Notes on state estimation

Luca Parolini

March 12, 2017

Computing speed from the measurements of position

Consider a car and a (quite inaccurate) model of it

The car we want to model.[1]

The (quite inaccurate) model

Position at time t: $(\xi(t), \zeta(t))$
Equations of motion:

$$\ddot{\xi}(t) = u_1(t)$$
$$\ddot{\zeta}(t) = u_2(t)$$

where

- $u_1(t)$ acceleration along first coordinate
- $u_2(t)$ acceleration along second coordinate

[1] Image by Edvvc from London, UK (2015 BMW i8)

Model rewriting

We want to rewrite the model in a form that better highlights *input*, *output*, and *state* of the system

$$\ddot{\xi}(t) = u_1(t)$$
$$\ddot{\zeta}(t) = u_2(t)$$

$$\begin{bmatrix} \dot{\xi} \\ \ddot{\xi} \\ \dot{\zeta} \\ \ddot{\zeta} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \xi \\ \dot{\xi} \\ \zeta \\ \dot{\zeta} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}$$

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \xi \\ \dot{\xi} \\ \zeta \\ \dot{\zeta} \end{bmatrix}$$

Model rewriting

We want to rewrite the model in a form that better highlights *input*, *output*, and *state* of the system

$$\ddot{\xi}(t) = u_1(t)$$
$$\ddot{\zeta}(t) = u_2(t)$$

$$\begin{bmatrix} \dot{\xi} \\ \ddot{\xi} \\ \dot{\zeta} \\ \ddot{\zeta} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \xi \\ \dot{\xi} \\ \zeta \\ \dot{\zeta} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}$$

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \xi \\ \dot{\xi} \\ \zeta \\ \dot{\zeta} \end{bmatrix}$$

$\boldsymbol{x}(t)$: state
$\boldsymbol{u}(t)$: input
$\boldsymbol{y}(t)$: output

State space form

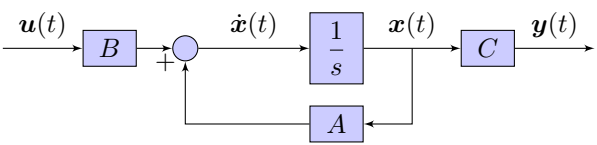
The names A, B, C are often used for the matrices in state space form. We can write

$$\dot{\boldsymbol{x}}(t) = A\boldsymbol{x}(t) + B\boldsymbol{u}(t)$$
$$\boldsymbol{y}(t) = C\boldsymbol{x}(t)$$

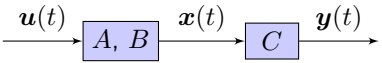
where

- $\boldsymbol{x}(t) = \begin{bmatrix} \xi \\ \dot{\xi} \\ \zeta \\ \dot{\zeta} \end{bmatrix}$ state of the system
- $\boldsymbol{y}(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix}$ measured output
- $\boldsymbol{u}(t) = \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}$ input

A graphical representation



Simplified form



Conversion to discrete time

Our estimator will be as a discrete time system. We need to convert our original model

- Δt sampling time interval
- t_0 initial time stamp
- $\boldsymbol{x}(k)$ is the k^{th} samples of the state, i.e., $\boldsymbol{x}(k) = \boldsymbol{x}(t_0 + k\Delta t)$. Similar notation for input and output
- A_d: discrete time version of the matrix A
- B_d: discrete time version of the matrix B

Continuous time

$$\dot{\boldsymbol{x}}(t) = A\boldsymbol{x}(t) + B\boldsymbol{u}(t)$$
$$\boldsymbol{y}(t) = C\boldsymbol{x}(t)$$

Discrete time

$$\boldsymbol{x}(k+1) = A_d\boldsymbol{x}(k) + B_d\boldsymbol{u}(k)$$
$$\boldsymbol{y}(k) = C\boldsymbol{x}(k)$$

Matrices for the discrete time system

Continuous time

$$\dot{\boldsymbol{x}}(t) = A\boldsymbol{x}(t) + B\boldsymbol{u}(t)$$
$$\boldsymbol{y}(t) = C\boldsymbol{x}(t)$$

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Discrete time [2]

$$\boldsymbol{x}(k+1) = A_d\boldsymbol{x}(k) + B_d\boldsymbol{u}(k)$$
$$\boldsymbol{y}(k) = C\boldsymbol{x}(k)$$

$$A_d = \begin{bmatrix} 1 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta t \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$B_d = \begin{bmatrix} \frac{\Delta t^2}{2} & 0 \\ \Delta t & 0 \\ 0 & \frac{\Delta t^2}{2} \\ 0 & \Delta t \end{bmatrix}$$

[2] For the discretization step see, e.g., G. Fredrik. *Sensor fusion* (slides)

Estimation approaches

We consider two approaches

1. *Numerical differentiation*: speed could be estimated by numerical differentiation from the measured position
2. *Observer*: we could estimate the *whole* state of the system and then, extract the speed of the car from the estimated state

1. Numerical differentiation

$$\hat{\dot{\xi}}(k) = \frac{y_1(k) - y_1(k-1)}{\Delta t}$$
$$\hat{\dot{\zeta}}(k) = \frac{y_2(k) - y_2(k-1)}{\Delta t}$$

2. Observer

1. Estimate the *whole* state of the system, $\hat{\boldsymbol{x}}(k)$
2. Extract speed information from the second and fourth elements of the state, i.e. $\hat{\dot{\xi}}(k) = \hat{x}_1(k)$, $\hat{\dot{\zeta}}(k) = \hat{x}_3(k)$

