# Lecture Notes in Inertial Sensor Fusion

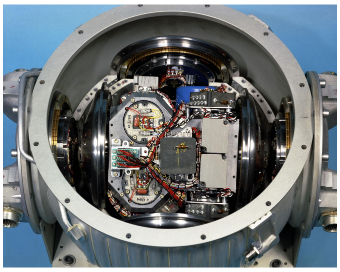
Thomas Seel

**Organizational Remarks:**

- introduced lecture in 2017 due to large demand and a lack of teaching in this technological field
- encouraged by colleagues from biomedical engineering, aerospace engineering, automotive engineering and robotics
- content matched with highly recommendable neighboring courses such as
  - "Lageregelung von Raumfahrzeugen" (FG Raumfahrttechnik),
  - "Rehabilitationstechnik II" (FG Medizintechnik),
  - "Stochastische Modelle" (FG Stochastische Analysis),
  - "Probabilistic and Bayesian Modelling in ML and AI" (FG Künstliche Intelligenz)
  - or several courses in control engineering taught by our group
- methods/tools being taught include
  - 3D gyroscopes, accelerometers and magnetometers
  - representations of 3D rotations: rotation matrices, Euler angles, quaternions
  - some very basic group theory
  - dead reckoning, strapdown integration, zero-velocity updates
  - orientation and joint angle estimation with IMUs
  - probabilistic modeling and linear kalman filter
  - nonlinear filters: extended and unscent kalman, particle filter
  - Bayesian state estimation: smooting and filtering
  - parameter identification in kinematic systems
- applications considered include:
  - human motion tracking
  - state estimation and navigation of aerial/ground vehicles
  - control of robotic actuators
  - ...
- strongly research-oriented course with computer exercises in which you make the methods work on real data and write extended abstracts about your methods and results
- taught by Daniel Laidig, Dustin Lehmann and me; supported by selected visiting lecturers
- taught in English unless a strong majority wants German
- feedback is highly appreciated, please help us to further improve the course for generations to come; talk to us or use anonymous form on ISIS
- ISIS course link is found on course website, registration password: quaternion

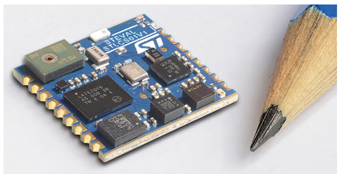
### What are Inertial Measurement Units?

- motion sensors that are used in a rapidly increasing number of applications systems
- IMUs use a combination of accelerometers, (rate) gyroscopes and magnetometers, typically 3 (linearly independent) axes each
- track the position and orientation of an object relative to a known starting point, orientation and velocity
- advantage: no interaction force (as in encoders), no line-of-sight restriction (as in optical methods)
- example: cars, missiles, aircrafts, submarines, robotic actuators, human body segments, crane systems
- size weight and cost have decreased dramatically since the first IMUs



**(a)** Apollo IMU

**(b)** Aircraft IMU (AHRS)



**(c)** MEMS IMU

### What is Sensor Fusion?

- variables/parameters of interest are not always directly measurable
- measurements are often subject to errors (noise, bias, nonlinearity, ...)
- multiple measurements / sources of information may be contradicting
- sensor fusion aims to merge data from multiple sources in a way that reduces uncertainty
- this fusion is often based on dynamical system models and/or probability theory

### What is Inertial Sensor Fusion?

- attach IMUs to objects of interest (single objects or kinematic chains)
- combine the measurements on an IMU and potentially other sensors (IMU+x)
- estimate orientations, velocities and positions of objects
- overcome the limitations of each individual IMU component
- some examples:
  - use IMU+GPS to estimate the orientation and position of an aircraft in foggy weather

- use IMU+barometer to let a drone hold its position automatically
- use IMUs to reduce overshoot in crane systems even for loading cargo on the high seas
- use IMU+camera+lidar+... to control and navigate autonomous cars
- use IMUs and kinematic models to track and control robotic actuators
- use wearable IMUs to track human motions and control neuroprostheses
- ...

# 1 Gyroscopes, Accelerometers and Magnetometers

## 1.1 Gyroscopes

- sensors that measure (change of) orientations
- only main types of gyroscope are presented, no exhaustive list

### 1.1.1 Mechanical Gyroscopes

**Idea:**

- isolate a platform from rotations of its surrounding shell
- such that platform remains aligned with the global frame even if shell rotates and moves

**Realization:**

- low-friction gimbals[1] and a rotating mass
- conservation of momentum assures that spin axis remains unrotated with respect to fixed frame
- two orientation angles can be measured, need a second rotation axis to measure third orientation angle.
- Examples: Mechanical Gyro, Walter Lewin Lecture

**Disadvantage:**

- typically large and bulky setups (mass-friction ratio)
- difficult to eliminate friction-related error
- need to speed-up the fly wheel before measurement

**Question:** What is the global frame? (frame of reference, fixed frame, earth frame)

- earth rotation is 15° per hour, i.e. 0.004°/s.
- proven (for example) with magnetically levitating mechanical gyro[2]
- note that also earth rotates around sun in 365 days and sun rotates around Sagittarius A (center of our galaxy (milky way)) in ∼250 million years
- every frame with constant (zero or non-zero) velocity and zero rotational velocity is an inertial frame
- $\exists$ absolute rotational velocity, BUT $\nexists$ absolute translational velocity
- please recall basic concepts of special relativity

[1] A gimbal is a pivoted support that allows the rotation of an object about a single axis. The gimbal suspension used for mounting compasses etc. is sometimes called a Cardan suspension after Italian mathematician and physicist Gerolamo Cardano (1501–1576).

[2] https://doi.org/10.1016/S0921-4526(00)00753-5

### 1.1.2 Optical Gyroscopes

**Idea:**

- light travels at the speed of light with respect to the inertial frame of its source
- it should take shorter to travel a circular path against the direction of rotation than to travel it with the direction of rotation (Sagnac effect)

**Realization:** Sagnac interferometer

- use a large coil of optical fibre with a large number of loops
- split a laser beam to travel the coil in both directions and then look at interference
- Visualization of the Sagnac effect: Sagnac effect
- note the difference between the Sagnac setup and the Michelson-Morley experiment

**Disadvantage:**

- long light path (>100m) required for good sensitivity
- typical dimensions ∼10cm, weight ∼0.2kg, price ∼1000 EUR

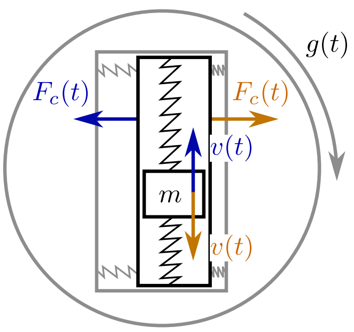
### 1.1.3 Micro-Electro-Mechanical-Systems (MEMS) Gyroscopes

**Idea:**

- use silicon micro-machining techniques
- exploit Coriolis effect (no spinning mass required)

**Realization:** (one of many options)

- make a mass vibrate along its x-axis and suspend it to allow deflection along its y-axis
- if the entire object rotates around the z-axis, the mass will vibrate along its y-axis
- after proper calibration, the absolute value of the angular velocity can be determined from the oscillation amplitude, while the phase difference between both oscillations yields the direction of rotation
- in MEMS chips, the same principle is exploited in tiny comb-like structures (video)



**Coriolis Effect:** A mass $m \in \mathbb{R}_{>0}$ moving with velocity $\mathbf{v}(t) \in \mathbb{R}^{3\times 1}$ with respect to a coordinate system that rotates at angular velocity $\mathbf{g}(t) \in \mathbb{R}^{3\times 1}$ experiences a force $\mathbf{F}_c(t) \in \mathbb{R}^{3\times 1}$, with

$$\mathbf{F}_c(t) = 2m(\mathbf{v}(t) \times \mathbf{g}(t)), \tag{1}$$

where the operator $\times$ denotes the vector (cross) product.

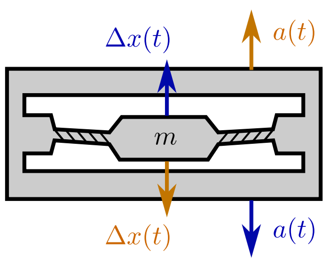
- acts on tradewinds that travel towards the equator and deflects them to the west (Congo–Ecuador faster flight than Ecuador–Congo)
- exploited 1851 by Léon Foucault to demonstrate earth rotation with a huge pendulum
- exploited in mass flow meters
- exploited by flies which have halteres to detect body angular velocity (video)

**Disadvantage:**

- less accurate, but major recent achievements
- continuous excitation required (energy consumption ∼1 mA at 3 V)

## 1.2 Accelerometers

- sensors that measure the specific force, which is a combination of the change of velocity and a gravity-related acceleration that is only zero in free fall
- typically a mass is suspended, for example, by a cantilever that is bent when the device is accelerated
- in MEMS chips, tiny flexible comb-like structures are used (video)
- when the accelerometer is at rest or when it is moving at constant velocity, the mass is only affected by gravity, i.e. it is drawn toward the center of earth, and the device measures the gravitational acceleration of approximately $9.8\,\mathrm{ms}^{-2}$ in vertical *upwards* direction
- thus, an accelerometer can yield information on the inclination (attitude) of an object with respect to the horizontal plane
- when the device moves, however, it measures the sum of this gravitational acceleration and the acceleration that is related to change of velocity

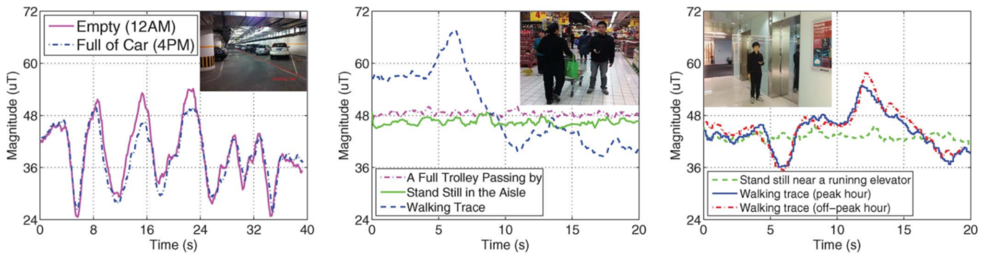


**Question:** What is the gravitational acceleration?

- equator: 9.78 $\mathrm{m/s^2}$, poles: 9.83 $\mathrm{m/s^2}$, at 45° latitude 9.806 $\mathrm{m/s^2}$
- 9.81693 $\mathrm{m/s^2}$ in Berlin[3],
  about 10% smaller in 408 km height, but practically zero on ISS (due to 15.5 orbits per day)
- pointing UP, due to measurement priciple (thought experiment: elevator in space)

## 1.3 Magnetometers

- sensors that measure direction and strength of the local magnetic field
- many measurement principles, exploiting for example the Lorentz force or the Hall effect
- main disadvantage: local magnetic field is easily disturbed in the proximity of ferromagnetic material and electronic devices
- this is a notorious problem in several other fields such as magnetic deviation in ships/airplanes




Variation in magnetic field strength by environmental influences. Source: Shu et al.

**Lorentz force:** A particle of charge $q$ moving with velocity $\mathbf{v}$ in the presence of an electric field $\mathbf{E}$ and a magnetic field $\mathbf{B}$ experiences a force

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}). \tag{2}$$

**Earth Magnetic Field:**

- strength varies within $[25, 65]\,\mu\mathrm{T} = [0.25, 0.65]\,\mathrm{Gs}$[4] ($\approx 50\,\mu$T in Berlin)
- dip angle is negative on northern hemisphere ($\approx -68°$ in Berlin)
- varies over the years and decades
  - ~6% strength per century,
  - ~50 km pole shift per year

[3]http://www.wolframalpha.com/input/?i=gravitation+Berlin
[4]Gauss is the cgs (centimeter-gram-second) unit of magnetic flux density.

# 2 Inertial Measurement Units

Two fundamentally different approaches to realize an IMU:

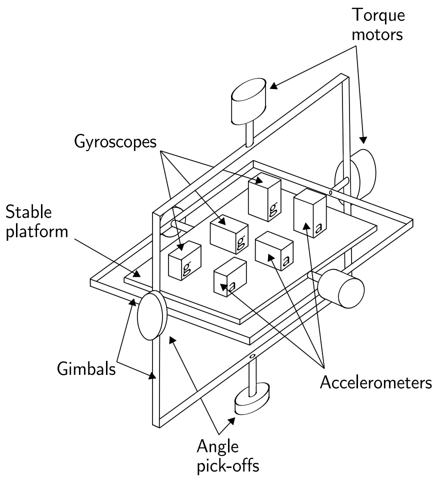
## 2.1 Stable-Platform Systems

**Idea:**

- isolate a platform from rotations of its surrounding shell
- such that platform remains aligned with the global frame even if shell rotates and moves

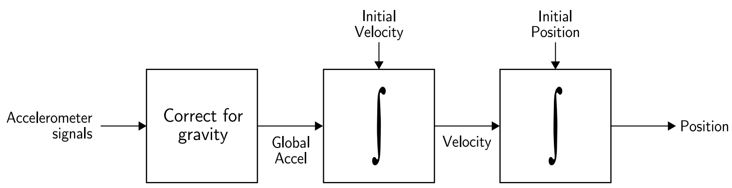
**Realization:**

- EITHER by combining two mechanical gyroscopes and adding accelerometers
- OR three actuated gimbals and feedback control based on (rate) gyroscope measurements




**Signal Processing:**

- orientation of the object can be measured by decoders in the gimbals
- velocity and position of the system can be calculated by integrating the accelerations measured on the stable platform (minus gravitational acceleration)



## 2.2 Strapdown Systems

**Idea:**

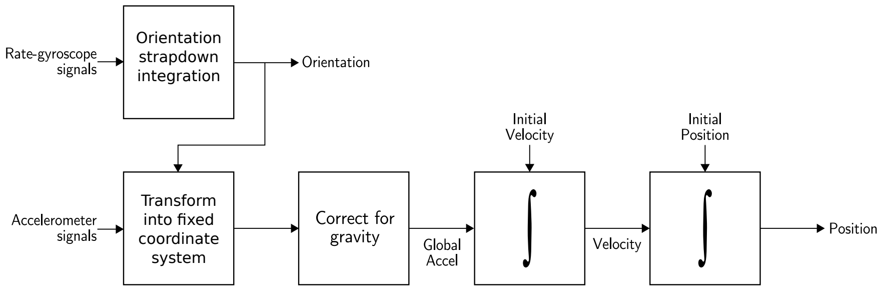
- reduce complexity of setup
- pay for it by increased complexity of signal processing

**Realization:**

- mount gyroscopes, accelerometers and magnetometers rigidly onto the moving body of interest
- a 9D MEMS IMU is a device that rigidly connects a 3D MEMS gyroscope, a 3D MEMS accelerometer, and a 3D MEMS magnetometer

**Signal Processing:**

- deduce orientation by strapdown integration of gyroscope
  (Orientation Strapdown Integration)
- transform into fixed frame, remove gravity and integrate accelerations
  (Position Strapdown Integration)

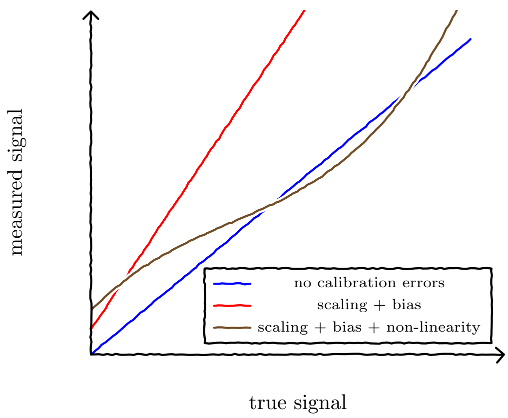


# 3 Error Characteristics of MEMS IMUs

## 3.1 Calibration Errors

- Calibration errors are deterministic constant errors that can be eliminated by calibration
- explanation uses example of gyroscope readings but equally holds for accelerometers and magnetometers
- assume a noiseless measurement for the moment and describe the difference between the true angular rate $\mathbf{g}_{\text{true}}$ and the measured angular rate $\mathbf{g}_{\text{measured}}$ by

$$\mathbf{g}_{\text{measured}} = {}^{\mathcal{A}}_{\mathcal{G}}\mathbf{M}\,\mathbf{C}_{\not\perp} \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & s_z \end{bmatrix} \mathbf{g}_{\text{true}} + \begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix} \tag{3}$$




Measured signal over the true signal. Blue: Reference course measurement with no errors.

### Bias

- bias is the constant offset $b_x, b_y, b_z$ error, i.e. it is what the sensor measures when the true value is zero
- when integrating measurements over time, bias leads to drift, i.e. an error that grows linearly with time
- when double-integrating measurements over time, bias leads to second-order drift, i.e. an error that grows quadratically with time
- orientations/velocities that are determined by integrating gyroscope/accelerometer readings are thus prone to drift
- positions that are determined by double integration of accelerometer readings are thus prone to second-order drift

→ (in noisy measurements) bias can be determined, for each sensing axis, as the (long-time average of the) sensor reading at a true value of zero

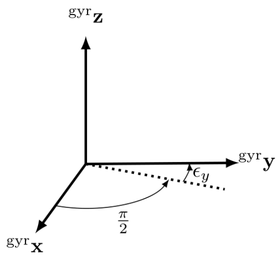
### Scaling and Nonlinearity

- scaling $s_x, s_y, s_z$ is the ratio between the true (non-zero) value and the measured (non-zero) value; ideally it is 1
- for scalings smaller than one, large true values are underestimated by the measurement
- nonlinearity is when the scaling $s_x, s_y, s_z$ depends on the values of $\mathbf{a}_{\text{true}}$

→ for each axis, scaling and nonlinearity can be determined by plotting measured values over real values and observing the slope of the curve

### Non-orthogonality and Misalignment

- non-orthogonality $\mathbf{C}_{\not\perp}$ refers to the measurement axes not being pairwise orthogonal to each other




Example for non-orthogonality. The angle $\epsilon_y$ is the deviation from the orthogonality

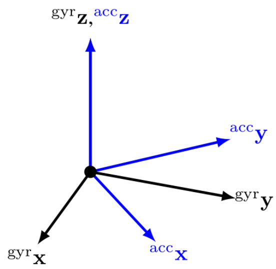
→ non-orthogonality can be observed, for example, by

  – attach the IMU to a precisely manufactured cuboid and first let it rest on each side, then rotate it on each side
  – check that the angle between the accelerometer/gyroscope readings taken at rest/rotation is actually $\frac{\pi}{2}$

- misalignment means that the measurement axes $\mathcal{G}$ of the gyroscopes do not coincide with the axes $\mathcal{A}$ of the accelerometers or those of the magnetometers, and so on

→ misalignment can be observed, for example, by rotating the IMU around a constant axis

  – determine the rotation axis coordinates by normalizing the (integral over the) measured angular rate (should be a constant direction, integrate only to remove noise)
  – check whether the scalar product of the constant rotation axis $\mathbf{j}$ and the measured acceleration $\mathbf{a}(t)$ is constant over time $t$




Example for misalignment between the gyroscope and the accelerometer

**Calibration**

- the aformentioned deterministic errors can be eliminated to a (very) large extent by (potentially laborious) calibration measurements with a reference measurement system (e.g. actuated gimbals)
- the measured datasets can be used to determine all calibration parameters by optimization:

$$\underset{{}^{\mathcal{A}}_{\mathcal{G}}\mathbf{M},\mathbf{C}_{\not\perp},\mathbf{S},\mathbf{b}}{\operatorname{argmin}} \sum_{i}^{N\gg 1} \left( \mathbf{a}_{\text{measured}}(t_i) - {}^{\mathcal{A}}_{\mathcal{G}}\mathbf{M}\,\mathbf{C}_{\not\perp} \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & s_z \end{bmatrix} \mathbf{a}_{\text{true}}(t_i) - \begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix} \right)^2 \tag{4}$$

- note that there are also calibration techniques that do not require a ground-truth reference measurement but exploit other facts
- for example, the accelerometer/magnetometer readings can be calibrated to have constant norm (i.e. to lie on a three-dimensional sphere) for measurements in which the IMU is slowly rotated to take very many different orientations

## 3.2 Turn-On Bias

- bias changes every time the sensor is powered-up (turn-on-to-turn-on bias), can thus only be removed on-the-fly (given some sufficiently long rest period)
- typical order of magnitude: $1°/\mathrm{s}$ for MEMS gyroscopes and $0.1\ \mathrm{m/s}^2$ for MEMS accelerometers

## 3.3 Thermo-Mechanical Noise

- fluctuations at frequencies much higher than sampling frequency
- leads to white noise in measurement output
- (double) integration of white noise yields a (second-order) random walk
- typical order of magnitude:
  - angle random walk of $5°/\sqrt{\mathrm{h}}$ (= e.g. 10° after 4 hours)
  - velocity random walk of $0.1\,\mathrm{m/s}/\sqrt{\mathrm{h}}$ (= e.g. $0.2\,\mathrm{m/s}$ after 4 hours)

**Integrating white noise over time**

- denote the measurements $m_i$ with sampling index i
- each $m_i$ is identically distributed with zero mean and with variance $\mathrm{Cov}(m_i, m_i) = \sigma^2 \ \forall i$
- by the definition of white noise, all $m_i$ are independent of each other, i.e.

$$\mathrm{Cov}(m_i, m_j) = 0 \ \forall i, j$$

- discrete-time integration yields the sum $(m_1 + m_2 + m_3 + \ldots)T_s$ with $T_s$ being the sampling period
- facts from stochastics, for two random variables $X, Y$ and two constant scalars $a, b$:

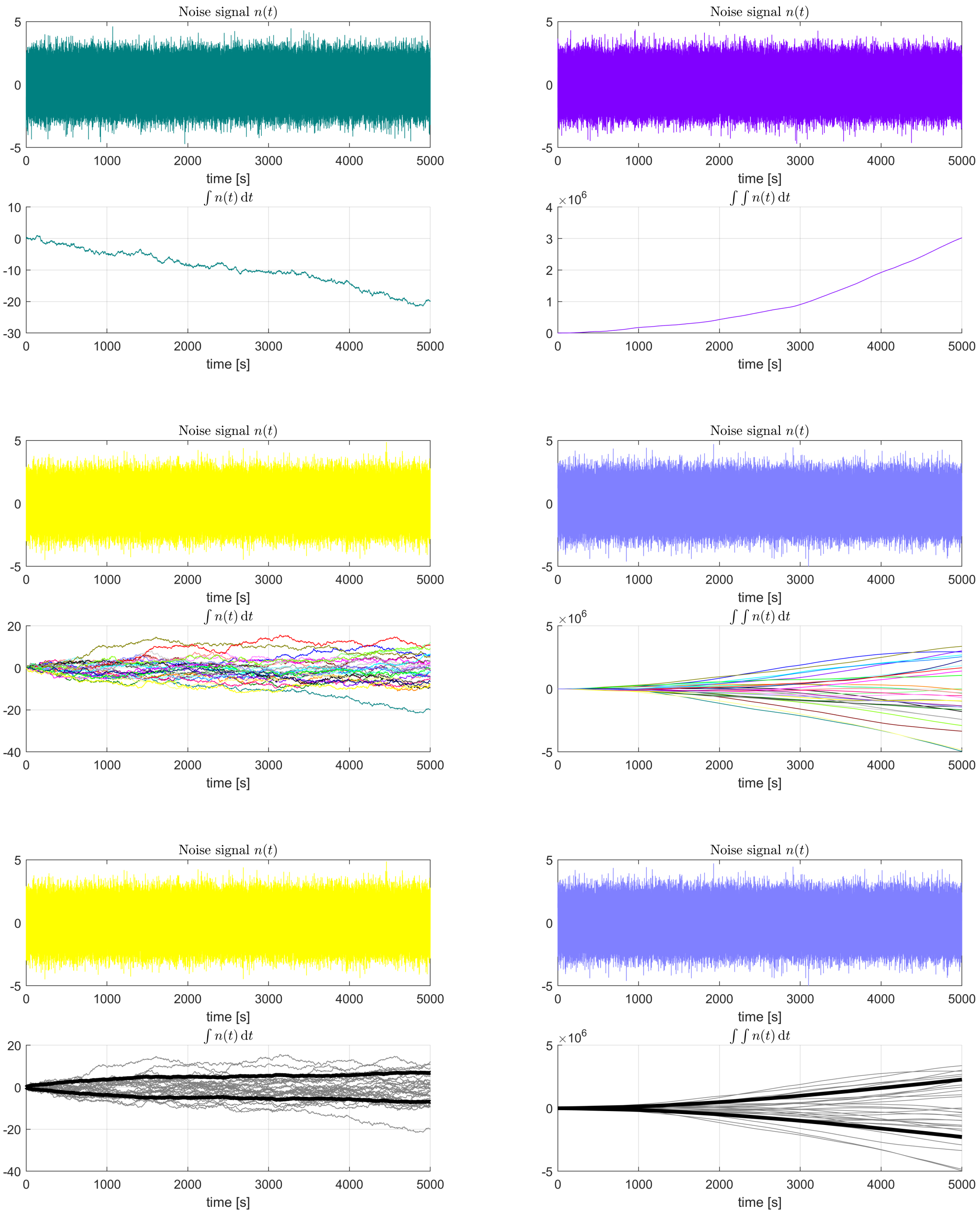
$$\mathrm{E}(aX + bY) = a\mathrm{E}(X) + b\mathrm{E}(Y)$$

$$\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + 2ab\mathrm{Cov}(X, Y)$$

- therefore, the expectation of the integral is zero, but its variance is

$$\mathrm{E}\left(\sum_{i=1}^{N} m_i T_s\right) = 0, \qquad \mathrm{Var}\left(\sum_{i=1}^{N} m_i T_s\right) = N\sigma^2 T_s^2 = \underbrace{NT_s}_{t} T_s \sigma^2,$$

  i.e. the standard deviation of the integral is $\sigma\sqrt{T_s}$ times the square root of $t$
- analogously, the standard deviation of the double integral is found to grow with $t^{3/2}$
- note the similarity with the concept of Wiener process used by physicists to explain Brownian motion; in this analogy, our white noise variance $\sigma^2$ is related to the diffusion coefficient of particles in a fluid

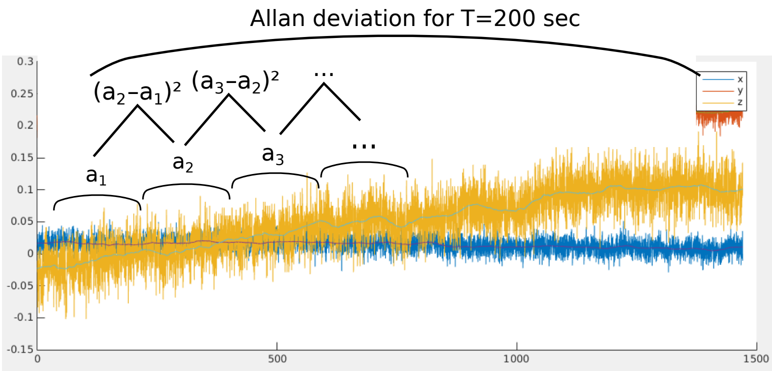



Single and double integration of white noise signal $n(t)$ for multiple trials. Left: Single integration $\int n(t)$. Right: Double integration $\iint n(t)$. Lowest images: Standard deviation (thick black lines) for all trials

## 3.4 Bias Instability

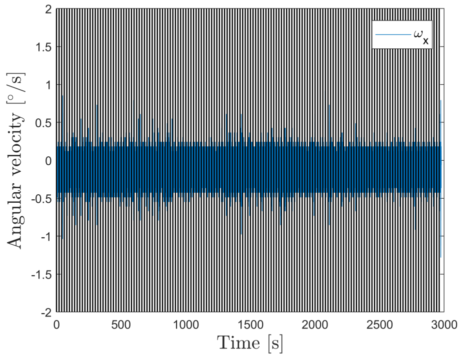
- due to flicker noise (pink noise) in the electronics and in other components susceptible to random flickering
- sometimes modeled by a random walk of the bias itself
- due to temperature changes

**Allan Variance:**

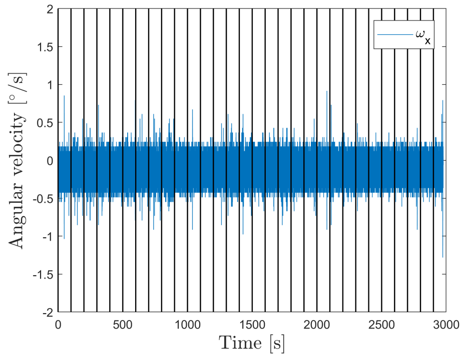



- take a long sequence of data and divide it into bins of length $t$
- average the data in each bin to obtain a list of averages $a_1, a_2, \ldots$
- then $a^2 = \text{avg}\left\{\frac{1}{2}(a_2 - a_1)^2, \frac{1}{2}(a_3 - a_2)^2, \ldots\right\}$ defines the Allan variance $a^2$

  and the Allan deviation $a$
- the $\frac{1}{2}$ assures that the variance $a^2$ is equal to the classical variance for a white noise signal
- See Stanley_1994 or AppNote 5087 for further informations on the derivation of the Allan variance
- repeat this for many very small to very large timespans and plot $\log a$ over $\log t$
- white noise appears on the plot as a slope with gradient $-0.5$ for small $t < 10$ s
- bias instability appears as a flat region around the minimum (typically at large $t > 100$ s)



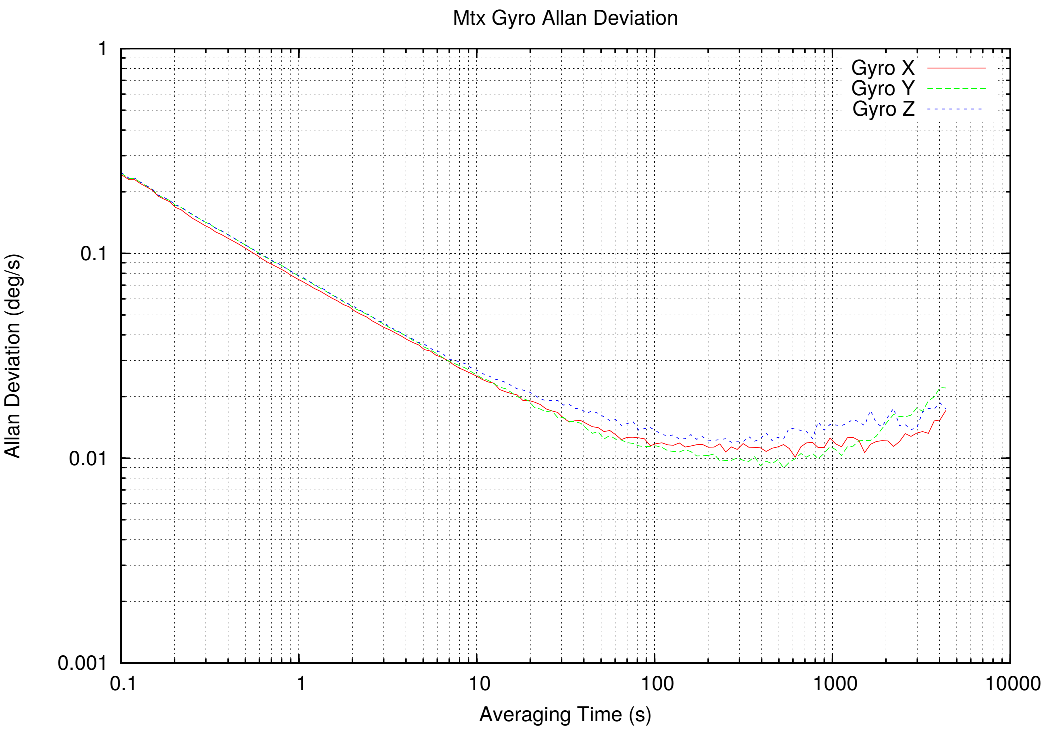

**(a)** $T_{\text{average}} = 20\,\text{s}$
$a^2 = 5.3 \cdot 10^{-6}$




**(b)** $T_{\text{average}} = 100\,\text{s}$
$a^2 = 4.2 \cdot 10^{-6}$

Different bin sizes for a long gyroscope measurement

**Example:**




- read at $t = 1$ s to obtain angle random walk of $0.075°/\sqrt{\text{s}} = 4.6°/\sqrt{\text{h}}$
- read at minimum to obtain bias instability of $0.01°/\text{s} = 36°/\text{h}$

### 3.5 Bias Correction Methods

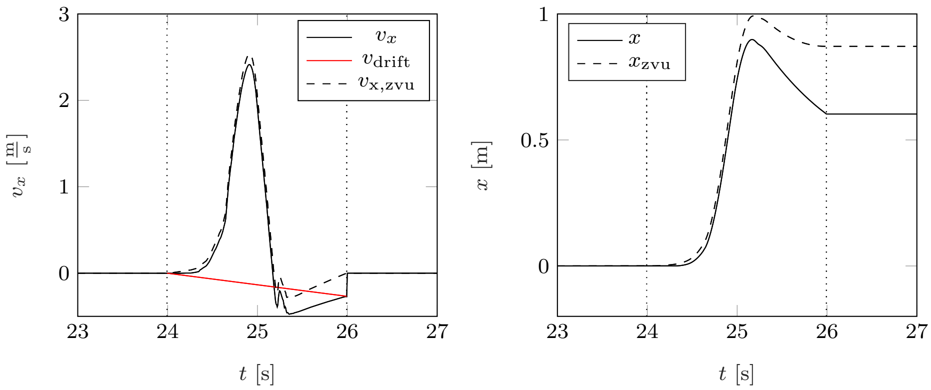
- static calibration can mitigate the effect of bias and drift drift only to a limited extend
- in some applications, additional assumptions on the performed motion can be exploited to compensate drift , e.g. regarding the true values of absolute/relative orientation/velocity/position at certain moments in time
- Examples:
  - in gait, we know that, after each stride, the foot is resting briefly (in the same inertial frame as before the step)
  - a robotic arm may always return to the same relative orientation of its segments after a certain amount of time
  - an object was moved and then placed in a different location but with the same vertical position (height)
- consider the case in which we know that the tracked object has the same velocity at two given points in time $t_\mathrm{A}$ and $t_\mathrm{B}$ and (without loss of generality) assume that velocity is zero, i.e. velocity $\mathbf{v}(t_\mathrm{A}) = \mathbf{v}(t_\mathrm{B}) = [0, 0, 0]^\intercal$
- we exploit this knowledge to determine the bias of the accelerometer measurements during the time span $t_\mathrm{A} < t < t_\mathrm{B}$

$$\mathbf{b}_\mathrm{acc} = \frac{1}{t_\mathrm{B} - t_\mathrm{A}} \int_{t_\mathrm{A}}^{t_\mathrm{B}} \mathbf{a}(t)\,\mathrm{d}t \tag{5}$$

- subtracting this bias from the measured acceleration prior to time-integration mitigates drift effects in the velocity
- alternatively, one can first integrate the measured acceleration and then correct the obtained velocity $\mathbf{v}$ by subtracting the drift component $\mathbf{v}_\mathrm{drift}$:

$$\mathbf{v}_\mathrm{drift} = \frac{\mathbf{v}(t_\mathrm{B}) - \mathbf{v}(t_\mathrm{A})}{t_\mathrm{B} - t_\mathrm{A}} (t - t_\mathrm{A}),\ \forall t \in [t_\mathrm{A}, t_\mathrm{B}], \quad \mathbf{v}_\mathrm{zvu} = \mathbf{v} - \mathbf{v}_\mathrm{drift}\ \forall t \in [t_\mathrm{A}, t_\mathrm{B}] \tag{6}$$

- this approach is commonly called *zero-velocity update*




Zero-velocity update for one of the components of $\mathbf{v}$ and the resulting change in position $x$ with and without correction in the time span $24\,\mathrm{s} < t < 26\,\mathrm{s}$ with known velocities $v_x(24\,\mathrm{s}) = v_x(26\,\mathrm{s}) = 0$.

