PHYSICAL HUMAN-ROBOT INTERACTION

Statistical filtering

Riccardo Muradore

UNIVERSITÀ di VERONA
Dipartimento di INFORMATICA

Outline

Statistical Estimation

Sampled-data systems
(State-space representation)

Speed estimation

Random variables

Minimum Variance Estimation

Stochastic models

Kalman filtering

PROJECT

Kalman smoother

PROJECT

Statistical Estimation

Given the measurement equation

$$y(t) = s(t) + n(t)$$

where:

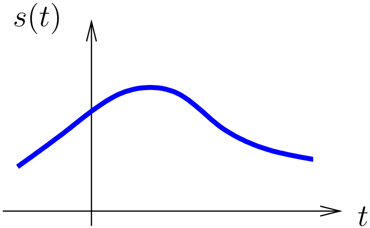
- $y(t)$ is the measurement at time t (ex. encoder)
- $s(t)$ is the signal we are interested in (ex. angular position)
- $n(t)$ is the additive measurement noise

We will tackle the following estimation problems:

- **Filtering**: optimal estimate $\hat{s}(t)$ of $s(t)$ using measurements $y(\cdot)$ till time t (i.e. $y(0), y(1), \ldots, y(t)$) --> causal signal (I'm using data until time t, current output does not depend on the future values, but only on the past values)
- **Prediction**: optimal estimate $\hat{s}(t+h)$ of $s(t+h)$ measurements y till instant t -> causal
- **Smoothing**: optimal estimate $\hat{s}(t-h)$ of $s(t-h)$ measurements y till instant t
 smoothing è per forza offline perchè ha bisogno di dati futuri --> non è causale
 ho registrato l'intero tempo fino a t e per stimare t-2sec utilizzo anche le misurazioni successive (=>uso tutte le misurazioni)

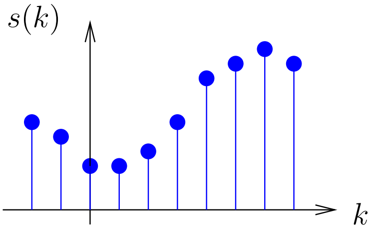
Continuous-time signal

$t \in \mathbb{R}$



Discrete-time signal

$t = kT_s, \quad k \in \mathbb{Z}$



Let θ be the angular position and ω be the corresponding angular velocity.

Assumption: When no information are available about the physical system that produces the signal $\theta(t)$, we set the derivative of the velocity $\omega(t)$ equal to a white stochastic process $w(\cdot)$

★ random walk of order 1

A stochastic process $w(t)$ is called *white noise* if its values $w(t_i)$ and $w(t_j)$ are uncorrelated $\forall i \neq j$, i.e.

$$\mathrm{Corr}\{w(t_i), w(t_j)\} = Q(t_i)\delta(t_i - t_j)$$

Speech estimation – Kinematic model

Let θ be the angular position and ω be the corresponding angular velocity.

Assumption: When no information are available about the physical system that produces the signal $\theta(t)$, we set the derivative of the velocity $\omega(t)$ equal to a white stochastic process $w(\cdot)$

★ random walk of order 1

A stochastic process $w(t)$ is called *white noise* if its values $w(t_i)$ and $w(t_j)$ are uncorrelated $\forall i \neq j$, i.e.

$$\mathrm{Corr}\{w(t_i), w(t_j)\} = Q(t_i)\delta(t_i - t_j)$$

We also assume the $w(t)$ is *Gaussian* with zero-mean and constant variance Q for all t.

The *kinematic model* is

$$\begin{aligned} \dot{\theta}(t) &= \omega(t) \\ \dot{\omega}(t) &= w(t) \end{aligned}$$

and the measurement equation is

$$y(t) = \theta(t) + v(t)$$

where $v(\cdot)$ is a white stochastic process and $v(t)$ is a Gaussian random variable with zero mean and variance equal to R.

Speed estimation – Kinematic model

Let $x(t) := \begin{bmatrix} \theta(t) \\ \omega(t) \end{bmatrix}$ be the *vector state* of the continuous-time state model

$$\begin{aligned} \dot{x}(t) &= \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w(t) \\ y(t) &= \begin{bmatrix} 1 & 0 \end{bmatrix} x(t) + v(t) \end{aligned}$$

Measurements are given at constant rate (i.e. *uniform sampling*): every T_s seconds a new measurements is available

By writing the *discretized* actual time as $t = kT_s$ within the measurement equation we get

$$y(kT_s) = \begin{bmatrix} 1 & 0 \end{bmatrix} x(kT_s) + v(kT_s)$$

or, using a more compact notation,

$$y_k = \begin{bmatrix} 1 & 0 \end{bmatrix} x_k + v_k$$

where the *sample time* T_s is implicit.