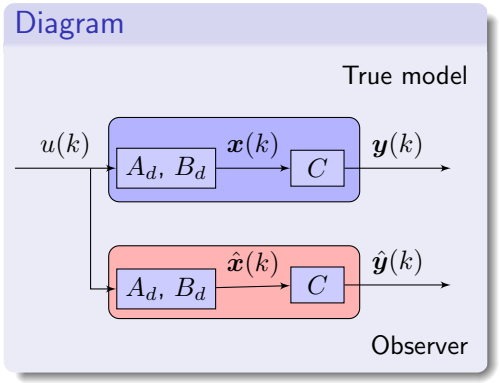
Observer derivation

Assume

- Model and parameters of the system are known
- Input and output are measured without noise
- An initial estimate of the state is known $\hat{\boldsymbol{x}}(0)$

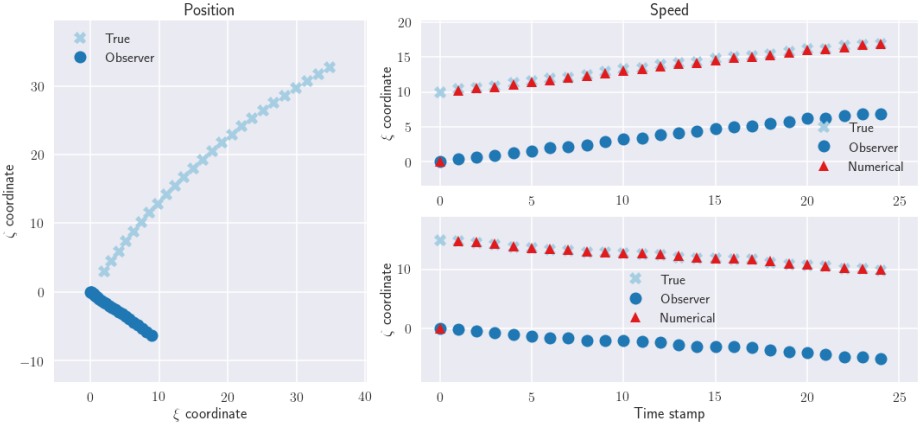
We could define the observer as

$$\hat{\boldsymbol{x}}(k+1) = A_d\hat{\boldsymbol{x}}(k) + B_d\boldsymbol{u}(k)$$
$$\hat{\boldsymbol{y}}(k) = C\hat{\boldsymbol{x}}(k)$$



Open loop observer - Example

Trends of velocities are well tracked. Input is known and system parameters are known.



Initial errors on speed and position are not corrected over time

A possible solution: Luenberger observer

Key idea: use the measured variable for correcting the estimated state

New equations of the observer

$$\hat{\boldsymbol{x}}(k+1) = A_d\hat{\boldsymbol{x}}(k) + B_d\boldsymbol{u}(k) + L(\boldsymbol{y}(k) - \hat{\boldsymbol{y}}(k)) \longrightarrow \text{Correcting term}$$

$$\hat{\boldsymbol{y}}(k) = C\hat{\boldsymbol{x}}(k)$$

A possible solution: Luenberger observer

Key idea: use the measured variable for correcting the estimated state

New equations of the observer

$$\hat{\boldsymbol{x}}(k+1) = A_d\hat{\boldsymbol{x}}(k) + B_d\boldsymbol{u}(k) + L(\boldsymbol{y}(k) - \hat{\boldsymbol{y}}(k)) \longrightarrow \text{Correcting term}$$

$$\hat{\boldsymbol{y}}(k) = C\hat{\boldsymbol{x}}(k)$$

Consider now the error on the estimated state: $\tilde{\boldsymbol{x}}(k) = \boldsymbol{x}(k) - \hat{\boldsymbol{x}}(k)$. We can write

$$\begin{aligned}
\tilde{\boldsymbol{x}}(k+1) &= \boldsymbol{x}(k+1) - \hat{\boldsymbol{x}}(k+1) \\
&= A_d\boldsymbol{x}(k) + B_d\boldsymbol{u}(k) - (A_d\hat{\boldsymbol{x}}(k) + B_d\boldsymbol{u}(k) + L(\boldsymbol{y}(k) - \hat{\boldsymbol{y}}(k))) \\
&= A_d\hat{\boldsymbol{x}}(k) - LC\boldsymbol{x}(k) - A_d\hat{\boldsymbol{x}}(k) + LC\hat{\boldsymbol{y}}(k) \\
&= (A_d - LC)\boldsymbol{x}(k) - (A_d - LC)\hat{\boldsymbol{x}}(k) \\
&= (A_d - LC)\tilde{\boldsymbol{x}}(k)
\end{aligned}$$

A possible solution: Luenberger observer

Key idea: use the measured variable for correcting the estimated state

New equations of the observer

$$\hat{\boldsymbol{x}}(k+1) = A_d\hat{\boldsymbol{x}}(k) + B_d\boldsymbol{u}(k) + L(\boldsymbol{y}(k) - \hat{\boldsymbol{y}}(k)) \longrightarrow \text{Correcting term}$$

$$\hat{\boldsymbol{y}}(k) = C\hat{\boldsymbol{x}}(k)$$

A possible solution: Luenberger observer

Key idea: use the measured variable for correcting the estimated state

New equations of the observer

$$\hat{\boldsymbol{x}}(k+1) = A_d\hat{\boldsymbol{x}}(k) + B_d\boldsymbol{u}(k) + L(\boldsymbol{y}(k) - \hat{\boldsymbol{y}}(k))$$

