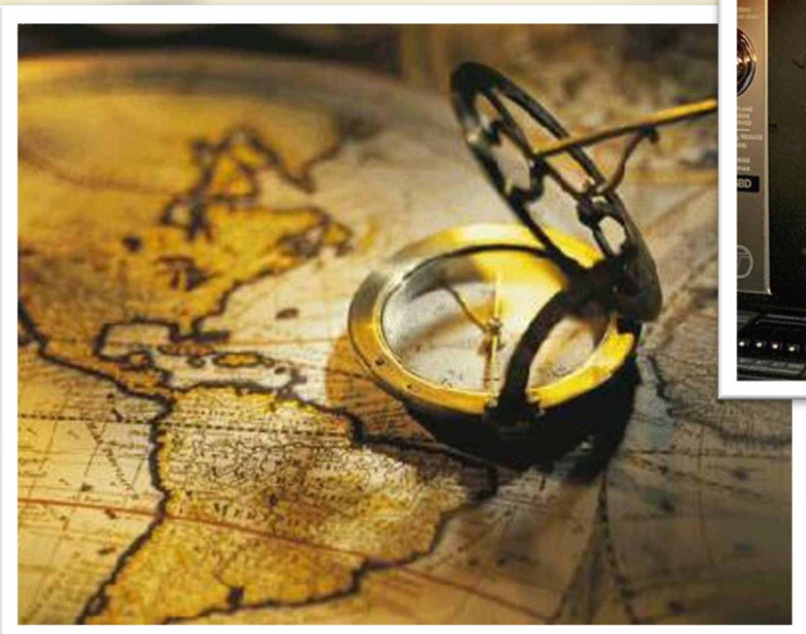
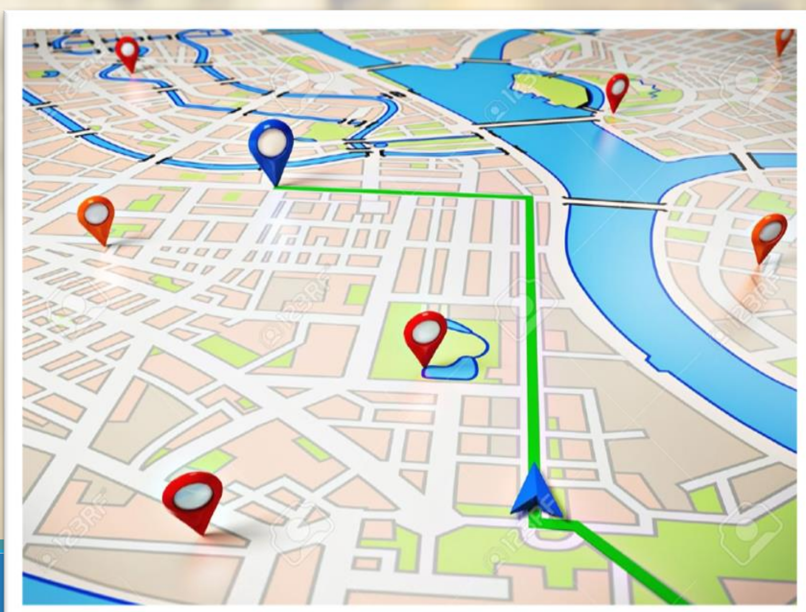




Lecture

KALMAN FILTERING

- practical considerations -

- dynamic detection -

Kalman filtering – divergence and effective cure

Continuous Kalman filters

- **rarely implemented in the real world**
  - **IF** measurements continuously available
  - **AND** state estimates needed continuously
  - **THEN** a computer is needed to solve the simultaneously following differential equations:

$$\dot{\hat{x}} = A\hat{x} + K[z - H\hat{x}];\ \hat{x}(0) = \hat{x}_0$$

$$K = PH^T R^{-1}$$

$$\dot{P} = AP + PA^T + GQG^T - KRK^T$$

- **are nevertheless useful**
  - for theoretical and analytical studies to illustrate filter behaviors.
- **provide solutions for limiting case of discrete filters when $T_s \rightarrow 0$**