- **Note**: Addition or integration of accelerations, (angular) velocities and positions is only possible in a common frame of reference!
- concept can also be applied to positions and orientations

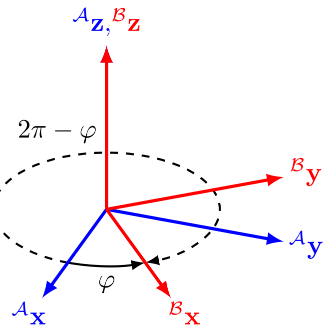
# 4 Mathematical Representations of 3D Orientations

- remember we want to measure the orientation of a rigid body
- when using strapdown IMUs, we also need the orientation to determine the position
- but how exactly can an orientation be put in numbers?!

## 4.1 Orientations and Rotations

- orientation is a state, rotation is the process/action that changes the state
- an orientation is the results of a (sequence of) rotation(s)
- an orientation is with respect to some reference coordinate system
- a rotation is with respect to whatever orientation it is applied to
- note the conceptual similarity with locations and directions in 3D space

- but note the difference in uniqueness of the concepts, which is caused by the fact that, unlike the $\mathbb{R}^3$, the space of orientation is cyclic:
  - for each direction there is one location that is obtained when starting in the origin
  - for each location there is one direction that connects it with the origin
  - for each rotation there is one orientation that is obtained when applying that rotation to the reference orientation
  - BUT for each orientation there are two rotations that both yield that orientation when applied to the reference orientation (example: $\frac{\pi}{2}$ around $x$ and $\frac{3\pi}{2}$ around $-x$)

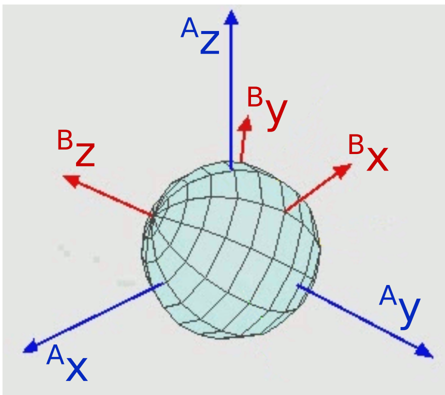



A rotation of $\varphi$ around $\mathbf{z}$ yields the same orientation as a rotation of $2\pi - \varphi$ around $-\mathbf{z}$

**Euler Rotation Theorem**

- original formulation: when a sphere is rotated around its centre it is always possible to find a diameter whose direction in the new orientation is the same as in the initial orientation
- meaning: apply arbitrary rotations to a rigid body while holding one point (called origin) fixed, the result is always equivalent to a rotation around a single fixed axis passing through the origin
- this implies that every orientation is (almost) uniquely determined by a rotation axis $\mathbf{j}$ and a rotation angle $\alpha \in [0, \pi]$
- why "almost" in the statement above? two "singular points" or ambiguities:
  - for the zero-rotation with zero angle, the axis is arbitrary and the result is trivial
  - 180°-rotation rotation around any axis is equivalent to 180° around the opposite axis
- why "$[0, \pi]$" in the statement above? if we consider $\alpha \in [0, 2\pi]$, then for every (non-zero) orientation, there are exactly two rotations that result in it: swap axis and use $2\pi - \alpha$.

**Example:** The red coordinate system B was obtained by a sequence of rotations starting with the blue initial (reference) coordinate system A. The resulting orientation between A and B can be obtained by a rotation around a (not easy to see) single axis.




## 4.2 Coordinates in Rotated Rigid Bodies

### 4.2.1 Coordinate System Definitions

- consider a rigid body and define an orthonormal coordinate system $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ onto it
- without loss of generality, choose a cuboid and put the CS in one of its corners
- choose a rotation axis $\mathbf{j}$ pointing through the origin of that coordinate system
- rotate (a virtual copy of) the rigid body by a rotation angle $\alpha \in [0, 2\pi)$
- note that now you have two coordinate systems: denote the CS before rotation by $\mathcal{A}$ and the CS after completion of the rotation by $\mathcal{B}$

- note that the coordinate system $\mathcal{B}$ is orthonormal, since rotations of rigid bodies leave angles and dimensions of the body unaffected
- note that the new orientation is uniquely determined by the three coordinate axes of $\mathcal{B}$ and vice versa

**Question:** How many corners of a (hidden) cuboid do you need to know before you know its (position and) orientation? How many axes of a (hidden) right-handed coordinate system do you need to know before you can tell where all three coordinate axes are?

### 4.2.2 Coordinate Definitions

- choose a point ${}^{\mathcal{A}}\mathbf{p}$ of the body $\mathcal{A}$, determine its coordinates in $\mathcal{A}$ and denote them by ${}^{\mathcal{A}}_{\mathcal{A}}\mathbf{p}$
- note that the corresponding point ${}^{\mathcal{B}}\mathbf{p}$ of the rotated body has the same coordinates in $\mathcal{B}$:

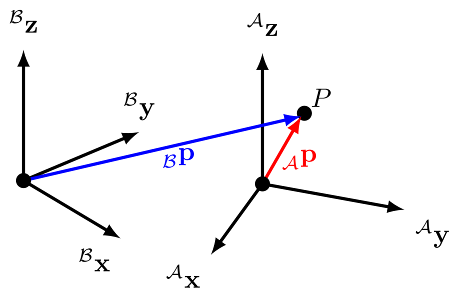
$$ {}^{\mathcal{A}}_{\mathcal{A}}\mathbf{p} = {}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p} $$

- however, for almost all points we find that

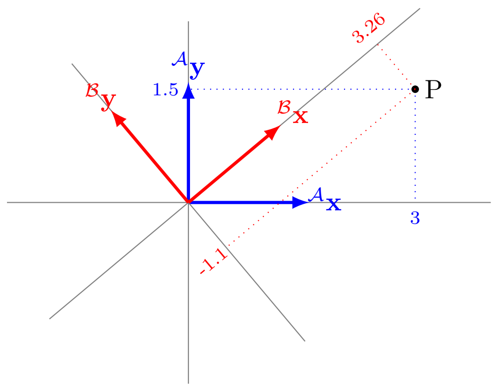
$$ {}^{\mathcal{A}}_{\mathcal{A}}\mathbf{p} \neq {}^{\mathcal{A}}_{\mathcal{B}}\mathbf{p} \quad \text{and} \quad {}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p} \neq {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{p} \quad \text{and} \quad {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{p} \neq {}^{\mathcal{A}}_{\mathcal{B}}\mathbf{p} $$

- here, ${}^{\mathcal{A}}_{\mathcal{B}}\mathbf{p}$ reads "p-$\mathcal{A}$ in $\mathcal{B}$-coordinates"

**Question:** For which points ${}^{\mathcal{A}}\mathbf{p}$ does ${}^{\mathcal{A}}_{\mathcal{B}}\mathbf{p} = {}^{\mathcal{A}}_{\mathcal{A}}\mathbf{p} = {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{p} = {}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}$ hold? Hint: This surely holds for the origin.



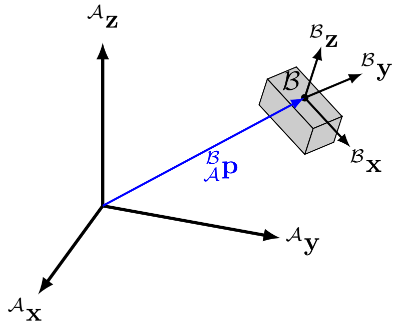

Visual representation of two coordinate systems $\mathcal{A}$ and $\mathcal{B}$ and the vectors ${}_{\mathcal{A}}\mathbf{p}$ and ${}_{\mathcal{B}}\mathbf{p}$ denoting the position of point $P$ in the coordinate systems




${}_{\mathcal{A}}\mathbf{p} = [3, 1.5, 0]^\intercal$ ${}_{\mathcal{B}}\mathbf{p} = [3.26, -1.1, 0]^\intercal$

## 4.3 Rotation Matrix Representation

- obviously, the coordinate axes $\{{}^{\mathcal{A}}_{\mathcal{A}}\mathbf{x}, {}^{\mathcal{A}}_{\mathcal{A}}\mathbf{y}, {}^{\mathcal{A}}_{\mathcal{A}}\mathbf{z}\}$ are $\{[1,0,0]^\intercal, [0,1,0]^\intercal, [0,0,1]^\intercal\}$




Visual representation of the coordinate system $\mathcal{A}$ as well as the rigid body and its fixed frame $\mathcal{B}$. The vector ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{p}$ describes the position of the origin of $\mathcal{B}$ in $\mathcal{A}$-coordinates

- likewise, the coordinate axes $\{{}^{\mathcal{B}}_{\mathcal{B}}\mathbf{x}, {}^{\mathcal{B}}_{\mathcal{B}}\mathbf{y}, {}^{\mathcal{B}}_{\mathcal{B}}\mathbf{z}\}$ are also $\{[1,0,0]^\intercal, [0,1,0]^\intercal, [0,0,1]^\intercal\}$
- the coordinate axes ${}^{\mathcal{B}}_{\mathcal{B}}\mathbf{x}, {}^{\mathcal{B}}_{\mathcal{B}}\mathbf{y}, {}^{\mathcal{B}}_{\mathcal{B}}\mathbf{z}$ can be expressed in the original $\mathcal{A}$-coordinates of the body
- denote these coordinates by ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{x}, {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{y}, {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{z}$, and the $\mathcal{A}$-axes in $\mathcal{B}$-CS by ${}^{\mathcal{A}}_{\mathcal{B}}\mathbf{x}, {}^{\mathcal{A}}_{\mathcal{B}}\mathbf{y}, {}^{\mathcal{A}}_{\mathcal{B}}\mathbf{z}$

**Problem:** Given ${}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}$, you want to determine ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{p}$. But all you have is ${}^{\mathcal{A}}_{\mathcal{B}}\mathbf{x}, {}^{\mathcal{A}}_{\mathcal{B}}\mathbf{y}, {}^{\mathcal{A}}_{\mathcal{B}}\mathbf{z}$.

- note the following relationships (in fact the very definition of what coordinates are):

$$ {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{p} = \left[{}^{\mathcal{A}}_{\mathcal{B}}\mathbf{x}, {}^{\mathcal{A}}_{\mathcal{B}}\mathbf{y}, {}^{\mathcal{A}}_{\mathcal{B}}\mathbf{z}\right]^\intercal {}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p} \qquad \text{(projection onto } \mathcal{A}\text{-coordinate axes)} $$

$$ {}^{\mathcal{A}}_{\mathcal{B}}\mathbf{p} = \left[{}^{\mathcal{B}}_{\mathcal{A}}\mathbf{x}, {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{y}, {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{z}\right]^\intercal {}^{\mathcal{A}}_{\mathcal{A}}\mathbf{p} \qquad \text{(projection onto } \mathcal{B}\text{-coordinate axes)} $$

  the scalar product on the right-hand side only makes sense if the lower-left indices agree

  the vector is transformed into the coordinate system whose axes are stacked in the matrix

- note that those two matrices must be inverses of each other ($\mathcal{A} \to \mathcal{B}$ and $\mathcal{B} \to \mathcal{A}$), and recall that the inverse of an orthonormal matrix is its transpose, which leads to

$$ {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{p} = \underbrace{\left[{}^{\mathcal{B}}_{\mathcal{A}}\mathbf{x}, {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{y}, {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{z}\right]}_{{}^{\mathcal{B}}_{\mathcal{A}}\mathbf{M}} {}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p} \qquad \text{and} \qquad {}^{\mathcal{A}}_{\mathcal{B}}\mathbf{p} = \underbrace{\left[{}^{\mathcal{A}}_{\mathcal{B}}\mathbf{x}, {}^{\mathcal{A}}_{\mathcal{B}}\mathbf{y}, {}^{\mathcal{A}}_{\mathcal{B}}\mathbf{z}\right]}_{{}^{\mathcal{A}}_{\mathcal{B}}\mathbf{M}} {}^{\mathcal{A}}_{\mathcal{A}}\mathbf{p} $$

  the lower-left index of the rotation matrix says which coordinates ($\mathcal{A}$ or $\mathcal{B}$) it yields

- note that ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{x} = {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{M}\, {}^{\mathcal{B}}_{\mathcal{B}}\mathbf{x}$ (selecting first column by multiplying $[1,0,0]^\intercal$)
- also note that trivially ${}^{\mathcal{A}}_{\mathcal{A}}\mathbf{M} = \mathbf{I}_{3\times 3}$ and ${}^{\mathcal{B}}_{\mathcal{B}}\mathbf{M} = \mathbf{I}_{3\times 3}$

### 4.3.1 Definition of Rotation Matrix

The rotation matrix that transforms an arbitrary vector ${}_{\mathcal{A}}\mathbf{p}$ from the right-handed coordinate system $\mathcal{A}$ to the right-handed coordinate system $\mathcal{B}$ is the unique matrix ${}^{\mathcal{A}}_{\mathcal{B}}\mathbf{M}$ that satisfies

$$
{}_{\mathcal{B}}\mathbf{p} = {}^{\mathcal{A}}_{\mathcal{B}}\mathbf{M}\,{}_{\mathcal{A}}\mathbf{p} \quad \forall\,{}_{\mathcal{A}}\mathbf{p} \in \mathbb{R}^3, \tag{7}
$$

and its columns are the $\mathcal{B}$-coordinates of the unit-length axis vectors of coordinate system $\mathcal{A}$, i.e.

$$
{}^{\mathcal{A}}_{\mathcal{B}}\mathbf{M} = \left[{}^{\mathcal{A}}_{\mathcal{B}}\mathbf{x}, {}^{\mathcal{A}}_{\mathcal{B}}\mathbf{y}, {}^{\mathcal{A}}_{\mathcal{B}}\mathbf{z}\right]. \tag{8}
$$

- note that it does not matter to which body or coordinate system the vector $\mathbf{p}$ is attached
- note furthermore that we can use a rotation matrix to describe an orientation in a bidirectionally unique way, i.e. every rotation matrix corresponds to exactly one orientation and vice versa
- consequently, every rotation matrix corresponds to two different rotations (both of which lead to the same orientation)
- the inverse of a matrix with orthonormal rows is its transpose, therefore

$$
{}_{\mathcal{A}}\mathbf{p} = {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{M}\,{}_{\mathcal{B}}\mathbf{p} = ({}^{\mathcal{A}}_{\mathcal{B}}\mathbf{M})^{-1}\,{}_{\mathcal{B}}\mathbf{p} = ({}^{\mathcal{A}}_{\mathcal{B}}\mathbf{M})^{\intercal}\,{}_{\mathcal{B}}\mathbf{p} \quad \forall\,{}_{\mathcal{A}}\mathbf{p} \in \mathbb{R}^3
$$

- note that the determinant is the triple product $({}^{\mathcal{B}}_{\mathcal{A}}\mathbf{x} \times {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{y}) \cdot {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{z}$ of the rows, which in this case is 1 by definition (would be $-1$ for left-handed coordinate systems)

**Question:** If ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{M} \neq \mathbf{I}_{3\times 3}$, which ${}_{\mathcal{B}}\mathbf{p}$ fulfill ${}_{\mathcal{A}}\mathbf{p} = {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{M}\,{}_{\mathcal{B}}\mathbf{p} = {}_{\mathcal{B}}\mathbf{p}$ ?

**Question:** Note the following two facts: Multiplication by a rotation matrix does not change the length of a vector. The determinant of a proper rotation matrix is one. What can you conclude about the eigenvalues and eigenvectors of rotation matrices?

**Answer:** One eigenvector is the rotation axis, it belongs to the eigenvalue one. The other two eigenvectors are complex and belong to the eigenvalues $e^{\pm i\alpha}$, where $\alpha$ is the rotation angle. Obviously, if $\alpha$ is zero, all eigenvalues are one and any vector is an eigenvector. Similarly, if $\alpha$ is $180°$, two eigenvalues are $-1$ and any vector in the plane perpendicular to the rotation axis is an eigenvector. See examples for basic rotations below.

## 4.4 Concatenation of Rotations

- consider a sequence of rotations $\mathcal{A} \to \mathcal{B} \to \mathcal{C}$
- any vector given in $\mathcal{A}$-coordinates can be transformed into $\mathcal{C}$-coordinates by ${}^{\mathcal{A}}_{\mathcal{C}}\mathbf{M} = {}^{\mathcal{B}}_{\mathcal{C}}\mathbf{M}\,{}^{\mathcal{A}}_{\mathcal{B}}\mathbf{M}$

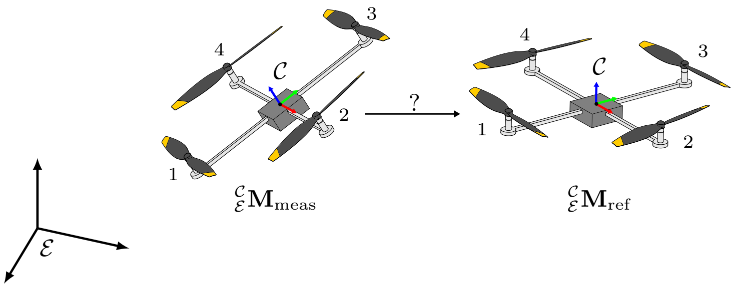
$$
{}_{\mathcal{C}}\mathbf{p} = {}^{\mathcal{B}}_{\mathcal{C}}\mathbf{M}\,{}_{\mathcal{B}}\mathbf{p} = {}^{\mathcal{B}}_{\mathcal{C}}\mathbf{M}\,{}^{\mathcal{A}}_{\mathcal{B}}\mathbf{M}\,{}_{\mathcal{A}}\mathbf{p} \quad \forall\,{}_{\mathcal{A}}\mathbf{p} \in \mathbb{R}^3
$$

- note that multiplication of rotation matrices is associative but not commutative, i.e.

$$
\left({}^{\mathcal{C}}_{\mathcal{D}}\mathbf{M}\,{}^{\mathcal{B}}_{\mathcal{C}}\mathbf{M}\right){}^{\mathcal{A}}_{\mathcal{B}}\mathbf{M} = {}^{\mathcal{C}}_{\mathcal{D}}\mathbf{M}\left({}^{\mathcal{B}}_{\mathcal{C}}\mathbf{M}\,{}^{\mathcal{A}}_{\mathcal{B}}\mathbf{M}\right), \quad \text{but in general} \quad {}^{\mathcal{B}}_{\mathcal{C}}\mathbf{M}\,{}^{\mathcal{A}}_{\mathcal{B}}\mathbf{M} \neq {}^{\mathcal{A}}_{\mathcal{B}}\mathbf{M}\,{}^{\mathcal{B}}_{\mathcal{C}}\mathbf{M}
$$

- this is in agreement with the observation that applying two rotations in a certain order does, in general, not yield the same result as applying them in the inverse order, as you can see yourself for the following basic rotations

**Question:** For a quadrocopter $\mathcal{C}$ the currently measured orientation ${}^{\mathcal{C}}_{\mathcal{E}}\mathbf{M}_{\text{meas}}$ with respect to a reference frame $\mathcal{E}$ is ${}^{\mathcal{C}}_{\mathcal{E}}\mathbf{M}_{\text{meas}} = \begin{bmatrix} 0.69 & 0.41 & 0.60 \\ -0.59 & -0.15 & 0.79 \\ 0.42 & -0.90 & 0.14 \end{bmatrix}$. The desired orientation ${}^{\mathcal{C}}_{\mathcal{E}}\mathbf{M}_{\text{ref}}$ in the same reference frame is given as ${}^{\mathcal{C}}_{\mathcal{E}}\mathbf{M}_{\text{ref}} = \begin{bmatrix} 0.69 & -0.13 & 0.72 \\ -0.59 & -0.67 & 0.45 \\ 0.42 & -0.73 & -0.54 \end{bmatrix}$. What control action (motor 1-4) does the quadrocopter need to perform to rotate towards the desired orientation?




**Answer:** The difference between ${}^{\mathcal{C}}_{\mathcal{E}}\mathbf{M}_{\text{meas}}$ and ${}^{\mathcal{C}}_{\mathcal{E}}\mathbf{M}_{\text{ref}}$ is the relative orientation

$$\mathbf{M}_{\text{rel}} = {}^{\mathcal{C}}_{\mathcal{E}}\mathbf{M}^{-1}_{\text{meas}}\, {}^{\mathcal{C}}_{\mathcal{E}}\mathbf{M}_{\text{ref}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \cong -\frac{\pi}{4} \text{ around } \mathbf{x}$$

The quadrocopter needs to increase the thrust on motor 1 and decrease the thrust on motor 3.

## 4.5 Rotation Matrices of Basic Rotations

There are several basic rotations that can be described by signed permutation matrices, which are a special subset of rotation matrices.

### 4.5.1 Half Turns

- a 180° rotation around $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ leads to ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{x} = -\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ and ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{y} = -\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, thus ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{M} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
- a 180° rotation around $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ leads to ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{z} = -\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ and ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{y} = -\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, thus ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$
- a 180° rotation around $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ leads to ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{x} = -\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ and ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{z} = -\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, thus ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{M} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$
- Note that each of these matrices has a unique eigenvector for the eigenvalue 1. It is the rotation axis.

### 4.5.2 Quarter Turns

- a 90° rotation around $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ leads to ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{z} = -\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ and ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{y} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, thus ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$

- a 90° rotation around $\begin{bmatrix}0\\1\\0\end{bmatrix}$ leads to ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{x} = -\begin{bmatrix}0\\0\\1\end{bmatrix}$ and ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{z} = \begin{bmatrix}1\\0\\0\end{bmatrix}$, thus ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{M} = \begin{bmatrix}0 & 0 & 1\\0 & 1 & 0\\-1 & 0 & 0\end{bmatrix}$
- a 90° rotation around $\begin{bmatrix}0\\0\\1\end{bmatrix}$ leads to ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{x} = \begin{bmatrix}0\\1\\0\end{bmatrix}$ and ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{y} = -\begin{bmatrix}1\\0\\0\end{bmatrix}$, thus ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{M} = \begin{bmatrix}0 & -1 & 0\\1 & 0 & 0\\0 & 0 & 1\end{bmatrix}$

### 4.5.3 Diagonal Turns

- 120° around $\begin{bmatrix}1\\1\\1\end{bmatrix}$ yield ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{x} = \begin{bmatrix}0\\1\\0\end{bmatrix}$ and ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{y} = \begin{bmatrix}0\\0\\1\end{bmatrix}$ and ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{z} = \begin{bmatrix}1\\0\\0\end{bmatrix}$, thus ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{M} = \begin{bmatrix}0 & 0 & 1\\1 & 0 & 0\\0 & 1 & 0\end{bmatrix}$
- 120° around $\begin{bmatrix}-1\\-1\\-1\end{bmatrix}$ yield ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{x} = \begin{bmatrix}0\\0\\1\end{bmatrix}$ and ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{y} = \begin{bmatrix}1\\0\\0\end{bmatrix}$ and ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{z} = \begin{bmatrix}0\\1\\0\end{bmatrix}$, thus ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{M} = \begin{bmatrix}0 & 1 & 0\\0 & 0 & 1\\1 & 0 & 0\end{bmatrix}$

### 4.5.4 Arbitrary Turns Around Coordinate Axes

- a rotation of $\alpha$ around $\begin{bmatrix}1\\0\\0\end{bmatrix}$ yield ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{M} = \begin{bmatrix}1 & 0 & 0\\0 & \cos\alpha & -\sin\alpha\\0 & \sin\alpha & \cos\alpha\end{bmatrix}$
- a rotation of $\alpha$ around $\begin{bmatrix}0\\1\\0\end{bmatrix}$ yield ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{M} = \begin{bmatrix}\cos\alpha & 0 & \sin\alpha\\0 & 1 & 0\\-\sin\alpha & 0 & \cos\alpha\end{bmatrix}$
- a rotation of $\alpha$ around $\begin{bmatrix}0\\0\\1\end{bmatrix}$ yield ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{M} = \begin{bmatrix}\cos\alpha & -\sin\alpha & 0\\\sin\alpha & \cos\alpha & 0\\0 & 0 & 1\end{bmatrix}$

## 4.6 Rotation Matrices of Arbitrary Rotations

### 4.6.1 Rotation Matrix from Axis and Angle

- consider a rotation of $\alpha$ around an arbitrary axis ${}_{\mathcal{A}}\mathbf{j} = {}_{\mathcal{B}}\mathbf{j}$
- define an auxiliary coordinate system with first axis ${}^{\hat{\mathcal{A}}}_{\mathcal{A}}\mathbf{x} := {}_{\mathcal{A}}\mathbf{j}$
- construct the second axis to be ${}^{\hat{\mathcal{A}}}_{\mathcal{A}}\mathbf{y} := \dfrac{{}_{\mathcal{A}}\mathbf{j} \times {}^{\mathcal{A}}_{\mathcal{A}}\mathbf{x}}{\left|\left|{}_{\mathcal{A}}\mathbf{j} \times {}^{\mathcal{A}}_{\mathcal{A}}\mathbf{x}\right|\right|_2}$
- construct the third axis to be ${}^{\hat{\mathcal{A}}}_{\mathcal{A}}\mathbf{z} := \dfrac{{}_{\mathcal{A}}\mathbf{j} \times {}^{\hat{\mathcal{A}}}_{\mathcal{A}}\mathbf{y}}{\left|\left|{}_{\mathcal{A}}\mathbf{j} \times {}^{\hat{\mathcal{A}}}_{\mathcal{A}}\mathbf{y}\right|\right|_2}$
- thus, the auxiliary coordinate transformation can be represented by ${}^{\hat{\mathcal{A}}}_{\mathcal{A}}\mathbf{M} = \left[{}^{\hat{\mathcal{A}}}_{\mathcal{A}}\mathbf{x}, {}^{\hat{\mathcal{A}}}_{\mathcal{A}}\mathbf{y}, {}^{\hat{\mathcal{A}}}_{\mathcal{A}}\mathbf{z}\right]$
- note that the auxiliary coordinate system is rotated with the rigid body
- however, in the auxiliary coordinates we find ${}_{\hat{\mathcal{A}}}\mathbf{j} = {}_{\hat{\mathcal{B}}}\mathbf{j} = [1, 0, 0]^\intercal$
- furthermore, ${}^{\hat{\mathcal{A}}}_{\hat{\mathcal{A}}}\mathbf{y}$ and ${}^{\hat{\mathcal{A}}}_{\hat{\mathcal{A}}}\mathbf{z}$ are rotated as in a rotation around the x-axis

$$
{}^{\hat{\mathcal{B}}}_{\hat{\mathcal{A}}}\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix}
$$

- by inverting the auxiliary coordinate transformation in the $\mathcal{B}$-frame, we obtain ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{M}$:

$$
{}^{\mathcal{B}}_{\mathcal{A}}\mathbf{M} \;=\; {}^{\hat{\mathcal{A}}}_{\mathcal{A}}\mathbf{M} \; {}^{\hat{\mathcal{B}}}_{\hat{\mathcal{A}}}\mathbf{M} \; {}^{\mathcal{B}}_{\hat{\mathcal{B}}}\mathbf{M} \;=\; {}^{\hat{\mathcal{A}}}_{\mathcal{A}}\mathbf{M} \; {}^{\hat{\mathcal{B}}}_{\hat{\mathcal{A}}}\mathbf{M} \left({}^{\hat{\mathcal{A}}}_{\mathcal{A}}\mathbf{M}\right)^{-1}
$$

- note that only the rotation matrix in the middle refers to an actual physical rotation of the rigid body
- doing the math results in the following formula, wherein ${}_{\mathcal{A}}\mathbf{j} = {}_{\mathcal{B}}\mathbf{j} =: [j_x, j_y, j_z]^\intercal$

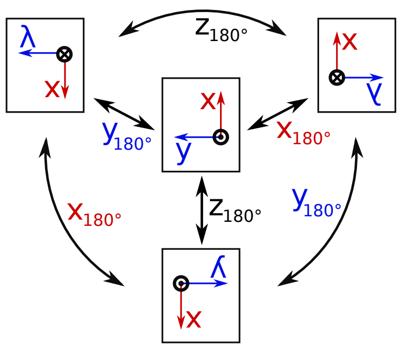
$$
{}^{\mathcal{B}}_{\mathcal{A}}\mathbf{M} = \begin{bmatrix} \cos\alpha + j_x^2(1-\cos\alpha) & j_x j_y(1-\cos\alpha) - j_z\sin\alpha & j_x j_z(1-\cos\alpha) + j_y\sin\alpha \\ j_x j_y(1-\cos\alpha) + j_z\sin\alpha & \cos\alpha + j_y^2(1-\cos\alpha) & j_y j_z(1-\cos\alpha) - j_x\sin\alpha \\ j_x j_z(1-\cos\alpha) - j_y\sin\alpha & j_y j_z(1-\cos\alpha) + j_x\sin\alpha & \cos\alpha + j_z^2(1-\cos\alpha) \end{bmatrix} \tag{9}
$$

### 4.6.2 Axis and Angle from Rotation Matrix

- obviously, the trace of ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{M}$ is $1 + 2\cos\alpha$, which yields $\alpha$ when ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{M}$ is given
- subsequently, $j_x$ can be determined from the first-row, first-column element (unless $\alpha = 0$) and $j_y$ and $j_z$ can be determined equivalently
- however, for small angles, this yields bad sensitivity (small element variations imply large axis variations)
- alternatively, note that ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{M}(3,2) - {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{M}(2,3) = 2j_x\sin\alpha$, which is less sensitive for $\alpha \ll 1$

## 4.7 Special Orthogonal Group SO(3)

- a group is a set of elements $\mathbb{S} := \{A, B, \ldots\}$ with an operation $\otimes$ that fulfills
  - closure $\quad \forall A, B: \ (A \otimes B) \in \mathbb{S}$
  - associativity $\quad (A \otimes B) \otimes C = A \otimes (B \otimes C)$
  - identity ($\exists I \in \mathbb{S}$ such that $A \otimes I = A \ \forall A \in \mathbb{S}$)
  - invertibility (for each $A \in \mathbb{S}$ there exists an $\hat{A} \in \mathbb{S}$ such that$A \otimes \hat{A} = I$)
- the set of $3 \times 3$ orthogonal ("orthonormal") matrices (pairwise orthogonal unit-length columns and rows) with the operation multiplication is the orthogonal group O(3)
- the subgroup of rotation matrices (determinant +1) is the special orthogonal group SO(3)
- SO(3) describes the set of 3D orientations (with the operation concatenation)
- it is nonabelian, since the concatenation is not commutative
- it is also a subgroup of the general linear group consisting of all invertible linear transformations of $\mathbb{R}^3$
- it has a few very remarkable properties
- applying the same half turn twice yields the original orientation, i.e. each of the three matrices (see above) is its own inverse
- applying the same diagonal turn three times yields the original orientation ($3 \cdot 120° = 360°$)
- applying a half turn around $x$, then around $y$, and then around $z$ also yields identity
- these half turns form the Klein four-group, which is a subgroup of SO(3):




Klein four-group

| | 1 | x | y | z |
|---|---|---|---|---|
| 1 | 1 | x | y | z |
| x | x | 1 | z | y |
| y | y | z | 1 | x |
| z | z | y | x | 1 |

multiplication table

**Question:** Are the following sets of actions a group? If yes, which properties do the groups have: Swapping tires of a car to distribute wear equally. Manipulating a Rubik's cube.