Correcting term

$$\hat{\boldsymbol{y}}(k) = C\hat{\boldsymbol{x}}(k)$$

Consider now the error on the estimated state: $\tilde{\boldsymbol{x}}(k) = \boldsymbol{x}(k) - \hat{\boldsymbol{x}}(k)$. We can write

$$\begin{aligned}
\tilde{\boldsymbol{x}}(k+1) &= \boldsymbol{x}(k+1) - \hat{\boldsymbol{x}}(k+1) \\
&= A_d\boldsymbol{x}(k) + B_d\boldsymbol{u}(k) - (A_d\hat{\boldsymbol{x}}(k) + B_d\boldsymbol{u}(k) + L(\boldsymbol{y}(k) - \hat{\boldsymbol{y}}(k))) \\
&= A_d\hat{\boldsymbol{x}}(k) - LC\boldsymbol{x}(k) - A_d\hat{\boldsymbol{x}}(k) + LC\hat{\boldsymbol{y}}(k) \\
&= (A_d - LC)\boldsymbol{x}(k) - (A_d - LC)\hat{\boldsymbol{x}}(k) \\
&= (A_d - LC)\,\tilde{\boldsymbol{x}}(k)
\end{aligned}$$

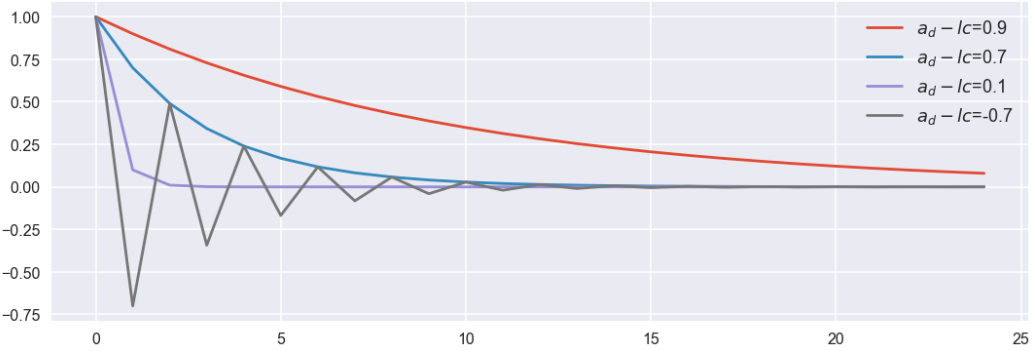
Dynamics of the estimation error

Input values have no effect on error dynamics

Exponential decay - scalar case

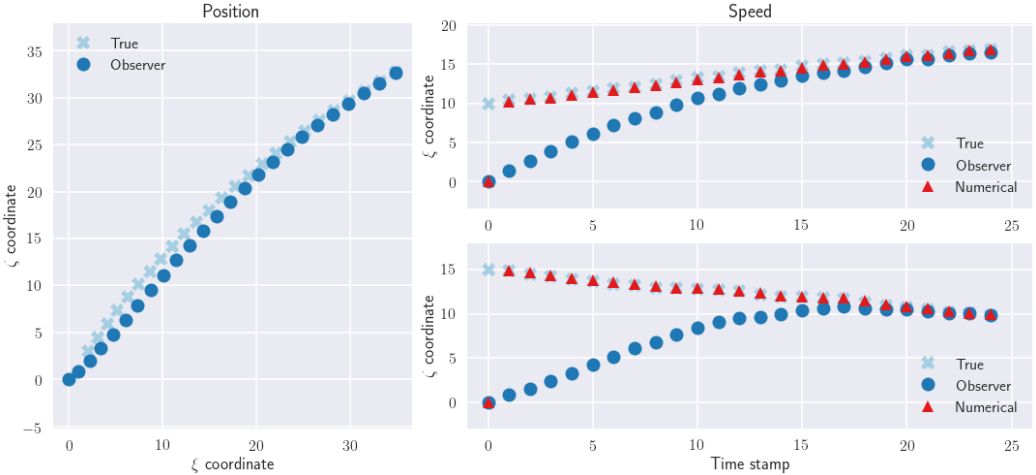
Focus on the scalar case: $\tilde{x}(k+1) = (a_d - lc)\tilde{x}(k)$

- If $|a_d - lc| < 1$, then $\tilde{x}(k)$ will tend to 0 for all values of $\tilde{x}(0)$



