Luca Baglivo, Mariolino De Cecco

# Autonomous Vehicles Navigation

## Fundamentals of "sensor fusion" for localization

Notes for the course on "Robotics and Sensor Fusion for Mechatronic Systems"
for Mechatronic Engineering

# TABLE OF CONTENTS

# 1. Introduction

The autonomous navigation is unerstood as the set of techniques by means of which a system is made able to move with a certain degree of autonomy in a certain type of environment (terrestrial, air, underwater, space). The problems in the field of autonomous navigation that are addressed concern first of all *location* of the system with respect to the environment, the planning of their duties and motion control.

An interesting class of systems covered by the study of autonomous navigation is the one of AGV systems (Autonomous Guided Vehicles). They are now widely used in industries, ports, airports, hospitals, etc, however, measuring the position and the balance of a vehicle it is still a problem of substantial interest.

The sensor fusion is the process that combines information from a number of different sources to provide a complete and robust description (measure) of a set of variables of interest. The sensor fusion is of particular utility in any application where many measures have to be combined together, blended and optimized in order to obtain quality information and integrity suitable for the purpose of the application. The sensor fusion techniques are used in many industrial, military, monitoring, civil surveillance, systems in the control processes and information technology systems. The problem of localisation of autonomous vehicles, in which in almost all cases, the individual transducers are insufficient to constitute a complete and robust localisation system for autonomous navigation, requires the use of sensor fusion techniques to combine the measurements from different types of transducers the characteristics of which, if fused together, allow to obtain a more reliable and accurate measure of the state of the system and the environment surrounding it.

The sensor fusion techniques are widely applied in the field of autonomous navigation, where you need to get a good estimate of the position measurement and balance (*layout*) of a mobile robot. Incremental or dead-reckoning measuring methods, which use encoders, ultrasound gyroscopes etc., have the considerable advantages of being self-contained within the robot, to be relatively simple to use and provide high refresh rate of the measure. However, because these measuring systems integrate relative increments, the errors grow significantly with the increasing of the integration time [3,5,6]. The navigation systems pertaining to the environment make use of references in the environment in which the robot moves. These systems have a lower refresh rate than the incremental ones and work only in certain conditions of visibility of the artificial references. However, they have considerable advantages: because the position and balance measurement is performed with respect to the fixed reference in the environment, the uncertainty is still limited and the repeatability of the measurement with respect to the environment is guaranteed.

To maintain flexibility in route planning and to obtain strength in the accuracy of the measurement with respect to working conditions, you can use a combination of incremental navigation systems and those related to the environment. This will ensure a faster refresh rate of the measurement and limited uncertainty in the estimate of the layout.

Below are the foundations on which the development of sensor fusion techniques is based, according to the quantification of uncertainty with probabilistic models (Bayes theorem, Kalman filter) and brief notes on alternative models to that of probability.

# 2. Fundamentals of "sensor fusion"

## *2.1. Probabilistic models. Bayes theorem.*

Each measuring process implies that in the information obtained from a sensor, a certain level of uncertainty is contained, and it must be, for practical purposes of use of the measure, and even more so for sensor fusion, some way represented analytically and quantified. Important directions are provided to this effect by the ISO standard [8] which defines the measurement as a set of elements: a value, a range of values associated with it, a measurement unit and a confidence level that defines the probability that the measure is within the indicated range. However, as it will be seen later, to combine multiple measures / information, or for the Sensor Fusion, what is needed is also the function of probability density associated with the measurement. This function is taken into account before the integration operation of probability density, which results in a simple interval, what is then associated to the measurement with the confidence interval. At the end of this operation, therefore, you lose the information on how the probability is distributed around the defined value. For example, if we assume a Gaussian type density probability function, the measurement will be constituted of the mean value and a range of values cantered around the average $\mu$ and of amplitude equal to a certain coefficient multiplied by the standard deviation $\sigma$ of the distribution. If you want to associate to the measurement a 95% confidence level (as is usual), the interval will have its extremes in $\pm 2\sigma$ from the mean value. After defining the elements of the measure, the probability distribution type used will be omitted but this information is critical in a Sensor / Data Fusion application. To understand the assertion just stated you will need to follow his train of thought developed in the present booklet on which, therefore, we will work with the probability density functions associated with the object of the measure (measurand) and not with the simple intervals seen in dedicated courses and suggested by ISO standard.

Although there are many methods to represent uncertainty, to date the most widely used models are the *probabilistic ones.* Their wide diffusion is justified by the relative simplicity in use compared to other methods and by their characteristic of being intuitive for the use in the fusion of information and in the decision logic (that comes before the fusion).

It is said, however, that probability is always the best way to represent uncertainty, in fact, reality sometimes suggests the use of alternative methods to obtain information that is more complete and adherent to the measure and to its uncertainty. The main alternative models to the probabilistic ones will be mentioned later.

Below is the description of the essential elements to probabilistic modelling for sensor fusion:

- probability density
- conditional probability
- Bayes theorem

Before passing on to the description of probabilistic models it is however necessary to introduce the concepts of state and observation of a system and related models that describe its evolution.

### 2.1.1. State Model and Observation

The concepts of state and observation of a system are behind the development of various theories, including, of interest to us, the control theory and the data fusion theory. Once defined a physical model of the nature of a system, more or less close to reality, the concept of *state* of a system is associated with the concept of true value of the variables of interest of that system, i.e. of values that can take the physical quantities or properties that generally define the model of the system. The true value of a state is also called *measurand*, i.e. physical quantity whose value is

being investigated in the measurement process. For example, you want to build a kinematics model of a vehicle. We start from the assumption that it is a rigid body, an in-built point of the vehicle is chosen and a reference axis, a position vector (in the plane, for example, two Cartesian coordinates and the trim angle) and a the speed vector (two linear speed and an angular speed) are connected to it. The state will be constituted by the position vector and the velocity vector. In the model is described how (the derivative of the state with respect to time) the system in function of its own state and of other possible parameters evolves over time.

Standard form of the linear model of a system status:

$$\dot{x}(t) = F(t)x(t) + B(t)u(t) + G(t)v(t) \quad (2\text{-}1)$$

Where

$x(t) \in \Re^n$ is the state vector, i.e. the vector of interest,

$u(t) \in \Re^s$ is known as a control vector,

$v(t) \in \Re^q$ is a random variable (called noise) from which depends the uncertainty in the evolution of the state in time (in the previous example it could be an acceleration error, i.e. of the derivative of the variable of speed state),

$F(t)$ is the state matrix (or model) and has dimension $n \times n$,

$B(t)$ is the input matrix and has dimension $n \times s$,

$G(t)$ is the noise matrix and has dimension $n \times q$

Standard form of the model of observation (or a measuring system model):

$$z(t) = H(t)x(t) + D(t)w(t) \quad (2\text{-}2)$$

Where

$z(t) \in \Re^m$ is the observation vector,

$w(t) \in \Re^r$ is a random variable that describes the uncertainty in the observation

$H(t)$ is the matrix of the measurement model and has dimension $m \times n$

$D(t)$ is the matrix of measurement noise and has dimension $m \times r$

The previous two equations define the evolution of a continuous-time system with continuous-time observations of the state. Clearly, for the implementation of the state model and observation of a control algorithm or a fusion algorithm (such as the Kalman filter), it is necessary to build discrete-time models. In this case the continuous time variable $t$ becomes the set of the instants of integration $t = \{t_0, t_1, ..., t_{k-1}, t_k\}$ up to the k-th step and the equations (2-1) and (2-2) can be rewritten in the following way:

$$x(t_k) = F(t_k)x(t_{k-1}) + B(t_k)u(t_k) + G(t_k)v(t_k) \quad (2\text{-}3)$$

$$z(t_k) = H(t_k)x(t_k) + D(t_k)w(t_k) \quad (2\text{-}4)$$

The dissertation that leads to calculate the discrete time matrices $F(t_k)$, $B(t_k)$, $G(t_k)$, $H(t_k)$, $D(t_k)$ is omitted starting from their corresponding continuous time. If the time interval $\Delta t(k) = t_k - t_{k-1}$ between two successive iterations remains constant it is common to write the equations (3-3) and (3-4) in the concise form:

$$x(k) = F(k)x(k-1) + B(k)u(k) + G(k)v(k) \quad (2\text{-}5)$$

$$z(k) = H(k)x(k) + D(k)w(k) \quad (2\text{-}6)$$

In which the index $k$ represents the variable $t_k = t_0 + k \cdot \Delta t$.

If the state model accurately reflected the nature of the system, it would provide a full understanding of its evolution, once the initial conditions and the input variables (control) are known. Unfortunately, due to various types of uncertainty factors, including the same inevitable approximation with which the model is built, often the knowledge of the variables of interest resulting from the state model is not sufficient. It is therefore necessary to obtain additional information by making some *observations* (measures) of the state. They are also based on a more or less complete model which describes how the observation is related to the state on which it depends on and an uncertainty (probabilistic or other) type can be associated to it. The fundamental objective of the sensor fusion, in the case of autonomous vehicles, is to improve the knowledge of the position of the vehicle in function of time in order to be able to use the vehicles for complex tasks that require a measure and a very accurate position control, trying to reduce the uncertainty introduced by the disturbing factors; among these, in addition to the sources of uncertainty that characterize all the sensors, it is to be considered also the not perfect respect of the non-holonomic constraints (wheel slip and other phenomena), which results in a model uncertainty in addition to the measurement.

### 2.1.2. Probabilistic Models

First, the concept of *fulfilment* of a value of the random variable is introduced $x$: in stochastic processes the values of $x \in X$ that occur are defined realizations of the random process. Below the terms "fulfil" and "occur" shall be construed as equivalent.

A probability density function (*pdf*) $P_y(y)$ (or simply $P(y)$ ) is defined on a random variable (random) $y$ which can be a scalar or a vector.

The *pdf* is considered a *probabilistic model* of the amount $y$, whether it is an observation or a state of the system (see §2.1.1)

The *pdf* is valid if:

1. $P(y) > 0, \forall y$;
2. the overall probability is: $\int_y P(y)dy = 1$.

in which the unit value indicates the certainty that the event associated with y will happen

The *joint distribution* $P_{xy}(x, y)$ is defined in the same manner of $P(y)$ and it represents the probability distribution of the joint fulfilment of two random variables.

The marginal probability density of $P_y(y)$ is associated with the joint probability, and is the function that for each value of $y$ expresses the probability that it is fulfilled in combination with all possible values of $x$. It is obtained by integrating the $P_{xy}(x, y)$ compared to $x$:

$$P_y(y) = \int_x P_{xy}(x, y)dx \quad (2\text{-}7)$$

and similarly integrating with respect to y the *marginal pdf* is obtained $P_x(x)$.

The *pdf* joint on $n$ variables $P(x_1,....,x_n)$ can be defined in the same way of the *joint pdf* of two variables.