Sampled-data systems (State-space representation)

How can we compute the discrete approximation of the dynamic equation?

$$\Sigma_c : \begin{cases} \dot{x}(t) & = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w(t) \\ \\ y(t) & = \begin{bmatrix} 1 & 0 \end{bmatrix} x(t) + v(t) \end{cases}$$

How can we compute the discrete approximation of the dynamic equation?

$$\Sigma_c : \begin{cases} \dot{x}(t) &= \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w(t) \\ y(t) &= \begin{bmatrix} 1 & 0 \end{bmatrix} x(t) + v(t) \end{cases} \qquad \Sigma_d : \begin{cases} x_{k+1} &= \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix} x_k + \begin{bmatrix} ? \\ ? \end{bmatrix} w_k \\ y_k &= \begin{bmatrix} 1 & 0 \end{bmatrix} x_k + v_k \end{cases}$$

Let Σ be a LTI system in its state space representation

$$\Sigma : \begin{cases} \dot{x}(t) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t) + Du(t) \end{cases} \qquad x(t_0) = x_0 \quad , \qquad t \geq t_0 \geq 0.$$

The time evolution of the state $x(t)$, $t \geq t_0$ is given by the sum of the free evolution $x_l(t)$ (i.e. initial conditions x_0) and of the forced evolution $x_f(t)$ (i.e. input $u(t)$)

$$\Sigma_l \quad : \quad \begin{cases} \dot{x}_l(t) &= Ax_l(t) \\ x_l(t_0) &= x_0 \end{cases} , \qquad t \geq t_0 \geq 0$$

$$\Sigma_f \quad : \quad \begin{cases} \dot{x}_f(t) &= Ax_f(t) + Bu(t) \\ x_f(t_0) &= 0 \end{cases} , \qquad t \geq t_0 \geq 0$$

explicitly,

$$\begin{aligned} x(t) &= x_l(t) + x_f(t) \\ &= e^{A(t-t_0)}x(t_0) + \int_{t_0}^{t} e^{A(t-\tau)}Bu(\tau)d\tau, \qquad t \geq t_0 \geq 0 \end{aligned}$$

The state at $x(t)$

$$t = (k+1)T_s, \qquad t_0 = kT_s$$

is given by

$$\begin{aligned} x((k+1)T_s) &= e^{AT_s}x(kT_s) + \int_{kT_s}^{(k+1)T_s} e^{A((k+1)T_s-\tau)}Bu(\tau)d\tau \\ &= e^{AT_s}x(kT_s) + \int_0^{T_s} e^{A\tau'}Bu(\tau')d\tau' \end{aligned}$$

where in the last expression the change of variable $\tau' = (k+1)T_s - \tau$ has been exploited. Assuming the input $u(t)$ constant between kT_s and $(k+1)T_s$, we get

$$x((k+1)T_s) = e^{AT_s}x(kT_s) + \left(\int_0^{T_s} e^{A\tau'}Bd\tau'\right)u(kT_s)$$

If B is time-invariant, we have

$$x((k+1)T_s) = e^{AT_s}x(kT_s) + \left(\int_0^{T_s} e^{A\tau'}d\tau'\right)Bu(kT_s)$$

and, finally,

$$\begin{aligned} x_{k+1} &= \underbrace{e^{AT_s}}_{\triangleq A_d} x_k + \underbrace{\left(\int_0^{T_s} e^{A\tau'}d\tau'\right)B}_{\triangleq B_d} u_k \\ &= A_d x_k + B_d u_k \end{aligned}$$

The measurement equation does not change (it is an algebraic expression)

$$\begin{aligned} y_k &= C_d x_k + D_d u_k \\ &= C x_k + D u_k \end{aligned}$$

Speed estimation

In the speed estimation case, the dynamic equation is

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w(t)$$

and so the matrices A_d, B_d of the discrete-time approximation are

$$A_d \triangleq e^{AT_s} = e^{\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} T_s} = \begin{bmatrix} 1 & T_s \\ 0 & 1 \end{bmatrix}$$

$$B_d \triangleq \left(\int_0^{T_s} e^{A\tau'} d\tau' \right) B = \left(\int_0^{T_s} e^{\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \tau'} d\tau' \right) \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{T_s^2}{2} \\ T_s \end{bmatrix}$$

the discrete time matrices depend on the sampling time Ts
if you compute the discrete time version of the same continuos time system with different sampling times, you end up with different Ad Bd matrices

The overall discrete-time model is

problem (why we did this)
from measurement yk we want to estimate the state x

$$y_k \longrightarrow \hat{x}_k = \begin{bmatrix} \hat{\theta}_k \\ w_k \end{bmatrix}$$

$$\begin{cases} x_{k+1} = \begin{bmatrix} 1 & T_s \\ 0 & 1 \end{bmatrix} x_k + \begin{bmatrix} \frac{T_s^2}{2} \\ T_s \end{bmatrix} w_k \\ y_k = \begin{bmatrix} 1 & 0 \end{bmatrix} x_k + v_k \end{cases}$$

if I know the state xk then I have an approximtion of theta (that I already measured but if I have an estimation using the kalman filter(for example) I can filter out the noise) and I have an estimation on the angular velocity that was completely unknown

Tuning

The variance matrix R (*noise variance*) depends on the sensors whereas the variance matrix Q (*model variance*) is chosen in order to “explain” the measurements as well as possible.