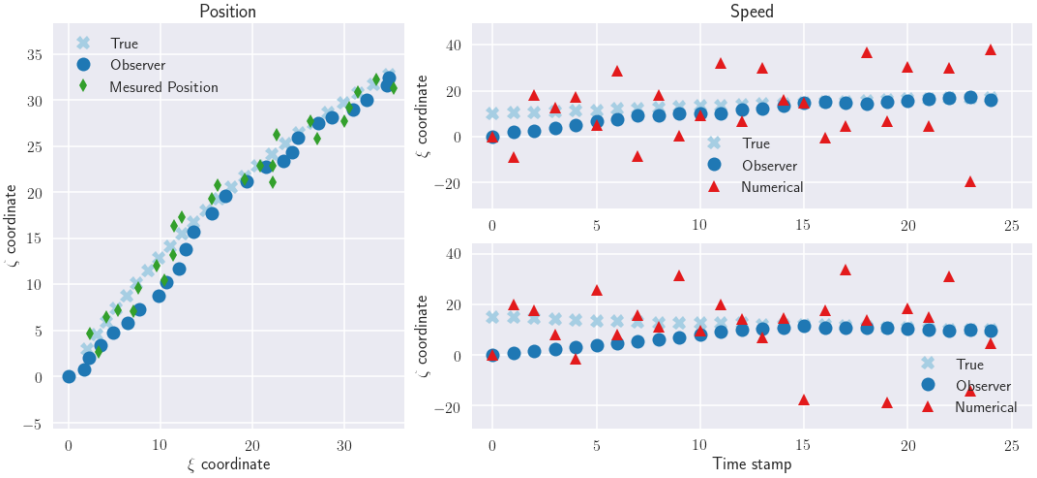
Example

True and estimated position of the car, along with the true and estimated speed



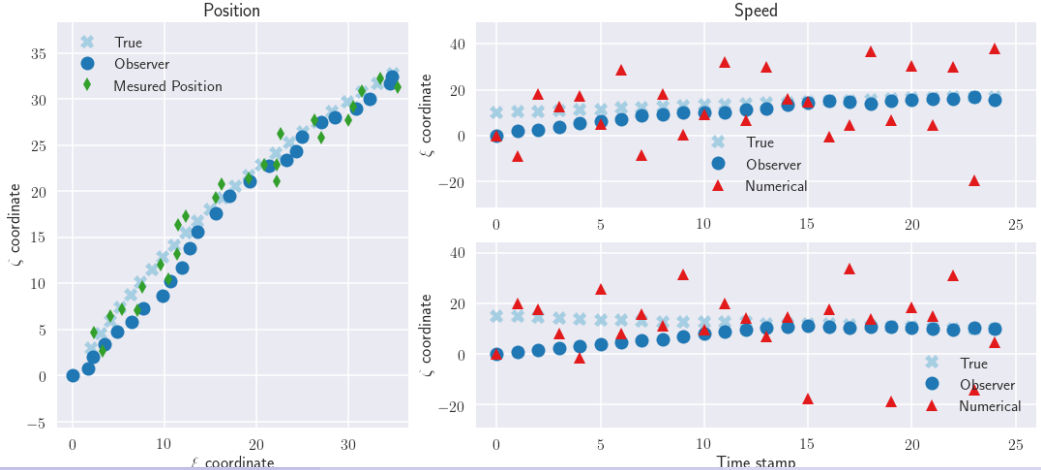
Estimation with position measurement noise

$\xi(k)$ coordinate is affected by noise uniformly distributed between -2 and 2
$\zeta(k)$ coordinate is affected by noise having Gaussian distribution with mean 0 and standard deviation 1



Estimation with position and acceleration measurement noise

$u_1(k)$ measurements are affected by noise with uniform distribution between -2 and 2
$u_2(k)$ measurements are affected by noise with Gaussian distribution with mean 0 and standard deviation 1



Summary

- Numerical difference is a viable approach only when measurement noise is negligible
- Definition of the observer requires knowledge of the system parameters
- Using feedback from the output can compensate initial estimation errors
- Multiple gains of the observer can be used
- All eigenvalues of $A_d - LC$ have absolute value smaller than 1
- For information see, among others, A. Bemporad. State estimation and linear observer

What would it be the best gain?

Random variables

Random variables

A variable quantity whose possible values depend, in random manner, on a set of random outcomes events[3]

Every random variables is defined over *probability space*: $(\Omega, \mathcal{F}, \mathcal{P})$ [4]

Consider the toss of a coin

- $\Omega = \{H, T\}$ is the set of possible outcomes. In this case, head or tail
- $\mathcal{F} = \{\{\}, \{H\}, \{T\}, \{H, T\}\}$ is the set of events we consider
- $\mathcal{P}$ probability function. It associates elements of $\mathcal{F}$ with a probability value. For example

$$\mathcal{P}(\{\}) = 0, \quad \mathcal{P}(\{H\}) = 0.5, \quad \mathcal{P}(\{T\}) = 0.5, \quad \mathcal{P}(\{H, T\}) = 1$$

[3] Wikipedia

[4] Additional information can be found, among others, in D.P. Bertsekas and J.N. Tsitsiklis. Introduction to Probability

Moments

X is a random variable , taking values in $\mathbb{R}$, and having probability density function $f(x)$

- Mean: $\mu = E[X] = \int_{\mathbb{R}} s f(s) ds$
- n^{th} moment: $\int_{\mathbb{R}} s^n f(s) ds$
- n^{th} central moment $E[(X - E[X])^n] = \int_{\mathbb{R}} (s - \mu)^n f(s) ds$
  - Variance: second central moment $E[(X - E[X])^2] = \int_{\mathbb{R}} (s - \mu)^2 f(s) ds$

Some variables have moments on infinite value. For example *heavy tailed distribution* [5]

$$f(x) = \begin{cases} \frac{1}{x^2} & x \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

[5] Wikipedia. Heavy-tailed distribution

Mean and mode (most likely outcome)

Mean and *mode* are different concept

- Mean: weighted sum of all of the possible outcomes
  - Mean value could lie outside the set of possible outcomes
- Mode: An outcome with the highest probability value

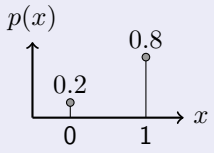
X is a random variable taking values

- 0 with probability 0.2
- 1 with probability 0.8

Mean value: $\mu = 0.2 \cdot 0 + 0.8 \cdot 1 = 0.8$

Mode: $\underset{x \in \{0,1\}}{\arg\max}\, p(x) = 1$

Probability of X



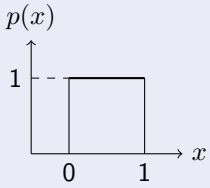
The mean value is not even an element of the possible outcomes

