Spacecraft Dynamics

Lesson 6: Attitude estimation. Kalman Filtering.

Rafael Vázquez Valenzuela

Departamento de Ingeniería Aeroespacial
Escuela Superior de Ingenieros, Universidad de Sevilla rvazquez1@us.es

June 30, 2022

Attitude estimation

- The (dynamic) estimation of attitude (classicaly known simply as attitude estimation) requires the use of kynematic models, gyro measurements, and Kalman filter, as well as complementary sensors (measuring a direction).
- Gyros measure the angular velocity $\omega^{B}_{B/I}$ w.r.t. the inertial frame. One can recover the attitude by using this measurement to integrate the kinematic differential equations. Unfortunately, small errors acumulate over time generating a certain drift in the estimation; thus it is always necessary to use additional sensors to improve the measurement.
- To better understand how errors acumulate, one needs to model it as an stochastic (random) process, and use the propagation equations.
- Notation: in this lesson, arrows will not be used for vectors.

Stochastic Processes

- A stochastic process (or stochastic variable) is a random variable $X(t)$ whose distribution evolves (changes) with time. Estimation errors are modelled as this.
- Thus, mean and covariance also change with time: $m(t)$, $\Sigma(t)$.
- For a process, one can define the autocorrelation as $R(t,\tau) = E[X(t)X(\tau)^T]$. Autocorrelation allows to model how the past history of X influences its present value.
- Gaussian process: A Gaussian process verifies $X(t) \sim N_n(m(t), \Sigma(t))$, this is, it is distributed as a multivariate normal whose mean and covariance evolve with time.

White noise.

- White noise: It is the process $\nu(t)$ verifying:
  - $E[\nu(t)] = 0$.
  - $E[\nu(t)\nu(t)^T] = Q$.
  - $R(t, \tau) = E[\nu(t)\nu(\tau)^T] = \delta(t - \tau)Q$, where $\delta(x)$ is 1 if $x = 0$ and 0 otherwise.
- The last condition means that the value of white noise at present is independent of its value in any previous instant.
- Gaussian white noise: It is a process verifying the previous conditions and in addition, being Gaussian.
- A good model for sensor errors is $\delta\epsilon(t) = b + D\nu$, where ν is Gaussian white noise. The value of b is the mean of the error (bias), which sometimes is also modelled as a process itself (albeit slowly varying).

Propagation of error. Continuous case

- Consider a differential equation such as

$$\dot{x} = Ax + D\nu,$$

where ν is Gaussian white noise with covariance Q, and the initial condition is also a Gaussian: $x_0 \sim N_n(m_0(t), P_0(t))$. This is called a *stochastic differential equation* (the simplest possible one). Then one has that x is a Gaussian process, $x \sim N_n(m(t), P(t))$, with mean and covariance evolving as follows:

$$\begin{aligned} \dot{m} &= Am, \\ \dot{P} &= AP + PA^T + DQD^T, \\ m(0) &= m_0, \\ P(0) &= P_0 \end{aligned}$$

Propagation of error. Discrete case

- Consider a discrete equation of the type

$$x_{k+1} = Ax_k + Db_k,$$

where b_k is Gaussian white noise with covariance Q_k, and the initial condition is also a Gaussian: $x_0 \sim N_n(m_0(t), P_0(t))$This is called a *stochastic discrete process* (the simplest possible one). Then one has that x_k is a Gaussian process, $x_k \sim N_n(m_k(t), P_k(t))$, with mean and covariance evolving as follows:

$$\begin{aligned} m_{k+1} &= Am_k, \\ P_{k+1} &= AP_kA^T + DQ_kD^T, \end{aligned}$$

1-D example: gyro drift

- When one has gyro measurement, one needs to integrate the kinematic differential equations with the measurement.
- To easily grasp the concept of "error as a process", let us analyze the easiest possible case: a single degree of freedom in rotation. Thus, there is a single angle θ, whose kinematic differential equation is

$$\dot{\theta} = \omega$$

- A gyro produces a measurement of ω which we can denote by $\hat{\omega}$; for simplification purposes, assume we have a continuous measurement (it will be fast but not really continuous). In reality, it will not be exactly ω, but it'd rather be corrupted by some noise (which we model as Gaussian white noise, with variance Q related to the quality of the gyro) ν:

