Kalman Filter: Multiplying Normal Distributions

Norbert Freier, mail@norbert-freier.de, ITVS, TU Dresden (17^{th} October 2013)

1 Abstract

This publication in the domain of sensor data fusion considers the Kalman filter, as one representative of fusion algorithms, in detail. Therefore, the widely spread description following [4] that is derived by orthogonal projections is presented first. Because this kind of explanation is hard to understand, especially by beginners, this paper introduces an other point of view. The new viewpoint is based on maximum likelihood methods. These methods use the product of probability density functions of different measurements to determine a likelihood function and its maximum as best fusion solution. Because the Kalman filter is based on normal distributions, these distributions are used in this paper as underlying probability density functions to calculate a maximum likelihood solution. For simplification, we start in the unidimensional space and calculate a likelihood function with the shape of a normal distribution. It is shown that the product of the underlying normal distributions is up to a constant factor a normal distribution as well and a Kalman filter equivalent solution also. To proof equality of the product and the Kalman filter solution in general, the problem is transferred to multiple dimensions. Finally an overview of the Kalman filter algorithm with variations is given using the multiplicative point of view.

2 Sensor Data Fusion

As stated in the abstract, this publication discusses sensor data fusion. The aim of data fusion is to determine one or more values of interest $\breve{x}_i$ as good as possible. The desired values are physical quantities in general. They are not named measurements, because later it is shown, that they do not need to be directly observable by a sensor. Values, occurring at one point in time t simultaneous, are cumulated in a state $\breve{\boldsymbol{x}}_t$. A state can be seen as a snapshot that freezes the specific quantities of the values of interest $\breve{x}_{t,i}$ at single point in time t.

To qualify a specific state, practical applications hold defective measurement devices only. So it has to be expected that the determined measurement $y_{t,j}$ contains an error $e_{t,j}$ always. Due these errors, the exact determination of a state is impossible. That is why the error containing measurements are used to determine an estimate $\boldsymbol{x}_t$, as good as possible, instead of the true state $\breve{\boldsymbol{x}}_t$ itself. To achieve this estimate, multiple measurements are combined proficient. This process is called fusion and an exemplary procedure is given by figure 1. A fusion algorithm uses measurements of multiple sensors of a single point in time or of a single sensor determining measurements as time goes by. Using a single sensor along time, it is necessary that the state $\breve{\boldsymbol{x}}$ is fixed. Otherwise it would not be possible to achieve a set of measurements belonging to exactly the same state. An alternative is, to take the occurring changes into account and calculate their influence on the estimate. A static state $\breve{\boldsymbol{x}}$ can not be guaranteed in most of the use cases generally. As a consequence, the majority of sensor data fusion algorithms can handle the shift of the state in time. One representative of these algorithms is the Kalman filter. It is introduced in detail by the following section.

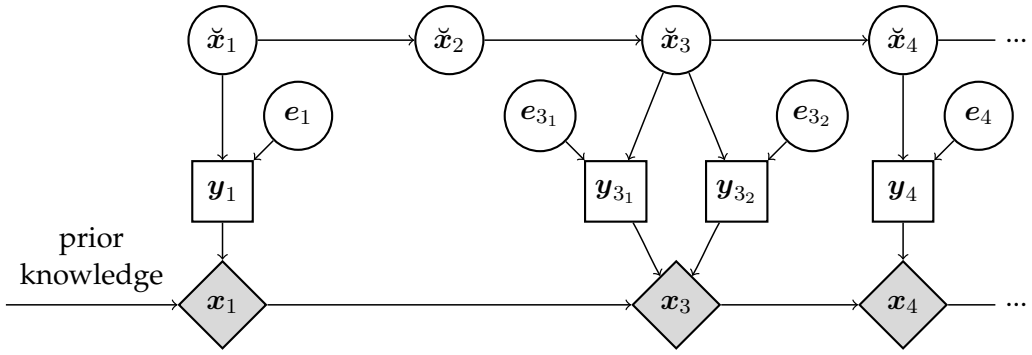

Figure 1: To achieve an estimate $\boldsymbol{x}_t$ of the true state $\breve{\boldsymbol{x}}_t$ at a specific point in time t, a sensor data fusion algorithm (gray) uses at least one observation $\boldsymbol{y}_t$. An observation $\boldsymbol{y}_t$ is a measurement of the true state $\breve{\boldsymbol{x}}_t$ and contains the error $\boldsymbol{e}_t$. Due to the usage of digital sensors, the measuring rate is bounded by a maximum frequency. Thats why not all states can be treated by a single sensor and do not cause a measurement. So is for example the second state $\breve{\boldsymbol{x}}_2$ not measurable. In contrast to this, the usage of multiple sensors allows to obtain two measurements of $\breve{\boldsymbol{x}}_3$. Changes in the state from $\breve{\boldsymbol{x}}_1$, at the first point in time, passing $\breve{\boldsymbol{x}}_2$ to $\breve{\boldsymbol{x}}_3$, at the third point in time, are used by the fusion algorithm, as good as possible to shift the knowledge of $\boldsymbol{x}_1$ in time. Because of that, the calculation of $\boldsymbol{x}_3$ is not only based the on two measurements $\boldsymbol{y}_{3_1}$ and $\boldsymbol{y}_{3_2}$. At the beginning, prior knowledge can be used for the initial estimate $\boldsymbol{x}_1$.

3 The Kalman Filter

The Kalman filter was published by R. E. Kalman in 1960 [4]. As a data fusion algorithm, it combines defective measurements to calculate a proper estimate of the state. The measurement influencing errors are seen as Gaussian white noise. So they are normally distributed around zero and errors of different points in time are assumed to be stochastically independent. In detail, a measurement $\boldsymbol{y}_t$ at the point in time t depends on the actual state $\breve{\boldsymbol{x}}_t$ and is affected by the error $\boldsymbol{e}_t$.

$$\boldsymbol{y}_t = \boldsymbol{H}_t * \breve{\boldsymbol{x}}_t + \boldsymbol{e}_t \tag{1}$$

Here $\boldsymbol{H}_t$ is the matrix of the actual observation function. $\boldsymbol{H}_t$ determines how the measurement $\boldsymbol{y}_t$ is gained from the true state $\breve{\boldsymbol{x}}_t$ at the point in time t. By the given equation, it can bee seen that a value of interest $\breve{x}_{t,i}$ not needs to be directly observable via $y_{t,i}$. Due to $\boldsymbol{H}_t$, a measurement can be a wighted sum of the components of $\breve{\boldsymbol{x}}_t$. However, the usage of $\boldsymbol{H}_t$ causes a limitation also. The observation function can only handle linear relationships between the real state $\breve{\boldsymbol{x}}_t$ and the observation $\boldsymbol{y}_t$. To extend the application field of the Kalman filter, week nonlinearities can be handled by the extended Kalman filter.

The advantage of the Kalman filter is the ability to calculate a proper estimate $\boldsymbol{x}$ of a variable state $\breve{\boldsymbol{x}}$ under usage of a single measuring device to generate observations. For this, a two stage iterative process that is shown in figure 2 is used. After an initial estimate, prediction and update steps alternate. Thereby observations, belonging to different steps in time, can be used as long as the shift in time of the real state is predictable and can be transferred on the estimate.

3.1 Prediction Step

Starting at the actual state estimate $\boldsymbol{x}_t$, the following state estimate $\hat{\boldsymbol{x}}_{t+1}$ is forecasted in the prediction step. This prediction tries to capture the real shift in time of the state from $\breve{\boldsymbol{x}}_t$ to $\breve{\boldsymbol{x}}_{t+1}$ in the state transition matrix $\boldsymbol{F}_t$. The estimated state transition is transferred on the estimate $\boldsymbol{x}_t$ to predict the next estimate $\hat{\boldsymbol{x}}_{t+1}$. Following [8] it is calculated as

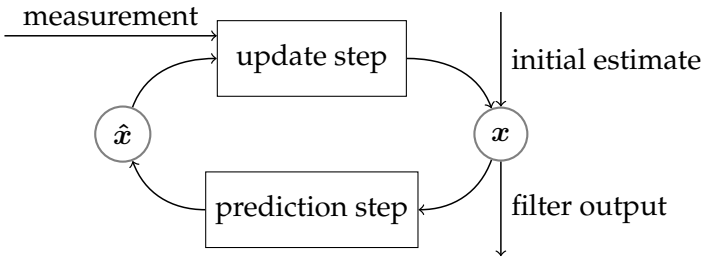

Figure 2: The Kalman filter algorithm starts with an initial estimate $\boldsymbol{x}$ of the state and an iterative process follows. During this process, prediction and update steps alternate. In the prediction step the change of the real state is forecasted and transferred on the estimate to calculate a prediction $\hat{\boldsymbol{x}}$ of the next estimate. The update step combines the predicted estimate and a measurement to gain an improved estimate.

$$\hat{\boldsymbol{x}}_{\boldsymbol{t+1}} = \underbrace{\boldsymbol{F}_t * \boldsymbol{x}_t}_{\substack{\text{deterministic} \\ \text{state transition}}} + \underbrace{\boldsymbol{B}_t * \boldsymbol{u}_t}_{\substack{\text{external} \\ \text{influence}}} + \underbrace{\boldsymbol{G}_t * \boldsymbol{w}_t}_{\substack{\text{model} \\ \text{error}}} . \tag{2}$$

Here is $\boldsymbol{u}_t$ the actual active control vector. With its help, the influence of external forces can be considered. This could be for example the usage of a cars brake to decelerate the vehicle. The matrix $\boldsymbol{B}_t$ spreads the external forces on the components of the predicted state estimate $\hat{\boldsymbol{x}}_{t+1}$. A break of a car has not only influence on the speed, the pitch angle is changed also. The matrix $\boldsymbol{G}_t$ acts equal to $\boldsymbol{B}_t$ but spreads uncertainties $\boldsymbol{w}_t$ in the model on the components of $\hat{\boldsymbol{x}}_{t+1}$. By that, simplifications and errors of the model $\boldsymbol{F}_t$ can be taken into consideration. Model errors occur when the shift in time of the real state is reproduced insufficient by $\boldsymbol{F}_t$. This inaccuracy leads to a changed state estimate prediction also. User intended model simplifications have the same effect. Because of the fact that errors are not determinable in general and simplifications are intended, both are left out of the prediction. Only the forecast quality is affected negative and the degradation has to be taken into account later. So the equation can be abbreviated to

$$\hat{\boldsymbol{x}}_{t+1} = \boldsymbol{F}_t * \boldsymbol{x}_t + \boldsymbol{B}_t * \boldsymbol{u}_t. \tag{3}$$

For this description you should keep in mind that the Kalman filter works with normal distributions only. Measurements are not seen as a single value. They underlie Gaussian white noise and are seen as a whole probability density distribution. To keep the normal distributions through the calculation process, the Kalman filter uses linear transformations only. The requirement of linearity restricts the set of possible transition and observation functions. As a result, reality can be modeled defective only. This has a negative influence on the estimate.