The *conditioned pdf* $P(x|y)$ is the probability density associated with the variable $x$ conditioned to the previous occurrence of a given value of $y$. In other words, when $y$ is already fulfilled, we are interested in calculating the probability density of the variable $x$. It is defined in the following way:

$$P(x|y) = \frac{P(x,y)}{P(y)} \quad (2\text{-}8)$$

and it has the same properties of a *pdf* with the variable $x$ dependent on a fixed value of the variable $y$. Similarly the conditioned *pdf* is defined $P(y|x)$.

Thus, to express a *pdf* joint in terms of conditional and marginal distributions you can use the chain rule of conditional distributions:

$$P(x,y) = P(x|y)P(y) \quad (2\text{-}9)$$

This equation is derived from the concept of the relationship between two variables. The variables $x$ and $y$ may belong to the same set or different sets, what matters for the purpose of joint distribution is whether the two variables are related to each other or not. $P(x,y) \neq P(x) \cdot P(y)$ as, after y has been fulfilled, if there are relations between $x$ and $y$ the collection to which $x$ belongs to is changed and therefore also its probability of fulfilment, then in general $P(x) \neq P(x|y)$ instead if between $x$ and $y$ there are no relations $P(x) = P(x|y)$.

The derivation rule can be extended to any number of variables in the following way:

$$P(x_1,....,x_n) = P(x_1|x_{2,}..,x_n)P(x_2|x_3,..,x_n)...P(x_{n-1}|x_n)P(x_n) \quad (2\text{-}10)$$

Substituting Equation (2-9) in Equation (2-7) the expression of the marginal distribution of a variable is obtained, in terms of the marginal distribution of a second variable:

$$P_y(y) = \int_x P_{y|x}(y|x)P_x(x)dx \quad (2\text{-}11)$$

This equation is known as *total probability theorem.* It states that the probability density of $y$ can be obtained by considering the ways in which $y$ can occur conditionally on the fulfilment of a specific value of $x$ ($P_{y|x}(y|x)$), weighted by the probability that each of these values of $x$ is fulfilled ($P_x(x)$).

If the knowledge of the value of $y$ does not add any information about the value of $x$ then $x$ and $y$ are called *independent* and

$$P(x) = P(x|y)$$

Therefore in this case:

$$P(x,y) = P(x)P(y) \quad (2\text{-}12)$$

Applying the chain rule to the joint probability density function of three variables:

$$P(x,y,z) = P(x,y|z)P(z)$$
$$= P(x|y,z)P(y|z)P(z)$$

Which together with the hypothesis of conditional independence between $x$ and $y$, once the value of $z$ is fixed, it leads to an intuitive result:

$$P(x,y,z) = P(x,y|z) \cdot P(z) = P(x|z) \cdot P(y|z) \cdot P(z)$$

And so:

$$P(x,y|z) = P(x|z) \cdot P(y|z) \qquad (2\text{-}13)$$

In other words, if $x$ is independent from $y$ given the knowledge of $z$ (since z contains all the information in $y$ about $x$) then the *joint pdf* of $x$ and $y$ conditioned to $z$ is simply the product of *marginal pdf* $x$ and $y$ each conditioned to $z$, as in Equation (2-12).

The concept of conditional independence is the basis of many of the data fusion algorithms. It is important to clarify with an example: consider the state $x$ of a system and two observations $z_1$ and $z_2$. The two observations are not independent from each other:

$$P(z_1, z_2) \neq P(z_1)P(z_2)$$

In fact, both depend on the state $x$ in common. On the other hand, it is reasonable to assume that the only thing that the two observations have in common is the system state, so the comments will be independent once it has been fulfilled; i.e. you can write that

$$P(z_1, z_2|x) = P(z_1|x)P(z_2|x)$$

### 2.1.3. Bayes theorem

Bayes theorem, however simple, is a result of fundamental importance in the study of probabilistic models and accordingly in the data fusion algorithm design.

Consider two random variables $x$ and $z$. In order to apply Bayes theorem to the problem of data fusion the variable $x$ is given the state of the system interest, or measurand, and to $z$ the observation of that state, or measure (for example the state of interest may be the linear speed of a vehicle and the observation can be a measured value of the angular speed of the wheels, that through the knowledge of the radius of the wheels, generally of the observation model, leads to the estimation of the linear velocity, or it can be a direct measure of linear velocity).

The conditional probability derivation rule can be used to express the density function in two different ways:

$$P(x,z) = P(x|z)P(z) = P(z|x)P(x)$$

Bayes theorem is obtained by expressing the $P(x|z)$ thanks to the last equation written above:

$$P(x|z) = \frac{P(z|x)P(x)}{P(z)} \qquad (2\text{-}14)$$

In the Equation (2-14) the marginal distribution $P(z)$ is simply needed to normalize the posterior distribution (it is a term that does not depend on the variable $x$ which is the one of interest, and so it can be considered ***constant***), in fact:

$$\int_x P(x|z)\cdot dx = \int_x \frac{P(z|x)P(x)}{P(z)}\cdot dx = \int_x \frac{P(x,z)}{P(z)}\cdot dx = \frac{P(z)}{P(z)} = 1$$

Equation (2-14) shows that $P(x|z)$ is a probability density function and it expresses what is typically sought in a measurement operation, or the probability associated with the measurand $x$**,** given the measurement $z$ with the instrument we have. After that we can refer to ISO standard to give a value and a confidence interval in addition to the unit of measure used.

Bayes theorem thus provides a direct method for combining the information observed with the previous estimate of the probability density associated with measuring (i.e. the $P(x)$) to obtain a new probability density, conditioned to the z measurement, therefore better than the previous one. It is therefore the basis for many of the data fusion algorithms.

In other words, the information contained in the observation is used to calculate the new probability distribution associated with the state $x$ with respect to the previous probability density function. The reshuffle is called **posterior distribution** $P(x|z)$ and it describes the probabilities associated with $x$ given the observation $z$.

To better fully understand the value and usefulness of this (seemingly) simple manipulation of the probability density functions it is necessary to analyze the meaning of functions $P(x|z)$, $P(z|x)$, $P(z)$ and $P(x)$.

The $pdf\, P(x)$ is **the prior probability density function** and quantifies the uncertainty with which the expected value of a state before the new observation $z$ is known.

In order to get more information about the state $x$**, an observation is carried out** $z$**.** The observations are modelled by the conditional probability density function $P(z|x)$ which describes, for each fixed state $x \in X$, the probability that the observation will be carried out $z \in Z$, otherwise known as the probability of $z$ given $x$.

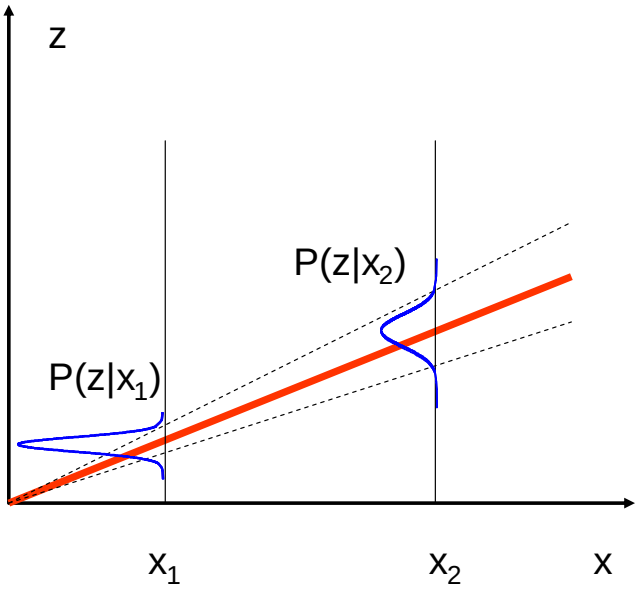
The conditional distribution $P(z|x)$ plays the role of what is called *sensor model.* In this model $x$ is the input / measurand and $z$ is the output / observation of the sensor. This distribution can be thought of in two distinct ways.

1. the first is referred to the *construction of the sensor model* (see Picture 2-1), in this case the distribution is obtained by setting a particular value of $x = x_p$ finding which results the *pdf* relative to the variable $z$ (it coincides with the calibration process in which determines the uncertainty of the instrument in function of the input). Therefore, in this case $P(z|x_p)$ is considered function of $z$. For example, suppose that the true value of the distance from a target is known ( $x_t$ ), so $P(z|x_t)$ will be the distribution of the observations (measurements) around this value (see Picture 2-2).
2. but if there is already a model of the sensor, suppose making an observation and therefore fix $z = z_p$. From this observation you want to infer the value of the state $x$. In this case the distribution $P(z_p|x)$ is a function of $x$ and it is known as **Likelihood Function** $\Lambda(x) = P(z_p \mid x)$ (see Picture 2-3).

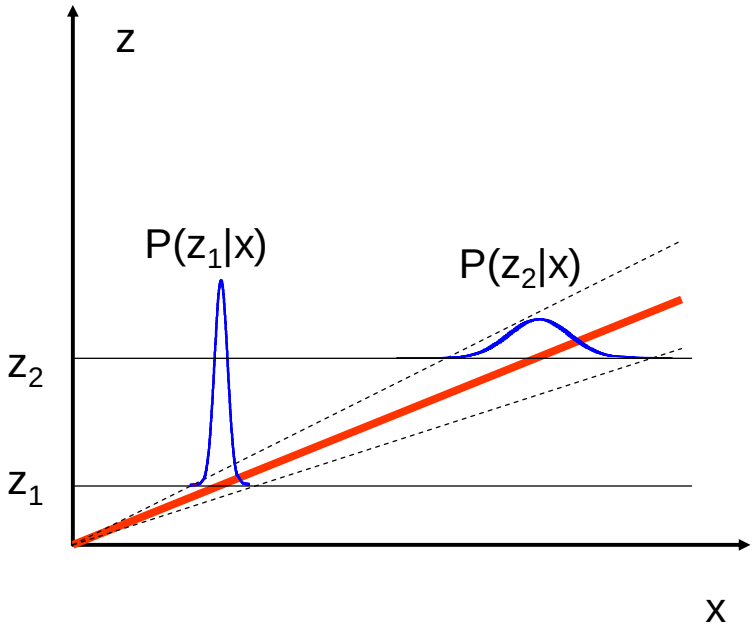

**Picture 2-1: Probabilistic model of the sensor.**

In the calibration phase the work is carried out in a direct way as in Picture 2-2, but, in reality, we are interested in the reverse process, that is the $P(x \mid z)$, i.e. the distribution of the value of measurand x based on measurement z.

In the practical implementation of Bayes theorem (Equation (2-14)), $P(z \mid x)$ is constructed as a function of both variables and in the case of discrete variables it assumes matrix form. For each value of $x$ fixed a distribution of $z$ is defined. So with the variation of $x$ a family of distributions is created in $z$.