Kalman filtering – divergence and effective cure

Discrete Kalman filters

- **almost always implemented**
  - **SINCE** large scale, high speed, cost effective digital computers are readily available
- can handle the continuous system form
  - with measurements available on either a continuous or discrete basis
  - as long as state estimates are not needed continuously, but only every T_s in (s)
    - T chosen sufficiently small that states cannot change much between two consequent estimates

Kalman filtering – divergence and effective cure

Discrete Kalman filters

- **problems occur in the real world by implementation of the digital filters**

Problems

1. Filter theory does not address **the scheduling problem** (when to take measurements)
2. Filter theory does not address **what kind of measurements**/sensors to select
3. Optimal filters have **undesirable characteristics**
4. Optimal filters are usually **unachievable,** filter may not converge to "the right answer"
5. the system **may be nonlinear** – not linear as assumed in Kalman filter theory – the real world is not linear and thus linearization (often by 1st order) needs to be perform
6. **suboptimal filters** may be preferable
7. **bad data** may cause bad estimates
8. filter may experience **numerical problems**

Scheduling problem

- theory assumes measurements to be regularly scheduled
- theory can handle case of "missing" measurement from a regular schedule
  - extrapolation with a state transition matrix or by setting K = 0
- the fewer the number of measurements, the more critical the schedule becomes

Measurement/sensor selection

- Kalman filter theory assumes that **the number, type and noise characteristics** of the measurements/sensors **<u>are known</u>** parameters (a priori defined)
  - noise parameters from all available sensors are optimally combined
  - BUT the theory does not directly address the question how to choose these parameters
- HOWEVER, generally there can be said:
  - the more the sensors, the smaller the estimation error
  - the lower the noise level in the sensors, the smaller the estimation error

<u>SOLUTION</u>

- construct a set of feasible sensor combination
- examine the cost and performance of each → choose the best tradeoff

Undesirable characteristics

- optimal filter often have undesirable characteristics
  - a heavy computer load
  - strong sensitivity to modelling errors
- computer load
  - truly optimum filters must model all states and random processes which are present
    - cannot ignore states or random processes representing “small effects”
    - cannot make approximation
    - consequently, the number of states may be very large → large amount of a computer time and memory
    - to avoid approximation and cover small effects the number of states are generally doubled

SOLUTION

- a suboptimal filter is often needed

Undesirable characteristics

- strong sensitivity
  - optimal filters require detailed and precise knowledge of the structure and parameters of the following model:
    - ❑ system state equations $\mathbf{\Phi}$
    - ❑ measurement equations **H**
    - ❑ system driving noise $\mathbf{G^*Q^*G^T}$
    - ❑ measurement noise **R**
    - ❑ a priori covariance $\mathbf{P_0(+)}$
  - to obtain lowest estimation error, the optimal filter makes maximum use of all stated above
  - as a result, the optimal filter becomes "supertuned" and highly sensitive to uncertainties (errors) in the modeling. **THESE UNCERTAINTIES ALWAYS EXIST!**

SOLUTION

- reduced estimation error is achieved in the optimal filter at the price of increased sensitivity
- tradeoff between estimation error and sensitivity must be considered

Optimal filters - unachievable

- a filter is optimal only if:
  - all the required models are precisely known
  - the real world behaves exactly as modelled
  - the entire optimal filter algorithm is implemented
- BUT:
  - models are not precisely known
    - not available
    - practically unobtainable
    - involve approximation
  - systems or measurements have some nonlinearities
  - computer load is excessive for optimal algorithm

CONSEQUENCES

- filters are almost always suboptimal
- we need methods for:
  - intentionally suboptimizing filters to achieve reduced computer load and parameter sensitivity
  - evaluating performance degradation
    - P_k is no longer correct
  - doing the sensitivity analyses
  - modifying the filter to handle nonlinearities