Despite this obvious weakness, the Kalman filter is very popular. A reason could be found in the simple evaluation of the estimation quality. It is given by the covariance of the underlying normal distribution. A low covariance means that the Kalman filter has achieved good solution.

The state estimate $\boldsymbol{x}_t$ and its forecast $\hat{\boldsymbol{x}}_{t+1}$ are the mean value vectors $\boldsymbol{\mu}_{x_t}$ and $\boldsymbol{\mu}_{\hat{x}_{t+1}}$ of the used normal distributions. The probability density function of a normal distribution unifies mean, mode and median in one point. So none of them has to be chosen as best estimate (see [5]).

Beside the estimated value itself, the accuracy of the estimation is necessary also. It is given by the covariance matrix of the normal distribution. The lower the covariance is the better is estimate. Not only the mean value needs to be forecasted by the prediction step. Also the accuracy as covariance matrix $\hat{\boldsymbol{P}}_{t+1}$ of the predicted estimate $\hat{\boldsymbol{x}}_{t+1}$ needs to be calculated. It

is gained from the covariance matrix $\boldsymbol{P}_t$ of the actual estimate $\boldsymbol{x}_t$. The covariance has to be calculated to get a complete normal distribution again.

$$\hat{\boldsymbol{P}}_{t+1} = \boldsymbol{F}_t * \boldsymbol{P}_t * \boldsymbol{F}_t^T + \boldsymbol{Q}_t \tag{4}$$

Errors in the control vector $\boldsymbol{u}_t$ and inaccuracies in the model $\boldsymbol{F}_t$ are considered by $\boldsymbol{Q}_t$. As opposed to the prediction of the estimate, model errors and simplifications have to be considered by the accuracy calculation. So the covariance of the error, occurring in the prediction step, has to be estimated and integrated into the accuracy of the estimation. Thats why model errors do not influence the prediction itself but their accuracy. The estimation accuracy decreases in the prediction step in general. Not only model errors are responsible for that, also combinations of the estimation errors so far can lead in the prediction step to an accuracy decrease.

3.2 Update Step

After calculating the prediction of the estimate an update step follows. In this update step the predicted estimate is corrected by a new measurement. The so improved estimate $\boldsymbol{x}_{t+1}$ is the filter output of the next point in time. The update step weights the prediction and the measurement by their covariance matrices. The more accurate a value is, the more influence it exerts over the adjustment of the estimate. Therefore a weighting factor $\boldsymbol{K}_t$, named Kalman gain, is calculated following [7] as

$$\boldsymbol{K}_{t+1} = \hat{\boldsymbol{P}}_{t+1} * \boldsymbol{H}_{t+1}^T * \left[\boldsymbol{H}_{t+1} * \hat{\boldsymbol{P}}_{t+1} * \boldsymbol{H}_{t+1}^T + \boldsymbol{R}_{t+1}\right]^{-1}. \tag{5}$$

Using the Kalman gain, the new state estimate $\boldsymbol{x}_{t+1}$ of the next point in time is calculated from the predicted estimate $\hat{\boldsymbol{x}}_{t+1}$ [7].

$$\boldsymbol{x}_{t+1} = \hat{\boldsymbol{x}}_{t+1} + \boldsymbol{K}_{t+1} * \left(\boldsymbol{y}_{t+1} - \boldsymbol{H}_{t+1} * \hat{\boldsymbol{x}}_{t+1}\right) \tag{6}$$

Finally the accuracy of the updated estimate is calculated as covariance matrix $\boldsymbol{P}_{t+1}$ [7].

$$\boldsymbol{P}_{t+1} = \hat{\boldsymbol{P}}_{t+1} - \boldsymbol{K}_{t+1} * \boldsymbol{H}_{t+1} * \hat{\boldsymbol{P}}_{t+1} \tag{7}$$

3.3 Iterative Process

After the first iteration of the Kalman filter that consists of a prediction and update step the following iterations can further improve the state estimate. Following Welch and Bishop [7], the algorithm can be summarized as figure 3 shows.

This level of detail is sufficient for the following work on the Kalman filter. Further information about and the derivation of the filter can be found in literature like [4], [7], [5] or [6].

4 Multiplicative Point of View

After presenting the Kalman filter in a traditional manner, an other point of view is introduced by the following section. For simplification, we use the one-dimensional space for the introduction. The estimated state $\breve{x}$ is set to be fixed and is except for an error e_t directly observable. Under these assumptions it can be shown commonly that the hereinafter introduced multiplicative solution equals the Kalman filter solution in the unidimensional space.

Later on, ideas gained in this first proof are used to show equality in the multidimensional space also. For that, the state $\breve{\boldsymbol{x}}$ is still set to be fixed and assumed to be directly observable. Afterwards, neglecting direct observability, it will be evaluated if the multiplicative point of

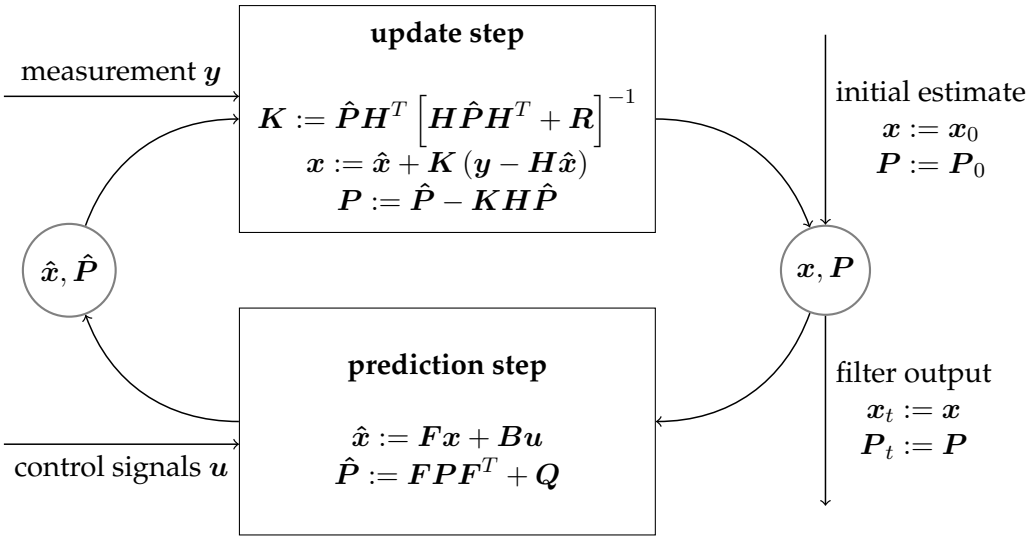

Figure 3: An overview of the single steps with their calculations within the Kalman filter, following [7]. Starting with the initial estimate $\boldsymbol{x}_0$ with the accuracy $\boldsymbol{P}_0$ at the point in time 0, the state estimate $\boldsymbol{x}$ with its accuracy $\boldsymbol{P}$ is calculated by an iterative process. During this process, a prediction and update step alternate. The calculated estimate prediction $\hat{\boldsymbol{x}}$ with its accuracy $\hat{\boldsymbol{P}}$ is corrected by a measurement to achieve an improved state estimate through the update step. At the end of each iteration step, the Kalman filter emits a valid state estimate $\boldsymbol{x}_t$ with its accuracy $\boldsymbol{P}_t$ at the point in time t.

view can handle indirect observations. Finally, an alternative description of the Kalman filter will be given, for a dynamic state, using the multiplicative point of view.

In the introduction it is stated, that the multiplicative point of view will be presented as product of normal distributions. Because of that, it is necessary to show which normal distributed values are used. Analogous to the Kalman filter, the measurement error $\boldsymbol{e}_t$ is assumed to be normally distributed here as well. So no bias occurs and the error can be written as

$$\boldsymbol{e}_t \sim N(0, \boldsymbol{\Sigma}_{e_t}). \tag{8}$$

The resulting normal distribution is described by the covariance matrix $\boldsymbol{\Sigma}_{e_t}$ that is first of all determined by the quality of the measuring device. The observation $\boldsymbol{y}_t$ is the sum of he real actual state $\breve{\boldsymbol{x}}_t$ and an error $\boldsymbol{e}_t$. So the observation $\boldsymbol{y}_t$ is normally distributed also. However the environment can still have an influence on the measurement. So $\boldsymbol{\Sigma}_{e_t}$ needs not to be constant in time for one measuring device. Thats why the covariance matrix $\boldsymbol{\Sigma}_{y_t} = \boldsymbol{\Sigma}_{e_t}$ is given for each observation separately.

$$\boldsymbol{y}_t \sim N(\breve{\boldsymbol{x}}_t, \boldsymbol{\Sigma}_{e_t}) = N(\breve{\boldsymbol{x}}_t, \boldsymbol{\Sigma}_{y_t}) \tag{9}$$