**Question:** What is the mathematical connection between the Klein four-group and a finite simple group of order two and which rotations can be described by the latter?

**Answer:** The direct product of two finite simple groups of order two yields a result which is isomorphic to the Klein four-group:

| · | **1** | **a** |
|---|---|---|
| **1** | 1 | a |
| **a** | a | 1 |

×

| · | **1** | **b** |
|---|---|---|
| **1** | 1 | b |
| **b** | b | 1 |

=

| · | **(1,1)** | **(a,1)** | **(1,b)** | **(a,b)** |
|---|---|---|---|---|
| **(1,1)** | (1,1) | (a,1) | (1,b) | (a,b) |
| **(a,1)** | (a,1) | (1,1) | (a,b) | (1,b) |
| **(1,b)** | (1,b) | (a,b) | (1,1) | (a,1) |
| **(a,b)** | (a,b) | (1,b) | (a,1) | (1,1) |

## 4.8 Unit Quaternion Representation

There must be a more elegant way to capture the group structure of 3D rotations than using the set of rotation matrices. Let us ask the man to whom we owe the Cayley-Hamilton theorem (Matrix fulfills its own characteristic equation) and the Hamilton-Jacobi-Bellman equation (PDE that is central in optimal control theory).

### 4.8.1 Quaternions:

- three-dimensional vectors well describe positions and translation, but not orientations and rotations
- Monday 16 October 1843 in Dublin, Sir William Rowan Hamilton on way to Royal Irish Academy, carved $\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{i}\,\mathbf{j}\,\mathbf{k} = -1$ into a stone on Brougham (Broom) Bridge
- this definitions implies $\mathbf{i}\,\mathbf{j} = \mathbf{k}$, $\mathbf{j}\,\mathbf{k} = \mathbf{i}$, $\mathbf{k}\,\mathbf{i} = \mathbf{j}$, BUT $\mathbf{j}\,\mathbf{i} = -\mathbf{k}$, $\mathbf{k}\,\mathbf{j} = -\mathbf{i}$, $\mathbf{i}\,\mathbf{k} = -\mathbf{j}$

| 1st \ 2nd | **1** | **i** | **j** | **k** |
|---|---|---|---|---|
| **1** | 1 | **i** | **j** | **k** |
| **i** | **i** | -1 | **k** | **-j** |
| **j** | **j** | **-k** | -1 | **i** |
| **k** | **k** | **j** | **-i** | -1 |

- same concept also discovered by Carl Friedrich Gauss and Olinde Rodrigues, but Hamilton dedicated most of his life to studying and promoting them
- define a quaternion as a complex number with three imaginary parts

$$q = a + b\,\mathbf{i} + c\,\mathbf{j} + d\,\mathbf{k}, \quad a, b, c, d \in \mathbb{R}$$

- note the multiplication rule that follows from the above definitions

$$\begin{aligned}
q_1 \otimes q_2 =& (a_1 + b_1\,\mathbf{i} + c_1\,\mathbf{j} + d_1\,\mathbf{k}) \otimes (a_2 + b_2\,\mathbf{i} + c_2\,\mathbf{j} + d_2\,\mathbf{k}) \\
=& \quad a_1a_2 + a_1b_2\,\mathbf{i} + a_1c_2\,\mathbf{j} + a_1d_2\,\mathbf{k} \\
&+ b_1a_2\,\mathbf{i} + b_1b_2\underbrace{\mathbf{i}\,\mathbf{i}}_{-1} + b_1c_2\underbrace{\mathbf{i}\,\mathbf{j}}_{\mathbf{k}} + b_1d_2\underbrace{\mathbf{i}\,\mathbf{k}}_{-\mathbf{j}} \\
&+ c_1a_2\,\mathbf{j} + c_1b_2\underbrace{\mathbf{j}\,\mathbf{i}}_{-\mathbf{k}} + c_1c_2\underbrace{\mathbf{j}\,\mathbf{j}}_{-1} + c_1d_2\underbrace{\mathbf{j}\,\mathbf{k}}_{\mathbf{i}} \\
&+ d_1a_2\,\mathbf{k} + d_1b_2\underbrace{\mathbf{k}\,\mathbf{i}}_{\mathbf{j}} + d_1c_2\underbrace{\mathbf{k}\,\mathbf{j}}_{-\mathbf{i}} + d_1d_2\underbrace{\mathbf{k}\,\mathbf{k}}_{-1} \\
=& \quad (a_1a_2 - b_1b_2 - c_1c_2 - d_1d_2) + (a_1b_2 + b_1a_2 \underline{+ c_1d_2 - d_1c_2})\,\mathbf{i} \\
&+ (a_1c_2 + c_1a_2 \underline{- b_1d_2 + d_1b_2})\,\mathbf{j} + (a_1d_2 + d_1a_2 \underline{+ b_1c_2 - c_1b_2})\,\mathbf{k}
\end{aligned}$$

- note that quaternion multiplication is not commutative, as seen in the underlined terms
- define the conjugate $q^* = a - b\,\mathbf{i} - c\,\mathbf{j} - d\,\mathbf{k}$ and note that $q \otimes q^* = a^2 + b^2 + c^2 + d^2 \in \mathbb{R}$
- define the norm $\|q\| = \sqrt{q \otimes q^*} = \sqrt{a^2 + b^2 + c^2 + d^2}$
- note that this norm is special, since it is multiplicative, i.e. $\|q_1 \otimes q_2\| = \|q_1\|\,\|q_2\|$

### 4.8.2 Unit Quaternions and Rotations:

- since $\|q_1 \otimes q_2\| = \|q_1\|\,\|q_2\|$, the set of unit quaternions is closed, and even a group itself
- consider only unit quaternions and associate each with a rotation axis and angle

$$\alpha \in [0,\pi] \;@\; \underbrace{\mathbf{j} = [j_x, j_y, j_z]^\intercal}_{\|j\|_2=1} \quad \cong \quad {}^{\mathcal{B}}_{\mathcal{A}}q = \cos(\frac{\alpha}{2}) + (j_x\,\mathbf{i} + j_y\,\mathbf{j} + j_z\,\mathbf{k})\sin(\frac{\alpha}{2})$$

- BUT note the most striking property that follows from this definition:

$${}^{\mathcal{B}}_{\mathcal{A}}q \otimes {}^{\mathcal{C}}_{\mathcal{B}}q = {}^{\mathcal{C}}_{\mathcal{A}}q \quad \forall\, {}^{\mathcal{B}}_{\mathcal{A}}q, {}^{\mathcal{C}}_{\mathcal{B}}q,$$

  i.e. it precisely captures the group properties of **arbitrary** rotations in 3D space!

- try it yourself: ${}^{\mathcal{B}}_{\mathcal{A}}q \;\cong\; \frac{\pi}{2} @ \begin{bmatrix}1\\0\\0\end{bmatrix}$ and ${}^{\mathcal{C}}_{\mathcal{B}}q \;\cong\; \frac{\pi}{2} @ \begin{bmatrix}1\\0\\0\end{bmatrix}$

$$\left(\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}\,\mathbf{i}\right) \otimes \left(\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}\,\mathbf{i}\right) = \mathbf{i} \quad \cong \quad \pi @ \begin{bmatrix}1\\0\\0\end{bmatrix} \quad \checkmark$$

- try it yourself: ${}^{\mathcal{B}}_{\mathcal{A}}q \;\cong\; \frac{\pi}{2} @ \begin{bmatrix}1\\0\\0\end{bmatrix}$ and ${}^{\mathcal{C}}_{\mathcal{B}}q \;\cong\; \frac{\pi}{2} @ \begin{bmatrix}0\\1\\0\end{bmatrix}$ (use $\cos(\frac{\pi}{3}) = \frac{1}{2}$, $\sin(\frac{\pi}{3}) = \frac{\sqrt{3}}{2}$)

$$\left(\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}\,\mathbf{i}\right) \otimes \left(\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}\,\mathbf{j}\right) = \left(\frac{1}{2} + \frac{1}{2}(\mathbf{i} + \mathbf{j} + \mathbf{k})\right) \quad \cong \quad \frac{2\pi}{3} @ \begin{bmatrix}\frac{1}{\sqrt{3}}\\ \frac{1}{\sqrt{3}}\\ \frac{1}{\sqrt{3}}\end{bmatrix} \quad \checkmark$$

### 4.8.3 Negative Real Parts

- the above definition yields unit quaternions with nonnegative real part ($\cos([0,\frac{\pi}{2}]) \geq 0$)
- BUT what if ${}^{\mathcal{B}}_{\mathcal{A}}q \;\cong\; \frac{3\pi}{4} @ \begin{bmatrix}1\\0\\0\end{bmatrix}$ and ${}^{\mathcal{C}}_{\mathcal{B}}q \;\cong\; \frac{3\pi}{4} @ \begin{bmatrix}1\\0\\0\end{bmatrix}$ (twice 135°around x)
- the result should be $\frac{3\pi}{2}$@x, which is $-\frac{\pi}{2}$@x in the above definition, i.e. $\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}\,\mathbf{i}$

$$\left(\frac{\sqrt{2-\sqrt{2}}}{2} + \frac{\sqrt{2+\sqrt{2}}}{2}\,\mathbf{i}\right) \otimes \left(\frac{\sqrt{2-\sqrt{2}}}{2} + \frac{\sqrt{2+\sqrt{2}}}{2}\,\mathbf{i}\right) = \left(-\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}\,\mathbf{i}\right) \quad \cong \quad -\frac{\pi}{2} @ \begin{bmatrix}1\\0\\0\end{bmatrix}$$

→ the unit quaternions with positive real part are not closed under $\otimes$ (i.e. useless!)

→ need all unit quaternions and generalize our definition from above to **arbitrary angles**:

$$\begin{array}{ccc} \alpha \;@\; \mathbf{j} = [j_x, j_y, j_z]^\intercal & \cong & {}^{\mathcal{B}}_{\mathcal{A}}q = \cos(\frac{\alpha}{2}) + (j_x\,\mathbf{i} + j_y\,\mathbf{j} + j_z\,\mathbf{k})\sin(\frac{\alpha}{2}) \\ \wr\Vert & & \wr\Vert \\ (2\pi - \alpha) \in [0, 2\pi) \;@\; (-\mathbf{j}) & \cong & \underbrace{\cos(\tfrac{2\pi-\alpha}{2})}_{-\cos(\frac{\alpha}{2})} - (j_x\,\mathbf{i} + j_y\,\mathbf{j} + j_z\,\mathbf{k})\underbrace{\sin(\tfrac{2\pi-\alpha}{2})}_{\sin(\frac{\alpha}{2})} = -\,{}^{\mathcal{B}}_{\mathcal{A}}q \end{array} \tag{10}$$

→ the space of unit quaternions is a double cover of SO(3)

→ each rotation is **uniquely** described by one unit quaternion

→ each orientation can be described by **two** unit quaternions with opposite signs ($q$ and $-q$)

| **Rotation** | **Quaternion** | **Orientation = RotMatrix** |
|---|---|---|
| $\frac{\pi}{2}$ @ $[1,0,0]^\intercal$ | $\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}\,\mathbf{i}$ | $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$ |
| $\frac{3\pi}{2}$ @ $[-1,0,0]^\intercal$ | $-\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}\,\mathbf{i}$ | |
| $-\frac{\pi}{2}$ @ $[-1,0,0]^\intercal$ | $\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}\mathbf{i}$ | |

**Question:** Which quaternion describes a rotation of $-\frac{\pi}{2}$ around the x-axis?

Answer: $-\frac{\pi}{2}$ @ $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \cong \cos(-\frac{\pi}{4}) + \sin(-\frac{\pi}{4})\,\mathbf{i} = \cos(\frac{\pi}{4}) - \sin(\frac{\pi}{4})\,\mathbf{i} \cong \frac{\pi}{2}$ @ $\begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}$

### 4.8.4 Vector Notation

- for convenience, write the quaternion in vector representation:

$$\alpha \in [0, 2\pi) \text{ @ } \mathbf{j} = [j_x, j_y, j_z]^\intercal \quad \cong \quad {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{q} = \begin{bmatrix} \cos(\frac{\alpha}{2}) \\ j_x \sin(\frac{\alpha}{2}) \\ j_y \sin(\frac{\alpha}{2}) \\ j_z \sin(\frac{\alpha}{2}) \end{bmatrix} = \begin{bmatrix} \cos(\frac{\alpha}{2}) \\ \sin(\frac{\alpha}{2}) \begin{bmatrix} j_x \\ j_y \\ j_z \end{bmatrix} \end{bmatrix}$$

- note that sometimes (rarely) a definition with the real part is the last entry is used
- note the following quaternions of simple rotations:
- Half Turns:

$$\begin{aligned} 180° \text{ @ } [1,0,0]^\intercal &\cong [0,1,0,0]^\intercal \cong [0,-1,0,0]^\intercal \\ 180° \text{ @ } [0,1,0]^\intercal &\cong [0,0,1,0]^\intercal \cong [0,0,-1,0]^\intercal \\ 180° \text{ @ } [0,0,1]^\intercal &\cong [0,0,0,1]^\intercal \cong [0,0,0,-1]^\intercal \end{aligned}$$

- Quarter Turns:

$$\begin{aligned} 90° \text{ @ } [1,0,0]^\intercal &\cong \left[\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}, 0, 0\right]^\intercal \cong \left[-\tfrac{1}{\sqrt{2}}, -\tfrac{1}{\sqrt{2}}, 0, 0\right]^\intercal \\ 90° \text{ @ } [0,1,0]^\intercal &\cong \left[\tfrac{1}{\sqrt{2}}, 0, \tfrac{1}{\sqrt{2}}, 0\right]^\intercal \cong \left[-\tfrac{1}{\sqrt{2}}, 0, -\tfrac{1}{\sqrt{2}}, 0\right]^\intercal \\ 90° \text{ @ } [0,0,1]^\intercal &\cong \left[\tfrac{1}{\sqrt{2}}, 0, 0, \tfrac{1}{\sqrt{2}}\right]^\intercal \cong \left[\tfrac{1}{-\sqrt{2}}, 0, 0, -\tfrac{1}{\sqrt{2}}\right]^\intercal \end{aligned}$$

- Diagonal Turns:

$$
\begin{aligned}
120° \ @ \ \left[1,1,1\right]^\intercal \quad &\cong \quad \left[\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2}\right]^\intercal \quad &\cong \quad \left[-\tfrac{1}{2},-\tfrac{1}{2},-\tfrac{1}{2},-\tfrac{1}{2}\right]^\intercal \\
-120° \ @ \ \left[1,1,1\right]^\intercal \quad &\cong \quad \left[-\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2}\right]^\intercal \quad &\cong \quad \left[\tfrac{1}{2},-\tfrac{1}{2},-\tfrac{1}{2},-\tfrac{1}{2}\right]^\intercal
\end{aligned}
$$

- furthermore, note how simple the quaternion multiplication looks in vector style:

$$
\begin{bmatrix} a \\ \mathbf{v} \end{bmatrix} \otimes \begin{bmatrix} b \\ \mathbf{w} \end{bmatrix} = \begin{bmatrix} ab - \mathbf{v}^\intercal \mathbf{w} \\ a\mathbf{w} + b\mathbf{v} + \mathbf{v} \times \mathbf{w} \end{bmatrix}
$$

### 4.8.5 Transformations between Rotated Frames

- need method equivalent to multiplication by rot. matrix, but for axis-angle / quaternions
- note that when transforming vectors, the projection on the rotation axis $\mathbf{j}$ never changes (see e.g. the 1-entries in the rotation matrices for rotations about coordinate axes)

$$({}_{\mathcal{A}}\mathbf{p}^{\intercal}\mathbf{j})\mathbf{j} \quad = \quad ({}_{\mathcal{B}}\mathbf{p}^{\intercal}\mathbf{j})\mathbf{j} \qquad \text{(projections on axis)}$$

- hence, decompose any vector into its invariant component (parallel to $\mathbf{j}$) and the rest:

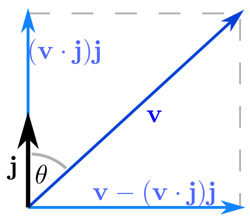
$${}_{\mathcal{A}}\mathbf{p} = \underbrace{({}_{\mathcal{A}}\mathbf{p}^{\intercal}\mathbf{j})\mathbf{j}}_{{}_{\mathcal{A}}\mathbf{p}_{\text{invar}}} + \underbrace{{}_{\mathcal{A}}\mathbf{p} - ({}_{\mathcal{A}}\mathbf{p}^{\intercal}\mathbf{j})\mathbf{j}}_{{}_{\mathcal{A}}\mathbf{p}_{\text{rest}}}\,, \qquad {}_{\mathcal{A}}\mathbf{p}_{\text{rest}} \perp \mathbf{j}$$

- note that ${}^{\mathcal{A}}\mathbf{p}_{\text{rest}}$ and ${}^{\mathcal{B}}\mathbf{p}_{\text{rest}}$ both lie in the plane perpendicular to $\mathbf{j}$;
  in that plane, you can transform ${}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}_{\text{rest}} = {}^{\mathcal{A}}_{\mathcal{A}}\mathbf{p}_{\text{rest}}$ to ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{p}_{\text{rest}}$ by

$${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{p}_{\text{rest}} \;=\; {}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}_{\text{rest}}\cos\alpha \;+\; \left(\mathbf{j}\times{}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}\right)\sin\alpha,$$

- this yields $\left(\mathbf{j}\times{}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}\right)$ for a quarter turn ($\alpha = \frac{\pi}{2}$), which makes sense, because the vector product is perpendicular to both factors and has exactly the length of the rest vector

Linear Algebra Basics:




$$\begin{aligned}
&\|\mathbf{j}\|_2 = 1\\
&\|(\mathbf{v}\cdot\mathbf{j})\mathbf{j}\|_2 = |\mathbf{v}\cdot\mathbf{j}| \quad = \|\mathbf{v}\|_2\,|\cos\theta|\\
&\|\mathbf{v}-(\mathbf{v}\cdot\mathbf{j})\mathbf{j}\|_2 = \|\mathbf{v}\times\mathbf{j}\|_2 \quad = \|\mathbf{v}\|_2\,|\sin\theta|\\
&\mathbf{v}-(\mathbf{v}\cdot\mathbf{j})\mathbf{j} \;=\; \mathbf{j}\times\mathbf{v}\times\mathbf{j} \quad \text{(triple product)}\\
&\|\mathbf{v}\|_2^2 - (\mathbf{v}\cdot\mathbf{j})^2 = \|\mathbf{v}\times\mathbf{j}\|_2^2 \quad \text{(Pythagoras)}
\end{aligned}$$

- conclude that transformation between ${}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p} = {}^{\mathcal{A}}_{\mathcal{A}}\mathbf{p}$ and ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{p}$ is obtained by

$$\begin{aligned}
{}^{\mathcal{B}}_{\mathcal{A}}\mathbf{p} \;&=\; ({}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}^{\intercal}\mathbf{j})\mathbf{j} \;+\; \left({}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p} - ({}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}^{\intercal}\mathbf{j})\mathbf{j}\right)\cos\alpha \;+\; \left(\mathbf{j}\times{}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}\right)\sin\alpha\\
&=\; \left(({}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}^{\intercal}\mathbf{j})\mathbf{j}\right)(1-\cos\alpha) \;+\; \left({}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}\right)\cos\alpha \;+\; \left(\mathbf{j}\times{}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}\right)\sin\alpha \quad \text{(use double angle formulas)}\\
&=\; \left(({}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}^{\intercal}\mathbf{j})\mathbf{j}\right)\left(2\sin^2\frac{\alpha}{2}\right) \;+\; \left({}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}\right)\left(\cos^2\frac{\alpha}{2}-\sin^2\frac{\alpha}{2}\right) \;+\; \left(\mathbf{j}\times{}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}\right)\left(2\sin\frac{\alpha}{2}\cos\frac{\alpha}{2}\right)
\end{aligned}$$

- note that (miraculously) this is the same as the imaginary part of

$$\begin{aligned}
&\begin{bmatrix}\cos\frac{\alpha}{2}\\ \mathbf{j}\sin\frac{\alpha}{2}\end{bmatrix}\otimes\begin{bmatrix}0\\ {}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}\end{bmatrix}\otimes\begin{bmatrix}\cos\frac{\alpha}{2}\\ -\mathbf{j}\sin\frac{\alpha}{2}\end{bmatrix} \quad = \begin{bmatrix}\cos\frac{\alpha}{2}\\ \mathbf{j}\sin\frac{\alpha}{2}\end{bmatrix}\otimes\begin{bmatrix}{}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}^{\intercal}\mathbf{j}\sin\frac{\alpha}{2}\\ {}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}\cos\frac{\alpha}{2}+\mathbf{j}\times{}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}\sin\frac{\alpha}{2}\end{bmatrix}\\
&= \begin{bmatrix}{}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}^{\intercal}\mathbf{j}\sin\frac{\alpha}{2}\cos\frac{\alpha}{2} - \left(\mathbf{j}\sin\frac{\alpha}{2}\right)^{\intercal}\left({}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}\cos\frac{\alpha}{2}+\mathbf{j}\times{}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}\sin\frac{\alpha}{2}\right)\\ \cos\frac{\alpha}{2}\left({}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}\cos\frac{\alpha}{2}+\mathbf{j}\times{}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}\sin\frac{\alpha}{2}\right) + \left(({}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}^{\intercal}\mathbf{j})\mathbf{j}\right)\sin^2\frac{\alpha}{2} + \left(\mathbf{j}\sin\frac{\alpha}{2}\right)\times\left({}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}\cos\frac{\alpha}{2} - {}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}\times\mathbf{j}\sin\frac{\alpha}{2}\right)\end{bmatrix}\\
&= \begin{bmatrix}0\\ {}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}\cos^2\frac{\alpha}{2} + \left(\mathbf{j}\times{}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}\right)\sin\frac{\alpha}{2}\cos\frac{\alpha}{2} + \left(({}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}^{\intercal}\mathbf{j})\mathbf{j}\right)\sin^2\frac{\alpha}{2} + \left(\mathbf{j}\times{}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}\right)\sin\frac{\alpha}{2}\cos\frac{\alpha}{2} - \underbrace{\mathbf{j}\times{}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}\times\mathbf{j}}_{{}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}-({}^{\mathcal{B}}_{\mathcal{B}}\mathbf{p}^{\intercal}\mathbf{j})\mathbf{j}}\sin^2\frac{\alpha}{2}\end{bmatrix}
\end{aligned}$$

- conclude that $\quad {}_{\mathcal{A}}\mathbf{p} = {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{q} \otimes {}_{\mathcal{B}}\mathbf{p} \otimes {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{q}^* \quad$ and agree on the short notation ${}_{\mathcal{A}}\mathbf{p} = \left[{}_{\mathcal{B}}\mathbf{p}\right]_{\mathcal{A}}$
- note that we have extended the definition of $\otimes$ to map from $\mathbb{R}^4 \times \mathbb{R}^3 \times \mathbb{R}^4$ to $\mathbb{R}^3$
- Nomenclature and indices vary from source to source, but note the following fact:
  The quaternion that corresponds to the axis and angle of the rotation from frame $\mathcal{A}$ to frame $\mathcal{B}$ is always on the left (right) side when transforming into $\mathcal{A}$- ($\mathcal{B}$-) coordinates.

#### 4.8.6 Rotation Matrix from/to Quaternion

Since we already related axis-angle to rotation matrices, it is easy to deduce that ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{q} = \begin{bmatrix} w & x & y & z \end{bmatrix}^\mathsf{T}$ is realized by

$$
{}^{\mathcal{B}}_{\mathcal{A}}\mathbf{M} = \begin{bmatrix} 1 - 2y^2 - 2z^2 & 2(xy - zw) & 2(xz + yw) \\ 2(xy + zw) & 1 - 2x^2 - 2z^2 & 2(yz - xw) \\ 2(xz - yw) & 2(yz + xw) & 1 - 2x^2 - 2y^2 \end{bmatrix}
$$

A quaternion can be determined from a given rotation matrix in several ways, for example using the trace $(4w^2 - 1)$ or the difference of transpose elements (i.e. $4zw$, $4yw$, $4xw$).

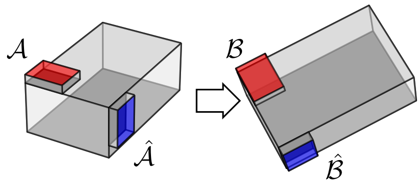
#### 4.8.7 Summary Rotation Matrix $\leftrightarrow$ Quaternion

| **Rotation Matrix** | **Unit Quaternions** |
|---|---|
| ${}_{\mathcal{B}}\mathbf{p} = {}^{\mathcal{A}}_{\mathcal{B}}\mathbf{M}\, {}_{\mathcal{A}}\mathbf{p}$ | ${}_{\mathcal{B}}\mathbf{p} = {}^{\mathcal{A}}_{\mathcal{B}}\mathbf{q} \otimes {}_{\mathcal{A}}\mathbf{p} \otimes {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{q}$ |
| ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{M} = \left({}^{\mathcal{A}}_{\mathcal{B}}\mathbf{M}\right)^{-1} = {}^{\mathcal{A}}_{\mathcal{B}}\mathbf{M}^{\mathrm{T}}$ | ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{q} = {}^{\mathcal{A}}_{\mathcal{B}}\mathbf{q}^{-1} = {}^{\mathcal{A}}_{\mathcal{B}}\mathbf{q}^*$ |
| ${}^{\mathcal{A}}_{\mathcal{C}}\mathbf{M} = {}^{\mathcal{B}}_{\mathcal{C}}\mathbf{M}\, {}^{\mathcal{A}}_{\mathcal{B}}\mathbf{M}$ | ${}^{\mathcal{A}}_{\mathcal{C}}\mathbf{q} = {}^{\mathcal{B}}_{\mathcal{C}}\mathbf{q} \otimes {}^{\mathcal{A}}_{\mathcal{B}}\mathbf{q}$ |
| axis is EV for $\lambda = 1$ | axis is $\mathbf{q}(2:4)$ |
| angle is "hidden" | angle is $2 \arccos \mathbf{q}(1)$ |
| new coordinate system axes are rows | new coordinate system axes are "hidden" |
| describes an orientation | two quaternions $\mathbf{q}$ and $-\,\mathbf{q}$ describe different rotations but **same** orientation |

both are associative but **not** commutative

## 4.9 Multiple Frames per Object

Consider a rigid object with one measurement frame $\mathcal{A}$ and one additional frame of interest $\hat{\mathcal{A}}$

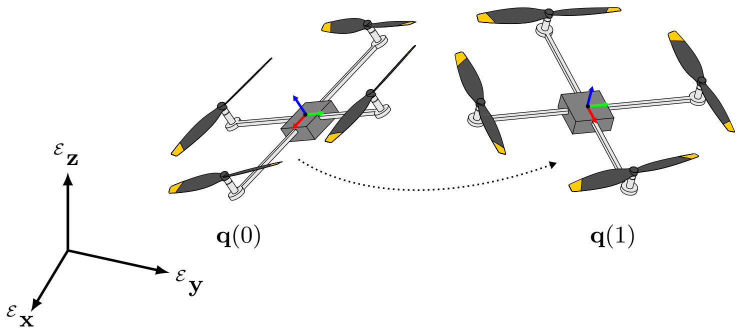



- note that the relative orientation between $\mathcal{A}$ and $\hat{\mathcal{A}}$ remains unchanged, i.e. ${}^{\mathcal{A}}_{\hat{\mathcal{A}}}\mathbf{q} = {}^{\mathcal{B}}_{\hat{\mathcal{B}}}\mathbf{q}$.
- assume that we know ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{q}$ but want to know ${}^{\hat{\mathcal{B}}}_{\hat{\mathcal{A}}}\mathbf{q}$; then the solution is obviously:

$$ {}^{\hat{\mathcal{B}}}_{\hat{\mathcal{A}}}\mathbf{q} \;=\; {}^{\mathcal{A}}_{\hat{\mathcal{A}}}\mathbf{q} \otimes {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{q} \otimes {}^{\hat{\mathcal{B}}}_{\mathcal{B}}\mathbf{q} \;=\; {}^{\mathcal{A}}_{\hat{\mathcal{A}}}\mathbf{q} \otimes {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{q} \otimes {}^{\mathcal{A}}_{\hat{\mathcal{A}}}\mathbf{q}^{*} \quad =: \left[{}^{\mathcal{B}}_{\mathcal{A}}\mathbf{q}\right]_{\hat{\mathcal{A}}} $$

- note the subtle difference to the previous use of this notation for transforming 3D vectors

**Question:** Assume that at time zero a vehicle with an intrinsic coordinate frame has some orientation $\mathbf{q}(0)$ with respect to an inertial navigation frame with vertical z-axis. Assume that a moment later the object has an orientation of $\mathbf{q}(1)$ with respect to the same inertial frame. It is easy to tell whether the object has rotated about any of its intrinsic coordinate axes. However, how can you tell whether the performed rotation was around the vertical axis?




**Answer 1:** Denote the inertial frame by $\mathcal{E}$ and note that $\mathbf{q}(0) = {}^{\mathcal{A}}_{\mathcal{E}}\mathbf{q}$ and $\mathbf{q}(1) = {}^{\mathcal{B}}_{\mathcal{E}}\mathbf{q}$. Calculate the relative orientation ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{q} = {}^{\mathcal{A}}_{\mathcal{E}}\mathbf{q}^{*} \otimes {}^{\mathcal{B}}_{\mathcal{E}}\mathbf{q}$ and transform this rotation into the $\mathcal{E}$-frame:

$$ \left[{}^{\mathcal{B}}_{\mathcal{A}}\mathbf{q}\right]_{\mathcal{E}} \;=\; {}^{\mathcal{A}}_{\mathcal{E}}\mathbf{q} \otimes {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{q} \otimes {}^{\mathcal{A}}_{\mathcal{E}}\mathbf{q}^{*} \;=\; {}^{\mathcal{A}}_{\mathcal{E}}\mathbf{q} \otimes {}^{\mathcal{A}}_{\mathcal{E}}\mathbf{q}^{*} \otimes {}^{\mathcal{B}}_{\mathcal{E}}\mathbf{q} \otimes {}^{\mathcal{A}}_{\mathcal{E}}\mathbf{q}^{*} \;=\; {}^{\mathcal{B}}_{\mathcal{E}}\mathbf{q} \otimes {}^{\mathcal{A}}_{\mathcal{E}}\mathbf{q}^{*} $$

Then check whether this has axis $\begin{bmatrix}0 & 0 & 1\end{bmatrix}^{\intercal}$. Note how unusual the concatenation ${}^{\mathcal{B}}_{\mathcal{E}}\mathbf{q} \otimes {}^{\mathcal{A}}_{\mathcal{E}}\mathbf{q}^{*}$ is!

**Answer 2:** You can also transform the ${}^{\mathcal{E}}\mathbf{z}$ into local coordinates of $\mathcal{C}$ at $t = 0$ and $t = 1$ and check if ${}^{\mathcal{E}}_{\mathcal{C}}\mathbf{z}(0) = {}^{\mathcal{E}}_{\mathcal{C}}\mathbf{z}(1)$. If so, then the global vertical axis is the rotation axis.

## 4.10 Euler Angles

One last representation of 3D orientations to talk about – the one closest to intuition but mathematically the worst.