Divergence/non-convergence/convergence

- Definition of symbols

$\hat{x}$ – filter estimate of the system state

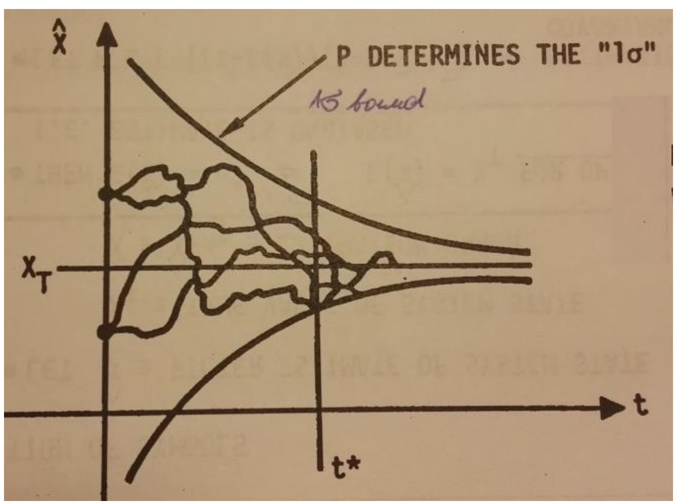
x_T – true value of the system state

$\tilde{x} = \hat{x} - x_T$ – estimation error

- THEN $E(\tilde{x}) = 0$ ⇒ $E(\hat{x}) = x_T$ for an optimal filter

i.e. the estimate is unbiased

$$P = E[(\tilde{x} - E(\tilde{x}))(\tilde{x} - E(\tilde{x}))^T] \text{ – estimation error covariance matrix}$$



NOTES:

- each time the filter is run, a different $\hat{x}(t)$ is generated no matter the initial conditions
- at time t^* the average (across the ensemble of estimates) estimation error is zero, $E(\hat{x} - x_T) = 0$

Divergence/non-convergence/convergence

- For an optimal filter the convergence/divergence is correctly and fully defined by P(t).
- in general, a suboptimal filter's estimate is biased
  - THEN, the behavior of P(t) is not enough to define the convergence

Suboptimal filters converge **<u>if and only if</u>**:

- $\lim_{t \to \infty} (\text{trace } P) = 0$

AND

- $\lim_{t \to \infty} E(\tilde{x}) = 0$

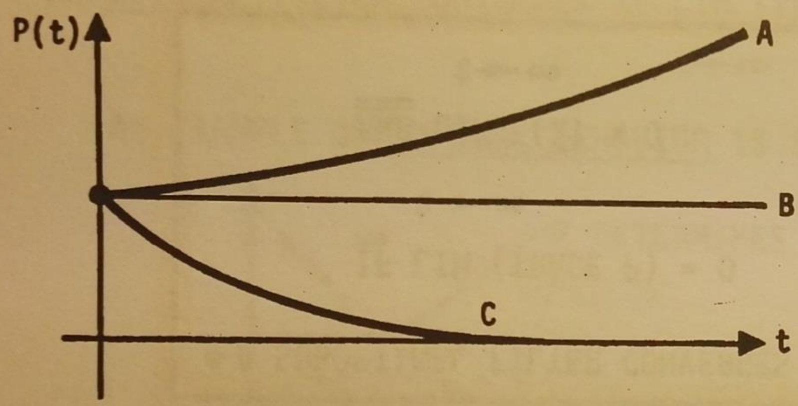
Non-convergence can mean:

- divergence: $\lim_{t \to \infty} (\text{trace } P) = \infty$ $\mathbf{OR}$ $\lim_{t \to \infty} \|E(\tilde{x})\| = \infty$
- neither convergence or divergence

Divergence/convergence

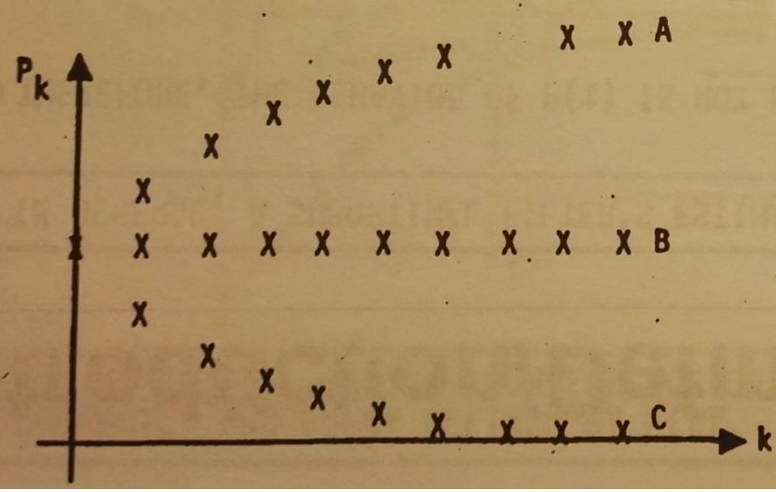
Some typical behavioral patterns for P(t)