**Picture 2-2: Sensor Model: the real state value is supposedly known and the distribution of the measures of this value is observed. It is equivalent to the phase of calibration of the sensor, in which the true value of the measurand is the one detected with a reference instrument of higher accuracy than the one to be calibrated**




**Picture 2-3Likelihood Functions: in order to make a measurement an output of the instrument is read ($z_1$ oz$_2$) and a possible distribution of the true state value is associated to it.**

## Example 1

Consider a constant value state $x$, for example the distance from a target, and an observation $z$ of this state. The model used for the distribution of the observations of the real state value, commonly used for this type of observations, is the Gaussian one, i.e. of the Normal Distribution, where the observations are distributed with mean $x$ and variance $\sigma_z^2$:

$$P(z|x) = \frac{1}{\sqrt{2\pi}\sigma_z}\exp(-\frac{1}{2}\frac{(z-x)^2}{\sigma_z^2}) \qquad (2\text{-}15)$$

This is considered to be a function of both variables $z$ and $x$. If you know the value of the state, then the distribution is a function only of $z$ and it describes a model of the sensor. If you make an observation, the distribution and function only of $x$ describes the probability of the state according to a Gaussian with mean $z$ and variance $\sigma_z^2$. In the latter case it is the Likelihood Function.

Assume that you have a previous estimate of the true state $x = x_p$ also based on the Normal probability model, function only of $x$

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma_x}\exp(-\frac{1}{2}\frac{(x-x_p)^2}{\sigma_x^2})$$

Bayes theorem can be applied directly to combine this previous information with the sensor information. In a practical implementation, first the observation is carried out $z = z_p$, its value is substituted to the random variable $z$ in the Equation (2-15). Then the two distributions are multiplied together to produce the posterior distribution

$$P(x|z) = K\frac{1}{\sqrt{2\pi}\sigma_z}\exp(-\frac{1}{2}\frac{(x-z_p)^2}{\sigma_z^2})\frac{1}{\sqrt{2\pi}\sigma_x}\exp(-\frac{1}{2}\frac{(x-x_p)^2}{\sigma_x^2})$$

$$= \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{1}{2}\frac{(x-\bar{x})^2}{\sigma^2})$$

Constant K is a constant independent from $x$ and chosen so that the posterior distribution is normalized (according to the hypothesis 2. of §2.1.2). The mean and variance of the posterior distribution turn out to be

$$\bar{x} = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_z^2} z_p + \frac{\sigma_z^2}{\sigma_x^2 + \sigma_z^2} x_p$$
$$\sigma^2 = \frac{\sigma_z^2 \sigma_x^2}{\sigma_x^2 + \sigma_z^2} = \left( \frac{1}{\sigma_x^2} + \frac{1}{\sigma_z^2} \right)^{-1}$$

The posterior distribution is therefore also Gaussian with a mean (state estimated value) that appears to be the weighted average between the average of the previous distribution and likelihood, and with a variance (uncertainty assigned to the estimate) equal to the parallel combination between the original variances.

The calculation of the normalization factor K does not have, in this case, a fundamental importance because, given that both distributions multiplied between them are of Gaussian type, for practical purposes the extended expression of distributions is overlooked, passing directly to the calculation, known, of the media and the resulting variance, given the averages and the starting variances. For the purposes of clarity, below is the calculation of the normalization factor K. By adding the exponents of the two distributions we have:

$$P(x|z) = K \frac{1}{2\pi\sigma_z\sigma_x} \exp\left( -\frac{1}{2} \left( \frac{\sigma_x^2 (x - z_p)^2 + \sigma_z^2 (x - x_p)^2}{\sigma_x^2 \sigma_z^2} \right) \right)$$

The argument of the exponential can be written as

$$-\frac{1}{2} \left( \frac{1}{\sigma^2} \left( x^2 - 2\sigma^2 \left( \frac{x_p}{\sigma_x^2} + \frac{z_p}{\sigma_z^2} \right) x + \sigma^2 \left( \frac{x_p^2}{\sigma_x^2} + \frac{z_p^2}{\sigma_z^2} \right) \right) \right)$$
$$= -\frac{1}{2} \frac{\left( x^2 - 2\bar{x}x + \bar{x}^2 - \bar{x}^2 + \sigma^2 \left( \frac{x_p^2}{\sigma_x^2} + \frac{z_p^2}{\sigma_z^2} \right) \right)}{\sigma^2}$$

Where the term $\bar{x}^2$ was added and subtracted in order to have:

$$-\frac{1}{2\sigma^2} \left( (x - \bar{x})^2 - \bar{x}^2 + \sigma^2 \left( \frac{x_p^2}{\sigma_x^2} + \frac{z_p^2}{\sigma_z^2} \right) \right)$$

Returning to the complete expression of the distribution:

$$P(x|z) = K \frac{1}{2\pi\sigma_z^2\sigma_x^2} \exp\left( -\frac{1}{2\sigma^2} \left( (x - \bar{x})^2 - \bar{x}^2 + \sigma^2 \left( \frac{x_p^2}{\sigma_x^2} + \frac{z_p^2}{\sigma_z^2} \right) \right) \right)$$

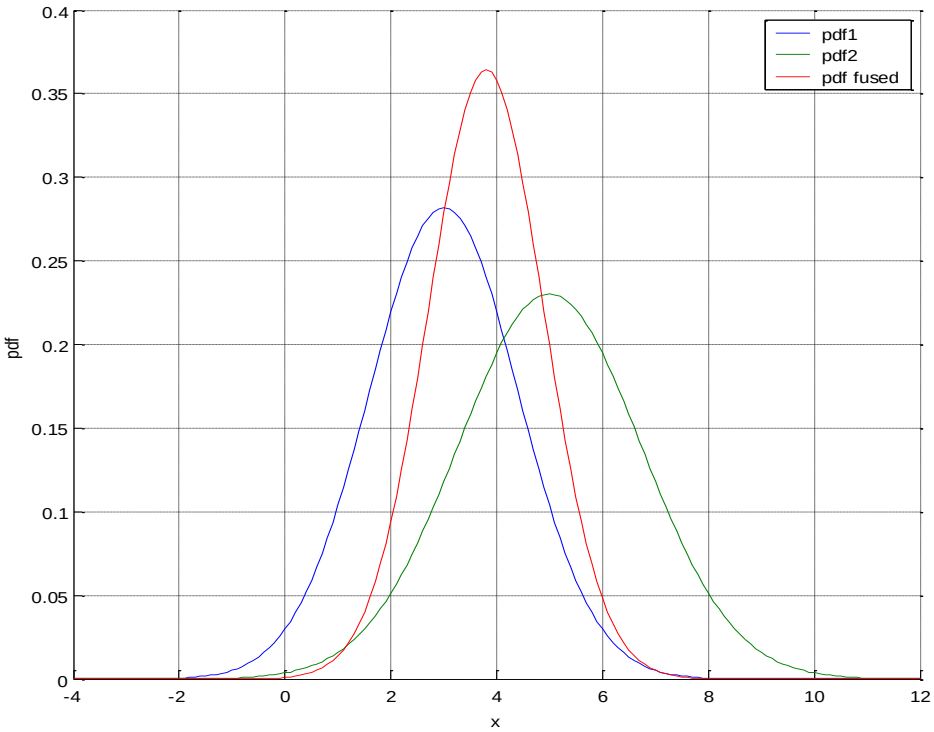
Thus the known term turns out to be (remembering that the factor in the denominator of the normal distribution must be $1/(\sqrt{2\pi}\sigma)$)

$$K = \sqrt{2\pi} \sqrt{\frac{\sigma_x^2 + \sigma_z^2}{\sigma_x \sigma_z}} \exp\left( \frac{\bar{x}^2}{2\sigma^2} - \frac{1}{2} \left( \frac{x_p^2}{\sigma_x^2} + \frac{z_p^2}{\sigma_z^2} \right) \right)$$

Once known, the symmetric property of the Gaussian distributions (as shown), namely that the product of two Gaussian distributions is still a Gaussian distribution with mean and variance expressed by the above

equations, it is not necessary to perform calculations every time on distributions, but it is sufficient to use directly the already known expressions of the resulting mean and variance. This result allows the real time calculation in an implementation of the data fusion based on these probabilistic models. Later we will explain the results of the combination, according to Bayes theorem applied to Gaussian distributions, of the information of vector values and not scalar, in which instead of the variance the covariance matrix is used.

In Picture 2-4 there is the graphic example of what could be, in the scalar case, the previous distribution, the likelihood function and the posterior distribution. It is clear that the posterior distribution is still of Gaussian type, as expected, it is situated in a position interposed between the other two distributions (compared to their respective medium) and is a narrower bell having a smaller variance than the two original variance.




**Picture 2-4: graphic example of what can be, in the scalar case, the previous distribution, the likelihood function and the posterior distribution. The posterior distribution has as average the average weighted among the original averages and has a lower variance than both the original ones, in fact, it has more information than the individual contributions.**

## Example 2

Consider the case in which the state *x* is a continuous vector variable and so is the observation vector *z*.

To quantitatively describe the dispersion of the random variable vector $y$ it is necessary to calculate the *covariance matrix* defined as

$$C = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})(y_i - \bar{y})^T$$

where $\bar{y}$ is the mean value of the $N$ samples $y_i$ of the variable $y$. If $y_i$ It is the i-th sample of $y$ type:

$$y_i = [\eta_i, \xi_i]$$

Then the covariance matrix takes the form

$$C = \begin{bmatrix} \sigma_\eta^2 & \mathrm{cov}(\eta,\xi) \\ \mathrm{cov}(\xi,\eta) & \sigma_\xi^2 \end{bmatrix}$$

$C$ is a symmetric matrix and its mixed terms are not void only if the components of the vector *y* are related to each other.
Assuming that the state follows a Normal distribution of probability, the previous distribution $\mathrm{P}(\mathrm{x})$ can be represented with a multi-dimensional Gaussian of the type

$$P(x) = \frac{1}{\sqrt{2\pi^d \det(C_x)}} e^{-\frac{1}{2}(x-x_p)^T C_x^{-1}(x-x_p)}$$

Where $d$ is the size of the vector and $x_p$ the mean value.