Sum of independent random variable

The distribution of the sum of two random variables must always be carefully computed

- X uniformly distributed between 0 and 1
- Y uniformly distributed between 0 and 1
- $Z = X + Y$ *is not* uniformly distributed between 0 and 1

Probability density functions

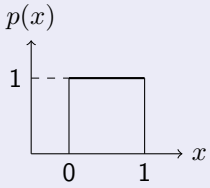


Sum of independent random variable

The distribution of the sum of two random variables must always be carefully computed

- X uniformly distributed between 0 and 1
- Y uniformly distributed between 0 and 1
- $Z = X + Y$ *is not* uniformly distributed between 0 and 1
- The distribution of Z depends on the joint distribution of X and Y

Probability density functions



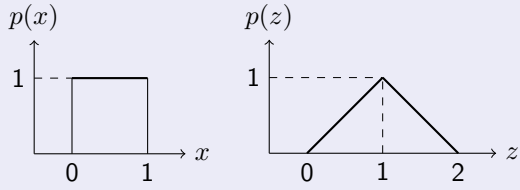
Sum of independent random variable

The distribution of the sum of two random variables must always be carefully computed

- X uniformly distributed between 0 and 1
- Y uniformly distributed between 0 and 1
- $Z = X + Y$ *is not* uniformly distributed between 0 and 1
- The distribution of Z depends on the joint distribution of X and Y
- If X and Y are independent variables, then Z has a triangular distribution
  - X and Y are independent if for all x and y

$$P(X \leq x, Y \leq y) = P(X \leq x) \cdot P(Y \leq y)$$

Probability density functions



Sum of variables normally distributed random variables

The Gaussian distribution is also called *Normal* distribution and is denoted with the symbol $\mathcal{N}$

- The probability density function of a normally distributed random variable X is

$$f_X(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^s}{2\sigma^2}}$$

Sum of variables normally distributed random variables

The Gaussian distribution is also called *Normal* distribution and is denoted with the symbol $\mathcal{N}$

- The probability density function of a normally distributed random variable X is

$$f_X(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^s}{2\sigma^2}}$$

- If $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$ and $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$ and are independent, then
- $Z = X + Y$ is also normally distributed; $Z \sim \mathcal{N}(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$

Note however that $X \cdot Y$ is *not* normally distributed

Sum of variables normally distributed random variables

The Gaussian distribution is also called *Normal* distribution and is denoted with the symbol $\mathcal{N}$

- The probability density function of a normally distributed random variable X is

$$f_X(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^s}{2\sigma^2}}$$

- If $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$ and $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$ and are independent, then
- $Z = X + Y$ is also normally distributed; $Z \sim \mathcal{N}(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$

- If $\boldsymbol{Z}$ has a multivariate normal distribution with mean $\boldsymbol{\mu}_z$ and covariance $L = E[(Z - \boldsymbol{\mu}_z)(Z - \boldsymbol{\mu}_z)^T]$ and A is a matrix, then
- The variable $\boldsymbol{S} = A\boldsymbol{Z}$ is normally distributed with mean $A\boldsymbol{\mu}_z$ and covariance matrix is ALA^T

Random variables and dynamic systems

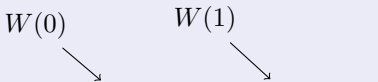
Gaussian distributions and linear systems

Assume

- $X(0)$ is normally distributed, $X(0) \sim \mathcal{N}(\mu_0, \sigma_0^2)$
- $W(k)$ is normally distributed, $W(k) \sim \mathcal{N}(\mu_{w,k}, \sigma_{w,k}^2)$
- $X(0)$ and $W(k)$ are independent for all k
- for $k \geq 0$ the following recursive equation holds: $X(k+1) = X(k) + W(k)$

What is the distribution of $X(1)$?

Graphical representation

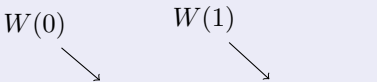


Gaussian distributions and linear systems

Assume

- $X(0)$ is normally distributed, $X(0) \sim \mathcal{N}(\mu_0, \sigma_0^2)$
- $W(k)$ is normally distributed, $W(k) \sim \mathcal{N}(\mu_{w,k}, \sigma_{w,k}^2)$
- $X(0)$ and $W(k)$ are independent for all k
- for $k \geq 0$ the following recursive equation holds: $X(k+1) = X(k) + W(k)$

Graphical representation



What is the distribution of $X(1)$?

- $X(1)$ is normally distributed (sum of two independent Gaussian variables)
- Mean $\mu_1 = \mu_0 + \mu_{w,0}$, variance $\sigma_1^2 = \sigma_0^2 + \sigma_{w,0}^2$

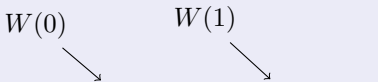
What is the distribution of $X(k)$?

Gaussian distributions and linear systems

Assume

- $X(0)$ is normally distributed, $X(0) \sim \mathcal{N}(\mu_0, \sigma_0^2)$
- $W(k)$ is normally distributed, $W(k) \sim \mathcal{N}(\mu_{w,k}, \sigma_{w,k}^2)$
- $X(0)$ and $W(k)$ are independent for all k
- for $k \geq 0$ the following recursive equation holds: $X(k+1) = X(k) + W(k)$

Graphical representation



What is the distribution of $X(1)$?