$$\hat{\omega} = \omega - \nu$$

1-D example: gyro drift

- If one tries to estimate θ (denote the estimation as $\hat{\theta}$) from $\hat{\omega}$ and assuming we know some estimation of its initial value $\hat{\theta}_0$, one would just write:

$$\dot{\hat{\theta}} = \hat{\omega}, \hat{\theta}(0) = \hat{\theta}_0$$

- Thus the estimation error $\delta\theta = \theta - \hat{\theta}$ verifies:

$$\delta\dot{\theta} = \omega - \hat{\omega} = \nu$$

- Assuming some initial error $\delta\theta(0) \approx N(0, P_0)$, one finds by applying the previous theory that the error $\delta\theta(t) \approx N(m(t), P(t))$, with:

$$\dot{m} = 0, m(0) = 0 \longrightarrow m(t) = 0, \quad \dot{P} = Q, P(0) = P_0 \longrightarrow P = P_0 + Qt$$

- Thus, even if the mean of the error is always zero, the variance grows linearly in time and eventually blows up, thus this estimator is useless in the medium-long term (but note error is small in the short term if P_0 was small to begin with).

External measurement

- Now assume one has external measurements of the angle with an additional sensor. When time $t = t_k$ (this is at certain time instants) one gets $\hat{\theta}(t_k)$, which we denote as $\hat{\theta}_k^m$, with some other device (which also should have some associated error, thus $\hat{\theta}_k^m = \theta_k - \epsilon$, where ϵ is white noise with variance R.
- Since time in-between measurements could be large, maybe it is not a good idea to ignore the gyro and say $\hat{\theta}(t) = \hat{\theta}_k^m$ for $t \in [t_k, t_{k+1})$.
- A possible idea is to *reset* the estimator of the previous slide when $t = t_k$, this is, combining the measures as follows:

$$\dot{\hat{\theta}} = \hat{\omega}, \quad \hat{\theta}(t_k) = \hat{\theta}_k^m, \quad t \in [t_k, t_{k+1}),$$

- Thus every new external measurement resets the initial condition of the differential equation and one integrates again.
- It is easy to see that the estimation error now verifies $\delta\theta \approx N(m(t), P(t))$, with $m(t) = 0$ and $\dot{P} = Q$, for $t \in [t_k, t_{k+1})$, with $P(t_k) = R$, thus $P = R + Q(t - t_k)$.

Kalman Filter

- The resetting idea makes the error maximum just before a measurement. The error would be $P = R + Q(t_{k+1} - t_k)$ right at that time instant.
- The problem with resetting is that it neglects the previous estimation from the differential equation, when in-between measurements it does not grow so large (it is short term). The idea of Kalman filtering is to *combine* the estimation from the differential equation (called the “a priori” estimation obtained from a “propagation step”) with the external measurement in an “update step” to obtain the “best possible combination” (called the “a posteriori” estimation). The combination is best in the sense that it minimizes the covariance.

Kalman Filter

- Some notation: estimation before the measurement is called a priori and denoted as $\hat{\theta}_k^-$.
- Estimation after the measurement is the a posteri estimation, denoted as $\hat{\theta}_k^+$ and it is computed as:

$$\hat{\theta}_k^+ = \hat{\theta}_k^- + K(\hat{\theta}_k^m - \hat{\theta}_k^-)$$

  where K is the Kalman gain and the parenthesis is the difference between the external measurement and the a priori estimation.
- K is computed to minimize the covariance of the a posteriori error.

Kalman Filter

- Covariance a priori is called P_k^-.
- Remember the formulas for combination of normals from Lesson 3 (slide 32).
- A posteriori, computing the covariance of θ_k^+:

$$P_k^+ = (1-K)^2 P_k^- + K^2 R$$

- Take derivative w.r.t. K and make it zero to find a minimizer: $0 = -2(1-K)P_k^- + 2KR$, thus $K = \frac{P_k^-}{P_k^- + R}$.
- Then covariance a posteriori becomes with that value of K:

$$P_k^+ = \frac{P_k^- R}{P_k^- + R}$$

- It can be seen that P_k^+ is less than both R and P_k^- (since both are positive numbers): thus one gets to improve estimation by using all the available information in the best way!