This perspective is based on $\boldsymbol{y}_t = \breve{\boldsymbol{x}}_t + \boldsymbol{e}_t$ and $\boldsymbol{y}_t \sim \breve{\boldsymbol{x}}_t + N(0, \boldsymbol{\Sigma}_{e_t})$. Because of the validity of $\breve{\boldsymbol{x}}_t = \boldsymbol{y}_t - \boldsymbol{e}_t$ and the symmetry of normal distributions relative to their mean value, the probability of a state to occur, under the condition that the measurement $\boldsymbol{y}_t$ is taken, is $\breve{\boldsymbol{x}}_t \sim \boldsymbol{y}_t + N(0, \boldsymbol{\Sigma}_{e_t}) = N(\boldsymbol{y}_t, \boldsymbol{\Sigma}_{y_t})$. This leads to a probability density distribution $P(\breve{\boldsymbol{x}}_t | \boldsymbol{y}_t)$ of the state $\breve{\boldsymbol{x}}_t$ depending on the actual observation $\boldsymbol{y}_t$. To estimate the state via a maximum likelihood method, the point of the maximum of this conditional probability density distribution is used as best estimate. The accuracy of this estimate is given by $\boldsymbol{\Sigma}_{y_t}$.

$$\boldsymbol{x}_t = \boldsymbol{\mu}_{y_t} = \boldsymbol{y}_t \qquad\qquad \boldsymbol{\Sigma}_{x_t} = \boldsymbol{\Sigma}_{y_t} = \boldsymbol{\Sigma}_{e_t} \tag{10}$$

So the state can be estimated by a single measurement only. But this estimate is as accurate as the used measurement. To increase the accuracy, sensor data fusion algorithms use a set of measurements. For each of these measurements a likelihood function can be given in principle. This leads, using the maximum as estimate, to a separate estimate for each measurement. But the wanted estimate is the maximum of the combined likelihood function. To achieve a likelihood function of all measurements in combination, the underlying likelihood functions have to be unified. The here chosen approach is to calculate the combined function as product of the functions derived from the measurements. This multiplicative point of view is shown next. Following, equivalence to the Kalman filter is proofed.

In the next proofs the three steps of the Kalman filter are discussed individually. These steps are the calculation of the initial estimate, the prediction and the update step. It is shown that the used initial estimate and the prediction are equivalent to the initialization and update step of the Kalman filter. After that, more work has to be spend on the update step. The proof of equivalence is more complex than the proofs of the initialization or the prediction step. So a set of assumptions is used in the beginning to proof equivalence. In further steps we get rid of these assumptions to reach generality.

4.1 Initialization

As initial estimate $\boldsymbol{x}_1$ the maximum likelihood solution $\boldsymbol{\mu_{y_1}}$ of the first measurement $\boldsymbol{y}_1$ is used without loss of generality. Alternatively prior knowledge can be used as long as the resulting estimate is normally distributed. Satisfying the previous assumption, prior knowledge can be seen as a pseudo measurement. The accuracy of the first estimate is given as covariance matrix $\boldsymbol{\Sigma}_{x_1}$. It is equal to the accuracy $\boldsymbol{\Sigma}_{y_1}$ of the first measurement or of prior knowledge. The first measurement can be chosen as initial estimate for the Kalman filter also.

$$\boldsymbol{x}_1 = \boldsymbol{\mu}_{x_1} = \boldsymbol{\mu}_{y_1} = \boldsymbol{y}_1 \qquad\qquad \boldsymbol{\Sigma}_{x_1} = \boldsymbol{\Sigma}_{y_1} = \boldsymbol{\Sigma}_{e_1} \tag{11}$$

4.2 Prediction Step for Static States

For simplification, the estimated real state $\breve{\boldsymbol{x}}$ is set to be static. In the prediction step the real shift in time of the state $\breve{\boldsymbol{x}}_t$ to $\breve{\boldsymbol{x}}_{t+1}$ is captured. It is transferred on the estimate $\boldsymbol{x}_t$ to calculate a prediction $\hat{\boldsymbol{x}}_{t+1}$ for the next point in time $t+1$. Thats why the estimate is assumed as static also. So the state transition matrix $\boldsymbol{F}_t$ is at each point the identity matrix $\boldsymbol{I}$ and a control vector $\boldsymbol{u}_t$ does not exist.

$$\hat{\boldsymbol{x}}_{t+1} = \boldsymbol{F}_t * \boldsymbol{x}_t + \boldsymbol{B}_t * \boldsymbol{u}_t = \boldsymbol{x}_t \tag{12}$$

$$\boldsymbol{\mu}_{\hat{x}_{t+1}} = \boldsymbol{\mu}_{x_t} \tag{13}$$

Because of the static real state $\breve{\boldsymbol{x}}$ and the absence of control inputs, the real shift could be reproduced by the prediction step without an error. So the accuracy $\boldsymbol{\Sigma}_{x_t}$ of the estimate is the accuracy of the predicted estimate $\boldsymbol{\Sigma}_{\hat{x}_{t+1}}$. A prediction calculated this way can be derived by the Kalman filter under usage of $\boldsymbol{F}_t = \boldsymbol{I}$ also.

$$\hat{\boldsymbol{P}}_{t+1} = \boldsymbol{F}_t * \boldsymbol{P}_t * \boldsymbol{F}_t^T + \boldsymbol{Q}_t \tag{14}$$

$$\boldsymbol{\Sigma}_{\hat{x}_{t+1}} = \boldsymbol{\Sigma}_{x_t} \tag{15}$$

4.3 Prediction Step for Dynamic States

The assumption of a static state can not be fulfilled in general. To find a Kalman filter equivalent solution at all, an other prediction step has to be found. If changes occur between $\breve{\boldsymbol{x}}_t$ and

$\breve{\boldsymbol{x}}_{t+1}$ they need to bee applied to the prediction. Control inputs have to be taken into account also. Analogous to the prediction step of the static state, the Kalman filter prediction step is used here. Because of the fact that control signals $\boldsymbol{u}_t$ underlie an error and can be determined inaccurate only, they are seen as mean value vector $\boldsymbol{\mu}_{u_t}$ after their transformation by $\boldsymbol{B}_t$. Model and control signal errors are seen as normally distributed. Their common covariance matrix is $\boldsymbol{\Sigma}_{Q_t}$. Using this for the Klaman filter, leads to

$$\begin{aligned} \hat{\boldsymbol{x}}_{t+1} &= \boldsymbol{F}_t * \boldsymbol{x}_t + \boldsymbol{B}_t * \boldsymbol{u}_t = \boldsymbol{x}_t & (16)\\ \boldsymbol{\mu}_{\hat{x}_{t+1}} &= \boldsymbol{F}_t * \boldsymbol{\mu}_{x_t} + \boldsymbol{\mu}_{u_t} & (17) \end{aligned}$$

and

$$\begin{aligned} \hat{\boldsymbol{P}}_{t+1} &= \boldsymbol{F}_t * \boldsymbol{P}_t * \boldsymbol{F}_t^T + \boldsymbol{Q}_t & (18)\\ \boldsymbol{\Sigma}_{\hat{x}_{t+1}} &= \boldsymbol{F}_t * \boldsymbol{\Sigma}_{x_t} * \boldsymbol{F}_t^T + \boldsymbol{\Sigma}_{Q_t}. & (19) \end{aligned}$$

4.4 Update Step in the Unidimensional Space with Direct Observability

For simplification, equivalence of the multiplicative point of view and the Kalman filter update step is shown in the one dimensional space first. Thats why the true state $\breve{x}_t$ consists of a single value. It is constant in time and can be observed directly, except an error e_t. So $y_t = \breve{x}_t + e_t$ is valid for an observation. The occurring error e_t is seen as Gaussian white noise.

The Kalman filter combines in the update step the predicted estimate $\hat{x}_{t+1}$ and the measurement y_{t+1} to get an improved estimate x_{t+1}. The incline of accuracy can be seen by the decline of the variance $\sigma^2_{x_{t+1}}$. The variance of the corrected estimate is smaller than the variance $\sigma^2_{y_{t+1}}$ of the measurement and even smaller than the variance $\sigma^2_{\hat{x}_{t+1}}$ of the old estimate. So the Kalman filter update step calculates a new normal distribution by combining the normal distributions of the predicted estimate and the measurement.

$$N(\mu_{x_{t+1}}, \sigma^2_{x_{t+1}}) = fusion\left(N(\mu_{\hat{x}_{t+1}}, \sigma^2_{\hat{x}_{t+1}}), N(\mu_{y_{t+1}}, \sigma^2_{y_{t+1}})\right) \quad (20)$$

To show that the product of the two normal distributions $N(\mu_{\hat{x}_{t+1}}, \sigma^2_{\hat{x}_{t+1}})$ and $N(\mu_{y_{t+1}}, \sigma^2_{y_{t+1}})$ is equivalent to the solution of the Kalman filter update step, it has to be proved first that a new normal distribution is the solution of the product. The next step is, to show that this normal distribution is equal to the solution of the Kalman filter update step.

Theorem 1. *The product of two normal distribution density functions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ around their mean values μ_1, μ_2 with the variances σ_1^2, σ_2^2 is up to a constant factor c equal to the normal distribution density function $N(\mu, \sigma^2)$ around the mean value μ with the variance σ^2.*

$$N(\mu_1, \sigma_1^2) * N(\mu_2, \sigma_2^2) = c * N(\mu, \sigma^2) \quad (21)$$

Proof. The following proof of equality is dived into two steps. In the first step, the exponent a of the calculated normal distribution $N(\mu, \sigma^2)$ will be found. For that μ and σ are calculated. Thats why the new exponent can be written as sum of the existing exponents a_1 and a_2.

$$\begin{aligned} \frac{1}{\sigma_1\sqrt{2\pi}} * e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} * \frac{1}{\sigma_2\sqrt{2\pi}} * e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2} &= c * \frac{1}{\sigma\sqrt{2\pi}} * e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\\ \frac{1}{\sigma_1\sqrt{2\pi}} * e^{a_1} * \frac{1}{\sigma_2\sqrt{2\pi}} * e^{a_2} &= c_1 * \frac{1}{\sigma\sqrt{2\pi}} * e^{(c_2+a)}\\ a_1 + a_2 &= c_2 + a \end{aligned}$$

In the second step, the constant scaling factor c belonging to the new exponent a is calculated. Thereafter universal equality will be shown. As later can be seen, the scaling factor c consists of two parts. Thats why $c = c_1 * e^{c_2}$ is used here.

