- Conceptually: run multiple Kalman filters in parallel, each with a different system model (e.g., one for straight flight, one for left, one for right turns)
- Have an HMM on top whose discrete hidden state $S^{(t)}$ acts as a (probabilistic) 'switch' variable
- For any discrete state i, the transition model $P(X^{(t+1)} \mid X^{(t)}, S^{(t)} = i)$ is a linear Gaussian model (as in a Kalman filter)
- Use *weighted sum of predictions* (depending on $P(S^{(t)} \mid \boldsymbol{y}^{(1:t)})$
- Specialised inference algorithms for this (Murphy, 1998).



Real-time Adaptive Piano Accompaniment with a Switching Kalman-HMM Model[6]



https://www.youtube.com/watch?v=KE6WhYxuWLk

I	...	Position in score (HMM; discrete, hidden)
V	...	Performance tempo (KF; continuous, hidden)
ΔT	...	Performed inter-onset intervals (KF/HMM; continuous, observed)
P	...	Performed notes (pitches) (HMM; discrete, observed)

[6] Cancino-Chacón et al., 2023; Durand, 2017.

FALLING WALLS **PRESS RELEASES**

FALLING WALLS ANNOUNCES SCIENCE BREAKTHROUGHS OF THE YEAR 2021

Berlin, 15 September 2021. Today the first Science Breakthroughs of the Year Awards are announced by the Falling Walls Foundation in Berlin. The Science Breakthroughs are awarded in 10 categories, the first seven recipients ranging from Life Sciences and Physical Sciences to Art and Science and Science and Innovation Management.

The first laureates of the prestigious "Science Breakthrough of the Year" award are:

José-Alain Sahel, University of Pittsburgh School of Medicine & Institut de la Vision (Sorbonne Université/Inserm/CNRS)| Life Sciences
Breaking the Wall to Restoring Vision for Retinal Degeneration

Elham Fadaly and Erik Bakkers, Eindhoven University of Technology | Physical Sciences
Breaking the Wall to Light-Emitting Silicon

Francesca Santoro, Instituto Italiano di Tecnologia | Engineering and Technology
Breaking the Wall to Biohybrid Synapses

Dilip Menon, University of Witwatersrand | Social Sciences and Humanities
Breaking the Wall to a Paracolonial Paradigm

Gerhard Widmer, Johannes Kepler University Linz | Art and Science
Breaking the Wall to Computational Expressivity in Music Performance

Advanced Topics (not treated here)

- Smoothing in Kalman Filters
- Variants of the basic model:
  - The Switching Kalman Filter
  - The Extended Kalman Filter
  - The Unscented Kalman Filter
- ... and a lot of other stuff from 50 years of scientific literature.

What you should remember of this section

What you need not know by heart:

- The detailed update equations for the filtering steps
- The exact definition of the Kalman Gain
- ...

What you should understand:

- The general concept of a Kalman filter
- How a Kalman filter model is specified
- The basic idea of the Kalman filtering algorithm
- What factors influence the belief state update, and in what way
- Some of the limitations of linear system models

Literature

Cancino-Chacón, C., Peter, S., Hu, P., Karystinaios, E., Henkel, F., Foscarin, F., Varga, N., and Widmer, G. (2023).
The ACCompanion: Combining Reactivity, Robustness, and Musical Expressivity in an Automatic Piano Accompanist. In *Proceedings of the 32nd International Joint Conference on Artificial Intelligence (IJCAI 2023)*, Macao, S.A.R.

Durand, A. (2017).
Modèles probabilistes pour suivi de partition à partir de données symboliques et monophoniques. Rapport de stage d'ingénieur. Inst. of Computational Perception, Johannes Kepler University Linz, Austria.

Koller, Daphne and Friedman, Nir (2009).
Probabilistic Graphical Models: Principles and Techniques. Cambridge, MA: MIT Press.

Murphy, Kevin (1998).
Switching Kalman Filters. Technical report, DEC/Compaq Cambridge Research Labs, 1998. `https://www.cs.ubc.ca/~murphyk/Papers/skf.pdf`

Russell, Stuart J. and Norvig, Peter (2003).
Artificial Intelligence: A Modern Approach. Englewood Cliffs, NJ: Prentice Hall.