Speed estimation – Kinematic model

prima rivedere ultimi 5 minuti lezione 24 ottobre phri

please do this

Exercise

Compute the discrete-time approximation of the system

$$
\begin{aligned}
\dot{x}(t) &= \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} w(t) \\
y(t) &= \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} x(t) + v(t)
\end{aligned}
$$

where the state vector $x(t)$ has as components *position, velocity, acceleration*.

★ random walk of order 2

please do this

Exercise

Compute the state space model of the second order system (mass–damper–spring)

$$m\ddot{y} + b\dot{y} + ky = \tau$$

and derive its discrete-time approximation.

Hint.

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{b}{m} \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ \frac{1}{m} \end{bmatrix} \tau(t)$$

Given the measurement y_k, $k = 0, 1, \ldots$ related to an unknown state variable x_k, we are interested in the following estimation problems

- **Filtering**

$$y_0, y_1, \ldots, y_k \longrightarrow \hat{x}_{k|k} \qquad \longrightarrow \hat{\omega}_{k|k}$$

- ***h*-step ahead Prediction**

$$y_0, y_1, \ldots, y_k \longrightarrow \hat{x}_{k+h|k} \qquad \longrightarrow \hat{\omega}_{k+h|k}$$

- ***h*-step backward Smoothing**

$$y_0, y_1, \ldots, y_k \longrightarrow \hat{x}_{k-h|k} \qquad \longrightarrow \hat{\omega}_{k-h|k}$$

- **Smoothing**

$$y_0, y_1, \ldots, y_N \longrightarrow \hat{x}_{k|N} \qquad \longrightarrow \hat{\omega}_{k|N}$$

Random variables

Definition (random variable (r.v.))

A random variable $\mathbf{x} : \Omega \to E$ is a measurable function from the set of possible outcomes Ω to some set E. Ω is a probability space and E is a measurable space.

Roughly speaking: A random variable $\mathbf{x}$ is a rule for assigning to every outcome ω of an experiments a number $\mathbf{x}(\omega)$

Definition (stochastic process)

Given a probability space $(\Omega, \mathcal{F}, P)$ and a measurable space (S, Σ), an S-valued stochastic process is a collection of S-valued random variables on Ω, indexed by a totally ordered set T ("time"). That is, a stochastic process is a collection $\{\mathbf{x}_t : t \in T\}$ where each $\mathbf{x}_t$ is an S-valued random variable on Ω. The space S is then called the state space of the process.

Roughly speaking: A stochastic process $\mathbf{x}_t$ is a rule for assigning to every outcome ω a function $\mathbf{x}(t, \omega)$

From r.v. to stochastic processes

$\{\mathbf{x}_t\}$ has the following interpretations:

- It is a family of functions $\mathbf{x}_t(\omega)$ when t and ω are variables.
- It is a single time function (or a realization of the given process) $\mathbf{x}_t(\bar{\omega})$ when t is a variable and $\omega = \bar{\omega}$ is fixed.
- It is a random variable if $t = \bar{t}$ is fixed and ω is variable, i.e. $\mathbf{x}_{\bar{t}}(\omega)$ state of the process at time t.
- It is a number if t and ω are fixed

If $T = \mathbb{R}$, $\{\mathbf{x}_t\}$ is a continuous-time process

If $T = \mathbb{Z}$, $\{\mathbf{x}_k\}$ is a discrete-time process

Even though the dynamics of the system is described by ODE, in the following we will consider discrete-time processes because the sensing system provides measurements at discrete moments.

Remark The r.v. $\mathbf{x}_{\bar{k}}(\omega)$ can be continuous even if $k \in T = \mathbb{Z}$

Minimum Variance Estimation

Gaussian filters

Gaussian filters assume that the undergoing phenomena can be modeled by Gaussian distributions.

This assumption allows to solve in recursive way the general Bayes filters' formulation

Why are Gaussian distributions so good?

- Gaussians are *unimodal*: they have a single maximum
- The statistics (mean, variance and higher order moments) of a Gaussian are described by two parameters: its *mean* μ *and variance* Σ
- The *linear combination* of Gaussians is still Gaussian

Definition (Gaussian r.v.)

An n-dimensional random variable $\mathbf{x}$ is Gaussian with mean $\mu \in \mathbb{R}^n$ and variance $\Sigma \in \mathbb{R}^{n \times n}, \Sigma = \Sigma^T > 0$, $\mathbf{x} \sim \mathcal{N}(\mu, \Sigma)$, if its probability density function (PDF) is given by

$$p(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}.$$

This means that

$$\begin{aligned} \mu &= \mathbb{E}[\mathbf{x}] \\ \Sigma &= \mathrm{Var}(\mathbf{x}) = \mathbb{E}[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])^T] \end{aligned}$$

where $\mathbb{E}[\,\cdot\,]$ is the expectation operator.