Step 1: Calculation of μ, σ and c_2 by comparing the exponents

$$
\begin{aligned}
c_2 - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 &= -\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - \frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2 \\
-2 * c_2 + \frac{x^2 - 2x\mu + \mu^2}{\sigma^2} &= \frac{x^2 - 2x\mu_1 + \mu_1^2}{\sigma_1^2} + \frac{x^2 - 2x\mu_2 + \mu_2^2}{\sigma_2^2} \\
&= \frac{\sigma_2^2 x^2 - 2\sigma_2^2 x\mu_1 + \sigma_2^2\mu_1^2 + \sigma_1^2 x^2 - 2\sigma_1^2 x\mu_2 + \sigma_1^2\mu_2^2}{\sigma_1^2\sigma_2^2} \\
&= \frac{\left(\sigma_2^2 + \sigma_1^2\right)x^2 - 2x\left(\sigma_2^2\mu_1 + \sigma_1^2\mu_2\right) + \sigma_2^2\mu_1^2 + \sigma_1^2\mu_2^2}{\sigma_1^2\sigma_2^2} \\
&= \frac{\left(\sigma_2^2 + \sigma_1^2\right)x^2 - 2x\left(\sigma_1^2 + \sigma_2^2\right)\left(\frac{\sigma_2^2\mu_1 + \sigma_1^2\mu_2}{\sigma_1^2 + \sigma_2^2}\right) + \left(\sigma_1^2 + \sigma_2^2\right)\left(\frac{\sigma_2^2\mu_1^2 + \sigma_1^2\mu_2^2}{\sigma_1^2 + \sigma_2^2}\right)}{\sigma_1^2\sigma_2^2} \\
&= \frac{x^2 - 2x\left(\frac{\sigma_2^2\mu_1 + \sigma_1^2\mu_2}{\sigma_1^2 + \sigma_2^2}\right) + \left(\frac{\sigma_2^2\mu_1^2 + \sigma_1^2\mu_2^2}{\sigma_1^2 + \sigma_2^2}\right)}{\left(\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right)} \\
&= \frac{x^2 - 2x\left(\frac{\sigma_2^2\mu_1 + \sigma_1^2\mu_2}{\sigma_1^2 + \sigma_2^2}\right) + \left(\frac{\sigma_2^2\mu_1^2 + \sigma_1^2\mu_2^2}{\sigma_1^2 + \sigma_2^2}\right)}{\left(\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right)} + \frac{\overbrace{\left(\frac{\sigma_2^2\mu_1 + \sigma_1^2\mu_2}{\sigma_1^2 + \sigma_2^2}\right)^2}^{=0}}{\left(\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right)} - \frac{\left(\frac{\sigma_2^2\mu_1 + \sigma_1^2\mu_2}{\sigma_1^2 + \sigma_2^2}\right)^2}{\left(\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right)} \\
&= \frac{x^2 - 2x\left(\frac{\sigma_2^2\mu_1 + \sigma_1^2\mu_2}{\sigma_1^2 + \sigma_2^2}\right) + \left(\frac{\sigma_2^2\mu_1 + \sigma_1^2\mu_2}{\sigma_1^2 + \sigma_2^2}\right)^2}{\left(\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right)} + \frac{\left(\frac{\sigma_2^2\mu_1^2 + \sigma_1^2\mu_2^2}{\sigma_1^2 + \sigma_2^2}\right)}{\left(\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right)} - \frac{\left(\frac{\sigma_2^2\mu_1 + \sigma_1^2\mu_2}{\sigma_1^2 + \sigma_2^2}\right)^2}{\left(\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right)} \\
-2 * c_2 + \underbrace{\frac{x^2 - 2x\mu + \mu^2}{\sigma^2}}_{a \text{ of } N(\mu,\sigma^2)} &= \underbrace{\frac{x^2 - 2x\mu + \mu^2}{\sigma^2}}_{a \text{ of } N(\mu,\sigma^2)} + \frac{\sigma_2^2\mu_1^2 + \sigma_1^2\mu_2^2}{\sigma_1^2\sigma_2^2} - \frac{\left(\sigma_2^2\mu_1 + \sigma_1^2\mu_2\right)^2}{\sigma_1^2\sigma_2^2\left(\sigma_1^2 + \sigma_2^2\right)} \\
-2 * c_2 &= \frac{\left(\sigma_1^2 + \sigma_2^2\right) * \left(\sigma_2^2\mu_1^2 + \sigma_1^2\mu_2^2\right) - \left(\sigma_2^2\mu_1 + \sigma_1^2\mu_2\right)^2}{\sigma_1^2\sigma_2^2\left(\sigma_1^2 + \sigma_2^2\right)} \\
&= \frac{\sigma_1^4\mu_2^2 + \sigma_1^2\sigma_2^2\mu_1^2 + \sigma_1^2\sigma_2^2\mu_2^2 + \sigma_2^4\mu_1^2 - \sigma_1^4\mu_2^2 - 2\sigma_1^2\sigma_2^2\mu_1\mu_2 - \sigma_2^4\mu_1^2}{\sigma_1^2\sigma_2^2\left(\sigma_1^2 + \sigma_2^2\right)} \\
&= \frac{\left(\sigma_1^2\sigma_2^2\right) * \left(\mu_1^2 + \mu_2^2 - 2\mu_1\mu_2\right)}{\sigma_1^2\sigma_2^2\left(\sigma_1^2 + \sigma_2^2\right)} \\
&= \frac{\mu_1^2 + \mu_2^2 - 2\mu_1\mu_2}{\left(\sigma_1^2 + \sigma_2^2\right)} \\
c_2 &= -\frac{\left(\mu_1 - \mu_2\right)^2}{2 * \left(\sigma_1^2 + \sigma_2^2\right)}
\end{aligned}
$$

Except of a constant summand c_2, the exponent a has the style of a normal distribution density functions exponent. The summand c_2 is later extracted from the exponent and can be found as constant factor in front of the resulting normal distribution. So the product differs only by a constant factor from the target normal distribution. Thereby the factor c_2 compensates the

distance $|\mu_2 - \mu_1|$ between the two mean values to keep the unbounded integral of the resulting function to one. Here μ, the mean value, and σ^2, the variance of the new normal distribution, are calculated as:

$$\mu = \frac{\sigma_2^2\mu_1 + \sigma_1^2\mu_2}{\sigma_1^2 + \sigma_2^2} \qquad \sigma^2 = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \tag{22}$$

Note that this equations are already known from Maybeck [5]. Even though a proof is missing, he states that this is the maximum likelihood estimate. Also others have examined this equations, e.g. for human cue integration, see [1], [2].

Step 2: Calculation of the Constant Factor c_1

Even if the two normal distributions used in the product and the target distribution are spread around the same mean value with $\mu_1 = \mu_2 = \mu$ a correction has to be made. The used factor c_1 is calculated in the following section.

$$\begin{aligned}
c * N(\mu, \sigma^2) = c_1 * e^{c_2} * N(\mu, \sigma^2) &= N(\mu_1, \sigma_1^2) * N(\mu_2, \sigma_2^2) \\
c_1 * e^{c_2} * \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} &= \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} * \frac{1}{\sigma_2\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2} \\
c_1 * \frac{1}{\sigma\sqrt{2\pi}} &= \frac{1}{\sigma_1\sqrt{2\pi}} * \frac{1}{\sigma_2\sqrt{2\pi}} \\
c_1 * \frac{1}{\frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}}\sqrt{2\pi}} &= \frac{1}{\sigma_1\sigma_2 2\pi} \\
c_1 &= \frac{1}{\sqrt{2\pi\left(\sigma_1^2 + \sigma_2^2\right)}}
\end{aligned}$$

In this second part the target distribution also differs by a constant factor c_1 from the product of the normal distributions. The square root can be used in the factors calculation because the values of a normal distribution are always positive. Thats why the constant factor is positive as well. As it can be seen, the factor depends only on σ_1^2 and σ_2^2, thats why it is constant. The factor keeps the unbounded integral to one when all normal distributions have the same mean value.

Finally both parts are merged. The constant factor c is the product of both factors c_1 and e^{c_2}.

$$N(\mu_1, \sigma_1^2) * N(\mu_2, \sigma_2^2) = \underbrace{\frac{1}{\sqrt{2\pi\left(\sigma_1^2 + \sigma_2^2\right)}} * e^{-\frac{(\mu_1-\mu_2)^2}{2\left(\sigma_1^2+\sigma_2^2\right)}}}_{c} * N\left(\left[\frac{\sigma_2^2\mu_1 + \sigma_1^2\mu_2}{\sigma_1^2 + \sigma_2^2}\right], \left[\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right]\right) \tag{23}$$

□

Knowing that the product of two normal distributions is except of a constant factor again a normal distribution, it can be proofed that the so gained solution is equal to the Kalman filter solution, see [5] also. For consistent naming, the Kalman filter equations are rewritten as normal distributions. Because of the direct observability H_{t+1} is 1. The multiplicative calculation generates a likelihood function but only the point of its maximum has our interest. The constant factor does not change the point of the maximum. Thats why this factor is neglected.

$$\mu_1 = \hat{x}_{t+1} \quad \sigma_1^2 = \hat{P}_{t+1} \qquad \mu_2 = y_{t+1} \quad \sigma_2^2 = R_{t+1} \qquad \mu = x_{t+1} \quad \sigma^2 = P_{t+1}$$

$$
\begin{aligned}
K_{t+1} &= \hat{P}_{t+1} H_{t+1}^T \left[H_{t+1} \hat{P}_{t+1} H_{t+1}^T + R_{t+1} \right]^{-1} \\
K_{t+1} &= \sigma_1^2 \left[\sigma_1^2 + \sigma_2^2 \right]^{-1} = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}
\end{aligned}
$$

$$
\begin{aligned}
x_{t+1} &= \hat{x}_{t+1} + K_{t+1} * (y_{t+1} - H_{t+1} \hat{x}_{t+1}) \\
\mu_{kalman} &= \mu_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} * (\mu_2 - \mu_1)
\end{aligned}
$$

$$
\begin{aligned}
P_{t+1} &= \hat{P}_{t+1} - K_{t+1} * H_{t+1} * \hat{P}_{t+1} \\
\sigma_{kalman}^2 &= \sigma_1^2 - \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} * \sigma_1^2.
\end{aligned}
$$

