On Kalman-Filtering

⋯ a technique introduced 1960 by *Rudolf Emil Kalman*, who was born in 1930.
The applications range widely: navigational and guidance systems, radar tracking, sonar ranging, and satellite orbit determination.
In this document we cover the
1) the discrete linear and
2) the extended Kalman Filter.

Discrete linear Kalman-Filter

May a system (alt.: process, state) $x : I \to \mathbb{R}^n$ be governed by the continuous linear stochastic differential equation
p1 • $\dot{x} = F\,x + G\,u + w$
where
$I \subset \mathbb{R}$ is a connected interval
$F \in \mathbb{R}^{n\times n}$ and $G \in \mathbb{R}^{n\times m}$ are constant matrices (coining "linear")
$u : I \to \mathbb{R}^m$ is used-specified control or drift
$w : I \to \mathbb{R}^n$ is white noise (see below).
A discrete linear Kalman-Filter estimates the system, i.e. $x : I \to \mathbb{R}^n$ at concrete time points $\{t_0, t_1, \ ...\} \subset I$, and to do so, the filter takes into account measurements z at times $\{t_1, t_2, \ ...\}$ of the form
p2 • $z = H\,x + v$
where
• $z : I \to \mathbb{R}^l$ is the measurement, computed via
• $H \in \mathbb{R}^{l\times n}$, a constant matrix, and
• $v : I \to \mathbb{R}^l$ as white noise.
We are dealing with stochastic equations, in the sense, that
$w \sim N(0, Q)$ and $v \sim N(0, R)$, where
• $Q \in \mathbb{R}^{n\times n}$, $R \in \mathbb{R}^{l\times l}$ are constant covariance matrices.

Let's go through an intuitive, historical approach.

Recursive Least Squares (of order 0 and 1)

Johann Carl Friedrich Gauss (1777-1855) was at the age of 17, when he discovered the formula for the best polynomial fit (in the least square sense). Besides the well known matrix equation, the solution can be found recursively:

Example 1: We suspect our process $x : \mathbb{R}^+ \to \mathbb{R}$ is governed by
$\dot{x} = 0\,x$,
i.e. we are chasing a constant, namely: x. We perform measurements at successive time points $\{k : k \in \mathbb{N}_0\} \subset \mathbb{R}$. Let each measurement z_k be the true state x polluted with white noise $v_k \sim N(0, \sigma^2)$ for some variance $\sigma^2 \in \mathbb{R}^+$, i.e.
$z_k = x + v_k$.
Taking previous measurements into account, (by the weak law of large numbers) we archive improving estimates $\hat{x}_k$ according to the following recursion
$\hat{x}_k = \hat{x}_{k-1} + K_k(z_k - \hat{x}_{k-1})$, where
• $K_k \in \mathbb{R}$ are coefficients, which will later be referred to as the *Kalman gain*.
To derive the ideal values K_k for $k \in \mathbb{N}$ for the above process, we reasonably have to set $\hat{x}_1 = z_1$, which $\Longrightarrow K_1 = 1$. That means also, we trusted our initial estimate, not at all (just $K_k \to 1\,!$), i.e. we expected the error $e_0 := \hat{x}_0 - x$ to have variance $P_0 := E({e_0}^2) = \infty$.

Use of the *Ricatti* Equations

$M_k = P_{k-1}$

$K_k = \frac{M_k}{M_k+R}$ (← simplified version)

$P_k = (I - K_k)\, M_k$

yields $P_1 = \sigma^2$. In general, $P_k = E(e_k\, e_k{}^T)$ is the covariance of the error $e_k := \hat{x}_k - x$ (i.e. after the k-th measurement is taken into account). Next, we iterate some more: our second guess is

$\hat{x}_2 = \hat{x}_1 + K_2(z_2 - \hat{x}_1)$, where $K_2 = \frac{\sigma^2}{\sigma^2+\sigma^2} = \frac{1}{2}$. Furthermore, $P_2 = \frac{\sigma^2}{2}$.

We conclude with a table

$$\begin{pmatrix} k & 0 & 1 & 2 & 3 & \cdots & k \\ K_k & / & 1 & \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{k} \\ P_k & \infty & \sigma^2 & \frac{\sigma^2}{2} & \frac{\sigma^2}{3} & \cdots & \frac{\sigma^2}{k} \end{pmatrix}$$

and with the observation

$$\begin{aligned} & \hat{x}_3 = \tfrac{2}{3}\, \hat{x}_2 + \tfrac{1}{3}\, z_3 \\ \Longleftrightarrow\quad & \hat{x}_3 = \tfrac{2}{3}[\hat{x}_1 + K_2(z_2 - \hat{x}_1)] + \tfrac{1}{3}\, z_3 \\ \Longleftrightarrow\quad & \hat{x}_3 = \tfrac{2}{3}[\tfrac{1}{2}\, z_1 + \tfrac{1}{2}\, z_2] + \tfrac{1}{3}\, z_3 \\ \Longleftrightarrow\quad & \hat{x}_3 = \tfrac{1}{3}\,(z_1 + z_2 + z_3). \end{aligned}$$