Therefore, if an observation is made $z = z_p$, two dimension vector for example, with a sensor whose relative likelihood function $z_p$ has covariance matrix $C_z$:

$$P(z|x) = \frac{1}{2\pi \det C_z} e^{-\frac{1}{2}(z_p-x)^T C_z^{-1}(z_p-x)}$$

It is finally possible to calculate the posterior distribution $P(x \mid z)$, in the same way of the previous example, but taking into account that the distributions are vector-valued. Therefore, the posterior distribution will be still Gaussian, with covariance matrix equal to the parallel of the starting covariance matrices, namely:

$$C = (C_x^{-1} + C_z^{-1})^{-1}$$

The expression of $C$ can also be written more efficiently by calculating only one matrix inversion:

$$C_x^{-1} + C_z^{-1} = C_x^{-1} C_z C_z^{-1} + C_z^{-1} = (C_x^{-1} C_z + I) C_z^{-1} =$$
$$= (C_x^{-1} C_z + C_x^{-1} C_x) C_z^{-1} = C_x^{-1}(C_z + C_x) C_z^{-1}$$

Whence

$$\left(C_x^{-1} + C_z^{-1}\right)^{-1} = \left(C_x^{-1}(C_x + C_z) C_z^{-1}\right)^{-1} = C_z (C_x + C_z)^{-1} C_x$$

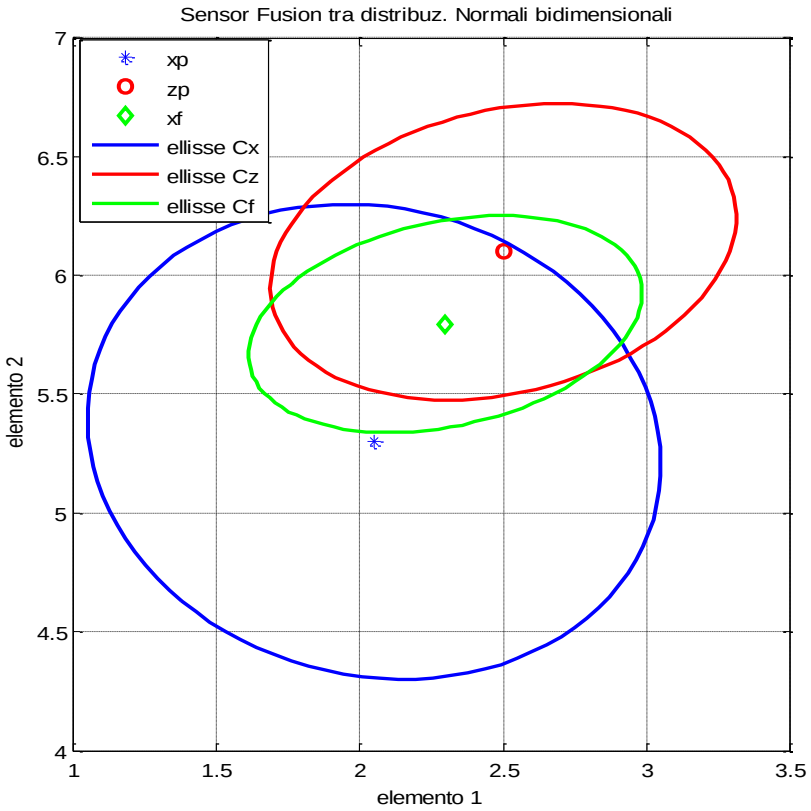
So

$$\mathbf{C = C_z (C_x + C_z)^{-1} C_x}$$

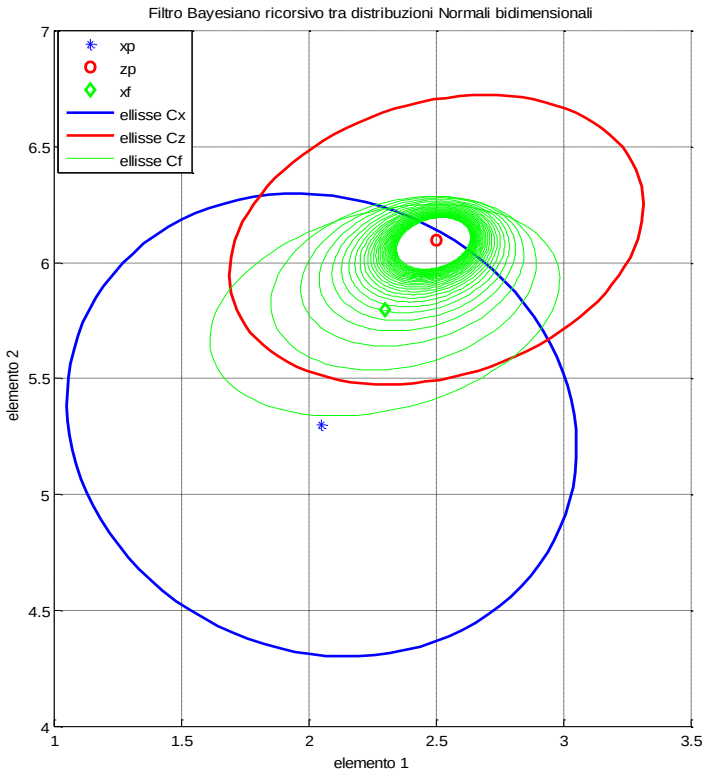
While the mean of the posterior distribution will be (similarly to the scalar case)

$$\mathbf{\bar{x} = C_x (C_x + C_z)^{-1} z_p + C_z (C_x + C_z)^{-1} x_p}$$

Graphically a two-dimensional Normal distribution is represented by a Gaussian bell whose cross sections are ellipses that represent the locus of points characterized by equal probability. In the covariance matrix there is the information about the size of the main axes of the ellipses (eigenvalues of matrix) and their orientation (eigenvectors). An example of what graphically corresponds to the just obtained equations is reported in Picture 2-5.




**Picture 2-5: Merger between the previous distribution of likelihood and probability distribution relative to the observation zp.**




**Picture 2-6: Equi-probability Ellipses in the case where the observation obtained and the relative uncertainty ellipse always remain the same. The ellipse of the posterior distribution converges with its centre toward the observed value and is increasingly being reduced more with every iteration.**

In Picture 2-6 the case in which the observation $z_p$ is repeated 20 times has been simulated. As expected, the ellipse of the posterior distribution converges with its centre toward the observed value and is increasingly being reduced more with every iteration despite the ellipse of equiprobability of the observation (obtained from the likelihood function) remains the same.

### 2.1.4. Application of Bayes theorem to the sensor fusion

Bayes theorem can be applied directly to the fusion of information coming from different sources (sensors). What is wanted is to obtain posterior distribution $P(x|Z^n)$ where $Z^n \equiv \{z_1 \in Z_1, ..., z_n \in Z_n\}$ is the set of observations.

The posterior distribution describes the probability of the various values of the state $x \in X$ given the information obtained from the $n$ observations. In principle, Bayes theorem used directly to calculate the distribution function provides

$$P(x|Z^n) = \frac{P(Z^n|x)P(x)}{P(Z^n)} = \frac{P(z_1, z_2, ..., z_n|x)P(x)}{P(z_1, z_2, ..., z_n)} \quad (2\text{-}16)$$

In practice it would be difficult to apply this calculation because the joint and conditioned distribution should be known completely $P(z_1, z_2, ..., z_n|x)$, that is, the joint distribution of all possible combinations of observations conditioned to the state. It is considered that, under certain assumptions, to assume that once given the state $x \in X$, the information obtained from the i-th source of information is independent from the information of the other sources. Thanks to this assumption, extending the $P(x, y|z) = P(x|z) \cdot P(y|z)$ **(2-13)**(2-13) to the case of $n$ variables,

$$P(z_i|x, z_1, ..., z_{i-1}, z_{i+1}, ..., z_n) = P(z_i|x)$$

$$P(z_1, ..., z_n|x) = P(z_1|x)...P(z_n|x) = \prod_{i=1}^{n} P(z_i|x) \quad (2\text{-}17)$$

Substituting the $P(z_1, ..., z_n|x) = P(z_1|x)...P(z_n|x) = \prod_{i=1}^{n} P(z_i|x)$ **(2-17)**(2-17) into (2-16),

$$P(x|Z^n) = \left[P(Z^n)\right]^{-1} P(x) \prod_{i=1}^{n} P(z_i|x) \quad (2\text{-}18)$$

So the posterior distribution on $x$, that is to say the updated probability in the state, are proportional to the product of the previous probability distribution and the probability distributions of each observation. The marginal distribution $P(Z^n)$ acts as a normalization constant. The equation (2-18) provides a simple method for the fusion of information from multiple sensors and is called *probability independent group*. In the implementation of this algorithm, first of all the conditional probabilities are stored $P(z_i|x)$ as functions of both variables $z_i$ and $x$. When a sequence of observations (of various sensors but also of the same sensor) is performed the values observed are

replaced in their respective probability distributions and the $n$ likelihood functions $\Lambda_i(x)$ are built. The product of these functions for the previous information $P(x)$, normalized by $P(Z^n)$, provides the update of information on the state $P(x|Z^n)$ which is a function of only the $x$ for a specific set of $n$ observations $\{z_1, z_2, ..., z_n\}$.

The effectiveness of this method is based on the assumption that the observations are independent among them when conditioned to the true value of the state. This assumption is reasonable if the state to which the observations relate is the only thing they have in common, so once the state has been specified it is reasonable to assume that the information is conditionally independent. This would certainly not be corrected without conditionality: it would be wrong to say that the information is unconditionally independent as

$$P(z_1, z_1, ...z_n) \neq \prod_{i=1}^{n} P(z_i)$$

each i-th content depends on the common state $x \in X$ .
It is not said that the conditional independence assumption i is always reasonable, for example in the case in which an observation can significantly change the status.

### Example 3

Consider an example of the application of Bayes' theorem to the sensor fusion between two sensors to estimate a discrete parameter. The environment of interest is modelled by a single state $x$ which can take one of the following three values:

$x_1$: *$x$ is a type 1 target;*

$x_2$: *$x$ is a type 2 target;*

$x_3$: *there are no visible targets;*

A single sensor observes $x$ and it returns three possible values:

$z_1$: *observation of a type 1 target;*

$z_2$: *observation of a type 2 target;*

$z_3$: *no target observed;*

The model of sensor A is described by the probability matrix:

$$P_A(z_A|x) = \begin{bmatrix} & z_1 & z_2 & z_3 \\ x_1 & 0.45 & 0.45 & 0.1 \\ x_2 & 0.45 & 0.45 & 0.1 \\ x_3 & 0.1 & 0.1 & 0.8 \end{bmatrix}$$

The model of sensor B is described by the probability matrix

$$P_B(z_B|x) = \begin{bmatrix} & z_1 & z_2 & z_3 \\ x_1 & 0.45 & 0.1 & 0.45 \\ x_2 & 0.1 & 0.45 & 0.45 \\ x_3 & 0.45 & 0.45 & 0.1 \end{bmatrix}$$

These probability matrices are a function both of $x$ and of $z$. For a fixed real value of the state they describe the probability that a particular observation is made (this is the model of the sensor and in that sense, the

matrix should be read along its lines). When a particular observation is made, the matrix describes the probability distribution of the true state value (this is the likelihood function and in that sense, the matrix should be read along its columns).

The posterior distribution of the state $x$ conditioned to the observation $z_A = z_i$, carried out for example by sensor A, is given by

$$P(x|z_{A,i}) = k \times P_A(z_i|x)P(x)$$

Where $k$ is the normalization factor defined by the condition that the sum of the probabilities of the variable $x$ is unitary.