Theorem 2. *The value μ_{kalman} and accuracy σ_{kalman}^2 of the Kalman filter solution are the same as the value μ and the accuracy σ^2 of the multiplicative solution.*

$$
\frac{\sigma_2^2 \mu_1 + \sigma_1^2 \mu_2}{\sigma_1^2 + \sigma_2^2} = \mu = \mu_{kalman} = \mu_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} * (\mu_2 - \mu_1)
$$

and

$$
\frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} = \sigma^2 = \sigma_{kalman}^2 = \sigma_1^2 - \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} * \sigma_1^2
$$

Proof.

$$
\begin{aligned}
\frac{\sigma_2^2 \mu_1 + \sigma_1^2 \mu_2}{\sigma_1^2 + \sigma_2^2} &= \mu_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} * (\mu_2 - \mu_1) \\
&= \frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2 + \sigma_2^2} * \mu_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} * \mu_2 - \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} * \mu_1 \\
&= \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} * \mu_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} * \mu_2 \\
\frac{\sigma_2^2 \mu_1 + \sigma_1^2 \mu_2}{\sigma_1^2 + \sigma_2^2} &= \frac{\sigma_2^2 \mu_1 + \sigma_1^2 \mu_2}{\sigma_1^2 + \sigma_2^2}
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} &= \sigma_1^2 - \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} * \sigma_1^2 \\
&= \frac{\left(\sigma_1^2 + \sigma_2^2\right) * \sigma_1^2 - \sigma_1^4}{\sigma_1^2 + \sigma_2^2} \\
&= \frac{\sigma_1^4 + \sigma_2^2 * \sigma_1^2 - \sigma_1^4}{\sigma_1^2 + \sigma_2^2} \\
\frac{\sigma_2^2 \sigma_1^2}{\sigma_1^2 + \sigma_2^2} &= \frac{\sigma_2^2 \sigma_1^2}{\sigma_1^2 + \sigma_2^2}
\end{aligned}
$$

□

For the one-dimensional case with direct observability, it could be shown that the multiplicative solution calculates the same improved state estimate as the Kalman filter does. As well, the accuracies of both algorithms update steps are equal. Comparing both methods as a whole, it can be seen that they are equal at all. They are initialized with the same values, prediction and update step are equal. Because that and the fact that both methods only consist of these steps, they gain the same solution with the same accuracy.

4.5 Update Step in the Multidimensional Space with Direct Observability

After equality of both methods is shown in the one-dimensional space, this is now proofed for the multidimensional space. The state $\breve{\boldsymbol{x}}$ is set to be static and except of an error $\boldsymbol{e}_t$ directly observable. So an observation can be defined as $\boldsymbol{y}_t = \boldsymbol{x}_t + \boldsymbol{e}_t$. The measurement error $\boldsymbol{e}_t = N(0, \Sigma_{e_t})$ is seen as Gaussian white noise. This leads to a likelihood function in the form of an multivariate normal distribution. Following [3] this normal distribution can be written as:

$$N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{d}{2}} \, |\boldsymbol{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})} \tag{24}$$

The shape of this function is determined by the mean value vector $\boldsymbol{\mu} = \boldsymbol{\mu}_{y_{t+1}}$ and by the covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{y_{t+1}}$. Here is d the dimensionality and $|\boldsymbol{\Sigma}|$ the determinant of the covariance matrix.

In opposite to the unidimensional space, commutativity can in general not be seen as given for the matrix multiplications of the update step. This complicates the search for $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ a lot. Thats why these values are derived from the Klaman filter in the following section. Later it is checked if they are suitable as multiplicative solution. The proof used here is vice versa to the proof used in the unidimensional space. The past proof derived the solution from multiplication and checked equality to the Kalman filter.

4.5.1 Suitable Representation of the Mean Value Vector

Starting with the solution $\boldsymbol{\mu}_{Kalman}$ of the Kalman filter, the multiplicative $\boldsymbol{\mu}$ is derived. To keep it understandable, the mean value vectors $\boldsymbol{\mu_1} = \hat{\boldsymbol{x}}_{\boldsymbol{t+1}}$, $\boldsymbol{\mu_2} = \boldsymbol{y}_{t+1}$ and covariance matrices $\boldsymbol{\Sigma_1} = \hat{\boldsymbol{P}}_{\boldsymbol{t+1}}$, $\boldsymbol{\Sigma_2} = \boldsymbol{R}_{t+1}$ are used.

$$\begin{aligned}
\boldsymbol{K}_{t+1} &= \hat{\boldsymbol{P}}_{\boldsymbol{t+1}} \boldsymbol{H}_t^T \left[\boldsymbol{H}_{t+1} \hat{\boldsymbol{P}}_{t+1} \boldsymbol{H}_{t+1}^T + \boldsymbol{R}_{t+1} \right]^{-1} \\
\boldsymbol{K}_{t+1} &= \boldsymbol{\Sigma}_1 \left[\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 \right]^{-1}
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{x}_{t+1} &= \hat{\boldsymbol{x}}_{t+1} + \boldsymbol{K}_{t+1} \left(\boldsymbol{y}_{t+1} - \boldsymbol{H}_{t+1} \hat{\boldsymbol{x}}_{t+1} \right) \\
\boldsymbol{\mu}_{kalman} &= \boldsymbol{\mu}_1 + \boldsymbol{K}_{t+1} \left(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1 \right)
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{\mu} &= \boldsymbol{\mu}_{kalman} \\
&= \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_1 \left[\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 \right]^{-1} \left[\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1 \right] \\
&= \left[\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 \right] \left[\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 \right]^{-1} \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_1 \left[\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 \right]^{-1} \left[\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1 \right] \\
&= \boldsymbol{\Sigma}_1 \left[\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 \right]^{-1} \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_2 \left[\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 \right]^{-1} \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_1 \left[\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 \right]^{-1} \boldsymbol{\mu}_2 - \boldsymbol{\Sigma}_1 \left[\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 \right]^{-1} \boldsymbol{\mu}_1 \\
\boldsymbol{\mu} &= \boldsymbol{\Sigma}_2 \left[\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 \right]^{-1} \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_1 \left[\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 \right]^{-1} \boldsymbol{\mu}_2
\end{aligned}$$

The mean value vector $\boldsymbol{\mu}$ seems to be suitable for the further proof. Also $\boldsymbol{\mu}^T$ is needed and derived from $\boldsymbol{\mu}$ in the following section. Therefore $\boldsymbol{\mu}^T$ is transferred into a $\boldsymbol{\mu}$ equivalent form.

For this transformation the symmetry of the covariance matrices is used. The symmetry can be written as $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_1^T$ and $\mathbf{\Sigma}_2 = \mathbf{\Sigma}_2^T$.

$$\begin{aligned}
\boldsymbol{\mu}^T &= \left(\mathbf{\Sigma}_2 \left[\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right]^{-1} \boldsymbol{\mu}_1 + \mathbf{\Sigma}_1 \left[\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right]^{-1} \boldsymbol{\mu}_2\right)^T \\
&= \left(\mathbf{\Sigma}_2 \left[\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right]^{-1} \boldsymbol{\mu}_1\right)^T + \left(\mathbf{\Sigma}_1 \left[\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right]^{-1} \boldsymbol{\mu}_2\right)^T \\
&= \boldsymbol{\mu}_1^T \left(\left[\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right]^{-1}\right)^T \mathbf{\Sigma}_2^T + \boldsymbol{\mu}_2^T \left(\left[\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right]^{-1}\right)^T \mathbf{\Sigma}_1^T \\
\boldsymbol{\mu}^T &= \boldsymbol{\mu}_1^T \left[\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right]^{-1} \mathbf{\Sigma}_2 + \boldsymbol{\mu}_2^T \left[\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right]^{-1} \mathbf{\Sigma}_1
\end{aligned}$$

4.5.2 Suitable Representation of the Covariance Matrix

Analogue to the mean value vector $\boldsymbol{\mu}$, a suitable representation of the covariance matrix $\mathbf{\Sigma}$ is necessary. Because of the in $\mathbf{\Sigma}$ included inverse sum $\left[\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right]^{-1}$, a proof of equivalence is used instead of a direct transformation from the Kalman filter solution. Therefore the relation $\mathbf{\Sigma}\mathbf{\Sigma}^{-1} = \boldsymbol{I}$ for a invertable covariance matrix $\mathbf{\Sigma}$ is used. Equality $\mathbf{\Sigma} = \mathbf{\Sigma}_{kalman}$ of both solutions is assumed.

$$\begin{aligned}
\boldsymbol{P}_{t+1} &= \hat{\boldsymbol{P}}_{t+1} - \boldsymbol{K}_{t+1}\boldsymbol{H}_{t+1}\hat{\boldsymbol{P}}_{t+1} \\
\mathbf{\Sigma}_{kalman} &= \mathbf{\Sigma}_1 - \boldsymbol{K}_{t+1}\mathbf{\Sigma}_1
\end{aligned}$$

Theorem 3. *The later used cavariance matrix $\mathbf{\Sigma} = \mathbf{\Sigma}_1 \left[\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right]^{-1} \mathbf{\Sigma}_2$ is equal to the solution of the Kalman filter update step $\mathbf{\Sigma}_{kalman}$.*

$$\mathbf{\Sigma}_1 \left[\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right]^{-1} \mathbf{\Sigma}_2 = \mathbf{\Sigma} = \mathbf{\Sigma}_{kalman} = \mathbf{\Sigma}_1 - \mathbf{\Sigma}_1 \left[\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right]^{-1} \mathbf{\Sigma}_1$$

Proof.