- BETWEEN MEASUREMENT SAMPLES



CASE	TYPICAL CAUSES
A	SYSTEM UNSTABLE $\dot{X} = AX \quad A > 0$ OR SYSTEM HAS DRIVING NOISE $\dot{X} = AX + W$
B	SYSTEM HAS CONSTANT STATE $\dot{X} = 0$
C	SYSTEM IS STABLE AND HAS NO DRIVING NOISE $\dot{X} = -AX \quad A > 0$

- IMMEDIATELY AFTER MEASUREMENTS ARE PROCESSED

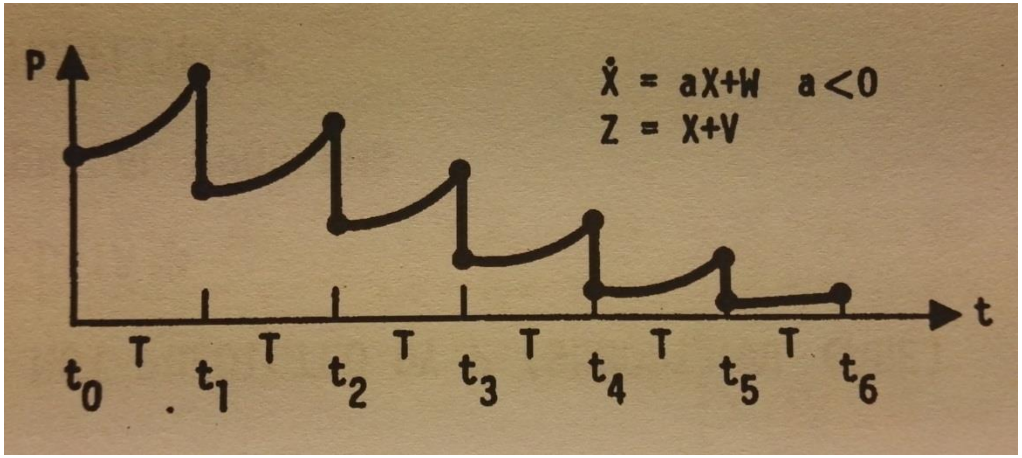


CASE	TYPICAL CAUSES
A	SYSTEM DRIVING NOISE AND MEASUREMENT NOISE ARE LARGE RELATIVE TO P_0
B	STATE IS UNOBSERVABLE AND UNCORRELATED WITH OTHER STATES
C	SYSTEM DRIVING NOISE AND MEASUREMENT NOISE ARE SMALL RELATIVE TO P_0

Divergence/convergence

An example of a combined behavioral pattern for $P(t)$

- between and at the time of measurements



NOTES:

- processing the measurement tends to reduce P
- the larger Q parameters are, the lower the overall estimation accuracy becomes
- system driving noise tends to increase P
- an unstable system tends to increase P
- with white measurement noise the time between samples can be shortened to reduce P
- the behavior of P represents a composite of all of these effects and often reaches a “statistical equilibrium”.