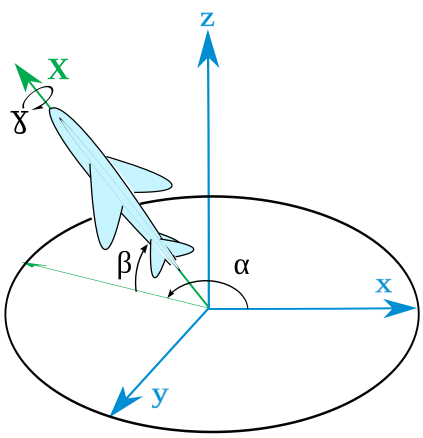
- every orientation can be produced by a concatenation of three rotations around major coordinate axes and can hence be described by the corresponding three angles, e.g.

$$\mathrm{Eu}^{int}_{xzy}(\alpha,\beta,\gamma) \cong \left\{\alpha \text{ @ } \begin{bmatrix}1\\0\\0\end{bmatrix}\right\} \otimes \left\{\beta \text{ @ } \begin{bmatrix}0\\0\\1\end{bmatrix}\right\} \otimes \left\{\gamma \text{ @ } \begin{bmatrix}0\\1\\0\end{bmatrix}\right\}$$

$$\mathrm{Eu}^{int}_{zyx}(\alpha,\beta,\gamma) \cong \left\{\alpha \text{ @ } \begin{bmatrix}0\\0\\1\end{bmatrix}\right\} \otimes \left\{\beta \text{ @ } \begin{bmatrix}0\\1\\0\end{bmatrix}\right\} \otimes \left\{\gamma \text{ @ } \begin{bmatrix}1\\0\\0\end{bmatrix}\right\}$$

- note that there are six possible conventions (axis orders): xyz, xzy, yxz, yzx, zxy, zyx

- however, the most commonly employed convention is zyx, especially in aerospace, where these angles are called yaw, pitch, roll (sometimes heading angle, elevation angle, bank angle). However, note that in aerospace the aircraft fixed coordinate system is defined as the x-axis pointing forward, the z-axis pointing down and y-axis pointing to the right wing.




- note that pitch is a rotation around the rotated y-axis (wing-to-wing axis) of the aircraft and that roll is a rotation around the rotated x-axis (longitudinal axis) of the aircraft

- if we use angle ranges $\alpha \in [-\pi, \pi]$, $\beta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, $\gamma \in [-\pi, \pi]$, then almost every orientation is described by a unique set of angles

**Question:** Consider a daring jet pilot who takes his plane to a pitch (elevation) angle of 90 degrees. If she then rotates the plane along the longitudinal axis, does the roll or the yaw angle change, or both?

**Answer:** This is undefined. It is a singularity of the Euler angle convention. Near the singularity, roll and yaw might fluctuate, while their sum remains almost constant.

### 4.10.1 Euler Angles from Quaternions

- the quaternion is obtained by concatenation of the single rotations

$$\mathrm{Eu}_{zyx}^{int}(\alpha,\beta,\gamma) \cong \begin{bmatrix}\cos\frac{\alpha}{2}\\0\\0\\\sin\frac{\alpha}{2}\end{bmatrix} \otimes \begin{bmatrix}\cos\frac{\beta}{2}\\0\\\sin\frac{\beta}{2}\\0\end{bmatrix} \otimes \begin{bmatrix}\cos\frac{\gamma}{2}\\\sin\frac{\gamma}{2}\\0\\0\end{bmatrix} \tag{11}$$

$$= \begin{bmatrix}\cos(\gamma/2)\cos(\beta/2)\cos(\alpha/2)+\sin(\gamma/2)\sin(\beta/2)\sin(\alpha/2)\\\sin(\gamma/2)\cos(\beta/2)\cos(\alpha/2)-\cos(\gamma/2)\sin(\beta/2)\sin(\alpha/2)\\\cos(\gamma/2)\sin(\beta/2)\cos(\alpha/2)+\sin(\gamma/2)\cos(\beta/2)\sin(\alpha/2)\\\cos(\gamma/2)\cos(\beta/2)\sin(\alpha/2)-\sin(\gamma/2)\sin(\beta/2)\cos(\alpha/2)\end{bmatrix} \tag{12}$$

- the inverse operation is performed by

$$\alpha = \mathrm{atan2}(2(wz+xy), 1-2(y^2+z^2)) \tag{13}$$

$$\beta = \arcsin(2(wy-zx)) \tag{14}$$

$$\gamma = \mathrm{atan2}(2(wx+yz), 1-2(x^2+y^2)) \tag{15}$$

- note that each Euler angle convention has its specific formulas to/from quaternion and rotation matrices

### 4.10.2 Extrinsic and Proper Euler Angles

- note that proper Euler angles use same axis for first and third rotation, e.g. zxz
- as before, six conventions are possible, all of which can describe every 3D orientation
- for clarity, the angles introduced above are sometimes called Tait-Bryan angles
- in literature, intrinsic Euler angle conventions are often denoted by $x$-$y'$-$z''$ and extrinsic by $x$-$y$-$z$
- note furthermore that one can also rotate the object around global/fixed axes
- such rotations are called extrinsic as opposed to the intrinsic ones above
- all proper Euler angle and Tait-Bryan angle conventions can be intrinsic or extrinsic
  → in total 24 possible meanings of the term "Euler angles"
- all of them have singularity points but each of them can capture every 3D orientation

**Question:** What are reasonable angle ranges for an extrinsic proper Euler angle convention zxz, and where are the singularities?

**Answer:** First and third angle in $(-\pi, \pi]$, second angle in $(0, \pi]$. The singularities occur at the second angle being a multiple of $\pi$.

- note that extrinsic and intrinsic Euler angles are related as follows:

$$\mathrm{Eu}^{\mathrm{int}}_{zyx}(\alpha, \beta, \gamma) \quad \cong \quad \mathrm{Eu}^{\mathrm{ext}}_{xyz}(\gamma, \beta, \alpha)$$

  which can be seen from the following arguments:

$$\mathrm{Eu}^{\mathrm{ext}}_{zyx}(\gamma, \beta, \alpha) \quad \cong \quad {}^{\mathcal{D}}_{\mathcal{A}}\mathbf{q} = \underbrace{\{\gamma \ @ \ {}^{\mathcal{A}}_{\mathcal{A}}\mathbf{x}\}}_{{}^{\mathcal{B}}_{\mathcal{A}}\mathbf{q}} \otimes \underbrace{\{\beta \ @ \ {}^{\mathcal{A}}_{\mathcal{B}}\mathbf{y}\}}_{{}^{\mathcal{C}}_{\mathcal{B}}\mathbf{q}} \otimes \{\alpha \ @ \ {}^{\mathcal{A}}_{\mathcal{C}}\mathbf{z}\} \tag{16}$$

$$\cong \quad {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{q} \otimes \left( {}^{\mathcal{A}}_{\mathcal{B}}\mathbf{q} \otimes \{\beta \ @ \ {}^{\mathcal{A}}_{\mathcal{A}}\mathbf{y}\} \otimes {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{q} \right) \otimes \{\alpha \ @ \ {}^{\mathcal{A}}_{\mathcal{C}}\mathbf{z}\} \tag{17}$$

$$\cong \quad \underbrace{\{\beta \ @ \ {}^{\mathcal{A}}_{\mathcal{A}}\mathbf{y}\} \otimes \{\gamma \ @ \ {}^{\mathcal{A}}_{\mathcal{A}}\mathbf{x}\}}_{{}^{\mathcal{C}}_{\mathcal{A}}\mathbf{q}} \otimes \{\alpha \ @ \ {}^{\mathcal{A}}_{\mathcal{C}}\mathbf{z}\} \tag{18}$$

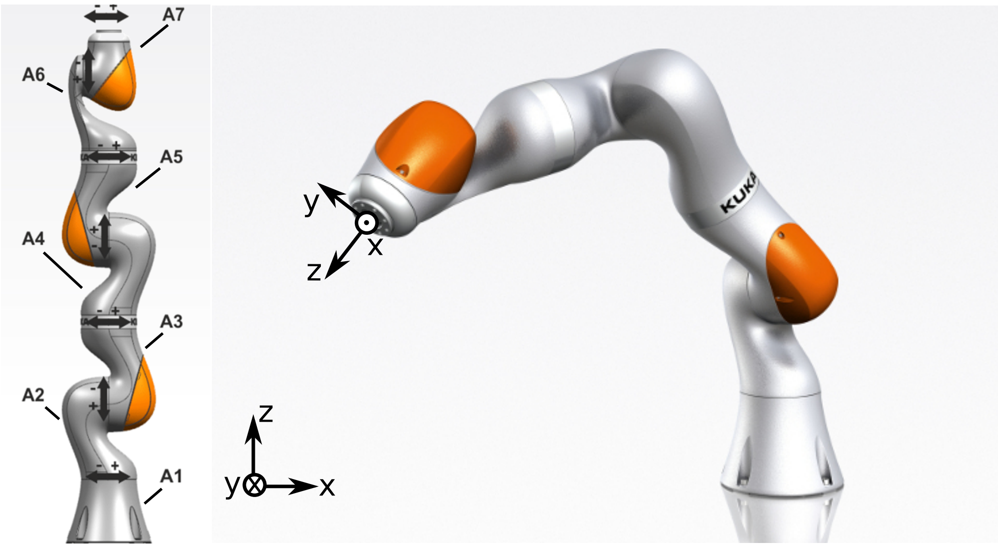
$$\cong \quad {}^{\mathcal{C}}_{\mathcal{A}}\mathbf{q} \otimes \left( {}^{\mathcal{A}}_{\mathcal{C}}\mathbf{q} \otimes \{\alpha \ @ \ {}^{\mathcal{A}}_{\mathcal{A}}\mathbf{z}\} \otimes {}^{\mathcal{C}}_{\mathcal{A}}\mathbf{q} \right) \tag{19}$$

$$\cong \quad \{\alpha \ @ \ {}^{\mathcal{A}}_{\mathcal{A}}\mathbf{z}\} \otimes \{\beta \ @ \ {}^{\mathcal{A}}_{\mathcal{A}}\mathbf{y}\} \otimes \{\gamma \ @ \ {}^{\mathcal{A}}_{\mathcal{A}}\mathbf{x}\} \qquad \text{q.e.d.} \tag{20}$$

- note that with a similar argument we can find the inverse of a set of Euler angles:

$$\left( \mathrm{Eu}^{\mathrm{int}}_{zyx}(\alpha, \beta, \gamma) \right)^{-1} \quad \cong \quad \mathrm{Eu}^{\mathrm{int}}_{xyz}(-\gamma, -\beta, -\alpha)$$

**Question:** Consider a robotic limb with a chain of joints as depicted below. On the left-hand side, all angles are zero, and all segment coordinate systems coincide. On the right-hand side, the first six joints have been set to known angles $\alpha_1, \alpha_2, ..., \alpha_6$. What is the orientation between the base segment coordinate system $\mathcal{A}_1$ and the coordinate system $\mathcal{A}_7$ of the seventh segment?




**Answer:** ${}^{\mathcal{A}_7}_{\mathcal{A}_1}\mathbf{q} \ = \ \mathrm{Eu}^{\mathrm{int}}_{zxz}(\alpha_1, \alpha_2, \alpha_3) \ \otimes \ \mathrm{Eu}^{\mathrm{int}}_{xzx}(-\alpha_4, \alpha_5, \alpha_6)$

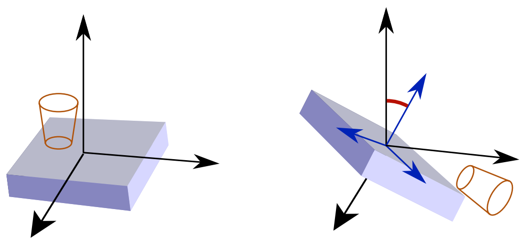
### 4.11 Inclination and Heading Angles

- consider the rotation $\mathbf{q} = \left\{\frac{\pi}{3} \,@\, \frac{1}{\sqrt{3}}[1,1,1]^\intercal\right\}$ of a flying quadrocopter in different Euler angle conventions:

$$\mathbf{q} = \left\{\frac{\pi}{3} \,@\, \frac{1}{\sqrt{3}}[1,1,1]^\intercal\right\} \quad \cong \quad \mathrm{Eu}^{\mathrm{int}}_{zyx}(45, 19.5, 45) \tag{21}$$

$$\mathbf{q} = \left\{\frac{\pi}{3} \,@\, \frac{1}{\sqrt{3}}[1,1,1]^\intercal\right\} \quad \cong \quad \mathrm{Eu}^{\mathrm{int}}_{zxy}(26.6, 41.8, 26.6) \tag{22}$$

- note how much the angles depend on the employed convention!
- but what is the "shortest" rotation that makes the rotor planes horizontal, i.e. how much is the quadrocopter inclined/tilted?
- it is the angle between the local z-axis and the vertical z-axis of the reference frame
- the axis of that rotation is the horizontal axis that is perpendicular to the local z-axis



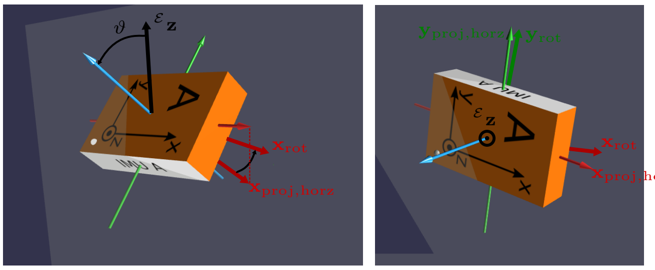
**Question:** You are traveling in a submarine at constant velocity. Your coffee cup is about to slide off the table, because the angle between the normal vector of the table and the vertical axis is not zero. Which Euler angle of which angle convention is this?

**Answer:** It is the second angle $\beta$ of $\mathrm{Eu}^{\mathrm{ext}}_{zxz}(\alpha,\beta,\gamma)$, of $\mathrm{Eu}^{\mathrm{ext}}_{zyz}(\alpha,\beta,\gamma)$, of $\mathrm{Eu}^{\mathrm{int}}_{zxz}(\alpha,\beta,\gamma)$, and of $\mathrm{Eu}^{\mathrm{int}}_{zyz}(\alpha,\beta,\gamma)$. The corresponding heading angle is $\alpha+\gamma$ (for all conventions).

- for almost each $\mathbf{q}$ there exists exactly one heading angle $\varphi$, one inclination angle $\vartheta$ and one horizontal axis $\mathbf{k}$ such that

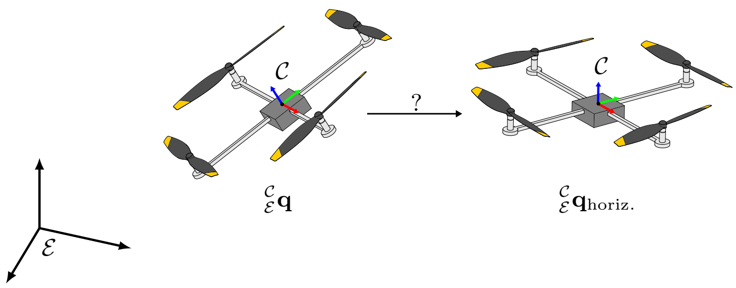
$$\mathbf{q} = \left\{\varphi \,@\, \begin{bmatrix}0\\0\\1\end{bmatrix}\right\} \otimes \{\vartheta \,@\, \mathbf{k}\}, \quad \mathbf{k} \perp \begin{bmatrix}0\\0\\1\end{bmatrix}. \tag{23}$$

- from now on, use the words "inclination" and "heading" in this sense and be aware that this heading is in general not the same as the yaw angle of any Euler angle convention
- summary: we can decompose orientations either into a set of three Euler angles or into a heading and an inclination (with a horizontal axis of rotation for the latter)



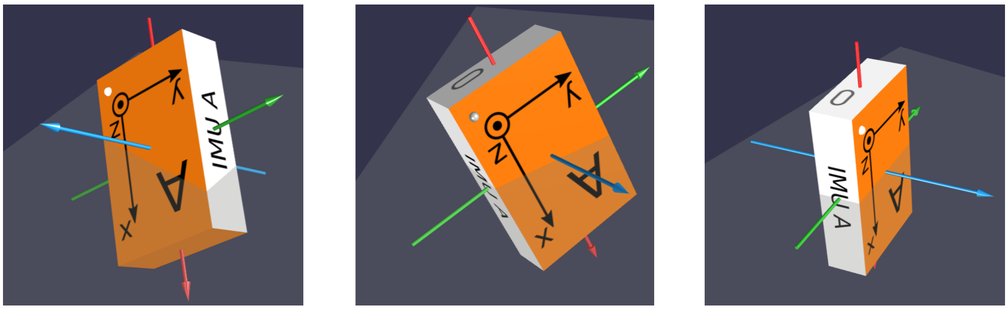

Difference between the projections of the axes into the horizontal plane (proj, horz) and the axes after being rotated the shortest way to horizontal around the rotation axis $\mathbf{k}$ by the inclination angle $-\vartheta$ (rot). Note that during the rotation around the axis $\mathbf{k}$ the heading of the object does not change.

**Question:** The orientation of a quadrocopter $\mathcal{C}$ with respect to a fixed inertial frame $\mathcal{E}$ at any given time is denoted as ${}^{\mathcal{C}}_{\mathcal{E}}\mathbf{q}$. The Euler angles of the $z$-$y'$-$x''$ convention are denoted as $\alpha$ (Yaw), $\beta$ (Pitch) and $\gamma$ (Roll). Around which axis $\mathbf{k}$ should the quadrocopter rotate to be in a horizontal orientation ($\beta = \gamma = 0$)?



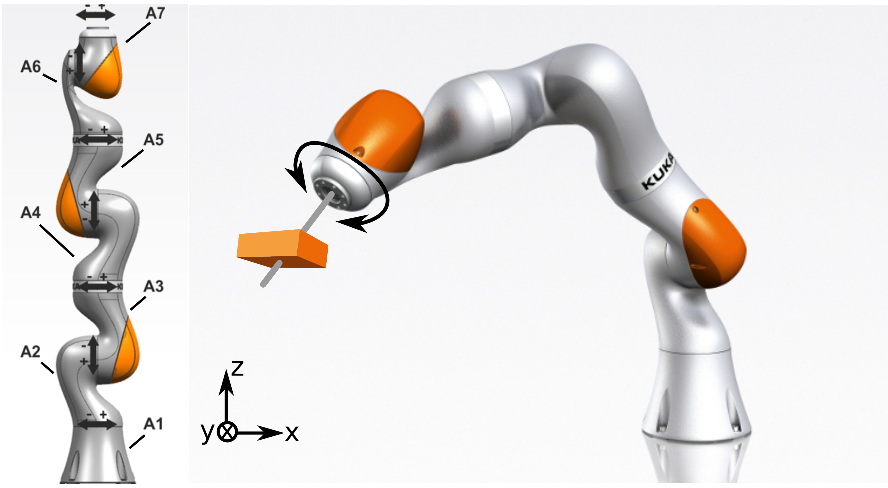

a) $\alpha = 30°$ $\beta = 45°$ $\gamma = 45°$
b) $\alpha = 90°$ $\beta = 70°$ $\gamma = 40°$
c) $\alpha = 180°$ $\beta = 90°$ $\gamma = 80°$

**Answer:** Transform the Euler angles into a quaternion with ${}^{\mathcal{C}}_{\mathcal{E}}\mathbf{q} = \mathrm{Eu}^{\text{int}}_{zyx}(\alpha, \beta, \gamma)$ and calculate the inclination angle $\vartheta$ and the axis $\mathbf{k}$. The heading does not influence the shortest path to a horizontal orientation.
a) $\vartheta = 60°$ $\mathbf{k} = [0.58, 0.82, 0]^\intercal$
b) $\vartheta = 75°$ $\mathbf{k} = [0.23, 0.97, 0]^\intercal$
c) $\vartheta = 90°$ $\mathbf{k} = [0, 1, 0]^\intercal$




**(a)** **(b)** **(c)**

## 4.12 Projection Angles of Rotations




**Question:** Consider a hinge joint connecting a rigid body to a fixed object, e.g. the ground. The body can move only around the known hinge joint axis with coordinates

$$ {}_{\mathcal{A}_1}\mathbf{j} = {}^{\mathcal{A}_7}_{\mathcal{A}_1}\mathbf{q} \otimes \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^\intercal \otimes {}^{\mathcal{A}_1}_{\mathcal{A}_7}\mathbf{q} $$

in the reference frame $\mathcal{A}_1 = \mathcal{E}$. Theoretically, the IMU $\mathcal{S}$ that is attached to the limb should measure an orientation around that axis. However, due to mechanical inaccuracies or measurement errors, the axis of the measured quaternion ${}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}$ is not precisely ${}_{\mathcal{E}}\mathbf{j}$. What is (without further knowledge) the best available estimate of the joint angle?

**Answer:** Find the angle $\varphi$ around $\mathbf{j}$ that minimizes the angle $\vartheta$ of the residual rotation.

- decompose a given rotation into one rotation around a given axis and one residual rotation (with an arbitrary axis)

$$ \mathbf{q} = [w, x, y, z]^\intercal = \underbrace{\{\varphi \,@\, \mathbf{j}\}}_{\mathbf{q}_{\text{proj}}} \otimes \underbrace{\{\vartheta \,@\, \mathbf{k}\}}_{\mathbf{q}_{\text{rest}}}, \tag{24} $$

- note how (the angle of) the residual rotation depends on $\varphi$:

$$ \mathbf{q}_{\text{rest}} = \left( \ \{\varphi \,@\, \mathbf{j}\} \ \right)^{-1} \otimes \ \mathbf{q} \tag{25} $$

$$ \tag{26} $$

- using the definition of quaternion multiplication, this yields

$$ \begin{aligned} \mathbf{q}_{\text{rest}}[1] &= w \cos\frac{\varphi}{2} + ([x, y, z]\,\mathbf{j}) \sin\frac{\varphi}{2} && (27) \\ &= \sqrt{w^2 + ([x, y, z]\,\mathbf{j})^2}\, \cos\left(\frac{\varphi}{2} - \operatorname{atan2}([x, y, z]\,\mathbf{j}\,,\, w)\right) && (28) \end{aligned} $$

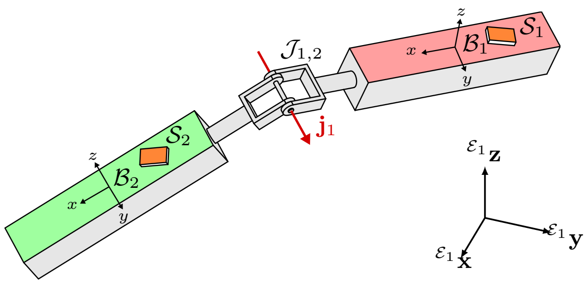
- choose the value for $\varphi$ that minimizes the angle of $\mathbf{q}_{\text{rest}}$:

$$\varphi = \arg\max_{\varphi} |\, \mathbf{q}_{\mathrm{rest}}[1]| \tag{29}$$

$$= 2\,\mathrm{atan2}([x, y, z]\,\mathbf{j}\,,\ w) \tag{30}$$

- note that $\mathbf{q}_{\mathrm{rest}}$ has a rotation axis perpendicular to the given axis, i.e. $\mathbf{j} \perp \mathbf{k}$
- we call $\mathbf{q}_{\mathrm{proj}}$ the projection of $\mathbf{q}$ onto the axis $\mathbf{j}$ and $\varphi$ the corresponding projection angle

**Question:** Consider a hinge joint connecting two rigid bodies $\mathcal{B}_1$ and $\mathcal{B}_2$. The relative rotation of the two bodies is limited to rotations around the joint axis $\mathbf{j}$. The orientations of the two bodies estimated in the same reference frame are ${}^{\mathcal{B}_1}_{\mathcal{E}}\mathbf{q}$ and ${}^{\mathcal{B}_2}_{\mathcal{E}}\mathbf{q}$. Due to measurement and estimation inaccuracies, the relative orientation ${}^{\mathcal{B}_2}_{\mathcal{B}_1}\mathbf{q}$ does not perfectly describe a rotation of the joint angle $\varphi$ around $\mathbf{j}$. How can we determine the joint angle $\varphi$?




**Answer:** By projecting the relative orientation ${}^{\mathcal{B}_2}_{\mathcal{B}_1}\mathbf{q} = {}^{\mathcal{B}_1}_{\mathcal{E}}\mathbf{q}^{-1} \otimes {}^{\mathcal{B}_2}_{\mathcal{E}}\mathbf{q}$ onto the axis ${}_{\mathcal{B}_1}\mathbf{j}$ in local coordinates of $\mathcal{B}_1$.

# 5 Orientation Estimation using Strapdown Systems

- consider an object (free to rotate and move in 3D) and a strapdown IMU attached to it
- Goal: track orientation of an IMU relative to a fixed inertial (reference) coordinate system
- describe this orientation, for example, by a rotation matrix ${}^{\mathcal{S}}_{\mathcal{E}}\mathbf{M}(t)$ or a quaternion ${}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}(t)$
- $\mathcal{E}$ should be a *common* or *global* reference frame in the sense that the orientation of multiple objects (IMUs) shall be determined with respect to the same coordinate system
- typical choice: ${}^{\mathcal{E}}\mathbf{z}$ points vertically up and ${}^{\mathcal{E}}\mathbf{x}$ (or ${}^{\mathcal{E}}\mathbf{y}$) points horizontally north (or south)

## 5.1 Orientation from Gravity and Magnetic Field Vector

- note that the orientation of an IMU is already known from rest measurements of the accelerometers $\mathbf{a}(t_{\text{rest}})$ and the magnetometers $\mathbf{m}(t_{\text{rest}})$, since they must fulfill

$$\begin{bmatrix}0 & 0 & \|\mathbf{a}(t_{\text{rest}})\|_2\end{bmatrix}^\intercal = {}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}(t_{\text{rest}}) \otimes \mathbf{a}(t_{\text{rest}}) \otimes {}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}(t_{\text{rest}})^* \tag{31}$$

$$\begin{bmatrix}?_{(>0)} & 0 & ?\end{bmatrix}^\intercal = {}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}(t_{\text{rest}}) \otimes \mathbf{m}(t_{\text{rest}}) \otimes {}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}(t_{\text{rest}})^*, \tag{32}$$

- which yields four (three) equations for the four (three) unknown entries of ${}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}(t_{\text{rest}})$.
- alternatively, one can construct/find the axes of $\mathcal{E}$ in $\mathcal{S}$-coordinates:

$${}^{\mathcal{E}}_{\mathcal{S}}\mathbf{z} = \frac{\mathbf{a}(t_{\text{rest}})}{\|\mathbf{a}(t_{\text{rest}})\|_2}, \qquad {}^{\mathcal{E}}_{\mathcal{S}}\mathbf{y} = \frac{{}^{\mathcal{E}}_{\mathcal{S}}\mathbf{z} \times \mathbf{m}(t_{\text{rest}})}{\|{}^{\mathcal{E}}_{\mathcal{S}}\mathbf{z} \times \mathbf{m}(t_{\text{rest}})\|_2}, \qquad {}^{\mathcal{E}}_{\mathcal{S}}\mathbf{x} = {}^{\mathcal{E}}_{\mathcal{S}}\mathbf{y} \times {}^{\mathcal{E}}_{\mathcal{S}}\mathbf{z}$$

- and stack them to obtain ${}^{\mathcal{E}}_{\mathcal{S}}\mathbf{M}(t_{\text{rest}}) = \begin{bmatrix}{}^{\mathcal{E}}_{\mathcal{S}}\mathbf{x} & {}^{\mathcal{E}}_{\mathcal{S}}\mathbf{y} & {}^{\mathcal{E}}_{\mathcal{S}}\mathbf{z}\end{bmatrix} = {}^{\mathcal{S}}_{\mathcal{E}}\mathbf{M}(t_{\text{rest}})^*$
- mind that this approch relies on noise-free measurements and on the assumption (31), which only holds if the IMU frame neither rotates nor accelerates (i.e. is an inertial frame)

## 5.2 Orientation Strapdown Integration

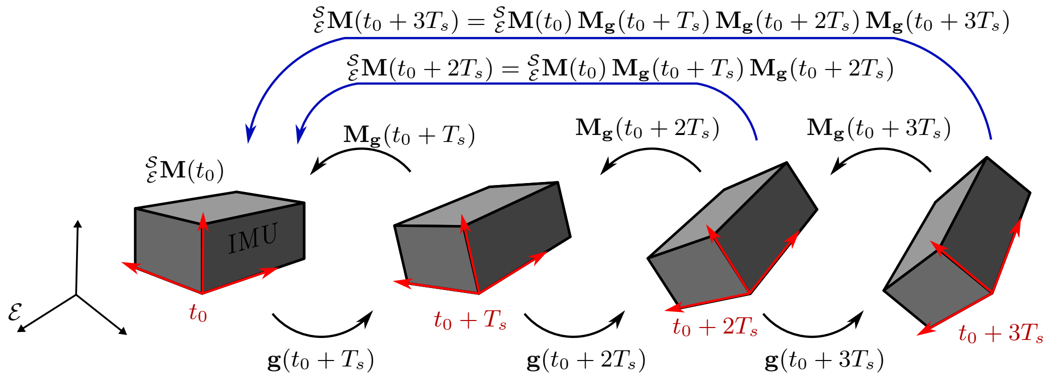
- Assumption: angular velocity signal $\mathbf{g}(t)$ with sufficiently small sample period $T_s$ well describes the change of orientation from sample instant $(t - T_s)$ to sample instant $t$
- this holds if (but not only if) rotation axis and velocity are sampling-period-wise constant

$$\text{for any sampling instant } t: \quad \mathbf{g}(\tau) = \mathbf{g}(t)\ \forall \tau \in (t - T_s, t] \tag{33}$$

- we define the previous-to-current-sample rotation $\mathbf{M}_\mathbf{g}(t)$ and find that

$$\mathbf{M}_\mathbf{g}(t) := {}^{\mathcal{S}}_{\mathcal{E}}\mathbf{M}(t-T_s)^*\,{}^{\mathcal{S}}_{\mathcal{E}}\mathbf{M}(t) \;\cong\; {}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}(t-T_s)^* \otimes {}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}(t) \;\cong\; \|\mathbf{g}(t)\|_2 T_s \;@\; \frac{\mathbf{g}(t)}{\|\mathbf{g}(t)\|_2} \tag{34}$$

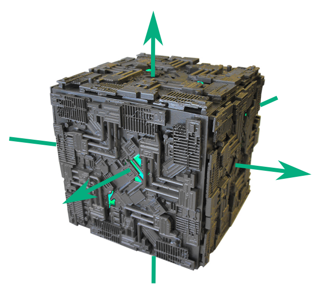
- assume there exists a sampling instant $t = t_0$, for which the orientation ${}^{\mathcal{S}}_{\mathcal{E}}\mathbf{M}(t_0)$ is known
- then ${}^{\mathcal{S}}_{\mathcal{E}}\mathbf{M}(t) = {}^{\mathcal{S}}_{\mathcal{E}}\mathbf{M}(t_0)\left(\mathbf{M}_\mathbf{g}(\tau + T_s)\,\mathbf{M}_\mathbf{g}(\tau + 2T_s)\ldots\,\mathbf{M}_\mathbf{g}(t)\right)$ and equivalently for ${}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}(t)$




- note that strapdown integration cannot be accurate on large timescales, for two reasons
  - the rotation might not be well described by discrete-time measurements
  - the measurement $\mathbf{g}(t)$ has non-zero bias and non-zero noise

→ combine orientation strapdown integration with accelerometer and magnetometer readings

**Question:** Consider a cubic spaceship moving at constant velocity in gravity-free space somewhere in the delta quadrant. It has three pairs of nozzles that exert a torque around the x-, y-, and z-axis of the spaceship. An extremely fast feedback controller sets these torques in real time to assure desired angular velocities $\mathbf{g} = [g_x, g_y, g_z]^\intercal$ around the principal axes of the spaceship. Formulate a state-space model that captures the influence of these angular velocities on the the spaceship's orientation $\mathcal{C}$ with respect to its initial orientation $\mathcal{C}_0$.