Theorem (Joint Gaussian r.v.)

Let $\mathbf{x} \in \mathbb{R}^n$ *and* $\mathbf{y} \in \mathbb{R}^m$ *be joint Gaussian*

$$p(x,y) \sim \mathcal{N}\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}\right)$$

Then

- *the r.v.* $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y}$ *is still Gaussian, i.e.* $\mathbf{z} \sim \mathcal{N}(\mu_z, \Sigma_z)$, *where*

$$\begin{aligned}
\mu_z &= \mathbb{E}[A\mathbf{x} + B\mathbf{y}] = A\mu_x + B\mu_y \\
\Sigma_z &= \mathbb{E}\left[(A\mathbf{x} + B\mathbf{y} - A\mu_x - B\mu_y)(A\mathbf{x} + B\mathbf{y} - A\mu_x - B\mu_y)^T\right] \\
&= \mathbb{E}\left[\left([A \;\; B]\begin{bmatrix} \mathbf{x} - \mu_x \\ \mathbf{y} - \mu_y \end{bmatrix}\right)\left([A \;\; B]\begin{bmatrix} \mathbf{x} - \mu_x \\ \mathbf{y} - \mu_y \end{bmatrix}\right)^T\right] \\
&= [A \;\; B]\mathbb{E}\left[\begin{bmatrix} \mathbf{x} - \mu_x \\ \mathbf{y} - \mu_y \end{bmatrix}\begin{bmatrix} \mathbf{x} - \mu_x \\ \mathbf{y} - \mu_y \end{bmatrix}^T\right][A \;\; B]^T = [A \;\; B]\begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}\begin{bmatrix} A^T \\ B^T \end{bmatrix}
\end{aligned}$$

Theorem (...)

- *the Gaussian random variable* **x** *conditioned on the Gaussian random variable* **y** *is still a Gaussian random variable. The PDF of* **x** *given* **y** *is*

$$p(x|y) = \mathcal{N}(\mu_{x|y}, \Sigma_{x|y}) \tag{1}$$

where

$$\begin{aligned} \mu_{x|y} &= \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y) && (2) \\ \Sigma_{x|y} &= \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} && (3) \end{aligned}$$

if $\Sigma_{yy} > 0$.

Theorem (Minimum Variance Estimator)

Let $\mathbf{x} \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}^m$ *be two r.v. (not necessarily Gaussian), and* $g : \mathbb{R}^m \rightarrow \mathbb{R}^n$ *a measurable function.*

We define $\hat{\mathbf{x}}_g = g(\mathbf{y})$ *as the estimator of* $\mathbf{x}$ *given* $\mathbf{y}$ *through the function* g*, and* $\mathbf{e}_g = \mathbf{x} - g(\mathbf{y}) = \mathbf{x} - \hat{\mathbf{x}}_g$ *the corresponding estimation error.*

The estimator $\hat{\mathbf{x}} = \mathbb{E}[\mathbf{x}|\mathbf{y}] = \hat{g}(\mathbf{y})$ *is optimal because it minimizes the error variance, i.e.*

$$Var(\mathbf{e}) = \mathbb{E}[(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^T] \leq \mathbb{E}[(\mathbf{x} - \hat{\mathbf{x}}_g)(\mathbf{x} - \hat{\mathbf{x}}_g)^T] = Var(\mathbf{e}_g), \quad \forall g(\cdot)$$

where $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$ *is the error of the optimal estimator.*

The error of the optimal estimator $\mathbf{e}$ *and the estimation* $\hat{g}(\mathbf{y})$ *are uncorrelated*

$$\mathbb{E}[\mathbf{e}\hat{g}(\mathbf{y})^T] = 0.$$

Stochastic models

LTI Stochastic model

Now we focus on the state-space representation of a generic Linear Time-Invariant (LTI) discrete-time stochastic model:

$$\begin{cases} x_{k+1} &= Ax_k + w_k \\ y_k &= Cx_k + v_k \end{cases}$$

where:

$$\begin{cases} v_k & \sim & \mathcal{N}(0, R), & \mathbb{E}[v_k v_h^T] = R\delta(k-h) \\ w_k & \sim & \mathcal{N}(0, Q), & \mathbb{E}[w_k w_h^T] = Q\delta(k-h) \\ x_0 & \sim & \mathcal{N}(\bar{x}_0, P_0) & \end{cases}$$

and v_k, w_k, x_0 are uncorrelated zero-mean Gaussian r.v.

$$\begin{aligned} \mathbb{E}[v_k w_h^T] &= 0 \\ \mathbb{E}[x_0 v_k^T] &= 0 \\ \mathbb{E}[x_0 w_k^T] &= 0 \end{aligned}$$