$$\begin{aligned}
\boldsymbol{I} &= \mathbf{\Sigma}^{-1} * \mathbf{\Sigma}_{kalman} \\
&= \left(\mathbf{\Sigma}_1 \left[\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right]^{-1} \mathbf{\Sigma}_2\right)^{-1} * \left(\mathbf{\Sigma}_1 - \mathbf{\Sigma}_1 \left[\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right]^{-1} \mathbf{\Sigma}_1\right) \\
&= \mathbf{\Sigma}_2^{-1} \left[\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right] \mathbf{\Sigma}_1^{-1} * \left(\mathbf{\Sigma}_1 - \mathbf{\Sigma}_1 \left[\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right]^{-1} \mathbf{\Sigma}_1\right) \\
&= \mathbf{\Sigma}_2^{-1} \left[\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right] * \left(\mathbf{\Sigma}_1^{-1}\mathbf{\Sigma}_1 - \mathbf{\Sigma}_1^{-1}\mathbf{\Sigma}_1 \left[\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right]^{-1} \mathbf{\Sigma}_1\right) \\
&= \mathbf{\Sigma}_2^{-1} \left[\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right] * \left(\boldsymbol{I} - \boldsymbol{I} \left[\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right]^{-1} \mathbf{\Sigma}_1\right) \\
&= \left(\mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_2\right) * \left(\boldsymbol{I} - \left[\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right]^{-1} \mathbf{\Sigma}_1\right) \\
&= \left(\mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1 + \boldsymbol{I}\right) * \left(\boldsymbol{I} - \left[\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right]^{-1} \mathbf{\Sigma}_1\right) \\
&= \mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1 + \boldsymbol{I} - \mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1 \left[\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right]^{-1} \mathbf{\Sigma}_1 - \left[\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right]^{-1} \mathbf{\Sigma}_1 \\
\boldsymbol{I} - \boldsymbol{I} &= \mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1 \left[\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right]^{-1} \mathbf{\Sigma}_1 - \left[\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right]^{-1} \mathbf{\Sigma}_1 \\
\mathbf{0} &= \mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1 \left[\left(\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right)^{-1} \mathbf{\Sigma}_1\right]^{-1} - \mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1\boldsymbol{I} - \boldsymbol{I} \\
&= \mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1\mathbf{\Sigma}_1^{-1} \left(\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right) - \mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1 - \boldsymbol{I} \\
&= \mathbf{\Sigma}_2^{-1} I \left(\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right) - \mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1 - \boldsymbol{I} \\
&= \mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_2 - \mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1 - \boldsymbol{I} \\
&= \mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1 + \boldsymbol{I} - \mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1 - \boldsymbol{I} \\
&= \mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1 \\
\mathbf{0} &= \mathbf{0}
\end{aligned}$$

□

Beside the found solution, an other representation of $\boldsymbol{\Sigma}$ is necessary for the later proof of equality. The second representation is derived from the first, using the symmetries $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_1^T$ and $\boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_2^T$.

Theorem 4.

$$\boldsymbol{\Sigma}_1 \left[\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2\right]^{-1} \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_2 \left[\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2\right]^{-1} \boldsymbol{\Sigma}_1$$

Proof.

$$\begin{aligned}
\boldsymbol{\Sigma} &= \boldsymbol{\Sigma}^T \\
&= \left(\boldsymbol{\Sigma}_1 \left[\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2\right]^{-1} \boldsymbol{\Sigma}_2\right)^T \\
&= \boldsymbol{\Sigma}_2^T \left(\left[\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2\right]^{-1}\right)^T \boldsymbol{\Sigma}_1^T \\
&= \boldsymbol{\Sigma}_2 \left[\boldsymbol{\Sigma}_1^T + \boldsymbol{\Sigma}_2^T\right]^{-1} \boldsymbol{\Sigma}_1 \\
\boldsymbol{\Sigma}_1 \left[\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2\right]^{-1} \boldsymbol{\Sigma}_2 &= \boldsymbol{\Sigma}_2 \left[\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2\right]^{-1} \boldsymbol{\Sigma}_1
\end{aligned}$$

□

Both representations of $\boldsymbol{\Sigma}$ are in the unidimensional space equal to the known σ^2. So it seems that they can be used in further proofs.

4.5.3 Multiplicative Point of View Using μ and Σ

Knowing the promising representations of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, it can be shown that in the multidimensional space the product of two normal distributions is except of a constant factor a normal distribution.

Theorem 5. *The product of two multivariate normal probability density distribution functions $N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ around their mean value vectors $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ with the covariance matrices $\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$ is up to an constant factor $\mathbf{c}$ equal to the multivariate normal probability density distribution function $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ around the mean value vector $\boldsymbol{\mu}$ with the covariance matrix $\boldsymbol{\Sigma}$.*

$$N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) * N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2) = c * N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

Proof. Here we start with the exponents also. The sum of the two exponents is a normal distribution exponent again.

$$\begin{aligned}
&-\frac{1}{2}\left(\boldsymbol{x} - \boldsymbol{\mu}_1\right)^T \boldsymbol{\Sigma}_1^{-1} \left(\boldsymbol{x} - \boldsymbol{\mu}_1\right) \\
&-\frac{1}{2}\left(\boldsymbol{x} - \boldsymbol{\mu}_2\right)^T \boldsymbol{\Sigma}_2^{-1} \left(\boldsymbol{x} - \boldsymbol{\mu}_2\right) &= -\frac{1}{2}\left(\boldsymbol{x} - \boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{x} - \boldsymbol{\mu}\right) + c_2
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{x}^T\boldsymbol{\Sigma}_1^{-1}\boldsymbol{x} + \boldsymbol{x}^T\boldsymbol{\Sigma}_2^{-1}\boldsymbol{x} &\quad & \boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x} \\
-\boldsymbol{x}^T\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{x}^T\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2 &= & -\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\
-\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}_1^{-1}\boldsymbol{x} - \boldsymbol{\mu}_2^T\boldsymbol{\Sigma}_2^{-1}\boldsymbol{x} &\quad & -\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x} \\
+\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2^T\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2 &\quad & +\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - 2 * c_2
\end{aligned}$$

$$
\begin{array}{rcl}
\boldsymbol{x}^T\boldsymbol{\Sigma}_1^{-1}\left([\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\right)^{-1}[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\boldsymbol{x} & & \\
+\boldsymbol{x}^T\boldsymbol{\Sigma}_2^{-1}\left([\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\right)^{-1}[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\boldsymbol{x} & & \boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x} \\
-\boldsymbol{x}^T\boldsymbol{\Sigma}_1^{-1}\left([\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\right)^{-1}[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\boldsymbol{\mu}_1 & & \\
-\boldsymbol{x}^T\boldsymbol{\Sigma}_2^{-1}\left([\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\right)^{-1}[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\boldsymbol{\mu}_2 & = & -\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\
-\boldsymbol{\mu}_1^T[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\left([\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\right)^{-1}\boldsymbol{\Sigma}_1^{-1}\boldsymbol{x} & & \\
-\boldsymbol{\mu}_2^T[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\left([\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\right)^{-1}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{x} & & -\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x} \\
+\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}_1^{-1}\left([\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\right)^{-1}[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\boldsymbol{\mu}_1 & & \\
+\boldsymbol{\mu}_2^T\boldsymbol{\Sigma}_2^{-1}\left([\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\right)^{-1}[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\boldsymbol{\mu}_2 & & +\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}-2*c_2
\end{array}
$$

$$
\begin{array}{rcl}
\boldsymbol{x}^T\boldsymbol{\Sigma}_1^{-1}\left([\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\right)^{-1}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_2[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\boldsymbol{x} & & \\
+\boldsymbol{x}^T\boldsymbol{\Sigma}_2^{-1}\left([\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\right)^{-1}\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_1[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\boldsymbol{x} & & \boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x} \\
-\boldsymbol{x}^T\boldsymbol{\Sigma}_1^{-1}\left([\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\right)^{-1}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_2[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\boldsymbol{\mu}_1 & & \\
-\boldsymbol{x}^T\boldsymbol{\Sigma}_2^{-1}\left([\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\right)^{-1}\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_1[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\boldsymbol{\mu}_2 & = & -\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\
-\boldsymbol{\mu}_1^T[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_2^{-1}\left([\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\right)^{-1}\boldsymbol{\Sigma}_1^{-1}\boldsymbol{x} & & \\
-\boldsymbol{\mu}_2^T[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_1^{-1}\left([\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\right)^{-1}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{x} & & -\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x} \\
+\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}_1^{-1}\left([\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\right)^{-1}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_2[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\boldsymbol{\mu}_1 & & \\
+\boldsymbol{\mu}_2^T\boldsymbol{\Sigma}_2^{-1}\left([\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\right)^{-1}\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_1[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\boldsymbol{\mu}_2 & & +\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}-2*c_2
\end{array}
$$

$$
\begin{array}{rcl}
\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_2[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\boldsymbol{x}+\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_1[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\boldsymbol{x} & & \boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x} \\
-\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_2[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\boldsymbol{\mu}_1-\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_1[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\boldsymbol{\mu}_2 & = & -\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\
-\boldsymbol{\mu}_1^T[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}^{-1}\boldsymbol{x}-\boldsymbol{\mu}_2^T[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}^{-1}\boldsymbol{x} & & -\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x} \\
+\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_2[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\boldsymbol{\mu}_1+\boldsymbol{\mu}_2^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_1[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\boldsymbol{\mu}_2 & & +\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}-2*c_2
\end{array}
$$

Validity of this equation is proofed by testing equality of each line.