Divergence/convergence

Non-convergence categories

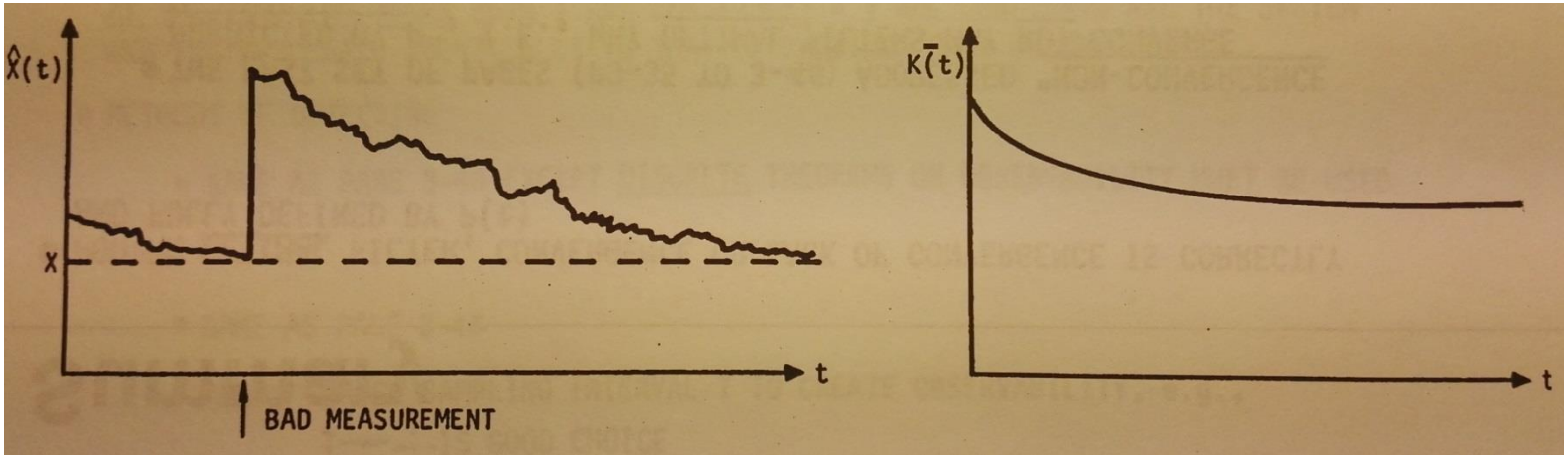
- **<u>predicted by P</u>** (optimal case)
  - ❑ as "natural behavior" - $\lim_{t \to \infty} P(t) \to P_{constant}$
  - ❑ due to non-observability - $\lim_{t \to \infty} P(t) = \infty \textbf{ OR } \neq 0$
    - CURES:
      - create observability by adding another type of measurement
      - increase the sampling rate of measurement (! aliasing problem)

- **<u>not predicted by P</u>** (suboptimal case)
  - ❑ due to bad data
  - ❑ due to numerical problems
  - ❑ due to mismodeling

Divergence/convergence

Bad data

- can occur in two forms:
  - ❑ $\hat{x}(0)$ the initial estimate is badly chosen, e.g. $|\hat{x}(0) - x_T| = |\tilde{x}| \gg 2\sigma(0)$
  - ❑ the measurement z_k has an error excessively large, e.g. $|v| \gg 2\sigma_v$
- in either case, if the system is truly linear, the Kalman filter (in theory) will recover as T goes large.



The best way is to prevent bad data from getting into the filter in the first place!

Divergence/convergence

Bad data:

- in the real world recovery often does not occur
  - ❑ the interval (0, T) of measurement available is fixed and may be too short to allow sufficient recovery
  - ❑ a normal behavior of the gain matrix K may be too rapid convergence towards its steady-state value
- DETECTION:
  - ❑ inspection of P(t) and K(t) is useless
  - ❑ inspect the state estimates for sudden jumps (after bad measurement has already been used by a filter)
  - ❑ inspect the innovations vector for sudden jumps or large entries (before bad measurement is processed by a filter)
- CURE:
  - ❑ implement "bad data detector" and reject bad data before use of a filter
  - ❑ remove bad data from filter estimates
    - the simplest way is to save the old estimate $\Phi\hat{x}_{k-1}(+)$ and skip the data point by setting K=0
  - ❑ keep filter alive by adding the process noise w in the system model – improve recovery

Divergence/convergence

Numerical problems may cause not converging to the “right answer”.

- DETECTION:
  - sometimes detected by impossible P_k behavior
    - terms on main diagonal of P_k go negative
    - terms on main diagonal of P_k are larger immediately after measurement is processed than immediately before, i.e. $\sigma(+) > \sigma(-)$
- CAUSES:
  - usually inadequate word length (precision) of a computer
  - problems become worse with large number of state variables
- CURES:
  - brute force
    - use higher precision
    - reduce number of states
      - merge or eliminate unobservable states
      - eliminate states representing “small effects”
      - use suboptimal filters, e.g. 3x 20-state filter instead of 1x 60-state filter
    - use square-root filtering to obtain more computational accuracy with same computer precision.