If we assume that there is no prior knowledge about the state:

$$P(x) = [1/3, 1/3, 1/3] \square [0.333, 0.333, 0.333]$$

Assuming that the observation will be made available $z_A = z_1$ from sensor A, the posterior distribution in this case will be given by the first column of the matrix of probabilities of sensor A (notation $\otimes$ indicates the product element by element), i.e. the likelihood function for the observation $z_A = z_1$

$$\begin{aligned} P(x|z_{A,1}) &= k \times P_A(z_1|x)P(x) \\ &= k \times (0.45, 0.45, 0.1) \otimes (0.333, 0.333, 0.333) = (0.45, 0.45, 0.1) \end{aligned}$$

If then we use the posterior distribution as previous distribution for a merging with a new observation of sensor A, $P(x) = (0.45, 0.45, 0.1)$ and a new observation is made available, we assume once again $z_A = z_1$, the new posterior distribution will be:

$$\begin{aligned} P(x|z_{A,1}) &= k \times P_A(z_1|x)P(x) \\ &= k \times (0.45, 0.45, 0.1) \otimes (0.45, 0.45, 0.1) = (0.488, 0.488, 0.024) \end{aligned}$$

In summary, in the case where two successive observations are made $z_1$ of sensor A, the fusion algorithm of the information tends to increase the probability that one of the two targets is present at the expense of the hypothesis of no visible target.

From the observation of probability matrices of the two sensors it is evident that while sensor A is more effective in estimating the presence of a target but less to observe its presence, B sensor instead is more effective at distinguishing the type of target but less effective to say if the target is visible or not. So it is natural to use the information from both sensors and combine it in order to obtain more comprehensive information and better estimating if the target is visible or not and what kind of target is.

Let's suppose, therefore, having a uniform prior distribution as in the previous case. If we look $z_A = z_1$ with sensor A and $z_B = z_1$ with sensor B, the posterior distribution in $x$ will be

$$\begin{aligned} P(x|z_{A,1}, z_{B,1}) &= k \times P_{AB}(z_1, z_1|x) \times P(x) \\ &= k \times P_A(z_1|x)P_B(z_1|x) \\ &= k \times (0.45, 0.45, 0.1) \otimes (0.45, 0.1, 0.45) = (0.6924, 0.1538, 0.1538) \end{aligned}$$

Comparing this result with the result of the previous case in which two measurements are performed with sensor A is evident how B sensor provides substantial additional information on the discrimination of the type of target at the expense of a slight loss of effectiveness in the identification of the target.

Repeating the calculation for each pair of observations, results in the following probability matrices

$z_A = z_1$

$$\begin{bmatrix} z_B = & z_1 & z_2 & z_3 \\ x_1 & 0.692 & 0.154 & 0.488 \\ x_2 & 0.154 & 0.692 & 0.488 \\ x_3 & 0.154 & 0.154 & 0.024 \end{bmatrix}$$

$z_A = z_2$

$$\begin{bmatrix} z_B = & z_1 & z_2 & z_3 \\ x_1 & 0.692 & 0.154 & 0.488 \\ x_2 & 0.154 & 0.692 & 0.488 \\ x_3 & 0.154 & 0.154 & 0.024 \end{bmatrix}$$

$z_A = z_3$

$$\begin{bmatrix} z_B = & z_1 & z_2 & z_3 \\ x_1 & 0.108 & 0.024 & 0.265 \\ x_2 & 0.0241 & 0.108 & 0.265 \\ x_3 & 0.868 & 0.868 & 0.471 \end{bmatrix}$$

From which results that the overall performance of the observation system has improved in each of the possible cases. If fact, a type 2 target with the second sensor is observed, after having observed a type 1 target with the first sensor, the posterior distribution is $P(x|z_{A,1}, z_{B,2}) = (0.154, 0.692, 0.154)$ which improves the identification of target 2. If, as a further confirmation of the effectiveness, we do not observe any target with the second sensor and observe a type 1 target with the first sensor, the posterior distribution is $(0.488, 0.488, 0.024)$. I.e. it is still well estimated that one target is visible although it is not known which one, as the second sensor is not able to distinguish it. If then sensor A observed that no target is visible while B sensor observed Type 2 target, the posterior distribution is $(0.108, 0.024, 0.868)$, which indicates that in any case it is estimated that there is no target despite the observation of sensor B, which contains little information as it is not able to perform a good identification of the target but is only able to distinguish between the two types of target.

### 2.1.5. Bayesian recursive filter

The merger of the information of more observations or more sensors which uses Equation (2-18) would require, in principle, to store all past information and, upon arrival of new k-th information in the form $P(z_k|x)$, to recalculate the updated overall probability. Bayes theorem easily lends itself to the recursive or incremental implementation. Considering in fact that $Z^k \square \{z_k, Z^{k-1}\}$ can be written and assuming conditional independence of the observations,

$$\begin{aligned} P(x, Z^k) &= P(x|Z^k)P(Z^k) \\ &= P(z_k, Z^{k-1}|x)P(x) \\ &= P(z_k|x)P(Z^{k-1}|x)P(x) \end{aligned}$$

Equalling the terms at the extremes of this equation you obtain:

$$P(x|Z^k)P(Z^k) = P(z_k|x)P(Z^{k-1}|x)P(x)$$
$$= P(z_k|x)P(x|Z^{k-1})P(Z^{k-1})$$

Recalling that $P(Z^k)/P(Z^{k-1}) = P(z_k|Z^{k-1})$,

$$P(x|Z^k) = \frac{P(z_k|x)P(x|Z^{k-1})}{P(z_k|Z^{k-1})} \qquad (2\text{-}19)$$

The advantage of this recursive algorithm is that at each iteration step it is sufficient to calculate and store only the posterior distribution $P(x|Z^k)$ which contains a comprehensive summary of all past information. When you have a new observation and so a new probability function $P(z_k|x)$, the posterior distribution becomes the previous one and the two functions, multiplied together, become, after having normalized the product, the new posterior distribution.

**Example 4**

Wanting to apply the recursive Bayesian filter to the data fusion case of Example 1, the mean and variance of the posterior distribution updated at step k, can be calculated immediately and are as follows:

$$x_k = \frac{\sigma_{k-1}^2}{\sigma_{k-1}^2 + \sigma^2} z_k + \frac{\sigma^2}{\sigma_{k-1}^2 + \sigma^2} x_{k-1}$$

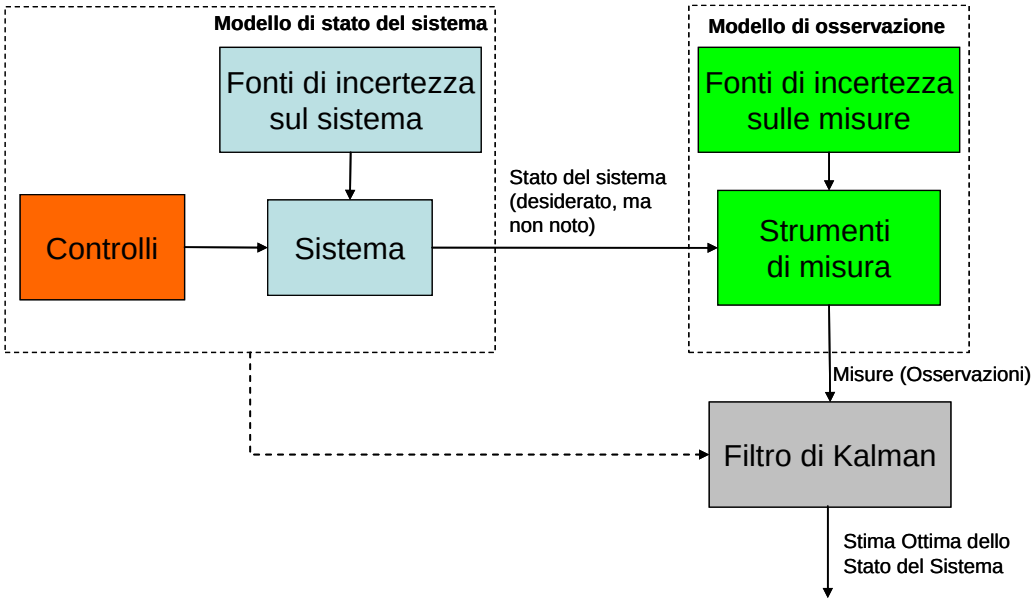
$$\sigma_k^2 = \frac{\sigma_{k-1}^2 \sigma^2}{\sigma_{k-1}^2 + \sigma^2} = \left( \frac{1}{\sigma_{k-1}^2} + \frac{1}{\sigma^2} \right)^{-1}$$

Where $\sigma^2$ is the variance associated with the new observation $z_k$, $\sigma_{k-1}^2$ is the variance associated to the previous distribution.

### 2.1.6. The Kalman filter

All data fusion problems include an estimation phase. Since a number of measures (observations) from a group of sensors, we want to use the information obtained to obtain an estimate that is better than each of the individual measures in probabilistic terms, of the state of interest. An estimator is a decision rule that, given as input a sequence of measurements of a quantity of interest, calculates the value of that quantity (parameter, state of a system, measuring). The Kalman filter is a *algorithm optimal data processing recursive, or recursive linear estimator*. The definition of good is referred, under the assumptions on which it is based, to the performance of the Kalman filter in respect to any other possible data fusion algorithm [10]. It processes all available measures, whatever their accuracy, to estimate the current value of the variables of interest. For example, suppose we want to measure the speed of a vehicle and to have the measurements from a laser triangulation or from a group of odometry sensors or from an inertial platform (accelerometers and gyroscopes). Rather than being forced to choose only one of these measures and ignore all the others, you can build a Kalman filter that takes input from all available measures and combines the information contained in each data, including the knowledge of the model of the instrument that supplied it. The output result will be a better estimate of the speed.

The term *recursive* is crucial for an implementation to the calculator of the Kalman filter, in fact it indicates that the algorithm does not require the storage and reworking of the whole data history every time a new measurement is available.




**Picture 2-7:** Block diagram of a possible application of the Kalman filter

Picture Picture 2-7 represents a typical application where you can use the Kalman filter. The observed system can be or not interlocked by some control input and the measuring instruments provide output information for the feedback, and in general to estimate certain variables of interest of the system. Knowledge of these inputs (controls) and outputs (measures) is the only thing available in the physical system that can be used to make an excellent estimate of the state of the system. Sources of uncertainty on the system or *model uncertainty* originate from incomplete knowledge of all the variables that affect the dynamics of the system. Sources of uncertainty on the measures include the imperfect knowledge of what are the relations between the state variables and the measured outputs (uncertainty of instrument model) and the uncertainty due to random and systematic errors of the various instruments. A method to extract useful information from noisy signals is therefore needed. There may be different instruments with different dynamic and accuracy characteristic which provide measures of the same variable; a method to combine the outputs of the various tools is desirable, if not necessary, to produce an estimate of the variables of interest in order to minimize errors in the statistical sense. The concept of optimal filter that defines the Kalman filter is understood in a statistical sense and it is as if to say that, if we apply the test filters many times for the same application, then the average of the results of the Kalman filter would be better than the average of the results of any test filter.