The state-space model is a way to describe the dynamic evolution of a stochastic process

LTI Stochastic model

From the evolution of the state and of the output at time $t = kT_s$ as sum of the free and forced terms

$$
\begin{aligned}
x_k &= A^{k-k_0}x_0 + \sum_{i=k_0}^{k-1} A^{k-i-1}w_i \\
y_k &= CA^{k-k_0}x_0 + \sum_{i=k_0}^{k-1} CA^{k-i-1}w_i + v_k
\end{aligned}
$$

we also have

$$
\begin{aligned}
\mathbb{E}[x_k w_h^T] &= 0, \quad \forall h \geq k \\
\mathbb{E}[x_k v_h^T] &= 0 \\
\mathbb{E}[y_k v_h^T] &= Q\delta(k-h)
\end{aligned}
$$

Kalman filtering

The Kalman filter (or minimum variance filter) is defined as:

$$\hat{x}_{k+1|k+1} = \mathbb{E}\left[x_{k+1}|y_0, \ldots, y_{k+1}\right] = \mathbb{E}\left[x_{k+1}|y_{k+1}, Y^k\right] \tag{4}$$

where $Y^k = (y_k, \ldots, y_1, y_0)$.

Goal: we need a recursive expression for $\hat{x}_{k+1|k+1}$ without using

$$\mu_{x|y} \quad = \quad \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y) \tag{5}$$

at any time instant k, i.e. when a new measurement is available.

The explicit expression for $\mathbb{E}[X|Y]$ is easy to derive from (5) if X e Y are joint Gaussian with means μ_X, μ_Y and variances $\begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix}$.

To rewrite

$$\hat{x}_{k+1|k+1} = \mathbb{E}\left[x_{k+1}|y_0,\ldots,y_{k+1}\right] = \mathbb{E}\left[x_{k+1}|y_{k+1},Y^k\right]$$

in the form $\mathbb{E}[X|Y]$ we introduce the following conditional random variables

$$\begin{aligned} X &= x_{k+1}|Y^k \\ Y &= y_{k+1}|Y^k \end{aligned}$$

and compute the following means, variances and covariances:

$$\begin{aligned} \mu_X &= \mathbb{E}\left[x_{k+1}|Y^k\right] \\ \mu_Y &= \mathbb{E}\left[y_{k+1}|Y^k\right] \\ P_{k+1|k} = \Sigma_{XX} &= Var\left[x_{k+1}|Y^k\right] \\ \Sigma_{YY} &= Var\left[y_{k+1}|Y^k\right] \\ \Sigma_{XY} = \Sigma_{YX}^T &= Cov\left[x_{k+1},y_{k+1}|Y^k\right]. \end{aligned}$$

The optimal estimator is given by

this is the goal
$$\mathbb{E}\left[X|\,Y\right] = \hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + \Sigma_{XY}\Sigma_{YY}^{-1}\left(y_{k+1} - \hat{y}_{k+1|k}\right) \tag{6}$$

and the variance of the estimation error is

$$\Sigma_{X|Y} = P_{k+1|k+1} = \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX} \tag{7}$$

Conditional means

Mean μ_X

$$
\begin{aligned}
\mu_X &= \mathbb{E}\left[x_{k+1} | Y^k\right] \\
&= \mathbb{E}\left[Ax_k + w_k | Y^k\right] \\
&= A\mathbb{E}\left[x_k | Y^k\right] + \mathbb{E}\left[w_k | Y^k\right] \\
&= A\hat{x}_{k|k} \\
&= \hat{x}_{k+1|k}
\end{aligned}
$$

Mean μ_Y

$$
\begin{aligned}
\mu_Y &= \mathbb{E}\left[y_{k+1} | Y^k\right] \\
&= \mathbb{E}\left[Cx_{k+1} + v_{k+1} | Y^k\right] \\
&= C\mathbb{E}\left[x_{k+1} | Y^k\right] + \mathbb{E}\left[v_{k+1} | Y^k\right] \\
&= C\hat{x}_{k+1|k} \\
&= CA\hat{x}_{k|k}
\end{aligned}
$$

Conditional variances

Variance Σ_{XX}

$$
\begin{aligned}
\Sigma_{XX} &= Var\left[x_{k+1} | Y^k\right] \\
&= \mathbb{E}\left[\left(x_{k+1} - \hat{x}_{k+1|k}\right)\left(x_{k+1} - \hat{x}_{k+1|k}\right)^T | Y^k\right] \\
&= \mathbb{E}\left[\left(Ax_k + w_k - A\hat{x}_{k|k}\right)\left(Ax_k + w_k - A\hat{x}_{k|k}\right)^T | Y^k\right] \\
&= A\mathbb{E}\left[\left(x_k - \hat{x}_{k|k}\right)\left(x_k - \hat{x}_{k|k}\right)^T | Y^k\right] A^T + \\
&\quad + A\mathbb{E}\left[\left(x_k - \hat{x}_{k|k}\right) w_k^T | Y^k\right] + \\
&\quad + \mathbb{E}\left[w_k \left(x_k - \hat{x}_{k|k}\right)^T | Y^k\right] A^T + \mathbb{E}\left[w_k w_k^T | Y^k\right] \\
&= AP_{k|k}A^T + Q \\
&= P_{k+1|k}
\end{aligned}
$$