1. line

$$
\begin{array}{rcl}
\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_2[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\boldsymbol{x}+\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_1[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}\boldsymbol{x} & = & \boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x} \\
\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_2[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1}+\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_1[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2]^{-1} & = & \boldsymbol{\Sigma}^{-1} \\
\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_2\boldsymbol{I}+\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_1\boldsymbol{I} & = & \boldsymbol{\Sigma}^{-1}[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2] \\
\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_2+\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_1 & = & \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_2
\end{array}
$$

2. line

$$
\begin{array}{rcl}
-\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_2\left[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2\right]^{-1}\boldsymbol{\mu}_1 & & \\
-\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_1\left[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2\right]^{-1}\boldsymbol{\mu}_2 & = & -\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\
& = & -\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Sigma}_2\left[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2\right]^{-1}\boldsymbol{\mu}_1+\boldsymbol{\Sigma}_1\left[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2\right]^{-1}\boldsymbol{\mu}_2\right) \\
\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_2\left[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2\right]^{-1}\boldsymbol{\mu}_1 & & \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_2\left[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2\right]^{-1}\boldsymbol{\mu}_1 \\
+\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_1\left[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2\right]^{-1}\boldsymbol{\mu}_2 & = & +\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_1\left[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2\right]^{-1}\boldsymbol{\mu}_2
\end{array}
$$

3. line

$$
\begin{array}{rcl}
-\boldsymbol{\mu}_1^T\left[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2\right]^{-1}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}^{-1}\boldsymbol{x} & & \\
-\boldsymbol{\mu}_2^T\left[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2\right]^{-1}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}^{-1}\boldsymbol{x} & = & -\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x} \\
& = & -\left(\boldsymbol{\mu}_1^T\left[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2\right]^{-1}\boldsymbol{\Sigma}_2+\boldsymbol{\mu}_2^T\left[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2\right]^{-1}\boldsymbol{\Sigma}_1\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{x} \\
\boldsymbol{\mu}_1^T\left[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2\right]^{-1}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}^{-1} & & \boldsymbol{\mu}_1^T\left[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2\right]^{-1}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}^{-1} \\
+\boldsymbol{\mu}_2^T\left[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2\right]^{-1}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}^{-1} & = & +\boldsymbol{\mu}_2^T\left[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2\right]^{-1}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}^{-1}
\end{array}
$$

4. line

$$
\begin{array}{rcl}
\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_2\left[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2\right]^{-1}\boldsymbol{\mu}_1+\boldsymbol{\mu}_2^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_1\left[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2\right]^{-1}\boldsymbol{\mu}_2 & = & \boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}-2*c_2 \\
+\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_2\left[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2\right]^{-1}\boldsymbol{\mu}_1+\boldsymbol{\mu}_2^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_1\left[\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2\right]^{-1}\boldsymbol{\mu}_2-\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} & = & -2*c_2
\end{array}
$$

The fourth line is in the multidimensional case a constant $\boldsymbol{c_2}$ also. This constant has no influence on the place of the maximum or on the covariances. It is neglected by the maximum likelihood solution. This holds for the constant c_1 also. It is a scaling factor outside of the exponent.

$$
\begin{array}{rcl}
N(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_1)*N(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_2) & = & c*N(\boldsymbol{\mu},\boldsymbol{\Sigma}) \\
\frac{1}{(2\pi)^{\frac{d}{2}}\left|\boldsymbol{\Sigma}_1\right|^{\frac{1}{2}}}e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)}*\frac{1}{(2\pi)^{\frac{d}{2}}\left|\boldsymbol{\Sigma}_2\right|^{\frac{1}{2}}}e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu_2})^T\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_2)} & = & \\
\frac{1}{(2\pi)^{\frac{d}{2}}\left|\boldsymbol{\Sigma}_1\right|^{\frac{1}{2}}}*\frac{1}{(2\pi)^{\frac{d}{2}}\left|\boldsymbol{\Sigma}_2\right|^{\frac{1}{2}}}*e^{c_2}*e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})} & = & \\
c_1*e^{c_2}*\frac{1}{(2\pi)^{\frac{d}{2}}\left|\boldsymbol{\Sigma}\right|^{\frac{1}{2}}}e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})} & = & c*N(\boldsymbol{\mu},\boldsymbol{\Sigma})
\end{array}
$$

□

So we have shown that the multiplicative solution is a normal distribution in the multidimensional case also. The update step of the Kalman filter generates the same solution with the same accuracy. Because of the fact that both methods are initialized equal and have the same prediction step, the whole methods are equivalent at all.

4.6 Update Step in the Multidimensional Space with Indirect Observability

In the sections before the state $\breve{\boldsymbol{x}}_t$ was seen as constant in time. It is except of an error $\boldsymbol{e}_t$ directly observable by a measurement $\boldsymbol{y}_t$. So the observation matrix $\boldsymbol{H}_t$ is the identity matrix $\boldsymbol{I}$. But direct observability is not given always. So it has to be checked if the multiplicative point of view can handle $\boldsymbol{H}\neq\boldsymbol{I}$.

Therefore we have to distinguish the state and observation space. Variables in the state space have the dimensionality and the structure of the state. Naturally the state is included by the state space. The observation space is derived from the state space, using $\boldsymbol{H}$ as linear operator. Variables in the in the observation space are marked by a dot. A measurement $\dot{\boldsymbol{y}}_t$ is always part of the observation space. It depends on the true state $\breve{\boldsymbol{x}}_t$ and can be calculated as $\dot{\boldsymbol{y}}_{\boldsymbol{t}} = \boldsymbol{H}_t\breve{\boldsymbol{x}}_t + \dot{\boldsymbol{e}}_t$ with the accuracy $\dot{\boldsymbol{R}}$, determined by the error $\dot{\boldsymbol{e}}_t$. The state estimate $\boldsymbol{x}_t$ with the accuracy $\boldsymbol{P}_t$ and the predicted state estimate $\hat{\boldsymbol{x}}_{t+1}$ with its accuracy $\hat{\boldsymbol{P}}_{t+1}$ are part of the state space.

To get a multiplicative point of view, using the state and observation space, the following section examines how the Kalman filter handles both spaces. The solution of the Kalman filter is part of the state space but calculated with values of both spaces. Thats why the calculation is transferred to the state space. Here $\boldsymbol{H}$ instead of $\boldsymbol{H}_t$ is used for simplification because it is constant within one step.

$$\boldsymbol{K} = \boldsymbol{\Sigma}_1\boldsymbol{H}^T\left[\boldsymbol{H}\boldsymbol{\Sigma}_1\boldsymbol{H}^T + \dot{\boldsymbol{\Sigma}}_2\right]^{-1}$$

$$\begin{aligned}
\boldsymbol{\mu} &= \boldsymbol{\mu}_1 + \boldsymbol{K}\left[\dot{\boldsymbol{\mu}}_2 - \boldsymbol{H}\boldsymbol{\mu}_1\right] \\
&= \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_1\boldsymbol{H}^T\left[\boldsymbol{H}\boldsymbol{\Sigma}_1\boldsymbol{H}^T + \dot{\boldsymbol{\Sigma}}_2\right]^{-1}\left[\dot{\boldsymbol{\mu}}_2 - \boldsymbol{H}\boldsymbol{\mu}_1\right] \\
&= \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_1\boldsymbol{H}^T\left[\boldsymbol{H}\boldsymbol{\Sigma}_1\boldsymbol{H}^T + \boldsymbol{H}\boldsymbol{\Sigma}_2\boldsymbol{H}^T\right]^{-1}\left[\boldsymbol{H}\boldsymbol{\mu}_2 - \boldsymbol{H}\boldsymbol{\mu}_1\right] \\
&= \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_1\boldsymbol{H}^T\left[\boldsymbol{H}\left(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2\right)\boldsymbol{H}^T\right]^{-1}\boldsymbol{H}\left[\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1\right] \\
&= \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_1\boldsymbol{H}^T\left[\boldsymbol{H}^T\right]^{-1}\left[\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2\right]^{-1}\boldsymbol{H}^{-1}\boldsymbol{H}\left[\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1\right] \\
\boldsymbol{\mu} &= \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_1\left[\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2\right]^{-1}\left[\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1\right]
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{\Sigma} &= \boldsymbol{\Sigma}_1 - \boldsymbol{K}\boldsymbol{H}\boldsymbol{\Sigma}_1 \\
&= \boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_1\boldsymbol{H}^T\left[\boldsymbol{H}\boldsymbol{\Sigma}_1\boldsymbol{H}^T + \dot{\boldsymbol{\Sigma}}_2\right]^{-1}\boldsymbol{H}\boldsymbol{\Sigma}_1 \\
&= \boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_1\boldsymbol{H}^T\left[\boldsymbol{H}\boldsymbol{\Sigma}_1\boldsymbol{H}^T + \boldsymbol{H}\boldsymbol{\Sigma}_2\boldsymbol{H}^T\right]^{-1}\boldsymbol{H}\boldsymbol{\Sigma}_1 \\
&= \boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_1\boldsymbol{H}^T\left[\boldsymbol{H}\left(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2\right)\boldsymbol{H}^T\right]^{-1}\boldsymbol{H}\boldsymbol{\Sigma}_1 \\
&= \boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_1\boldsymbol{H}^T\left[\boldsymbol{H}^T\right]^{-1}\left[\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2\right]^{-1}\boldsymbol{H}^{-1}\boldsymbol{H}\boldsymbol{\Sigma}_1 \\
\boldsymbol{\Sigma} &= \boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_1\left[\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2\right]^{-1}\boldsymbol{\Sigma}_1
\end{aligned}$$

Here you can see that the Kalman filter solution can be transformed from the observation space to the already known equations in the state space. For these equations we had shown that the multiplicative solution is equivalent. Starting with the multiplicative solution in the state space, it is transferred to the observation space. For simplification an other representation of $\boldsymbol{\Sigma}$ is used.