To minimize errors in a probabilistic sense, i.e., Bayesian sense, the filter should propagate the density of conditional probability of the measurand (quantity of interest), conditional to the knowledge of current measures of the instruments. For example, the probability density indicated as $f_{x(i)|\boldsymbol{z}(1),\boldsymbol{z}(2),\ldots,\boldsymbol{z}(i)}(x\,|\,\boldsymbol{z}_1,\boldsymbol{z}_2,\ldots,\boldsymbol{z}_i)$ is the function that describes the probability that the measurand (supposing it is scalar) $x \in X$ assumes a particular value within the $X$ (for example, $X \equiv$ ) at instant $i$ ( $x(i)$ ); this probability is conditional to the knowledge that the carrier of the measures $\boldsymbol{z}(1)$ at instant 1 takes the value $\boldsymbol{z}_1$ and so on for instant 2 up to $i$. Suppose, for example, that $x(i)$ is the one-dimensional position of a vehicle at an instant in time $i$, and that $\boldsymbol{z}(j)$ is the two-dimensional vector of measurements of the position at the time $j$ coming from two laser sensors. This density of conditional probability contains all the available information on $x(i)$: it indicates, for each given value of all measurements obtained up to the instant $i$, what is the probability that the measurand

has a given value or range of values $x(i)$. The shape of the graph of function $f$ contains all the information about the uncertainty on the value of $x$: if it is configured as a narrow and high peak then most content of probability is concentrated in a narrow range of values $x$, if it is instead configured with a little slanting trend, the uncertainty of $x$ is higher and its probability is distributed in a wider range of values.
The optimal estimate is obtained directly once the conditional density has been propagated as this contains the necessary information to choose an excellent value for the measurand. Possible candidates for the choice are *average*, the *trend* and the *median.* These values, in the case of the Kalman filter, coincide; in fact, the propagation of the conditional probability density takes place under the assumptions that:

- the system is described by a *linear model*
- the noise associated with the system and the measures is *white* and *Gaussian* (i.e., zero-mean, with spectral power density distributed uniformly over the entire band of frequencies and with Normal probability distribution).

Under these three restrictive assumptions, the Kalman filter is defined excellent in the sense described above.

*Too restrictive assumptions?*

Only at first sight the three assumptions on which we build the Kalman filter may seem too restrictive and not very close to reality.

The linear system model is justifiable for various reasons. Often a linear model is well suited for the purpose and, in the presence of non-linearity, the most widely used engineering approach is that of the linearization of the model around some system configuration, such as a point or a trajectory. Linear systems are more desirable because they are more easily managed with engineering tools and the theory of linear systems is much more comprehensive and practical than the non-linear ones. Then there are some methods to extend the application of the Kalman filter to non linear systems, where linear models should prove inadequate.

The white noise assumption implies that the noise values are not related to each other in time. Basically, if you know how much the noise is worth at the present instant, this does not add any information for the purpose of a forecast of what will be its value in another instant. A white noise has a spectrum with equal power density for all frequencies. This implies that such a signal has infinite power, so a white noise cannot exist in nature. However, since each physical system has a limited bandwidth in frequency space and is typically affected by a broadband noise, from the system point of view it is equivalent to assuming that there is white noise since the latter, within the passband of the system, is identical to the real broadband noise and outside of the passband the noise has no influence on the system. The white noise assumption greatly simplifies the mathematics used in the Kalman filter and does not add restrictions to the model of the system if this is affected by broadband noise. If, however, within the passband of the system the noise has non-uniform spectral density or is correlated in time, through an additional filter any form of correlated noise can be played, starting from a white noise, (the filter consists of a small linear system called "*shaping filter*").

While the "white" attribute for a noise is reported to its temporal characteristics (or frequency), the "Gaussian" attribute is reported to its amplitude. That is, for each time instant, the amplitude probability density of a Gaussian noise has the known form of a bell. This assumption is justified the physical sense by the fact that, typically, there is a large number of small sources that contribute to create the noise of measurement. Mathematically this phenomenon is described by the central limit theorem: the higher the number of independent random variables that are add up among them, whatever is the probability distribution of each, the probability distribution of the sum tends to be one Gaussian. In addition to this justification a practical one can be added. Usually of a variable affected by noise, the statistics of the first and second order are only known, respectively, the mean and the variance (or the standard deviation), and in that case, that is, when other statistics

of a higher order than the second are not known, there is no other distribution to assume that is better than the Gaussian one. A Gaussian is determined completely once the two mean and variance parameters are known, unlike many distributions that require the knowledge of an even infinite number of statistics. The Kalman filter, which propagates the mean and the variance, includes all the information contained in the conditional probability density rather than only some of it, as it would happen with another form of probability distribution. The assumptions on which the Kalman filter is based would be overly restrictive in some cases if the goal was to want to build a model of the system which is as close to reality as possible, therefore abandoning the assumptions of linearity and white noise and Gaussian. The aim is instead to build a good estimator algorithm (or control), in which case the assumptions made often allow to obtain a representation of the system which is suitable for the aim and to use easily manageable mathematical instruments. The need is therefore that the model is efficient to make a good estimate of the parameters of interest, and this requirement also affects the extension of the model in cases of greatest applicability and that the above mentioned assumptions are no longer valid.

### Example 5

This is a simple example of application of the Kalman filter to the case of the estimation of a single variable, suppose it is the one-dimensional position of a car or its position in the case of rectilinear trajectory; this example allows to treat the problem in an intuitive way also by including the dynamic case where the measurand varies in time.

Suppose you are travelling on a straight path and you do not know your position. At a certain moment $t_1$ you make an observation to $z_1$ estimate your position, more or less accurate, for example on the basis of knowledge, the time elapsed from the start, and the average speed. The result of this measure will certainly be affected by an uncertainty that is estimated and represented by the standard deviation (the statistic of the second order is the variance $\sigma_{z_1}^2$). There is therefore the availability of function $f_{x(t_1)|z(t_1)}(x \mid z_1)$ which represents the conditional probability distribution of the position at the time $t_1$, $x(t_1)$, conditioned to the observed value $z_1$. If the distribution is Gaussian, the standard deviation $\sigma_{z_1}$ is a direct measure of the uncertainty of measurement $z_1$, and it is such that the probability that the true value of x is within the range cantered at $z_1$ and with half-width $\sigma_{z_1}$ it is 68.3%. Speaking in terms of Bayesian probability, if it is assumed that the *previous distribution* of $x(t_1)$ is rectangular with equal probability for any value of $x$, the *posterior distribution* after the measurement (observation) $z_1$, will be Gaussian with the variance of the estimation:

$$\sigma_x^2(t_1) = \sigma_{z_1}^2$$

And the best estimate of the position will be:

$$\hat{x}(t_1) = z_1$$

Suppose then to dispose of the position measurement using GPS or, better yet, DGPS, at the time $t_2 \cong t_1$ (so that the position of the vehicle does not change to any appreciable extent), obtaining a measure $z_2$ with variance $\sigma_{z_2}^2$ less than $\sigma_{z_1}^2$. At this point you have a further observation of the state, that is, a second, more accurate, measurement of the position. As seen in the previous sections, a method for combining with each other, probabilistically, two measurements of the same variable with the respective probability distributions, is to blend them the Bayesian way. As seen in Example 1, when you consider the estimate made at instant $t_1$ as a previous estimate ($P(x)$) and you want to get a better estimate after the observation $t_2$, in order to obtain the posterior distribution, i.e. the updated estimate, just multiply between them the distribution function of the previous and the current observation. In the case of Gaussian distributions it is not necessary to multiply between them the original density functions, as the first and

second order statistics, the mean and variance, are obtained directly from the ones of the starting estimates. Alternatively you can apply the recursive filtering method according to Example 3. The result of the location of the probability density $x(t_2)$, conditioned to both measurements made at the instants $t_1 \cong t_2$, $z_1$ and $z_2$, is the Gaussian distribution with mean $\mu$ and variance $\sigma^2$ given by:

$$\mu = \frac{\sigma_{z_2}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2} z_1 + \frac{\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2} z_2$$

$$\sigma^2 = \left( \frac{1}{\sigma_{z_1}^2} + \frac{1}{\sigma_{z_2}^2} \right)^{-1}$$

The resulting variance is less than both the original variances which amounts to saying that the uncertainty in estimating the location decreased after the combination of the starting measures.
The best current estimate is:

$$\hat{x}(t_2) = \mu$$

with associated variance

$$\sigma_x^2(t_2) = \sigma^2 .$$

This estimate is the one with the highest probability, the optimized value to the ordinary least squares. Note that, if the measure $z_2$ is more accurate than $z_1$, that is that $\sigma_{z_2}^2 \square \sigma_{z_1}^2$, the variance of this combined estimate is however less than $\sigma_{z_2}^2$; it follows that, even if a measure contains little information, however, it increases the accuracy of the result of the filter.

By manipulating the last equation given above, we can write,

$$\hat{x}(t_2) = \frac{\sigma_{z_2}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2} z_1 + \frac{\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2} z_2$$

$$= z_1 + \frac{\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2} (z_2 - z_1)$$

and, remembering that $\hat{x}(t_1) = z_1$, in the classical form in which the Kalman filter is implemented:

$$\hat{x}(t_2) = \hat{x}(t_1) + K(t_2)[z_2 - \hat{x}(t_1)] \quad (2\text{-}20)$$

where

$$K(t_2) = \frac{\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2} \quad (2\text{-}21)$$

Looking at the last two equations we can see that the optimal estimate at the time $t_2$, $\hat{x}(t_2)$, is equal to the best prediction of its value before the measure is available $z_2$, $\hat{x}(t_1)$, added to a correction term which consists of the product of a coefficient of excellent combination for the difference between $z_2$ and the best prediction of its value before it is actually available, $\hat{x}(t_1)$. This filter therefore has a feature called "predictor-corrector". Based on all the above information, a forecast is made of the value that the variables of interest and the measures associated to them will have in the next measurement instant. When

the next measurement is done the difference between it and its expected value is used to correct the estimate of the variables of interest.

The variance of the optimal estimate can be rewritten as:

$$\sigma_x^2(t_2) = \sigma_x^2(t_1) - K(t_2)\sigma_x^2(t_1)$$

It is important to note that the values $\hat{x}(t_2)$ and $\sigma_x^2(t_2)$ include all the information of function $f_{x(t_2)|\boldsymbol{z}(t_1),\boldsymbol{z}(t_2)}(x \mid \boldsymbol{z}_1, \boldsymbol{z}_2)$, which is equivalent to say that the conditional probability density is obtained by propagating these two variables. Furthermore, the solution to this problem of static case fusion with the Kalman filter is exactly equivalent to the one which would be obtained by applying a recursive Bayesian filter.

We consider the same issue in the *dynamic case*.

Assume that the vehicle in question travels a certain distance before a new position measurement of any type is available. A constant speed state model for the system of the form is assumed:

$$\frac{dx}{dt} = v_0 + \xi$$

Where $v_0$ is a nominal speed and $\xi$ is a noise term used to represent the uncertainty in the knowledge of the actual value of the speed because of disturbance from which it is affected, of non-nominal conditions, of effects not considered in the first-order simple differential equation, etc. The noise $\xi$ is modelled as a Gaussian white noise with zero mean and variance $\sigma_\xi^2$.