Kalman Filter

- Summarizing the algorithm:
- Initialization: For $t_0 = t = 0$ start with $\theta_0^+ = \hat{\theta}_0$ and $P_0^+ = P_0$.
- Propagation: For $t \in [t_k, t_{k+1})$, $k = 0, \ldots,$ one integrates from the last a posteriori estimation both the estimation and the covariance of the error

$$\dot{\hat{\theta}} = \hat{\omega}, \quad \hat{\theta}(t_k) = \hat{\theta}_k^+, \qquad \dot{P} = Q, \quad P(t_k) = P_k^+,$$

- Update: When $t = t_{k+1}$ set $\hat{\theta}_{k+1}^- = \hat{\theta}(t_{k+1})$ and $P_{k+1}^- = P(t_{k+1})$, and one gets the external measurement $\hat{\theta}_{k+1}^m$. Apply the KF:

$$\hat{\theta}_{k+1}^+ = \hat{\theta}_{k+1}^- + K(\hat{\theta}_{k+1}^{med} - \hat{\theta}_{k+1}^-),$$

where $K = \frac{P_{k+1}^-}{P_{k+1}^- + R}$, also $P_{k+1}^+ = \frac{P_{k+1}^- R}{P_{k+1}^- + R}$.
- Increase k and repeat the propagation step.

Kalman Filter: dependence on process/measurement noise

- If the measurement of the gyro is of very bad quality (Q is very large) then $P_k^- \to \infty$, one can see tha then $P_k^+ \to R$, $K \to 1$, and therefore $\hat{\theta}_k^+ \to \hat{\theta}_k^m$ (this is the resetting method: one takes the external measurement ignoring the result of integrating the differential equation).
- If the external measurement is of very bad quality (R is very large), then $P_k^+ \to P_k^-$, $K \to 0$, and thus $\hat{\theta}_k^+ \to \hat{\theta}_k^-$ (the estimation is just the result of integrating as if there was no external measurement).
- If it happens that $P_k^- \to R$, this is, the a priori estimation and the external measurement have the same level of error, then $P_k^+ \to R/2$, $K \to 1/2$, and then $\hat{\theta}_k^+ \to \frac{\hat{\theta}_k + \hat{\theta}_k^-}{2}$ (one takes the average between the integration step and the external measurement; note that the error is halved).

Kalman Filter: additional considerations

- This is a considerable simplification because only a 1-D linear case has been considered.
- Next the n-D linear case will be studied, the the nonlinear case (addressed by linearization), and finally a special case involving quaternions.
- In any case, conceptually all are the same: one integrates the kinematic differential equation with the gyroscopes and when obtaining an external measurement, the Kalman algorithm is used to weight the a priori estimation and the measurement.
- In aircraft and missiles Kalman Filtering is used to integrate the use of IMUs (gyros+ accelerometers) with external measurements such as GPS.

Kalman Filter for linear systems

- Next the KF will be explained for linear systems which are continuous with discrete measurement.
- All systems are in practice discrete, however, this explanation is simpler conceptually speaking and can be easily implemented in a lab setting.
- In the nomenclature of KF, a system is known as a "process".
- Note that the following development is conceptually very similar to the 1-D example, but more abtruse in terms of notation (and the number of involved matrices).
- KF is used in many engineering contexts (e.g. navigation, orbital mechanics, tracking...). It is a very useful tool to know.

System model (linear case)

- PROCESS: The process is continuous $\dot{x}(t) = A(t)x(t) + B(t)u(t) + D(t)\epsilon(t)$, where x is a Gaussian process of dimension n_x, $A(t)$ is a matrix (that could be time-varying) of dimension $n_x \times n_x$, $\epsilon(t)$ is Gaussian white noise of dimension n_ϵ and convariance $Q(t)$ (process noise), and $D(t)$ is a matrix es una matriz (that could be time-varying) of dimension $n_x \times n_\epsilon$. $u(t)$ if it exists is some input (e.g. gyro measurement) of dimension n_u and $B(t)$ is of dimension $n_x \times n_u$.
- MEASUREMENT: In discrete times $t = t_k$ a measurement z is taken, defined as follows: $z(t_k) = H_k x(t_k) + \nu(t_k)$, where z is of dimension n_z, H_k is a matrix (that could be time-varying) of dimension $n_z \times n_x$, and $\nu(t_k)$ is Gaussian white noise of dimension n_ν and convariance R_k (measurement noise).
- In addition $\nu(t_k)$ and $\epsilon(t)$ should be independent and the initial condition of x is $x(t_0) \sim N_{n_x}(\hat{x}_0, P_0)$.