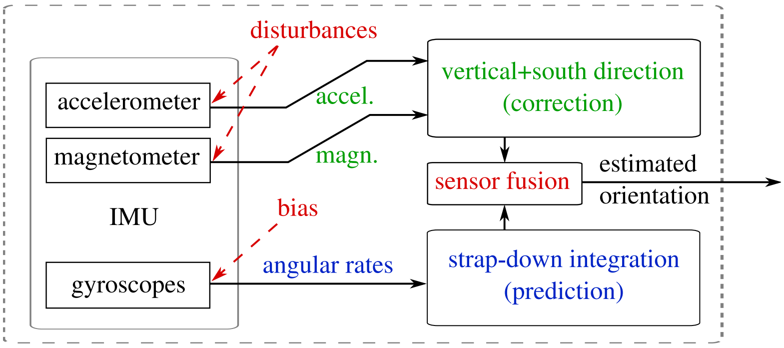


**Answer:** The orientation cannot be obtained by element-wise integration of the angular velocities. Instead, we choose (for example) a quaternion state and find that, for sufficiently high sampling rates:

$$ {}^{\mathcal{C}}_{\mathcal{C}_0}\mathbf{q}(t) = {}^{\mathcal{C}}_{\mathcal{C}_0}\mathbf{q}(t-T_s) \otimes \begin{bmatrix} \cos(\frac{1}{2}\|\mathbf{g}(t)\|_2 T_s) \\ \sin(\frac{1}{2}\|\mathbf{g}(t)\|_2 T_s)\frac{\mathbf{g}(t)}{\|\mathbf{g}(t)\|_2} \end{bmatrix}. \tag{35} $$

## 5.3 Inertial Sensor Fusion

- note the following general scheme of sensor fusion in orientation estimation:




- consider an arbitrary sampling instant $t$ and assume that the following are given:
  - an estimate ${}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}(t-T_s)$ of the IMU orientation at the previous sampling instant
  - a gyroscope-based measurement ${}^{\mathcal{S}_t}_{\mathcal{S}_{t-T_s}}\mathbf{q}$ of the change in orientation according to (34)
  - an estimate ${}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}_{\text{acc,mag}}(t)$ of the IMU orientation based on current acc/mag readings
- Task: correct the prediction ${}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}_{\text{gyr}}(t) = {}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}(t-T_s)\,{}^{\mathcal{S}_t}_{\mathcal{S}_{t-T_s}}\mathbf{q}$ if it disagrees with ${}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}_{\text{acc,mag}}(t)$
- Solution: weighted average between both quaternions

$$\alpha_{\text{disag}} \ @\ \mathbf{j}_{\text{disag}} \cong \left({}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}_{\text{gyr}}(t)\right)^* \otimes {}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}_{\text{acc,mag}}(t) \tag{36}$$

$${}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}(t) = {}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}_{\text{gyr}}(t) \otimes \mathbf{q}\left\{\underbrace{k\,\alpha_{\text{disag}}}_{\alpha_{\text{corr}}} \ @\ \mathbf{j}_{\text{disag}}\right\}, \quad k \in [0,1] \tag{37}$$

- this interpolation between two orientations is known as spherical linear interpolation (slerp)
- note that, for $\mathbf{g}(t) = 0$, the estimate ${}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}(t)$ converges exponentially to ${}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}_{\text{acc,mag}}(t)$
- the time constant of that convergence depends on $T_s$ and the fusion weight $k$
- if the IMU is at rest, ${}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}_{\text{acc,mag}}(t)$ and $\mathbf{g}(t)$ are constant but (due to bias) $\mathbf{g}(t) \neq 0$, thus the estimate ${}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}(t)$ converges to a constant orientation ${}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}(t \to \infty) \neq {}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}_{\text{acc,mag}}(t)$
- this residual error depends on the bias as well as on the fusion weight $k$

**Example:** $${}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}_{\text{gyr}}(t) = \begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\\ 0\\ 0\end{bmatrix},\ {}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}_{\text{acc,mag}}(t) = \begin{bmatrix}\frac{1}{2}\\ \frac{1}{2}\\ \frac{1}{2}\\ \frac{1}{2}\end{bmatrix},\ k = \tfrac{1}{3} \ \to\ {}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}(t) = \begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\\ 0\\ 0\end{bmatrix} \otimes \underbrace{\begin{bmatrix}\frac{\sqrt{6}+\sqrt{2}}{4}\\ 0\\ \frac{\sqrt{6}-\sqrt{2}}{4}\\ 0\end{bmatrix}}_{30^\circ\ @\ {}^{\mathcal{S}}\mathbf{y}}$$

### 5.3.1 Adaptive Fusion Weights

- the weight $k$ can be adjusted for each sample based on the norm of $\mathbf{a}(t)$, $\mathbf{m}(t)$
- e.g. when $\|\mathbf{a}(t)\|$ is not close enough to 9.8 then $k$ might be decreased automatically
- since $\|\mathbf{a}(t)\|$ might (shortly) be close to 9.8 even during accelerated motions, require

$$||\mathbf{a}(\tau)\| - 9.8| < \epsilon \quad \forall \tau \in [t - nT_s, t] \quad \text{for some } \epsilon, n$$

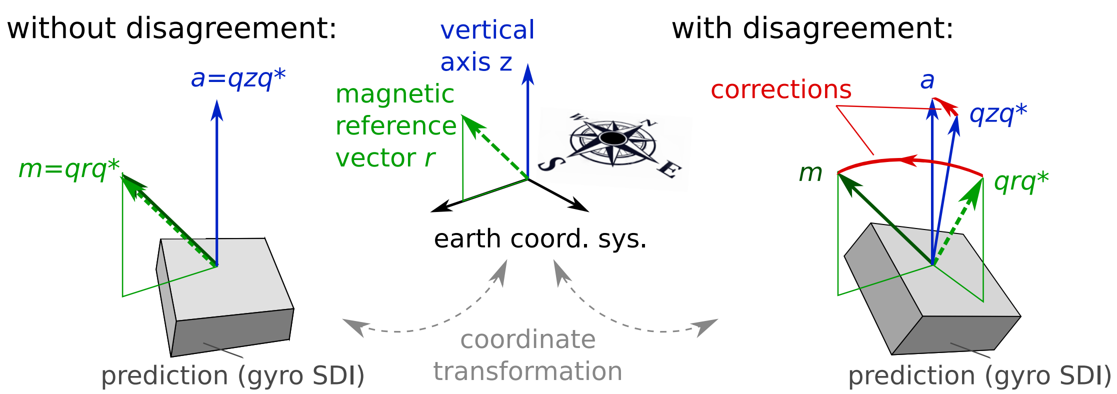
## 5.4 Advanced Sensor Fusion for Orientation Estimation

- in the simple sensor fusion algorithm above, we can only balance between acc/mag and gyr, but choose trust weights for acc and mag independently
- often, the magnetic field is good while the velocity changes are large, or the velocity is constant but the magnetic field is disturbed
- consider the task of tracking the orientation of a sensor $\mathcal{S}$ with respect to a fixed reference frame $\mathcal{E}$ with vertical axis ${}^{\mathcal{E}}\mathbf{r}_{\text{acc}}$ and horizontally northbound axis ${}^{\mathcal{E}}\mathbf{r}_{\text{mag}}$
- assume that initialization ${}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}_{\text{gyr}}(t_0) = \texttt{quat\_from\_accmag}(\mathbf{a}(t_0), \mathbf{m}(t_0))$ and gyroscope-based prediction ${}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}_{\text{gyr}}(t)$ are performed as above
- define the following disagreements

$$f_{\text{mag}}({}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}) := \left|\left| {}^{\mathcal{E}}_{\mathcal{S}}\mathbf{q} \; {}^{\mathcal{E}}_{\mathcal{E}}\mathbf{r}_{\text{mag}} \; {}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q} - {}_{\mathcal{S}}\mathbf{m}(t) \right|\right|_2, \tag{38}$$

$$f_{\text{acc}}({}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}) := \left|\left| {}^{\mathcal{E}}_{\mathcal{S}}\mathbf{q} \; {}^{\mathcal{E}}_{\mathcal{E}}\mathbf{r}_{\text{acc}} \; {}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q} - {}_{\mathcal{S}}\mathbf{a}(t) \right|\right|_2. \tag{39}$$

- if the gyroscope-based prediction ${}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}_{\text{gyr}}(t)$ was perfect and if the acc/mag measurements were undisturbed, then both disagreements would be zero
- if integration drift or magnetic disturbances occur, the measured magnetic field vector ${}_{\mathcal{S}}\mathbf{m}(t)$ and the transformed reference vector ${}^{\mathcal{E}}_{\mathcal{S}}\mathbf{r}_{\text{mag}}$ do not coincide
- the prediction is then corrected by a small rotation into the direction of the measurement
- these corrections can be applied for acc and mag measurements separately and with different (time-dependent) weights



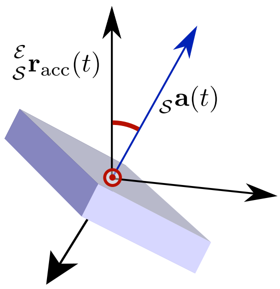
### 5.4.1 Accelerometer-based Correction

**Question:** Consider a reference frame with ${}^{\mathcal{E}}_{\mathcal{E}}\mathbf{r}_{\mathrm{acc}} = [0, 0, 9.8]^\intercal$ and a sample instant $t$ at which the prediction yields ${}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}(t) = [1, 0, 0, 0]^\intercal$ and the measured acceleration is ${}_{\mathcal{S}}\mathbf{a}(t) = [2, 2, 9.4]^\intercal$. Which rotation $\mathbf{q}_{\mathrm{acc}}(t)$ rotates the measurement onto the transformed reference?

$$ {}^{\mathcal{E}}_{\mathcal{S}}\mathbf{r}_{\mathrm{acc}}(t) \;=\; \mathbf{q}_{\mathrm{acc}}(t) \otimes {}_{\mathcal{S}}\mathbf{a}(t) \otimes \mathbf{q}_{\mathrm{acc}}(t)^* $$

**Answer:** It is the rotation with the direct angle $\alpha_{\mathrm{err,acc}}(t) = \sphericalangle\left({}_{\mathcal{S}}\mathbf{a}(t), {}^{\mathcal{E}}_{\mathcal{S}}\mathbf{r}_{\mathrm{acc}}(t)\right)$ between both vectors and the axis $\mathbf{x}_{\mathrm{corr,acc}} = \frac{{}_{\mathcal{S}}\mathbf{a}(t) \times {}^{\mathcal{E}}_{\mathcal{S}}\mathbf{r}_{\mathrm{acc}}}{\left\| {}_{\mathcal{S}}\mathbf{a}(t) \times {}^{\mathcal{E}}_{\mathcal{S}}\mathbf{r}_{\mathrm{acc}} \right\|_2}$ that is perpendicular to both vectors.

$$ \mathbf{q}_{\mathrm{acc}}(t) = \begin{bmatrix} \cos\left(\frac{1}{2}\alpha_{\mathrm{err,acc}}(t)\right) \\ \sin\left(\frac{1}{2}\alpha_{\mathrm{err,acc}}(t)\right) \mathbf{x}_{\mathrm{corr,acc}}(t) \end{bmatrix} $$




- note that the concatenation of ${}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}_{\mathrm{gyr}}(t) \otimes \mathbf{q}_{\mathrm{acc}}(t)$ minimizes the cost function, since

$$ {}^{\mathcal{E}}_{\mathcal{E}}\mathbf{r}_{\mathrm{acc}} = {}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}_{\mathrm{gyr}}(t)\, {}^{\mathcal{E}}_{\mathcal{S}}\mathbf{r}_{\mathrm{acc}}\, {}^{\mathcal{E}}_{\mathcal{S}}\mathbf{q}_{\mathrm{gyr}}(t) $$

- as discussed above, it is desirable to balance between drift compensation and disturbance rejection
- therefore, use only a small portion of $\alpha_{\mathrm{err,acc}}(t)$ in every sampling interval

$$ \begin{aligned} {}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}_{\mathrm{gyracc}}(t) &= {}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}_{\mathrm{gyr}}(t) \otimes \mathbf{q}_{\mathrm{corr,acc}}(t), \\ \mathbf{q}_{\mathrm{corr,acc}}(t) &= \begin{bmatrix} \cos\left(\frac{1}{2}k_{\mathrm{acc}}\,\alpha_{\mathrm{err,acc}}(t)\right) \\ \sin\left(\frac{1}{2}k_{\mathrm{acc}}\,\alpha_{\mathrm{err,acc}}(t)\right) \mathbf{x}_{\mathrm{corr,acc}}(t) \end{bmatrix}, \end{aligned} \tag{40} $$

- $k_{\mathrm{acc}} \in [0, 1]$ is the adjustable sensor fusion weight for the accelerometer readings
- a similar approach was published by Mahony et al.[1]

### 5.4.2 Magnetometer-based Correction

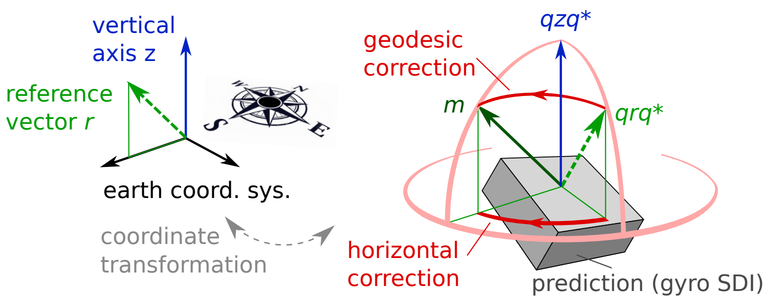
- it is desirable to correct only the heading but not the inclination of ${}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}_{\text{gyracc}}(t)$
- thereby, we assure that magnetic disturbances can only influence the heading accuracy
- use ${}^{\mathcal{E}}_{\mathcal{S}}\mathbf{r}_{\text{acc}}$ to project the measured magnetic field vector ${}_{\mathcal{S}}\mathbf{m}(t)$ into the horizontal plane:

$$ {}_{\mathcal{S}}\bar{\mathbf{m}}(t) = {}_{\mathcal{S}}\mathbf{m}(t) - \left({}_{\mathcal{S}}\mathbf{m}(t) \cdot {}^{\mathcal{E}}_{\mathcal{S}}\mathbf{r}_{\text{acc}}\right) {}^{\mathcal{E}}_{\mathcal{S}}\mathbf{r}_{\text{acc}} $$

- and use a horizontal magnetic reference vector, e.g. ${}^{\mathcal{E}}_{\mathcal{E}}\mathbf{r}_{\text{mag}} = [1, 0, 0]^\intercal$ or $[0, -1, 0]^\intercal$
- use the angle between both, and their cross product, to apply the correction step

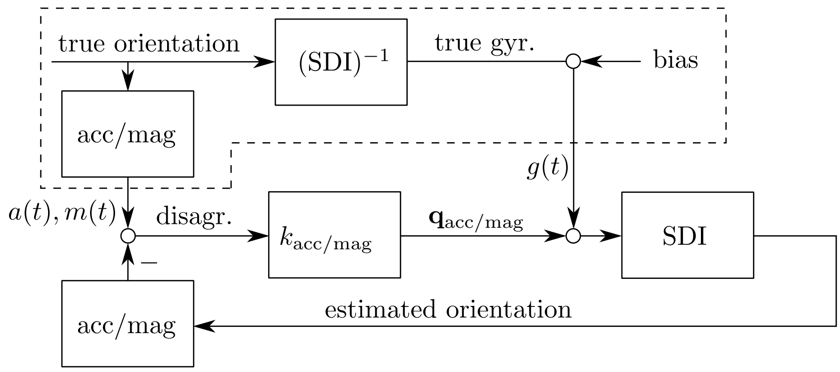
$$
\begin{aligned}
{}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}_{\text{gyraccmag}}(t) &= {}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}_{\text{gyracc}}(t) \otimes \mathbf{q}_{\text{corr,mag}}(t), \\
\mathbf{q}_{\text{corr,mag}}(t) &= \begin{bmatrix} \cos\left(\frac{1}{2} k_{\text{mag}}\, \alpha_{\text{err,mag}}(t)\right) \\ \sin\left(\frac{1}{2} k_{\text{mag}}\, \alpha_{\text{err,mag}}(t)\right) \mathbf{x}_{\text{corr,mag}}(t) \end{bmatrix}, \\
\alpha_{\text{err,mag}}(t) &= \sphericalangle\left(\bar{\mathbf{m}}(t), {}^{\mathcal{E}}_{\mathcal{S}}\mathbf{r}_{\text{mag}}\right), \\
\mathbf{x}_{\text{corr,mag}}(t) &= \frac{\bar{\mathbf{m}}(t) \times {}^{\mathcal{S}}\mathbf{r}_{\text{mag}}(t)}{\|\bar{\mathbf{m}}(t) \times {}^{\mathcal{E}}_{\mathcal{S}}\mathbf{r}_{\text{mag}}(t)\|_2} = \frac{{}^{\mathcal{E}}_{\mathcal{S}}\mathbf{r}_{\text{acc}}}{\left\|{}^{\mathcal{E}}_{\mathcal{S}}\mathbf{r}_{\text{acc}}\right\|_2},
\end{aligned}
\tag{41}
$$

- $k_{\text{mag}} \in [0, 1]$ is the adjustable sensor fusion weight of the magnetometer readings
- the coordinate system is rotated around a vertical axis, which leaves the inclination portion of the orientation estimate unchanged



### 5.4.3 Bias Compensation by Integral Action

- integration drift cannot be compensated completely by a correction that is proportional to the observed disagreement
- this becomes evident if bias is interpreted as a step-type input disturbance on a plant with integral dynamics, for which we have so far only designed a proportional controller




- propose the following bias estimation that exhibits integral action and is based on both observed disagreements:

$$\begin{aligned}\mathbf{b}(t) = \mathbf{b}(t - T_s) &- k_{\text{bias,acc}}\, \alpha_{\text{err,acc}}(t)\, \mathbf{x}_{\text{corr,acc}}(t) \\ &- k_{\text{bias,mag}}\, \alpha_{\text{err,mag}}(t)\, \mathbf{x}_{\text{corr,mag}}(t),\end{aligned} \tag{42}$$

- $k_{\text{bias,acc}}, k_{\text{bias,mag}} \in [0, 1]$ are adjustable gains
- for each sampling interval, the estimated bias is used to calculate corrected angular rates

$${}^{\mathcal{S}}\breve{\mathbf{g}}(t) = {}^{\mathcal{S}}\mathbf{g}(t) - \mathbf{b}(t - T_s). \tag{43}$$

- we obtain an improved prediction ${}^{\mathcal{S}}_{\mathcal{E}}\breve{\mathbf{q}}_{\text{gyr}}(t)$, which is then corrected as described above
- note that (in the linear, continuous-time case) the closed-loop dynamics of an integrator plant and a PI-controller are asymptotically stable for all combinations of positive proportional and integral gain – at least as long as there are no delays

# 6 Magnetometer-free orientation estimation

## 6.1 Motivation

- in indoor environments or near ferromagnetic material the magnetic field can be heavily disturbed/inhomogenous

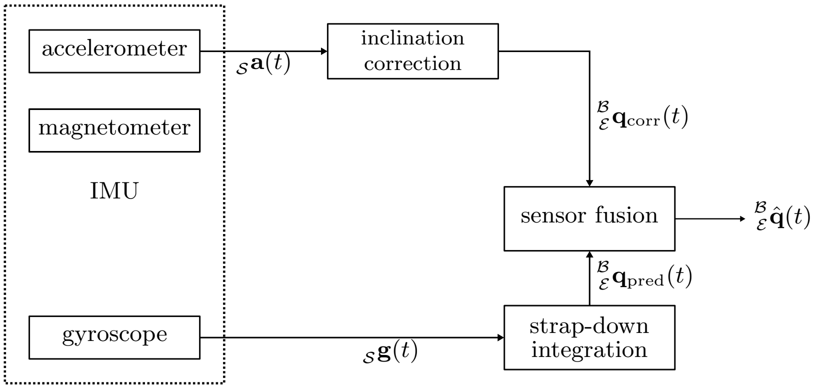
Magnetic field in indoor environment [2]

- this affects the measurement of the magnetometer that is supposed to measure directional information from the earth's magnetic field

## 6.2 6D sensor fusion without magnetometers

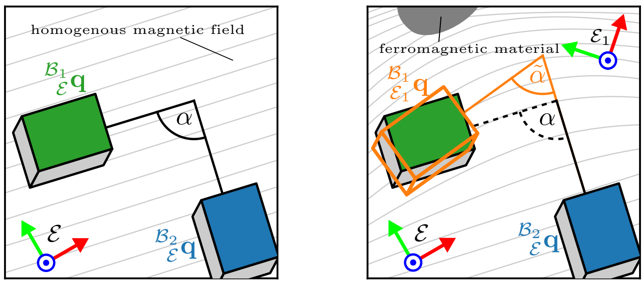
- measuring an inhomogeneous or disturbed magnetic field yields no valid/useful information
- an approach is to omit the magnetometer readings completely and only use the gyroscope and accelerometer for orientation estimation
- this is based on strap-down integration of the gyroscope to obtain a prediction of the orientation and using the accelerometer to correct the inclination to get an estimate ${}^{\mathcal{B}_1}_{\mathcal{E}}\hat{\mathbf{q}}$ of the orientation ${}^{\mathcal{B}_1}_{\mathcal{E}}\mathbf{q}$ of a body $\mathcal{B}_1$ with respect to an inertial frame $\mathcal{E}$ with one vertical and one horizontally northbound axis

- in 9D sensor fusion, the magnetometer-based correction accounts for errors in the heading component of the estimated orientation ${}^{\mathcal{B}_1}_{\mathcal{E}}\hat{\mathbf{q}}$ (see Sec. 5.4.2)
- in 6D sensor fusion, the heading component of ${}^{\mathcal{B}_1}_{\mathcal{E}}\hat{\mathbf{q}}$ is only dependent on the initial orientation ${}^{\mathcal{B}_1}_{\mathcal{E}}\hat{\mathbf{q}}(t = 0)$ and the strap-down integration of the gyroscope
- this implies the heading component of ${}^{\mathcal{B}_1}_{\mathcal{E}}\hat{\mathbf{q}}(t)$ is incorrect and won't converge towards the true heading because of the lack of a correction algorithm for the following reasons
  - the heading of the initial orientation ${}^{\mathcal{B}_1}_{\mathcal{E}}\hat{\mathbf{q}}(t = 0)$ is incorrect
  - the strap-down integration of the gyroscope does not yield a perfect representation of the underlying motion (as it is usually the case because of noise, bias or insufficient sampling)
  - the lack of a heading correction algorithm




General approach to a 6D sensor fusion algorithm

- this can be described as if the estimated orientation $^{\mathcal{B}_1}_{\mathcal{E}}\hat{\mathbf{q}}(t)$ of the body $\mathcal{B}_1$ differs from the true orientation $^{\mathcal{B}_1}_{\mathcal{E}}\mathbf{q}(t)$ by being estimated in a different, unknown reference frame $\mathcal{E}_1 \neq \mathcal{E}$. So $^{\mathcal{B}_1}_{\mathcal{E}}\hat{\mathbf{q}}(t) = {}^{\mathcal{B}_1}_{\mathcal{E}_1}\mathbf{q}(t)$




**(a)** Homogenous magnetic field

**(b)** Disturbed magnetic field

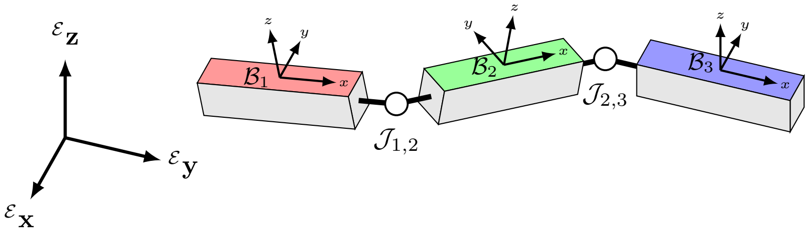
The true orientation $^{\mathcal{B}_1}_{\mathcal{E}}\mathbf{q}$ as well as the wrong orientation $^{\mathcal{B}_1}_{\mathcal{E}_1}\mathbf{q}$ caused by an inhomogenous magnetic field in proximity to a ferromagnetic object

- using the algorithm explained in Sec. 5.4.1 we can ensure that the lack of magnetometer correction only affects the heading of the orientation and does not affect the inclination. The difference between the frames $\mathcal{E}$ and $\mathcal{E}_1$ can therefore be described by a rotation around the global $z$-axis
- the difference quaternion $^{\mathcal{E}_1}_{\mathcal{E}}\mathbf{q}$ is not constant, since uncorrected integration of gyroscope bias around the global vertical axis will lead to a time-variant difference $^{\mathcal{E}_1}_{\mathcal{E}}\mathbf{q}(t)$ between the reference frames

## 6.3 Orientations of connected multi-body systems

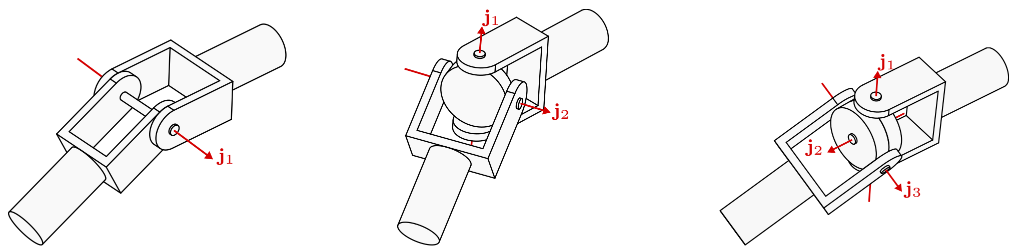
- one approach to correct for unknown heading components with 6D sensor fusion is using the relationships between bodies in connected multi-body systems

- Consider a system of $N$ rigid bodies connected by $N-1$ joints. Rigid bodies are denoted $\mathcal{B}_{\mathrm{i}}$, $i \in [1, N]$. The joint connecting the bodies $\mathcal{B}_{\mathrm{i}}$ and $\mathcal{B}_{\mathrm{i+1}}$ is denoted as $\mathcal{J}_{\mathrm{i,i+1}}$. Joints only connect two adjacent bodies to form a kinematic pair




Model of a kinematic chain with the segments $\mathcal{B}_1$, $\mathcal{B}_2$ and $\mathcal{B}_3$ and their respective coordinate systems as well as the joints $\mathcal{J}_{1,2}$ and $\mathcal{J}_{2,3}$.

- joints allow for the relative motion of two adjacent segments. We only consider rotational joints. We also use a simple joint model to describe this rotation by rotations around joint axes $\mathbf{j}_{\mathrm{p}}$. The number of distinct joint axes is called degrees of freedom of a joint, short $\mathrm{dof}(\,\mathcal{J}\,)$




**(a)** $\mathcal{J}_1$ **(b)** $\mathcal{J}_2$ **(c)** $\mathcal{J}_3$

Joints $\mathcal{J}_1$, $\mathcal{J}_2$ and $\mathcal{J}_3$ and their joint axes with $\mathrm{dof}(\,\mathcal{J}_1\,) = 1$, $\mathrm{dof}(\,\mathcal{J}_2\,) = 2$ and $\mathrm{dof}(\,\mathcal{J}_3\,) = 3$.

- assume the orientations of two adjacent segments $\mathcal{B}_{\mathrm{i}}$ and $\mathcal{B}_{\mathrm{i+1}}$ being estimated in a common reference frame $\mathcal{E}$. Their relative orientation can then be determined by

$$
{}^{\mathcal{B}_{\mathrm{i+1}}}_{\mathcal{B}_{\mathrm{i}}}\mathbf{q} = {}^{\mathcal{B}_{\mathrm{i}}}_{\mathcal{E}}\mathbf{q}^{-1} \otimes {}^{\mathcal{B}_{\mathrm{i+1}}}_{\mathcal{E}}\mathbf{q}. \tag{44}
$$

- We model the rotation ${}^{\mathcal{B}_{\mathrm{i+1}}}_{\mathcal{B}_{\mathrm{i}}}\mathbf{q}$ from $\mathcal{B}_{\mathrm{i}}$ to $\mathcal{B}_{\mathrm{i+1}}$ as consecutive rotations around the joint axes $\mathbf{j}_{\mathrm{p}} \in \mathbb{R}^3$, $\|\mathbf{j}_{\mathrm{p}}\| = 1$, by the *joint angles* $\varphi_{\mathrm{p}} \in \mathbb{R}$, $\mathrm{p} \in [1 \ldots \mathrm{dof}(\,\mathcal{J}_{\mathrm{i,i+1}}\,)]$ of the corresponding joint $\mathcal{J}_{\mathrm{i,i+1}}$

- the joint angles can be restricted to a range of values due to range of motion constraints

$$
\varphi_{\mathrm{p}} \in \{\varphi \in \mathbb{R} | \varphi_{\mathrm{p,min}} \leq \varphi \leq \varphi_{\mathrm{p,max}}\} \tag{45}
$$

- The relative orientation is then modeled as consecutive rotations around the joint axes

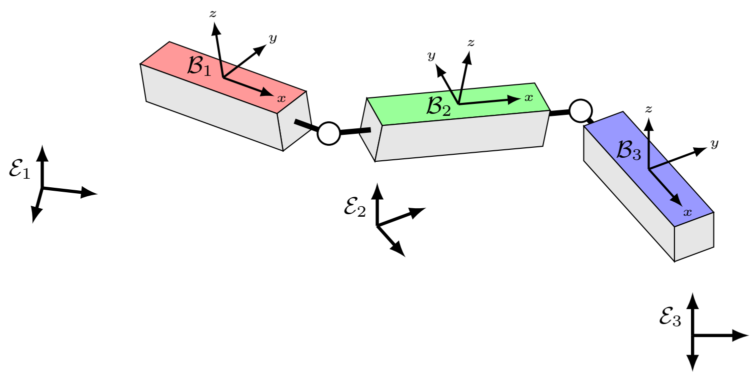
$$ {}^{\mathcal{B}_{\mathrm{i+1}}}_{\mathcal{B}_{\mathrm{i}}}\mathbf{q} = \prod_{p=1}^{\mathrm{dof}(\,\mathcal{J}_{\mathrm{i,i+1}}\,)} (\varphi_{\mathrm{p}} \,@\, \mathbf{j}_{\mathrm{p}}) \tag{46} $$

- due to restrictions in the degrees of freedom of a joint or restricted range of motions, the set P describing all possible relative orientations ${}^{\mathcal{B}_{\mathrm{i+1}}}_{\mathcal{B}_{\mathrm{i}}}\mathbf{q}$ between the two bodies is only a subset of all possible orientations $\mathbb{H}$. Using the generic joint model we can give an approximation of that set by

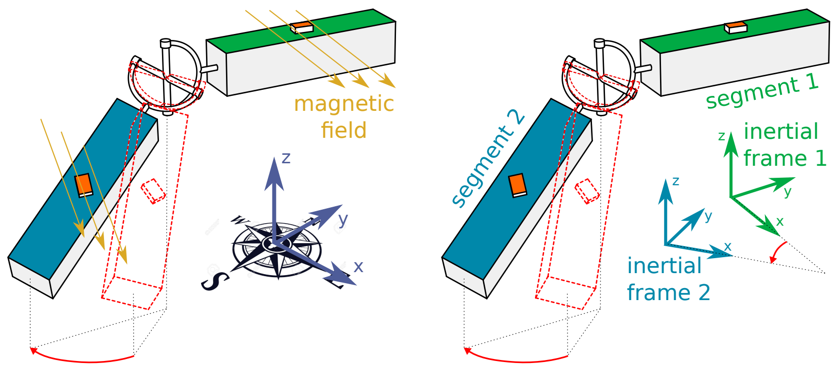
$$ \mathrm{P} = \left\{ \mathbf{q} \in \mathbb{H} \,\middle|\, \mathbf{q} = \prod_{\mathrm{p=1}}^{\mathrm{dof}(\,\mathcal{J}_{\mathrm{i,i+1}}\,)} (\varphi_{\mathrm{p}} \,@\, \mathbf{j}_{\mathrm{p}})\,, \varphi_{\mathrm{p}} \in [\varphi_{\mathrm{p,min}}, \varphi_{\mathrm{p,max}}] \right\} \tag{47} $$

- properties that hold for all orientations described by this set are called orientation-based kinematic constraints
- other kinematic constraints can be formulated by considering the positional relationships of bodies [3, 4] or properties of the underlying raw measurements [5]

## 6.4 Magnetometer-free orientation estimation in kinematic chains

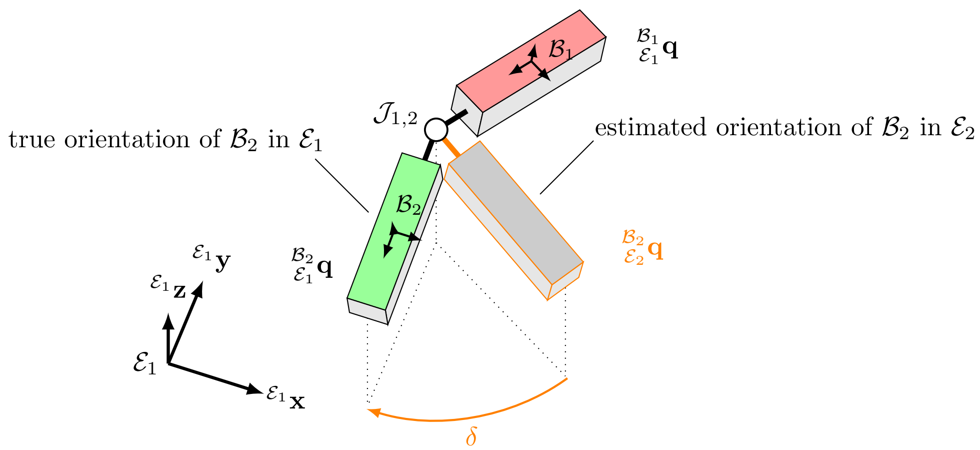



- Consider the kinematic chain with multiple segments $\mathcal{B}_{\mathrm{i}}$ and multiple sensors $\mathcal{S}_{\mathrm{i}}$ and assume the sensor-to-segment orientations ${}^{\mathcal{B}_{\mathrm{i}}}_{\mathcal{S}_{\mathrm{i}}}\mathbf{q}$ to be known.
- we use 6D sensor fusion to estimate the orientation ${}^{\mathcal{B}_{\mathrm{i}}}_{\mathcal{E}}\mathbf{q}$ of each segment with respect the inertial frame $\mathcal{E}$. However, due to the previously explained reasons, the estimated orientation has an incorrect heading component and this is modeled as if the estimated orientation ${}^{\mathcal{B}_{\mathrm{i}}}_{\mathcal{E}_{\mathrm{i}}}\mathbf{q}$ of $\mathcal{B}_{\mathrm{i}}$ is estimated in the reference frame $\mathcal{E}_{\mathrm{i}}$, which differs from $\mathcal{E}$ by a rotation around the global vertical axis




- Consider the two segments $\mathcal{B}_1$ and $\mathcal{B}_2$. Since at each moment in time the difference between $\mathcal{E}_1$ and $\mathcal{E}$ and between $\mathcal{E}_2$ and $\mathcal{E}$ is only a non-constant rotation around the global vertical axis, the difference between $\mathcal{E}_1$ and $\mathcal{E}_2$ is also a time-variant rotation around the global vertical axis, described by the quaternion ${}^{\mathcal{E}_2}_{\mathcal{E}_1}\mathbf{q}(t)$.
- the angle describing this quaternion as a rotation around $[0, 0, 1]^\intercal$ is called *heading offset* and is denoted by $\delta$

$$
{}^{\mathcal{E}_2}_{\mathcal{E}_1}\mathbf{q} = \left(\delta \,@\, {}^{\mathcal{E}}_{\mathcal{E}}\mathbf{z}\right) = \begin{bmatrix}\cos\left(\frac{\delta}{2}\right) & 0 & 0 & \sin\left(\frac{\delta}{2}\right)\end{bmatrix}^T \tag{48}
$$




Orientations of the segments $\mathcal{B}_1$ and $\mathcal{B}_2$ in the reference frame $\mathcal{E}_1$. The angle $\delta$ describes the difference between the reference frames $\mathcal{E}_1$ and $\mathcal{E}_2$.

