State Space Models - Filtering Basics

Contents

Filtering problems

In words, filtering problems are as follows:

- you have some sort of dynamical system (physical or engineered)
- you observe a noisy measurement of (part of) the state of the system
- you want to figure out what is the actual state of the system

Here are some examples of this:

- Object tracking:
  - you have a vehicle (e.g., a rocket) the motion of which is described by Newton's laws
  - you have some noisy measurements of the state of the system (e.g., GPS, acceleratometers)
  - you want to figure out where exactly the rocket is, so that you can control it
- Biomedical imaging:
  - you have a biological organ (e.g., a heart) which consists of tissue and fluids that can be described with physical variables (the state) such as strains, stresses, velocity, and pressure which are governed by physical equations (elasticity equations and Navier-Stokes equations)
  - you have some noisy measurements of the state of the system (e.g., accoustic doppler shifts, average velocity in small voxels (MRI))
  - you want to figure out the underlying physical states because they may be indicative of disease

In what follows we are going to define the filtering problem mathematically. To achieve this, we first need to explain the Markov property. **Note:** All the material in these notes is explained in the video lectures as well. Feel free to skip the following sections if they seem very difficult at the moment. You can revisit them later.

Markov model without external disturbances

Assume that you have some sort of a stochastic dynamical system with state variables $\mathbf{x}_t$. For simplicity, assume also that the time runs in discrete steps, $t = 0, 1, 2, \ldots, n$. If the system doesn't have any external disturbances, i.e., if the system is closed, then the following property is valid:

$$p(\mathbf{x}_{t+1}|\mathbf{x}_{0:t}) = p(\mathbf{x}_{t+1}|\mathbf{x}_t).$$

This simply says that the probability of the state at time $t+1$ given all the states up to the timestep t depends only on the state of the system at time t. That is, the past states before t do not affect the future states directly. This property is the *Markov property* and $p(\mathbf{x}_{t+1}|\mathbf{x}_t)$ is called the *transition probability*.

If you have the transition probability, you can write the joint probability density of all states as follows:

$$p(\mathbf{x}_{0:n}) = p(\mathbf{x}_0) \prod_{t=0}^{n-1} p(\mathbf{x}_{t+1}|\mathbf{x}_t),$$

See proof below for justification

where $p(\mathbf{x}_0)$ is the probability density of the initial conditions $\mathbf{x}_0$. Once you spend a bit of time on this formula, you will start seeing why its valid. The proof goes like this. Let's start with the special case $n = 1$. From the product rule we have that:

$$p(\mathbf{x}_{0:1}) = p(\mathbf{x}_0)p(\mathbf{x}_1|\mathbf{x}_0),$$ --> directly from Bayes

which is the desired formula. For $n = 2$, we have by applying the product rule again: https://en.wikipedia.org/wiki/Bayes'_theorem

$$p(\mathbf{x}_{0:2}) = p(\mathbf{x}_{0:1})p(\mathbf{x}_2|\mathbf{x}_{0:1}).$$ --> also in the form of Bayes

Now, we can use the formula we derived for $n = 1$ and the Markov property to write this as:

$$p(\mathbf{x}_{0:2}) = p(\mathbf{x}_0)p(\mathbf{x}_1|\mathbf{x}_0)p(\mathbf{x}_2|\mathbf{x}_1).$$ substitute, expand, and use Markov property

So the formula holds for $n = 2$. Similarly, you can prove this for an arbitrary n assuming that it holds for $n - 1$. Then the proof is complete.

Markov model with observations

Let's now add some observations to our system. We assume that we have some sensors that measure something on every timestep. Say that they measure the variables $\mathbf{y}_t$. The assumption is that $\mathbf{y}_t$ can only depend on the state of the system $\mathbf{x}_t$ and on nothing else. Mathematically, the assumption is: $p(\mathbf{y}_t|\mathbf{x}_{0:t}) = p(\mathbf{y}_t|\mathbf{x}_t). So, to describe the statistics of \mathbf{y}_t we need to define the so-called *emission probability* p(\mathbf{y}_t|\mathbf{x}_t)$.   *emission probability*
$Now, we can write the joint probability density of the system states and the observations as:$
$p(\mathbf{x}_{0:n}, \mathbf{y}_{1:n}) = p(\mathbf{x}_0) \prod_{t=0}^{n-1} p(\mathbf{x}_{t+1}|\mathbf{x}_t)p(\mathbf{y}_t|\mathbf{x}_t).$ This is an important formula that you should spend some time on. You can prove it by induction very easily.