System model (linear case)

- Summarizing:

$$\begin{aligned} \dot{x}(t) &= A(t)x(t) + B(t)u(t) + D(t)\epsilon(t), \\ z(t_k) &= H_k x(t_k) + \nu(t_k), \\ E[\epsilon(t)] &= E[\nu(t_k)] = 0, \\ E[\epsilon(t)\epsilon^T(\tau)] &= \delta(t-\tau)Q(t), \\ E[\nu(t_k)\nu^T(t_j)] &= \delta_{kj}R_k, \\ E[\epsilon(t)\nu^T(t_j)] &= 0, \\ x(t_0) &\sim N_{n_x}(\hat{x}_0, P_0). \end{aligned}$$

- Define the estimation (in t) of $x(t)$ as $\hat{x}(t)$.
- Define the covariance of the estimation error as $P(t) = E[(x(t) - \hat{x}(t))(x(t) - \hat{x}(t))^T]$.
- The goal of KF is, using the above model, and from the measurements $z(t_k)$, obtain the best possible estimation, this is, the value of $\hat{x}(t)$ that minimizes $P(t)$.

KF I

- If there are no measurements one can take $\hat{x}$ as the mean of the process; then, $x(t) \sim N_{n_x}(\hat{x}(t), P_k)$, where:

$$\begin{aligned} \dot{\hat{x}}(t) &= A(t)\hat{x}(t) + B(t)u(t), \\ \dot{P} &= A(t)P + PA^T(t) + D(t)Q(t)D^T(t). \end{aligned}$$

- The idea of the KF is that this is the best we can do until we get a new measurement at $t = t_k$, $z(t_k)$. Denote the estimation until then (the "a priori" estimation) as $\hat{x}^-(t_k)$ and the covariance of the error as P_k^-.
- Now if the estimation and measurement were perfect, one would have $z(t_k) = H_k\hat{x}^-(t_k)$. However, since this is not the case, one updates the estimation (obtaining the "a posteriori" estimation) proportionally to the discrepance between what we expect to measure and what we really measure: $\hat{x}^+(t_k) = \hat{x}^-(t_k) + K_k(z(t_k) - H_k\hat{x}^-(t_k))$.

KF II

- In $\hat{x}^{+}(t_k) = \hat{x}^{-}(t_k) + K_k(z(t_k) - H_k\hat{x}^{-}(t_k))$ we don't know K_k, which is the Kalman gain. This is determined to guarantee that he covariance of $\hat{x}^{+}(t_k)$, P_k^{+}, is as small as possible.
- Compute P_k^{+}: $P_k^{+} = E[(x(t_k) - \hat{x}^{+}(t_k))(x(t_k) - \hat{x}^{+}(t_k))^T]$, and replacing $\hat{x}^{+}(t_k)$:

$$
\begin{aligned}
P_k^{+} &= E\left[\left(x(t_k) - \hat{x}^{+}(t_k)\right)\left(x(t_k) - \hat{x}^{+}(t_k)\right)^T\right] \\
&= E\left[\left(x(t_k) - \hat{x}^{-}(t_k) - K_k(z(t_k) - H_k\hat{x}^{-}(t_k)\right)\right. \\
&\quad \left.\times \left(x(t_k) - \hat{x}^{-}(t_k) - K_k(z(t_k) - H_k\hat{x}^{-}(t_k))\right)^T\right]
\end{aligned}
$$

- Substituting $z(t_k) = H_k x(t_k) + \nu(t_k)$:

$$
\begin{aligned}
P_k^{+} &= E\left[\left(x(t_k) - \hat{x}^{-}(t_k) - K_k(H_k x(t_k) + \nu(t_k) - H_k\hat{x}^{-}(t_k)\right)\right. \\
&\quad \left.\times \left(x(t_k) - \hat{x}^{-}(t_k) - K_k(H_k x(t_k) + \nu(t_k) - H_k\hat{x}^{-}(t_k))\right)^T\right]
\end{aligned}
$$

KF III

- Simplifying terms:

$$
\begin{aligned}
P_k^+ &= E\Big[\Big((I - K_k H_k)(x(t_k) - \hat{x}^-) - K_k \nu(t_k)\Big) \\
&\quad \times \Big((I - K_k H_k)(x(t_k) - \hat{x}^-) - K_k \nu(t_k)\Big)^T\Big] \\
&= (I - K_k H_k) P_k^- (I - K_k H_k)^T + K_k R_k K_k^T
\end{aligned}
$$