- Since the relative orientation of the two reference frames is not constant, the angle $\delta$ is not constant either and is drifting, i.e. $\delta(t)$
- The initial heading offset $\delta(t = 0)$ is the constant offset due to the lack of absolute heading information. Since the change in $\delta(t)$ is caused by integration of the bias in the gyroscope readings, the time derivative can be approximated constant

$$\frac{\partial \delta}{\partial t} \approx \text{const.} \tag{49}$$

- The heading offset $\delta(t)$ can therefore be approximated linearly by

$$\delta(t) \approx \underbrace{\frac{\partial \delta}{\partial t} \cdot t}_{\text{heading drift}} + \underbrace{\delta(t=0)}_{\text{constant offset in heading}} \tag{50}$$

- This approximation only holds true if the gyro bias is similar for all three gyroscope axes or for sufficiently small time windows. For a small bias, this approximation is valid and has been proven adequate

- The heading offset $\delta$ only describes the difference between the segments' reference frames, but not the difference between the reference frames $\mathcal{E}_\mathrm{i}$ and the global reference frame $\mathcal{E}$

- why is this angle chosen and not the difference angles between $\mathcal{E}_\mathrm{i}$ and the global reference frame $\mathcal{E}$?

- with 6D sensor fusion, we estimate the orientations ${}^{\mathcal{B}_\mathrm{i}}_{\mathcal{E}_\mathrm{i}}\mathbf{q}$ of the segments. As shown in (44) the relative orientation of two adjacent segments ${}^{\mathcal{B}_\mathrm{i+1}}_{\mathcal{B}_\mathrm{i}}\mathbf{q}$ can only be determined if the orientations of the segments can be expressed in a common reference frame. This is possible by knowing the heading offset $\delta_{i+1,i}(t)$ at each moment in time

$${}^{\mathcal{B}_\mathrm{i+1}}_{\mathcal{B}_\mathrm{i}}\mathbf{q} = {}^{\mathcal{B}_\mathrm{i}}_{\mathcal{E}_\mathrm{i}}\mathbf{q}^{-1} \otimes {}^{\mathcal{E}_\mathrm{i+1}}_{\mathcal{E}_\mathrm{i}}\mathbf{q} \otimes {}^{\mathcal{B}_\mathrm{i+1}}_{\mathcal{E}_\mathrm{i+1}}\mathbf{q} = {}^{\mathcal{B}_\mathrm{i}}_{\mathcal{E}_\mathrm{i}}\mathbf{q}^{-1} \otimes \left(\delta_{i+1,i} \,@\, {}^{\mathcal{E}}_{\mathcal{E}}\mathbf{z}\right) \otimes {}^{\mathcal{B}_\mathrm{i+1}}_{\mathcal{E}_\mathrm{i+1}}\mathbf{q} \tag{51}$$

- note that the latter part describes the orientation of one segment in the reference frame of the other:

$${}^{\mathcal{E}_\mathrm{i+1}}_{\mathcal{E}_\mathrm{i}}\mathbf{q} \otimes {}^{\mathcal{B}_\mathrm{i+1}}_{\mathcal{E}_\mathrm{i+1}}\mathbf{q} = {}^{\mathcal{B}_\mathrm{i+1}}_{\mathcal{E}_\mathrm{i}}\mathbf{q} \tag{52}$$

- this formulation of the problem is reducing the lack of heading correction to a problem of knowing one scalar value $\delta(t)$ at each moment in time for each kinematic pair in the chain

- formulating the problem by using the relative orientation connects to the model of a kinematic chain and the set of limited possible relative orientations and the derived orientation-based kinematic constraints

- **Goal**: find a formulation of a kinematic constraint for a kinematic pair that incorporates $\delta$ and formulate it in a way that

$$f({}^{\mathcal{B}_\mathrm{i}}_{\mathcal{E}_\mathrm{i}}\mathbf{q}, {}^{\mathcal{B}_\mathrm{i+1}}_{\mathcal{E}_\mathrm{i+1}}\mathbf{q}, \delta, t) \stackrel{!}{=} 0 \;\forall t \tag{53}$$

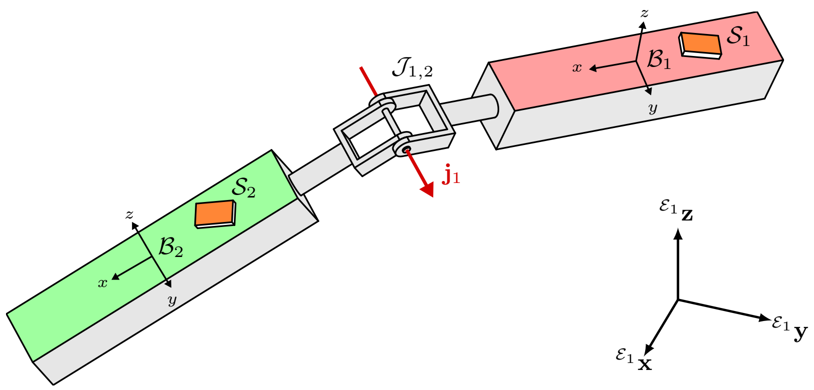
- this way we can either find an analytical expression for $\delta(t)$ or use an optimization-based approach to find an estimate $\hat{\delta}(t)$ that fulfills the constraint the best

## 6.5 Kinematic constraints for different joint configurations

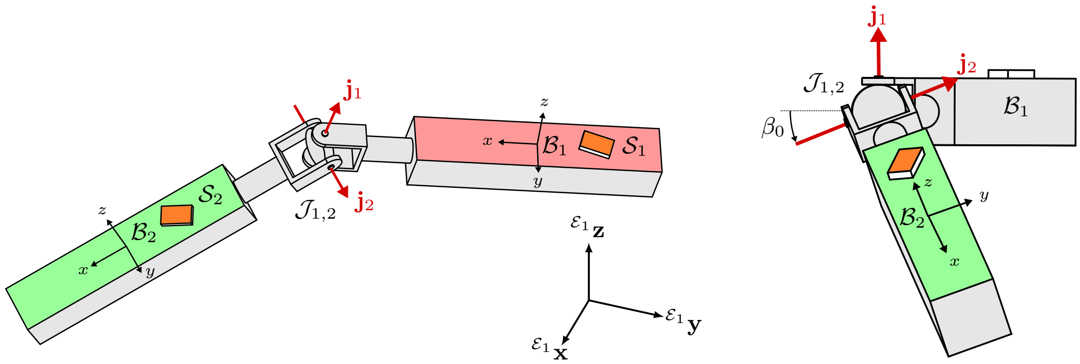
### 6.5.1 One-dimensional joints

### 6.5.2 Two-dimensional joints

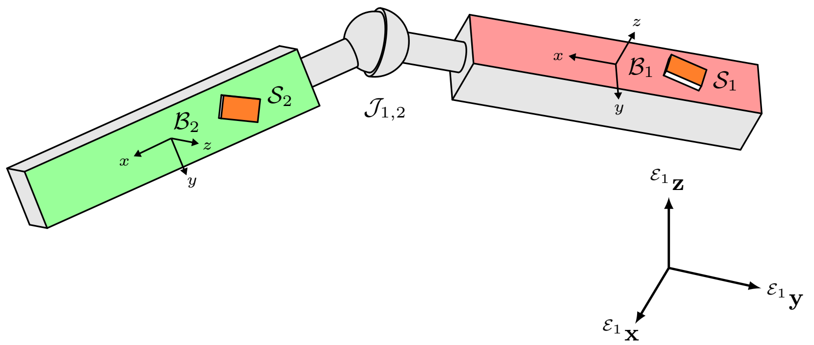
### 6.5.3 Arbitrary joints with limited range of motion




Kinematic model for a one-dimensional joint. The two segments $\mathcal{B}_1$ and $\mathcal{B}_2$ are connected by the one-dimensional joint $\mathcal{J}_{1,2}$ with the joint axis $\mathbf{j}_1$.




Kinematic model of a two-dimensional joint with the segments $\mathcal{B}_1$ and $\mathcal{B}_2$ connected by the two-dimensional joint $\mathcal{J}_{1,2}$ with the joint axes $\mathbf{j}_1$ and $\mathbf{j}_2$.




Kinematic model for a three-dimensional joint with the segments $\mathcal{B}_1$ and $\mathcal{B}_2$ connected by the three-dimensional joint $\mathcal{J}_{1,2}$.

# 7 Information Fusion Methods

## 7.1 Random Variables

- Let $\mathbf{x} \in \mathbb{R}^n$ be a random variable and $\mathrm{E}(\mathbf{x}) \in \mathbb{R}^n$ its expected value
- the *covariance matrix* $\mathbf{P} \in \mathbb{R}^{n \times n}$ is defined as

$$\mathbf{P} = \mathrm{E}\left((\mathbf{x} - \mathrm{E}(\mathbf{x}))(\mathbf{x} - \mathrm{E}(\mathbf{x}))^\intercal\right) \tag{54}$$

- the diagonal elements of $\mathbf{P}$ are the variances of the entries of $\mathbf{x}$
- $\mathbf{P}$ is diagonal if and only if the individual elements of $\mathbf{x}$ are uncorrelated
- $\mathbf{P}$ is positive semidefinite (positive definite if the variance of each entry of $\mathbf{x}$ is non-zero)

### 7.1.1 Gaussian Random Variables

- Let $\mathbf{x} \in \mathbb{R}^n$ be a Gaussian random variable characterized by a mean $\bar{\mathbf{x}}$ and covariance $\mathbf{P}$
- i.e. in short notation: $\mathbf{x} \sim \mathcal{N}(\bar{\mathbf{x}}, \mathbf{P})$ ($\mathcal{N}$ as in "normal" distribution)
- the *probability density function* (pdf) is denoted $p(\mathbf{x})$ and is then given by

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}} \frac{1}{\sqrt{\det \mathbf{P}}} \exp\left(\frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}})^\intercal \mathbf{P}^{-1}(\mathbf{x} - \bar{\mathbf{x}})\right) \tag{55}$$

- note that, in many cases, the distribution of the (normalized) sum of $m$ independent random variables tends toward such a normal distribution for $n \to \infty$ (for further details, please study the Central Limit Theorem)

Gaussian random variables exhibit a few remarkable and useful mathematical properties:

- Let $\mathbf{x}_1 \in \mathbb{R}^n \sim \mathcal{N}(\bar{\mathbf{x}}_1, \mathbf{P}_1)$ and $\mathbf{x}_2 \in \mathbb{R}^n \sim \mathcal{N}(\bar{\mathbf{x}}_2, \mathbf{P}_2)$ be two *independent* random variables. The addition of $\mathbf{x}_1$ and $\mathbf{x}_2$ yields

$$\mathbf{x}_1 + \mathbf{x}_2 \sim \mathcal{N}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2, \mathbf{P}_1 + \mathbf{P}_2) \tag{56}$$

- if $\mathbf{x} \in \mathbb{R}^n$ is a random variable with $\mathbf{x} \sim \mathcal{N}(\bar{\mathbf{x}}, \mathbf{P})$, then multiplication of $\mathbf{x}$ by a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ yields a random variable with distribution

$$\mathbf{A}\,\mathbf{x} \sim \mathcal{N}(\mathbf{A}\bar{\mathbf{x}}, \mathbf{APA}^\intercal) \tag{57}$$

**Examples for random variables and covariance matrices:**

$\mathbf{x}_k \in \mathbb{R}^2 \quad \mathbf{x}_k \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 5\end{bmatrix}\right) \quad \forall k$

| k | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $\mathbf{x}_k[1]$ | 0.7 | -0.2 | -0.1 | 1.5 | 1.4 | 1.3 |
| $\mathbf{x}_k[2]$ | -2.7 | 1.6 | 3.6 | 1.1 | 2.3 | 1.6 |

$\mathbf{x}_k \in \mathbb{R}^2 \quad \mathbf{x}_k \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}5 & -4\\-4 & 5\end{bmatrix}\right) \quad \forall k$

| k | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $\mathbf{x}_k[1]$ | 0.7 | -1.8 | 2.0 | -2.6 | -2.5 | -1.8 |
| $\mathbf{x}_k[2]$ | 1.4 | 1.8 | -2.6 | 3.9 | -0.4 | 1.3 |

**Example for $\mathbf{v} + \mathbf{w} \sim \mathcal{N}(\bar{v} + \bar{w}, \mathbf{V} + \mathbf{W})$**

$\mathbf{v}_k \in \mathbb{R}^2 \quad \mathbf{v}_k \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 5\end{bmatrix}\right) \quad \forall k$ $\qquad \mathbf{w}_k \in \mathbb{R}^2 \quad \mathbf{w}_k \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}5 & -4\\-4 & 5\end{bmatrix}\right) \forall k$

| k | 1 | 2 | 3 | 4 | 5 | 6 | $P_i$ |
|---|---|---|---|---|---|---|---|
| $\mathbf{v}_k[1]$ | 0.7 | -0.2 | -0.1 | 1.5 | 1.4 | 1.3 | 1 |
| $\mathbf{v}_k[2]$ | -2.7 | 1.6 | 3.6 | 1.1 | 2.3 | 1.6 | 5 |
| $\mathbf{w}_k[1]$ | 0.7 | -1.8 | 2.0 | -2.6 | -2.5 | -1.8 | |
| $\mathbf{w}_k[2]$ | 1.4 | 1.8 | -2.6 | 3.9 | -0.4 | 1.3 | |
| $\mathbf{v}_k[1] + \mathbf{w}_k[1]$ | -2 | 1.4 | 3.5 | 2.6 | 3.7 | 2.9 | 6 |

**Example for $\mathbf{A}\,\mathbf{x}_k \sim \mathcal{N}(\mathbf{A}\bar{\mathbf{x}}, \mathbf{A}\mathbf{P}\mathbf{A}^\intercal)\ \forall k$ with $\mathbf{A} = \begin{bmatrix}2 & 0\\0 & 2\end{bmatrix}$**

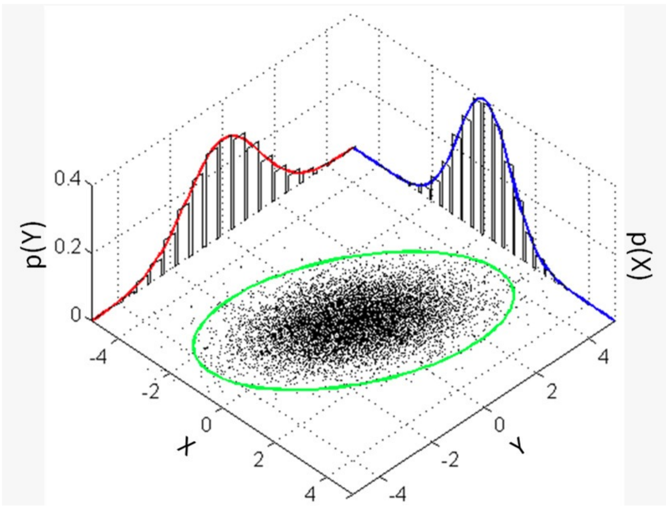
$\mathbf{A}\,\mathbf{x}_k \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}4 & 0\\0 & 20\end{bmatrix}\right)$

| k | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $\mathbf{x}_k[1]$ | 1.4 | -0.4 | -0.2 | 3.0 | 2.8 | 2.6 |
| $\mathbf{x}_k[2]$ | -5.4 | 3.2 | 7.2 | 2.2 | 4.8 | 3.2 |

$\mathbf{A}\,\mathbf{x}_k \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}20 & -16\\-16 & 20\end{bmatrix}\right)$

| k | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $\mathbf{x}_k[1]$ | 1.3 | -3.5 | 4.0 | -5.1 | -4.8 | -3.6 |
| $\mathbf{x}_k[2]$ | 2.8 | 3.7 | -5.2 | 7.8 | -0.8 | 2.6 |

## 7.2 Multivariate probability distributions



### 7.3 Simple fusion of Gaussian random variables

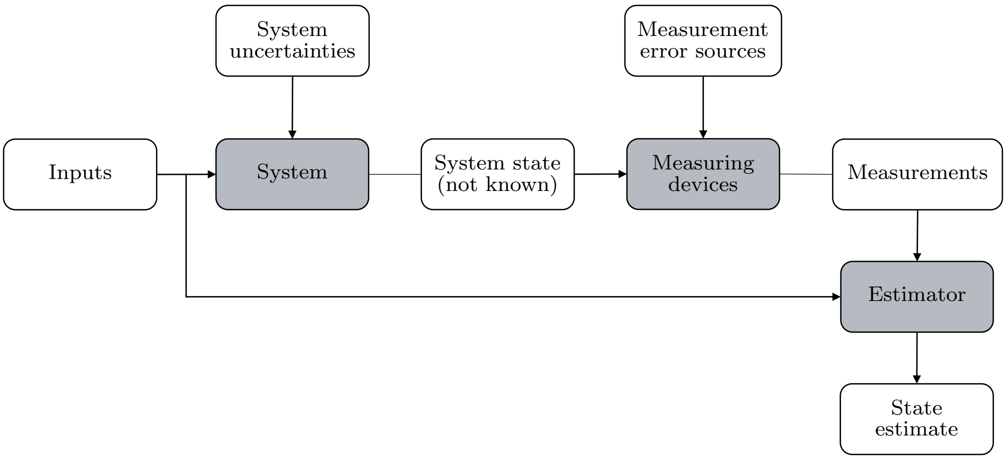
| | 1 | 2 | 3 | ... | 1000 | $\bar{x}$ | $P$ |
|---|---|---|---|---|---|---|---|
| $x_1$ | 26 | 36 | 22 | | 27 | 35 | $10^2$ |
| $x_2$ | 13 | 40 | 21 | | 43 | 25 | $20^2$ |
| $0.5x_1 + 0.5x_2$ | 19.5 | 38 | 21.5 | | 35 | 30 | $11.2^2$ |
| $0.7x_1 + 0.3x_2$ | 22.1 | 37.2 | 21.7 | | 31.8 | 32 | $9.2^2$ |
| $0.8x_1 + 0.2x_2$ | 23.4 | 36.8 | 21.8 | | 30.2 | 33 | $8.9^2$ |

$$\bar{x} = a\bar{x}_1 + (1-a)\bar{x}_2 \tag{58}$$

$$P = a^2 P_1 + (1-a)^2 P_2 \tag{59}$$

## 7.4 Estimation problem

An abstraction of a general estimation problem can be illustrated as follows:




- the "System state" is some inner variable of a system that shall be estimated by the "Estimator"
- "Inputs" are known signal, which can often (but not always) be manipulated, while the "Measuements" are direcly measurable outputs
- "System" and "Measuring devices" represent (dynamic) functions or mechanism for which some sort of model exists
- "System uncertainties" and "Measurement error sources" are unknown input (disturbance) signals or representations of model uncertainties; they are not known to the "Estimator"

### 7.4.1 Nonlinear system

- For a large number of applications, the "System" and "Measuring devices" can be modeled by a nonlinear discrete-time state space model

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}, \mathbf{u}_k) + \mathbf{v}_k \tag{60}$$
$$\mathbf{y}_k = h(\mathbf{x}_k) + \mathbf{w}_k \tag{61}$$

- with $\mathbf{x} \in \mathbb{R}^n$ being the *states* of the system, $\mathbf{u} \in \mathbb{R}^p$ the *inputs* to the system and $\mathbf{y} \in \mathbb{R}^q$ the *measurements*
- $f(\cdot)$ describes the dynamics of the system, $h(\cdot)$ is the measurement equation
- the variables $\mathbf{v}_k \in \mathbb{R}^n$ and $\mathbf{w}_k \in \mathbb{R}^q$ are, for each sampling instant $k$, some *unknown* random variables
- common simplification/assumptions are
  a for any two sampling instants $k_1$ and $k_2 \neq k_1$, the random variables $\mathbf{v}_{k_1}$ and $\mathbf{v}_{k_2}$ are independent (and thus uncorrelated) and have the same pdf

b for any two sampling instants $k_1$ and $k_2 \neq k_1$, the random variables $\mathbf{w}_{k_1}$ and $\mathbf{w}_{k_2}$ are independent (and thus uncorrelated) and have the same pdf

c these identical distributions are Gaussian distributions

d for any two sampling instants $k_1$ and $k_2$, the random variables $\mathbf{v}_{k_1}$ and $\mathbf{w}_{k_2}$ are independent (and thus uncorrelated)

- (a)/(b) is often written "the samples of $\mathbf{v}_k/\mathbf{w}_k$ are independent and identically distributed" (i.i.d.), which implies that $\mathbf{v}_k$ and $\mathbf{w}_k$ are white-noise sequences
- (a+c)/(b+c) is often written "$\mathbf{v}_k$ / $\mathbf{w}_k$ is a Gaussian white-noise sequence"
- (a+b+c+d) is often written "$\mathbf{v}_k$ and $\mathbf{w}_k$ are independent Gaussian white-noise sequences"

### 7.4.2 Linear system

- a special case is the linear discrete-time system

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{B}\mathbf{u}_k + \mathbf{v}_k \tag{62}$$
$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{w}_k \tag{63}$$

- with $\mathbf{x} \in \mathbb{R}^n$ being the *states* of the system, $\mathbf{u} \in \mathbb{R}^p$ the *inputs* to the system and $\mathbf{y} \in \mathbb{R}^q$ the *measurements* and $\mathbf{v}_k$ and $\mathbf{w}_k$ as described above. $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are matrices with $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{B} \in \mathbb{R}^{n \times p}$ and $\mathbf{C} \in \mathbb{R}^{q \times n}$
- the estimate of the state $\mathbf{x}_k$ is denoted by $\hat{\mathbf{x}}_k$ and the covariance matrix of $\hat{\mathbf{x}}_k$ by $\mathbf{P}_k$
- the expected value of the initial state $\mathrm{E}(\mathbf{x}_0)$ and its covariance $\mathbf{P}_0$ are known
- the input $\mathbf{u}_k$ and the output $\mathbf{y}_k$ of the system are known
- $\mathbf{v}_k$ and $\mathbf{w}_k$ are independent Gaussian white-noise sequences with $\mathbf{v}_k \sim N(0, \mathbf{V})$ $\forall k$ and $\mathbf{w}_k \sim N(0, \mathbf{W})$ $\forall k$

**Question:** Assume you have $\hat{\mathbf{x}}_0 = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$ with $\mathbf{P}_0 = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}^\mathsf{T}$ for the dynamics

$$\mathbf{x}_k = \mathbf{x}_{k-1} + \mathbf{u}_k + \mathbf{v}_k, \qquad \mathbf{y}_k = \mathbf{x}_k + \mathbf{w}_k, \tag{64}$$

with $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, but the measurement $\mathbf{y}_1 = \begin{bmatrix} 10 \\ 10 \end{bmatrix}$. What is a good estimate for $\mathbf{x}_1$ if $\mathbf{V} \gg \mathbf{W}$? What is a good estimate for $\mathbf{x}_1$ if $\mathbf{V} \ll \mathbf{W}$? What is a good estimate for $\mathbf{x}_1$ if $\mathbf{V} \approx \mathbf{W}$?

**Answer:** Employ a measurement-based correction of a model-based prediction:

$$\hat{\mathbf{x}}_1 = \underbrace{\begin{bmatrix} 6 \\ 6 \end{bmatrix}}_{\text{pred.}} + \mathbf{K}_1 \underbrace{\begin{bmatrix} 10-6 \\ 10-6 \end{bmatrix}}_{\text{disagreement}} = (\mathbf{I} - \mathbf{K}_1) \underbrace{\begin{bmatrix} 6 \\ 6 \end{bmatrix}}_{\text{pred.}} + \mathbf{K}_1 \underbrace{\begin{bmatrix} 10 \\ 10 \end{bmatrix}}_{\text{meas.}},$$

where $\mathbf{K}_1$ is chosen (almost) zero if $\mathbf{V} \ll \mathbf{W}$ and (almost) one if $\mathbf{V} \gg \mathbf{W}$. Note that the covariance of the prediction is $(\mathbf{P}_0 + \mathbf{V})$, while the covariance of the measurement is $\mathbf{W}$. Therefore, a good balance is achieved by choosing $\mathbf{K}_1 = (\mathbf{P}_0 + \mathbf{V})(\mathbf{W} + (\mathbf{P}_0 + \mathbf{V}))^{-1}$

- this idea can be generalized and used at each timestep recursively, i.e. it can be implemented as an *estimation filter* that solves the following problem for $k = 1, 2, \dots$
  - given $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{V}, \mathbf{W}, \hat{\mathbf{x}}_0, \mathbf{P}_0$ as well as $\mathbf{u}_k, \mathbf{y}_k$
  - find best estimate $\hat{\mathbf{x}}_k$ of the state $\mathbf{x}_k$
- the generalized idea is then to determine $\hat{\mathbf{x}}_k$ as a weighted combination of an uncertain model-based prediction $\hat{\mathbf{x}}_k^-$ and a measurement-based correction:

$$\hat{\mathbf{x}}_k = \underbrace{\hat{\mathbf{x}}_k^-}_{\mathbf{A}\hat{\mathbf{x}}_{k-1}+\mathbf{B}\mathbf{u}_k} + \mathbf{K}_k \overbrace{(\mathbf{y}_k - \mathbf{C}\,\hat{\mathbf{x}}_k^-)}^{\text{disagreement}} = (\mathbf{I} - \mathbf{K}_k\mathbf{C})\hat{\mathbf{x}}_k^- + \mathbf{K}_k\mathbf{y}_k \tag{65}$$

- $\mathbf{K}_k$ is the gain of the fusion of prediction and measurement
- ideally, it should be chosen such that it minimizes the expected error $\mathrm{E}\left[(\mathbf{x}_k - \hat{\mathbf{x}}_k)(\mathbf{x}_k - \hat{\mathbf{x}}_k)^\intercal\right]$

## 7.5 Kalman Filter

- assume a given prior estimate $\hat{\mathbf{x}}_{k-1}$ with $\mathrm{E}(\hat{\mathbf{x}}_{k-1}) = \mathrm{E}(\mathbf{x}_{k-1})$ as well as its uncertainty $\mathbf{P}_{k-1}$
- determine the predicted a-priori estimate $\hat{\mathbf{x}}_k^-$ of the state

$$\hat{\mathbf{x}}_k^- = \mathbf{A}\hat{\mathbf{x}}_{k-1} + \mathbf{B}\mathbf{u}_k \tag{66}$$

- then determine the a-priori uncertainty of the prediction $\mathbf{P}_k^-$ using (56) and (57)

$$\mathbf{P}_k^- = \mathbf{A}\mathbf{P}_{k-1}\mathbf{A}^\intercal + \mathbf{V} \tag{67}$$

- the optimal filter gain $\mathbf{K}_k$ is then determined with

$$\mathbf{K}_k = \mathbf{P}_k^-\mathbf{C}^\intercal(\overbrace{\mathbf{W} + \mathbf{C}\mathbf{P}_k^-\mathbf{C}^\intercal}^{\text{cov. of disagreement}})^{-1} \tag{68}$$

- the current estimate $\hat{\mathbf{x}}_k$ is then determined as

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{y}_k - \mathbf{C}\hat{\mathbf{x}}_k^-) \tag{69}$$

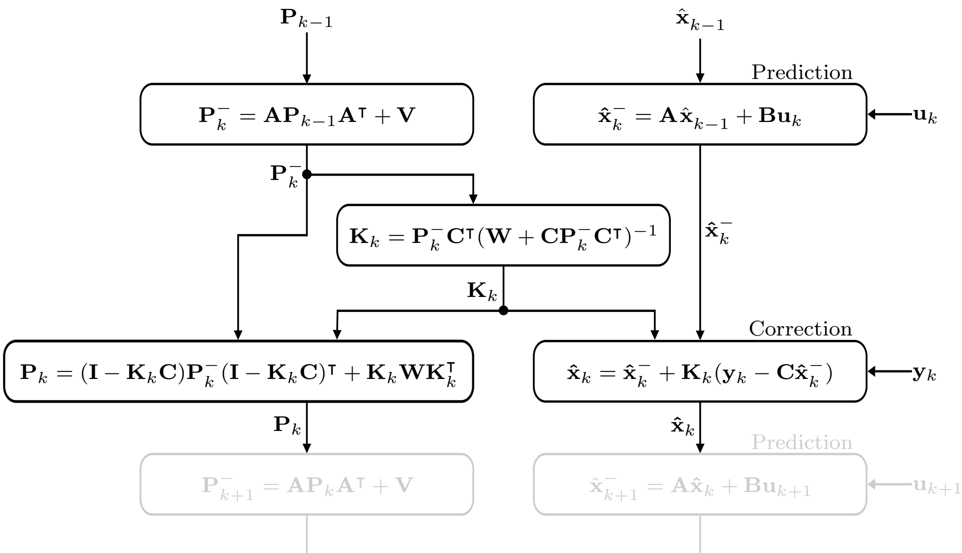
- the uncertainty $\mathbf{P}_k$ of the current estimate is then given by

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k\mathbf{C})\mathbf{P}_k^-(\mathbf{I} - \mathbf{K}_k\mathbf{C})^\intercal + \mathbf{K}_k\mathbf{W}\mathbf{K}_k^\intercal = (\mathbf{I} - \mathbf{K}_k\mathbf{C})\mathbf{P}_k^- \tag{70}$$

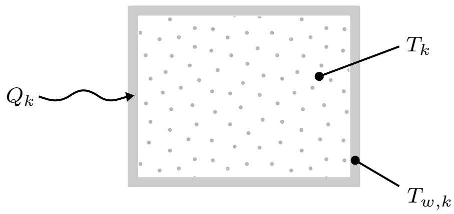
- the procedure is then performed repeatedly for each time step
- note that the dynamics of $\hat{\mathbf{x}}_k$ and $\mathbf{P}_k$ are governed by $(\mathbf{I} - \mathbf{K}_k\mathbf{C})$, which implies that
  - if $\mathbf{K}_k = 0$, i.e. if $\mathbf{W} \gg \mathbf{P}_{k,\text{pre}}$, then $\hat{\mathbf{x}}_k \approx \hat{\mathbf{x}}_k^-$ (we rely only on the prediction), and $\mathbf{P}_k \approx \mathbf{P}_{k,\text{pre}} = \mathbf{A}\mathbf{P}_{k-1}\mathbf{A}^\intercal + \mathbf{V}$ (the uncertainty grows by integration of $\mathbf{V}$)
  - if $\mathbf{W} \ll \mathbf{P}_{k,\text{pre}}$ and $\exists \mathbf{C}^{-1}$, then $(\mathbf{I} - \mathbf{K}_k\mathbf{C}) = 0$, which implies $\hat{\mathbf{x}}_k \approx \mathbf{C}^{-1}\mathbf{y}_k$ (invert measurem.), and the uncertainty $\mathbf{P}_k$ jumps to zero
- note furthermore that the values $\mathbf{P}_k, \mathbf{K}_k, k = 1, 2, \dots$, do not depend on the inputs or the measured values, i.e. they can be calculated (from $\mathbf{V}, \mathbf{W}$ and $\mathbf{P}_0$) before the actual trial

**Algorithm Overview:**

Given: $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{W}, \mathbf{V}, \mathbf{P}_0, \hat{\mathbf{x}}_0, \mathbf{u}_k, \mathbf{y}_k$




**Application example:** Consider a heating chamber with an internal temperature $T_k$. The input to the system is the heat $Q_k$ with the proportionality factor $a = 1.5$. The internal temperature cannot be measured, but the wall temperature $T_{w,k}$. The initial temperatures are $T_0 = 300\,\mathrm{K}$ and $T_{w,0} = 310\,\mathrm{K}$.