Divergence/convergence

Mismodeling

- P matrix is correct only if model used in filter is correct
- with mismodeling P matrix:
  - is erroneous
  - is of little use in detecting non-convergence
  - P can converge to zero with state estimate diverging
- numerous kinds of mismodeling can cause non-convergence, e.g.:
  - unmodelled states
  - unmodelled system driving noise
  - wrong parameters in system description matrices
  - nonlinearities
  - bad initial conditions
  - wrong correlation time use

Divergence/convergence

Unmodelled states

- can cause non-convergence as long as $K(t) \rightarrow 0$ as $t \rightarrow \infty$ and thus filter is unable to track changes $\rightarrow$ "sleepy filter syndrome"
- DETECTION:
  - ❑ P(t) will not indicate any problem exists
  - ❑ $\hat{x}(t)$ will not look particularly unusual
  - ❑ innovation vector will show divergence and presence of “signals” other than white noise
- CURES:
  - ❑ properly model all states
  - ❑ add “fictitious” process noise to the system model
    - the gain K(t) does not go to zero as before
    - the filter improves its capability to track $P(\infty) = \sqrt{rq}; K(\infty) = \sqrt{\frac{q}{r}}$
    - if results biased, add “fictitious” noise as a state of the model
      - transient response may become more sluggish

Divergence/convergence

Wrong parameters

- DETECTION:
  - ❑ P(t) will not indicate any problem exists
  - ❑ $\hat{x}(t)$ will not look particularly unusual
  - ❑ innovation vector will show divergence/non-convergence
- CURES:
  - ❑ make sure right parameter values are used in the filter model
    - in the real world it is often impossible to do it precisely since right values are not known and can be only estimated
  - ❑ if the amount of degradation obtained is unacceptable, consider letting unknown parameter itself become a state variable and thus estimated

Divergence/convergence

Nonlinearities can cause in the real world system non-convergence/divergence.

- a Kalman filter formulation assumes the real world system state and output equations to be linear
- HOWEVER, the real world is almost always nonlinear, although linearity can be a good approximation
- the number of possible nonlinearities are infinite
- IMPACT:
  - the Kalman filter estimate is
    - not optimum
    - not unbiased
    - non-convergent or divergent
  - P matrix is
    - incorrect
    - optimistic
- the impact of nonlinearity and the character of the convergence/divergence depends on the distance from the state variable from its expected value
- CURES:
  - use of
    - extended Kalman filter
    - higher-order approximations

Divergence/convergence

Extended Kalman filter

- do not linearize about some a priori defined fixed point
- do continuously linearize about the best available estimate
- although the EKF may show greatly improved convergence relative to the linearized filter
  - for large deviations between linearization point and the state expected value non-convergence can occur
  - the P matrix is totally unreliable as an indication of estimation error for large deviations between linearization point and the state expected value

Divergence/convergence

Evaluating filter performance

- the Kalman filter P matrix does not address for nonlinear cases
  - region of convergence
  - the estimation error covariance
  - the estimation error mean value $E(\tilde{x}) \neq 0$

- approximation techniques are commonly used to handle above stated problem
  - for the case of "state as close as possible" the linearization point or trajectory-treat P matrix as the "approximate estimation error covariance matrix"
    - generally, this only provides a lower bound
    - no information is provided on the region of convergence
    - no information is provided on the estimation error mean value

Suboptimal filtering

Techniques for suboptimizing linear filters

- modify the Kalman gain K
- modify the filter's model of the real world

BUT

- regardless of the method used
  - the P matrix is not the true estimation error covariance
  - the estimation error, in general, becomes biased

Suboptimal filtering

Modified Kalman gain

- Wiener filters
  - although the gain vector K is time-varying, sometimes it quickly reaches a constant non-zero value
  - $K(t) = K(\infty)$
  - if in addition, the matrices A, C are time-invariant, the matrix of transfer functions characterizing the Wiener filter can easily be computed
  - DISADVANTAGES:
    - cannot use this approach if $K(\infty) \neq$ constant OR $K(\infty) = 0$
    - typical penalty is poorer transient response
- Use of approximation function for the modification
  - ADVANTAGES over Wiener filters
    - can handle cases where $K(\infty) \neq$ constant AND $K(\infty) = 0$
    - close to optimal performance