- One needs to find K_k to minimize the previous expression. However one cannot “minimize a matrix” (what does that even mean?). However, since the diagonal of the covariance matrix is the individual variances, one idea is to minimize the trace of the matrix.
- The following mathematical relations help a lot:

$$
\frac{\partial \mathrm{Tr}[\mathrm{ABA^T}]}{\partial A} = 2BA^T, \quad \frac{\partial \mathrm{Tr}[\mathrm{AB}]}{\partial A} = B
$$

KF III

- Using these relations:

$$\mathrm{Tr}[P_k^+] = \mathrm{Tr}[K_k(R_k + H_k P_k^- H_k^T)K_k^T] - 2\mathrm{Tr}[K_k H_k P_k^-]$$

- Thus:

$$\frac{\partial \mathrm{Tr}[P_k^+]}{\partial K_k} = 2(R_k + H_k P_k^- H_k^T)K_k^T - 2H_k P_k^-$$

- Equating to zero:

$$K_k^T = (R_k + H_k P_k^- H_k^T)^{-1} H_k P_k^-$$

- Therefore we find an expression for the optimal Kalman gain

$$K_k = P_k^- H_k^T (R_k + H_k P_k^- H_k^T)^{-1}$$

- And substituting in P_k^+ to find the minimum we get

$$P_k^+ = (I - K_k H_k) P_k^-$$

KF algorithm

- Summarizing the algorithm:
  1. (Initialization): In $t = t_k$, we start from $\hat{x}^+(t_k)$ and $P^+(t_k)$. If $k = 0$ we take $\hat{x}^+(t_0) = \hat{x}_0$ y $P_0^+ = P_0$.
  2. (Propagation): For $t \in (t_k, t_{k+1})$, use the process model:

$$\begin{aligned}\dot{\hat{x}} &= A(t)\hat{x} + B(t)u(t), \quad \hat{x}(t_k) = \hat{x}^+(t_k) \\ \dot{P} &= A(t)P + PA^T(t) + D(t)Q(t)D^T(t), \quad P(t_k) = P^+(t_k)\end{aligned}$$

  3. (Update): In $t = t_{k+1}$ we get $z(t_{k+1})$, call $\hat{x}^-(t_{k+1}) = \hat{x}(t_{k+1})$ and $P^-(t_{k+1}) = P(t_{k+1})$. Compute the Kalman gain: $K_{k+1} = P_{k+1}^- H_{k+1}^T \left(H_{k+1} P_{k+1}^- H_{k+1}^T + R_{k+1}\right)^{-1}$. With $z(t_{k+1})$ compute the a posteriori estimation:

$$\begin{aligned}\hat{x}^+(t_{k+1}) &= \hat{x}^-(t_{k+1}) + K_{k+1}(z(t_{k+1}) - H_{k+1}\hat{x}^-(t_{k+1})), \\ P_{k+1}^+ &= (I - K_{k+1}H_{k+1})P_{k+1}^-.\end{aligned}$$

  4. Iterate for the next value of k.

About measurements

- Note: Measurements may change in the different t_k's (more or less measurements).
- This is reflected in changes in H_k (it can even change dimension).

Kalman Filter for nonlinear systems

- Next the EKF will be explained for nonlinear systems which are continuous with discrete measurement.
- The main tool is to linearize *around the estimation*.
- Unfortunately convergence is not guaranteed.
- If the initial estimation is good, the errors are not too large, and the measurements are of decent quality, it should work. However it is very dependent on the quality of the matrices Q and R.

System model (nonlinear case)

- The model is more general:

$$
\begin{aligned}
\dot{x}(t) &= f(x, u, t) + D(t)\epsilon(t), \\
z_k &= h(x_k, t_k) + \nu(t_k), \\
E[\epsilon(t)] &= E[\nu(t_k)] = 0, \\
E[\epsilon(t)\epsilon^T(\tau)] &= \delta(t-\tau)Q(t), \\
E[\nu(t_k)\nu^T(t_j)] &= \delta_{kj}R_k, \\
E[\epsilon(t)\nu^T(t_j)] &= 0, \\
x(t_0) &\sim N_{n_x}(\hat{x}_0, P_0).
\end{aligned}
$$

- Define the matrices and vectors: $F(\hat{x}(t), t) = \left.\frac{\partial f(x,u,t)}{\partial x}\right|_{x=\hat{x},u}$,

$\delta z_k = z_k - h(\hat{x}_k, t_k)$, $H_k(\hat{x}_k) = \left.\frac{\partial h(x,t_k)}{\partial x}\right|_{x=\hat{x}_k}$.