Conditional variances

Variance Σ_{YY}

$$
\begin{aligned}
\Sigma_{YY} &= Var\left[y_{k+1}|Y^k\right] \\
&= \mathbb{E}\left[\left(y_{k+1}-\hat{y}_{k+1|k}\right)\left(y_{k+1}-\hat{y}_{k+1|k}\right)^T|Y^k\right] \\
&= \mathbb{E}\left[\left(Cx_{k+1}+v_{k+1}-C\hat{x}_{k+1|k}\right)\left(Cx_{k+1}+v_{k+1}-C\hat{x}_{k+1|k}\right)^T|Y^k\right] \\
&= C\mathbb{E}\left[\left(x_{k+1}-\hat{x}_{k+1|k}\right)\left(x_{k+1}-\hat{x}_{k+1|k}\right)^T|Y^k\right]C^T+ \\
&\quad +C\mathbb{E}\left[\left(x_{k+1}-\hat{x}_{k+1|k}\right)v_{k+1}^T|Y^k\right]+ \\
&\quad +\mathbb{E}\left[v_{k+1}\left(x_{k+1}-\hat{x}_{k+1|k}\right)^T|Y^k\right]C^T+\mathbb{E}\left[v_{k+1}v_{k+1}^T|Y^k\right] \\
&= CP_{k+1|k}C^T+R
\end{aligned}
$$

Conditional variances

Covariance $\Sigma_{XY} = \Sigma_{YX}^T$

$$
\begin{aligned}
\Sigma_{XY} &= Cov\left[x_{k+1}, y_{k+1} | Y^k\right] \\
&= \mathbb{E}\left[\left(x_{k+1} - \hat{x}_{k+1|k}\right)\left(y_{k+1} - \hat{y}_{k+1|k}\right)^T | Y^k\right] \\
&= \mathbb{E}\left[\left(Ax_k - A\hat{x}_{k|k} + w_k\right)\left(CAx_k - CA\hat{x}_{k|k} + v_{k+1} + Cw_k\right)^T | Y^k\right] \\
&= A\mathbb{E}\left[\left(x_k - \hat{x}_{k|k}\right)\left(x_k - \hat{x}_{k|k}\right)^T | Y^k\right] A^T C^T + \mathbb{E}\left[w_k w_k^T | Y^k\right] C^T \\
&= AP_{k|k}A^T C^T + QC^T \\
&= P_{k+1|k}C^T
\end{aligned}
$$

Summing up

The random variable $z = \begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix}$ conditioned on Y^k, has the following pdf

$$p\left(z|Y^k\right) \sim \mathcal{N}\left(\begin{bmatrix} \hat{x}_{k+1|k} \\ C\hat{x}_{k+1|k} \end{bmatrix}, \begin{bmatrix} P_{k+1|k} & P_{k+1|k}C^T \\ CP_{k+1|k} & CP_{k+1|k}C^T + R \end{bmatrix}\right)$$

with:

$$\begin{aligned} \hat{x}_{k+1|k} &= A\hat{x}_{k|k} \\ P_{k+1|k} &= AP_{k|k}A^T + Q \end{aligned}$$

Recursive formulation

The last step is to compute

$$p\left(x_{k+1} | Y^{k+1}\right) \sim \mathcal{N}\left(\hat{x}_{k+1|k+1}, P_{k+1|k+1}\right)$$

where the mean $\hat{x}_{k+1|k+1}$ is the optimal estimation we are looking for and $P_{k+1|k+1}$ the variance of the corresponding estimation error.

Substituting the previous expression we end up with

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + P_{k+1|k} C^T \left(C P_{k+1|k} C^T + R\right)^{-1} \left(y_{k+1} - C\hat{x}_{k+1|k}\right)$$

The **Kalman gain** is the matrix

$$K_{k+1} = P_{k+1|k} C^T \left(C P_{k+1|k} C^T + R\right)^{-1}$$

mapping the output estimation error into the correction of the prediction state

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_{k+1} \left(y_{k+1} - C\hat{x}_{k+1|k}\right)$$

The variance of the estimation error is

$$P_{k+1|k+1} = P_{k+1|k} - P_{k+1|k} C^T \left(C P_{k+1|k} C^T + R\right)^{-1} C P_{k+1|k}$$