- $X(1)$ is normally distributed (sum of two independent Gaussian variables)
- Mean $\mu_1 = \mu_0 + \mu_{w,0}$, variance $\sigma_1^2 = \sigma_0^2 + \sigma_{w,0}^2$

What is the distribution of $X(k)$?

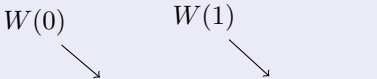
- $X(k)$ is Gaussian distributed
- Mean $\mu_k = \mu_0 + \sum_{j=0}^{k-1} \mu_{w,j}$, variance $\sigma_k^2 = \sigma_0^2 + \sum_{j=0}^{k-1} \sigma_{w,j}^2$

Non Gaussian distributions and linear systems

Assume

- $X(0)$ is uniformly distributed, $X(0) \sim \mathcal{U}(0,1)$
- $W(k)$ is uniformly distributed, $W(k) \sim \mathcal{U}(0,1)$
- $X(0)$ and $W(k)$ are independent for all k
- for $k \geq 0$ the following recursive equation holds: $X(k+1) = X(k) + W(k)$

Graphical representation



[6] See also here

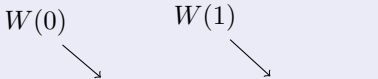
Non Gaussian distributions and linear systems

Assume

- $X(0)$ is uniformly distributed, $X(0) \sim \mathcal{U}(0,1)$
- $W(k)$ is uniformly distributed, $W(k) \sim \mathcal{U}(0,1)$
- $X(0)$ and $W(k)$ are independent for all k
- for $k \geq 0$ the following recursive equation holds: $X(k+1) = X(k) + W(k)$

What is the distribution of $X(1)$?

Graphical representation



[6] See also here

Non Gaussian distributions and linear systems

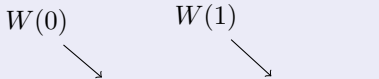
Assume

- $X(0)$ is uniformly distributed, $X(0) \sim \mathcal{U}(0,1)$
- $W(k)$ is uniformly distributed, $W(k) \sim \mathcal{U}(0,1)$
- $X(0)$ and $W(k)$ are independent for all k
- for $k \geq 0$ the following recursive equation holds: $X(k+1) = X(k) + W(k)$

What is the distribution of $X(1)$?

- $X(1)$ has a triangular distribution
- Mean $\mu_1 = \mu_0 + \mu_{w,0}$, variance $\sigma_1^2 = \sigma_0^2 + \sigma_{w,0}^2$ [6]

Graphical representation



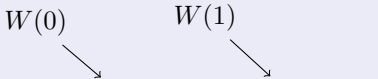
[6] See also here

Non Gaussian distributions and linear systems

Assume

- $X(0)$ is uniformly distributed, $X(0) \sim \mathcal{U}(0,1)$
- $W(k)$ is uniformly distributed, $W(k) \sim \mathcal{U}(0,1)$
- $X(0)$ and $W(k)$ are independent for all k
- for $k \geq 0$ the following recursive equation holds: $X(k+1) = X(k) + W(k)$

Graphical representation



What is the distribution of $X(1)$?

- $X(1)$ has a triangular distribution
- Mean $\mu_1 = \mu_0 + \mu_{w,0}$, variance $\sigma_1^2 = \sigma_0^2 + \sigma_{w,0}^2$ [6]

What is the distribution of $X(k)$?

- The distribution of $X(k)$ depends on the distribution of $X(k-1)$ and of $W(k-1)$
- Mean $\mu_k = \mu_0 + \sum_{j=0}^{k-1} \mu_{w,j}$, variance $\sigma_k^2 = \sigma_0^2 + \sum_{j=0}^{k-1} \sigma_{w,j}^2$

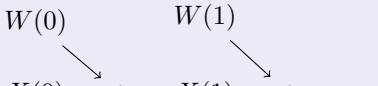
[6] See also here

Gaussian distributions and non linear systems

Assume

- $X(0)$ is normally distributed $X(0) \sim \mathcal{N}(0,1)$
- $W(k)$ is normally distributed $W(k) \sim \mathcal{N}(0,1)$
- $X(0)$ and $W(k)$ are independent for all k
- for $k \geq 0$ we have $X(k+1) = X^2(k) + W(k)$

Graphical representation



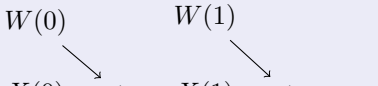
Gaussian distributions and non linear systems

Assume

- $X(0)$ is normally distributed $X(0) \sim \mathcal{N}(0, 1)$
- $W(k)$ is normally distributed $W(k) \sim \mathcal{N}(0, 1)$
- $X(0)$ and $W(k)$ are independent for all k
- for $k \geq 0$ we have $X(k+1) = X^2(k) + W(k)$

What is the distribution of $X(1)$?

Graphical representation



Gaussian distributions and non linear systems

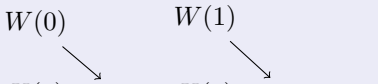
Assume

- $X(0)$ is normally distributed $X(0) \sim \mathcal{N}(0,1)$
- $W(k)$ is normally distributed $W(k) \sim \mathcal{N}(0,1)$
- $X(0)$ and $W(k)$ are independent for all k
- for $k \geq 0$ we have $X(k+1) = X^2(k) + W(k)$

What is the distribution of $X(1)$?