EKF algorithm

- The EKF is as follows:
  1. (Initialization): In $t = t_k$, we start from $\hat{x}^+(t_k)$ and $P^+(t_k)$. If $k = 0$ we take $\hat{x}^+(t_0) = \hat{x}_0$ y $P_0^+ = P_0$.
  2. (Propagation): For $t \in (t_k, t_{k+1})$, use the (nonlinear) process model:

$$\begin{aligned} \dot{\hat{x}} &= f(\hat{x}, u, t), \quad \hat{x}(t_k) = \hat{x}^+(t_k) \\ \dot{P} &= F(\hat{x}(t), t)P + PF^T(\hat{x}(t), t) + D(t)Q(t)D^T(t), \quad P(t_k) = P^+(t_k) \end{aligned}$$

  3. (Update): In $t = t_{k+1}$ we get $z(t_{k+1})$, call $\hat{x}^-(t_{k+1}) = \hat{x}(t_{k+1})$ and $P^-(t_{k+1}) = P(t_{k+1})$. Compute $\delta z_{k+1} = z_{k+1} - h(\hat{x}_{k+1}^-, t_{k+1})$ and $H_{k+1} = H_k(\hat{x}_{k+1}^-, t_{k+1})$. Compute the Kalman gain: $K_{k+1} = P_{k+1}^- H_{k+1}^T \left(H_{k+1} P_{k+1}^- H_{k+1}^T + R_{k+1}\right)^{-1}$. Then:

$$\begin{aligned} \hat{x}^+(t_{k+1}) &= \hat{x}^-(t_{k+1}) + K_{k+1}\delta z_{k+1}, \\ P_{k+1}^+ &= (I - K_{k+1}H_{k+1})P_{k+1}^-. \end{aligned}$$

  4. Iterate for the next value of k.

Multiplicative Extended Kalman Filter (MEKF)

- This is specific for attitude estimation.
- The EKF can be altered to take into account that the quaternions cannot be linearized in the standard way, but rather using the quaternion error (in a multiplicative way). Then one gets the MEKF.
  1. Assume one has gyros in the 3 axis, so that angular velocity $\hat{\omega}^B_{B/N}$ is estimated, with white noise error of covariance Q. This is assumed as continouous.
  2. At instants t_k one gets measurements of n directions in body axes $\hat{v}_i^B$, so that $v_i^B = C_N^B v_i^N$ and $v_i^B = \hat{v}_i^B + \epsilon_i$ for $i = 1, \dots, n$. ϵ_i is Gaussian white noise with covariance R_i.
- With only measurements one could use TRIAD or the q algorithm.
- With only gyros the estimation would be $\dot{\hat{q}} = \frac{1}{2} q \star q_{\hat{\omega}}$.

Multiplicative Extended Kalman Filter (MEKF)

- To linearize kinematics remember the quaternion error $q = \hat{q} \star \delta q$, with

$$\delta q(a) = \frac{1}{\sqrt{4 + \|a\|^2}} \begin{bmatrix} 2 \\ a \end{bmatrix}, \quad \dot{a} \approx \nu + a \times \hat{\omega} = -\hat{\omega}^\times a + \nu.$$

- Thus one can study the covariance of the vector a which represents the error:

$$\dot{P} = -\hat{\omega}^\times P + P\hat{\omega}^\times + Q, \quad P(0) = P_0$$

Multiplicative Extended Kalman Filter (MEKF)

- From the estimated quaternion $\hat{q}$ one can get $\hat{C}_N^B(\hat{q})$ (Euler-Rodrigues).
- Call δz_i the discrepance between measurement and expected measurement: $\delta z_i = \hat{v}_i^B - \hat{C}_N^B(\hat{q}) v_i^N$. If everything was perfect then $\delta z_i = 0$.
- Measurement is not pefect: $\hat{v}_i^B = v_i^B - \epsilon_i$.
- Estimation is not perfect: $\hat{C}_N^B = C_N^{\hat{B}} = C_B^{\hat{B}} C_N^B$.
- Thus $\delta z_i = v_i^B - C_B^{\hat{B}} v_i^B - \epsilon_i$.
- Remember that from the relationship between the error quaternion and the small angles DCM: $C_B^{\hat{B}} = I - a^\times$, thus $\delta z_i = -a^\times v_i^B - \epsilon_i = (v_i^B)^\times a - \epsilon_i$.
- Thus we have n measurements of error in the form $\delta z_i = H_i a - \epsilon_i$, where $H_i \approx (\hat{v}_i^B)^\times$. (NOTE: take only two rows to avoid invertibility issues). The covariance of the measurement is R_i.