Recursive formulation

Prediction step / A priori estimation

$$
\begin{aligned}
\hat{x}_{k+1|k} &= A\hat{x}_{k|k} \\
P_{k+1|k} &= AP_{k|k}A^T + Q
\end{aligned}
$$

Estimation step / A posteriori estimation

$$
\begin{aligned}
\hat{x}_{k+1|k+1} &= \hat{x}_{k+1|k} + K_{k+1}(y_{k+1} - C\hat{x}_{k+1|k}) \\
P_{k+1|k+1} &= P_{k+1|k} - P_{k+1|k}C^T(CP_{k+1|k}C^T + R)^{-1}CP_{k+1|k}
\end{aligned}
$$

Initial conditions

$$
\begin{aligned}
\hat{x}_{0|-1} &= \bar{x}_0 \\
P_{0|-1} &= P_0
\end{aligned}
$$

Kalman filter

$$\begin{aligned}\hat{x}_{k+1|k+1} &= A\hat{x}_{k|k} + K_{k+1}(y_{k+1} - CA\hat{x}_{k|k})\\ P_{k+1|k} &= AP_{k|k-1}A^T - AP_{k|k-1}C^T(CP_{k|k-1}C^T + R)^{-1}CP_{k|k-1}A^T + Q\end{aligned}$$

with

$$K_{k+1} = P_{k+1|k}C^T\left(CP_{k+1|k}C^T + R\right)^{-1}$$

The matrix recursive equation $P_{k+1|k} = \dots$ is called Riccati equation.

Kalman predictor

$$\begin{aligned}\hat{x}_{k+1|k} &= A\hat{x}_{k|k-1} + \bar{K}_k(y_k - C\hat{x}_{k|k-1})\\ P_{k+1|k} &= AP_{k|k-1}A^T - AP_{k|k-1}C^T(CP_{k|k-1}C^T + R)^{-1}CP_{k|k-1}A^T + Q\end{aligned}$$

with

$$\bar{K}_k = AP_{k|k-1}C^T\left(CP_{k|k-1}C^T + R\right)^{-1}$$

Observations:

1. All the information in $y(i)$, $i \in [0, k-1]$ is "embedded" in the estimation state $\hat{x}_{k-1|k-1}$: the following conditional expectations are equal

$$\mathbb{E}[x_k | y_k, y_{k-1}, \ldots, y_0] = \mathbb{E}[x_k | \hat{x}_{k-1|k-1}, y_k]$$

2. The optimal gain K_k is time-varying even if the stochastic model is LTI.
3. There is a more general formulation of the Kalman filter where w_k and v_k are correlated.
4. The same recursive equation for the Kalman filter can be used with linear time-varying stochastic systems.

What's happen when $k \to \infty$?
Does the estimation error converge to zero with minimal variance?

Theorem

Given the stochastic LTI model

$$\begin{cases} x_{k+1} = Ax_k + w_k \\ y_k = Cx_k + v_k \end{cases} \tag{8}$$

with

$$\begin{cases} v_k \sim \mathcal{N}(0, R), & \mathbb{E}[v_k v_h^T] = R\delta(k-h) \\ w_k \sim \mathcal{N}(0, Q), & \mathbb{E}[w_k w_h^T] = Q\delta(k-h) \\ x_0 \sim \mathcal{N}(\bar{x}_0, P_0) \end{cases} \tag{9}$$

where v_k, w_k, x_0 are zero mean uncorrelated Gaussian random variables.

Theorem (...)

Then

1. *The Algebraic Riccati Equation (ARE):*

$$P_\infty = AP_\infty A^T - AP_\infty C^T(CP_\infty C^T + R)^{-1}CP_\infty A^T + Q$$

   has a unique positive definite symmetric matrix solution $P_\infty = P_\infty^T > 0$

2. P_∞ *is stabilizable, i.e.* $(A - K_\infty C)$ *is asymptotically stable with*

$$K_\infty = P_\infty C^T \left(CP_\infty C^T + R\right)^{-1}.$$

3. $\lim_{k\to\infty} P(k|k-1) = P_\infty$ *holds for all initial conditions* $P(0|-1) = P_0 = P_0^T \geq 0$,

if and only if

1. (A, C) *is detectable,*
2. $(A, Q^{1/2})$ *is stabilizable.*

Kalman filter (LTI)

$$
\begin{aligned}
\hat{x}_{k+1|k+1} &= A\hat{x}_{k|k} + K_\infty(y_{k+1} - CA\hat{x}_{k|k}) \\
P &= APA^T - APC^T(CPC^T + R)^{-1}CPA^T + Q
\end{aligned}
$$

with

$$
K_\infty = PC^T\left(CPC^T + R\right)^{-1}
$$

Kalman predictor (LTI)

$$
\begin{aligned}
\hat{x}_{k+1|k} &= A\hat{x}_{k|k-1} + \bar{K}_\infty(y_k - C\hat{x}_{k|k-1}) \\
P &= APA^T - APC^T(CPC^T + R)^{-1}CPA^T + Q
\end{aligned}
$$

with

$$
\bar{K}_\infty = APC^T\left(CPC^T + R\right)^{-1}.
$$


Project

PROJECT – Assignment # 3

To do

- Implement the Kalman filter/predictor and estimate the velocity and acceleration from noisy position measurements (see .mat file)
- Implement the steady-state Kalman filter/predictor and estimate the velocity and acceleration from noisy position measurements (see .mat file)