- $X(1)$ is not normally distributed.
- Mean $\mu_1 = E[X(0)^2] + \mu_{w,0}$
- Variance $\sigma_1^2 = var(X(0)) + \sigma_{w,0}^2$

Graphical representation



Few remarks

- Mean of the sum is always equal to the sum of the means (linear operator)
- If two random variables are independent, then the variance of the sum is equal to the sum of the variances
- The knowledge of mean and variance rarely completely characterize the distribution of a random variable

Marginal and conditional density functions

Marginal and conditional density functions

Assume X and Y are two random variables taking values in the interval $[0, 1]$ and $f_{X,Y}(x, y)$ is the joint probability density function of X and Y

- Marginal probability density functions are given by

$$f_X(x) = \int_0^1 f_{X,Y}(x, z)dz \ , \qquad f_Y(y) = \int_0^1 f_{X,Y}(z, y)dz$$

Marginal and conditional density functions

Assume X and Y are two random variables taking values in the interval $[0, 1]$

and $f_{X,Y}(x, y)$ is the joint probability density function of X and Y

- Marginal probability density functions are given by

$$f_X(x) = \int_0^1 f_{X,Y}(x, z)dz \ , \qquad f_Y(y) = \int_0^1 f_{X,Y}(z, y)dz$$

- Conditional probability density function: $f_{X|Y}(x; y) = \dfrac{f_{X,Y}(x, y)}{f_Y(y)}$, if $f_Y(x) > 0$

Marginal and conditional density functions

Assume X and Y are two random variables taking values in the interval $[0, 1]$

and $f_{X,Y}(x, y)$ is the joint probability density function of X and Y

- Marginal probability density functions are given by

$$f_X(x) = \int_0^1 f_{X,Y}(x, z)dz \ , \qquad f_Y(y) = \int_0^1 f_{X,Y}(z, y)dz$$

- Conditional probability density function: $f_{X|Y}(x; y) = \dfrac{f_{X,Y}(x, y)}{f_Y(y)}$, if $f_Y(x) > 0$
- Conditional mean and variance of X given Y

$$E[X|Y = y] = \int_0^1 z f_{X|Y}(z; y)dz, \quad var(X|Y = y) = \int_0^1 (z - E[X|Y = y])^2 f_{X|Y}(z; y)dz$$

Conditional distribution and expected conditional risk

Consider the case of estimating the value of a variable X from its measurement Y

- For example $Y = X + W$, where W is random noise independent of X
- The conditional distribution of $X|Y$ provides information about the probability of different values of X given that $Y = y$

Conditional distribution and expected conditional risk

Consider the case of estimating the value of a variable X from its measurement Y

- For example $Y = X + W$, where W is random noise independent of X
- The conditional distribution of $X|Y$ provides information about the probability of different values of X given that $Y = y$

How could we select a best value for X?

Conditional distribution and expected conditional risk

Consider the case of estimating the value of a variable X from its measurement Y

- For example $Y = X + W$, where W is random noise independent of X
- The conditional distribution of $X|Y$ provides information about the probability of different values of X given that $Y = y$
- For selecting a best value, we need a cost function, e.g., the *expected conditional risk*

$$R(t, y) = E[c(t - X)|Y = y] = \int_{\mathbb{R}} c(t - z) f_{X|Y}(z; y) dz$$

If

- $c(t - x)$ is symmetric, e.g., $c(z - x) = (z - x)^2$ and
- the conditional density function $f_{X|Y}(z; y)$ is symmetric around $E(X|Y = y)$,

then

- the minimum of $R(t, y)$ does not depends on the specific cost function $c(t - X)$
- *the minimum of* $R(t, y)$ *is given by* $E[X|Y = y]$

The Maximum a posteriori estimator

Another approach for selecting a best value for the estimation of X given $Y = y$ could be to take the maximum of $f_{X|Y}(z; y)$

- We can define
$$\hat{x}(y) = \arg\max_{z \in \mathbb{R}} f_{X|Y}(z; y)$$
- This estimator is called *Maximum a posteriori* (MAP)
- For unimodal and symmetric distributions, e.g., Gaussian distribution, then MAP estimator and the expected conditional value estimator coincide

In general, MAP and the expected conditional value are different

Kalman filter

State estimation and dynamical system

Consider once again the problem of estimating the speed of the car from the measurements of its position

$$\hat{\boldsymbol{x}}(k+1) = A_d\hat{\boldsymbol{x}}(k) + B_d\boldsymbol{u}(k) + B_d\boldsymbol{w}(k)$$
$$\hat{\boldsymbol{y}}(k) = C\hat{\boldsymbol{x}}(k) + \boldsymbol{v}(k)$$

where $\boldsymbol{w}(k)$ and $\boldsymbol{v}(k)$ are two random processes

We want to find an estimate of $\boldsymbol{x}(k)$ given the sequence of measurements $\mathcal{Y}_t = \{\boldsymbol{y}(0), \boldsymbol{y}(1), \ldots, \boldsymbol{y}(k)\}$

State estimation and dynamical system

Consider once again the problem of estimating the speed of the car from the measurements of its position

$$\hat{\boldsymbol{x}}(k+1) = A_d\hat{\boldsymbol{x}}(k) + B_d\boldsymbol{u}(k) + B_d\boldsymbol{w}(k)$$
$$\hat{\boldsymbol{y}}(k) = C\hat{\boldsymbol{x}}(k) + \boldsymbol{v}(k)$$

where $\boldsymbol{w}(k)$ and $\boldsymbol{v}(k)$ are two random processes

We want to find an estimate of $\boldsymbol{x}(k)$ given the sequence of measurements $\mathcal{Y}_t = \{\boldsymbol{y}(0), \boldsymbol{y}(1), \ldots, \boldsymbol{y}(k)\}$
A possible approach