After we have confirmed our guess with an inductive proof, we understand, that our estimate after k steps, matches the least square solution, namely the mean of all measurements up to k. P_k computes as

$$P_k = E(e_k{}^2) = E\Big((\hat{x}_k - x)^2\Big) = E\Big(\big(\tfrac{1}{k}\, \textstyle\sum_{i=1}^k (x + v_k) - x\big)^2\Big) = E\Big(\big(\tfrac{1}{k}\, \textstyle\sum_{i=1}^k v_k\big)^2\Big)$$
$$= \tfrac{1}{k^2} \sum\sum E(v_i\, v_j) = \tfrac{1}{k^2} \sum\sum \sigma^2\, \delta_{\mathrm{ij}} = \tfrac{1}{k}\, \sigma^2$$

Example 2: Feeling mature, we suspect our process is governed by

$$\dot{x} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} x = :F\, x,$$

i.e. we are chasing a polynomial of degree 1. We remind the fanciless reader, that policemen intend that, when pointing their old-fashioned-non-doppler-radar-cannon towards your moving car. Allow us to talk henceforth in position $\hat{\boldsymbol{x}}_{\boldsymbol{k}}$ and speed $\hat{\dot{\boldsymbol{x}}}_{\boldsymbol{k}}$ notion; we estimate both:

$$\hat{x}_k = \begin{pmatrix} \hat{\boldsymbol{x}}_{\boldsymbol{k}} \\ \hat{\dot{\boldsymbol{x}}}_{\boldsymbol{k}} \end{pmatrix}$$

In view of the iteration, we are doing right to propagate the previous estimate

$\hat{x}_k = \Phi(\Delta\, t) \cdot \hat{x}_{k-1} + K_k(z_k - H \cdot \Phi(\Delta\, t) \cdot \hat{x}_{k-1}),$ eq. 1

by multiplication with the fundamental matrix

$$\Phi(t) = e^{F\, t} = I + F\, t + \frac{F^2\, t^2}{2} + \ldots + \frac{F^m\, t^m}{m!} + \ldots = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$$

which is precisely Explicit-Euler-Integration, where $\Delta\, t$ is the time between measurements $k - 1$ and k.

A (projection) matrix H comes in handy, because it is non unlikely, that we can only measure a subset of our state variables. In case of the non-doppler-radar, we measure position, but not velocity, so here we choose

$H = (\,1 \quad 0\,).$

Note, that in eq. 1 we "compare" the measurement $z_k = x + v_k$, which is as usual polluted with noise $N(0, \sigma^2)$, with the propagated estimate.

Whenever unambigous: $\Phi = \Phi(\Delta\, t)$. E.g. now, the *Ricatti* Equations are modified to

$M_k = \Phi\, P_{k-1}\, \Phi^T$

$K_k = \frac{M_k\, H^T}{H\, M_k\, H^T + R}$ (yet incomplete version)

$P_k = (I - K_k\, H)\, M_k$

The transformation $\Phi\, P_{k-1}\, \Phi^T$, takes care that e.g. error covariance in velocity, increases error variance in position.

It turns out, for a fixed $\Delta\, t$,

$$K_k = \begin{pmatrix} \frac{2(2k-1)}{k(k+1)} \\ \frac{6}{k(k+1)\Delta t} \end{pmatrix} \quad \text{and} \quad P_k = \begin{pmatrix} \frac{2(2k-1)}{k(k+1)}\sigma^2 & 0 \\ 0 & \frac{12\sigma^2}{k(k^2+1)\Delta t^2} \end{pmatrix}$$

and unsurprisingly, K_k coincides with the recursive least square gains for linear regression.

■ Kalman-Filter

We state the full body of the Kalman-Filter iteration

k1 • $\hat{x}_k = [\Phi\,\hat{x}_{k-1} + G_k\,u_{k-1}] + K_k(z_k - H\,[\Phi\,\hat{x}_{k-1} + G_k\,u])$, eq. 2

which shall correspond to the stochastic diff. equation (see beginning)

$$\dot{x} = F\,x + G\,u + w$$

in which u is the *control* or *drift* and w is *process noise* where $p(w) \sim N(0, Q)$. Q is an input parameter (!).