The temperatures are modeled as

$$T_k = T_{k-1} + aQ_k \tag{71}$$

$$T_{w,k} = 0.9\,T_{w,k-1} + 0.1\,T_{k-1} \tag{72}$$

The linear time-discrete system is modeled as

$$x_{1,k} = x_{1,k-1} + au_k + v_{1,k} \tag{73}$$

$$x_{2,k} = 0.9x_{2,k-1} + 0.1x_{1,k-1} + v_{2,k} \tag{74}$$

$$y_k = x_{2,k} + w_k \tag{75}$$

with

$$\mathbf{x}_k = \begin{bmatrix} T_k & T_{w,k} \end{bmatrix}^\intercal \quad u_k = Q_k \tag{76}$$

and $v_{1,k}, v_{2,k}$ describing the model uncertainties as well as $w_k$ describing the measurement uncertainty.

To use a Kalman Filter, the model is written in state-space notation

$$\mathbf{x}_{k+1} = \underbrace{\begin{bmatrix} 1 & 0 \\ 0.1 & 0.9 \end{bmatrix}}_{\mathbf{A}} \mathbf{x}_k + \underbrace{\begin{bmatrix} a \\ 0 \end{bmatrix}}_{\mathbf{B}} u_k + \mathbf{v}_k \tag{77}$$

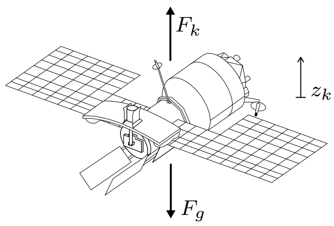
$$y_k = \underbrace{\begin{bmatrix} 0 & 1 \end{bmatrix}}_{\mathbf{C}} \mathbf{x}_k + w_k \tag{78}$$

The covariances are chosen as

$$\mathbf{V} = \begin{bmatrix} 10^2 & 0 \\ 0 & 10^2 \end{bmatrix} \quad W = 1 \quad \mathbf{P}_0 = \begin{bmatrix} 10^2 & 0 \\ 0 & 10^{-3} \end{bmatrix}. \tag{79}$$

Then the algorithm as described above can be used to find an estimate $\hat{T}_k$ for the internal temperature.

**Application example:** Consider a satellite in space. The satellite is far away from earth and only experiences an uncertain gravitational acceleration of $g \approx 0.5\,\frac{\mathrm{m}}{\mathrm{s}^2}$. The mass of the satellite is known with $m = 2.6\,\mathrm{t}$. The position $z_k$ of the satellite is measurable with a rate of $f = 1000\,\mathrm{Hz}$. The position at $t = 0$ is set to $z_0 = 0\,\mathrm{m}$. The propulsion force $F_k$ is an input to the system, but is only known approximately. At $t = 0\,\mathrm{s}$ the satellite is moving with $v_0 \approx 1\,\frac{\mathrm{m}}{\mathrm{s}}$. To estimate the satellite's velocity in z direction $v_k$ a Kalman filter is implemented.




The system dynamics are modeled as

$$\begin{aligned} z_k &= z_{k-1} + \frac{1}{f} v_{k-1} \\ v_k &= v_{k-1} + \frac{1}{f}\frac{1}{m}(F_k - F_g) \end{aligned} \tag{80}$$

The first idea for a state-space model is

$$\mathbf{x}_k = \begin{bmatrix} z_k & v_k \end{bmatrix}^\intercal \quad u_k = F_k \quad y_k = z_k \tag{81}$$

$$\mathbf{x}_k = \begin{bmatrix} 1 & \frac{1}{f} \\ 0 & 1 \end{bmatrix} \mathbf{x}_{k-1} + \begin{bmatrix} 0 \\ \frac{1}{fm} \end{bmatrix} u_k + \mathbf{v}_k \tag{82}$$

$$y_k = \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{x}_k + w_k \tag{83}$$

Note that $\mathbf{v}_k$ and $w_k$ denote the system and measurement noise. However, in this state-space model, the gravitational force $F_g$ does not appear.

Solution: Add $F_g$ as additional state:

$$\mathbf{x}_k = \begin{bmatrix} z_k & v_k & F_g \end{bmatrix}^\mathsf{T} \quad u_k = F_k \quad y_k = z_k \tag{84}$$

$$\mathbf{x}_k = \underbrace{\begin{bmatrix} 1 & \frac{1}{f} & 0 \\ 0 & 1 & -\frac{1}{fm} \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{A}} \mathbf{x}_{k-1} + \underbrace{\begin{bmatrix} 0 \\ \frac{1}{fm} \\ 0 \end{bmatrix}}_{\mathbf{B}} u_k + \mathbf{v}_k \tag{85}$$

$$y_k = \underbrace{\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}}_{\mathbf{C}} \mathbf{x}_k + w_k \tag{86}$$

As covariances choose

$$\mathbf{V} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 100 & 0 \\ 0 & 0 & 5 \end{bmatrix} \quad W = 1 \quad \mathbf{P}_0 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 100 & 0 \\ 0 & 0 & 100 \end{bmatrix} \text{ for } \mathbf{x}_0 = \begin{bmatrix} 0 \\ 1 \\ 1300 \end{bmatrix} \tag{87}$$

$\mathbf{V}[2,2]$ is chosen high because the input $u_k$ is not perfectly known as stated in the description. $\mathbf{V}[1,1]$ is chosen very low, since (80) is only the relationship between velocity and position. $\mathbf{V}[3,3]$ allows the Kalman Filter to estimate $F_g$ not perfectly constant but to allow for slight dynamics.

**Application example:** Consider the task of measuring the knee joint angle $\alpha$ of a leg orthosis, prosthesis or exoskeleton. Assume that two IMUs are attached to the thigh and the shank segment.




Assume that you can determine a very good estimate $\dot{\alpha}_{\text{gyr}}$ of the time derivative of the knee angle directly from the measured angular rates. Assume furthermore that you can get an approximate, very noisy but driftless estimate of the joint angle $\alpha_{\text{acc}}$ from the accelerometer readings of the IMUs. How does this sensor fusion problem fit the Kalman filter framework?

**Answer:** Define the states $x_1 = \dot{\alpha}$+bias, $x_2 = \alpha$, $x_3 = b$ (bias) and find the dynamics:

$$
\begin{aligned}
x_{1,k} &= x_{1,k-1} & &+ v_1 \quad \text{(random angular acceleration)},\\
x_{2,k} &= x_{2,k-1} + x_{1,k}T_s - x_{3,k} & &+ v_2 \quad \text{(integrate angular rate w/o bias)},\\
x_{3,k} &= x_{3,k-1} & &+ v_3 \quad \text{(almost constant bias)},\\
y_1 = \dot{\alpha}_{\text{gyr}} &= x_{1,k} & &+ w_1 \quad \text{(measure ang. rate with bias)},\\
y_2 = \alpha_{\text{acc}} &= x_{2,k} & &+ w_2 \quad \text{(measure angle with large noise)}.
\end{aligned}
$$

Alternatively, we can define states $x_1 = \alpha$, $x_2 = b$ (bias) and find the dynamics:

$$
\begin{aligned}
x_{1,k} &= x_{1,k-1} + \dot{\alpha}_{\text{gyr}}T_s - x_{2,k} & &+ v_1 \quad \text{(measurement as input)},\\
x_{2,k} &= x_{2,k-1} & &+ v_2 \quad \text{(almost constant bias)},\\
\alpha_{\text{acc}} &= x_{1,k} & &+ w \quad \text{(measure angle with large noise)},
\end{aligned}
$$

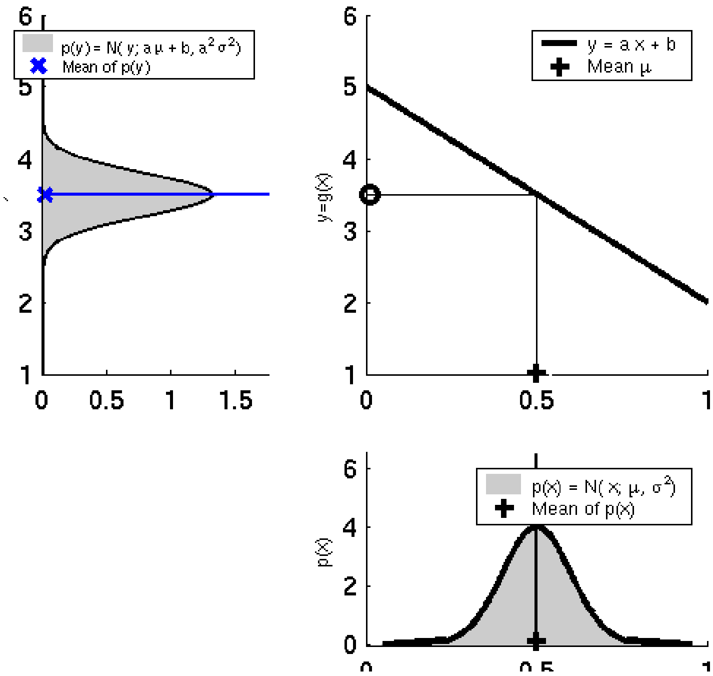
where $v_1$ and $v_2$ are small and uncorrelated noises, i.e. $\mathbf{V}$ is diagonal, while $w$ has a considerably larger standard deviation.

## 7.6 The Extended Kalman Filter

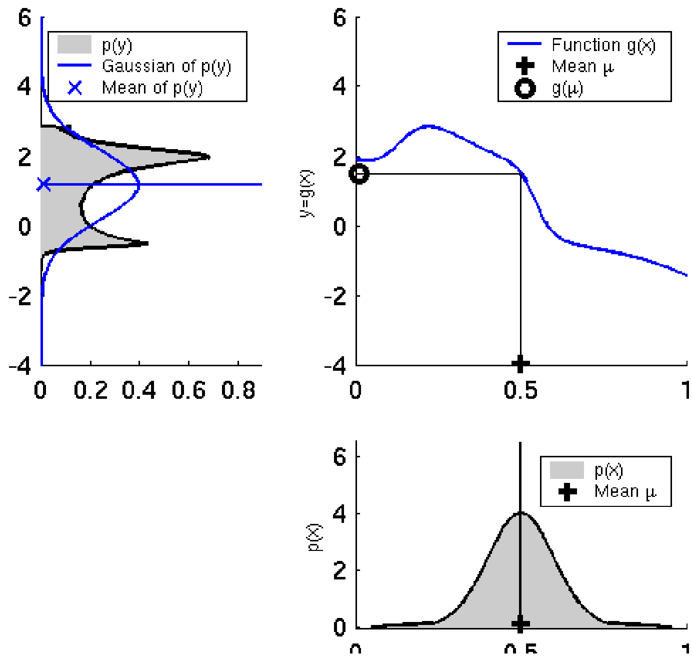
- the original Kalman Filter uses a linear system model
- most practical, real-world systems are nonlinear, e.g. orientation estimation
- if the nonlinearities are mild, i.e. a linearized model is a sufficiently good approximation of the true dynamics, then a linear Kalman Filter can still employed based on that linearization
- however, in case of systems with strong nonlinearities, that approach might yield bad results
- in that case, it might be advantageous to use the *Extended Kalman Filter*
- Consider the nonlinear system

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}, \mathbf{u}_k) + \mathbf{v}_k \tag{88}$$
$$\mathbf{y}_k = h(\mathbf{x}_k) + \mathbf{w}_k \tag{89}$$

- with $\mathbf{x} \in \mathbb{R}^n$ being the *states* of the system, $\mathbf{u} \in \mathbb{R}^p$ the *inputs* to the system and $\mathbf{y} \in \mathbb{R}^q$ the *measurements*
- $\mathbf{v}_k \sim \mathcal{N}(0, \mathbf{V})\ \forall k$ denotes the system noise (model and input uncertainties), and $\mathbf{w}_k \sim \mathcal{N}(0, \mathbf{W})\ \forall k$ denotes the measurement noise
- $f(\cdot)$ describes the dynamics of the system, $h(\cdot)$ is the measurement function
- note a crucial difference between linear and nonlinear systems:
  - *linear system*: if the system dynamics $\mathbf{A}\,\mathbf{x} + \mathbf{B}\,\mathbf{u}$ are applied to the state $\mathbf{x}$ with a Gaussian pdf, the pdf **remains Gaussian**
  - *nonlinear system*: if the system dynamics $f(\mathbf{x}_k, \mathbf{u}_k)$ are applied to the state $\mathbf{x}$ with a Gaussian pdf, the pdf **does not remain Gaussian**




**(a)** linear transformation of a Gaussian pdf [6]




**(b)** nonlinear transformation of a Gaussian pdf [6]

- this implies that the nonlinear prediction $\hat{\mathbf{x}}_k^- = f(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_k)$ does not yield a Gaussian distribution

- solution approach: linearize the dynamics around the last estimate $\hat{\mathbf{x}}_{k-1}$ and approximate the distribution of the nonlinear prediction by Gaussian
- the covariance of that Gaussian is then determined as in the linear case but using the linearization

$$f(\mathbf{x}_{k-1}, \mathbf{u}_k) \approx f(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_k) + \underbrace{\left.\frac{\partial f}{\partial \mathbf{x}_{k-1}}\right|_{\hat{\mathbf{x}}_{k-1}, \mathbf{u}_k}}_{\mathbf{F}_{k-1}} (\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}) + \underbrace{\left.\frac{\partial f}{\partial \mathbf{u}_k}\right|_{\hat{\mathbf{x}}_{k-1}, \mathbf{u}_k} (\mathbf{u}_k - \mathbf{u}_k)}_{=0} \tag{90}$$

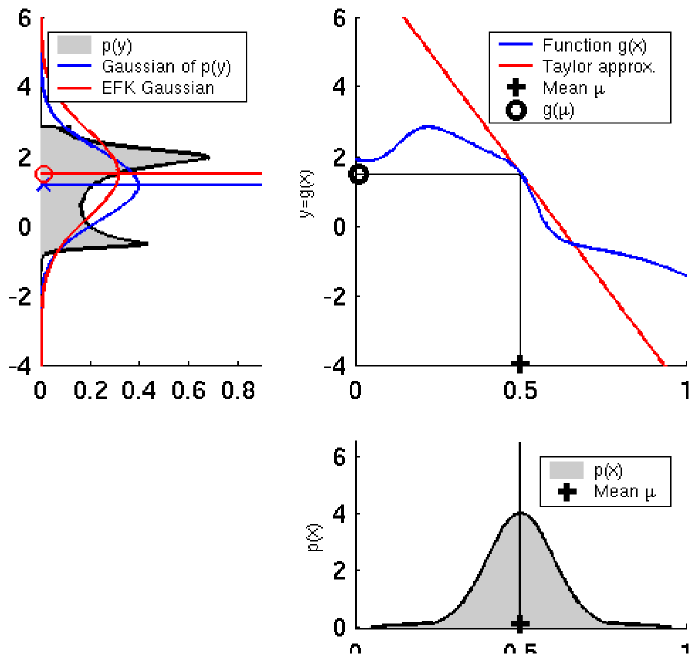
$$= \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \underbrace{f(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_k) - \mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1}}_{\tilde{\mathbf{u}}_k} \tag{91}$$

$$= \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \tilde{\mathbf{u}}_k \tag{92}$$

$$h(\mathbf{x}_k) \approx h(\hat{\mathbf{x}}_k^-) + \underbrace{\left.\frac{\partial h}{\partial \mathbf{x}_k}\right|_{\hat{\mathbf{x}}_k^-}}_{\mathbf{H}_k} (\mathbf{x}_k - \hat{\mathbf{x}}_k^-) \tag{93}$$

$$= \mathbf{H}_k\mathbf{x}_k + \underbrace{h(\hat{\mathbf{x}}_k^-) - \mathbf{H}_k\hat{\mathbf{x}}_k^-}_{\mathbf{z}_k} = \mathbf{H}_k\mathbf{x}_k + \mathbf{z}_k \tag{94}$$

- where $\mathbf{F}_{k-1}$ and $\mathbf{H}_k$ are the Jacobian matrices of the system dynamics $f$ and the measurement function $h$




linear approximation of nonlinear transformation of a Gaussian pdf [6]

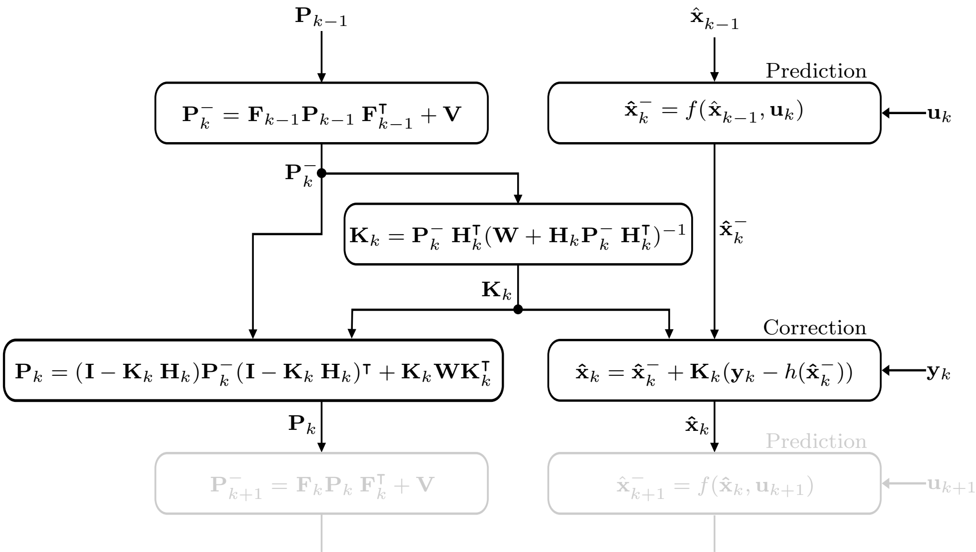
- the linearized model then takes the following form

$$\mathbf{x}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \tilde{\mathbf{u}}_k + \mathbf{v}_k \tag{95}$$
$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{z}_k + \mathbf{w}_k \tag{96}$$

- note that the only nonlinearities (in $\tilde{\mathbf{u}}_k$ and $\mathbf{z}_k$) depend on $\hat{\mathbf{x}}_{k-1}$ and $\hat{\mathbf{x}}_k^-$ and are thus known terms, while the dependency on the variable $\mathbf{x}_{k-1}$ is indeed linear

- using this linearization with the standard Kalman filter equations gives the *extended* Kalman filter:

$$\hat{\mathbf{x}}_k^- = \mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1} + \tilde{\mathbf{u}}_k \quad = f(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_k) \tag{97}$$
$$\mathbf{P}_k^- = \mathbf{F}_{k-1}\mathbf{P}_{k-1}\,\mathbf{F}_{k-1}^\intercal + \mathbf{V} \tag{98}$$
$$\mathbf{K}_k = \mathbf{P}_k^-\,\mathbf{H}_k^\intercal(\mathbf{W} + \mathbf{H}_k\mathbf{P}_k^-\,\mathbf{H}_k^\intercal)^{-1} \tag{99}$$
$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^- - \mathbf{z}_k) \quad = \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{y}_k - h(\hat{\mathbf{x}}_k^-)) \tag{100}$$
$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k\,\mathbf{H}_k)\mathbf{P}_k^-(\mathbf{I} - \mathbf{K}_k\,\mathbf{H}_k)^\intercal + \mathbf{K}_k\mathbf{W}\mathbf{K}_k^\intercal \tag{101}$$




- note that the prediction and correction are done with the nonlinear functions $f$ and $h$

- however the propagation of the covariance uses the linearizations

- Limitations of the EKF
  - for complex systems, the calculation of the Jacobians is computationally expensive
  - truncation errors due to the linearization
  - large estimation errors render the linearization senseless
  - with large uncertainty the linearization error increases
  - the estimator can diverge when dealing with strong nonlinearities

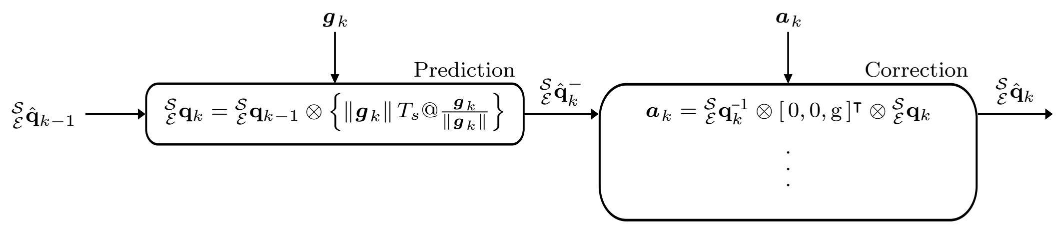
## 7.7 Orientation estimation with an Extended Kalman Filter

Consider an IMU $\mathcal{S}$ measuring the angular velocity $\boldsymbol{g}(t_k)$, the acceleration $\boldsymbol{a}(t_k)$ as well as the magnetic field strength $\boldsymbol{m}(t_k)$ in local coordinates at regular time intervals $t_k$ with a rate $f$ corresponding to a sample time $T_s$. The orientation of the IMU with respect to a reference frame $\mathcal{E}$ at the time instant $t_k$ is denoted by ${}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}_k = {}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}(t_k)$. Equal to the previous approaches, the base of the algorithm is the prediction of the current orientation ${}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}_k$ by strapdown integration of the gyroscope measurement (see (34))

$$ {}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}_k = {}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}_{k-1} \otimes \left\{ \|\boldsymbol{g}_k\|\, T_s @ \frac{\boldsymbol{g}_k}{\|\boldsymbol{g}_k\|} \right\} \tag{102} $$

The measurement and correction step uses the assumption, that the measured acceleration $\boldsymbol{a}_k$ corresponds to the gravitational acceleration transformed into the coordinate system of the sensor:

$$ \boldsymbol{a}_k = {}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}_k^{-1} \otimes [\,0, 0, \mathrm{g}\,]^{\intercal} \otimes {}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}_k \tag{103} $$




In the first approach, we only use the accelerometer to correct the orientation and omit the magnetometer. This results in 6D sensor fusion without heading correction.
Since (102) and (103) are highly nonlinear, an Extended Kalman Filter has to be implemented to solve this sensor fusion task.
We first define the state $\mathbf{x}$, the output $\mathbf{y}$ as well as the input $\mathbf{u}$ according to (102) and (103) as

$$ \mathbf{x}_k := {}^{\mathcal{S}}_{\mathcal{E}}\mathbf{q}_k = \begin{bmatrix} w_k & x_k & y_k & z_k \end{bmatrix}^{\intercal} \tag{104} $$
$$ \mathbf{y}_k := \boldsymbol{a}_k = \begin{bmatrix} a_{x,k} & a_{y,k} & a_{z,k} \end{bmatrix}^{\intercal} \tag{105} $$
$$ \mathbf{u}_k := \boldsymbol{g}_k = \begin{bmatrix} g_{x,k} & g_{y,k} & g_{z,k} \end{bmatrix}^{\intercal} . \tag{106} $$

The nonlinear model $f(\mathbf{x}_{k-1}, \mathbf{u}_k)$ and measurement equation $h(\mathbf{x}_k)$ are

$$ \mathbf{x}_k = f(\mathbf{x}_{k-1}, \mathbf{u}_k) = \mathbf{x}_{k-1} \otimes \begin{bmatrix} \cos(\frac{1}{2}\,\|\mathbf{u}_k\|\, T_s) \\ \sin(\frac{1}{2}\,\|\mathbf{u}_k\|\, T_s)\frac{\mathbf{u}_k}{\|\mathbf{u}_k\|} \end{bmatrix} + \mathbf{v}_k \tag{107} $$
$$ \mathbf{y}_k = h(\mathbf{x}_k) = \mathbf{x}_k^{-1} \otimes [\,0, 0, \mathrm{g}\,]^{\intercal} \otimes \mathbf{x}_k + \mathbf{w}_k \tag{108} $$

For the propagation of the covariance, the Jacobi matrices $\mathbf{F} = \frac{\partial f}{\partial \mathbf{x}}$ and $\mathbf{H} = \frac{\partial h}{\partial \mathbf{x}}$ of (107) and (108) have to be determined. For the sake of a more compact notation we introduce virtual inputs $\tilde{u}_1, \tilde{u}_2, \tilde{u}_3, \tilde{u}_4$, which are the components of the quaternion that is given by the input $\mathbf{u}_k$:

$$ [\,\tilde{u}_1, \tilde{u}_2, \tilde{u}_3, \tilde{u}_4\,]^{\intercal} := \begin{bmatrix} \cos(\frac{1}{2}\,\|\mathbf{u}_k\|\, T_s) \\ \sin(\frac{1}{2}\,\|\mathbf{u}_k\|\, T_s)\frac{\mathbf{u}_k}{\|\mathbf{u}_k\|} \end{bmatrix} \tag{109} $$

The model (107) can then be written as

$$f(\mathbf{x}_{k-1}, \mathbf{u}_k) = \begin{bmatrix} w_{k-1} \\ x_{k-1} \\ y_{k-1} \\ z_{k-1} \end{bmatrix} \otimes \begin{bmatrix} \tilde{u}_1 \\ \tilde{u}_2 \\ \tilde{u}_3 \\ \tilde{u}_4 \end{bmatrix} = \begin{bmatrix} w_{k-1}\tilde{u}_1 - x_{k-1}\tilde{u}_2 - y_{k-1}\tilde{u}_3 - z_{k-1}\tilde{u}_4 \\ w_{k-1}\tilde{u}_2 + x_{k-1}\tilde{u}_1 + y_{k-1}\tilde{u}_4 - z_{k-1}\tilde{u}_3 \\ w_{k-1}\tilde{u}_3 + y_{k-1}\tilde{u}_1 - x_{k-1}\tilde{u}_4 + z_{k-1}\tilde{u}_2 \\ w_{k-1}\tilde{u}_4 + z_{k-1}\tilde{u}_1 + x_{k-1}\tilde{u}_3 - y_{k-1}\tilde{u}_2 \end{bmatrix}. \tag{110}$$

The Jacobi matrix $\mathbf{F}_{k-1}$ of (110) is

$$\mathbf{F}_{k-1} = \frac{\partial f(\mathbf{x}_{k-1}, \mathbf{u}_k)}{\partial \mathbf{x}_{k-1}} = \begin{bmatrix} \tilde{u}_1 & -\tilde{u}_2 & -\tilde{u}_3 & -\tilde{u}_4 \\ \tilde{u}_2 & \tilde{u}_1 & \tilde{u}_4 & -\tilde{u}_3 \\ \tilde{u}_3 & -\tilde{u}_4 & \tilde{u}_1 & \tilde{u}_2 \\ \tilde{u}_4 & \tilde{u}_3 & -\tilde{u}_2 & \tilde{u}_1 \end{bmatrix}. \tag{111}$$

The measurement equation (108) can be written as follows:

$$h(\mathbf{x}_k) = \begin{bmatrix} w_k \\ -x_k \\ -y_k \\ -z_k \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 0 \\ \mathrm{g} \end{bmatrix} \otimes \begin{bmatrix} w_k \\ x_k \\ y_k \\ z_k \end{bmatrix} = \quad \dots \quad = \mathrm{g} \begin{bmatrix} 2x_k z_k - 2y_k w_k \\ 2y_k z_k + 2w_k x_k \\ z_k^2 + w_k^2 - y_k^2 - x_k^2 \end{bmatrix}. \tag{112}$$

The Jacobi matrix $\mathbf{H}_k$ of (112) is then determined as

$$\mathbf{H}_k = \frac{\partial h(\mathbf{x}_k)}{\partial \mathbf{x}_k} = 2\mathrm{g} \begin{bmatrix} -y_k & z_k & -w_k & x_k \\ x_k & w_k & z_k & y_k \\ w_k & -x_k & -y_k & z_k \end{bmatrix}. \tag{113}$$

The accelerometer measurements are more noisy than the gyroscope measurements. Furthermore, during motion, the accelerometer does not provide accurate information on the attitude of the sensor. Therefore we choose for the covariances $\mathbf{V} \in \mathbb{R}^{4\times4}$ and $\mathbf{W} \in \mathbb{R}^{3\times3}$, for example

$$\mathbf{V} = \begin{bmatrix} 10^{-1} & 0 & 0 & 0 \\ 0 & 10^{-1} & 0 & 0 \\ 0 & 0 & 10^{-1} & 0 \\ 0 & 0 & 0 & 10^{-1} \end{bmatrix} \quad \text{and} \quad \mathbf{W} = \begin{bmatrix} 10^6 & 0 & 0 \\ 0 & 10^6 & 0 \\ 0 & 0 & 10^6 \end{bmatrix}. \tag{114}$$

To add an accelerometer rating, i.e. make the accelerometer-based correction dependent on the current norm of the acceleration to mitigate the influence of unwanted accelerations, we can make the matrix $\mathbf{W}$ non-constant and a function of $\|\boldsymbol{a}\|$.
As initial orientation $\hat{\mathbf{x}}_0$ we can either choose a known orientation with a small covariance $\mathbf{P}_0 \in \mathbb{R}^{4\times4}$ or the identity quaternion $[\,1, 0, 0, 0\,]^\intercal$ with a large covariance $\mathbf{P}_0$. The EKF is then implemented as described in (97)-(101).

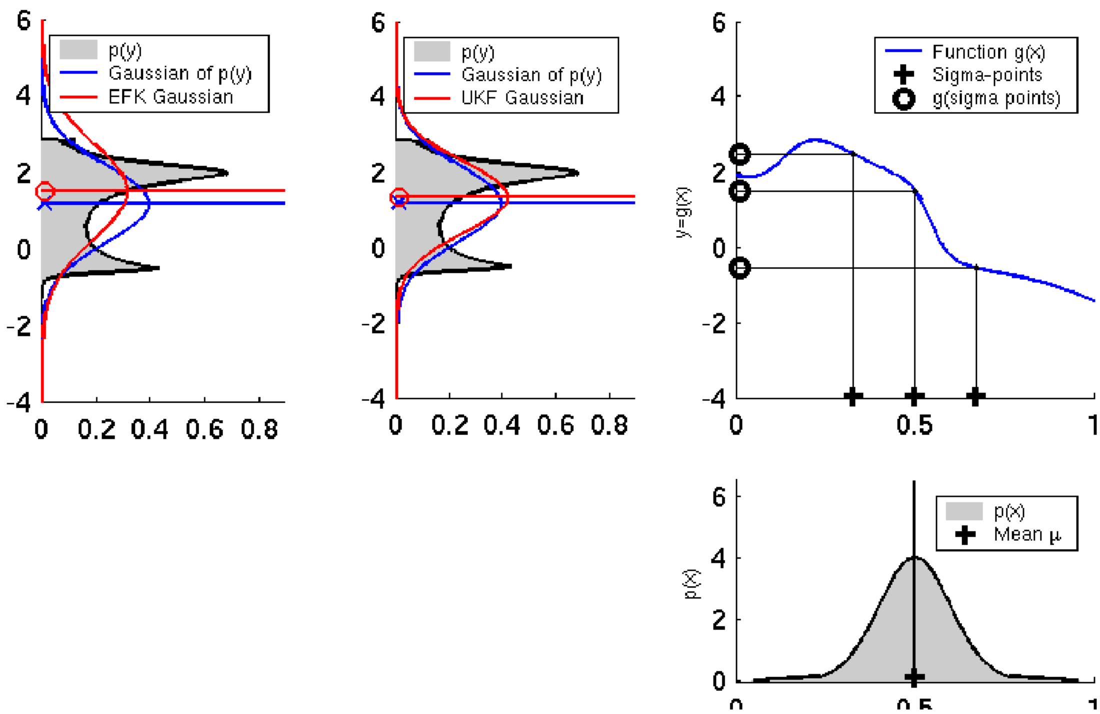
Note that the presented approach, which uses only the accelerometer for correction, cannot correct the heading component of the orientation. To account for this, we can extend the output vector $\mathbf{y}_k$ to also incorporate the magnetometer measurements and find a similar expression like (103):

$$\mathbf{y}_k = \begin{bmatrix} a_{x,k} & a_{y,k} & a_{z,k} & m_{x,k} & m_{y,k} & m_{z,k} \end{bmatrix} \tag{115}$$

We can further extend the state vector $\mathbf{x}_k$ and the model $f(\mathbf{x}_{k-1}, \mathbf{u}_k)$ by the gyroscope bias $\mathbf{b}_k = \begin{bmatrix} b_{x,k} & b_{y,k} & b_{z,k} \end{bmatrix}$ to let the Extended Kalman Filter estimate the remaining bias in the gyroscope measurements.