The conditional probability density given $z_1$ and $z_2$, at the instant $t_2$ is identical to the one obtained above. As time passes while the vehicle is travelling at a nominal speed $v_0$, the uncertainty due to noise on the speed increases and the density function is a Gaussian one centred in the values of $x(t)$ gradually estimated by the state model, but is characterized by a greater variance, that is represented by an increasingly large and flattened bell. In other words, the probability density, imagined scrolling on the axis of $x$ as time increase as, starts at the instant $t_2$ from the optimal estimation and moves with nominal speed according to the dynamic model widening in time because we are less and less sure of the exact position of the vehicle owing to the accumulation of uncertainty in time. At the instant $t_3^-$, immediately before a new measure at the instant available $t_3$, the probability density $f_{x(t_3)|\boldsymbol{z}(t_1),\boldsymbol{z}(t_2)}(x \mid \boldsymbol{z}_1, \boldsymbol{z}_2)$ is a Gaussian one characterized by the following mean and variance:

$$\hat{x}(t_3^-) = \hat{x}(t_2) + v_0(t_3 - t_2) \qquad (2\text{-}22)$$

$$\sigma_x^2(t_3^-) = \sigma_x^2(t_3^-) + \sigma_\xi^2(t_3 - t_2) \qquad (2\text{-}23)$$

The average is obtained by simple integration of the nominal speed over time, depending on the model. The variance $\sigma_x^2(t_3^-)$ is obtained by propagating the variance of the optimal estimation $\sigma_x^2(t_2)$ between the instant $t_2$ and the instant $t_3$ or rather by calculating the uncertainty on the value of $\hat{x}(t_3^-)$ via the Kleine and McClintock formula applied to the first of the two equations written above and assuming that the time measurement is not affected by errors.

$\hat{x}(t_3^-)$ is the perfect **forecast** of which the value of $x$ will be at the instant $t_3^-$ before a measure at the instant is carried out $t_3$, and $\sigma_x^2(t_3^-)$ is the expected variance in this forecast.

At the instant $t_3$ a measurement is taken (for example with GPS) $z_3$ with variance $\sigma_{z_3}^2$. At this point you have two probability densities, one comprising all the previous information available before the last measurement was carried out, the other containing the information of the measurement itself. According to the assumptions made in §2.1.4 the measures are conditionally independent from each other. Similarly to the static case, the two conditional probability densities are combined still producing a Gaussian distribution with mean

$$\hat{x}(t_3) = \hat{x}(t_3^-) + K(t_3)[z_3 - \hat{x}(t_3^-)] \text{ ( 2-24)}$$

and variance

$$\sigma_x^2(t_3) = \sigma_x^2(t_3^-) - K(t_3)\sigma_x^2(t_3^-) \text{ ( 2-25)}$$

Where the advantage $K(t_3)$ it is given by

$$K(t_3) = \frac{\sigma_x^2(t_3^-)}{\sigma_x^2(t_3^-) + \sigma_{z_3}^2} \text{ ( 2-26)}$$

This second filter stage, in which we calculate the estimate and its variance downstream of all available information, that is, when you have the latest measure of its variance and the previous forecast of estimates and variance, is called **updating**.

The expression of optimal estimation $\hat{x}(t_3)$ in the dynamic case has the same shape as the one which is obtained in the static case, the difference between the two applications of the filter is that in the dynamic case one introduces the estimate obtained from the model of the system to make the prediction on the variables of interest and on the measures in the instants in which they are not available, in this time interval the uncertainty of optimal estimation, corrected at the instant in which a measure is obtained, is propagated until the instant at which a new measurement is available.

It is interesting to observe the expression of advantage $K(t_3)$. If $\sigma_{z_3}^2$, the variance of the noise measurement is large, then the advantage $K(t_3)$ is small; this simply means that what little confidence is given in a very noisy measure that therefore has little weight in the filter correction phase. If the limit $\sigma_{z_3}^2$ tended to infinity the filter advantage would be zero, the measure would be ignored and only the estimated model would be taken into account. Conversely, if the noise of the model estimate $\sigma_{\xi}^2$ has high variance, or if it has been too long since the last measurement even if the model is accurate, $\sigma_x^2(t_3^-)$ it will be high and so will the gain, therefore, a lot of confidence will be given to the measure. If at the limit $\sigma_x^2(t_3^-) \to \infty$ then $K(t_3) \to 1$ and therefore $\hat{x}(t_3) = \hat{x}(t_3^-) + 1[z_3 - \hat{x}(t_3^-)] = z_3$. In this case the model output is ignored and the optimal estimation coincides with the measurement.

### 2.1.6.1. The Kalman filter algorithm

It was said that the Kalman filter is a recursively linear estimator: it calculates, at consecutive instants, the estimate of a state (by measuring, with scalar or vector values) to constant values that evolves over time, based on periodic observations (measurements) of the state. The filter uses a statistical model of how the vector of interest, the state $x(t)$ evolves over time, and a statistical model of how the observations $z(t)$ are linked to this vector of interest. The first of these models is built from the system status model, linear or linearised (Kalman filter "extended"); the second statistical model is built starting from the observation model that includes the model of the measurement instrument /s employed. The advantages of the Kalman filter are chosen so that, under the assumptions made about the observations and the status process, the resulting estimate $\hat{x}(t)$ is such as to minimize the mean square error defined by

$$L(t) = \int_{-\infty}^{+\infty} (x(t) - \hat{x}(t))^T (x(t) - \hat{x}(t)) P(x(t) \mid Z^t) dx$$

If $L(t)$ differs compared to $\hat{x}(t)$ and the zero expression is set you get the value of $\hat{x}(t)$ which minimizes $L(t)$.

$$\frac{dL(t)}{d\hat{x}(t)} = \int_{-\infty}^{+\infty} \frac{d}{d\hat{x}(t)}[(x(t)-\hat{x}(t))^T(x(t)-\hat{x}(t))P(x(t)\mid Z^t)]dx = 0$$

$$\Rightarrow \int_{-\infty}^{+\infty} 2(\hat{x}(t)-x(t))P(x(t)\mid Z^t)]dx = 0 \Rightarrow \hat{x}(t)\int_{-\infty}^{+\infty} P(x(t)\mid Z^t)dx = \int_{-\infty}^{+\infty} x(t)P(x(t)\mid Z^t)dx$$

And since $\int_{-\infty}^{+\infty} P(x(t)\mid Z^t)dx = 1$,

$$\hat{x}(t) = \int_{-\infty}^{+\infty} x(t)P(x(t)\mid Z^t)dx$$

So the value obtained is not more than the conditioning average $E(x(t)\mid Z^t)$, that is, the expected value of the status at the instant $t$ conditioned to the sequence of observations $Z^t$ carried out up to that instant.

In §2.1.1 the state and observation in continuous time models were describes. However, the Kalman filter, being implemented in a calculator, is constructed in discrete steps of time in an iterative manner, so also the state and observation models must be transformed into discrete form prior to implementation of the filter. Assuming integration steps (sampling, for the measurement), constants, equations of model (2-5) and (2-6) are used.

The key assumption for the Kalman filter in which it is assumed that the noise of the process of evolution of the state and the measurement noise are in *zero mean* and *temporally uncorrelated* is translated respectively with the equations

$$E\{v_k\} = E\{w_k\} = 0, \ \forall k$$

$$E\{v(i)v(j)^T\} = \delta_{ij}\cdot Q(i), \quad E\{w(i)w(j)^T\} = \delta_{ij}\cdot R(i)$$

in which the operator $E\{\cdot\}$ is the one that calculates the expected value of a random variable (ie the average). $\delta_{ij}$ is the Kronecker operator and indicates that the *covariance matrices* $Q$ *and* $R$, of noise respectively $v$ and $w$, are non-void only if the expected value is calculated with respect to a single instant in time $i$ and not at two different instants.

It is indicated by the estimate of the state at the instant $k$ conditioned to all information obtained up to the instant $k$. The estimate status at the instant $k$ conditioned only to the information obtained up to the instant $k-1$ is called *forecast* a step forward and it is indicated with $\hat{x}(k\mid k-1)$ this symbolism is extended to all other parameters of the algorithm. It is assumed that at the k-th step the following parameters are known:

- the estimate $\hat{x}(k-1\mid k-1)$
- the conditional covariance (variance for scalar variables) $P(k-1\mid k-1)$

The Kalman filter consists in the recursion of two stages:

**Forecast**

A prediction is calculated $\hat{x}(k\mid k-1)$ of status at the instant $k$ and its covariance $P(k\mid k-1)$:

$$\hat{x}(k|k-1) = F(k)\hat{x}(k-1|k-1) + B(k)u(k) \quad (2\text{-}27)$$

$$P(k|k-1) = F(k)P(k-1|k-1)F^T(k) + G(k)Q(k)G^T(k) \quad (2\text{-}28)$$

**Updating**

An observation is made available $k$ at the instant $z(k)$. Therefore the updated estimate is calculated $\hat{x}(k|k)$ of the state $x(k)$ together with the updated covariance $P(k|k)$ from the prediction just calculated:

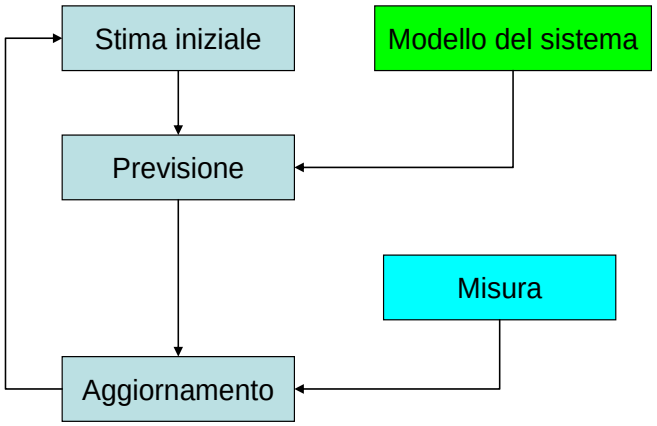
$$\hat{x}(k|k) = \hat{x}(k|k-1) + W(k)[z(k) - H(k)\hat{x}(k|k-1)] \quad (2\text{-}29)$$

$$P(k|k) = [1 - W(k)H(k)]P(k|k-1)[1 - W(k)H(k)]^T + W(k)R(k)W^T(k) \quad (2\text{-}30)$$

where the advantage matrix $W(k)$ is given by

$$W(k) = P(k|k-1)H(k)[H(k)P(k|k-1)H^T(k) + R(k)]^{-1} \quad (2\text{-}31)$$

The Kalman filter is a recursive loop (see Picture 2-8). The starting condition is an estimate obtained in some way through the knowledge that you have of measurand up to the initial instant of the cycle, after which a prediction is made, lastly the measurement to update the prediction and to produce a new estimate is used.