- Define an optimality criterion
- For all t
  - Compute the conditional density function of $X|\mathcal{Y}_t$
  - Select the "best" value for X given $\mathcal{Y}_t$ based on the optimal criterion

Problems with the proposed approach

Typically not feasible to compute $f_{X|\mathcal{Y}_t}(x;y)$ at every t

The problem is that $\mathcal{Y}_t$ grows over time

- At time 0, we need to compute $f_{X|Y_0}(x;y)$
- At time 1, we need to compute $f_{X|\{Y_1,Y_0\}}(x;y)$
- $\dots$

We can simplify a little this problem by relying on conditional independence among variables

The particle filter aims at approximating $f_{X|Y_t}(x;y)$

For some particular case however, the computation of $f_{X|Y_t}(x;y)$ is relatively simple

A particular case

Our system is linear, this is already a very particular case.

$$\hat{\boldsymbol{x}}(k+1) = A_d\hat{\boldsymbol{x}}(k) + B_d\boldsymbol{u}(k) + B_d\boldsymbol{w}(k)$$
$$\hat{\boldsymbol{y}}(k) = C\hat{\boldsymbol{x}}(k) + \boldsymbol{v}(k)$$

Further assume

- $\boldsymbol{w}(k)$ is a white noise process and for all k $\boldsymbol{w}(k) \sim \mathcal{N}(0, Q)$
- $\boldsymbol{v}(k)$ is a white noise process and for all k $\boldsymbol{v}(k) \sim \mathcal{N}(0, R)$
- $\boldsymbol{w}(k)$ and $\boldsymbol{v}(k)$ are correlated with correlation matrix S, i.e.

$$E\left[\begin{bmatrix}\boldsymbol{w}(k)\\ \boldsymbol{v}(k)\end{bmatrix}\begin{bmatrix}\boldsymbol{w}(k)^T & \boldsymbol{v}(k)^T\end{bmatrix}\right] = \begin{bmatrix}Q & S\\ S^T & R\end{bmatrix}$$

- An initial estimate of $\boldsymbol{x}(0)$ is available
- The initial estimate is normally distributed with mean $\boldsymbol{x}(0)$ and covariance $P(0)$

Equations of the Kalman filter

The Kalman filter

Developed by Rudolf Emil Kálmán (May 19, 1930 - July 2, 2016)
Typically divided into two steps: prediction and update
Multiple implementations are possible, with different numerical accuracy/complexity trade-off [7]

- Prediction step:

$$\hat{\boldsymbol{x}}(k|k-1) = A_d\hat{\boldsymbol{x}}(k-1|k-1) + SR^{-1}\boldsymbol{y}(k-1) + B_d\boldsymbol{u}(k-1)$$
$$P(k|k-1) = A_dP(k-1|k-1)A_d^T + (Q - SR^{-1}S^T)$$

- Update step:

$$S(k) = CP(k|k-1)C^T + R$$
$$K(k) = P(k|k-1)C^TS(k)^{-1}$$
$$P(k|k) = (I - K(k)C)P(k|k-1)(I - K(k)C)^T + K(k)RK(k)^T$$
$$\boldsymbol{\epsilon}(k) = \boldsymbol{y}(k) - C\hat{\boldsymbol{x}}(k|k-1)$$
$$\hat{\boldsymbol{x}}(k|k) = \hat{\boldsymbol{x}}(k|k-1) + K(k)\boldsymbol{\epsilon}(k)$$

[7] See for example G. Welch, G. Bishop. "An Introduction to the Kalman Filter"

Some properties of the Kalman filter

- Under the previous assumptions, the Kalman filter is optimal
  - Even among non linear filters
  - The couple $(\hat{\boldsymbol{x}}(k|k), P(k|k))$ completely describes $f_{X|Y_k}(x; y)$
- If $w(k)$ and $v(k)$ are not normally distributed, then the Kalman filter is the best *linear* filter
  - Some other non linear filter could estimate of the state with lower uncertainty
- The covariance of the state estimate and the gain of the filter *do not depend on measurements*
  - They could be pre-computed
- Under mild hypothesis
  - The Kalman gain converges to a constant $\bar{K}$
  - All the eigenvalues of $A - \bar{K}C$ have absolute value smaller than 1

Some points to consider

The Kalman filter is a recursive algorithm. Its implementation is often relatively simple

We need however, to make sure that numerical errors remain negligible between on step and the following

Some of the approaches used for improving the numerical stability of the algorithm are

- Pre and post-scaling
- Square root implementation of the Kalman filter [8]

Additionally, when debugging a software implementation of the filter, consider that

- The evolution of the Kalman filter depends on both the estimated state and its covariance
- Both elements should be visualized when debugging the implementation

[8] See for example hier

Estimation on non-linear filters

Many approaches exist. The goal is always to reach a good-enough approximation of $f_{X|Y_t}$

Among recursive filters, often EKF and UKF are used

- If possible, prefer UKF over EKF
- In both cases, the filter gain will likely depend on the estimated state value
  - The gain of the filter cannot be computed off-line
- Verify that the state remains within "realistic" regions of the state space
  - Often, additional steps to constraint the estimate state are used[9]

[9] See e.g., D. Simon "Kalman filtering with state constraints: a survey of linear and nonlinear algorithms"
S. Kolåsa et al. "Constrained nonlinear state estimation based on the UKF approach"