Kalman smoother

Kalman Smoother

Model: $\{A, C, Q, R\}$

$$
\begin{aligned}
x_{k+1} &= Ax_k + w_k \\
y_k &= CX_k + v_k
\end{aligned}
$$

Data: sequence of N samples of the output

$$
y_0, y_1, \ldots, y_N
$$

STEP 1: **forward step**

"Standard" Kalman filtering

$$
\begin{aligned}
\hat{x}^f_{k+1|k+1} &= A\hat{x}^f_{k|k} + K_{k+1}(y_{k+1} - CA\hat{x}^f_{k|k}) \\
\hat{x}^f_{0|0} &= \bar{x}_0 \\
P^f_{k|k} &= \ldots \\
P^f_{k+1|k} &= \ldots \\
P^f_{0|0} &= P_0
\end{aligned}
$$

STEP 2: **backward step**

Smoothing

$$
\begin{aligned}
\hat{x}^s_{k|N} &= \hat{x}^f_{k|k} + \breve{K}_k \left[\hat{x}^s_{k+1|N} - \hat{x}^f_{k+1|t} \right] \\
\hat{x}^s_{N|N} &= \hat{x}^f_{N|N}
\end{aligned}
$$

where the conditional covariance matrix $P(t|N)$ satisfies the time-backward matrix equation

$$
\begin{aligned}
P_{k|N} &= P^f_{k|k} + \breve{K}_k \left[P_{k+1|N} - P^f_{k+1|k} \right] \\
P_{N|N} &= P^f_{N|N}.
\end{aligned}
$$

with

$$
\breve{K}_k = P^f_{k|k} A^T \left(P^f_{k+1|k} \right)^{-1}
$$

To design the Kalman filter|predictor|smoother we have to find out the relationship between two state space models: the discrete-time approximation of the kinematics (left) and the “standard stochastic model” (right)

$$\begin{cases} x_{k+1} = \begin{bmatrix} 1 & T_s \\ 0 & 1 \end{bmatrix} x_k + \begin{bmatrix} \frac{T_s^2}{2} \\ T_s \end{bmatrix} n_k \\ y_k = \begin{bmatrix} 1 & 0 \end{bmatrix} x_k + v_k \end{cases}, \qquad \begin{cases} x_{k+1} = Ax_k + w_k \\ y_k = Cx_k + v_k \end{cases}$$

We have

$$\begin{aligned} A &:= \begin{bmatrix} 1 & T_s \\ 0 & 1 \end{bmatrix}, & w_k &:= \begin{bmatrix} \frac{T_s^2}{2} \\ T_s \end{bmatrix} n_k \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{T_s^2}{2} \\ T_s \end{bmatrix} \begin{bmatrix} \frac{T_s^2}{2} \\ T_s \end{bmatrix}^T Q \right) \\ C &:= \begin{bmatrix} 1 & 0 \end{bmatrix}, & v_k &:= v_k \end{aligned}$$

Tuning of the filter

The variance R depends on the encoder resolution (we can read it on the datasheet) whereas the matrix Q is chosen by the designer to try to “explain” the measurements in the best way.

If an input command u_k enters within the stochastic model

$$\begin{cases} x_{k+1} & = & Ax_k + \boxed{Bu_k} + w_k \\ y_k & = & Cx_k + v_k \end{cases},$$

how do the filter equations change?

Fortunately if u_k is a function of past measurements (e.g. $u_k = f(y_{0:k})$ or $u_k = Kx_k$) then we can simple add the term $\boxed{Bu_k}$ in the recursive equations:

$$\begin{aligned} \hat{x}_{k+1|k} &= A\hat{x}_{k|k-1} + \boxed{Bu_k} + K_k(y_k - C\hat{x}_{k|k-1}) \\ P_{k+1|k} &= AP_{k|k-1}A^T - AP_{k|k-1}C^T(CP_{k|k-1}C^T + R)^{-1}CP_{k|k-1}A^T + Q \end{aligned}$$

or

$$\begin{aligned} \hat{x}_{k+1|k} &= A\hat{x}_{k|k-1} + \boxed{Bu_k} + \bar{K}_\infty(y_k - C\hat{x}_{k|k-1}) \\ P &= APA^T - APC^T(CPC^T + R)^{-1}CPA^T + Q \end{aligned}$$


Project

PROJECT – Assignment # 4

To do

- Implement the Kalman smoother and estimate the velocity and acceleration from noisy position measurements (see .mat file)