**Picture 2-8: Block diagram of the recursion of the Kalman filter**

The forecast is generated by the filter using the model of evolution of the measurand while the update makes use of the measurement model. The update stage is that of a typical linear filter of the first order: a weighted mean is calculated between the prediction and the measurement, in which the weight $W(k)$ is associated to $z(k)$ and the weight $1 - W(k)H(k)$ is associated to prediction $\hat{x}(k|k-1)$. Additionally, the filter propagates the covariance both in the forecast, by spreading the uncertainty due to the noise associated to the in measurand (a simple example is the equation (2-23)), and in updating the estimate, spreading uncertainty due to the noise associated to the measure.

### 2.1.6.2. Measurement of the performance of the Kalman filter

In order to monitor the performance of the Kalman filter a measure that is indicative of how well the filter is able to "follow" the evolution of the state in time must be used. It is possible to make a prediction of what observation will be done instantly $k$ based on observations made up to

the instant $k-1$ by calculating the expected value from the observation given from the model associated with it (Equation (2-4)) subject to previous observations:

$$\begin{aligned}\hat{z}(k\,|\,k-1) &= E\left\{z(k)\,|\,Z^{k-1}\right\}\\ &= E\left\{H(k)x(k)+W(k)\,|\,Z^{k-1}\right\}\\ &= E\left\{H(k)x(k)\,|\,Z^{k-1}\right\}\\ &= H(k)E\left\{x(k)\,|\,Z^{k-1}\right\}\\ &= H(k)\hat{x}(k\,|\,k-1)\end{aligned}$$

then, as you could guess, observation prediction in step $k$ is obtained simply by applying the model of observation of the state forecast.

The difference between the observation $z(k)$ and observation forecasting $H(k)\hat{x}(k\,|\,k-1)$ is called **innovation** or residue $\nu(k)$:

$$\rho(k) = z(k) - H(k)\hat{x}(k\,|\,k-1) \quad (2\text{-}32)$$

As it is not possible to directly compare the estimates obtained with the filter and the real state value, because very often this value is not known, innovation is an important measure of the deviation between the filter estimates and observations and can be used as a tool to evaluate the efficiency of the filter. In particular, for the purpose of testing the hypotheses on which the state and observation models are based and the same filter, you can check whether the main *properties of the innovations, are respected* that is that

$$E\left\{\rho(k)\,|\,Z^{k-1}\right\} = 0$$

$$E\left\{\rho(i)\rho^{T}(j)\right\} = \delta_{ij}S(i)$$

respectively the two equations indicate that the sequence of the innovations is white and uncorrelated in time. The matrix $S(k)$ and is called innovation covariance

$$S(k) = R(k) + H(k)P(k\,|\,k-1)H(k)^{T}$$

Innovation and innovation covariance can be used to express in an alternative and more concise way the update equations:

$$\hat{x}(k\,|\,k) = \hat{x}(k\,|\,k-1) + W(k)\rho(k)$$

$$P(k\,|\,k) = P(k\,|\,k-1) - W(k)S(k)W^{T}(k)$$

Where

$$W(k) = P(k\,|\,k-1)H(k)S^{-1}(k)$$

These newly written update equations have the same form as those of the simple continuous value scalar case of Example 5 (Equations (2-24, 2-25, 2-26)), in which the observation matrix is the

identity matrix, instead of the matrix $P$ appears $\sigma_x^2$, Instead of R appears $\sigma_z^2$, in place of the advantage matrix $W$ the advantage is shown $K$.

## 2.2. Alternatives to probability models

It is evident that the uncertainty representation is essential to provide the information on the observation of the state necessary for the fusion process. Several models of uncertainty representation have been proposed in alternative to the probabilistic ones to remedy the limitations of probabilistic methods.

From the literature four perceived limitations in modelling techniques according to the probabilistic approach can be seen:

- Complexity: It is necessary to specify a large number of probability to be able to apply correctly the methods of probabilistic reasoning
- Inconsistency: Because of the difficulties in specifying a coherent set of probabilistic estimates and use them to get a consistent deduction on the state of interest
- Accuracy of the models: The need to be accurate in the analytical definition of the probability for the amount of which there is little knowledge
- Uncertainty definition: The difficulty in assigning probabilities to the uncertainty on the source of information.

There are three main modelling techniques that try to address these limitations:

1. *Interval calculation*
2. *Fuzzy logic*
3. *Evidence theory*

Below there is a brief overview of these techniques and their advantages and disadvantages.

### 2.2.1.1. Interval calculation

In the technique of the intervals the uncertainty of a size $x$ is simply described by the statement that it is known that the true value of state $x$ is limited in a superior or inferior way, that is, that it belongs to an interval, $x \in [a,b]$. No other probabilistic information is added, in particular not even the hypothesis that $x$ is uniformly distributed within the interval (rectangular distribution). There are some basic rules for handling uncertainty intervals [11].

$$[a,b]+[c,d]=[a+c,b+d]$$
$$[a,b]-[c,d]=[a-d,b-c]$$
$$[a,b]\times[c,d]=[\min(ac,ad,bc,bd),\max(ac,ad,bc,bd)]$$
$$[a,b]/[c,d]=[a,b]\otimes[1/d,1/c],\quad 0\notin[c,d]$$

The operations of addition and multiplication of intervals are both associative and commutative.

The main advantages of this technique:

- the intervals provide a good measure in the case of scarcity of probabilistic information on the state
- they can be used as measures of detection of objects

The disadvantages:

- is difficult to obtain a convergence value for the measure, i.e. the estimate of the uncertainty intervals is too pessimistic
- it is difficult to represent the dependence between variables, which is important for many data fusion algorithms

### 2.2.1.2. Fuzzy Logic

Fuzzy Logic is widely used to represent the uncertainty especially in the supervision and sensor fusion algorithms to a hierarchically high level. There is a vast literature on this method [7].
Distinctive feature of the fuzzy reasoning, as well as a unique analytical tool available to it, are the *membership functions* (Membership functions).
If you consider a certain set of elements $X = \{x\}$ and a subset $A \subseteq X$ such that

$$A = \{x | x \text{ ha una specifica proprietà}\}$$

In conventional logic systems, a membership function $\mu_A(x)$ that determines whether a given element $x \in X$ is a member or not of this subset is defined:

$$A \quad \mu_A(x) = \begin{cases} 1 \ se \ x \in A \\ 0 \ se \ x \notin A \end{cases} \qquad (2\text{-}33)$$

In the fuzzy logic instead, a fuzzy set is defined in such a way that there is a degree of membership of the elements of the set, between 0 and 1. Therefore the membership function assigns a value between 0 and 1 that indicates the degree of membership of an element $x$ to a given set $A$.
Formally

$$A \quad \mu_A(x) \mapsto [0,1]$$

The dialling rules for fuzzy sets (fuzzy sets) follow the same composition processes of crisp sets (those described by Equation (2-33)), in particular

the AND operation is implemented as a minimum:

$$A \cap B \quad \mu_{A \cap B}(x) = \min([\mu_A(x), \mu_B(x)])$$

the OR operation is implemented as maximum:

$$A \cup B \quad \mu_{A \cup B}(x) = \max([\mu_A(x), \mu_B(x)])$$

the NOT operation is implemented as a complement:

$$\bar{A} \quad \mu_{\bar{A}}(x) = 1 - \mu_A(x)$$

### 2.2.1.3. Evidence theory

The evidence theory was born from the ideas of Dempster and Shafter. It is more complex than other methods cited in the sense that a probability may be associated not only to an element $x$ belonging to a set $X$ or a subset $A \subseteq X$, but also to sets of sets. While the dominion of probabilistic methods is the one of all possible sets, the dominion of evidence theory is the one of all sets of all subsets.

**Example 6**

Consider the set $X = \{piove, non\ piove\}$. In the theory of probability, by means of a forecast, we can assign a probability to each possible event. Eg, $P(piove) = 0.8$ and then $P(piove) = 0.2$. According to the evidence theory instead, the set of all subsets is built (called power set):

$$2^X = \{\{piove, non\ piove\}, \{piove\}, \{non\ piove\}, \{\ \}\}$$

A probability is assigned to each of the elements of this set

$$m(\{piove, non\ piove\}) = 0.2$$
$$m(\{piove\}) = 0.7$$
$$m(\{non\ piove\}) = 0.1$$
$$m(\{\ \}) = 0.0$$

To put it better, there is a 70% chance that it will rain, the 10% that it does not rain and 20% chance that it rains and does not rain together, i.e. a measure of ignorance that we have in making the prediction and distinguish between the two alternatives.

The evidence theory provides a method for quantifying the inability to distinguish between alternatives. In the probability theory this would result with assigning even probabilities to each alternative. But of course, to say that there is a 50% chance of rain is not equivalent to saying that it is unknown whether it rains or not. It can be immediately observed that this involves an increased complexity. For if the original set has $n$ elements, there will be $2^n$ possible sets of subsets each of which must be assigned a probability. For $n$ great this is unfeasible. In addition, if the set is continuous, the set of subsets are not measurable.

# 3. WORKS CITED


[1] J.O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer Verlag, 1985.J.

[2] S. Blackman and R. Popoli. *Design and Analysis of Modern Tracking Systems*. Artec House, 1999.

[3] Borenstein, L Feng, Measurement and Correction of Systematic Odometry Errors in Mobile Robots, IEEE Transactions on Robotics and Automation, Vol 12, No 5, Oct. 1996.

[4] M. De Cecco. Self-Calibration of AGV Inertial-Odometric Navigation Using Absolute-Reference Measurements. IEEE Instrumentation and Measurement Technology Conference, Anchorage, AK, USA, 21-23 May 2002.

[5] M De Cecco. Sensor fusion of inertial-odometric navigation as a function of the actual manoeuvres of autonomous guided vehicles. Measurement Science and Technology, vol. 14, pp 643-653, 2003.

[6] De Cecco M., Baglivo L., Angrilli F., Real-Time Uncertainty Estimation of Autonomous Guided Vehicles Trajectory taking into account Correlated and Uncorrelated Effects. Special Section in IEEE Transaction on Instrumentation & Measurement, accepted for publication on June 2007.

[7] D. Dubois and H. Prade. *Fuzzy Sets and Systems: Theory and Applications*. Academic Press, 1980

[8] *Guide to Expression of Uncertainty in Measurements* ISO 1993.
[9] L. Jetto, S. Longhi, D. Vitali, Localization of a wheeled mobile robot by sensor data fusion based on a fuzzy logic adapted Kalman filter, Control Engineering Practice, vol 7, pp. 763-771, 1999.
[10] Maybeck P.S., "The Kalman Filter. An introduction to concepts". In *Autonomous Robot Vehicles*, I.J. Cox, G.T Wilfong Editors, Springer-Verlag.
[11] R.E. Moore. *Interval Analysis*. Prentice Hall, 1966.
[12] S Shoval, I Zeitoun, E Lenz, Implementation of a Kalman filter in positioning for Autonomous Vehicles, and its Sensitivity to the Process Parameters, Int J Adv Manuf Technol, Vol 13, pp 738-746, 1997.
[13] R. C. Smith, P. Cheeseman, On the representation and Estimation of Spatial Uncertainty, int. J. Of Robotics Research, Vol 5, no 4, pp 56-68, 1986.