Suboptimal filtering

Modified filter's model

- usually the intent is to make the filter's model less complex than an actual system is
- ways to do so:
  - ignore some states
  - prefilter to attenuate some states
  - decouple states when their coupling is weak
  - combine two states if they are "close functions" over the entire measurement interval

Prefiltering and data rejection

Prefiltering may be beneficial for several purposes

- to allow a discrete Kalman filter to be used on a continuous system without throwing away measurements
- to attenuate some states so they can be ignored then in a suboptimal filter
- to reduce the iteration rate in the discrete filter thus saving “computer time”
- to reject bad data so the estimate is not degraded

Data rejection filter

- assuming that adequate knowledge of the innovation exists
- THEN data rejection filter can be implemented as:
  - excess amplitude
    - $if\ |(z - H\hat{x})| > A_{max} \rightarrow \text{reject data}$
  - excess rate of change
    - $if\ |(z - H\hat{x})_i - (z - H\hat{x})_{i-1}| > \delta A_{max} \rightarrow \text{reject data}$
- when data is intended to reject than K=0

Prefiltering and data rejection

Chi-squared statistic

- detecting anomalous sensor data
- Kalman gain $K_k = P_k(-)H^T \underbrace{(HP_k(-)H^T + R_k)^{-1}}_{Y_{vk}}$
- Y_{vk} is the matrix of innovations
- associated likelihood function for innovations $L(v_k) = \exp(-\frac{1}{2}v^T Y_{vk} v)$
- log-likelihood is then $log[L(v_k)] = -v^T Y_{vk} v$, which can be easily calculated
- THEN detecting can be performed based on

$$\chi^2 = \frac{v^T Y_{vk} v}{No.of\ measurements}$$

- an upper limit threshold value on χ^2 can be used to detect anomalous sensor data , but its practical value should be determined according to designer system observation.
- its range should be determined by monitoring the system in operation.

Dynamics detection

There exist several approaches; however, one of the most common can be done according to evaluating conditions of acceleration or angular rate changes.

CONDITION when accelerometers' reading can be used for attitude compensation

$$\|\vec{a}\| = \sqrt{{a_x}^2 + {a_y}^2 + {a_z}^2} = 1 \ \ (\text{in } g) \rightarrow \text{only gravity is present in the readings.}$$

Above statement can be kept true in theory, but in real applications where vibration is affecting the readings a **threshold** should be set as follows

$$abs\ (\|\vec{a}\| - 1) < a_t, \text{where } a_t \text{ can be chosen according to specific conditions.}$$

Above statement can be accompanied by following **CONDITION**

$$\|\vec{\omega}\| = \sqrt{{\omega_x}^2 + {\omega_y}^2 + {\omega_z}^2} < \omega_t,$$

where ω_t is unbiased value and can be chosen according to specific application.

Be aware that all above is the simplest approach, however, commonly used, but has limitations.

Dynamics detection - implementation

Commonly there are three approaches:

- **FIRST** – when $abs\ (\|\vec{a}\| - 1) > a_t$ set Kalman gain K=0 for particular row.
  This approach is too rough and does not allow adaptation of the fusion.

- **SECOND** – before each correction update there is needed to modify a covariance matrix of the measurement – R in particular rows where ACC readings are put the way as suggested below

  $$R_{ii} = {\sigma_a}^2 + k(\|\vec{a}\| - 1)^2$$

  Additive function noise $k(\|\vec{a}\| - 1)^2$ is optionally included to increase the ACC noise proportionally to the increments of the body frame acceleration. The increment of measurement noise decreases the influence of ACC corrections when conditions are not totally stationary.
  k can be any positive number appropriately chosen to increase R and thus decrease the effect of ACC based compensation the way that for large $(\|\vec{a}\| - 1)$ the compensation effect comes to approx. zero.

- **THIRD** – a combination of above two approaches.