Markov model with observations and controls

Now assume that at every timestep t we can pass a control command to the system $\mathbf{u}_t$. For example, if we are dealing with a rocket, we can decide which thrusters to activate. The control command will affect where the system state goes in the next timestep. We say that the system is Markovian if the following equation holds:

$$p(\mathbf{x}_{t+1}|\mathbf{x}_{0:t}, \mathbf{u}_{0:t}) = p(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{u}_t).$$

As before, the probability density $p(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{u}_t)$ is known as the *transition probability*. Remember that the entire history of controls $\mathbf{u}_{0:n-1}$ is known (we pick it). So, when we write down the joint probability density of states and observations we can condition on it. We have:

$$p(\mathbf{x}_{0:n}, \mathbf{y}_{1:n}|\mathbf{u}_{0:n-1}) = p(\mathbf{x}_0) \prod_{t=0}^{n-1} p(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{u}_t)p(\mathbf{y}_t|\mathbf{x}_t).$$ recall that x_0 are the initial conditions

The proof is similar to the case without any controls.

Mathematical definition of the filtering problem

We have all the ingredients to define the filtering problem mathematically. The problem is to characterize the states $\mathbf{x}_{0:n}$ given the available observations $\mathbf{y}_{1:n}$ and the applied controls $\mathbf{u}_{0:n-1}$. The best you can say about the states comes from applying Bayes' rule. It is:

$$p(\mathbf{x}_{0:n}|\mathbf{y}_{1:n}, \mathbf{u}_{0:n-1}) = \frac{p(\mathbf{x}_{0:n}, \mathbf{y}_{1:n}|\mathbf{u}_{0:n-1})}{p(\mathbf{y}_{1:n}|\mathbf{u}_{0:n-1})} \propto p(\mathbf{x}_0) \prod_{t=0}^{n-1} p(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{u}_t)p(\mathbf{y}_t|\mathbf{x}_t).$$ This looks like the form for Bayes with three events

This the formal answer and it is the starting point for developing specific algorithms.

> ℹ️ **Note**
>
> Strictly speaking the filtering problem is actually to estimate the current state given all data, i.e., $p(\mathbf{x}_n|\mathbf{y}_{1:n}, \mathbf{u}_{0:n-1})$. The estimation of all states (including the past) is known as *smoothing*.

Linear systems with Gaussian probabilities

Let's now give a specific example of transition and emission probabilities. All transitions will be linear and all probabilities will be Gaussian. There are two possible ways to write this down. The equations and the probabilistic way. We will give both of them, starting from the equations.

The equations way

- **Intial conditions:**

$$\mathbf{x}_0 = \boldsymbol{\mu}_0 + \mathbf{z}_0,$$

where $\boldsymbol{\mu}_0$ is a parameter and $\mathbf{z}_0 \sim N(\mathbf{0}, \mathbf{V}_0)$ with $\mathbf{V}_0$ an properly sized covariance matrix.

- **Transitions:**

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}_t + \mathbf{z}_t,$$

where $\mathbf{A}$ and $\mathbf{B}$ are matrices and $\mathbf{z}_t \sim N(\mathbf{0}, \mathbf{Q})$ with $\mathbf{Q}$ a properly sized covariance matrix (known as process covariance).

- **Emissions:**

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{w}_t,$$

where $\mathbf{C}$ is a matrix and $\mathbf{w}_t \sim N(\mathbf{0}, \mathbf{R})$ with $\mathbf{R}$ a properly sized covariance matrix (known as process covariance).

The probabilistic way

- **Initial conditions:**

$$p(\mathbf{x}_0) = N(\mathbf{x}_0 | \boldsymbol{\mu}_0, \mathbf{V}_0),$$

- **Transitions:**

$$p(\mathbf{x}_{t+1} | \mathbf{x}_t, \mathbf{u}_t) = N(\mathbf{x}_{t+1} | \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}_t, \mathbf{Q}).$$

- **Emissions:**

$$p(\mathbf{y}_t | \mathbf{x}_t) = N(\mathbf{y}_t | \mathbf{C}\mathbf{x}_t, \mathbf{R}).$$

Note: The parameters $\mathbf{A}, \mathbf{B}, \mathbf{C}, \boldsymbol{\mu}_0, \mathbf{V}_0, \mathbf{Q}$ and $\mathbf{R}$ are assumed to be known in the context of our class. However, in realistic situations they maybe unknown (or partially known). If that's the case, they would have to be estimated along with the states. In that case, we say that we have a *parameter estimation and filtering problem.* These are much harder.

By Ilias Bilionis (ibilion[at]purdue.edu)

© Copyright 2021.