From eq. 2, we can see, that G_k is to propagate the control, assumed to be permanently applied during the time step from $k-1$ to k. If $u_k = u(t_k)$, we yield from stoch. analysis

$$G_k(t) = \int_0^t \Phi(\tau)\,G\,d\tau$$

The process noise affects the Ricatti eqs.

$$M_k = \Phi\,P_{k-1}\,\Phi^T + Q_k(\Delta t)$$

$$K_k = \frac{M_k\,H^T}{H\,M_k\,H^T + R} \qquad \text{(final version)}$$

$$P_k = (I - K_k\,H)\,M_k$$

where

$$Q_k(t) = \int_0^t \Phi(\tau)\,Q\,\Phi^T(\tau)\,d\tau.$$

Remarks:

• If we estimate position and velocity, then u could be the known (!) driver-caused acceleration of a car or the gravity constant influencing an airplain (flying of couse). The scenario suggests $G = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and we compute

$$G_k(\Delta t) = \int_0^{\Delta t} \begin{pmatrix} 1 & \tau \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} d\tau = \int_0^{\Delta t} \begin{pmatrix} \tau \\ 1 \end{pmatrix} d\tau = \begin{pmatrix} \frac{\Delta t^2}{2} \\ \Delta t \end{pmatrix}.$$

Note, that the result coincides with the excerpt of the 3×3 fundamental matrix

$$\begin{pmatrix} 1 & t & \frac{t^2}{2} \\ 0 & 1 & t \\ 0 & 0 & 1 \end{pmatrix}, \text{ i.e. the state propagation of the 2nd derivative.}$$

• The process noise models the inaccuracy of the fundamental matrix, and choosing η^2 large is therefore a possibility to make the filter robust agaist unexpected changes in the state. In Example 1, one can imagine that the state is constant, but the constant swaps to different values on the macro time scale. Usually, one assumes the process noise to enter the highest derivative. Picking up the Example 2, in the stochastic diff. equation,

$$\dot{x} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} x + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u + w,$$

where $p(w) \sim N\left[0, \begin{pmatrix} 0 & 0 \\ 0 & \eta^2 \end{pmatrix}\right]$, such that effectively only acceleration is affected.

$$Q_k(t) = \int_0^t \begin{pmatrix} 1 & \tau \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 0 & \eta^2 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ \tau & 1 \end{pmatrix} d\tau = \eta^2 \int_0^t \begin{pmatrix} \tau^2 & \tau \\ \tau & 1 \end{pmatrix} d\tau = \eta^2 \begin{pmatrix} \frac{t^3}{3} & \frac{t^2}{2} \\ \frac{t^2}{2} & t \end{pmatrix}.$$

• The Kalman iteration differs from rec. least squares, if the underlying process is not assumed to be governed by a polynomial.

• Precomputation is welcome, whenever storage is available and computation is expensive. We note, that if the measurements are taken in constant intervals, say each time step elapses after Δt, the Kalman gains K_k can be precomputed and Φ and G_k are constant.

Derivation of the Riccati Eqs.

see attached without-permission-copy from <u>Fundamentals of Kalman Filtering</u> by Paul Zarchan & Howard Musoff 2000 - only reference used.

Discrete non-linear Kalman-Filter (alt.: Extended)

For most applications we are dealing with

$\dot{x} = f(x, u) + w$and

$z = h(x) + v$.

The iteration keeps its form, but Φ_k, G_k and H_k vary over k. To yield Φ_k at each step we compute

$F_k = \partial_x\, f(x, u)\,|_{x=\hat{x}_{k-1}}$

$\Phi_k = e^{F_k\, t} = I + F_k\, t + \ldots$

and $H_k = \partial_x\, h(x)\,|_{x=\hat{x}_{k-1}}$.

such that

$\hat{x}_k = [\Phi_k\, \hat{x}_{k-1} + G_k\, u_{k-1}] + K_k(z_k - H_k\, [\Phi_k\, \hat{x}_{k-1} + G_k\, u])$. eq. 3

Example $\lfloor\pi\rfloor$:

We would like to estimate hight, velocity and ballistic coefficient β of an object, that is falling towards earth. System is assumed to be governed by

$\ddot{x} = \beta\, e^{-\lambda\, x}\, \dot{x}^2 - g$ and $\dot{\beta} = 0$

which encounters drag.

$$F = \begin{pmatrix} 0 & 1 & 0 \\ \partial_x\, \ddot{x} & \partial_{\dot{x}}\, \ddot{x} & \partial_\beta\, \ddot{x} \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ -\beta\, \lambda\, e^{-\lambda\, x}\, \dot{x}^2 & 2\, \beta\, e^{-\lambda\, x}\, \dot{x} & e^{-\lambda\, x}\, \dot{x}^2 \\ 0 & 0 & 0 \end{pmatrix}$$

For this example $\Phi_k(t) = e^{F\, t}\,|_{x=\hat{x}_{k-1}}$ has closed form, otherwise we could have used the Taylor approximation.