Multiplicative Extended Kalman Filter (MEKF)

- Use the a priori $(-)$ and a posteriori $(+)$ notation. From integration we had $\hat{q}^-$ with error a^- whose covariance is P^-.
- With the measurements available from

$$\delta z = \begin{bmatrix} \delta z_1 \\ \vdots \\ \delta z_n \end{bmatrix}, H = \begin{bmatrix} H_1 \\ \vdots \\ H_n \end{bmatrix}, R = \begin{bmatrix} R_1 & & \\ & \ddots & \\ & & R_n \end{bmatrix}$$

- Using the measurements $a^+ = a^- + K(z - Ha^-)$, but since the mean of the error is zero, $a^- = 0$, thus: $a^+ = K\delta z$, where K is the Kalman gain, computed as $K = P^- H^T (HP^- H^T + R)^{-1}$. Covariance is updated as $P^+ = P^- - KHP^-$.
- With a^+ update $\hat{q}$:$\hat{q}^+ = \hat{q}^- \star \delta q = \hat{q}^- \star \begin{bmatrix} 2 \\ a^+ \end{bmatrix} \frac{1}{\sqrt{4+\|a^+\|^2}}$
- This procedure is iterated.

Multiplicative Extended Kalman Filter (MEKF)

- Summary. Initial data: $\hat{q}_0$, P_0, Q, R_i. One considers $\hat{\omega}$ continuous. Ocassionally, one gets measurements and thus can compute $\delta z_i = \hat{v}_i^B - \hat{C}_N^B(\hat{q}) v_i^N$.
  1. Initialize and compute $\hat{q}$ and P:
  $$\begin{aligned} \dot{\hat{q}} &= \frac{1}{2} q \star q_{\hat{\omega}}\,, \quad q(0) = q_0\,, \\ \dot{P} &= -\hat{\omega}^{\times} P + P \hat{\omega}^{\times} + Q, \quad P(0) = P_0 \end{aligned}$$
  2. At time $t = t_k$ one gets measurements, call $\hat{q}^- = \hat{q}(t_k)$ and $P^- = P(t_k)$. Compute δz, H, R. Compute $K = P^- H^T (H P^- H^T + R)^{-1}$. Compute $a^+ = K \delta z$. Update $\hat{q}^+ = \hat{q}^- \star \delta q = \hat{q}^- \star \begin{bmatrix} 2 \\ a^+ \end{bmatrix} \frac{1}{\sqrt{4 + \|a^+\|^2}}$, $P^+ = P^- - K H P^-$.
  3. Keep integrating the equations from the a posteri estimations until more measurements arrive:
  $$\begin{aligned} \dot{\hat{q}} &= \frac{1}{2} q \star q_{\hat{\omega}}\,, \quad q(t_k) = q^+, \\ \dot{P} &= -\hat{\omega}^{\times} P + P \hat{\omega}^{\times} + Q, \quad P(t_k) = P^+ \end{aligned}$$
  4. When new measurements arrive, go back to 2.

Multiplicative Extended Kalman Filter (MEKF)

- Additional ideas:
  - Don't forget to renormalize $\hat{q}(t)$ if modulus goes away from unity.
  - The covariance matrix $P(t)$ must be symmetric. One can "symmetrize" by forcing $P = 1/2(P + P^T)$, or compute only a triangular matrix and impose the rest is the transpose.
  - The Kalman gain is optimal only for the linearized system. If estimation has large errors, the filter may diverge.
  - One can and should include gyro bias in the estimation.
  - In practice it is not so easy to obtain Q and R so some simulation/experiments are required.
- Other filtering algorithms exist. MEKF is "simple" and flexible but not necessarily the best (this is a research field).
- In a lab we will test the MEKF with a cell phone.