Theorem 6.

$$\boldsymbol{\Sigma} = \left[\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1}\right]^{-1} = \boldsymbol{\Sigma}_1\left[\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2\right]^{-1}\boldsymbol{\Sigma}_2$$

Proof.

$$\begin{aligned}
\left[\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1}\right]^{-1} &= \boldsymbol{\Sigma}_1 \left[\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2\right]^{-1} \boldsymbol{\Sigma}_2 \\
\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1} &= \left(\boldsymbol{\Sigma}_1 \left[\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2\right]^{-1} \boldsymbol{\Sigma}_2\right)^{-1} \\
&= \boldsymbol{\Sigma}_2^{-1} \left[\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2\right] \boldsymbol{\Sigma}_1^{-1} \\
&= \boldsymbol{\Sigma}_2^{-1} \boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1} \boldsymbol{\Sigma}_2 \boldsymbol{\Sigma}_1^{-1} \\
\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1} &= \boldsymbol{\Sigma}_2^{-1} + \boldsymbol{\Sigma}_1^{-1}
\end{aligned}$$

□

With this shorter representation of $\boldsymbol{\Sigma}$ the multiplicative solution can be transferred to the observation space easier.

$$\begin{aligned}
\boldsymbol{\Sigma} &= \left[\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1}\right]^{-1} \\
&= \left[\boldsymbol{\Sigma}_1^{-1} + \left(\boldsymbol{H}^{-1} \boldsymbol{H} \boldsymbol{\Sigma}_2 \boldsymbol{H}^T \left[\boldsymbol{H}^T\right]^{-1}\right)^{-1}\right]^{-1} \\
\boldsymbol{\Sigma} &= \left[\boldsymbol{\Sigma}_1^{-1} + \left(\boldsymbol{H}^{-1} \dot{\boldsymbol{\Sigma}}_2 \left[\boldsymbol{H}^{-1}\right]^T\right)^{-1}\right]^{-1}
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{\mu} &= \boldsymbol{\Sigma}_2 \left[\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2\right]^{-1} \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_1 \left[\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2\right]^{-1} \boldsymbol{\mu}_2 \\
&= \boldsymbol{H}^{-1} \boldsymbol{H} \boldsymbol{\Sigma}_2 \boldsymbol{H}^T \left(\boldsymbol{H}^T\right)^{-1} \left[\boldsymbol{\Sigma}_1 + \boldsymbol{H}^{-1} \boldsymbol{H} \boldsymbol{\Sigma}_2 \boldsymbol{H}^T \left(\boldsymbol{H}^T\right)^{-1}\right]^{-1} \boldsymbol{\mu}_1 \\
&\quad + \boldsymbol{\Sigma}_1 \left[\boldsymbol{\Sigma}_1 + \boldsymbol{H}^{-1} \boldsymbol{H} \boldsymbol{\Sigma}_2 \boldsymbol{H}^T \left(\boldsymbol{H}^T\right)^{-1}\right]^{-1} \boldsymbol{H}^{-1} \boldsymbol{H} \boldsymbol{\mu}_2 \\
\boldsymbol{\mu} &= \boldsymbol{H}^{-1} \dot{\boldsymbol{\Sigma}}_2 \left(\boldsymbol{H}^{-1}\right)^T \left[\boldsymbol{\Sigma}_1 + \boldsymbol{H}^{-1} \dot{\boldsymbol{\Sigma}}_2 \left(\boldsymbol{H}^{-1}\right)^T\right]^{-1} \boldsymbol{\mu}_1 \\
&\quad + \boldsymbol{\Sigma}_1 \left[\boldsymbol{\Sigma}_1 + \boldsymbol{H}^{-1} \dot{\boldsymbol{\Sigma}}_2 \left(\boldsymbol{H}^{-1}\right)^T\right]^{-1} \boldsymbol{H}^{-1} \dot{\boldsymbol{\mu}}_2
\end{aligned}$$

This shows that the multiplicative point of view can be used without direct observability. But the measurement has to be transferred from the observation to the state space first.

5 Multiplicative Kalman Filter

Initialization, prediction and update step were examined separately. Now they are integrated into an overview of the Kalman filter.

Figure 1 shows that a measurement needs not to be available at each point in time. In contrast to that, multiple measurements can be available using a set of sensors. But the Kalman filter does not consider both cases directly. The update step can handle one measurement only. To use a second measurement at the same point in time, a prediction step need to be performed first. But this prediction calculates a forecast of the same point in time. So it can be left out. If no measurement is available, the Kalman filter can not calculate an estimate. The update step follows instead.

Also the names update and prediction step are imperfect. A prediction $\hat{\boldsymbol{x}}_{t+1}$ is often calculated at $t+1$. Because the calculation is triggered by the measurement $\boldsymbol{y_{t+1}}$. Thats why $\hat{\boldsymbol{x}}_{t+1}$ is a forecast of the present point in time. So it is better to say that the knowledge of $\boldsymbol{x}_t$ is kept up to date instead of calculating a prediction. Changes applied to $\breve{\boldsymbol{x}}_t$ are caught in $\boldsymbol{F}_t$ and the

control signal vector $\boldsymbol{u}_t$ in real time. So the prediction step is a synchronization step. It is often said that the update step corrects the predicted estimate. But it is not a real correction because $\boldsymbol{x}_{t+1} = \breve{\boldsymbol{x}}_{t+1}$ can not be reached. A bad measurement can lead the update into the wrong direction all the worse. The update step could be called fusion step. The information is combined and no values are corrected.

Figure 5 shows the Kalman filter, using the renamed steps. The presented algorithm allows to skip single steps. The predicted estimate $\hat{\boldsymbol{x}}$ and the corrected estimate $\boldsymbol{x}$ con not be distinguished anymore. So they are conjoint to $\boldsymbol{x}$.

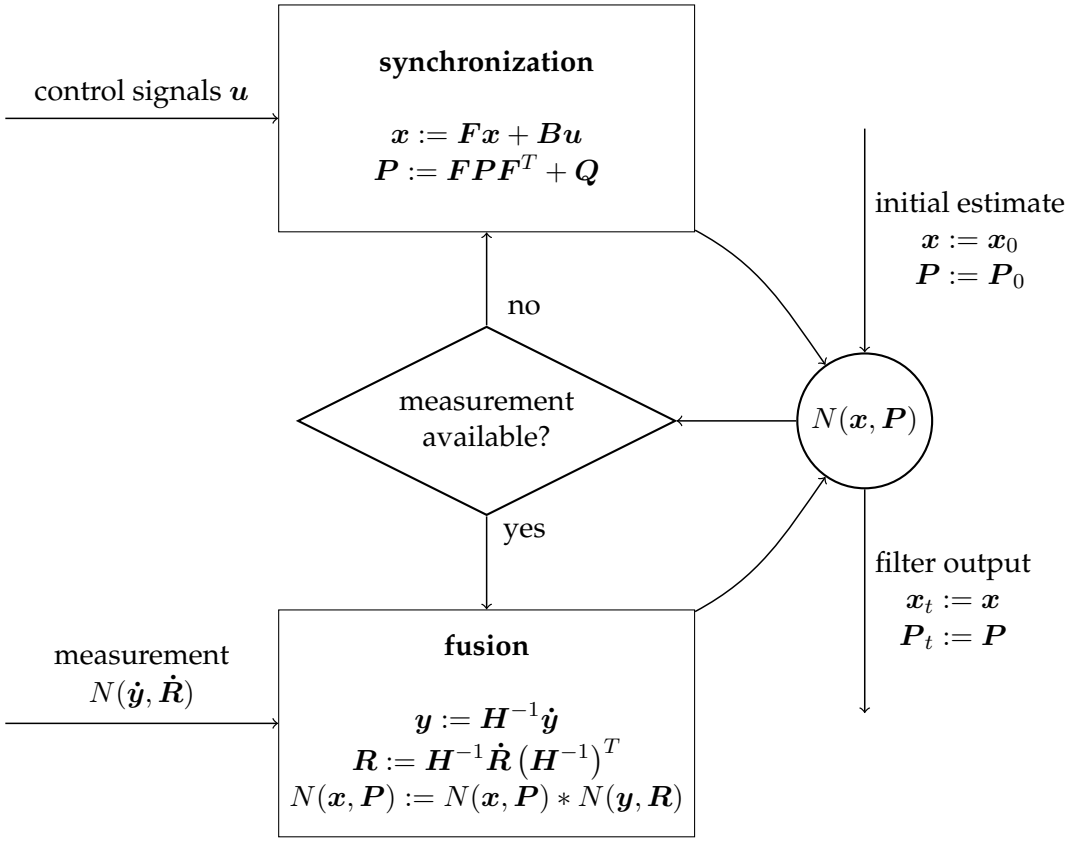

Figure 4: Overview of the steps of the multiplicative Kalman filter. Actual estimate $\boldsymbol{x}$ and measurement $\dot{\boldsymbol{y}}$ are seen as normal distributions. In the fusion step, the product of the underlying normal distributions is calculated. In contrast to the classical Kalman filter there is no prediction update structure. Depending on the availability of a measurement, a fusion or synchronization step is executed.

6 Summary

Starting with the classical description of the Klaman filter, we derived a multiplicative point of view. It was shown that in a simple case with direct observability of the state, the Kalman filter solution is equal to the product of the normal distributed measurements. Therefore it was necessary to proof that the product of normal distributions is up to an constant factor a normal distribution again. We showed that the multiplicative viewpoint is usable when direct observability is not given and that it is still equivalent to the Kalman filter solution in the multidimensional space with no limitation of dimensions. Finally, using the multiplicative solution, a modified overview of the Kalman filter was given.

References

[1] Vikranth R. Bejjanki, Meghan Clayards, David C. Knill, and Richard N. Aslin. Cue integration in categorical tasks: Insights from audio-visual speech perception. *PLoS ONE*, 6:e19812+, 2011.

[2] Marc O. Ernst and Martin S. Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415:429–433, 2002.

[3] Joachim Hartung and Bärbel Elpelt. *Multivariate Statistik / Lehr- und Handbuch der angewandten Statistik*. Oldenbourg, München ; Wien, 7., unveränd. aufl. edition, 2007.

[4] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME-Journal of Basic Engineering*, 82:35 – 45, 1960.

[5] Peter S. Maybeck. *Stochastic models, estimation, and control. Volume 1*. Academic Press, Arlington, Va., 1979.

[6] H. W. Sorenson. Least-squares estimation: from gauss to kalman. *IEEE spectrum JULY*, 7:63 – 68, 7 1970.

[7] Greg Welch and Gary Bishop. An introduction to the kalman filter, 1995.

[8] Jan Wendel. *Integrierte Navigationssysteme: Sensordatenfusion, GPS und Inertiale Navigation*. Oldenbourg Wissenschaftsverlag GmbH, 2011.