### 7.8 Unscented Kalman Filter

- for nonlinear systems, the EKF is a better approach for information fusion than a linearization of the model and a standard Kalman Filter
- if the nonlinearities are severe, the linearized propagation of the covariances might not yield a sufficiently accurate approximation of the true probability density function of the state estimation
- this can lead to high inaccuracies or divergence of the filter
- there exist different, more computational expensive methods for approximation of the pdf
- the Unscented Kalman Filter (UKF) uses the *unscented transform* to provide a better estimate of the resulting pdf
- it uses so called *sigma points* of the original Gaussian pdf and evaluates the nonlinear function at these points
- the sigma points are located at the mean and symetrically along the main axes of the original Gaussian




Propagation of the probability density function $p(x)$ by the nonlinear function $y = g(x)$ with an Unscented Kalman Filter [6]

### 7.8.1 Prediction using sigma points

- Consider the state $\mathbf{x}_k \sim \mathcal{N}(\bar{\mathbf{x}}_k, \mathbf{P}_k)$, $\mathbf{x}_k \in \mathbb{R}^n$, which shall be estimated
- the non-linear system dynamics are given as

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}, \mathbf{u}_k) + \mathbf{v}_k \tag{116}$$

$$\mathbf{y}_k = h(\mathbf{x}_k) + \mathbf{w}_k \tag{117}$$

- At time instance $t_k$, the estimate of the last time step $\hat{\mathbf{x}}_{k-1}$ and its covariance $\mathbf{P}_{k-1}$ are given
- to find a prediction $\hat{\mathbf{x}}_k$ as well as its covariance $\mathbf{P}_k$ of the current time step $t_k$, a set of *sigma points* $\boldsymbol{\mathcal{X}}^-$ is determined around the mean $\hat{\mathbf{x}}_{k-1}$ of the previous estimate. The number of sigma points is usually chosen as $2n+1$. The sigma points sample the non-linear mapping of the pdf around the previous estimate to get a better representation of the resulting pdf.

1. Sigma points around previous estimate $\hat{\mathbf{x}}_{k-1}$

$$\boldsymbol{\mathcal{X}}_{k-1} = \Sigma(\hat{\mathbf{x}}_{k-1}, \mathbf{P}_{k-1}) \in \mathbb{R}^{n\times(2n+1)} \tag{118}$$

$$\mathcal{X}_{k-1,i} = \begin{cases} \hat{\mathbf{x}}_{k-1}, & \text{if } i = 0 \\ \hat{\mathbf{x}}_{k-1} + \left(\sqrt{(n+\lambda)\mathbf{P}_{k-1}}\right)_i, & \text{if } 1 \le i \le n \\ \hat{\mathbf{x}}_{k-1} - \left(\sqrt{(n+\lambda)\mathbf{P}_{k-1}}\right)_{i-n}, & \text{if } n+1 \le i \le 2n \end{cases} \tag{119}$$

2. Mapping of sigma points $\boldsymbol{\mathcal{X}}_{k-1}$ through non-linear function $f$ to get the transformed sigma points $\xi_{k,i}$ in the state-space

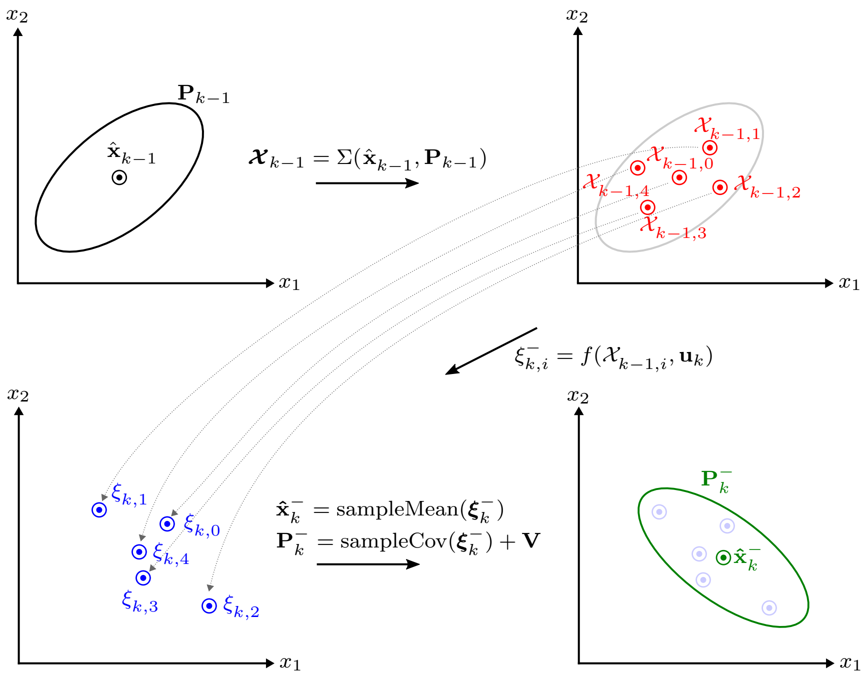
$$\xi_{k,i} = f(\mathcal{X}_{k-1,i}, \mathbf{u}_k) \quad \forall i \in [0, 1, \dots, 2n] \tag{120}$$

3. A-priori prediction $\hat{\mathbf{x}}_k^-$ and covariance $\mathbf{P}_k^-$ as sample mean and covariance of all $\xi_{k,i}$

$$\hat{\mathbf{x}}_k^- = \text{sampleMean}(\boldsymbol{\xi}_k) = \sum_{i=0}^{2n} w_{\text{m},i}\xi_{k,i} \tag{121}$$

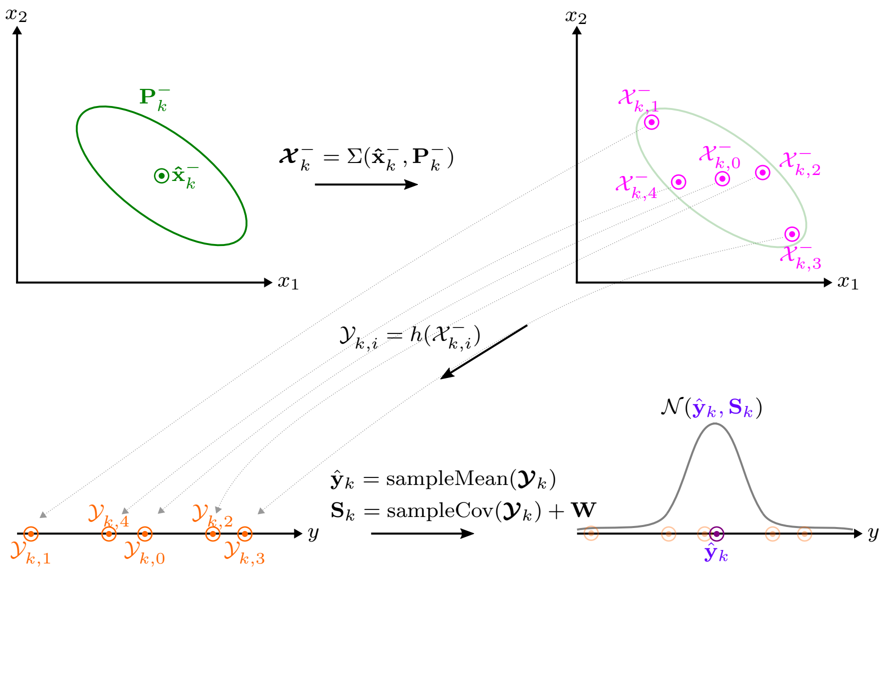
$$\mathbf{P}_k^- = \text{sampleCov}(\boldsymbol{\xi}_k) + \mathbf{V} = \sum_{i=0}^{2n} w_{\text{c},i}(\xi_{k,i} - \hat{\mathbf{x}}_k^-)(\xi_{k,i} - \hat{\mathbf{x}}_k^-)^\mathsf{T} + \mathbf{V} \tag{122}$$

Sigma points do not have to lie on the principal axes of the ellipsoid representing $\mathbf{P}_{k-1}$!




**Prediction $\hat{\mathbf{y}}_k$ of the measurement**

1. Sigma points $\boldsymbol{\mathcal{X}}^-_k$ around a-priori prediction $\hat{\mathbf{x}}^-_k$ with covariance $\mathbf{P}^-_k$
2. Propagation of sigma points $\boldsymbol{\mathcal{X}}^-_k$ through non-linear measurement equation $h$ to get transformed sigma points $\boldsymbol{\mathcal{Y}}_k$ in the measurement space
3. A-priori prediction $\hat{\mathbf{y}}_k$ of $\mathbf{y}_k$ and covariance $\mathbf{S}_k$ as sample mean and covariance of $\boldsymbol{\mathcal{Y}}_k$



### 7.8.2 Original choice of sigma points

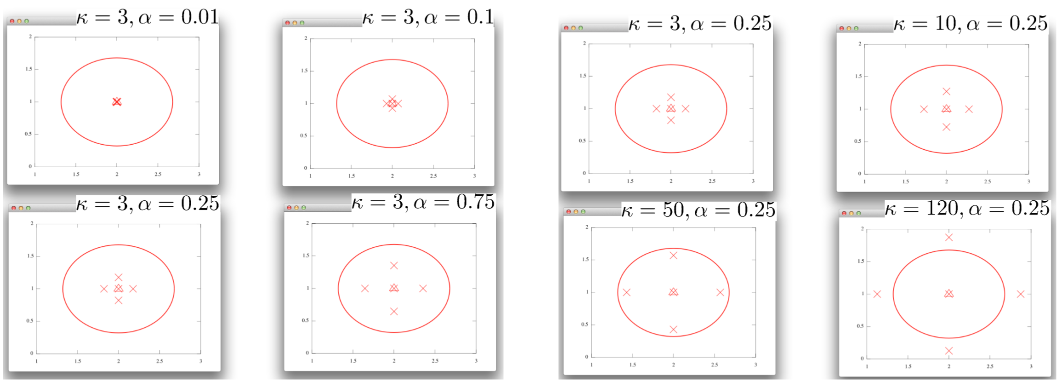
Julier and Uhlmann[7] originally proposed the following sigma points:

$$\mathcal{X}_0^- = \hat{\mathbf{x}}_{k-1} \tag{123}$$

$$\mathcal{X}_i^- = \hat{\mathbf{x}}_{k-1} + \left(\sqrt{(n+\lambda)\mathbf{P}_{k-1}}\right)_i \quad \text{for } i = 1, \ldots, n \tag{124}$$

$$\mathcal{X}_i^- = \hat{\mathbf{x}}_{k-1} - \left(\sqrt{(n+\lambda)\mathbf{P}_{k-1}}\right)_{i-n} \quad \text{for } i = n+1, \ldots, 2n \tag{125}$$

- with $\lambda = \alpha^2(n+\kappa) - n$. The parameters $\alpha$ and $\kappa$ are scaling parameters for the distribution of the points. Usually $0 < \alpha < 1$ and $\kappa = 0$
- $\left(\sqrt{\cdot}\right)_i$ denotes the $i$–th row of the matrix square root (Matlab: `sqrtm()` or with Cholesky decomposition. Note: $\mathbf{P}$ has to be positive definite!)




Influence of scaling parameters $\alpha$ and $\kappa$. The middle point is the previous estimate $\mathbf{x}_{k-1}$, the ellipse illustrates the covariance $\mathbf{P}_{k-1}$. The optimal distance of the sigma points to the previous estimate is dependent of the type of non-linearities in the model.

- each sigma point has two weights $w_\mathrm{m}$ and $w_\mathrm{c}$ associated with

$$w_{\mathrm{m},0} = \frac{\lambda}{n+\lambda} \tag{126}$$

$$w_{\mathrm{c},0} = \frac{\lambda}{n+\lambda} + (1 - \alpha^2 + \beta) \tag{127}$$

$$w_{\mathrm{m},i} = w_{\mathrm{c},i} = \frac{1}{2(n+\lambda)} \tag{128}$$

- the parameter $\beta$ incorporates knowledge of the prior distribution. For Gaussian pdf $\beta = 2$ is optimal
- the weights fulfill $\sum_{i=0}^{2n} w_{\mathrm{m},i} = \sum_{i=0}^{2n} w_{\mathrm{c},i} = 1$. In general, the first weights $w_{\mathrm{m},0}$ and $w_{\mathrm{c},0}$ are high negative numbers
- $w_c$ and $w_m$ weigh the individual sigma points. Sigma points that are located further away from the previous estimate are weighted less
- a trivial choice of weights is $w_{m,i} = w_{c,i} = \frac{1}{n}$

### 7.8.3 The UKF algorithm

- for each timestep the UKF algorithm starts with generating the sigma points $\boldsymbol{\mathcal{X}}_{k-1}$ from the a-posteriori expectation $\hat{\mathbf{x}}_{k-1}$ and covariance $\mathbf{P}_{k-1}$ of the previous timestep
- the sigma points are then passed through the nonlinear model function $f$ to estimate the weighted mean $\hat{\mathbf{x}}_k^-$ of the transformed sigma points as well as the resulting covariance $\mathbf{P}_k^-$

$$\xi_{k,i} = f(\mathcal{X}_{k-1,i}, \mathbf{u}_k) \tag{129}$$

$$\hat{\mathbf{x}}_k^- = \sum_{i=0}^{2n} w_{\mathrm{m},i}\xi_{k,i} \tag{130}$$

$$\mathbf{P}_k^- = \sum_{i=0}^{2n} w_{\mathrm{c},i}(\xi_{k,i} - \hat{\mathbf{x}}_k^-)(\xi_{k,i} - \hat{\mathbf{x}}_k^-)^\intercal + \mathbf{V} \tag{131}$$

- $\hat{\mathbf{x}}_k^-$ is not $f(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_k)$ as in the EKF, but a weighted mean of all transformed sigma points $\xi_{k,i}$!
- for the measurement equation, a new set of sigma points $\boldsymbol{\mathcal{X}}_k^-$ is determined around the a-priori prediction $\hat{\mathbf{x}}_k^-$ and its covariance $\mathbf{P}_k^-$

$$\mathcal{X}_{k,0}^- = \hat{\mathbf{x}}_k^- \tag{132}$$

$$\mathcal{X}_{k,i}^- = \hat{\mathbf{x}}_k^- + \left(\sqrt{(n+\lambda)\mathbf{P}_k^-}\right)_i \qquad \text{for } i = 1, \dots, n \tag{133}$$

$$\mathcal{X}_{k,i}^- = \hat{\mathbf{x}}_k^- - \left(\sqrt{(n+\lambda)\mathbf{P}_k^-}\right)_{i-n} \qquad \text{for } i = n+1, \dots, 2n \tag{134}$$

- these sigma points are then passed through the measurement equation $h$ and the mean $\hat{\mathbf{y}}_k$ and its covariance $\mathbf{S}_k$ are determined

$$\mathcal{Y}_{k,i} = h(\mathcal{X}_{k,i}^-) \tag{135}$$

$$\hat{\mathbf{y}}_k = \sum_{i=0}^{2n} w_{\mathrm{m},i}\mathcal{Y}_{k,i} \tag{136}$$

$$\mathbf{S}_k = \sum_{i=0}^{2n} w_{\mathrm{c},i}(\mathcal{Y}_{k,i} - \hat{\mathbf{y}}_k)(\mathcal{Y}_{k,i} - \hat{\mathbf{y}}_k)^\intercal + \mathbf{W} \tag{137}$$

- Additionally, the cross-covariance $\mathbf{P}_k^{xy}$ has to be calculated

$$\mathbf{P}_k^{xy} = \sum_{i=0}^{2n} w_{c,i}(\xi_{k,i} - \hat{\mathbf{x}}_k^-)(\mathcal{Y}_{k,i} - \hat{\mathbf{y}}_k)^\intercal \tag{138}$$

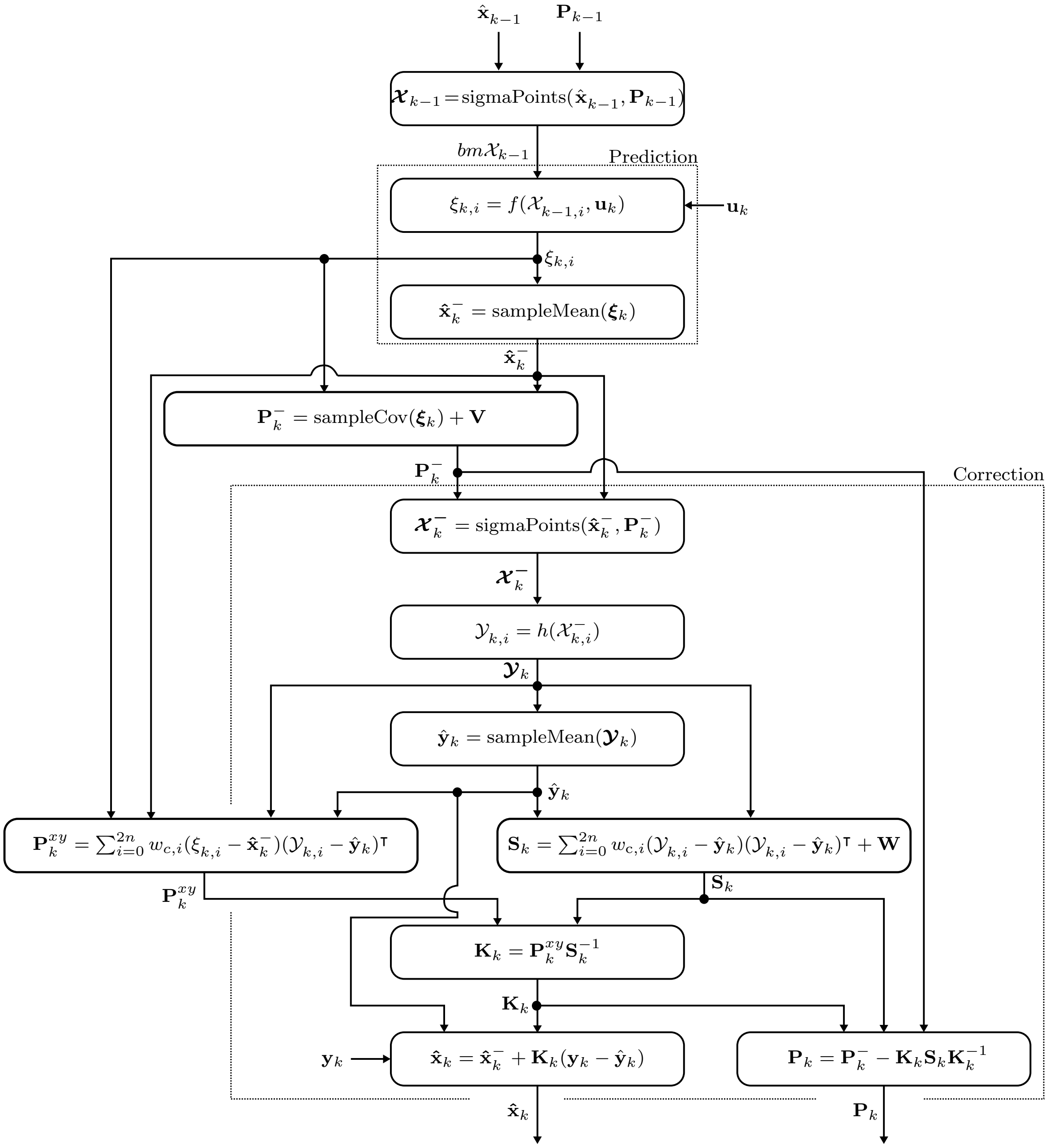
- With this, the Kalman Filter Gain $\mathbf{K}_k$, the current estimate $\hat{\mathbf{x}}_k$ and its covariance can be determined

$$\mathbf{K}_k = \mathbf{P}_k^{xy}\mathbf{S}_k^{-1} \tag{139}$$

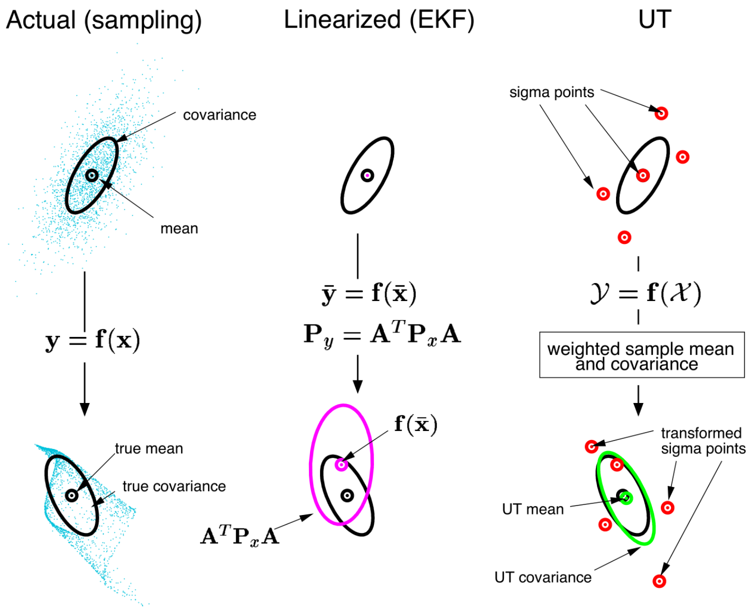
$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{y}_k - \hat{\mathbf{y}}_k) \tag{140}$$

$$\mathbf{P}_k = \mathbf{P}_k^- - \mathbf{K}_k\mathbf{S}_k\mathbf{K}_k^\intercal = \mathbf{P}_k^- - \mathbf{P}_k^{xy}\mathbf{K}_k^\intercal \tag{141}$$

- Analogous to the linear and extended Kalman Filter, the algorithm is shown as a block diagram in the following figure:



### 7.8.4 UKF vs EKF




Comparison between the true distribution, the approximation by the EKF and UKF. It can be seen, that the UKF approximates the true distribution significantly more accurate [8]

- in comparison to the EKF, for the UKF the Jacobi matrices do not need to be calculated and therefore provides a more "plug & play" solution
- the computation time increases with the number of sigma points. In general, more sigma points result in a better approximation of the real pdf
- however, the UKF still makes the assumption of Gaussian pdf for the state, so for highly non-linear systems, this approximation might lead to high deviations from the true state

### 7.8.5 Basic example

Consider the highly non-linear system of equations

$$x_{1,k} = 1 - \frac{1}{2}x_{1,k-1} + x_{2,k-1}\sin(u_k) + v_{1,k} \tag{142}$$

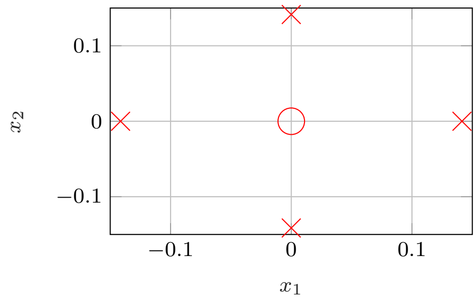
$$x_{2,k} = \cos(x_{1,k-1}) - \frac{1}{3}x_{1,k-1}x_{2,k-1} + v_{2,k} \tag{143}$$

$$y_k = x_{1,k}^3 + w_k \tag{144}$$

We simulate the system of equations to obtain a timeseries of the state $\mathbf{x}_k$ and the measurement $\mathbf{y}_k$ with the covariances $\mathbf{V} = \mathrm{diag}(10^{-1})$ and $\mathbf{W} = \mathrm{diag}(10^1)$. We further choose $\mathbf{x}_0 = [1, 2]^\intercal$ and $u_k = \frac{k}{10}$.

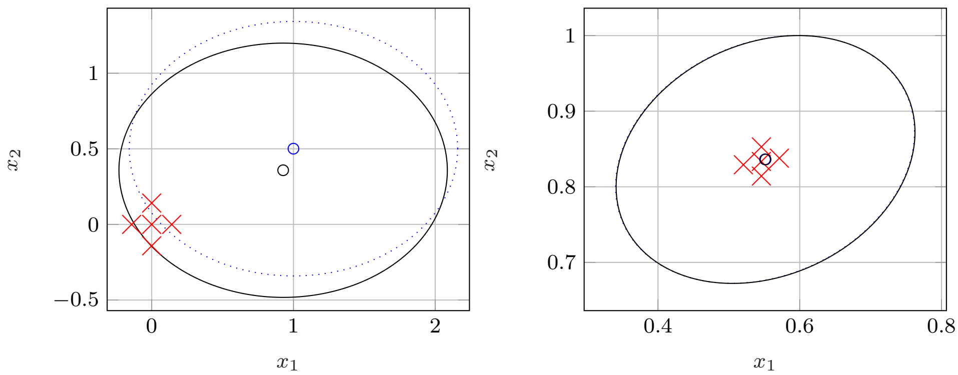
To estimate the state $\mathbf{x}_k$ we use an unscented Kalman Filter with $\hat{\mathbf{x}}_0 = [0, 0]^\intercal$ and $\mathbf{P}_0 = \mathrm{diag}(1)$. For the filter we choose common settings with $\alpha = 10^{-3}, \kappa = 0, \beta = 2$.

For the first time step $k = 1$, the previous estimate $\hat{\mathbf{x}}_{k-1} = [0, 0]^\intercal$ and the corresponding sigma points $\mathcal{X}_i^-$ can be shown in the $x_1 - x_2$-plane:



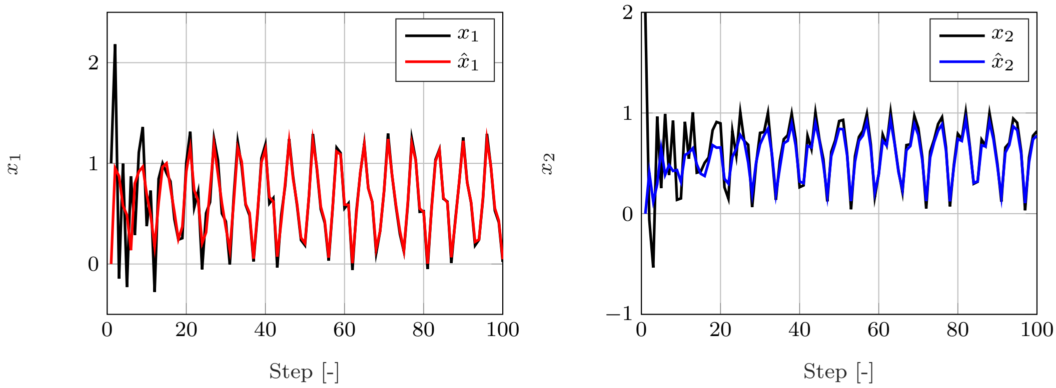

Previous estimate $\hat{\mathbf{x}}_{k-1} = \mathcal{X}_0^-$ (dot) and sigma points $\mathcal{X}_i^-$ $i \in [1 \dots 4]$ (cross)

In the following figure, the sigma points, and the a-priori prediction $\hat{\mathbf{x}}_k^-$ as well as the corrected estimate $\hat{\mathbf{x}}_k$ alongside their covariances $\mathbf{P}_k^-$ and $\mathbf{P}_k$ are shown. At $k = 1$ the difference between the prior $\hat{\mathbf{x}}_{k-1}$ and current estimate $\hat{\mathbf{x}}_k$ is significant, while at $k = 50$ the difference is small.



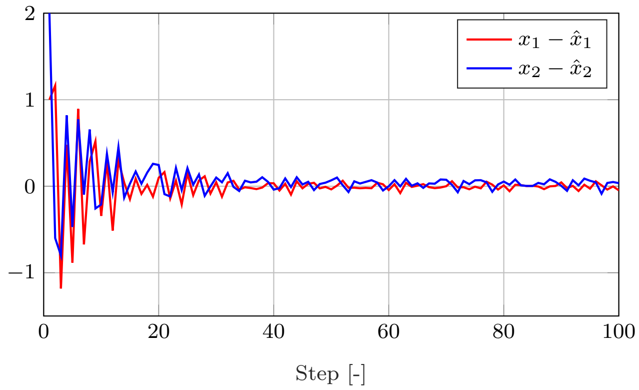

Sigma points (red crosses), a-priori prediction $\hat{\mathbf{x}}_k^-$ and $\mathbf{P}_k^-$ (blue) as well as estimate $\hat{\mathbf{x}}_k$ with its covariance $\mathbf{P}_k$ (black) for $k = 1$ (left) and $k = 50$ (right)

The estimated time series as well as the true values for the two states are shown in the following figure:




True values $x_1$ and $x_2$ as well as estimates $\hat{x}_1$ and $\hat{x}_2$

After a short convergence phase, the UKF can follow the trajectories quite accurately. The difference between the true and estimated values is shown in the following figure:




Error between the true and estimated values of the states

### 7.8.6 Orientation estimation with the unscented Kalman Filter

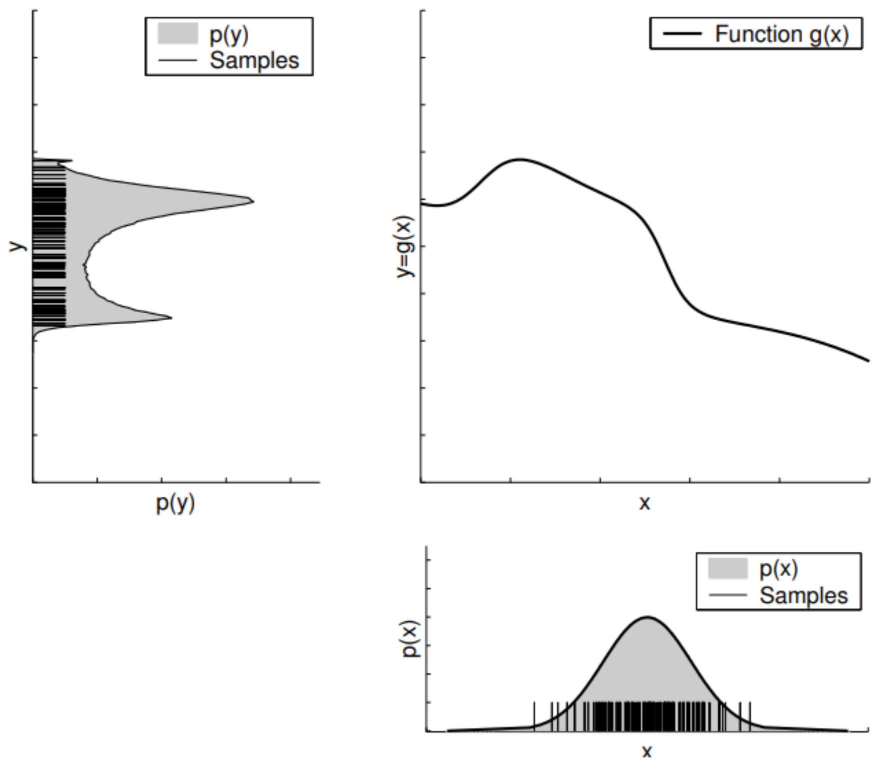
- to use the UKF for orientation estimation, we choose the same model as for the EKF in (107) and (108)
- one method to use the UKF "as is" needs a modification of the correction step to account for non-unit-quaternion-based vector transformation

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k \left( \frac{\mathbf{y}_k}{||\mathbf{y}_k||} - \frac{\hat{\mathbf{y}}_k}{||\hat{\mathbf{y}}_k||} \right) \tag{145}$$

- furthermore, the state $\hat{\mathbf{x}}_k$ has to be normalized before passing it to the next iteration
- another, more advanced adaption of the UKF a quaternion-based orientation tracking is Kraft_2003

## 7.9 Particle Filter

- parametric filter that does not make assumptions on the form of the pdf
- can represent arbitrarily shaped pdf
- particles are drawn from the original distribution and passed through the nonlinear function
- each particle is individually weighted with an *importance factor* based on the knowledge of the prior distribution
- based on the importance factors the set of particles for the next iteration is chosen




Propagation of the probability density function $p(x)$ by the nonlinear function $y = g(x)$ with a particle filter [6]

- good accuracy when the number of particles is sufficiently big
- computationally expensive
- particle deprevation: possibility of particles drifting to low-likelihood regions of the distribution leaving the high likelihood rergions underrepresented

## References


[1] R. Mahony, T. Hamel, and J. . Pflimlin, "Complementary filter design on the special orthogonal group so(3)," in *Proceedings of the 44th IEEE Conference on Decision and Control*, Dec 2005, pp. 1477–1484.

[2] A. Solin, S. Särkkä, J. Kannala, and E. Rahtu, "Terrain navigation in the magnetic landscape: Particle filtering for indoor positioning," 05 2016, pp. 1–9.

[3] M. Kok, J. D. Hol, and T. B. Schön, "An optimization-based approach to human body motion capture using inertial sensors," *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 79–85, Jan. 2014.

[4] B. Taetz, G. Bleser, and M. Miezal, "Towards self-calibrating inertial body motion capture," in *2016 19th International Conference on Information Fusion (FUSION)*, Jul. 2016, pp. 1751–1759.

[5] D. Laidig, P. Müller, and T. Seel, "Automatic anatomical calibration for imu-based elbow angle measurement in disturbed magnetic fields," *Current Directions in Biomedical Engineering*, vol. 3, 09 2017.

[6] D. F. Sebastian Thrun, Wolfram Burgard, *Probabilistic Robotics.* The MIT Press, 2005.

[7] E. A. Wan and R. Van Der Merwe, "The unscented kalman filter for nonlinear estimation," in *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373)*, Oct 2000, pp. 153–158.

[8] ——, "The unscented kalman filter for nonlinear estimation," in *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373)*, Oct 2000, pp. 153–158.